ESTIMATION AND TESTS OF FIT FOR TWO COMPONENT WEIBULL MIXTURES

by

W. Chandanie Wijayalatha Perera

B.Sc.(Mathematics Special), University of Colombo, Sri Lanka, 1985

Postgraduate Diploma(Applied Statistics), University of Colombo, Sri Lanka, 1987

THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in the Department

of

Mathematics and Statistics

Canada

APPROVAL

Name: W. C. W. Perera

Degree: Master of Science

Title of thesis: Estimation and tests of fit for two component Weibull mixtures

Examining Committee:

Chairman: Dr. A.H. Lachlan

_____
Dr. R. A. Lockhart
Senior Supervisor

_____
Dr. C. B. Dean

_____
Dr. M. A. Stephens

_____
Dr. D. M. Eaves
External Examiner
Department of Mathematics and Statistics
Simon Fraser University

Date Approved : August 6, 1992

ii

Title of Thesis/Project/Extended Essay

Estimation and tests of fit for Two component

Weibull mixtures

Author:

(signature)

W. C. W. Perera

(name)

August 6, 1992

(date)

Abstract

Mixture models are encountered in a wide range of real world phenomena. The complexity of the likelihood function of mixture models gives rise to difficulties in parameter estimation.

An iterative scheme and the relevant software are developed to estimate the parameters of a two parameter Weibull mixture population.

Likelihood equations for Monte Carlo samples from Weibull mixture populations with poorly separated components were found to have more than one solution point; techniques are proposed to find these solution points. Each provided a very good fit to the data.

The Cramér-von Mises statistic is employed to test the goodness of fit of the fitted mixture model. The complexity of the mixture distribution function makes it impossible to find a closed form solution for the covariance kernel. Suggestions are made and software is developed to estimate the covariance of the asymptotic process and to find estimates for the eigenvalues of the covariance kernel. The p-values for the test statistic are computed using these estimates. The usual asymptotic theory related to the distribution of the Cramér-von Mises statistic is verified by a Monte Carlo study.

Acknowledgements

Dedication

To my parents.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Mixture density functions are an important area of statistics with a wide range of appli-
cations. For example, in Biology, it is often required to measure a certain characteristic
of a population consisting of individuals with different ages. The characteristic of inter-
est may show a marked difference between individuals of different ages which is difficult
to ascertain and thus gives rise to a mixture population problem.

Another important area is often met in Engineering where the observations are times
to failure of a sample of items. Often failures can occur for more than one reason and
each cause of failure yields a subpopulation with an underlying distribution that can
be estimated by an Exponential distribution or a Weibull Distribution. Therefore, the
population as a whole can be treated as a mixture population.

We begin our study with a discussion of the mixture problem from a mathematical
point of view. In chapter 2, we discuss the problem of estimating the parameters of
a mixture population, with special emphasis on mixtures of Weibull populations. We
end chapter 2 with a brief discussion of the problems encountered in the estimation
procedure due to the behaviour of the likelihood function.

Chapter 3 is devoted to testing goodness of fit based on the Empirical Distribution

Function(EDF). Section 3.1 introduces the empirical distribution function. Section 3.2 provides an introduction to the EDF statistics. Section 3.3 briefly outlines how EDF statistics are used in testing goodness of fit. Distributional properties of quadratic EDF statistics are discussed in section 3.4. For ease of exposition, we restrict our attention to the Cramér-von Mises statistic. The complexity of the mixture distribution function makes it impossible to find the exact eigenvalues of the covariance kernel of the Gaussian process associated with the Cramér-von Mises statistic. Section 3.5 introduces what we call the brute force approach to obtain estimates for the eigenvalues of the covariance kernel. Computational formulas for EDF statistics are presented in section 3.6. Computation of the p-values for the Cramér-von Mises statistic is discussed in section 3.7. In chapter 4, we present some applications of the procedure discussed in section 3.5.

Chapter 5 is concerned with the software developed for the computation of p-values of the Cramér-von Mises statistic in Weibull mixture populations.

In chapter 6 we describe a Monte Carlo study that was performed to test the accuracy of the asymptotic theory related to the distribution of the Cramér-von Mises statistic.

Finally, a summary of our study and a discussion of the open problems related to this study are offered in chapter 7.

Chapter 2

Discussion of the mixture problem and estimation of the unknown parameters

We often come across the problem of fitting a model to a population which is not homogeneous, but is made up of two or more sub-populations. The mixing proportions of the sub-populations may or may not be known but are treated as fixed. A finite mixture is one which consists of only a finite number of components and is the topic of interest here.

Section 2.1 provides a description of the mixture problem from a mathematical point of view. Section 2.2 describes the notion of identifiability. In section 2.3 we discuss the problem of estimating the unknown parameters of mixture populations. Three iterative schemes proposed by Kaylan and Carl[10] to estimate the unknown parameters of a mixture of two parameter Weibull distributions are also described in section 2.3. The problems that we encounter with the available iterative schemes motivated us to look for other possible iterative procedures. Section 2.4 summarises these problems. Section 2.5 offers an alternative iterative scheme that we developed in this study for the above

3

purpose. We end this chapter with a brief description of the behaviour of likelihood functions for samples from Weibull mixture populations.

2.1 Mixture problem from a mathematical point of view

In a more mathematical sense, we shall say that a random variable or vector X has a finite mixture distribution, if its distribution can be represented by a probability density function(or probability mass function in the case of discrete X) of the form,

$$f(x) = p_1 f_1(x, \underline{\theta}_1) + p_2 f_2(x, \underline{\theta}_2) + ... + p_k f_k(x, \underline{\theta}_k) \text{ where } p_i \geq 0 \quad i = 1, 2, ..., k,$$

$\sum_{i=1}^{k} p_i = 1$, and $f_i(.) \geq 0$, $\int f_i(x) dx = 1$, $i = 1, 2, ..., k$. The density $f(.)$ is called the finite mixture density function.

The parameters $p_1, p_2, ..., p_k$ are called the mixing weights and

$f_1(.), f_2(.), ..., f_k(.)$ the component densities of the mixture.

Given a random sample from the mixture population, our interest is to decompose the mixture by estimating the unknown parameters, $\{p_1, p_2, ..., p_k, \underline{\theta}_1, \underline{\theta}_2, ..., \underline{\theta}_k\}$ where $\underline{\theta}_1, \underline{\theta}_2, ..., \underline{\theta}_k$ are the vectors of parameters of each component distribution. Here k is the number of components that constitute the mixture.

Mixture problems, with k and the families of the component distributions known, can basically be divided into two classes.

1. The component distributions are completely known and only the proportion parameters $p_1, p_2, ..., p_k$ are to be estimated.

2. The component distributions are known apart from the parameters $\underline{\theta}_1, \underline{\theta}_2, ..., \underline{\theta}_k$. In this case vectors of parameters $\underline{\theta}_1, \underline{\theta}_2, ..., \underline{\theta}_k$ and the proportion parameters $p_1, p_2, ..., p_k$ are to be estimated. It is noted that in this case, to be useful in

practice, k has to be a small number such as 2 and the $f_i's$ have to be all members of some small parametric family such as the exponential, or normal, or two parameter Weibull families.

In this study we are concerned with the second class and from here onwards we will assume that $k = 2$, and each component density is a member of the two parameter Weibull family.

So, in our problem of interest the mixture density can be written as,

$$f(x, c_1, \theta_1, c_2, \theta_2, p) = p \ f_1(x, c_1, \theta_1) + (1 - p) \ f_2(x, c_2, \theta_2) \qquad (2.1)$$

The mixing weights are $p_1 = p$ and $p_2 = (1 - p)$ and each f_i is a two parameter Weibull density;

$$f_i(x, c_i, \theta_i) = \frac{c_i}{\theta_i} \left(\frac{x - x_0}{\theta_i} \right)^{c_i - 1} \exp \left\{ - \left(\frac{x - x_0}{\theta_i} \right)^{c_i} \right\}$$

The parameters c_1, c_2 are called shape parameters and θ_1, θ_2 are called scale parameters. The location parameters are assumed to be known and are further assumed to be the same for both components.

2.2   Identifiability of the mixture

Before moving on to the problem of estimating the parameters, it is important to discuss the notion of "identifiability". In general, the parametric family of probability density functions $f(x, \underline{\alpha})$ where $\underline{\alpha}$ is the vector of unkown parameters, is said to be identifiable, if distinct values of α determine distinct members of the family.

In a more mathematical sense we shall say that a class of finite mixture densities is identifiable, if for any two members, $f(x, \underline{\alpha}) = \sum_{i=1}^{k} p_i f_i(x, \underline{\theta}_i)$, and $f(x, \underline{\alpha}') =$

$\sum_{i=1}^{k'} p'_i f_i(x, \underline{\theta}'_i)$, $f(x, \underline{\alpha}) \equiv f(x, \underline{\alpha}')$ if and only if $k = k'$, and we can permute the component labels so that $p_i = p_{i'}$ and $f_i(x, \underline{\theta}_i) \equiv f(x, \underline{\theta}'_i)$ (McLachlan and Basford [13].)

It is important to consider the identifiability in practice because, without it, estimation procedures are not likely to be well defined. Furthermore, identifiability is a necessary requirement for the usual asymptotic theory to hold for the estimation. (Titterington, Smith and Markov(1985) [18]).

Now let us turn to our problem of interest. We note that in the mixture density defined by equation 2.1, if we switch $\underline{\theta}_1, \underline{\theta}_2$ and p and $(1 - p)$ accordingly we get the same density. Therefore, we can only estimate the parameters if we decide to say that the first population will be the one with the smaller shape parameter or if they have the same shape parameter, the first population will be the one with the smaller scale parameter. This convention identifies the parameters.

Technically, also, part of the difficulty arises from the fact that when p is 0 or 1, then the corresponding $\underline{\theta}_i$ is meaningless.

2.3   Estimation of the unknown parameters

In the literature, one can find several methods for estimating unknown parameters such as the method of moments, maximum likelihood estimation, Bayesian methods, graphical methods, method of weighted least squares etc. We will confine our attention to the method of maximum likelihood estimation.

At this point we would like to mention that our interest in mixture problems was initiated by a problem that was brought to the consulting service in the Department of Mathematics and Statistics. That problem was a case with grouped data. We will present the likelihood estimation procedure for ungrouped data first and touch briefly on grouped data.

Maximum Likelihood Estimation:

We now describe the maximum likelihood estimation procedure briefly. The likelihood function based on a sample of n observations from the mixture can be written as,

$$L(\underline{\alpha}) = \prod_{j=1}^{n} \sum_{i=1}^{k} p_i f(x_j, \underline{\theta}_i)$$

where $\underline{\alpha}$ is the vector of unknown parameters.

Basically, maximisation of $L(\underline{\alpha})$ with respect to $\underline{\alpha}$ for given data $\underline{x}$, yields the maximum likelihood estimate of $\underline{\alpha}$. Equivalently, the quantity maximised is the log-likelihood function,

$$l(\underline{\alpha}) = \log L(\underline{\alpha}),$$

which is also called the objective function.

In estimation problems related to mixtures, one has to take into account a set of constraints in addition to the objective function. For example, the mixing proportions p_i's have to satisfy the condition that, $0 < p_i < 1$. There may exist other constraints related to the parameters of sub-populations depending on the situation of interest. The constraints are generally of a linear type, and hence the maximum likelihood estimation problem can be formulated as a mathematical programming problem with a non linear objective function with linear constraints.

Now we give a more detailed discussion of the maximum likelihood procedure for the estimation of five parameters of a two component mixture of two parameter Weibull distributions. Equation 2.1 defines the mixture density function. This is the mixture that is used in the Monte Carlo study described in chapter 6.

Case 1: Ungrouped data

The likelihood function is,

$$L(\underline{\alpha}) = \prod_{j=1}^{n} \sum_{i=1}^{2} p_i f(x_j, \underline{\theta}_i),$$

where $p_1 = p$, $p_2 = (1 - p)$, $\underline{\alpha}^T = (c_1, c_2, \theta_1, \theta_2, p)$ is the vector of unknown parameters and n is the number of observations in the sample.

The constraints on the parameters are, $0 < p < 1$, $c_i > 0$ and $\theta_i > 0$ for $i = 1, 2$. It is also noted that in order to make the problem identifiable we have to impose other restrictions on the parameters such as $c_1 \leq c_2$ and, if $c_1 = c_2$ then $\theta_1 < \theta_2$.

Therefore, maximum likelihood estimation can be formulated as,

$$\max_{\underline{\alpha} \epsilon S} L(\underline{\alpha}) \quad \text{where,} \quad \underline{\alpha}^T = (c_1, c_2, \theta_1, \theta_2, p) \text{ and}$$

$$S = \{\underline{\alpha} \mid 0 < p < 1, \ c_i > 0, \ \theta_i > 0 \ for \ i = 1, 2, \ c_1 \leq c_2 \ and, \ if \ c_1 = c_2 \ then \ \theta_1 < \theta_2\}$$

An estimate of the parameter vector $\underline{\alpha}$ is said to be a "feasible estimate", if it belongs to S. The conditions which determine whether or not a solution is feasible are called "feasibility conditions".

Following the notation used by Kaylan and Carl [10] we employ f_{ij} and f_j instead of $f_i(x_j, c_i, \theta_i)$, $f(x_j, \alpha)$ respectively.

The log-likelihood function based on a random sample of n observations is,

$$l(\underline{x}, \underline{\alpha}) = \sum_{j=1}^{n} \log f(x_j, \underline{\alpha})$$

where $f(x_j, \underline{\alpha}) = pf_{1j} + (1 - p)f_{2j}$ and $f_{ij} = f_i(x_j, c_i, \theta_i)$.

One can easily obtain the partial derivatives $\frac{\partial l}{\partial c_i}$, $\frac{\partial l}{\partial \theta_i}$, $\frac{\partial l}{\partial p_i}$ given below.

$$\frac{\partial l}{\partial c_i} = \sum_{j=1}^{n} \frac{1}{f_j} p_i \frac{\partial f_{ij}}{\partial c_i} \quad \text{where} \quad p_1 = p \quad \text{and} \quad p_2 = 1 - p, \tag{2.2}$$

$$\frac{\partial f_{ij}}{\partial c_i} = f_{ij} \left[\frac{1}{c_i} + \ln\left(\frac{x_j - x_0}{\theta_i}\right) - \left(\frac{x_j - x_0}{\theta_i}\right)^{c_i} \ln\left(\frac{x_j - x_0}{\theta_i}\right) \right]$$

$$\frac{\partial l}{\partial \theta_i} = \sum_{j=1}^{n} \frac{1}{f_j} p_i \frac{\partial f_{ij}}{\partial \theta_i}, \tag{2.3}$$

$$\frac{\partial f_{ij}}{\partial \theta_i} = f_{ij} \left[\frac{c_i}{\theta_i} \left(\frac{x_j - x_0}{\theta_i}\right)^{c_i} - \frac{c_i}{\theta_i} \right]$$

$$\frac{\partial l}{\partial p} = \sum_{j=1}^{n} \frac{1}{f_j} [f_{1j} - f_{2j}] \tag{2.4}$$

8

The maximum likelihood estimates are given by,

$$\frac{\partial l}{\partial c_i} = 0, \quad \frac{\partial l}{\partial \theta_i} = 0, \quad i = 1, 2 \tag{2.5}$$

$$\frac{\partial l}{\partial p} = 0 \tag{2.6}$$

Equations 2.5- 2.6 are called likelihood equations. Numerical methods are employed to solve the likelihood equations simultaneously.

2.3.1 Solution of the likelihood equations

The surface determined by the log-likelihood function in the space of unknown parameters is called the likelihood surface. The likelihood surface for samples from Weibull mixture populations is flat over a wide range of the parameter space. Therefore, popular methods like the Newton Raphson method fail to converge and the need arises to develop special techniques to solve the likelihood equations.

It is also noted that at any point on the likelihood surface the gradient vector to the surface, usually denoted by $\bigtriangledown l$, is defined as $\bigtriangledown l = (\frac{\partial l}{\partial c_1}, \frac{\partial l}{\partial c_2}, \frac{\partial l}{\partial \theta_1}, \frac{\partial l}{\partial \theta_2}, \frac{\partial l}{\partial p})$.

Kaylan and Carl[10] have suggested three iterative schemes that can be used to solve the above equations numerically, which are briefly described below.

If we employ the notation $c_i^v, \theta_i^v, p_i^v, f_{ij}^v$ and f_j^v to indicate the values of c_i, θ_i, p, f_{ij} and f_j at step v of the iteration process then the set of equations at the step $v + 1$ for an iterative scheme to find the maximum likelihood estimates of the parameters are,

$$c_i^{v+1} = [\sum_{j=1}^{n} \frac{f_{ij}^v}{f_j^v}]/[\sum_{j=1}^{n} \frac{f_{ij}^v}{f_j^v} \ln(\frac{x_j - x_0}{\theta_i^v})([\frac{x_j - x_0}{\theta_i^v}]^{c_i^v} - 1)] \tag{2.7}$$

$$\theta_i^{v+1} = [\sum_{j=1}^{n} \frac{f_{ij}^v}{f_j^v}(x_j - x_0)^{c_i^v})/(\sum_{j=1}^{n} \frac{f_{ij}^v}{f_j^v})]^{\frac{1}{c_i^v}} \tag{2.8}$$

$$p^{v+1} = \frac{p^v}{n} \sum_{j=1}^{n} \frac{f_{ij}^v}{f_j^v} \tag{2.9}$$

A "good guess" $\alpha^0 = (c_1^0, \theta_1^0, c_2^0, \theta_2^0, p^0)$ is to be chosen as an initial estimate of the

9

parameters. In this study, we do not deal with the problem of finding an initial guess. However, some met¹ . ·hat are available in the literature are cited at the end of section 2.5.

An alternative iterative scheme also proposed by Kaylan and Carl[10] replaces equation 2.9 above by,

$$p^{v+1} = p^v + \sum_{j=1}^{n} \frac{f_{1j}^v - f_{2j}^v}{f_j^v} / \sum_{j=1}^{n} (\frac{f_{1j}^v - f_{2j}^v}{f_j^v})^2 \qquad (2.10)$$

The second iterative scheme suggested is to iterate using equations 2.7, 2.8 and 2.10.

The third suggested scheme is to iterate with equations 2.7, 2.8 and 2.9 to begin with, and then shifting to iterate with equations 2.7, 2.8 and 2.10, when the absolute values of the components of the gradient vector are all less than 10(5) for a sample of size 500(100). They refer to this iterative scheme as the two phase method. According to Kaylan and Carl, the two phase scheme performs better.

Feasibility condition:

A set of parameter values is called a feasible solution if both shape and scale parameters are positive and the mixing proportion parameter p is between 0 and 1(exclusive). At each step of the iteration this feasibility condition is checked. According to Kaylan and Carl [10] if the initial point is feasible, the feasibility conditions hold true at every iteration for the algorithm built on equations 2.7, 2.8 and 2.9. They also suggest that it is required to check the mixing proportion constraint for the algorithms that use equation 2.10. Kaylan and Carl suggest that if the new solution point does not satisfy the mixing proportion constraints, then one should move back to the boundary hyperplane $p_1 + p_2 + ... + p_{k-1} = 1$.

In particular, for a mixture of two components their suggestion is as follows. If $p^{(v+1)}$ does not satisfy the proportion constraints $0 \leq p \leq 1$, then replace p^{v+1} by 1.

Improvement Condition:

At each step of the iteration the condition of improvement $l^{(v+1)} \geq l^v$, is also checked. If at a certain step v this condition is not satisfied, then the algorithm step size is bisected, until an improvement is observed in l. That is, if $l^{v+1} < l^v$ then try halfway from $\underline{\alpha}^v$ to $\underline{\alpha}^{v+1}$, and if l is still not improved then try one quarter of that distance from $\underline{\alpha}^v$ and so on.

The algorithm is terminated if either of the following two conditions is satisfied.

1. The absolute values of the gradient elements are all less than a small value ϵ_1(e.g. $\epsilon_1 = 0.001$) and, the absolute relative difference in l, that is $|(l^{v+1} - l^v)/l^v)|$ is less than a small value ϵ_2 (e.g. $\epsilon_2 = 0.0001$) or,

2. A maximum number of iterations(e.g.250) is reached.

2.3.2 Estimation of the variances of the parameter estimates

The estimated variances of the parameter estimates are the diagonal elements of the matrix $-H^{-1}$ evaluated at the solution point α, where H is the Hessian matrix. Thus, we need the following second derivatives of l, which are the elements of H.

$$\frac{\partial^2 l}{\partial c_i^2} = \sum_{j=1}^{n}[\frac{p_j}{f_j}\frac{\partial^2 f_{ij}}{\partial c_i^2} - \frac{p_i^2}{f_j^2}(\frac{\partial f_{ij}}{\partial c_i})^2], \tag{2.11}$$

$$\frac{\partial^2 f_{ij}}{\partial c_i^2} = \frac{\partial f_{ij}}{\partial c_i}[\frac{1}{c_i} + \ln(\frac{x_j - x_0}{\theta_i}) - (\frac{x_j - x_0}{\theta_i})^{c_i}\ln(\frac{x_j - x_0}{\theta_i})]$$
$$+ f_{ij}[\frac{-1}{c_i^2} - (\frac{x_j - x_0}{\theta_i})^{c_i}\ln^2(\frac{x_j - x_0}{\theta_i})]$$

$$\frac{\partial^2 l}{\partial \theta_i^2} = \sum_{j=1}^{n}[\frac{p_j}{f_j}\frac{\partial^2 f_{ij}}{\partial \theta_i^2} - \frac{p_i^2}{f_j^2}(\frac{\partial f_{ij}}{\partial \theta_i})^2], \tag{2.12}$$

$$\frac{\partial^2 f_{ij}}{\partial \theta_i^2} = \frac{\partial f_{ij}}{\partial \theta_i}[\frac{c_i}{\theta_i}(\frac{x_j - x_0}{\theta_i})^{c_i} - \frac{c_i}{\theta_i}] + f_{ij}[\frac{c_i}{\theta_i^2} - \frac{c_i(c_i+1)}{\theta_i^2}(\frac{x_j - x_0}{\theta_i})^{c_i}]$$

$$\frac{\partial^2 l}{\partial p^2} = -\sum_{j=1}^{n}\frac{(f_{1j} - f_{2j})^2}{f_j^2} \tag{2.13}$$

11

$$\frac{\partial^2 l}{\partial c_1 \partial c_2} = -\sum_{j=1}^{n} \frac{p_2}{f_j^2} \frac{\partial f_{2j}}{\partial c_2} p_1 \frac{\partial f_{1j}}{\partial c_1} \tag{2.14}$$

$$\frac{\partial^2 l}{\partial c_i \partial \theta_i} = \sum_{j=1}^{n} [\frac{p_i}{f_j} \frac{\partial^2 f_{ij}}{\partial c_i \partial \theta_i} - \frac{p_i^2}{f_j^2} \frac{\partial f_{ij}}{\partial c_i} \frac{\partial f_{ij}}{\partial \theta_i}], \tag{2.15}$$

$$\frac{\partial^2 f_{ij}}{\partial c_i \partial \theta_i} = \frac{\partial f_{ij}}{\partial c_i} [\frac{c_i}{\theta_i} (\frac{x_j - x_0}{\theta_i})^{c_i} - \frac{c_i}{\theta_i}]$$
$$+ f_{ij} [\frac{1}{\theta_i} (\frac{x_j - x_0}{\theta_i})^{c_i} - \frac{1}{\theta_i} + \frac{c_i}{\theta_i} (\frac{x_j - x_0}{\theta_i})^{c_i} \ln(\frac{x_j - x_0}{\theta_i})]$$

$$\frac{\partial^2 l}{\partial c_i \partial \theta_k} = \sum_{j=1}^{n} -\frac{p_1 p_2}{f_j^2} \frac{\partial f_{ij}}{\partial c_i} \frac{\partial f_{ij}}{\partial \theta_k} \quad \text{for } i \neq k \tag{2.16}$$

$$\frac{\partial^2 l}{\partial c_i \partial p} = \sum_{j=1}^{n} [-\frac{p_i}{f_j^2} (f_{1j} - f_{2j}) \frac{\partial f_{ij}}{\partial c_i} + \frac{(-1)^{i+1}}{f_j} \frac{\partial f_{ij}}{\partial c_i}] \tag{2.17}$$

$$\frac{\partial^2 l}{\partial \theta_1 \partial \theta_2} = -\sum_{j=1}^{n} \frac{p_2}{f_j^2} \frac{\partial f_{2j}}{\partial \theta_2} p_1 \frac{\partial f_{1j}}{\partial \theta_1} \tag{2.18}$$

$$\frac{\partial^2 l}{\partial \theta_i \partial p} = \sum_{j=1}^{n} [-\frac{p_i}{f_j^2} (f_{1j} - f_{2j}) \frac{\partial f_{ij}}{\partial \theta_i} + \frac{(-1)^{i+1}}{f_j} \frac{\partial f_{ij}}{\partial \theta_i}] \tag{2.19}$$

Case 2: Grouped Data:

Now we briefly mention the alternatives available in literature for grouped data. However, in this study, our emphasis is on ungrouped data.

Kaylan has suggested a method to deal with grouped data. It is simply to replace in the above equations suggested for case 1, x_j by $\overline{x_j}$, which is the midpoint of the j th class $j = 1, 2, ..., k'$ and $\sum_{j=1}^{n}$ by $\sum_{j=1}^{k'} n_j$ where n_j is the number of observations in the jth class and k' the total number of classes.

Another way to deal with grouped data is to find the parameters that maximise the log-likelihood function, $l(\underline{x}, \underline{\alpha}) = \sum_{j=1}^{n} n_j \log(P_j)$, where $P_j = \int_{a_{j-1}}^{a_j} f(x, \underline{\alpha}) dx$ and a_j represents the upper boundary of the jth class.

2.4 Motivation for an alternative scheme

We first tried to estimate the five parameters of the Weibull mixture distribution, using the two phase method described in Kaylan and Carl [10](section 2.3.1). For all of the data sets we tried, the bisection method used when the improvement condition was not satisfied failed to produce an increased value of l. Kaylan and Carl [10] report that the subpopulations must be well separated for the convergence to be assured.

For example, Kaylan and Carl [10] have considered specific examples of mixtures of exponentials with population parameters $(1, 1, 2, 4, 0.5)$ and $(1, 1, 0.4, 2.5, 0.5)$ in the order two shape, two scale and proportion. They report that the algorithm based on the equations 2.7, 2.8 and 2.9 failed to converge for half of the computer runs using 250 iterations which is the maximum number of iterations suggested. The convergence rate of the second and third iterative schemes are not reported. They have also reported their computational experience with each of the three iterative schemes. According to their report the second scheme requires less computational effort compared to the first scheme. They also report that there is substantial improvement if the two phase method is employed, provided that the shift to the second phase is not made too early. (This was described in section 2.3.1.) However, they have not reported their experience on the performance of the two phase method in the case of Weibull mixtures.

We were interested in looking for other possible methods which can be applicable in more general situations. We started by using the Newton Raphson method which takes $\underline{\alpha}^{v+1} = \underline{\alpha}^v - (H^v)^{-1} \bigtriangledown l^v$, where $\underline{\alpha}^v$ is the parameter vector at the step v of the iteration and H is the Hessian matrix(i.e. the matrix of second derivatives). However, it was found that due to the flat shape of the likelihood surface the iterative scheme that was built on the Newton Raphson method often fails to converge. The divergence of the scheme was so bad that in some cases the estimates even failed to satisfy the feasibility conditions.

This motivated us to transform the problem into one which is an unconstrained problem, so that, at each step of the iteration, the parameters satisfy the feasibility conditions.

2.5 An alternative scheme for the solution of likelihood equations

In this section, we describe another method that we developed for the purpose of obtaining the solution of the likelihood equations.

The problem is to maximise,

$$\log L(\alpha) = \sum_{j=1}^{n} \log(f(x_j, \alpha)),$$

where $f(x_j, \alpha) = p\, f_{1j} + (1 - p)\, f_{2j}$, and $f_{ij} = f_i(x_j, c_i, \theta_i)$ is the two parameter Weibull density, which was given in equation 2.1.

The constraints of the problem are:

i). The mixing proportion p has to satisfy $0 < p < 1$.

ii). The shape parameters c_1, c_2 and the scale parameters θ_1, θ_2 are positive.

iii). $c_1 \le c_2$ and if $c_1 = c_2$ then $\theta_1 < \theta_2$.

We begin by transforming the problem into one which is an unconstrained problem. This is done as follows.

Define a_i, b_i such that $c_i = a_i^2$ and $\theta_i = b_i^2$ for $i = 1, 2$. In other words a_i and b_i were taken as the positive square roots of c_i, and θ_i respectively. Let $m = \log\left(\frac{p}{1-p}\right)$. i.e. $p = \frac{e^m}{1+e^m}$ so that at each a_i, b_i, m, the shape and the scale parameters, $c_i, \theta_i > 0$ and the proportion parameter p satisfies, $0 < p < 1$, for all finite m. We do not deal

14

with the constraints due to the identifiability at this stage. Once we find the solution, the components of the parameter vector can easily be permuted so as to meet these constraints.

We note that if l is the log-likelihood function,

$$\frac{\partial l}{\partial a_i} = \frac{\partial l}{\partial c_i}\frac{\partial c_i}{\partial a_i} = \frac{\partial l}{\partial c_i}2a_i$$

$$\frac{\partial l}{\partial b_i} = \frac{\partial l}{\partial \theta_i}2b_i$$

$$\frac{\partial l}{\partial m} = \frac{\partial l}{\partial p}\frac{\partial p}{\partial m} = \frac{\partial l}{\partial p}\frac{e^m}{(1+e^m)^2} = \frac{\partial l}{\partial p}p(1-p)$$

So for c_i, $\theta_i > 0$ and $0 < p < 1$, the two sets of equations

$$\{\frac{\partial l}{\partial c_i} = 0, \frac{\partial l}{\partial \theta_i} = 0, \frac{\partial l}{\partial p} = 0\}$$

and

$$\{\frac{\partial l}{\partial a_i} = 0, \frac{\partial l}{\partial b_i} = 0, \frac{\partial l}{\partial m} = 0\}$$

will have zeros at the corresponding points.

So, we deal with the unconstrained problem that,

$$\frac{\partial l}{\partial a_i} = 0, \quad \frac{\partial l}{\partial b_i} = 0, \quad \frac{\partial l}{\partial m} = 0.$$

It is noted that the components of $\alpha^T = (c_1, c_2, \theta_1, \theta_2, p)$ are called untransformed variables and the components of $\beta^T = (a_1, a_2, b_1, b_2, m)$ are called transformed variables.

Method of solution:

The method we use is based on the Newton Raphson method, (on the space of transformed variables) which uses the iterative scheme, $\beta^{v+1} = \beta^v - (H^v)^{-1}\bigtriangledown l^v$, where β^v is the parameter vector at step v of the iteration, H^v is the matrix of second derivatives of l^v(i.e. the Hessian matrix) and $\bigtriangledown l^v = (\frac{\partial l}{\partial a_1}, \frac{\partial l}{\partial a_2}, \frac{\partial l}{\partial b_1}, \frac{\partial l}{\partial b_2}, \frac{\partial l}{\partial m})^T$ is the gradient vector at step v of the iteration.

In the case of mixture populations the behaviour of the likelihood function, that is, the flat shape of the likelihood surface over a wide range of the parameter space makes the estimation problem quite difficult. There is no guarantee that the Newton Raphson method works in such a case. However, if we carefully examine the step size taken and the direction of movement at each step, making changes whenever it is necessary, the method converges in our problem. Now we describe how this is done in our procedure.

At each step check whether the value of the log-likelihood l has increased or not.

- If the likelihood has increased, or in other words if $l^{v+1} > l^v$, we assume that the step size taken and the direction of movement are both appropriate and we proceed to the next step of the iteration.

- If $l^{v+1} \leq l^v$ then it is viewed as due to one of the following reasons, and the relevant change is made before proceeding to the next step of the iteration.

  1. We are moving in the correct direction, in the sense that our step, $\beta^{v+1} - \beta^v$ is in a direction along which the likelihood increases initially, but we have taken a step so large that we have already passed the region of increase.

     If the Hessian is negative definite then for all small positive ϵ, $l(\beta^v + \epsilon(\beta^{v+1} - \beta^v))$ is larger than $l(\beta^v)$. Thus, when the Hessian is found to be negative definite, we cut the step size in half repeatedly until we find an increase in the value of the likelihood. Once we observe an increase in l we go back to the orginal iterative scheme.

  2. We are not moving in the correct direction.

     If the matrix of second derivatives is not negative definite there is no guarantee that the method of bisection will eventually produce an increased value for l. In such a case, we replace the matrix H by $H^* = H - kI$ where I is the identity matrix with the same order as H and do one step of Newton Raphson. Here, k is a scalar chosen so that H^* is negative definite. In our

program we select k to be $m + 5$, where m is the largest eigenvalue of H. This results in a change of the direction of movement.

However, after replacing H by H^* in the Newton Raphson formula $\beta^{v+1} = \beta^v - (H^v)^{-1} \bigtriangledown l^v$ we may still not observe an increase in the likelihood function. This is due to taking too large a step. Again we cut the step size in half repeatedly (the method of bisection) until we observe an increase in the likelihood. It is worthwhile to note that, in this case the method of bisection is assured to give an improved estimate, but the bisection has to be done in the space of transformed variables. Once we observe an increase in l, we go back to the original Newton Raphson iterative scheme.

We will also describe here another possibility which we investigated to change the direction of movement. This is to replace the matrix H by, $H^{**} = PDP^T$ where P is the orthogonal matrix with columns as eigenvectors of H and D is the diagonal matrix whose diagonal entries are $-\lambda$, where the λ's are the absolute values of the eigenvalues of H. H^{**} is also negative definite and thus we are moving in the correct direction. Again, it is possible that the step size $\beta^{v+1} - \beta^v = (-H^{**})^{-1} \bigtriangledown l^v$ is too large and in such a case we cut the step size in half repeatedly until an increase in l is observed. This alternative was also found to work well. However, in the software we developed we used the previous method.

Feasibility conditions:

The transformation equations guarantee that at each step of iteration the estimates of all parameters are positive. However, if at a certain step of iteration m is very large then the estimated proportion p becomes 1 due to rounding off. Then according to our notation the parameter vector of the second component density $\underline{\theta_2}$ is meaningless. To avoid this, if at a certain step v of iteration p^v is 1, then we replace p^v by 0.99, which we assume to be a number close enough to retain the accuracy of the estimates and

proceed.

The restriction we imposed to clarify the problem of identifiability of the mixture, that is, $c_1 \leq c_2$ and if $c_1 = c_2$ then $\theta_1 < \theta_2$ is not directly touched in our iterative scheme. However, it seems to happen that this restriction is taken care of when we give initial guesses which have the first component having smaller shape, or if both shape parameters are the same giving the first component as the one which has the smaller scale parameter.

The algorithm that we developed is terminated if either of the following two conditions is satisfied.

1. The absolute values of all the components of the gradient vector are less than a small value ϵ (in our program we used $\epsilon = 10^{-6}$) and, the eigenvalues of the Hessian matrix at the solution point are all negative.

2. A maximum number of iterations(e.g. 300) is reached. It is also noted that according to our experience with the simulated data sets the program does converge after about 5-30 iterations, depending on the problem.)

We were able to use the algorithm outlined above to solve the likelihood equations. However, we discovered that for some of the data sets the likelihood surface has saddle points and multiple local maxima.

In section 2.6 we present a more detailed description of the likelihood surface of Weibull mixture distributions and offer two simulated data sets which have likelihood surfaces with multiple maxima.

Finding maxima starting from a saddle point

Existence of saddle points forces us to check whether the solution point is really a maximum. At the solution point the eigenvalues of the Hessian matrix are also

computed. If all the eigenvalues are negative indicating that the Hessian matrix is negative definite, the solution point is identified as a local maximum. If not, the solution point is identified as a saddle point. If the solution point is a local maximum it is difficult to examine whether it is the global maximum and therefore, it is accepted as the maximum likelihood estimate. It is also noted that one can get a rough idea about the global behaviour of the likelihood surface by starting with several different initial points and examining each solution point. However, we did not explore this possibility.

If the solution point is a saddle point it does not give estimates that can be accepted as maximum likelihood estimates. At a saddle point all the components of the gradient vector are almost zero and the Newton Raphson algorithm does not help us to take a further step. Therefore, the need arises to develop techniques to find maxima starting from a saddle point. Now we describe the algorithm that we developed to find maxima starting from saddle points.

Suppose that at the step v we observe the saddle point β^v and let $\underline{v}_j$ be the eigenvector corresponding to the positive eigenvalue (we never observed more than one positive eigenvalue) of the Hessian matrix at β^v. Consider,

$$f(a) = l(\beta^v + a\underline{v}_j) = f(\beta^v) + a\underline{v}_j^T \bigtriangledown l + \frac{a^2}{2}\underline{v}_j^T H \underline{v}_j + \text{negligible terms}.$$

We note that in a small neighbourhood of β^v, the gradient of the surface $\bigtriangledown l$ is almost zero, and $\underline{v}_j^T H \underline{v}_j$ is positive so that the function $f(a)$ increases in both directions of $\underline{v}_j$ away from β^v. Therefore, we use the algorithm

$$\beta^{v+1} = \beta^v \pm a\underline{v}_j \tag{2.20}$$

to move away from the saddle point. This is repeated as long as the likelihood is increased after each step. Suppose at a certain step v', $l^{v'+1} \leq l^{v'}$. Even if the Hessian at that point is negative definite there is no guarantee that we are moving in the correct direction. This is because now the sign of $(\beta^{v'+1} - \beta^{v'})^T \bigtriangledown l = (\pm a\underline{v}_j)^T \bigtriangledown l$ is not determined by H. It is also noted that this can happen only after moving out

from the saddle point and therefore now we can use the Newton Raphson method $\beta^{v'+1} = \beta^{v'} - (H)^{-1} \bigtriangledown l^{v'}$. It is to be noted that if H is not negative definite at step v', $H^* = H - kI$ takes the place of H in the above equation. (Again k is a scalar chosen to make H^* negative definite.) Now, an increase in l is guaranteed unless the step size is too large. Therefore, if l is not increased, we cut the step size in half repeatedly until we observe an increase in the likelihood. Once we observe an increase in l we go back to the iteration defined by the equation 2.20 which requires less computational effort. This is repeated until we observe maxima on both directions of $\underline{v}_j$ and finally we choose the one which corresponds to the largest value of the likelihood function.

It is also noted that if the first observed solution point is a saddle point we are almost always bound to see three solution points from our iterative scheme: two local maxima and a saddle point.

Finding an initial guess:

To use the iterative schemes presented, initial estimates of the parameters are required. Method of moments is suggested as an applicable method to find an initial guess. An estimation procedure using the method of moments is presented in Falls [7]. Macdonald [12] provides an estimation procedure using a weighted least squares estimator. Bayesian estimation of the parameters of a mixture of Weibull distributions is offered in Sinha [14].

2.6 Behaviour of the likelihood surface for samples from Weibull mixture populations

A well separated two component mixture population is one in which almost all the small observations belong to one component and almost all the large observations belong to

the other component. Figure 2.15 illustrates one such population. Figures 2.1, 2.2 are two examples of mixture populations which do not have well separated components.

In section 2.3.1 we have already noted that the likelihood surface for samples of Weibull mixture populations appear to be flat over a wide range of the parameter space. The other striking feature is that when the component densities of the Weibull mixture populations are not very well separated the likelihood surfaces for samples of such populations possess more than one maximum, and saddle points. It is also worthwhile to note that we were not interested to see if there are any minima on the likelihood surface. Each of these appear as solutions of the likelihood equations. The differences in likelihood values at these different solution points are so small that they would not be judged significant by a likelihood ratio test. For convenience we will refer to the local maximum which corresponds to the larger likelihood value as local maximum(1) and the other as local maximum(2).

To illustrate this now we offer two data sets(Appendix A) which are simulated from the populations with parameters given in table 2.1.

data set	parameter				
	shape(1)	shape(2)	scale(1)	scale(2)	proportion
1	2	3	3	0.9	0.5
2	1	3	2	4	0.5

Table 2.1: Population parameters of data sets given in Appendix A

Figures 2.1, 2.2 show the component densities and the mixture density in each case. In the figures, the solid line denotes the mixture density while the dotted lines denote the component densities.
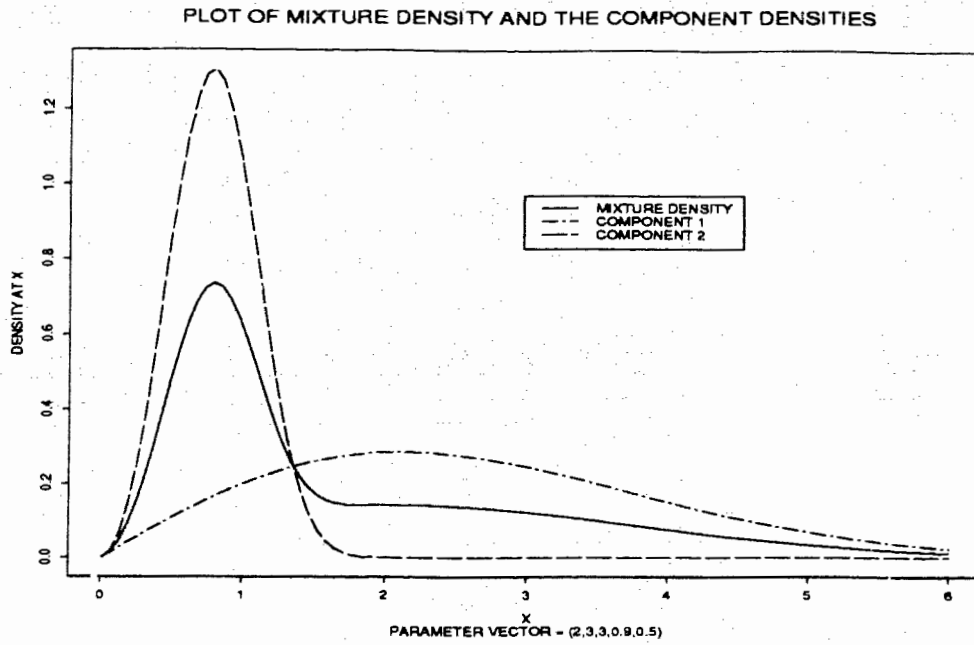
Figure 2.1: Plot of mixture density and the component densities: Data set (1)
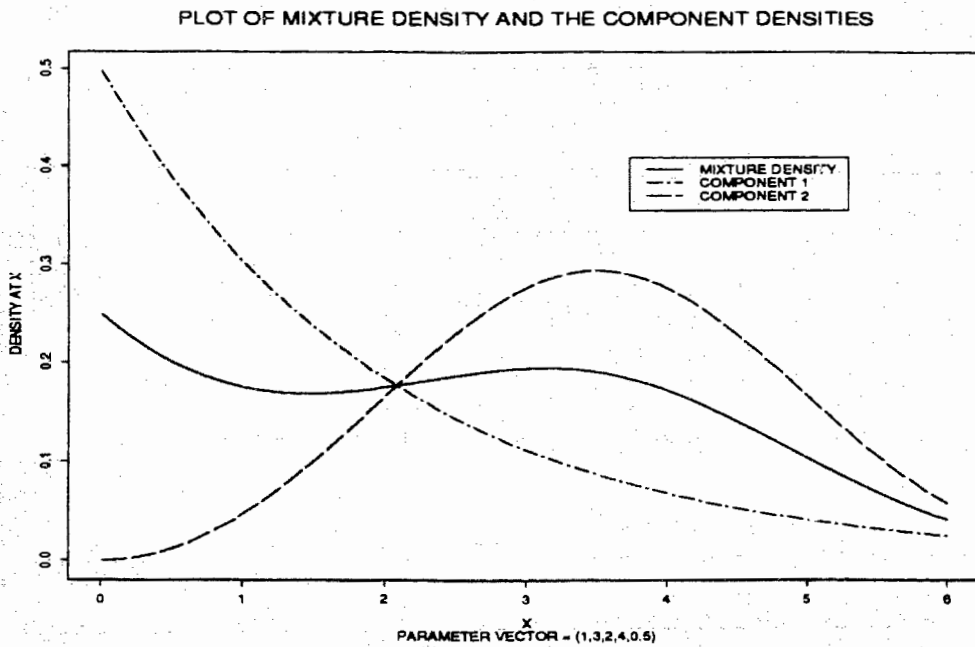
Figure 2.2: Plot of mixture density and the component densities: Data set (2)

22

Data set (1): (Appendix A.1)

Data set(1), a sample of size 100, which is attached in Appendix (A), was generated from a two component mixture population. Each component distribution is from the two parameter Weibull family. The algorithm which we developed was used to find the estimates of the parameters. The estimates at different solution points together with the parameter values that were used to generate the data set are given in table 2.2.

	parameter					l
	shape(1)	shape(2)	scale(1)	scale(2)	proportion	
true parameters	2	3	3	0.9	0.5	-139.9548
local(1)	1.271	4.154	2.279	1.035	0.631	-136.5221
local(2)	2.160	2.414	3.422	0.979	0.375	-137.3578
saddle	1.693	2.628	2.967	0.968	0.457	-137.4828

Table 2.2: Parameter estimates and the value of the log-likelihood(l) at the three solution points: Data set(1)

Remarks on table 2.2:

- It is seen that the individual parameter estimates provided by the three solution points are quite different. This motivates us to look at the fitted distributions provided by the three solution points.

- The differences in the log-likelihood values at the three solution points are so small that they would not be judged significant by a likelihood ratio test.

- According to table 2.2 the local maximum(2) seems to provide a better fit compared to the local maximum(1) even though the latter one corresponds to the larger value of the likelihood.

Now we look at the fitted distributions using the estimates at each of the three solution points.

	parameter					l
	shape(1)	shape(2)	scale(1)	scale(2)	proportion	
true parameters	2	3	3	0.9	0.5	-139.9548
local(1)	1.271	4.154	2.279	1.035	0.631	-136.5221
standard errors	0.141	0.988	0.298	0.066	0.090	
gradient	0.024	0.002	-0.009	-0.021	-0.026	

Table 2.3: Parameter estimates at the local maximum(1): Data set(1)

Remarks on tables 2.3- 2.9:

- In tables 2.3- 2.9, the roots have been rounded to three decimal places. Values of the gradient presented in the table were calculated at the root achieved by the algorithm after rounding. Near the roots the gradient vector has a rapidly changing size; the gradients at the roots actually presented in the table (that is, rounded to 3 decimal places) have a much larger magnitude compared to what we really observe from the algorithm.

- It is noted that the variances of the estimates are computed by inverting the negative Hessian matrix at the maximum likelihood estimate. However, in the case when there is more than one local maximum the estimates are less reliable and the actual variances of the estimates are expected to be higher than those recorded in the tables.

- Another estimate for the variances of the estimates would be to use the inverse of $-H'$ where H' is the average of the Hessian matrices evaluated at the maximum likelihood estimate using several samples(e.g. 500). This may be a better estimate

than the one we have provided. Due to the time constraint we had to be satisfied with the estimates provided because the latter requires more computational effort.

- In the case of saddle points the estimates are not the maximum likelihood estimates. Thus, the theory of maximum likelihood does not provide estimates for the standard errors.

Figure 2.3 shows the fitted distribution using the estimates at the local maximum(1) together with the true distribution and the empirical distribution of the data.

TRUE PARAMETER VECTOR = (2,3,3,0.9,0.5),FITTED PARAMETER VECTOR = (1.2713612,4.1541081,2.2788484,1.0349146,0.6308951)
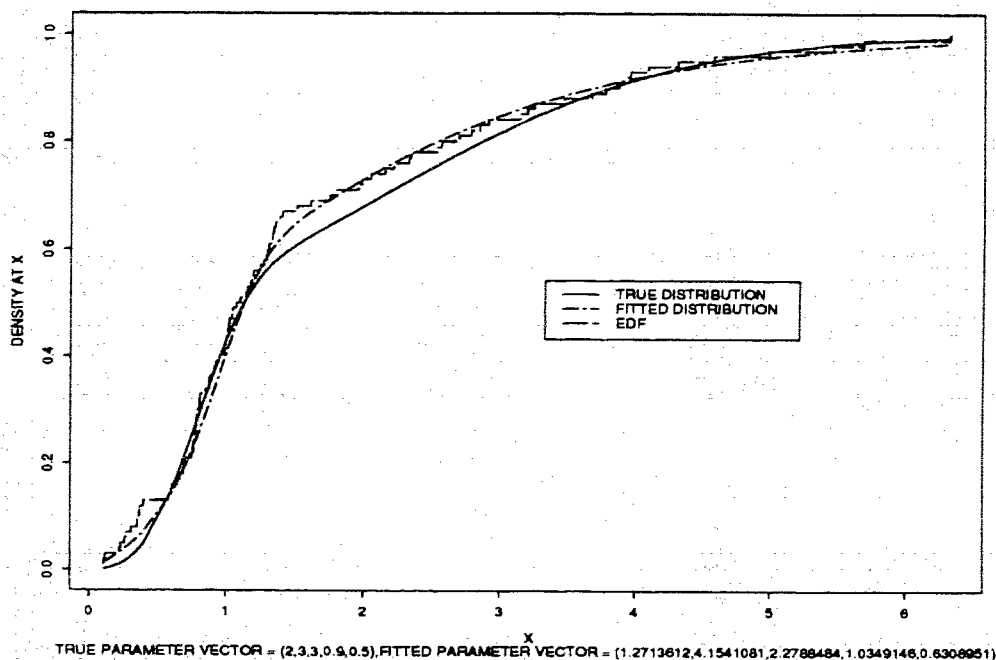
Figure 2.3: Fitted distribution at the local maximum(1): Data set (1)

As a further test of goodness of fit the observed cumulatives were plotted against the fitted cumulatives at each of the solution point. Figure 2.4 illustrates this plot using the estimates at the local maximum(1). For ease of comparison, the straight line through the origin with slope 1 is also displayed.

Figure 2.4: P-P plot using parameters at the local maximum(1): Data set(1)

The estimates of the parameters and their standard errors corresponding to the local maximum(2) are given in table 2.4.

	parameter					
	shape(1)	shape(2)	scale(1)	scale(2)	proportion	l
true parameters	2	3	3	0.9	0.5	-139.9548
local(2)	2.160	2.414	3.422	0.979	0.375	-137.3578
standard errors	0.670	0.345	0.577	0.074	0.099	
gradient	0.004	-0.000	0.010	0.049	0.118	

Table 2.4: Parameter estimates at the local maximum(2): Data set(1)

Figure 2.5 shows the fitted distribution using the estimates at the local maximum(2) together with the true distribution and the empirical distribution of the data.

TRUE PARAMETER VECTOR = (2,3,3,0.9,0.5), FITTED PARAMETER VECTOR = (2.1601438,2.4141572,3.4224680,0.9791251,0.3753321)

Figure 2.5: Fitted Distribution at the local maximum(2) : Data set(1)

Figure 2.6 illustrates the plot of observed cumulative against the fitted cumulative using the estimates at the local maximum. Again this indicates a good fit.

PLOT OF OBSERVED CUMULATIVE AGAINST THE FITTED CUMULATIVE

Figure 2.6: P-P plot using parameters at the local maximum(2): Data set(1)

The estimates of the parameters corresponding to the saddle point are given in table 2.5.

	parameter					l
	shape(1)	shape(2)	scale(1)	scale(2)	proportion	
true parameters	2	3	3	0.9	0.5	-139.9548
saddle	1.693	2.628	2.967	0.968	0.457	-137.4828
gradient	-0.012	0.002	0.005	-0.052	-0.028	

Table 2.5: Parameter estimates at the saddle point: Data set(1)

Figure 2.7 shows the fitted distribution using the estimates at the saddle point together with the true distribution and the empirical distribution of the data.

TRUE PARAMETER VECTOR = (2,3,3,0.9,0.5),FITTED PARAMETER VECTOR = (1.692728,2.628158,2.967274,0.967826,0.456922)

Figure 2.7: Fitted Distribution at the saddle point: Data set(1)

Figure 2.8 illustrates the plot of observed cumulative against the fitted cumulative using the estimates at the saddle point.

PLOT OF OBSERVED CUMULATIVE AGAINST THE FITTED CUMULATIVE
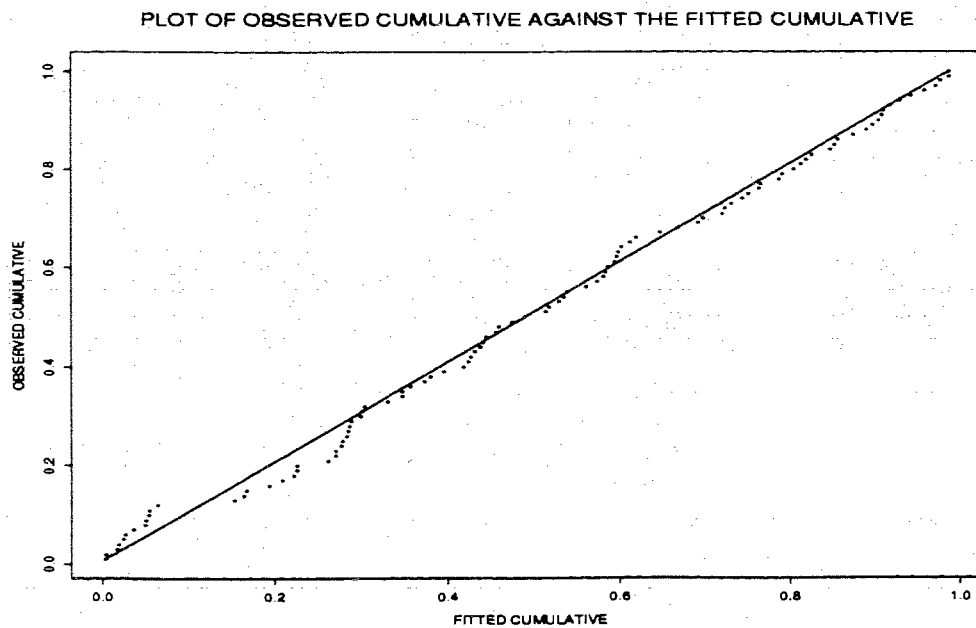
Figure 2.8: P-P plot using parameters at the saddle point: Data set(1)

Remarks:

- The individual parameter estimates at the three solution points are quite different yet each provides a good fit to the data.

- The estimates provided by the saddle point seem to be closer to the true parameter values. This motivated us to check whether the log-likelihood function is concave at the true parameter value. The Hessian matrix and the eigenvalues of the Hessian matrix at the true parameter values were computed. One of the eigenvalues was found to be positive indicating that the log-likelihood function is not concave at the true parameter value.

Data set (2): (Appendix 1.1)

Data set(2), again a sample of size 100, which is attached in Appendix (A), was generated from a two component mixture with one exponential component and one non-exponential component. The estimates at different solution points together with the parameter values that were used to generate the data set are given in table 2.6.

	parameter					l
	shape(1)	shape(2)	scale(1)	scale(2)	proportion	
true parameters	1	3	2	4	0.5	-190.655
local(1)	1.330	2.502	0.320	4.018	0.180	-186.1819
local(2)	0.831	3.310	1.300	4.302	0.365	-187.3228
saddle	0.831	3.160	1.162	4.299	0.341	-187.3294

Table 2.6: Parameter estimates and the value of the log-likelihood(l) at the three solution points: Data set(2)

Remarks on table 2.6:

- Unlike the previous case the estimates provided by the saddle point and the local maximum(2) appear to be very close.

- Again the local maximum(2) seems to provide a better fit compared to the local maximum(1) even though the latter one corresponds to the larger value of the likelihood.

- The differences in the log-likelihood values at the three solution points is again very small and would not be judged significant by a likelihood ratio test.

Now we examine the fits provided by each of these solution points. The estimates of the parameters and their standard errors corresponding to the observed local maximum(1) are given in table 2.7.

	parameter					l
	shape(1)	shape(2)	scale(1)	scale(2)	proportion	
true parameters	1	3	2	4	0.5	-190.655
local(1)	1.330	2.502	0.320	4.018	0.180	-186.1819
standard errors	0.372	0.280	0.102	0.211	0.047	
gradient	-0.004	0.011	0.034	-0.012	-0.235	

Table 2.7: Parameter estimates at the local maximum(1) : Data set(2)

Figure 2.9 shows the fitted distribution using the estimates at the local maximum(1) together with the true distribution and the empirical distribution of the data.

TRUE PARAMETER VECTOR = (1,3,2,4,0.5), FITTED PARAMETER VECTOR = (1.3301374,2.5019904,0.3199733,4.0175225,0.1796013)
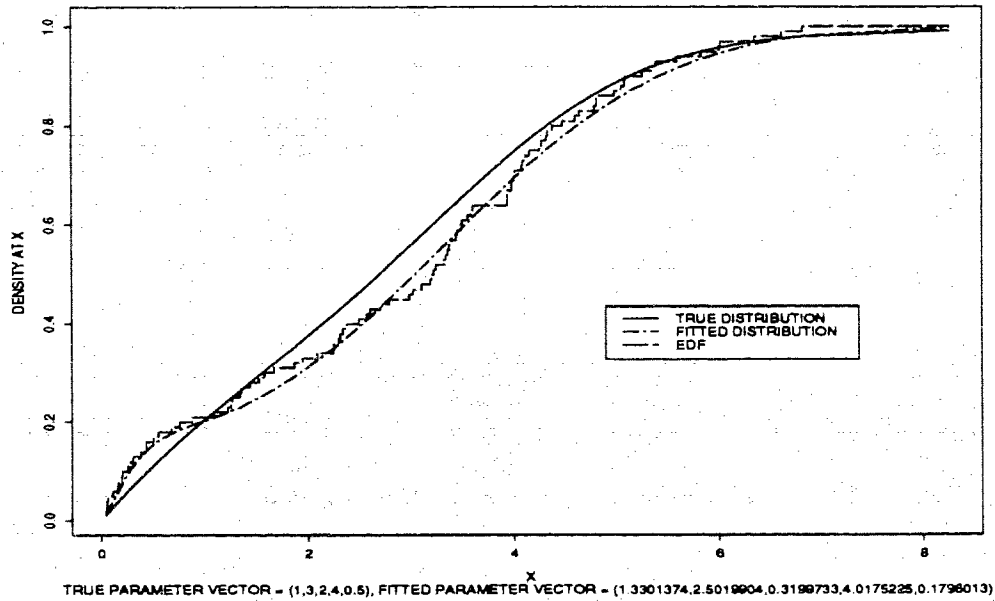
Figure 2.9: Fitted Distribution at the local maximum(1) : Data set(2)

Figure 2.10 illustrates the plot of observed cumulative against the fitted cumulative using the estimates at the local maximum(1).
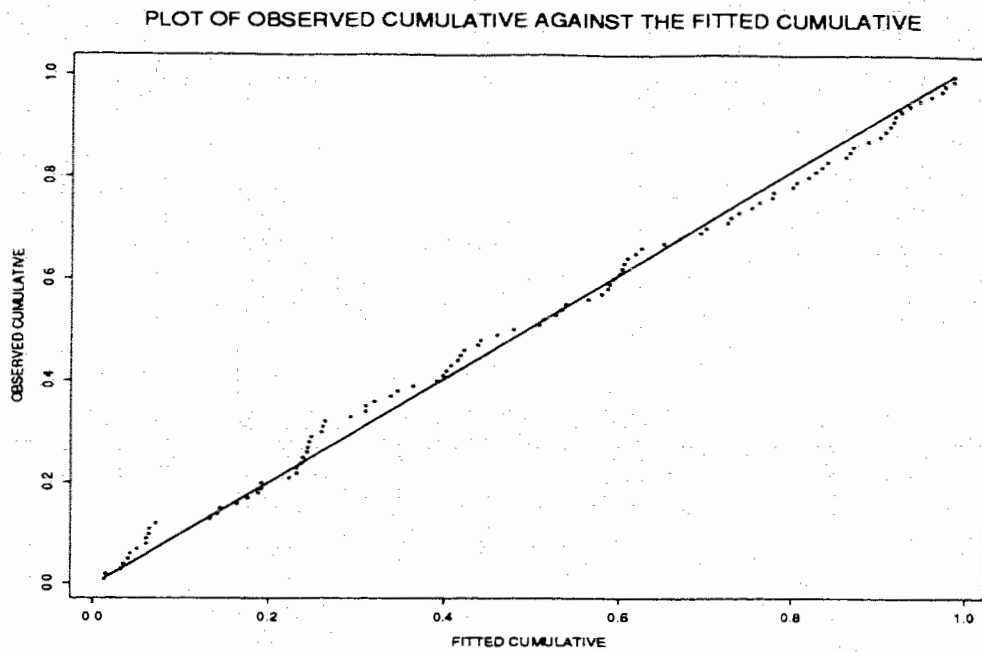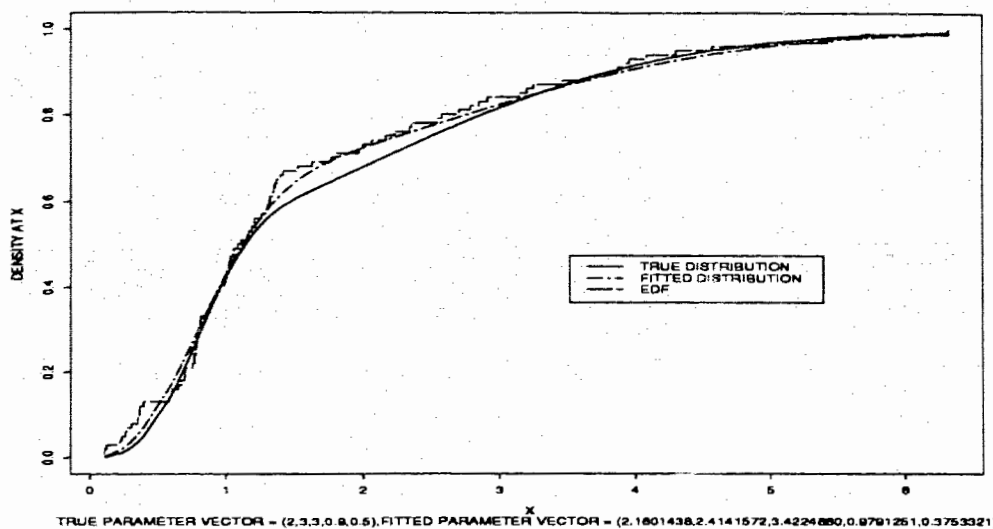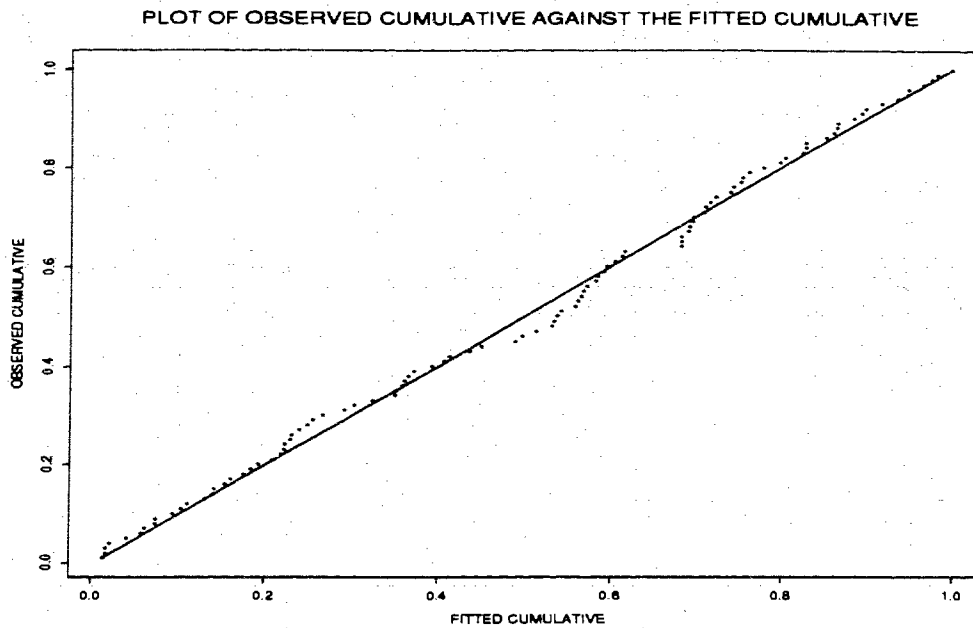

Figure 2.10: P-P plot using parameters at the local maximum(1) : Data set(2)

The estimates of the parameters and their standard errors corresponding to the local maximum(2) are given in table 2.8.

	parameter					l
	shape(1)	shape(2)	scale(1)	scale(2)	proportion	
true parameters	1	3	2	4	0.5	-190.655
local(2)	0.831	3.310	1.300	4.302	0.365	-187.3228
standard errors	0.124	0.876	0.747	0.220	0.139	
gradient	-0.036	-0.004	0.002	0.001	0.028	

Table 2.8: Parameter estimates at the local maximum(2) : Data set(2)

Figure 2.11 shows the fitted distribution using the estimates at the local maximum(2) together with the true distribution and the empirical distribution of the data.
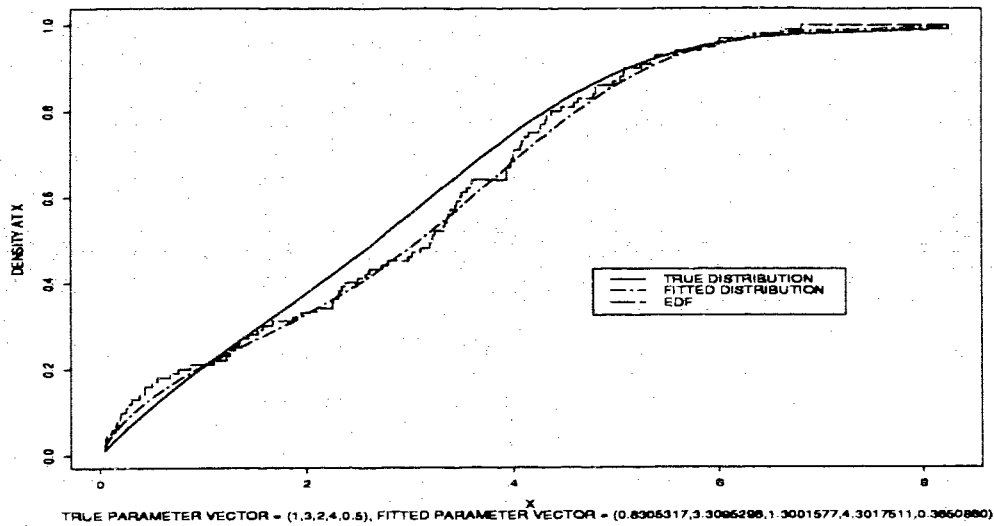
Figure 2.11: Fitted Distribution at the local maximum(2) : Data set(2)

Figure 2.12 illustrates the plot of observed cumulative against the fitted cumulative using the estimates at the local maximum(2).
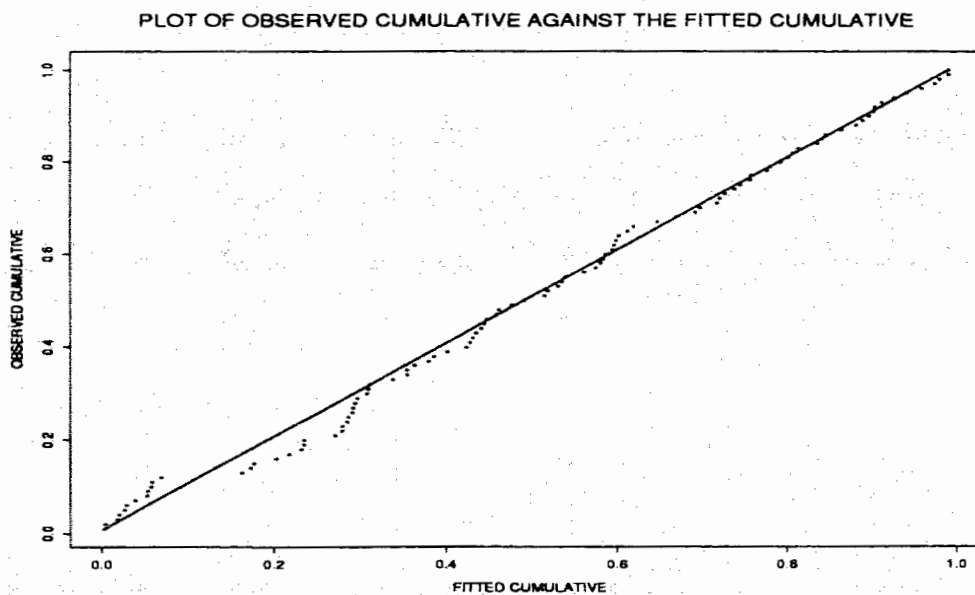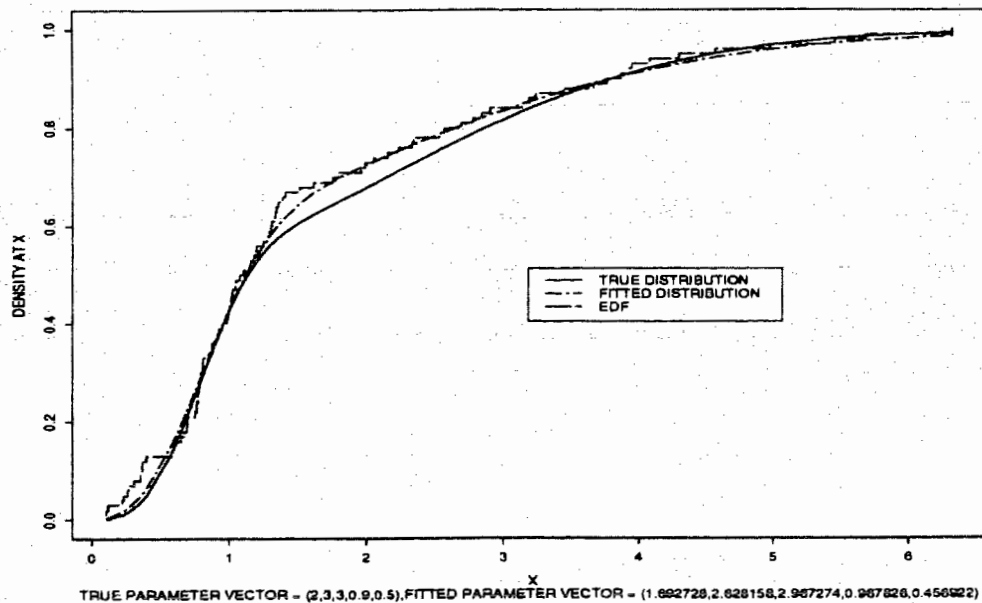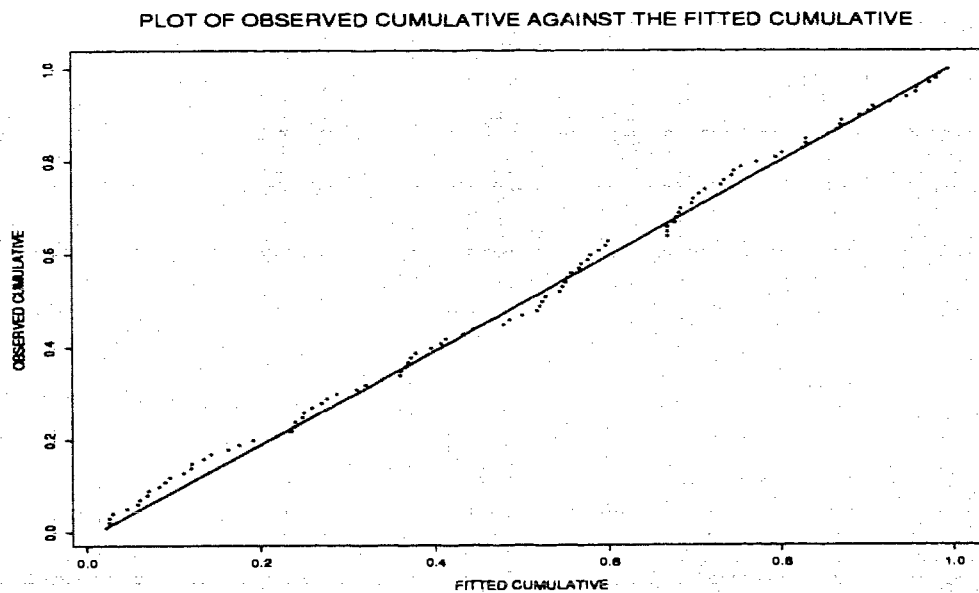

Figure 2.12: P-P plot using parameters at the local maximum(2): Data set(2)

The estimates of the parameters corresponding to the saddle point are given in table 2.9.

	parameter					log
	shape(1)	shape(2)	scale(1)	scale(2)	proportion	likelihood
true parameters	1	3	2	4	0.5	-190.655
saddle	0.831	3.160	1.162	4.299	0.341	-187.3294
gradient	0.010	0.003	0.003	-0.002	-0.023	

Table 2.9: Parameter estimates at the saddle point : Data set(2)

Figure 2.13 shows the fitted distribution using the estimates at the saddle point together with the true distribution and the empirical distribution of the data.

TRUE PARAMETER VECTOR - (1,3,2,4,0.5), FITTED PARAMETER VECTOR - (0.8311558,3.1598606,1.1619223,4.2990128,0.3408734)
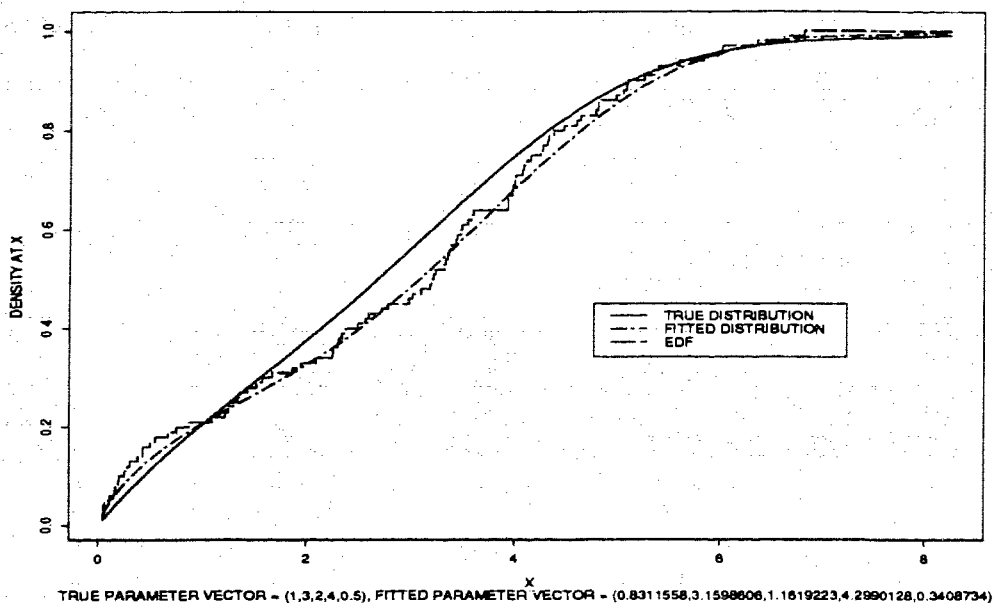
Figure 2.13: Fitted Distribution at the saddle point : Data set(2)

Figure 2.14 illustrates the plot of observed cumulative against the fitted cumulative using the estimates at the saddle point.

PLOT OF OBSERVED CUMULATIVE AGAINST THE FITTED CUMULATIVE
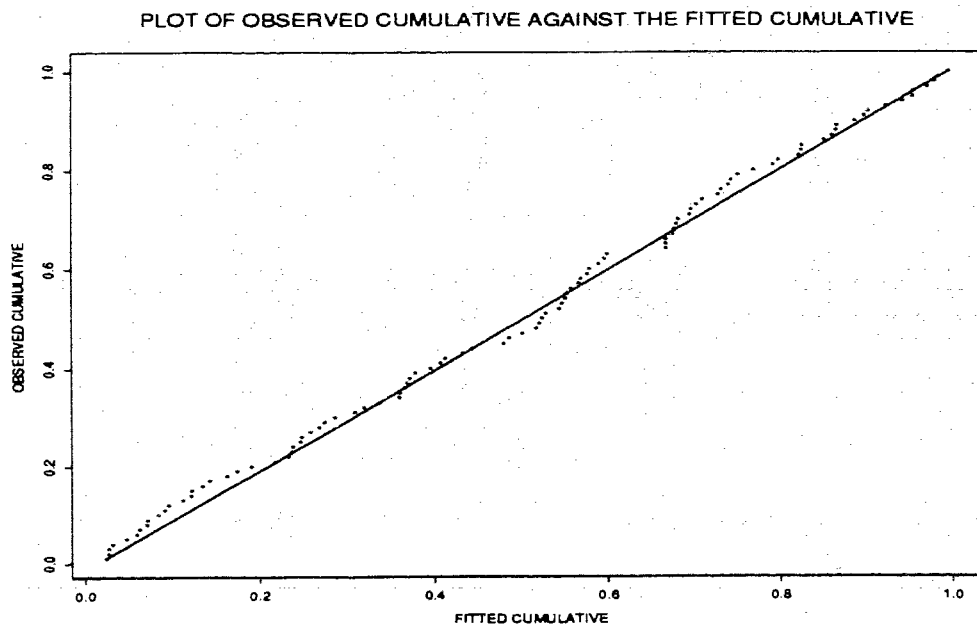
Figure 2.14: P-P plot using parameters at the saddle point: Data set(2)

Remarks:

- The parameter estimates provided by the local maximum(1), which is the local maximum corresponding to the largest log-likelihood value, are quite different from the true parameter values. The estimates provided by the other two solution points are closer to each other and are also closer to the true parameter values compared to those provided by the local maximum(1). However, each solution point seems to provide a good fit to the data.

- In this case all the eigenvalues of the Hessian matrix at the true parameter values were found to be negative indicating that the log-likelihood is concave in a small neighbourhood of the true parameter.

A population with well separated components:

parameter vector = (2,8,1,4,0.5)

To make the comparison easier, the component densities and the mixture density of a population with well separated components are shown in figure 2.15. The plot of the fitted distribution, the true distribution and the empirical distribution function is also given for one data set from this population. This is the population which is labelled population number '5' in chapter 6.

PLOT OF MIXTURE DENSITY AND THE COMPONENT DENSITIES
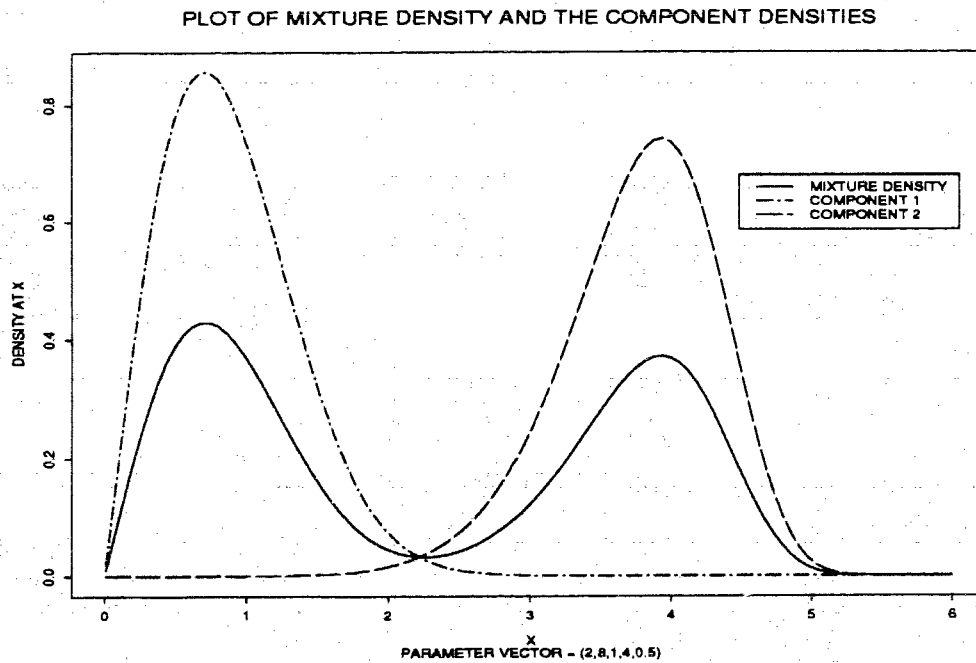
Figure 2.15: Plot of densities: well separated components
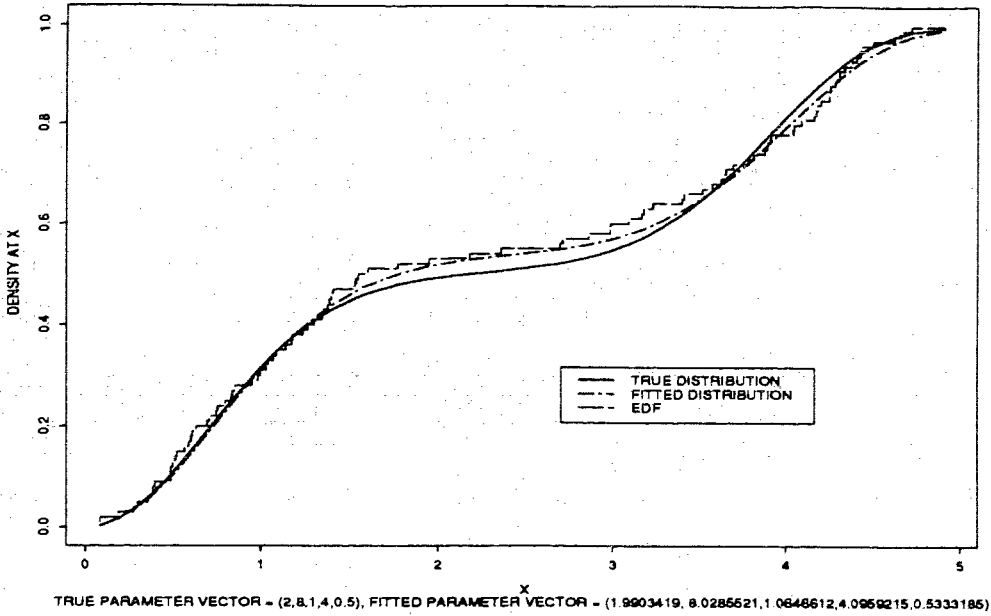
Figure 2.16: Fitted distribution : population (5)

Remarks:

- In the case of well separated components we never witnessed multiple maxima and saddle points.

- Even though the three data sets mentioned were all generated from mixture populations, the empirical distribution functions of the former cases do not give as clear evidence of the presence of a mixture as in the latter case. Therefore, we also fitted a two parameter Weibull distribution instead of a mixture distribution in each case. However, we found that a two parameter Weibull distribution cannot give a satisfactory fit in any of the two cases.

37

Chapter 3

Theory of EDF statistics

Suppose we are given a random sample of n observations from a continuous distribution $F(x)$. The goodness of fit problem could simply be stated as a test of the null hypothesis H_0 that, the distribution F belongs to some parametric family, in this case, the family of all two component Weibull mixtures. In this chapter, we propose tests based on the empirical distribution function(EDF) of the sample.

3.1 The Empirical Distribution Function

Let $x_1 \leq x_2 \leq \ldots \leq x_n$ be the order statistics, that is, the observations arranged in increasing order.

$F(x)$ is the probability that the random variable takes a value less than or equal to x. The empirical distribution function is a (non-parametric) estimate of F.

In a more mathematical sense, the empirical distribution function(EDF) is $F_n(x)$ defined by,

$$F_n(x) = \frac{Number \ of \ observations \leq x}{n} = \frac{1}{n} \sum_{i=1}^{n} 1[x_i \leq x] \qquad -\infty < x < \infty$$

where $1[a \le b] = 1$ if $a \le b$ and $1[a \le b] = 0$ if $a > b$. Thus, $F_n(x)$ is the proportion of observations less than or equal to x and we expect $F_n(x)$ to be close to $F(x)$.

3.2   EDF statistics

Any statistic which measures the difference between $F_n(x)$ and $F(x)$ is called an EDF statistic. EDF statistics are basically divided into two classes, supremum statistics and quadratic or integral statistics.

3.2.1   The Supremum Statistics

These are based on the largest difference between $F_n(x)$ and $F(x, \widehat{\theta})$ where $\widehat{\theta}$ is an estimate of θ. Four such statistics are,

$$
\begin{aligned}
D^+ &= \sup_{-\infty < x < \infty} (F_n(x) - F(x, \widehat{\theta})) \\
D^- &= \sup_{-\infty < x < \infty} (F(x, \widehat{\theta}) - F_n(x)) \\
D &= \max(D^+, D^-) = \sup_{-\infty < x < \infty} |F_n(x) - F(x, \widehat{\theta})|
\end{aligned}
$$

which is the Kolmogorov-Smirnov statistic and

$$
V = D^+ + D^-
$$

and is called the Kuiper statistic. If θ is completely specified as θ_0, then θ_0 is used to replace $\widehat{\theta}$ in the above expressions.

3.2.2   The Quadratic Statistics(or the Integral Statistics)

Quadratic statistics are based on the weighted and integrated squared discrepancies between F_n and F given by the Cramér-von Mises family,

$$
Q = n \int_{-\infty}^{\infty} \{F_n(x) - F(x)\}^2 \Psi(x) dF(x)
$$

39

where $\Psi(x)$ is called the weight.

1). When $\Psi(x) = 1$ the statistic is called the Cramér-von Mises statistic;

$$W^2 = n \int_{-\infty}^{\infty} \{F_n(x) - F(x)\}^2 dF(x).$$

Our particular interest here is in this statistic which is discussed in detail later.

2). When $\Psi(x) = \frac{1}{F(x)(1-F(x))}$ the statistic is called the Anderson-Darling statistic;

$$A^2 = n \int_{-\infty}^{\infty} \{F_n(x) - F(x)\}^2 \frac{1}{F(x)[1 - F(x)]} dF(x).$$

The other well-known statistic is the Watson statistic U^2 defined by,

$$U^2 = n \int_{-\infty}^{\infty} \{F_n(x) - F(x) - \int_{-\infty}^{\infty} [F_n(t) - F(t)] dF(t)\}^2 dF(x)$$

3.3 How to use EDF statistics in testing goodness of fit

The key features of any hypothesis testing problem can be summarised as follows.

1. Settle on an appropriate test statistic with a known distribution.

2. Compute the value of the test statistic.

3. Compute the p-values based on the distribution of the statistic and either reject or accept the null hypothesis by comparing the p-values with the tolerance level α.

Stephens(1986)[17] provides some key facts on how to settle on an appropriate test statistic. We only focus on W^2 and will explain below how to compute the asymptotic distribution of W^2. The remaining two steps are dealt with in sections 3.6 and 3.7.

3.4 Distributional properties of quadratic EDF statistics

It is known that the asymptotic distribution of any of the three statistics W^2, A^2 and U^2 is that of

$$\int_0^1 Y^2(t)dt,$$

where $Y(t)$ is an appropriate Gaussian Process, whose mean is 0 and whose covariance function $\rho(s, t)$ depends on the statistic, on $F(x, \alpha)$ and on the parameters to be estimated. We will discuss the case of the Cramér-von Mises statistic in more detail.

3.4.1 Limiting distribution of the Cramér-von Mises test statistic

Various authors have shown that under suitable regularity conditions
(Cramér [3], Durbin [4]), the limiting distribution of W^2 under the null hypothesis is that of $W^2 = \sum_{j=1}^{\infty} \lambda_j Z_j^2$ where the $Z_j's$ are independent $N(0, 1)$ variables and the $\lambda_j's$ are the eigenvalues of the covariance kernel ρ, namely, the solutions of the eigenvalue equation,

$$\int_0^1 \rho(s, t)f(t)dt = \lambda f(s).$$

For W^2, the appropriate covariance ρ is $\rho(s, t) = \rho_0(s, t) = \min(s, t) - st$, if the null hypothesis is a simple hypothesis and $\widehat{\rho}(s, t) = \min(s, t) - st - \psi(s)'I^{-1}\psi(t)$, where $\psi(s) = \frac{\partial F}{\partial \alpha}[F^{-1}(s, \widehat{\alpha}), \alpha]$, if the null hypothesis is composite.

Durbin and Knott [5] provides an extensive discussion of the case when the null hypothesis is simple. The case when the null hypothesis is composite is discussed in Durbin, Knott and Taylor [6].

While it is possible to compute the λ's in closed form when the null hypothesis is

simple this seems impossible in the case at hand. In the next section we discuss the problem of computing the eigenvalues.

3.5 Computation of the eigenvalues of the covariance kernel

When the null hypothesis is simple the eigenvalues are the solutions of

$$\lambda f(s) = \int_0^1 \rho_0(s,t) f(t) dt,$$

$$\rho_0(s,t) = \min(s,t) - st.$$

With some algebra one can prove that the eigenvalues are $1/\pi^2 j^2$ $j = 1, 2, \ldots$ and the corresponding eigenfunctions are $\sqrt{2} \sin(\pi j s)$ $j = 1, 2, \ldots$.

When the null hypothesis is composite, we have to solve

$$\lambda f(s) = \int_0^1 \widehat{\rho}(s,t) f(t) dt,$$

$$\widehat{\rho}(s,t) = \rho_0(s,t) - \psi(s)' I^{-1} \psi(t),$$

$$\psi(s) = \frac{\partial F}{\partial \alpha} [F^{-1}(s, \widehat{\alpha}), \alpha]$$

where I is the information of a single observation, $\widehat{\alpha}$ is an asymptotically efficient estimate of α and F is the distribution function.

For covariances of this form Stephens [15] has shown how the eigenvalues and eigenfunctions may be calculated from those of ρ_0. The method requires that, ϕ, which is defined by $\phi(s)\phi(t) = \psi(s)' I^{-1} \psi(t)$ be expanded in a Fourier series in the eigenfunctions of ρ_0. The method of finding such ϕ is also given. Stephens [15] provides eigenvalues which are computed using this method for the cases when the underlying distribution is normal or exponential. This is extended to the extreme value distribution in Stephens [16]. However, in our problem this seems impossible because of the difficulty

associated with inverting the mixture distribution function and the difficulty associated with finding a closed form for the information matrix.

In this study we use what we call a brute force approach to find the eigenvalues of the covariance kernel $\rho(s,t)$. The procedure is as follows.

If λ is an eigenvalue of $\rho(s,t)$ and $f(s)$ is an eigenfunction corresponding to the eigenvalue λ then $\lambda f(s) = \int_0^1 \rho(s,t)f(t)dt$. Divide the interval $[0,1]$ into $(m+1)$ sub-intervals each of length $1/(m+1)$. Then,

$$
\begin{aligned}
\lambda f(i/(m+1)) &= \int_0^1 \rho(i/(m+1),t)f(t)dt \\
&\approx \frac{1}{m}\sum_{j=1}^{m} \rho(i/(m+1),j/(m+1))f(j/(m+1)), \quad \text{for sufficiently large } m
\end{aligned}
$$

Let

$$
V' = \begin{pmatrix} f(1/(m+1)) \\ \vdots \\ f(m/(m+1)) \end{pmatrix}
$$

Let Q be the $m \times m$ matrix whose (i,j)th element is $Q_{ij} = \frac{1}{m}\rho(i/(m+1),j/(m+1))$. Then the above set of equations can be written as, $\lambda V' = QV'$, and finding the eigenvalues of ρ reduces to the discretised problem of finding the eigenvalues of the matrix Q. In our analysis, we wrote a program to create the matrix Q using an estimate of $\rho(s,t)$ at each point (s,t); s, $t = \frac{k'}{(m+1)}$, $k' = 1,2,...,m$. Then we use the S-plus function "eigen" to compute the eigenvalues of the matrix Q. This way we can find estimates $\hat{\lambda}$ for the eigenvalues λ of ρ. The application of the above procedure will be discussed in chapter 4.

3.6 Computing formulas for EDF statistics

It is quite cumbersome to use the previous definitions to compute the values of the EDF statistics.

Fortunately, the fact that the probability integral transformation $z = F(x, \theta)$, when $F(x, \theta)$ is the true distribution, yields a new set of random variables which are uniformly distributed between 0 and 1 makes the computation problem quite simple. Without loss of generality let us assume that $x_1, x_2, ..., x_n$ are the order statistics of the original sample and $z_i = F(x_i, \theta)$.

As is explained in Stephens(1986)[17] the hypothesis testing problem is then equivalent to that of testing the hypothesis that $z_1, z_2, ..., z_n$ is an ordered sample of independent uniform[0,1] variables.

This leads the following formulae for computing EDF statistics using the z-values.

$$
\begin{aligned}
D^+ &= \max_{1 \le i \le n} \{\frac{i}{n} - z_i\} \\
D^- &= \max_{1 \le i \le n} \{z_i - \frac{(i-1)}{n}\} \\
D &= \max_{1 \le i \le n} (D^+, D^-) \\
V &= D^+ + D^- \\
W^2 &= \sum_{i=1}^{n} \{z_i - \frac{(2i-1)}{2n}\}^2 + \frac{1}{12n} \\
A^2 &= -n - \frac{1}{n} \sum_{i=1}^{n} \{(2i-1) \ln z_i + (2n+1-2i) \ln(1 - z_i)\} \\
U^2 &= W^2 - n\{0.5 - \frac{1}{n} \sum_{i=1}^{n} z_i\}^2.
\end{aligned}
$$

It is noted that when the null hypothesis is composite we actually use $z_i = F(x_i, \widehat{\theta})$ to compute the test statistics.

3.7 Computation of the p-values for the Cramér-von Mises test statistic

In section 3.4 we have already noted that when the unknown parameters are estimated by a method which is asymptotically efficient, the test statistic of interest,

$$W_n^2 = \sum_{i=1}^{n} \left\{ z_i - \frac{(2i-1)}{2n} \right\}^2 + \frac{1}{12n}$$

(where $z_i = F(x_i, \hat{\theta})$ are the probability integral transforms of the ordered statistics x_i) has asymptotically a weighted chi square distribution. The weights λ_i are the eigenvalues of the covariance kernel of the Gaussian process $\{\hat{Y}_n(z)\}$ where $\hat{Y}_n(z) = \sqrt{n}[F_n(z) - z]$. In section 3.6 we dealt with the problem of finding estimates $\hat{\lambda}$ for the eigenvalues λ.

Thus, the computation of the p-values hinges on the calculation of

$$P(\sum_{i=1}^{\infty} \hat{\lambda}_i \chi_{(i)}^2 \geq x),$$

where $x(\geq 0)$ is the value of the test statistic.

Chen[2] provides literature related to the computation of the above probability together with a comparison of the accuracy of the methods. We will only discuss Imhof's method(Imhof [8]) which we used in this study.

Computation of the p-values using Imhof's method:

If $V_1, V_2, ..., V_m$ are independent random variables with V_i having a non-central chi-square distribution on h_i degrees of freedom with non-centrality parameter δ_i then Imhof[8] shows

$$P(\sum_{i=1}^{m} \lambda_i V_i > x) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \frac{\sin \theta(u)}{u\rho(u)} du,$$

where

$$\theta(u) = \frac{1}{2}\sum_1^m [h_r \tan^{-1}(\lambda_r u) + \delta_r^2 \lambda_r u (1 + \lambda_r^2 u^2)^{-1}] - \frac{1}{2}xu,$$

$$\rho(u) = \prod_i^m (1 + \lambda_r^2 u^2)^{h_r/4} \exp\{\frac{1}{2} \sum_{i=1}^m (\delta_r \lambda_r u)^2 / (1 + \lambda_r^2 u^2)\}.$$

The weights $\widehat{\lambda}_i$ are computed as discussed in section 3.5. We have software to evaluate the integral numerically when the degrees of freedom are all equal to 1 and the non-centrality parameters are all 0.

Chapter 4

Computation of the eigenvalues of the covariance kernel: Applications

In this chapter, we use the brute force approach discussed in chapter 3 to compute eigenvalues of the covariance kernel from a given estimate of the parameter vector. As a preview of the degree of accuracy of our approach we compute the eigenvalues in some cases considered in Stephens [15] and compare with the tables provided. In section 4.1, we discuss the application of the above approach for the case when the underlying distribution is normal. In section 4.2 we extend the above ideas for the exponential distribution, and in section 4.3, for Weibull and for extreme value distributions. Section 4.4 deals with the case when the data come from a Weibull mixture population.

Outline of the procedure:

- If the covariance kernel ρ of the distribution of interest is unknown, find an estimate of ρ.

- Decide on a suitable number of subdivisions m in the brute force approach, which is a compromise between the degree of accuracy and the computational effort.

47

The accuracy of the results is increased when the number of subdivisions of [0,1] is increased. We used 200 subdivisions.

- Create the matrix Q whose elements are $Q_{s,t} = \rho(s,t)$, for $s,t = 1/(m+1), 2/(m+1), ..., m/(m+1)$.

- The eigenvalues of Q are taken as estimates for the eigenvalues of ρ. We wrote a program in 'Splus' to create the matrix Q and to compute the eigenvalues.

4.1 Eigenvalues of the covariance kernel for the normal distribution

In this section we closely follow the notation and procedure discussed in Stephens[15]. Let $F(x,\theta)$ be the distribution function of the normal distribution where $\theta = (\mu, \sigma^2)$. Following Stephens[15] we discuss four cases based on the prior knowledge about the parameter vector.

Case 0: the underlying distribution is completely specified by the null hypothesis. (i.e. both μ, σ^2 are known under H_0)

Case 1: $\theta = (\mu, \sigma^2)$, with σ^2 is known and μ is to be estimated by the sample mean $\bar{x}$

Case 2: μ is known, but σ^2 is to be estimated by $\sum_{i=1}^n \frac{(x_i - \mu)^2}{n}$

Case 3: μ, σ^2 are both unknown, and are respectively to be estimated using $\bar{x}$ and $s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{(n-1)}$

Stephens [15] presents the covariance kernel for each of the above cases which takes the form;

Case 0: $\rho(s,t) = \rho_0(s,t) = \min(s,t) - st$

Case 1: $\rho(s,t) = \rho_0(s,t) - \phi_1(s)\phi_1(t)$,

where $\phi_1(s) = \frac{\partial \Phi}{\partial \theta}(\Phi^{-1}(s,\theta),\theta) = \frac{1}{\sqrt{(2\pi)}}\exp(-(\Phi^{-1}(s))^2/2)$ and Φ denotes the distribution function of the standard normal distribution.

Case 2: $\rho(s,t) = \rho_0(s,t) - \phi_2(s)\phi_2(t)$,

where $\phi_2(s) = \frac{\partial \Phi}{\partial \theta}(\Phi^{-1}(s,\theta),\theta) = \frac{1}{\sqrt{(2\pi)}}\frac{\Phi^{-1}(s)}{2^{1/2}}\exp\{-(\Phi^{-1}(s))^2/2\}$

Case 3: $\rho(s,t) = \rho_0(s,t) - \phi_1(s)\phi_1(t) - \phi_2(s)\phi_2(t)$

The procedure outlined above was used to find the estimates for the eigenvalues. The estimates for the first ten eigenvalues are given in Table 4.1.

i	1	2	3	4	5	6	7	8	9	10
Case 1	1.844	0.539	0.254	0.147	0.096	0.067	0.050	0.039	0.031	0.025
Case 2	1.351	0.439	0.217	0.130	0.086	0.061	0.045	0.035	0.028	0.023

Table 4.1: Estimated eigenvalues of the covariance kernel for the normal distribution

Remarks on table 4.1:

- In the above table the values recorded are the estimated eigenvalues multiplied by 100.

- We have only recorded the estimates of the eigenvalues($\widehat{\lambda}$) corresponding to those provided in Stephens [15]. There are also eigenvalues of the form $1/\pi^2 j^2$ with j even for Case 1 or odd for Case 2. Our estimates of these eigenvalues also agree well.

- The estimates of the eigenvalues for Case 3, are found to be the same as those given by the two sets arising in Cases 1 and 2 excluding those of the form $1/\pi^2 j^2$.

- On comparison with the eigenvalues of ρ provided in Stephens [15], we can say that the tabulated estimates provided by the eigenvalues of Q are quite close to the

eigenvalues of ρ. The sum of our estimated eigenvalues which is also the trace of Q, provides a good estimate for $\int_0^1 \rho(s,s)ds$. This is different from the infinite sum of the eigenvalues provided by Stephens [15]. Since the first few eigenvalues agree quite well the rest of the estimated eigenvalues can expected to be substantially different from those values of Stephens [15] procedure.

• It is also worthwhile to note that the precision of the above estimates can be improved by computing the eigenvalues for several different m and using a method like Richardson extrapolation or by plotting the eigenvalues vs. $1/m$ and choosing the values corresponding to $m = 0$. However, we did not try doing this because our main concern was to find the estimates of the eigenvalues corresponding to the two parameter Weibull mixture distribution.

4.2 Eigenvalues of the covariance kernel for the exponential distribution

Following Stephens [15] when the underlying distribution is exponential with parameter unknown, the covariance kernel is, for W^2,

$$\rho(s,t) = \rho_0(s,t) - \Phi(s)\Phi(t), \text{ where } \Phi(s) = (1-s)\ln(1-s).$$

The first ten estimates of the eigenvalues we got in this case using the brute force approach are given in table 4.2. The eigenvalues agree well with those given in Stephens [15].

i	1	2	3	4	5	6	7	8	9	10
$\hat{\lambda}$	4.223	1.721	0.820	0.512	0.335	0.243	0.181	0.142	0.113	0.093

Table 4.2: Estimated eigenvalues of the covariance kernel for the exponential distribution

4.3   Eigenvalues of the covariance kernel for the Weibull and extreme value distributions

To make the discussion complete we briefly include the Weibull and extreme value distributions as well. For more details one may refer to Stephens [17].

We limit our discussion to two parameter Weibull distributions. For the case of test of fit for the three parameter Weibull distribution, one may refer to Lockhart and Stephens [11].

As is explained in Chandra et.al. [1] we first note that when the sample is from the two parameter Weibull distribution, $F_1(x) = 1 - \exp[-\{x/\delta\}^\gamma]$, $x \geq 0$, we can make use of the transformation $y = -\log(x)$ to transform the data to those coming from the extreme value distribution, $F(y) = \exp[-\exp-\{(y-\zeta)/\theta\}]$, $-\infty < y < \infty$ with $\zeta = -\log\delta$ and $\theta = 1/\gamma$. Thus, the problem is reduced to that of testing that the data come from the extreme value distribution, with ζ or θ or both unknown. So we will only discuss the case that the data come from the extreme value distribution.

Following Stephens [16] we discuss four test situations:

Case 0 : both ζ and θ are known.

Case 1 : θ is known and ζ is to be estimated.

Case 2 : ζ is known and θ is to be estimated.

Case 3 : both ζ and θ are unknown and must be estimated.

The covariance kernel for each of the above cases takes the following form.

Case 0 : $\rho(s,t) = \rho_0(s,t) = \min(s,t) - st$.

Case 1 : $\rho(s,t) = \rho_0(s,t) - \Phi_1(s)\Phi_1(t)$, where, $\Phi_1(s) = s\log(s)$.

Case 2 : $\rho(s,t) = \rho_0(s,t) - \Phi_2(s)\Phi_2(t)$, where

$$\Phi_2(s) = s(\log(s))\{-\log(\log(s))\}/1.350437$$

Case 3 : In this study we only used the brute force approach to compute the estimates for the eigenvalues of ρ in cases 1 and 2. For case 3, one may refer to Stephens [16] or Stephens [17].

As is explained in Stephens [17] case 1 above reduces to testing for the exponential distribution. So, we only tabulate the first ten estimates of the eigenvalues corresponding to case 2.

i	1	2	3	4	5	6	7	8	9	10
$\widehat{\lambda}$	9.869	1.456	1.108	0.469	0.396	0.229	0.202	0.136	0.123	0.090

Table 4.3: Estimated eigenvalues of the covariance kernel for the extreme value distribution

Remarks:

- It is noted that we only computed estimates for the eigenvalues of the covariance kernel in each of the distributions discussed above. It would be useful to compute the critical points and compare them with the published values in each case j, but due to lack of time we had to limit our scope. Our primary concern in this study is to compute the p-values for the Cramér-von Mises statistic in two component Weibull mixtures.

- For the percentage points in each of the above cases one might refer to Stephens [16].

4.4 Eigenvalues of the covariance kernel for a mixture of two Weibull distributions

In this section we deal with the case that the data come from a population which is made up of two Weibull component populations mixed together in fixed proportions.

The data come from the distribution,

$$f(x, c_1, \theta_1, c_2, \theta_2, p) = p \, f_1(x, c_1, \theta_1) + (1 - p) \, f_2(x, c_2, \theta_2),$$

where f_i is a two parameter Weibull density.

$$f_i(x, c_i, \theta_i) = \frac{c_i}{\theta_i} \left(\frac{x}{\theta_i}\right)^{c_i-1} \exp[-(\frac{x}{\theta_i})^{c_i}], \; 0 \le x < \infty, \; c_i, \theta_i > 0$$

We will only discuss the case of the full composite hypothesis.

As we discussed in the previous chapter, the covariance kernel of the limiting process Y is

$$
\begin{aligned}
\widehat{\rho}(s, t) &= \rho_0(s, t) - \Psi(s)' I^{-1} \Psi(t), \quad \text{where} \\
\rho_0(s, t) &= \min(s, t) - st \quad \text{and} \\
\Psi(s) &= \frac{\partial F}{\partial \alpha}[F^{-1}(s, \widehat{\alpha}), \alpha]
\end{aligned}
$$

Unlike the previous cases, due to the complexity of the mixture distribution function, we cannot find a closed form solution for $\widehat{\rho}$. Therefore, it is not possible to create the matrix Q as described in the previous cases. Instead we estimate the covariance kernel at any given point and thus create the matrix Q using these estimates. In the next chapter we describe the problems that we encounter in computing the eigenvalues and the software that we developed to find estimates for the eigenvalues.

Chapter 5

Computation of the p-values:

Development of software

In the previous chapters we discussed the theory behind the computation of the p-values of the Cramér-von Mises statistic in Weibull mixture populations. However, the complexity of the mixture population problem makes it difficult to find a closed form for the covariance kernel. Therefore, the method described in Stephens [15] for the normal case cannot be extended to find the eigenvalues of the covariance kernel for the mixture model at hand. We begin this chapter with a brief description of the problems that we encounter in computing the eigenvalues and the p-values for the test statistics in Weibull mixture populations. Then, we describe the software that we developed to find an estimate of the covariance kernel at a given point and henceforth to find estimates for the eigenvalues using brute force approach. These estimated eigenvalues are used to find an approximate p-value for the test statistic.

The covariance kernel of the limiting process is,

$$\hat{\rho}(s,t) = \rho_0(s,t) - \Psi(s)'I^{-1}\Psi(t), \quad \text{where}$$

$$\rho_0(s,t) = \min(s,t) - st \quad \text{and}$$

$$\Psi(s) = \frac{\partial F}{\partial \alpha}[F^{-1}(s, \widehat{\alpha}), \alpha]$$

The problems associated with the computation of p-values in the mixture case are:

- In the above formula, I is the information matrix corresponding to a single observation. Since the information matrix is not available in closed form we were forced to estimate it.

- Computation of $\Psi(s)$ requires the inverse of the mixture distribution function and this has to be done numerically. In normal and other cases considered before we were able to solve for $\Psi(s)'I^{-1}\Psi(t)$ algebraically.

- In the previous cases the eigenvalues and the critical points do not depend on the values of the parameters. Thus critical points could be given in a small table. In the mixture case, critical points would depend on all five parameters.

These problems make direct computation of eigenvalues and preparation of tables of critical points impractical. Instead we developed software to compute an approximate p-value based on the asymptotic distribution for the covariance using an estimate of the information matrix.

5.1 An estimate of the information matrix

The information matrix for a sample of size n, I_n, is given by,

$$I_n = E\{[\frac{\partial \log L}{\partial \alpha}][\frac{\partial \log L}{\partial \alpha}]'\}.$$

If the estimated parameters are close to the actual parameters, the elements of I_n are almost the same as the elements of the matrix $-H$, where H is the matrix of second derivatives with respect to the parameters and is called the Hessian matrix. In our

procedure we estimate the information matrix of a single observation I, by the matrix $-H/n$, evaluated at the maximum likelihood estimate. We therefore find p-values from the distribution of $\int_0^1 Y^2(t)dt$; Y is a mean 0 Gaussian process with covariance function $\min(s,t) - st - \Psi(s)^T(-H/n)^{-1}\Psi(t)$, where $-H/n$ is the Hessian matrix evaluated at the maximum likelihood estimate. It is also noted that for the normal and exponential distributions the matrix $-H/n$ gives the exactly correct form of ρ.

5.2 Covariance kernel under composite hypothesis

The covariance function is

$$
\begin{aligned}
\rho(s,t) &= \min(s,t) - st - \Psi(s)^T(-H/n)^{-1}\Psi(t), \quad \text{where,} \\
\Psi(s) &= \frac{\partial F}{\partial \alpha}(F^{-1}(s,\widehat{\alpha}),\alpha).
\end{aligned}
$$

Computation of $\Psi(s)$ requires the inverse of the distribution function at the point s. This again cannot be found algebraically and now we describe the software that we developed to find this numerically.

Inverse of the mixture distribution function

The distribution function is,

$$
F(X,\underline{\alpha}) = p\left[1 - \exp\left\{-\left(\frac{x}{\theta_1}\right)^{c_1}\right\}\right] + (1-p)\left[1 - \exp\left\{-\left(\frac{x}{\theta_2}\right)^{c_2}\right\}\right]
$$

where $\alpha^T = (c_1, c_2, \theta_1, \theta_2, p)$. We first compute the inverse of the mixture distribution function at a given point numerically. Given t, we want to find x s.t. $F(x,\underline{\alpha}) = t$. This is exactly the same as finding zeros of $g(x) = F(x,\underline{\alpha}) - t$.

In our procedure we use the "Secant method", that is, the iterative scheme

$$
\begin{aligned}
x_{n+1} &= \frac{x_n g(x_{n-1}) - x_{n-1}g(x_n)}{g(x_n) - g(x_{n-1})}, \quad \text{where} \\
g(x) &= p\left[1 - \exp\left\{-\left(\frac{x}{\theta_1}\right)^{c_1}\right\}\right] + (1-p)\left[1 - \exp\left\{-\left(\frac{x}{\theta_2}\right)^{c_2}\right\}\right] - t.
\end{aligned}
$$

Two initial points are required to start with.

Finding an initial guess:

From equation (3.4) if we put p=0, g(x)=0 when,

$$1 - \exp\left\{\left(\frac{x}{\theta_2}\right)^{c_2}\right\} = t$$

$$\exp\left\{-\left(\frac{x}{\theta_2}\right)^{c_2}\right\} = 1 - t$$

$$-\left(\frac{x}{\theta_2}\right)^{c_2} = \log(1 - t)$$

We know that $\log(1 - t) < 0$, since $t > 0$. So, we can write the above equation as,

$$-\left(\frac{x}{\theta_2}\right)^{c_2} = -|\log(1 - t)|$$

$$\text{and thus,} \quad x_2 = \theta_2(|\log(1 - t)|)^{1/c_2}.$$

Note that we simply escape from the problem of $log(1 - t)$ being negative by using the absolute value. This step is very important for x_2 to be meaningful.

Similarly, when $p = 1$, $x_1 = \theta_1(|\log(1 - t)|)^{1/c_1}$.

We take x_1, x_2 as initial points. We found that in all the examples we tried it is the case that x_1 and x_2 are on the opposite sides of the root, and this scheme serves the purpose well with a satisfactory degree of accuracy.

The Newton Raphson method which uses,

$$x_{n+1} = x_n - \frac{g(x_n)}{g'(x_n)} \quad (\text{if } g'(x_n) \neq 0)$$

with the initial guess, $x_0 = px_1 + (1 - p)x_2$, was also found to work well except in a few cases where $g'(x), g''(x)$ are vanishingly small at certain points.

Here,

$$g(x) = p\left[1 - \exp\left\{-\left(\frac{x}{\theta_1}\right)^{c_1}\right\}\right] + (1 - p)\left[1 - \exp\left\{-\left(\frac{x}{\theta_2}\right)^{c_2}\right\}\right] - t.$$

$$g'(x) = p\frac{c_1}{\theta_1}\left(\frac{x}{\theta_1}\right)^{c_1-1}\exp\left\{-\left(\frac{x}{\theta_1}\right)^{c_1}\right\} + (1 - p)\frac{c_2}{\theta_2}\left(\frac{x}{\theta_2}\right)^{c_2-1}\exp\left\{-\left(\frac{x}{\theta_2}\right)^{c_2}\right\}.$$

In our problem we have to evaluate the inverse of the mixture distribution at all the points corresponding to the subdivisions of the interval $[0,1]$. In other words we have to evaluate the inverse of a vector of points; the components of the vector being the subdivisions of the interval $[0,1]$. It is noted that rather less computational time is required if we can extend the above program to compute the inverse of a given vector of points instead of evaluating the inverse of one point at a time. However, the difficulty is that different points in the vector require a different number of iterations to converge. Fortunately the language 'Splus' that we use has the ability to select those components which satisfy a certain given condition called the convergence criterion and do iterations with the selected ones for which the convergence criterion is not yet satisfied thereafter by recalling the same function. In our problem we take the condition to be: the difference between two consecutive iterated points is less than a small number ϵ (for example we used $\epsilon = 5 \times 10^{-6}$). If this condition is met for a particular point in the vector we consider that the convergence criterion has been met for that particular point and iterate with rest of the points, until the convergence criterion is met with all the points in the vector. The advantage of using the aforementioned procedure to compute the inverse is that we can save lot of computational time and effort by adopting this procedure. The function we wrote to find the inverse of a vector of points ("inv.Weib") is attached in Appendix B.

The elements of $\frac{\partial F}{\partial \alpha}$ are $\frac{\partial F}{\partial c_1}$, $\frac{\partial F}{\partial c_2}$, $\frac{\partial F}{\partial \theta_1}$, $\frac{\partial F}{\partial \theta_2}$, and $\frac{\partial F}{\partial p}$. These are given by,

$$\frac{\partial F}{\partial c_i} = p_i \left(\frac{x}{\theta_i}\right)^{c_i} \log\left(\frac{x}{\theta_i}\right) \exp\left\{-\left(\frac{x}{\theta_i}\right)^{c_i}\right\}$$

$$\frac{\partial F}{\partial \theta_i} = -p_i \left(\frac{c_i}{\theta_i}\right) \left(\frac{x}{\theta_i}\right)^{c_i} \exp\left\{-\left(\frac{x}{\theta_i}\right)^{c_i}\right\}, \quad \text{for } i = 1, 2$$

$$\frac{\partial F}{\partial p} = \exp\left\{-\left(\frac{x}{\theta_2}\right)^{c_2}\right\} - \exp\left\{-\left(\frac{x}{\theta_1}\right)^{c_1}\right\}$$

In the above equations $p_1 = p$ and $p_2 = (1 - p)$. Now we can approximate $\Psi(s)$ by, $\frac{\partial F}{\partial \alpha}(F^{-1}(s, \widehat{\alpha}), \alpha)$ evaluated at $\alpha = \widehat{\alpha}$ and $x = F^{-1}(s, \widehat{\alpha})$. This way we have an estimate for the covariance kernel at any given point (s, t).

5.3 Eigenvalues of the covariance kernel

We create the matrix Q whose $(s,t)th$ element is $\hat{\rho}(s,t) = \rho_0(s,t) - \Psi(s)'I^{-1}\Psi(t)$, for $s,t = 1,2,...,n$. Then as described in the previous chapter the eigenvalues of the covariance kernel ρ can be estimated by the eigenvalues of the matrix Q. In our program we used 100 subdivisions of the interval [0,1], in the brute force approach, and so Q is a 100×100 matrix. This way we find 100 eigenvalues of Q and thus estimates for 100 eigenvalues of ρ.

5.4 p-values based on the asymptotic distribution of the test statistic

Now we outline the procedure that is used to compute an approximate p-value based on the asymptotic distribution of the Cramér-von Mises statistic.

1. Estimate the parameters by the method of maximum likelihood.

2. Compute z_i by the probability integral transform. Use $z_i = F(x_i, \hat{\theta})$, where $\hat{\theta}$ is the maximum likelihood estimate of θ, when θ is not completely specified by the null hypothesis.

3. Compute W^2 using the computing formulas.

4. Estimate I by $-H/n$, where H is the Hessian matrix evaluated at the maximum likelihood estimate.

5. Compute Ψ numerically at a grid of points.

6. Evaluate the matrix Q using the results of 4 and 5.

7. Find the eigenvalues of Q. These are the weights to be used in the asymptotic distribution of the Cramér-von Mises statistic, which is a weighted chi-square distribution.

8. Find the probability that a linear combination of chi-squares (the coefficients being the weights mentioned in 7), exceeds the value of the test statistic in 3, using the procedure described in section 3.7.

We developed software using the language 'Splus' to implement the above procedure.

Chapter 6

Monte Carlo study on mixtures of two parameter Weibull distributions

6.1 Objective

The objective of the Monte Carlo study is to test the accuracy of the asymptotic results for the Cramér-von Mises statistic applied to Weibull mixture populations and to study further the behaviour of the likelihood function in Weibull mixture populations.

We begin this study by showing that, if the Cramér-von Mises statistic has the weighted chi square distribution mentioned in section 3.4 then the p-values for the test statistic are uniformly distributed on [0,1].

Let w be the value of the test statistic W_n computed using a sample of size n. Let $G(t, \theta, I)$ be the distribution of $\int_0^1 Y^2(t)dt$, where Y is a mean zero Gaussian process with covariance,

$$\rho(s,t) = \min(s,t) - st - \Psi(s,\theta)^T \; I^{-1} \; \Psi(t,\theta)$$

If θ is the true parameter value W_n converges in distribution to G.

That is, $P(W_n \leq w) \to G(w, \theta, I(\theta))$, where I is the information at θ.

If G_n is the exact distribution of W_n then,

$$
\begin{aligned}
P(1 - G_n(W_n) \leq u) &= P(G_n(W_n) > 1 - u) \\
&= P(W_n > G_n^{-1}(1 - u)) \\
&= 1 - G_n(G_n^{-1}(1 - u)) = u
\end{aligned}
$$

So that $1 - G_n(W_n)$ has a uniform distribution.

The asymptotic theory is that, $G_n(x) \to G(x, \theta, I(\theta))$ and also, $G(x, \widehat{\theta}, \widehat{I}) \to G(x, \theta, I(\theta))$ so that, $p = 1 - G(W_n, \widehat{\theta}, \widehat{I})$ is close to $1 - G_n(W_n)$ and therefore has a nearly uniform distribution. The validity of the asymptotic theory can thus be tested by checking to see if the p-values are approximately uniformly distributed when the null hypothesis is true.

6.2   Description

We chose five different Weibull mixture populations for the study. From each population, we generated 500 samples each of size 100. For each sample we estimated the parameters by maximum likelihood and the information matrix using $-H/n$, evaluated at the maximum likelihood estimate. Then we computed an approximate p-value following the method outlined in section 5.4 above. The p-values obtained for each population were tested for uniformity. In section 6.3 we describe the 5 mixture populations used. Results are in section 6.4. Conclusions are in section 6.5.

6.3 Populations used in the Monte Carlo study

Parameter values of the populations that are used in this study are given in table 6.1. Populations number 1, 2, 4 and 5 are mixtures of Weibull components, while population 3 is a mixture of an exponential and a Weibull component. Populations were chosen to range from poorly separated components to very well separated components. As we mentioned in section 2.6 if the component densities are not well separated, the likelihood functions for the data sets taken from such populations sometimes have more than one maximum and saddle points. In the figures which show the densities of the populations we have also marked whether or not saddle points are observed when sampling from that population.

| population | parameter | | | | |
	shape(1)	shape(2)	scale(1)	scale(2)	proportion
1	2	3	3	0.9	0.5
2	1.5	3	2	4	0.5
3	1	3	2	4	0.5
4	2	4	0.5	3	0.5
5	2	8	1	4	0.5

Table 6.1: Populations used in the Monte Carlo study

To ease the comparison of results in each case the plots of densities are given in the next section (Figures 6.1- 6.10) together with the results. In each figure, the solid line denotes the mixture density, while the dotted lines denote the component densities.

6.4 Results

As a preliminary look at the p-values histograms and plots of quantiles of the uniform distribution against the ordered p-values(Q-Q plots) were considered.

Figure 6.1: Density of population 1 : *(saddle points were observed)*
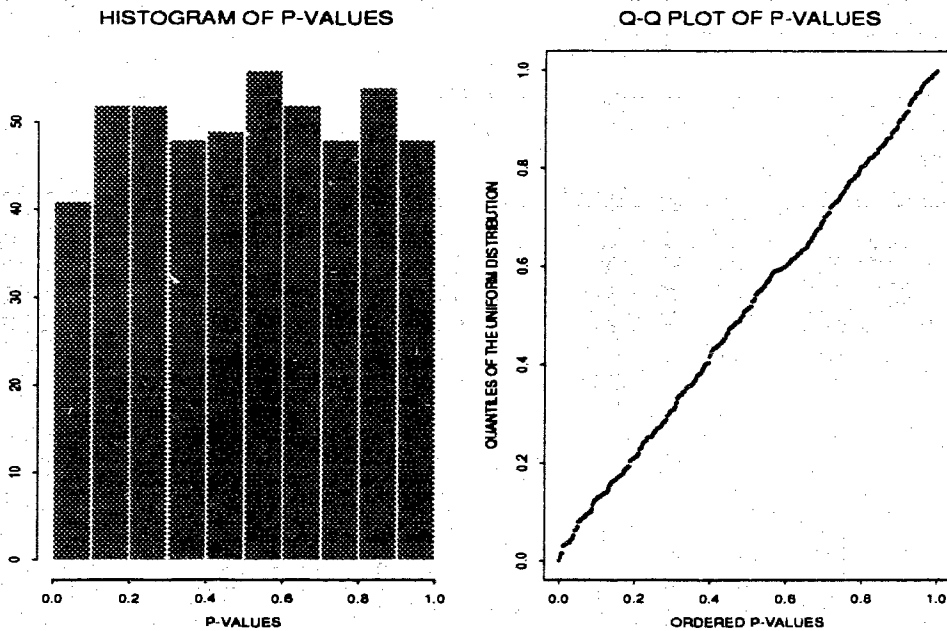
Figure 6.2: Histogram and Q-Q plot of p-values: Population 1
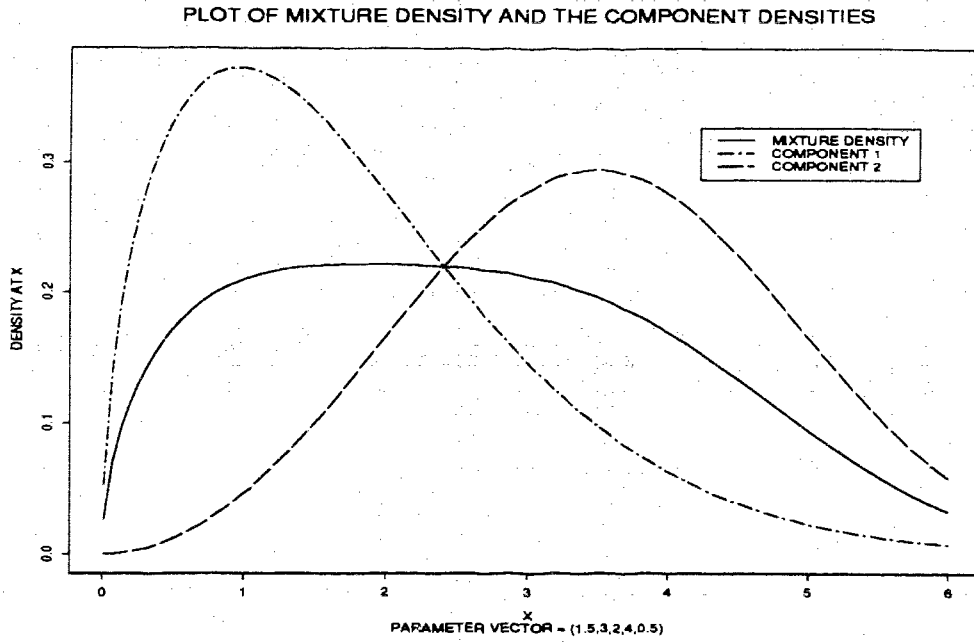
Figure 6.3: Density of population 2: *(saddle points were observed)*
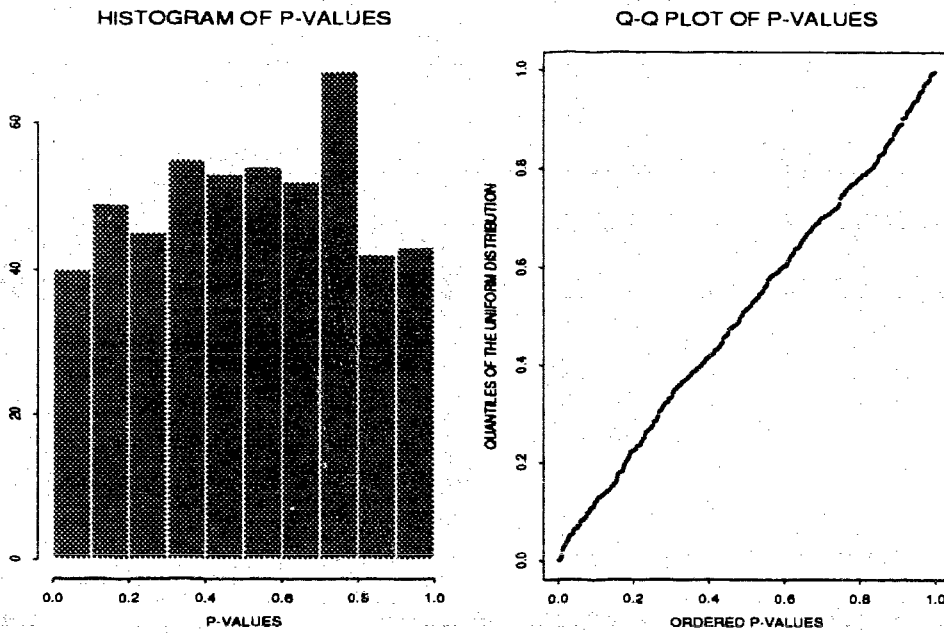
Figure 6.4: Histogram and Q-Q plot of p-values: Population 2
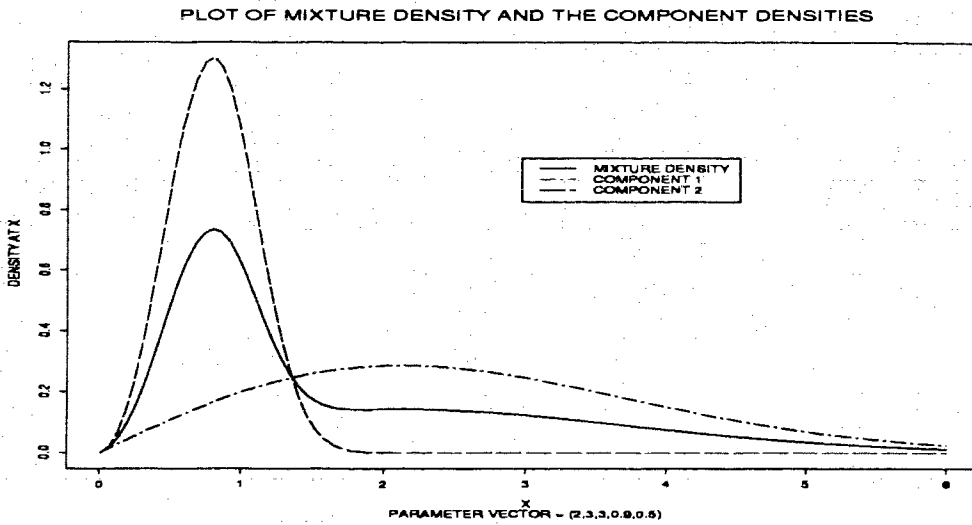
65

PLOT OF MIXTURE DENSITY AND THE COMPONENT DENSITIES
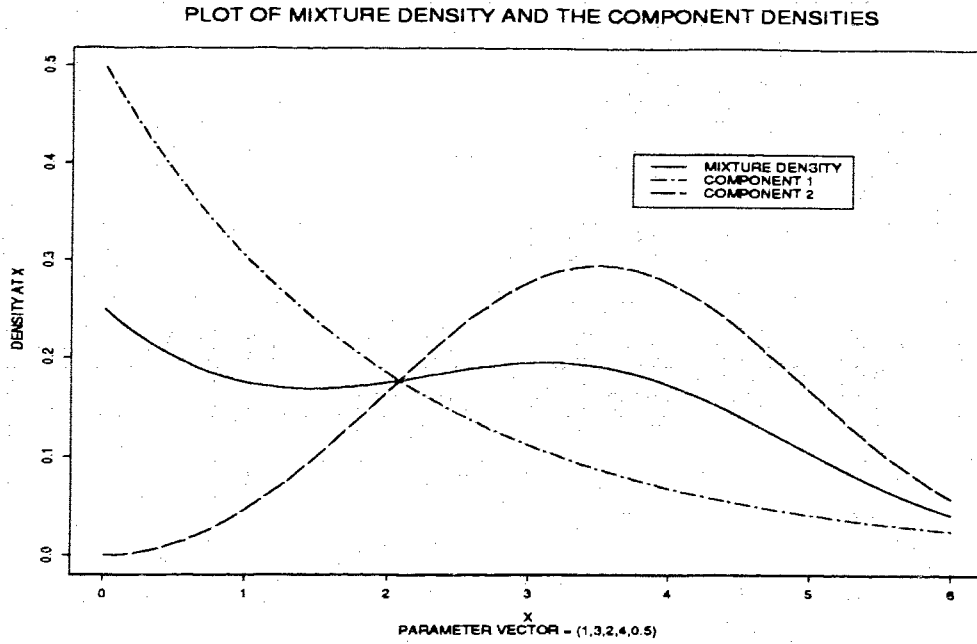
Figure 6.5: Density of population 3 : *(saddle points were observed)*
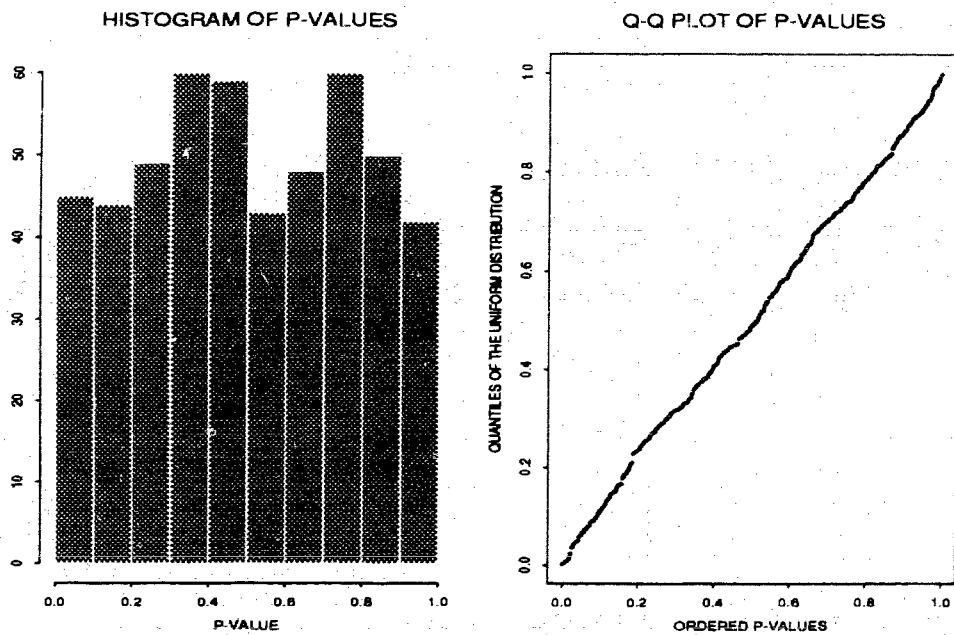
Figure 6.6: Histogram and Q-Q plot of p-values: Population 3

PLOT OF MIXTURE DENSITY AND THE COMPONENT DENSITIES
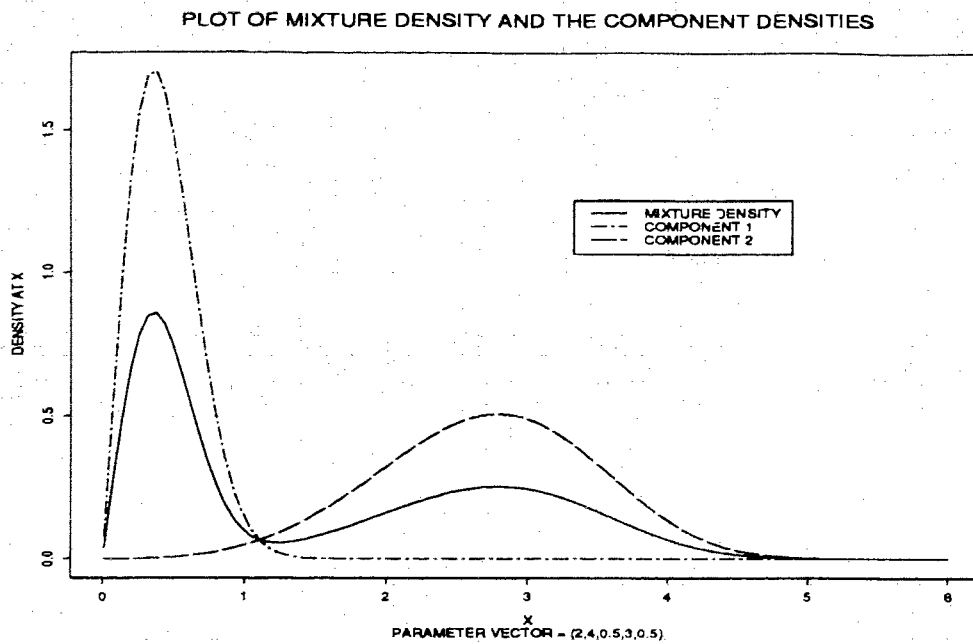
Figure 6.7: Density of population 4 : *(no saddle points were observed)*

HISTOGRAM OF P-VALUES   Q-Q PLOT OF P-VALUES
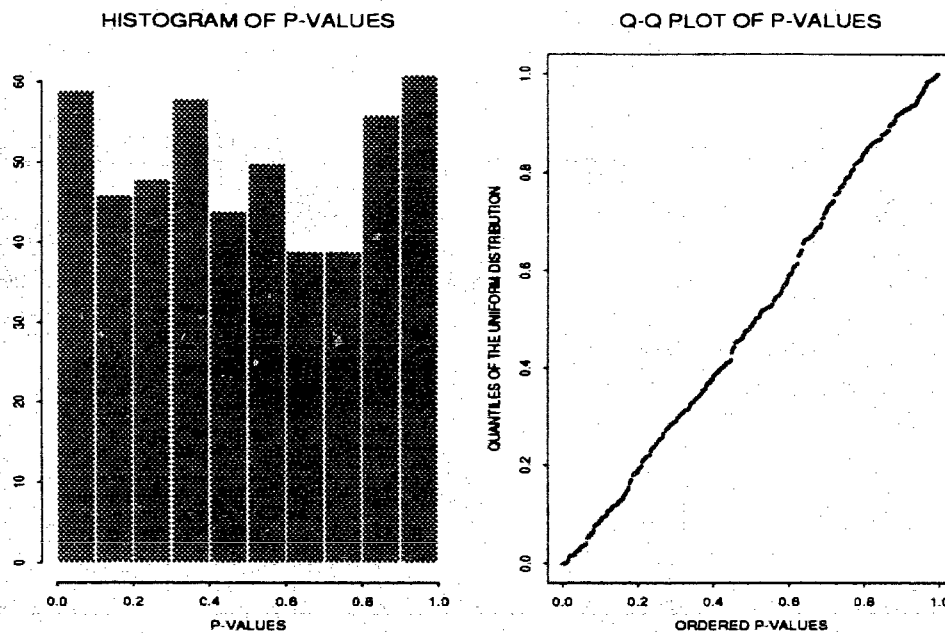
Figure 6.8: Histogram and Q-Q plot of p-values: Population 4

PLOT OF MIXTURE DENSITY AND THE COMPONENT DENSITIES
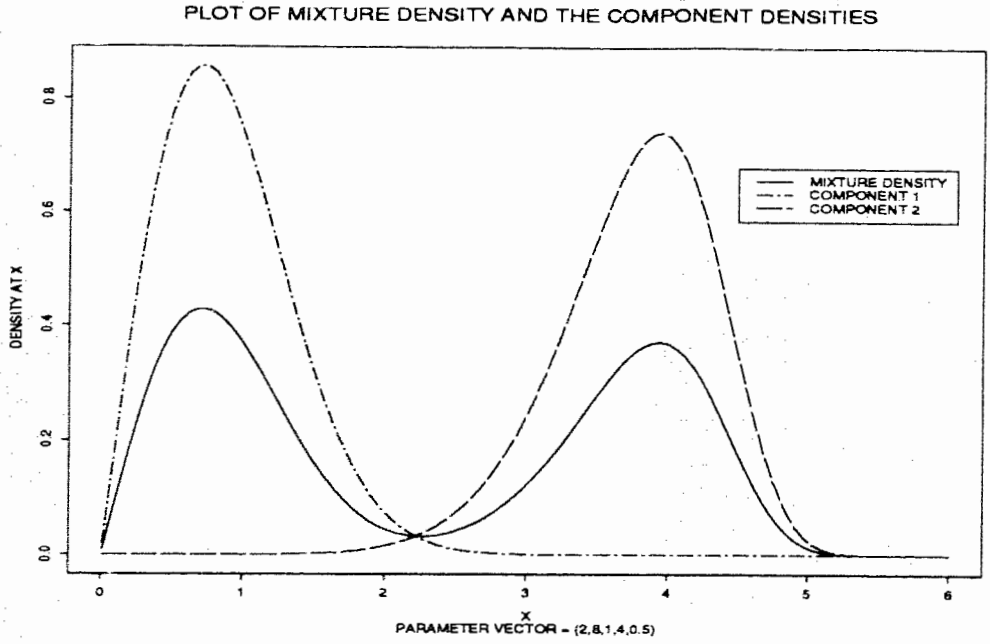
Figure 6.9: Density of population 5 : *(no saddle points were observed)*

HISTOGRAM OF P-VALUES   Q-Q PLOT OF P-VALUES
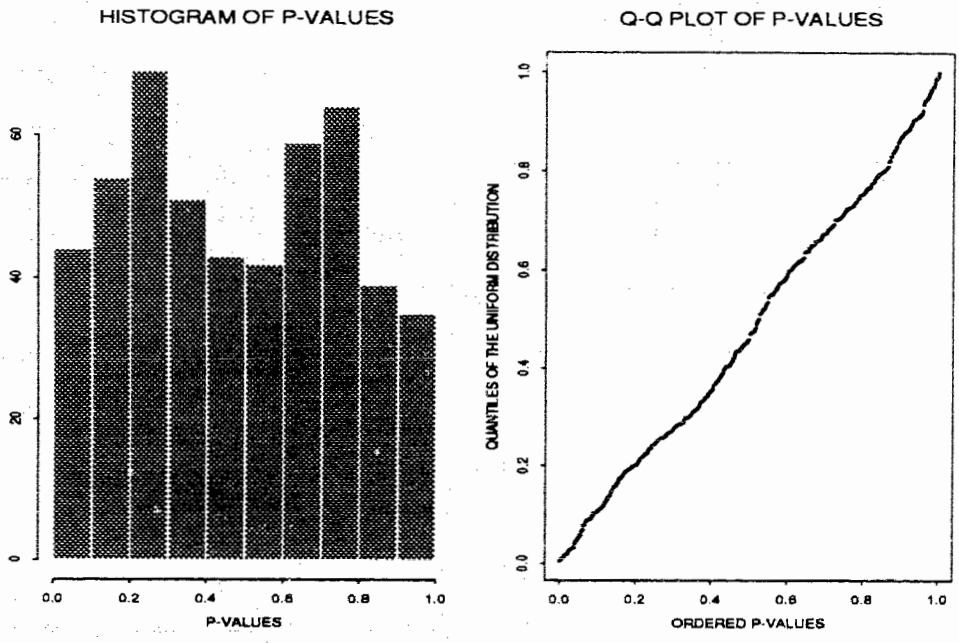
Figure 6.10: Histogram and Q-Q plot of p-values: Population 5

Figures 6.2- 6.10 suggest that p-values observed in each case are uniformly distributed. We used the Anderson-Darling statistic as a measure of uniformity of p-values. The null hypothesis to be tested is, H_0 : the p values are uniformly distributed on $[0,1]$. Thus under H_0 the distribution is completely specified. According to Stephens(1986) it falls under Case 0. Percentage points are given in Table 4.2(page 105) of Stephens(1986). We arrange the p-values in ascending order and we note that the probability integral transforms z_i are exactly the values itself because the distribution of interest is the uniform distribution. The value of the test statistic is computed using the formula for A^2 given in section 3.6. The value of the test statistic and the corresponding p-value of the Anderson-Darling test in each case are given in table 6.2. The p-values for the Anderson-Darling test given in the table are computed by first estimating the eigenvalues of the covariance kernel, for A^2,

$$\rho_0(s,t) = \frac{\min(s,t) - st}{\sqrt{t(1-t)s(1-s)}}$$

using the brute force approach described in section 3.5 and then using these eigenvalues as weights to compute the p-values.

| pop. | parameter | | | | | And. | |
	shape(1)	shape(2)	scale(1)	scale(2)	prop.	test stat.	p-value
1	2	3	3	0.9	0.5	0.78	0.49
2	1.5	3	2	4	0.5	1.28	0.24
3	1	3	2	4	0.5	0.95	0.38
4	2	4	0.5	3	0.5	1.47	0.18
5	2	8	1	4	0.5	2.43	0.05

Table 6.2: Table of Anderson Darling test statistics and p-values

6.5 Conclusions

- According to the above results, the Anderson-Darling test does not reject the null hypothesis that p-values are uniformly distributed in any of the above cases.

- Therefore, it is reasonable to assume that the usual asymptotic theory for the Cramér-von Mises statistic is valid in the Weibull mixture case.

- This justifies the use of the approximate p-value suggested in section 5.4 to test the goodness of fit of the fitted Weibull mixture model.

Chapter 7

Conclusions and summary of open problems

7.1 Concluding remarks

1. The likelihood surfaces for samples of Weibull mixture distributions appear to be flat over a wide range of the parameter space. This gives rise to difficulties in maximum likelihood estimation. For example, the Newton Raphson method fails to converge in such a case and special techniques have to be implemented to find the solution.

2. Likelihood functions for samples of Weibull mixture distributions whose components are not well separated sometimes have more than one maximum; it is hard to find the global maximum with certainty.

3. In some cases several very different roots of the likelihood equations all gave good fits. This problem arose only when the components are not well separated.

4. Surprisingly, the values of the log-likelihood function at each of these roots were similar.

5. P-values for the Cramér-von Mises statistic computed on the basis of our asymptotic distribution have the predicted uniform distribution. Therefore, our approximate asymptotic distribution may be used to compute an approximate p-value for the test of fit of the fitted Weibull mixture model.

7.2 Summary of open problems and suggestions

1. In this study we developed software to compute the p-values of the Cramér-von Mises test statistic applicable to Weibull mixture populations. These can very easily be extended to other mixture distributions as well as to cover other test statistics which are based on the empirical distribution.

2. In this study, we restricted ourselves to ungrouped data. One could extend the above ideas to grouped data as well as to censored data.

Appendix A

Data Sets

A.1 Data set(1)

1.0223828	1.0177587	1.2011735	2.5710518	2.8460915	0.1069041
1.2504893	0.9950643	0.7833413	5.6836319	0.6721168	1.6171286
1.3060745	0.9868945	0.7655612	2.1595537	0.8788154	3.8687286
0.6935191	3.1426349	0.7916427	0.7769707	2.5357167	0.6057939
4.9822883	6.3284718	6.3209155	1.0002028	5.4567053	3.4622927
1.3421827	2.2220684	3.9372875	0.8543468	3.6672644	1.7983155
0.2601858	1.0476711	1.1927935	4.0856903	0.9174970	0.2188122
1.3588268	1.0786055	0.7658786	4.3017336	1.0074364	0.3545173
2.6971382	3.7689321	0.3520612	1.1589643	2.7830480	0.7530154
0.8787339	0.7855622	0.5815194	2.0529836	1.0518789	0.3955225
1.2992692	0.6983452	1.1057061	1.9616712	2.3353226	0.9271150
1.4110422	0.8139170	1.3478895	0.6982980	3.9535968	1.3889530
0.8928673	3.2456759	4.5736033	1.2828072	0.8081162	0.2699175
0.7881533	0.9514808	0.5999366	1.3146535	2.3457089	2.9094415
0.3657555	0.2313114	1.5166873	1.7566231	3.1964081	0.3675468
1.1502145	0.3070789	1.3349492	0.8103855	0.1184989	0.7745366
0.6484203	1.1819143	1.9878050	1.0272995		

A.2   Data set(2)

5.3042269	2.6350808	1.2531883	3.0837725	3.9388123	0.1950201	1.0304685
1.3744614	3.1815707	3.3786843	1.6643244	7.0453781	0.5569809	0.9384950
5.1281041	0.6424147	0.8965324	3.7326078	0.5616477	2.6328054	0.9321107
4.6326301	4.0407108	4.4729154	2.9519342	2.0761575	2.0717230	0.8072693
2.5395409	2.9170521	5.3826999	2.7898499	1.6345463	1.5631035	1.6419326
3.0749187	1.5481462	5.3976012	3.0693061	0.2379801	2.9776892	3.1170207
2.9082141	0.4268044	3.5837500	0.3376380	2.6252562	1.2038543	1.8331696
2.7439493	2.1922736	1.9830151	4.3473678	4.7277217	0.4119260	0.7855037
1.2149359	3.3253510	3.4293134	2.2921864	2.6141537	3.6185271	0.8654889
2.5747383	2.7976565	3.6060248	3.7215747	3.0944627	2.5974006	2.5554961
1.1919997	1.7530734	3.3257847	3.5877874	2.6251822	2.5897634	1.0296417
2.2877457	1.3918966	4.0280519	0.7155959	5.9203678	1.9450702	3.5671604
2.9074556	1.1480348	1.5701826	2.8480250	0.8028973	2.2337762	0.4703399
4.6829135	1.1444424	1.4796428	4.0830916	1.7523359	2.5985532	2.7180774
3.3873932	4.4156611					

Appendix B

B.1 Inverse of the Weibull mixture distribution function

The function which we call "inv.Weib" computes the inverse of a vector of points using less computaional time and effort compared to finding the inverse at each point separately.

Description:

If $c1$ is the vector which contains the inverses of the points of the vector $t1$ then $F(c1, \underline{\alpha}) = t1$. So $c1$ is a root of the equation $g(x) = F(c1, \underline{\alpha}) - t1$. We use the Secant Method to compute the root at each point of the vector $c1$.

The input variables for the function are;

1. u = parameter vector in the order (shape1,shape2,scale1,scale2,proportion).

2. $t1$ = vector of points that are to be inverted

3. a, b = two initial points which are on the opposite side of the root.

The output variable $c1$ contains the inverse points of the vector $t1$.

A more detailed description of this function is offered in section 5.2.

```
inv.Weib_

function(u,t1,a,b){

        f_function(u,t1,y){

          u[5] * (1 - exp(-((y/u[3])^u[1])))

                + (1 - u[5]) * (1 -exp(-((y/u[4])^u[2]))) - t1}

        fa_f(u,t1,a)

        fb_f(u,t1,b)

        c1_(b*fa-a*fb)/(fa-fb)

        fc1_f(u,t1,c1)

        recur_(abs(fc1)>0.5e-05)

        if(any(recur)) {

        c2_c1[recur]

        fc2_f(u,t1[recur],c2)

        b[recur][fa[recur]*fc2<0]_c2[fa[recur]*fc2<0]

        a[recur][fb[recur]*fc2<0]_c2[fb[recur]*fc2<0]

        c1[recur]_inv.Weib(u,t1[recur],a[recur],b[recur])}

        c1}
```

Bibliography

[1] Chandra, M., Singpurwalla, N.D. , and Stephens, M. A. (1981). Kolmogorov statistics for tests of fit for the extreme-value and Weibull distributions. *J. Amer. Statist. Assoc.* 76, 729-731.

[2] Chen, G.(1991). *Empirical processes based on regression residuals: Theory and applications.* Ph. D. Thesis, Simon Fraser University.

[3] Cramér, H.(1964). *Mathematical methods of statistics.* Princeton University Press.

[4] Durbin, J.(1973). Weak convergence of the sample distribution function when parameters are estimated. *The Annals of Statistics,* Vol.1, No.2, 279-290.

[5] Durbin, J. and Knott, M. (1972). Components of Cramér-von Mises Statistics I. *J. Roy. Statist. Soc.* B, 34, 290-307.

[6] Durbin, J., Knott, M. and Taylor, C. C. (1975). Components of Cramér-von Mises Statistics II. *J. Roy. Statist. Soc.,* B, 37, 216-237.

[7] Falls, L. W.(1970). Estimation of parameters in compound Weibull distributions, *Technometrics,* Vol.12, No. 2, 399-407.

[8] Imhof, J.P.(1961). Computing the distribution of the quadratic forms in normal variables. *Biometrika,* 48, 3 and 4, 419-426.

[9] Johnson, N.L., Kotz, S.(1970),*Continuous univariate distributions*, Vol. 2, New York: Hougton Mifflin.

[10] Kaylan, A.R. and Carl, M. H.(1981). Efficient algorithms to derive maximum likelihood estimates for finite exponential and Weibull mixtures. *Comput. and Ops. Res.*, Vol.8, 97-104.

[11] Lockhart, R.A. and Stephens, M. A. (1989). Tests of fit for the three-parameter Weibull distribution. *Technical Report,* Simon Fraser University.

[12] Macdonald, P. D. M.(1975). Estimation of finite distribution mixtures. *Applied Statistics: Proc. Conf. Dalhousie University, Halifax, May 2-4*, 231-245, North Holland Publishing Company.

[13] McLachlan G.J. and Basford, K.E.(1988). *Mixture models: Inference and applications to clustering.* New York: Marcel Dekker Inc.

[14] Sinha, S. K.(1987). Bayesian estimation of the parameters and reliability function of a mixture of Weibull life distributions. *Journal of Statistical Planning and Inference*, 16, 377-387.

[15] Stephens, M. A. (1976). Asymptotic results for goodness of fit statistics with unknown parameters. *The Annals of Statistics*, Vol.4, No. 2, 357-369.

[16] Stephens, M. A. (1977). Goodness of fit for the extreme value distribution. *Biometrika*, 64, 3, 583-588.

[17] Stephens, M. A.(1986). Chapter 4 in *Goodness of fit techniques.* edited by D'Agostino, R. B. and Stephens, M. A. New York: Marcel Dekker Inc.

[18] Titterington, D.M., Smith A.F.M. and Markov, U. E.(1985). *Statistical analysis of finite mixture distributions.* New York: Wiley.