Maximum Likelihood Estimation for a Heavy-Tailed Mixture Distribution

Philippe Kponbogan Dovoedo
philippe.dovoedo@hotmail.com

Follow this and additional works at: https://huskiecommons.lib.niu.edu/allgraduate-thesesdissertations

Part of the Statistics and Probability Commons

Recommended Citation
https://huskiecommons.lib.niu.edu/allgraduate-thesesdissertations/6983

This Dissertation/Thesis is brought to you for free and open access by the Graduate Research & Artistry at Huskie Commons. It has been accepted for inclusion in Graduate Research Theses & Dissertations by an authorized administrator of Huskie Commons. For more information, please contact jschumacher@niu.edu.
ABSTRACT

MAXIMUM LIKELIHOOD ESTIMATION FOR A HEAVY-TAILED MIXTURE DISTRIBUTION

Philippe Dovoedo, Ph.D.
Department of Mathematical Sciences
Northern Illinois University, 2019
Dr. Alan Polansky, Director

In an increasingly connected global environment, “high-impact, low-probability” (HILP) events can have devastating consequences and result in large insurance losses with a heavy-tailed distribution. Examples of such events include Hurricane Katrina, the Deepwater Horizon oil disaster and the Japanese nuclear crisis and tsunami. According to the 2012 Blackett Review of HILP Risks from the UK Government Office for Science, the identification of low-probability risks, and the subsequent development of mitigation plans, is complicated by their rare or conjectural nature, and their potential for causing impacts beyond everyday experience. Extremal mixture models and more generally extreme value analysis help assess HILP risks. In this dissertation, we introduce various classes of heavy-tailed distributions before moving on to mixture models. In particular, we are interested in the mixture of a heavy-tailed distribution and a light-tailed distribution. Estimation of the mixture distribution is based on the expectation-maximization (EM) algorithm and model selection is achieved using information criteria. Our results indicate that one of the components of our mixture may provide us with a good model for modeling nonnegative, heavy-tailed data.
MAXIMUM LIKELIHOOD ESTIMATION FOR A HEAVY-TAILED MIXTURE DISTRIBUTION

BY

PHILIPPE DOVOEDO
© 2019 Philippe Dovoedo

A DISSERTATION SUBMITTED TO THE GRADUATE SCHOOL IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE DOCTOR OF PHILOSOPHY

DEPARTMENT OF MATHEMATICAL SCIENCES

Dissertation Director:
Dr. Alan Polansky
ACKNOWLEDGEMENTS

First and foremost, I would like to thank the Holy Trinity for His bountiful blessings.

Also, I owe a huge debt of gratitude to my advisor, Dr. Alan Polansky for his patience, encouragement, and faith in me.

Special thanks are due to Dr. Duchwan Ryu for his assistance, to Dr. Barbara Gonzalez for her insight, to Dr. Sanjib Basu for admitting me to NIU. Many thanks also to Dr. Nader Ebrahimi, Dr. Yanbin Yin, Dr. Victor Gensini, and Dr. Anders Linner for their support.

Many thanks as well to all of the following people: Shelley Harold, Carolyn Atkins, Tom Kapraun and Katelyn Kozinski, Rexford Akakpo, Paramahansa Pramanik, Hermiole Atcha, Mr. & Mrs. Simon Fassinou, Mr. & Mrs. Yann Vodounou, Mr. & Mrs. Tim Travis, Mr. & Mrs. Obi Osunkwo, Mrs. Willie Mae Stewart, Mr. & Mrs. Tim Hegberg, Rev. & Mrs. John Oladiran, Mr. & Mrs. Kola Olatunbosu, Rev. Ike & Mrs. Okafor, Mr. & Mrs. Sam Olakpade, Mr. & Mrs. Jones Onigbinde, Mr. & Mrs. Stephen Jamam, Mr. & Mrs. Elias Enianku, Rev. Karen & Mr. Eddie Storey, Mr. & Mrs. Gabriel Barrios, Mr. & Mrs. Ade Owotuyi, Mr. & Mrs. Theo Oji, Mrs. Joan Hughes, Rev. & Mrs. Josh Idowu, Mr. & Mrs. Steve Pace, Mr. & Mrs. Rob Ensor, Mrs. Sylvia Bray, Mr. & Mrs. Bunmi Omitoyin, and Taiwo Omitoyin, Mr. & Mrs. Roland Ajanaku, Dr. Margaret & Mr. Anietie King, Mr. & Mrs. Sam Oyedokun, Abigail and Rebecca Oyedokun, Mr. & Mrs. Folu Olusanya, Mr. & Mrs. Wole Ogunyemi, Mr. & Mrs. Sunday Balogun, and Mr. & Mrs. Shane Wagnaar.

Last, but not least, I must express my deep appreciation for my family, especially for my wife, Debora, my daughter, Victoria, my mother, Patience, my siblings, Herve, Olivier, Liliane, Charles and Josue, as well as my uncle, Honore, my aunt, Leontine, and my cousins, Isabelle and Gad.
DEDICATION

To the memories of my father Joseph Dovoedo and the Reverend Amos Adeoye.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vi</td>
</tr>
<tr>
<td><strong>Chapter</strong></td>
<td></td>
</tr>
<tr>
<td>1 PRELIMINARIES</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Motivation</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Classes of Heavy-Tailed Distributions</td>
<td>2</td>
</tr>
<tr>
<td>1.3 Hazard Rate</td>
<td>7</td>
</tr>
<tr>
<td>1.4 Extreme Value Theory (EVT)</td>
<td>8</td>
</tr>
<tr>
<td>2 MIXTURE MODELS</td>
<td>14</td>
</tr>
<tr>
<td>2.1 A Family of Pure Mixture Distributions</td>
<td>14</td>
</tr>
<tr>
<td>2.2 Parameter Estimation</td>
<td>45</td>
</tr>
<tr>
<td>2.3 A Simulation Study</td>
<td>47</td>
</tr>
<tr>
<td>2.4 Application to Real Data</td>
<td>52</td>
</tr>
<tr>
<td>2.5 Extremal Mixture Models</td>
<td>56</td>
</tr>
<tr>
<td>2.6 A Family of Dynamically Weighted Mixture Models</td>
<td>58</td>
</tr>
<tr>
<td>2.7 Concluding Remarks and Future Work</td>
<td>60</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>61</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Absolute value of estimated bias (EB), estimated standard error (ESE), and mean square error (MSE) of the MLE for different sample sizes and ((\theta, \rho, \lambda) = (1, 1, 1)).</td>
<td>48</td>
</tr>
<tr>
<td>2.2</td>
<td>Absolute value of estimated bias (EB), estimated standard error (ESE), and mean square error (MSE) of the MLE for different sample sizes and ((\theta, \rho, \lambda) = (2, 1, 1)).</td>
<td>49</td>
</tr>
<tr>
<td>2.3</td>
<td>Absolute value of estimated bias (EB), estimated standard error (ESE), and mean square error (MSE) of the MLE for different sample sizes and ((\theta, \rho, \lambda) = (1, 2, 1)).</td>
<td>50</td>
</tr>
<tr>
<td>2.4</td>
<td>Absolute value of estimated bias (EB), estimated standard error (ESE), and mean square error (MSE) of the MLE for different sample sizes and ((\theta, \rho, \lambda) = (1, 1, 2)).</td>
<td>51</td>
</tr>
<tr>
<td>2.5</td>
<td>Models used in modeling remission times.</td>
<td>53</td>
</tr>
<tr>
<td>2.6</td>
<td>Results of goodness of fit tests with their test statistics and p-values.</td>
<td>55</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Heavy-tailed distribution and light-tailed distribution.</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>Extreme value distributions, Weibull in red, Gumbel is in green, and Frechet in blue.</td>
<td>10</td>
</tr>
<tr>
<td>2.1</td>
<td>Mixture density for the full parameter $(0.25, 2, 1, 1)$ in solid blue, $(0.5, 1, 2, 1)$ in dotted red, and $(0.75, 1, 1, 2)$ in longdash green.</td>
<td>45</td>
</tr>
<tr>
<td>2.2</td>
<td>Histogram, empirical pdf and cdf for remission times data</td>
<td>54</td>
</tr>
<tr>
<td>2.3</td>
<td>QQ plot of remission times data against the exponential distribution</td>
<td>54</td>
</tr>
<tr>
<td>2.4</td>
<td>Sample mean excess plot of remission times data</td>
<td>55</td>
</tr>
</tbody>
</table>
CHAPTER 1
PRELIMINARIES

1.1 Motivation

The understanding of financial asset return distributions is an important issue in risk management. Traditionally, academics and practitioners have used models based on the assumption of normally distributed returns. However, in 1963, Benoit Mandelbrot conjectured that financial returns are better described by a non-normal stable distribution often called a stable Paretian distribution [6, 13, 14, 24, 25 & 26]. In particular, he established that the change in cotton prices was heavy-tailed. Heavy tails characterize distributions that significantly put more probability on larger values [51]. The reference distribution for classifying the tails of a distribution is usually taken to be the exponential distribution. Examples of heavy-tail phenomena include income distributions, record-breaking insurance losses, financial log-returns, foreign exchange rates, file sizes stored on a server, transmission rates of files, service time and input in queuing models, flood levels of rivers, wave heights during a storm, extreme levels of ozone concentration, and high wind-speed values [1, 3 & 4].

The intuition associated with heavy-tailedness includes the sparsity of observations in the tail of the distribution (so differences between successively larger observations increases and the ratio of successive record values does not decrease), the possible nonexistence of some moments, and the slower-than-exponential decay of the tail [4 & 16].

Heavy-tailed distributions arise in areas as diverse as biometry, economics, ecological systems, sociology, finance, business, physics, epidemiology, and geoscience [2 & 4]. Therefore,
it is important to study these distributions. In this dissertation, we are mostly interested in the right tail of the distribution. An important subclass of heavy-tailed distributions, widely used in insurance and telecommunications, is that of subexponential distributions. The notion of subexponentiality was introduced in 1964 by Chistyakov [81 & 83], in connection with its application to branching processes. Subexponential distributions are also important for the modeling of large claims [29]. The best known subclass of subexponential distributions is that of distributions with regularly varying tails.

1.2 Classes of Heavy-Tailed Distributions

An r.v. $X$ (or its df $F$) on $\mathbb{R}$ is said to have right-unbounded support if $F(x) > 0$, for all $x$. An r.v. $X$ (or its df $F$) is right heavy-tailed or has a heavy right tail ($F \in \mathcal{K}$) if the moment generating function of $F$, $M_F(\varepsilon)$, is infinite for all $\varepsilon > 0$, that is, $E[\exp(\varepsilon X)] = \infty$, for all $\varepsilon > 0$. An r.v. $X$ (or its df $F$) is left heavy-tailed (or has a heavy left tail) if the r.v. $-X$ is right heavy-tailed. A distribution that is not heavy-tailed is called light-tailed.

![Figure 1.1: Heavy-tailed distribution and light-tailed distribution.](image)
A nonnegative function $f$ is (right) heavy-tailed if

$$\lim_{x \to \infty} \sup \exp(\delta x) f(x) = \infty,$$

for all $\delta > 0$. An r.v. $X$ (or its d.f $F$) is long-tailed ($F \in \mathcal{L}$) if

$$\lim_{x \to \infty} \frac{F(x + t)}{F(x)} = 1,$$

for all $t > 0$, where we are implicitly assuming that $F(x) > 0$, for all $x$. That is to say $F(x + t) \sim F(x)$, for all $t > 0$. An r.v. $X$ concentrated on $[0, \infty)$ (or its d.f $F$) is subexponential ($F \in \mathcal{S}$) if $F^{*2}(x) \sim 2F(x)$, or equivalently if $P(X_1 + X_2 > x) \sim 2P(X_1 > x)$, where $X_1$ and $X_2$ are two independent copies of $X$. More generally, an r.v. $X$ distributed by $F$ is subexponential if $F_+$ is subexponential, where $F_+ = FI([0, \infty))$.

A measurable and ultimately positive function $L$ is called regularly varying at infinity with index $\alpha \in \mathbb{R}$ if

$$\lim_{x \to \infty} \frac{L(tx)}{L(x)} = t^\alpha,$$

for all $t > 0$. The parameter $\alpha$ is called the index of variation or exponent of variation. If $\alpha = 0$, then $L$ is said to be slowly varying. Roughly speaking, regularly varying functions behave asymptotically like power functions.

An r.v. (or its d.f. $F$) has a regularly varying (right) tail ($F \in \mathcal{R}$) if there exists $\alpha \geq 0$ such that $F$ is regularly varying (at infinity) with index $-\alpha$. That is,

$$\lim_{x \to \infty} \frac{F(tx)}{F(x)} = t^{-\alpha},$$
for all $t > 0$. An r.v. $X$ (or its df $F$) has a dominated varying (right) tail ($F \in \mathcal{D}$) if

$$\limsup_{x \to \infty} \frac{F(tx)}{F(x)} < \infty,$$

for any $0 < t < 1$. Two distributions $F$ and $G$ with right-unbounded supports are said to be tail equivalent if $F(x) \sim \bar{G}(x)$ as $x \to \infty$, that is,

$$\lim_{x \to \infty} \frac{F(x)}{\bar{G}(x)} = 1.$$

Two distributions $F$ and $G$ with right-unbounded supports are said to be weakly tail equivalent if

$$0 < \liminf_{x \to \infty} \frac{F(x)}{\bar{G}(x)} \leq \limsup_{x \to \infty} \frac{F(x)}{\bar{G}(x)} < \infty.$$

The symbols $\mathcal{K}$, $\mathcal{L}$, $\mathcal{S}$, $\mathcal{D}$ and $\mathcal{R}$ will be used to denote respectively the classes of heavy-tailed distributions, long-tailed distributions, subexponential distributions, dominated varying distributions, and regularly varying distributions. Note that the classes $\mathcal{K}$, $\mathcal{L}$, $\mathcal{D}$, $\mathcal{R}$, and $\mathcal{S}$ are all closed under tail equivalence. For $\mathcal{K}$, $\mathcal{L}$, $\mathcal{D}$, and $\mathcal{R}$, the proofs are trivial; for the proof on $\mathcal{S}$, see [7]. Moreover, if a df $F$ concentrated on $[0, \infty)$ is subexponential, then for each natural number $n \geq 2$, we have that $\frac{F^m(x)}{\overline{F}(x)} \sim n \overline{F}(x)$ [7]. In particular, $\overline{F^m}$ is subexponential. The converse also holds, that is, if $\overline{F^m}(x) \sim n \overline{F}(x)$ for some $n \geq 2$, then $F$ is subexponential [2]. Under weak tail equivalence, neither $\mathcal{R}$, $\mathcal{L}$ or $\mathcal{S}$ are closed, while $\mathcal{D}$ and $\mathcal{K}$ are [70 & 71]. Finally, the classes $\mathcal{D} \cap \mathcal{L}$, $\mathcal{L}$, and $\mathcal{R}$ are closed under convolutions, respectively, whereas the class $\mathcal{S}$ is not [2, 54, 65 & 66].

Several results provide characterizations of distribution functions in various situations.

(i) An r.v. $X$ (or its df $F$) is (right) heavy-tailed if and only if the tail function $\overline{F}$ is (right) heavy-tailed, that is,

$$\limsup_{x \to \infty} \exp(\delta x) F(x) = \infty,$$
for all $\delta > 0$. Details can be found in [2].

(ii) An r.v. $X$ (or its df $F$) is long-tailed if and only if

$$\lim_{x \to \infty} \frac{F(x + 1)}{F(x)} = 1.$$ 

This can be proved using the fact that $F$ is monotone decreasing [2].

(iii) An r.v. $X$ (or its df $F$) has a dominated varying (right) tail if and only if

$$\limsup_{x \to \infty} \frac{F(x)}{F(2x)} < \infty \text{ [50].}$$

(iv) Let $F$ be a df on the nonnegative half-line. Then $P[\max(X_1, X_2) > x] \sim 2F(x)$, where $X_1$ and $X_2$ are independent random variables with common df $F$ [7].

(v) Assume that the df $F$ is subexponential on the real line, and that the df $G$ is long-tailed. If $F$ and $G$ are weakly tail equivalent, then $G$ is subexponential on the real line [2]. In light of part (iv), a df $F$ belongs to the class $S$ if and only if the probability of the set $X_1 + X_2 > x$ is asymptotically the same as the probability of its subset $\max(X_1, X_2) > x$. Colloquially, in the subexponential case, the only way in which $X_1 + X_2$ can get large is roughly by $X_1$ or $X_2$ becoming large. This is known as the principle of the single big jump or the catastrophe principle.

There are also results that regulate the relationships between classes of heavy-tailed distributions. Every long-tailed df is heavy-tailed, that is $L \subseteq K$. However, the converse does not hold [2]. Every subexponential df is long-tailed, that is $S \subseteq L$. Nevertheless, a long-tailed distribution need not be subexponential [9, 18 & 20]. Every regularly varying df is subexponential, that is $R \subseteq S$, but the converse is not true [7 & 29]. If $F$ is long-tailed, then $F$ is heavy-tailed, but the converse is false [2]. If $F$ is dominated varying distribution and long-tailed then $F$ is subexponential $(D \cap L \subseteq S)$ [10 & 29]. The inclusion $R \subset D$ is strict.
We also have that $D \not\subseteq S$, and $S \not\subseteq D$ [29]. Finally, we mention a result on heavy-tailed densities. Let the df $F$ be absolutely continuous with pdf $f$. If the df $F$ is heavy-tailed, then the function $f$ is heavy-tailed [2]. In practice, most commonly used heavy-tailed distributions are subexponential [2 & 69]. Examples of distributions with a heavy right tail include the Pareto distribution, which has tail function

$$F(x) = \left(\frac{\kappa}{x + \kappa}\right)^\alpha,$$

for some shape parameter $\alpha > 0$, some scale parameter $\kappa > 0$, and $x \geq 0$. The Pareto distribution is a power-law distribution that has all moments of order less than $\alpha$ finite, whereas moments of order greater than or equal than $\alpha$ are infinite. Among actuaries, it is a popular model for modeling fire claim data [80]. Originally, the Pareto distribution was used to model the distribution of wealth [40]. The smaller the value of $\alpha$, the more concentrated wealth will be in the hands of a small minority. For a particular set of parameters, the Pareto distribution follows the 80/20 rule, also called the Pareto principle. Another example is the log-normal distribution, which has pdf

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\log x - \mu)^2}{2\sigma^2}\right],$$

where $\mu \in \mathbb{R}$ is a location parameter, $\sigma \in \mathbb{R}^+$ is a shape parameter, and $x \in \mathbb{R}^+$. The lognormal distribution is used to model phenomena whose relative growth rate is independent of size. It is also useful for depicting claim amount distributions (e.g., motor insurance claim data) [11 & 80]. All moments of the log-normal distribution are finite. A final example is given by the Weibull distribution whose tail function is

$$F(x) = \exp\left[-\left(\frac{x}{\beta}\right)^\alpha\right],$$
where $\alpha > 0$ is a shape parameter, $\beta > 0$ is a scale parameter, and $x \geq 0$. This distribution originates from reliability theory and is a generalization of the exponential distribution. The Weibull distribution is widely used to fit distributions of strengths of materials and of failure times [82]. This distribution is heavy-tailed if and only if $\alpha < 1$. The moments of the Weibull distribution are all finite.

The normal distribution, exponential distribution, any mixture of exponentials, gamma distribution, and Weibull distribution (with shape parameter $\alpha > 1$) are all light-tailed [12].

### 1.3 Hazard Rate

Given a nonnegative r.v. $X$ with absolutely continuous df $F$ and pdf $f$, the hazard rate is defined as $$q(x) = \lim_{\Delta x \to 0} \frac{P(x < X \leq x + \Delta x | X > x)}{\Delta x} = \frac{f(x)}{1 - F(x)},$$ and the cumulative hazard function as $$Q(x) = \int_0^x q(x)dx = - \log [F(x)].$$

For details, refer to [19 & 43]. In some applications, $X$ corresponds to insurance losses or a lifetime model. The distribution of $X$ in such cases may be studied in terms of these functions because of the ease of interpretation, modeling simplifications, and mathematical convenience. Properties of these functions are also related to tail heaviness. In particular, $F$ is heavy-tailed if and only if $$\liminf_{x \to \infty} \frac{Q(x)}{x} = 0.$$
For details, see [2]. Similarly, $F$ is long-tailed if and only if $q(x) = o(1)$ as $x \to \infty$. A sufficient condition for $F \in \mathcal{D} \cap \mathcal{L}$ is that $xq(x) = O(1)$ as $x \to \infty$ [61 & 62]. If the hazard rate $q$ is regularly varying with index $\alpha \in (-1, 0)$, then $F$ is subexponential [62].

### 1.4 Extreme Value Theory (EVT)

Unlike most traditional statistical theory which deals with the “average” behavior of processes, extreme value analysis deals with the behavior of a process at unusually large (or small) values [72]. EVT was originally motivated by a practical question: How high should a barrier to be built to prevent flooding, for instance, with a probability of 0.9999 [55]? For proper risk management, tail events (high impact, low probability) are vitally important. Since inference on extremes is challenging due to scarcity of data, we will study the behavior of the tail of the distribution using extreme value models. Let $X_1, \ldots, X_n$ be a random sample from a parent population with df $F$. The df of $M_n = \max(X_1, \ldots, X_n)$ is given by

$$F_{M_n}(x) = P[\max(X_1, \ldots, X_n \leq x] = F^n(x).$$

Then, given a real number $x$, the limit of $F_{M_n}(x)$ is such that

$$\lim_{n \to \infty} F_{M_n}(x) = \begin{cases} 1, & \text{if } F(x) = 1; \\ 0, & \text{if } F(x) < 1. \end{cases}$$

Thus the limit distribution is degenerate. To avoid degeneracy, we seek linear transformations $x \mapsto a_n x + b_n$ such that the limit distributions

$$\lim_{n \to \infty} F_{M_n}(a_n x + b_n) = \lim_{n \to \infty} F^n(a_n x + b_n) = H(x),$$
for all $x \in \mathbb{R}$, are not degenerate. The next theorem specifies the possible limit distributions for standardized maxima of i.i.d. r.v.s.

**Theorem 1** Under certain regularity conditions on the df $F$, there exist suitable constants $a_n > 0$ and $b_n \in \mathbb{R}$, such that

$$
\frac{M_n - b_n}{a_n} \xrightarrow{D} Y \text{ as } n \to \infty,
$$

for some non-degenerate r.v. $Y$. Then the d.f. $H$ of $Y$ is one of the following:

1. Fréchet type ($\gamma > 0$): $\Phi_{\gamma}(x) = \exp(-x^{-1/\gamma})$ for $x > 0$;
2. Weibull type ($\gamma < 0$): $\Psi_{\gamma}(x) = \exp[-(-x)^{-1/\gamma}]$ for $x < 0$;
3. Gumbel (or double-exponential) type ($\gamma = 0$): $\Lambda(x) = \exp[-\exp(-x)]$.

This result is known as Extremal Types Theorem (ETT) or Fisher-Tippett-Gnedenko Theorem and its converse is also valid [5, 19, 44 & 55]. Furthermore, the use of the term ‘type’ in the previous theorem refers to the fact that $H$ is given up to scaling and location constants. The three distributions above are called standard extreme value distributions, and the following three statements are equivalent.

(i) The d.f. of $X$ is $\Phi_{\gamma}$.
(ii) The d.f. of $-1/X$ is $\Psi_{\gamma}$.
(iii) The d.f. of $\log X^\alpha$ is $\Lambda$.

Following von Mises and Jenkinson, we can use one parametrization for all three types. We obtain the class of generalized extreme value distributions (GEV) defined as

$$
H_{\gamma}(x) = \exp\left[-(1 + \gamma x)^{-1/\gamma}\right],
$$

where $\gamma$ is real-valued, $1 + \gamma x > 0$, and when $\gamma = 0$, the right-hand side is interpreted as $\exp[-\exp(-x)]$. It is worth noting that all these extreme value dfs are continuous on $\mathbb{R}$. 
Figure 1.2: Extreme value distributions, Weibull in red, Gumbel is in green, and Frechet in blue.
In the context of the ETT, the parameter $\gamma$ is called the extreme value index (EVI), and the parameter $\alpha = 1/\gamma$ is called the tail index. The tail index allows us to draw inference about the tail behaviour of a distribution, and plays a central role in the analysis of heavy-tailed data. The tail index can also be used for estimation of extreme quantiles and probability densities [47 & 48]. When the df of a standardized maxima from an i.i.d. sequence of r.v.s with df $F$ converges to an extreme value df $H_\gamma$ for some $\gamma \in \mathbb{R}$, the df $F$ is said to belong to the maximum domain of attraction of $H_\gamma$ and we write $F \in \text{MDA}(H_\gamma)$. The following result is useful. If $F$ belongs to $\text{MDA}(\Phi_\gamma)$, then we have that $E(X^{\beta}_+ < \infty$ if and only if $\beta < 1/\gamma$ [59]. Moreover, it can be shown that $F \in \text{MDA}(\Phi_\gamma)$ if and only if $F$ is regularly varying with index $-1/\gamma$ ($F \in \mathcal{R}_{-1/\gamma}$) [8, 29 & 36]. Similarly, a necessary and sufficient condition for $F$ to belong to $\text{MDA}(\Psi_\gamma)$ is $x_F < \infty$ and $F(x_F - x^{-1}) \in \mathcal{R}_{-1/\gamma}$. We have that $F \in \text{MDA}(\Lambda)$ if and only if $F(x) = c(x) \exp \left(-\int_0^x 1/a(t)dt\right)$, $x \in (z, x_F)$, where $c(x)$ is a measurable function for which $c(x) \to c > 0$, as $x \uparrow x_F$, $a(x)$ is a positive and differentiable function such that $a'(x) = o(1)$ as $x \uparrow x_F$, and if $x_F < \infty$, then $a(x) = o(1)$ as $x \uparrow x_F$. If $x_F = \infty$, then for any such function $a(x)$, $a(x) = o(x)$ as $x \to \infty$. On the other hand, if $x_F < \infty$, then any such function $a(x)$ satisfies $a(x) = o(x_F - x)$ as $x \uparrow x_F$. A possible choice for the function $a(x)$ is the mean excess function $a(x) = E(X - x|X > x)$. Another characterization of $\text{MDA}(\Lambda)$ is given by $F \in \text{MDA}(\Lambda)$ if and only if $\lim_{x \uparrow x_F} F(x) = \exp(-t)$, $t \in \mathbb{R}$, where $\tilde{a}(x)$ is a suitable positive function. Any such function $\tilde{a}(x)$ necessarily satisfies the relation $\tilde{a}(x) \sim a(x)$ as $x \uparrow x_F$, with $a(x)$ differentiable and $a'(x) = o(1)$ as $x \uparrow x_F$. A possible choice for the function $\tilde{a}(x)$ is the mean excess function [29, 55 & 63]. Regarding the possible limits of the normalized maxima, a Frechet limit (domain) occurs if and only if $F$ has regularly varying tails. Examples include the Cauchy, Pareto, Burr, log-gamma, $t$, and $F$ distributions. A Weibull limit occurs for light-tailed distributions with a finite right end point. Examples include the continuous uniform and beta distributions. The
Gumbel limit is the intermediate case which include moderately heavy-tailed distributions such as the log-normal distribution and Weibull distribution with shape parameter $\alpha < 1$, and light-tailed distributions such as the normal, exponential, gamma, and logistic. For extreme value analysis, one approach is the block maxima (BM) or Gumbel method, for which the observation period is divided into blocks (nonoverlapping periods of equal size) and the maximum observation is taken from each block [59]. Under domains of attraction conditions, the new observations follow a GEV distribution. Although the BM approach seems to waste data, there are practical reasons to use it [74]. Another approach is the peaks-over-threshold (POT) method, where only observations that exceed a high threshold are selected [75]. Peaks-over-threshold (POT) modeling has been used in hydrology, actuarial science, survival analysis and environmental science to predict extreme events [56 & 57]. The distribution of the selected observations is approximately a generalized Pareto distribution. As with the GEV, the shape parameter $\gamma$ determines the tail behavior. If $\gamma > 0$, a Pareto distribution is obtained, while $\gamma < 0$ gives a beta distribution, whereas $\gamma = 0$ leads to an exponential distribution [21 & 79]. The standard generalized Pareto distribution (GPD) is defined as $G_{\gamma}(x) = 1 - (1 + \gamma x)^{-1/\gamma}$, where $\gamma$ is real-valued, $x > 0$, $1 + \gamma x > 0$, and for $\gamma = 0$, the right-hand side is interpreted as $1 - \exp(-x)$. One can define the related location-scale family $G_{\gamma,\mu,\sigma}$ by replacing the argument $x$ above by $(x - \mu)/\sigma$, where $\mu \in \mathbb{R}$ and $\sigma > 0$ are location and scale parameter, respectively. The Pickands-Balkema-de Haan Theorem relates $F_u$, the excess d.f. over a threshold $u$, to a GPD $G_{\gamma,0,\sigma_u}$ through the maximum domain of attraction of a GEV distribution [27 & 28].

**Theorem 2** Let $X$ be a r.v. with df $F$. Then for every $\gamma \in \mathbb{R}$,

$$X \in MDA(H_{\gamma}) \text{ if and only if } \lim_{\mu \to x_F} \sup_{0 < x < x_F - \mu} |F_u(x) - G_{\gamma,\sigma_u}(x)| = 0,$$
where $\sigma > 0$, $F_u(x) = P(X - u \leq x | X > u)$, and the generalized Pareto distribution is defined as

$$G_{\gamma,\sigma_u}(x) = G_{\gamma,0,\sigma_u}(x) = 1 - \left(1 + \frac{\gamma x}{\sigma_u}\right)^{-1/\gamma},$$

where $x > 0$, $1 + \gamma x/\sigma_u > 0$, $\sigma_u$ is a function of $u$ and for $\gamma = 0$, the right-hand side is interpreted as $1 - \exp(-x/\sigma_u)$.

For details, see [4 & 29]. The parameters of the GPD are uniquely determined by those of the corresponding GEV: both have the shape parameter, but the scale parameter of the GPD depends on the location and scale parameter of the GEV. Specifically, we have that $\sigma_u = \sigma + \gamma(u - \mu)$. For $x \leq u$, we may estimate $F$ using a non-parametrically empirical d.f. or a smooth d.f. Specifically, if $x_{i,n} \leq x < x_{i+1,n}$, and $x \leq u$, we may choose the estimate using

$$\hat{F}(x) = \frac{i - 0.5}{n}.$$

The choice of the threshold $u$ is a big hurdle and requires judgment and expertise, as there is a bias-variance trade-off. If the threshold is too low, the asymptotic assumptions of the model may be violated, while a too high threshold gives fewer extreme events and the resulting estimator will have greater variance. Dupuis proposed a threshold selection method based on robust procedures [92]. We also mention approaches based on visual inspection and approaches based on goodness of fit tests [73]. Once the threshold $u$ has been chosen, the parameters $\gamma$ and $\sigma$ can be estimated by the maximum likelihood method [30], the method of probability-weighted moments (PWM) [31], the elementary percentile method (EPM) [32] or Bayesian methods [33].
CHAPTER 2
MIXTURE MODELS

2.1 A Family of Pure Mixture Distributions

The beginnings of finite mixture (FM) models can be traced back to the 19th century with the works of Quetelet, Bertillon, Newcomb and Pearson [86, 87, 88 & 89]. One of the motivations for considering FM models is to envisage a dataset as made of several latent (that is, missing, unobserved) strata or subpopulations. The mixture structure appears because the origin of each observation (that is, the allocation to a specific stratum or subpopulation) is lost [105]. Analyzing the heights of 9002 conscripts, Bertillon observed that the graphical representation of these heights gave two modes which constituted a surprise. Then he contended that this phenomenon was due to the presence of two distinct ethnic groups [90]. FM models have received increasing attention in recent years from both a theoretical as well as practical point of view. They are useful in approximating general distribution functions in a semi-parametric way and accounting for unobserved heterogeneity. In actuarial modeling, FM models have been used to fit loss data. For instance, they arise when the risk class of a policyholder is uncertain and the number of risks classes is discrete [84]. One of the advantages of FM models is that they can be estimated by various techniques such as the method of moments, graphical methods, minimum-distance methods, the maximum likelihood (ML) method and Bayesian approaches [113]. It is worth mentioning that the weights are associated with the missing-data structure whereas the other parameters are related to the observations [105]. In this section, we wish to study the FM of two random variables $U$
and $V$, where $U$ is heavy-tailed and $V$ light-tailed. Assume that $U$ is the maximum of two independent random variables $U_1$ and $U_2$, with $U_2$ heavy-tailed. For the df $F$ of the discrete mixture, we obtain

$$F(x, p, \theta, \rho, \lambda) = pM(x, \theta)H(x, \rho) + (1 - p)L(x, \lambda),$$

where $x \geq 0$, $0 < p < 1$, $H(x, \rho)$ is a heavy-tailed df with parameter vector $\rho \in \mathbb{R}^d$, $L(x, \lambda)$ is a light-tailed df with parameter vector $\lambda \in \mathbb{R}^e$, and $M(x, \theta)$ is a df (that regulates the influence of $L(x, \lambda)$) with parameter vector $\theta \in \mathbb{R}^c$. From now on, we assume that the mixture $F$ is identifiable. We note that an important special case is obtained when the df $M$ is the Heaviside function.

**Proposition 1** $F$ is a heavy-tailed df.

**Proof** We first show that $F$ is a df. Fix $p \in (0, 1)$, $\theta \in \mathbb{R}^c$, $\rho \in \mathbb{R}^d$, and $\lambda \in \mathbb{R}^e$ throughout the proof. We know that $M(x, \theta)$, and $H(x, \rho)$ are dfs, hence the product $M(x, \theta)H(x, \rho)$ is a df. Since $L(x, \lambda)$ is a d.f. and $0 < p < 1$, it stands to reason that $F(x, p, \theta, \rho, \lambda)$ is a df.

Next, we check that $F$ has right unbounded support. Let $x \geq 0$ be given. Since $H$ is heavy-tailed, we have $H(x, \rho) < 1$. We obtain

$$F(x, p, \theta, \rho, \lambda) \leq pH(x, \rho) + (1 - p)L(x, \lambda) \leq pH(x, \rho) + (1 - p) < p + (1 - p) = 1,$$

where we have used the fact that $M(x) \leq 1$, and $L(x) \leq 1$.

Finally, we show that $F$ is heavy-tailed. Observe that

$$F(x, p, \theta, \rho, \lambda) = 1 - pM(x, \theta)H(x, \rho) - (1 - p)L(x, \lambda)$$

$$= 1 - pM(x, \theta)H(x, \rho) - (1 - p)L(x, \lambda)$$

$$= pM(x, \theta) + (1 - p)L(x, \lambda) + pM(x)H(x, \rho)$$
For each \( \delta > 0 \), we have that
\[
\limsup_{x \to \infty} \exp(\delta x) \left[ pM(x, \theta)H(x, \rho) \right] = \infty,
\]
since \( H \) is heavy-tailed, and there exists a point \( x_0 \) such that \( M(x, \theta) > 0 \) for all \( x > x_0 \).

Moreover,
\[
\liminf_{x \to \infty} \exp(\delta x) \left[ pM(x, \theta) + (1 - p)L(x, \lambda) \right] \geq 0,
\]
as a limit inferior of a nonnegative function. Consequently, the df \( F \) is heavy-tailed.

Henceforward, we will write \( F(x) \) instead of \( F(x, p, \theta, \rho, \lambda) \). Similarly for \( M(x, \theta) \), \( H(x, \rho) \), and \( L(x, \lambda) \).

**Proposition 2** Assume that \( M/H \) and \( L/H \) have finite limits \( \beta_1 \) and \( \beta_2 \), respectively, at (positive) infinity and set \( \beta = p\beta_1 + (1 - p)\beta_2 \).

(i) If \( H \) is long-tailed, then so is \( F \). Further, if \( H \) is dominated varying and long-tailed then \( F \) is subexponential.

(ii) Let \( M \), \( H \) and \( L \) be absolutely continuous dfs with pdfs \( m \), \( h \) and \( l \), respectively. If \( \frac{xq_H(x)}{h(x)} = O(1) \) as \( x \to \infty \) and \( m/q_H \) has a finite limit \( \gamma \) at infinity, then \( F \) is subexponential.

(iii) Provided that \( p(1 - \gamma) + \beta \neq 0 \), the previous result holds if the condition that the function \( xq_H(x) \) is bounded for sufficiently large values of \( x \) is replaced by the condition that the function \( q_H \) is regularly varying with index \( \alpha \in (-1, 0) \).

(iv) If \( H \) belongs to the maximal domain of attraction of the Frechet distribution with index \( \alpha > 0 \), then so does \( F \), and hence \( F \) is subexponential.

(v) If \( H \) belongs to the maximal domain of attraction of the Gumbel distribution, then so does \( F \).
Proof Part (i). We check that the df $F$ has asymptotically locally constant tails. Note that

$$
\frac{F(x + 1)}{F(x)} = \frac{p M(x + 1) H(x + 1)}{H(x)} + \frac{p M(x + 1) + (1 - p) L(x + 1)}{H(x)} \cdot \frac{p M(x) + (1 - p) L(x)}{H(x)}
$$


hence

$$
\frac{F(x + 1)}{F(x)} = \frac{p M(x + 1) H(x + 1)}{H(x)} + \left\{ \frac{p M(x + 1) + (1 - p) L(x + 1)}{H(x + 1)} \right\} \left\{ \frac{H(x + 1)}{H(x)} \right\} \cdot \frac{p M(x) + (1 - p) L(x)}{H(x)}
$$

$$
= \left\{ \frac{H(x + 1)}{H(x)} \right\} \left\{ p M(x + 1) + \frac{p M(x + 1) + (1 - p) L(x + 1)}{H(x + 1)} \right\} \left\{ p M(x) + \frac{p M(x) + (1 - p) L(x)}{H(x)} \right\}
$$

and so

$$
\lim_{x \to \infty} \frac{F(x + 1)}{F(x)} = 1,
$$

since

$$
\lim_{x \to \infty} \frac{H(x + 1)}{H(x)} = 1,
$$

$$
\lim_{x \to \infty} M(x) = 1,
$$

and

$$
\lim_{x \to \infty} \frac{p M(x) + (1 - p) L(x)}{H(x)} = \beta.
$$
where $\beta$ is finite. Therefore, $F$ is long-tailed.

Similarly, we have that

$$
\frac{F(x)}{F(2x)} = \frac{pM(x)\overline{H}(x)}{H(2x)} + \frac{p\overline{M}(x) + (1-p)\overline{L}(x)}{H(2x)}
$$

Therefore, given that $F$ is long-tailed, $F$ is subexponential.

Part (ii). Observe that

$$
xq_F(x) = xq_H(x) \left\{ \frac{pM(x) + \frac{pm(x)H(x) + (1-p)L(x)}{h(x)}}{pM(x) + \frac{p\overline{M}(x) + (1-p)\overline{L}(x)}{\overline{H}(x)}} \right\}
$$

and so

$$
xq_F(x) = xq_H(x) \left\{ \frac{pM(x) + \frac{pm(x) + (1-p)L(x)}{h(x)} - \frac{pm(x)}{q_H(x)}}{pM(x) + \frac{p\overline{M}(x) + (1-p)\overline{L}(x)}{\overline{H}(x)}} \right\}.
$$
By assumption,

$$\lim_{x \to \infty} \frac{pM(x) + (1 - p)L(x)}{H(x)} = \beta,$$

where $\beta$ is finite, so that by l'Hopital's Rule,

$$\lim_{x \to \infty} \frac{pm(x) + (1 - p)l(x)}{h(x)} = \beta.$$

Moreover,

$$\lim_{x \to \infty} \frac{m(x)}{qH(x)} = \gamma,$$

where $\gamma$ is finite. Then

$$0 \leq \lim_{x \to \infty} \left\{ \frac{pM(x) + \frac{pm(x) + (1 - p)l(x)}{h(x)} - \frac{pm(x)}{qH(x)}}{pM(x) + \frac{pM(x) + (1 - p)L(x)}{H(x)}} \right\} < \infty.$$

Since

$$\lim_{x \to \infty} xqH(x) < \infty,$$

we get

$$\lim_{x \to \infty} xqF(x) < \infty,$$
showing that the df $F$ is long-tailed and has a dominated varying right tail function. Therefore, $F$ is subexponential.

Part (iii) We have that

$$
\frac{q_F(tx)}{q_F(x)} = \frac{q_H(tx)}{q_H(x)} \left\{ \frac{pM(tx) + \frac{pm(tx) + (1 - p)l(tx)}{h(tx)} - \frac{pm(tx)}{q_H(tx)}}{pM(x) + \frac{pm(x) + (1 - p)l(x)}{h(x)} - \frac{pm(x)}{q_H(x)}} \right\} \times \left\{ \frac{pM(x) + \frac{pM(x) + (1 - p)L(x)}{H(x)}}{pM(tx) + \frac{pM(tx) + (1 - p)L(tx)}{H(tx)}} \right\}.
$$

Furthermore,

$$
\lim_{x \to \infty} \frac{pM(x) + (1 - p)L(x)}{H(x)} = \beta,
$$

$$
\lim_{x \to \infty} \frac{pm(x) + (1 - p)l(x)}{h(x)} = \beta,
$$

and

$$
\lim_{x \to \infty} \frac{m(x)}{q_H(x)} = \gamma,
$$

with $p(1 - \gamma) + \beta \neq 0$. Then

$$
\lim_{x \to \infty} \frac{q_F(tx)}{q_F(x)} = t^\alpha, \text{ for each } t > 0, \text{ where } -1 < \alpha < 0,
$$

since

$$
\lim_{x \to \infty} \frac{q_H(tx)}{q_H(x)} = t^\alpha.
$$
The claim follows.

Part (iv). Note that

\[
\frac{F(tx)}{F(x)} = \left\{ \frac{H(tx)}{H(x)} \right\} \left\{ pM(tx) + \frac{pM(tx) + (1-p)L(tx)}{H(tx)} \right\} \left\{ pM(x) + \frac{pM(x) + (1-p)L(x)}{H(x)} \right\}.
\]

Moreover,

\[
\lim_{x \to \infty} \frac{H(tx)}{H(x)} = t^{-\alpha},
\]

for each \( t > 0 \), where \( \alpha > 0 \), and

\[
\lim_{x \to \infty} \frac{\mu(x) + (1-p)\mu(x)}{H(x)} = \beta,
\]

where \( \beta \) is finite. Then

\[
\lim_{x \to \infty} \frac{F(tx)}{F(x)} = t^{-\alpha},
\]

so that \( F \) belongs to the domain of attraction of the Frechet distribution with index \( \alpha \).

Part (v). By assumption,

\[
\lim_{x \to \infty} \frac{H[x + t\tilde{a}(x)]}{H(x)} = \exp(-t),
\]

for each real \( t \), and some positive function \( \tilde{a} \). We know that

\[
\frac{F[x + t\tilde{a}(x)]}{F(x)} = \left\{ \frac{H[x + t\tilde{a}(x)]}{H(x)} \right\} \left\{ pM[x + t\tilde{a}(x)] + \frac{p\mu[x + t\tilde{a}(x)] + (1-p)\mu[x + t\tilde{a}(x)]}{H(x + t\tilde{a}(x))} \right\} \left\{ pM(x) + \frac{p\mu(x) + (1-p)\mu(x)}{H(x)} \right\}.
\]

Additionally, \( \tilde{a}(x) = o(x) \) as \( x \to \infty \). Given that

\[
\lim_{x \to \infty} \frac{p\mu(x) + (1-p)\mu(x)}{H(x)} = \beta,
\]
where $\beta$ is finite, we obtain

$$\lim_{x \to \infty} \frac{F[x + t\tilde{a}(x)]}{F(x)} = \exp(-t),$$

for all real $t$. This proves that $F$ belongs to the domain of attraction of the Gumbel distribution.

**Proposition 3** Let $M$, $H$ and $L$ be absolutely continuous dfs with pdfs $m$, $h$ and $l$, respectively. If $H$ is subexponential, the function $l/(mH + hM)$ is bounded above, and the function $M/H$ has a finite nonzero limit at infinity, then the mixture df $F$ is subexponential.

**Proof** According to Pitman [68], if the function $l/(mH + hM)$ is bounded, then a sufficient condition for the df $F$ to be subexponential is that the product $MH$ is subexponential. Now, we will prove that $MH$ and $H$ are weakly tail equivalent. Note that

$$\frac{1 - M(x)H(x)}{1 - H(x)} = \frac{1 - M(x)\left[1 - \overline{H}(x)\right]}{\overline{H}(x)} = M(x) + \frac{\overline{M}(x)}{\overline{H}(x)},$$

for each $x \geq 0$. By assumption,

$$0 < \liminf_{x \to \infty} \frac{\overline{M}(x)}{\overline{H}(x)} \leq \limsup_{x \to \infty} \frac{\overline{M}(x)}{\overline{H}(x)} < \infty.$$

We also know that

$$\lim_{x \to \infty} M(x) = 1.$$

Therefore,

$$1 < \liminf_{x \to \infty} \left\{ \frac{1 - M(x)H(x)}{1 - H(x)} \right\} \leq \limsup_{x \to \infty} \left\{ \frac{1 - M(x)H(x)}{1 - H(x)} \right\} < \infty,$$
showing that the d.f.s MH and H are weakly tail equivalent. Moreover, by an extension of Proposition 2, MH is long-tailed. Since H is subexponential, MH is subexponential. This establishes the subexponentiality of the mixture d.f. F.

In mixture problems, the log-likelihood function often does not have an upper bound and therefore a global maximum does not always exist. Redner and Walker provide sufficient conditions for the existence, consistency and asymptotic normality of the maximum likelihood estimator [15 & 17].

**Proposition 4** For the sake of brevity, we denote the elements of the parameter space Ω by Ψ = (Ψ₁, ..., Ψₙ), where d is the dimension of Ω. Assume that X₁, ..., Xₙ is a sequence of i.i.d. r.v.s with common pdf f(x, Ψ) (strictly positive and thrice differentiable in Ψ for all x in the common support S), and that the parameter space Ω is an open subset of Rᵈ, with the parameter Ψ being identifiable. Following [15, 23, 34, 41, 42, 49 & 106], we state the following conditions.

**Condition 1.** There exists a neighborhood U of the true parameter value Φ*, an integrable function g, and a function g* such that

\[ \left\| D_{Ψ} f(x, Ψ) \right\| + \left\| D^{2}_{Ψ} f(x, Ψ) \right\| \leq g(x) \text{, } \left\| D^{2}_{Ψ} s(x, Ψ) \right\| \leq g^{*}(x) \text{ and } E_{Ψ^{*}} g^{*}(X) < \infty, \]

for all x ∈ S and all Ψ ∈ U, where D_{Ψ} denotes the first derivative operator, w.r.t. Ψ, D^{2}_{Ψ} the second derivative operator, w.r.t. Ψ and s(x, Ψ) = [D_{Ψ} log f(x, Ψ)]ᵀ the score vector.

**Condition 2.** The (expected) Fisher information matrix \( I(Ψ) = E_{Ψ} [s(X, Ψ)s(X, Ψ)ᵀ] \) is well defined and positive definite at Ψ*.

For each Ψ ∈ Ω and each r > 0, let \( \overline{B}_r(Ψ) \) be the closed ball of radius r about Ψ in Ω and define \( f(x, Ψ, r) = \sup \{ f(x, Ψ'), Ψ' ∈ \overline{B}_r(Ψ) \} \), and \( f^{*}(x, Ψ, r) = \max \{ 1, f(x, Ψ, r) \} \).

**Condition 3.** For each Ψ ∈ Ω and sufficiently small r > 0, \( E_{Ψ^{*}} [\log f^{*}(X, Ψ, r)] < \infty. \)
Condition 4. $E_{\Psi^*} [\log f(X, \Psi^*)] < \infty$.

If the first two conditions hold, then for each sufficiently small neighborhood of $\Psi^*$ in $\Omega$ and sufficiently large $n$, there exists, with probability 1, a unique solution $\Psi_n$ of the likelihood equations, which locally maximizes the likelihood function. Furthermore, $\sqrt{n} (\Psi_n - \Psi^*)$ is asymptotically normally distributed with mean 0 and covariance matrix $I(\Psi^*)^{-1}$. On the other hand, under conditions 3 and 4, we have that

$$
\lim_{n \to \infty} \sup \left\{ \prod_{i=1}^{n} f(x_i | \Psi) / \prod_{i=1}^{n} f(x_i | \Psi^*), \Psi \in D \right\} = 0 \text{ a.s.,}
$$

for each compact subset $\Omega'$ of $\Omega$ which contains $\Psi^*$ in its interior and each closed subset $D$ of $\Omega'$. Roughly speaking, if $\Omega'$ is a compact subset of $\Omega$ which contains $\Psi^*$ in its interior, then with probability 1, $\Psi_n$ behaves like a maximum likelihood estimate in $\Omega'$ for $n$ large enough.

**Example 1.** Consider the mixture $F(x, p, \theta, \rho, \lambda) = pM(x, \theta)H(x, \rho) + (1 - p)L(x, \lambda)$, where $x \geq 0$, $0 < p < 1$, $\theta > 0$, $\rho > 0$, $\lambda > 0$ and the d.f.s $M(x, \theta)$, $H(x, \rho)$ and $L(x, \lambda)$ are given by $M(x, \theta) = 1 - \exp (-x/\theta)$, $H(x, \rho) = 1 - (x + 1)^{-\rho}$ and $L(x, \lambda) = 1 - \exp (-x/\lambda)$.

Claim 1. The parameter $\Psi = (p, \theta, \rho, \lambda)$ is identifiable.

Claim 2. $F$ is subexponential.

Claim 3. Our mixture satisfies Conditions 1 to 4.

Proof of Claim 1. Assume that two quadruples $\Psi = (p, \theta, \rho, \lambda)$ and $\Psi' = (p', \theta', \rho', \lambda')$ in $\Omega = (0, 1) \times (0, \infty) \times (0, \infty) \times (0, \infty)$ correspond to the same d.f. $F$. Then

$$
p \left[ 1 - \exp \left( -\frac{x}{\theta} \right) \right] [1 - (x + 1)^{-\rho}] + (1 - p) \left[ 1 - \exp \left( -\frac{x}{\lambda} \right) \right] = p' \left[ 1 - \exp \left( -\frac{x}{\theta'} \right) \right] [1 - (x + 1)^{-\rho'}] + (1 - p') \left[ 1 - \exp \left( -\frac{x}{\lambda'} \right) \right] \quad (2.1)
$$
for all $x \geq 0$. Differentiating with respect to $x$ yields

\[
p\left\{ \frac{1}{\theta} \exp \left( -\frac{x}{\theta} \right) \right\} \left[ 1 - (x + 1)^{-\rho} \right] + \left[ 1 - \exp \left( -\frac{x}{\theta} \right) \right] \left[ \rho (x + 1)^{-\rho - 1} \right] + (1 - p) \left[ \frac{1}{\lambda} \exp \left( -\frac{x}{\lambda} \right) \right] = p'\left\{ \frac{1}{\theta'} \exp \left( -\frac{x}{\theta'} \right) \right\} \left[ 1 - (x + 1)^{-\rho'} \right] + \left[ 1 - \exp \left( -\frac{x}{\theta'} \right) \right] \left[ \rho' (x + 1)^{-\rho' - 1} \right] + (1 - p') \left[ \frac{1}{\lambda'} \exp \left( -\frac{x}{\lambda'} \right) \right] \quad (2.2)
\]

for all $x > 0$. In Equation (2.2), taking the (right-hand) limit of both sides as $x \to 0$, we obtain $(1 - p)/\lambda = (1 - p')/\lambda'$.

Equation (2.2) becomes

\[
p\left\{ \frac{1}{\theta} \exp \left( -\frac{x}{\theta} \right) \right\} \left[ 1 - (x + 1)^{-\rho} \right] + \left[ 1 - \exp \left( -\frac{x}{\theta} \right) \right] \left[ \rho (x + 1)^{-\rho - 1} \right] - p'\left\{ \frac{1}{\theta'} \exp \left( -\frac{x}{\theta'} \right) \right\} \left[ 1 - (x + 1)^{-\rho'} \right] \left[ 1 - \exp \left( -\frac{x}{\theta'} \right) \right] \left[ \rho' (x + 1)^{-\rho' - 1} \right] = \frac{1 - p}{\lambda} \left[ \exp \left( -\frac{x}{\lambda} \right) - \exp \left( -\frac{x}{\lambda'} \right) \right],
\]

for all $x > 0$.

Assume that $\rho > \rho'$. Multiplying each side by $(x + 1)^{\rho+1}$ yields

\[
p\left\{ \frac{1}{\theta} \exp \left( -\frac{x}{\theta} \right) \right\} \left[ 1 - (x + 1)^{-\rho} \right] (x + 1)^{\rho+1} + \rho \left[ 1 - \exp \left( -\frac{x}{\theta} \right) \right] (x + 1)^{\rho+1} - p'\left\{ \frac{1}{\theta'} \exp \left( -\frac{x}{\theta'} \right) \right\} \left[ 1 - (x + 1)^{-\rho'} \right] (x + 1)^{\rho+1} + \rho' \left[ 1 - \exp \left( -\frac{x}{\theta'} \right) \right] (x + 1)^{\rho'-\rho'} = \frac{1 - p}{\lambda} \left[ \exp \left( -\frac{x}{\lambda} \right) - \exp \left( -\frac{x}{\lambda'} \right) \right] (x + 1)^{\rho+1},
\]
for all $x > 0$. As $x \to \infty$, the left-hand side tends to $-\infty$, whereas the right-hand side tends to 0. This is a contradiction. Observe that we made use of the fact that for each $k \in \mathbb{R}$, $u^k \exp(-u)$ tends to zero as $u$ tends to $\infty$. Similarly, the assumption $\rho < \rho'$ leads to a contradiction. Thus $\rho = \rho'$. We rewrite Equation (2.2) as

$$
\left[ \frac{p}{\theta} \exp \left( -\frac{x}{\theta} \right) - \frac{p'}{\theta'} \exp \left( -\frac{x}{\theta'} \right) \right] \left[ 1 - (x + 1)^{-\rho} \right]
+ \rho \left\{ p \left[ 1 - \exp \left( -\frac{x}{\theta} \right) \right] - p' \left[ 1 - \exp \left( -\frac{x}{\theta'} \right) \right] \right\} (x + 1)^{-\rho - 1}
= \frac{1 - p}{\lambda} \left[ \exp \left( -\frac{x}{\lambda} \right) - \exp \left( -\frac{x}{\lambda'} \right) \right],
$$

for all $x > 0$. Multiplying by $(1/\rho) [1 - (x + 1)^{-\rho}]^{-1} (x + 1)^{\rho+1}$ yields

$$
\frac{1}{\rho} \left[ \frac{p}{\theta} \exp \left( -\frac{x}{\theta} \right) - \frac{p'}{\theta'} \exp \left( -\frac{x}{\theta'} \right) \right] (x + 1)^{\rho+1}
+ \frac{\rho \left\{ p \left[ 1 - \exp \left( -\frac{x}{\theta} \right) \right] - p' \left[ 1 - \exp \left( -\frac{x}{\theta'} \right) \right] \right\}}{1 - (x + 1)^{-\rho}}
= \frac{(1 - p)}{\rho \lambda} \left[ \exp \left( -\frac{x}{\lambda} \right) - \exp \left( -\frac{x}{\lambda'} \right) \right] (x + 1)^{\rho+1},
$$

for all $x > 0$. As $x \to \infty$, the left-hand side tends to $p - p'$, while the right-hand side tends to 0. Thus $p = p'$. Since $(1 - p)/\lambda = (1 - p')/\lambda'$, it follows that $\lambda = \lambda'$. Using Equation (2.1), we obtain $\theta = \theta'$.

Proof of Claim 2. We have that

$$
\lim_{x \to \infty} \frac{M(x)}{H(x)} = \lim_{x \to \infty} (x + 1)^{p} \exp \left( -\frac{x}{\theta} \right) = 0.
$$

Similarly,

$$
\lim_{x \to \infty} \frac{L(x)}{H(x)} = 0.
$$
Clearly, $H$ and $L$ have right unbounded support $[0, \infty)$. Observe that

$$\lim_{x \to \infty} \frac{H(x + 1)}{H(x)} = \lim_{x \to \infty} \frac{(x + 1)^\rho}{(x + 2)^\rho} = 1,$$

and that for $0 < \lambda < 1$

$$\lim_{x \to \infty} \exp(\lambda x) L(x) = \exp((\lambda - 1)x) = 0.$$

Thus $H$ is long-tailed and $L$ is light-tailed. Moreover,

$$\lim_{x \to \infty} \frac{H(x)}{H(2x)} = \lim_{x \to \infty} \frac{(2x + 1)^\rho}{(x + 1)^\rho} = 2^\rho.$$

By part (i) of Proposition 2, $F$ is subexponential.

Proof of Claim 3. We check that Conditions 1-4 hold.

Step 1. For each $x \geq 0$ and each $\Psi = (p, \theta, \rho, \lambda) \in \Omega$, we have that

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

and so

$$\frac{\partial}{\partial p} f(x, p, \theta, \rho, \lambda) = \frac{1}{\theta} \exp \left( -\frac{x}{\theta} \right) \left[ 1 - (x + 1)^{-\rho} \right] + \rho(x + 1)^{-\rho - 1} \left[ 1 - \exp \left( -\frac{x}{\theta} \right) \right] - \frac{1}{\lambda} \exp \left( -\frac{x}{\lambda} \right).$$
Hence

\[ \left| \frac{\partial}{\partial p} f(x, p, \theta, \rho, \lambda) \right| \leq \frac{1}{\theta} \exp \left( -\frac{x}{\theta} \right) + \rho(x + 1)^{-\rho - 1} + \frac{1}{\lambda} \exp \left( -\frac{x}{\lambda} \right), \]

which is an integrable function on \([0, \infty)\).

Similarly,

\[ \frac{\partial}{\partial \theta} f(x, p, \theta, \rho, \lambda) = -\frac{p}{\theta^2} \exp \left( -\frac{x}{\theta} \right) \left\{ [1 - (x + 1)^{-\rho}] \left( 1 - \frac{x}{\theta} \right) + \rho x (x + 1)^{-\rho - 1} \right\}, \]

and so

\[ \left| \frac{\partial}{\partial \theta} f(x, p, \theta, \rho, \lambda) \right| \leq \frac{1}{\theta^2} \left[ 1 + \left( \frac{1}{\theta} + \rho \right) x \right] \exp \left( -\frac{x}{\theta} \right). \]

Next,

\[ \frac{\partial}{\partial \rho} f(x, p, \theta, \rho, \lambda) = p \left\{ [1 - \exp \left( -\frac{x}{\theta} \right)] [1 - \rho \log(x + 1)] 
\quad + \frac{(x + 1) \log(x + 1)}{\theta} \exp \left( -\frac{x}{\theta} \right) \right\} (x + 1)^{-\rho - 1}, \]

thus

\[ \left| \frac{\partial}{\partial \rho} f(x, p, \theta, \rho, \lambda) \right| \leq [1 + \rho \log(x + 1)] (x + 1)^{-\rho - 1} + \frac{x}{\theta} \exp \left( -\frac{x}{\theta} \right). \]

Finally,

\[ \frac{\partial}{\partial \lambda} f(x, p, \theta, \rho, \lambda) = -\frac{1 - p}{\lambda^2} \left( 1 - \frac{x}{\lambda} \right) \exp \left( -\frac{x}{\lambda} \right), \]
hence
\[ \left| \frac{\partial}{\partial \lambda} f(x, p, \theta, \rho, \lambda) \right| \leq \frac{1}{\lambda^2} \left( 1 + \frac{x}{\lambda} \right) \exp \left( -\frac{x}{\lambda} \right). \]

Thus for \( \Psi = (p, \theta, \rho, \lambda) \) ranging over a sufficiently small neighborhood of \( \Psi^* = (p^*, \theta^*, \rho^*, \lambda^*) \), each of the absolute values of the first order partial derivatives of \( f \) w.r.t. the parameters \( (p, \theta, \rho, \lambda) \) is dominated by a function of the form
\[
g_1(x) = (a_1 + a_2 x) \exp \left( -\frac{x}{a_3} \right) + (b_1 + b_2 x) \exp \left( -\frac{x}{b_3} \right) + [c_1 + c_2 \log(x + 1)] (x + 1)^{-c_3 - 1},
\]
where \( a_3, b_3 \) and \( c_3 \) are positive. It is easily seen that \( g_1 \) is integrable on \([0, \infty)\).

For each \( x \geq 0 \) and each \( \Psi = (p, \theta, \rho, \lambda) \in \Omega \), we have that
\[
\frac{\partial^2}{\partial p^2} f(x, p, \theta, \rho, \lambda) = 0.
\]

Similarly,
\[
\frac{\partial^2}{\partial \theta^2} f(x, p, \theta, \rho, \lambda) = \frac{p}{\theta^3} \left\{ \left[ 1 - (x + 1)^{-\rho} \right] \left( \frac{x^2}{\theta^2} - \frac{4x}{\theta} + 2 \right) + \rho x \left( 2 - \frac{x}{\theta} \right) (x + 1)^{-\rho - 1} \right\} \exp \left( -\frac{x}{\theta} \right),
\]
and so
\[
\left| \frac{\partial^2}{\partial \theta^2} f(x, p, \theta, \rho, \lambda) \right| \leq \frac{1}{\theta^3} \left[ \left( \frac{x^2}{\theta^2} + \frac{4x}{\theta} + 2 \right) + \rho x \left( 2 + \frac{x}{\theta} \right) \right] \exp \left( -\frac{x}{\theta} \right).
Also,

\[
\frac{\partial^2}{\partial \rho^2} f(x, p, \theta, \rho, \lambda) = -p \left\{ [2 - \rho \log(x + 1)] \left[ 1 - \exp \left( -\frac{x}{\rho} \right) \right] \\
+ \frac{(x + 1) \log(x + 1)}{\theta} \exp \left( -\frac{x}{\theta} \right) \right\} (x + 1)^{-\rho-1} \log(x + 1),
\]

thus

\[
\left| \frac{\partial^2}{\partial \rho^2} f(x, p, \theta, \rho, \lambda) \right| \leq [2 + \rho \log(x + 1)] (x + 1)^{-\rho-1} \log(x + 1) + \frac{x^2}{\theta} \exp \left( -\frac{x}{\theta} \right).
\]

Next,

\[
\frac{\partial^2}{\partial \lambda^2} f(x, p, \theta, \rho, \lambda) = \frac{1}{\lambda^3} \left( \frac{x^2}{\lambda^2} - \frac{4x}{\lambda} + 2 \right) \exp \left( -\frac{x}{\lambda} \right),
\]

and so

\[
\left| \frac{\partial^2}{\partial \lambda^2} f(x, p, \theta, \rho, \lambda) \right| \leq \frac{1}{\lambda^3} \left( \frac{x^2}{\lambda^2} + \frac{4x}{\lambda} + 2 \right) \exp \left( -\frac{x}{\lambda} \right).
\]

In addition,

\[
\frac{\partial^2}{\partial \rho \partial \theta} f(x, p, \theta, \rho, \lambda) = -\frac{1}{\theta^2} \left\{ [1 - (x + 1)^{-\rho}] \left( 1 - \frac{x}{\rho} \right) + \rho x (x + 1)^{-\rho-1} \right\} \exp \left( -\frac{x}{\theta} \right),
\]

hence

\[
\left| \frac{\partial^2}{\partial \rho \partial \theta} f(x, p, \theta, \rho, \lambda) \right| \leq \frac{1}{\theta^2} \left( 1 + \frac{x}{\theta} + \rho x \right) \exp \left( -\frac{x}{\theta} \right).
Similarly,

\[
\frac{\partial^2}{\partial p \partial \rho} f(x, p, \theta, \rho, \lambda) = \left\{ \left[ 1 - \exp \left( -\frac{x}{\theta} \right) \right] \left[ 1 - \rho \log(x + 1) \right] + \left[ \frac{(x + 1) \log(x + 1)}{\theta} \exp \left( -\frac{x}{\theta} \right) \right] \right\} (x + 1)^{-\rho - 1},
\]

then

\[
\left| \frac{\partial^2}{\partial p \partial \theta} f(x, p, \theta, \rho, \lambda) \right| \leq \left[ 1 - \rho \log(x + 1) \right] (x + 1)^{-\rho - 1} + \frac{x}{\theta} \exp \left( -\frac{x}{\theta} \right).
\]

Also,

\[
\frac{\partial^2}{\partial p \partial \lambda} f(x, p, \theta, \rho, \lambda) = -\frac{1}{\lambda^2} \left( 1 - \frac{x}{\lambda} \right) \exp \left( -\frac{x}{\lambda} \right),
\]

hence

\[
\left| \frac{\partial^2}{\partial p \partial \lambda} f(x, p, \theta, \rho, \lambda) \right| \leq \frac{1}{\lambda^2} \left( 1 + \frac{x}{\lambda} \right) \exp \left( -\frac{x}{\lambda} \right).
\]

Likewise,

\[
\frac{\partial^2}{\partial \theta \partial \rho} f(x, p, \theta, \rho, \lambda) = -\frac{p}{\theta^2} \left\{ x \left[ 1 - \rho \log(x + 1) \right] + (x + 1) \left( 1 - \frac{x}{\theta} \right) \log(x + 1) \right\} 
\times (x + 1)^{-\rho - 1} \exp \left( -\frac{x}{\theta} \right),
\]

thus

\[
\left| \frac{\partial^2}{\partial \theta \partial \rho} f(x, p, \theta, \rho, \lambda) \right| \leq \frac{x}{\theta^2} \left[ 2 + \left( \rho + \frac{1}{\theta} \right) x \right] \exp \left( -\frac{x}{\theta} \right).
\]
Next,

$$\frac{\partial^2}{\partial \theta \partial \lambda} f(x, p, \theta, \rho, \lambda) = 0,$$

and finally,

$$\frac{\partial^2}{\partial \rho \partial \lambda} f(x, p, \theta, \rho, \lambda) = 0.$$

Thus for $\Psi = (p, \theta, \rho, \lambda)$ ranging over a sufficiently small neighborhood of $\Psi^* = (p^*, \theta^*, \rho^*, \lambda^*)$, each of the absolute values of the second order partial derivatives of $f$ w.r.t. the parameters $(p, \theta, \rho, \lambda)$ is dominated by a function of the form

$$g_2(x) = (a'_1 + a'_2 x + a'_3 x^2) \exp \left( -\frac{x}{a'_4} \right) + (b'_1 + b'_2 x + b'_3 x^2) \exp \left( -\frac{x}{b'_4} \right) + \left[ c'_1 + c'_2 \log(x + 1) + c'_3 \log^2(x + 1) \right] (x + 1)^{-c'_3 - 1},$$

where $a_3, b_3$ and $c_3$ are positive. Clearly, $g_2$ is integrable on $[0, \infty)$. Now, we compute the third order partial derivatives of $f$ w.r.t. $(p, \theta, \rho, \lambda)$ and then bound each of the third order partial derivatives of $\log f$. For each $x \geq 0$ and each $\Psi = (p, \theta, \rho, \lambda) \in \Omega$, the third order partial derivatives of $f$ w.r.t. the parameters $(p, \theta, \rho, \lambda)$ are

$$\frac{\partial^3}{\partial p^3} f(x, p, \theta, \rho, \lambda) = 0,$$

$$\frac{\partial^3}{\partial \theta^2} f(x, p, \theta, \rho, \lambda) = \frac{p}{\theta^4} \left\{ \left[ 1 - (x + 1)^{-\rho} \right] \left( \frac{x^3}{\theta^3} - \frac{9 x^2}{\theta^2} + \frac{18 x}{\theta} - 6 \right) \right.$$

$$\left. - \rho (x + 1)^{-\rho - 1} \left( \frac{x^3}{\theta^2} + \frac{6 x^2}{\theta} - 6 x \right) \right\} \exp \left( -\frac{x}{\theta} \right),$$
\[
\frac{\partial^3}{\partial \rho^3} f(x, p, \theta, \rho, \lambda) = p \left\{ \left[ 1 - \exp \left( -\frac{x}{\theta} \right) \right] [\rho \log(x + 1) - 3] \log(x + 1) + \frac{1}{\theta} \exp \left( -\frac{x}{\theta} \right) \left[ 2 - (x + 1) \log^2(x + 1) \right] \right\} (x + 1)^{-\rho - 1} \log(x + 1),
\]

\[
\frac{\partial^3}{\partial \lambda^3} f(x, p, \theta, \rho, \lambda) = \frac{1 - p}{\lambda^4} \left( \frac{x^3}{\lambda^3} - \frac{9x^2}{\lambda^2} + \frac{18x}{\lambda} - 6 \right) \exp \left( -\frac{x}{\lambda} \right),
\]

\[
\frac{\partial^3}{\partial p^3} f(x, p, \theta, \rho, \lambda) = 0,
\]

\[
\frac{\partial^3}{\partial p^2 \partial \theta} f(x, p, \theta, \rho, \lambda) = 0,
\]

\[
\frac{\partial^3}{\partial p^2 \partial \rho} f(x, p, \theta, \rho, \lambda) = 0,
\]

\[
\frac{\partial^3}{\partial \theta^2 \partial \rho} f(x, p, \theta, \rho, \lambda) = \frac{1}{\theta^3} \left\{ \left[ 1 - (x + 1)^{-\rho} \right] \left( \frac{x^2}{\theta^2} - \frac{4x}{\theta} + 2 \right) + \rho (x + 1)^{-\rho - 1} \left( 2x - \frac{x^2}{\theta} \right) \right\} \exp \left( -\frac{x}{\theta} \right),
\]

\[
\frac{\partial^3}{\partial \theta^2} f(x, p, \theta, \rho, \lambda) = \frac{p}{\theta^3} \left\{ x [1 - \rho \log(x + 1)] \left( 2 - \frac{x}{\theta} \right) + (x + 1) \left( \frac{x^2}{\theta^2} - \frac{4x}{\theta} + 2 \right) \log(x + 1) \right\} (x + 1)^{-\rho - 1} \exp \left( -\frac{x}{\theta} \right),
\]

\[
\frac{\partial^3}{\partial \theta^2 \partial \lambda} f(x, p, \theta, \rho, \lambda) = 0,
\]
\[
\frac{\partial^3}{\partial \rho^2 \partial p} f(x, p, \theta, \rho, \lambda) = \left\{ \rho \log^2 (x+1) - 2 \log (x+1) \right\} \left[ 1 - \exp \left( -\frac{x}{\theta} \right) \right] \\
- \frac{x+1}{\theta} \log^2 (x+1) \exp \left( -\frac{x}{\theta} \right) \right\} (x+1)^{-\rho-1},
\]

\[
\frac{\partial^3}{\partial \rho^2 \partial \theta} f(x, p, \theta, \rho, \lambda) = \frac{p}{\theta^2} \left\{ -x \left[ \rho \log^2 (x+1) - 2 \log (x+1) \right] \\
+ (x+1) \left( 1 - \frac{x}{\theta} \right) \log^2 (x+1) \right\} (x+1)^{-\rho-1} \exp \left( -\frac{x}{\theta} \right),
\]

\[
\frac{\partial^3}{\partial \rho^2 \partial \lambda} f(x, p, \theta, \rho, \lambda) = 0,
\]

\[
\frac{\partial^3}{\partial \lambda^2 \partial p} f(x, p, \theta, \rho, \lambda) = 0,
\]

\[
\frac{\partial^3}{\partial \rho \partial \theta \partial \rho} f(x, p, \theta, \rho, \lambda) = -\frac{1}{\theta^2} \left\{ x \left[ 1 - \rho \log (x+1) \right] \\
- (x+1) \left( 1 - \frac{x}{\theta} \right) \log (x+1) \right\} (x+1)^{-\rho-1} \exp \left( -\frac{x}{\theta} \right),
\]

\[
\frac{\partial^3}{\partial \rho \partial \theta \partial \lambda} f(x, p, \theta, \rho, \lambda) = 0,
\]

\[
\frac{\partial^3}{\partial \rho \partial \rho \partial \lambda} f(x, p, \theta, \rho, \lambda) = 0,
\]
\[
\frac{\partial^3}{\partial \theta \partial \rho \partial \lambda} f(x, p, \theta, \rho, \lambda) = 0.
\]

By the chain rule,
\[
\frac{\partial^3 \log f}{\partial \Psi_k \partial \Psi_j \partial \Psi_i} = \frac{1}{f} \frac{\partial^3 f}{\partial \Psi_k \partial \Psi_j \partial \Psi_i} + \frac{2}{f^3} \frac{\partial^2 f}{\partial \Psi_k \partial \Psi_j} \frac{\partial f}{\partial \Psi_i} + \frac{2}{f^3} \frac{\partial^2 f}{\partial \Psi_j \partial \Psi_i} \frac{\partial f}{\partial \Psi_k} + \frac{\partial^2 f}{\partial \Psi_k \partial \Psi_i} \frac{\partial f}{\partial \Psi_j} [17].
\]

For each \( x \geq 0 \) and each \( \Psi = (p, \theta, \rho, \lambda) \in \Omega \), we have that
\[
f(x, p, \theta, \rho, \lambda) \geq p \rho \left[ 1 - \exp \left( -\frac{x}{\theta} \right) \right] (x + 1)^{-\rho - 1} \tag{2.3}
\]

Thus for each \( \Psi \) in a sufficiently small neighborhood of \( \Psi^* \), and each \( 1 \leq i, j, k \leq 4 \), we have that
\[
\frac{1}{f} \left| \frac{\partial^3 f}{\partial \Psi_k \partial \Psi_j \partial \Psi_i} \right| \leq g^*_3,
\]

where \( g^*_3 \) is of the form
\[
g^*_3(x) = \frac{P_3(x) \exp \left( -\frac{x}{a_3} \right)}{1 - \exp \left( -\frac{x}{b_3} \right)} + Q_3 \left[ \log(x + 1) \right],
\]
\( P_3 \) and \( Q_3 \) are polynomials, and \( \bar{a}_3, \bar{b}_3 \) positive numbers. Similarly, for each \( \Psi \) in a sufficiently small neighborhood of \( \Psi^* \), and each \( 1 \leq i \leq 4 \), the following holds.

\[
\left| \frac{1}{f} \left| \frac{\partial f}{\partial \Psi_i} \right| \right| \leq g^*_1,
\]

where

\[
g^*_1(x) = \frac{P_1(x) \exp \left( -\frac{x}{\bar{a}_1} \right)}{1 - \exp \left( -\frac{x}{\bar{b}_1} \right)} + \frac{Q_1 [\log(x + 1)]}{1 - \exp \left( -\frac{x}{\bar{c}_1} \right)} + R_1 [\log(x + 1)],
\]

\( P_1, Q_1 \) and \( R_1 \) are polynomials, and \( \bar{a}_1, \bar{b}_1, \bar{c}_1 \) positive numbers. Finally, for each \( \Psi \) in a sufficiently small neighborhood of \( \Psi^* \), and for each \( 1 \leq i, j \leq 4 \), it holds that

\[
\left| \frac{1}{f} \left| \frac{\partial^2 f}{\partial \Psi_j \partial \Psi_i} \right| \right| \leq g^*_2,
\]

where

\[
g^*_2(x) = \frac{P_2(x) \exp \left( -\frac{x}{\bar{a}_2} \right)}{1 - \exp \left( -\frac{x}{\bar{b}_2} \right)} + \frac{Q_2 [\log(x + 1)]}{1 - \exp \left( -\frac{x}{\bar{c}_2} \right)},
\]

\( P_2, Q_2 \) are polynomials, and \( \bar{a}_2, \bar{b}_2 \) and \( \bar{c}_2 \) positive numbers. Therefore, for each \( 1 \leq i, j, k \leq 4 \), provided that \( \Psi \) in a sufficiently small neighborhood of \( \Psi^* \), we have that

\[
\left| \frac{\partial^3 \log f}{\partial \Psi_k \partial \Psi_j \partial \Psi_i} \right| \leq g^*.
\]
where $g^*$ is such that $E_{\Psi^*} [g^*(X)] < \infty$. Therefore, Condition C1 is satisfied.

Step 2. We first prove that for all $1 \leq i, j \leq 4$

$$
\int_0^\infty \left| f \frac{\partial \log f}{\partial \Psi_i} \frac{\partial \log f}{\partial \Psi_j} \right| < \infty,
$$

or, equivalently,

$$
\int_0^\infty \frac{1}{f} \left| \frac{\partial f}{\partial \Psi_i} \frac{\partial f}{\partial \Psi_j} \right| < \infty.
$$

For each $x \geq 0$ and each $\Psi = (p, \theta, \rho, \lambda) \in \Omega$, the convexity of the square function entails

$$
\left[ \frac{\partial}{\partial p} f(x, p, \theta, \rho, \lambda) \right]^2 \leq 3 \left[ \frac{1}{\theta^2} \exp \left( -\frac{2x}{\theta} \right) + \rho^2 (x + 1)^{-2\rho - 2} + \frac{1}{\lambda^2} \exp \left( -\frac{2x}{\lambda} \right) \right].
$$

Combining this with Equation (2.3), we see that

$$
\frac{1}{f(x, p, \theta, \rho, \lambda)} \left[ \frac{\partial}{\partial p} f(x, p, \theta, \rho, \lambda) \right]^2 \leq 3 \left\{ \frac{\exp \left( -\frac{x}{\theta} \right)}{p \theta [1 - (x + 1)^{-\rho}] } + \frac{\rho (x + 1)^{-\rho - 1}}{p \left[ 1 - \exp \left( -\frac{x}{\theta} \right) \right]} + \frac{\exp \left( -\frac{x}{\lambda} \right)}{(1 - p) \lambda} \right\},
$$

hence

$$
\int_0^\infty \frac{1}{f} \left| \frac{\partial f}{\partial p} \right|^2 < \infty.
$$

Similarly, for each $x \geq 0$ and each $\Psi = (p, \theta, \rho, \lambda) \in \Omega$

$$
\left[ \frac{\partial}{\partial \theta} f(x, p, \theta, \rho, \lambda) \right]^2 \leq 2 \left[ \frac{\rho^2 x^2}{\theta^4} + \frac{1}{\theta^2} \left( 1 + \frac{x}{\theta} \right)^2 \right] \exp \left( -\frac{2x}{\theta} \right).
$$
hence
\[ \frac{1}{f(x,p,\theta,\rho,\lambda)} \left[ \frac{\partial}{\partial \theta} f(x,p,\theta,\rho,\lambda) \right]^2 \leq 2 \left\{ \frac{\rho^2 x^2 + \theta^2 \left( 1 + \frac{x}{\theta} \right)}{p\theta^3 \left[ 1 - (x + 1)^{-\rho} \right]} \right\} \exp \left( -\frac{2x}{\theta} \right), \]

thus
\[ \int_0^\infty \frac{1}{f} \left| \frac{\partial f}{\partial \theta} \right|^2 < \infty. \]

Next, for each \( x \geq 0 \) and each \( \Psi = (p,\theta,\rho,\lambda) \in \Omega \)
\[ \left[ \frac{\partial}{\partial \rho} f(x,p,\theta,\rho,\lambda) \right]^2 \leq 2 \left\{ [1 + \rho \log(x + 1)]^2 (x + 1)^{-2\rho - 2} + \frac{x^2}{\theta^2} \exp \left( -\frac{2x}{\theta} \right) \right\}, \]

hence
\[ \frac{1}{f(x,p,\theta,\rho,\lambda)} \left[ \frac{\partial}{\partial \rho} f(x,p,\theta,\rho,\lambda) \right]^2 \leq 2 \left\{ \frac{1 + \rho \log(x + 1)]^2 (x + 1)^{-\rho - 1}}{p\rho \left[ 1 - \exp \left( -\frac{x}{\theta} \right) \right]} + \frac{x^2 \exp \left( -\frac{x}{\theta} \right)}{p\theta^3 \left[ 1 - (x + 1)^{-\rho} \right]} \right\}, \]

and so
\[ \int_0^\infty \frac{1}{f} \left| \frac{\partial f}{\partial \rho} \right|^2 < \infty. \]

Finally, for each \( x \geq 0 \) and each \( \Psi = (p,\theta,\rho,\lambda) \in \Omega \)
\[ \left[ \frac{\partial}{\partial \lambda} f(x,p,\theta,\rho,\lambda) \right]^2 \leq \frac{1}{\lambda^4} \left( 1 + \frac{x}{\lambda} \right)^2 \exp \left( -\frac{2x}{\lambda} \right), \]

then
\[ \frac{1}{f(x,p,\theta,\rho,\lambda)} \left[ \frac{\partial}{\partial \lambda} f(x,p,\theta,\rho,\lambda) \right]^2 \leq \frac{\left( 1 + \frac{x}{\lambda} \right)^2}{(1 - p)\lambda^3} \exp \left( -\frac{2x}{\lambda} \right), \]
and so
\[ \int_0^\infty \frac{1}{f} \left| \frac{\partial f}{\partial \lambda} \right|^2 \, \, d\lambda < \infty. \]

When $1 \leq i \leq 4$ and $1 \leq j \leq 4$ are distinct, we deal with
\[ \int_0^\infty \frac{1}{f} \left| \frac{\partial f}{\partial \Psi_i} \frac{\partial f}{\partial \Psi_j} \right| \, \, d\lambda, \]
by using Holder’s Inequality.
\[ \int_0^\infty \frac{1}{f} \left| \frac{\partial f}{\partial \Psi_i} \frac{\partial f}{\partial \Psi_j} \right| \leq \sqrt{\int_0^\infty \frac{1}{f} \left| \frac{\partial f}{\partial \Psi_i} \right|^2 \, \, d\lambda} \sqrt{\int_0^\infty \frac{1}{f} \left| \frac{\partial f}{\partial \Psi_j} \right|^2 \, \, d\lambda}. \]

Therefore, the Fisher information matrix (FIM) is well-defined. Now, let us check that the FIM is positive definite. It suffices to prove that the first order partial derivatives of log $f$ are linearly independent as a function of $x$ in the support $[0, \infty)$ of the d.f. $F$. Assume that
\[ \alpha \frac{\partial \log f}{\partial p} + \beta \frac{\partial \log f}{\partial \theta} + \gamma \frac{\partial \log f}{\partial \rho} + \delta \frac{\partial \log f}{\partial \lambda} = 0, \]
for some real numbers $\alpha, \beta, \gamma$ and $\delta$. Hence
\[ \alpha \frac{\partial f}{\partial p} + \beta \frac{\partial f}{\partial \theta} + \gamma \frac{\partial f}{\partial \rho} + \delta \frac{\partial f}{\partial \lambda} = 0. \]

Taking the limit of the left-hand side as $x \to \infty$, we obtain $\alpha = 0$. Taking $x = 0$ yields $\alpha + \delta (1 - p) / \lambda = 0$, thus $\delta = 0$. Our equation becomes
\[ \frac{\beta}{\theta^2} \left\{ \rho x (x + 1)^{-\rho - 1} + [1 - (x + 1)^{-\rho}] \left(1 - \frac{x}{\theta}\right) \right\} \exp \left(-\frac{x}{\theta}\right) - \gamma \left\{ 1 - \exp \left(-\frac{x}{\theta}\right) \left[1 - \rho \log(x + 1) + \frac{(x + 1) \log(x + 1)}{\theta} \exp \left(-\frac{x}{\theta}\right) \right] \right\} (x + 1)^{-\rho - 1} = 0, \quad (2.4) \]
for all $x > 0$. Assuming that $\beta \neq 0$, we obtain

$$
\rho = \frac{\gamma}{\beta} \left\{ 1 + \theta [1 - \rho(e - 1)] + \frac{\theta(e - 1)}{\log(\theta + 1)} \right\} \log(\theta + 1).
$$

As $\theta \to \infty$, the left-hand side tends to $\rho$, whereas the right-hand side tends to 0 when $\gamma = 0$ and its absolute value tends to $\infty$ when $\gamma \neq 0$. This is impossible, thus $\beta = 0$. Equation (2.4) becomes

$$
\gamma \left\{ [1 - \exp(-\frac{x}{\theta})] [1 - \rho \log(x + 1)] + \frac{(x + 1) \log(x + 1)}{\theta} \exp\left(-\frac{x}{\theta}\right) \right\} = 0.
$$

Suppose that $\gamma = 0$. Then

$$
\frac{1}{\theta} = \left[ \frac{\exp\left(\frac{x}{\theta}\right) - 1}{x + 1} \right] \left[ \rho - \frac{1}{\log(x + 1)} \right].
$$

As $x \to \infty$, the left-hand side tends to $1/\theta$, whereas the right-hand side tends to $\infty$. This is a contradiction. Then $\gamma = 0$. Therefore, the Fisher information matrix is positive-definite. Hence condition C2 is fulfilled.

Step 3. For brevity’s sake, we introduce a new notation. For each $\Psi \in \Omega$ and each $r > 0$, we now write

$$
f_{\Psi,r} = \sup \left\{ f_{\Psi'}, \Psi' \in B_r(\Psi) \right\}, \text{ and } f_{\Psi,r}^* = \max \left\{ 1, f_{\Psi,r} \right\}.
$$

Fix $\Psi \in \Omega$ and $r > 0$. We have that

$$
\int_0^\infty \log f_{\Psi,r}^* f_{\Psi'} = \int_{\{f_{\Psi,r} \geq 1\}} \log f_{\Psi,r}^* f_{\Psi'} + \int_{\{f_{\Psi,r} < 1\}} \log f_{\Psi,r}^* f_{\Psi'}
= \int_{\{f_{\Psi,r} \geq 1\}} \log f_{\Psi,r} f_{\Psi'}.
$$
We know that for each \( u > 0 \)

\[
\log u \leq u - 1,
\]

hence

\[
0 \leq I(\{f_{\Psi,r} \geq 1\}) \log f_{\Psi,r} f_{\Psi^*} \leq f_{\Psi,r} f_{\Psi^*}.
\]

For each \( x \geq 0 \) and each \( \Psi' = (p', \theta', \rho', \lambda') \in \Omega \)

\[
f_{\Psi'}(x) = p' \left\{ \left[ \frac{1}{\theta'} \exp \left( -\frac{x}{\theta'} \right) \right] \left[ 1 - (x + 1)^{-\rho'} \right] 
+ \left[ 1 - \exp \left( -\frac{x}{\theta'} \right) \right] \left[ \rho'(x + 1)^{-\rho'-1} \right] \right\} 
+ (1 - p') \left[ \frac{1}{\lambda'} \exp \left( -\frac{x}{\lambda'} \right) \right],
\]

and so

\[
f_{\Psi'}(x) \leq p' \left( \frac{1}{\theta'} + \rho' \right) + (1 - p') \left( \frac{1}{\lambda'} \right),
\]

Thus if \( r > 0 \) is sufficiently small, there exists \( M > 0 \) such that

\[
f_{\Psi,r} \leq M,
\]

hence

\[
0 \leq I(\{f_{\Psi,r} \geq 1\}) \log f_{\Psi,r} f_{\Psi^*} \leq M f_{\Psi^*}.
\]
and so

$$\int_{\{f_{\psi,r} \geq 1\}} \log f_{\psi,r} f_{\psi^*} < \infty.$$  

Therefore,

$$\int_{0}^{\infty} \log f_{\psi,r}^* f_{\psi^*} < \infty,$$

provided that $r > 0$ is small enough. Hence Condition C3 holds.

Step 4. We want to check that $\int_{0}^{\infty} |\log f_{\psi^*}| f_{\psi^*} < \infty$.

For each $u > 0$, we have that

$$|\log u| \leq \frac{|u - 1|}{\sqrt{u}} \leq \sqrt{u} + \frac{1}{\sqrt{u}},$$

thus $|\log f_{\psi^*}| f_{\psi^*} \leq \sqrt{f_{\psi^*}} f_{\psi^*} + \sqrt{f_{\psi^*}}$.

By the AM-GM Inequality,

$$\sqrt{f_{\psi^*}} f_{\psi^*} \leq \frac{1}{2} (f_{\psi^*} + f_{\psi^*}^2).$$

We know that for each $x \geq 0$

$$f_{\psi^*}(x) = p^* f_{1\psi^*}(x) + (1 - p^*) f_{2\psi^*}(x),$$

where

$$f_{1\psi^*}(x) = \frac{1}{\theta^*} \exp \left( -\frac{x}{\theta^*} \right) [1 - (x + 1)^{-\rho^*}] + \rho^*(x + 1)^{\rho^*-1} \left[ 1 - \exp \left( -\frac{x}{\theta^*} \right) \right].$$
and

\[ f_{2\psi^*}(x) = \frac{1}{\lambda^*} \exp \left( -\frac{x}{\lambda^*} \right). \]

Therefore,

\[ [f_{\psi^*}(x)]^2 \leq 3 \left[ \left( \frac{p^*}{\theta^*} \right)^2 \exp \left( -\frac{2x}{\theta^*} \right) + (p^* \rho^*)^2 (x + 1)^{-2\rho^* - 2} + \left( \frac{1 - p^*}{\lambda^*} \right)^2 \exp \left( -\frac{2x}{\lambda^*} \right) \right], \]

and so

\[ \int_{0}^{\infty} f_{\psi^*}^2 < \infty. \]

Since

\[ \int_{0}^{\infty} f_{\psi^*} = 1, \]

it follows that

\[ \int_{0}^{\infty} \sqrt{f_{\psi^*} f_{\psi^*}} < \infty. \]

Now, it remains to prove that

\[ \int_{0}^{\infty} \sqrt{f_{\psi^*}} < \infty. \]
For each $x \geq 0$, we have that

$$\frac{f_{1\Psi^*}(x)}{\sqrt{f_{\Psi^*}(x)}} \leq \frac{f_{1\Psi^*}(x)}{\sqrt{p^* f_{1\Psi^*}(x)}} \leq \left\{ \frac{p^*}{\theta^*} \exp \left( -\frac{x}{\theta^*} \right) \left[ 1 - (x + 1)^{-\rho} \right] \right\}^{-1/2} f_{1\Psi^*}(x).$$

Recalling that

$$f_{1\Psi^*}(x) = \frac{1}{\theta^*} \exp \left( -\frac{x}{\theta^*} \right) \left[ 1 - (x + 1)^{-\rho} \right] + \rho^*(x + 1)^{\rho^* - 1} \left[ 1 - \exp \left( -\frac{x}{\theta^*} \right) \right],$$

and noting that

$$\int_0^\infty (x + 1)^{-(\rho^*+1)} \exp \left( \frac{x}{2\theta^*} \right) dx = 4 \theta^* \int_1^\infty \left[ 1 + 4 \theta^* \log(u) \right]^{-(\rho^*+1)} u du = (4 \theta^*)^{-\rho^*} \int_1^\infty \left[ 1 + \frac{1}{4 \theta^* \log(u)} \right]^{-\rho^*/-(\rho^*+1)} \log(u)^{-(\rho^*+1)} u du < \infty,$$

entail

$$\int_0^\infty \frac{f_{1\Psi^*}(x)}{\sqrt{f_{\Psi^*}(x)}} < \infty.$$
we get

$$\int_0^\infty \frac{f_{2\Psi^*}(x)}{\sqrt{f_{\Psi^*}(x)}} < \infty.$$ 

Therefore, $f_0^\infty \sqrt{f_{\Psi^*}} < \infty$, showing that

$$\int_0^\infty |\log f_{\Psi^*}| f_{\Psi^*} < \infty.$$ 

This completes the proof of Claim 3.

![Figure 2.1: Mixture density for the full parameter (0.25, 2, 1, 1) in solid blue, (0.5, 1, 2, 1) in dotted red, and (0.75, 1, 1, 2) in longdash green](image)

**2.2 Parameter Estimation**

The pdf of the mixture can be written as

$$f(x, \Psi) = pf_1(x, \Psi_1) + (1 - p)f_2(x, \Psi_2), \quad (2.5)$$

where $\Psi = (p, \theta, \rho, \lambda)$ is the full parameter, $\Psi_1 = (\theta, \rho)$, $\Psi_2 = \lambda$,

$$f_1(x, \Psi_1) = \frac{1}{\theta} \exp \left( -\frac{x}{\theta} \right) \left[ 1 - (x + 1)^{-\rho} \right] + \rho(x + 1)^{-1-\rho} \left[ 1 - \exp \left( -\frac{x}{\theta} \right) \right],$$
and
\[ f_2(x, \Psi_2) = \frac{1}{\lambda} \exp \left( -\frac{x}{\lambda} \right). \]

With the growing use of computers in the 1960s, the maximum likelihood (ML) method has become one of the most common methods for mixture density estimation. The customary way of finding ML estimates is to first determine a system of equations called the likelihood equations and then attempt to solve these equations. For mixture density problems, the likelihood equations are almost certainly nonlinear and beyond hope of solutions by analytic means. Because of these computational difficulties, one must seek an approximate solution via some iterative procedure [15]. The expectation-maximization (EM) algorithm provides an iterative procedure for computing ML estimates of the parameter values of a mixture distribution. Assume that \((x_1, \ldots, x_n)\) is an observed sample of size \(n\) from the density in Equation (2.5). For \(i = 1, \ldots, n\), let \(z_i\) be the indicator function of cluster membership. In other words, let \(z_i = 1\) if the observation \(x_i\) did arise from the first component mixture \(f_1(x, \Psi_1)\) and \(z_i = 0\) if the observation \(x_i\) did arise from the second component mixture \(f_2(x, \Psi_2)\). For this specification, the complete data likelihood is
\[
L_c(\Psi) = \prod_{i=1}^{n} \left[ pf_1(x_i, \Psi_1) \right]^{z_i} \left[ (1 - p) f_2(x_i, \Psi_2) \right]^{1 - z_i}.
\]

Denoting \(\Psi^{(k)}\) the \(k\)-th iterative solution, the E-step is given by
\[
E[\log L_c(\Psi)|x, \Psi^{(k)}] = \sum_{i=1}^{n} E(z_i|x, \Psi^{(k)}) \log f_1(x_i, \Psi_1) + \sum_{i=1}^{n} \left[ 1 - E(z_i|x, \Psi^{(k)}) \right] \log f_2(x_i, \Psi_2)
\]
\[
+ \sum_{i=1}^{n} E(z_i|x, \Psi^{(k)}) \log p + \sum_{i=1}^{n} \left[ 1 - E(z_i|x, \Psi^{(k)}) \right] \log(1 - p) \tag{2.6}
\]
where

\[ E(z_i|x, \Psi^{(k)}) = \frac{p^{(k)} f_1(x_i, \Psi_1)}{p^{(k)} f_1(x_i, \Psi_1) + (1 - p^{(k)}) f_2(x_i, \Psi_2)}. \]

For the M-step, we differentiate Equation (2.6) with respect to \( p \) and equating it to zero to obtain the updated estimate of \( p \) as

\[ p^{(k+1)} = \frac{1}{n} \sum_{i=1}^{n} E(z_i|x, \Psi^{(k)}). \]

Similarly, the updated estimate of \( \lambda \) is obtained as

\[ \lambda^{(k+1)} = \frac{1}{\sum_{i=1}^{n} [1 - E(z_i|x, \Psi^{(k)})]} \sum_{i=1}^{n} [1 - E(z_i|x, \Psi^{(k)})] x_i. \]

We obtain the updated estimates \((\theta^{(k+1)}, \rho^{(k+1)})\) of \((\theta, \rho)\) by maximizing \( E[\log L_c(\Psi)|x, \Psi^{(k)}] \).

As stopping criterion, we use the relative change in log-likelihood and set the tolerance to \( 10^{-6} \). The likelihood function is given by

\[ L(\Psi) = \prod_{i=1}^{n} [p f_1(x_i, \Psi_1) + (1 - p) f_2(x_i, \Psi_2)], \]

and the maximum number of iterations is set to 1000.

### 2.3 A Simulation Study

For each set of parameter values, we simulate a sample of size \( n = 100, 1000, 10000 \) from the mixture distribution and perform \( b = 1000 \) replications in each case. The parameters are estimated using the EM algorithm and the replicate estimates are used to compute the bias and standard error. The results are the following:
Table 2.1: Absolute value of estimated bias (EB), estimated standard error (ESE), and mean square error (MSE) of the MLE for different sample sizes and \((\theta, \rho, \lambda) = (1, 1, 1)\).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>(n)</th>
<th>EB</th>
<th>ESE</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>100</td>
<td>0.004017</td>
<td>0.130531</td>
<td>0.017054</td>
</tr>
<tr>
<td>1</td>
<td>0.129834</td>
<td>0.121364</td>
<td>0.031586</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.013960</td>
<td>0.094229</td>
<td>0.009074</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.009467</td>
<td>0.204002</td>
<td>0.041706</td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>1000</td>
<td>0.002097</td>
<td>0.037230</td>
<td>0.001390</td>
</tr>
<tr>
<td>1</td>
<td>0.137490</td>
<td>0.039158</td>
<td>0.020437</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.000071</td>
<td>0.027997</td>
<td>0.000784</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.021939</td>
<td>0.058806</td>
<td>0.003940</td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>10000</td>
<td>0.002872</td>
<td>0.011846</td>
<td>0.000149</td>
</tr>
<tr>
<td>1</td>
<td>0.138188</td>
<td>0.012386</td>
<td>0.019249</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.001795</td>
<td>0.009174</td>
<td>0.000872</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.021876</td>
<td>0.019276</td>
<td>0.000850</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>100</td>
<td>0.077124</td>
<td>0.127659</td>
<td>0.022245</td>
</tr>
<tr>
<td>1</td>
<td>0.085909</td>
<td>0.147818</td>
<td>0.029231</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.044735</td>
<td>0.087107</td>
<td>0.009589</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.056635</td>
<td>0.276511</td>
<td>0.079666</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>1000</td>
<td>0.056436</td>
<td>0.033657</td>
<td>0.004318</td>
</tr>
<tr>
<td>1</td>
<td>0.076099</td>
<td>0.047552</td>
<td>0.008052</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.057053</td>
<td>0.025536</td>
<td>0.003907</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.037352</td>
<td>0.080185</td>
<td>0.007825</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>10000</td>
<td>0.057252</td>
<td>0.007144</td>
<td>0.003329</td>
</tr>
<tr>
<td>1</td>
<td>0.075201</td>
<td>0.015038</td>
<td>0.005881</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.058516</td>
<td>0.008431</td>
<td>0.003495</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.040179</td>
<td>0.020338</td>
<td>0.002028</td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>100</td>
<td>0.079115</td>
<td>0.135018</td>
<td>0.024489</td>
</tr>
<tr>
<td>1</td>
<td>0.030133</td>
<td>0.145988</td>
<td>0.022221</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.033993</td>
<td>0.080818</td>
<td>0.007687</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.132117</td>
<td>0.428093</td>
<td>0.200718</td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>1000</td>
<td>0.060991</td>
<td>0.035917</td>
<td>0.005010</td>
</tr>
<tr>
<td>1</td>
<td>0.015470</td>
<td>0.045357</td>
<td>0.002297</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.041348</td>
<td>0.024245</td>
<td>0.002297</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.122116</td>
<td>0.138748</td>
<td>0.034163</td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>10000</td>
<td>0.064610</td>
<td>0.010149</td>
<td>0.004278</td>
</tr>
<tr>
<td>1</td>
<td>0.014393</td>
<td>0.014343</td>
<td>0.000413</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.042101</td>
<td>0.007979</td>
<td>0.001836</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.133845</td>
<td>0.037863</td>
<td>0.019348</td>
<td></td>
</tr>
</tbody>
</table>
Table 2.2: Absolute value of estimated bias (EB), estimated standard error (ESE), and mean square error (MSE) of the MLE for different sample sizes and \((\theta, \rho, \lambda) = (2, 1, 1)\).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>n</th>
<th>EB</th>
<th>ESE</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>100</td>
<td>0.036425</td>
<td>0.097470</td>
<td>0.010827</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.004741</td>
<td>0.281410</td>
<td>0.079214</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.047621</td>
<td>0.086731</td>
<td>0.009790</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.048857</td>
<td>0.235603</td>
<td>0.057896</td>
</tr>
<tr>
<td>0.25</td>
<td>1000</td>
<td>0.021432</td>
<td>0.028731</td>
<td>0.001285</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.017385</td>
<td>0.083788</td>
<td>0.007323</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.060492</td>
<td>0.025791</td>
<td>0.004324</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.027774</td>
<td>0.073213</td>
<td>0.006132</td>
</tr>
<tr>
<td>0.25</td>
<td>10000</td>
<td>0.020706</td>
<td>0.008868</td>
<td>0.000507</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.019142</td>
<td>0.028080</td>
<td>0.001155</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.062082</td>
<td>0.008457</td>
<td>0.003926</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.026666</td>
<td>0.023061</td>
<td>0.001243</td>
</tr>
<tr>
<td>0.5</td>
<td>100</td>
<td>0.076132</td>
<td>0.088873</td>
<td>0.013694</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.077585</td>
<td>0.287712</td>
<td>0.088797</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.200522</td>
<td>0.094796</td>
<td>0.049195</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.132086</td>
<td>0.302958</td>
<td>0.109230</td>
</tr>
<tr>
<td>0.5</td>
<td>1000</td>
<td>0.094564</td>
<td>0.017239</td>
<td>0.009239</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.074299</td>
<td>0.089136</td>
<td>0.013466</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.188611</td>
<td>0.028402</td>
<td>0.036381</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.188569</td>
<td>0.083142</td>
<td>0.042471</td>
</tr>
<tr>
<td>0.5</td>
<td>10000</td>
<td>0.095869</td>
<td>0.005358</td>
<td>0.009220</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.074401</td>
<td>0.029399</td>
<td>0.006400</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.187133</td>
<td>0.009284</td>
<td>0.035105</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.193832</td>
<td>0.024947</td>
<td>0.038193</td>
</tr>
<tr>
<td>0.75</td>
<td>100</td>
<td>0.058099</td>
<td>0.064567</td>
<td>0.007544</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.136275</td>
<td>0.275876</td>
<td>0.094678</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.147818</td>
<td>0.081009</td>
<td>0.028413</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.179138</td>
<td>0.417017</td>
<td>0.205994</td>
</tr>
<tr>
<td>0.75</td>
<td>1000</td>
<td>0.059040</td>
<td>0.011733</td>
<td>0.003623</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.124184</td>
<td>0.089252</td>
<td>0.023387</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.137411</td>
<td>0.024590</td>
<td>0.019486</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.233726</td>
<td>0.108334</td>
<td>0.066364</td>
</tr>
<tr>
<td>0.75</td>
<td>10000</td>
<td>0.058842</td>
<td>0.002639</td>
<td>0.003469</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.123264</td>
<td>0.028709</td>
<td>0.016018</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.136129</td>
<td>0.007918</td>
<td>0.018594</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.240665</td>
<td>0.030255</td>
<td>0.058835</td>
</tr>
</tbody>
</table>
Table 2.3: Absolute value of estimated bias (EB), estimated standard error (ESE), and mean square error (MSE) of the MLE for different sample sizes and \((\theta, \rho, \lambda) = (1, 2, 1)\).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>(n)</th>
<th>EB</th>
<th>ESE</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>100</td>
<td>0.013174</td>
<td>0.235517</td>
<td>0.055642</td>
</tr>
<tr>
<td>0.25</td>
<td>1000</td>
<td>0.021766</td>
<td>0.091055</td>
<td>0.008765</td>
</tr>
<tr>
<td>0.25</td>
<td>10000</td>
<td>0.022320</td>
<td>0.028074</td>
<td>0.001286</td>
</tr>
<tr>
<td>0.5</td>
<td>100</td>
<td>0.089767</td>
<td>0.234626</td>
<td>0.063107</td>
</tr>
<tr>
<td>0.5</td>
<td>1000</td>
<td>0.002857</td>
<td>0.114842</td>
<td>0.013197</td>
</tr>
<tr>
<td>0.5</td>
<td>10000</td>
<td>0.000554</td>
<td>0.037377</td>
<td>0.001397</td>
</tr>
<tr>
<td>0.75</td>
<td>100</td>
<td>0.063045</td>
<td>0.121189</td>
<td>0.018661</td>
</tr>
<tr>
<td>0.75</td>
<td>1000</td>
<td>0.101594</td>
<td>0.119763</td>
<td>0.024665</td>
</tr>
<tr>
<td>0.75</td>
<td>10000</td>
<td>0.057989</td>
<td>0.016877</td>
<td>0.003648</td>
</tr>
</tbody>
</table>
Table 2.4: Absolute value of estimated bias (EB), estimated standard error (ESE), and mean square error (MSE) of the MLE for different sample sizes and $(\theta, \rho, \lambda) = (1, 1, 2)$.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$n$</th>
<th>EB</th>
<th>ESE</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>100</td>
<td>0.075846</td>
<td>0.127417</td>
<td>0.021988</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.052220</td>
<td>0.146530</td>
<td>0.024198</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.151390</td>
<td>0.064290</td>
<td>0.027052</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.022409</td>
<td>0.254496</td>
<td>0.065270</td>
</tr>
<tr>
<td>0.25</td>
<td>1000</td>
<td>0.072928</td>
<td>0.038276</td>
<td>0.006784</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.068155</td>
<td>0.044974</td>
<td>0.006668</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.155405</td>
<td>0.019867</td>
<td>0.024545</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.004948</td>
<td>0.078074</td>
<td>0.006120</td>
</tr>
<tr>
<td>0.25</td>
<td>10000</td>
<td>0.071816</td>
<td>0.012624</td>
<td>0.005317</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.069763</td>
<td>0.014125</td>
<td>0.005066</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.155775</td>
<td>0.006403</td>
<td>0.024307</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.005391</td>
<td>0.026007</td>
<td>0.007050</td>
</tr>
<tr>
<td>0.5</td>
<td>100</td>
<td>0.093928</td>
<td>0.166588</td>
<td>0.038552</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.008745</td>
<td>0.146821</td>
<td>0.021633</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.114091</td>
<td>0.064107</td>
<td>0.017126</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.083000</td>
<td>0.340674</td>
<td>0.122498</td>
</tr>
<tr>
<td>0.5</td>
<td>1000</td>
<td>0.091777</td>
<td>0.050548</td>
<td>0.011052</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.007229</td>
<td>0.045005</td>
<td>0.002078</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.118396</td>
<td>0.019536</td>
<td>0.014399</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.046569</td>
<td>0.010124</td>
<td>0.012456</td>
</tr>
<tr>
<td>0.5</td>
<td>10000</td>
<td>0.091133</td>
<td>0.016471</td>
<td>0.008577</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.008840</td>
<td>0.014133</td>
<td>0.000278</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.118817</td>
<td>0.006383</td>
<td>0.014158</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.044625</td>
<td>0.033844</td>
<td>0.003137</td>
</tr>
<tr>
<td>0.75</td>
<td>100</td>
<td>0.153047</td>
<td>0.160349</td>
<td>0.049135</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.068209</td>
<td>0.147090</td>
<td>0.026288</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.132055</td>
<td>0.066673</td>
<td>0.021884</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.204200</td>
<td>0.459912</td>
<td>0.253217</td>
</tr>
<tr>
<td>0.75</td>
<td>1000</td>
<td>0.105209</td>
<td>0.046931</td>
<td>0.013271</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.052036</td>
<td>0.045018</td>
<td>0.004734</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.113628</td>
<td>0.020599</td>
<td>0.013336</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.186114</td>
<td>0.153132</td>
<td>0.058088</td>
</tr>
<tr>
<td>0.75</td>
<td>10000</td>
<td>0.096499</td>
<td>0.015401</td>
<td>0.009549</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.050410</td>
<td>0.014135</td>
<td>0.002741</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.111835</td>
<td>0.006792</td>
<td>0.012553</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.190056</td>
<td>0.050757</td>
<td>0.038697</td>
</tr>
</tbody>
</table>
Our observations are as follows: As the sample size increases, the estimated standard error (ESE) and the mean square error (MSE) of each parameter decrease. However, this is not true for the bias. From Claim 3, the MLE exists and is globally consistent in every compact sub-parameter space containing the true parameter in its interior [15 & 23]. Despite this elegant Redner’s result, this method suffers, at least theoretically, from the risk that the true mixing distribution may not satisfy the constraint imposed [110 & 112]. This possibly explains the behavior of the bias in our simulation study. Interestingly, Holmer found that the MLE for mixture of two normal distributions may not perform well with regards to bias in the small sample case, especially when the two distributions are poorly separated [107, 108 & 109]. To address the limitations of the MLE, we could use estimation methods such as the penalized ML and the penalized minimum-distance [111, 112 & 114].

2.4 Application to Real Data

Our dataset represents remission times (in months) of a random sample of 128 bladder cancer patients reported in [93]. We fit our mixture distribution to the data. With starting values \((0.5, \text{median}(x), 1 + 1/\text{median}(x), \text{quantile}(x, \text{probs} = 0.10))\), we obtain as MLE estimates: \(\hat{p} = 0.8110, \hat{\theta} = 6.280, \hat{\rho} = 1.159\) and \(\hat{\lambda} = 14.32\) after 18 iterations. For our dataset, the summary statistics include: 0.080 (minimum), 3.348 (first quartile), 6.28 (median), 9.209 (mean), 11.68 (third quartile) and 79.05 (maximum). We can observe that the mean is higher than the median and the third quartile is far away from the median while the first quartile is closer to it. Hence we could consider the fit of right-skewed distributions such as Weibull, gamma, log-normal and Pareto. It is worth noting that the estimated skewness coefficient is 3.359 and the estimated kurtosis 16.09. With the R libraries fitdistrplus and mixtools, we fit different models to the data and perform discrimination among them using the Akaike
information criterion (AIC) [100], the corrected Akaike information criterion (AICc) [101],
the consistent Akaike information criterion (CAIC) [102], the Bayesian information criterion
(BIC) [103] and the Hannan-Quinn information criterion (HQC) [104]. The results are given
as follows, with the understanding that the Lomax distribution referred to in the following
table is the Lomax distribution with scale parameter 1.

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>AICC</th>
<th>CAIC</th>
<th>BIC</th>
<th>HQC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed mixture</td>
<td>826.25</td>
<td>826.58</td>
<td>841.66</td>
<td>837.66</td>
<td>830.89</td>
</tr>
<tr>
<td>Weibull distribution</td>
<td>827.78</td>
<td>827.88</td>
<td>835.49</td>
<td>833.49</td>
<td>830.10</td>
</tr>
<tr>
<td>Gamma distribution</td>
<td>826.17</td>
<td>826.27</td>
<td>833.87</td>
<td>831.87</td>
<td>828.49</td>
</tr>
<tr>
<td>2-parameter Pareto distribution</td>
<td>827.27</td>
<td>827.37</td>
<td>834.98</td>
<td>832.98</td>
<td>829.59</td>
</tr>
<tr>
<td>Log-normal distribution</td>
<td>829.31</td>
<td>829.42</td>
<td>837.02</td>
<td>835.02</td>
<td>831.63</td>
</tr>
<tr>
<td>Lomax distribution</td>
<td>942.08</td>
<td>942.11</td>
<td>945.94</td>
<td>944.94</td>
<td>943.24</td>
</tr>
<tr>
<td>Exponential distribution</td>
<td>826.38</td>
<td>826.41</td>
<td>830.23</td>
<td>829.23</td>
<td>827.53</td>
</tr>
<tr>
<td>Mixture Lomax-exponential</td>
<td>830.36</td>
<td>830.46</td>
<td>841.92</td>
<td>838.92</td>
<td>833.84</td>
</tr>
<tr>
<td>Mixture of two gamma distributions</td>
<td>823.87</td>
<td>824.36</td>
<td>843.13</td>
<td>838.13</td>
<td>829.67</td>
</tr>
<tr>
<td>Mixture of two normal distributions</td>
<td>860.50</td>
<td>860.99</td>
<td>879.76</td>
<td>874.76</td>
<td>866.30</td>
</tr>
<tr>
<td>Distribution of the maximum of two independent exponential and Lomax random variables</td>
<td>822.30</td>
<td>822.39</td>
<td>830</td>
<td>828</td>
<td>824.61</td>
</tr>
</tbody>
</table>
Figure 2.2: Histogram, empirical pdf and cdf for remission times data

Figure 2.3: QQ plot of remission times data against the exponential distribution
Table 2.6: Results of goodness of fit tests with their test statistics and p-values

<table>
<thead>
<tr>
<th>GOF Test</th>
<th>Value of test statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>0.05640</td>
<td>0.8078</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>0.4387</td>
<td>0.8094</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>0.06928</td>
<td>0.7569</td>
</tr>
<tr>
<td>Chi-square</td>
<td>0.1121</td>
<td>0.9455</td>
</tr>
</tbody>
</table>

Figure 2.4: Sample mean excess plot of remission times data

From Table 2.5, we note that the best model is the distribution of the maximum of independent exponential and Lomax. For the remission times data, the parameter estimates for this model are: $\tilde{\theta} = 7.539$, and $\tilde{\rho} = 1.141$ and the results of the goodness of fit (GOF) tests are displayed in Table 2.6, where ties were removed for the Kolmogorov-Smirnov GOF test. Moreover, the best model is given by one of the components of the proposed mixture. On the QQ plot, we observe a concave departure from typical exponential behavior. Moreover, the points show an upward trend on the sample mean excess plot. Then it seems safe to conclude that the remission times data has a (right) tail heavier than the exponential.
The results of the goodness of fit tests indicate that our fitted distribution is an appropriate representation of the population as hypothesized under the null hypothesis.

### 2.5 Extremal Mixture Models

The modeling of the tail is of vital importance when dealing with large claims. As mentioned earlier, the threshold estimation can be difficult. One of the motivations for the development of EVMMs is that they reduce the subjectivity of threshold estimation while accounting for any induced uncertainty. Extremal mixture models, also called extreme value mixture models (EVMMs), include threshold selection (or bypass it by using a transition function) and capture the bulk of the distribution below the threshold as well as the tail of the distribution above the threshold. Typically, the tail model is a GPD whereas the bulk model is parametric (e.g., normal, gamma, Weibull, beta, log-normal), semi-parametric (e.g., mixture of normal, gamma, log-normal), or non-parametric (e.g., kernel density estimation, smoothing polynomials). A key issue with these models is the sensitivity of the tail fit to that of the bulk.

Behrens et al [76] proposed the common mixture model specification

\[
F(x|\eta, u, \sigma_u, \xi) = \begin{cases} 
H(x|\eta), & \text{if } x \leq u; \\
H(u|\eta) + [1 - H(u|\eta)]G(x|u, \sigma_u, \xi), & \text{if } x > u,
\end{cases}
\]

where \(H(x|\eta)\) is a gamma, Weibull or normal d.f. and \(G(x|u, \sigma_u, \xi)\) the GPD d.f. The threshold \(u\) is treated as a parameter to be estimated and \(u\) is the point where the density has a discontinuity. Carreau and Bengio [77] introduced a hybrid Pareto model (a combination of normal and GPD tail) with continuity constraints on the pdf and its first derivative at
the threshold, including extension to a mixture of these hybrid Pareto’s to capture possible asymmetry, multi-modality and tail heaviness. The pdf of the hybrid Pareto is given by

$$f(x|\mu, \beta, \xi) = \begin{cases} \frac{1}{\tau} h(x|\mu, \beta), & \text{if } x \leq u; \\ \frac{1}{\tau} g(x|u, \sigma_u, \xi), & \text{if } x > u, \end{cases}$$

where $\tau = 1 + \Phi[(u - \mu)/\sigma]$ is a normalization constant, $h(x|\mu, \beta)$ the normal density with mean $\mu$ and standard deviation $\beta$ and $g(x|u, \sigma_u, \xi)$ the GPD density. Frigessi et al [78] suggested a dynamically weighted mixture model where the weight function varies over the range of the support, shifting the weights from a light-tailed density (Weibull) to the GPD which will dominate the upper tail. The pdf is

$$f(x|\theta, \beta, u, \sigma_u, \xi) = \frac{[1 - p(x|\theta)]h(x|\beta) + p(x|\theta)g(x|u = 0, \sigma_u, \xi)}{Z(\theta, \beta, u = 0, \sigma_u, \xi)},$$

where $x \geq 0$, $\theta = (\mu, \tau)$, $\tau > 0$, $p(x|\theta) = \arctan[(x - \mu)/\tau] + 1/2$ denotes the Cauchy df with location parameter $\mu$ and scale parameter $\tau$, $h(x|\beta)$ the Weibull pdf and $g(x|u = 0, \sigma_u, \xi)$ the GPD pdf. One of the drawbacks with the aforementioned approaches is the prior specification of a parametric model for the bulk of the distribution (and the associated weight function where appropriate) and the complicated inference (and sample properties) for the mixture of hybrid Pareto distributions [91]. Assuming that the observations below the threshold follow a non-parametric density $h(.|\lambda, X)$ which is dependent on a parameter $\lambda$ and the observation vector $X$, MacDonald et al [91] suggested the model

$$F(x|\lambda, u, \sigma_u, \xi, X) = \begin{cases} (1 - \phi_u) \frac{H(x|\lambda, X)}{H(u|\lambda, X)}, & \text{if } x \leq u; \\ (1 - \phi_u) + \phi_u G(x|u, \sigma_u, \xi), & \text{if } x > u, \end{cases}$$

where $H(x|\lambda, X)$ and $G(x|u, \sigma_u, \xi)$ are the CDFs of the hybrid Pareto and GPD densities, respectively.
where $X = \{x_i : i = 1, ..., n\}$ is a sequence of $n$ iid observations, $\phi_u$ is the probability of being above the threshold and $\phi_u G(x|u, \sigma_u, \xi)$ the unconditional GPD function given by the survival probability. The last two EVMMs (models proposed by Frigessi et al [78] and MacDonald et al [91]) can be easily written as pure mixture models. It should be noted that Behrens et al [76] and Frigessi et al [76] presented EVMMs with fully parametric bulk models, which is not the case of Carreau and Bengio [77] or MacDonald et al [91]. For a review of EVMMs, we refer to [73].

2.6 A Family of Dynamically Weighted Mixture Models

Following Frigessi et al [78], we propose the following general mixture model

$$f(x, \theta, \rho, \lambda) = \frac{M(x, \theta) h(x, \rho) + [1 - M(x, \theta)] l(x, \lambda)}{C(\theta, \rho, \lambda)},$$

where $x \geq 0$, $M(x, \theta)$ is a df with parameter vector $\theta$, $h(x, \rho)$ a heavy-tailed pdf with parameter vector $\rho$, $l(x, \lambda)$ a light-tailed pdf with parameter vector $\lambda$ and $C(\theta, \rho, \lambda)$ a normalization constant. Henceforth, we suppose that the mixture $f$ is identifiable.

**Proposition 5** $f$ is a heavy-tailed pdf.

**Proof** Let $\delta > 0$ be given, and fix the parameters $\theta, \rho, \lambda$. For every $x \geq 0$, we have that

$$\exp(\delta x) f(x, \theta, \rho, \lambda) \geq \exp(\delta x) h(x, \rho) \left\{ \frac{M(x, \theta)}{C(\theta, \rho, \lambda)} \right\}.$$

Since

$$\limsup_{x \to \infty} \exp(\delta x) h(x, \rho) = \infty,$$
and
\[ \lim_{x \to \infty} \frac{M(x, \theta)}{C(\theta, \rho, \lambda)} = \frac{1}{C(\theta, \rho, \lambda)} \in (0, \infty), \]
we obtain
\[ \limsup_{x \to \infty} \exp(\delta x) f(x, \theta, \rho, \lambda) = \infty, \]
so that \( f \) is heavy-tailed.

**Proposition 6** Assume that \( h \) is long-tailed and \( l/h \) has a finite limit at \( \infty \).
Then \( f \) is long-tailed.

**Proof** For each \( x \geq 0 \), the following holds.
\[
\frac{f(x + 1)}{f(x)} = \frac{M(x + 1)h(x + 1) + [1 - M(x + 1)] l(x + 1)}{M(x)h(x) + [1 - M(x)] l(x)}
\]
\[
= \frac{h(x + 1)}{h(x)} \left\{ \frac{M(x + 1) + [1 - M(x + 1)] l(x + 1)}{M(x) + [1 - M(x)] l(x)} \right\}
\]
Since the limit of \( M \) at infinity is 1, the limit of \( l/h \) is finite and \( h(x + 1) \sim h(x) \), we have that \( f(x + 1) \sim f(x) \). Therefore \( f \) is long-tailed.

**Proposition 7** Suppose that \( h \) is subexponential and that \( l/h \) has a finite limit at \( \infty \).
Then \( f \) is subexponential.

**Proof** For each \( x \geq 0 \), we have that
\[
\frac{f(x)}{h(x)} = \frac{1}{C} \left\{ M(x) + [1 - M(x)] \frac{l(x)}{h(x)} \right\}
\]
Since the limit of \( M \) at infinity is 1 and the limit of \( l/h \) is finite, the limit of \( f/h \) is a strictly positive number. Hence \( f \) and \( h \) are weakly tail-equivalent. By assumption, \( h \) is subex-
ponential. Therefore $f$ is long-tailed by Proposition 6. It follows that $f$ is subexponential [2].

2.7 Concluding Remarks and Future Work

Finite mixture models have been widely used in modeling various data with reasonably good results. In this dissertation, we study a composite mixture of two distributions. Some structural properties of the resulting distribution are discussed, including tail properties and maximum likelihood (ML) estimation via the EM algorithm. In many instances, the behavior of the mixture (e.g., heavy-tailedness, long-tailedness, subexponentiality, domain of attraction) is primarily governed by the Lomax distribution. Furthermore, our results indicate that one of the components of our mixture could be useful in modeling non-negative, heavy-tailed data. It can be proved that this component exhibits many of the properties of the composite mixture. It could be of interest to study the component closely, e.g., mathematical properties such as limiting behaviors of the pdf and the hazard rate. It would also be interesting to change its form and/or make-up. We use the exponential distribution as the other component, though this is just one of the possible choices. Finding better ways of choosing starting values in the EM algorithm would also be useful. In addition, estimation methods such as the penalized ML and the penalized minimum-distance are worth considering. Other future research projects relate to composite models, folded models, models spliced with extremal tails, continuous mixtures, geometric mixtures, robust estimation of heavy-tailed distributions, risk measures, heavy-tailed time series, stable distributions, Bayesian estimation of finite mixtures and ruin probabilities.
REFERENCES


## Appendix A: Initial Values for the Simulations

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\theta(0)$</th>
<th>$\rho(0)$</th>
<th>$\lambda(0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.25,1,1,1)</td>
<td>median($x$)</td>
<td>TME ($b_1 = 0, b_2 = 0.10$)</td>
<td>MPE-II</td>
</tr>
<tr>
<td>(0.50,1,1,1)</td>
<td>median($x$)</td>
<td>TME ($b_1 = 0, b_2 = 0.05$)</td>
<td>MPE-I</td>
</tr>
<tr>
<td>(0.75,1,1,1)</td>
<td>$x.40$</td>
<td>$n/\text{sum} \log(1+x)$</td>
<td>$x.25$</td>
</tr>
<tr>
<td>(0.25,2,1,1)</td>
<td>($x.70+x.80)/2$</td>
<td>TME ($b_1 = 0, b_2 = 0.10$)</td>
<td>MPE-II</td>
</tr>
<tr>
<td>(0.50,2,1,1)</td>
<td>($x.60+x.70)/2$</td>
<td>Combination of $g$ and $h$</td>
<td>$x.25$</td>
</tr>
<tr>
<td>(0.75,2,1,1)</td>
<td>($x.50+x.60)/2$</td>
<td>Combination of $g$ and $h$</td>
<td>$x.10$</td>
</tr>
<tr>
<td>(0.25,1,2,1)</td>
<td>($x.50+x.60)/2$</td>
<td>$1+(1/\text{mean}(x))$</td>
<td>MPE-II</td>
</tr>
<tr>
<td>(0.50,1,2,1)</td>
<td>($x.50+x.60)/2$</td>
<td>$1+(1/\text{mean}(x))$</td>
<td>MPE-I</td>
</tr>
<tr>
<td>(0.75,1,2,1)</td>
<td>($x.50+x.60)/2$</td>
<td>$1+(1/\text{median}(x))$</td>
<td>$x.10$</td>
</tr>
<tr>
<td>(0.25,1,1,2)</td>
<td>($x.30+x.40)/2$</td>
<td>PITSE ($t = 0.883$)</td>
<td>MPE-II</td>
</tr>
<tr>
<td>(0.50,1,1,2)</td>
<td>($x.30+x.40)/2$</td>
<td>PITSE ($t = 0.324$)</td>
<td>MPE-I</td>
</tr>
<tr>
<td>(0.75,1,1,2)</td>
<td>($x.30+x.40)/2$</td>
<td>$(n-3)/\text{sum} \log(1+x))</td>
<td>$x.25$</td>
</tr>
</tbody>
</table>

### Notes
1. For the trimmed mean estimator (TME), see [94 & 95].
2. The notation $x.40$ stands for the empirical 40th percentile.
3. Combining the geometric mean $g$ of ($x+1$), i.e., $g = \prod(x+1)^{1/n}$ and the harmonic mean $h$ of ($x+1$), i.e., $h = 1/\text{mean}(1/(x+1))$, we obtain as estimator of the shape parameter of the one-parameter Pareto distribution: $(nh)/\text{sum} \log(1+x) - 1$.
4. $1+(1/\text{median}(x))$ is a modified method of moments estimator of the shape parameter of the Pareto distribution.
5. For the probability integral transform statistic estimator (PITSE), refer to [96 & 97].
6. \((n-3)/\text{sum}(\log(1+x))\) is a modified maximum likelihood estimator of the shape parameter of the Pareto distribution. It was studied by A. M. Johnson [99].

7. MPE-I and MPE-II are modified percentile estimators for parameter estimation of the exponential distribution. They were computed using the method presented in [98]. An expression for MPE-I is given by \((x_{.50} + x_{.75})/(3 \log(2))\), and an expression for MPE-II is \((x_{.075} + \exp(\text{mean}(\log(x))))/(\log(4) + \exp(\psi(1)))\), where \(\psi\) is the digamma function. The proof proceeds as follows:

For the first part of the statement, the cdf of the exponential distribution with scale \(\theta\) is:
\[ F(x) = 1 - \exp(-x/\theta), \]
thus \(1 - \exp(-x_{.50}/\theta) = 0.50\) and \(1 - \exp(-x_{.75}/\theta) = 0.75\). Solving both equations simultaneously, we get the first modified percentile estimator (MPE-I) of \(\theta\) as:
\[ \hat{\theta} = (x_{.50} + x_{.75})/(3 \log(2)). \]

For the second part, we know that \(x_{.75} = -\theta \log(0.25)\). On the other hand, the geometric mean of the exponential distribution with scale \(\theta\) is:
\[ GM_x = \theta \exp(\text{digamma}(1)). \]
Solving both equations simultaneously, we obtain the second modified percentile estimator (MPE-II) of \(\theta\) as:
\[ \hat{\theta} = (x_{.75} + GM_x)/(\log(4) + \exp(\psi(1))). \]

.2 Appendix B: Computer Codes for Simulations

Example 1. Simulating from the mixture d.f. with full parameter \((0.25, 1, 1, 1)\)

# 1. Some help functions
n <- 100 # Set of sample sizes is \{100,1000,10000\}
tol <- 1e-6
max.it <- 1000

# Heavy-tailed mixture df
phi <- c(0.25,1,1,1)
pmix <- function(x,phi) {
    M <- 1-exp(-1*x/phi[2])
    H <- 1-(x+1)^(-1*phi[3])
    L <- 1-exp(-1*x/phi[4])
    return(phi[1]*M*H+(1-phi[1])*L)
}

# Suggested quantile function
qmix <- function(q,phi,tol) {
    lower <- 0
    upper <- 1
    while(pmix(upper,phi) < q) upper <- 2*upper
    error <- upper-lower
    while (error > tol){
        midpoint <- (lower+upper)/2
        if (pmix(midpoint,phi) < q) lower <- midpoint
        else upper <- midpoint
        error <- upper-lower
    }
    return((lower+upper)/2)
}

set.seed(1126)

# We simulate a sample of size n from the heavy-tailed df
rmix <- function(n,phi,tol){
    u <- runif(n)
    m <- matrix(0,n,1)
    for (i in 1:n) m[i] <- qmix(u[i],phi,tol)
return(m)
}

# 2. EM algorithm function
em.mix <- function(x, tol, max.it) {
  q <- median(x)
  b2 <- n-floor(0.10*n)
  g2 <- function(n) {
    b <- b2-1
    c <- 0
    for (k in 0:b) {
      for (j in 0:k) {
        c <- c + 1/(n-j)
      }
    }
    return(c)
  }
  t_2 <- function(n) {
    d =0
    for (i in 1:b2) {
      d <- d + log(1+x[i,1])
    }
    return(d)
  }
  r <- quantile(x, probs=0.75)
  s <- g2(n)/t_2(n)
  start <- c(q, s)
phi <- numeric(4)
phi[1] <- 0.5
phi[2] <- q
phi[3] <- s
phi[4] <- (r+exp(mean(log(x))))/(log(4)+exp(digamma(1)))

mysum <- function(a) {sum(a[is.finite(a)])}

f1 <- function(phi,x) {
t1 <- exp(-1*x/phi[2])*(1-(x+1)^(-1*phi[3]))/phi[2]
t2 <- phi[3]*(x+1)^(-1-phi[3])*(1-exp(-1*x/phi[2]))
return(t1+t2)
}

f2 <- function(phi,x) {
exp(-1*x/phi[4])/phi[4]
}

loglik<- rep(NA, 1000)
loglik[1]<- 0

loglik[2] <- mysum(log(phi[1]*f1(phi,x)+ (1-phi[1])*f2(phi,x)))

k <- 2
# loop
while(abs(loglik[k]-loglik[k-1])> tol*abs(loglik[k-1]) & k <= max.it) {
# E step
tau1 <- (phi[1]*f1(phi,x))/((phi[1]*f1(phi,x))+(1-phi[1])*f2(phi,x))
tau2 <- 1-tau1
# M step
phi[1] <- mysum(tau1)/n
phi[4] <- mysum(tau2*x)/mysum(tau2)
f <- function (phi, tau1,k) {
    return(-mysum(tau1*log(f1(phi, x)))-mysum(tau2*log(f2(phi, x))))
}

z <- optim(start,f,lower=0.01,tau1=tau1,k=k, method="L-BFGS-B")$par
phi <- c(phi[1],phi[2], phi[3],phi[4])
loglik[k+1]<- mysum(log(phi[1]*f1(phi,x)+(1-phi[1])*f2(phi,x)))
k <- k+1
}

phihat <- phi
return(phihat)
}
est <- matrix(0,1000,4)
for(i in 1:1000) {
x <- rmix(n,phi,tol)
est[i,] <- em.mix(x,tol,max.it)
}
est.bias <- apply(est,2,mean) - c(0.25,1,1,1)
est.bias
est.se <- apply(est,2,sd)
est.se
est.mse <- est.bias^2 + est.se^2

Example 2. Simulating from the mixture d.f. with full parameter (0.5, 1, 1, 2)
# 1. Some help functions

def n <- 100  # Set of sample sizes is {100,1000,10000}
def tol <- 1e-6
def max.it <- 1000

# Heavy-tailed mixture df

def phi <- c(0.5,1,1,2)
def pmix <- function(x,phi) {
    M <- 1-exp(-1*x/phi[2])
    H <- 1-(x+1)^(-1*phi[3])
    L <- 1-exp(-1*x/phi[4])
    return(phi[1]*M*H+(1-phi[1])*L)
}

# Suggested quantile function

def qmix <- function(q,phi,tol) {
    lower <- 0
    upper <- 1
    while(pmix(upper,phi) < q) upper <- 2*upper
    error <- upper-lower
    while (error > tol){
        midpoint <- (lower+upper)/2
        if (pmix(midpoint,phi) < q) lower <- midpoint
        else upper <- midpoint
        error <- upper-lower
    }
    return(((lower+upper)/2)
}
```r
set.seed(1126)

# We simulate a sample of size n from the heavy-tailed df
rmix <- function(n,phi,tol){
  u <- runif(n)
  m <- matrix(0,n,1)
  for (i in 1:n) m[i] <- qmix(u[i],phi,tol)
  return(m)
}

# 2. EM algorithm function
em.mix <- function(x,tol, max.it) {
  q <- (quantile(x,probs=0.3)+quantile(x,probs=.4))/2
  t <- 0.324
  fn <- function(alpha){
    sum((1/(x+1))^(alpha*t))/n-(1/(t+1))
  }
  w <- median(x)
  r <- quantile(x,probs=0.75)
  s <- uniroot(fn, lower=0, upper=2)$root
  start <- c(q,s)
  phi <- numeric(4)
  phi[1] <- 0.5
  phi[2] <- q
  phi[3] <- s
  phi[4] <- (w+r)/(3*log(2))
  mysum <- function(a) {sum(a[is.finite(a)])}
  f1 <- function(phi,x) {
```
t1 <- exp(-1*x/phi[2])*(1-(x+1)^(-1*phi[3]))/phi[2]
t2 <- phi[3]*(x+1)^(-1-phi[3])*(1-exp(-1*x/phi[2]))
return(t1+t2)
}

def2 <- function(phi,x) {
    exp(-1*x/phi[4])/phi[4]
}

loglik <- rep(NA, 1000)
loglik[1] <- 0
loglik[2] <- mysum(log(phi[1]*f1(phi,x)+(1-phi[1])*f2(phi,x)))
k <- 2
# loop
while(abs(loglik[k]-loglik[k-1])> tol*abs(loglik[k-1]) & k <= max.it) {
    # E step
tau1 <- (phi[1]*f1(phi,x))/((phi[1]*f1(phi,x))+((1-phi[1])*f2(phi,x)))
tau2 <- 1-tau1
    # M step
    phi[1] <- mysum(tau1)/n
    phi[4] <- mysum(tau2*x)/mysum(tau2)
    f <- function (phi, tau1,k){
        return(-mysum(tau1*log(f1(phi, x)))-mysum(tau2*log(f2(phi, x))))
    }
    z <- optim(start,f,lower=0.01,tau1=tau1,k=k, method="L-BFGS-B")$par
    phi <- c(phi[1],phi[2],phi[3],phi[4])
loglik[k+1] <- mysum(log(phi[1]*f1(phi,x)+(1-phi[1])*f2(phi,x)))
k <- k+1
}
phihat <- phi
return(phihat)
}
ests <- matrix(0,1000,4)
for(i in 1:1000) {
x <- rmix(n,phi,tol)
est[i,] <- em.mix(x,tol,max.it)
}
est.bias <- apply(est,2,mean) - c(0.5,1,1,2)
est.bias
est.se <- apply(est,2,sd)
est.se
est.mse <- est.bias^2 + est.se^2
est.mse

.3 Appendix C: Computer Codes for Real Data Application

1. Summary statistics, histogram, exploratory QQ plot and sample mean excess plot

dat<-read.csv("C:/Users/pkd20/Desktop/Excel/Remissiontimes.csv",header=F)
x <- dat[, 1]# Remission times in months
library(e1071)
skewness(x)
kurtosis(x)
summary(x)
hist(x, freq=F, main="Remission times in months")
library(fitdistrplus)
plotdist(x, histo = TRUE, demp = TRUE)
library(evir)
qplot(x)
meplot(x)

2. Fitting the proposed mixture

n <- length(x)
tol <- 1e-6
max.it <- 1000
phi <- numeric(4)
# 2. EM algorithm function
em.mix <- function(x, tol, max.it) {
  q <- median(x)
  r <- quantile(x, probs=0.75)
  s <- 1 + (1/median(x))
  start <- c(q, s)
  phi <- numeric(4)
  phi[1] <- 0.5
  phi[2] <- q
  phi[3] <- s
  phi[4] <- quantile(x, probs=0.10)
  mysum <- function(a) {sum(a[is.finite(a)])}
f1 <- function(phi, x) {
  t1 <- exp(-1*x/phi[2])/(1-(x+1)^(-1*phi[3]))/phi[2]
  t2 <- phi[3]*(x+1)^(-1-phi[3])*(1-exp(-1*x/phi[2]))
  return(t1+t2)
}

f2 <- function(phi, x) {
  exp(-1*x/phi[4])/phi[4]
}

loglik <- rep(NA, 1000)
loglik[1] <- 0
loglik[2] <- mysum(log(phi[1]*f1(phi, x)+ (1-phi[1])*f2(phi, x)))

k <- 2
# loop
while(abs(loglik[k]-loglik[k-1])> tol*abs(loglik[k-1]) & k<=max.it) {
  # E step
  tau1 <- phi[1]*f1(phi, x)/(phi[1]*f1(phi, x)+(1-phi[1])*f2(phi, x))
  tau2 <- 1-tau1
  # M step
  phi[1] <- mysum(tau1)/n
  phi[4] <- mysum(tau2*x)/mysum(tau2)
  f <- function(phi, tau1, k){
    return(-mysum(tau1*log(f1(phi, x)))-mysum(tau2*log(f2(phi, x))))
  }
  z <- optim(start, f, lower=0.01, tau1=tau1, k=k, method="L-BFGS-B")$par
phi <- c(phi[1], phi[2], phi[3], phi[4])
loglik[k+1] <- mysum(log(phi[1]*f1(psi, x) + (1-phi[1])*f2(phi, x)))
k <- k+1
}

phihat <- phi
return(phihat)
}

em.mix(x, tol, max.it)

3. Fitting various distributions

library(fitdistrplus)

# Fitting the Weibull distribution
fw <- fitdist(x, "weibull")
summary(fw)

# Fitting the gamma distribution
fg <- fitdist(x, "gamma", lower = 0.01, optim.method="L-BFGS-B")
summary(fg)

# Fitting the 2-parameter Pareto distribution
library(actuar)
s = 1 + (1/median(x))
fp <- fitdist(x, "pareto", start = list(shape=s, scale = mean(x)))
summary(fp)

# Fitting the log-normal distribution
fln <- fitdist(x, "lnorm")
summary(fln)

# Fitting the Lomax distribution (with scale parameter 1)
dl <- function(x, shape) shape*(x+1)^(-shape-1)
pl <- function(q, shape) 1-(q+1)^(-shape)
ql <- function(p,shape) ((1-p)^(-1/shape))-1
s=1+(1/median(x))
fl <- fitdist(x,"l",lower=0.01,optim.method="L-BFGS-B",list(shape=s))
summay(fl)
# Fitting the exponential distribution
de <- function(x, scale) exp(-x/scale)/scale
pe <- function(q, scale) 1-exp(-q/scale)
qe <- function(p, scale) -scale*log(1-p)
fe <- fitdist(x,"e",lower=0.001,optim.method="L-BFGS-B", list(scale=mean(x)))
summary(fe)
# Fitting a pure mixture of Lomax(scale=1) and exponential
n <- length(x)
tol <- 1e-6
max.it <- 1000
phi <- numeric(3)
# 2. EM algorithm function
em.mix <- function(x,tol, max.it) {
  r <- quantile(x,probs=0.75)
s <- 1 +(1/median(x))
  phi[2] <- s
phi[3] <- (r+exp(mean(log(x))))/(log(4)+exp(digamma(1)))
}
mysum <- function(a) {sum(a[is.finite(a)])}

f1 <- function(phi,x) {
  phi[2]*((x+1)^(-1-phi[2]))
}
f2 <- function(phi,x) {
  exp(-1*x/phi[3])/phi[3]
}

loglik<- rep(NA, 1000)
loglik[1]<- 0
loglik[2] <- mysum(log(phi[1]*f1(phi,x)+ (1-phi[1])*f2(phi,x)))
k <- 2
# loop
while(abs(loglik[k]-loglik[k-1])> tol*abs(loglik[k-1]) & k<=max.it) {
  # E step
  tau1 <- phi[1]*f1(phi,x)/(phi[1]*f1(phi,x)+(1-phi[1])*f2(phi,x))
  tau2 <- 1-tau1
  # M step
  phi[1] <- mysum(tau1)/n
  phi[2] <- mysum(tau1)/mysum(tau1*log(x+1))
  phi[3] <- mysum(tau2*x)/mysum(tau2)
  phi <- c(phi[1],phi[2], phi[3])
  loglik[k+1]<- mysum(log(phi[1]*f1(phi,x)+(1-phi[1])*f2(phi,x)))
  k <- k+1
  cat("k = ",k," loglik = ",loglik[k],"\n")
}

phihat <- phi
# Fitting a mixture of two gamma distributions
library(mixtools)
set.seed(1126)
mixmdl <- gammamixEM(x, epsilon=1e-06)
mixmdl[c("lambda", "gamma.pars","loglik")]

# Fitting a mixture of two normal distributions
library(mixtools)
set.seed(1126)
mixmdl <- normalmixEM(x, epsilon=1e-06)
mixmdl[c("lambda", "mu", "sigma", "loglik")]

# Fitting the maximum of two independent exponential and Lomax r.v.s
n <- length(x)
alpha <- numeric(2)
fn <- function(alpha){
  t1 <- log(alpha[2]*(x+1)^(-alpha[2]-1)*(1-exp(-1*x/alpha[1])))
  t2 <- exp(-1*x/alpha[1])*(1-(x+1)^(-alpha[2]))/alpha[1])
  return(-sum(t1+t2))
}
start <- c(mean(x), 1+(1/median(x)))
res <- optim(start, fn, gr=NULL, method="L-BFGS-B", lower=0.01, hessian=F)
res

4. Goodness of fit tests
# Kolmogorov-Smirnov test
theta<-7.539
rho<-1.141
fn<-function(x){
  (1-exp(-x/theta))*(1-(x+1)^(-rho))
}
ks.test(unique(x),"fn", exact=T)

# Anderson-Darling test
library(goftest)
ad.test(x, "fn")

# Cramer-von Mises test
library(goftest)
cvm.test(x, "fn")

# Chi-square test
library(zoo)
hs <- hist(x,breaks=nclass.Sturges(x),right=FALSE)
breaks_cdf <- fn(hs$breaks)
null.probs <- rollapply(breaks_cdf, 2, function(x) x[2]-x[1])
hs$counts

# Some of the expected frequencies are less than 5.
# We obtain 3 bins by combining the last 5 bins
# Observed frequency of the new bin
f3 <- sum(hs$counts)-hs$counts[1]-hs$counts[2]
# Expected proportion of observations in the new bin
p3 <- sum(null.probs)-null.probs[1]-null.probs[2]
# Chi-square test statistic and corresponding p-value
probs = c(0.686956262, 0.213779955, 0.09259549)

res <- chisq.test(c(90, 26, 12), p = probs, rescale.p = TRUE)

res