NEW STATISTICAL MODELS FOR EXTREME VALUES

A THESIS SUBMITTED TO THE UNIVERSITY OF MANCHESTER

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

IN THE FACULTY OF ENGINEERING AND PHYSICAL SCIENCES

2013

Sumaya Saleh Eljabri

School of Mathematics

Contents

III A New Family of Extreme Value Distributions 148

6 On Chen *et al.*'s Extreme Value Distribution 149

7 A New Distribution for Extreme Values 170

IV Conclusions and Future Work   189

Word count 22173

List of Tables

List of Figures

The University of Manchester

Sumaya Saleh Eljabri
Doctor of Philosophy
New Statistical models for Extreme Values
April 24, 2013

Extreme value theory (EVT) has wide applicability in several areas like hydrology, engineering, science and finance. Across the world, we can see the disruptive effects of flooding, due to heavy rains or storms. Many countries in the world are suffering from natural disasters like heavy rains, storms, floods, and also higher temperatures leading to desertification. One of the best known extraordinary natural disasters is the 1931 Huang He flood, which led to around 4 millions deaths in China; these were a series of floods between Jul and Nov in 1931 in the Huang He river.

Several publications are focused on how to find the best model for these events, and to predict the behaviour of these events. Normal, log-normal, Gumbel, Weibull, Pearson type, 4-parameter Kappa, Wakeby and GEV distributions are presented as statistical models for extreme events. However, GEV and GP distributions seem to be the most widely used models for extreme events. In spite of that, these models have been misused as models for extreme values in many areas.

The aim of this dissertation is to create new modifications of univariate extreme value models. The modifications developed in this dissertation are divided into two parts: in the first part, we make generalisations of GEV and GP, referred to as the Kumaraswamy GEV and Kumaraswamy GP distributions. The major benefit of these models is their ability to fit the skewed data better than other models.

The other idea in this study comes from Chen, which is presented in Proceedings of the International Conference on Computational Intelligence and Software Engineering, pp. 1-4. However, the cumulative and probability density functions for this distribution do not appear to be valid functions. The correction of this model is presented in chapter 6.

The major problem in extreme event models is the ability of the model to fit tails of data. In chapter 7, the idea of the Chen model with the correction is combined with the GEV distribution to introduce a new model for extreme values referred to as new extreme value (NEV) distribution. It seems to be more flexible than the GEV distribution.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns any copyright in it (the "Copyright") and s/he has given The University of Manchester the right to use such Copyright for any administrative, promotional, educational and/or teaching purposes.

ii. Copies of this thesis, either in full or in extracts, may be made **only** in accordance with the regulations of the John Rylands University Library of Manchester. Details of these regulations may be obtained from the Librarian. This page must form part of any such copies made.

iii. The ownership of any patents, designs, trade marks and any and all other intellectual property rights except for the Copyright (the "Intellectual Property Rights") and any reproductions of copyright works, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property Rights and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and exploitation of this thesis, the Copyright and any Intellectual Property Rights and/or Reproductions described in it may take place is available from the Head of the School of Mathematics.

Notation

Symbol	Description
M_n	The maximum value in the set of data
m_n	The minimum value in the set of data
$D(G)$	Domain of attraction of G
μ	Location parameter
σ	Scale parameter
ξ	Shape parameter
a	Shape parameter
b	Shape parameter
A^2	Anderson-Darling goodness of fit statistic
W^2	Cramér-von Mises goodness of fit statistic
$Bias$	Bias of the estimator
T	Return period in years
Γ	Gamma function
γ	Incomplete Gamma function
δ_1	The mean deviation about the mean
δ_2	The mean deviation about the median
B(., .)	Beta function
E[.]	Expectation of a random variable
$B_x(., .)$	Incomplete beta function
$\kappa_3(X)$	The skewness
$\kappa_4(X)$	The kurtosis
$\psi(x)$	Euler's psi function

$-\mathbf{J}$ The observed information matrix

$-E\mathbf{J}$ The Expected information matrix.

Abbreviations

GEV	Generalised Extreme Value Distribution
GP	Generalised Pareto Distribution
EV1	Gumbel Distribution
EV2	Frechét Distribution
MLE	Maximum Likelihood Estimation
EVT	Extreme Value Theory
MSe	Mean Square Error
SE	Standard Error
AIC	Akaike Information Criterion
CAIC	Consistent Akaike Information Criterion
AICc	Akaike Information Criterion with a Correction
BIC	Bayesian Information Criterion
Kum	Double Bounded Probability Density Function (Kumaraswamy Distribution)
pdf	Probability Density Function
cdf	Cumulative Distribution Function
hrf	Hazard Rate Function

Dedication

To my parents

with my pure Love.

Acknowledgements

Living far away from your family and friends for four years, would have been unbearable without the kindness of some people who made it less painful and more influential in my heart. Thanks to my friends and to anyone who put a smile on my face.

Before I attended the University of Manchester as a postgraduate student, I had some wonderful guides in the Department of Statistics in the University of Tripoli. Without their encouragement, I would not have been able to reach the level of a doctoral degree. Also, I would like to thank all my maths teachers at school who let me know that maths not just learning how to add up numbers. Really, I am a lucky to have been surrounded by great mathematicians during my life.

The huge support made available by the Libyan government from the financial support, encouragement and guidance made my studying period easier, so I would like to take this opportunity to say thanks to everyone who works there.

Part I

Introduction and Background

Chapter 1

Introduction

Climate change is considered to be one of the most pressing issues of the last two decades. It can be thought of as long-term changes in the statistical distribution of weather patterns. Changes in the climate pose one of the greatest threats to humans. Floods, due to maximum rainfall events and increases in global average sea levels, due to higher temperatures, will continue to have an enormous impact on our lives. Nowadays, the world is increasingly seeing extreme effects from rainfall events, so the accurate prediction for these events will significantly aid in policy planning. Studying and analysing rainfall events has a significant impact on hydraulic designs for flood protection such as flood barriers, floodgates, dams and dykes.

In section 1.1 we review the literature in rainfall and floods analysis. The motivation of this dissertation is declared next, in the section 1.2. Section 1.3 presents the structure of this research. Finally, section 1.4 presents related published papers and conferences.

1.1 The Historical Review of Rainfall and Floods Analysis

Extreme value models are widely used in several areas of scientific research, engineering, and medicine. The most common applications of extreme value distributions

are annual maximum rainfall events and floods. In this section, we review the background of annual maximum rainfall and floods data.

Around the world much research has been conducted on maximum rainfall distributions and most are focused on the identification of the best fitting probability distribution function (pdf) for the data. Models used include normal, log-normal, Gumbel, Weibull, Pearson type, 4-parameter Kappa, Wakeby, two-component GEV, GEV, and Gamma distributions. Since the first application of extreme values in flood flows, presented by Fuller (1914), several published works are focused on the applications of extreme values in hydrology such as floods, rainfall, earthquakes, etc. The applications of univariate extreme value models to maximum rainfall and floods data are applied to several regions around the world. As a result of the different topography (plain areas to mountainous zones), the records of maximum rainfall events are different and they will belong to different types of distributions. In this section, we review some of the applications of extreme value distributions to maximum rainfall and floods data from several places in the world.

Houghton (1977) introduced a new five-parameter distribution known as the Wakeby distribution. He clarified the advantages of this model over the lognormal distribution in fitting U.S flood records depending on several goodness of fit tests. Also, he showed the ability of the Wakeby distribution to explain the split effect,which is not evident in other known distributions. Haktanir and Horlacher (1993) examined nine different distributions that are commonly employed for flood frequency analysis. These models were applied to two datasets recorded from different locations, the first dataset is flood data collected from 11 streams in the Rhine Basin, Germany, while the other is from 2 streams in Scotland. Results for this study can be summarised into some main points, which are:

- The GEV and 3-parameter lognormal distributions can predict the flood with the return period of 100-years better than the others,

- For the maximum likelihood estimator, they found that the log-logistic distribution overestimated return period floods greatly,

- Sometimes the Frechét (EV2) and log-Pearson type 3 (with positively skewed) yield slightly conservative peaks.

Cannarozzo et al. (1995) proposed the two component extreme value distribution (TCEV) to model data from rainfall and floods. To explore the usefulness of this model, the authors used annual maximum rainfalls with several durations: 1, 3, 6, 12 and 24 hours which were recorded at 172 gauges from 1928 to 1981 with sample size n from 10 to 45 years. For the floods, they used annual maximum peak flood data collected from 27 stream gauges between 1936 and 1982. According to Karim and Chowdhury (1995), lognormal, Gumbel and log-Pearson type 3 distributions were disqualified as best models to annual maximum floods series in Bangladesh. Also, they explored the GEV distribution as being better than the others for flood frequency analysis. To assessment the return period of the peak discharges and changes in hydrological response due to the dam break, there are three different methods, which are: Annual maximum series extreme value analysis (AMS), partial duration series (PDS) and regional analysis. In order to compare these three methods, Hoybye and Iritz (1997) applied log-Pearson type III (LP3), Gumbel (EVD1), Pearson type III (P3), 3-parameter lognormal (LN3), and GEV distributions to a monsoon climate catchment in Hongru River, China. They found that the GEV distribution was the only model that accepted all catchments when they were tested within the AMS method.

The GP distribution has been found to be an appropriate distribution for flood frequency analysis (see El-Jabi et al. (1998)). They applied this model to a number of hydrometric stations in the province of New Brunswick, Canada. The 4-parameter Kappa distribution (by Parida (1999)) is the best-given model for the Indian summer monsoon. This model has been applied to the Indian summer monsoon rainfall (ISMR) data observed over a common period (June-September) during (1940-1980) at 50 gauging stations across India. The benefit of using the 4-parameter Kappa

distribution, as mentioned by the author, is the fact that it has sub-models such as GEV, GP, generalised logistic, exponential, Gumbel, logistic, uniform, and reverse exponential distributions.

Yue et al. (1999) introduced the Gumbel mixed model, the bivariate extreme value model with Gumbel marginals, to model the joint probability of dependent flood peaks and volume, as well as flood volume and durations. In this study, they applied this model to flood data from the Ashuapmushuan river, Quebec, Canada. Results show that this model is an appropriate model for both kinds of data. Koutsoyiannis and Baloutsos (2000) used three extreme distributions, which are: extreme value type I (EV1 or Gumbel), 2-parameter extreme value type II (EV2(2)), and the generalised extreme value (GEV) distributions to model the annual maximum series of daily rainfall depths in Athens during the period (1860-1995). The authors concluded that the GEV distribution is a more appropriate model than the others for long records of the annual maximum rainfall (136 years), whereas the EV1 seems to be a suitable model if fewer years of measurements are used (34 years of this sample were considered).

To clarify some of the characteristics of hydrologic extremes, Katz et al. (2002) treated two examples of hydrologic extremes; the first example was concerned with estimating the best model for the annual maximum of daily rainfall, while the second was concerned with estimating the model of the annual peak streamflow. While, the GEV distribution was fitted for the first example, the GP distribution was considered for the second. To detect the trend of the mean of the Pearson linear correlation coefficient and MannKendall tests, Crisci et al. (2002) collected rainfall data at durations of (1, 3, 6, 12 and 24 hours) at 81 rain gauges located in Tuscany. The reason for choosing a variety of rain gauges comes from the fact that the Tuscany area is characterized by an extreme variety of topography, from plain areas to mountainous and hilly zones. The authors fit the data to the GEV distribution to identify the areas that were affected by the heaviest rainfall at any storm duration. Park et al. (2001) used the Wakeby distribution (WAD) and estimated the parameters by the L-moments method (L-M) to fit the summer annual maximum daily and bi-daily rainfall data at 61 gauges in South Korea. They used different lengths of time series,

since the records began at different dates, but all of them ran until 1999.

Park and Jung (2002) applied the 4 parameter Kappa distribution (K4D) to fit the same data with maximum likelihood estimation (MLE) to estimate the 4 parameters. Koutsoyiannis (2004) compared two types of extreme value distributions, Gumbel (EV1), and Frechét (EV2), by applying them to a collection of 169 ganges of the available maximum rainfall records worldwide, with each record having 100-154 years of data. He showed that the (EV2) distribution is more appropriate than (EV1) for a longer-record of maximum rainfall data.

Li et al. (2005) analysed daily extreme rainfall at five geographical stations in the Southwest Western Australia (SWWA). They used the generalised Pareto (GP) distribution to model the extreme rainfall from daily rainfall data. The authors call a daily rainfall 'extreme' if it exceeds a given threshold. Nadarajah (2005) applied the GEV distribution to the annual maxima of daily rainfall data in the period 1901-2003 at 14 locations in West Central Florida. To find the best model to fit flood events of the Pachang river, Taiwan, Nadarajah and Shiau (2005) employed the Gumbel and GEV distributions as models of extreme values. They showed that the Gumbel distribution can be considered over the GEV distribution for both flood volume and flood peak. Feng et al. (2007) analysed four-time durations: daily, 2 days, 5 days and 10 days for annual maximum rainfall from 651 weather stations in China over the period 1951-2000. They used the GEV distribution to model this type of data. Also, they modified the GEV distribution to model linear trends in extreme values. According to Hanson and Vogel (2008), the Pearson Type-III (P3) distribution makes a better fit for the full record of daily rainfall, with length 24,657 days at 237 stations in the U.S. The Kappa (KAP) distribution has been considered by Hanson and Vogel (2008) as the best model to describe only-wet-days daily rainfall data, which was constructed from the above data by eliminating zero and trace values.

The Hawaiian Islands have frequently experienced heavy rainfall, and floods, a cause for a lot of damages to agriculture and properties as well as social problems due to the effect on tourism. Chu et al. (2009) applied the method of L-moments to fit the 3-parameter GEV model to 20 years of maximum rainfall records up to 2005 from

158 stations in the Hawaiian Islands. Ranchi, in India, has a subtropical climate in the summer (March to June) with the temperature between $20 - 37°C$, while the winter temperature is between $2 - 22°C$. Temperatures are quite different in both seasons, though rainfall is very little in both summer and winter except for extreme maximum rainfall in the monsoon season (July to September) with a daily average of 1,100 mms. Shukla et al. (2012) clarify that the GEV distribution can be considered as the best model to fit the subtropical monsoon region in India by applying this model to 51 years of the maximum rainfall data during 1956-2006. This distribution was compared with the Gumbel, Frechet and Weibull distributions.

Panthou et al. (2012) compared two approaches for spatial estimation, a local-fit and interpolation (LFI) as well as spatial maximum likelihood estimation (SMLE). They applied five LFI and three SMLE methods to the GEV distribution on the 126 daily rainfall series covering the period 1950-1990 in Sub-Saharan West Africa. The maximum daily rainfall data from four different stations in Pakistan: Islamabad, Murree, Lahore and Sialkot between 1954 and 2005 were studied by Abbas et al. (2012) to find the appropriate model to fit the data. They compared the gamma, GEV and GP distributions, and concluded that GEV and GP distributions are more suitable for the annual maximum rainfall data in several places in Pakistan, more so than the gamma distribution.

1.2 Motivation

Nowadays, the generalised extreme value distribution has become one of the most widely applied distributions in univariate extreme value theory. It has several applications covering most areas of research in science, engineering and medicine such as floods, wind speed, annual maximum rainfall, earthquake records, high level ozone concentrations, extreme maximum and minimum temperatures, etc. For more applications of generalised extreme value distribution in engineering problems we refer the reader to Castillo et al. (2005) and Beirlant et al. (1996). However, the GEV distribution has been misused in too many areas, as can be seen from the list given

in Chapter 4. Consider the following problems:

a. Many hydrological engineering planning, design, and management problems require a detailed knowledge of flood event characteristics. Flood frequency analysis often uses the GEV distribution to model flood peak values, which provides an assessment of flood events.

b. Corrosion science has been based mainly upon deterministic approaches, particularly the electrochemical theory of corrosion. Localized corrosion, however, cannot be explained without statistical and stochastic approaches because of the large scatter in data that is common in the laboratory and the field. The GEV distribution has been used in many successful applications of statistical approaches to localized corrosion in engineering data, to estimate the maximum pit depth that would be found in a large-area installation by using a small number of samples with a small area.

c. In the time series of extreme dynamic pressures (i.e. of the squares of extreme wind speeds), the GEV distribution has also been shown to present good fits to this type of data.

d. Each of the problems above is concerned with the tail behavior of one or more variables. So, by capturing the tail behavior more accurately, one could obtain improved estimation and prediction.

Our proposed models, which we present in Chapters 4, 5, 6 and 7 provide one way of doing this. We create two new modifications for the GEV distribution: the first one depends on the double-bounded probability density function (Kumaraswamy distribution). While the idea of the second comes from the most asked question in extreme value models "How can we control the thickness of shapes of extreme value models". Then we apply these two models to annual maximum rainfall data to show the flexibility and advantages of these models over the ones that are commonly used to describe the behaviour of extremes. Another modification is given later, but it depends on the excess over the threshold model (generalised Pareto distribution). As

mentioned before, the Pareto distribution is used to model large loss data in insurance due to the fact that insurance payment data is positively skewed with large upper tails. The weakness of using a Pareto model for insurance claims is that even though it covers the large losses well, it fails to cover the behaviour of the small ones.

In Chapter 5, we combine two known models: the GP distribution and the Kum distribution to create a new model with the aim of attracting wider applicability in insurance claims. The main motivation in this thesis is to create new models for univariate extreme values, which can be considered instead of the known distributions.

1.3   Outline

In this thesis we focus on developing the current extreme models (GEV and GP distributions).

This work is organized as follows:

Chapter 2 can be regarded as a review chapter of extreme value modeling. The first part of this chapter starts with the classical extreme value theory and Pickands Balkema de Haan theorem. The two major distributions in extreme value theory, the GEV and GP distributions are given and followed by some applications of extreme value theory in some fields of research such as ocean engineering, finance and insurance, structural engineering, etc. The last part is dedicated to discussing the graphical tools for data analysis and estimation methods, which are used in subsequent chapters.

Chapter 3 covers the double-bounded probability density function. In the beginning, we start with the introduction of the Kumaraswamy (DB-PDF) distribution and present its pdf and cdf. The next section presents the variety of shapes that the Kum distribution has, to illustrate the flexibility of this model, including some examples on applying it to real data. Then, we discuss main properties of the Kum distribution. The relation with other known distributions and some examples of applications are given in the last parts of this chapter. Finally, an overview is given as a summary of the Kumaraswamy distribution.

Chapters 4 and 5 introduce the univariate Kum-G distributions as modified models of the extreme value distributions. In Chapter 4 a new distribution - referred to as the Kumaraswamy Generalised Extreme Value (KumGEV) distribution is introduced. Some mathematical properties of this distribution are studied. We derive analytical shapes of the density and hazard rate functions and review some sub-models of the KumGEV distribution depending on the shape parameters. We calculate explicit closed form expressions for moments and the moment generating function. Skewness and Kurtosis are also examined for the distribution. We estimate its parameters by the method of maximum likelihood and provide the observed information matrix. To illustrate the potential of the new model, we apply this model to daily rainfall maxima in millimetres from 1938 to 1972 at Uccle, Belgium. The Rainfall data is contained in the evd package (Stephenson (2002)) in the R programme. Finally, some bivariate generalizations of the model are proposed.

In Chapter 5, the same approach used in chapter 4 to modify the GEV distribution is used here to generate a new distribution, the so-called Kumaraswamy GP distribution, which includes some sub-models that we present. The mathematical properties of this distribution are studied. We derive moments, the moment generating function and mean deviations. Two measures of entropy are derived. Maximum likelihood estimation is used to estimate the parameters including the information matrix. To illustrate the benefit of this distribution over the others, we apply KumGP distribution to dataset consisting of 154 exceedances of the threshold $65m^3s^{-1}$ by the River Nidd at Hunsingore Weir from 1934 to 1969. This data is taken from NERC (1975). In this chapter, we used the **evir** Pfaff et al. (2010) and Ribatet (2009) packages in the R programme.

Chen et al. (2010) claim to have proposed a new extreme value distribution, but the formulas given for the distribution do not form a valid probability distribution. In Chapter 6, we correct their formulas to form a valid probability distribution. For this valid distribution, we provide a comprehensive treatment of mathematical properties, estimate parameters by the method of maximum likelihood and provide the observed information matrix. The flexibility of the distribution is illustrated using a real data

set.

In Chapter 7, a new family of distributions is introduced to model univariate extreme values. We extend the GEV model by adding one shape parameter to control the tail and the mode. Some known distributions are presented as related distributions to the new model. The statistical properties of this model are derived with details of the variety of shapes. The parameter estimation by the method of maximum likelihood is given. Finally, to illustrate the flexibility of this model, we use the same data used in chapter 4, which is the annual maximum rainfall data in Uccle, Belgium.

The last part of this research is focused on the main results that we obtained from this work. Chapter 8 concludes the work in this thesis and reviews some thoughts that can be considered as future work for each model.

1.4 Thesis Related Publications and Papers

The main result in Chapter 4 has been presented in the Young Researcher Meeting in **Bristol April/2012** and in Mathematics Research Students' Conference (MRSc) **Manchester 28th September/2012**. The works in Chapters 4 and 5 are currently under consideration for the *Extremes* journal. The main result in Chapter 6 is currently under consideration for *Statistics: A Journal of Theoretical and Applied Statistics*.

Chapter 2

Extreme Value Theory

2.1   Introduction

Extreme value theory (EVT) is a very important theory in probability and statistics devoted to study of the behavior of extreme values. Even though these values have a very low chance to appear, they can turn out to have a very high impact to the observed system. Finance and insurance are the best fields of research to observe the importance of extreme events. EVT can be considered as a developing area of research. It has been started in the last century as an equivalent theory to the central limit theory, which is dedicated to study of the asymptotic distribution of the average of a sequence. The EVT focuses on the behavior of block maxima or minima. The extreme value theory was introduced first by M.Fréchet (1927) and Fisher and Tippett (1928) then followed by Von Mises (1936), which is translated and reprinted in Von Mises (1964), and completed by Gnedenko (1943), which is translated and reprinted in Johnson (1992). The latter gave the Extremal Types Theorem. All of the following research focused on finding the limiting behavior of the maxima of sequence of iid random variables.

In this Chapter, we present a general review of the main theories in univariate extreme value analysis (Sections 2.3 and 2.4). The rest of this Chapter concentrates on extreme value models and their applications in hydrology, sciences, engineering and medicine. We then move to graphical tools for data analysis, maximum likelihood

estimation method, goodness of fit tests and information criteria.

2.2 Classic Extreme Value Theory

The central limit theorem states that the sum and the mean of an arbitrary finite distribution are normally distributed under the condition that the sample size is sufficiently large. However, in some practical studies we are looking for the limiting distribution of maximum or minimum values rather than the average of the data. Assume that $X_1, X_2, \ldots, X_n$ is a sequence of *iid* random variables distributed with cdf denote F. One of the most interesting statistics in a research is the sample maximum $M_n = max\{X_1, X_2, \ldots, X_n\}$. This theory studied the behaviour of M_n as the sample size n increases to infinity.

$$
\begin{aligned}
P_r\{M_n \leq x\} &= P_r\{X_1 \leq x, X_2 \leq x, \ldots, X_n \leq x, \} \\
&= P_r\{X_1 \leq x\}P_r\{X_2 \leq x\}\ldots\ldots P_r\{X_n \leq x\} \\
&= F^n(x).
\end{aligned}
$$

The concept of finding a limiting distribution for the block maxima is similar to the motivation of the central limit theorem, where the unknown distribution of sums leads to the normal distribution (Beirlant et al. (1996)). The distribution function F cannot be found practically, but the Fisher-Tippet-Gnedenko theorem provides the asymptotic result.

Theorem 2.1. *Suppose there are sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that:*

$$
Pr\left\{\frac{(M_n - b_n)}{a_n} \leq x\right\} \to G(x) \qquad as \qquad n \to \infty. \tag{2.1}
$$

Then if G is a non-degenerate *distribution function then it will belong to one of the three following fundamental types of Extreme value family:*

*Type I (**Gumbel**):*

$$
F(x) = \exp\{-e^{-(x-\mu)/\sigma}\}, \qquad x \in \mathbb{R}
$$

Type II **(Frechét):**

$$F(x) = \begin{cases} 0, & x \le 0, \\ \exp\{-((x-\mu)/\sigma)^{-\alpha}\}, & x > \mu, \alpha > 0 \end{cases}$$

Type III **(Weibull):**

$$F(x) = \begin{cases} 1, & x > 0, \\ \exp\{-(-(x-\mu)/\sigma)^{\alpha}\}, & x < \mu, \alpha > 0 \end{cases}$$

for the parameters $\sigma > 0, \quad -\infty < \mu < \infty$.

2.2.1 Domain of Attraction

Definition 2.1. The distribution function F is in the domain of attraction of an extreme value distribution if the relation (2.1) is satisfied. We write $F \in D(G)$.

2.2.2 Max-Stable Distributions

Definition 2.2. A non-generate distribution function G is max-stable distribution if for each $n = 2, 3, 4, \ldots$ there are a constant $a_n > 0$ and b_n such that

$$G^n(a_n x + b_n) = G(x) \qquad as \qquad n \to \infty. \tag{2.2}$$

Now, the following theorem combines the definition of the max-stable distribution with the limiting distribution of extreme values.

Theorem 2.2. *Suppose G is a nondegenerate distribution function. Then It is a max-stable distribution if and only if there is a sequence of distribution functions $\{F_n\}$ and constants $a_n > 0$ and b_n such that*

$$F_n\left(a_{nk}^{-1}x + b_{nk}\right) = G^{1/k}(x) \qquad as \qquad n \to \infty \tag{2.3}$$

for all $k = 1, 2, 3, \ldots$.

The proof of this theorem can be found in Leadbetter et al. (1983).

2.3   Pickands Balkema De Haan Theorem

Let $X_1, X_2, \ldots X_n$ be a sequence of iid random variables with distribution function $F(x)$. We are interested in estimating the distribution function $F(u)$ of extreme values that exceed a threshold u. The distribution function is given by:

$$Pr\{X > u + y / X > u\} = \frac{1 - F(u + y)}{1 - F(u)}, \ y > 0. \tag{2.4}$$

The problem of values exceeding a certain threshold may be important to several problems such as a harbor being damaged if a flood exceeds a certain level, or the glass of oven breaking under high temperature. For more applications we recommend Castillo et al. (2005). Pickands III (1975), Balkema and De Haan (1974) stated the limiting distribution of extreme values that exceed a threshold u, which is presented in the following theorem.

Theorem 2.3. *Let $X_1, X_2, \ldots$ be a sequence of iid random variables with distribution function, F. a large class of underlying distribution functions F and large enough u, the liming distribution of $F(u)$ is*

$$H(y) = 1 - (1 + \frac{\xi y}{\tilde{\sigma}})^{-1/\xi} \tag{2.5}$$

where $1 + \frac{\xi y}{\tilde{\sigma}} > 0$ and $\tilde{\sigma} = \sigma + \xi(u - \mu)$.

The family of distributions in Eq(2.5) is called the Generalised Pareto distribution.

2.4   Univariate Extreme Value Modeling

Extreme value theory (EVT) has been used to develop two classes of extreme value distributions. The first class is the asymptotic distribution of a sequence of maximum or minimum values known as one of the three following distributions: Gumbel, Frechét and Weibull distributions (Coles (2001)). These distributions can be represented as members of the Generalised extreme value (GEV) distribution. The second class is the distribution of all exceedances over a high threshold, which is called the Generalised Pareto distribution.

2.4.1 Generalised Extreme Value Distribution

In practice, it is very difficult to choose which of the three families (Gumbel, Frechét and Weibull) is the most appropriate for real data. Therefore, a better analysis of block maxima (minima) is offered by combining these distributions into a single family of models called the generalised extreme value (GEV) distribution. Jenkinson (1955) proposed a formula for (GEV) distribution.

The *cdf* of the generalised extreme value (GEV) distribution is defined as:

$$
F(x, \mu, \sigma, \xi) = \begin{cases} \exp[-[1 + \xi(x - \mu)/\sigma]^{-1/\xi}] & \text{if} \quad \xi \neq 0, \\ \exp[-exp(-(x - \mu)/\sigma)] & \text{if} \quad \xi = 0. \end{cases}
$$

where $\sigma > 0$ and $-\infty < \mu < \infty$ are the scale and location parameters respectively. The other parameter ξ is the shape parameter, representing the behavior of the tail. Sub-models can be defined by $\xi \to 0$, $\xi > 0$, and $\xi < 0$, corresponding respectively to the Gumbel, Frechét and Weibull distributions, which are mentioned above.

The density function of the GEV distribution is:

$$
f(x, \mu, \sigma, \xi) \;=\; \sigma^{-1}[1 + \xi(x - \mu)/\sigma]^{-1/\xi - 1} \exp[-[1 + \xi(x - \mu)/\sigma]^{-1/\xi}]. \quad (2.6)
$$

2.4.2 Generalised Pareto Distribution

As mentioned above, the GEV distribution is the asymptotic distribution for the maxima or minima. Suppose we are interested not only in the maxima of observations, but also in the behavior of large observations that exceed a high threshold. The generalised Pareto (GP) distribution was introduced by Pickands III (1975), and Balkema and De Haan (1974).

We assume that the excess variables $X_1, X_2, \ldots, X_n$ are iid, then the definition of the GP distribution can be define as:

Definition 2.1. (del Castillo and Daoudi (2009)) The distribution and probability

density function of the GP distribution can be written as

$$
F_{(\sigma,\xi)}(x) = \begin{cases} 1 - [1 + \xi\ x/\sigma]^{-1/\xi} & \text{if} \quad \xi \neq 0, \\[2mm] 1 - \exp{-x/\sigma} & \text{if} \quad \xi = 0. \end{cases}
$$

and

$$
f_{(\sigma,\xi)}(x) = \begin{cases} \sigma^{-1}[1 + \xi\ x/\sigma]^{-(1+\xi)/\xi} & \text{if} \quad \xi \neq 0, \\[2mm] \sigma^{-1}\exp{-x/\sigma} & \text{if} \quad \xi = 0. \end{cases}
$$

for $x \geq 0$, when $\xi \geq 0$, or $x \geq 0$, and $x \leq -\frac{\sigma}{\xi}$ when $\xi < 0$, where $\sigma > 0$ is the scale parameter and $\xi \in \mathbb{R}$ is the shape parameter.

The shape parameter of the GP distribution is dominant in determining the qualitative behavior of the tail. So, the following values of the parameter ξ are of interest:

- When $\xi \to 0$, the GP distribution converges to the exponential distribution with mean σ.

- When $\xi = -1$, the GP distribution becomes the uniform distribution $U(0, \sigma)$.

- When $\xi = 1/2$, the GP distribution becomes the triangular distribution.

- The Pareto distribution is obtained when $\xi > 0$.

- When $\xi \leq -\frac{1}{2}, var(X) = \infty$. The rth central moment exists only if $\xi > -1/r$.

2.5   Applications of Extreme Value Theory

In many statistical applications, the interest is centered on estimating some population parameters. For example, the average temperature, the median income, the average rainfall, etc; based on samples taken from the same population. Sometimes, the most important values are not the average, but the maximum or minimum values (see Castillo (1994), Weibull et al. (1951), and Galambos (1987)). For example, the maximum flood height, maximum earthquake intensity, largest wildfire, the amounts of large insurance losses, etc. Largest values, such as loads, earthquakes, winds,

floods, waves, etc., and the smallest values, such as strength, stress, etc. are the key to failure of engineering works, so construction engineering should be based on extremes. Some publications related to extremes from fields such as ocean engineering, structural engineering, material strength, fatigue strength, etc., can be found in Leadbetter et al. (1983), Ferro and Segers (2003), Court (1953), Battjes (1978), Borgman (1963, 1970, 1973). Bretschneider (1959), Bryant (1983), Castillo and Sarabia (1992, 1994), Cavanie et al. (1976), Chakrabarti and Cooley (1977), Draper (1963), Earle et al. (1974), Goodknight and Russell (1963), Günbak (1978), Houmb and Overvik (1977), Longuet-Higgins (1952, 1975), Onorato et al. (2002), Putz (1952), Sellars (1975), Davenport (1963, 1967, 1972, 1978), Grigoriu (1984), Hasofer (1972), Hasofer and Sharpe (1969), Mistéth (1973), and Moses (1974), etc.

In this section we review some applications of the two most widely used distributions for univariate extreme values. Possible applications of these models cover most areas of science, engineering and medicine. Here, we brief some published applications such as:

a. In hydrology, the GEV and GP models are applied to flood frequency distributions, wind speed distributions, regional analysis of annual maximum rainfall, analysis of extreme floods, downscaling of future rainfall extreme events, analysis of regional earthquake records, analysis of daily discharge records, analysis of high level ozone concentrations, analysis of extreme maximum and minimum temperatures, the conductor icing distribution, analysis of air pollution data, mapping snow depth return levels, rainfall temporal patterns for urban drainage design, analysis of ocean climate,and extreme wave climate variability.

b. In engineering, there are some applications for these models such as pressure extreme in the energy-dissipating structure, assessment of corrosion-based failure in stainless steel containers, study of friction stir welded copper canisters, analysis of mobile networks, analysis of extreme loads for design of wind turbine components, and electricity price modeling.

c. In science, the GEV and GP models are applied for the spatial prediction of

soil properties, analysis of particle tracking data, estimation of depth-duration-frequency curves, fisheries research, volume fluctuations in a confined one-dimensional gas, the role of attractive methane-water interactions, and extreme electron fluxes.

d. In economics, the GEV and GP models are applied for retail and wholesale market power, risk management, analysis of risk dependence for foreign exchange data, value-at-risk for financial returns, and the estimation of interest-rate volatility.

e. In medicine, they applied to the testing multiple gene interactions, characteristics of Alzheimer's disease, modeling for controlled drug delivery systems.

For more details on the GEV distribution, its theory and further applications, we refer readers to Leadbetter et al. (1983), Embrechts et al. (1997), Castillo et al. (2005), and Resnick (2007).

2.6 Graphical Tools for Data Analysis

In extreme value analysis, to treat real data: first, we have to use graphics that will illustrate clearly the features of the data. In this section, we present some common graphs used in later chapters such as P-P plots, Q-Q plots, return level plots and mean residual life plots. In the rest of this dissertation, they will help us to decide which model will fit a certain kind of data better than others.

2.6.1 Probability-Probability Plots

Let $x_1, ..., x_n$ be a random sample from the cdf F and let $\hat{F}$ be the estimated cdf. Then a plot of

$$\hat{F}(x_{i:n}) \quad \text{against} \quad p_{i:n}; \quad i = 1, 2, ..., n,$$

is called the P-P plot, where $x_{i:n}$ is the ith order statistic and $p_{i:n}$ is the plotting position, which is defined as

$$p_{i:n} = \frac{i - \alpha}{n + \beta}; \quad i = 1, 2, ..., n.$$

Here α, $\beta \geq 0$ can be chosen empirically based on the behavior of data, the type of distribution, the estimation method used to estimate the parameters, etc. We choose $\alpha = 0.375$ and $\beta = 0.25$. If the model fit the data well, then the pattern of points will be very close to the 45-degree line.

2.6.2 Quantile-Quantile Plots

Let $\hat{F(x)}$ be the estimate of the distribution function F. The quantile-quantile plot is similar to the probability-probability plot. We plot the estimate of the inverse cdf

$$\hat{F}^{-1}(p_{i:n}) \quad \text{versus} \quad x_{i:n}; \quad i = 1, 2, ..., n.$$

The model will fit the data well if the points of the scatter plot are very close to the 45-degree line.

2.6.3 Return Level Plots

The return level can be defined as the level which is expected to be exceeded once every $1/p$ period, which is known as **a return period**. Then the return level, say x_T, exceeded on average once in T years can be written as

$$
\begin{aligned}
T &= \frac{1}{P} \\
P(X > x) &= 1 - F(x) = \frac{1}{T}, \\
\therefore x_T &= F^{-1}(1 - \frac{1}{T}),
\end{aligned}
\tag{2.7}
$$

which is given by the quantile function F^{-1}. The return level is very important to determine, for example, the heights of the sea walls, water dams, etc.

2.6.4 Mean Residual Life Plots

The empirical mean residual life plot is the locus of points

$$\left(u, \frac{1}{n_u} \sum_{i=1}^{n_u} (x_i - u)\right)$$

where $x_{(1)}, ..., x_{(n_u)}$ are the n_u observations that exceed the threshold u. If the exceedances of a threshold u_0 follow the generalised Pareto distribution, the empirical mean residual life plot should be approximately linear for $u > u_0$.

2.7 Maximum Likelihood Estimation

Assume that, we have $X_1, X_2, ..., X_n$ a random sample with pdf $f(x, \Theta)$, where $\Theta = (\theta_1, \theta_2, ..., \theta_k)$. Then the joint probability density function can be written as

$$f(x/\Theta) = \prod_{i=1}^{n} f(x_i; \Theta). \tag{2.8}$$

After the random sample is collected, the joint pdf will become a function of Θ and this function is known as a likelihood function, denoted by $L(\Theta)$

$$L(\Theta/x) = \prod_{i=1}^{n} f(x_i; \Theta) \tag{2.9}$$

The log-likelihood function is the equivalent formula of the likelihood function and is given by

$$l(\Theta/x) = \sum_{i=1}^{n} \log f(x_i; \Theta). \tag{2.10}$$

Then the estimator $\hat{\Theta}$ are the values of the Θ that maximise the likelihood function (or the log-likelihood function) with respect to Θ and are obtained by solving the system of equations

$$\frac{\partial l(\hat{(\Theta)}/x)}{\partial \theta_i} = 0, \quad i = 1, 2, ...k \tag{2.11}$$

2.8 Goodness of Fit

A 'goodness-of-fit' test is a method used to determine whether a random sample $X_1, . . . , X_n$, with size n, came from a certain distribution. Three goodness of fit tests: likelihood ratio test (LRT), Cramér-von Mises criterion and Anderson-Darling test are discussed in this section. All of these tests are used in Chapters 4, 5, 6 and 7 to fit models to real data.

2.8.1   The Likelihood Ratio Test

The likelihood ratio test is used to clarify how well a model fits a certain dataset. When we use this test to compare two models, they should be nested. The principle for this test is very simple to follow. Suppose that X has a pdf denoted by $f(x, \theta)$ with unknown parameter θ. We interested to test the hypothesis $H_0 : \theta$ is in Θ_1 versus $H_1 : \theta$ is in Θ_2, where Θ_1 and Θ_2 are the parameters for models 1 and 2, respectively. The log-likelihood ratio statistic denoted by LRT can be written as

$$LRT = -2 \ln \left(\frac{L_1(\widehat{\theta})}{L_2(\widehat{\theta})} \right) \qquad (2.12)$$

where L_1 and L_2 are the likelihood functions for models 1 and 2, respectively. Model 1 has fewer parameters than model 2. The log-likelihood ratio statistic (LRT) is distributed asymptotically as a chi-square Rv with degrees of freedom equal to the difference between the number of free parameters of the two models. We prefer model 1 if $LRT > \chi^2_{0.95, p_1 - p_2}$, where p_1 and p_2 are the free parameters.

2.8.2   Cramér-Von Mises Criterion and Anderson-Darling Tests

The Cramér-von Mises criterion and Anderson-Darling tests are common tests used to test if a random sample $x_1, x_2, ..., x_n$ came from a specific distribution. The former was proposed in (1928-1930) by Cramér and von Mises, while the latter was introduced by Anderson and Darling (1952). They can be treated as modifications of the Kolmogorov-Smirnov (K-S) test, but they give more weight to the tails than the Kolmogorov-Smirnov test does. These tests depend on the specific distribution to calculate the critical values. This point can be considered as an advantage in allowing a more sensitive test, or a disadvantage that critical values must be calculated for each distribution.

From now on, we denote the Cramér-von Mises and Anderson-Darling tests by A-D and C-VM, respectively.

Let $F(x; \theta)$ be the cdfs and W^* and A^* are the Cramér-von Mises and Anderson-Darling test statistics, respectively. We follow this procedure:

- Calculate the distribution function $v_i = F(x_i, \theta)$, where the data are in ascending order and then compute $y_i = \Phi^{-1}(v_i)$, where $\Phi^{-1}(.)$ is the quantile function of the standard normal $N(0, 1)$.

- Compute the standard values of y_i, which can be obtained as

$$u_i = \{(y_i - \bar{y})/s_y\}.$$

  where $(\bar{y})$ is the mean of y_i, $i = 1, 2, ..., n$ and s_y is the standard deviation of y_i.

- Then the C-VM and A-D test statistics can be written as

$$W^2 = \sum_{i=1}^{n} \left\{ u_i - \frac{(2i-1)}{2n} \right\}^2 + \frac{1}{12n},$$

  and

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^{n} \left\{ (2i-1) \log(u_i) + (2n+1-2i) \log(1-u_i) \right\}.$$

  where the logarithm is the natural logarithm.

- To compare these models we use the modified version of the statistics, W^* and A^*, which can be written as

$$W^* = W^2 \left(1 + \frac{0.5}{n} \right),$$

  and

$$A^* = A^2 \left(1 + \frac{0.75}{n} + \frac{2.25}{n^2} \right).$$

When comparing between two models by these tests, we prefer the model with the smaller values of these statistics.

2.9 Information Criteria

Undoubtedly, the construction of statistical models is strongly dependent on the results of theoretical analysis of the observed data. Sometimes there is a big gap

between the theoretical results and practical procedures. In the analysis of a given set of data, the selection of the perfect model to describe the data is still a considerable issue. In the previous section, we presented some goodness of fit tests that are used to compare which model is a better fit to data, but if these models are not nested, then we have to use other methods. One of these methods is information criteria. Here, we present the most widely used information criteria: AIC, AICc, BIC and CAIC, which are used in Chapters 4, 5, 6 and 7 in this dissertation.

2.9.1   Akaike Information Criterion

Akaike Information Criterion was first introduced by Akaike (1973) and developed further in Akaike (1974). It is the most widely used model selection tool among researchers. To apply AIC, we start with some optional models, which are regarded as proper models for certain data. The Akaike information criterion can be calculated from this formula:

$$AIC = -2 \log L(\hat{\theta}) + 2k \tag{2.13}$$

where k is the number of estimated parameters for the model. Thereafter, we can say that the proper model to fit the data is the one with the minimum value of AIC compared to others.

For a large sample, it is possible to prove that AIC introduces good model selections, see Shawky and Abu-Zinadah (2008). Nevertheless, the bias seems to be a basic issue of study (Findley (1985)).

AICc was proposed by Sugiura (1978) with a view towards bias reduction. Thereafter, Hurvich and Tsai (1989) proved that it improved model selections also in small samples.

If the sample size of the data n is small, $n/k < 40$, or the model has a large number of parameters, then we prefer the AICc correction:

$$AICc = AIC + \frac{2k(k+1)}{n-k-1} \tag{2.14}$$

AICc is suitable for strongly skewed models.

2.9.2 Consistent Akaike Information Criterion

Bozdogan (1987) provided the analytical extension of AIC without harming the basic principles of the AIC. It is called consistent Akaike Information Criterion (CAIC). The formula of CAIC can be written as

$$CAIC = -2\log L(\hat{\theta}) + k\left[\log(n) + 1\right]. \tag{2.15}$$

2.9.3 Bayesian Information Criterion

Schwarz (1978) developed a new criterion, known as Schwarz Information Criterion (SIC) or Bayesian Information Criterion (BIC). The main idea of BIC comes from approximating the Bayes factor with the assumption that the data is independent and identically distributed. Even though this criteria is derived within the Bayesian framework, unlike AIC, its formula is very close to the Akaike information criterion (AIC). It can be defined as

$$BIC = -2\log L(\hat{\theta}) + k\log(n). \tag{2.16}$$

where n is a sample size and $\log L(\hat{\theta})$ is the natural logarithm of the likelihood function. Like AIC, the appropriate model for a certain data is the one with minimum BIC compared to others.

Chapter 3

The Kumaraswamy distribution

3.1 Introduction

For hydrological random variables, such as daily rainfall, daily stream flow, etc, both classical probability and empirical distributions do not faithfully fit the data. However, this problem appears in other random processes when the element values for these processes are bounded both at the lower and upper ends. In these problems, the main measures such as mean, variance, skewness and kurtosis cannot be calculated exactly. In order to find an appropriate model for this type of data, Kumaraswamy (1980) proposed a new pdf, known as the Double-bounded probability density function (Kumaraswamy distribution).

Nowadays, the Kumaraswamy distribution is widely used in the field of hydrology to determine the most probable rainfall (see Sundar and Subbiah (1989), Fletcher and Ponnambalam (1996), Seifi et al. (2000) and Ganji et al. (2006)). Kumaraswamy's double bounded distribution is a family of continuous probability distributions defined on an interval (c, d). It is denoted by $Kum(\alpha, \beta, c, d)$.

The distribution and probability density functions are given by:

$$G_{Kum}(z, \alpha, \beta, c, d) = 1 - \left[1 - \left(\frac{z - c}{d - c}\right)^{\alpha}\right]^{\beta}; \qquad x \in (c, d) \tag{3.1}$$

$$g_{Kum}(z, \alpha, \beta, c, d) = \frac{1}{(d - c)}\alpha\ \beta\ \left(\frac{z - c}{d - c}\right)^{\alpha - 1}\left[1 - \left(\frac{z - c}{d - c}\right)^{\alpha}\right]^{\beta - 1}; \ x \in (c, d) \tag{3.2}$$

where $\alpha > 0$ and $\beta > 0$ are the shape parameters and c and d are boundary parameters. The standard form of the probability density function of the Kumaraswamy distribution can be obtained by using the linear transformation $X = (z - c) / (d - c)$. So, the variable X having the standard Kumaraswamy distribution $Kum(\alpha, \beta)$ has the cdf and pdf defined as:

$$G_{Kum}(x, \alpha, \beta) = 1 - (1 - x^{\alpha})^{\beta}; \qquad 0 < x < 1. \tag{3.3}$$

$$g_{Kum}(x, \alpha, \beta) = \alpha \; \beta \; x^{\alpha-1}(1 - x^{\alpha})^{\beta-1}; \qquad 0 < x < 1. \tag{3.4}$$

From (3.3) the quantile function x_p can be written as

$$x_p = \left[1 - (1 - p)^{\beta} \right]^{1/\alpha}. \tag{3.5}$$

In this chapter, we review the Kumaraswamy distribution, its properties and applications. In section 3.2 we illustrate the features of Kumaraswamy distribution. The properties of this distribution are discussed in section 3.2. We clarify the similarity and differences between the Kumaraswamy (Kum) and beta distributions. Its relation to other distributions are presented in sections 3.4 and 3.5 respectively. Some of its applications are given in section 3.6. Finally, a summary of the Kum distribution is given in section 3.7.

3.2 Features of the Kum Distribution

As mentioned above the shape parameters α and β can take any positive value or zero. In Figure 3.1, we illustrate the variety of shapes of the pdf. It is clear that when α and β cross the line 1 there is a huge change in the shape. These changes can be broken into three cases:

- **Cases when $\alpha < 1$:**

    - When $\beta < 1$ "**Graph A**": Both $f(0)$ and $f(1)$ are infinite. That means the shape of the pdf will take a bathtub shape: This is common in reliability engineering data, for example, switching voltage binary devices.

> – When $\beta = 1$ "**Graph B**" : $f(0)$ is infinite and $f(1)$ is finite.
>
> – When $\beta > 1$ "**Graph C**": The curve is similar to when b is equal 1, but $f(1) = 0$, the best example of this kind of shapes is the daily rainfall data.

- **Cases when $\alpha = 1$:**

  > – When $\beta < 1$ "**Graph D**": $f(0)$ is finite and $f(0) = c$, where c is a constant, and $f(1)$ is infinite.
  >
  > – When $\beta = 1$ "**Graph E**": The pdf is uniformly distributed with $f(x) = 1$.
  >
  > – When $\beta > 1$ "**Graph F**": $f(x)$ is monotonically decreasing from a finite value at $x = 0$, to 0 at $x = 1$.

- **Cases when $\alpha > 1$:**

  > – When $\beta < 1$ "**Graph G**": $f(x)$ increases monotonically form zero to infinity at $x = 1$.
  >
  > – When $\beta = 1$ "**Graph H**": The curve starts from zero at $x = 0$. Thereafter, increases to a finite value at $x = 1$.
  >
  > – When $\beta > 1$ "**Graph I**": The shape of the pdf takes an unimodal shape, which is common in statistical analysis and many distributions have this shape.

According to Kumaraswamy (1980) "The DB-PDF in graphs C, F and I (see a Figure 3.1) have been adopted by the rainfall process".
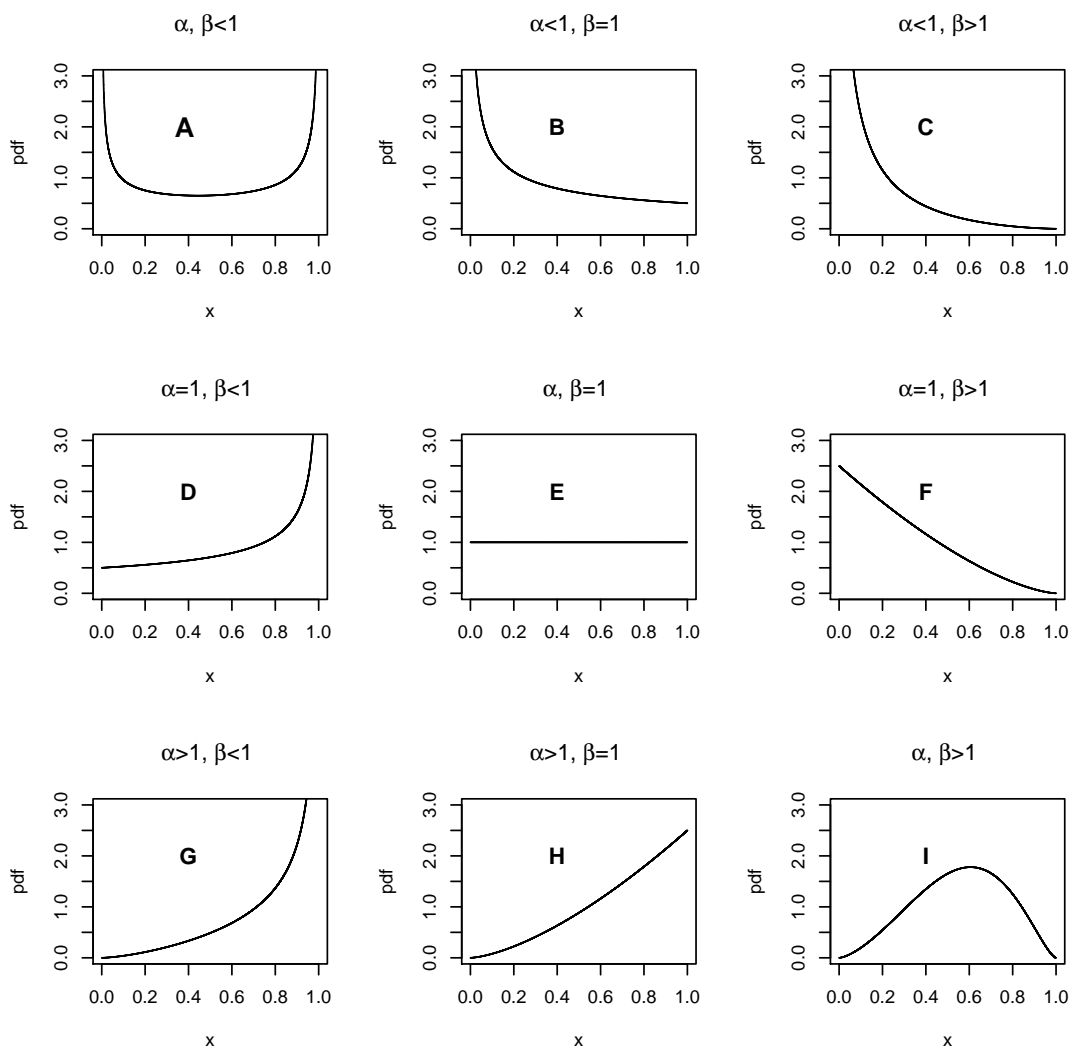
Figure 3.1: Shapes of DB-PDF for various combinations of α and β.

3.3 Properties of The Kum Distribution

In this section we present some known properties of the Kum distribution.

According to Mitnik (2008), the Kum distribution shares some properties of the beta distribution such as closed under linear transformation. This means, if X has Kum distribution then any linear transformation will be Kum distributed. However, with the same shape parameters unlike the beta distribution. Also, Kum distribution is closed under positive exponentiation. The explanation of these two properties are presented by the following theorems.

3.3.1 Closed Under Linear Transformation

Theorem 3.1. Let X be a random variable distributed as Kum with α, and β shape parameters, $X \sim Kum(\alpha, \beta, c, d)$. Let $Y = a\,X + b$. Then:

$$Y \sim Kum(\alpha, \beta, ac + b, ad + b).$$

Corollary 3.2. Let $Y = a - X$, where a is a real number. Then

$$Y \sim Kum(\alpha, \beta, a - c, a - d)$$

Corollary 3.3. If $X \sim Kum(\alpha, \beta)$ then the linear transformation $a + (b - a)X$ will move the random variable X from $(0, 1)$ to (a, b); under the condition that a and b do not depend on the shape parameters (α, β).

Proof: We proof just Theorem 3.1. Corollaries 3.2-3.3 can be proved similarly.
Let X be distributed as Kum distribution. Then the pdf of the Kum distribution can be written as

$$f_X(x, \alpha, \beta, c, d) = \frac{1}{(d-c)} \alpha\,\beta \left(\frac{x-c}{d-c}\right)^{\alpha-1} \left[1 - \left(\frac{x-c}{d-c}\right)^{\alpha}\right]^{\beta-1} ;\ x \in (c, d)$$

Suppose that $Y = a\,X + b$, $a \neq 0$ is the linear transformation of the random variable X. Then pdf of Y can be written as

$$
\begin{aligned}
f_Y(y) &= f_X\left(\frac{y-b}{a}\right)\left|\frac{dx}{dy}\right| \\[2mm]
&= \frac{1}{(d-c)}\alpha\,\beta\,\left(\frac{\left(\frac{y-b}{a}\right)-c}{d-c}\right)^{\alpha-1}\left[1-\left(\frac{\left(\frac{y-b}{a}\right)-c}{d-c}\right)^{\alpha}\right]^{\beta-1}\frac{1}{a} \\[2mm]
&= \frac{1}{(a\,d+b)-(a\,c+b)}\alpha\,\beta\,\left(\frac{y-(a\,c+b)}{(a\,d+b)-(a\,c+b)}\right)^{\alpha-1} \\[2mm]
&\quad \times\left[1-\left(\frac{y-(a\,c+b)}{(a\,d+b)-(a\,c+b)}\right)^{\alpha}\right]^{\beta-1}
\end{aligned}
$$

Then, we obtain   (3.6)

$$
f_Y(y) = \frac{1}{d^*-c^*}\alpha\,\beta\,\left(\frac{y-c^*}{d^*-(c^*)}\right)^{\alpha-1}\left[1-\left(\frac{y-c^*}{d^*-c^*}\right)^{\alpha}\right]^{\beta-1} \qquad (3.7)
$$

where, $d^* = a\,d+b$ and $c^* = a\,c+b$. Then Y is distributed as Kum with parameters $(\alpha,\ \beta,\ ac+b,\ ad+b)$ and it can be written as

$$
Y \sim Kum(\alpha,\ \beta,\ ac+b,\ ad+b)
$$

□

3.3.2   Closed Under Exponentiation

Theorem 3.1. *Let X be a random variable distributed as standard Kum with α and β shape parameters; $X \sim Kum(\alpha,\ \beta)$. Let $Y = X^m$, $m > 0$. Then*

$$
Y \sim Kum(\frac{\alpha}{m},\ \beta)
$$

From this theorem we can prove that $Y \sim B(1,\beta)$, where $B(.,.)$ is the beta function defined as

$$
B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.
$$

And $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$ is the gamma function.

Proof: Suppose that X is a random variable distributed as standard Kum with parameters α and β. From Eq(3.4) the pdf is:

$$
f_X(x,\alpha,\beta) = \alpha\,\beta\,x^{\alpha-1}(1-x^\alpha)^{\beta-1}; \qquad 0 < x < 1.
$$

Let $Y = X^m$. Using the transformation method, it is easy to prove that $Y \sim Kum(\frac{\alpha}{m}, \beta)$:

$$
\begin{aligned}
f_Y(y) &= f_X\left(y^{1/m}\right)\left|\frac{dx}{dy}\right| \\
&= \alpha\,\beta\,\left(y^{1/m}\right)^{\alpha-1}\left(1-(y^{1/m})^\alpha\right)^{\beta-1}\frac{1}{m}y^{1/m-1}, \\
\therefore f_Y(y) &= \frac{\alpha}{m}\beta\,y^{\alpha/m-1}\left(1-y^{\alpha/m}\right)^{\beta-1}, \tag{3.8}
\end{aligned}
$$

which is the *pdf* of the standard kum distribution with parameters $(\frac{\alpha}{m}, \beta)$. $\square$

3.3.3   The Limit Behaviour

According to Jones (2009), the asymptotic distribution of the pdf of the Kum distribution can be written as

$$
g(x) \sim x^{\alpha-1} \quad as \quad x \to 0,
$$

$$
g(x) \sim (1-x)^{\beta-1} \quad as \quad x \to 1,
$$

The behaviour is similar to the beta distribution.

3.3.4   Unimodality

Theorem 3.2. *The Kumaraswamy distribution is unimodal in the case when both α and β greater than 1, and the mode x_0 can be expressed as*

$$
x_0 = \left(\frac{\alpha-1}{\alpha\,\beta-1}\right)^{1/\alpha}
$$

Proof: The first derivative of $\log\left(f(x)\right)$ is

$$
\frac{d\log f(x)}{dx} = \frac{\alpha-1}{x} - \frac{\beta-1}{1-x},
$$

Then, equaling this equation to zero and solving for x, we obtain

$$
\frac{\alpha-1}{x} = \frac{\beta-1}{1-x},
$$

Then the mode of x can be written as

$$x_0 = \left(\frac{\alpha - 1}{\alpha \ \beta - 1} \right)^{1/\alpha}.$$

$\square$

3.3.5 The Moments of the Kum Distribution

The following theorem gives the moments of the Kum distribution

Theorem 3.3. *Let X be a Kum distributed random variable with pdf given in (3.4). Then the moments around zero can be written as*

$$\mu'_r(X) = \beta \ B \left(1 + \frac{r}{\alpha}, \beta \right) \tag{3.9}$$

We note that the rth moment of the Kum distribution exist if $r > -\alpha$.

Proof: We can write

$$E(X^r) = \int_{-\infty}^{\infty} x^r f(x) dx,$$

$$= \alpha \ \beta \int_0^1 x^r x^{\alpha-1} (1 - x^\alpha)^{\beta-1} dx,$$

$$= \alpha \ \beta \int_0^1 x^{r+\alpha-1} (1 - x^\alpha)^{\beta-1} dx,$$

Then substituting x by $y = x^\alpha$, we obtain

$$E(X^r) = \beta \int_0^1 y^{r/\alpha} (1 - y)^{\beta-1} dy,$$

$$= \beta B \left(\frac{r}{\alpha} + 1, \beta \right), \tag{3.10}$$

$\square$

From this theorem, the mean and variance can be easily obtained as

$$E(X) = \beta \ B \left(1 + \frac{1}{\alpha}, \beta \right), \tag{3.11}$$

$$Var(X) = \beta \ B \left(1 + \frac{2}{\alpha}, \beta \right) - \left[\beta \ B \left(1 + \frac{1}{\alpha}, \beta \right) \right]^2 \tag{3.12}$$
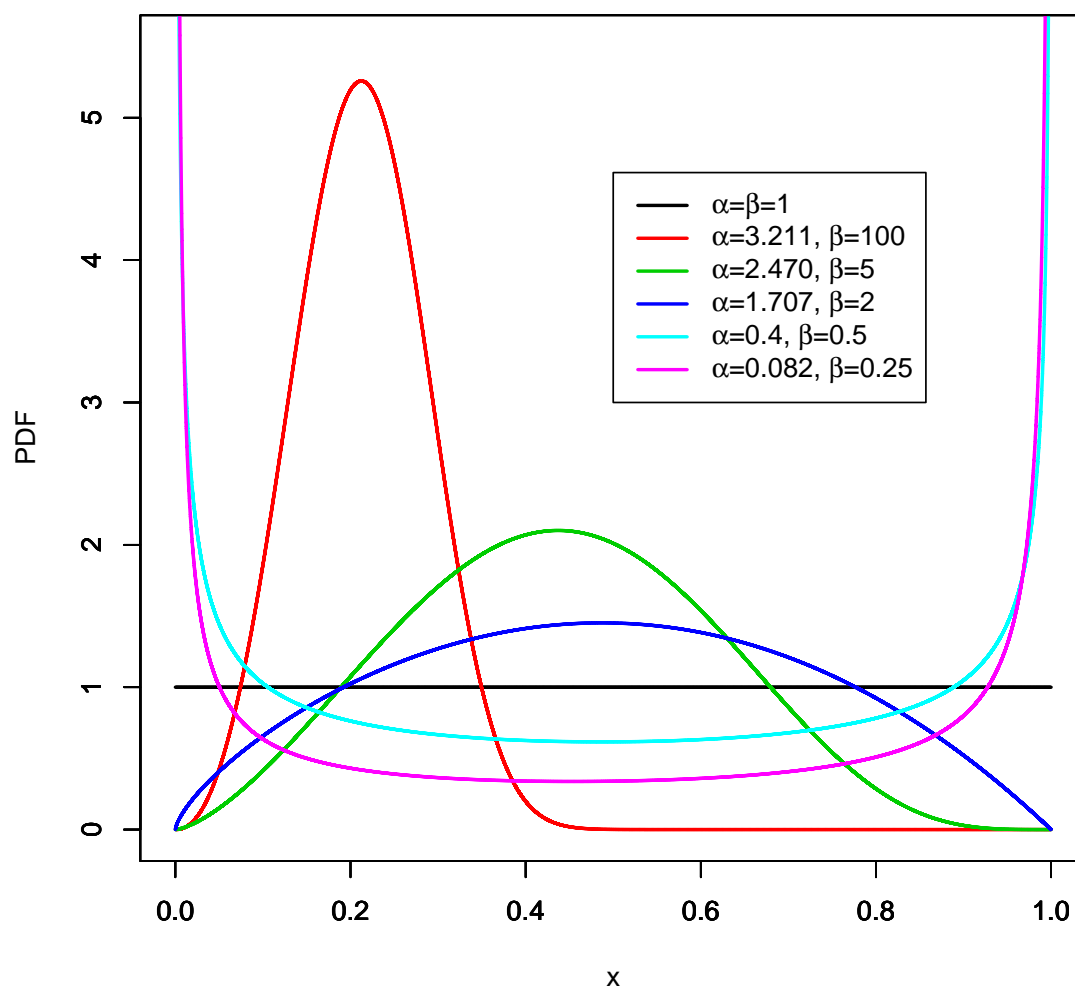
Figure 3.2: Plots of some almost-symmetric pdfs of the Kum distribution for several values of shape parameters.

Symmetry and Almost-Symmetry

Unlike the beta distribution, the Kum distribution does not have symmetric special cases except the uniform case when $\alpha = \beta = 1$. However, there are some special cases which can be considered as almost-symmetric. It is worth mentioning that when both α and β greater than 1 it takes unimodel shape, while uniantimodel when α and β less than 1. The Figure 3.2 presents some of these cases for certain values of the shape parameters.

3.4   Kum and Beta Distributions

In probability theory, the Kum distribution is very similar to the beta distribution, but the former has some advantages. Jones (2009) clarified the similarities and differences between Kum and beta distributions. For example, Kum densities are also unimodal, uniantimodal, increasing, decreasing or constant, depending on the values of shape parameters just like beta distribution.

- When $\alpha = \beta = 1$ the Kum pdf is uniformly distributed.

- It is unimodel if $\alpha > 1$ and $\beta > 1$.

- If both α and $\beta < 1$ then pdf is uniantimodal.

- If $\beta < 1$ and $\alpha \geq 1$ the pdf is increasing.

- Finally, when $\alpha l1$ and $\beta \geq 1$ the pdf is decreasing.

He also provided some advantages of the Kum distribution over the beta distribution such as it has a simple explicit formula, a simple formula for a random variate generation, an explicit formula for L-moments and a simpler formula for moments of order statistics. His study also provided an examination of some advantages of the beta distribution over Kum distribution such as a simpler formula for moments and moment generating function, a simpler moment estimation and more ways of generating the distribution via physical processes. According to Sundar and Subbiah (1989) "The major advantage of this distribution is the ability to reproduce Gaussian type distribution as well as extreme value distributions using the same equation, of course, represented by different values of parameters of the distribution."

3.5   Relation to Other Distributions

Jones (2009) found that there are three limiting distributions to the Kum distribution, which can be presented as:

- If $X \sim Kum(\alpha, \beta)$ the transformation $Y = \beta^{1/\alpha} X$ is distributed $Kum(\alpha, \beta, 0, \beta^{1/\alpha})$, and its probability density function tends to Weibull distribution as $\beta \to \infty$.

- Similarly, the distribution of $Z = \alpha(1 - X)$ tends to the negative logarithm of the $Beta(1, \beta)$ distribution as $\alpha \to \infty$.

- Finally, in the case when both α and $\beta \to \infty$ the limiting distribution of $Y = \alpha(1 - \beta^{1/\alpha} X)$ is the extreme value type I distribution with density $e^{-x} \exp(-e^{-x})$.

In addition, the author mentioned some other relations by transformation for example: $B(a, 1)$ and $Kum(\alpha, 1)$ are both the power law distributed; both $B(1, 1)$ and $Kum(1, 1)$ are uniformly distributed; beta and Kum distributions can be treated as special cases of the generalised beta distribution (G-Beta). In other words, the special cases of this model can be defined as

$$Beta(\gamma, \delta) = G - beta(1, \gamma, \delta)$$

and

$$Kum(p, \delta) = G - beta(p, 1, \delta)$$

where the G-beta distribution has its pdf defined as:

$$g(x) = \frac{p}{B(\gamma, \delta)} x^{\gamma\ p - 1} (1 - x^p)^{\delta - 1}, \qquad 0 < x < 1. \tag{3.13}$$

For more details we refer readers to see Jones (2009).

3.6   Application of Kum Distribution

The Kum distribution does not seem to be very familiar to statisticians due to the fact that it is sadly not yet widely used. The best example of its applications is the model of the storage volume of the reservoir, see Fletcher and Ponnambalam (1996). In hydrology and related areas the Kum distribution has received considerable interest: Sundar and Subbiah (1989) applied the Kum distribution to ocean wave data;

Seifi et al. (2000) applied the Kum distribution to data taken from a simple voltage divider with two resistors; Ponnambalam et al. (2001) used the Kum distribution to "approximate optimal tolerance ranges and yield of distributions which can be non-symmetrical" and Ganji et al. (2006) applied the Kum distribution to fit the weekly soil moisture. According to Nadarajah (2008), many papers in the hydrological literature have used this distribution because it is deemed as a "better alternative" to the beta distribution, see for example, Koutsoyiannis and Xanthopoulos (1989).

3.7 Summary

Some main points about the Kum distribution are:

- The Kum distribution is closed under linear transformation and exponentiation.

- pdf of the Kum distribution is flexible. That means it has a variety of shapes. Table 3.1 shows shapes of the Kum distribution and its behaviour at the boundaries.

- In spite of the fact that it is very similar to the beta distribution, there are some advantages and disadvantages over beta distribution, which are mentioned in this chapter.

Table 3.1: Shapes of the Kum distribution for different values of its shape parameters.

	$\beta < 1$	$\beta = 1$	$\beta > 1$
$\alpha < 1$	Antimodal distribution $\lim_{x\to 0} F(x) = \infty$ $\lim_{x\to 1} F(x) = \infty$	Monotonically decreasing $\lim_{x\to 0} F(x) = \infty$ $\lim_{x\to 1} F(x) = c$	Monotonically decreasing $\lim_{x\to 0} F(x) = \infty$ $\lim_{x\to 1} F(x) = 0$
$\alpha = 1$	Monotonically increasing $\lim_{x\to 0} F(x) = c$ $\lim_{x\to 1} F(x) = \infty$	Uniform distribution $\lim_{x\to 0} F(x) = 1$ $\lim_{x\to 1} F(x) = 1$	Monotonically decreasing $\lim_{x\to 0} F(x) = c$ $\lim_{x\to 1} F(x) = 0$
$\alpha > 1$	Monotonically increasing $\lim_{x\to 0} F(x) = 0$ $\lim_{x\to 1} F(x) = \infty$	Monotonically increasing $\lim_{x\to 0} F(x) = 0$ $\lim_{x\to 1} F(x) = c$	Unimodal distribution $\lim_{x\to 0} F(x) = 0$ $\lim_{x\to 1} F(x) = 0$

Part II

The Univariate Kum-G

Distributions

Chapter 4

KumGEV distribution

4.1 Introduction

The GEV distribution has been widely used in many areas, as can be seen from the list given in chapter 2. However, it does not give adequate fits in many applications. For example, Buishand (1991) says that the GEV distribution is not "flexible enough" for extreme rainfall estimation for various durations. Kharin and Zwiers (2000) show that the GEV distribution does not fit the distribution of longest annual dry periods well. Martins et al. (2000) say that sometimes GEV distributions do not have density functions consistent with flood flows and rainfall. Katz et al. (2002) state that "the GEV distribution does not appear to be acceptable", for some annual peak flow time series. For data on traffic loading, Caprani et al. (2008) say that "the assumption of convergence to a single GEV distribution is not valid". Tolikas and Gettinby (2009) find that "the popular GEV distribution is not the best model for both the extreme minima and maxima daily returns of the Singapore stock market". Tolikas (2011) finds that among emerging African stock markets "the popular GEV distribution is not the best model for the extreme minima in all but the Egyptian stock market". Zwiers et al. (2011) find that the GEV distribution does not fit well for daily temperature extremes at regional scales. Dupuis and Field (1998) propose robust estimation of extremes, arguing that the GEV distribution may not always yield good fits. But robust estimation can be time consuming and costly. So, there

is a need for generalizations of the GEV distribution.

In this chapter, we give a simple generalization of the GEV distribution. We provide a motivation for this simple generalization. This is based on the definition of the GEV distribution. The GEV distribution arises as the limiting distribution of normalized maxima: if $X_1, X_2, \ldots, X_n$ is a random sample and if $M_n = \max(X_1, X_2, \ldots, X_n)$ then there may be norming constants $a_n > 0$, b_n such that

$$\Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) \to G_{\xi,\mu,\sigma}(x) = \exp(-u)$$

as $n \to \infty$, where $1 + \xi(x - \mu)/\sigma > 0$, $-\infty < \xi < \infty$, $-\infty < \mu < \infty$, $\sigma > 0$ and $u = \{1 + \xi(x - \mu)/\sigma\}^{-1/\xi}$. We shall refer to ξ as the shape parameter, μ as the location parameter, and σ as the scale parameter.

In practice, there may be situations where the distribution of M_n is heterogeneous (see, for example, Caprani et al. (2008)). One possible way to describe this situation is to model the distribution of M_n as a mixture, say

$$\Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) = \Pr\left(M_n \leq a_n x + b_n\right) = \sum_{i=1}^{p} w_i \Pr\left(M_{n,i} \leq a_n x + b_n\right), \quad (4.1)$$

where $M_{n,i}$ is a random variable representing the ith component of the mixture and w_i are nonnegative weights summing to one. Under suitable conditions, the limiting distribution of (4.1) may be

$$\Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) \to \sum_{i=1}^{p} w_i G_{\xi_i,\mu_i,\sigma_i}(x) \qquad (4.2)$$

as $n \to \infty$.

But mixtures of the form (4.2) are notoriously difficult to handle not just because of the complicated mathematical form. Inferences and fitting of (4.2) are also difficult. Indeed, applications of mixtures of GEV distributions have been very limited.

A way around is to rewrite (4.2) in a simple mathematical form. There are many choices for the mathematical form.

A choice motivated by the works of Kumaraswamy (1980) and Cordeiro and de Castro (2011) is

$$F(x) = 1 - \{1 - G(x)^a\}^b, \qquad (4.3)$$

where $G(\cdot)$ denotes a GEV cumulative distribution function and $a > 0$, $b > 0$ are two additional parameters. Note that the right hand side of (4.3) can be expanded as

$$1 - \{1 - G(x)^a\}^b = \sum_{i=1}^{\infty} c_i G^{ai}(x), \tag{4.4}$$

a mixture taking the form of (4.2). The coefficients c_i are functions of b. For instance, $c_1 = b$. The parameter a dictates the tail behaviors of the mixture components. The parameter b dictates the mixture coefficients.

Following the terminology used in Cordeiro and de Castro (2011), we shall refer to the distribution given by (4.3) as the KumGEV distribution. The probability density function corresponding to (4.3) is

$$f(x) = ab \, g(x) \, G(x)^{a-1} \left\{1 - G(x)^a\right\}^{b-1}, \tag{4.5}$$

where

$$g(x) = g_{\xi,\mu,\sigma}(x) = \frac{dG_{\xi,\mu,\sigma}(x)}{dx} = \sigma^{-1} u^{1+\xi} \exp(-u) \tag{4.6}$$

is the probability density function of the GEV distribution. Because $g(\cdot)$ and $G(\cdot)$ are tractable, the KumGEV distribution can be used quite effectively even if the data are censored. Moreover, existing software for the GEV distribution (say, to compute probability density function, cumulative distribution function, quantile function, moments, maximum likelihood estimates, random numbers, etc) can be easily adapted for the KumGEV distribution. Clearly, the GEV distribution is a special case of the KumGEV distribution for $a = b = 1$.

The role of the two additional parameters, $a > 0$ and $b > 0$, is to govern skewness and generate distributions with heavier/ligther tails. If $a < 1$ then the tails of $f(\cdot)$ will be heavier than those of $g(\cdot)$. Similarly, if $b < 1$ then the tails of $f(\cdot)$ will be heavier than those of $g(\cdot)$. On the other hand, if $a > 1$ then the tails of $f(\cdot)$ will be lighter than those of $g(\cdot)$. Similarly, if $b > 1$ then the tails of $f(\cdot)$ will be lighter than those of $g(\cdot)$. Further description of the role of a and b is given in Sections 4.4, 4.5, and 4.15.

One major benefit of the class of Kum generalised distributions is its ability to fit skewed data that cannot be properly fitted by existing distributions.

There are other ways to generalise the GEV distribution. One of the most recent approaches uses beta generated distributions, see Eugene et al. (2002). But beta generated distributions involve the incomplete beta function ratio, a special function requiring numerical routines. There are some other distributions in the literature that have been used to model extreme values. These include: the three-parameter kappa distribution due to Mielke Jr (1973) and the four-parameter kappa distribution due to Hosking (1994). But neither of these actually generalise the GEV distribution.

In this chapter, we study the mathematical properties of the KumGEV distribution with the hope it will attract wider applicability. From now on, we write the cumulative distribution function and probability density function of the GEV distribution, respectively, by:

$$G(x, \mu, \sigma, \xi) = \exp(-u), \tag{4.7}$$

$$g(x, \mu, \sigma, \xi) = \sigma^{-1} u^{1+\xi} \exp(-u), \tag{4.8}$$

Where $-\infty < x < \infty$, $-\infty < \xi < \infty$, $-\infty < \mu < \infty$, $\sigma > 0$, and throughout this chapter, we use $u = \{1 + \xi(x - \mu)/\sigma\}^{-1/\xi}$.

4.2 The KumGEV Distribution and its Sub-Models

The cumulative distribution function and probability density function of the KumGEV distribution are given by

$$F(x) = 1 - \{1 - \exp(-au)\}^b, \tag{4.9}$$

and

$$f(x) = \sigma^{-1} a \, b \, u^{1+\xi} \, \exp(-au) \, \{1 - \exp(-au)\}^{b-1}, \tag{4.10}$$

respectively, where a and b are positive parameters. They can be reduced to some sub-models for some special values of a and b.

1. **GEV distribution**:

   If $a = b = 1$ then the KumGEV distribution reduces to

   $$f_{(\mu,\sigma,\xi)}(x) = \sigma^{-1}u^{1+\xi} \, \exp(-u).$$

   which is the pdf of the GEV distribution with the location parameter $\mu \in \mathbf{R}$, the scale parameter $\sigma > 0$, the shape parameter $-\infty < \xi < \infty$ and $u = \{1 + \xi(x - \mu)/\sigma\}^{-1/\xi}$. It can be reduced to Gumbel, Frechét and Weibull distributions when $\xi \to 0$, $\xi > 0$ and $\xi < 0$ respectively.

2. **Gumbel distribution**:

   For $a = b = 1$ and $\xi \to 0$ the KumGEV density function yields

   $$f(x) = \sigma^{-1}u \, \exp(-u).$$

   which is the pdf of the Gumbel distribution with the location parameter $\mu \in \mathbf{R}$, the scale parameter $\sigma > 0$ and $u = \exp\{-(x - \mu)/\sigma\}$.

3. **Frechét distribution**:

   When $a = b = 1$ and $\xi > 0$ the KumGEV density reduces to

   $$f(x) = \sigma^{-1}\alpha \, u^{-\alpha-1} \, \exp\{-u^{-\alpha}\}; \quad x > \mu, \; \alpha > 0.$$

   which is the pdf of the Frechét distribution, where $u = (x - \mu)/\sigma$.

4. **Weibull distribution**:

   For $a = b = 1$ and $\xi < 0$, the KumGEV density function yields

   $$f(x) = \sigma^{-1}\alpha \, u^{\alpha-1} \exp\{-u^{\alpha}\}; \quad x > \mu, \; \alpha > 0.$$

   which is the pdf of the Weibull distribution, where $u = (x - \mu)/\sigma$.

5. **KumG distribution(Cordeiro et al. (2012))**:

   When $\xi \to 0$, the KumGEV distribution can be reduced to Kum Gumbel distribution with the *pdf* defined as

   $$f(x) = \sigma^{-1}a \, b \, u \, \exp(-au) \, \{1 - \exp(-au)\}^{b-1}.$$

where $-\infty < x$, $\mu < \infty$, $\sigma > 0$ and $u = \exp\{-(x - \mu)/\sigma\}$. This model was proposed by Cordeiro et al.(2012) and it can be reduced to Gumbel distribution when $a = b = 1$.

6. **KumW distribution (Cordeiro et al. (2010))**:

   When $\xi < 0$ we obtain

$$f(x) = \sigma^{-1} a\ b\ \alpha\ u^{\alpha-1}\ \exp(-u^{\alpha})\ \{1 - \exp(-u^{\alpha})\}^{a-1}$$

$$\times (1 - \{1 - \exp(-u^{\alpha})\}^{a})^{b-1}.$$

which is the Kum Weibull distribution introduced by Cordeiro et al. (2010), where $-\infty < x$, $\mu < \infty$, σ, $\alpha > 0$ and $u = \exp\{(x - \mu)/\sigma\}$. When $a = b = 1, b = 1$ and $c = b = 1$ KumW distribution reduces to the Weibull, exponentiated Weibull(EW) and exponentiated exponential(EE) distributions, respectively. For more details, see Cordeiro et al. (2010).

These sub-models are defined in Table 4.1. The KumGEV distribution has three shape parameters (a, b, ξ) allowing it to be highly flexibility.

Distribution	Parameters				
	a	b	μ	σ	ξ
Gumbel	1	1	-	-	$\to 0$
Frechét	1	1	-	-	> 0
Weibull	1	1	-	-	< 0
GEVD	1	1	-	-	-
KumG	-	-	-	-	$\to 0$
KumW	-	-	-	-	< 0
KumGEV	-	-	-	-	-

Table 4.1: Some sub-models of the KumGEV distribution

In the next section we shall see that KumGEV distribution has all five major shapes.

A physical interpretation of the KumGEV distribution given by (4.9) and (4.10) (when a and b are positive integers) is as follows. Suppose a system is made of b independent components and that each component is made up of a independent subcomponents. Assume that the system fails if any of the b components fails and that each component fails if all of the a subcomponents fail. Let $X_{j1}, X_{j2}, \ldots, X_{ja}$ denote the lifetimes of the subcomponents within the jth component, $j = 1, 2, \ldots, b$ with common cumulative distribution function $G_{\xi,\mu,\sigma}(x)$. Let X_j denote the lifetime of the jth component, $j = 1, \ldots, b$, and let X denote the lifetime of the entire system. So, the cumulative distribution function of X is

$$\Pr(X \leq x)$$

$$= 1 - \Pr\left(X_1 > x, X_2 > x, ..., X_b > x\right)$$

$$= 1 - \Pr^b\left(X_1 > x\right) = 1 - \left\{1 - \Pr\left(X_1 \leq x\right)\right\}^b$$

$$= 1 - \left\{1 - \Pr\left(X_{11} \leq x, X_{12} \leq x, \ldots, X_{1a} \leq x\right)\right\}^b$$

$$= 1 - \left\{1 - \Pr^a\left(X_{11} \leq x\right)\right\}^b = 1 - \left\{1 - G_{\mu,\sigma}^a(x)\right\}^b.$$

So, it follows that the KumGEV distribution given by (4.10)and (4.9) is precisely the time to failure distribution of the entire system.

The KumGEV probability density function (4.10) is much more flexible than the GEV distribution. This generalization can allow for greater flexibility of its tail. Plots of the probability density function (4.10) and (4.8) for some parameter values are given in Figures 4.1-4.2.

If X is a random variable with probability density function (4.10), we write $X \sim$ KumGEV(a, b, μ, σ, ξ). The KumGEV quantile function is obtained by inverting (4.9)

$$x = Q(z) = F^{-1}(z) = \mu + \frac{\sigma}{\xi}\left\{\left[-\frac{1}{a}\log\left\{1 - (1-z)^{1/b}\right\}\right]^{-\xi} - 1\right\}. \quad (4.11)$$

So, one can generate KumGEV variates from (4.11) by $X = Q(Z)$, where Z is a uniform variate on the unit interval $(0, 1)$.

Our second method for simulation from the KumGEV distribution is based on the rejection method. It holds if $a \geq 1$ and $b \geq 1$. Define a constant M by

$$M = \frac{a^b b(a-1)^{1-1/a}(b-1)^{b-1}}{(ab-1)^{b-1/a}}.$$
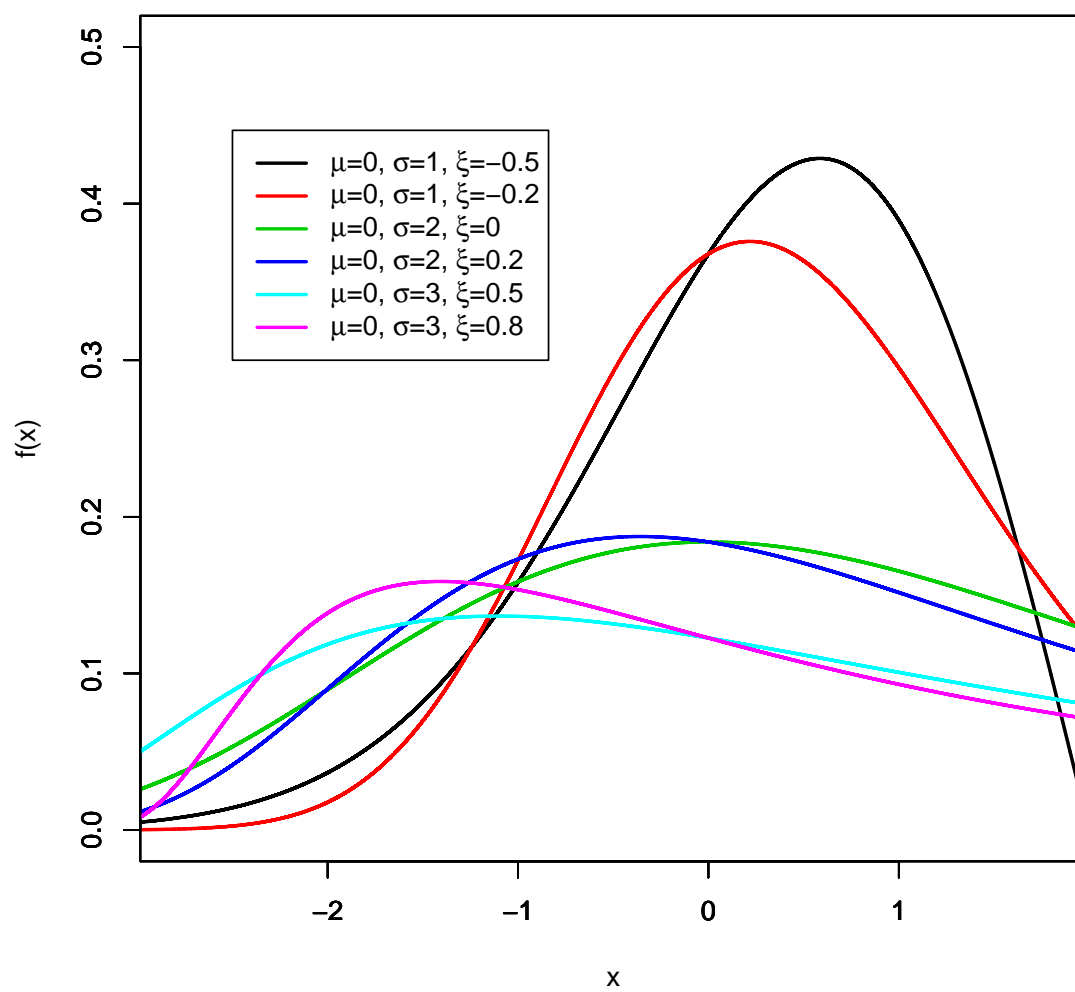
Figure 4.1: Plots of the pdf of the GEV distribution for $\mu = 0$, and several values of σ and ξ.
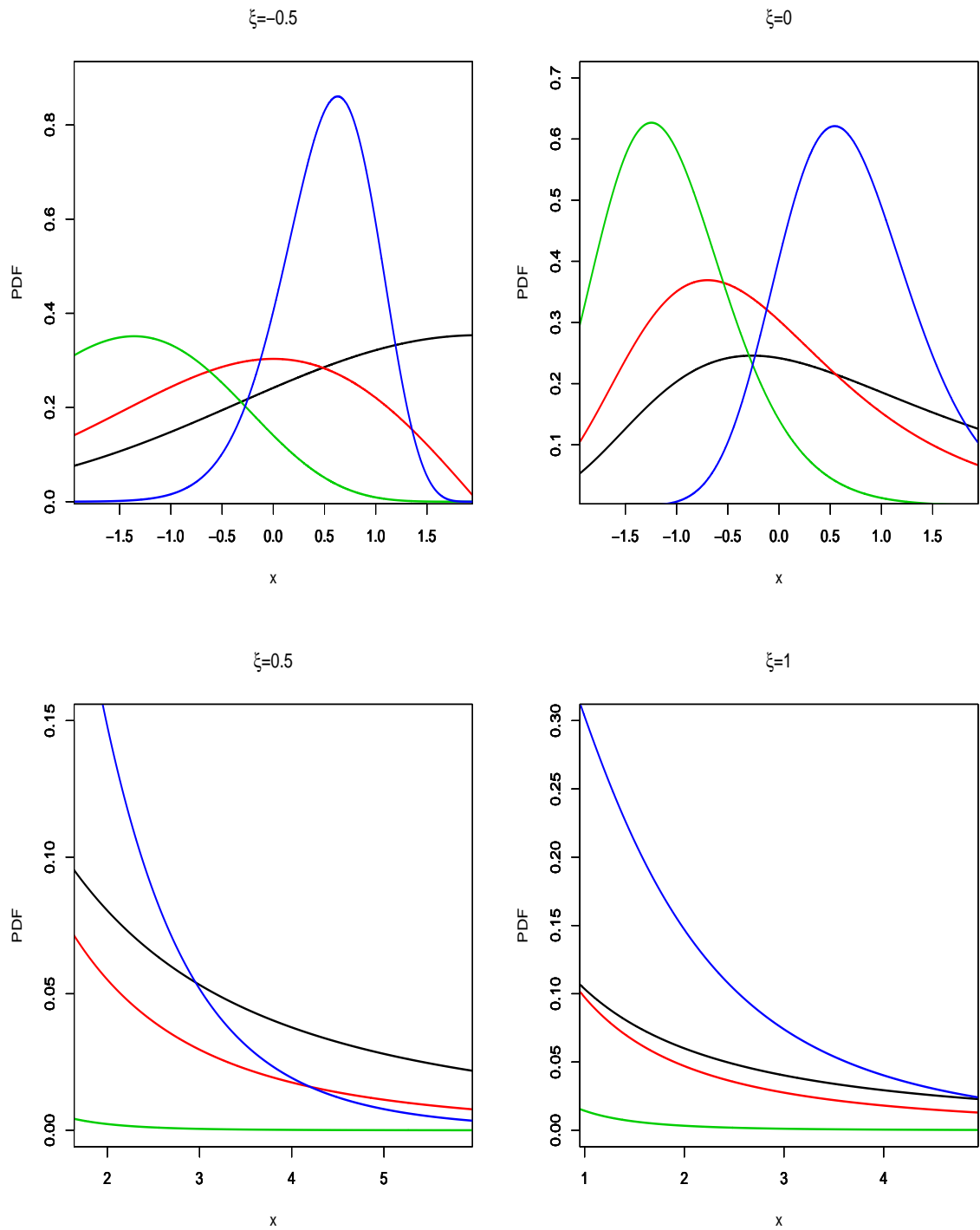
Figure 4.2: Plots of the pdf of the KumGEV distribution for $\mu = 0$, $\sigma = 1$, $\xi = -0.5, 0, 0.5, 1$, $(a, b) = (0.5, 0.5)$ (black curve), $(a, b) = (0.5, 1)$ (red curve), $(a, b) = (0.5, 3)$ (green curve) and $(a, b) = (3, 3)$ (blue curve).

Then, the following scheme holds for simulating KumGEV variates:

1. simulate $X = x$ from the probability density function $g_{\xi,\mu,\sigma}(x)$;

2. simulate $Y = UMx$, where U is a uniform variate on the unit interval $(0, 1)$;

3. accept $X = x$ as a KumGEV variate if $Y < (x)$. If $Y \geq f(x)$ return to step 2, where $g(x)$ is the pdf of GEV distribution, while $f(x)$ is the pdf of KumGEV distribution.

In the rest of this chapter, we provide a comprehensive description of the mathematical properties of (4.10). Expansions for the cumulative distribution function and probability density function of the KumGEV distribution are given in section 4.3. We examine the shape of (4.10) and its associated hazard rate function in sections 4.4 and 4.5, respectively. We derive expressions for the moments and characteristic function in sections 4.6 and 4.7. Mean deviations are derived in section 4.8. Order statistics, their moments and L moments are calculated in section 4.9. Asymptotic distributions of the extreme values are provided in section 4.10. Rényi and Shannon entropies are derived in section 4.11. Estimation by the method of maximum likelihood, including the observed information matrix, is presented in section 4.12. A simulation study is presented in section 4.13 to assess the performance of the maximum likelihood estimators. Applications of the KumGEV distribution to real data sets are illustrated in section 4.14. Bivariate generalizations of (4.10) are discussed in section 4.15. Finally, conclusions remarks are given in section 4.16.

4.3 Expansions for Distribution and Density Functions

We now derive expansions for the cumulative distribution function and probability density function of the KumGEV distribution,which are useful to study its mathematical properties. Consider the series representation (for α real, non-integer)

$$(1+z)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} z^k.$$

For b real, non-integer, the KumGEV cumulative distribution function (4.9) can be expressed as

$$F(x) = 1 - \sum_{k=0}^{\infty} \binom{b-1}{k} (-1)^k \exp(-kau) \tag{4.12}$$

and the KumGEV probability density function (4.10) follows as

$$f(x) = \sigma^{-1} abu^{1+\xi} \sum_{k=0}^{\infty} \binom{b-1}{k} (-1)^k \exp\left\{-(k+1)au\right\}, \tag{4.13}$$

where $u = \{1 + \xi(x-\mu)/\sigma\}^{-1/\xi}$. If b is an integer, the index k in equations (4.12) and (4.13) stop at b and $b-1$, respectively.

Note that (4.13) can be rewritten as a linear combination of GEV probability density functions. Let

$$\mu^* = \frac{\sigma(k+1)^\xi a^\xi}{\xi} \left[1 - \frac{1}{(k+1)^\xi a^\xi}\left(1 - \frac{\xi\mu}{\sigma}\right)\right],$$

$$\sigma^* = \sigma(k+1)^\xi a^\xi,$$

and

$$u^* = \left\{1 + \xi\frac{x-\mu^*}{\sigma^*}\right\}^{-1/\xi}.$$

Then, (4.13) reduces to

$$f(x) = \frac{b}{\sigma^*} \sum_{k=0}^{\infty} \binom{b-1}{k} \frac{(-1)^k}{k+1} (u^*)^{1+\xi} \exp(-u^*), \tag{4.14}$$

a linear combination of GEV probability density functions with parameters μ^*, σ^* and ξ. It follows that (4.12) can be rewritten as

$$F(x) = b \sum_{k=0}^{\infty} \binom{b-1}{k} \frac{(-1)^k}{k+1} \exp(-u^*),$$

a linear combination of GEV cumulative distribution functions with parameters μ^*, σ^* and ξ.

4.4 Shape of the Probability Density Function

To study the possible shapes of the KumGEV distribution, we derive the pdf and the hazard rate function of the KumGEV distribution. The first derivative of $\log\{f(x)\}$ for the KumGEV distribution is:

$$\frac{d\log f(x)}{dx} = -\frac{u^{1+\xi}}{\sigma}\left\{(1+\xi)u^{-1} - a + \frac{a(b-1)}{\exp(au) - 1}\right\},$$

where $u = \{1 + \xi(x-\mu)/\sigma\}^{-1/\xi}$. So, the modes of $f(x)$ are the roots of the equation

$$\frac{a(b-1)}{\exp(au) - 1} = a - (1+\xi)u^{-1}. \tag{4.15}$$

There may be more than one root to Eq(4.15). If $x = x_0$ is a root of Eq(4.15) then it corresponds to a local maximum, a local minimum or a point of inflexion depending on whether $\lambda(x_0) < 0$, $\lambda(x_0) > 0$ or $\lambda(x_0) = 0$, where $\lambda(x) = d^2\log f(x)/dx^2$ is given by

$$\lambda(x) = \frac{(1+\xi)u^{1+2\xi}}{\sigma^2}\left\{(1+\xi)u^{-1} - a + \frac{a(b-1)}{\exp(au)-1}\right\}$$
$$- \frac{u^{2(1+\xi)}}{\sigma^2}\left\{(1+\xi)u^{-2} + \frac{a^2(b-1)\exp(au)}{[\exp(au)-1]^2}\right\}.$$

Plots of the shapes of (4.10) for $\mu = 0$, $\sigma = 1$ and selected values of (a, b, ξ) are given in the Figure 4.2.

We can summarize the shape of pdf as follows:

a) **First case** when $\xi < 0$:

pdf of the KumGEV distribution is monotonically increasing for $b \leq 1$, while unimodal when $b > 1$.

b) **Second case** when $\xi \to 0$:

In this case pdf takes unimodal shapes and the mode x_{mode} can be calculated by solving Eq(4.15).

c) **Last case** when $\xi > 0$:

Monotonically decreasing shapes appear in the case when $\xi > 0$.

Furthermore, the asymptotes of $f(x)$ and $F(x)$ as $u \to 0, \infty$ are given by

$$f(x) \sim \sigma^{-1}abu^{1+\xi}\exp(-au) \qquad \text{as } u \to \infty,$$

$$f(x) \sim \sigma^{-1}a^b bu^{\xi+b} \qquad \text{as } u \to 0,$$

$$F(x) \sim b\exp(-au) \qquad \text{as } u \to \infty, \text{ and}$$

$$1 - F(x) \sim (au)^b \qquad \text{as } u \to 0.$$

Note that the upper tail of $f(x)$ is of exponential type while the lower tail is of double exponential type. Larger values of a correspond to lighter lower tails and heavier upper tails of f. Larger values of b correspond to heavier lower tails and lighter upper tails of f.

4.5 Shape of the Hazard Rate Function

The hazard rate function (hrf) can be defined as $h(x) = f(x)/\{1 - F(x)\}$. It is an important quantity characterizing the life time phenomena of a system. For the KumGEV distribution, $h(x)$ takes the form

$$h(x) = \frac{abu^{1+\xi}\exp(-au)}{\sigma\{1 - \exp(-au)\}}, \tag{4.16}$$

where $u = \{1 + \xi(x-\mu)/\sigma\}^{-1/\xi}$. The first derivative of $\log h(x)$ is:

$$\frac{d\log h(x)}{dx} = -\frac{u^{1+\xi}}{\sigma}\left\{(1+\xi)u^{-1} - a - \frac{a}{\exp(au) - 1}\right\}.$$

So, the modes of $h(x)$ are the roots of the equation

$$\frac{a}{\exp(au) - 1} = (1+\xi)u^{-1} - a. \tag{4.17}$$

There may be more than one root to (4.17). If $x = x_0$ is a root of (4.17) then it corresponds to a local maximum, a local minimum, or a point of inflexion; depending
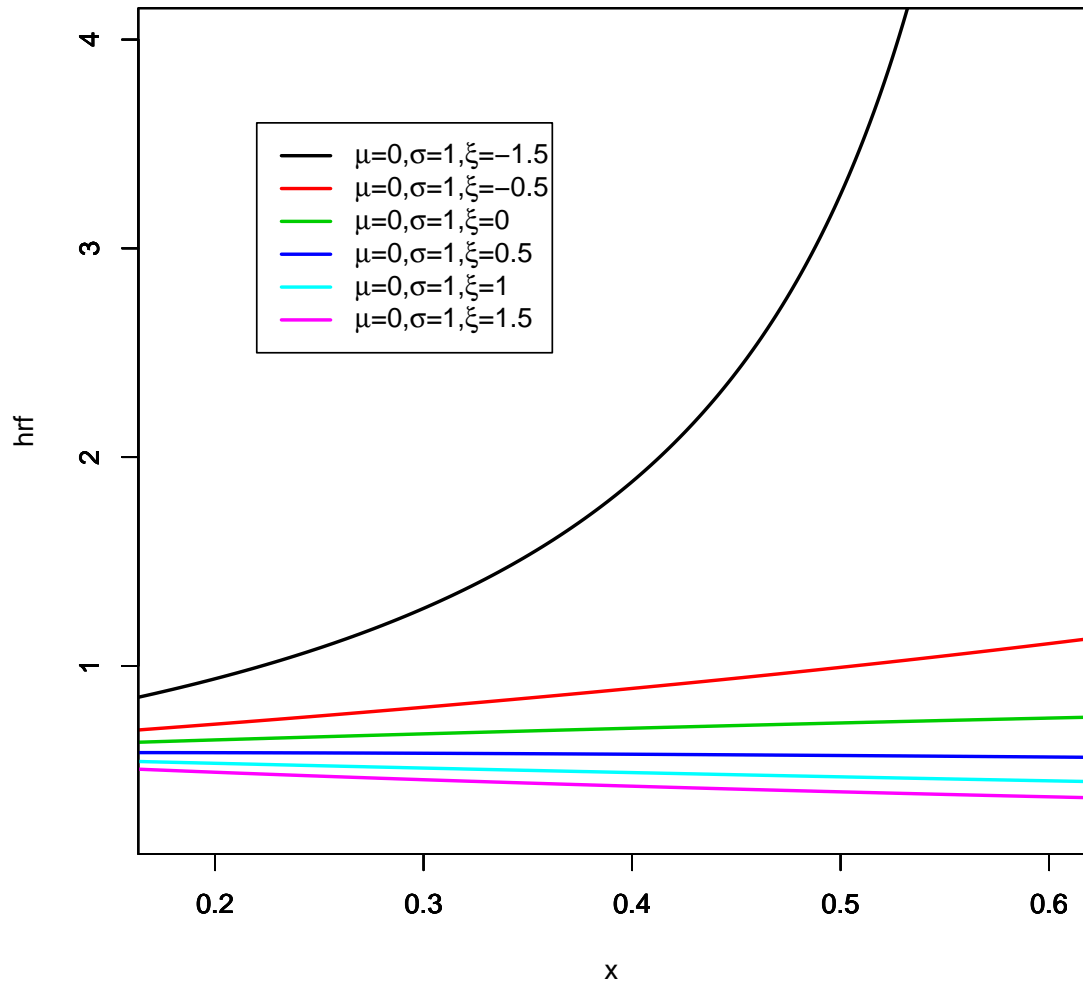
Figure 4.3: Plots of the hazard rate function of the GEV distribution for $\mu = 0$, $\sigma = 1$,and several values of ξ.

on whether $\lambda(x_0) < 0$, $\lambda(x_0) > 0$ or $\lambda(x_0) = 0$, where here $\lambda(x) = d \log^2 h(x)/dx^2$ is given by

$$
\lambda(x) = \frac{(1+\xi)u^{1+2\xi}}{\sigma^2} \left\{ (1+\xi)u^{-1} - a - \frac{a}{\exp(au) - 1} \right\}
$$
$$
- \frac{u^{2(1+\xi)}}{\sigma^2} \left\{ (1+\xi)u^{-2} - \frac{a^2 \exp(au)}{[\exp(au) - 1]^2} \right\}.
$$

Plots of the hrf for the GEV and KumGEV distributions for selected parameter values are shown in Figures 4.3-4.4.
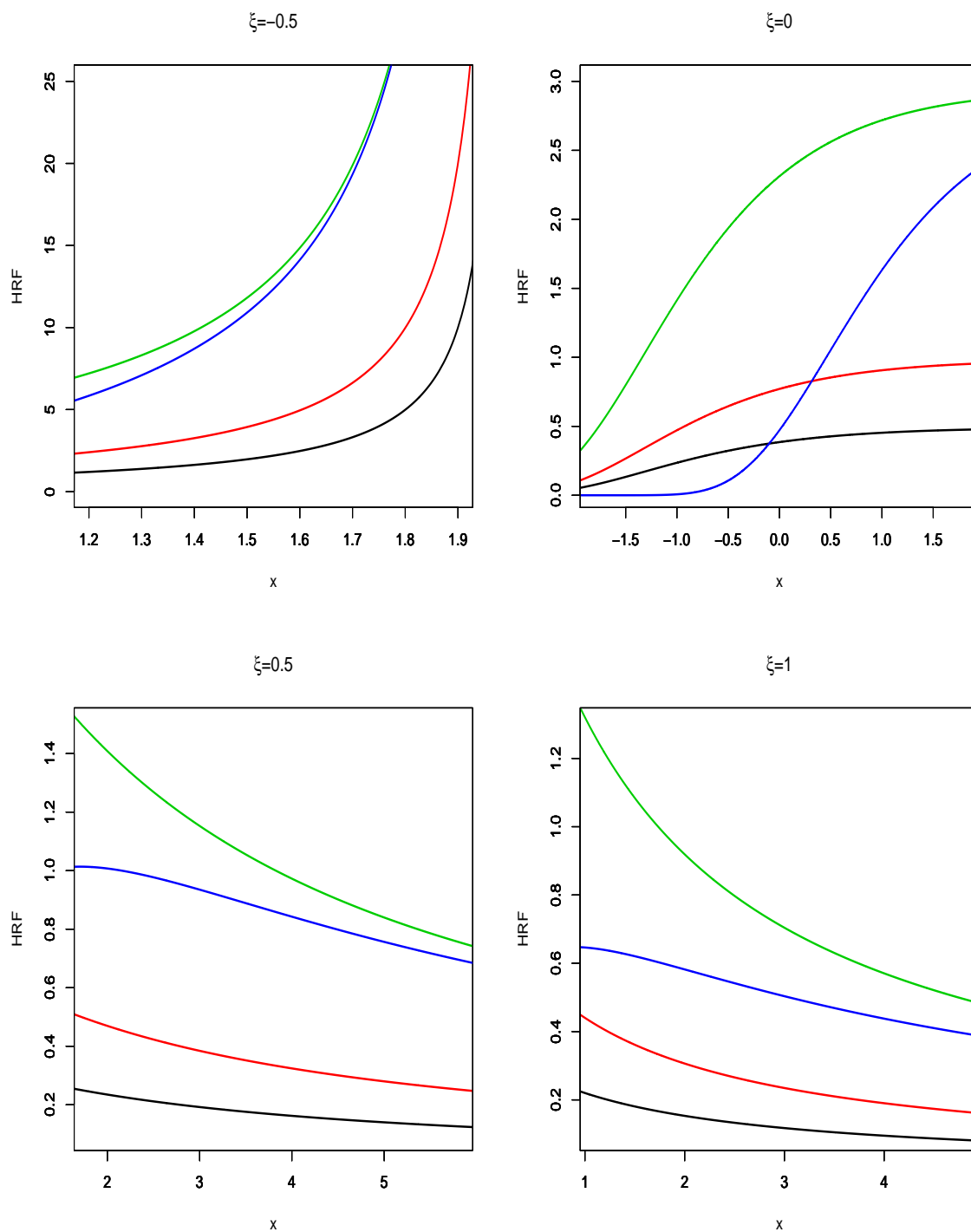
Figure 4.4: Plots of the hazard rate function of the KumGEV distribution for $\mu = 0$, $\sigma = 1$, $\xi = -0.5, 0, 0.5, 1$, $(a, b) = (0.5, 0.5)$ (black curve), $(a, b) = (0.5, 1)$ (red curve), $(a, b) = (0.5, 3)$ (green curve) and $(a, b) = (3, 3)$ (blue curve).

Furthermore, the asymptotes of $h(x)$ as $u \to 0, \infty$ are given by

$$h(x) \sim \sigma^{-1} a b u^{1+\xi} \exp(-au); \quad \text{as } u \to \infty \text{ and}$$

$$h(x) \sim \sigma^{-1} b u^{\xi}; \quad \text{as } u \to 0.$$

So, the ultimate hazard rates behave exponentially while the initial hazard rates behave double exponentially. Larger values of a correspond to lighter lower tails of h. Larger values of b correspond to heavier lower tails and heavier upper tails of h.

Figure 4.4 illustrates some of the possible shapes of $h(x)$ for $\mu = 0$, $\sigma = 1$ and selected values of (a, b, ξ). Both monotonically increasing and monotonically decreasing shapes appear possible. Monotonically increasing shapes appear when $\xi \leq 0$. Monotonically decreasing shapes appear when $\xi > 0$.

4.6 Moments

The moments of the KumGEV distribution can be presented in the following theorem.

Theorem 4.1. *Let X distributed KumGEV(a, b, μ, σ, ξ). Then the nth moments can be written as*

$$E(X^n) = \sigma^{-1} a b \sum_{k=0}^{\infty} \binom{b-1}{k} (-1)^k I(n, 1+\xi, (k+1)a)$$

Using the representation, (4.13), we can write

$$E(X^n) = \sigma^{-1} a b \sum_{k=0}^{\infty} \binom{b-1}{k} (-1)^k \int_{-\infty}^{\infty} x^n u^{1+\xi} \exp\{-(k+1)au\} \, dx, \quad (4.18)$$

where $u = \{1 + \xi(x - \mu)/\sigma\}^{-1/\xi}$. Applying Lemma 1 in appendix **A** to calculate integral in (4.18), we obtain

$$E(X^n) = \sigma^{-1} a b \sum_{k=0}^{\infty} \binom{b-1}{k} (-1)^k I(n, 1+\xi, (k+1)a)$$

where $n \geq 1$. The first four moments can be defined as

$$E(X) = b \sum_{k=0}^{\infty} \binom{b-1}{k} \frac{(-1)^k}{k+1} \left\{ \left(\mu - \frac{\sigma}{\xi} \right) + \left(\frac{\sigma}{\xi} \right) [(k+1)a]^{\xi} \Gamma(1-\xi) \right\}, \quad (4.19)$$

$$E\left(X^2\right) = b \sum_{k=0}^{\infty} \binom{b-1}{k} \frac{(-1)^k}{k+1} \left\{ \left(\mu - \frac{\sigma}{\xi}\right)^2 + 2\left(\mu - \frac{\sigma}{\xi}\right)\left(\frac{\sigma}{\xi}\right) \right.$$
$$\left. \left[(k+1)a\right]^\xi \Gamma\left(1-\xi\right) + \left(\frac{\sigma}{\xi}\right)^2 \left[(k+1)a\right]^{2\xi} \Gamma\left(1-2\xi\right) \right\}, \qquad (4.20)$$

$$E\left(X^3\right) = b \sum_{k=0}^{\infty} \binom{b-1}{k} \frac{(-1)^k}{k+1} \left\{ \left(\mu - \frac{\sigma}{\xi}\right)^3 + 3\left(\mu - \frac{\sigma}{\xi}\right)^2 \left(\frac{\sigma}{\xi}\right) \right.$$
$$\left[(k+1)a\right]^\xi \Gamma\left(1-\xi\right) + 3\left(\mu - \frac{\sigma}{\xi}\right)\left(\frac{\sigma}{\xi}\right)^2 \left[(k+1)a\right]^{2\xi}$$
$$\left. \Gamma\left(1-2\xi\right) + \left(\frac{\sigma}{\xi}\right)^3 \left[(k+1)a\right]^{3\xi} \Gamma\left(1-3\xi\right) \right\}, \qquad (4.21)$$

and

$$E\left(X^4\right) = b \sum_{k=0}^{\infty} \binom{b-1}{k} \frac{(-1)^k}{k+1} \left\{ \left(\mu - \frac{\sigma}{\xi}\right)^4 + 4\left(\mu - \frac{\sigma}{\xi}\right)^3 \left(\frac{\sigma}{\xi}\right) \right.$$
$$\left[(k+1)a\right]^\xi \Gamma\left(1-\xi\right) + 6\left(\mu - \frac{\sigma}{\xi}\right)^2 \left(\frac{\sigma}{\xi}\right)^2 \left[(k+1)a\right]^{2\xi}$$
$$\Gamma\left(1-2\xi\right) + 4\left(\mu - \frac{\sigma}{\xi}\right)\left(\frac{\sigma}{\xi}\right)^3 \left[(k+1)a\right]^{3\xi} \Gamma\left(1-3\xi\right)$$
$$\left. + \left(\frac{\sigma}{\xi}\right)^4 \left[(k+1)a\right]^{4\xi} \Gamma\left(1-4\xi\right) \right\}. \qquad (4.22)$$

provided that $1-\xi$, $1-2\xi$, $1-3\xi$ and $1-4\xi$ are not integers. The infinite series in (4.19)-(4.22) all converge.

The expressions given by (4.19)-(4.22) can be used to compute the mean, variance, skewness and kurtosis of X. Values of these four quantities versus ξ are plotted in the figure (4.5) for $\mu = 0$, $\sigma = 1$ and selected values of (a, b). It is evident each of the quantities is an increasing function of ξ for most choices of (a, b).
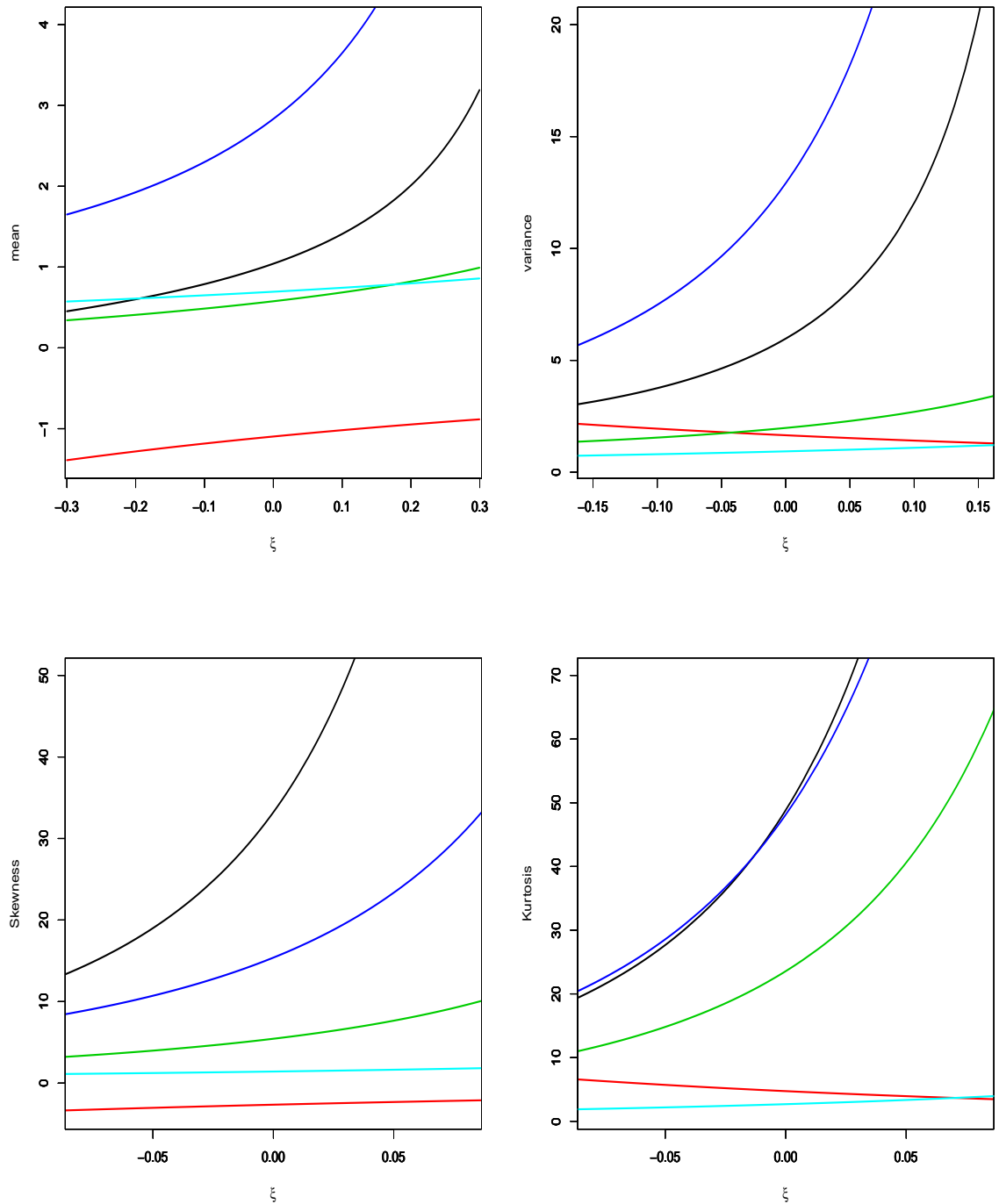
Figure 4.5: Plots of the mean, variance, skewness and kurtosis versus ξ for $\mu = 0$, $\sigma = 1$, $(a, b) = (0.5, 0.5)$ (black curve), $(a, b) = (0.5, 3)$ (red curve), $(a, b) = (1, 1)$ (green curve) and $(a, b) = (3, 0.5)$ (blue curve), and $(a, b) = (3, 3)$ (turquoise curve).

4.7 Characteristic Function

Let $X \sim \text{KumGEV}(a, b, \mu, \sigma, \xi)$. By using the representation, (4.14), the characteristic function of X is $\phi(t) = E[e^{(\text{i}tX)}]$, where $\text{i} = \sqrt{-1}$, can be expressed as

$$\phi(t) = b \sum_{k=0}^{\infty} \binom{b-1}{k} \frac{(-1)^k}{k+1} \phi^*(t), \tag{4.23}$$

where $\phi^*(\cdot)$ denotes the characteristic function of a GEV random variable with parameters μ^*, σ^* and ξ.

The characteristic function of GEV random variables has been derived only recently, see Nadarajah and Pogány (2012). It involves Fox's $H_{0,2}^{2,0}$ function and the Wright generalised confluent hypergeometric $_1\Psi_0$-function. For details on these special functions, we refer the readers to Wright (1935), (Mathai and Saxena, 1978, chap. 1), (Srivastava et al., 1982, chap. 1) and Kilbas et al. (2006).

The moment generating and cumulant generating functions of X can be deduced from (4.23). The moment generating function of X is

$$\phi(-\text{i}t) = b \sum_{k=0}^{\infty} \binom{b-1}{k} \frac{(-1)^k}{k+1} \phi^*(-\text{i}t).$$

The cumulant generating function of X is

$$\log \phi(t) = \log b + \log \left[\sum_{k=0}^{\infty} \binom{b-1}{k} \frac{(-1)^k}{k+1} \phi^*(t) \right].$$

The latter can be used to deduce the cumulants of $X \sim \text{KumGEV}(a, b, \mu, \sigma, \xi)$ from those of a GEV random variable.

4.8 Mean Deviations

The amount of scatter in a population is evidently measured to some extent by the totality of deviations from the mean and median. These are known as the mean deviation about the mean and the mean deviation about the median – defined respectively, by

$$\delta_1(X) = \int_{-\infty}^{\infty} |x - \mu_0| \, f(x) dx$$

and

$$\delta_2(X) = \int_{-\infty}^{\infty} |x - M| \, f(x) dx,$$

where $\mu_0 = E(X)$ and $M = Median(X)$ denotes the median. The measures, $\delta_1(X)$ and $\delta_2(X)$, can be calculated using the relationships

$$\delta_1(X) = \int_{-\infty}^{\mu_0} (\mu_0 - x) \, f(x) dx + \int_{\mu_0}^{\infty} (x - \mu_0) \, f(x) dx$$

$$= \mu_0 F(\mu_0) - \int_{-\infty}^{\mu_0} x f(x) dx - \mu_0 \{1 - F(\mu_0)\} + \int_{\mu_0}^{\infty} x f(x) dx$$

$$= 2\mu_0 F(\mu_0) - 2\mu_0 + 2 \int_{\mu_0}^{\infty} x f(x) dx$$

and

$$\delta_2(X) = \int_{-\infty}^{M} (M - x) f(x) dx + \int_{M}^{\infty} (x - M) f(x) dx$$

$$= MF(M) - \int_{-\infty}^{M} x f(x) dx - M \{1 - F(M)\} + \int_{M}^{\infty} x f(x) dx$$

where M is the median then $F(M) = \frac{1}{2}$. Therefore $\delta_2(X)$ can be obtain as

$$= 2 \int_{M}^{\infty} x f(x) dx - \mu_0.$$

Let $X \sim \text{KumGEV}(a, b, \mu, \sigma, \xi)$. Using the representation, (4.13), we can write

$$\int_{y}^{\infty} x f(x) dx = \sigma^{-1} ab \sum_{k=0}^{\infty} \binom{b-1}{k} (-1)^k \int_{y}^{\infty} x u^{1+\xi} \exp\left\{-(k+1)au\right\} dx$$

$$= \sigma^{-1} ab \sum_{k=0}^{\infty} \binom{b-1}{k} (-1)^k J\left(y, (k+1)a\right),$$

where $u = \{1 + \xi(x - \mu)/\sigma\}^{-1/\xi}$ and the final step follows by Lemma 2 in appendix **A**. It follows that

$$\delta_1(X) = 2\mu_0 \left\{1 - [1 - \exp(-au)]^b\right\} - 2\mu_0 + 2\sigma^{-1} ab \sum_{k=0}^{\infty} \binom{b-1}{k} (-1)^k J\left(\mu_0, (k+1)a\right),$$

and

$$\delta_2(X) = 2\sigma^{-1} ab \sum_{k=0}^{\infty} \binom{b-1}{k} (-1)^k J\left(M, (k+1)a\right) - \mu_0,$$

where μ_0 is given by (**??**) and

$$M = \mu + \frac{\sigma}{\xi} \left\{\left[-\frac{1}{a} \log\left\{1 - 2^{-1/b}\right\}\right]^{-\xi} - 1\right\}.$$

4.8.1 Bonferroni and Lorenz Curves

Bonferroni and Lorenz Curves, which are widely used in some fields like insurance, reliability, medicine and economics. The Lorenz curve was introduced by Lorenz (1905), and it is used in economics to represent income distribution. It was developed later by Gastwirth (1971). It can be defined as:

$$L(p) = \frac{1}{\mu} \int_{-\infty}^{q} x f(x)\, dx,$$

Another curve used widely in economics and reliability is the Bonferroni Curve, which was proposed by Bonferroni (1930). It can be written as:

$$B(p) = \frac{1}{p\mu} \int_{-\infty}^{q} x f(x)\, dx,$$

where, $\mu = E(x)$ and $q = \mu + \frac{\sigma}{\xi}\{\log[1 - (1-p)^{-\frac{1}{a}}]^{-\xi} - 1\}$.

4.9 Order Statistics

Order statistics make their appearance in many areas of statistical theory and practical applications. Let $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$ denote the order statistics for a random sample $X_1, X_2, \ldots, X_n$ from (4.10). Then the probability density function and the cumulative distribution function of the kth order statistic, say $Y = X_{k:n}$, are given, respectively, by

$$f_Y(y) = \frac{abn!}{\sigma(k-1)!(n-k)!} u^{1+\xi} \exp(-au)\left[1 - \exp(-au)\right]^{b(1+n-k)-1}$$
$$\left\{1 - [1 - \exp(-au)]^b\right\}^{k-1}$$

and

$$F_Y(y) = \sum_{j=k}^{n} \binom{n}{j} \left\{1 - [1 - \exp(-au)]^b\right\}^{j} [1 - \exp(-au)]^{b(n-j)},$$

Where $u = \{1 + \xi(x - \mu)/\sigma\}^{-1/\xi}$. The moments of Y can be found as follows, by expanding the probability density function:

$$E\left(Y^p\right) = \frac{abn!}{\xi(k-1)!(n-k)!} \sum_{i=0}^{k-1} \sum_{j=0}^{\infty} \binom{k-1}{i} \binom{b(i+1+n-k)-1}{j} (-1)^{i+j}$$

$$\times \int_{-\infty}^{\infty} x^p u^{1+\xi} \exp\left\{-(j+1)au\right\} dx$$

$$= \frac{abn!}{\xi(k-1)!(n-k)!} \sum_{i=0}^{k-1} \sum_{j=0}^{\infty} \binom{k-1}{i} \binom{b(i+1+n-k)-1}{j} (-1)^{i+j}$$

$$\times I\left(p, 1 + \xi, (j+1)a\right), \tag{4.24}$$

where the final step follows by Lemma 1 in appendix **A**.

L-moments are summary statistics for probability distributions and data samples Hosking (1990). They are analogous to ordinary moments but are computed from linear functions of the ordered data values. The rth L moment is defined by

$$\lambda_r = \sum_{j=0}^{r-1} (-1)^{r-1-j} \binom{r-1}{j} \binom{r-1+j}{j} \beta_j,$$

where $\beta_j = E\{XF(X)^j\}$. In particular, $\lambda_1 = \beta_0$, $\lambda_2 = 2\beta_1 - \beta_0$, $\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0$ and $\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0$. In general, $\beta_r = (r+1)^{-1}E(X_{r+1:r+1})$, so it can be computed using (4.24). The L moments have several advantages over ordinary moments: for example, they apply for any distribution having finite mean; no higher-order moments need be finite.

4.10   Extreme Values

If $X_1, \ldots, X_n$ is a random sample from (4.10) and if $\bar{X} = (X_1 + \cdots + X_n)/n$ denotes the sample mean, then by the usual central limit theorem $\sqrt{n}(\bar{X} - \mathrm{E}(X))/\sqrt{\mathrm{Var}(X)}$ approaches the standard normal distribution as $n \to \infty$. Sometimes one would be interested in the asymptotic of the extreme order statistics $M_n = \max(X_1, \ldots, X_n)$ and $m_n = \min(X_1, \ldots, X_n)$.

Take the probability density function and the cumulative distribution function as given by (4.10) and (4.9), respectively. It is easy to see that

$$\Pr\left(M_n \le x\right) = \left\{1 - \left[1 - \exp(-au)\right]^b\right\}^n$$

and

$$\Pr\left(m_n \le x\right) = 1 - \left[1 - \exp(-au)\right]^{nb},$$

where $u = \{1 + \xi(x - \mu)/\sigma\}^{-1/\xi}$. Let

$$\omega_n = \mu + \frac{\sigma}{\xi}\left[\left(\frac{u}{na} + \frac{\log b}{a}\right)^{-\xi} - 1\right]$$

and

$$\epsilon_n = \mu + \frac{\sigma}{\xi}\left[\left(\frac{u}{a} + \frac{\log(nb)}{a}\right)^{-\xi} - 1\right].$$

Then,

$$\lim_{n\to\infty} \Pr\left(M_n \le \omega_n\right) = \lim_{n\to\infty}\left\{1 - \left[1 - \exp\left(-\frac{u}{n} - \log b\right)\right]^b\right\}^n$$

$$= \lim_{n\to\infty}\left\{1 - \left[1 - \frac{1}{b}\exp\left(-\frac{u}{n}\right)\right]^b\right\}^n$$

$$= \lim_{n\to\infty}\left\{\exp\left(-\frac{u}{n}\right)\right\}^n$$

$$= \exp(-u)$$

and

$$\lim_{n\to\infty} \Pr\left(m_n \le \epsilon_n\right) = \lim_{n\to\infty} 1 - \left[1 - \exp\left(-u - \log n - \log b\right)\right]^{nb}$$

$$= \lim_{n\to\infty} 1 - \left[1 - \frac{1}{nb}\exp\left(-u\right)\right]^{nb}$$

$$= 1 - \left[1 - \exp\left(-u\right)\right]$$

$$= \exp(-u).$$

Note that both ω_n and ϵ_n are non-linear transformations of x. One can also find linear transformations that converge in distribution to GEV random variables as $n \to \infty$. For details, we refer the readers to (Leadbetter et al., 1983, chap. 1).

4.11 Entropies

An entropy of a random variable X is a measure of variation of the uncertainty. Let $X \sim \text{KumGEV}(a, b, \mu, \sigma, \xi)$. Here, we derive explicit forms for two most popular entropies: Rényi entropy and Shannon entropy.

Firstly, consider Rényi entropy Rényi (1961) defined by

$$\mathcal{J}_R(\gamma) = \frac{1}{1-\gamma} \log \left\{ \int f^\gamma(x) dx \right\}, \tag{4.25}$$

where $\gamma > 0$ and $\gamma \neq 1$. For the probability density function given by (4.10),

$$\int_{-\infty}^{\infty} f^\gamma(x) dx = \left(\frac{ab}{\sigma}\right)^\gamma \int_{-\infty}^{\infty} u^{\gamma+\gamma\xi} \exp(-a\gamma u) \left[1 - \exp(-au)\right]^{b\gamma-\gamma} dx$$

$$= \left(\frac{ab}{\sigma}\right)^\gamma \sum_{k=0}^{\infty} \binom{b\gamma-\gamma}{k} (-1)^k \int_{-\infty}^{\infty} u^{\gamma+\gamma\xi} \exp\left\{-(\gamma+k)au\right\} dx$$

$$= \left(\frac{ab}{\sigma}\right)^\gamma \sum_{k=0}^{\infty} \binom{b\gamma-\gamma}{k} (-1)^k I\left(0, \gamma+\gamma\xi, (\gamma+k)a\right),$$

where $u = \{1 + \xi(x-\mu)/\sigma\}^{-1/\xi}$. So, (4.25) yields the expression

$$\mathcal{J}_R(\gamma) = \frac{\gamma}{1-\gamma} \log\left(\frac{ab}{\sigma}\right) + \frac{1}{1-\gamma} \log\left\{ \sum_{k=0}^{\infty} \binom{b\gamma-\gamma}{k} (-1)^k I\left(0, \gamma+\gamma\xi, (\gamma+k)a\right) \right\},$$

where $\gamma > 0$ and $\gamma \neq 1$.

Shannon entropy (Shannon (1951)) defined by $E[-\log f(X)]$ is the particular case of (4.25) for $\gamma \uparrow 1$. However, its expression can be derived more easily without using this fact. Let $U = \{1 + \xi(X-\mu)/\sigma\}^{-1/\xi}$. Then, using the series expansion for $\log(1-z)$, we can write

$$E\left[-\log f(X)\right] = -\log\left(\frac{ab}{\sigma}\right) - (1+\xi)E(\log U) + aE(U) + (1-b)$$

$$\times E\left\{\log\left[1 - \exp(-aU)\right]\right\}$$

$$= -\log\left(\frac{ab}{\sigma}\right) - (1+\xi)E(\log U) + aE(U) + (b-1)$$

$$\times \sum_{i=1}^{\infty} \frac{1}{i} E\left[\exp(-iaU)\right]. \tag{4.26}$$

By (4.13) and Lemma 1 in appendix **A**,

$$E\left[\exp(-iaU)\right] = \sigma^{-1}ab \sum_{k=0}^{\infty} \binom{b-1}{k} (-1)^k \int_{-\infty}^{\infty} u^{1+\xi} \exp\left\{-(k+i+1)au\right\} dx$$

$$= \sigma^{-1}ab \sum_{k=0}^{\infty} \binom{b-1}{k} (-1)^k I\left(0, 1+\xi, (k+i+1)a\right). \tag{4.27}$$

By (4.13) and Lemma 3 in appendix **A**,

$$E\left[\log U\right] = \sigma^{-1}ab \sum_{k=0}^{\infty} \binom{b-1}{k} (-1)^k \int_{-\infty}^{\infty} \log u \, u^{1+\xi} \exp\left\{-(k+1)au\right\} dx$$

$$= -b \sum_{k=0}^{\infty} \binom{b-1}{k} \frac{(-1)^k}{k+1} \left\{\log\left[(k+1)a\right] + C\right\}, \tag{4.28}$$

where C is Euler's constant. Combining (4.26)-(4.28), we obtain

$$E\left[-\log f(X)\right] = -\log\left(\frac{ab}{\sigma}\right) + b(1+\xi)\sum_{k=0}^{\infty}\binom{b-1}{k}\frac{(-1)^k}{k+1}\left\{\log\left[(k+1)a\right] + C\right\}$$

$$+\sigma^{-1}a^2b\sum_{k=0}^{\infty}\binom{b-1}{k}(-1)^k I\left(0, 2+\xi, (k+1)au\right)$$

$$+\sigma^{-1}ab(b-1)\sum_{i=1}^{\infty}\frac{1}{i}\sum_{k=0}^{\infty}\binom{b-1}{k}(-1)^k I\left(0, 1+\xi, (k+i+1)a\right).$$

4.12   Maximum Likelihood Estimation

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from (4.10). Let $u_i = \{1 + \xi(x_i - \mu)/\sigma\}^{-1/\xi}$ for $i = 1, 2, \ldots, n$. Then the log-likelihood function for the vector of parameters (a, b, μ, σ, ξ) can be written as

$$\log L(a, b, \mu, \sigma, \xi) = -n\log\sigma + n\log(ab) + (1+\xi)\sum_{i=1}^{n}\log u_i - a\sum_{i=1}^{n}u_i$$

$$+(b-1)\sum_{i=1}^{n}\log\left[1 - \exp\left(-au_i\right)\right]. \tag{4.29}$$

The first-order partial derivatives of (4.29) with respect to the five parameters are:

$$\frac{\partial\log L}{\partial a} = \frac{n}{a} - \sum_{i=1}^{n}u_i + (b-1)\sum_{i=1}^{n}\frac{u_i}{\exp\left(au_i\right)-1}, \tag{4.30}$$

$$\frac{\partial\log L}{\partial b} = \frac{n}{b} + \sum_{i=1}^{n}\log\left[1 - \exp\left(-au_i\right)\right], \tag{4.31}$$

$$\frac{\partial\log L}{\partial\mu} = \frac{1+\xi}{\sigma}\sum_{i=1}^{n}u_i^{\xi} - \frac{a}{\sigma}\sum_{i=1}^{n}u_i^{1+\xi} + \frac{a(b-1)}{\sigma}\sum_{i=1}^{n}\frac{u_i^{1+\xi}}{\exp\left(au_i\right)-1}, \tag{4.32}$$

$$\frac{\partial\log L}{\partial\sigma} = -\frac{n}{\sigma} + \frac{1+\xi}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)u_i^{\xi} - \frac{a}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)u_i^{1+\xi}$$

$$+\frac{a(b-1)}{\sigma^2}\sum_{i=1}^{n}\frac{(x_i-\mu)u_i^{1+\xi}}{\exp\left(au_i\right)-1}, \tag{4.33}$$

and

$$\frac{\partial\log L}{\partial\xi} = \sum_{i=1}^{n}\log u_i - \frac{1+\xi}{\xi}\sum_{i=1}^{n}\left(\log u_i + u_i^{\xi}\frac{x_i-\mu}{\sigma}\right) + \frac{a}{\xi}\sum_{i=1}^{n}u_i\left(\log u_i + u_i^{\xi}\frac{x_i-\mu}{\sigma}\right)$$

$$-\frac{a(b-1)}{\xi}\sum_{i=1}^{n}\frac{u_i}{\exp\left(au_i\right)-1}\left(\log u_i + u_i^{\xi}\frac{x_i-\mu}{\sigma}\right). \tag{4.34}$$

The maximum likelihood estimates of (a, b, μ, σ, ξ), $(\widehat{a}, \widehat{b}, \widehat{\mu}, \widehat{\sigma}, \widehat{\xi})$, are the simultaneous solutions of the equations $\partial \log L / \partial a = 0$, $\partial \log L / \partial b = 0$, $\partial \log L / \partial \mu = 0$, $\partial \log L / \partial \sigma = 0$ and $\partial \log L / \partial \xi = 0$. As $n \to \infty$, $(\widehat{a} - a, \widehat{b} - b, \widehat{\mu} - \mu, \widehat{\sigma} - \sigma, \widehat{\xi} - \xi)$ approaches a multivariate normal vector with zero means and variance-covariance matrix $-(E[\mathbf{J}])^{-1}$, where

$$
\mathbf{J} = \begin{pmatrix}
\dfrac{\partial^2 \log L}{\partial a^2} & \dfrac{\partial^2 \log L}{\partial a \partial b} & \dfrac{\partial^2 \log L}{\partial a \partial \mu} & \dfrac{\partial^2 \log L}{\partial a \partial \sigma} & \dfrac{\partial^2 \log L}{\partial a \partial \xi} \\
\dfrac{\partial^2 \log L}{\partial b \partial a} & \dfrac{\partial^2 \log L}{\partial b^2} & \dfrac{\partial^2 \log L}{\partial b \partial \mu} & \dfrac{\partial^2 \log L}{\partial b \partial \sigma} & \dfrac{\partial^2 \log L}{\partial b \partial \xi} \\
\dfrac{\partial^2 \log L}{\partial \mu \partial a} & \dfrac{\partial^2 \log L}{\partial \mu \partial b} & \dfrac{\partial^2 \log L}{\partial \mu^2} & \dfrac{\partial^2 \log L}{\partial \mu \partial \sigma} & \dfrac{\partial^2 \log L}{\partial \mu \partial \xi} \\
\dfrac{\partial^2 \log L}{\partial \sigma \partial a} & \dfrac{\partial^2 \log L}{\partial \sigma \partial b} & \dfrac{\partial^2 \log L}{\partial \sigma \partial \mu} & \dfrac{\partial^2 \log L}{\partial \sigma^2} & \dfrac{\partial^2 \log L}{\partial \sigma \partial \xi} \\
\dfrac{\partial^2 \log L}{\partial \xi \partial a} & \dfrac{\partial^2 \log L}{\partial \xi \partial b} & \dfrac{\partial^2 \log L}{\partial \xi \partial \mu} & \dfrac{\partial^2 \log L}{\partial \xi \partial \sigma} & \dfrac{\partial^2 \log L}{\partial \xi^2}
\end{pmatrix}.
$$

The matrix, $-E[\mathbf{J}]$, is known as the expected information matrix. The matrix, $-\mathbf{J}$, is known as the observed information matrix.

In practice, n is finite. The literature (see, Cox and Hinkley (1979)) suggests that it is best to approximate the distribution of $(\widehat{a} - a, \widehat{b} - b, \widehat{\mu} - \mu, \widehat{\sigma} - \sigma, \widehat{\xi} - \xi)$ by a multivariate normal distribution with zero means and variance-covariance matrix given by $-\mathbf{J}^{-1}$, inverse of the observed information matrix, with (a, b, μ, σ, ξ) replaced $(\widehat{a}, \widehat{b}, \widehat{\mu}, \widehat{\sigma}, \widehat{\xi})$.

The $5 * 5$ unit expected information matrix $(k_{i,j} = -E[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln f(x_i, \theta)])$ can be written as:

$$
k(\theta) = \begin{bmatrix}
k_{a,a}(\theta) & k_{a,b}(\theta) & k_{a,\mu}(\theta) & k_{a,\sigma}(\theta) & k_{a,\xi}(\theta) \\
& k_{b,b}(\theta) & k_{b,\mu}(\theta) & k_{b,\sigma}(\theta) & k_{b,\xi}(\theta) \\
& & k_{\mu,\mu}(\theta) & k_{\mu,\sigma}(\theta) & k_{\mu,\xi}(\theta) \\
& & & k_{\sigma,\sigma}(\theta) & k_{\sigma,\xi}(\theta) \\
& & & & k_{\xi,\xi}(\theta)
\end{bmatrix}
$$

Elements in the information matrix $\mathbf{J}$ and the expected information matrix $-E[\mathbf{J}]$ are given in appendix $\mathbf{A}$.

The multivariate normal approximation can be used to construct approximate confidence intervals and confidence regions for the individual parameters and for the hazard and survival functions.

4.13   Simulation Study

Here, we assess the performance of the maximum likelihood estimates given by (4.30)-(4.34) with respect to sample size n. The assessment is based on a simulation study:

1. Generate a thousand samples of size n from (4.10). The inversion method is used to generate samples, i.e variates of the KumGEV distribution are generated using (4.11).

2. Compute the maximum likelihood estimates for the thousand samples, say $(\widehat{a}_i, \widehat{b}_i, \widehat{\mu}_i, \widehat{\sigma}_i, \widehat{\xi}_i)$ for $i = 1, 2, \ldots, 1000$.

3. Compute the biases and mean squared errors given by

$$bias_h(n) = \frac{1}{1000} \sum_{i=1}^{1000} \left(\widehat{h}_i - h \right)$$

and

$$MSE_h(n) = \frac{1}{1000} \sum_{i=1}^{1000} \left(\widehat{h}_i - h \right)^2$$

for $h = a, b, \mu, \sigma, \xi$.

We repeat these steps for $n = 10, 20, \ldots, 1000$ with $a = 3$, $b = 3$, $\mu = 0$, $\sigma = 1$ and $\xi = 0.5$, so computing $bias_a(n)$, $bias_b(n)$, $bias_\mu(n)$, $bias_\sigma(n)$, $bias_\xi(n)$ and $MSE_a(n)$, $MSE_b(n)$, $MSE_\mu(n)$, $MSE_\sigma(n)$, $MSE_\xi(n)$ for $n = 10, 20, \ldots, 1000$.

Figures 4.6 and 4.7 show how the five biases and the five mean squared errors vary with respect to n. The broken line in Figure 4.6 corresponds to the biases being zero. The following observations can be made:

1. The biases for each parameter are generally positive.

2. Although they appear volatile, the biases for each parameter decrease to zero as $n \to \infty$.

3. The biases appear largest for the parameter b.

4. The biases appear smallest for the parameter a.

5. Although they appear volatile, the mean squared errors for each parameter decrease to zero as $n \to \infty$.

6. The mean squared errors appear largest for the parameters μ, σ and b.

7. the mean squared errors appear smallest for the parameters a and ξ.

We have presented results for only one choice for (a, b, μ, σ, ξ), namely that $(a, b, \mu, \sigma, \xi) = (3, 3, 0, 1, 0.5)$. But the results are analogous for other choices.
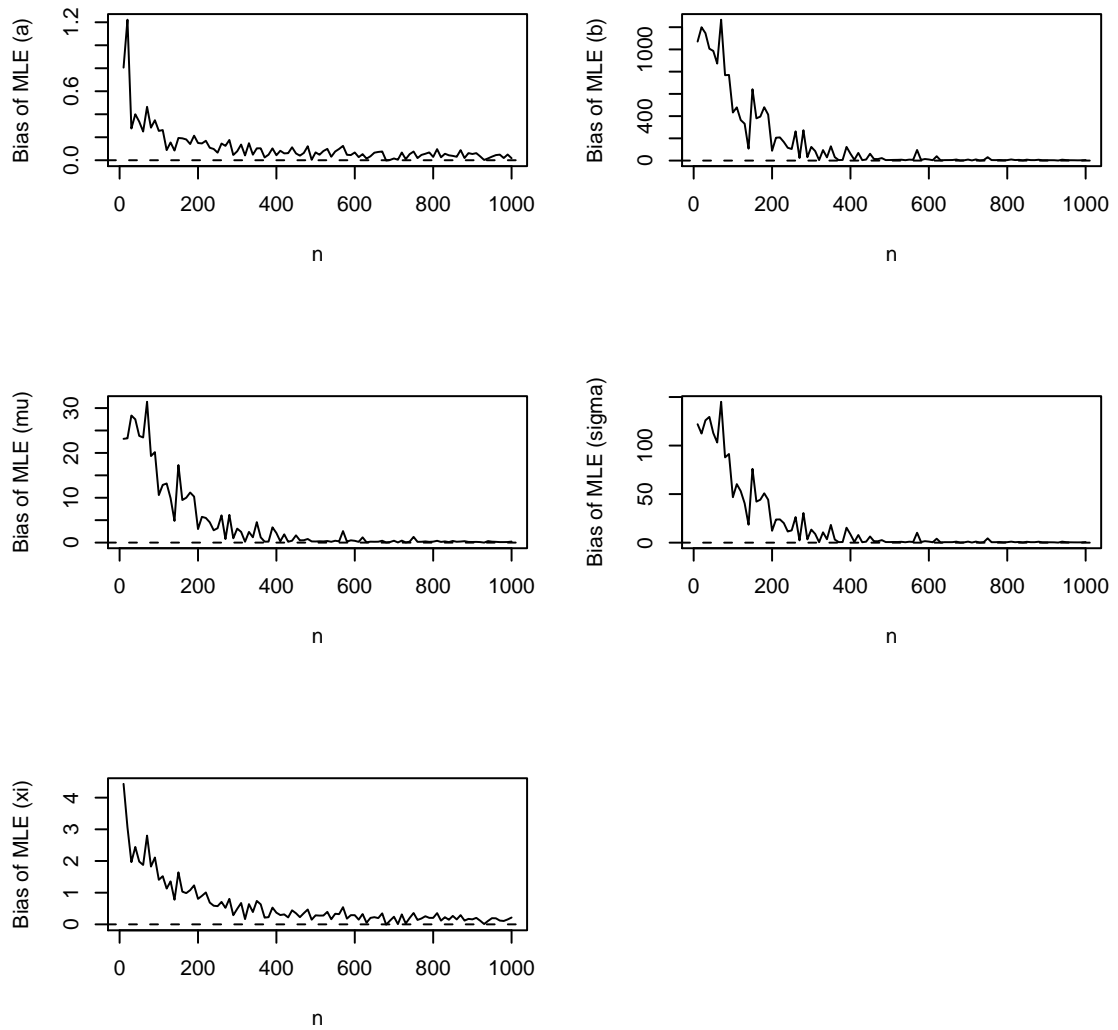
Figure 4.6: $bias_a(n)$ (top left), $bias_b(n)$ (top right), $bias_\mu(n)$ (middle left), $bias_\sigma(n)$ (middle right) and $bias_\xi(n)$ (bottom left) versus $n = 10, 20, \ldots, 1000$.
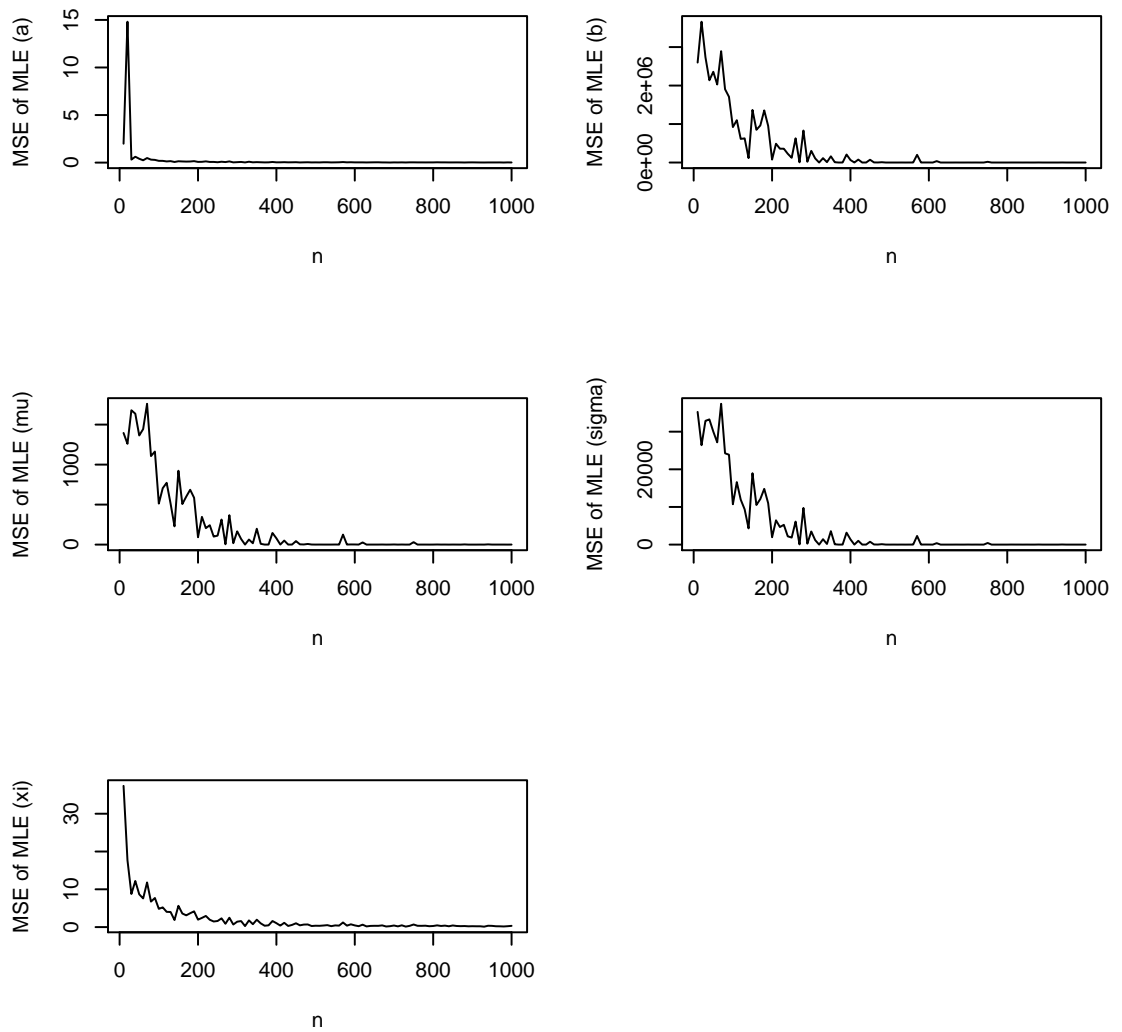
Figure 4.7: $MSE_a(n)$ (top left), $MSE_b(n)$ (top right), $MSE_\mu(n)$ (middle left), $MSE_\sigma(n)$ (middle right) and $MSE_\xi(n)$ (bottom left) versus $n = 10, 20, \ldots, 1000$.

4.14   Application

4.14.1   The Annual Rainfall Maxima Data-Uccle

In this section, we illustrate the flexibility of the KumGEV distribution using a real data set. We use the annual rainfall maxima in millimetres from 1938 to 1972 at Uccle, Belgium, over the duration of one day. This data set is contained as part of the evd contributed package in the R package (R Development (2011)). The data was collected by Sneyers (1977).

We fitted (4.8) and (4.10) to the annual daily maximum rainfall. While fitting (4.10), we have considered the particular case for $a = 1$, in order not to use the full flexibility of the proposed distribution. The maximum likelihood procedure described in section 4.12 was used for fitting (4.10). The results are: $\widehat{\mu} = 28.3824(1.9204)$, $\widehat{\sigma} = 9.0291(1.5793)$, $\widehat{\xi} = 0.2316(0.2133)$ with $-\log L = 136.9071$ for (4.8); $\widehat{b} = 0.1358(0.0233)$, $\widehat{\mu} = 20.5080(1.0703)$, $\widehat{\sigma} = 2.5521(0.0599)$, $\widehat{\xi} = -0.0277(0.0021)$ with $-\log L = 134.9920$ for (4.10). The numbers within brackets are the standard errors computed by inverting the observed information matrix (see Section (4.12)).

We can see that (4.10) gives much smaller standard errors for μ, σ and ξ, suggesting that the proposed distribution can be more accurate. To illustrate the fitting to maximum rainfall data, we calculate the four types of information criteria, which are **AIC**, **BIC**, **AICc**, and **CAIC**. For more details on these statistics see section 2.9.3. Table 4.2 shows the values of theses information criteria. Therefore, the best model is the one with minimum information criteria.

Model	a	b	μ	σ	ξ	AIC	BIC	CAIC	AICc
KumGEV	1	0.1358	20.5080	2.5521	-0.0277	277.98	284.2	288.2	279.6
	(-)	(0.0233)	(1.0703)	(0.0599)	(0.0021)				
GEV	1	1	28.3824	9.0291	0.2316	279.81	284.5	287.5	280.8
	(-)	(-)	(1.9204)	(1.5793)	(0.2133)				

Table 4.2: MLEs of the model parameters for the annual rainfall maxima data, the corresponding SEs (given in parentheses) and the statistics AIC, BIC and CAIC.

The two fitted models, (4.8) and (4.10), are nested. So, they can be compared by the likelihood ratio test (see Table 4.3). Comparing the likelihood values, we see that (4.10) provides a significantly better fit than (4.8). Furthermore, chi-square goodness of fit tests give the p-values of 0.0676 and 0.0373 for (4.8) and (4.10), respectively, suggesting again that (4.10) provides a significantly better fit. Also, the value of the Anderson-Darling and Cramér-von Mises test for the KumGEV is smaller that those of the GEV distribution, suggesting that the KumGEV distribution is more significance that the other.

Model	Hypotheses	Statistics w	P-value
KumGEV vs GEV	$H_o : a = b = 1 vs$ $H_1 : H_o$ is false	3.8302	0.0503

Table 4.3: LR statistics for the annual rainfall maxima data

Model	Anderson-Darling	Cramér-von Mises
KumGEV	0.1695225	0.02260322
GEV	0.3105545	0.04270863

Table 4.4: The Anderson-Darling and Cramr-von Mises statistics for the annual rainfall maxima data

The conclusion based on the likelihood values and the chi-square goodness of fit tests can be verified by means of probability-probability plots, quantile-quantile plots and density plots. For the model given by (4.8), $\exp\{-[1 + \widehat{\xi}(x_{(j)} - \widehat{\mu})/\widehat{\sigma}]^{-1/\widehat{\xi}}\}$ are plotted versus $(j - 0.375)/(n + 0.25)$, $j = 1, 2, \ldots, n$ (as recommended by Blom (1958) and Chambers *et al.* (1983)), where $x_{(j)}$ are the sorted values of the annual daily maximum rainfall and n is the number of observations. For the model given by (4.8), $\widehat{\mu} + (\widehat{\sigma}/\widehat{\xi})\{[-\log((j - 0.375)/(n + 0.25))]^{-\widehat{\xi}} - 1\}$ are plotted versus $x_{(j)}$, $j = 1, 2, \ldots, n$ (as recommended by Blom (1958) and Chambers (1983)).

The probability-probability plots and quantile-quantile plots for the two fitted models

are shown in Figures 4.8 and 4.9. We can see that the model given by (4.10) has the points closer to the diagonal line especially in the upper tail.

The density plots are shown in the Figure 4.10. Again the fitted probability density function for (4.10) appears to capture the general pattern of the empirical histogram better.

Quantities of interest for practitioners of extreme value models are the *return levels*. A T year return level, x_T, is defined as the level that is exceeded on average every T years. For the GEV model given by (4.8),

$$x_T = \mu + \frac{\sigma}{\xi} \left\{ \left[-\log\left(1 - \frac{1}{T}\right) \right]^{-\xi} - 1 \right\}. \tag{4.35}$$

For the KumGEV model given by (4.10),

$$x_T = \mu + \frac{\sigma}{\xi} \left\{ \left[-\frac{1}{a} \log\left\{ 1 - T^{-1/b} \right\} \right]^{-\xi} - 1 \right\}. \tag{4.36}$$

Plots of (4.35) and (4.36) for $T = 2, 3, \ldots, 50$ along with 95 confidence intervals computed by the delta method ((Rao, 1973, pages. 387-389)) are shown in Figure 4.11.

Return levels are important quantities. They are used to determine, for example, dimensions of sea walls, water dams, flood defences, etc. Figure 4.11 suggests that the return levels given by (4.35) and (4.36) do not differ so much. However, the confidence bands for (4.36) appear much narrower and much more realistic. The confidence bands for (4.35) are wide and even take negative values. So, if one were to use the model (4.10) instead of (4.8), significant savings with respect to cost and time could be made.
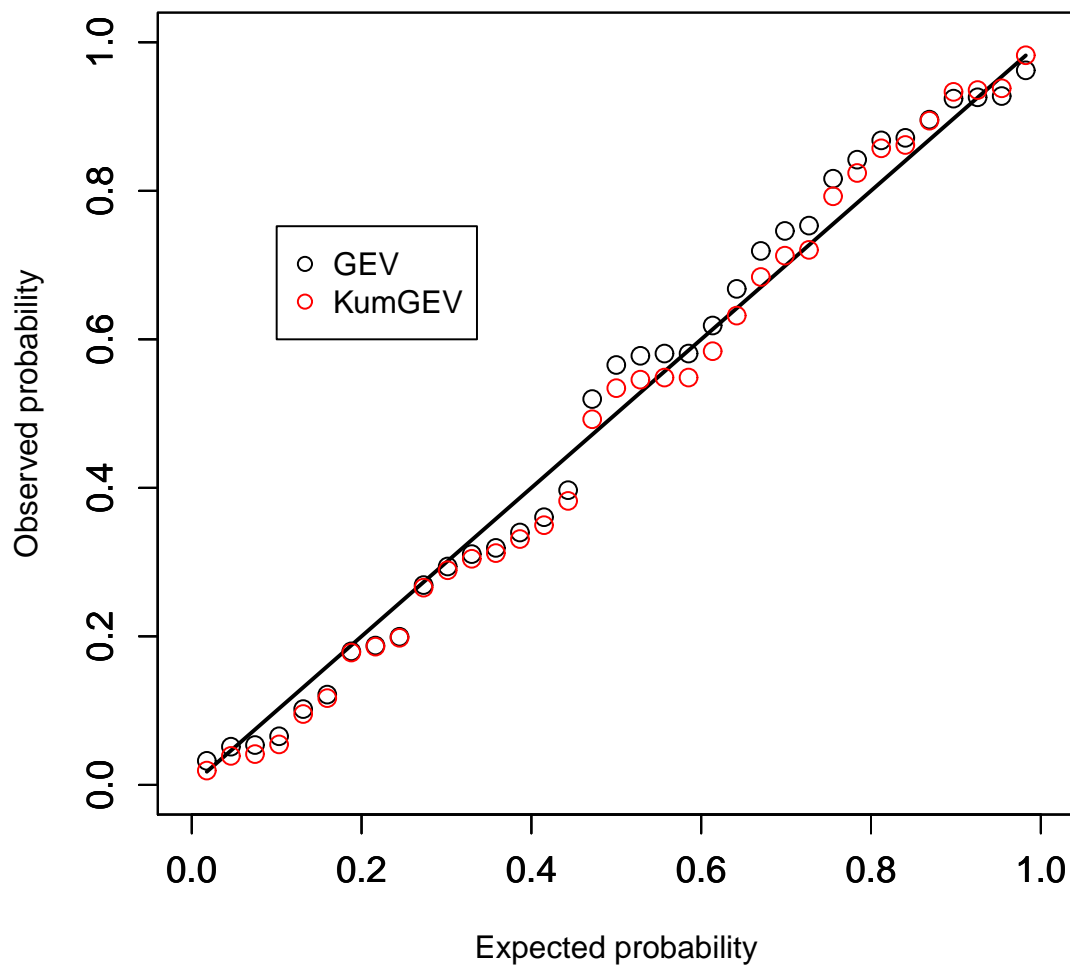
Figure 4.8: Probability plots for the fits of the pdf of the GEV distribution and the pdf of the KumGEV distribution for annual daily rainfall maxima from Uccle, Belgium.
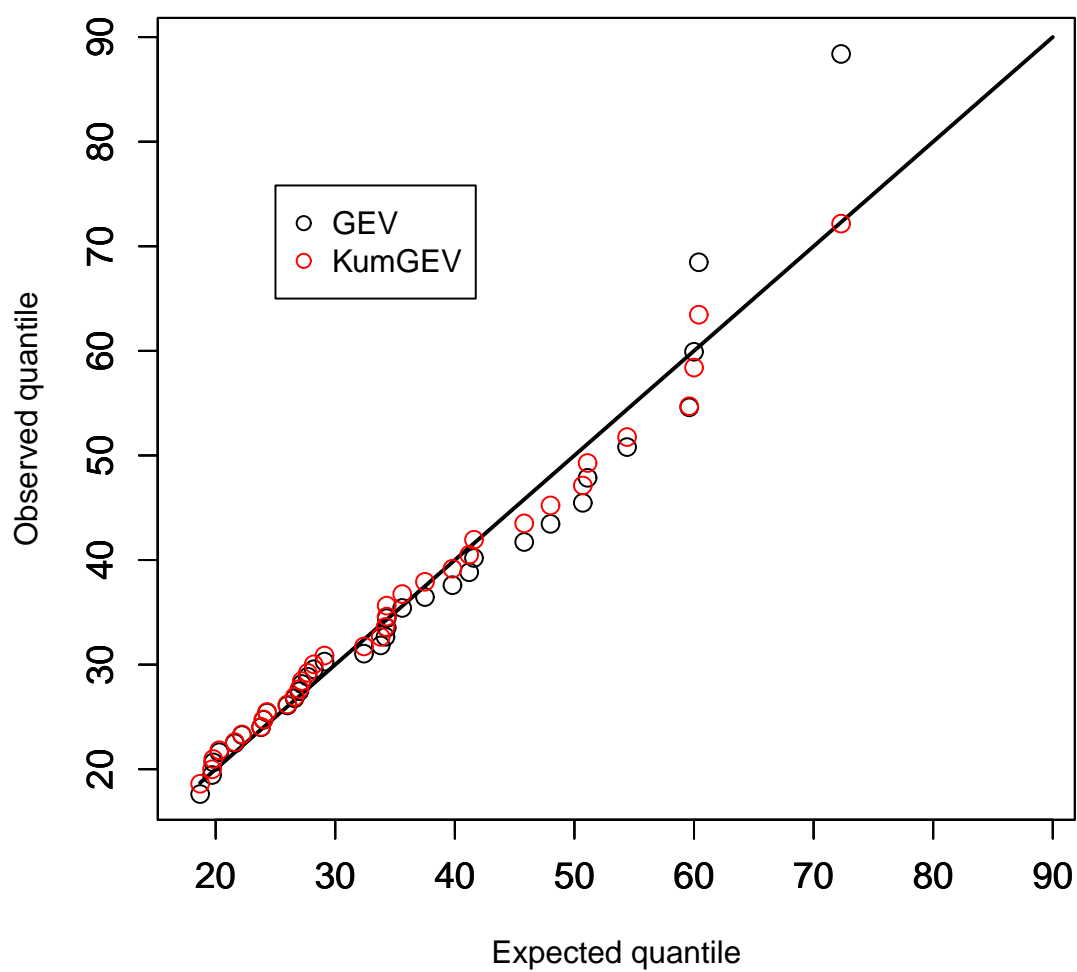
Figure 4.9: Quantile plots for the fits of the pdf of the GEV distribution and the pdf of the KumGEV distribution for annual daily rainfall maxima from Uccle, Belgium.
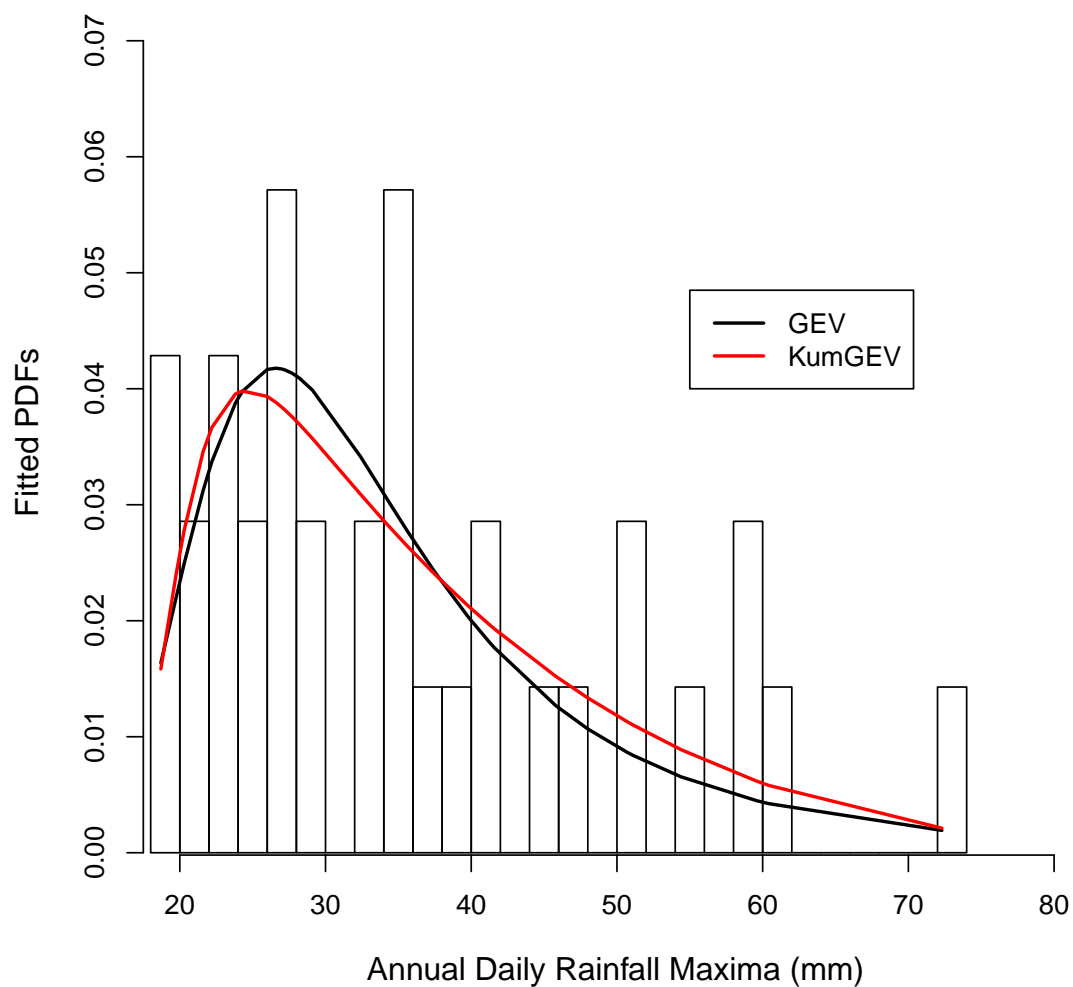
Figure 4.10: Fitted probability density functions to the GEV and KumGEV distributions for annual daily rainfall maxima from Uccle, Belgium.
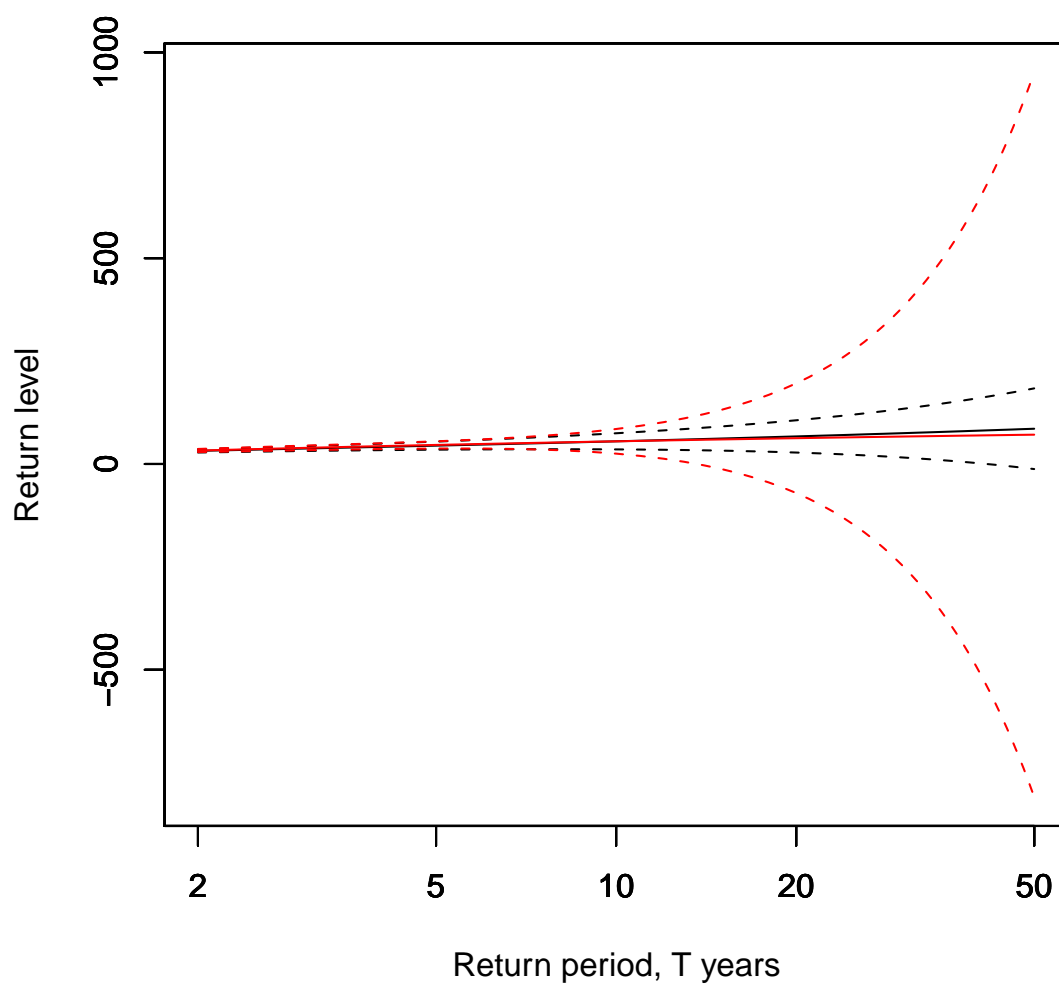
Figure 4.11: Return levels for annual daily rainfall maxima from Uccle, Belgium and their 95 percent confidence intervals for the fits of the pdf of the GEV distribution (in red) and the pdf of the KumGEV distribution (in black).

4.15   The Bivariate KumGEV Distribution

The Kum-G distribution defined by (5.5) can be generalised to the bivariate and multivariate cases in a natural way. Consider the bivariate case for simplicity. Let G denote a bivariate cumulative distribution function with joint probability density function g, marginal probability density functions g_i, $i = 1, 2$ and marginal cumulative distribution functions G_i, $i = 1, 2$. There are various forms of the bivariate Kumaraswamy Generalised extreme value (KumGEV) distributions, but the most common form was introduced by Pickands (1981).

A sensible bivariate generalization of (4.9) is:

$$F\left(x, y\right) = 1 - \left\{1 - G^a\left(x, y; \boldsymbol{\phi}\right)\right\}^b, \tag{4.37}$$

where $a > 0$, $b > 0$ and $\boldsymbol{\phi}$ are specified by G. The generalization is sensible because the two motivations presented in Section 1 also apply to (4.37): this time, motivation could be based on mixtures of bivariate extreme value distributions (not mixtures of GEV distributions) and failures of a system due to two different causes (and not failures of a system due to a single cause).

The marginal probability density functions f_i, $i = 1, 2$ and the marginal cumulative distribution functions F_i, $i = 1, 2$ of F are

$$f_i\left(x\right) = abg_i\left(x\right)G_i^{a-1}\left(x\right)\left\{1 - G_i^a\left(x\right)\right\}^{b-1}$$

and

$$F_i\left(x\right) = 1 - \left\{1 - G_i^a\left(x\right)\right\}^b.$$

The conditional cumulative distribution functions of F are

$$F\left(y \mid x\right) = \frac{1 - \left\{1 - G^a\left(x, y\right)\right\}^b}{1 - \left\{1 - G_1^a\left(x\right)\right\}^b}$$

and

$$F\left(x \mid y\right) = \frac{1 - \left\{1 - G^a\left(x, y\right)\right\}^b}{1 - \left\{1 - G_2^a\left(y\right)\right\}^b}.$$

The joint probability density function of F is

$$f(x, y) = abA(x, y),$$

where

$$
\begin{aligned}
A(x, y) = {} & -a(b-1)\left\{1 - G^a(x, y)\right\}^{b-2} G^{2a-2}(x, y)\frac{\partial G(x, y)}{\partial x}\frac{\partial G(x, y)}{\partial y} \\
& + (a-1)\left\{1 - G^a(x, y)\right\}^{b-1} G^{a-2}(x, y)\frac{\partial G(x, y)}{\partial x}\frac{\partial G(x, y)}{\partial y} \\
& + \left\{1 - G^a(x, y)\right\}^{b-1} G^{a-1}(x, y)\frac{\partial^2 G(x, y)}{\partial x \partial y}.
\end{aligned}
$$

4.15.1 Maximum Likelihood Estimation for the Bivariate Kum-GEV Distribution

Finally, we consider maximum likelihood estimation of the parameters of (4.37). Suppose (x_1, y_1), (x_2, y_2), $\ldots$, (x_n, y_n) is a random sample of size n from (4.37). Then the log-likelihood function for the vector of parameters $(a, b, \boldsymbol{\phi})$ can be written as

$$\log L(a, b, \boldsymbol{\phi}) = n\log(ab) + \sum_{i=1}^{n} \log A(x_i, y_i). \tag{4.38}$$

The first-order partial derivatives of (4.38) with respect to the parameters are:

$$
\frac{\partial \log L}{\partial a} = \frac{n}{a} + \sum_{i=1}^{n} \frac{B(x_i, y_i)}{A(x_i, y_i)},
$$

$$
\frac{\partial \log L}{\partial b} = \frac{n}{b} + \sum_{i=1}^{n} \frac{C(x_i, y_i)}{A(x_i, y_i)},
$$

$$
\frac{\partial \log L}{\partial \boldsymbol{\phi}} = \sum_{i=1}^{n} \frac{\mathbf{D}(x_i, y_i)}{A(x_i, y_i)},
$$

where

$$B\left(x_i, y_i\right) = (1-b)\left[1 - G^a\left(x_i, y_i\right)\right]^{b-2} G^{2a-2}\left(x_i, y_i\right) \frac{\partial G\left(x_i, y_i\right)}{\partial x_i} \frac{\partial G\left(x_i, y_i\right)}{\partial y_i}$$

$$+ a(b-1)(b-2)\left[1 - G^a\left(x_i, y_i\right)\right]^{b-3} G^{3a-2}\left(x_i, y_i\right) \log G\left(x_i, y_i\right)$$

$$\times \frac{\partial G\left(x_i, y_i\right)}{\partial x_i} \frac{\partial G\left(x_i, y_i\right)}{\partial y_i} - 2a(b-1)\left[1 - G^a\left(x_i, y_i\right)\right]^{b-2} G^{2a-2}\left(x_i, y_i\right)$$

$$\times \log G\left(x_i, y_i\right) \frac{\partial G\left(x_i, y_i\right)}{\partial x_i} \frac{\partial G\left(x_i, y_i\right)}{\partial y_i} + \left[1 - G^a\left(x_i, y_i\right)\right]^{b-1} G^{a-1}\left(x_i, y_i\right)$$

$$\times \frac{\partial G\left(x_i, y_i\right)}{\partial x_i} \frac{\partial G\left(x_i, y_i\right)}{\partial y_i} - (a-1)(b-1)\left[1 - G^a\left(x_i, y_i\right)\right]^{b-2} G^{2a-1}\left(x_i, y_i\right)$$

$$\times \log G\left(x_i, y_i\right) \frac{\partial G\left(x_i, y_i\right)}{\partial x_i} \frac{\partial G\left(x_i, y_i\right)}{\partial y_i} + (a-1)\left[1 - G^a\left(x_i, y_i\right)\right]^{b-1}$$

$$\times G^{a-1}\left(x_i, y_i\right) \log G\left(x_i, y_i\right) \frac{\partial G\left(x_i, y_i\right)}{\partial x_i} \frac{\partial G\left(x_i, y_i\right)}{\partial y_i}$$

$$- (b-1)\left[1 - G^a\left(x_i, y_i\right)\right]^{b-2} G^{2a-1}\left(x_i, y_i\right) \log G\left(x_i, y_i\right) \frac{\partial^2 G\left(x_i, y_i\right)}{\partial x_i \partial y_i}$$

$$+ \left[1 - G^a\left(x_i, y_i\right)\right]^{b-1} G^{a-1}\left(x_i, y_i\right) \log G\left(x_i, y_i\right) \frac{\partial^2 G\left(x_i, y_i\right)}{\partial x_i \partial y_i},$$

$$C\left(x_i, y_i\right) = -a\left[1 - G^a\left(x_i, y_i\right)\right]^{b-2} G^{2a-2}\left(x_i, y_i\right) \frac{\partial G\left(x_i, y_i\right)}{\partial x_i} \frac{\partial G\left(x_i, y_i\right)}{\partial y_i}$$

$$- a(b-1)\left[1 - G^a\left(x_i, y_i\right)\right]^{b-2} G^{2a-2}\left(x_i, y_i\right) \log\left[1 - G^a\left(x_i, y_i\right)\right]$$

$$\times \frac{\partial G\left(x_i, y_i\right)}{\partial x_i} \frac{\partial G\left(x_i, y_i\right)}{\partial y_i} + (a-1)\left[1 - G^a\left(x_i, y_i\right)\right]^{b-1}$$

$$\times G^{a-1}\left(x_i, y_i\right) \log\left[1 - G^a\left(x_i, y_i\right)\right] \frac{\partial G\left(x_i, y_i\right)}{\partial x_i} \frac{\partial G\left(x_i, y_i\right)}{\partial y_i}$$

$$+ \left[1 - G^a\left(x_i, y_i\right)\right]^{b-1} G^{a-1}\left(x_i, y_i\right) \log\left[1 - G^a\left(x_i, y_i\right)\right] \frac{\partial^2 G\left(x_i, y_i\right)}{\partial x_i \partial y_i},$$

and

$$
\mathbf{D}\left(x_i, y_i\right) = \left[1 - G^a\left(x_i, y_i\right)\right]^{b-1} G^{a-1}\left(x_i, y_i\right) \frac{\partial^3 G\left(x_i, y_i\right)}{\partial x_i \partial y_i \partial \phi}
$$

$$
-2a(a-1)(b-1)\left[1 - G^a\left(x_i, y_i\right)\right]^{b-2} G^{2a-3}\left(x_i, y_i\right) \frac{\partial G\left(x_i, y_i\right)}{\partial x_i} \frac{\partial G\left(x_i, y_i\right)}{\partial y_i}
$$

$$
\times \frac{\partial G\left(x_i, y_i\right)}{\partial \phi} - a(b-1)\left[1 - G^a\left(x_i, y_i\right)\right]^{b-2} G^{2a-2}\left(x_i, y_i\right) \frac{\partial G\left(x_i, y_i\right)}{\partial y_i}
$$

$$
\times \frac{\partial^2 G\left(x_i, y_i\right)}{\partial x_i \partial \phi} - a(b-1)\left[1 - G^a\left(x_i, y_i\right)\right]^{b-2} G^{2a-1}\left(x_i, y_i\right) \frac{\partial G\left(x_i, y_i\right)}{\partial x_i}
$$

$$
\times \frac{\partial^2 G\left(x_i, y_i\right)}{\partial y_i \partial \phi} - a(a-1)(b-1)\left[1 - G^a\left(x_i, y_i\right)\right]^{b-2} G^{2a-2}\left(x_i, y_i\right)
$$

$$
\times \frac{\partial G\left(x_i, y_i\right)}{\partial x_i} \frac{\partial G\left(x_i, y_i\right)}{\partial y_i} \frac{\partial G\left(x_i, y_i\right)}{\partial \phi} + (a-1)^2 \left[1 - G^a\left(x_i, y_i\right)\right]^{b-1}
$$

$$
\times G^{a-2}\left(x_i, y_i\right) \frac{\partial G\left(x_i, y_i\right)}{\partial x_i} \frac{\partial G\left(x_i, y_i\right)}{\partial y_i} \frac{\partial G\left(x_i, y_i\right)}{\partial \phi}
$$

$$
+(a-1)\left[1 - G^a\left(x_i, y_i\right)\right]^{b-1} G^{a-1}\left(x_i, y_i\right) \frac{\partial G\left(x_i, y_i\right)}{\partial y_i} \frac{\partial^2 G\left(x_i, y_i\right)}{\partial x_i \partial \phi}
$$

$$
+(a-1)\left[1 - G^a\left(x_i, y_i\right)\right]^{b-1} G^{a-1}\left(x_i, y_i\right) \frac{\partial G\left(x_i, y_i\right)}{\partial x_i} \frac{\partial^2 G\left(x_i, y_i\right)}{\partial y_i \partial \phi}
$$

$$
-a(b-1)\left[1 - G^a\left(x_i, y_i\right)\right]^{b-2} G^{2a-2}\left(x_i, y_i\right) \frac{\partial^2 G\left(x_i, y_i\right)}{\partial x_i \partial y_i} \frac{\partial G\left(x_i, y_i\right)}{\partial \phi}
$$

$$
+(a-1)\left[1 - G^a\left(x_i, y_i\right)\right]^{b-1} G^{a-2}\left(x_i, y_i\right) \frac{\partial^2 G\left(x_i, y_i\right)}{\partial x_i \partial y_i} \frac{\partial G\left(x_i, y_i\right)}{\partial \phi}
$$

$$
+a^2(b-1)(b-2)\left[1 - G^a\left(x_i, y_i\right)\right]^{b-3} G^{3a-3}\left(x_i, y_i\right) \frac{\partial G\left(x_i, y_i\right)}{\partial x_i}
$$

$$
\times \frac{\partial G\left(x_i, y_i\right)}{\partial y_i} \frac{\partial G\left(x_i, y_i\right)}{\partial \phi}.
$$

The maximum likelihood estimators of $(a, b, \boldsymbol{\phi})$, say $(\widehat{a}, \widehat{b}, \widehat{\boldsymbol{\phi}})$, are the simultaneous solutions of the equations $\partial \log L / \partial a = 0$, $\partial \log L / \partial b = 0$ and $\partial \log L / \partial \boldsymbol{\phi} = \mathbf{0}$. As $n \to \infty$, $(\widehat{a} - a, \widehat{b} - b, \widehat{\boldsymbol{\phi}} - \boldsymbol{\phi})$ approaches a multivariate normal vector with zero means and variance-covariance matrix, $-(E\mathbf{J})^{-1}$, where

$$
\mathbf{J} = \begin{pmatrix} \dfrac{\partial^2 \log L}{\partial a^2} & \dfrac{\partial^2 \log L}{\partial a \partial b} & \dfrac{\partial^2 \log L}{\partial a \partial \phi} \\[2mm] \dfrac{\partial^2 \log L}{\partial b \partial a} & \dfrac{\partial^2 \log L}{\partial b^2} & \dfrac{\partial^2 \log L}{\partial b \partial \phi} \\[2mm] \dfrac{\partial^2 \log L}{\partial \phi \partial a} & \dfrac{\partial^2 \log L}{\partial \phi \partial b} & \dfrac{\partial^2 \log L}{\partial \phi^2} \end{pmatrix}.
$$

As suggested in Section 4.12, it is best to approximate the distribution of $(\widehat{a} - a, \widehat{b} - b, \widehat{\boldsymbol{\phi}} - \boldsymbol{\phi})$ by a multivariate normal distribution with zero means and variance-covariance matrix given by $-\mathbf{J}^{-1}$, inverse of the observed information matrix, with

(a, b, ϕ) replaced $(\widehat{a}, \widehat{b}, \widehat{\phi})$. Explicit expressions for the elements of $\mathbf{J}$ are too long to be presented here.

Two most common extreme value forms for G (Gumbel and Mustafi (1967)) are

$$G\left(x, y\right) = G_1(x)G_2(y) \exp\left\{-\theta \left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-1}\right\} \qquad (4.39)$$

for $0 \leq \theta \leq 1$, and

$$G\left(x, y\right) = \exp\left\{-\left[(\log G_1(x))^\theta + (\log G_2(y))^\theta\right]^{1/\theta}\right\} \qquad (4.40)$$

for $1 \leq \theta \leq \infty$. To compute the maximum likelihood estimators as well as the observed information matrix, one needs the partial derivatives $\partial G/\partial x$, $\partial G/\partial y$, $\partial G/\partial \theta$, $\partial^2 G/\partial x \partial y$, $\partial^2 G/\partial x \partial \theta$, $\partial^2 G/\partial y \partial \theta$, and $\partial^3 G/\partial x \partial y \partial \theta$. Explicit expressions for these partial derivatives for (4.39) and (4.40) are given in appendix $\mathbf{A}$.

We saw in Sections 4.1, 4.4 and 4.5 how the parameters a and b control the marginal tails of (4.37). It is also of interest to know how the parameters control the joint tail of (4.37). There are several measures for dependence in the joint tail. One popular measure due to Coles et al. (1999) is

$$\chi = \lim_{y \to \infty} \frac{1 - F(y, \infty) - F(\infty, y) + F(y, y)}{1 - F(y, \infty)}. \qquad (4.41)$$

It is difficult to derive an expression for (4.41) for the general form in (4.37). However, one can show that (4.41) reduces to $2 - (2 - \theta/2)^b$ and $2 - 2^{b/\theta}$ for the particular cases in (4.39) and (4.40), respectively, if $G_1(\cdot)$ and $G_2(\cdot)$ are identical. Hence, only the parameter b appears to control the dependence in the joint tail. As b increases the amount of dependence decreases. Asymptotic independence for (4.39) corresponds to $b = \log 2/\log(2 - \theta/2)$. Asymptotic independence for (4.40) corresponds to $b = \theta$.

4.16   Summary

In this chapter, we study some mathematical properties of a generalisation of the GEV distribution, the Kumaraswamy GEV (KumGEV) distribution. This can be quite flexible in the analysis of continuous data in some area of research. We provide

its moments and characteristics function. Additionally, explicit expressions are derived for the mean deviations and Rényi and Shannon Entropies. We obtain explicit expressions for the moments of order statistics and examine the asymptotic distributions of extreme order statistics. The estimation of parameters is performed by MLE method. Also, the usefulness of the KumGEV distribution is illustrated by applying it to annual maximum rainfall data at Uccle, Belgium.

Based on some statistical measures, we conclude that the KumGEV distribution appears to be more appropriate than the GEV distribution for the annual maximum rainfall data.

Chapter 5

KumGP distribution

5.1 Introduction

In extreme value theory, we are not only interested in maximum and minimum values, we use other extremes data if there are available. Let $X_1, X_2, ..., X_n$ be the original collected data, one technique to extract higher extremes from a set of data is to take the exceedances y_i over high threshold t. So, we interested in a set of observations $x_1, x_2, ...$ that exceed the threshold t, $x_i > t$. In some instances the observations below the threshold are not recorded, due to they not being considered important in the end. Pickands III (1975) and Balkema and De Haan (1974) introduced the generalised Pareto (GP) distribution as the limiting distribution for exceedances (shortfalls) over (below) a higher (lower) threshold. This has since been widely used to model geophysical phenomena such as floods or extreme windstorms, due to its close relationship with the extreme value distributions. Also, it is a most popular distribution in a financial risk modelling. The GP distribution has a distribution function

$$F_{(\sigma, \xi)}(x) = \begin{cases} 1 - [1 + \xi(x-t)/\sigma]^{-1/\xi} & \text{if} \quad \xi \neq 0, \\ 1 - \exp{-(x-t)/\sigma} & \text{if} \quad \xi = 0. \end{cases}$$

Where σ is the scale parameter and ξ is the shape parameter. The domain of x depends on ξ; $x \geq t$, when $\xi \geq 0$, or $x \geq t$, and $x \leq t - \frac{\sigma}{\xi}$ when $\xi < 0$. Some known distributions occur as special cases, depending on the value of the shape parameter

ξ. Cases $\xi = 0$, $t = 0$ and $\xi = 1$, $t = 0$ yield, exponential distribution with mean σ, and uniform distribution on the interval $(0, \sigma)$ respectively. The Pareto distribution is obtained when $\xi > 0$, $t = \sigma/\xi$, and it becomes the triangular distribution when $\xi = -1/2$, $t = 0$.

The GP distribution is one of the most widely applied models for univariate extreme values such as lifetime data analysis, the coupon collector's problem, analysis of radio audience data, analysis of rainfall time series, comparing investment risk between Chinese and American stock markets, regional flood frequency analysis, drought modeling, value at risk, analysis of turbine steady-state, second-order material property closures, wind extremes, analysis of motor liability insurance, analysis of finite buffer queues, river flow modeling, measuring liquidity risk of open-end funds, modeling of extreme earthquake events, estimation of the maximum inclusion size in clean steels, and modeling of high-concentrations in short-range atmospheric dispersion.

For more details on the GP distribution, its theory and further, we refer the reader to Leadbetter et al. (1983), Embrechts (1999), Castillo et al. (2005), and Resnick (2007). Even though, it is widely applied in some fields of research, it has also been misused in many areas. It does not give adequate fits in many areas, as can be seen in some of the research.

For example, Madsen and Rosbjerg (1998) find that the GP distribution does not give a good fit to drought deficit volumes due to many small drought events. In an illustrative example of the SAS/ETS SEVERITY procedure, Joshi (2010) finds that "both plots indicate that the Exp , Pareto, and GP distributions are a poor fit".

We propose a simple generalization of the GP distribution. We provide a motivation for this simple generalization. This is based on the definition of the GP distribution. The GP distribution arises as the conditional distribution of exceedances of a process over a large threshold (Pickands III (1975)). If $F(\cdot)$ denotes the cumulative distribution function of the process then we can write

$$1 - F(x) \approx p \left(1 + \xi \frac{x - t}{\sigma} \right)^{-1/\xi} \tag{5.1}$$

for $x > t$ and some large t, where $p = 1 - F(t)$, $x > t$ if $\xi \geq 0$, $t < x \leq t - \sigma/\xi$ if $\xi < 0$, $-\infty < \xi < \infty$ is a shape parameter and $\sigma > 0$ is a scale parameter. One way to improve on (5.1) is to take a mixture of GP cumulative distribution functions. That is, write

$$1 - F(x) \approx \sum_{i=1}^{k} w_i \left(1 + \xi_i \frac{x - t}{\sigma_i} \right)^{-1/\xi_i} \tag{5.2}$$

for $x > t$ and some large t. But mixtures of the form (5.2) are notoriously difficult to handle not just because of the complicated mathematical form. Inferences and fitting of (5.2) are also difficult. For example, on the subject of estimating a mixture of Pareto distributions, Bee et al. (2009) say "Application of standard techniques to a mixture of Pareto is problematic". Indeed, applications of mixtures of Pareto distributions have been very limited.

A way around is to rewrite (5.2) in a simple mathematical form. There are many choices for the mathematical form. A choice motivated by the works of Kumaraswamy (1980) and Cordeiro and de Castro (2011) is

$$1 - F(x) = \left\{ 1 - G(x)^a \right\}^b, \tag{5.3}$$

where $G(\cdot)$ denotes a GP cumulative distribution function and $a > 0$, $b > 0$ are two additional parameters whose role is partly to introduce skewness and to vary tail weights. Note that the right hand side of (5.3) can be expanded as

$$\left\{ 1 - G(x)^a \right\}^b = \sum_{i=0}^{\infty} c_i \left[1 - G(x) \right]^{b+i}, \tag{5.4}$$

a mixture taking the form of (5.2). The coefficients c_i are functions of a and b. For instance, $c_0 = a^b$. The parameter b mainly dictates the tail behaviors of the mixture components. The parameter a mainly dictates the mixture coefficients.

Following the terminology used in Cordeiro and de Castro (2011), we shall refer to the distribution given by (5.3) as the KumGP distribution. The probability density function corresponding to (5.3) is

$$f(x) = ab\, g(x)\, G(x)^{a-1} \left\{ 1 - G(x)^a \right\}^{b-1}, \tag{5.5}$$

where $g(x) = dG(x)/dx$ is a GP probability density function. Because $g(\cdot)$ and $G(\cdot)$ are tractable, the KumGP distribution can be used quite effectively even if the data are censored. Moreover, existing software for the GP distribution (say, to compute probability density function, cumulative distribution function, quantile function, moments, maximum likelihood estimates, random numbers, etc) can be easily adapted for the KumGP distribution. Clearly, the GP distribution is a special case of the KumGP distribution for $a = b = 1$ with a continuous crossover towards cases with different shapes (for example, a particular combination of skewness and kurtosis).

The role of the two additional parameters, $a > 0$ and $b > 0$, is to govern skewness and generate distributions with heavier/ligther tails. If $a < 1$ then the tails of $f(\cdot)$ will be heavier than those of $g(\cdot)$. Similarly, if $b < 1$ then the tails of $f(\cdot)$ will be heavier than those of $g(\cdot)$. On the other hand, if $a > 1$ then the tails of $f(\cdot)$ will be lighter than those of $g(\cdot)$. Similarly, if $b > 1$ then the tails of $f(\cdot)$ will be lighter than those of $g(\cdot)$. Further description of the role of a and b is given in Section 5.4.

Another physical interpretation for the KumGP distribution when a and b are positive integers is as follows. Suppose a system is made of b independent components and that each component is made up of a independent subcomponents. Suppose the system fails if any of the b components fails and that each component fails if all of the a subcomponents fail. Let $X_{j1}, X_{j2}, \ldots, X_{ja}$ denote the lifetimes of the subcomponents within the jth component, $j = 1, 2, \ldots, b$ with a common GP cumulative distribution function. Let X_j denote the lifetime of the jth component, $j = 1, \ldots, b$, and let X denote the lifetime of the entire system. So, the cumulative distribution function of X is

$$\Pr(X \le x) = 1 - \Pr^b(X_1 > x) = 1 - \{1 - \Pr(X_1 \le x)\}^b$$

$$= 1 - \{1 - \Pr(X_{11} \le x, X_{12} \le x, \ldots, X_{1a} \le x)\}^b$$

$$= 1 - \{1 - \Pr^a(X_{11} \le x)\}^b = 1 - \{1 - G_{\xi,\sigma}^a(x)\}^b.$$

So, it follows that the KumGP distribution given by (5.3) and (5.5) is precisely the time to failure distribution of the entire system. The GP distribution has been widely

used to model lifetimes: see, for example, Mahmoudi (2011).

There are other ways to generalise the GP distribution; the most recent generalizations of the GP distribution were proposed by Papastathopoulos and Tawn (2012), a paper currently in press for the *Journal of Statistical Planning and Inference.* They referred to their generalizations as EGP1, EGP2 and EGP3 distributions. The EGP1 distribution is specified by the cumulative distribution function

$$F(x) = \frac{1}{B\left(\kappa, 1/\mid \xi \mid\right)} B_{1-\left(1+\xi\frac{x}{\sigma}\right)^{-|\xi|/\xi}}\left(\kappa, 1/\mid \xi \mid\right) \tag{5.6}$$

for $x > 0$ (if $\xi \geq 0$), $0 < x \leq -\sigma/\xi$ (if $\xi < 0$), $\sigma > 0$, $\kappa > 0$ and $-\infty < \xi < \infty$, where $B_x(\cdot, \cdot)$ denotes the incomplete beta function defined by

$$B_x(a, b) = \int_0^x t^{a-1}(1-t)^{b-1}dt.$$

The EGP2 distribution is specified by the cumulative distribution function

$$F(x) = \frac{1}{\Gamma(\kappa)}\gamma\left[\frac{1}{\xi}\ln\left(1+\xi\frac{x}{\sigma}\right), \kappa\right] \tag{5.7}$$

for $x > 0$ (if $\xi \geq 0$), $0 < x \leq -\sigma/\xi$ (if $\xi < 0$), $\sigma > 0$, $\kappa > 0$ and $-\infty < \xi < \infty$, where $\Gamma(\cdot)$ denotes the gamma function defined by

$$\Gamma(a) = \int_0^\infty t^{a-1}\exp(-t)dt,$$

and $\gamma(\cdot, \cdot)$ denotes the incomplete gamma function defined by

$$\gamma(a, x) = \int_0^x t^{a-1}\exp(-t)dt.$$

The EGP3 distribution is specified by the cumulative distribution function

$$F(x) = \left\{1 - \left(1+\xi\frac{x}{\sigma}\right)^{-1/\xi}\right\}^{\kappa} \tag{5.8}$$

for $x > 0$ (if $\xi \geq 0$), $0 < x \leq -\sigma/\xi$ (if $\xi < 0$), $\sigma > 0$, $\kappa > 0$ and $-\infty < \xi < \infty$.

Unfortunately, none of the distributions given by (5.6)-(5.8) are new. There have been many papers published proposing distributions same as (5.6)-(5.8) or containing (5.6)-(5.8) as special cases. However, none of these papers have been cited by Papastathopoulos and Tawn (2012). Besides, (5.6) and (5.7) involve the incomplete beta

function and the incomplete gamma function, special functions requiring numerical routines. We shall also see later that none of (5.6)-(5.8) provide significant improvements over the GP distribution for the data set considered in Papastathopoulos and Tawn (2012).

Now, let us explain why the distributions given by (5.6)-(5.8) are not new. Firstly, (5.6) is a special case of the class of *beta-G distributions* introduced by Eugene et al. (2002), and followed by Jones (2004) and many others. The beta-G distribution is specified by the cumulative distribution function

$$F(x) = \frac{1}{B(a,b)} \int_0^{G(x)} t^{a-1}(1-t)^{b-1} dt \tag{5.9}$$

for $a > 0$ and $b > 0$, where $B(\cdot, \cdot)$ denotes the beta function defined by

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt.$$

Note that (5.6) is a special case of (5.9) for $G(\cdot)$ specified by

$$G(x) = 1 - \left(1 + \xi \frac{x}{\sigma}\right)^{-|\xi|/\xi}.$$

This a special case is considered in detail by (Akinsete et al., 2008, sec. 2.2), Mahmoudi (2011) and many others.

Secondly, (5.7) is a special case of the class of *gamma-G distributions* introduced by Zografos and Balakrishnan (2009), and followed by Ristić and Balakrishnan (2012) and many others. The gamma-G distribution is specified by the cumulative distribution function

$$F(x) = \frac{\gamma(a, -\log[1 - G(x)])}{\Gamma(a)} \tag{5.10}$$

for $a > 0$. Note that (5.7) is a special case of (5.10) when $G(\cdot)$ is a GP cumulative distribution function. Furthermore, the formula for the cumulative distribution function of the EGP2 distribution given in Papastathopoulos and Tawn (2012) is not a cumulative distribution function at all.

Finally, (5.8) is identical to the exponentiated Pareto distribution studied by Adeyemi and Adebanji (2005), Shawky and Abu-Zinadah (2008, 2009), Afify (2010) and many others.

5.2   The KumGP Distribution and Sub-Models

Let $G_{(\sigma,\xi)}$ and $g_{(\sigma,\xi)}$ denote the cdf and the pdf of the GP distribution, respectively. From now on, the G and g functions will be written as

$$G_{(\sigma,\xi)}(x) = 1 - u, \tag{5.11}$$

and

$$g_{(\sigma,\xi)}(x) = \sigma^{-1}u^{1+\xi}, \tag{5.12}$$

respectively, for $x > t$ if $\xi \geq 0$, $t < x \leq t - \sigma/\xi$ if $\xi < 0$, $\sigma > 0$, $-\infty < \xi < \infty$, and $u = \{1 + \xi(x - t)/\sigma\}^{-1/\xi}$. The cdf and the pdf of the KumGP distribution are given by

$$F(x) = 1 - \left\{1 - (1 - u)^a\right\}^b, \tag{5.13}$$

and

$$f(x) = \sigma^{-1}abu^{1+\xi}(1 - u)^{a-1}\left\{1 - (1 - u)^a\right\}^{b-1}, \tag{5.14}$$

respectively. The EGP3 distribution given by (5.8) is a particular case of the KumGP distribution. Unlike the EGP1 and EGP2 distributions, the KumGP distribution does not involve special functions. So, one can expect that the KumGP distribution will attract wider applicability than the EGP1, EGP2 and EGP3 distributions.

As has been indicated in the introduction, the KumGP distribution has many sub-models. In this section, we introduce some distributions, which can arise as special cases from KumGP distribution.

- **GP distribution**:

  If $a = b = 1$ then the KumGP distribution reduces to

  $$f_{(\sigma,\xi)}(x) = \sigma^{-1}[1 + \xi(x - t)/\sigma]^{-(1+\xi)/\xi},$$

  which is the pdf of the GP distribution with the scale parameter $\sigma > 0$ and the shape parameter $-\infty < \xi < \infty$.

- **Pareto distribution**:

  For $a = b = 1$, $\xi > 0$ and $t = \sigma/\xi$ the KumGP density function yields

  $$f_{(\sigma,\xi)}(x) = \frac{\xi}{\sigma}(\frac{x-t}{\sigma})^{-(\xi+1)},$$

  which is the pdf of the Pareto distribution with the scale parameter $\sigma > 0$ and the shape parameter $\xi > 0$.

- **Exponential distribution**:

  For $a = b = 1$, $\xi = 0$ and $t = 0$ the KumGP distribution becomes

  $$f_{\sigma}(x) = \frac{1}{\sigma}\exp{-x/\sigma},$$

  which is the pdf of the exponential distribution with mean σ

- **Uniform distribution**:

  For $a = b = 1$, $\xi = -1$ and $t = 0$ the KumGP distribution reduces to

  $$f_{\sigma}(x) = \frac{1}{\sigma},$$

  the pdf of the uniform distribution defined on the interval $(0, \sigma)$

- **Triangular distribution**:

  For $a = b = 1$, $\xi = -1/2$ and $t = 0$ the KumGP distribution becomes

  $$f_{\sigma}(x) = \begin{cases} \frac{2x}{\sigma c} & 0 \geq x \leq c, \\ \frac{2(\sigma-x)}{\sigma(\sigma-c)} & c < x < \sigma, \end{cases}$$

  which is the triangular distribution on $[0, \sigma]$ with mode c, where $0 < c < \sigma$.

- **Lomax distribution**:

  For $a = b = 1$, $\xi = 1/\alpha, \sigma = \lambda/\alpha$ and $t = 0$ the KumGP distribution becomes

  $$f_{\alpha,\lambda}(x) = \frac{\alpha}{\lambda}(1 + \frac{x}{\lambda})^{-(\alpha+1)}$$

  which is the Lomax distribution with parameters α and λ. The Lomax distribution is also known as Pareto type II distribution.

	Parameters				
Distribution	a	b	t	σ	ξ
Exponential	1	1	0	-	0
Lomax	1	1	0	λ/α	$1/\alpha$
Uniform	1	1	0	-	-1
Triangular	1	1	0	-	$-1/2$
Pareto	1	1	σ/ξ	-	>0
GPD	1	1	-	-	-
KumGPD	-	-	-	-	-

Table 5.1: Some sub-models of the KumGP distribution

In the Table 5.1 we review these sub-models for the KumGP distribution.

The KumGP distribution given by (5.14) is much more flexible than the GP distribution and can allow for greater flexibility of tails. Plots of the pdf for the GP and KumGP distributions for selected parameter values are shown in the Figures 5.1-5.2

To generate a random sample from the KumGP distribution, assume that X has probability density function (**??**). The quantile function of the KumGP distribution can be written as

$$x = Q(z) = t + \frac{\sigma}{\xi} \left\{ \left\{ 1 - \left[1 - (1-z)^{1/b} \right]^{1/a} \right\}^{-\xi} - 1 \right\}. \qquad (5.15)$$

So, we can generate a random sample from the KumGP distribution by $X = Q(z)$, where Z is uniformly distributed on $(0,1)$.

In the rest of this chapter, we provide a comprehensive description of the mathematical properties of (5.14). We examine the shape of (5.14) and its associated hazard rate function in Section 5.4. We derive expressions for moments and moment generating function in sections 5.5 and 5.6, respectively. Order statistics, their moments and L moments are calculated in section 5.8. Asymptotic distributions of the extreme values are provided in section 5.9. Estimation by the method of maximum likelihood – including the observed information matrix – is presented in section 5.11.

Figure 5.1: Plot of the pdf of the GP distribution.
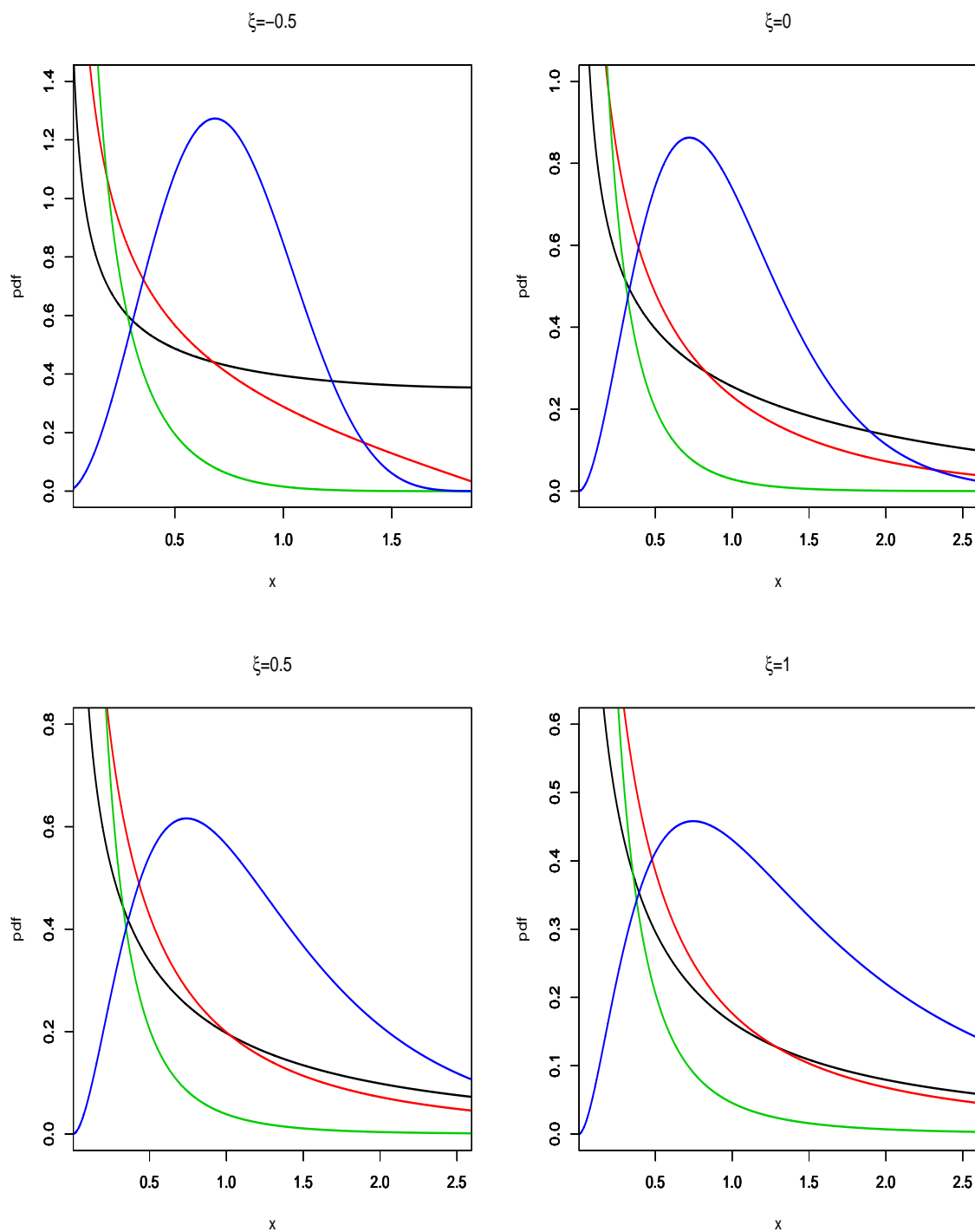
Figure 5.2: Plots of the pdf of the KumGP distribution for $t = 0$, $\sigma = 1$, $\xi = -0.5, 0, 0.5, 1$, $(a, b) = (0.5, 0.5)$ (black curve), $(a, b) = (0.5, 1)$ (red curve), $(a, b) = (0.5, 3)$ (green curve) and $(a, b) = (3, 3)$ (blue curve).

A simulation study is presented in section 5.12 to assess the performance of the maximum likelihood estimators. Application of the KumGP distribution to a real data set is illustrated in Section 5.13.

5.3 Hazard Rate Function

The hazard rate function (hrf)(also known as the failure rate, hazard rate, or force of mortality) is defined by $h(x) = \frac{f(x)}{1-F(x)}$. For the KumGP distribution the hrf takes the formula:

$$h(x) = \frac{ab\sigma^{-1}u^{\xi+1}(1-u)^{a-1}}{1-(1-u)^a} \tag{5.16}$$

$u = \{1 + \xi(x-t)/\sigma\}^{-1/\xi}$. Plots of the hrf for the GP and KumGP distribution for selected parameter values are shown in the Figures 5.3-5.4. We can see the flexibility of the hrf of the KumGP distribution over the hrf of the GP distribution.

5.4 Shapes of the PDF and HRF

1. **Shape of the pdf**:

   To study the shape of the KumGP distribution, we derive the pdf of the KumGP distribution. The first derivative of $\log f(x)$ for the KumGP distribution is:

   $$\frac{d}{dx}\log f(x) = -\frac{u^\xi}{\sigma}\left\{(\xi+1) - \frac{u}{1-u}\left[(a-1) - \frac{a(b-1)(1-u)^a}{[1-(1-u)^a]}\right]\right\}$$

   So, the modes of $f(x)$ are the roots of the following equation:

   $$\frac{a(b-1)(1-u)^{a-1}}{1-(1-u)^a} = \frac{a-1}{1-u} - \frac{1+\xi}{u}. \tag{5.17}$$

   where, $u = \{1 + \xi(x-t)/\sigma\}^{-1/\xi}$. There may be more than one root to Eq(5.17). So, if $x = x_0$ is the root of Eq(5.17) then it may correspond to a local maximum, local minimum or the point of inflexion depending on whether $\lambda(x_0) < 0$, $\lambda(x_0) > 0$ or $\lambda(x_0) = 0$, where $\lambda(x) = \frac{d^2}{dx^2}\log f(x)$ is :
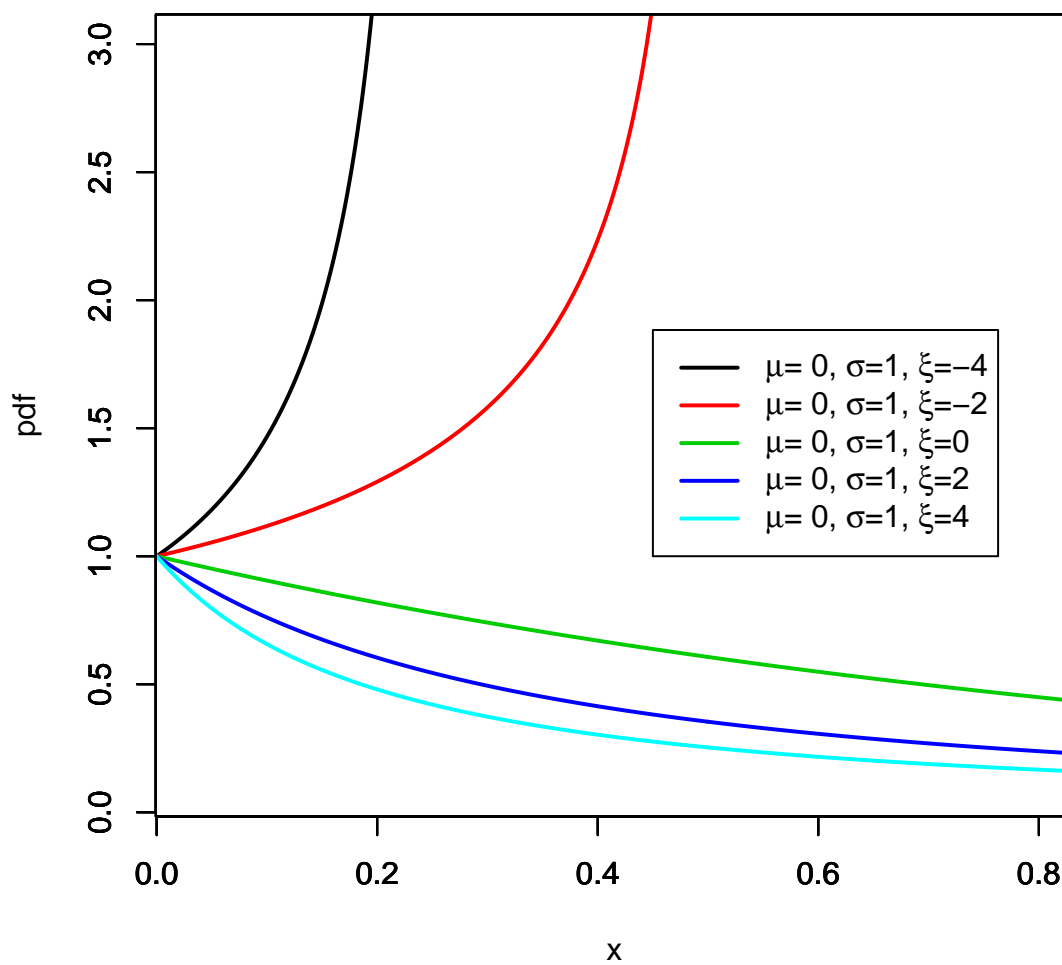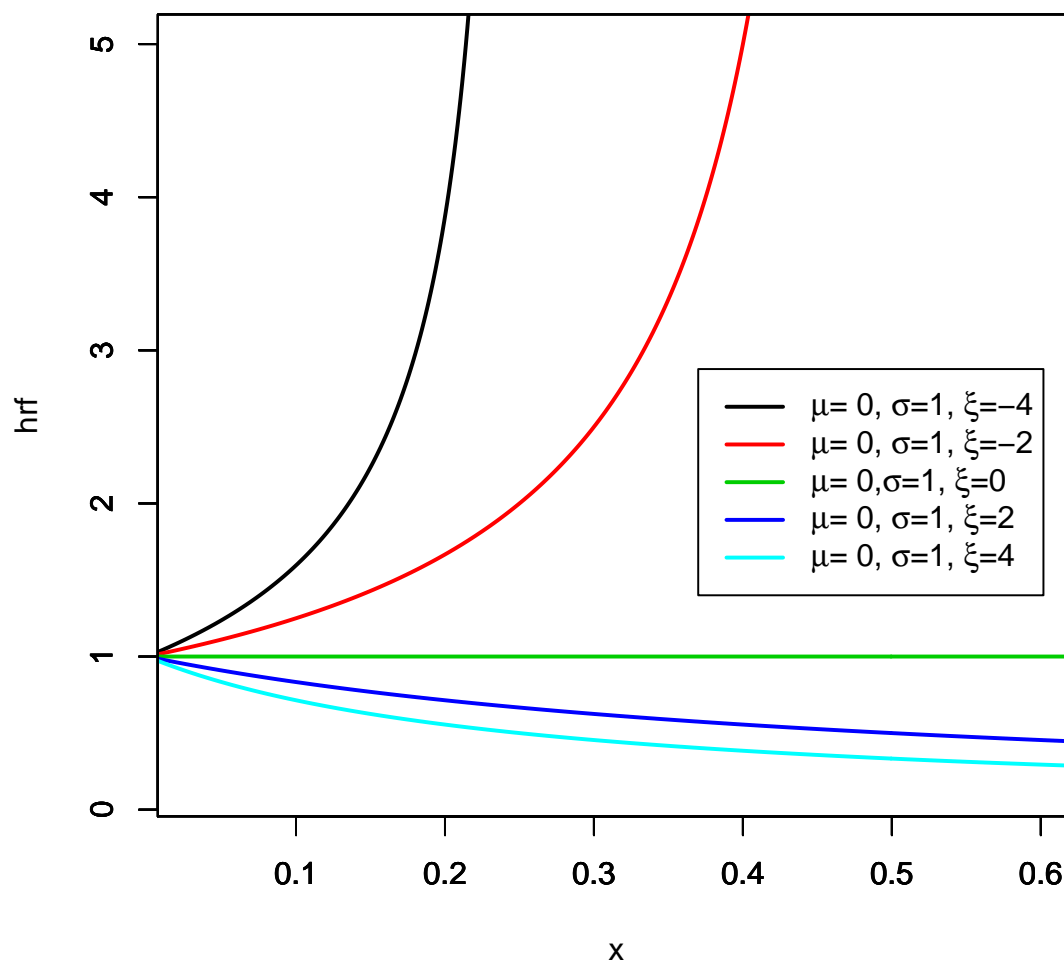
Figure 5.3: Plot of the hrf of the GP distribution.
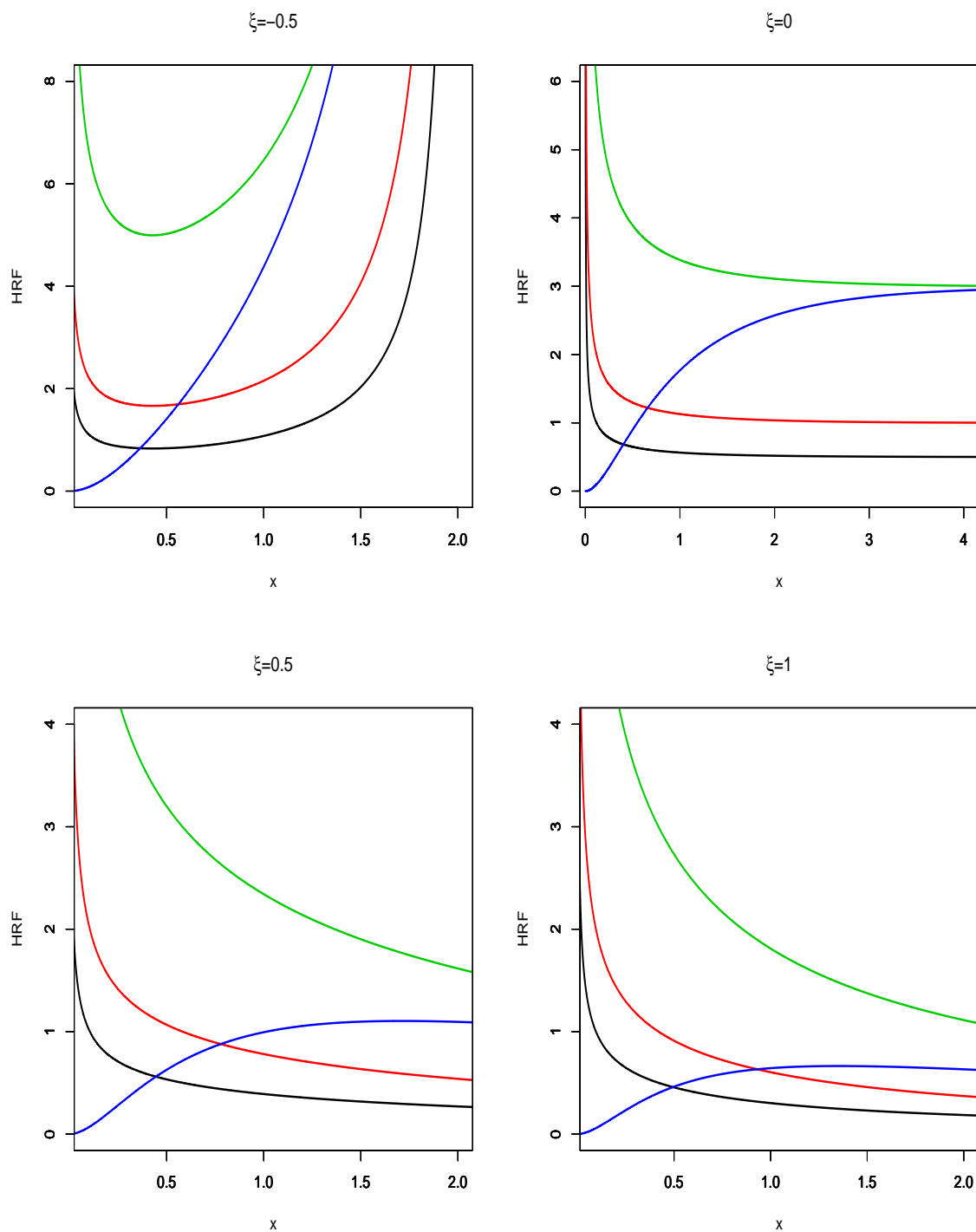
Figure 5.4: Plots of the hrf of the KumGP distribution for $t = 0$, $\sigma = 1$, $\xi = -0.5, 0, 0.5, 1$, $(a, b) = (0.5, 0.5)$ (black curve), $(a, b) = (0.5, 1)$ (red curve), $(a, b) = (0.5, 3)$ (green curve) and $(a, b) = (3, 3)$ (blue curve).

$$\lambda(x) = \frac{u^{2\xi}}{\sigma^2} \left(\xi\left(\xi+1\right)\right) - \frac{(a-1)u^{2/\xi+1}}{\sigma^2} \left[\frac{\left(\xi+1\right)\left(1-u\right)+u}{\left(1-u\right)^2}\right] - \frac{a(b-1)}{\sigma^2}$$

$$\times \left\{\frac{u^{2\xi+1}\left(1-u\right)^{a-2}\left[(a-1)u - (1+\xi)\left(1-u\right)\right]}{\left[1-\left(1-u\right)^a\right]} + \frac{au^{1+\xi}\left(1-u\right)^{2a-1}}{\left[1-\left(1-u\right)^a\right]^2}\right\}.$$

According to Nadarajah et al. (2012), the asymptotes of $f(x)$ and $F(x)$ as $u \to 0, 1$ are given by

$$f(x) \sim a^b b \sigma^{-1} u^{b+\xi} \exp(-au); \quad \text{as } u \to 0,$$

$$f(x) \sim ab\sigma^{-1}(1-u)^{a-1}; \quad \text{as } u \to 1,$$

$$1 - F(x) \sim (au)^b; \quad \text{as } u \to 0,$$

and

$$F(x) \sim b(1-u)^a; \quad \text{as } u \to 1.$$

Note that both the upper and lower tails of $f(x)$ are polynomials with respect to u. Larger values of a correspond to lighter lower tails and lighter upper tails of f. Larger values of b correspond to heavier lower tails and heavier upper tails of f.

2. **Shape of the hrf**:

The first derivative of $\log h(x)$ is given by:

$$\frac{d}{dx} \log h(x) = -\frac{u^{\xi}}{\sigma} \left\{\left(\xi+1\right) - \frac{(a-1)u}{\left[1-u\right]} - \frac{au\left(1-u\right)^{a-1}}{\left[1-\left(1-u\right)^a\right]}\right\}$$

The modes of the $h(x)$ are the roots of the following equation:

$$\left(\xi+1\right) = u \left\{\frac{(a-1)}{(1-u)} - \frac{a\left(1-u\right)^{a-1}}{\left[1-\left(1-u\right)^a\right]}\right\} \tag{5.18}$$

The local maximum, local minimum or the point of inflexion of the shape depends on whether $\lambda(x_0) < 0$, $\lambda(x_0) > 0$ or $\lambda(x_0) = 0$, where $\lambda(x) = \frac{d^2}{dx^2} \log f(x)$

is given by:

$$\lambda(x) = \frac{d^2 \log h(x)}{dx^2}$$

$$= \frac{u^{2\xi}}{\sigma^2} \left\{ \xi(\xi+1) - (a-1)u\frac{(\xi+1)(1-u)+u}{(1-u)^2} \right\} + \frac{au^{2\xi+1}(1-u)^{a-1}}{\sigma^2[1-(1-u)^a]}$$

$$\left\{ \frac{(a-1)u - (\xi+1)(1-u)}{(1-u)} + \frac{au(1-u)^a}{[1-(1-u)^a]^2} \right\}.$$

The asymptotes of $h(x)$ as $u \to 0, 1$ are given by

$$h(x) \sim b\sigma^{-1}u^{\xi} \qquad \text{as } u \to 0,$$

and

$$h(x) \sim ab\sigma^{-1}(1-u)^{a-1} \qquad \text{as } u \to 1.$$

Note that both the upper and lower tails of $h(x)$ are polynomials with respect to u. Larger values of a correspond to lighter lower tails. Larger values of b correspond to heavier lower tails and heavier upper tails of h. Bathtub shaped hazard rates are the most realistic ones in terms of practical applications. It is interesting to note that the KumGP distribution can exhibit this shape. The GP distribution cannot exhibit bathtub shaped hazard rates.

For more details of the shapes of the pdf and the hrf, we present the following three cases:

a) **First case** when $\xi < 0$:

Both the *pdf* and *hrf* of the KumGEV distribution are bathtub shaped for $a < 1$ and monotonically increasing for $a \geq 1$.

b) **Second case** when $\xi \to 0$:

Constant, monotonically decreasing and increasing shapes appear possible in this case. Both the pdf and hrf are monotonically decreasing for $a < 1$ and the pdf will be unimodal with mode x_{mode} for $a > 1$, while the hrf will be constant for $a = 1, b \leq 1$, slightly decreasing for $a = 1, \quad b > 1$ and monotonically increasing for $a > 1$.

c) **Last case** when $\xi > 0$:

Monotonically decreasing shapes appear for $a \leq 1$, while unimodal shape for $a > 1$. For both the pdf and hrf, we can find the mode by solving Eq (5.17).

5.5 Moments

Let $X \sim \text{KumGP}(a, b, \sigma, \xi)$. Using the transformation $u = \{1 + \xi(x - t)/\sigma\}^{-1/\xi}$, we can write

$$
\begin{aligned}
E\left(X^{n}\right) &= ab \int_{0}^{1} \left[\frac{\sigma}{\xi}\left(u^{-\xi} - 1\right) + t\right]^{n} (1 - u)^{a-1} \left[1 - (1 - u)^{a}\right]^{b-1} du \\
&= ab \sum_{i=0}^{n} \binom{n}{i} \left(\frac{\sigma}{\xi}\right)^{i} \left(t - \frac{\sigma}{\xi}\right)^{n-i} \int_{0}^{1} u^{-i\xi} (1 - u)^{a-1} \left[1 - (1 - u)^{a}\right]^{b-1} du \\
&= ab \sum_{i=0}^{n} \binom{n}{i} \left(\frac{\sigma}{\xi}\right)^{i} \left(t - \frac{\sigma}{\xi}\right)^{n-i} \sum_{j=0}^{\infty} \binom{b-1}{j} (-1)^{j} \int_{0}^{1} u^{-i\xi} (1 - u)^{a+aj-1} du \\
&= ab \sum_{i=0}^{n} \binom{n}{i} \left(\frac{\sigma}{\xi}\right)^{i} \left(t - \frac{\sigma}{\xi}\right)^{n-i} \sum_{j=0}^{\infty} \binom{b-1}{j} (-1)^{j} B\left(1 - i\xi, a + aj\right) \quad (5.19)
\end{aligned}
$$

for $n \geq 1$ provided that $1 - i\xi$ is not an integer for all $i = 0, 1, \ldots, n$. The first four moments are:

$$
E\left(X\right) = ab \sum_{j=0}^{\infty} \binom{b-1}{j} (-1)^{j} \left[\left(t - \frac{\sigma}{\xi}\right) \frac{1}{a + aj} + \frac{\sigma}{\xi} B\left(1 - \xi, a + aj\right)\right], \quad (5.20)
$$

$$
\begin{aligned}
E\left(X^{2}\right) = ab \sum_{j=0}^{\infty} \binom{b-1}{j} (-1)^{j} &\left[\left(t - \frac{\sigma}{\xi}\right)^{2} \frac{1}{a + aj} + 2 \left(\frac{\sigma}{\xi}\right) \left(t - \frac{\sigma}{\xi}\right)\right. \\
&\left.\times B\left(1 - \xi, a + aj\right) + \left(\frac{\sigma}{\xi}\right)^{2} B\left(1 - 2\xi, a + aj\right)\right],
\end{aligned} \quad (5.21)
$$

$$
\begin{aligned}
E\left(X^{3}\right) = ab \sum_{j=0}^{\infty} \binom{b-1}{j} (-1)^{j} &\left[\left(t - \frac{\sigma}{\xi}\right)^{3} \frac{1}{a + aj} + 3 \left(\frac{\sigma}{\xi}\right) \left(t - \frac{\sigma}{\xi}\right)^{2}\right. \\
&\times B\left(1 - \xi, a + aj\right) + 3 \left(t - \frac{\sigma}{\xi}\right) \left(\frac{\sigma}{\xi}\right)^{2} B\left(1 - 2\xi, a + aj\right) \\
&\left.+ \left(\frac{\sigma}{\xi}\right)^{3} B\left(1 - 3\xi, a + aj\right)\right],
\end{aligned} \quad (5.22)
$$

and

$$E\left(X^4\right) = ab\sum_{j=0}^{\infty}\binom{b-1}{j}(-1)^j\left[\left(t-\frac{\sigma}{\xi}\right)^4\frac{1}{a+aj}+4\left(\frac{\sigma}{\xi}\right)\left(t-\frac{\sigma}{\xi}\right)^3\right.$$

$$\times B\left(1-\xi, a+aj\right)+6\left(t-\frac{\sigma}{\xi}\right)^2\left(\frac{\sigma}{\xi}\right)^2 B\left(1-2\xi, a+aj\right)$$

$$\left.+4\left(t-\frac{\sigma}{\xi}\right)\left(\frac{\sigma}{\xi}\right)^3 B\left(1-3\xi, a+aj\right)+\left(\frac{\sigma}{\xi}\right)^4 B\left(1-4\xi, a+aj\right)\right].(5.23)$$

provided that $1-\xi$, $1-2\xi$, $1-3\xi$ and $1-4\xi$ are not integers. The infinite series in (5.20)-(5.23) all converge.

The mean, variance, skewness and kurtosis can be calculated from the first four nth moments, which is given by (5.20)-(5.23). Plots of these measures versus ξ are shown in Figure 5.5 for $\mu = 0$, $\sigma = 1$ and selected values of (a, b). These plots show that each of the measures are increasing of ξ, for most choice values of parameters a, b.
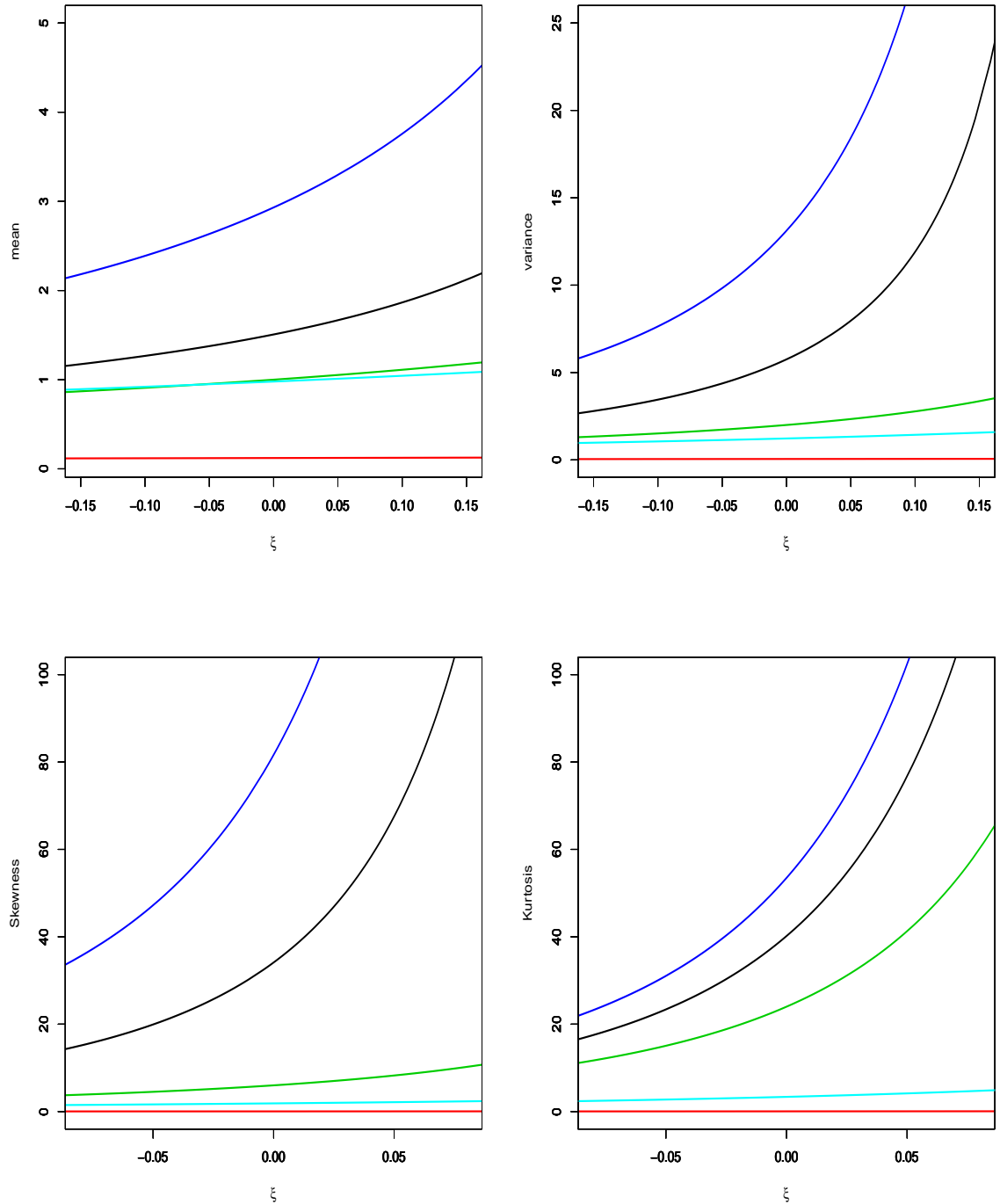
Figure 5.5: Plots of the mean, variance, skewness and kurtosis versus ξ for $\mu = 0$, $\sigma = 1$, $(a, b) = (0.5, 0.5)$ (black curve), $(a, b) = (0.5, 3)$ (red curve), $(a, b) = (1, 1)$ (green curve) and $(a, b) = (3, 0.5)$ (blue curve), and $(a, b) = (3, 3)$ (turquoise curve).

5.6 Moment Generating and Characteristic Functions

Suppose X is a random variable having the KumGP pdf in Eq(**??**). We obtain the moment generating function by using the following formula $M(t) = E[e^{(tX)}]$. So,

$$M(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx$$

$$= \sigma^{-1} a\, b \sum_{k=0}^{\infty} (-1)^k \binom{b-1}{k} \int_{-\infty}^{\infty} e^{tx} u^{\xi+1} (1-u)^{a(k+1)-1} dx,$$

$$= \sigma^{-1} a\, b \sum_{k=0}^{\infty} (-1)^k \binom{b-1}{k} J_3(t, a(k+1)),$$

where $u = \{1 + \xi(x-\mu)/\sigma\}^{-1/\xi}$ and μ is the threshold. Using **Lemma 1** in appendix **B**, we obtain

$$M(t) = a\, b e^{t_1^*} \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} (-1)^k \binom{b-1}{k} \frac{(t_2^*)^j}{j!} B(1 - j\xi, a(k+1)),$$

where $\xi < 1/j$. Also, the characteristic function, $\phi(t) = E[\exp(itX)]$, where $i = \sqrt{-1}$, of the KumGP distribution can be written as

$$\phi(t) = a\, b e^{it_1^*} \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} (-1)^k \binom{b-1}{k} \frac{(it_2^*)^j}{j!} B(1 - j\xi, a(k+1)),$$

where $t_1^* = t(\mu - \frac{\sigma}{\xi})$ and $t_2^* = t\frac{\sigma}{\xi}$.

5.7 Mean Deviations

The mean deviation about the mean and the mean deviation about median are defined as:

$$\delta_1(x) = \int_{-\infty}^{\infty} |x - \mu_0| f(x)\, dx,$$

$$\delta_2(x) = \int_{-\infty}^{\infty} |x - M| f(x)\, dx$$

respectively, where μ_0 and M denote to the mean and median, respectively. Then the measures, $\delta_1(X)$ and $\delta_2(X)$, can be calculated by using these relationships:

$$\delta_1(X) = \int_{-\infty}^{\mu_0} (\mu_0 - x) \, f(x) dx + \int_{\mu_0}^{\infty} (x - \mu_0) \, f(x) dx$$

$$= \mu_0 F(\mu_0) - \int_{-\infty}^{\mu_0} x f(x) dx - \mu_0 \{1 - F(\mu_0)\} + \int_{\mu_0}^{\infty} x f(x) dx$$

$$= 2 \int_{\mu_0}^{\infty} x f(x) dx + 2\mu_0 F(\mu_0) - 2\mu_0,$$

and

$$\delta_2(X) = \int_{-\infty}^{M} (M - x) f(x) dx + \int_{M}^{\infty} (x - M) f(x) dx$$

$$= MF(M) - \int_{-\infty}^{M} x f(x) dx - M \{1 - F(M)\} + \int_{M}^{\infty} x f(x) dx$$

$$= 2 \int_{M}^{\infty} x f(x) dx - \mu_0.$$

Let $X \sim \text{KumGP}(a, b, \sigma, \xi)$. Then

$$\int_{y}^{\infty} x f(x) dx = \sigma^{-1} a \, b \sum_{k=0}^{\infty} (-1)^k \binom{b-1}{k} \int_{y}^{\infty} x u^{\xi+1} (1-u)^{a(k+1)-1} dx$$

$$= \sigma^{-1} a \, b \sum_{k=0}^{\infty} (-1)^k \binom{b-1}{k} J_4 \left(y, a(k+1) \right).$$

The final step follows by Lemma 2 in appendix **B**. Then the mean deviation about the mean can be written as

$$\delta_1(X) = 2\sigma^{-1} a \, b \sum_{k=0}^{\infty} \sum_{j=0}^{ak-1} (-1)^k \binom{b-1}{k} J_4 \left(\mu_0, a(k+1) \right)$$

$$+ 2\mu_0 \left\{ 1 - \left[1 - (1 - u_0)^a \right]^b \right\} - 2\mu_0,$$

and

$$\delta_2(X) = 2\sigma^{-1} a \, b \sum_{k=0}^{\infty} (-1)^k \binom{b-1}{k} J_4 \left(M, a(k+1) \right) - \mu_0,$$

where $\mu_0^* = \left[1 + \xi(\mu_0 - t)/\sigma \right]^{-1/\xi}$. From Eq (5.20) we get μ_0, and the median can be calculated from the quantile function

$$M = t + \frac{\sigma}{\xi} \left\{ \left\{ 1 - \left[1 - 2^{-1/b} \right]^{1/a} \right\}^{-\xi} - 1 \right\}.$$

5.8   Order Statistics

The main purpose of this section is to derive the order statistics, their moments, probability weighted moments and L-moments for the KumGP distribution.

- **Order Statistics**:

  Assume that $X_1, X_2, \ldots, X_n$ is a random sample from the KumGP distribution with pdf (**??**). Let $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$ be the order statistics for this sample. Then the probability density and distribution function of the kth order statistics, say $Y = X_{k:n}$, are given by

$$
\begin{aligned}
f_Y(y) &= \frac{n!}{(k-1)!(n-k)!} \left[F(x)\right]^{k-1} \left[1 - F(x)\right]^{n-k} f(x), \\
&= \frac{a\, b\, n!}{\sigma(k-1)!(n-k)!} u^{1+\xi}(1-u)^{a-1} \left[1 - (1-u)^a\right]^{b(n-k+1)-1} \\
&\quad \left\{1 - [1 - (1-u)^a]^b\right\}^{k-1} \\
&= \frac{abn!}{\sigma(k-1)!(n-k)!} \sum_{i=0}^{\infty} \binom{k-1}{i} (-1)^i u^{1+\xi}(1-u)^{a-1} \left[1 - (1-u)^a\right]^{b(i+n-k+1)-1} \\
&= \frac{abn!}{\sigma(k-1)!(n-k)!} \sum_{i=0}^{\infty} \binom{k-1}{i} (-1)^i f_{a,b(i+n-k+1)-1,\sigma,\xi}(x),
\end{aligned}
$$

where $u = \{1 + \xi(x-t)/\sigma\}^{-1/\xi}$ and $f_{a,b,\sigma,\xi}(\cdot)$ denotes the probability density function of $X_{a,b,\sigma,\xi} \sim \mathrm{KumGP}(a, b, \sigma, \xi)$. So, the probability density function of Y is a linear combination of probability density functions of $\mathrm{KumGP}(a, b, \sigma, \xi)$. Hence, other properties of Y can be easily derived. For instance, the cumulative distribution function of Y can be expressed as

$$
\begin{aligned}
F_Y(y) &= \sum_{j=k}^{n} \binom{n}{j} \left\{F(x)\right\}^j \left\{1 - F(x)\right\}^{(n-j)}, \\
&= \frac{abn!}{\sigma(k-1)!(n-k)!} \sum_{i=0}^{\infty} \binom{k-1}{i} (-1)^i F_{a,b(i+n-k+1)-1,\sigma,\xi}(x), .
\end{aligned}
$$

where $F_{a,b,\sigma,\xi}(\cdot)$ denotes the cumulative distribution function corresponding to $f_{a,b,\sigma,\xi}(\cdot)$. The qth moment of Y can be expressed as

$$E\left(Y^q\right) = \int_{-\infty}^{\infty} x^q f_Y(y) dx$$

$$= \frac{abn!}{\sigma(k-1)!(n-k)!} \sum_{i=0}^{\infty} \binom{k-1}{i} (-1)^i \int_{-\infty}^{\infty} x^q f_{a,b(i+n-k+1)-1,\sigma,\xi}(x) dx,$$

$$= \frac{abn!}{\sigma(k-1)!(n-k)!} \sum_{i=0}^{\infty} \binom{k-1}{i} (-1)^i E\left[X_{a,b(i+n-k+1)-1,\sigma,\xi}^q\right], \quad (5.24)$$

where the last part can be calculated from (5.19). Then the qth moment of Y can be written as

$$E\left(Y^q\right) = \frac{a^2 b\, b^* n!}{\sigma(k-1)!(n-k)!} \sum_{i=0}^{\infty} \sum_{r=0}^{n} \binom{k-1}{i} \binom{n}{r} (-1)^i \left(\frac{\sigma}{\xi}\right)^i$$

$$\times \left(t - \frac{\sigma}{\xi}\right)^{n-i} \sum_{j=0}^{\infty} \binom{b-1}{j} (-1)^j B\left(1 - r\xi, a + aj\right). \quad (5.25)$$

- **Probability weighted moments of order statistics**:

  An alternative form for the moments of the KumGP order statistics depend on the probability weighted moments. Barakat and Abdelkader (2004) used a new formula to present the sth moments which can be written as

$$E(X_{i:n}^s) = s \sum_{j=n-i+1}^{n} (-1)^{j-n+i-1} \binom{j-1}{n-i} \binom{n}{j} I_j(s)$$

  where $I_j(s)$ is the integral

$$I_j(s) = \int_{-\infty}^{\infty} x^{s-1} \{1 - F(x)\}^j dx,$$

$$= \int_{-\infty}^{\infty} x^{s-1} \{1 - (1-u)^a\}^{bj} dx,$$

  By using the binomial expansion, we obtain

$$I_j(s) = \sigma \sum_{l=0}^{\infty} \sum_{k=0}^{s-1} (-1)^l \binom{s-1}{k} \binom{bj}{l} \left(\frac{\sigma}{\xi}\right)^k \left(\frac{t-\sigma}{\xi}\right)^{s-k-1} B(-ks - \xi, al - 1),$$

  and then

$$E(X_{i:n}^s) = s\sigma \sum_{j=n-i+1}^{n} \sum_{l=0}^{\infty} \sum_{k=0}^{s-1} \binom{j-1}{n-i} \binom{n}{j} \binom{s-1}{k} \binom{bj}{l} (-1)^{j-n+l+i-1}$$

$$\times \left(\frac{\sigma}{\xi}\right)^k \left(\frac{t-\sigma}{\xi}\right)^{s-k-1} B(-ks - \xi, al - 1), \quad (5.26)$$

  where $\xi < -ks$.

- **L-moments of order statistics**:

  L-moments are used to summaries the shape of the probability distribution. They are similar to the ordinary moments, but they are calculated from linear combinators of the order statistics. Hosking (1990) gave some advantages of the L-moments over the ordinary moments such as: for L-moments to be finite, we only assume that the distribution has finite mean; unlike the ordinary moments, which request that the higher-order moments are finite. The $r$ht L-moments can be given by

  $$\lambda_r = \sum_{j=0}^{r-1} (-1)^{r-1-j} \binom{r-1}{j} \binom{r-1+j}{j} \beta_j,$$

  where $\beta_j = E\{XF(X)^j\}$. In particular, $\lambda_1 = \beta_0$, $\lambda_2 = 2\beta_1 - \beta_0$, $\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0$ and $\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0$. In general, $\beta_r = (r+1)^{-1}E(X_{r+1:r+1})$, so it can be computed using (5.25).

5.9   Extreme Values

Suppose $X_1, \ldots, X_n$ is a random sample from the KumGP distribution (**??**). If $\bar{X}$ denotes the sample mean, then by the central limit theorem, $\sqrt{n}(\bar{X} - E(X))/\sqrt{\mathrm{Var}(X)}$ is distributed normally as $n \to \infty$, provided that $\xi < 1/2$. Sometimes one would be interested in the asymptotic of the extreme order statistics $M_n = \max(X_1, \ldots, X_n)$ and $m_n = \min(X_1, \ldots, X_n)$.

Firstly, suppose that G, the cdf of the GP distribution (5.11), belongs to the max domain of attraction of the Gumbel extreme value distribution. Then, by (Leadbetter et al., 1983, chap. 1), there must exist a strictly positive function $h(t)$, such that:

$$\lim_{t \to \infty} \frac{1 - G(t + xh(t))}{1 - G(t)} = \exp(-x)$$

for every $x \in (-\infty, \infty)$. Using L'Hopital's rule, we note that

$$\lim_{t \to \infty} \frac{1 - F(t + xh(t))}{1 - F(t)} = \lim_{t \to \infty} \left\{ \frac{1 - G^a(t + xh(t))}{1 - G^a(t)} \right\}^b$$

$$= \lim_{t \to \infty} \left\{ \frac{1 - G(t + xh(t))}{1 - G(t)} \right\}^b$$

$$= \exp(-bx)$$

for every $x \in (-\infty, \infty)$. So, it follows that F also belongs to the max domain of attraction of the Gumbel extreme value distribution with

$$\lim_{n \to \infty} \Pr\left\{ a_n (M_n - b_n) \leq x \right\} = \exp\left\{ -\exp(-bx) \right\}$$

for some suitable norming constants $a_n > 0$ and b_n.

Secondly, suppose that G, belongs to the max domain of attraction of the Fréchet extreme value distribution. Then by (Leadbetter et al., 1983, chap. 1), there must exist a $\beta > 0$ such that

$$\lim_{t \to \infty} \frac{1 - G(tx)}{1 - G(t)} = x^\beta$$

for every $x > 0$. Using L'Hopital's rule, we note that

$$\lim_{t \to \infty} \frac{1 - F(tx)}{1 - F(t)} = \lim_{t \to \infty} \left\{ \frac{1 - G^a(tx)}{1 - G^a(t)} \right\}^b$$

$$= \lim_{t \to \infty} \left\{ \frac{1 - G(tx)}{1 - G(t)} \right\}^b$$

$$= x^{b\beta}$$

for every $x > 0$. So, it follows that F also belongs to the max domain of attraction of the Fréchet extreme value distribution with

$$\lim_{n \to \infty} \Pr\left\{ a_n (M_n - b_n) \leq x \right\} = \exp\left(-x^{b\beta} \right)$$

for some suitable norming constants $a_n > 0$ and b_n.

Thirdly, suppose that G, belongs to the max domain of attraction of the Weibull extreme value distribution. Then by (Leadbetter et al., 1983, chap. 1), there must exist a $\alpha > 0$ such that

$$\lim_{t \to 0} \frac{G(tx)}{G(t)} = x^\alpha$$

for every $x < 0$. Using L'Hopital's rule, we note that

$$\lim_{t \to 0} \frac{F(tx)}{F(t)} = \lim_{t \to 0} \frac{1 - [1 - G^a(tx)]^b}{1 - [1 - G^a(t)]^b}$$

$$= \lim_{t \to 0} \left[\frac{G(tx)}{G(t)} \right]^a$$

$$= x^{a\beta}.$$

So, it follows that F also belongs to the max domain of attraction of the Weibull extreme value distribution with

$$\lim_{n \to \infty} \Pr\{a_n (M_n - b_n) \leq x\} = \exp\{-(-x)^{a\alpha}\}$$

for some suitable norming constants $a_n > 0$ and b_n.

The same argument applies to min domains of attraction. That is, F belongs to the same min domain of attraction as that of G.

5.10 Rényi and Shannon Entropies

In this section, we calculate the Rényi and Shannon entropies for the KumGP distribution. The Rényi entropy is defined as

$$H(\gamma) = \frac{1}{1 - \gamma} \log \left\{ \int_{-\infty}^{\infty} f^{\gamma}(x) dx \right\}.$$

where $\gamma > 0$ and $\gamma \neq 1$. If X a random variable has pdf of the KumGP distribution given in (6.5) then we can write,

$$\int_{-\infty}^{\infty} f^{\gamma}(x) dx = \left(\frac{a\,b}{\sigma} \right)^{\gamma} \int_{-\infty}^{\infty} u^{\gamma + \gamma\xi}(1 - u)^{\gamma a - \gamma} [1 - (1 - u)^a]^{b\gamma - \gamma} dx$$

$$= \left(\frac{a\,b}{\sigma} \right)^{\gamma} \sum_{k=0}^{\infty} \binom{b\gamma - \gamma}{k} (-1)^k \int_{-\infty}^{\infty} u^{\gamma + \gamma\xi}(1 - u)^{a(\gamma + k) - \gamma} dx$$

$$= \left(\frac{a\,b}{\sigma} \right)^{\gamma} \sigma \sum_{k=0}^{\infty} \binom{b\gamma - \gamma}{k} (-1)^k \int_{0}^{1} u^{(\gamma + k)(1 + \xi)}(1 - u)^{a(\gamma + k) - \gamma} du,$$

$$= \left(\frac{a\,b}{\sigma} \right)^{\gamma} \sigma \sum_{k=0}^{\infty} \binom{b\gamma - \gamma}{k} (-1)^k B(\gamma^*, a^*),$$

where $\gamma^* = (\gamma - 1)(\xi - 1) + 1$ and $a^* = (a - 1)\gamma + ak + 1$ Then the Rényi entropy can written as:

$$H(\gamma) = \frac{\sigma}{1 - \gamma} \log \left\{ \int_{-\infty}^{\infty} f^{\gamma}(x) dx \right\},$$

$$= \frac{1}{1 - \gamma} \log \left\{ \left(\frac{a\,b}{\sigma} \right)^{\gamma} \sum_{k=0}^{\infty} \binom{b\gamma - \gamma}{k} (-1)^k B(\gamma^*, a^*) \right\}.$$

Using Maclaurin series expansion for $\log(1 - z)$, the Shannon entropy can be calculated:

$$E\left[-\log f(X)\right] = -\log \left(\frac{a\,b}{\sigma} \right) - (1 + \xi) E \log U - (a - 1) E \log(1 - U)$$

$$-(b - 1) E \log \left[1 - (1 - U)^a \right]$$

$$= -\log \left(\frac{a\,b}{\sigma} \right) - (1 + \xi) E \log U + (a - 1) \sum_{i=1}^{n} \frac{1}{i} E(U^i)$$

$$+(b - 1) \sum_{i=1}^{n} \frac{1}{i} E\left[(1 - U)^{ai} \right]. \tag{5.27}$$

The three components in this equation can be derived as follows:

$$E\left[\log U\right] = \left(\frac{a\,b}{\sigma} \right) \sum_{k=0}^{\infty} (-1)^k \binom{b - 1}{k} \int_{-\infty}^{\infty} \log u\; u^{\xi+1}(1 - u)^{a(k+1)-1} d\,x,$$

$$= ab \sum_{k=0}^{\infty} (-1)^k \binom{b - 1}{k} \int_{0}^{1} \log u \quad (1 - u)^{a(k+1)-1} d\,u,$$

by using Eq (4.2531) in Gradshteyn and Ryzhik (1994), we obtain

$$E\left[\log U\right] = a\,b \sum_{k=0}^{\infty} (-1)^k \binom{b - 1}{k} B(1, a(k + 1)) \left[\psi(1) - \psi(a(k + 1) + 1) \right],$$

where $\psi(x)$ is the Euler's psi function.

$$E(U^i) = \left(\frac{a\,b}{\sigma} \right) \sum_{k=0}^{\infty} (-1)^{k+j} \binom{b - 1}{k} \int_{-\infty}^{\infty} u^i u^{\xi+1}(1 - u)^{a(k+1)-1} d\,x,$$

$$= ab \sum_{k=0}^{\infty} (-1)^k \binom{b - 1}{k} \int_{0}^{1} u^i (1 - u)^{a(k+1)-1} d\,u,$$

$$= a\,b \sum_{k=0}^{\infty} (-1)^k \binom{b - 1}{k} B(i + 1, a(k + 1)).$$

The last component can be calculated as

$$E(1-U)^{ai} = \left(\frac{a\,b}{\sigma}\right)\sum_{k=0}^{\infty}(-1)^{k+j}\binom{b-1}{k}\int_{-\infty}^{\infty}(1-u)^{ai}u^{\xi+1}(1-u)^{a(k+1)-1}d\,x,$$

$$= a\,b\sum_{k=0}^{\infty}(-1)^{k}\binom{b-1}{k}\int_{0}^{1}(1-u)^{a(k+i+1)-1}d\,u,$$

$$= a\,b\sum_{k=0}^{\infty}\frac{(-1)^{k}}{a(k+i+1)}\binom{b-1}{k}.$$

So, the Shannon entropy can be written as:

$$E\left[-\log f(X)\right] = -\log\left(\frac{a\,b}{\sigma}\right) + a(a-1)b\sum_{i=1}^{n}\sum_{k=0}^{\infty}\sum_{j=0}^{ak-1}\frac{(-1)^{k+j}}{i(j+i+1)}$$

$$\times\binom{b-1}{k}\binom{ak-1}{j} + ab(b-1)\sum_{i=1}^{n}\sum_{k=0}^{\infty}\sum_{j=0}^{a\,k-1}\frac{(-1)^{k+j}}{i}$$

$$\times\binom{b-1}{k}\binom{ak-1}{j}B(j+1,ai+1) - a\,b(1+\xi)$$

$$\times\sum_{k=0}^{\infty}\sum_{j=0}^{a\,k-1}(-1)^{k+j}\binom{b-1}{k}\binom{ak-1}{j}\left[\psi(j+1)\right.$$

$$\left.-\psi(j+2)\right]B(j+1,1). \tag{5.28}$$

5.11   Maximum Likelihood Estimation

Suppose $X_1, X_2, \ldots, X_n$ is a random sample of size n from (5.14). Let $u_i = \{1 + \xi(x_i - t)/\sigma\}^{-1/\xi}$ for $i = 1, 2, \ldots, n$. Then the log-likelihood function for the vector of parameters (a, b, σ, ξ) can be written as

$$\log L(a,b,\sigma,\xi) = n\log(ab) - n\log\sigma + (1+\xi)\sum_{i=1}^{n}\log u_i + (a-1)\sum_{i=1}^{n}\log\left(1-u_i\right)$$

$$+(b-1)\sum_{i=1}^{n}\log\left[1-(1-u_i)^a\right]. \tag{5.29}$$

The first-order partial derivatives of (5.29) with respect to the four parameters are:

$$\frac{\partial\log L}{\partial a} = \frac{n}{a} + \sum_{i=1}^{n}\log\left(1-u_i\right) - (b-1)\sum_{i=1}^{n}\frac{(1-u_i)^a\log\left(1-u_i\right)}{1-(1-u_i)^a}, \tag{5.30}$$

$$\frac{\partial\log L}{\partial b} = \frac{n}{b} + \sum_{i=1}^{n}\log\left[1-(1-u_i)^a\right], \tag{5.31}$$

$$\frac{\partial \log L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1+\xi}{\sigma^2} \sum_{i=1}^{n} u_i^{\xi}(x_i - t) - \frac{a-1}{\sigma^2} \sum_{i=1}^{n} \frac{u_i^{1+\xi}(x_i - t)}{1 - u_i}$$

$$+ \frac{a(b-1)}{\sigma^2} \sum_{i=1}^{n} \frac{u_i^{1+\xi}(1 - u_i)^{a-1}(x_i - t)}{1 - (1 - u_i)^a}, \tag{5.32}$$

and

$$\frac{\partial \log L}{\partial \xi} = \sum_{i=1}^{n} \log u_i + \frac{1+\xi}{\xi^2} \sum_{i=1}^{n} \left\{ \log\left[1 + \xi\frac{x_i - t}{\sigma}\right] - \frac{x_i \xi}{\sigma}\left[1 + \xi\frac{x_i - t}{\sigma}\right]^{-1} \right\} - \frac{a-1}{\xi^2}$$

$$\times \sum_{i=1}^{n} \frac{u_i}{1 - u_i} \left\{ \log\left[1 + \xi\frac{x_i - t}{\sigma}\right] - \frac{x_i \xi}{\sigma}\left[1 + \xi\frac{x_i - t}{\sigma}\right]^{-1} \right\} + \frac{b-1}{\xi^2}$$

$$\times \sum_{i=1}^{n} \frac{a u_i (1 - u_i)^{a-1}}{1 - (1 - u_i)^a} \left\{ \log\left[1 + \xi\frac{x_i - t}{\sigma}\right] - \frac{x_i \xi}{\sigma}\left[1 + \xi\frac{x_i - t}{\sigma}\right]^{-1} \right\}. \tag{5.33}$$

The maximum likelihood estimates of (a, b, σ, ξ), say $(\widehat{a}, \widehat{b}, \widehat{\sigma}, \widehat{\xi})$, are the simultaneous solutions of the equations $\partial \log L/\partial a = 0$, $\partial \log L/\partial b = 0$, $\partial \log L/\partial \sigma = 0$ and $\partial \log L/\partial \xi = 0$. As $n \to \infty$, $(\widehat{a} - a, \widehat{b} - b, \widehat{\sigma} - \sigma, \widehat{\xi} - \xi)$ approaches a multivariate normal vector with zero means and variance-covariance matrix $-(E[\mathbf{J}])^{-1}$, where

$$\mathbf{J} = \begin{pmatrix} \frac{\partial^2 \log L}{\partial a^2} & \frac{\partial^2 \log L}{\partial a \partial b} & \frac{\partial^2 \log L}{\partial a \partial \sigma} & \frac{\partial^2 \log L}{\partial a \partial \xi} \\ \frac{\partial^2 \log L}{\partial b \partial a} & \frac{\partial^2 \log L}{\partial b^2} & \frac{\partial^2 \log L}{\partial b \partial \sigma} & \frac{\partial^2 \log L}{\partial b \partial \xi} \\ \frac{\partial^2 \log L}{\partial \sigma \partial a} & \frac{\partial^2 \log L}{\partial \sigma \partial b} & \frac{\partial^2 \log L}{\partial \sigma^2} & \frac{\partial^2 \log L}{\partial \sigma \partial \xi} \\ \frac{\partial^2 \log L}{\partial \xi \partial a} & \frac{\partial^2 \log L}{\partial \xi \partial b} & \frac{\partial^2 \log L}{\partial \xi \partial \sigma} & \frac{\partial^2 \log L}{\partial \xi^2} \end{pmatrix}.$$

The matrix, $-E[\mathbf{J}]$, is known as the expected information matrix. The matrix, $-\mathbf{J}$, is known as the observed information matrix.

In simulations and real data applications described later on, we maximized the log-likelihood function using the nlm function in the R statistical package. For each maximization, the nlm function was executed for a wide range of initial values. At least one maximum was identified each time. In cases of more than one maximum occurring, we took the maximum likelihood estimates to correspond to the largest of the maxima.

The literature (see, for example, Cox and Hinkley (1979)) suggests that it is best to approximate the distribution of $(\widehat{a} - a, \widehat{b} - b, \widehat{\sigma} - \sigma, \widehat{\xi} - \xi)$ by a multivariate normal distribution with zero means and variance-covariance matrix given by $-\mathbf{J}^{-1}$, inverse

of the observed information matrix, with (a, b, σ, ξ) replaced $(\widehat{a}, \widehat{b}, \widehat{\sigma}, \widehat{\xi})$. So, it is useful to have explicit expressions for the elements of $\mathbf{J}$, which are shown in appendix $\mathbf{B}$.

The matrix, $-\mathbf{J}$, is known as the observed information matrix. The matrix, $-E\mathbf{J}$, is known as the expected information matrix. The $4 * 4$ unit expected information matrix $k(\theta)$ is:

$$\mathbf{k}(\theta) = \begin{pmatrix} k_{a,a}(\theta) \ k_{a,b}(\theta) \ k_{a,\sigma}(\theta) \ k_{a,\xi}(\theta) \\ k_{a,b}(\theta) \ k_{b,b}(\theta) \ k_{b,\sigma}(\theta) \ k_{b,\xi}(\theta) \\ k_{a,\sigma}(\theta) \ k_{b,\sigma}(\theta) \ k_{\sigma,\sigma}(\theta) \ k_{\sigma,\xi}(\theta) \\ k_{a,\xi}(\theta) \ k_{b,\xi}(\theta) \ k_{\sigma,\xi}(\theta) \ k_{\xi,\xi}(\theta) \end{pmatrix}.$$

5.12 Simulation Study

The aim of this section is to show that the estimators $(\hat{a}, \hat{b}, \hat{\sigma} \text{ and } \hat{\xi})$ are unbiased and have a minimum mean square error when the sample size, n, is large. We follow the simulation procedure:

1. Generate a thousand samples of size n from the KumGP distribution (5.14). The inversion method given earlier, (5.15), is used to generate samples.

2. Compute the maximum likelihood estimates for the thousand samples, say $(\widehat{a}_i, \widehat{b}_i, \widehat{\sigma}_i, \widehat{\xi}_i)$ for $i = 1, 2, \ldots, 1000$.

3. Compute the biases and mean squared errors given by

$$\text{bias}_h(n) = \frac{1}{1000} \sum_{i=1}^{1000} \left(\widehat{h}_i - h \right)$$

and

$$\text{MSE}_h(n) = \frac{1}{1000} \sum_{i=1}^{1000} \left(\widehat{h}_i - h \right)^2$$

for $h = a, b, \sigma, \xi$.

We repeat these steps for $n = 10, 20, \ldots, 1000$ with $a = 3$, $b = 3$, $t = 0$, $\sigma = 1$ and $\xi = 0.5$, so computing $\text{bias}_a(n)$, $\text{bias}_b(n)$, $\text{bias}_\sigma(n)$, $\text{bias}_\xi(n)$ and $\text{MSE}_a(n)$, $\text{MSE}_b(n)$, $\text{MSE}_\sigma(n)$, $\text{MSE}_\xi(n)$ for $n = 10, 20, \ldots, 1000$.
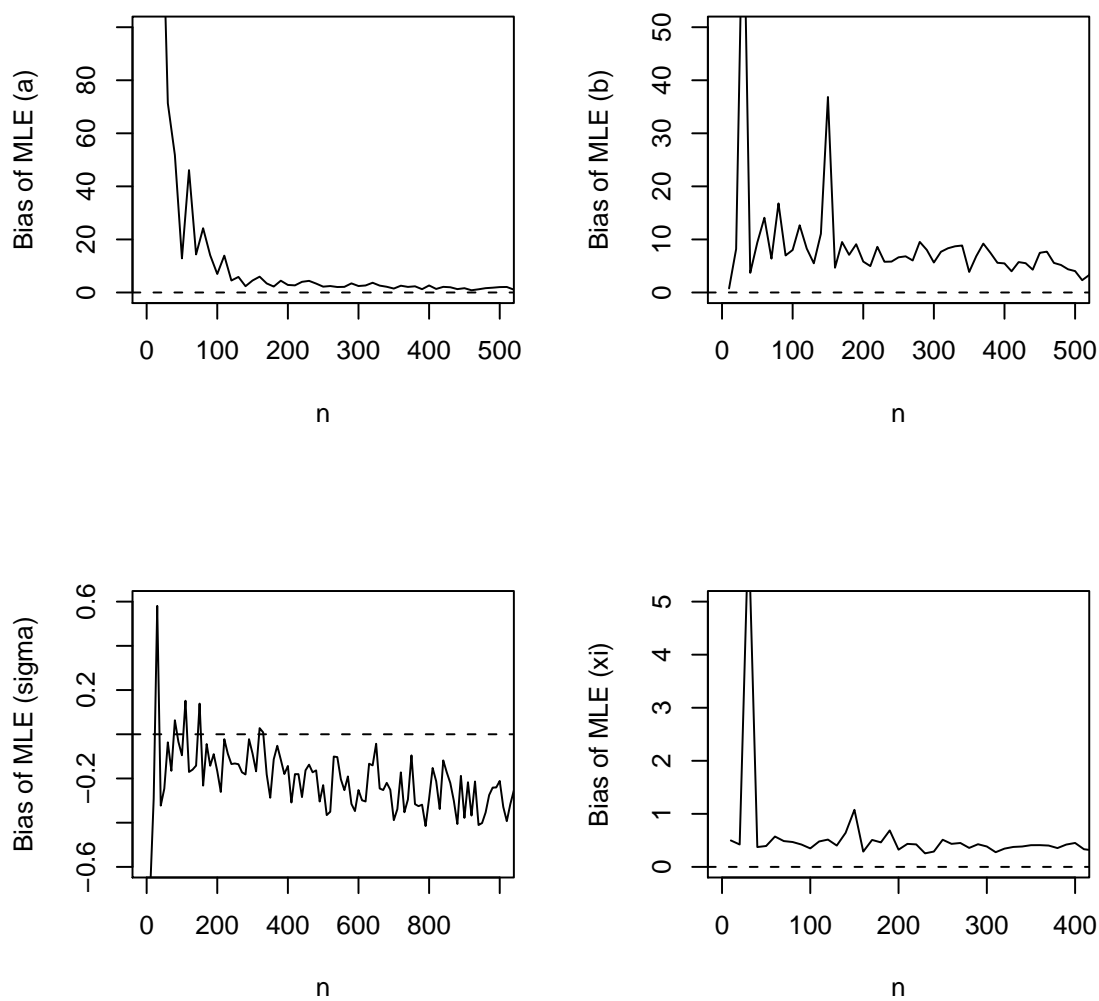
Figure 5.6: $bias_a(n)$ (top left), $bias_b(n)$ (top right), $bias_\mu(n)$ (middle left), $bias_\sigma(n)$ (middle right) and $bias_\xi(n)$ (bottom left) versus $n = 10, 20, \ldots, 1000$.
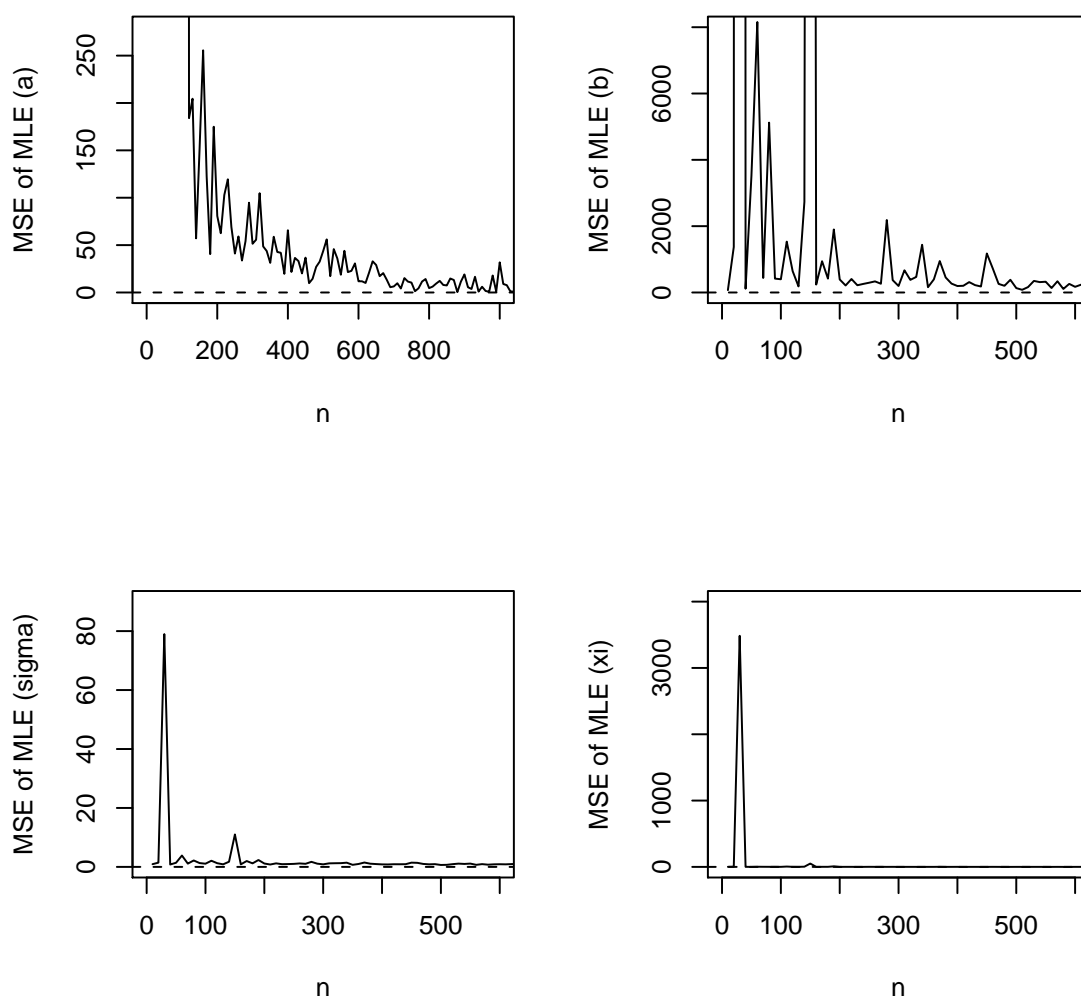
Figure 5.7: $MSE_a(n)$ (top left), $MSE_b(n)$ (top right), $MSE_\mu(n)$ (middle left), $MSE_\sigma(n)$ (middle right) and $MSE_\xi(n)$ (bottom left) versus $n = 10, 20, \ldots, 1000$.

Figures 5.6 and 5.7 show how the four biases and the four mean squared errors vary with respect to n. The broken line in Figure 5.6 corresponds to the biases being zero. The broken line in Figure 5.7 corresponds to the mean squared errors being zero. We know from theory that maximum likelihood estimates have biases of the order $O(1/n)$ and mean squared errors of the order $O(1/n)$. With this in mind, we have shown in Figures 5.6 and 5.7 how the four biases and the four mean squared errors vary with respect to n. The following observations can be made:

1. The biases for the parameters a, b, and ξ are generally positive, while they are negative for σ.

2. Although they appear volatile, the biases for each parameter decrease to zero as $n \to \infty$.

3. The biases appear largest for parameters b and σ, while it is smallest for parameters a and ξ.

4. Although they appear volatile, the mean squared errors for each parameter decrease to zero as $n \to \infty$.

5. The mean squared errors appear largest for the parameter b.

6. The mean squared errors appear smallest for the parameters a, σ and ξ.

7. The biases for a appear to level out for all $n > 150$;

8. The biases for b appear to level out for all $n > 500$;

9. The biases for ξ appear to level out for all $n > 400$;

10. The mean squared errors for a appear to level out for all $n > 800$;

11. The mean squared errors for b appear to level out for all $n > 500$;

12. The mean squared errors for σ appear to level out for all $n > 180$;

13. The mean squared errors for ξ appear to level out for all $n > 80$.

Here, we have presented results for only one choice for (a, b, σ, ξ), namely that $(a, b, \sigma, \xi) = (3, 3, 1, 0.5)$, but results are analogous for other choices. In addition to computing the biases and mean squared errors, we also computed p values to check for multivariate normality and validity of likelihood ratio tests. The p values for multivariate normality were based on the Shapiro-Wilk test Royston (1982). The p values for the validity of likelihood ratio tests were based on the chi-square goodness of fit test. Plots of the p values versus n showed that they remained above 0.05 for all values of n greater than 200.

5.13   An Application

To illustrate the better flexibility of our model compared to the others, we apply this model to floods data.

5.13.1   Floods Data in the River Nidd at Hunsingore Weir

Here, we illustrate the flexibility of the KumGP distribution using a real data set, analysed in Papastathopoulos and Tawn (2012). The data set consists of 154 exceedances of the threshold level $65m^3s^{-1}$ by the River Nidd at Hunsingore Weir from 1934 to 1969. The data is taken from NERC (1975).

Empirical Mean Residual Life Plot:

A mean residual life plot of the data is shown in Figure 5.8. From this plot we choose $t = 72.7m^2s^{-1}$. This threshold shown in red seems appropriate for this.

Estimate the Parameters:

We fit the KumGP, GP, EGP1, EGP2, and EGP3 distributions to the data. The latter three distributions are those considered by Papastathopoulos and Tawn (2012). The maximum likelihood procedure described in section 5.11 was used for fitting Eq(**??**). The MLEs of the parameters, the Log-Lik and information criteria are listed in Tables 5.2- 5.3.

Model	a	b	κ	σ	ξ	-Log-Lik
KumGP	4.659	111.634	-	2.467	10.073	551.4591
GP	1	1	-	19.823	0.393	556.2307
EGP1	-	-	1.875	7.190	0.687	553.7751
EGP2	-	-	2.220	3.911	0.601	553.5387
EGP3	-	-	1.834	8.057	0.696	553.858

Table 5.2: MLEs of the model parameters for Floods data.

None of the three-parameter distributions (EGP1, EGP2 and EGP3) provide a significant improvement over the GP distribution. From these three distributions, the EGP1 distribution has the largest likelihood value and the smallest AIC value,
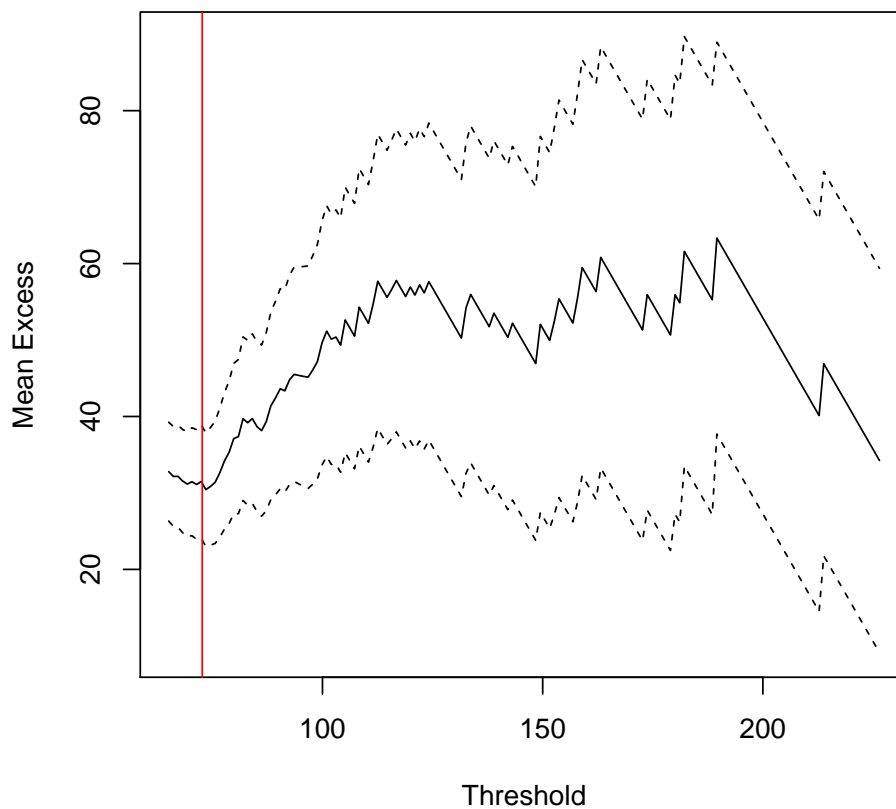
Figure 5.8: The Mean residual life plot for exceedances of the levels of River Nidd over the threshold $65m^3s^{-1}$.

Model	AIC	AICc
KumGP	1110.918	1111.243
GP	1116.461	1116.568
EGP1	1113.55	1113.752
EGP2	1113.077	1113.279
EGP3	1113.716	1113.917

Table 5.3: The statistics AIC and AICc.

but the fit of the EGP2 distribution is not significantly better than that of the GP distribution. The proposed four-parameter distribution provides a significant improvement over the GP distribution and the three three-parameter distributions. It has the largest likelihood value and the smallest AIC value amongst all of the fitted distributions. Furthermore, chi-square goodness of fit tests give the *p*-values of 0.0373, 0.0461, 0.048, 0.041 and 0.068 for GP, EGP1, EGP2, EGP3, and KumGP distributions, respectively, suggesting that actually the KumGP distribution provides the only adequate fit. In the Table presents values of the Anderson-Darling and Cramér-von Miseswe tests. According to these tests, we can conclude that the KumGP distribution give better fit than the GP distribution.

Model	Anderson-Darling	Cramér-von Mises
KumGP	0.1695225	0.02260322
GP	0.3105545	0.04270863

Table 5.4: The Anderson-Darling and Cramér-von Mises statistics for the annual rainfall maxima data

Graphs:

The conclusion drawn, based on the likelihood values, AIC , AICc values and the chi-square goodness of fit tests can be verified by means of probability-probability plots, quantile-quantile plots and density plots.

A probability-probability plot consists of plots of the observed probabilities against probabilities predicted by the fitted model. For example, for the GP model, the values $1 - [1 + \widehat{\xi}(x_{(j)} - t)/\widehat{\sigma}]^{-1/\widehat{\xi}}$ are plotted versus $(j - 0.375)/(n + 0.25)$, $j = 1, 2, \ldots, n$. This is method was recommended by Blom (1958) and Chambers (1983), where $x_{(j)}$ are the sorted values of the data in ascending order, and n is the number of observations. A quantile-quantile plot consists of plots of the observed quantile against quantile predicted by the fitted model. For example, for the GP model, the values $t + (\widehat{\sigma}/\widehat{\xi})\{(1 - (j - 0.375)/(n + 0.25))^{-\widehat{\xi}} - 1\}$ are plotted versus $x_{(j)}$, $j = 1, 2, \ldots, n$. This is the method was recommended by Blom (1958) and Chambers (1983). Probability-probability plots and quantile-quantile plots for the five fitted models are shown in

Figures 5.9 and 5.10. We can see that the KumGP model has the points closest to the diagonal line, especially in the upper tail, thus showing the closest agreement between expected and observed values. In fact, the sum of the absolute differences in probabilities are 2.28 for GP distribution, 2.61 for EGP1 distribution, 2.64 for EGP2 distribution, 2.59 for EGP3 distribution, and 1.83 for KumGP distribution. The sum of the absolute differences in quantiles are 430.6654, 623.6487, 639.2003, 623.0249, and 361.5818 for GP, EGP1, EGP2, and KumGP distribution, respectively.

A density plot compares the fitted probability density functions of the models with the empirical histogram of the observed data. The density plots are shown in Figure 5.11. Again the fitted probability density function for KumGP distribution appears to capture the general pattern of the empirical histogram best.

Quantities of interest for users of extreme value models are the *return levels*. A T year return level, say x_T, is defined as the level that is exceeded on average every T years. For the GP model given by (5.12),

$$x_T = t + \frac{\sigma}{\xi} \left\{ (T)^{\xi} - 1 \right\}, \tag{5.34}$$

where m is the average number of exceedances per year. For the KumGP model given by (**??**),

$$x_T = t + \frac{\sigma}{\xi} \left\{ \left[1 - \left\{ 1 - (T)^{-1/b} \right\}^{1/a} \right]^{-\xi} - 1 \right\}, \tag{5.35}$$

where m is, again, the average number of exceedances per year. Plots of (5.34) and (5.35) for $T = 2, 3, \ldots, 50$ along with 95 confidence intervals computed by the delta method ((Rao, 1973, pages. 387-389)) are shown in Figure 5.12.

Return levels are important quantities. They are used to determine, for example, dimensions of sea walls, water dams, flood defences, etc. Figure 5.12 suggests that the return levels given by (5.34) and (5.35) do not differ so much. However, the confidence bands for (5.35) appear much narrower and much more realistic. The confidence bands for (5.34) are so wide that they do not appear in the figure! So, if one were to use the KumGP model instead of the GP model, there could be significant savings with respect to cost and time.
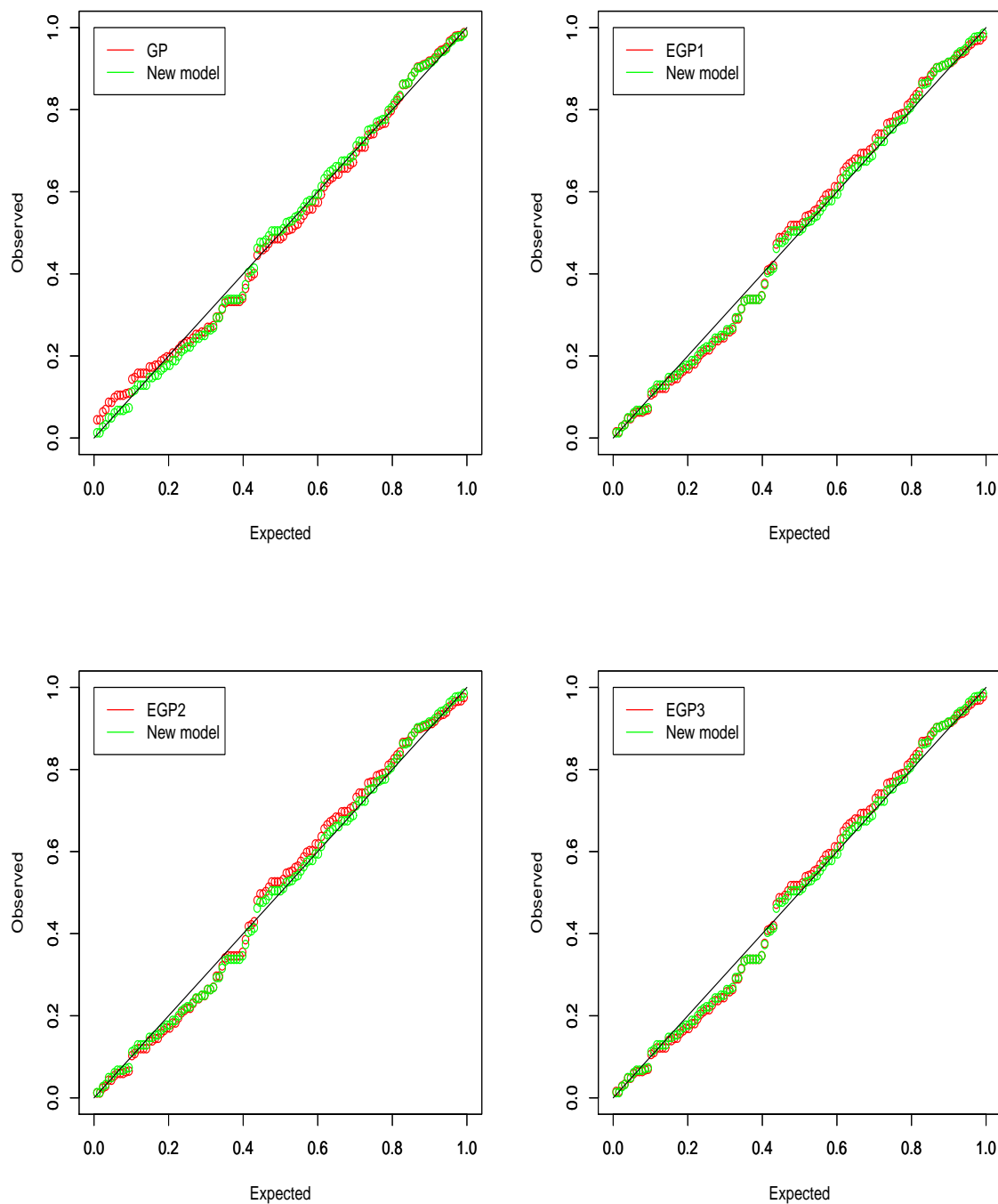
Figure 5.9: Probability plots for the fits of GP, EGP1, EGP2, EGP3, and KumGP (New model) for exceedances of the levels of River Nidd over the threshold $t = 72.7 m^3 s^{-1}$.
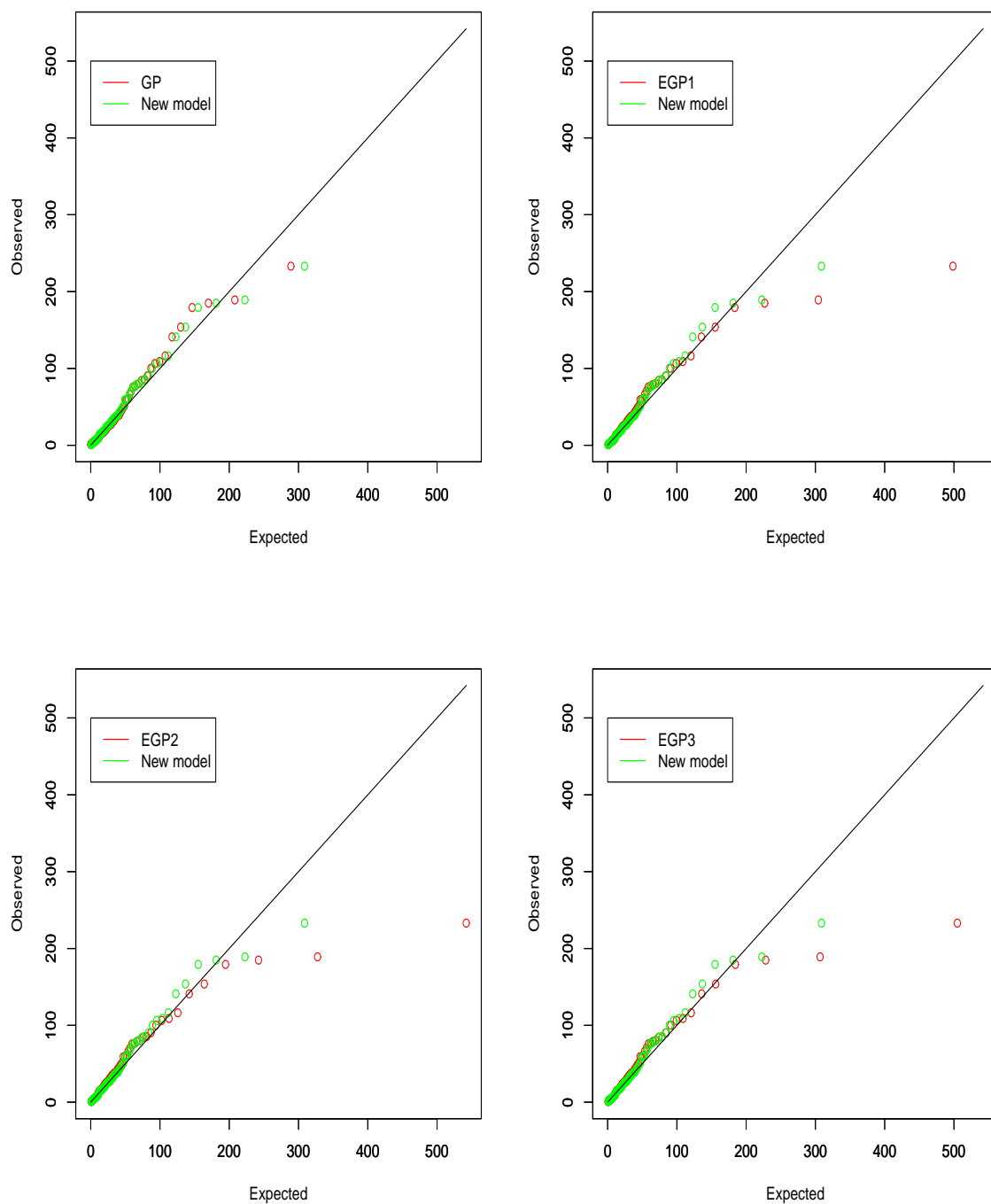
Figure 5.10: Quantile plots for the fits of GP, EGP1, EGP2, EGP3, and KumGP (New model) for exceedances of the levels of River Nidd over the threshold $t = 72.7m^3s^{-1}$.
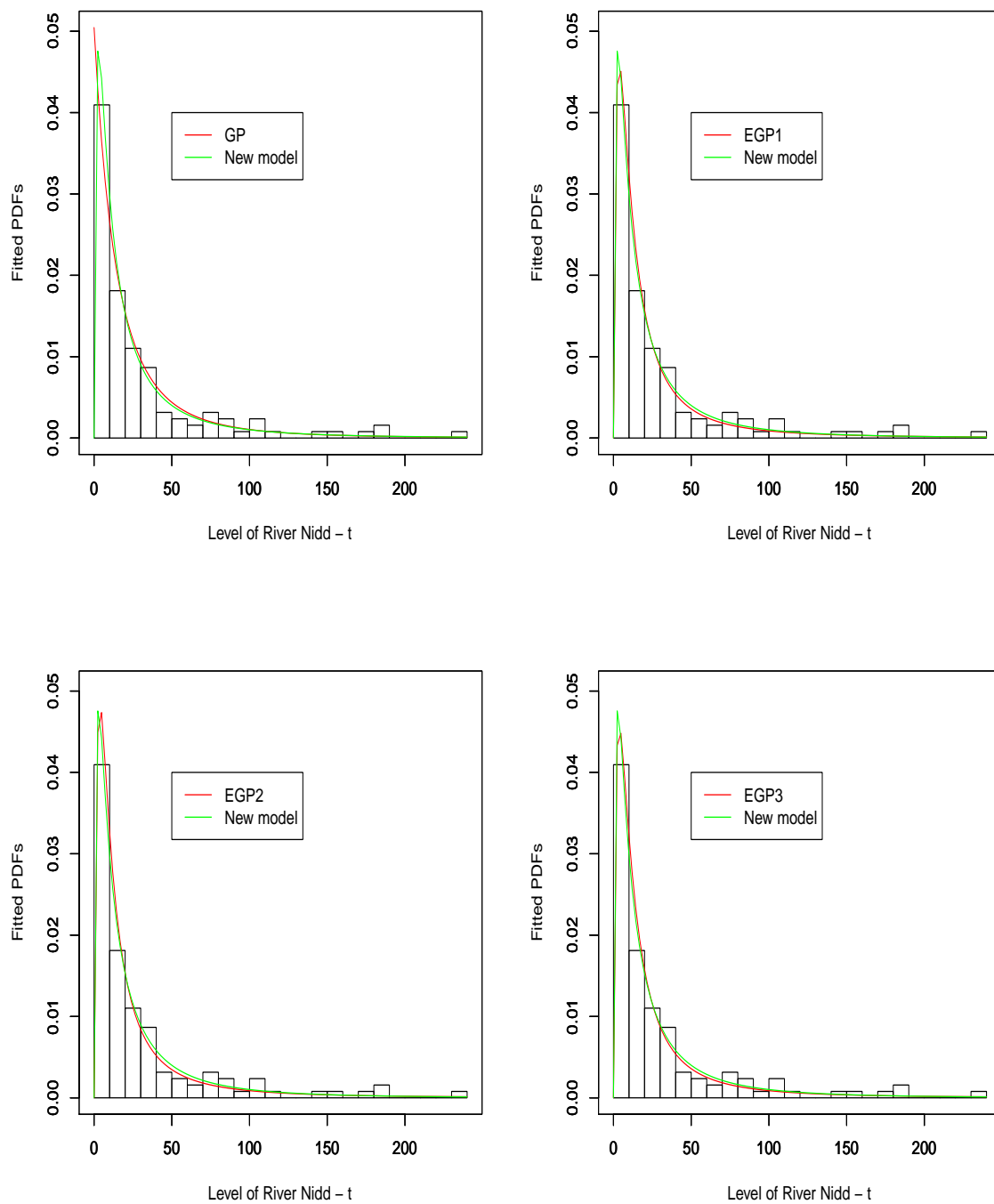
Figure 5.11: Fitted probability density functions of GP, EGP1, EGP2, EGP3, and KumGP (New model) for exceedances of the levels of River Nidd over the threshold $t = 72.7 m^3 s^{-1}$.
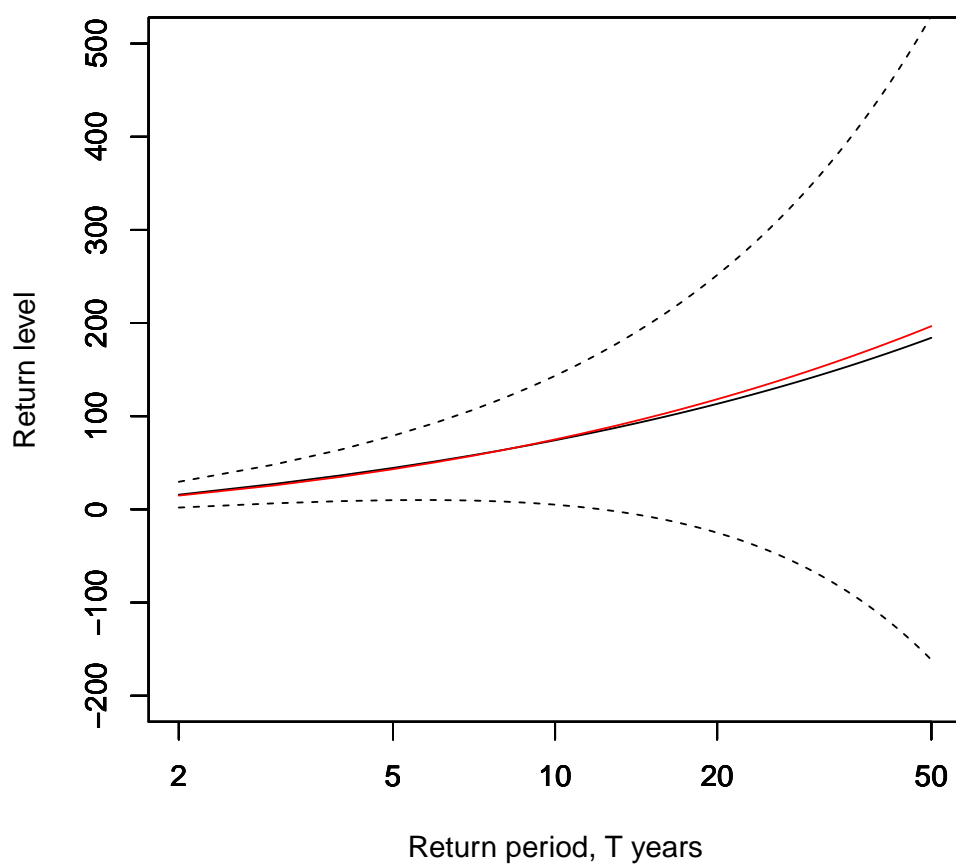
Figure 5.12: Return levels for exceedances of the levels of River Nidd and their 95 percent confidence intervals for the fits of the pdf of the GP distribution (in red) and the pdf of KumGP distribution (in black).

5.14 Summary

In this chapter, we discuss the most recent extensions of the GP distribution proposed by Papastathopoulos and Tawn. Here, we point out that Papastathopoulos and Tawns generalisations are in fact not new and then go on to propose a tractable generalization of the GP distribution dependent on the Kumarasswamy distribution. For the latter generalisation, we provide a comprehensive treatment of mathematical properties, estimate parameters by the method of maximum likelihood and provide the observed information matrix.

Some main points about the KumGP distribution are:

- pdf of the KumGP distribution is flexible. That means it can take a variety of shapes; monotonically decreasing, monotonically increasing, unimodal.

- Bathtub shaped hazard rates are the most realistic ones in terms of practical applications. It is interesting to note that the KumGP distribution can exhibit this shape. The GP distribution cannot exhibit bathtub shaped hazard rates. The KumGP distribution's hrf has a variety of shapes: monotonically decreasing, monotonically increasing, and unimodal.

- Finally, by applying the KumGP distribution to floods data in the river Nidd at Hunsingore weir, we conclude that the KumGP distribution is more flexible compared to the GP, EGP1, EGP2, and EGP3 distributions.

Part III

A New Family of Extreme Value Distributions

Chapter 6

On Chen *et al.*'s Extreme Value Distribution

6.1 Introduction

The GEV distribution is one of the most widely applied models for univariate extreme values. Its cumulative distribution function and probability density function are specified by

$$F(x) = \exp(-u)$$

and

$$f(x) = \sigma^{-1} u^{1+\xi} \exp(-u), \tag{6.1}$$

respectively, where $1 + \xi(x - \mu)/\sigma > 0$, $-\infty < \xi < \infty$, $-\infty < \mu < \infty$, $\sigma > 0$ and $u = \{1 + \xi(x - \mu)/\sigma\}^{-1/\xi}$, which is used throughout this chapter. Possible applications of the GEV distribution cover most areas of science, engineering and medicine. Some published applications are mentioned in previous chapters.

In recent years, several extensions of the GEV distribution have been proposed. The most recent of these is due to Chen et al. (2010). Earlier generalizations include the three-parameter kappa distribution due to Mielke Jr (1973) and the four-parameter kappa distribution due to Hosking (1994). Chen's generalization has the

cumulative distribution function and the probability density function given by

$$F(x) = \left\{ 1 + \exp\left[-\frac{1}{\delta}(x - \mu)^\alpha \right] \right\}^{-1/\beta}, \qquad (6.2)$$

and

$$f(x) = \alpha(\delta\beta)^{-1}(x - \mu)^{\alpha-1} \exp\left[-\frac{1}{\delta}(x - \mu)^\alpha \right] \left\{ 1 + \exp\left[-\frac{1}{\delta}(x - \mu)^\alpha \right] \right\}^{-1/\beta-1} \quad (6.3)$$

respectively, for $-\infty < x < \infty$, $\alpha > 0$, $\beta > 0$, $\delta > 0$ and $-\infty < \mu < \infty$. An excellent motivation for introducing (6.2) and (6.3) is described in Chen et al. (2010). However, neither of (6.2) and (6.3) appear to be valid functions since $(x - \mu)^\alpha$ is undefined for $x < \mu$. Here, we provide a modification of Chen et al. (2010)'s generalization to correct this error. We specify the cumulative distribution function and the probability density function by

$$F(x) = \left(1 - 2^{-1/\beta} \right)^{-1} \left[\left\{ 1 + \exp\left[-\frac{1}{\delta}(x - \mu)^\alpha \right] \right\}^{-1/\beta} - 2^{-1/\beta} \right], \qquad (6.4)$$

and

$$f(x) = \frac{\alpha(x - \mu)^{\alpha-1}}{\delta\beta(1 - 2^{-1/\beta})} \exp\left[-\frac{1}{\delta}(x - \mu)^\alpha \right] \left\{ 1 + \exp\left[-\frac{1}{\delta}(x - \mu)^\alpha \right] \right\}^{-1/\beta-1} \quad (6.5)$$

respectively, for $\alpha > 0$, $\beta > 0$, $\delta > 0$ and $x > \mu > -\infty$. Clearly, both (6.4) and (6.5) are valid functions. We shall refer to the distribution given by (6.4) and (6.5) as the Chen distribution.

If X is a random variable with probability density function (6.5), we write $X \sim$ Chen$(\alpha, \beta, \delta, \mu)$. The Chen quantile function is obtained by inverting (6.4)

$$x = Q(z) = F^{-1}(z) = \mu + \left[-\delta \ln\left\{ \left[2^{-1/\beta} + \left(1 - 2^{-1/\beta} \right) q \right]^{-\beta} - 1 \right\} \right]^{1/\alpha}. \quad (6.6)$$

So, one can generate Chen variates from (6.6) by $X = Q(U)$, where U is a uniform variate on the unit interval $(0, 1)$.

In the rest of this Chapter, we provide a comprehensive description of the mathematical properties of (6.5). We examine the shape of (6.5) and its associated hazard rate function in sections 6.2 and 6.3, respectively. We derive expressions for the moments in the section 6.4. Order statistics, their moments and L moments are

calculated in the section 6.5. Asymptotic distributions of the extreme values are provided in the section 6.6. Estimation by the method of maximum likelihood, including the observed information matrix, is presented in the section 6.7. A simulation study is presented in the section 6.8 to assess the performance of the maximum likelihood estimators. Application of the Chen distribution to a real data set is illustrated in the section 6.9.

Results in the section 6.4 involve infinite series representations. The terms of these infinite series are elementary, so infinite series can be computed by truncation using any standard package.

6.2   Shape of the Probability Density Function

The first derivative of $\ln(f(x))$ for the Chen distribution is:

$$\frac{d \ln f(x)}{dx} = -\frac{\alpha}{\delta}(x-\mu)^{\alpha-1} + \frac{\alpha-1}{x-\mu} + \frac{\alpha}{\delta}\left(\frac{1}{\beta}+1\right)\frac{(x-\mu)^{\alpha-1}}{1+\exp\left[\frac{1}{\delta}(x-\mu)^{\alpha}\right]}.$$

So, modes of $f(x)$ are the roots of the equation

$$\frac{\alpha}{\delta}(x-\mu)^{\alpha} - \frac{\alpha}{\delta}\left(\frac{1}{\beta}+1\right)\frac{(x-\mu)^{\alpha}}{1+\exp\left[\frac{1}{\delta}(x-\mu)^{\alpha}\right]} = \alpha - 1. \tag{6.7}$$

There may be more than one root to (6.7). If $x = x_0$ is a root of (6.7) then it corresponds to a local maximum if $d\ln f(x)/dx > 0$ for all $x < x_0$ and $d\ln f(x)/dx < 0$ for all $x > x_0$. It corresponds to a local maximum if $d\ln f(x)/dx < 0$ for all $x < x_0$ and $d\ln f(x)/dx > 0$ for all $x > x_0$. It corresponds to a point of inflexion if either $d\ln f(x)/dx > 0$ for all $x \neq x_0$ or $d\ln f(x)/dx < 0$ for all $x \neq x_0$.

Plots of shapes of (6.5) for $\mu = 0$, $\delta = 1$ and selected values of (α, β) are given in Figure (6.1). Both unimodal and monotonically decreasing shapes appear possible. Unimodal shapes appear for large α. Monotonically decreasing shapes appear for small α.

Furthermore, the asymptotes of $f(x)$ and $F(x)$ as $x \to 0, \infty$ are given by

$$f(x) \sim \alpha(\delta\beta)^{-1}\left(1-2^{-1/\beta}\right)^{-1} x^{\alpha-1} \exp\left[-\frac{1}{\delta}(x-\mu)^{\alpha}\right]; \quad \text{as } x \to \infty,$$

Figure 6.1: Plots of the pdf of Chen distribution for $\mu = 0$, $\delta = 1$, $\alpha = 0.5, 1, 2, 5$, $\beta = 0.5$ (black curve), $\beta = 1$ (red curve), $\beta = 2$ (green curve) and $\beta = 5$ (blue curve).

$$f(x) \sim \alpha(\delta\beta)^{-1}2^{-1/\beta-1}\left(1-2^{-1/\beta}\right)^{-1}(x-\mu)^{\alpha-1}; \qquad \text{as } x \to \mu,$$

$$1 - F(x) \sim \beta^{-1}\left(1-2^{-1/\beta}\right)^{-1}\exp\left[-\frac{1}{\delta}(x-\mu)^{\alpha}\right]; \qquad \text{as } x \to \infty,$$

and

$$F(x) \sim (\delta\beta)^{-1}2^{-1/\beta-1}\left(1-2^{-1/\beta}\right)^{-1}(x-\mu)^{\alpha}; \qquad \text{as } x \to \mu.$$

Note that the upper tail of $f(x)$ is that of a Weibull distribution. The lower tail is polynomial.

6.3   Shape of the Hazard Rate Function

For the Chen distribution, $h(x)$ takes the form

$$h(x) = \alpha \frac{(x-\mu)^{\alpha-1}\exp\left[-\frac{1}{\delta}(x-\mu)^{\alpha}\right]\left\{1+\exp\left[-\frac{1}{\delta}(x-\mu)^{\alpha}\right]\right\}^{-1/\beta-1}}{\delta\beta\left[1-\left\{1+\exp\left[-\frac{1}{\delta}(x-\mu)^{\alpha}\right]\right\}^{-1/\beta}\right]}. \qquad (6.8)$$

The first derivative of $\ln h(x)$ is:

$$\frac{d\ln h(x)}{dx} = -\frac{\alpha}{\delta}(x-\mu)^{\alpha-1} + \frac{\alpha-1}{x-\mu} + \frac{\alpha}{\delta}\left(\frac{1}{\beta}+1\right)\frac{(x-\mu)^{\alpha-1}}{1+\exp\left[\frac{1}{\delta}(x-\mu)^{\alpha}\right]}$$

$$+\alpha\frac{(x-\mu)^{\alpha-1}\exp\left[-\frac{1}{\delta}(x-\mu)^{\alpha}\right]\left\{1+\exp\left[-\frac{1}{\delta}(x-\mu)^{\alpha}\right]\right\}^{-1/\beta-1}}{\delta\beta\left[1-\left\{1+\exp\left[-\frac{1}{\delta}(x-\mu)^{\alpha}\right]\right\}^{-1/\beta}\right]}.$$

So, modes of $h(x)$ are the roots of the equation

$$\alpha - 1 = \frac{\alpha}{\delta}(x-\mu)^{\alpha} - \frac{\alpha}{\delta}\left(\frac{1}{\beta}+1\right)\frac{(x-\mu)^{\alpha}}{1+\exp\left[\frac{1}{\delta}(x-\mu)^{\alpha}\right]} - \alpha(x-\mu)^{\alpha}$$

$$\times\frac{\exp\left[-\frac{1}{\delta}(x-\mu)^{\alpha}\right]\left\{1+\exp\left[-\frac{1}{\delta}(x-\mu)^{\alpha}\right]\right\}^{-1/\beta-1}}{\delta\beta\left[1-\left\{1+\exp\left[-\frac{1}{\delta}(x-\mu)^{\alpha}\right]\right\}^{-1/\beta}\right]}. \qquad (6.9)$$

There may be more than one root to (6.9). If $x = x_0$ is a root of (6.9) then it corresponds to a local maximum if $d \ln h(x)/dx > 0$ for all $x < x_0$ and $d \ln h(x)/dx < 0$ for all $x > x_0$. It corresponds to a local maximum if $d \ln h(x)/dx < 0$ for all $x < x_0$ and $d \ln h(x)/dx > 0$ for all $x > x_0$. It corresponds to a point of inflexion if either $d \ln h(x)/dx > 0$ for all $x \neq x_0$ or $d \ln h(x)/dx < 0$ for all $x \neq x_0$.

Furthermore, the asymptotes of $h(x)$ as $x \to 0, \infty$ are given by

$$h(x) \sim \alpha \delta^{-1} x^{\alpha-1}; \qquad \text{as } x \to \infty,$$

and

$$h(x) \sim \alpha(\delta\beta)^{-1} 2^{-1/\beta-1} \left(1 - 2^{-1/\beta}\right)^{-1} (x - \mu)^{\alpha-1}; \quad \text{as } x \to \mu.$$

Note that both the upper and lower tails of $h(x)$ behave polynomially with respect to x.

Figure (6.2) illustrates some of the possible shapes of $h(x)$ for $\mu = 0$, $\delta = 1$ and selected values of (α, β). Both monotonically increasing, monotonically decreasing and upside down bathtub shapes appear possible.

1. Upside down bathtub shapes appear for small values of α and β;

2. Monotonically decreasing shapes appear for small values of α;

3. Monotonically increasing shapes appear for large values of α.

Upside down bathtub shaped hazard rates are a widely spread shape in reliability and survival analysis. Silva et al. (2010) presented an example for such hazard rates,which can be observed in the course of a disease whose mortality reaches a peak after some finite period, and then declines gradually . For other practical examples yielding upside down bathtub hazard rates, see Singh and Misra (1994).

It is interesting to note that the Chen distribution can exhibit upside down bathtub shapes. However, the GEV distribution cannot exhibit upside down bathtub shaped hazard rates.
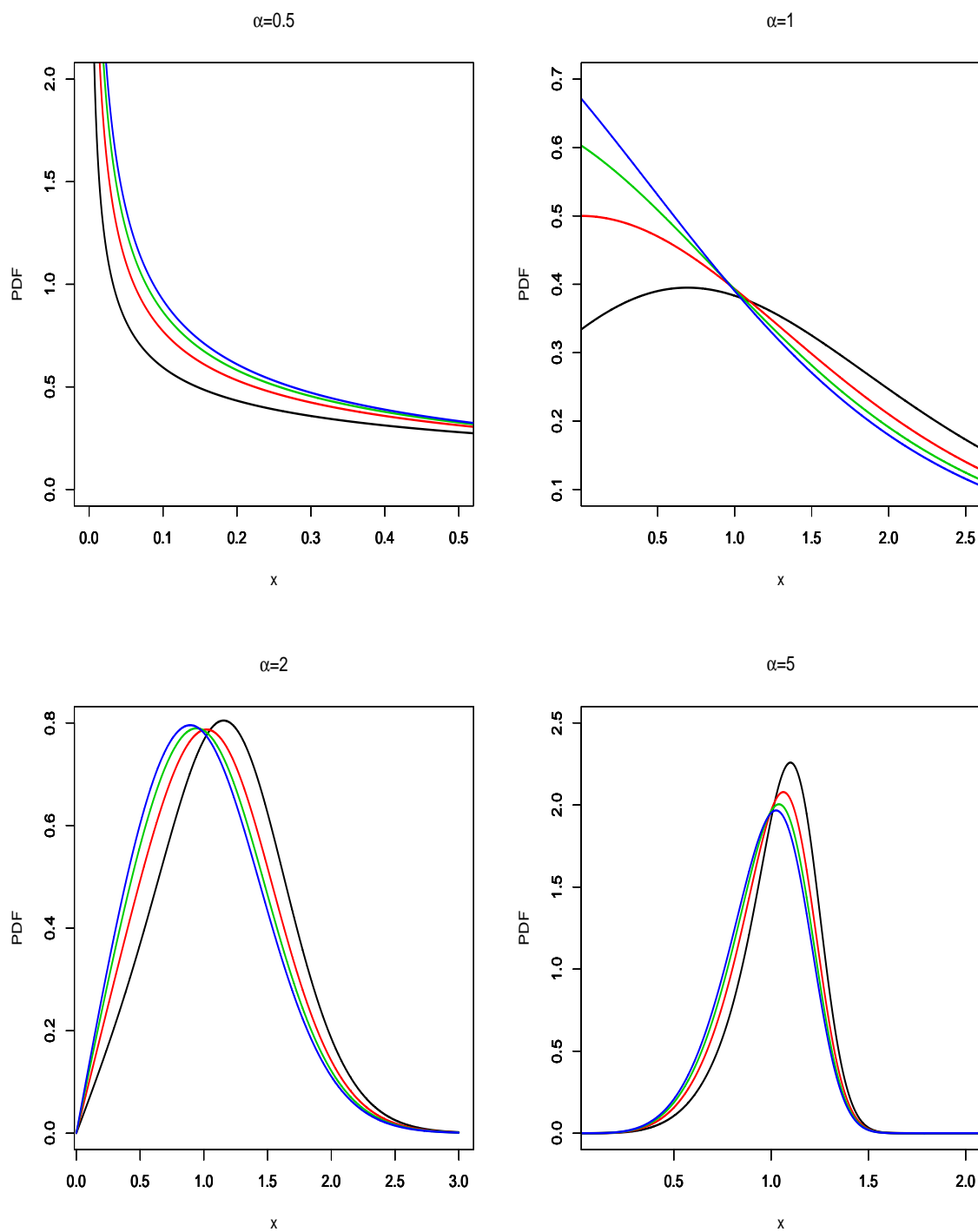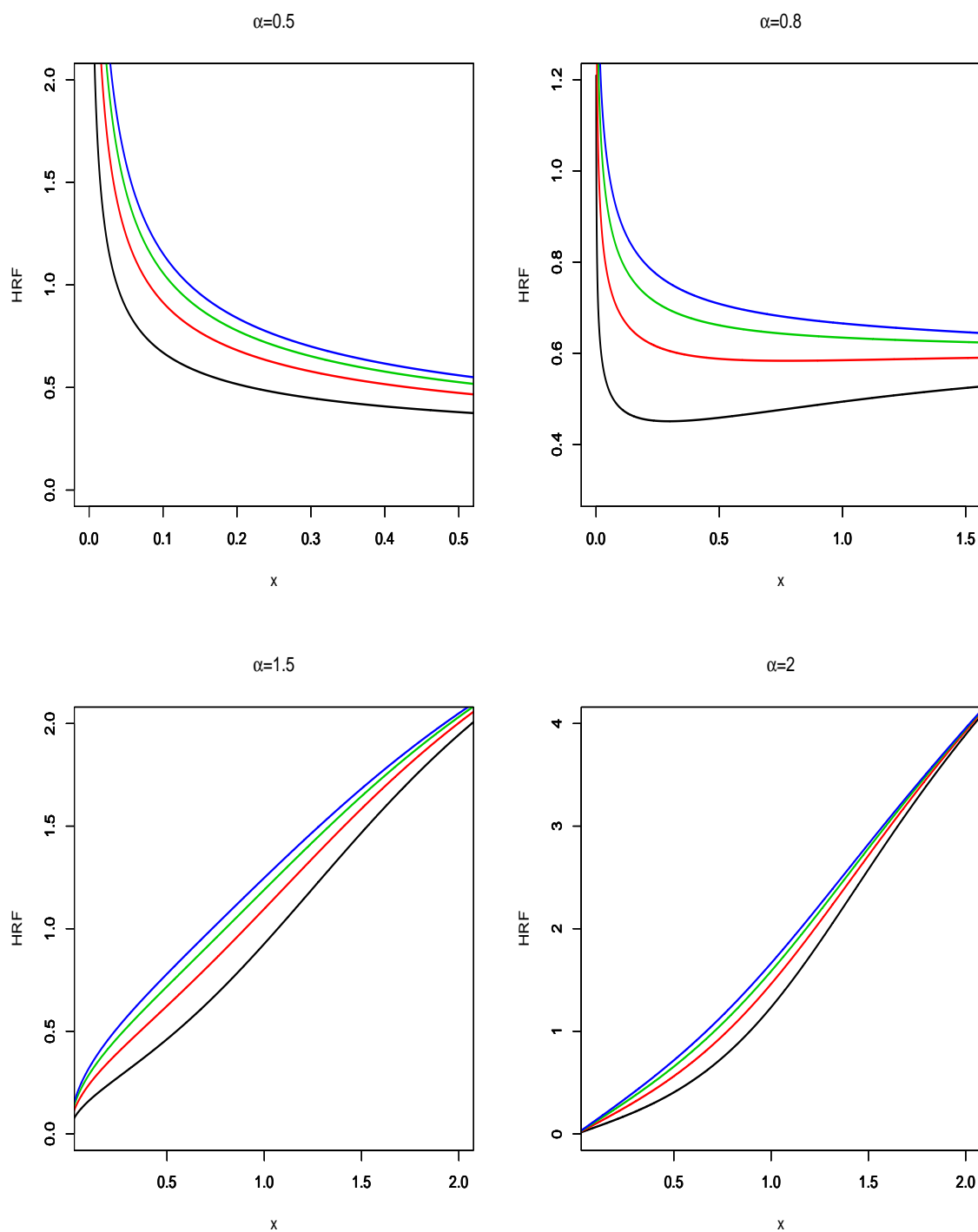
Figure 6.2: Plots of the hrf of the Chen distribution for $\mu = 0$, $\delta = 1$, $\alpha = 0.5, 0.8, 1.5, 2$, $\beta = 0.5$ (black curve), $\beta = 1$ (red curve), $\beta = 2$ (green curve) and $\beta = 5$ (blue curve).

6.4 Moments

Let $X \sim \text{Chen}(\alpha, \beta, \delta, \mu)$. Using the binomial expansion, we can write

$$
\begin{aligned}
E\left(X^n\right) &= E\left((X - \mu + \mu)^n\right) \\
&= \sum_{m=0}^{n} \binom{n}{m} \mu^{n-m} E\left((X - \mu)^m\right) \\
&= \alpha(\delta\beta)^{-1} \left(1 - 2^{-1/\beta}\right)^{-1} \sum_{m=0}^{n} \binom{n}{m} \mu^{n-m} \int_{\mu}^{\infty} (x - \mu)^{m+\alpha-1} \\
&\quad \times \exp\left[-\frac{1}{\delta}(x - \mu)^\alpha\right] \left\{1 + \exp\left[-\frac{1}{\delta}(x - \mu)^\alpha\right]\right\}^{-1/\beta-1} dx \\
&= \alpha(\delta\beta)^{-1} \left(1 - 2^{-1/\beta}\right)^{-1} \sum_{m=0}^{n} \binom{n}{m} \mu^{n-m} \int_{\mu}^{\infty} (x - \mu)^{m+\alpha-1} \\
&\quad \times \exp\left[-\frac{1}{\delta}(x - \mu)^\alpha\right] \sum_{k=0}^{\infty} \binom{-1/\beta - 1}{k} \exp\left[-\frac{k}{\delta}(x - \mu)^\alpha\right] dx \\
&= \alpha(\delta\beta)^{-1} \left(1 - 2^{-1/\beta}\right)^{-1} \sum_{m=0}^{n} \binom{n}{m} \sum_{k=0}^{\infty} \mu^{n-m} \binom{-1/\beta - 1}{k} \\
&\quad \times \int_{\mu}^{\infty} (x - \mu)^{m+\alpha-1} \exp\left[-\frac{k+1}{\delta}(x - \mu)^\alpha\right] dx \\
&= \beta^{-1} \left(1 - 2^{-1/\beta}\right)^{-1} \sum_{m=0}^{n} \binom{n}{m} \sum_{k=0}^{\infty} \mu^{n-m} \binom{-1/\beta - 1}{k} \\
&\quad \times \delta^{m/\alpha}(k+1)^{-m/\alpha-1} \int_{0}^{\infty} y^{m/\alpha} \exp(-y) \, dy \\
&= \beta^{-1} \left(1 - 2^{-1/\beta}\right)^{-1} \sum_{m=0}^{n} \binom{n}{m} \sum_{k=0}^{\infty} \mu^{n-m} \binom{-1/\beta - 1}{k} \\
&\quad \times \delta^{m/\alpha}(k+1)^{-m/\alpha-1} \Gamma\left(m/\alpha + 1\right)
\end{aligned}
\tag{6.10}
$$

for $m > 0$ any real number. The first four moments are given by:

$$
\begin{aligned}
E(X) = \beta^{-1} \left(1 - 2^{-1/\beta}\right)^{-1} & \left[\sum_{k=0}^{\infty} \mu \binom{-1/\beta - 1}{k} (k+1)^{-1} \right. \\
& \left. + \sum_{k=0}^{\infty} \binom{-1/\beta - 1}{k} \delta^{1/\alpha}(k+1)^{-1/\alpha-1} \Gamma\left(1/\alpha + 1\right) \right],
\end{aligned}
\tag{6.11}
$$

$$E\left(X^2\right) = \beta^{-1}\left(1-2^{-1/\beta}\right)^{-1}\left[\sum_{k=0}^{\infty}\mu^2\binom{-1/\beta-1}{k}(k+1)^{-1}\right.$$

$$+2\sum_{k=0}^{\infty}\mu\binom{-1/\beta-1}{k}\delta^{1/\alpha}(k+1)^{-1/\alpha-1}\Gamma\left(1/\alpha+1\right)$$

$$\left.+\sum_{k=0}^{\infty}\binom{-1/\beta-1}{k}\delta^{2/\alpha}(k+1)^{-2/\alpha-1}\Gamma\left(2/\alpha+1\right)\right], \qquad (6.12)$$

$$E\left(X^3\right) = \beta^{-1}\left(1-2^{-1/\beta}\right)^{-1}\left[\sum_{k=0}^{\infty}\mu^3\binom{-1/\beta-1}{k}(k+1)^{-1}\right.$$

$$+3\sum_{k=0}^{\infty}\mu^2\binom{-1/\beta-1}{k}\delta^{1/\alpha}(k+1)^{-1/\alpha-1}\Gamma\left(1/\alpha+1\right)$$

$$+3\sum_{k=0}^{\infty}\mu\binom{-1/\beta-1}{k}\delta^{2/\alpha}(k+1)^{-2/\alpha-1}\Gamma\left(2/\alpha+1\right)$$

$$\left.+\sum_{k=0}^{\infty}\binom{-1/\beta-1}{k}\delta^{3/\alpha}(k+1)^{-3/\alpha-1}\Gamma\left(3/\alpha+1\right)\right], \qquad (6.13)$$

and

$$E\left(X^4\right) = \beta^{-1}\left(1-2^{-1/\beta}\right)^{-1}\left[\sum_{k=0}^{\infty}\mu^4\binom{-1/\beta-1}{k}(k+1)^{-1}\right.$$

$$+4\sum_{k=0}^{\infty}\mu^3\binom{-1/\beta-1}{k}\delta^{1/\alpha}(k+1)^{-1/\alpha-1}\Gamma\left(1/\alpha+1\right)$$

$$+6\sum_{k=0}^{\infty}\mu^2\binom{-1/\beta-1}{k}\delta^{2/\alpha}(k+1)^{-2/\alpha-1}\Gamma\left(2/\alpha+1\right)$$

$$+4\sum_{k=0}^{\infty}\mu\binom{-1/\beta-1}{k}\delta^{3/\alpha}(k+1)^{-3/\alpha-1}\Gamma\left(3/\alpha+1\right)$$

$$\left.+\sum_{k=0}^{\infty}\binom{-1/\beta-1}{k}\delta^{4/\alpha}(k+1)^{-4/\alpha-1}\Gamma\left(4/\alpha+1\right)\right]. \qquad (6.14)$$

The infinite series in (6.10)-(6.14) all converge.

The expressions given by (6.11)-(6.14) can be used to compute the mean, variance, skewness and kurtosis of X. The values of these four quantities versus α are plotted in Figure (6.3) for $\mu = 0$, $\delta = 1$ and selected values of β.
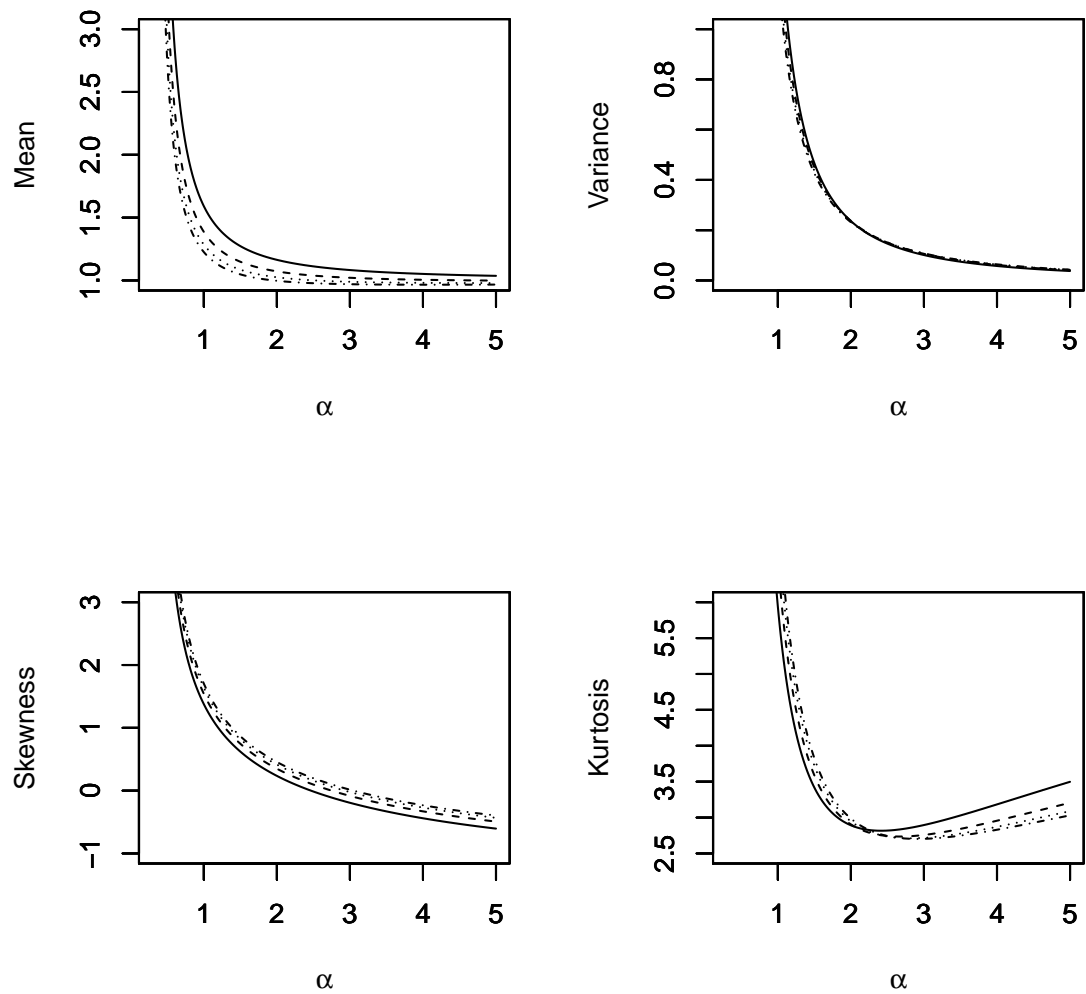
Figure 6.3: Mean, variance, skewness, and kurtosis of the Chen distribution, versus α for $\mu = 0$, $\delta = 1$, $\beta = 0.5$ (solid curve), $\beta = 1$ (curve of dashes), $\beta = 2$ (curve of dots) and $\beta = 5$ (curve of dots and dashes).

From these measures, we can see that:

I. Mean, variance and skewness are monotonic decreasing functions of α.

II. Kurtosis initially decreases before increasing with respect to α.

III. Mean is a monotonic decreasing function of β.

IV. Skewness is a monotonic increasing function of β.

6.5 Order Statistic

Order statistic is the most statistical tools that appear in many areas of statistical theory and practice. Let $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$ denote the order statistics for a random sample $X_1, X_2, \ldots, X_n$ from (6.5). Then the pdf of the kth order statistic, say $Y = X_{k:n}$, can be expressed as

$$
\begin{aligned}
f_Y(y) &= \frac{\alpha \left(1 - 2^{-1/\beta}\right)^{-n-1} n!}{\delta \beta (k-1)!(n-k)!} (x-\mu)^{\alpha-1} \left\{ 1 + \exp\left[-\frac{1}{\delta}(x-\mu)^\alpha \right] \right\}^{-1/\beta-1} \\
&\quad \times \exp\left[-\frac{1}{\delta}(x-\mu)^\alpha \right] \left[\left\{ 1 + \exp\left[-\frac{1}{\delta}(x-\mu)^\alpha \right] \right\}^{-1/\beta} - 2^{-1/\beta} \right]^k \\
&\quad \times \left[1 - \left\{ 1 + \exp\left[-\frac{1}{\delta}(x-\mu)^\alpha \right] \right\}^{-1/\beta} \right]^{n-k} \\
&= \frac{\alpha \left(1 - 2^{-1/\beta}\right)^{-n-1} n!}{\delta \beta (k-1)!(n-k)!} \sum_{i=0}^{k-1} \sum_{j=0}^{n-k} \binom{k-1}{i} \binom{n-k}{j} (-1)^{k-i+j} 2^{(i+1-k)/\beta} \\
&\quad \times (x-\mu)^{\alpha-1} \exp\left[-\frac{1}{\delta}(x-\mu)^\alpha \right] \left\{ 1 + \exp\left[-\frac{1}{\delta}(x-\mu)^\alpha \right] \right\}^{-(i+j+1)/\beta-1} \\
&= \frac{\alpha \left(1 - 2^{-1/\beta}\right)^{-n-2} n!}{\delta \beta (k-1)!(n-k)!} \sum_{i=0}^{k-1} \sum_{j=0}^{n-k} \binom{k-1}{i} \binom{n-k}{j} \frac{2^{(i+1-k)/\beta} - 2^{-j/\beta}}{(-1)^{i-k-j}(i+j+1)} \\
&\quad \times f_{\alpha, \beta/(i+j+1), \delta, \mu}(x),
\end{aligned}
$$

where $f_{a,b,\sigma,\xi}(\cdot)$ denotes the probability density function of Chen (a, b, σ, ξ). So, the probability density function of Y is a finite linear combination of probability density functions of Chen random variables. Hence, other properties of Y can be easily

derived. For instance, the cumulative distribution function of Y can be expressed as

$$F_Y(y) = \frac{\alpha \left(1 - 2^{-1/\beta}\right)^{-n-2} n!}{\delta \beta (k-1)!(n-k)!} \sum_{i=0}^{k-1} \sum_{j=0}^{n-k} \binom{k-1}{i} \binom{n-k}{j}$$

$$\times \frac{2^{(i+1-k)/\beta} - 2^{-j/\beta}}{(-1)^{i-k-j}(i+j+1)} F_{\alpha,\beta/(i+j+1),\delta,\mu}(x),$$

where $F_{a,b,\sigma,\xi}(\cdot)$ denotes the cumulative distribution function corresponding to $f_{a,b,\sigma,\xi}(\cdot)$.
The qth moment of Y can be expressed as

$$E\left[Y^q\right] = \frac{\alpha \left(1 - 2^{-1/\beta}\right)^{-n-2} n!}{\delta \beta (k-1)!(n-k)!} \sum_{i=0}^{k-1} \sum_{j=0}^{n-k} \binom{k-1}{i} \binom{n-k}{j}$$

$$\times \frac{2^{(i+1-k)/\beta} - 2^{-j/\beta}}{(-1)^{i-k-j}(i+j+1)} E\left[X^q_{\alpha,\beta/(i+j+1),\delta,\mu}\right],$$

where $X_{a,b,\sigma,\xi} \sim \text{Chen}\,(a,b,\sigma,\xi)$.

Hosking (1990) explained that "L-moments are summary statistics for probability distributions and data samples . They are analogous to ordinary moments but are computed from linear functions of the order statistics". The rth L moment is defined by

$$\lambda_r = \sum_{j=0}^{r-1} (-1)^{r-1-j} \binom{r-1}{j} \binom{r-1+j}{j} \beta_j,$$

where $\beta_j = E\{XF(X)^j\}$. In particular, $\lambda_1 = \beta_0$, $\lambda_2 = 2\beta_1 - \beta_0$, $\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0$ and $\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0$. In general, $\beta_r = (r+1)^{-1} E(X_{r+1:r+1})$, so it can be computed using (6.15). Hosking (1990) clarified that "the L moments have several advantages over ordinary moments: for example, they apply for any distribution having finite mean; no higher-order moments need be finite".

6.6   Extreme Values

This section mirrors section (5.9). Suppose again $X_1, \ldots, X_n$ is a random sample from (6.5). $\overline{X}$ denotes the sample mean, then by using the central limit theorem, $\sqrt{n}(\overline{X} - \text{E}(X))/\sqrt{\text{Var}(X)}$ approaches the standard normal distribution as $n \to \infty$.

Here, we determine the max and min domains of attraction of the cumulative distribution function given by (6.4).

Let $g(t) = (\delta/\alpha)(t - \mu)^{1-\alpha}$. Then,

$$
\begin{aligned}
\lim_{t \to \infty} \frac{1 - F(t + xg(t))}{1 - F(t)} &= \lim_{t \to \infty} \exp\left\{ \frac{1}{\delta} [(t - \mu)^\alpha - (t + xg(t) - \mu)^\alpha] \right\} \\
&= \lim_{t \to \infty} \exp\left\{ \frac{1}{\delta}(t - \mu)^\alpha \left[1 - \left(1 + \frac{xg(t)}{t - \mu} \right)^\alpha \right] \right\} \\
&= \lim_{t \to \infty} \exp\left\{ -\frac{\alpha}{\delta}(t - \mu)^\alpha g(t)x \right\} \\
&= \exp(-x)
\end{aligned}
$$

for every $x \in (-\infty, \infty)$. So, it again follows by (Leadbetter et al., 1983, chap. 1) that F belongs to the max domain of attraction of the Gumbel extreme value distribution with

$$
\lim_{n \to \infty} \Pr\{a_n (M_n - b_n) \le x\} = \exp\{-\exp(-x)\}
$$

for some suitable norming constants $a_n > 0$ and b_n.

Also, by using Corollary 1.6.3 in Leadbetter et al. (1983), we can determine the form of the norming constants. One can see that $b_n = F^{-1}(1 - 1/n)$ and $a_n = (\alpha/\delta)(b_n - \mu)^{\alpha-1}$, where $F^{-1}(\cdot)$ denotes the inverse function of $F(\cdot)$.

For the min domain of attraction, we note that

$$
\lim_{t \to 0} \frac{F(tx + \mu)}{F(t + \mu)} = \lim_{t \to 0} \left(\frac{tx}{t} \right)^\alpha = x^\alpha.
$$

So, F belongs to the min domain of attraction of the Weibull extreme value distribution.

6.7 Maximum Likelihood Estimation

Suppose once more that $X_1, X_2, \ldots, X_n$ is a random sample of size n from (6.5). Then the log-likelihood function for the vector of parameters $(\alpha, \beta, \delta, \mu)$ can be written as

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

$$
\times \sum_{i=1}^{n} \ln(x_i - \mu) - \left(\frac{1}{\beta} + 1 \right) \sum_{i=1}^{n} \ln\left\{ 1 + \exp\left[-\frac{1}{\delta}(x_i - \mu)^\alpha \right] \right\} \tag{6.15}
$$

The first-order partial derivatives of (6.15) with respect to the four parameters are:

$$\frac{\partial \ln L}{\partial \alpha} = \frac{n}{\alpha} - \frac{n}{\delta} \sum_{i=1}^{n} (x_i - \mu)^{\alpha} \ln (x_i - \mu) + \sum_{i=1}^{n} \ln (x_i - \mu)$$

$$+ \frac{1}{\delta} \left(\frac{1}{\beta} + 1 \right) \sum_{i=1}^{n} \frac{\exp \left[-\frac{1}{\delta} (x_i - \mu)^{\alpha} \right] (x_i - \mu)^{\alpha} \ln (x_i - \mu)}{1 + \exp \left[-\frac{1}{\delta} (x_i - \mu)^{\alpha} \right]}, \quad (6.16)$$

$$\frac{\partial \ln L}{\partial \alpha} = -\frac{n}{\beta} + \frac{n \ln 2}{\beta^2 \left(2^{1/\beta} - 1 \right)} + \frac{1}{\beta^2} \sum_{i=1}^{n} \ln \left\{ 1 + \exp \left[-\frac{1}{\delta} (x_i - \mu)^{\alpha} \right] \right\}, \quad (6.17)$$

$$\frac{\partial \ln L}{\partial \delta} = -\frac{n}{\delta} + \frac{1}{\delta^2} \sum_{i=1}^{n} (x_i - \mu)^{\alpha} - \frac{1}{\delta^2} \left(\frac{1}{\beta} + 1 \right)$$

$$\times \sum_{i=1}^{n} \frac{\exp \left[-\frac{1}{\delta} (x_i - \mu)^{\alpha} \right] (x_i - \mu)^{\alpha}}{1 + \exp \left[-\frac{1}{\delta} (x_i - \mu)^{\alpha} \right]}, \quad (6.18)$$

and

$$\frac{\partial \ln L}{\partial \mu} = \frac{\alpha}{\delta} \sum_{i=1}^{n} (x_i - \mu)^{\alpha-1} - (\alpha - 1) \sum_{i=1}^{n} (x_i - \mu)^{-1} - \frac{\alpha}{\delta} \left(\frac{1}{\beta} + 1 \right)$$

$$\times \sum_{i=1}^{n} \frac{\exp \left[-\frac{1}{\delta} (x_i - \mu)^{\alpha} \right] (x_i - \mu)^{\alpha-1}}{1 + \exp \left[-\frac{1}{\delta} (x_i - \mu)^{\alpha} \right]}. \quad (6.19)$$

The maximum likelihood estimates of $(\alpha, \beta, \delta, \mu)$, say $(\widehat{\alpha}, \widehat{\beta}, \widehat{\delta}, \widehat{\mu})$, are the simultaneous solutions of the equations $\partial \ln L / \partial \alpha = 0$, $\partial \ln L / \partial \beta = 0$, $\partial \ln L / \partial \delta = 0$ and $\partial \ln L / \partial \mu = 0$. As $n \to \infty$, $(\widehat{\alpha} - \alpha, \widehat{\beta} - \beta, \widehat{\delta} - \delta, \widehat{\mu} - \mu)$ approaches a multivariate normal vector with zero means and variance-covariance matrix $-(E[\mathbf{J}])^{-1}$, where

$$\mathbf{J} = \begin{pmatrix} \dfrac{\partial^2 \ln L}{\partial \alpha^2} & \dfrac{\partial^2 \ln L}{\partial \alpha \partial \beta} & \dfrac{\partial^2 \ln L}{\partial \alpha \partial \delta} & \dfrac{\partial^2 \ln L}{\partial \alpha \partial \mu} \\[2mm] \dfrac{\partial^2 \ln L}{\partial \beta \partial \alpha} & \dfrac{\partial^2 \ln L}{\partial \beta^2} & \dfrac{\partial^2 \ln L}{\partial \beta \partial \delta} & \dfrac{\partial^2 \ln L}{\partial \beta \partial \mu} \\[2mm] \dfrac{\partial^2 \ln L}{\partial \delta \partial \alpha} & \dfrac{\partial^2 \ln L}{\partial \delta \partial \mu} & \dfrac{\partial^2 \ln L}{\partial \delta^2} & \dfrac{\partial^2 \ln L}{\partial \delta \partial \mu} \\[2mm] \dfrac{\partial^2 \ln L}{\partial \mu \partial \alpha} & \dfrac{\partial^2 \ln L}{\partial \mu \partial \beta} & \dfrac{\partial^2 \ln L}{\partial \mu \partial \delta} & \dfrac{\partial^2 \ln L}{\partial \delta^2} \end{pmatrix}.$$

The matrix, $-E[\mathbf{J}]$, is known as the expected information matrix. The matrix, $-\mathbf{J}$, is known as the observed information matrix.

In simulations, and real data applications described later on, we maximized the log-likelihood function using the nlm function in the R (R Development (2011)) statistical package. For each maximization, the nlm function was executed for a wide range of initial values. At least one maximum was identified each time. In cases of more than one maximum, we took the maximum likelihood estimates to correspond to the largest of the maxima.

In practice, n is finite. The literature (see, for example, Efron and Hinkley (1978) suggests that it is best to approximate the distribution of $(\widehat{\alpha} - \alpha, \widehat{\beta} - \beta, \widehat{\delta} - \delta, \widehat{\mu} - \mu)$ by a multivariate normal distribution with zero means and variance-covariance matrix given by $-\mathbf{J}^{-1}$, inverse of the observed information matrix, with $(\alpha, \beta, \delta, \mu)$ replaced $(\widehat{\alpha}, \widehat{\beta}, \widehat{\delta}, \widehat{\mu})$. So, it is useful to have explicit expressions for the elements of $\mathbf{J}$. They are given in appendix A.

The multivariate normal approximation can be used to construct approximate confidence intervals and confidence regions for the individual parameters and for the hazard and survival functions.

6.8 Simulation Study

In this section, we assess the performance of the maximum likelihood estimates given by (6.16)-(6.19) with respect to sample size n. The assessment is based on a simulation study:

1. Generate ten thousand samples of size n from (6.5). The inversion method is used to generate samples, i.e variates of the Chen distribution are generated using (6.6).

2. Compute the maximum likelihood estimates for the thousand samples, say $(\widehat{\alpha}_i, \widehat{\beta}_i, \widehat{\delta}_i, \widehat{\mu}_i)$ for $i = 1, 2, \ldots, 10000$.

3. Compute the biases and mean squared errors given by

$$\text{bias}_h(n) = \frac{1}{10000} \sum_{i=1}^{10000} \left(\widehat{h}_i - h \right)$$
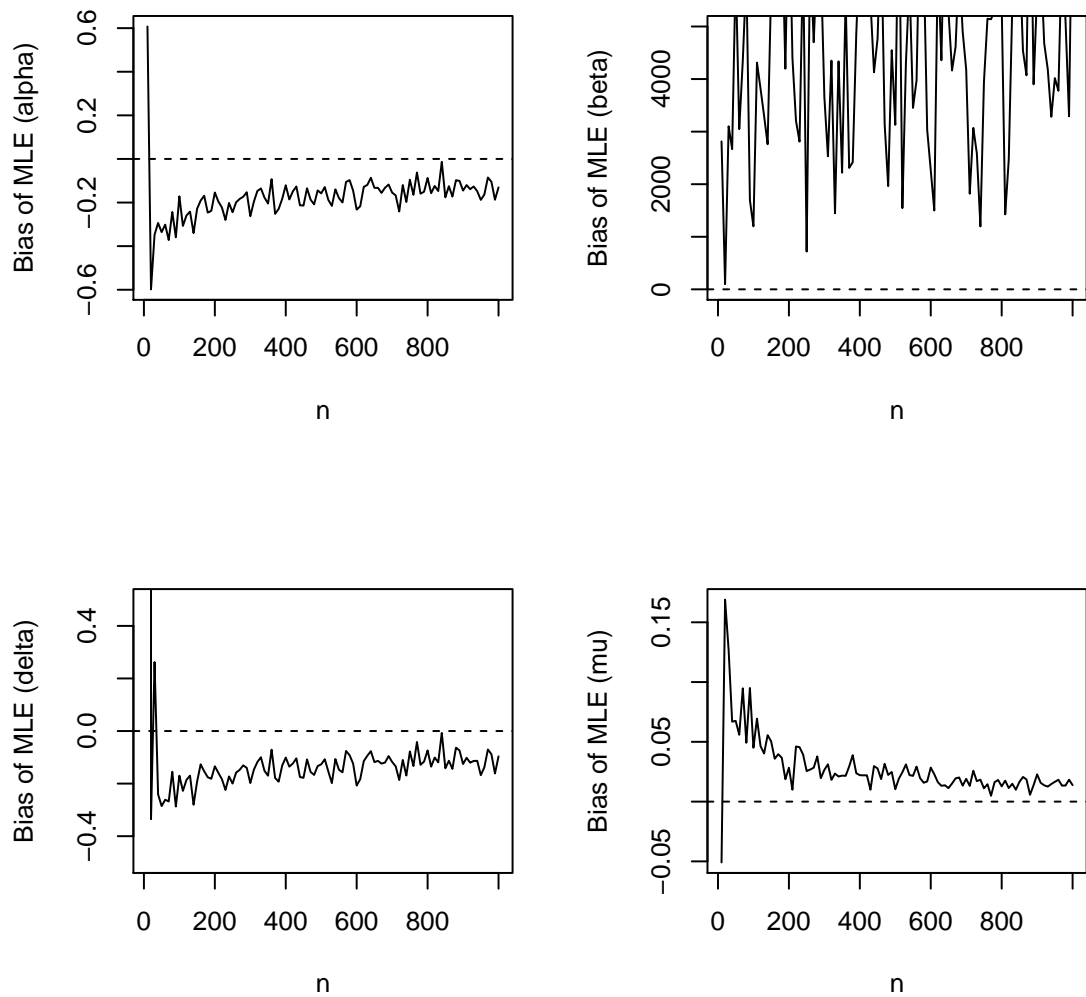
Figure 6.4: $\text{bias}_\alpha(n)$ (top left), $\text{bias}_\beta(n)$ (top right), $\text{bias}_\delta(n)$ (middle right) and $\text{bias}_\mu(n)$ (bottom left) versus $n = 10, 20, \ldots, 1000$.

and

$$\text{MSE}_h(n) = \frac{1}{10000} \sum_{i=1}^{10000} \left(\widehat{h}_i - h \right)^2$$

for $h = \alpha, \beta, \delta, \mu$.

We repeat these steps for $n = 10, 20, \ldots, 1000$ with $\alpha = 2$, $\beta = 2$, $\delta = 1$ and $\mu = 0$, so computing $\text{bias}_\alpha(n)$, $\text{bias}_\beta(n)$, $\text{bias}_\delta(n)$, $\text{bias}_\mu(n)$ and $\text{MSE}_\alpha(n)$, $\text{MSE}_\beta(n)$, $\text{MSE}_\delta(n)$, $\text{MSE}_\mu(n)$ for $n = 10, 20, \ldots, 1000$.
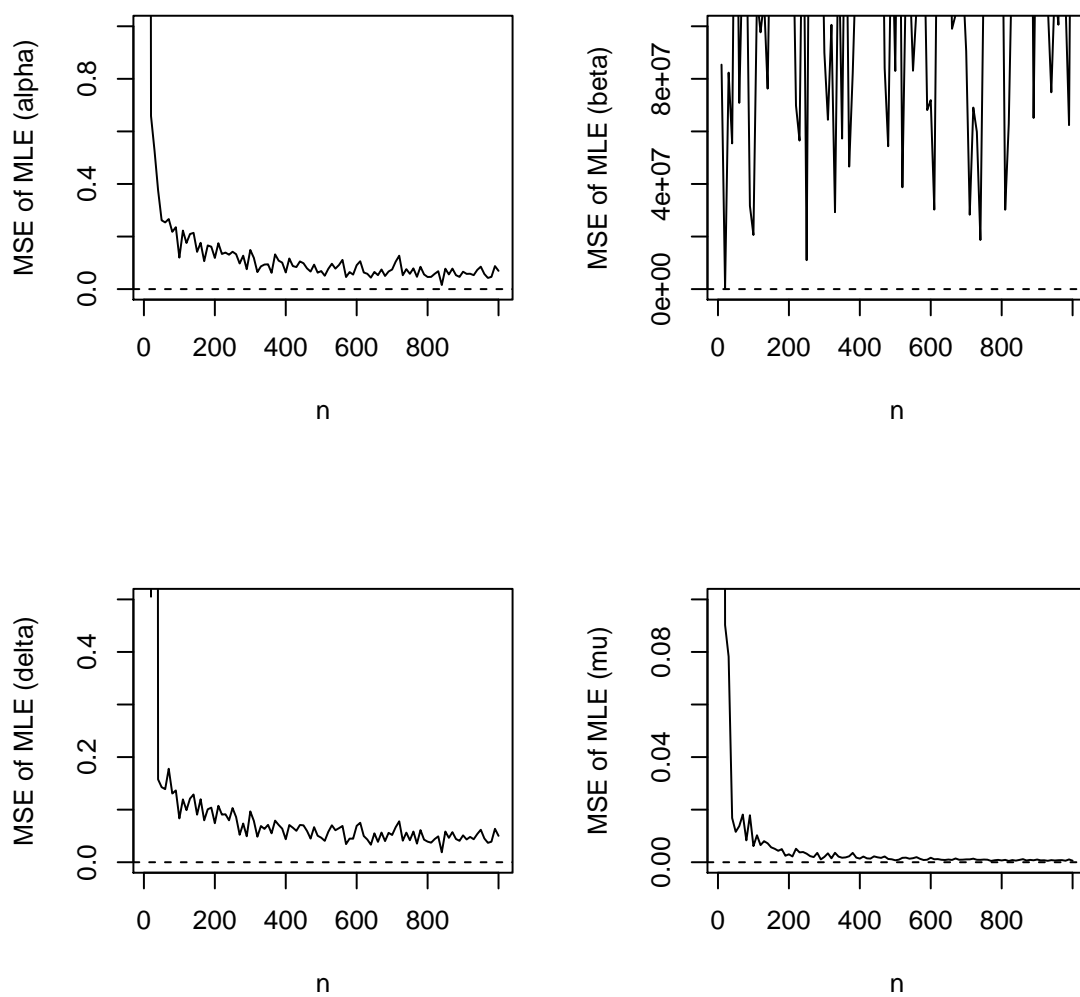
Figure 6.5: $\mathrm{MSE}_\alpha(n)$ (top left), $\mathrm{MSE}_\beta(n)$ (top right),$\mathrm{MSE}_\delta(n)$ (middle right) and $\mathrm{MSE}_\mu(n)$ (bottom left) versus $n = 10, 20, \ldots, 1000$.

Figures 6.4 and 6.5 show how the four biases and the four mean squared errors vary with respect to n. The broken line in Figure 6.4 corresponds to the biases being zero. The broken line in Figure 6.5 corresponds to the mean squared errors being zero. The following observations can be made:

1. Biases for α and δ are generally negative;

2. Biases for β and μ are generally positive;

3. Biases for each parameter generally approach zero as $n \to \infty$; exclude biases of the parameter b;

4. Biases appear largest for β; they appear exceedingly large;

5. Biases appear smallest for μ;

6. Mean squared errors for each parameter generally decrease to zero as $n \to \infty$; exclude Mean squared errors of the parameter b;

7. Mean squared errors appear largest for β; again they appear exceedingly large;

8. Mean squared errors appear smallest for μ.

We have presented results for only one choice for $(\alpha, \delta, \delta, \mu)$, namely that $(\alpha, \delta, \delta, \mu) = (2, 2, 1, 0)$, but results were similar for other choices.

6.9   An Application

To illustrate the flexibility of the Chen distribution, we fitted this model to the same real data set used in Chen et al. (2010). The data used by Chen et al. (2010) are present in appendix **C**.

We fitted the GEV and Chen distributions to the data. The maximum likelihood procedure described in the section 6.7 was used for fitting the Chen distribution. The fitted estimates for the Chen distribution were: $\widehat{\alpha} = 1.700(0.069)$, $\widehat{\beta} = 0.055(0.045)$, $\widehat{\delta} = 17125.09(10190.82)$, $\widehat{\mu} = -33.524(193.361)$ with $-\ln L = 324.443$ and $AIC =$
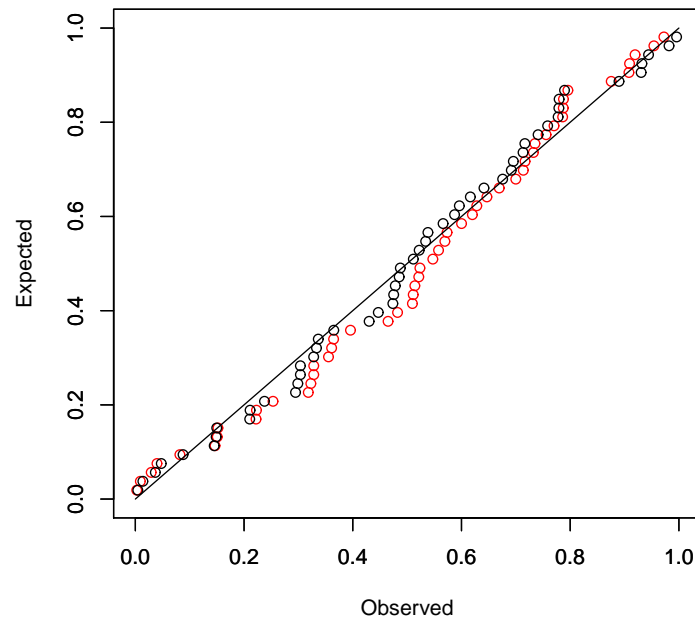
Figure 6.6: Probability plots for the fits of the GEV distribution (in red) and the Chen distribution (in black) for annual maximum rainfall from Maple Ridge in British Columbia.

656.885. The fitted estimates for the GEV distribution were: $\widehat{\mu} = 552.016(19.520)$, $\widehat{\sigma} = 129.755(13.307)$, $\widehat{\xi} = -0.308(0.072)$ with $-\ln L = 327.932$ and $AIC = 661.864$. The numbers within brackets are standard errors obtained by inverting the observed information matrix, see section 6.7.

We can see that the negative log-likelihood values and the AIC values are smaller for the Chen distribution. So, for the data set used in Chen et al. (2010), the Chen distribution provides a better fit. This is confirmed by the probability-probability plots, quantile-quantile plots and density plots shown in Figures 6.6- 6.8. The points in Figures 6.6 and 6.7 are closer to the diagonal lines for the Chen distribution. The fitted probability density function for the Chen distribution appears to better capture the histogram in Figure 6.8.

Furthermore, chi-square goodness of fit tests give the p-values of 0.039 and 0.071 for the GEV distribution and the Chen distribution, respectively, suggesting that
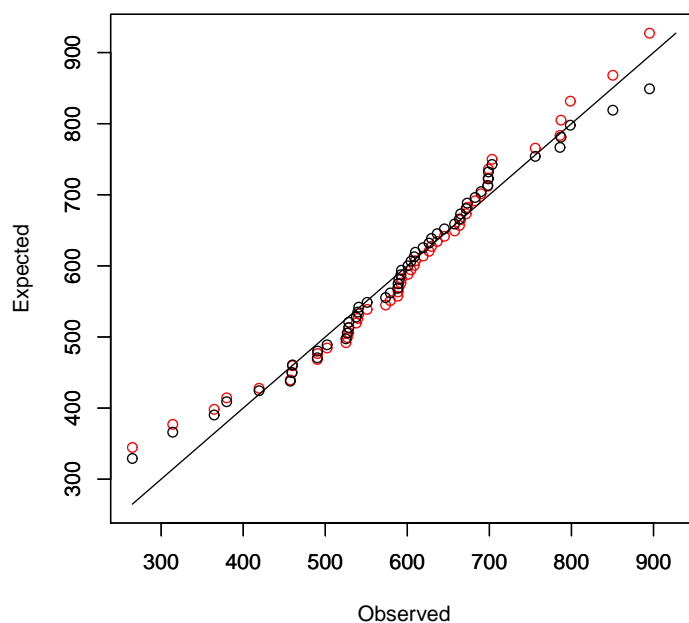
Figure 6.7: Quantile plots for the fits of the GEV distribution (in red) and the Chen distribution (in black) for annual maximum rainfall from Maple Ridge in British Columbia.
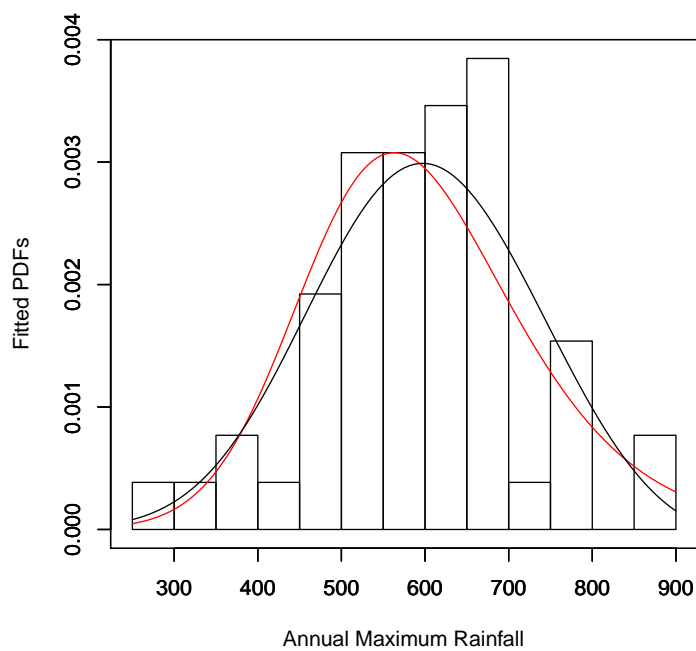
Figure 6.8: Density plots for the fits of the GEV distribution (in red) and the Chen distribution (in black) for annual maximum rainfall from Maple Ridge in British Columbia.

Eq(6.5) provides the only adequate fit.

6.10   Summary

This chapter discussed the most recent generalisation of GEV distribution introduced by Chen et al. (2010). This generalisation does not appear to be valid since $(x - \mu)^{\alpha}$ is undefined for $x < \mu$. In this study, we provide a modification of Chen et al. (2010)'s generalization to correct this error. The mathematical properties of this model are presented. Parameters of the Chen model are estimated by the maximum likelihood estimation method. To study the bias and mean square error (MSE) for the estimator, a simulation study is used.

Finally, using the same data as in Chen et al. (2010), we conclude that the corrected form of the Chen distribution appears to be an appropriate model for the annual maximum rainfall data compared to the GEV distribution.

Chapter 7

A New Distribution for Extreme Values

In this chapter, we introduce a new modification of the GEV distribution dependent on the quantile function. Because of its flexibility in the different pdf forms, this new model can be widely used to model some natural phenomena. The main idea of this new model is motivated by the question "How to make control on the thickness of the shapes of the extreme value models?". Chen et al. (2010) introduced a new distribution for the return period of antecedent precipitation by taking the power of the quantile function of the logistic distribution. According to Chen this new model has a variety of shapes and has a closed form density and distribution function. Also, it fits the real extreme value data better than known distributions of univariate extreme values, but the given formulas for the distribution do not form a valid probability distribution since $(x - \mu)^\alpha$ is undefined for $x < \mu$. Therefore, we corrected their formulas to form a valid probability distribution and presented them in the previous chapter. Here, we use the corrected formula to improve the GEV distribution. In section 7.1 the pdf and cdf of a new family are introduced. The relation to other distributions is presented in section 7.2. Some of the important statistical properties such as closeness under linear transformation, the variety of shapes of the pdf and hrf, moments and moment generating function and the analytical expression for the mean deviations about the mean and the median are introduced in section 7.3. We

derive the order statistics and extreme values for a new extreme value distribution in sections 7.4 and 7.5. In section 7.6, three methods to estimate the parameters are investigated. Finally, to illustrate the benefit of the new extreme value distribution over other models we fit these distributions to the real data.

7.1 Introduction

The distribution function of the GEV distribution is defined as

$$G(x, \mu, \sigma, \xi) = \exp(-u),$$

where $u = \{1 + \xi(x - \mu)/\sigma\}^{-1/\xi}$. So, the quantile function of GEV distribution can be written as

$$x_p = F^{-1}(p) = \mu + \frac{\sigma}{\xi} \left\{ \left[\log(\frac{1}{p}) \right]^{-\xi} - 1 \right\}, \tag{7.1}$$

According to Chen *et al.*(2010), the quantile function of the new distribution can be defined as

$$x_p = F^{-1}(p) = \mu + \left(\frac{\sigma}{\xi} \right)^{1/a} \left\{ \left[\log(\frac{1}{p}) \right]^{-\xi} - 1 \right\}^{1/a}, \tag{7.2}$$

where $p \in (0, 1)$ and the shape parameters α and ξ satisfy $a > 0$, $\xi \in (-\infty, \infty)$ and μ and σ are the location and scale parameters respectively. The major reason for adding another shape parameter is to control the thickness of the tail. This family of distributions has a closed form of its pdf, cdf and quantile function. In order to obtain the pdf and cdf we first derive two following functions

$$F(x) = \exp \left\{ - \left[1 + \xi \frac{(x - \mu)^a}{\sigma} \right]^{-1/\xi} \right\}, \tag{7.3}$$

$$f(x) = \frac{a}{\sigma} (x - \mu)^{a-1} \left[1 + \xi \frac{(x - \mu)^a}{\sigma} \right]^{-1/\xi-1} \exp \left\{ - \left[1 + \xi \frac{(x - \mu)^a}{\sigma} \right]^{-1/\xi} \right\}. \tag{7.4}$$

since $(x - \mu)^\alpha$ is undefined for $x < \mu$, we only consider the case when $x > \mu$ and hence ((7.5)) and ((7.6)) define the cdf and pdf for a new extreme value distribution, respectively. Also, we consider $t = \mu$ as a certain threshold. From now on let the new distribution of extremes values be denoted by NEV. Then the NEV distribution has

three parameters a, σ, and ξ. Throughout this chapter, we set $u = \left[1 + \xi \frac{(x-t)^a}{\sigma}\right]^{-1/\xi}$. Then the cdf and pdf can be written as

$$F(x) = \frac{1}{1 - e^{-1}} \left\{\exp\{-u\} - \exp\{-1\}\right\}, \tag{7.5}$$

$$f(x) = (1 - e^{-1})^{-1} \frac{a}{\sigma}\, u^{1+\xi}\, \left(\frac{\sigma(u^{-\xi} - 1)}{\xi}\right)^{1-1/a} \exp(-u); \quad x \in (t, \infty), \tag{7.6}$$

where $\sigma > 0$ and $\xi \in \mathbf{R}$. Let $C = (1 - e^{-1})^{-1}$. Then we can rewrite the cdf and pdf in the more simple forms:

$$F(x) = C \left\{\exp\{-u\} - \exp\{-1\}\right\}, \tag{7.7}$$

$$f(x) = \sigma^{-1} C\, a\, u^{1+\xi} \left(\sigma \left(u^{-\xi} - 1\right)/\xi\right)^{1-1/a} \exp(-u); \tag{7.8}$$

The hazard rate function takes the following form

$$h(x) = \frac{f(x)}{1 - F(x)},$$

$$h(x) = \sigma^{-1} C\, a\, u^{\xi+1} \frac{\left(\sigma \left(u^{-\xi} - 1\right)/\xi\right)^{1-1/a} \exp(-u)}{1 - C \left\{\exp\{-u\} - \exp\{-1\}\right\}}. \tag{7.9}$$

7.2 Relation to Other Distributions

From the NEV distribution, we can deduce some well known distributions. The following theorem gives the relation between NEV distribution and GEV distribution.

Theorem 7.1. *Let X be a r.v following the new distribution with a, ξ shape parameters; with $t = 0$. And $Y = X^{1/a}$, then*

$$X \sim NEVD(a,\, \sigma,\, \xi) \longleftrightarrow Y \sim GEV(0,\, \sigma,\, \xi).$$

7.3 Statistical Properties

In this section, we present some properties of the new distribution of extreme values.

7.3.1   Closeness Under Linear Transformation

Theorem 7.2. *Let X be a r.v following the new distribution with a, ξ shape parameters. And $Y = b\,X \pm d$, then*

$$X \sim NEV(a,\ t,\ \sigma,\ \xi) \longleftrightarrow Y \sim NEV(a,\ d \pm bt,\ b^a\sigma,\ \xi)$$

Proof:

We only proof in one direction:

$$f_Y(y) = f_X\left(\frac{y-d}{b}\right)|\frac{dx}{dy}|, \tag{7.10}$$

$$= \sigma^{-1}C\ a\ \left(\frac{y-d}{b} - t\right)^{a-1}\left[1 + \frac{\xi}{\sigma}\left(\frac{y-d}{b} - t\right)^a\right]^{-1/\xi-1}$$

$$\exp\left\{-\left[1 + \frac{\xi}{\sigma}\left(\frac{y-d}{b} - t\right)^a\right]^{-1/\xi}\right\}(\frac{1}{b}),$$

$$\therefore f(y) = \sigma^{*-1}C\ a\ (y - t^*)^{a-1}\left[1 + \frac{\xi}{\sigma^*}(y - t^*)^a\right]^{-1/\xi-1}$$

$$\exp\left\{-\left[1 + \frac{\xi}{\sigma^*}(y - t^*)^a\right]^{-1/\xi}\right\}, \tag{7.11}$$

where $\sigma^* = b^a\sigma$ and $t^* = d + bt$. Then Y is distributed with the NEV distribution with parameters $(a,\ b^a\sigma,\ \xi)$ and $t^* = d + bt$. We note that the random variable Y has the same shape parameters as the random variable X. In other words, the shape parameter of the NEV distribution is fixed, while the location and scale parameters change under the transformation. As a result of this, the skewness and the kurtosis for this model are constants and they are independent of the location and scale parameters.

7.3.2   Shapes of PDF and HRF

The first derivative of $\log f(x)$ is:

$$\frac{d}{dx}\log f(x) = \frac{a}{\xi}\left(\frac{\sigma(u^{-\xi} - 1)}{\xi}\right)^{-1/a}\left\{\frac{(a-1)}{a} - (1 - u^\xi)\left[(1 + \xi) - u\right]\right\}. \tag{7.12}$$

So, the modes are the roots of the equation

$$\frac{(a-1)}{a} = (1 - u^\xi)\left[(1 + \xi) - u\right]. \tag{7.13}$$

As it seems, this equation has no explicit solution. Therefore, we discuss two main points before clarifying the shapes of pdf and hrf.

a) **First case** when $a < 1$:

In this case the pdf has exponential shapes which means that there is no mode.

b) **Second case** when $a = 1$:

The pdf of the NEV distribution reduces to the pdf of the GEV distribution and equation (7.13) becomes $(1 - u^\xi)[(1 + \xi) - u] = 0$. However, we know that, the modes of the GEV distribution are defined as

$$mode = \begin{cases} \mu + \sigma \frac{(1+\xi)^{-\xi} - 1}{\xi} \ if \ \xi \neq 0, \\ \\ \quad\mu \qquad if \ \xi = 0. \end{cases} \tag{7.14}$$

The first derivative of $\log h(x)$ is:

$$\frac{d}{dx} \log h(x) = \left(\frac{\sigma(u^{-\xi} - 1)}{\xi} \right)^{-1/a} \left\{ (a - 1) - \frac{a(1 - u^\xi)}{\xi} \left[(1 + \xi) - \frac{u}{1 - e^u} \right] \right\}. \tag{7.15}$$

So, the modes are the roots of the equation

$$\xi(1 - 1/a)(1 - u^\xi)^{-1} = (1 + \xi) - \frac{u}{1 - e^u}. \tag{7.16}$$

Figures 7.1 and 7.2 show the different shapes for the pdf and hrf respectively. From these figures, we can see that pdf and hrf for the new distribution take shapes depending on the values of the shape parameters. For more details we consider three cases.

a) **First case** when $\xi < 0$:

The pdf of the NEV distribution is monotonically decreasing for $a < 1$ and bimodal for $a \geq 1$. The hrf takes bathtub shape when $a < 1$ while, monotonically increasing for $a \geq 1$.

b) **Second case** when $\xi \to 0$:

Monotonically decreasing and unimodal shapes appear possible in the pdf for $a \leq 1$ and $a > 1$, respectively. The hrf is monotonically decreasing for $a < 1$, slightly increasing in the case $a = 1$ and monotonically increasing for $a > 1$.

c) **Last case** when $\xi > 0$:

Monotonically decreasing shapes appear for $a \leq 1$, while unimodal for $a > 1$ for both the pdf and hrf. We can find the mode by solving Eq(7.13).

7.3.3   The Quantile Function

If X is a random variable with pdf of the new distribution, then we can write the quantile function as

$$x_p = F^{-1}(p) = t + \left\{ \frac{\sigma}{\xi} \left[-\log(p(1 - e^{-1})) \right]^{-\xi} - 1 \right\}^{1/a}. \tag{7.17}$$

From (7.17) we can generate random samples by $X = F^{-1}(p)$, where p is a uniform variate on the unit interval $(0, 1)$. In addition, we can calculate the median of the NEV distribution from the quantile function by setting $p = \frac{1}{2}$. The median is:

$$x_{0.5} = F^{-1}(p) = t + \left\{ \frac{\sigma}{\xi} [1.151822]^{-\xi} - 1 \right\}^{1/a}. \tag{7.18}$$

7.3.4   Moments

The moments of the NEV distribution are given by the following theorem.

Theorem 7.3. *Let X be a random variable with the pdf given in (7.8). Then the moments can be written as*

$$E(X^n) = C \sum_{k=0}^{n} \sum_{j=0}^{k/a} (-1)^{k/a-j} \binom{n}{k} \binom{k/a}{j} \left(\frac{\sigma}{\xi} \right)^{k/a} t^{n-k} \gamma \left(1 - \xi j, 1 \right). \tag{7.19}$$

Proof: We can write

$$E(x^n) = \int_{-\infty}^{\infty} x^n f(x) dx,$$

$$= \sigma^{-1} C \, a \, \left(\frac{\sigma}{\xi} \right)^{1-1/a} \int_{t}^{\infty} x^n u^{1+\xi} \left(u^{-\xi} - 1 \right)^{1-1/a} \exp(-u) dx \tag{7.20}$$

$$\therefore E(X^n) = C \, a \sigma^{-1} \left(\frac{\sigma}{\xi} \right)^{1-1/a} K_1(n). \tag{7.21}$$

Using Lemma 1 in appendix **C**, the moments of the NEV distribution can be written as

$$E(X^n) = C \sum_{k=0}^{n} \sum_{j=0}^{k/a} (-1)^{k/a-j} \binom{n}{k} \binom{k/a}{j} \left(\frac{\sigma}{\xi} \right)^{k/a} t^{n-k} \gamma \left(1 - \xi j, 1 \right).$$
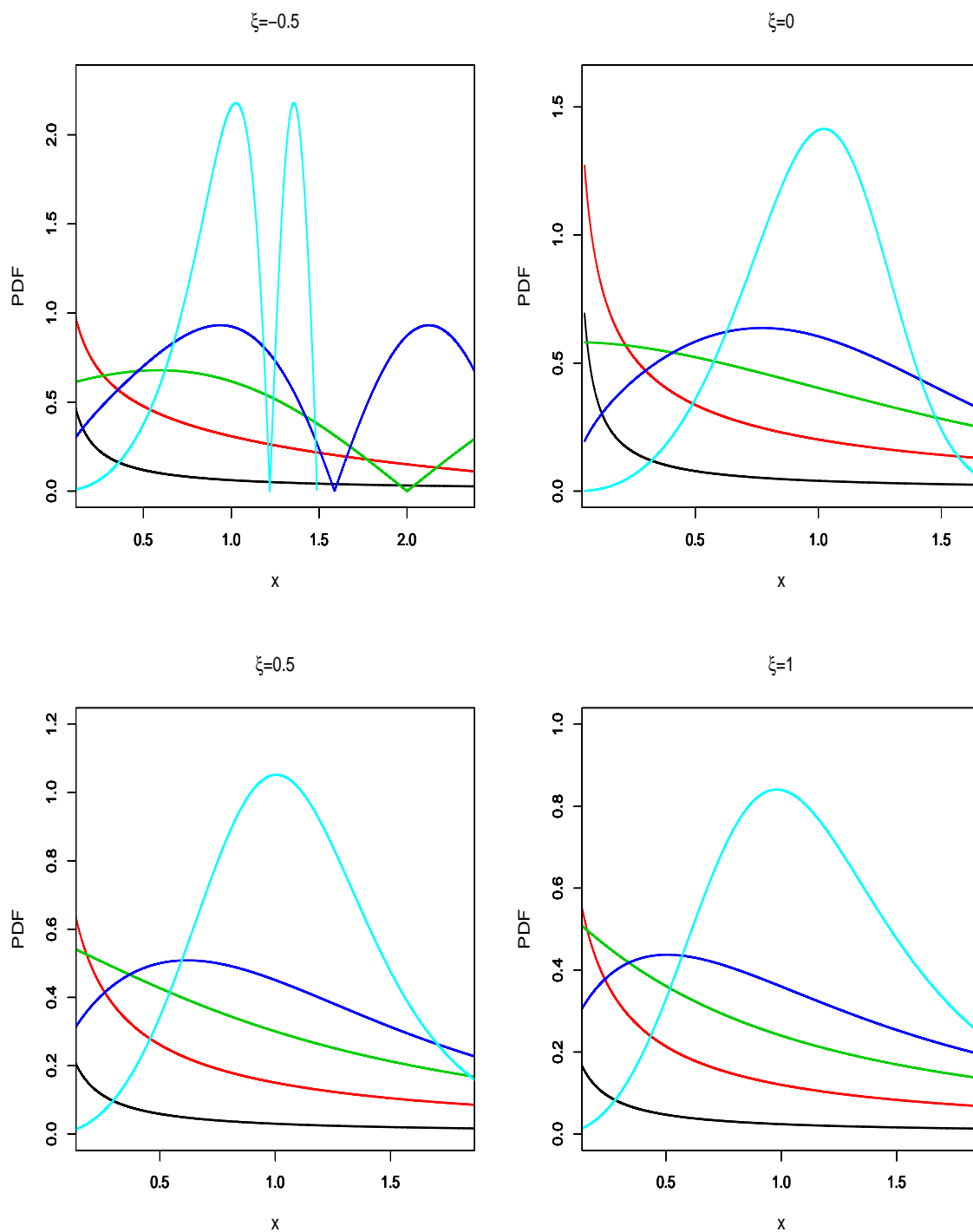
Figure 7.1: Plots of the pdf of the NEV distribution for $t = 0$, $\sigma = 1$, $\xi = -0.5, 0, 0.5, 1$, $a = 0.1$ (black curve), $a = 0.5$ (red curve), $a = 1$ (green curve), $a = 1.5$ (blue curve), and $a = 3.5$ (turquoise curve).
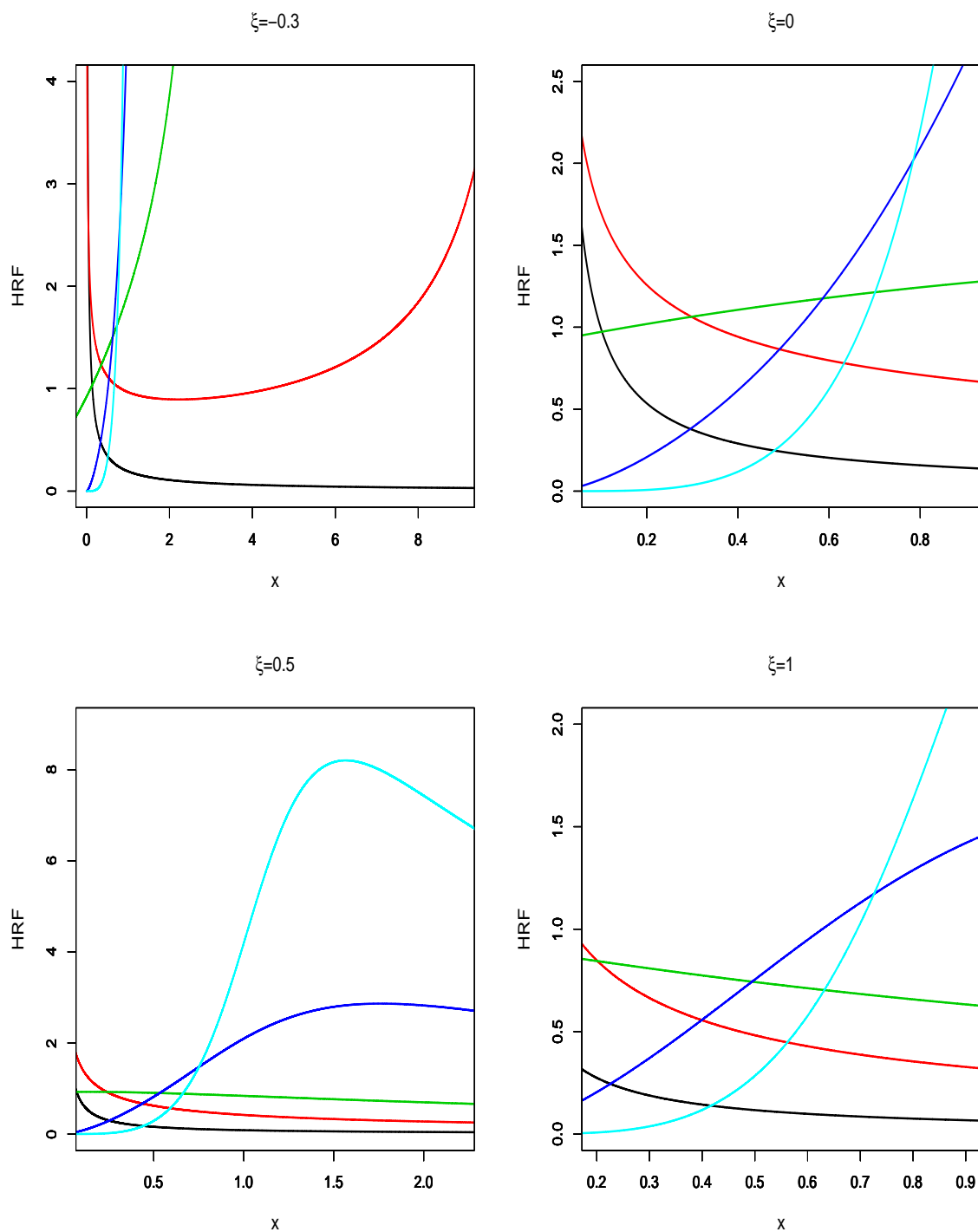
Figure 7.2: Plots of the hrf of the NEV distribution for $t = 0$, $\sigma = 1$, $\xi = -0.5, 0, 0.5, 1$, $a = 0.1$ (black curve), $a = 0.5$ (red curve), $a = 1$ (green curve), $a = 2.5$ (blue curve), and $a = 5$ (turquoise curve).

□

Based on this theorem, we can calculate the four main measures. First, compute the first four moments as

$$E\left(X\right) = C\left\{(-1)^{j}t\ \gamma(1,1) + \sum_{j=0}^{1/a}(-1)^{1/a-j}\binom{1/a}{j}\left(\frac{\sigma}{\xi}\right)^{1/a}\gamma(1-\xi j,1)\right\}, \qquad (7.22)$$

$$E\left(X^{2}\right) = C\left\{(-1)^{j}t^{2}\ \gamma(1,1) + 2\sum_{j=0}^{1/a}(-1)^{1/a-j}\binom{1/a}{j}\left(\frac{\sigma}{\xi}\right)^{1/a}t\ \gamma(1-\xi j,1)\right.$$
$$\left. + \sum_{j=0}^{2/a}(-1)^{2/a-j}\binom{2/a}{j}\left(\frac{\sigma}{\xi}\right)^{2/a}\gamma(1-\xi j,1)\right\}, \qquad (7.23)$$

$$E\left(X^{3}\right) = C\left\{(-1)^{j}t^{3}\ \gamma(1,1) + 3\sum_{j=0}^{1/a}(-1)^{1/a-j}\binom{1/a}{j}\left(\frac{\sigma}{\xi}\right)^{1/a}t^{2}\ \gamma(1-\xi j,1)\right.$$
$$\left. + 3\sum_{j=0}^{2/a}(-1)^{2/a-j}\binom{2/a}{j}\left(\frac{\sigma}{\xi}\right)^{2/a}t\ \gamma(1-\xi j,1) + \sum_{j=0}^{3/a}(-1)^{3/a-j}\binom{3/a}{j}\right.$$
$$\left. \times \left(\frac{\sigma}{\xi}\right)^{3/a}\gamma(1-\xi j,1)\right\}, \qquad (7.24)$$

$$E\left(X^{4}\right) = C\left\{(-1)^{j}t^{4}\ \gamma(1,1) + 4\sum_{j=0}^{1/a}(-1)^{1/a-j}\binom{1/a}{j}\left(\frac{\sigma}{\xi}\right)^{1/a}t^{3}\ \gamma(1-\xi j,1)\right.$$
$$\left. + 6\sum_{j=0}^{2/a}(-1)^{2/a-j}\binom{2/a}{j}\left(\frac{\sigma}{\xi}\right)^{2/a}t^{2}\ \gamma(1-\xi j,1) + 4\sum_{j=0}^{3/a}(-1)^{3/a-j}\binom{3/a}{j}\right.$$
$$\left. \times t\left(\frac{\sigma}{\xi}\right)^{3/a}\gamma(1-\xi j,1) + \sum_{j=0}^{4/a}(-1)^{4/a-j}\binom{4/a}{j}\left(\frac{\sigma}{\xi}\right)^{4/a}\gamma(1-\xi j,1)\right\}. \qquad (7.25)$$

As mentioned in previous chapters, we can use the first four moments to calculate the mean, variance, skewness and kurtosis.

7.3.5 Moment Generating Function

Let X be a random variable having the NEV distribution with $t = \mu$. Then the moment generating function can be calculated as

$$
\begin{aligned}
M(t) &= E(e^{tx}) \\
&= \int_{-\infty}^{\infty} e^{tx} f(x) dx \\
&= \sigma^{-1} C \, a \, \left(\frac{\sigma}{\xi}\right)^{1-1/a} \int_{\mu}^{\infty} e^{tx} u^{1+\xi} \left(u^{-\xi} - 1\right)^{1-1/a} \exp(-u) dx \\
&= \sigma^{-1} C \, a \, \left(\frac{\sigma}{\xi}\right)^{1-1/a} K_2(\mu, t),
\end{aligned}
\tag{7.26}
$$

where $t^* = t \left(\frac{\sigma}{\xi}\right)^{1/a}$. Using Lemma 2 in appendix **C**, the moment generating function can be written as

$$
M(t) = C \, e^{t\mu} \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \frac{(-1)^{j/a-i} t^{*j}}{j!} \binom{j/a}{i} \gamma(1 - \xi i, 1).
\tag{7.27}
$$

Using the relation between the moments and moment generating function, we can calculate moments from the moment generating function.

7.3.6 Mean Deviations

The mean deviations about mean and median, denoted by $\delta_1(x)$ and $\delta_2(x)$, respectively, are defined as

$$
\delta_1(x) = \int_{-\infty}^{\infty} |x - \mu_0| f(x) \, dx,
$$

and

$$
\delta_2(x) = \int_{-\infty}^{\infty} |x - M| f(x) \, dx
$$

where μ_0, M are the mean and median . These measures can be calculated by using the following relations:

$$
\begin{aligned}
\delta_1(X) &= \int_\mu^\infty |x - \mu_0| f(x)\, dx, \\
&= \int_\mu^{\mu_0} (\mu_0 - x)\, f(x) dx + \int_{\mu_0}^\infty (x - \mu_0)\, f(x) dx \\
&= \mu_0 F(\mu_0) - \int_\mu^{\mu_0} x f(x) dx - \mu_0 \{1 - F(\mu_0)\} + \int_{\mu_0}^\infty x f(x) dx \\
&= 2\mu_0 F(\mu_0) - 2\mu_0 + 2 \int_{\mu_0}^\infty x f(x) dx
\end{aligned}
$$

and

$$
\begin{aligned}
\delta_2(X) &= \int_\mu^\infty |x - M| f(x)\, dx, \\
&= \int_\mu^M (M - x) f(x) dx + \int_M^\infty (x - M) f(x) dx \\
&= 2 \int_M^\infty x f(x) dx - \mu_0.
\end{aligned}
$$

If $u = \{1 + \xi/\sigma (x - \mu)^\alpha\}^{-1/\xi}$ and $C = (1 - e^{-1})^{-1}$ then

$$
\begin{aligned}
\int_z^\infty x f(x) dx &= C \frac{a\, (\sigma/\xi)^{1-1/a}}{\sigma} \int_z^\infty x\, u^{1+\xi}\, (u^{-\xi} - 1)^{1-1/a} \exp(-u) dx \\
&= C \frac{a\, (\sigma/\xi)^{1-1/a}}{\sigma} K_3(z),
\end{aligned}
$$

where the final step follows by Lemma 3 in appendix **C**. Then we can write the mean deviations about mean and median as

$$
\delta_1(X) = 2\mu_0 C \left\{ e^{-u_0} - e^{-1} \right\} - 2\mu_0 + 2\sigma^{-1} ab \sum_{k=0}^\infty \binom{b-1}{k} (-1)^k J(\mu_0, (k+1)a),
$$

and

$$
\delta_2(X) = 2\sigma^{-1} ab \sum_{k=0}^\infty \binom{b-1}{k} (-1)^k J(M, (k+1)a) - \mu_0,
$$

where μ_0 is given by (7.22) and

$$
M = \mu + \frac{\sigma}{\xi} \left\{ \left[-\frac{1}{a} \log\left\{1 - 2^{-1/b}\right\} \right]^{-\xi} - 1 \right\}.
$$

7.4 Maximum Likelihood Estimation Method

Suppose that $X_1, X_2, ..., X_n$ are a random sample from the NEV distribution and assume that the vector θ is defined as $\theta = (a, \sigma, \xi)$. Note that the pdf of the new distribution is:

$$
f(x) = \sigma^{-1} C \, a \, (x - t)^{a-1} \left[1 + \frac{\xi}{\sigma} (x - t)^a \right]^{-1/\xi - 1}
$$

$$
\exp \left\{ - \left[1 + \frac{\xi}{\sigma} (x - t)^a \right]^{-1/\xi} \right\} \tag{7.28}
$$

Then the likelihood function of the parameters can be written as

$$
\prod_{i=1}^{n} f(x; \theta) = \sigma^{-n} (C \, a)^n \prod_{i=1}^{n} (x_i - t)^{a-1} \prod_{i=1}^{n} \left[1 + \frac{\xi}{\sigma} (x_i - t)^a \right]^{-1/\xi - 1}
$$

$$
\exp \left\{ - \sum_{i=1}^{n} \left[1 + \frac{\xi}{\sigma} (x_i - t)^a \right]^{-1/\xi} \right\} \tag{7.29}
$$

The Log-Likelihood function takes the form

$$
\log L(x; \theta) = -n \log \sigma + n \log(C \, a) + \sum_{i=1}^{n} \log \left[1 + \frac{\xi}{\sigma} (x_i - t)^a \right]^{-1/\xi - 1}
$$

$$
+ \sum_{i=1}^{n} \log (x_i - t)^{a-1} - \sum_{i=1}^{n} \left[1 + \frac{\xi}{\sigma} (x_i - t)^a \right]^{-1/\xi} \tag{7.30}
$$

By differentiating Eq (7.30) with respect to a, σ, ξ and equating them to zero, we obtain the maximum likelihood estimates (MLEs) $\hat{\theta} = (\hat{a}, \hat{\sigma}, \hat{\xi})$ for $\theta = (a, \sigma, \xi)$ as the solutions of the following equations

$$
\frac{\partial \log L}{\partial a} = \frac{n}{a} + \frac{n}{a^2} \log(\frac{\sigma}{\xi}) + \frac{1}{a^2} \sum_{i=1}^{n} \left(u_i^{-\xi} - 1 \right) + (1 - 1/a) \frac{\xi}{\sigma} \sum_{i=1}^{n} (x_i - t)
$$

$$
\times \log (x_i - t) - \frac{(1 + \xi)}{\sigma} \sum_{i=1}^{n} u_i^{\xi} (x_i - t)^a \log (x_i - t)
$$

$$
+ \frac{1}{\sigma} \sum_{i=1}^{n} u_i^{\xi+1} (x_i - t)^a \log (x_i - t), \tag{7.31}
$$

$$\frac{\partial \log L}{\partial \sigma} = -\frac{n}{a\,\sigma} - (1 - 1/a)\frac{\xi}{\sigma^2}\sum_{i=1}^{n}(x_i - t)^a + \frac{(1+\xi)}{\sigma^2}\sum_{i=1}^{n}u_i^{\xi}(x_i - t)^a$$

$$-\frac{1}{\sigma^2}\sum_{i=1}^{n}u_i^{\xi+1}(x_i - t)^a, \tag{7.32}$$

and

$$\frac{\partial \log L}{\partial \xi} = -\frac{n}{\xi}(1 - 1/a) + (1 - 1/a)\sum_{i=1}^{n}\frac{(x_i - t)^a}{\sigma^2} + \sum_{i=1}^{n}\log u_i - \frac{1+\xi}{\xi}$$

$$\times \sum_{i=1}^{n}\left(\log u_i + u_i^{\xi}\frac{(x_i - t)^a}{\sigma}\right) + \sum_{i=1}^{n}\frac{u_i}{\xi}\left(\log u_i + u_i^{\xi}\frac{(x_i - t)^a}{\sigma}\right). \tag{7.33}$$

For interval estimation and testing of hypothesis for the parameters $\theta = (a, \sigma, \xi)$, the fisher information matrix is required. The observed information matrix is denoted by $-J$, where J takes the following form

$$\mathbf{J} = \begin{pmatrix} \dfrac{\partial^2 \log L}{\partial a^2} & \dfrac{\partial^2 \log L}{\partial a \partial \sigma} & \dfrac{\partial^2 \log L}{\partial a \partial \xi} \\ \dfrac{\partial^2 \log L}{\partial \sigma \partial a} & \dfrac{\partial^2 \log L}{\partial \sigma^2} & \dfrac{\partial^2 \log L}{\partial \sigma \partial \xi} \\ \dfrac{\partial^2 \log L}{\partial \xi \partial a} & \dfrac{\partial^2 \log L}{\partial \xi \partial \sigma} & \dfrac{\partial^2 \log L}{\partial \xi^2} \end{pmatrix}.$$

The elements of this matrix are given in appendix $\mathbf{D}$.

The asymptotic distribution of the vector $\sqrt{(n)}\left(\hat{(\theta)} - \theta\right)$ is multivariate normal distribution with mean vector 0 and the variance-covariance matrix $K(\theta)^{-1}$, where $K(\theta)$ is the expected information matrix written as

$$k(\theta) = \begin{pmatrix} k_{a,a}(\theta) & k_{a,\sigma}(\theta) & k_{a,\xi}(\theta) \\ & k_{\sigma,\sigma}(\theta) & k_{\sigma,\xi}(\theta) \\ & & k_{\xi,\xi}(\theta) \end{pmatrix}.$$

7.5   Applications

To carry out the comparison between NEV distribution and other distributions, we have to determine which statistical distributions can potentially be used to analyse annual maximum rainfall data. As shown in chapter 1, there are several models that can be considered as best models for rainfall events. The most popular model for this kind of data is the GEV distribution which is the most widely used model in univariate extreme value analysis. In this section we apply these two models to maximum rainfall data in Uccle, Belgium.

7.5.1   Maximum Rainfall Data in Uccle, Belgium

To construct the best model for maximum rainfall data, we fitted GEV and NEV distributions to the annual maximum rainfall data in Uccle presented in chapter 4. Figure 7.3 shows the maximum rainfall data.
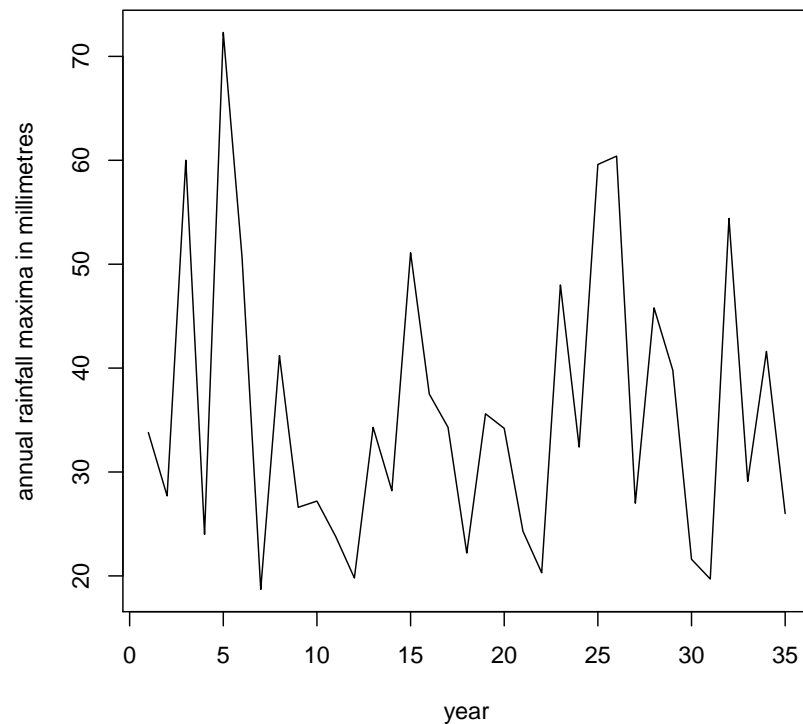
Figure 7.3: Plot of the maximum rainfall data in Uccle.

Empirical Mean Residual Life Plot:

To determine the threshold for the maximum rainfall data in Uccle, we plot a mean residual life plot of the data. From Figure 7.4 we choose $t = 15.9014mm$. This threshold shown in red line seems appropriate.

Mean Residual Life Plot
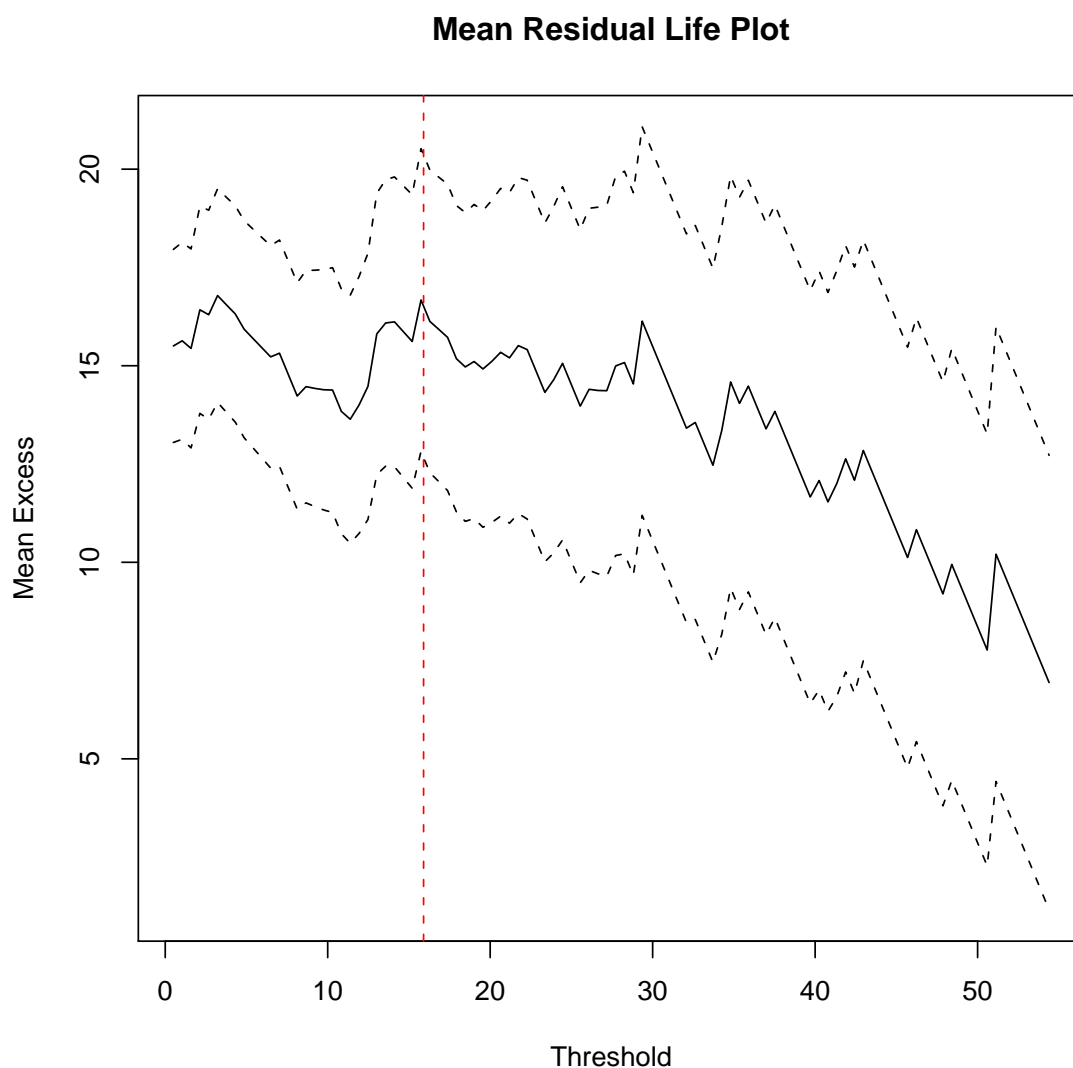
Figure 7.4: Mean residual life plot for exceedances of the levels of maximum rainfall in Uccle over the threshold 15.9014.

Estimates of the parameters:

The estimators for two models' parameters are shown in the table below:

Since the models are not nested and cannot be compared by the -log-Likelihood, we compute information criteria, which are presented in Table 7.2

Model	a	μ	σ	ξ	-Log-Lik
GEV	1	28.3824	9.0291	0.2316	136.9071
NEV	1.326973874	15.9014	46.065499897	0.006723105	136.0909

Table 7.1: MLEs of the model parameters and Log-Lik for Uccle maximum rainfall data.

Model	AIC	BIC	CAIC	AICc
GEV	279.8142	284.4802	287.4802	280.8142
NEV	278.1818	282.8478	285.8478	279.1818

Table 7.2: Information criteria AIC, BIC, CAIC and AICc.

Graphs:

After estimating the parameters for both models, we can visualize them. In this application, we employ probability-probability plots, quantile-quantile plots and density plots to confirm results shown by information criteria (AIC, BIC, CAIC and AICc).

As shown in Tables 7.1-7.2 the negative log-likelihood values and the information criteria values are smaller for the NEV distribution. Therefore, the NEV distribution provides a better fit for annul maximum rainfall data. This result is proven by the probability-probability plots, quantile-quantile plots and density plots shown in Figures 7.5- 7.7. The points in Figures 7.5 and 7.6 are closer to the diagonal lines for the NEV distribution. The fitted probability density function for the NEV distribution appears to better capture the upper tail of the histogram in Figure 7.7.
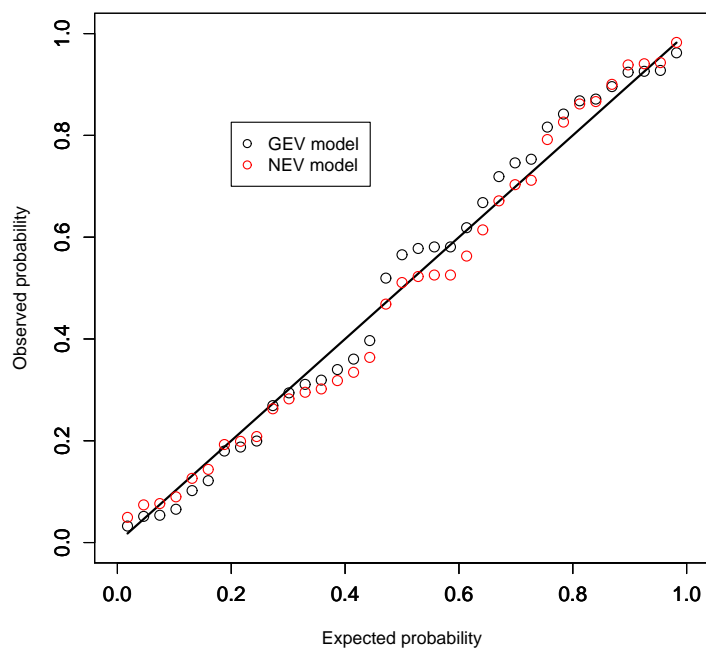
Figure 7.5: Probability plots for the fits of the NEV and GEV distributions for annual maximum rainfall from Uccle, Belgium.
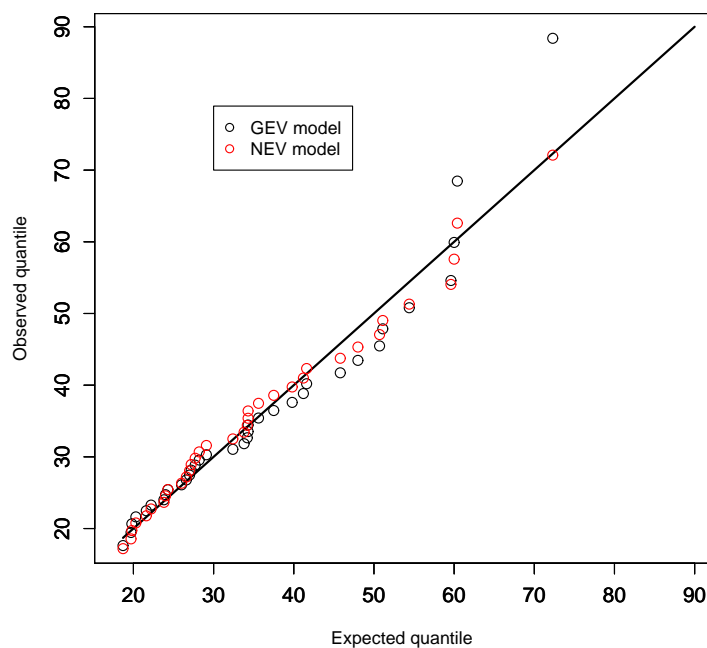
Figure 7.6: Quantile plots for the fits of the NEV and GEV distributions for annual maximum rainfall from Uccle, Belgium.
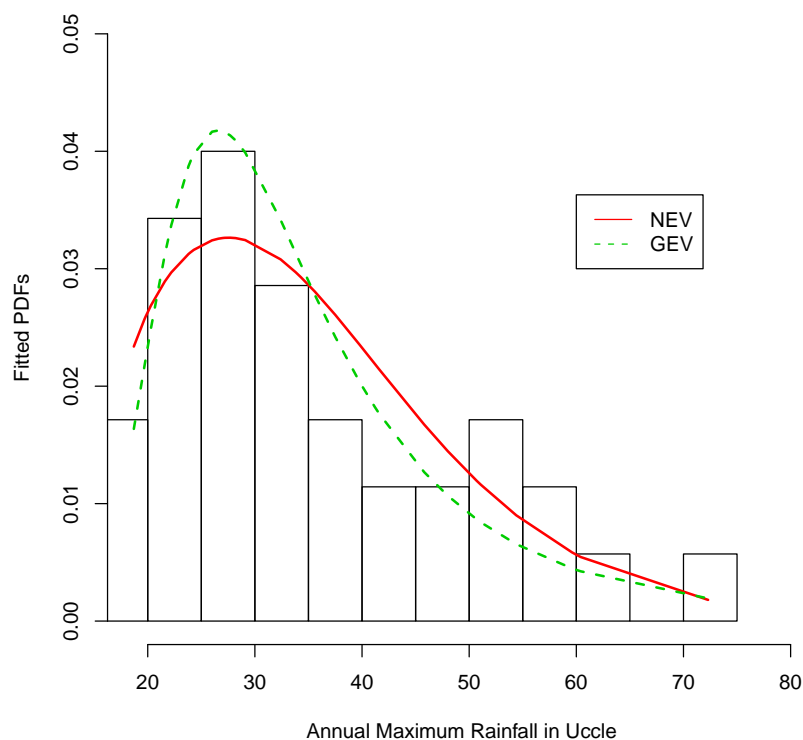
Figure 7.7: Density plots for the fits of the NEV distribution and the GEV distribution for annual maximum rainfall from Uccle, Belgium.

To sum up, we see that all statistical measures based on Log-likelihood, AIC, BIC, CAIC, AICc, P-P and Q-Q plots show that the new distribution of extreme values is a better model than the GEV distribution for the annual maximum rainfall data especially in the tails. This confirms our assumption that the shape parameter a controls the tails and distribution modes as well. Also, the NEV distribution has closed form for density, distribution and quantile functions.

7.6   Summary

The work in this chapter can be summarised into the following main points:

- pdf of the NEV distribution is flexible. That means it can take a variety of shapes: monotonically decreasing, monotonically increasing, unimodal; with some almost-symmetric cases when $\xi > 0$, and bimodal, depending on the shape parameters a, ξ.

- Bathtub shaped hazard rates are the most realistic ones in terms of practical applications. The NEV distribution's hrf has a variety of shapes: monotonically decreasing, monotonically increasing, and unimodal.

- Based on the annual maximum rainfall data in Uccle, Belgium, the NEV distribution and the GEV distribution are compared. The analysis shows that NEV is better model than the GEV distribution for the annual maximum rainfall data especially in the tails.

Part IV

Conclusions and Future Work

Chapter 8

Conclusions and Future Work

8.1 Conclusion of Thesis

This study has introduced four modifications for GEV and GP distributions.

In chapter 4, we studied some mathematical properties of a generalisation of the GEV distribution; the so-called Kumaraswamy GEV (KumGEV) distribution, which can be quite flexible in analyzing continuous data in some areas of engineering including: flood frequency analysis, network engineering, nuclear engineering, offshore engineering, risk-based engineering, space engineering, software reliability engineering, structural engineering and wind engineering. Its moments and characteristic function are given. Explicit expressions are derived for the mean deviations and two well known entropies, the Rényi and Shannon entropies. We examine moments of order statistics and asymptotic distributions of extreme order statistics.

The proposed model has been applied to the annual maximum rainfall data in Uccle, Belgium. Using the maximum likelihood estimation method to fit GEV and KumGEV distributions, it seems that KumGEV distribution is more accurate than GEV. These results are confirmed by P-P plots, Q-Q plots, and density plots, which indicate that KumGEV distribution gives a better fit for the annual maximum rainfall data, especially in the upper tail. Finally, the bivariate KumGEV models are presented.

The generalised Pareto (GP) distribution is the most popular model for extreme values; especially for floods and value at risk analysis. Recently, Papastathopoulos and Tawn have proposed some generalisations of the GP distribution for improved modeling.

This study pointed that Papastathopoulos and Tawn's generalizations are in fact not new, and then we went on to propose a tractable generalisation of the GP distribution depending on the Kumaraswamy distribution, referred to as the KumGP distribution. For the latter generalisation, we provide a comprehensive treatment of mathematical properties and the flexibility of this model is shown. We suggest more applications for the KumGP distribution for datasets that have increasing, decreasing and unimodel shapes. In addition, we found that the hazard rate function of the KumGP distribution can exhibit a bathtub shape that can not be reached by the original distribution.

Dataset consisting 154 exceedances of the threshold level $65m^3s^{-1}$ by the River Nidd at Hunsingore Weir from 1934 to 1969 is tested to illustrate the flexibility and to show its ability to fit that datasets. In this case, we compare the proposal model with GP, EP1, EP2 and EP3. The three last models were introduced by Papastathopoulos and Tawn (2012). These models were examined by the log-likelihood and information criteria (AIC and AICc). They suggested that the KumGP distribution provides a significant improvement over the others, even when we used other measures such as Cramér-von Mises criterion and Anderson-Darling tests. It seemed to be that the KumGP gives better results. Finally, P-P plots, Q-Q plots, and density plots confirmed these results.

In chapter 6, we investigated the model introduced by Chen, Bunce and Jiang. The formulas given for the distribution do not form a valid probability distribution. We corrected their formulas to form a valid probability distribution. For this valid distribution, we provided a comprehensive treatment of mathematical properties, estimated parameters by the method of maximum likelihood, and provided the observed information matrix. The flexibility of the distribution is illustrated using a real dataset.

In chapter 7, we used the corrected formula for the Chen distribution and applied it to the GEV distribution. The resulting model has a shape parameter a that controls the thickness of the tail. Its statistical properties were discussed. Also, we found that the NEV distribution is the exponential transformation of the GEV distribution when $\mu = 0$. The pdf of the NEV distribution can be monotonically decreasing, bimodal, unimodel with some almost-symmetric cases when $\xi > 0$; while, the hazard rate function can take monotonically decreasing, monotonically increasing, bathtub, unimodel shapes. To illustrate the flexibility of this model over the GEV distribution, we applied these two models to the maximum rainfall data in Uccle, Belgium. The analysis showed that NEV is a better model than the GEV distribution for the annual maximum rainfall data especially in the tails.

8.2   Discussion of Future Research

We now recommend some outlines for future work to extend these results. The future recommendations are:

1. Many situations in structural engineering require consideration of the extremes of more than one variable (eg floods, wind). So, it is not enough to study the univariate extreme behaviour. We need to study the bivariate and multivariate extreme value behaviours. The KumGEV, and the KumGP distributions can be extended to the bivariate and multivariate cases.

2. In this study, we used the maximum likelihood estimation method to estimate parameters for each model. For this method, the redundancy problem will appear, especially for models with many parameters. For such situations, we suggest using other methods to estimate the parameters, such as the probability weighted moments (PWM) method, and minimum distance estimation (MDM) method.

3. In chapter 5, we mention that the results reported must be treated conservatively because of the sample size. For $n = 154$, some of the biases and mean squared

errors reported in Figures (5.6) and (5.7) appear large. Furthermore, asymptotic normality does not appear to have been reached. Better estimation methods (for example, bias-corrected estimation methods or bootstrapping based methods) will be needed to draw more sensible results.

4. The Pareto distribution has been widely used in the insurance industry to model payment data, due to the fact that the payment data is highly positively skewed. Many researchers used the GP distribution, especially for data on large losses. The main reason for using the Pareto distribution to model the larger losses data is the shape of the Pareto distribution, having a long and thick upper tail. Despite advantages of these distributions to model highly skewed data, they fail to cover small losses. Because of the flexibility of the KumGP distribution, we suggest applying this model to insurance claims with small losses.

5. The existing software for the GEV and GP distribution (say, to compute pdf, cdf, quantile function, moments, MLEs, random numbers, etc) can be easily adapted for the KumGEV and KumGP distributions. Clearly, the GEV and GP distributions are special cases of the KumGEV and KumGP distributions, respectively, for $a = b = 1$.

Bibliography

K. Abbas, S.A. Khan, D.M. Khan, and A. Ali. Modeling the distribution of annual maximum rainfall in pakistan. *European Journal of Scientific Research*, 79(3): 418–429, 2012.

S. Adeyemi and T. Adebanji. The exponentiated generalized pareto distribution. *Ife Journal of Science*, 6(2):127–133, 2005.

WM Afify. On estimation of the exponentiated pareto distribution under different sample schemes. *Statistical Methodology*, 7(2):77–83, 2010.

H. Akaike. Information theory and an extension of the maximum likelihood principle. In *International Symposium on Information Theory, 2 nd, Tsahkadsor, Armenian SSR*, pages 267–281, 1973.

H. Akaike. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723, 1974.

A. Akinsete, F. Famoye, and C. Lee. The beta-pareto distribution. *Statistics*, 42(6): 547–563, 2008.

T.W. Anderson and D.A. Darling. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *The Annals of Mathematical Statistics*, 23 (2):193–212, 1952.

A.A. Balkema and L. De Haan. Residual life time at great age. *The Annals of Probability*, 2(5):792–804, 1974.

HM Barakat and YH Abdelkader. Computing the moments of order statistics from nonidentical random variables. *Statistical Methods & Applications*, 13(1):15–26, 2004.

J. A. Battjes. Probabilistic aspects of ocean waves. Technical Report 77-2, Laboratory of fluid mechanics, Department of civil engineering, Delft University of Technology, 1978.

M. Bee, R. Benedetti, and G. Espa. A note on maximum likelihood estimation of a pareto mixture. *Department of Economics, University of Trento, Italia Department of Economics Working Papers*, 2009.

J. Beirlant, J.L. Teugels, and P. Vynckier. *Practical analysis of extreme values.* Leuven University Press, 1996.

G. Blom. *Statistical estimates and transformed beta-variables.* PhD thesis, Stockholm, 1958.

C.E. Bonferroni. *Elementi di statistica generale.* Universitacommerciale Bocconi, 1930.

L.E. Borgman. Risk criteria. *Journal of the Waterways and Harbors Division, ASCE*, 89(3):1–35, 1963.

L.E. Borgman. Maximum wave height probabilities for a random number of random intensity storms. Technical report, University of California. Hydraulic Engineering Laboratory, 1970.

L.E. Borgman. Probabilities for highest wave in hurricane. *Journal of the Waterways, Harbors and Coastal Engineering Division*, 99(2):185–207, 1973.

H. Bozdogan. Model selection and akaike's information criterion (aic): The general theory and its analytical extensions. *Psychometrika*, 52(3):345–370, 1987.

C.L. Bretschneider. Wave variability and wave spectra for wind-generated gravity waves, 1959.

P.J. Bryant. Cyclic gravity waves in deep water. *The Journal of the Australian Mathematical Society. Series B. Applied Mathematics*, 25(01):2–15, 1983.

TA Buishand. Extreme rainfall estimation by combining data from several sites. *Hydrological Sciences Journal*, 36(4):345–365, 1991.

M. Cannarozzo, F. D'asaro, and V. Ferro. Regional rainfall and flood frequency analysis for sicily using the two component extreme value distribution. *Hydrological sciences journal*, 40(1):19–42, 1995.

C.C. Caprani, E.J. OBrien, and G.J. McLachlan. Characteristic traffic load effects from a mixture of loading events on short to medium span bridges. *Structural Safety*, 30(5):394–404, 2008.

E. Castillo. Extremes in engineering applications. *Extreme Value Theory and Applications, eds. J. Galambos, J. Lechner, E. Simiu, N. Macri (Dordecht: Kluwer Academic Publishers, 1994)*, pages 15–42, 1994.

E. Castillo and J.M. Sarabia. Engineering analysis of extreme value data: selection of models. *Journal of waterway, port, coastal, and ocean engineering*, 118(2):129–146, 1992.

E. Castillo and J.M. Sarabia. Extreme value analysis of wave heights. *Journal of Research of the National Institute of Standards and Technology*, 99(4):445–454, 1994.

E. Castillo, A.S. Hadi, N. Balakrishnan, and J.M. Sarabia. *Extreme value and related models with applications in engineering and science*. Wiley Chichester, 2005.

A. Cavanie, X. Arhan, and X. Ezraty. A statistical relationship between individual heights and periods of storm waves. *Behaviour of off-shore structures*, 1976.

S.K. Chakrabarti and R.P. Cooley. Ocean wave statistics for 1961 north atlantic storm. *Journal of the Waterway Port Coastal and Ocean Division*, 103(4):433–448, 1977.

J.M. Chambers. *Graphical methods for data analysis.* Duxbury Resource Center, 1983.

G. Chen, C. Bunce, and W. Jiang. A new distribution for extreme value analysis. In *Computational Intelligence and Software Engineering (CiSE), 2010 International Conference on*, pages 1–4. IEEE, 2010.

P.S. Chu, X. Zhao, Y. Ruan, and M. Grubbs. Extreme rainfall events in the hawaiian islands. *Journal of Applied Meteorology and Climatology*, 48(3):502–516, 2009.

S. Coles. *An introduction to statistical modeling of extreme values.* Springer, 2001.

S. Coles, J. Heffernan, and J. Tawn. Dependence measures for extreme value analyses. *Extremes*, 2(4):339–365, 1999.

G.M. Cordeiro and M. de Castro. A new family of generalized distributions. *Journal of statistical computation and simulation*, 81(7):883–898, 2011.

G.M. Cordeiro, E.M.M. Ortega, and S. Nadarajah. The kumaraswamy weibull distribution with application to failure data. *Journal of the Franklin Institute*, 347 (8):1399–1429, 2010.

G.M. Cordeiro, S. Nadarajah, and E.M.M. Ortega. The kumaraswamy gumbel distribution. *Statistical Methods & Applications*, pages 1–30, 2012.

Arnold Court. Wind extremes as design factors. *Journal of the Franklin Institute*, 256(1):39 – 56, 1953. ISSN 0016-0032. URL `http://www.sciencedirect.com/science/article/pii/0016003253909004`.

D.R. Cox and D.V. Hinkley. *Theoretical statistics.* Chapman & Hall/CRC, 1979.

A. Crisci, B. Gozzini, F. Meneguzzo, S. Pagliara, and G. Maracchi. Extreme rainfall in a changing climate: regional analysis and hydrological implications in tuscany. *Hydrological Processes*, 16(6):1261–1274, 2002.

A.G. Davenport. *The relationship of wind structure to wind loading.* National Physical Laboratory, 1963.

A.G Davenport. The dependence of wind loads on meteorological parameters. In *Proc. Int. Res. Seminar, Wind Effects on Buildings and Structures, Ottawa, Univ. of Toronto Press*, pages 19–82, 1967.

A.G Davenport. Structural safety and reliability under wind action. In *International Conference on Structural Safety and Reliability, Smithsonian Institution, Museum of History and Technology, Washington, DC, April 9, 10 and 11, 1969*, page 131. Pergamon Press, 1972.

A.G Davenport. Wind structure and wind climate. *Safety of structures under dynamic loading*, 1:209–256, 1978.

J. del Castillo and J. Daoudi. Estimation of the generalized pareto distribution. *Statistics & Probability Letters*, 79(5):684–688, 2009.

L. Draper. Derivation of a design wave from instrumental records of sea waves. In *ICE Proceedings*, pages 291–304. Ice Virtual Library, 1963.

DJ Dupuis and CA Field. Robust estimation of extremes. *Canadian Journal of Statistics*, 26(2):199–215, 1998.

M.D. Earle, C.C. Ebbesmeyer, and D.J. Evans. Height-period joint probabilities in hurricane camille. *Journal of the Waterways, Harbors and Coastal Engineering Division*, 100(3):257–264, 1974.

B. Efron and D.V. Hinkley. Assessing the accuracy of the maximum likelihood estimator: Observed versus expected fisher information. *Biometrika*, 65(3):457–483, 1978.

N. El-Jabi, F. Ashkar, and S. Hebabi. Regionalization of floods in new brunswick (canada). *Stochastic Hydrology and Hydraulics*, 12(1):65–82, 1998.

P. Embrechts. Extreme value theory in finance and insurance. *Manuscript, Department of Mathematics, ETH, Swiss Federal Technical University*, 1999.

P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events: for insurance and finance.* Springer, Berlin, 1997.

N. Eugene, C. Lee, and F. Famoye. Beta-normal distribution and its applications. *Communications in statistics-Theory and methods*, 31(4):497–512, 2002.

S. Feng, S. Nadarajah, and Q. Hu. Modeling annual extreme precipitation in china using the generalized extreme value distribution. *Journal of the Meteorological Society in Japan*, 85(5):599–613, 2007.

C.A.T. Ferro and J. Segers. Inference for clusters of extreme values. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):545–556, 2003.

D.F. Findley. On the unbiasedness property of aic for exact or approximating linear stochastic time series models. *Journal of Time Series Analysis*, 6(4):229–252, 1985.

R.A. Fisher and L.H.C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, pages 180–190. Cambridge Univ Press, 1928.

SG Fletcher and K. Ponnambalam. Estimation of reservoir yield and storage distribution using moments analysis. *Journal of Hydrology*, 182(1):259–275, 1996.

M. Fréchet. Sur la loi de probabilité de lécart maximum. *Ann. de la Soc. polonaise de Math*, 6:93–116, 1927.

W.E. Fuller. Flood flows. *Transactions of the American Society of Civil Engineers*, 77(1):564–617, 1914.

J. Galambos. *The asymptotic theory of extreme order statistics.* Krieger Pub Co, 1987.

A. Ganji, K. Ponnambalam, D. Khalili, and M. Karamouz. Grain yield reliability analysis with crop water demand uncertainty. *Stochastic Environmental Research and Risk Assessment*, 20(4):259–277, 2006.

J.L. Gastwirth. A general definition of the lorenz curve. *Econometrica: Journal of the Econometric Society*, pages 1037–1039, 1971.

B. Gnedenko. Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, 44, 1943.

RC Goodknight and TL Russell. Investigation of the statistics of wave heights. *ASCE Journal of the Waterways and Harbors Division*, 89:29–52, 1963.

I.S. Gradshteyn and I.M. Ryzhik. Table of integrals, series and products. *New York: Academic Press,— c1994, 5th ed. completely reset, edited by Jeffrey, Alan*, 1, 1994.

M. Grigoriu. Estimates of extreme winds from short records. *Journal of Structural Engineering*, 110(7):1467–1484, 1984.

E.J. Gumbel and C.K. Mustafi. Some analytical properties of bivariate extremal distributions. *Journal of the American Statistical Association*, 62(318):569–588, 1967.

AR Günbak. Statistical analysis of 21 wave records off the danish white sand coast during one storm in january 1976. *Division of Port and Ocean Engineering*, 1978.

T. Haktanir and H.B. Horlacher. Evaluation of various distributions for flood frequency analysis. *Hydrological sciences journal*, 38(1):15–32, 1993.

L.S. Hanson and R. Vogel. The probability distribution of daily rainfall in the united states. In *ASCEEWRI, World Water & Environmental Resources Congress*, 2008.

A.M Hasofer. Wind-load design based on statistics. In *Inst Civil Engineers Proc, London/UK/*, volume 51, 1972.

A.M Hasofer and K. Sharpe. The analysis of wind gusts(stochastic analysis of wind gusts applied to prediction of long term maximum velocities). *Australian Meteorological Magazine*, 17:198–214, 1969.

J.R.M. Hosking. L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 105–124, 1990.

JRM Hosking. The four-parameter kappa distribution. *IBM Journal of Research and Development*, 38(3):251–258, 1994.

J.C. Houghton. *Robust estimation of the frequency of extreme events in a flood frequency context*. PhD thesis, Harvard University, 1977.

O.G. Houmb and T. Overvik. *On the statistical properties of 115 wave records from the Norwegian continental shelf*. Division of Port and Ocean Engineering, the University of Trondheim, the Norwegian Institute of Technology, 1977.

J. Hoybye and L. Iritz. Analysis of extreme hydrological events in a monsoon climate catchment: The hongru river, china. *Hydrological Sciences Journal/Journal des Sciences Hydrologiques*, 42(3):343–355, 1997.

C.M. Hurvich and C.L. Tsai. Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307, 1989.

AF Jenkinson. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81:158–171, 1955.

N.L. Johnson. *Breakthroughs in Statistics: Foundations and basic theory*, volume 1. Springer, 1992.

MC Jones. Families of distributions arising from distributions of order statistics. *Test*, 13(1):1–43, 2004.

MC Jones. Kumaraswamys distribution: A beta-type distribution with some tractability advantages. *Statistical Methodology*, 6(1):70–81, 2009.

M.V. Joshi. Modeling the severity of random events with the sas/ets® severity procedure. *SAS Global Forum 2010*, 2010.

M.D.A. Karim and J.U. Chowdhury. A comparison of four distributions used in flood frequency analysis in bangladesh. *Hydrological sciences journal*, 40(1):55–66, 1995.

R.W. Katz, M.B. Parlange, and P. Naveau. Statistics of extremes in hydrology. *Advances in water resources*, 25(8):1287–1304, 2002.

V.V. Kharin and F.W. Zwiers. Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere-ocean gcm. *Journal of Climate*, 13 (21):3760–3788, 2000.

A.A. Kilbas, H.M. Srivastava, and J.J. Trujillo. *Theory and Applications of Fractional Differential Equations*, volume 204. North-Holland Mathematics Studies, Elsevier, Amsterdam, 2006.

D. Koutsoyiannis. Statistics of extremes and estimation of extreme rainfall: Ii. empirical investigation of long rainfall records. ii. empirical investigatio. *Hydrological Sciences Journal*, 49(4), 2004.

D. Koutsoyiannis and G. Baloutsos. Analysis of a long record of annual maximum rainfall in athens, greece, and design rainfall inferences. *Natural Hazards*, 22(1): 29–48, 2000.

D. Koutsoyiannis and T. Xanthopoulos. On the parametric approach to unit hydrograph identification. *Water resources management*, 3(2):107–128, 1989.

P. Kumaraswamy. A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, 46(1):79–88, 1980.

MR Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and related properties of random sequences and processes*, volume 11. Springer Verlag, 1983.

Y. Li, W. Cai, and EP Campbell. Statistical modeling of extreme rainfall in southwest western australia. *Journal of Climate*, 18(6):852–863, 2005.

MS Longuet-Higgins. On the statistical distribution of the heights of sea waves. *Journal of Marine Research*, 9:245–266, 1952.

MS Longuet-Higgins. On the joint distribution of the periods and amplitudes of sea waves. *Journal of Geophysical Research*, 80(18):2688–2694, 1975.

M.O. Lorenz. Methods of measuring the concentration of wealth. *Publications of the American Statistical Association*, 9(70):209–219, 1905.

H. Madsen and D. Rosbjerg. A regional bayesian method for estimation of extreme streamflow droughts. *Statistical and Bayesian Methods in Hydrological Sciences*, 20:327–339, 1998.

E. Mahmoudi. The beta generalized pareto distribution with application to lifetime data. *Mathematics and Computers in Simulation*, 81(11):2414–2430, 2011.

E.S. Martins, J.R. Stedinger, et al. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, 36(3):737–744, 2000.

A.M. Mathai and R.K. Saxena. *The H-function with applications in statistics and other disciplines.* Wiley New York, 1978.

P.W. Mielke Jr. Another family of distributions for describing and analyzing precipitation data. *Journal of Applied Meteorology*, 12:275–280, 1973.

E. Mistéth. Determination of the critical loads considering the anticipated durability of structures. *Acta Technica Academiae Scientiarum Hungaricae*, 74(1-2):21–38, 1973.

P. Mitnik. New properties of the kumaraswamy distribution. *Communications in Statistics-Theory and Methods, Forthcoming*, 2008.

F. Moses. Reliability of structural systems. *Journal of the Structural Division*, 100 (9):1813–1820, 1974.

S. Nadarajah. Extremes of daily rainfall in west central florida. *Climatic change*, 69 (2):325–342, 2005.

S. Nadarajah. On the distribution of kumaraswamy. *Journal of Hydrology*, 348: 568–569, 2008.

S. Nadarajah and T.K. Pogány. On the characteristic functions for extreme value distributions. *Extremes*, pages 1–12, 2012.

S. Nadarajah and JT Shiau. Analysis of extreme flood events for the pachang river, taiwan. *Water resources management*, 19(4):363–374, 2005.

S. Nadarajah, G.M. Cordeiro, and E.M.M. Ortega. General results for the kumaraswamy-g distribution. *Journal of Statistical Computation and Simulation*, 82(7):951–979, 2012.

NERC. *Flood Studies Report in Five Volumes*. NERC, 1975.

M. Onorato, AR Osborne, and M. Serio. Extreme wave events in directional, random oceanic sea states. *Physics of Fluids*, 14(4):L25–L28, 2002.

G. Panthou, T. Vischel, T. Lebel, J. Blanchet, G. Quantin, and A. Ali. Extreme rainfall in west africa: A regional modeling. *Water Resources Research*, 48(8): W08501, 2012.

I. Papastathopoulos and J.A. Tawn. Extended generalised pareto models for tail estimation. *Journal of Statistical Planning and Inference*, 2012.

B.P. Parida. Modelling of indian summer monsoon rainfall using a four-parameter kappa distribution. *International Journal of Climatology*, 19(12):1389–1398, 1999.

J.S. Park and H.S. Jung. Modelling korean extreme rainfall using a kappa distribution and maximum likelihood estimate. *Theoretical and applied climatology*, 72(1):55–64, 2002.

J.S. Park, H.S. Jung, R.S. Kim, and J.H. Oh. Modelling summer extreme rainfall over the korean peninsula using wakeby distribution. *International journal of climatology*, 21(11):1371–1384, 2001.

B. Pfaff, A. McNeil, and M.B. Pfaff. Package evir, 2010.

J. Pickands III. Statistical inference using extreme order statistics. *the Annals of Statistics*, pages 119–131, 1975.

K. Ponnambalam, A. Seifi, and J. Vlach. Probabilistic design of systems with general distributions of parameters. *International journal of circuit theory and applications*, 29(6):527–536, 2001.

R.R. Putz. Statistical distribution for ocean waves. *Trans. AGU*, 33(5):685–692, 1952.

CORE R Development. Team. 2008. r: A language and environment for statistical computing. r foundation for statistical computing, vienna, austria. *ISBN*, pages 3–900051, 2011.

C.R Rao. Linear statistical inference and its applications. *Wiley series in probability and mathematical statistics*, 1973.

A. Rényi. On measures of entropy and information. In *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, pages 547–561, 1961.

S.I. Resnick. *Extreme values, regular variation, and point processes*. Springer, 2007.

M. Ribatet. Pot: Generalized pareto distribution and peaks over threshold. *R package verions*, pages 1–1, 2009.

M.M. Ristić and N. Balakrishnan. The gamma-exponentiated exponential distribution. *Journal of Statistical Computation and Simulation*, 82(8):1191–1206, 2012.

JP Royston. Algorithm as 181: The w test for normality. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 31(2):176–180, 1982.

G. Schwarz. Estimating the dimension of a model. *The annals of statistics*, 6(2): 461–464, 1978.

A. Seifi, K. Ponnambalam, and J. Vlach. Maximization of manufacturing yield of systems with arbitrary distributions of component values. *Annals of Operations research*, 99(1):373–383, 2000.

F. Sellars. Maximum heights of ocean waves. *Journal of Geophysical Research*, 80 (3):398–404, 1975.

C.E. Shannon. Prediction and entropy of printed english. *Bell System Technical Journal*, 30(1):50–64, 1951.

AI Shawky and H.H. Abu-Zinadah. Characterizations of the exponentiated pareto distribution based on record values. *Applied Mathematical Sciences*, 2(26):1283–1290, 2008.

AI Shawky and H.H. Abu-Zinadah. Exponentiated pareto distribution: different method of estimations. *Int. J. Contemp. Math. Sci*, 4(14):677–693, 2009.

R.K. Shukla, M. Trivedi, and M. Kumar. On the proficient use of gev distribution: a case study of subtropical monsoon region in india. *arXiv preprint arXiv:1203.0642*, 2012.

R.B. Silva, W. Barreto-Souza, and G.M. Cordeiro. A new distribution with decreasing, increasing and upside-down bathtub failure rate. *Computational Statistics & Data Analysis*, 54(4):935–944, 2010.

H. Singh and N. Misra. On redundancy allocations in systems. *Journal of applied probability*, pages 1004–1014, 1994.

R. Sneyers. L'intensité maximale des précipitations en belgique. *Publications. Serie B-Institut Royal Meteorologique de Belgique*, 1977.

H.M. Srivastava, K.C. Gupta, and S.P Goyal. *The H-functions of one and two variables, with applications*, volume 36. South Asian Publishers New Delhi and Madras, 1982.

A. G. Stephenson. evd: Extreme value distributions. *R News*, 2(2):0, June 2002. URL `http://CRAN.R-project.org/doc/Rnews/`.

N. Sugiura. Further analysts of the data by akaike's information criterion and the finite corrections. *Communications in Statistics-Theory and Methods*, 7(1):13–26, 1978.

V. Sundar and K. Subbiah. Application of double bounded probability density function for analysis of ocean waves. *Ocean engineering*, 16(2):193–200, 1989.

K. Tolikas. The rare event risk in african emerging stock markets. *Managerial Finance*, 37(3):275–294, 2011.

K. Tolikas and G.D. Gettinby. Modelling the distribution of the extreme share returns in singapore. *Journal of Empirical Finance*, 16(2):254–263, 2009.

R. Von Mises. La distribution de la plus grande de n valeurs. *Rev. math. Union interbalcanique*, 1(1), 1936.

R. Von Mises. *Selected papers of Richard von Mises*, volume 2. American mathematical society, 1964.

W. Weibull et al. A statistical distribution function of wide applicability. *Journal of applied mechanics*, 18(3):293–297, 1951.

E.M. Wright. The asymptotic expansion of the generalized hypergeometric function. *Journal of the London Mathematical Society*, 1(4):286–293, 1935.

S. Yue, T. Ouarda, B. Bobée, P. Legendre, and P. Bruneau. The gumbel mixed model for flood frequency analysis. *Journal of Hydrology*, 226(1):88–100, 1999.

K. Zografos and N. Balakrishnan. On families of beta-and generalized gamma-generated distributions and associated inference. *Statistical Methodology*, 6(4): 344–362, 2009.

F.W. Zwiers, X. Zhang, and Y. Feng. Anthropogenic influence on long return period daily temperature extremes at regional scales. *Journal of Climate*, 24(3):881–892, 2011.

Appendix A

Appendix to Chapter 4

The calculations in chapter 4 require the following lemmas.

Lemma 1 *Let*

$$I(n, \alpha, \beta) = \int_{-\infty}^{\infty} x^n u^{\alpha} \exp(-\beta u) dx,$$

where $u = \{1 + \xi(x - \mu)/\sigma\}^{-1/\xi}$. Then

$$I(n, \alpha, \beta) = \sigma \sum_{k=0}^{n} \binom{n}{k} \left(\mu - \frac{\sigma}{\xi}\right)^{n-k} \left(\frac{\sigma}{\xi}\right)^{k} \beta^{\xi + k\xi - \alpha} \Gamma\left(\alpha - \xi - k\xi\right),$$

Proof: *We can write*

$$\begin{aligned}
I(n, \alpha, \beta) &= \int_{-\infty}^{\infty} x^n u^{\alpha} \exp(-\beta u) dx \\
&= \sigma \int_{0}^{\infty} \left[\frac{\sigma}{\xi}\left(u^{-\xi} - 1\right) + \mu\right]^{n} u^{\alpha - \xi - 1} \exp(-\beta u) du \\
&= \sigma \sum_{k=0}^{n} \binom{n}{k} \left(\mu - \frac{\sigma}{\xi}\right)^{n-k} \left(\frac{\sigma}{\xi}\right)^{k} \int_{0}^{\infty} u^{\alpha - \xi - 1 - k\xi} \exp(-\beta u) du.
\end{aligned}$$

So, the result follows from the definition of gamma function. □

Lemma 2 *Let*

$$J(y, \alpha) = \int_{y}^{\infty} x u^{1+\xi} \exp(-\alpha u) dx,$$

where $u = \{1 + \xi(x - \mu)/\sigma\}^{-1/\xi}$. Then

$$\begin{aligned}
J(y, \alpha) = \frac{\sigma^2}{\xi}(j+1)^{\xi-1}\alpha^{\xi-1}\gamma\left(1 - \xi, (j+1)\alpha z\right) + \sigma\left(\mu - \frac{\sigma}{\xi}\right)\frac{1}{(j+1)\alpha} \\
\{1 - \exp\left[-(j+1)\alpha z\right]\},
\end{aligned}$$

where $z = \{1 + \xi(y - \mu)/\sigma\}^{-1/\xi}$ and $\gamma(a, x)$ denotes the incomplete gamma function defined by

$$\gamma(a, x) = \int_0^x t^{a-1} \exp(-t)dt.$$

Proof: *We can write*

$$J(y, \alpha) = \int_y^\infty xu^{1+\xi} \exp(-\alpha u)dx$$

$$= \sigma \int_0^z \left[\frac{\sigma}{\xi} \left(u^{-\xi} - 1\right) + \mu\right] \exp(-\alpha u)du.$$

So, the result follows from the definition of incomplete gamma function. $\square$

Lemma 3 *We have*

$$\int_{-\infty}^\infty \log u \ u^{1+\xi} \ \exp(-\beta u)dx = -\frac{\sigma}{\beta} \left[\log \beta + C\right],$$

where $u = \{1 + \xi(x - \mu)/\sigma\}^{-1/\xi}$ and C is Euler's constant.

Proof: *We can write*

$$\int_{-\infty}^\infty \log u \ u^{1+\xi} \ \exp(-\beta u)dx = \sigma \int_0^\infty \log u \ \exp(-\beta u)du.$$

The second integral can be calculated using equation (4.331.1) in Gradshteyn and Ryzhik (2000) to yield the required result. $\square$

Information Matrix

It is useful to have explicit expressions for the elements of **J**. *These are now given:*

$$J_{11} = -\frac{n}{a^2} + (1 - b) \sum_{i=1}^n \frac{u_i^2 \exp(au_i)}{\left(\exp(au_i) - 1\right)^2},$$

$$J_{12} = \sum_{i=1}^{n} \frac{u_i}{\exp(au_i) - 1},$$

$$J_{13} = -\frac{1}{\sigma} \sum_{i=1}^{n} u_i^{1+\xi} + \frac{b-1}{\sigma} \sum_{i=1}^{n} u_i^{1+\xi} \frac{(1-au_i)\exp(au_i) - 1}{\left(\exp(au_i) - 1\right)^2},$$

$$J_{14} = -\frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) u_i^{1+\xi} + \frac{b-1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) u_i^{1+\xi} \frac{(1-au_i)\exp(au_i) - 1}{\left(\exp(au_i) - 1\right)^2},$$

$$J_{15} = \frac{1}{\xi} \sum_{i=1}^{n} u_i \left(\log u_i + u_i^{\xi} \frac{x_i - \mu}{\sigma}\right) - \frac{b-1}{\xi} \sum_{i=1}^{n} u_i \left(\log u_i + u_i^{\xi} \frac{x_i - \mu}{\sigma}\right)$$
$$\frac{(1-au_i)\exp(au_i) - 1}{\left(\exp(au_i) - 1\right)^2},$$

$$J_{22} = -\frac{n}{b^2},$$

$$J_{23} = \frac{a}{\sigma} \sum_{i=1}^{n} \frac{u_i^{1+\xi}}{\exp(au_i) - 1},$$

$$J_{24} = \frac{a}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) \frac{u_i^{1+\xi}}{\exp(au_i) - 1},$$

$$J_{25} = -\frac{a}{\xi} \sum_{i=1}^{n} \left(\log u_i + u_i^{\xi} \frac{x_i - \mu}{\sigma}\right) \frac{u_i}{\exp(au_i) - 1},$$

$$J_{33} = \frac{a(b-1)}{\sigma^2} \sum_{i=1}^{n} u_i^{1+2\xi} \frac{\exp(au_i)\left(1+\xi-au_i\right) - (1+\xi)}{\left(\exp(au_i) - 1\right)^2}$$
$$- \frac{a(1+\xi)}{\sigma^2} \sum_{i=1}^{n} u_i^{1+2\xi} + \frac{\xi(\xi+1)}{\sigma^2} \sum_{i=1}^{n} u_i^{2\xi}$$

$$J_{34} = \frac{a}{\sigma^2} \sum_{i=1}^{n} u_i^{1+\xi} - \frac{a(1+\xi)}{\sigma^3} \sum_{i=1}^{n} u_i^{2\xi+1} (x_i - \mu) - \frac{1+\xi}{\sigma^2} \sum_{i=1}^{n} u_i^{\xi}$$
$$+ \frac{\xi(1+\xi)}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) u_i^{2\xi} - \frac{a(b-1)}{\sigma^2} \sum_{i=1}^{n} \frac{u_i^{1+\xi}}{\exp(au_i) - 1}$$
$$+ \frac{a(b-1)}{\sigma^3} \sum_{i=1}^{n} (x_i - \mu) u_i^{1+2\xi} \frac{\exp(au_i)\left(1+\xi-au_i\right) - (1+\xi)}{\left(\exp(au_i) - 1\right)^2}$$

$$J_{35} = -\frac{a}{\sigma} \sum_{i=1}^{n} u_i^{1+\xi} \left\{\log u_i - \frac{1+\xi}{\xi}\left(\log u_i + u_i^{\xi} \frac{x_i - \mu}{\sigma}\right)\right\}$$
$$+ \frac{a(b-1)}{\sigma} \sum_{i=1}^{n} \frac{u_i^{1+\xi} \log u_i}{\exp(au_i) - 1} + \frac{1}{\sigma} \sum_{i=1}^{n} u_i^{\xi} - \frac{1+\xi}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) u_i^{2\xi}$$

$$
+\frac{a^2(b-1)}{\xi\sigma}\sum_{i=1}^{n}\frac{u_i^{2+\xi}}{\left(\exp\left(au_i\right)-1\right)^2}\left(\log u_i+u_i^{\xi}\frac{x_i-\mu}{\sigma}\right)\exp\left(au_i\right)
$$

$$
-\frac{(1+\xi)a(b-1)}{\xi\sigma}\sum_{i=1}^{n}\frac{u_i^{1+\xi}}{\exp\left(au_i\right)-1}\left(\log u_i+u_i^{\xi}\frac{x_i-\mu}{\sigma}\right)
$$

$$
J_{44}=\frac{n}{\sigma^2}-\frac{2a(b-1)}{\sigma^3}\sum_{i=1}^{n}\left(x_i-\mu\right)\frac{u_i^{1+\xi}}{\exp\left(au_i\right)-1}-\frac{2(1+\xi)}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)u_i^{\xi}
$$

$$
+\frac{\xi(1+\xi)}{\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2u_i^{2\xi}+\frac{2a}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)u_i^{1+\xi}-\frac{a(1+\xi)}{\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2u_i^{1+2\xi}
$$

$$
+\frac{a(b-1)}{\sigma^4}\sum_{i=1}^{n}\left(x_i-\mu\right)^2u_i^{1+2\xi}\frac{(1+\xi)\left(\exp\left(au_i\right)-1\right)-au_i\exp\left(au_i\right)}{\left(\exp\left(au_i\right)-1\right)^2},
$$

$$
J_{45}=\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)u_i^{\xi}-\frac{1+\xi}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)^2u_i^{2\xi}-\frac{a}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)u_i^{1+\xi}\log u_i
$$

$$
+\frac{a(1+\xi)}{\xi\sigma^2}\sum_{i=1}^{n}(x_i-\mu)u_i^{1+\xi}\left(\log u_i+u_i^{\xi}\frac{x_i-\mu}{\sigma}\right)+\frac{a(b-1)}{\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)
$$

$$
\frac{u_i^{1+\xi}\log u_i}{\exp\left(au_i\right)-1}+\frac{a^2(b-1)}{\xi\sigma^2}\sum_{i=1}^{n}\frac{(x_i-\mu)\,u_i^{2+\xi}\exp\left(au_i\right)}{\left(\exp\left(au_i\right)-1\right)^2}\left(\log u_i+u_i^{\xi}\frac{x_i-\mu}{\sigma}\right)
$$

$$
-\frac{(1+\xi)a(b-1)}{\xi\sigma^2}\sum_{i=1}^{n}\frac{(x_i-\mu)\,u_i^{1+\xi}}{\exp\left(au_i\right)-1}\left(\log u_i+u_i^{\xi}\frac{x_i-\mu}{\sigma}\right)
$$

and

$$
J_{55}=-\frac{2}{\xi^2}\sum_{i=1}^{n}\left(\log u_i+u_i^{\xi}\frac{x_i-\mu}{\sigma}\right)+\frac{1+\xi}{\xi\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2u_i^{2\xi}
$$

$$
-\frac{a}{\xi^2}\sum_{i=1}^{n}u_i\left(\log u_i+u_i^{\xi}\frac{x_i-\mu}{\sigma}\right)-\frac{a}{\xi^2}\sum_{i=1}^{n}u_i\left(\log u_i+u_i^{\xi}\frac{x_i-\mu}{\sigma}\right)^2
$$

$$
-\frac{a}{\xi}\sum_{i=1}^{n}u_i\left(\frac{1}{\xi}+u_i^{\xi}\frac{x_i-\mu}{\sigma}\right)\left(\log u_i+u_i^{\xi}\frac{x_i-\mu}{\sigma}\right)+\frac{a}{\xi\sigma}\sum_{i=1}^{n}\left(x_i-\mu\right)u_i^{1+\xi}\log u_i
$$

$$
+\frac{a(b-1)}{\xi^2}\sum_{i=1}^{n}\frac{u_i}{\exp\left(au_i\right)-1}\left(\log u_i+u_i^{\xi}\frac{x_i-\mu}{\sigma}\right)
$$

$$
+\frac{a(b-1)}{\xi^2}\sum_{i=1}^{n}\frac{u_i\left((1-au_i)\exp\left(au_i\right)-1\right)}{\left(\exp\left(au_i\right)-1\right)^2}\left(\log u_i+u_i^{\xi}\frac{x_i-\mu}{\sigma}\right)^2
$$

$$+\frac{a(b-1)}{\xi}\sum_{i=1}^{n}\frac{u_i}{\exp\left(au_i\right)-1}\left(\frac{1}{\xi}+u_i^{\xi}\frac{x_i-\mu}{\sigma}\right)\left(\log u_i+u_i^{\xi}\frac{x_i-\mu}{\sigma}\right)$$

$$-\frac{a(b-1)}{\xi\sigma}\sum_{i=1}^{n}(x_i-\mu)\frac{u_i^{1+\xi}\log u_i}{\exp\left(au_i\right)-1}.$$

Expected Information Matrix

As mentioned in chapter 4, the $(5*5)$ unit expected information matrix $(k_{i,j} = -E[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln f(x_i,\theta)])$ can be written as:

$$\mathbf{k}(\theta)=\begin{pmatrix} k_{a,a}(\theta)\ k_{a,b}(\theta)\ k_{a,\mu}(\theta)\ k_{a,\sigma}(\theta)\ k_{a,\xi}(\theta) \\ k_{b,b}(\theta)\ k_{b,\mu}(\theta)\ k_{b,\sigma}(\theta)\ k_{b,\xi}(\theta) \\ k_{\mu,\mu}(\theta)\ k_{\mu,\sigma}(\theta)\ k_{\mu,\xi}(\theta) \\ k_{\sigma,\sigma}(\theta)\ k_{\sigma,\xi}(\theta) \\ k_{\xi,\xi}(\theta) \end{pmatrix}$$

Elements of the expected information matrix can be defined as:

$$k_{a,a}=-\frac{1}{a^2}-(b-1)\Big(\beta_{0,1,1,0}-\beta_{0,0,2,0}\Big),\quad k_{a,b}=\beta_{0,0,1,0},$$

$$k_{a,\mu}=-\frac{1}{\sigma^2}\beta_{1,1,0,0}+\frac{b-1}{\sigma}\Big(\beta_{1,0,1,0}+a(\beta_{1,1,1,0}-\beta_{1,0,2,0})\Big),$$

$$k_{a,\sigma}=-\frac{1}{\sigma}\beta_{1,1,0,1}-(b-1)\Big(\frac{a}{\sigma^2}\beta_{1,1,1,1}-\frac{1}{\sigma^2}\beta_{1,0,1,1}+a\beta_{1,0,2,1}\Big),$$

$$k_{a,\xi}=-\gamma_{0,1,0,1}-(b-1)\Big(a(\gamma_{0,1,1,1}+\gamma_{0,0,2,1})-\gamma_{0,0,1,1}\Big),$$

$$k_{b,b}=-\frac{1}{b^2},\quad k_{b,\mu}=\frac{a}{\sigma}\beta_{1,0,1,0},\quad k_{b,\sigma}=\frac{a}{\sigma^2}\beta_{1,1,1,1},\quad k_{b,\xi}=a\gamma_{0,0,1,1},$$

$$k_{\mu,\mu}=-\frac{\xi(\xi+1)}{\sigma^2}\beta_{2,0,0,0}-\frac{a(1+\xi)}{\sigma^2}\beta_{2,1,0,0}+\frac{a(b-1)}{\sigma^2}\Big(a\beta_{2,1,1,0}+(1+\xi)\beta_{2,0,1,0}$$
$$-\beta_{2,0,2,0}\Big),$$

$$k_{\mu,\sigma}=\frac{(1+\xi)}{\sigma^2}\Big(\frac{\xi}{\sigma}\beta_{2,0,0,1}-\beta_{1,0,0,0}\Big)-\frac{a}{\sigma^2}\Big(\frac{(\xi+1)}{\sigma}\beta_{2,1,0,1}-\beta_{1,1,0,0}\Big)+a(b+1)$$
$$\Big(\frac{(\xi+1)}{\sigma^3}\beta_{2,0,1,1}+\frac{a}{\sigma^3}\beta_{2,1,1,1}-\frac{1}{\sigma^2}\beta_{1,0,1,0}-a\beta_{2,0,2,1}\Big),$$

$$k_{\mu,\xi}=\frac{1}{\sigma}\delta_{0,1,0,1,0}-\frac{(\xi+1)}{\sigma^2}\beta_{0,2,0,1,0}-\frac{a}{\sigma}\Big(\rho_{1,1,0,0,1,0}+\rho_{2,1,0,0,1,0}-\frac{1}{\sigma}\beta_{1,2,0,1,0}\Big)$$
$$-\frac{a(b-1)}{\sigma}\Big((a-1)\rho_{2,1,1,0,1,0}-\rho_{1,1,1,0,1,0}+a\rho_{2,1,2,0,1,0}+\frac{1}{\sigma}\beta_{1,2,1,1,0}\Big),$$

$$k_{\sigma,\sigma} = \frac{1}{\sigma^2} - \frac{(\xi+1)}{\sigma^3}\Big(2\beta_{0,1,0,1,0} - \frac{\xi}{\sigma}\beta_{0,2,0,2,0}\Big) - \frac{a}{\sigma^3}\Big(\frac{(\xi+1)}{\sigma}\beta_{1,2,0,2,0} - 2\beta_{1,1,0,1,0}$$
$$+ \frac{1}{\sigma}\beta_{2,2,0,1,0}\Big) - \frac{a(b-1)}{\sigma^3}\Big(\frac{a(a-1)}{\sigma}\beta_{2,2,1,2,0} - \frac{a(\xi+1)}{\sigma}\beta_{1,2,1,2,0}$$
$$+ 2\beta_{1,1,1,1,0} + \frac{a^2}{\sigma}\beta_{2,2,2,2,0}\Big),$$

$$k_{\sigma,\xi} = \frac{1}{\sigma^2}\beta_{1,0,0,1} - \frac{(\xi+1)}{\sigma^3}\beta_{2,0,0,2} - \frac{a}{\sigma^3}\Big(\delta_{1,1,0} - \frac{1}{\sigma}\beta_{2,1,0,2}\Big)$$
$$- \frac{a(b-1)}{\sigma^2}\Big(a\delta_{1,1,1} - \delta_{1,0,1} + \frac{1}{\sigma}\beta_{2,0,1,2} + \frac{a}{\sigma}\delta_{1,0,2}\Big),$$

and

$$k_{\xi,\xi} = -\frac{2}{\xi}\rho + \frac{2}{\xi^2}\beta_{1,0,0,1} + \frac{1+\xi}{\xi} - a\Big(\gamma_{0,1,0,2} - \frac{2}{\xi}\gamma_{0,1,0,1} + \frac{1}{\xi\sigma^2}\beta_{2,1,0,2}\Big)$$
$$- a(b-1)\Big(a\gamma_{0,1,1,2} - \gamma_{0,0,1,2} + \frac{2a}{\xi}\gamma_{0,0,1,1} - \frac{1}{\xi\sigma^2}\beta_{2,0,2,2} + a\gamma_{0,0,2,2}\Big).$$

where,

$$u = 1 + \frac{(x-\mu)}{\sigma}, \quad \rho = E(\ln U)$$

$$\beta_{i,j,k,r} = E\left\{(x-\mu)^r U^{\frac{k+j}{\xi}-i}\exp(-akU^{-1/\xi})\Big(1 - \exp(-aU^{-1/\xi})\Big)^{-k}\right\},$$

$$\gamma_{i,j,k,r} = E\left\{U^{\frac{k+j}{\xi}-i}\exp(-akU^{-1/\xi})\Big(1 - \exp(-aU^{-1/\xi})\Big)^{-k}\Big(\frac{\ln U}{\xi^2} - \frac{(x-\mu)}{\xi\sigma U}\Big)^r\right\},$$

and

$$\delta_{i,j,k} = E\left\{(x-\mu)U^{\frac{k+j}{\xi}-i}\exp(-akU^{-1/\xi})\Big(1 - \exp(-aU^{-1/\xi})\Big)^{-k}\Big(\frac{\ln U}{\xi^2} - \frac{(x-\mu)}{\xi\sigma U}\Big)\right\}.$$

Explicit expressions for the remaining elements of **J** *follow by symmetry.*

Two-Types of Kum-GEV Distribution

Here, we give explicit expressions for the partial derivatives needed in Section (4.15)
for the models given by (4.39) and (4.40). In the case of (4.39), they are

$$\frac{\partial G(x,y)}{\partial \theta} = -G(x,y)\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-1},$$

$$\frac{\partial^2 G(x,y)}{\partial \theta^2} = G(x,y)\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-2},$$

$$\frac{\partial G(x,y)}{\partial x} = g_1(x)G_2(y)\exp\left\{-\theta\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-1}\right\}$$
$$- \frac{\theta g_1(x)G_2(y)}{(\log G_1(x))^2}\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-2}\exp\left\{-\theta\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-1}\right\},$$

$$\frac{\partial^2 G\left(x,y\right)}{\partial x \partial \theta} = -g_1(x)G_2(y) \left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-1} \exp\left\{-\theta\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-1}\right\}$$

$$-\frac{g_1(x)G_2(y)}{(\log G_1(x))^2}\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-2} \exp\left\{-\theta\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-1}\right\},$$

$$+\frac{\theta g_1(x)G_2(y)}{(\log G_1(x))^2}\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-3} \exp\left\{-\theta\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-1}\right\},$$

$$\frac{\partial G\left(x,y\right)}{\partial y} = g_2(y)G_1(x) \exp\left\{-\theta\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-1}\right\}$$

$$-\frac{\theta g_1(x)G_2(y)}{(\log G_2(y))^2}\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-2} \exp\left\{-\theta\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-1}\right\},$$

$$\frac{\partial^2 G\left(x,y\right)}{\partial y \partial \theta} = -g_2(y)G_1(x) \left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-1} \exp\left\{-\theta\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-1}\right\}$$

$$-\frac{g_2(y)G_1(x)}{(\log G_2(y))^2}\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-2} \exp\left\{-\theta\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-1}\right\},$$

$$+\frac{\theta g_2(y)G_1(x)}{(\log G_2(y))^2}\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-3} \exp\left\{-\theta\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-1}\right\},$$

$$\frac{\partial^2 G\left(x,y\right)}{\partial x \partial y} = g_1(x)g_2(y) \exp\left\{-\theta\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-1}\right\}$$

$$-\frac{\theta g_1(x)g_2(y)}{(\log G_2(y))^2}\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-2} \exp\left\{-\theta\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-1}\right\}$$

$$-\frac{\theta g_1(x)}{(\log G_1(x))^2}\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-4} \exp\left\{-\theta\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-1}\right\}$$

$$\times \left\{g_2(y)\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^2 + \frac{g_2(y)}{(\log G_2(y))^2}\left[\frac{2}{\log G_1(x)} + \frac{2}{\log G_2(y)} - \theta\right]\right\},$$

and

$$\frac{\partial^3 G(x,y)}{\partial x \partial y \partial \theta} = -g_1(x)g_2(y)\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-1} \exp\left\{-\theta\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-1}\right\}$$

$$- \frac{g_1(x)g_2(y)}{(\log G_2(y))^2}\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-2} \exp\left\{-\theta\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-1}\right\}$$

$$+ \frac{\theta g_1(x)g_2(y)}{(\log G_2(y))^2}\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-3} \exp\left\{-\theta\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-1}\right\}$$

$$- \frac{g_1(x)}{(\log G_1(x))^2}\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-4} \exp\left\{-\theta\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-1}\right\}$$

$$\times \left\{g_2(y)\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^2 + \frac{g_2(y)}{(\log G_2(y))^2}\left[\frac{2}{\log G_1(x)} + \frac{2}{\log G_2(y)} - \theta\right]\right\}$$

$$+ \frac{\theta g_1(x)}{(\log G_1(x))^2}\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-5} \exp\left\{-\theta\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-1}\right\}$$

$$\times \left\{g_2(y)\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^2 + \frac{g_2(y)}{(\log G_2(y))^2}\left[\frac{2}{\log G_1(x)} + \frac{2}{\log G_2(y)} - \theta\right]\right\}$$

$$+ \frac{\theta g_1(x)g_2(y)}{(\log G_1(x)\log G_2(y))^2}\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-4} \exp\left\{-\theta\left[\frac{1}{\log G_1(x)} + \frac{1}{\log G_2(y)}\right]^{-1}\right\}.$$

In the case of (4.40), they are

$$\frac{\partial G(x,y)}{\partial \theta} = G(x,y)\left\{\theta^{-2}\log\left[(\log G_1(x))^\theta + (\log G_2(y))^\theta\right] - \theta^{-1}\left[(\log G_1(x))^\theta + (\log G_2(y))^\theta\right]^{-1}\right.$$

$$\left.\times \left[(\log G_1(x))^\theta \log\log G_1(x) + (\log G_2(y))^\theta \log\log G_2(y)\right]\right\},$$

$$\frac{\partial G(x,y)}{\partial x} = -\frac{g_1(x)}{G_1(x)}(\log G_1(x))^{\theta-1}\left[(\log G_1(x))^\theta + (\log G_2(y))^\theta\right]^{1/\theta-1}G(x,y),$$

$$\frac{\partial^2 G(x,y)}{\partial x \partial \theta} = -\frac{g_1(x)}{G_1(x)}(\log G_1(x))^{\theta-1}\log\log G_1(x)\left[(\log G_1(x))^\theta + (\log G_2(y))^\theta\right]^{1/\theta-1}G(x,y)$$

$$+ \theta^{-2}\frac{g_1(x)}{G_1(x)}(\log G_1(x))^{\theta-1}\left[(\log G_1(x))^\theta + (\log G_2(y))^\theta\right]^{1/\theta-1}G(x,y)$$

$$\times \log\left[(\log G_1(x))^\theta + (\log G_2(y))^\theta\right] - (\theta^{-1} - 1)\frac{g_1(x)}{G_1(x)}(\log G_1(x))^{\theta-1}G(x,y)$$

$$\times \left[(\log G_1(x))^\theta \log\log G_1(x) + (\log G_2(y))^\theta \log\log G_2(y)\right]$$

$$\times \left[(\log G_1(x))^\theta + (\log G_2(y))^\theta\right]^{1/\theta-2} - \frac{g_1(x)}{G_1(x)}(\log G_1(x))^{\theta-1}$$

$$\times \left[(\log G_1(x))^\theta + (\log G_2(y))^\theta\right]^{1/\theta-1}\frac{\partial G(x,y)}{\partial \theta},$$

$$\frac{\partial G(x,y)}{\partial y} = -\frac{g_2(y)}{G_2(y)}(\log G_2(y))^{\theta-1}\left[(\log G_1(x))^\theta + (\log G_2(y))^\theta\right]^{1/\theta-1}G(x,y),$$

$$\frac{\partial^2 G\left(x, y\right)}{\partial y \partial \theta} = -\frac{g_2(y)}{G_2(y)} \left(\log G_2(y)\right)^{\theta-1} \log \log G_2(y) \left[\left(\log G_2(y)\right)^{\theta} + \left(\log G_1(x)\right)^{\theta}\right]^{1/\theta-1} G(x, y)$$

$$+ \theta^{-2} \frac{g_2(y)}{G_2(y)} \left(\log G_2(y)\right)^{\theta-1} \left[\left(\log G_1(x)\right)^{\theta} + \left(\log G_2(y)\right)^{\theta}\right]^{1/\theta-1} G(x, y)$$

$$\times \log\left[\left(\log G_1(x)\right)^{\theta} + \left(\log G_2(y)\right)^{\theta}\right] - \left(\theta^{-1} - 1\right) \frac{g_2(y)}{G_2(y)} \left(\log G_2(y)\right)^{\theta-1} G(x, y)$$

$$\times \left[\left(\log G_1(x)\right)^{\theta} \log \log G_1(x) + \left(\log G_2(y)\right)^{\theta} \log \log G_2(y)\right]$$

$$\times \left[\left(\log G_1(x)\right)^{\theta} + \left(\log G_2(y)\right)^{\theta}\right]^{1/\theta-2} - \frac{g_2(y)}{G_2(y)} \left(\log G_2(y)\right)^{\theta-1}$$

$$\times \left[\left(\log G_1(x)\right)^{\theta} + \left(\log G_2(y)\right)^{\theta}\right]^{1/\theta-1} \frac{\partial G(x, y)}{\partial \theta},$$

$$\frac{\partial^2 G\left(x, y\right)}{\partial x \partial y} = -\frac{g_1(x) g_2(y)}{G_2(x) G_2(y)} \left(\log G_1(x)\right)^{\theta-1} \left(\log G_2(y)\right)^{\theta-1} \left[\left(\log G_1(x)\right)^{\theta} + \left(\log G_2(y)\right)^{\theta}\right]^{1/\theta-2}$$

$$\times G(x, y) \left\{ -\left[\left(\log G_1(x)\right)^{\theta} + \left(\log G_2(y)\right)^{\theta}\right]^{1/\theta} + (1-\theta) \right\},$$

and

$$\frac{\partial^3 G\left(x,y\right)}{\partial x \partial y \partial \theta} = \frac{g_1(x)g_2(y)}{G_2(x)G_2(y)} \log \log G_1(x) \left(\log G_1(x)\right)^{\theta-1} \left[\left(\log G_1(x)\right)^\theta + \left(\log G_2(y)\right)^\theta\right]^{2/\theta-2} G(x,y)$$

$$\times \left(\log G_2(y)\right)^{\theta-1} - (1-\theta)\frac{g_1(x)g_2(y)}{G_2(x)G_2(y)} \log \log G_1(x) \left(\log G_1(x)\right)^{\theta-1} \left(\log G_2(y)\right)^{\theta-1}$$

$$\times \left[\left(\log G_1(x)\right)^\theta + \left(\log G_2(y)\right)^\theta\right]^{1/\theta-2} G(x,y) + \frac{g_1(x)g_2(y)}{G_2(x)G_2(y)} \log \log G_2(y)$$

$$\times \left(\log G_1(x)\right)^{\theta-1} \left(\log G_2(y)\right)^{\theta-1} \left[\left(\log G_1(x)\right)^\theta + \left(\log G_2(y)\right)^\theta\right]^{2/\theta-2} G(x,y)$$

$$-(1-\theta)\frac{g_1(x)g_2(y)}{G_2(x)G_2(y)} \log \log G_2(y) \left(\log G_1(x)\right)^{\theta-1} \left(\log G_2(y)\right)^{\theta-1}$$

$$\times \left[\left(\log G_1(x)\right)^\theta + \left(\log G_2(y)\right)^\theta\right]^{1/\theta-2} G(x,y) - \theta^{-2}\frac{g_1(x)g_2(y)}{G_2(x)G_2(y)} \left(\log G_1(x)\right)^{\theta-1}$$

$$\times \left(\log G_2(y)\right)^{\theta-1} \left[\left(\log G_1(x)\right)^\theta + \left(\log G_2(y)\right)^\theta\right]^{2/\theta-2} G(x,y) \log \left[\left(\log G_1(x)\right)^\theta\right.$$

$$\left. + \left(\log G_2(y)\right)^\theta\right] + \theta^{-2}(1-\theta)\frac{g_1(x)g_2(y)}{G_2(x)G_2(y)} \left(\log G_1(x)\right)^{\theta-1} \left(\log G_2(y)\right)^{\theta-1}$$

$$\times \left[\left(\log G_1(x)\right)^\theta + \left(\log G_2(y)\right)^\theta\right]^{1/\theta-2} G(x,y) \log \left[\left(\log G_1(x)\right)^\theta + \left(\log G_2(y)\right)^\theta\right]$$

$$+ \left(\theta^{-1} - 2\right)\frac{g_1(x)g_2(y)}{G_2(x)G_2(y)} \left(\log G_1(x)\right)^{\theta-1} \left(\log G_2(y)\right)^{\theta-1} \left[\left(\log G_1(x)\right)^\theta + \left(\log G_2(y)\right)^\theta\right]^{2/\theta-3}$$

$$\times G(x,y) \left[\left(\log G_1(x)\right)^\theta \log \log G_1(x) + \left(\log G_2(y)\right)^\theta \log \log G_2(y)\right]$$

$$-(1-\theta)\left(\theta^{-1} - 2\right)\frac{g_1(x)g_2(y)}{G_2(x)G_2(y)} \left(\log G_1(x)\right)^{\theta-1} \left(\log G_2(y)\right)^{\theta-1} \left[\left(\log G_1(x)\right)^\theta\right.$$

$$\left. + \left(\log G_2(y)\right)^\theta\right]^{1/\theta-3} G(x,y) \left[\left(\log G_1(x)\right)^\theta \log \log G_1(x) + \left(\log G_2(y)\right)^\theta \log \log G_2(y)\right]$$

$$+ \frac{g_1(x)g_2(y)}{G_2(x)G_2(y)} \left(\log G_1(x)\right)^{\theta-1} \left(\log G_2(y)\right)^{\theta-1} \left[\left(\log G_1(x)\right)^\theta + \left(\log G_2(y)\right)^\theta\right]^{2/\theta-2}$$

$$\times \frac{\partial G(x,y)}{\partial \theta} - (1-\theta)\frac{g_1(x)g_2(y)}{G_2(x)G_2(y)} \left(\log G_1(x)\right)^{\theta-1} \left(\log G_2(y)\right)^{\theta-1}$$

$$\times \left[\left(\log G_1(x)\right)^\theta + \left(\log G_2(y)\right)^\theta\right]^{1/\theta-2} \frac{\partial G(x,y)}{\partial \theta} - \theta^{-2}\frac{g_1(x)g_2(y)}{G_2(x)G_2(y)}$$

$$\times \left(\log G_1(x)\right)^{\theta-1} \left(\log G_2(y)\right)^{\theta-1} \left[\left(\log G_1(x)\right)^\theta + \left(\log G_2(y)\right)^\theta\right]^{2/\theta-2}$$

$$\times G(x,y) \log \left[\left(\log G_1(x)\right)^\theta + \left(\log G_2(y)\right)^\theta\right]$$

$$+\theta^{-1}\frac{g_1(x)g_2(y)}{G_2(x)G_2(y)} \left(\log G_1(x)\right)^{\theta-1} \left(\log G_2(y)\right)^{\theta-1} \left[\left(\log G_1(x)\right)^\theta + \left(\log G_2(y)\right)^\theta\right]^{2/\theta-3}$$

$$\times G(x,y) \left[\left(\log G_1(x)\right)^\theta \log \log G_1(x) + \left(\log G_2(y)\right)^\theta \log \log G_2(y)\right]$$

$$+ \frac{g_1(x)g_2(y)}{G_2(x)G_2(y)} \left(\log G_1(x)\right)^{\theta-1} \left(\log G_2(y)\right)^{\theta-1} \left[\left(\log G_1(x)\right)^\theta + \left(\log G_2(y)\right)^\theta\right]^{1/\theta-2} G(x,y).$$

Appendix B

Appendix to Chapter 5

In this section we present main Lemmas that are repeatedly used in chapter 5.

Lemma 1 *Let*

$$J_3(t, \beta) = \int_{-\infty}^{\infty} e^{tx} u^{\xi+1} (1-u)^{\beta-1} dx,$$

where $u = \{1 + \xi(x-\mu)/\sigma\}^{-1/\xi}$. Then

$$J_3(t, \beta) = \sigma e^{t_1^*} \sum_{i=0}^{\infty} \frac{(t_2^*)^i}{i!} B(1 - i\xi, \beta).$$

Where $t_1^* = t(\mu - \frac{\sigma}{\xi})$ and $t_2^* = t\frac{\sigma}{\xi}$.

Proof: *We can write*

$$J_3(t, \beta) = \int_{-\infty}^{\infty} e^{tx} u^{\xi+1} (1-u)^{\beta-1} dx,$$

$$= \sigma \int_0^1 e^{t\left((\sigma/\xi)u^{-\xi} + (\mu - \sigma/\xi)\right)} (1-u)^{\beta-1} du,$$

$$= \sigma e^{t_1^*} \int_0^1 e^{t_2^* u^{-\xi}} (1-u)^{\beta-1} du,$$

$$= \sigma e^{t_1^*} \sum_{i=0}^{\infty} \frac{(t_2^*)^i}{i!} \int_0^1 u^{-i\xi} (1-u)^{\beta-1} du,$$

$$\therefore J_3(t, \beta) = \sigma e^{t_1^*} \sum_{i=0}^{\infty} \frac{(t_2^*)^i}{i!} B(1 - i\xi, \beta).$$

□

Lemma 2 *Let*

$$J_4(y, \beta) = \int_y^{\infty} x u^{\xi+1} (1-u)^{a(k+1)} dx,$$

where $u = \{1 + \xi(x-t)/\sigma\}^{-1/\xi}$. Then

$$J_4(y, \beta) = \frac{\sigma^2}{\xi} B(y^*; 1-\xi, \beta) + \sigma(t - \sigma/\xi)\left[1 - (1-y^*)^\beta\right].$$

Proof: *We can write*

$$J_4(y, \beta) = \int_y^\infty xu^{\xi+1}(1-u)^\beta dx,$$

$$= \sigma \int_0^{y^*} \left[\frac{\sigma}{\xi}u^{-\xi} + (t - \frac{\sigma}{\xi})\right](1-u)^{\beta-1}du,$$

$$= \frac{\sigma^2}{\xi} \int_0^{y^*} u^{-\xi}(1-u)^{\beta-1}du + \sigma(t - \frac{\sigma}{\xi})$$

$$\times \int_0^{y^*} (1-u)^{\beta-1}du,$$

$$\therefore J_4(y, \beta) = \frac{\sigma^2}{\xi} B(y^*; 1-\xi, \beta) + \sigma(t - \sigma/\xi)\left[1 - (1-y^*)^\beta\right].$$

where $y^* = \{1 + \xi(y-t)/\sigma\}^{-1/\xi}$. and $B(x; a, b)$ denotes the incomplete beta function defined by

$$B(x, a, b) = \int_0^x z^{a-1}(1-z)^{b-1}dz.$$

$\square$

Information Matrix

As shown in the chapter 5, the information matrix $\mathbf{J}$ can be written as:

$$\mathbf{J} = \begin{pmatrix} \dfrac{\partial^2 \log L}{\partial a^2} & \dfrac{\partial^2 \log L}{\partial a \partial b} & \dfrac{\partial^2 \log L}{\partial a \partial \sigma} & \dfrac{\partial^2 \log L}{\partial a \partial \xi} \\ \dfrac{\partial^2 \log L}{\partial b \partial a} & \dfrac{\partial^2 \log L}{\partial b^2} & \dfrac{\partial^2 \log L}{\partial b \partial \sigma} & \dfrac{\partial^2 \log L}{\partial b \partial \xi} \\ \dfrac{\partial^2 \log L}{\partial \sigma \partial a} & \dfrac{\partial^2 \log L}{\partial \sigma \partial b} & \dfrac{\partial^2 \log L}{\partial \sigma^2} & \dfrac{\partial^2 \log L}{\partial \sigma \partial \xi} \\ \dfrac{\partial^2 \log L}{\partial \xi \partial a} & \dfrac{\partial^2 \log L}{\partial \xi \partial b} & \dfrac{\partial^2 \log L}{\partial \xi \partial \sigma} & \dfrac{\partial^2 \log L}{\partial \xi^2} \end{pmatrix}.$$

The explicit expressions for the elements of **J** *are given as*

$$J_{11} = -\frac{n}{a^2} + (1-b)\sum_{i=1}^{n} \frac{(1-u_i)^a \log^2(1-u_i)}{1-(1-u_i)^a} + (1-b)\sum_{i=1}^{n} \frac{(1-u_i)^{2a}\log^2(1-u_i)}{[1-(1-u_i)^a]^2},$$

$$J_{12} = -\sum_{i=1}^{n} \frac{(1-u_i)^a \log(1-u_i)}{1-(1-u_i)^a},$$

$$J_{13} = -\frac{1}{\sigma^2}\sum_{i=1}^{n} \frac{u_i^{1+\xi}(x_i-t)}{1-u_i} + \frac{b-1}{\sigma^2}\sum_{i=1}^{n} u_i^{1+\xi}(x_i-t)(1-u_i)^{a-1}[1-(1-u_i)^a]$$

$$\times [a\log(1-u_i)+1] + \frac{a(b-1)}{\sigma^2}\sum_{i=1}^{n} u_i^{1+\xi}(x_i-t)(1-u_i)^{2a-1}\log(1-u_i),$$

$$J_{14} = -\frac{1}{\sigma^2}\sum_{i=1}^{n} \frac{u_i}{1-u_i}\left\{\log\left[1+\xi\frac{x_i-t}{\sigma}\right] - \frac{x_i\xi}{\sigma}\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1}\right\} + \frac{b-1}{\sigma^2}\sum_{i=1}^{n} u_i$$

$$\times (1-u_i)^{a-1}[a\log(1-u_i)+1]\left\{\log\left[1+\xi\frac{x_i-t}{\sigma}\right] - \frac{x_i\xi}{\sigma}\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1}\right\}$$

$$-\frac{b-1}{\sigma^2}\sum_{i=1}^{n} u_i[a\log(1-u_i)+1]\left\{\log\left[1+\xi\frac{x_i-t}{\sigma}\right] - \frac{x_i\xi}{\sigma}\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1}\right\}$$

$$\times (1-u_i)^{2a-1} + \frac{a(b-1)}{\sigma^2}\sum_{i=1}^{n} u_i(1-u_i)^{2a-1}\log(1-u_i)\left\{\log\left[1+\xi\frac{x_i-t}{\sigma}\right]\right.$$

$$\left. -\frac{x_i\xi}{\sigma}\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1}\right\},$$

$$J_{22} = -\frac{n}{b^2},$$

$$J_{23} = \frac{a}{\sigma^2}\sum_{i=1}^{n} \frac{u_i^{1+\xi}(1-u_i)^{a-1}(x_i-t)}{1-(1-u_i)^a},$$

$$J_{24} = \frac{a}{\xi^2}\sum_{i=1}^{n} \frac{u_i(1-u_i)^{a-1}}{1-(1-u_i)^a}\left\{\log\left[1+\xi\frac{x_i-t}{\sigma}\right] - \frac{x_i\xi}{\sigma}\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1}\right\},$$

$$J_{33} = \frac{n}{\sigma^2} - \frac{2(1+\xi)}{\sigma^3}\sum_{i=1}^{n} u_i^{\xi}(x_i-t) + \frac{\xi(1+\xi)}{\sigma^4}\sum_{i=1}^{n} u_i^{2\xi}(x_i-t)^2 - \frac{a-1}{\sigma^4}$$

$$\times \sum_{i=1}^{n} \frac{u_i^{2\xi+2}(x_i-t)^2}{(1-u_i)^2} - \frac{(a-1)(1+\xi)}{\sigma^4}\sum_{i=1}^{n} \frac{u_i^{2\xi+1}(x_i-t)^2}{1-u_i} + \frac{2(a-1)}{\sigma^3}$$

$$\times \sum_{i=1}^{n} \frac{u_i^{\xi+1}(x_i-t)}{1-u_i} - \frac{a(a-1)(b-1)}{\sigma^4} \sum_{i=1}^{n} \frac{u_i^{2\xi+2}(1-u_i)^{a-2}(x_i-t)^2}{1-(1-u_i)^a}$$

$$-\frac{a^2(b-1)}{\sigma^4} \sum_{i=1}^{n} \frac{u_i^{2\xi+2}(1-u_i)^{2a-2}(x_i-t)^2}{[1-(1-u_i)^a]^2} + \frac{a(b-1)(1+\xi)}{\sigma^2}$$

$$\times \sum_{i=1}^{n} \frac{u_i^{2\xi+1}(1-u_i)^{a-1}(x_i-t)}{1-(1-u_i)^a} - \frac{2a(b-1)}{\sigma^3} \sum_{i=1}^{n} \frac{u_i^{\xi+1}(1-u_i)^{a-1}(x_i-t)}{1-(1-u_i)^a},$$

$$J_{34} = \frac{1}{\sigma^2} \sum_{i=1}^{n} u_i^{\xi}(x_i-t) + \frac{1+\xi}{\xi^2\sigma^2} \sum_{i=1}^{n} u_i^{2+\xi} \log u_i (x_i-t) \left\{ \log\left[1+\xi\frac{x_i-t}{\sigma}\right] - \frac{x_i\xi}{\sigma} \right.$$

$$\times \left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1} \right\} - \frac{a-1}{\xi^2\sigma^2} \sum_{i=1}^{n} \frac{u_i^{2+\xi}(x_i-t)}{(1-u_i)^2} \left\{ \log\left[1+\xi\frac{x_i-t}{\sigma}\right] - \frac{x_i\xi}{\sigma} \right.$$

$$\times \left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1} \right\} - \frac{(a-1)(1+\xi)}{\xi^2\sigma^2} \sum_{i=1}^{n} \frac{u_i^{1+\xi}(x_i-t)}{1-u_i} \left\{ \log\left[1+\xi\frac{x_i-t}{\sigma}\right] \right.$$

$$-\frac{x_i\xi}{\sigma} \left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1} \right\} - \frac{a(a-1)(b-1)}{\xi^2\sigma^2} \sum_{i=1}^{n} \frac{u_i^{2+\xi}(1-u_i)^{a-2}(x_i-t)}{1-(1-u_i)^a}$$

$$\times \left\{ \log\left[1+\xi\frac{x_i-t}{\sigma}\right] - \frac{x_i\xi}{\sigma} \left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1} \right\} - \frac{a^2(b-1)}{\xi^2\sigma^2} \sum_{i=1}^{n} (x_i-t)$$

$$\times \frac{u_i^{2+\xi}(1-u_i)^{2a-2}}{[1-(1-u_i)^a]^2} \left\{ \log\left[1+\xi\frac{x_i-t}{\sigma}\right] - \frac{x_i\xi}{\sigma} \left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1} \right\}$$

$$+\frac{a(b-1)(1+\xi)}{\xi^2\sigma^2} \sum_{i=1}^{n} \frac{u_i^{1+\xi}(1-u_i)^{a-1}(x_i-t)}{1-(1-u_i)^a} \left\{ \log\left[1+\xi\frac{x_i-t}{\sigma}\right] \right.$$

$$-\frac{x_i\xi}{\sigma} \left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1} \right\}$$

and

$$
\begin{aligned}
J_{44} = {} & \frac{1}{\xi} \sum_{i=1}^{n} \left\{ \log\left[1 + \xi\frac{x_i - t}{\sigma}\right] - \frac{x_i\xi}{\sigma}\left[1 + \xi\frac{x_i - t}{\sigma}\right]^{-1} \right\} \\
& - \frac{2 + \xi}{\xi^3} \sum_{i=1}^{n} \left\{ \log\left[1 + \xi\frac{x_i - t}{\sigma}\right] - \frac{x_i\xi}{\sigma}\left[1 + \xi\frac{x_i - t}{\sigma}\right]^{-1} \right\} \\
& + \frac{1 + \xi}{\sigma^2\xi^2} \sum_{i=1}^{n} \left\{ -t\sigma\left[1 + \xi\frac{x_i - t}{\sigma}\right]^{-1} + \xi x_i\,(x_i - t)\left[1 + \xi\frac{x_i - t}{\sigma}\right]^{-2} \right\} \\
& + \frac{2(a - 1)}{\xi^3} \sum_{i=1}^{n} \frac{u_i}{1 - u_i} \left\{ \log\left[1 + \xi\frac{x_i - t}{\sigma}\right] - \frac{x_i\xi}{\sigma}\left[1 + \xi\frac{x_i - t}{\sigma}\right]^{-1} \right\} \\
& - \frac{a - 1}{\xi^4} \sum_{i=1}^{n} \frac{u_i}{(1 - u_i)^2} \left\{ \log\left[1 + \xi\frac{x_i - t}{\sigma}\right] - \frac{x_i\xi}{\sigma}\left[1 + \xi\frac{x_i - t}{\sigma}\right]^{-1} \right\}^2 \\
& - \frac{a - 1}{a\xi^2} \sum_{i=1}^{n} \frac{u_i}{1 - u_i}\left[1 + \xi\frac{x_i - t}{\sigma}\right]^{-1} \left\{ -t + \frac{x_i\xi}{\sigma}\,(x_i - t)\left[1 + \xi\frac{x_i - t}{\sigma}\right]^{-1} \right\} \\
& - \frac{2a(b - 1)}{\xi^3} \sum_{i=1}^{n} \frac{u_i\,(1 - u_i)^{a-1}}{1 - (1 - u_i)^a} \left\{ \log\left[1 + \xi\frac{x_i - t}{\sigma}\right] - \frac{x_i\xi}{\sigma}\left[1 + \xi\frac{x_i - t}{\sigma}\right]^{-1} \right\} \\
& + \frac{a(b - 1)}{\xi^2} \sum_{i=1}^{n} \frac{(1 - u_i)^{a-2}\,(1 - au_i)}{1 - (1 - u_i)^a} \left\{ \log\left[1 + \xi\frac{x_i - t}{\sigma}\right] - \frac{x_i\xi}{\sigma}\left[1 + \xi\frac{x_i - t}{\sigma}\right]^{-1} \right\}^2 \\
& - \frac{a^2(b - 1)}{\xi^2} \sum_{i=1}^{n} \frac{u_i\,(1 - u_i)^{2a-2}}{[1 - (1 - u_i)^a]^2} \left\{ \log\left[1 + \xi\frac{x_i - t}{\sigma}\right] - \frac{x_i\xi}{\sigma}\left[1 + \xi\frac{x_i - t}{\sigma}\right]^{-1} \right\}^2 \\
& + \frac{a(b - 1)}{\sigma\xi^2} \sum_{i=1}^{n} \frac{u_i\,(1 - u_i)^{a-1}}{1 - (1 - u_i)^a}\left[1 + \xi\frac{x_i - t}{\sigma}\right]^{-1} \left\{ -t + \frac{x_i\xi}{\sigma}\,(x_i - t)\left[1 + \xi\frac{x_i - t}{\sigma}\right]^{-1} \right\}.
\end{aligned}
$$

Explicit expressions for the remaining elements of **J** *follow by symmetry.*

Appendix C

Appendix to Chapter 6

Here, we give explicit expressions for the elements of the information matrix $\mathbf{J}$ *defined in section (6.7):*

$$J_{11} = -\frac{n}{\alpha^2} - \frac{1}{\delta} \sum_{i=1}^{n} (x_i - \mu)^\alpha \ln^2 (x_i - \mu) + \frac{1}{\delta} \left(\frac{1}{\beta} + 1 \right) \sum_{i=1}^{n} \frac{(x_i - \mu)^\alpha \ln^2 (x_i - \mu)}{1 + \exp\left[\frac{1}{\delta} (x_i - \mu)^\alpha \right]}$$

$$- \frac{1}{\delta^2} \left(\frac{1}{\beta} + 1 \right) \sum_{i=1}^{n} \frac{(x_i - \mu)^{2\alpha} \ln^2 (x_i - \mu)}{\left\{ 1 + \exp\left[\frac{1}{\delta} (x_i - \mu)^\alpha \right] \right\}^2},$$

$$J_{12} = -\frac{1}{\delta \beta^2} \sum_{i=1}^{n} \frac{(x_i - \mu)^\alpha \ln (x_i - \mu)}{1 + \exp\left[\frac{1}{\delta} (x_i - \mu)^\alpha \right]},$$

$$J_{13} = \frac{1}{\delta^2} \sum_{i=1}^{n} (x_i - \mu)^\alpha \ln (x_i - \mu) - \frac{1}{\delta^2} \left(\frac{1}{\beta} + 1 \right) \sum_{i=1}^{n} \frac{(x_i - \mu)^\alpha \ln (x_i - \mu)}{1 + \exp\left[\frac{1}{\delta} (x_i - \mu)^\alpha \right]}$$

$$+ \frac{1}{\delta^3} \left(\frac{1}{\beta} + 1 \right) \sum_{i=1}^{n} \frac{(x_i - \mu)^{2\alpha} \ln (x_i - \mu) \exp\left[\frac{1}{\delta} (x_i - \mu)^\alpha \right]}{\left\{ 1 + \exp\left[\frac{1}{\delta} (x_i - \mu)^\alpha \right] \right\}^2},$$

$$J_{14} = \frac{\alpha}{\delta} \sum_{i=1}^{n} (x_i - \mu)^{\alpha-1} \ln (x_i - \mu) + \frac{1}{\delta} \sum_{i=1}^{n} (x_i - \mu)^{\alpha-1} - \sum_{i=1}^{n} (x_i - \mu)^{-1}$$

$$- \frac{\alpha + 1}{\delta} \left(\frac{1}{\beta} + 1 \right) \sum_{i=1}^{n} \frac{(x_i - \mu)^{\alpha-1}}{1 + \exp\left[\frac{1}{\delta} (x_i - \mu)^\alpha \right]}$$

$$+\frac{\alpha}{\delta^2}\left(\frac{1}{\beta}+1\right)\sum_{i=1}^{n}\frac{(x_i-\mu)^{2\alpha-1}\ln(x_i-\mu)\exp\left[\frac{1}{\delta}(x_i-\mu)^{\alpha}\right]}{\left\{1+\exp\left[\frac{1}{\delta}(x_i-\mu)^{\alpha}\right]\right\}^2},$$

$$J_{22}=\frac{n}{\beta^2}-\frac{2\ln 2n}{\beta^2\left(2^{1/\beta}-1\right)}+\frac{2(\ln 2)^2 n}{\beta^4\left(2^{1/\beta}-1\right)^2}-\frac{2}{\beta^3}\sum_{i=1}^{n}\ln\left\{1+\exp\left[\frac{1}{\delta}(x_i-\mu)^{\alpha}\right]\right\},$$

$$J_{23}=\frac{\alpha}{\beta^2\delta}\sum_{i=1}^{n}\frac{(x_i-\mu)^{\alpha-1}}{1+\exp\left[\frac{1}{\delta}(x_i-\mu)^{\alpha}\right]},$$

$$J_{24}=\frac{n}{\delta^2}-\frac{2}{\delta^3}\sum_{i=1}^{n}(x_i-\mu)^{\alpha}+\frac{2}{\delta^3}\left(\frac{1}{\beta}+1\right)\sum_{i=1}^{n}\frac{(x_i-\mu)^{\alpha}}{1+\exp\left[\frac{1}{\delta}(x_i-\mu)^{\alpha}\right]}$$

$$-\frac{1}{\delta^4}\left(\frac{1}{\beta}+1\right)\sum_{i=1}^{n}\frac{(x_i-\mu)^{2\alpha}\exp\left[\frac{1}{\delta}(x_i-\mu)^{\alpha}\right]}{\left\{1+\exp\left[\frac{1}{\delta}(x_i-\mu)^{\alpha}\right]\right\}^2},$$

$$J_{34}=-\frac{\alpha}{\delta^2}\sum_{i=1}^{n}(x_i-\mu)^{\alpha-1}+\frac{\alpha}{\delta^2}\left(\frac{1}{\beta}+1\right)\sum_{i=1}^{n}\frac{(x_i-\mu)^{\alpha-1}}{1+\exp\left[\frac{1}{\delta}(x_i-\mu)^{\alpha}\right]}$$

$$-\frac{\alpha}{\delta^3}\left(\frac{1}{\beta}+1\right)\sum_{i=1}^{n}\frac{(x_i-\mu)^{2\alpha-1}}{\left\{1+\exp\left[\frac{1}{\delta}(x_i-\mu)^{\alpha}\right]\right\}^2}$$

and

$$J_{44}=-\frac{\alpha(\alpha-1)}{\delta}\sum_{i=1}^{n}(x_i-\mu)^{\alpha-2}+(\alpha-1)\sum_{i=1}^{n}(x_i-\mu)^{-2}$$

$$+\frac{\alpha(\alpha-1)}{\delta}\left(\frac{1}{\beta}+1\right)\sum_{i=1}^{n}\frac{(x_i-\mu)^{\alpha-2}}{1+\exp\left[\frac{1}{\delta}(x_i-\mu)^{\alpha}\right]}$$

$$-\frac{\alpha^2}{\delta^2}\left(\frac{1}{\beta}+1\right)\sum_{i=1}^{n}\frac{(x_i-\mu)^{2\alpha-2}}{\left\{1+\exp\left[\frac{1}{\delta}(x_i-\mu)^{\alpha}\right]\right\}^2}.$$

Explicit expressions for the remaining elements of **J** *follow by symmetry.*

Rainfall data from Maple Ridge in British Columbia, Canada

fifty-two ordered annual maximum antecedent rainfall measurements in mm from Maple Ridge in British Columbia, Canada: "264.9, 314.1, 364.6, 379.8, 419.3, 457.4, 459.4, 460.0, 490.3, 490.6, 502.2, 525.2, 526.8, 528.6, 528.6, 537.7, 539.6, 540.8, 551.0, 573.5, 579.2, 588.2, 588.7, 589.7, 592.1, 592.8, 600.8, 604.4, 608.4, 609.8,

619.2, 626.4, 629.4, 636.4, 645.2, 657.6, 663.5, 664.9, 671.7, 673.0, 682.6, 689.8, 698.0, 698.6, 698.8, 703.2, 755.9, 786.0, 787.2, 798.6, 850.4, 895.1".

Appendix D

Appendix to Chapter 7

The calculations in chapter 7 require the following lemmas.

Lemma 1 *Let*

$$K_1(\mu, n) = \int_\mu^\infty x^n u^{1+\xi} \left(u^{-\xi} - 1 \right)^{1-1/a} \exp(-u) dx,$$

where $u = \{1 + \xi/\sigma(x - \mu)^a\}^{-1/\xi}$. Then

$$K_1(\mu, n) = \frac{\xi}{a} \sum_{k=0}^n \sum_{j=0}^{k/a} (-1)^{k/a-j} \binom{n}{k} \binom{k/a}{j} \left(\frac{\sigma}{\xi} \right)^{(k+1)/a} \mu^{n-k} \gamma \left(1 - \xi j, 1 \right),$$

Proof: *We can write*

$$K_1(\mu, n) = \int_\mu^\infty x^n u^{1+\xi} \left(u^{-\xi} - 1 \right)^{1-1/a} \exp(-u) dx,$$

$$= \frac{\xi}{a} \left(\frac{\sigma}{\xi} \right)^{1/a} \int_0^1 \left[\frac{\sigma}{\xi} \left(u^{-\xi} - 1 \right)^{1/a} + \mu \right]^n \exp(-u) du$$

$$= \frac{\xi}{a} \sum_{k=0}^n \sum_{j=0}^{k/a} (-1)^{k/a-j} \binom{n}{k} \binom{k/a}{j} \left(\frac{\sigma}{\xi} \right)^{(k+1)/a} \mu^{n-k} \int_0^1 u^{-\xi j} \exp(-u) du.$$

Using the binomial expansion so, the result follows from the definition of incomplete gamma function. $\square$

Lemma 2 *Let*

$$K_2(\mu, t) = \int_\mu^\infty e^{tx} u^{1+\xi} \left(u^{-\xi} - 1 \right)^{1-1/a} \exp(-u) dx,$$

where $u = \{1 + \xi/\sigma(x-\mu)^a\}^{-1/\xi}$. *Then*

$$K_2(\mu, t) = \frac{\xi}{a} e^{t\mu} \left(\frac{\sigma}{\xi}\right)^{1/a} \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \frac{(-1)^{j/a-i} t^{*j}}{j!} \binom{j/a}{i} \gamma(1 - \xi i, 1).$$

Proof: *We can write*

$$K_2(\mu, t) = \int_{\mu}^{\infty} e^{tx} u^{1+\xi} \left(u^{-\xi} - 1\right)^{1-1/a} \exp(-u) dx,$$

$$= \frac{\xi}{a} e^{t\mu} \left(\frac{\sigma}{\xi}\right)^{1/a} \int_0^1 \exp\{t^* \left(u^{-\xi} - 1\right)^{1/a} - u\} du,$$

The exponential part in this integral can be represent by the power series

$e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$

$$K_2(\mu, t) = \frac{\xi}{a} e^{t\mu} \left(\frac{\sigma}{\xi}\right)^{1/a} \sum_{j=0}^{\infty} \frac{t^{*j}}{j!} \int_0^1 \left(u^{-\xi} - 1\right)^{j/a} \exp(-u) du.$$

Using the result from Lemma 1 in appendix **D**, *then*

$$K_2(\mu, t) = \frac{\xi}{a} e^{t\mu} \left(\frac{\sigma}{\xi}\right)^{1/a} \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \frac{(-1)^{j/a-i} t^{*j}}{j!} \binom{j/a}{i} \gamma(1 - \xi i, 1).$$

So, the result follows from the definition of incomplete gamma function. $\square$

Lemma 3 *Let*

$$K_3(z) = \int_z^{\infty} x \, u^{1+\xi} \left(u^{-\xi} - 1\right)^{1-1/a} \exp(-u) dx,$$

where $u = \{1 + \xi/\sigma(x-\mu)^a\}^{-1/\xi}$. *Then*

$$K_3(z) = \frac{\xi}{a} \left(\frac{\sigma}{\xi}\right)^{1/a} \left\{ \sum_{j=0}^{1/a} (-1)^{1/a-j} \binom{1/a}{j} \left(\frac{\sigma}{\xi}\right)^{1/a} \gamma(1 - \xi i, z^*) + \left(1 - e^{-z^*}\right) \right\}$$

where $z^* = \{1 + \xi(z-\mu)/\sigma\}^{-1/\xi}$.

Proof: *We can write*

$$K_3(z) = \int_z^{\infty} x \, u^{1+\xi} \left(u^{-\xi} - 1\right)^{1-1/a} \exp(-u) dx,$$

$$= \frac{\xi}{a} \left(\frac{\sigma}{\xi}\right)^{1/a} \int_0^{z^*} \left[\frac{\sigma}{\xi} \left(u^{-\xi} - 1\right)^{1/a} + \mu\right]^n \exp(-u) du,$$

$$\therefore K_3(z) = \frac{\xi}{a} \left(\frac{\sigma}{\xi}\right)^{1/a} \left\{ \sum_{j=0}^{1/a} (-1)^{1/a-j} \binom{1/a}{j} \left(\frac{\sigma}{\xi}\right)^{1/a} \gamma(1 - \xi i, z^*) + \left(1 - e^{-z^*}\right) \right\}.$$

So, the result follows from the definition of incomplete gamma function. $\square$

Information Matrix

The information matrix for NEV distribution can be written as:

$$
\mathbf{J} = \begin{pmatrix}
\dfrac{\partial^2 \log L}{\partial a^2} & \dfrac{\partial^2 \log L}{\partial a \partial \sigma} & \dfrac{\partial^2 \log L}{\partial a \partial \xi} \\[2ex]
\dfrac{\partial^2 \log L}{\partial \sigma \partial a} & \dfrac{\partial^2 \log L}{\partial \sigma^2} & \dfrac{\partial^2 \log L}{\partial \sigma \partial \xi} \\[2ex]
\dfrac{\partial^2 \log L}{\partial \xi \partial a} & \dfrac{\partial^2 \log L}{\partial \xi \partial \sigma} & \dfrac{\partial^2 \log L}{\partial \xi^2}
\end{pmatrix}.
$$

Then the 4×4 elements of information matrix are given by

$$
J_{11} = -\frac{n}{a^2} - \frac{2n}{a^3} \log\left(\frac{\sigma}{\xi}\right) - \frac{2}{a^3} \sum_{i=1}^{n} \left(u_i^{-\xi} - 1 \right) + \frac{2\xi}{a^2 \sigma} \sum_{i=1}^{n} (x_i - \mu)^a \log (x_i - \mu)
$$

$$
- \left(1 - \frac{1}{a}\right) \frac{\xi}{\sigma} \sum_{i=1}^{n} (x_i - \mu)^a \log^2 (x_i - \mu) - \frac{(1+\xi)\,\xi}{\sigma^2} \sum_{i=1}^{n} u_i^{2\xi} \ (x_i - \mu)^{2a}
$$

$$
\times \log^2 (x_i - \mu) + \frac{(1+\xi)}{\sigma} \sum_{i=1}^{n} u_i^{\xi} (x_i - \mu)^a \log^2 (x_i - \mu) - \frac{(1+\xi)}{\sigma^2} \sum_{i=1}^{n} u_i^{2\xi+1}
$$

$$
\times (x_i - \mu)^{2a} \log^2 (x_i - \mu) + \frac{1}{\sigma} \sum_{i=1}^{n} u_i^{\xi+1} (x_i - \mu)^a \log^2 (x_i - \mu),
$$

$$
J_{12} = -\frac{\xi}{a\,\sigma} \sum_{i=1}^{n} (x_i - \mu)^a - \left(1 - \frac{1}{a}\right) \frac{\xi}{\sigma} \sum_{i=1}^{n} (x_i - \mu)^{a-1} \left(a \log(x_i - \mu) + 1 \right)
$$

$$
- \frac{(1+\xi)}{\sigma} \sum_{i=1}^{n} \left[\frac{a\,\xi}{\sigma} u_i^{2\xi} (x_i - \mu)^{2a-1} \log (x_i - \mu) - a \ u_i^{\xi} (x_i - \mu)^{a-1} \right.
$$

$$
\times \log (x_i - \mu) - u_i^{\xi} (x_i - \mu)^{a-1} \right] + \frac{1}{\sigma} \sum_{i=1}^{n} \left[\frac{a(\xi+1)}{\sigma} u_i^{2\xi+1} (x_i - \mu)^{2a-1} \right.
$$

$$
\times \log (x_i - \mu) - a \ u_i^{\xi+1} (x_i - \mu)^{a-1} \log (x_i - \mu) - u_i^{\xi+1} (x_i - \mu)^{a-1} \Big],
$$

$$J_{13} = -\frac{n}{a^2\sigma} - \frac{\xi}{a^2\,\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^a - (1 - \frac{1}{a})\frac{\xi}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^a\log(x_i - \mu)$$

$$+\frac{(1+\xi)}{\sigma^2}\sum_{i=1}^{n}u_i^{\xi}(x_i - \mu)^a\log(x_i - \mu) - \frac{(1+\xi)\xi}{\sigma^3}\sum_{i=1}^{n}u_i^{2\xi}(x_i - \mu)^{2a}$$

$$\times\log(x_i - \mu) - \frac{1}{\sigma^2}\sum_{i=1}^{n}u_i^{\xi+1}(x_i - \mu)^a\log(x_i - \mu) + \frac{(1+\xi)}{\sigma^3}\sum_{i=1}^{n}u_i^{2\xi+1}$$

$$\times(x_i - \mu)^{2a}\log(x_i - \mu),$$

$$J_{14} = -\frac{n}{a^2\xi} + \frac{1}{a^2\,\sigma}\sum_{i=1}^{n}(x_i - \mu)^a + (1 - \frac{1}{a})\frac{1}{\sigma}\sum_{i=1}^{n}(x_i - \mu)^a\log(x_i - \mu)$$

$$-\frac{1}{\sigma}\sum_{i=1}^{n}u_i^{\xi}(x_i - \mu)^a\log(x_i - \mu) - \frac{(1+\xi)}{\sigma^2}\sum_{i=1}^{n}u_i^{2\xi}(x_i - \mu)^{2a}\log(x_i - \mu)$$

$$-\frac{1}{\sigma\xi}\sum_{i=1}^{n}u_i^{\xi+1}(x_i - \mu)^a\log(x_i - \mu)\Big(\log u_i + (1+\xi)u_i^{\xi}\frac{(x_i - \mu)^a}{\sigma}\Big),$$

$$J_{22} = \frac{\xi(a-1)^2}{\sigma}\sum_{i=1}^{n}(x_i - \mu)^{a-2} - \frac{a(1+\xi)}{\sigma}\sum_{i=1}^{n}u_i^{\xi}(x_i - \mu)^{a-2}\Big[(a-1) - \frac{a\,\xi}{\sigma}u_i^{\xi}$$

$$\times(x_i - \mu)^a] + \frac{a}{\sigma}\sum_{i=1}^{n}u_i^{\xi+1}(x_i - \mu)^{a-2}\Big[(a-1) - \frac{a\,(\xi+1)}{\sigma}u_i^{\xi}(x_i - \mu)^a\Big],$$

$$J_{23} = \frac{\xi(a-1)}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^{a-1} - \frac{a(1+\xi)}{\sigma^2}\sum_{i=1}^{n}u_i^{\xi}(x_i - \mu)^{a-1} + \frac{a(1+\xi)\xi}{\sigma^3}\sum_{i=1}^{n}u_i^{2\xi}$$

$$\times(x_i - \mu)^{2a-1} + \frac{a}{\sigma^2}\sum_{i=1}^{n}u_i^{\xi+1}(x_i - \mu)^{a-1} - \frac{a(1+\xi)}{\sigma^3}\sum_{i=1}^{n}u_i^{2\xi+1}(x_i - \mu)^{2a-1},$$

$$J_{24} = -\frac{(a-1)}{\sigma}\sum_{i=1}^{n}(x_i - \mu)^{a-1} + \frac{a}{\sigma}\sum_{i=1}^{n}u_i^{\xi}(x_i - \mu)^{a-1} - \frac{a(1+\xi)}{\sigma^2}\sum_{i=1}^{n}u_i^{2\xi}$$

$$\times(x_i - \mu)^{2a-1} + \frac{a}{\sigma\xi}\sum_{i=1}^{n}u_i^{\xi+1}(x_i - \mu)^{a-1}\Big(\log u_i + (1+\xi)u_i^{\xi}(x_i - \mu)^a\Big),$$

$$J_{33} = \frac{n}{a\sigma^2} - (1 - \frac{1}{a})\frac{2\xi}{\sigma^3}\sum_{i=1}^{n}(x_i - \mu)^a - \frac{2(1+\xi)}{\sigma^3}\sum_{i=1}^{n}u_i^{\xi}(x_i - \mu)^a + \frac{(1+\xi)\xi}{\sigma^4}\sum_{i=1}^{n}u_i^{2\xi}$$

$$\times(x_i - \mu)^{2a} + \frac{2}{\sigma^2}\sum_{i=1}^{n}u_i^{\xi+1}(x_i - \mu)^a - \frac{(1+\xi)}{\sigma^4}\sum_{i=1}^{n}u_i^{2\xi+1}(x_i - \mu)^{2a},$$

$$J_{34} = \Big(1 - \frac{1}{a}\Big)\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^a + \frac{1}{\sigma^2}\sum_{i=1}^{n}u_i^{\xi}(x_i - \mu)^a - \frac{(1+\xi)}{\sigma^3}\sum_{i=1}^{n}u_i^{2\xi}$$

$$\times(x_i - \mu)^{2a} + \frac{1}{\sigma^2\xi}\sum_{i=1}^{n}u_i^{\xi+1}(x_i - \mu)^a\Big(\log u_i + (1+\xi)u_i^{\xi}(x_i - \mu)^a\Big),$$

And

$$J_{44} = -\frac{n}{\xi^2}(1 - 1/a) + \frac{2}{\xi^2} \sum_{i=1}^{n} \left(\log u_i + u_i^{\xi} \frac{(x_i - \mu)^a}{\sigma} \right) - \frac{1}{\sigma^2 \xi} \sum_{i=1}^{n} u_i^{2\xi+1} (x_i - \mu)^{2a}$$

$$- \frac{2}{\xi^2} \sum_{i=1}^{n} u_i \left(\log u_i + u_i^{\xi} \frac{(x_i - \mu)^a}{\sigma} \right) + \frac{(1 + \xi)}{\sigma^2 \xi} \sum_{i=1}^{n} (x_i - \mu)^{2a} u_i^{2\xi} - \frac{1}{\xi^2} \sum_{i=1}^{n} u_i$$

$$\times \left(\log u_i + u_i^{\xi} \frac{(x_i - \mu)^a}{\sigma} \right)^2.$$

Where $u = [1 + \frac{\xi}{\sigma}(x - \mu)^a]^{-1/\xi}$.