UNIVERSITY MOHAMED KHIDER, BISKRA
FACULTY OF EXACT SCIENCES AND NATURE SCIENCES
DEPARTMENT OF MATHEMATICS

A Thesis Presented For The Degree Of

DOCTOR OF SCIENCES

In The Field Of Statistics

BY

Abdelaziz RASSOUL

Estimate Some Tools Probabilistic in Insurance and Finance

Examination Committee Members:

Pr.	Brahim	Mezerdi	U. Biskra	Chairman
Pr.	Abdelhakim	Necir	U. Biskra	Advisor
Pr.	Hocine	Fellag	U. Tizi-ouzou	Examinator
Pr.	Zaher	Mohdeb	U. Constantine	Examinator
Pr.	Abderrahmane	Yousfate	U. Sidi Bel-Abbes	Examinator
Dr.	Djamel	Meraghni	U. Biskra	Examinator

<u>September 20, 2010</u>

Dedicace

To the memory of our Prophet Muhammad Peace be upon him

To the memory of my father Ahmed ben Salem.

To my Daughters Marwa and Khouloud

Acknowledgments

I am very grateful to my advisor Professor Abdelhakim Necir for his guidance, patience and continuous encouragement. I really appreciate his comments and suggestions which have been extremely valuable for the completion of this task.
I thank Dr. Djamel Meraghni for his help in finishing this thesis.
Also, I thank the members of my Examination Committee for their comments, criticisms and suggestions.
Further thanks to the people of the Laboratory of Applied Mathematics, in particular Brahim and Djabrane.
As always my sincere thanks to my mother, my sisters and my brothers.
Huge personal thanks to my wife.
Finally, I would like to thank my family in Blida, Alger and Médéa for all their support throughout my
career.

Abstract

Extreme Value Theory (EVT) originated, in 1928, in the work of Fisher and Tippett describing the behavior of the maximum of independent and identically distributed random variables. Various applications have been implemented successfully in many fields such as: actuarial science, finance, economics, hydrology, climatology, telecommunications and engineering sciences. In this thesis, we give an overview on the extreme value theory and the different methods of estimation of the tail index and the extreme quantiles.

This thesis contains two applications of the extreme value theory, in particulary when the extreme value index is positive, which corresponds to the class of heavy-tailed distributions frequently used to model real data sets. The first is an application in the actuarial domain, to estimate one of the most popular risk measures, called the conditional tail expectation (CTE). The second contribution is an important application in fields of industrial reliability, warranty systems and telecominucations, to approximate the renewal function.

Abbreviations

rv	random variable
iid	independent identically distributed
CLT	Central Limit Theorem
EVI	Extreme Value Index
EVT	Extreme Value Theory
SLLN	slowly law number
WLLN	weakly law number
df	Distribution Function
cdf	cumulative distribution function
pdf	probability density function
GEVD	Generalized Extreme Value Distribution
GPD	Generalized Pareto Distribution
MLE	Maximum Likelihood Estimator
iff	if and only if
MSE	Mean Squared Error
RMSE	root mean squared error
edf	excess distribution function
e.g.	for example
POT	Peaks Over Threshold
PWM	Probability-Weighted Moments
mef	mean excess function
rf	renewal function
CTE	conditional tail expectation
VaR	Value-at-Risk
$TVaR$	Tail Value-at-Risk
ES	Expected Shortfall
sv	Slowly varying functions
WWW	World Wide Web

Notations

F_n Empirical cdf
$F^{\leftarrow}$ generalized inverse of F;
$\mathbb{Q}$ quantile function
$\overline{F}$ Survival Function, tail of F
I_A indicator function of set A
$\mathcal{N}(\mu, \sigma^2)$ normal or Gaussian distribution with mean μ and variance σ^2
$\mathcal{N}(0, 1)$ standard normal or standard Gaussian distribution
$\mathcal{N}_2(\omega, \Sigma)$ bivariate normal distribution with mean vector ω and covariance matrix Σ
$\mathbb{Q}_n$ empirical quantile function
$\mathbb{R}_+$ set of positive real numbers
$\mathbb{R}$ set of real numbers
$\overset{a.s.}{\to}$ almost surely convergence
$\overset{d}{\to}$ convergence in distribution
$\overset{P}{\to}$ convergence in probability
$\overset{d}{=}$ equal in distribution
$\vee$ max
$\wedge$ min
$[x]$ integer part of x
$\mathcal{D}(.)$ domain of attraction
$\mathcal{R}_\rho$ regular variation function at ∞ with index ρ
$\mathcal{R}_0$ slowly regular variation function
$(x)_+$ indicator of positiveness
$o(.)$ $f(x) = o(g(x))$ as $x \to x_0$: $\lim_{x \to x_0} f(x)/g(x) = 0$
$O(.)$ $f(x) = O(g(x))$ as $x \to x_0$: $\exists M > 0 : \lim_{x \to x_0} |f(x)/g(x)| \leq M$
$o_P(.)$ $f(x) = o_P(g(x))$ as $x \to x_0$: $P\left[\lim_{x \to x_0} f(x)/g(x) = 0\right] = 1$
$O_P(.)$ $f(x) = O_P(g(x))$ as $x \to x_0$: $\exists M > 0 : P\left[\lim_{x \to x_0} \left|\frac{f(x)}{g(x)}\right| \leq M\right] = 1$
$\sim$ $f(x) \sim g(x)$ as $x \to x_0$: $\lim_{x \to x_0} f(x)/g(x) = 1$
$\mathcal{E}(\theta)$ exponential rv with parameter θ.

Contents

List of Figures

List of Tables

Introduction

0.1 Problem description

The topic of research in this doctoral dissertation is situated within the domain of extreme value statistics. Here, in contrast to classical statistics, emphasis lies on the modelling of extreme events, i.e. events with low frequency (rare events), but mostly with high and often disastrous impact.

Statistics of extremes, help us "*to learn from almost disastrous events*". Thus, the domains of application of statistics of extremes are quite diversified. We could mention the fields of hydrology (river discharges, floods,...), environmental research and meteorology (air pollution, heavy rainfalls, windspeeds, storms,...), geology (earthquake modeling, value of diamonds,...), (re)insurance (premium calculations), finance (Value-at-Risk, conditional tail expectation,...), computer science (network traffic data, server waiting times,...), structural engineering, telecommunications and biostatistics, among others (see, for instance, Reiss and Thomas, 2001; Beirlant et al., 2004 and Castillo et al., 2005). Although it is possible to find some historical papers with applications related to extreme events, the field dates back to Gumbel, in his papers from 1935, summarized in his book (Gumbel, 1958).

Extreme value statistics deals with the estimation of quantities that are related to the tail of a distribution, or equivalently, with the analysis of the largest observations in a sample. Analysis of extreme values is firmly based on the so-called extreme value distributions and their domains of attraction. These distributions arise as the only possible limiting forms for the distributions of maxima in samples of independent and identically distributed random variables. Important topics in the field of extreme value statistics are the estimation of the tail index, extreme quantiles or small tail probabilities.

However, extreme value statistics is based on the fact that under rather mild conditions a class of functions can be considered to fit the distribution of the largest observation in a sample. From this limit theorem, it can be seen that the tail behavior of a distribution function can be completely characterized by a single real-valued parameter γ, called extreme value index or shape parameter. Based on the sign of this parameter, the domain of attraction of the extreme value distribution can be divided into three subclasses, namely Fréchet ($\gamma > 0$), Gumbel ($\gamma = 0$) and Weibull ($\gamma < 0$) classes. We will concentrate on the case when the extreme value index is positive, corresponding to heavy-tailed or Pareto-type distributions.

The analysis of heavy-tailed distributions requires special methods of estimation because of their specific features:

- slower than the exponential decay (to zero).
- violation of Cramér's condition.
- possible nonexistence of some moments.
- sparse observations at the tail domain.

For example, the central limit theorem, which states the convergence of sums of iid rv's to a Gaussian limit distribution, holds for a large variety of distributions provided that the variance of the summands is finite. If the variance is infinite, then we get the so-called stable distributions as limit distributions of the normalized sums (Lévy, 1925 and Khintchine and Lévy, 1936).

The statistical analysis of heavy-tailed distributions requires special methods that differ from classical tools due to the sparse observations in the tail domain of the distribution. For example, usually, quantiles can be estimated by means of an empirical distribution function or weighted estimators based on sample order statistics. However, high quantiles (e.g., of order 99% or 99.9%) cannot be calculated in the usual way, since the empirical distribution function is equal to 1 outside the range of the sample. Ignoring heavy tails in the data may lead to serious distortions of the estimation and errors in system control. This thesis focuses mainly on nonparametric and semiparametric methods of the statistical analysis of univariate heavy-tailed iid. rv's from samples of moderate sizes.

Traditionally, in EVT there are two main approaches with their own strength and weakness.
The first one is based on modelling the maximum of a sample (or a few largest values of a sample, called the upper order statistics) over a time period. This approach is rigorously formulated by the Fisher-Tippett theorem going back to 1928, where the block maxima (i.e., a set of maximal values selected in the blocks of data) are modelled by a generalized extreme value (GEV) distribution

$$\mathcal{H}_\gamma(x) = \exp\left\{ -\left(1 + \gamma \left(\frac{x - \mu}{\sigma}\right)\right)_+^{-1/\gamma} \right\}.$$

The second approach is based on modelling excess values of a sample over a threshold within a time period. This approach is called Peaks Over Threshold

(POT) method and has been suggested originally by hydrologists. In the POT method the values which are larger than some thresholds are modelled by the generalized Pareto distribution (GPD)

$$\mathbb{G}_{\sigma,\gamma}(x) = 1 - \left(1 + \left(\frac{\gamma - \mu}{\sigma}\right) x\right)_{+}^{-1/\gamma}.$$

The connection between both approaches is made by Theorem 3.4.5 of Pickands-Balkema-de Haan, presented in [50].

Statistics based on EVT has to use the largest (or smallest) values of a sample. They can be selected in different ways and we assume that we have iid data. The parameters in these two models (in particular the tail index $\alpha = 1/\gamma$) are estimated from a sample, using semiparametric methods (e.g., Hill's method) or parametric methods (e.g., maximum likelihood).

For practical needs, it is more important to provide estimates of high quantiles for heavy-tailed distributions. These quantities are applied to determine the values of characteristics of observed objects that may lead to rare but large losses. High quantiles indicate the $VaRs$ in finance or the thresholds of parameters in complex systems such as the Internet (e.g., the 99.9% quantile can provide the maximal threshold for the file size) or atomic power stations. In this thesis, we discuss some of the well known high quantile estimators.

The tail index is a key characteristic of heavy-tailed data. It shows the shape of the tail of the distribution without making any assumption regarding the parametric of the whole distribution. All characteristics of heavy-tailed rv's are based on this crucial parameter. In this thesis, many well-known estimators of the tail index such as Hill's, Pickands's, POT, moment, and kernel type estimators are considered.

One of the most popular actuarial risk measures is the conditional tail expectation (CTE) (see, e.g., Denuit *et al.,* 2005), which is the average amount of loss given that the loss exceeds a specified quantile. Hence, the CTE provides a measure of the capital needed due to the exposure to the loss, and thus serves as a risk measure. Not surprisingly, therefore, the CTE continues to receive increased attention in the actuarial and financial literature, where we also find numerous generalizations and extensions to the CTE (see, e.g., Landsman and Valdez, 2003; Hardy and Wirch, 2004; Cai and Li, 2005; Manistre and Hancock, 2005; Furman and Landsman, 2006; Furman and Zitikis, 2008 and the references therein).

For the purposes of warranty control, reliability analysis of technical systems, and particularly of telecommunication networks, one often needs to estimate the renewal function (RF). This function is equal to the mean number of arrivals of the relevant events before a fixed time. Usually, measurement facilities count the events of interest, for example, the number of requested and transferred Web pages, incoming or outgoing calls in consecutive time intervals of fixed length. To estimate the RF, several realizations of the counting process (e.g., observations of number of calls over several days) may be required, with further averaging inside the corresponding time interval. However, it may be that the RF has to be estimated using only one set of inter occurrence times of events. This applies particularly to warranty control or when it would be too expensive to obtain numerous observations of the process. Explicit forms of the RF are obtained only for a few inter-arrival time distributions such as the uniform, exponential, Erlang or normal (Asmussen, 1996). In this thesis, we are interested in the case where the inter-arrival times of the process are heavy-tailed.

Finally, we mention that the examples presented throughout this thesis are treated with the packages of the statistical software **R**, freely downloadable at www.r-project.cran.

0.2 Disposition

In this thesis, we study extreme value statistics from the theoretical development to its applications. This thesis gives a detailed survey of classical results and recent developments in the theory of nonparametric estimation of the tail distribution, tail index and high quantiles assuming the data come from iid. random variables with heavy-tailed distributions. Both asymptotic results (such as convergence rates of the estimates) and results for samples of moderate sizes are supported by the EVIR package which is written in the software program **R**.

The thesis bundles two papers and is partitioned into four chapters.:

Paper 1: ESTIMATING THE CONDITIONAL TAIL EXPECTATION IN THE CASE OF HEAVY-TAILED LOSSES.

Paper 2: POT-BASED ESTIMATION OF THE RENEWAL FUNCTION Of INTER-OCCURRENCE TIMES Of HEAVY-TAILED RISKS

In Chapter 1 we give an overview of the definitions and basic properties of probabilty theory, distribution of sum and limit central theorem, distribution of maxima and Fisher-Tippet theorem, domain of attraction, classes of regulary variation functions, GPD distribution and POT method, distributions of some particular order statistic. We also study the caracteristics of the intermediate order statistic and the second order condition of regular variation.

Chapter 2 is devoted to the study of different estimators of the extreme value index. In the literature, a minimal requirement is that any estimator should be consistent under the extreme value condition. It has been proved that for most known estimators a more restrictive but natural condition (the second order condition) leads to the asymptotic normality. Roughly speaking, the second order condition specifies the speed of convergence in the extreme value condition, see de Haan and Stadtmüller (1996). Tail index estimation and methods of selection of the number of largest order statistics in Hill's estimator are presented, estimators of the high quantiles, Weissman's estimate to the true value of the quantile is proved to be asymptotically normal. Finally, we present two estimators of the location parameter in the heavy-tailed case.

Chapter 3 presents some definitions of common risk measures which provide the general background for practical applications. We discuss some measures of risk and study the different axioms relative to these measures. We present some existing estimates and we elaborate a new estimation method for the conditional tail expectation function in the heavy-tailed case. We study the asymptotic behavior of the new estimator and we construct confidence bounds of this estimator. Finally illustrations and results of simulations are given with a comparison of the empirical estimator.

Finally, Chapter 4 includes the renewal process, renewal equation and its limiting theorems, in particular in cases where the second moment is infinite, this means that the inter arrival distribution are heavy-tailed with a shape parameter less than one. We interpret our contribution to represent a semiparametric estimation of the renewal function within infinite time intervals and infinite second moment, asymptotic theoretical properties are considered with results of simulation for this estimator.

Chapter 1

Introduction to Extreme Value Theory

1.1 History

An historical survey on extreme value distributions can be found in Kotz & Nadarajah (2000). The history goes back to 1709 and Nicolas Bernoulli discussing the mean of the largest distance among points lying at random on a line. The notion of the distribution of the largest value is more modern and was first introduced by von Bortkiewicz (1922). Fréchet (1927) identified one possible limit distribution for largest order statistics and, in the next year, Fisher & Tippett (1928) showed that these distributions can only be of three types. von Mises (1936) presented sufficient conditions for the convergence toward each of these types and Gnedenko (1943) gave a rigorous foundation of extreme value theory with necessary and sufficient conditions for weak convergence.
The late 1930s and 1940s were marked by a number of papers dealing with practical applications of extreme value theory, among which are Weibull (1939) studying strength of materials and Gumbel with a large number of papers culminating with his book, Gumbel (1958). As pointed out by Kotz & Nadarajah (2000), the literature in extreme value analysis is now enormous and growing very quickly. To the authors, "while this extensive literature serves as a testimony to the great vitality and applicability of the extreme value distributions and processes, it also unfortunately reflects on the lack of coordination between researchers and the inevitable duplication of results appearing in a wide range of diverse publica-

tions". This lack of unification was already mentioned by Pickands (1971) where the author links extreme value theory with the convergence of point processes. Statistical inference is developed in Pickands (1975) which justifies the use of the generalized Pareto distribution in threshold methods, commonly used by hydrologists. In parallel, methods based on several largest order statistics were proposed by Weissman (1978). These methods were developed afterward by several contributors, see Davison & Smith (1990). Galambos's (1978) monograph is one of the first reference books specifically dedicated to statistical models and treating also multivariate extremes.
It is followed by Leadbetter, Lindgren & Rootzén (1983), a key reference, in which is formally presented extreme value theory for stationary sequences.
Below are presented some main concepts in extreme value theory and its applications. These concepts can be found in the numerous reference books available, among which are Leadbetter et al. (1983), Tiago de Oliveira (1984), Resnick (1987), Embrechts, Klüppelberg & Mikosch (1997), Kotz & Nadarajah (2000), Reiss & Thomas (2001), Coles (2001), Finkenstädt & Rootzén (2004), Beirlant et al (2005) and de Haan & Ferreira (2006).

1.2 Extremal Events

Extremal events are also called "rare" events. Extremal events share three characteristics:

1. relatively rareness,
2. huge impact,
3. statistical unexpectness.

From the name of it, extremal events are extreme cases, that is, the chance of occurrence is very small. But over the last decades, we have observed several. We list some recent events with a large impact in time order. Hurricane Andrew (1992), Northridge earthquake (1994), terrorism attack (2001), disintegration of space shuttle Columbia (2003), winter blizzard of the northeast (2003). If we assume those events are in different fields and independent of each other, the probability of the occurrence of each event is small. The terminology $t-$year event has been used by hydrologists to describe how frequent a certain event will occur.

In general, suppose for every year, the probability of a certain event is p, where p is relatively small. Given that the events of each year are independent of each other, the number of the year when the event occurs follows a geometric distribution with parameter p, and thus has an expectation of $\frac{1}{p}$.

On the other words, we say Hurricane Andrew is a $30-$year event 1 means that for every year the probability of having a hurricane which is more severe than Hurricane Andrew is approximately $\frac{1}{30}$.

The enormous impact of catastrophic events on our society is deep and long. Not only we need to investigate the causes of such events and develop plans to protect against them, but also we will have to resolve the resulting huge financial loss. Financial plans have been established for the reduction of the impact of extremal events. For example, in the insurance industry of, the magnitude of property casualty risk has become a major topic of research and discussion. Traditionally, insurers buy reinsurance to hedge against catastrophic risk. For a catastrophic event causing more than 5$ billion of insured losses, the overwhelming majority is not covered by reinsurance. Table 1.1 lists 10 catastrophic events that caused the largest losses in history of the United States. Since the population and fortune grow at a relatively constant rate, we expect larger catastrophic events entering the list in the future. The task of understanding and resolving the underlying risk is still a big challenge for researchers.

year	Catastrophic Event	Estimated Losses (in billions)
1989	Hurricane Hugo	4.2
1992	Hurricane Andrew	15.5
1993	Midwest blizzard	1.8
1994	Northridge earthquake	12.5
1995	Hurricane Opal	2.1
1998	Hurricane Georges	3.0
1999	Hurricane Floyd	2.0
2001	St.Louis hailstorm	1.9
2001	Tropical Storm Allison	2.5
2001	Fire and explosion	20.3

Table 1.1: 10 Largest Catastrophe Events in US (Source: ISO's PCS)

1.3 Definitions and basic properties

We start with the common definitions in probability theory and inference statistics.

Definition 1.1 *The set* $(\Omega, \mathcal{A}, P)$ *is called the probability space, where* Ω *is the space of elementary events,* $\mathcal{A}$ *is a* $\sigma-$*algebra of subsets of* Ω*, and* P *is a probability measure on* $\mathcal{A}$*.*

Let $(\Omega, \mathcal{A})$ be some measurable space, $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be the real line with the $\sigma-$algebra $\mathcal{B}(\mathbb{R})$ of Borelian on $\mathbb{R}$.

Definition 1.2 *The real valued function* $X = X(\omega)$ *defined on* $(\Omega, \mathcal{A})$*, is called a random variable (rv), if for any* $B \subseteq \mathcal{B}(\mathbb{R}) \{\omega : X(\omega) \in B\} \subseteq \mathcal{A}$ *holds.*

Definition 1.3 *The function*

$$F_X(x) = P\{\omega : X(\omega) \leq x\}, x \in \mathbb{R},$$

is called the distribution function (cdf) of the rv X*, and survival function or tail distribution is*

$$\overline{F}_X(x) = 1 - F_X(x).$$

Definition 1.4 *Let a nonnegative real-valued function* $f(t), t \in \mathbb{R}$*, exist such that for all* $x \in \mathbb{R}$*,*

$$F_X(x) = \int_{-\infty}^{x} f(t)\, dt,$$

and survival function is

$$\overline{F}_X(x) = \int_{x}^{\infty} f(t)\, dt.$$

The function $f(t), t \in \mathbb{R}$*, is called the probability density function (pdf) of the rv* X*.*

Definition 1.5 *The rv's* $X_1, X_2, ..., X_n$ *with common cdf* $F(X_i \in B_i \subseteq \mathbb{R}$*,* B_i *is a finite set) are called independent identically distributed (iid) if, for any* $B_1, B_2, ..., B_n \in \mathcal{B}(\mathbb{R})$*,*

$$P(X_1 \in B_1, ..., X_n \in B_n) = P(X_1 \in B_1) ... P(X_n \in B_n).$$

or

$$F(x_1, x_2, ..., x_n) = F(x_1)F(x_2)...F(x_n).$$

Definition 1.6 *The empirical cdf (or sample cdf) of the sample* $(X_1, X_2, ..., X_n)$ *is defined by*

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq x\}}, x \in \mathbb{R}. \tag{1.1}$$

Definition 1.7 (Order statistics) *The order statistics pertaining to a sample* $(X_1, X_2, ..., X_n)$ *are the* X_i'*s arranged in non-decreasing order. They are denoted by* $X_{1,n}, X_{2,n}, ..., X_{n,n}$ *and for* $k = 1, 2, ..., n$*, the rv* $X_{n-k+1,n}$ *is called the* $k-th$ *upper order statistic. Order statistics satisfy* $X_{1,n} \leq X_{2,n} \leq ... \leq X_{n,n}$*. Thus*

$$X_{1,n} = \min(X_1, X_2, ..., X_n) \;\; and \; X_{n,n} = \max(X_1, X_2, ..., X_n).$$

Definition 1.8 (L-statistics) *For* $(a_1, a_2, ...a_n) \in \mathbb{R}^n$*, the statistic*

$$T_n = \sum_{i=1}^{n} a_i X_{i,n}, \tag{1.2}$$

is called L-statistic. It is a linear combinations of order statistics.

L-statistics play a major role in non-parametric statistics by providing robust estimators for location and scale parameters.For convenience in the study of the asymptotic behavior of T_n, the weights a_i are usually defined as $a_i = (1/n)J(i/(n+1))$, where J is a real application on $(0, 1)$.
The empirical cdf of the sample $(X_1, X_2, ..., X_n)$ is evaluated using order statistics as follows:

$$F_n(x) = \begin{cases} 0, & x < X_{1,n} \\ \frac{i-1}{n}, \; X_{1,n} \leq x \leq X_{n,n} \; for \; i = 1, 2, ..., n & \\ 1, & x \geq X_{n,n} \end{cases}.$$

Definition 1.9 (Quantile and tail quantile functions) *The quantile function of cdf* F *is the generalized inverse function of* F *defined by*

$$\mathbb{Q}(s) = F^{\leftarrow}(s) := \inf\{x \in \mathbb{R} : F(x) \geq x\}, 0 < s < 1. \tag{1.3}$$

with the convention that the infimum of the empty set is ∞*. In the theory of extremes, a function, denoted by* $\mathbb{U}$ *and (sometimes) called tail quantile function, is used quite often. It is defined by*

$$\mathbb{U}(t) := \mathbb{Q}(1 - 1/t) = \left(1/\overline{F}\right)^{\leftarrow}(t), 1 < t < \infty. \tag{1.4}$$

We shall see in this section, the function $\mathbb{U}$ plays a role in extreme value theory comparable to the role of the characteristic function in the theory of the stable distributions.

Definition 1.10 (Empirical quantile and tail quantile functions) *The empirical (or sample) quantile function of the sample $(X_1, X_2, ..., X_n)$ is defned by*

$$\mathbb{Q}_n(s) := \inf\{x \in \mathbb{R} : F_n(x) \geq s\}, 0 < s < 1. \tag{1.5}$$

$\mathbb{Q}_n$ may be expressed as a simple function of the order statistics pertaining to the sample $(X_1, X_2, ..., X_n)$. Namely, we have

$$\mathbb{Q}_n(s) := X_{n-k+1,n} \text{ for } \frac{n-1}{n} \leq s \leq \frac{n-i+1}{n}, \ i = 1, 2, ..., n.$$

Note that for $0 < p < 1$, $X_{[np]+1,n}$ is the sample quantile of order p.

Definition 1.11 (tail quantile process) *Let $X_{1,n} < X_{2,n} <...< X_{n,n}$ be the nth order statistics and $k = k(n)$ satisfying $k \to \infty, k/n \to 0$, as $n \to \infty$. We define the tail (empirical) quantile process to be the stochastic process $\{X_{n-[ks],n}\}_{s \geq 0}$.*

Definition 1.12 *The upper (or right) endpoint of cdf F is defined as follows:*

$$x^* = \sup\{x : F(x) < 1\}.$$

The lower (or left) endpoint of cdf F is defined as follows:

$$x_* = \inf\{x : F(x) > 0\}.$$

Clearly, x_* and x^* represent, respectively, the minimum and maximum attainable values of the rv X associated with F. Obviously, $x_* = -\infty$ in case of a lower-unbounded rv, and $x^* = +\infty$ for an upper-unbounded one.

Definition 1.13 (Brownian motion and Brownian bridge) *A stochastic process $\{\mathbb{W}_t, t \geq 0\}$ is said to be a Brownian motion or a Weiner process if*
(i) $\mathbb{W}_0 = 0$,
(ii) $(\mathbb{W}_t)$ *has independent and stationary increments,*
(iii) *for every $t > 0$, $\mathbb{W}_t$ is normally distributed with mean 0 and variance $\sigma^2 t$ for some positive constant σ.*

1. Parameter σ^2 is known as the variance parameter. When $\sigma^2 = 1$, the process is called standard Brownian motion. Since any Brownian motion $(\mathbb{W}_t)$ can always be converted to a standard Brownian motion, through the scaling $\mathbb{W}_t/\sigma$, the variance parameter is often set to 1.

2. A Brownian motion $(\mathbb{W}_t)$ is a Gaussian process with $E(\mathbb{W}_t) = 0$ and $Cov(\mathbb{W}_t, \mathbb{W}_s) = E(\mathbb{W}_t\mathbb{W}_s) = t \wedge s$. We also have that $\mathbb{W}_t - \mathbb{W}_s$ is $\mathcal{N}(0, t-s)$ for all $t \geq s \geq 0$.

3. Let $\{\mathbb{B}_t, t \geq 0\}$ be a Brownian motion. The conditional process
$$\{\mathbb{B}_t, 0 \leq t \leq 1 \mid \mathbb{B}_1 = 0\}$$
is called a **Brownian bridge**. It is a Gaussian process with mean 0 and covariance function $s(1-t)$, $s \leq t$. An alternative approach to obtaining such a process is to set
$$\mathbb{B}_t = \mathbb{W}_t - t\mathbb{W}_1, 0 \leq t \leq 1,$$
where $(\mathbb{W}_t)$ is a Brownian motion.

Definition 1.14 (Empirical process) *The empirical process of the iid sample $X_1, X_2, ..., X_n$ is defined by*
$$e_n(x) = \sqrt{n}\left(F_n(x) - F(x)\right), \ x \in \mathbb{R},$$
where F_n is the empirical distribution function F based upon $X_1, X_2, ..., X_n$.

Definition 1.15 (Uniform empirical process and uniform quantil process) *Let $U_1, U_2, ...$be a sequence of independent uniform $(0,1)$ rv. For ach integer $n \geq 1$, let*
$$G_n(s) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{U_i \leq s\}} \text{ and } H_n(s) = \inf\{t : G_n(t) \geq s\},$$
denotes the right-continuious uniform empirical distribution function and quantil function of G. Let
$$\alpha_n(s) = \sqrt{n}\left(G_n(s) - s\right), \ s \in \mathbb{R},$$
be the uniform empirical process, and
$$\beta_n(s) = \sqrt{n}\left(H_n(s) - s\right), \ s \in \mathbb{R},$$
the uniform quantile function.

Note that, the process $R_n(s) = \alpha_n(s) + \beta_n(s)$ is often called the Bahadur-Kiefer process.

Definition 1.16 (Vervaat process) *the process of Vervaat (1972) or integrated Bahadur-Kiefer process is define as follow*

$$V_n(t) = \int_0^t (\alpha_n(s) + \beta_n(s))\, ds, \ 0 \le t \le 1.$$

The following theorem descibes the limiting distribution of the Vervaat process.

Theorem 1.1 *For $t \in (0,1)$, we have*

$$2nV_n(t) \xrightarrow{d} \mathbb{B}_t^2, \ \text{as } n \to \infty,$$

where $\mathbb{B}_t$ denote a Brownian bridge on $[0,1]$.

1.4 Sums of iid rv's

Definition 1.17 *Let $X_1, X_2, ..., X_n$ a sequence of iid rv's with common cdf F. For an integer $n > 1$, define the partial sum by*

$$S_n = \sum_{i=1}^{n} X_i,$$

and the corresponding arithmetic mean $\overline{X}_n$ by

$$\overline{X}_n = S_n/n.$$

is then called sample mean or empirical mean.

1.4.1 WLLN and SLLN Theorems

Theorem 1.2 (WLLN and SLLN) *If $(X_1, X_2, ..., X_n)$ is a sample of a rv's X with cdf F and mean $\mathbb{E}(X) = \mu < \infty$, then*

$$\textbf{\textit{WLLN}}\text{:}\ \overline{X}_n \xrightarrow{P} \mu \ \text{as } n \to \infty.$$

$$\textbf{\textit{SLLN}}\text{:}\ \overline{X}_n \xrightarrow{a.s.} \mu \ \text{as } n \to \infty.$$

Applying the **SLLN** on $F_n(x)$ yields the following result.

Corollary 1.1

$$F_n(x) \overset{a.s.}{\to} F(x) \text{ as } n \to \infty, \text{ for every } x \in \mathbb{R}.$$

Theorem 1.3 (Glivenko-Cantelli)

$$\underset{x\in\mathbb{R}}{Sup} |F_n(x) - F(x)| \overset{a.s.}{\to} 0, \text{ as } n \to \infty.$$

1.4.2 Central Limit Theorem

Theorem 1.4 (CLT) *If $X_1, X_2, ..., X_n$ is a sequence of iid rv's with mean μ and finite variance σ^2, then*

$$\sqrt{n}\left(\frac{\overline{X}_n - \mu}{\sigma}\right) \overset{d}{\to} \mathcal{N}(0,1) \quad \text{as } n \to \infty.$$

Note that a necessary condition for the CLT is that the variance be finite. That is, if the finite variance assumption is dropped, the limit distribution in Theorem (1.4) is no longer normal. In the case of infinite variance, there exists a result known as the generalized CLT which states that stable laws.

1.5 Fluctuation of maxima

Suppose that $X_1, X_2, ..., X_n$ is a sequence of iid rv's with cdf F. Though we will generalize this argument later one simple way of characterizing the behaviour of extremes is by considering the behaviour of the maximum order statistic

$$M_1 = X_1, M_n = \max\{X_1, X_2, ..., X_n\}, n \geq 2.$$

Corresponding results for minima can easily be obtained from those for maxima by using the identity

$$\min\{X_1, X_2, ..., X_n\} = -\max\{-X_1, -X_2, ..., -X_n\}.$$

In principle this is trivial since

$$P(M_n \leq x) = P(X_1 \leq x, X_2 \leq x, ..., X_n \leq x) = F^n(x), x \in \mathbb{R}.$$

The difficulty arises in practice that the cdf F is unknown. Some bounds are available for the behaviour of M_n but these are too broad for practical application. This leads to an approach based on asymptotic argument. Specifically we shall see what possible limit distributions are possible for M_n as $n \to \infty$, then use this family as an approximation to the distribution of M_n for finite (but large) n.

1.5.1 The classical limit laws

The question is then what possible distributions can arise for the distribution of M_n as $n \to \infty$. Furthermore is it possible to formulate this set of limit distributions into a single class $\mathcal{H}$ which is independent of F? If so, then we can estimate the distribution of M_n directly using the family $\mathcal{H}$ without reference to F at all.
Before proceeding we must recognise that the solution to the problem as posed is trivial and degenerate. Necessarily with probability the distribution of M_n converges to the upper end point of F.
We adopt the same approach of CLT in obtaining limits of the distribution of M_n looking instead for limiting distributions of $(M_n - b_n)/a_n$ where a_n and b_n are sequences of normalizing coefficients. The solution to the range of possible limit distributions is given by Theorem(1.5). Before stating the result it is convenient to define an equivalence class of distributions the distributions $F^n(ax+b)$ such that

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = \lim_{n\to\infty} F^n(a_n x + b_n) = \mathcal{H}_\gamma(x). \tag{1.6}$$

Theorem 1.5 (Fisher-Tippet 1928) *The class of extreme value distributions is $\mathcal{H}_\gamma(ax+b)$ with $a>0$, b real, where*

$$\mathcal{H}_\gamma(x) = \exp\left\{-(1+\gamma x)^{-1/\gamma}\right\} \text{ where } 1+\gamma x > 0, \tag{1.7}$$

with γ real and for $\gamma = 0$ the right-hand side is interpreted as $\exp(-e^{-x})$.

$\mathcal{H}_\gamma(x)$ is called a standard generalized extreme value (GEVD) distribution.

Definition 1.18 *The parameter γ is called the extreme value index (EVI) and defines the shape of the tail of the r.v. X. The parameter $\alpha = 1/\gamma$ is called the tail index.*

The GEVD $\mathcal{H}_\gamma$ can be written in a more general form by replacing the argument x by $(x-\mu)/\sigma$ in the right hand side of equation (1.7), where $\mu \in \mathbb{R}$ and $\sigma > 0$ are respectively the location and scale parameters.

Remark 1.1 *Let $X_1, X_2, ..., X_n$ be iid rv's with cdf F. The cdf F is called max-stable if for some choice of constants $a_n > 0$ and b_n real,*

$$P\left\{(M_n - b_n)/a_n \leq x\right\} = P\left\{X_1 \leq x\right\}$$

for all x and $n = 1, 2, \ldots$.Note that the class of max-stable distributions is the same as the class of extreme value distributions

The parametrization in Theorem (1.5) is due to von Mises (1936) and Jenkinson (1955). This theorem is an important result in many ways. It shows that the limit distribution functions form a simple explicit one-parameter family apart from the scale and location parameters.

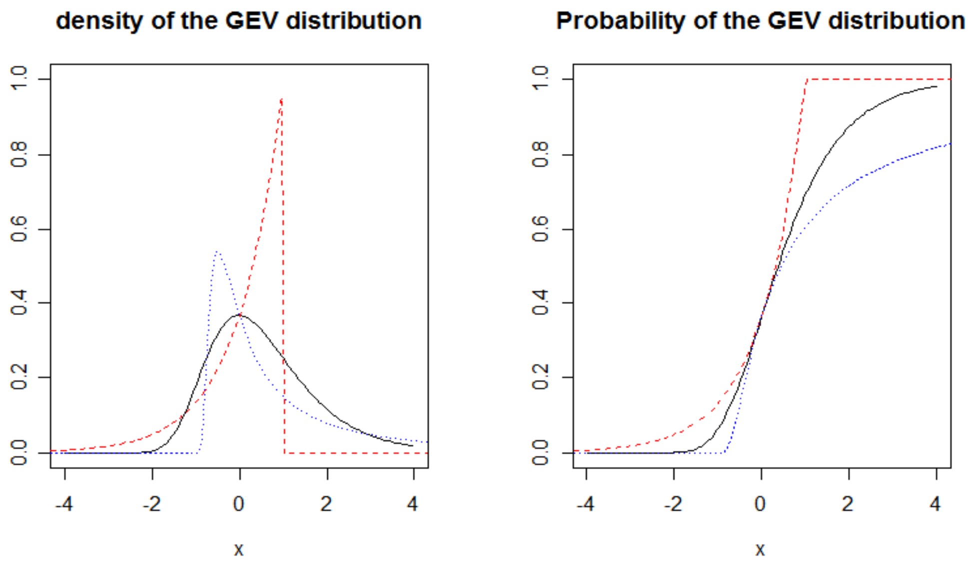

Figure 1.1: Density and Probability Plots of Generalized Extreme Distributions

Figure (1.1) illustrates the GEV family for some values of γ. Moreover, it shows that the class contains distributions with quite different features.
Let us consider the subclasses $\gamma > 0$, $\gamma = 0$, and $\gamma < 0$ separately:

a For $\gamma > 0$ use $\mathcal{H}_\gamma((x-1)/\gamma)$ and get with $\alpha = 1/\gamma > 0$,

$$\phi_\alpha(x) = \begin{cases} 0 & x \le 0 \\ \exp\left(-x^{-\alpha}\right), & x > 0, \end{cases} ;$$

This class is often called the Fréchet class of distributions (Fréchet (1927)

b The distribution function with $\gamma = 0$,

$$\mathcal{H}_0(x) = \exp\left\{-\exp\left(-x\right)\right\}, x \in \mathbb{R};$$

for all real x, is called the double-exponential or Gumbel distribution.

c For $\gamma < 0$ use $\mathcal{H}_\gamma(-(x+1)/\gamma)$ and get with $\alpha = -1/\gamma > 0$

$$\Psi_\alpha(x) = \begin{cases} \exp\{-(-x^\alpha)\}, x < 0, \\ 1 \qquad\qquad\qquad x \geq 0 \end{cases}.$$

This class is sometimes called the reverse-Weibull class of distributions.

Remark 1.2 *γ is a shape parameter determining the rate of tail decay, with:*

1. $\gamma > \mathbf{0}$: *giving the **heavy-tailed** (Fréchet) case, the class of distributions of this type includes the Pareto, Burr, Cauchy, Stable laws with exponent $\alpha < 2$, log-gamma, log-hyperbolic, log-logistic and t-distributions. The properties and the accompanied slowly varied functions of these distributions can be found in Beirlant, Teugels, and Vynckier (1996).*

2. $\gamma = \mathbf{0}$: *giving the **light-tailed** (Gumbel) case, the class of distributions of this type characterises an exponentially decreasing tail and includes the normal, exponential, gamma, and log-normal distributions.*

3. $\gamma < \mathbf{0}$: *giving the **short-tailed** (negative Weibull) case, the class of distributions with a finite upper bound, like the uniform in $(0,1)$ and beta distributions.*

Remark 1.3 *We state the explicit form of EV quantile function in both parameterizations starting with EV qfs in their $\alpha-$parameterization.*

$$\begin{array}{ll} \textit{Gumbel } \gamma = 0 & \mathcal{H}_0^{-1}(s) = -\log(-\log(s)), \\ \textit{Fréchet } \gamma > 0 & \phi_\alpha^{-1}(s) = (-\log(s))^{-1/\alpha}, \\ \textit{Weibull } \gamma < 0 & \Psi_\alpha^{-1}(s) = -(-\log(s))^{-1/\alpha}, \end{array}$$

Next, we present GEV qfs in their $\gamma-$representation:

$$\textit{for } \gamma \neq 0 : \mathcal{H}_\gamma^{-1}(s) = \left((-\log(s))^{-\gamma} - 1\right)/\gamma.$$

The Central Limit Theorem (CLT) is one of the most important tools in probability theory and statistics, stating that the normal distribution is the only distribution of iid sums under certain conditions. As an analog, EVT states the three types (standard GEV) are the only possibilities for the limiting distribution of the maxima. We have a brief comparison of these two theorem in the table (1.2).

Analog	CLT	EVT
Conditions on X_n	iid with finite second moment	iid and cdf is regularly varying
Study Object	S_n (Sum)	M_n (Maxima)
Limiting Distribution	Normal	GEV

Table 1.2: Analog between CLT and EVT

The relationship between the exceedance probability $\overline{F}(x)$ and the distribution of the maxima M_n will become clear with the following theorem.

Theorem 1.6 *For $0 < \tau < \infty$ and every sequence of real numbers u_n, $n \geq 1$, it holds for $n \to \infty$ that*

$$n\overline{F}(u_n) \to \tau \text{ iff } P\left(M_n \leq u_n\right) \to e^{-\tau}.$$

The result of theorem (1.5) leads to the following theorem.

Theorem 1.7 *For $\gamma \in \mathbb{R}$ the following statements are equivalent:*
1) There exist real constants $a_n > 0$ and b_n real such that

$$\lim_{n\to\infty} F^n\left(a_n x + b_n\right) = \mathcal{H}_\gamma(x) = \exp\left(-\left(1+\gamma x\right)^{-1/\gamma}\right), \tag{1.8}$$

for all x with $1+\gamma x > 0$.
2) There is a positive function a such that for $x > 0$,

$$\lim_{t\to\infty} \frac{\mathbb{U}(tx) - \mathbb{U}(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}, \tag{1.9}$$

where for $\gamma = 0$ the right-hand side is interpreted as $\log x$.
3) There is a positive function a such that

$$\lim_{t\to\infty} t\left(1 - F\left(a\left(t\right)x + \mathbb{U}\left(t\right)\right)\right) = \left(1+\gamma x\right)^{-1/\gamma}, \tag{1.10}$$

for all x with $1+\gamma x > 0$.
4) There exists a positive function f such that

$$\lim_{t\uparrow x^*} \frac{1 - F\left(t + xf(t)\right)}{1 - F(t)} = \left(1+\gamma x\right)^{-1/\gamma}, \tag{1.11}$$

for all x with $1+\gamma x > 0$.

Moreover **(1)** holds with $b_n := \mathbb{U}(n)$ and $a_n := a(n)$. Also **(4)** holds with $f\left(t\right) = a\left(1/\left(1-F\left(t\right)\right)\right)$.

1.5.2 Domains of Attraction

In statistical applications then we will not give any consideration to the population distribution F but will fit the GEV family $\mathcal{H}$ to series of maxima M_n. This parallels much of standard statistical inference in which tests are based on the asymptotic normality of X without concern for the parent distribution. It should be stressed however that there is substantial interest in extreme value theory for probabilistic research as well as statistical research. A major field of study in this respect has been the characterization of domains of attraction for the extreme value limits. That is given a particular limit distribution from the GEV class characterizing the set of distributions F for which the normalized M_n have the limit distribution $\mathcal{H}$. Alternatively for a given F how can the a_n and b_n be found such that a limit for M_n is obtained and what precisely is that limit? At its greatest level of generality this is a difficult question. We will give here a characterization which works with absolutely continuous distribution functions F with density f. In this case the reciprocal hazard function $\mathcal{H}$ is defined as

$$h_F(x) = \frac{1 - F(x)}{f(x)}. \tag{1.12}$$

Note that h_F is the derivative of the cumulative hazard function

$$H_F(x) = -\log(1 - F(x)) \tag{1.13}$$

Definition 1.19 *We say that the rv X and its cdf F belong to the maximum domain of attraction of $\mathcal{H}_\gamma(x)$ if equation (1.6) is fulfilled. We write $X \in \mathcal{D}(\mathcal{H}_\gamma)$ or $F \in \mathcal{D}(\mathcal{H}_\gamma)$.*

Theorem 1.8 *The df F belongs to the maximum domain of attraction of the GEV distribution $\mathcal{H}$ with the standardised sequences a_n, b_n exactly when $n \to \infty$*

$$\lim_{n\to\infty} n\left(1 - F(a_n x + b_n)\right) = -\ln \mathcal{H}_\gamma(x) \text{ for all } x \in \mathbb{R}. \tag{1.14}$$

When $\mathcal{H}_\gamma(x) = 0$ the limit is interpreted as ∞

The following theorem states a sufficient condition for belonging to a domain of attraction. The condition is called von Mises'condition.

Theorem 1.9 *Let F be a cdf and x^* its right endpoint. Suppose $F''(x)$ exists and $F'(x)$ is positive for all x in some left neighborhood of x^*. If*

$$\gamma = \lim_{x \uparrow x^*} h'(x), \tag{1.15}$$

or equivalentely,

$$-\gamma - 1 = \lim_{x \uparrow x^*} \frac{(1 - F(x))\, F''(x)}{(F'(x))^2}, \tag{1.16}$$

then F is in the domain of attraction of $\mathcal{H}_\gamma$.

Remark 1.4 *Under (1.15) we have (1.6) with $b_n = F^{-1}(1 - \frac{1}{n})$ and $a_n = h(b_n)$.*

Simpler conditions are possible for $\gamma \neq 0$.

Theorem 1.10 *1) $\gamma > 0$: Suppose $x^* = \infty$ and F' exists. If*

$$\lim_{t \to \infty} \frac{tF'(t)}{1 - F(t)} = \frac{1}{\gamma} \tag{1.17}$$

for some positive γ, then F is in the domain of attraction of $\mathcal{H}_\gamma$.
2) $\gamma < 0$: Suppose $x^* < \infty$ and F' exists for $x < x^*$. If

$$\lim_{t \uparrow x^*} \frac{(x^* - t)\, F'(t)}{1 - F(t)} = -\frac{1}{\gamma} \tag{1.18}$$

for some negative γ, then F is in the domain of attraction of $\mathcal{H}_\gamma$.

We shall establish necessary and sufficient conditions for a distribution function F to belong to the domain of attraction of $\mathcal{H}_\gamma$.

Theorem 1.11 *The cdf F is in the domain of attraction of the EVD $\mathcal{H}_\gamma$ iff*
***1)** for $\gamma > 0$: x^* is infinite and*

$$\lim_{t \to \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-1/\gamma}, \tag{1.19}$$

for all $x > 0$. This means that the function $1 - F$ is regularly varying at infinity with index $-1/\gamma$.
***2)** for $\gamma < 0$: x^* is finite and*

$$\lim_{t \downarrow 0} \frac{1 - F(x^* - tx)}{1 - F(x^* - t)} = x^{-1/\gamma}, \tag{1.20}$$

for all $x > 0$*;*
3) *for* $\gamma = 0$*:* x^* *can be finite or infinite and*

$$\lim_{t \uparrow x^*} \frac{1 - F(t + xf(t))}{1 - F(t)} = e^{-x}, \tag{1.21}$$

for all real x*, where* f *is a suitable positive function, if (1.21) holds for some* f*, then* $\int_t^{x^*} (1 - F(s))ds < \infty$ *for* $t < x^*$ *and (1.21) holds with*

$$f(t) = \frac{\int_t^{x^*} (1 - F(s))ds}{1 - F(t)}. \tag{1.22}$$

Theorem 1.12 *The cdf* F *is in the domain of attraction of the EVD* $\mathcal{H}_\gamma$ *iff*
1) *for* $\gamma > 0$*:* $F(x) < 1$ *for all* x*,* $\int_1^\infty [(1 - F(s))/s]\, ds < \infty$*, and*

$$\lim_{t \to \infty} \frac{\int_t^\infty (1 - F(s))ds/s}{1 - F(t)} = \gamma; \tag{1.23}$$

2) *for* $\gamma < 0$*: there is* $x^* < \infty$ *such that* $\int_{x^*-t}^{x^*} (1 - F(s))/(x^* - t)\, ds < \infty$ *and*

$$\lim_{t \downarrow 0} \frac{\int_{x^*-t}^{x^*} (1 - F(x))/(x^* - x)\, dx}{1 - F(x^* - t)} = -\gamma; \tag{1.24}$$

3) *for* $\gamma = 0$ *(*x^* *can be finite or infinite and):* $\int_x^{x^*} \int_t^{x^*} (1 - F(s))\, dsdt < \infty$ *and the function* h *defined by*

$$\hbar(x) := \frac{(1 - F(x)) \int_x^{x^*} \int_t^{x^*} (1 - F(s))\, dsdt}{\left(\int_x^{x^*} (1 - F(s))\, ds \right)^2}; \tag{1.25}$$

satisfies

$$\lim_{t \uparrow x^*} h(t) = 1.$$

Next we show how to find the normalizing constants $a_n > 0$ and b_n in the basic limit relation (1.6).

Corollary 1.2 *If the cdf* F *is in the domain of attraction of* $\mathcal{H}_\gamma$*, then*
1) *for* $\gamma > 0$*:*

$$\lim_{n \to \infty} F^n(a_n x) = \exp\left(-x^{-1/\gamma}\right) \tag{1.26}$$

holds for $x > 0$ *with* $a_n := \mathbb{U}(n)$*;*
2) *for* $\gamma < 0$*:*

$$\lim_{n \to \infty} F^n(a_n x + x^*) = \exp\left(-(-x)^{-1/\gamma}\right) \tag{1.27}$$

holds for $x < 0$ *with* $a_n := x^* - \mathbb{U}(n)$;
3) for $\gamma = 0$*:*

$$\lim_{n\to\infty} F^n(a_n x + b_n) = \exp\left(-e^{-x}\right) \tag{1.28}$$

holds for all x *with* $a_n := f(\mathbb{U}(n)), b_n := \mathbb{U}(n)$, *and* f *defined by equation (1.22).*

We reformulate Theorem 1.12 in a seemingly more uniform way.

Theorem 1.13 *The cdf* F *is in the domain of attraction of the EVD* $\mathcal{H}_\gamma$ *iff for some positive function* f,

$$\lim_{t\uparrow x^*} \frac{1 - F(t + xf(t))}{1 - F(t)} = (1 + \gamma x)^{-1/\gamma}, \tag{1.29}$$

for all x *with* $1 + \gamma x > 0$*. If (1.29) holds for some* $f > 0$*, then it also holds with*

$$f(t) = \begin{cases} \gamma t & \gamma > 0, \\ -\gamma(x^* - t) & \gamma < 0, \\ \int_t^{x^*} (1 - F(s))ds / (1 - F(t)), & \gamma = 0. \end{cases} .$$

1.5.3 Functions of Regular Variation

In this section, we treat a class of functions that shows up in a vast number of applications in the whole of mathematics and that is intimately related to the class of power functions. We first give some generalities. Then we state a number of fundamental properties.We continue with properties that are particularly important for us.

Definition 1.20 *Let* f *be an ultimately positive and measurable function on* $\mathbb{R}_+$*. We will say that* f *is regularly varying if and only if there exists a real constant* ρ *for which*

$$\lim_{x\to\infty} \frac{f(tx)}{f(x)} = t^\rho \text{ for all } t > 0.$$

We write $f \in \mathcal{R}_\rho$ *and we call* ρ *the index of regular variation. In the case* $\rho = 0$*, the function will be called slowly varying (s.v.) or of slow variation. We will reserve the symbol for such functions. The class of all regularly varying functions is denoted by* $\mathcal{R}$*.*

Example 1.1 *Consider the following examples to fix ideas.*
(1) *The canonical regularly varying functions are power functions* $\mathbb{U}(x) = x^\rho \in$

$\mathcal{R}_\rho$ for $x > 0$ and $\rho \in \mathbb{R}$.
(2) The canonical slowly varying function is $\log(1+x) \in \mathcal{R}_0$.
(3) If $\lim_{x\to\infty} \mathbb{U}(x) = \mathbb{U}(\infty)$ exists and is finite, then $\mathbb{U} \in \mathcal{R}_0$. So for instance, every probability distribution function on $\mathbb{R}_+$ is slowly varying at ∞.
(4) If X is a Pareto random variable with distribution F, so that

$$1 - F(x) =: \overline{F}(x) = x^{-\alpha}, x \geq 1, \alpha > 0,$$

then $\overline{F} \in \mathcal{R}_{-\alpha}$.

Proposition 1.1 *The function $f \in \mathcal{R}_\rho$ if and only if $f(x) = x^\rho \ell(x)$, where $\ell \in \mathcal{R}_0$.*

The following theorem is used to restate the definition of regular variation and to introduce a new concept called Π-variation.

Theorem 1.14 *If $f : (0,\infty) \to \mathbb{R}_+$ measurable function such that $\lim_{t\to\infty} \frac{f(tx)-f(t)}{a(t)}$, (where $x > 0$ and a is a positive function), exists and is not constant, then*

$$\lim_{t\to\infty} \frac{f(tx) - f(t)}{a(t)} = c\frac{x^\rho - 1}{\rho}, x > 0$$

for some $\rho \in \mathbb{R}$ and $c \neq 0$, with the convention that the right hand side reads $c \log x$ if $\rho = 0$.

The case $\rho = 0$ defined the so-called Π-varying functions.

Definition 1.21 (Π-varying function) *A positive, measurable function f on $(0,1)$ is Π-varying at infinity with auxiliary function $a > 0$; notation $f \in \Pi$, if*

$$\lim_{t\to\infty} \frac{f(tx) - f(t)}{a(t)} = \log x, x > 0.$$

f is said to be Π-varying at 0, notation $f \in \Pi^0$, if $f(1/x)$ is Π-varying at infinity.

Some of the basic properties of regularly varying functions are given in the following theorems.

Proposition 1.2 *Slowly varying functions have the following properties:*
i) $\mathcal{R}_0$ is closed under addition, multiplication and division.
ii) If ℓ is s.v. then ℓ^α is s.v. for all $\alpha \in \mathbb{R}$.
iii) If $\rho \in \mathbb{R}$, then $f \in \mathcal{R}_\rho$ iff $f^{-1} \in \mathcal{R}_{-\rho}$.

Mathematically, the two most important results about functions in $\mathcal{R}_0$ are given in the following theorem due to Karamata.

Theorem 1.15 ***(i) Uniform Convergence Theorem.*** *If $\ell \in \mathcal{R}_0$, then the convergence*

$$\lim_{x\to\infty} \frac{\ell(tx)}{\ell(x)} = 1$$

is uniform for $t \in [a, b]$ where $0 < a < b < \infty$.
(ii) Representation Theorem. *$\ell \in \mathcal{R}_\rho$ if and only if it can be represented in the form*

$$\ell(x) = c(x) \exp\left\{ \int_1^x \frac{\epsilon(u)}{u} du \right\}$$

where $c(x) \to c \in (0, \infty)$ and $\epsilon(x) \to 0$ as $x \to \infty$.

Proposition 1.3 *Let ℓ be slowly varying. Then*
1)

$$\lim_{x\to\infty} \frac{\log \ell(x)}{\log x} = 0.$$

2) *For each $\delta > 0$ there exists a x_δ so that for all constants $A > 0$ and $x > x_\delta$*

$$Ax^{-\delta} < \ell(x) < Ax^{\delta}.$$

3)*If $\ell \in \mathcal{R}_\rho$ with $\rho > 0$, then $f(x) \to \infty$, while for $\rho < 0$, $f(x) \to 0$ as $x \uparrow \infty$.*
4) Potter Bounds. *Given $A > 1$ and $\delta > 0$ there exists a constant $x_o(A, \delta)$ such that*

$$\frac{\ell(y)}{\ell(x)} \le A \max\left\{ \left(\frac{y}{x}\right)^{\delta}, \left(\frac{y}{x}\right)^{-\delta} \right\}, \quad x, y \ge x_o.$$

The following result, known as Karamata's theorem, says that one can take slowly varying functions out of integrals.

Theorem 1.16 *Let $\ell \in \mathcal{R}_0$ be locally bounded in $[A, \infty)$ for some $A > 0$. Then*
1) *For $\alpha > -1$,*

$$\int_A^x t^{\alpha} \ell(t)\, dt \sim (\alpha + 1)^{-1} x^{\alpha+1} \ell(x) \quad as\ x \to \infty,$$

2) *For $\alpha < -1$,*

$$\int_x^{\infty} t^{\alpha} \ell(t)\, dt \sim -(\alpha + 1)^{-1} x^{\alpha+1} \ell(x) \quad as\ x \to \infty,$$

Corollary 1.3 *Let $f \in \mathcal{R}_\rho$ for some $\rho \neq -1$, be locally bounded in $[a, \infty)$ for some $a > 0$. Then as $x \to \infty$,*
1) *For $\alpha > -1$,*

$$\frac{\int_a^x f(t)\,dt}{xf(x)} \to \frac{1}{\rho+1},$$

2) *For $\alpha < -1$,*

$$\frac{\int_x^\infty f(t)\,dt}{xf(x)} \to -\frac{1}{\rho+1}.$$

The definition of heavy-tailed distributions.

Definition 1.22 *A cdf $F(x)$ (or the rv X) is called heavy-tailed if its tail $\overline{F}(x) = 1 - F(x) > 0, x \geq 0$, satisfies,*

$$\lim_{x\to\infty} P(X > x + y \mid X > x) = \lim_{x\to\infty} \frac{\overline{F}(x+y)}{\overline{F}(x)} = 1, \text{ for all } y \geq 0, .$$

Definition 1.23 *The cdf $F(x)$ (or the rv X), defined on $(0, \infty)$, is called subexponential ($F \in \mathcal{S}$ ($X \in \mathcal{S}$)), if*

$$P(S_n > x) \sim nP(S_n > x) \sim P(M_n > x) \text{ as } x \to \infty.$$

The class of heavy-tailed distributions comprises the subexponential class of distributions ($\mathcal{S}$) and its subset, that is, distributions with regularly varying tails.

Definition 1.24 *The cdf $F(x)$ (or rv X) is called a regularly varying distribution at infinity of index $\alpha = 1/\gamma$, $\gamma > 0$ ($X \in \mathcal{R}_{-1/\gamma}$), if*

$$P(X > x) = x^{-1/\gamma}\ell(x), \forall x > 0,$$

where $\ell(x)$ is called a slowly varying function ($\ell(x) \in \mathcal{R}_0$).

Proposition 1.4 (Regular variation for distribution tails) *Assume that F is a continuous cdf (with pdf f) such that $F(x) < 1$ for all $x \geq 0$:*
a) *If $\lim_{x\to\infty} xf(x)/\overline{F}(x) = \rho > 0$, then $f \in \mathcal{R}_{-1-\rho}$ and consequently $\overline{F} \in \mathcal{R}_{-\rho}$.*
b) *If $f \in \mathcal{R}_{-1-\rho}(\rho > 0)$, then $\lim_{x\to\infty} xf(x)/\overline{F}(x) = \rho$.*
c) *If X is a non-negative r.v. with distribution tail $\overline{F} \in \mathcal{R}_{-\rho}$ $(\rho > 0)$, then*

$$EX^p < \infty \text{ if } p < \rho,$$
$$EX^p = \infty \text{ if } p > \rho.$$

d) *If* $\overline{F} \in \mathcal{R}_{-\rho}$ $(\rho > 0)$ *,then for* $\nu \geq \rho$

$$\lim_{x\to\infty} \frac{x^{\nu}\overline{F}(x)}{\int_0^x t^{\nu} dF(t)} = \frac{\nu - \rho}{\rho}.$$

The converse also holds when $\nu > \rho$*. If* $\rho = \nu$ *one can only conclude that* $\overline{F}(x) = o(x^{-\rho}\ell(x))$ *for some* $\ell \in \mathcal{R}_0$*.*

In practice, a tail function $\overline{F}(x)$ is often fitted by the generalized Pareto distribution. The latter is based on Pickands' theorem (Pickands, 1975).

1.5.4 Generalized Pareto Distributions (GPD)

As the GEV in the previous section describes the limit distribution of normalized maxima, the Generalized Pareto Distribution (GPD) is the limit distribution of scaled excess of high thresholds. The main connection is in the following theorem.

Theorem 1.17 *Suppose* $X_1, X_2, ..., X_n$ *are iid rv's with cdf* F*. As in Theorem(1.5),*

$$\mathcal{H}_{\gamma}(x) = \exp\left\{-[1+\gamma x]^{-1/\gamma}\right\} \text{ where } 1+\gamma x > 0,$$

is the limit distribution of the maxima M_n*.Then for a large enough threshold* u*, the conditional distribution function of* $Y = (X - u \mid X > u)$*, is approximately*

$$P\{X - u < x \mid X > u\} \sim \mathbb{G}_{\gamma}(x)$$

where

$$\mathbb{G}_{\gamma}(x) = \begin{cases} 1-(1+\gamma x)^{-1/\gamma} & \text{if } \gamma \neq 0 \\ 1-\exp(-x), & \text{if } \gamma = 0 \end{cases}, \tag{1.30}$$

with

$$\begin{array}{c} x \geq 0 \quad \text{if} \quad \gamma \geq 0 \\ 0 \leq x \leq -1/\gamma \text{ if } \gamma < 0. \end{array}$$

Then, $\mathbb{G}_{\gamma}$ *is called standard GPD and* γ *its shape parameter.*

The family of distributions defined by equation(1.30) is called the General Pareto Distribution family (GPD). For a fixed high threshold u. The GPD distribution has many good properties, (see, for instance, Embrechts, Klüppelberg and Mikosch, 1997, Section 3.4, and Reiss and Thomas (2007), Section 1.4, for more details). For the purpose of this paper, we particularly compute the expectation of the GPD.

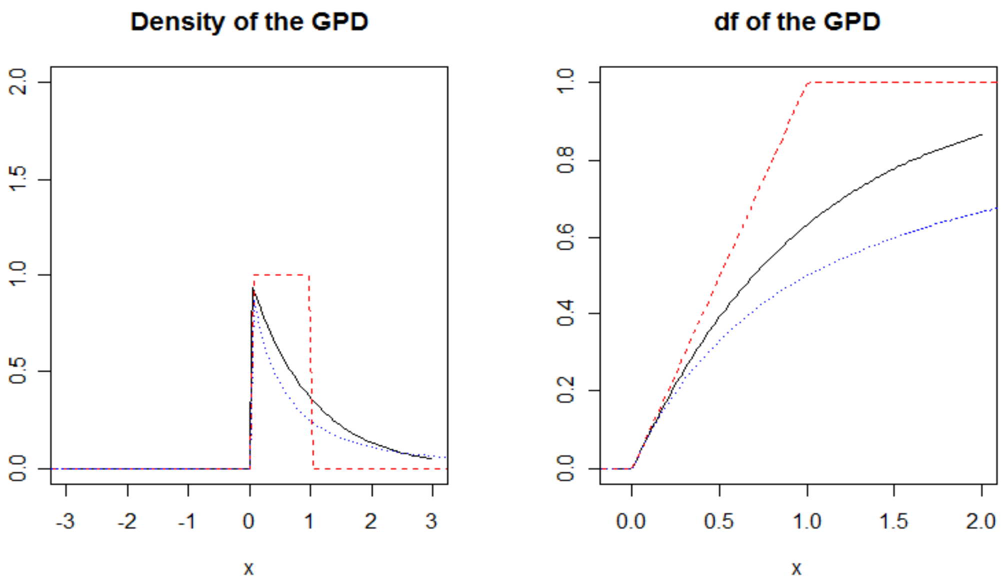

Figure 1.2: Density and distribution Plots of the GPD

Definition 1.25 *Let X be a rv with cdf F, for $u < x^*$*

$$\begin{aligned} F_u(x) &= P\{X - u \leq x \mid X > u\} \\ &= 1 - \frac{\overline{F}(x+u)}{\overline{F}(u)}, \text{ for } x > 0, \end{aligned} \tag{1.31}$$

is the excess distribution (edf) of rv X over the threshold u.
Define

$$e(u) = E(X - u \mid X > u) \tag{1.32}$$

and $e(u)$ is called the mean excess function (mef) of rv X.

Some useful forms of the mef are given in the following proposition (see Example 3.4.8 of [46]):

Proposition 1.5 (Useful forms of emf) ***a)***

$$e(u) = \frac{1}{\overline{F}(u)} \int_u^{x^*} \overline{F}(x)\, dx, u < x^*.$$

b) *If $\overline{F} \in \mathcal{R}_{-\gamma}$ for some $\gamma > 1$, then*

$$e(u) = u/(\gamma - 1) \text{ as } u \to \infty.$$

c) *If* F *is continuous and* B *is the left limit of its support, then*

$$\overline{F}(x) = \frac{e(B)}{e(x)} \exp\left\{-\int_B^x \frac{du}{e(u)}\right\}, x > B.$$

The following proposition consists of some probabilistic properties of the GPD.

Proposition 1.6 (Properties of the GPD) *(a) If* Y *is a rv having a GPD with parameters* $\gamma \in \mathbb{R}$ *and* $\sigma > 0$*. Then* $EY < \infty$ *iff* $\gamma < 1$*. In this case, the mef is linear. More precisely, for* $u < x^*$

$$e(u) = \frac{\sigma + \gamma u}{1 - \gamma}, \sigma + \gamma u > 0,$$

(b) If $(Y_n)_{n \geq 1}$ *is an iid sequence having a GPD with parameters* $\gamma \in \mathbb{R}$ *and* $\sigma > 0$ *and if for some* $\lambda > 0$*,* N *is* $P(\lambda)$ *independent of that sequence, then*

$$P(Y_{N,N} \leq y) = \exp\left\{-\lambda\left(1 + \frac{y\gamma}{\sigma}\right)^{-1/\gamma}\right\} = \mathcal{H}_{\gamma,\mu,\delta}$$

where $\mu = \frac{\sigma}{\gamma}(\lambda^\gamma - 1)$ *and* $\delta = \sigma\lambda^\gamma$*.*

The following result, due independently to Balkema and de Haan [8] and Pickands [108], is one of the most useful concepts in the statistical methods for extremes. It is known as GPD approximation or (as hydrologists call it) POT method and says that for a large threshold u, the cdf F_u is likely to be well approximated by a GPD with shape parameter γ (equal to the tail index of cdf F) and scale parameter $\sigma = \sigma(u)$.

Theorem 1.18 (GPD approximation) *For every* $\gamma \in \mathbb{R}$*.* $F \in \mathcal{D}(\mathcal{H}_\gamma)$ *iff*

$$\lim_{u \to x^*} \sup_{0 < y < u - x^*} \left|F_u(y) - \mathbb{G}_{\gamma,\sigma(u)}(y)\right| = 0, \tag{1.33}$$

for some positive function σ*.*

Results of both Proposition 1.6 and Theorem(1.18) are to be found as parts of Theorem 3.4.13 in [46], which may be consulted for the proofs.

1.6 Extreme and Intermediate Order Statistics

The subject of order statistics deals with the properties and applications of these ordered rv's and of functions involving them. Examples are the minimum $X_{1,n}$, the maximum, $X_{n,n}$, the range $W = X_{n,n} - X_{1,n}$, the extreme deviate (from the sample mean) $X_{n,n} - \overline{X}$. All these statistics have important applications. The extremes are the critical values used in engineering, physics, medicine, etc, see, for example, Castillo and Hadi (1997), and in the statistical study of floods and droughts, in problems of breaking strength and fatigue failure, and in auction theory (Krishna, 2002).

1.6.1 Distributions of Order Statistics

We discuss the distributions and density of order statistics when $(X_I, X_2, ..., X_n)$ is an iid sample of known size n drawn from a common cdf $F(x)$.

Distribution of a Single Order Statistic

Let $F_{(r)}(x)(r = 1, ..., n)$ denote the cdf of the rth order statistic $X_{r,n}$. Then the cdf of the largest order statistic $X_{n,n}$ is given by

$$F_{(n)}(x) = P(X_{n,n} \leq x) = P(\text{all } X_i \leq x) = F^n(x).$$

Likewise we have

$$F_{(1)}(x) = 1 - [1 - F(x)]^n .$$

These are important special cases of the general result for $F_{(r)}(x)$:

$$\begin{aligned} F_{(r)}(x) &= P(X_{r,n} \leq x) \\ &= Pr\{\text{at least } r \text{ of the } X_i \text{ are less than or equal to } x \} \\ &= \sum_{i=r}^{n} C_n^i F^i(x) \left[1 - F(x)\right]^{n-i} , \end{aligned}$$

or, the alternative form:

$$F_{(r)}(x) = F^r(x) \sum_{j=0}^{n-r} C_{r+j-1}^{r-1} \left[1 - F(x)\right]^j .$$

We shall now assume that X_i is continuous with probability density function (pdf) $f(x) = F'(x)$. If $f_{(r)}(x)$ denotes the pdf of $X_{(r)}$ we have

$$\begin{aligned} f_{(r)}(x) &= \frac{1}{B(r, n-r+1)} \frac{d}{dx} \int_0^{F(x)} t^{r-1} (1-t)^{n-r} dt \\ &= \frac{1}{B(r, n-r+1)} F^{r-1}(x) [1 - F(x)]^{n-r} f(x), \end{aligned}$$

where $\mathcal{B}$ is the Beta function $\mathcal{B}(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt, a > 0, b > 0.$

Proposition 1.7 *The random interval* $\left(X_{(i)}, X_{(j)}\right), i < j$, *include the quantile* $x_p, p \in (0, 1)$ *with probability*

$$\begin{aligned} \mathbb{P}\left(X_{(i)} \leq x_p \leq X_{(j)}\right) &= F_{(i)}(x_p) - F_{(j)}(x_p) \\ &= \sum_{k=i}^{j} C_n^k p^k (1-p)^{n-k}. \end{aligned}$$

Joint Distribution of two Order Statistics

The joint density function of $X_{(r)}$ and $X_{(s)}$ $(l < r < s < n)$ is denoted by $f_{(r)(s)}(x, y)$, with $x \leq y$, such that

$$f_{(r)(s)}(x, y) = \frac{n!}{(r-1)!(s-r-1)!(n-s)!} F^{r-1}(x) [F(y) - F(x)]^{s-r-1} f(x).$$

By integration, we obtain the joint cdf $F_{(r)(s)}(x, y)$ of $X_{(r)}$ and $X_{(s)}$. We have for $x \leq y$

$$F_{(r)(s)}(x, y) = \sum_{j=s}^{n} \sum_{i=r}^{j} \frac{n!}{i!(j-i)!(n-j)!} F^i(x) [F(y) - F(x)]^{j-i}.$$

Distribution of the range

From the joint pdf of k order statistics we can by standard transformation methods derive the pdf of any well-behaved function of the order statistics. For example, to find the pdf of $W_{rs} = X_{(s)} - X_{(r)}$ we set $w_{rs} = y - x$, we have

$$\begin{aligned} f_{W_{rs}}(w_{rs}) = C_{rs} \int_{-\infty}^{+\infty} & F^{r-1}(x) f(x) [F(x + w_{rs}) - F(x)]^{s-r-1} \\ & \times f(x + w_{rs}) [1 - F(x + w_{rs})]^{n-s} dx. \end{aligned}$$

Of special interest is the case $r = 1$, $s = n$, when W_{rs} becomes the range of W and

$$f_W(w) = n(n-1)\int_{-\infty}^{+\infty} f(x)\left[F(x+w) - F(x)\right]^{n-2} f(x+w)dx.$$

The cdf of W is somewhat simpler. On interchanging the order of integration we have

$$F_W(w) = n\int_{-\infty}^{+\infty} f(x)\left[F(x+w) - F(x)\right]^{n-1} dx.$$

Uniform Order Statistics and Simulation

For a rv X with arbitrary cdf F, let

$$F^{-1}(u) = \inf\{x : F(x) \geq u\}, 0 < u < 1.$$

Then, with U is standard uniform, we have

$$(X_{1,n}, X_{2,n}, ..., X_{n,n}) \stackrel{d}{=} \left(F^{-1}(U_{1,n}), F^{-1}(U_{2,n}), ..., F^{-1}(U_{n,n})\right).$$

Order Statistic and Markov Property

When the cdf F is continuous, the ordered sample $(X_{1,n}, X_{2,n}, ..., X_{n,n})$ forms a Markov Chain. In other words, we have for $i = 2, ..., n$

$$P(X_{i,n} \leq x \mid X_{1,n} = x_1, ..., X_{i-1,n} = x_{i-1}) = P(X_{i,n} \leq x \mid X_{i-1,n} = x_{i-1}).$$

1.6.2 Extreme Order Statistics and Poisson Point Processes

Let us start to derive the result for the exponential distribution. Suppose $E_1, E_2, ..., E_n$ are iid standard exponential and $E_{1,n}, E_{2,n}, ..., E_{n,n}$ are the nth order statistics. By Renyi's (1953) representation we have for fixed $k \leq n$,

$$E_{1,n}, E_{2,n}, ..., E_{n,n} \stackrel{d}{=} \left(\frac{E_1^*}{n}, \frac{E_1^*}{n} + \frac{E_2^*}{n-1}, ..., \frac{E_1^*}{n} + ... + \frac{E_k^*}{n-k+1}\right)$$

with $E_1^*, E_2^*, ..., E_k^*$ are iid standard exponential. Hence

$$n(E_{1,n}, E_{2,n}, ..., E_{n,n}) \stackrel{d}{\to} (E_1^*, E_1^* + E_2^*, ..., E_1^* + E_2^* + ... + E_k^*). \qquad (1.34)$$

This suggests that, the point process of normalized lower extreme-order statistics converges to a homogeneous Poisson process.
Next we generalize the result (1.34) to the entire domain of attraction, and as usual, we formulate it for upper order statistics rather than lower ones.

Theorem 1.19 *Let $X_1, X_2, ..., X_n$ be iid rv's with cdf F. Suppose F is in the DA of $\mathcal{H}_\gamma$ for some $\gamma \in \mathbb{R}$. Let $X_{1,n}, X_{2,n}, ..., X_{n,n}$ be the nth order statistics. Then with the normalizing constants $a_n > 0$ and b_n from (1.6) and fixed $k \in \mathbb{N}$.*

$$\left(\frac{X_{n,n}-b_n}{a_n}, \frac{X_{n-1,n}-b_n}{a_n}, ..., \frac{X_{n-k,n}-b_n}{a_n}\right)$$

converges in distribution to

$$\left(\frac{(E_1^*)^{-\gamma}-1}{\gamma}, \frac{(E_1^*+E_2^*)^{-\gamma}-1}{\gamma}, ..., \frac{(E_1^*+E_2^*+...+E_k^*)^{-\gamma}-1}{\gamma}\right)$$

where $E_1^*, E_2^*, ..., E_k^*$ are iid standard exponential.

Under the conditions of Theorem (1.19), consider the random collection

$$\left\{\left(\frac{i}{n}, \frac{X_i-b_n}{a_n}\right)\right\}_{i=1}^{\infty}$$

of points in $\mathbb{R}_+ \times \mathbb{R}$ and define a point process (random measure) N_n as follows: for each Borel set $B \subset \mathbb{R}_+ \times \mathbb{R}$,

$$N_n(B) = \sum_{i=1}^{\infty} 1_{\left\{\left(\frac{i}{n}, \frac{X_i-b_n}{a_n}\right)\in B\right\}}.$$

Moreover, consider a Poisson point process N on $\mathbb{R}_+ \times (x_*, x^*]$, where x_* and , $x^*$are the lower and upper endpoints of the cdf $\mathcal{H}_\gamma$, with mean measure v given by

$$v([a,b]\times[c,d]) = (b-a)\left[(1+\gamma c)^{-1/\gamma} - (1+\gamma d)^{-1/\gamma}\right],$$

with $0 < a < b$, $x_* < c < d < x^*$. The following limit relation holds.

Theorem 1.20 *The sequence of point processes N_n converges in distribution to the Poisson point process N, i.e., for any Borel sets $B_1, ..., B_r \subset \mathbb{R}_+ \times (x_*, x^*]$ with $v(\partial B_i) = 0$ for $i = 1, 2, ..., r$, where ∂B is the boundary of B, then*

$$(N_n(B_1), ..., N_n(B_r)) \xrightarrow{d} (N(B_1), ..., N(B_r)).$$

1.6.3 Intermediate Order Statistics

In the previous section we studied the asymptotic behavior of order statistics, that is, $X_{n-k,n}$ when $n \to \infty$ and k is fixed, along with an approximation by a Poisson point process. One can also consider $X_{n-k,n}$ with $k = k(n) \to \infty$ as $n \to \infty$. A commonly considered case is $k(n)/n \to p \in (0, 1)$ (the so-called central order statistics, see, e.g., Arnold, Balakrishnan, and Nagaraja (1992)). The normal distribution is then an appropriate limit distribution, and in fact, the stochastic process $X_{[ns],n}$, for some $0 < s < 1$, properly normalized, can be approximated by a Brownian bridge. But there is a case in between these two. Consider the order statistics $X_{n-k,n}$ with $n \to \infty$, $k = k(n) \to \infty$ and $k(n)/n \to 0$ as $n \to \infty$.
Those are called intermediate order statistics. Their behavior can be connected with extreme value theory, and the stochastic process $X_{n-[ks],n}$, properly normalized, can be approximated by Brownian motions, as we shall see.
The following result is given by (Smirnov, 1949, 1967; Falk, 1989) it shows that there is a connection between intermediate order statistics and extreme value theory.

Theorem 1.21 *Suppose von Mises condition for the DA of an evd $\mathcal{H}_\gamma$ holds section (1.5.2). Then, if $k = k(n) \to \infty$ and $k(n)/n \to 0$ as $n \to \infty$,*

$$\sqrt{k}\frac{X_{n-k,n} - \mathbb{U}\,(n/k)}{(n/k)\,\mathbb{U}'\,(n/k)}$$

is asymptotically standard normal.

In view of applications later on we state the following immediate corollary:

Corollary 1.4 *For $F_Y\,(y) = 1 - 1/y, y \geq 1$ as $n \to \infty$, $k \to \infty$, $k/n \to 0$,*

$$\sqrt{k}\left(\frac{k}{n}Y_{n-k,n} - 1\right) \tag{1.35}$$

is asymptotically standard normal.

So we see that the normal distribution is a natural limit distribution for intermediate order statistics. As in the case of extreme order statistics, where we made the connection with point processes, we want to put the present limit result in a wider framework, which in this case will be convergence toward a Brownian

motion. However, for this result we need more than just the domain of attraction condition. One can consider the domain of attraction condition as a special kind of asymptotic expansion of $\mathbb{U}$ near infinity. For the approximation by Brownian motion, as well as for many statistical results as we shall see later on, it is very useful to have a higher-order expansion.
We call this the second-order condition. This condition will be discussed in the next section.

1.6.4 Second-Order Condition

We are going to develop a second-order condition, we are begin by this proposition.

Proposition 1.8 (First Order Regular Variation Condition) *The following assertions are equivalents:*
(a) F heavy-tailed

$$F \in \mathcal{D}\left(\Phi_{1/\gamma}\right), \gamma > 0.$$

(b) $\overline{F}$ regularly varying at ∞ with index $-1/\gamma$

$$\lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = x^{-1/\gamma}, x > 0. \tag{1.36}$$

(c) $\mathbb{Q}(1-s)$ regularly varying at 0 with index $-\gamma$

$$\lim_{s\to 0} \frac{\mathbb{Q}(1-sx)}{\mathbb{Q}(1-s)} = x^{-\gamma}, x > 0. \tag{1.37}$$

(d) $\mathbb{U}$ regularly varying at ∞ with index γ

$$\lim_{t\to\infty} \frac{\mathbb{U}(tx)}{\mathbb{U}(t)} = x^{\gamma}, x > 0. \tag{1.38}$$

Definition 1.26 (Second Order Regular Variation Assumption) *We say that (the tail of) $F \in \mathcal{D}\left(\Phi_{1/\gamma}\right), \gamma > 0$, is second order regularly varying at infinity if it satisfies one of the following (equivalent) conditions :*
(a) *There exist some parameter $\rho \leq 0$ and a function A^*, tending to 0 and not changing sign near infinity, such that for all $x > 0$*

$$\lim_{t\to\infty} \frac{(1-F(tx))\,/\,(1-F(t)) - x^{-1/\gamma}}{A^*(t)} = x^{-1/\gamma}\frac{x^{\rho}-1}{\rho}. \tag{1.39}$$

(b) *There exist some parameter $\rho \leq 0$ and a function A^{**}, tending to 0 and not changing sign near 0, such that for all $x > 0$*

$$\lim_{s\to 0} \frac{\mathbb{Q}(1-sx)/\mathbb{Q}(1-s) - x^{-\gamma}}{A^{**}(s)} = x^{-\gamma}\frac{x^{\rho}-1}{\rho}. \tag{1.40}$$

(c) *There exist some parameter $\rho \leq 0$ and a function A, tending to 0 and not changing sign near infinity, such that for all $x > 0$*

$$\lim_{t\to\infty} \frac{\mathbb{U}(tx)/\mathbb{U}(t) - x^{\gamma}}{A(t)} = x^{\gamma}\frac{x^{\rho}-1}{\rho}. \tag{1.41}$$

If $\rho = 0$,interpret $\frac{x^{\rho}-1}{\rho}$ as $\log x$.

Note that A, A^{*} and A^{**} are regularly varying functions with $A^{*}(t) = A(1/(1 - F(t)))$ and $A^{**}(s) = A(1/s)$. Their role is to control the speed of convergence in (1.41), (1.39) and (1.40) respectively. More precisely we have $A \in \mathcal{R}_{\rho}$, $A^{*} \in \mathcal{R}_{\rho/\gamma}$ and $A^{**} \in \mathcal{R}^{0}_{-\rho}$. The relations above may be reformulated as respectively

$$\lim_{t\to\infty} \frac{\log(1-F(tx)) - \log(1-F(t)) + (1/\gamma)\log x}{A^{*}(t)} = \frac{x^{\rho}-1}{\rho}. \tag{1.42}$$

$$\lim_{s\to 0} \frac{\log \mathbb{Q}(1-sx) - \log \mathbb{Q}(1-s) + \gamma \log x}{A^{**}(s)} = \frac{x^{\rho}-1}{\rho}. \tag{1.43}$$

$$\lim_{t\to\infty} \frac{\log \mathbb{U}(tx) - \log \mathbb{U}(t) - \gamma \log x}{A(t)} = \frac{x^{\rho}-1}{\rho}. \tag{1.44}$$

1.6.5 Hall's class of cdf's

As an example of heavy-tailed distributions satisfying the second order assumption, we have the so called and frequently used Hall's model introduced in [74] which is a class of cdf's

$$F(x) = 1 - cx^{-1/\gamma}\left(1 + dx^{\rho/\gamma} + o\left(x^{\rho/\gamma}\right)\right) \text{ as } x \to \infty \tag{1.45}$$

where $\gamma > 0$, $\rho \leq 0$, $c > 0$, and $d \in \mathbb{R}^{*}$.
This sub-class of heavy-tailed distributions contains the Pareto, Burr, Fréchet and $t-$Student cdf's usually used, in insurance mathematics, as models for dangerous risks. Relation (1.45) may be reformulated in terms of functions $\mathbb{Q}$ and $\mathbb{U}$ as follows:

$$\mathbb{Q}(1-s) = c^{\gamma}s^{-\gamma}\left(1 + \gamma d c^{\rho} s^{-\rho} + o\left(s^{-\rho}\right)\right) \text{ as } s \to 0, \tag{1.46}$$

Distribution	c	d	γ	ρ	β	$A(z)$
Fréchet	1	$-\gamma/2$	γ	-1	$1/2$	$(\gamma/2)z^{-1}$
Burr	1	γ/ρ	γ	ρ	1	γz^{ρ}
Gener. Pareto	$1/\gamma$	0	γ	$-\gamma$	1	$\gamma z^{-\gamma}$
Student's t_ν	$\frac{\sqrt{\nu}}{c_\nu}$	$-\frac{\nu(1+\nu)(3+\nu)c_\nu^4}{8(2+\nu)^2(4+\nu)}$	$1/\nu$	$-2/\nu$	$\frac{(\nu+1)c_\nu^2}{2+\nu}$	$\frac{(\nu+1)c_\nu^2}{(2+\nu\nu)}z^{-2/\nu}$

Table 1.3: A specimen of distributions in Hall's class. The constant c_ν in the bottom row is equal to $(\nu\mathfrak{B}(\nu/2,1/2))^{1/\nu}$, where $\mathfrak{B}$ is the complete Beta function; the case $\nu = 1$ corresponds to the Cauchy cdf.

and

$$\mathbb{U}(t) = c^{\gamma}t^{\gamma}\left(1 + \gamma dc^{\rho}t^{\rho} + o\left(t^{\rho}\right)\right) \text{ as } t \to \infty.$$

Straightforward computations show that, in the case of Hall model, functions $A(t)$ and $A^*(t)$ are respectively equivalent to $d\rho\gamma c^{\rho}t^{\rho}$ and $d\rho\gamma t^{\rho/\gamma}$ as $t \to \infty$, whereas function $A^{**}(t)$ is equivalent to $d\rho\gamma c^{\rho}t^{-\rho}$ as $t \to 0$.

For this class of cdf's the second-order condition (1.44) holds with a function $A(t)$ such that $A(t) = \rho dt^{\rho} = \gamma\beta t^{\rho}$. In table (1.3), we specify the constants γ, ρ, β, c and d for several popular heavy-tailed distributions (i.e., Fréchet, Burr, generalized Pareto, and Student's t_ν) along with formulas of their corresponding functions $A(t)$.

1.6.6 Intermediate Order Statistics and Brownian Motion

We continue the discussion on the behavior of intermediate order statistics under extreme value conditions. We have seen that a sequence of intermediate order statistics is asymptotically normal (when properly normalized) under von Mises's extreme value condition. However, we want to consider many intermediate order statistics at the same time, hence we want to consider the tail (empirical) quantile process.

It is instructive to start by proving the main result of Theorem (1.21), i.e., the asymptotic normality of a sequence of intermediate order statistics, again, now not under von Mises' conditions but under the second-order condition.

Theorem 1.22 *Let $X_{1,n} \leq X_{2,n} \leq ... \leq X_{n,n}$ be the nth order statistics from an iid sample with df F. Suppose that the second-order condition (1.41), holds for* *some* $\gamma \in \mathbb{R}$, $\rho \leq 0$. *Then*

$$\sqrt{k}\frac{X_{n-k,n} - \mathbb{U}\left(n/k\right)}{a\left(n/k\right)}, \tag{1.47}$$

is asymptotically standard normal provided that, the sequence $k = k(n) \to \infty$ *and* $k(n)/n \to 0$ *as* $n \to \infty$, *and* $\lim_{n\to\infty} \sqrt{k}A\left(n/k\right)$ *exists and is finite.*

The last result can be vastly generalized and yields the following, relating the tail quantile process to Brownian motion in a strong sense.

Theorem 1.23 (Drees (1998), Theorem 2.1) *Suppose* $X_1, X_2, \ldots$ *are iid rv's with cdf* F. *Suppose that* F *satisfies the second-order condition (1.41) for some* $\gamma \in \mathbb{R}$ *and* $\rho \leq 0$. *Let* $X_{1,n} < X_{2,n} <\ldots< X_{n,n}$ *be the nth order statistics. We can define a sequence of Brownian motions* $\{\mathbb{W}_n(s)\}_{s>o}$ *such that for suitably chosen functions* a_o *and* A_0 *and each* $\varepsilon > 0$,

$$\sup_{K^{-1}\leq s\leq 1} s^{\gamma+1/2+\varepsilon} \left| \begin{array}{c} \sqrt{k}\left(\frac{X_{n-[ks],n}-\mathbb{U}(n/k)}{a_0(n/k)} - \frac{s^{-\gamma}-1}{\gamma}\right) \\ -s^{-\gamma-1}\mathbb{W}_n(s) - \sqrt{k}A_0\left(n/k\right)\Psi_{\gamma,\rho}\left(s^{-1}\right) \end{array} \right| \overset{P}{\to} 0, \tag{1.48}$$

$n\to \infty$, *provided* $k = k\left(n\right) \to \infty, k/n \to 0$ *and* $\sqrt{k}A_0\left(n/k\right) = O(1)$.

Chapter 2

Estimation of the Extreme Value Index and High Quantile

2.1 General Diagnostic Plots

In the real world, the extreme value theory as we described in chapter (1) needs to be applied through statistical data, the observed sea level, major insurance claims, large variation of security market values over a certain time period, daily records of temperature and precipitation at a certain location, etc. We hope the modeling of the empirical data through extremes would manifest most of these variations. Appropriate models can be used to manage financial risks, to set up prevention procedure or to obtain estimations and predictions.
Before getting into any statistical analysis, we want to learn our best knowledge of the data set. How is the data collected? Is there any missing or unreported data? Since it is particularly important that the data is error-free for extreme observations, we should give special attentions to data that are outliers.
The first step in exploratory analysis is to have a scatter plot of all observations. It is important just by looking at data to obtain the following information: range of data, several extremes, trends and seasonalities, any violation of independence and stationarity conditions. If trends and seasonalities are suspected, we recommend to fit the location, scale and shape parameters with a time variable t. See for example, chapter 6 in Coles [21] and references therein. Embrechts, Kluppelberg and Mikosch's book [46] and its references give a thorough list of both graphical and analytical methods, and in Bassi [9], a survival kit of some

terminologies on quantile estimation is provided. We have introduced the mean excess function in section (1.5.4), which is crucial in the later analysis on the determination of choice of transforms. We will emphasize two other methods used in this thesis: QQ plots, the return period and return levels.

2.1.1 Probability and Quantile Plots (QQ-plot)

Suppose $X_1, X_2, ..., X_n$ are continuous rv's and $x_1, x_2, ..., x_n$ are independent observations from a common population with unknown df F. An estimate of F, say $\widehat{F}$, has been obtained. The probability and quantile plots provide a graphical accessment to the fitted distribution $\widehat{F}$. It follows from the Quantile Transformation Lemma ([46] Lemma 4.1.9) that $F(X_i)$ has a uniform distribution on $(0, 1)$ for $i = 1, ..., n$. Furthermore, if $x_{1,n}, x_{2,n}, ..., x_{n,n}$ are the ordered sample, then the expectation of these quantiles can be computed as:

$$E\left(F\left(X_{i,n}\right)\right) = \frac{i}{n+1}, \text{ for } i = 1, ..., n,$$

and this leads to the following definition.

Definition 2.1 *Given an ordered sample of independent observations*

$$x_{1,n} \leq x_{2,n} \leq ... \leq x_{n,n}$$

from a population with estimated distribution function $\widehat{F}$, a graph

$$\left\{\widehat{F}\left(\left(X_{i,n}\right)\right), \frac{i}{n+1}\right\}, i = 1, ..., n$$

*is called a probability plot (**PP-plot**).*

Definition 2.2 *Given an ordered sample of independent observations*

$$x_{1,n} \leq x_{2,n} \leq ... \leq x_{n,n}$$

from a population with estimated distribution function $\widehat{F}$, a graph

$$\left\{\widehat{F}^{-1}\left(\frac{i}{n+1}\right), X_{i,n}\right\}, i = 1, ..., n$$

*is called a quantile plot (**QQ-plot**).*

The quantities $x_{i,n}$ is the empirical $\left(\frac{i}{n+1}\right)$-quantile of the population distribution F while $\widehat{F}^{-1}\left(\frac{i}{n+1}\right)$ is the estimation. If $\widehat{F}$ is a reasonable estimation of F, the quantile plot should look roughly linear. This remains true if the data come from a linear transformation of the distribution. And since that, the change of location and scale parameters only change the plot through y-intercept and slope.
Outliers can be easily identified on the QQ**-plot** in a general statistical analysis. While the subject of the extreme value study concerns the upper tail, we should be particularly cautious about any point that substantially deviates from the model on the large observation end. Since the shape parameter γ determines how heavy the tail distribution is, some difference in distributional shape may be deduced from the plot. In general, an over estimation of (heavy tail) will result in a concave down curve in the QQ-plot, and an under estimation of (light tail) will result in a concave up curve in the QQ-plot.

2.1.2 The Return Period and The Return Level

The return period and the return level are very silimar to the t-year event we mentioned in section (1.2). We make the definition precise as below:

Definition 2.3 *Let (X_i) be a sequence of iid rvs with continuous cdf F and u a given threshold. Consider the sequence $(1_{(X_i>u)})$ of iid Bernoulli rv's with success probability $p = 1 - F(u)$. Then*

$$E(L(u)) = 1/p, \tag{2.1}$$

*is called the **return period** of the event $(X_i > u)$, where $L(u) = min\{i \geq 1 : X_i > u\}$ is the time of first exceedance of the threshold u. $L(u)$ follows a geometric distribution with parameter p.*

From the definition, return period is the expected revisiting period corresponding to the threshold u. This leads to the other definition:

Definition 2.4 *Let (X_i) be a sequence of iid rv's with continuous cdf F. If*

$$F(z_p) = 1 - p$$

*then z_p is called the **return level** associated with the return period $1/p$.*

2.1.3 Mean Excess Function (mef) Plot

Some elementary properties of the *mef* could be used to distinguish between light-tailed and heavy-tailed models. The *mef* of a $\mathcal{E}(\theta)$ rv is equal to $1/\theta$ and for the Pareto case, the *mef* is linear (see 1.5 (a)). Thus, the *mef* of a heavy-tailed cdf, for large arguments, typically appears to be between a constant function and a straight line (with positive slope). A graphical investigation of tail behavior can now be based on the empirical *mef* e_n defined on $\mathbb{R}_+$ as follows:

$$e_n(u) = \frac{1}{\overline{F}_n(u)} \int_u^{x^*} \overline{F}_n(x)\,dx = \frac{1}{N_u} \sum_{i=1}^{n} (X_i - u)\, 1_{\{X_i > u\}}, \tag{2.2}$$

where N_u is the number of observations that exceed u.

Definition 2.5 (mef-plot) *The graph*

$$\{(X_{i,n}, e_n(X_{i,n})) : i = 1, ..., n\},$$

is called mef-plot.

When the mef-plot is close to a straight line, the underlying distribution may be modelled by a Pareto-like cdf .

2.1.4 Pareto Quantile Plot

This is another graphical tool for testing the hypothesis of tail heaviness of a given set of data. It is known that Pareto-type distributions satisfy $\mathbb{U}(x) = x^{1/\gamma} L_{\mathbb{U}}(x), x > 0$ with $L_{\mathbb{U}}$ being a slowly varying function at infinity (in particular $\log L_{\mathbb{U}}(x)/\log x \to 0$ as $x \to \infty$). This yields that as $x \to \infty$, $\log L_{\mathbb{U}}(x) \sim \gamma \log x$. Consequently, for heavy-tailed data the Pareto quantile plot

$$\left(\log\left(\frac{n+1}{i}\right), \log X_{n-i+1,n}\right) : i = 1, ..., n$$

should show an approximate linear behavior with slope equal to γ.

2.1.5 Gumbel's Method of Excesses

This analytical method concerns the number of values, among future observations, that exceed past records. If we take the k-th upper order statistic $X_{n-k+1,n}$

as a (random) threshold, then the number of its exceedances among the next r observations $X_{n+1},...,X_{n+r}$ is an hypergeometric rv. If we denote this number by $S_r^n(k)$ then $S_r^n(k) = \sum_{i=1}^{r} 1_{\{X_{n+1}>X_{n-k+1,n}\}}$ and we have

$$P\left(S_r^n(k) = j\right) = \left[C_{r+k-1-j}^{k-1} C_{j+n-k}^{n-k}\right] / C_{r+n}^{n}, \ j = 1, 2, ..., r.$$

With this formula one can compute the probabilities of several future events related to a given high threshold such as no exceedances, exceeding at least once,... The mean number of exceedances of the level $X_{n-k+1,n}$ is equal to $r(n-k+1)/(n+1)$.

2.2 Parameter Estimation for the GEV

Once the model has been set up, parameters in the model need to be estimated using appropriate procedures. Suppose $X_1, X_2, ..., X_n$ is a sequence of rv's of iid GEVD with parameters γ, μ and $\beta > 0$, and $x_1, x_2, ..., x_n$ are the recorded observations. As in 1.5, the df $\mathcal{H}(x)$ is

$$\mathcal{H}_\gamma(x) = \begin{cases} \exp\left\{-\left[1+\gamma\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\gamma}\right\} & \text{when } \gamma \neq 0, \ 1+\gamma\left(\frac{x-\mu}{\sigma}\right) > 0 \\ \exp\left\{-\exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right\} & \text{when } \gamma = 0, x \in \mathbb{R} \end{cases}.$$

In this section, we introduce two methods: maximum likelihood method (ML) and method of probability-weighted moments.(PWM)

2.2.1 Maximum Likelihood Method

The method of maximum likelihood is, by far, the most popular technique for deriving estimators.

Definition 2.6 *Let $f(X \mid \theta)$ denote the joint pdf of the sample $X = (X_1, X_2, ..., X_n)$. Given that $X = x = (x_1, x_2, ..., x_n)$ is observed, the function defined by*

$$L(\theta \mid X) = f(X \mid \theta),$$

is called the ***likelihood function****, and*

$$l(\theta \mid X) = \ln L(\theta \mid X),$$

defines the ***log-likelihood function****. The* ***maximum likelihood estimator*** *(MLE) $\widehat{\theta}(x) := \arg\max_{\theta\in\Theta} l(\theta \mid X)$ is the function of X that maximizes $L(\theta \mid X)$ or $l(\theta \mid X)$ over an appropriate parameter space Θ .*

2.2.2 Method of Probability-Weighted Moments

The technique of moments method is to equate the model-moments with empirical moments. Even though the general properties of obtained estimators can be unreliable, the method of moments can be very useful in obtaining an initial approximation of the intended statistics. The class of probability-weighted moments stands out to be more promising in obtaining a good first estimate. The results are used as the initial values for other methods when numerical techniques are applied. Let $\theta = (\gamma; \mu, \sigma)$ and define

$$w_r(\theta) = E\left[X\mathcal{H}_\theta^r(x)\right] \text{ for } r = 1, 2. \tag{2.3}$$

In case $\gamma < 1$, calculation yields

$$w_r(\theta) = \frac{1}{r+1}\left\{\mu - \frac{\sigma}{\gamma}\left[1 - \Gamma(1-\gamma)(1+r)^\gamma\right]\right\} \tag{2.4}$$

where Γ denotes the Gamma function $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$ for $x > 0$. Choosing $r = 0, 1, 2$, we can then immediately obtain

$$\begin{cases} w_0(\theta) = \mu - \frac{\sigma}{\gamma}\left[1 - \Gamma(1-\gamma)\right] \\ 2w_1(\theta) - w_0(\theta) = \frac{\sigma}{\gamma}\Gamma(1-\gamma)(2^\gamma - 1) \\ 3w_2(\theta) - w_0(\theta) = \frac{\sigma}{\gamma}\Gamma(1-\gamma)(3^\gamma - 1) \end{cases} \tag{2.5}$$

$(\gamma; \mu, \sigma)$ can be explicitly solved from the above system (2.5). For example

$$\frac{2^\gamma - 1}{3^\gamma - 1} = \frac{2w_1(\theta) - w_0(\theta)}{3w_2(\theta) - w_0(\theta)}. \tag{2.6}$$

Parameter estimation is obtained by replacing the model-moments $w_r(\theta)$ in (2.6) by empirical moments $\widehat{w}_r(\theta)$. To obtain empirical moments, notice

$$\widehat{w}_r(\theta) = \frac{1}{n}\sum_{j=1}^n X_{j,n}\mathcal{H}_\theta^r(X_{j,n}) \text{ for } r = 0, 1, 2. \tag{2.7}$$

Again it follows from the Quantile Transformation Lemma, that

$$\left(\mathcal{H}_\theta(X_{1,n}), ..., \mathcal{H}_\theta(X_{n,n})\right) = (U_{1,n}, ..., U_{n,n}),$$

where $U_{1,n}, ..., U_{n,n}$ are the order statistics of an iid sequence $U_1, U_2, ..., U_n$ uniformly distributed on $(0, 1)$. Thus, 2.7 can be rewritten as:

$$\widehat{w}_r(\theta) = \frac{1}{n}\sum_{j=1}^n X_{j,n}^r U_{j,n}^r \text{ for } r = 0, 1, 2, \tag{2.8}$$

where $U_{j,n}^r$ are often approximated by their expectations.

2.3 Estimation of the Extreme Value Index

In this section, we give some estimators of the EVI γ constructed under maximum domain of attraction conditions. That is, the data $(X_1, X_2, ..., X_n)$ are assumed to be drawn from a population X with cdf F in $\mathcal{D}(\mathcal{H}_\gamma)$. As opposed to the parametric methods of the previous section, the semi-parametric statistical procedures, appropriate to this situation, don't assume the knowledge of the whole distribution but only focus on the distribution tails. The case $\gamma > 0$ has got more interest because data sets in most real-life applications, exhibit heavy tails. For a detailed review of some of the firrst works done in this matter, see [96] amongst others. Classical estimators are based on the largest order statistics $X_{n-k,n}, ..., X_{n,n}$, where k is an intermediate sequence of integers related to the sample size n in the following way:

$$k = k_n \to \infty \text{ and } k/n \to 0 \text{ as } n \to \infty. \tag{2.9}$$

The statistic $X_{n-k,n}$ is then said to be intermediate order statistic. EVT-based estimators rely heavily on k.

2.3.1 A Simple Estimator for the Tail Index ($\gamma > 0$): The Hill Estimator

In order to introduce the Hill estimator, a simple and widely used estimator, let us start from theorem 1.11:

$$F \in \mathcal{D}(\mathcal{H}_\gamma) \text{ for } \gamma > 0 \text{ iff} \lim_{t\to\infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-1/\gamma}, x > 0.$$

In this case the parameter $\alpha = 1/\gamma > 0$ is called the tail index of F. Theorem 1.12 gives an equivalent form of this condition:

$$\lim_{t\to\infty} \frac{\int_t^\infty (1 - F(s))ds/s}{1 - F(t)} = \gamma. \tag{2.10}$$

Now partial integration yields

$$\int_t^\infty (1 - F(s))\frac{ds}{s} = \int_t^\infty (\log \mathbb{U} - \log t)\, dF(u).$$

Hence we have

$$\lim_{t\to\infty} \frac{\int_t^\infty (\log u - \log t)\, dF(u)}{1 - F(t)} = \gamma.$$

In order to develop an estimator based on this asymptotic result, replace in equation (2.10) the parameter t by the intermediate order statistic $X_{n-k,n}$ and F by the empirical distribution function F_n. We then get Hill's (1975) estimator $\widehat{\gamma}_n^H$, defined by

$$\widehat{\gamma}_n^H = \frac{1}{k}\sum_{i=1}^{k}\left(\log X_{n-i,n} - \log X_{n-k,n}\right). \tag{2.11}$$

Hill's estimator is usual and easy-to-explain. It can be derived through several other approaches see [46].

In his original paper [78], Hill did not investigate the asymptotic behavior of the estimator. It was Mason who proved the weak consistency in [94]. The strong consistency was proved in [36] by Deheuvels, Häusler and Mason who gave an optimal rate of convergence for an appropriately chosen sequence k_n. The asymptotic normality was established, under some extra condition on F, in several papers such as, e.g. [23], [31], [73] and [77].

The asymptotic behavior of Hill's estimator is given in the follows theorem

Theorem 2.1 *Let $X_1, X_2, ..., X_n$ be a iid rv with cdf F. Suppose that $F \in \mathcal{D}(\mathcal{H}_\gamma)$ with $\gamma > 0$. Then as $n \to \infty$, $k \to \infty$ and $k/n \to 0$*

(a) ***Weak Consistency:***

$$\widehat{\gamma}_n^H \xrightarrow{P} \gamma \text{ as } n \to \infty.$$

(b) ***Strong consistency:*** *If $k/\log\log n \to \infty$ as $n \to \infty$, then*

$$\widehat{\gamma}_n^H \xrightarrow{a.s.} \gamma \text{ as } n \to \infty.$$

(c) ***Asymptotic normality:*** *Suppose that the cdf F satisfies the second-order condition (1.39). Then*

$$\sqrt{k}\left(\widehat{\gamma}_n^H - \gamma\right) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda}{1-\rho}, \gamma^2\right),$$

provided $k = k(n) \to \infty$ and $k(n)/n \to 0$ as $n \to \infty$, and

$$\lim_{n\to\infty} \sqrt{k} A\left(n/k\right) = \lambda.$$

The second-order framework provides the most natural approach to the asymptotic normality of estimators like Hill's estimator.

2.3.2 The ratio estimator

A generalization of Hill's estimator in the sense that we use an arbitrary threshold level u_n instead of an order statistic $X_{n-k,n}$ in the relation (2.11), we obtain the ratio estimator

$$\widehat{\gamma}_n^R = \sum_{i=1}^{n} \log\left(X_i/u_n\right) 1_{\{X_i > u_n\}} / \sum_{i=1}^{n} 1_{\{X_i > x_n\}} \tag{2.12}$$

See (Goldie and Smith, 1987).
Note that Hill's estimator and ratio estimator may also be applied to dependent data (Novak, 2002; Resnick and Stârică, 1999). Hill's estimator is very sensitive with respect to dependence in the data (see Embrechts et al., 1997).

2.3.3 Reduced-Bias Tail Index

The Hill estimator still stays one of the most important estimators even though the graphs of the estimates as a function of k are not smooth and in spite of the fact that its bias increases quickly with k. For k small they have a high variance, and for large k a high bias. Several authors have recognized and exploited the importance of bias reduction and the use of quantile plots in estimating $\gamma > 0$ (Kratz and Resnick, 1996; Schultze and Steinebach, 1996; Beirlant, Dierckx, Goegebeur and Matthys, 1999; Feuerverger and Hall, 1999; Gomes and Martins, 2002) among others.
New second-order "shape" and "scale" estimators allowed the development of second-order reduced-bias estimators, which are much less sensitive to the choice of k.
Geluk and de Haan, 1987 to be able to reduce the bias of these estimators, it is quite useful to assume that we are working in Hall's class of heavy-tailed models (Hall, 1982; Hall and Welsh), where the second order condition (1.41) holds with $A(t) := \gamma\beta t^{\rho}$. de Haan and Peng, 1998 write the Hill estimator as follows

$$\widehat{\gamma}_n^H \stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}} Z_k + \frac{A(n/k)}{(1-\rho)}(1 + o_P(1)),$$

with $Z_k = \sqrt{k}\left(\sum_{i=1}^{k} E_i/k - 1\right)$, and $\{E_i\}$ iid. standard exponential rv's. Consequently, if we choose k such that $\sqrt{k}A(n/k) \to \lambda \neq 0$, finite as $n \to \infty$,

$$\sqrt{k}\left(\widehat{\gamma}_n^H - \gamma\right) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda}{1-\rho}, \gamma^2\right).$$

We see that the dominant component of the bias of Hill's estimator is $A(n/k)/(1-\rho) = \gamma\beta(n/k)^{\rho}/(1-\rho)$. This component can be easily estimated and removed from Hill's estimator, leading to any of the asymptotically equivalent estimators (Caeiro et al., 2005),

$$\widehat{\gamma}^{RB}_{\widehat{\rho},\widehat{\beta}} := \widehat{\gamma}^{H}_{n}\left(1-\frac{\widehat{\beta}}{1-\widehat{\rho}}\left(\frac{n}{k}\right)^{\widehat{\rho}}\right), \tag{2.13}$$

where $\widehat{\rho}$ and $\widehat{\beta}$ need to be adequate consistent estimators of the second order parameters ρ and β.

We shall consider here particular members of the class of estimators of the second order parameter proposed by Fraga Alves et al. (2003). Such a class of estimators may be parameterized by a tuning real parameter $\tau \in \mathbb{R}$ (Caeiro and Gomes, 2004). These estimators depend on the statistics

$$T_n^{(\tau)}(k) = \begin{cases} \dfrac{\left(\mathbb{M}_n^{(1)}(k)\right)^{\tau} - \left(\mathbb{M}_n^{(2)}(k)/2\right)^{\tau/2}}{\left(\mathbb{M}_n^{(2)}(k)/2\right)^{\tau/2} - \left(\mathbb{M}_n^{(3)}(k)/6\right)^{\tau/3}}, & \text{if } \tau \neq 0 \\ \dfrac{\ln\left(\mathbb{M}_n^{(1)}(k)\right)^{\tau} - \frac{1}{2}\ln\left(\mathbb{M}_n^{(2)}(k)/2\right)^{\tau/2}}{\frac{1}{2}\ln\left(\mathbb{M}_n^{(2)}(k)/2\right)^{\tau/2} - \frac{1}{3}\ln\left(\mathbb{M}_n^{(3)}(k)/6\right)^{\tau/3}}, & \text{if } \tau = 0 \end{cases}, \tag{2.14}$$

converge towards $3(1-\rho)/(3-\rho)$, independently of the tuning parameter, whenever the second order condition (1.41) holds and k is such that relation (2.9) holds and $\sqrt{k}\ A(n/k) \to \infty$, as $n \to \infty$, where

$$\mathbb{M}_n^{(r)}(k) = \frac{1}{k}\sum_{i=1}^{k}\left(\log(X_{n,n-i+1}) - \log(X_{n,n-k+1})\right)^{r}. \tag{2.15}$$

The ρ-estimators considered have the functional expression,

$$\widehat{\rho}_n^{(\tau)}(k) = -\min\left(0, \frac{3T_n^{(\tau)}(k)-1}{T_n^{(\tau)}(k)-3}\right). \tag{2.16}$$

For the estimation of β we shall here consider the estimator developed in Gomes and Martins (2002), with the functional expression,

$$\widehat{\beta}_{\widehat{\rho}}(k) = \left(\frac{k}{n}\right)^{\widehat{\rho}} \frac{\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\widehat{\rho}} N_n^{(1)}(k) - N_n^{(1-\widehat{\rho})}(k)}{\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\widehat{\rho}} N_n^{(1-\widehat{\rho})}(k) - N_n^{(1-2\widehat{\rho})}(k)}, \tag{2.17}$$

where

$$N_n^{(\alpha)}(k) = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{\alpha-1}\mathbf{U}_i,$$

with $\widehat{\rho} = \widehat{\rho}_n^{(\tau)}(k)$ and

$$\mathbf{U}_i = i\left(\log X_{n-i+1,n} - \log X_{n-k,n}\right).$$

The asymptotic behavior of reduction of bias Hill's estimator is given in the follows theorem

Theorem 2.2 *Under the second order conditions, let us consider the tail index estimators $\widehat{\gamma}_{\widehat{\rho},\widehat{\sigma}}(k)$ with $\widehat{\beta}$ and $\widehat{\rho}$ consistent for the estimation of β and ρ, respectively, both computed at the level k_1 of a larger order than the level k at which we compute the tail index, and such that $\widehat{\rho} - \rho = o_P(1/\ln n)$. Then*

$$\sqrt{k}\left(\widehat{\gamma}_{\widehat{\rho},\widehat{\beta}}^{RB}(k) - \gamma\right) \xrightarrow{d} \mathcal{N}\left(0, \gamma^2\right),$$

even if $\sqrt{k}A(n/k) \to \lambda \neq 0$.

2.3.4 General Case $\gamma \in \mathbb{R}$: The Pickands Estimator

The simplest and oldest estimator for γ is the Pickands estimator (1975):

$$\widehat{\gamma}_n^P := \frac{1}{\log 2}\log\frac{X_{n-k,n} - X_{n-2k,n}}{X_{n-2k,n} - X_{n-4k,n}}. \tag{2.18}$$

We shall give weak consistency and asymptotic normality of $\widehat{\gamma}_n^P$.

Theorem 2.3 *Let $X_1, X_2, ..., X_n$ be a iid rv's with cdf F. Suppose that $F \in \mathcal{D}(\mathcal{H}_\gamma)$ with $\gamma \in \mathbb{R}$. Then as $n \to \infty$, $k \to \infty$ and $k/n \to 0$.*
(a) ***Weak Consistency:***

$$\widehat{\gamma}_n^P \xrightarrow{P} \gamma \text{ as } n \to \infty.$$

(b) ***Strong consistency:*** *If $k/\log\log n \to \infty$ as $n \to \infty$, then*

$$\widehat{\gamma}_n^P \xrightarrow{a.s.} \gamma \text{ as } n \to \infty.$$

(c) ***Asymptotic normality:*** *Suppose that $\mathbb{U}$ has a positive derivative $\mathbb{U}'$ and that $\pm t^{1-\gamma}\mathbb{U}'(t)$ (with either choice of sign) is $\Pi-$varying at infinity with auxiliary function a. If $k = o\left(n/g^{\leftarrow}(n)\right)$, where $g(t) := t^{3-2\gamma}\left(\mathbb{U}'(t)/a(t)\right)^2$, then*

$$\sqrt{k}\left(\widehat{\gamma}_n^P - \gamma\right) \xrightarrow{d} \mathcal{N}\left(0, \eta^2\right) \text{ as } n \to \infty,$$

where

$$\eta^2 := \frac{\gamma^2\left(2^{2\gamma+1}+1\right)}{\left(2\left(2^\gamma - 1\right)\log 2\right)^2}.$$

A detailed account on Pickands' estimator (with different conditions on cdf F and various examples) as well as the proofs of the results of Theorem 2.1 are to be found in [38]. A full generalization of Pickands' estimator has been introduced in [126] as follows:

$$\widehat{\gamma}_n^{PG} := \widehat{\gamma}^{PG}(k; u, v) := \frac{1}{\log v}.\log \frac{X_{n-k,n} - X_{n-[uk],n}}{X_{n-[vk],n} - X_{n-[uvk],n}}, \tag{2.19}$$

where u, v are positive real numbers different from 1 such that $[uk]$, $[vk]$ and $[uvk]$ don't exceed n. For $u = v = 2$, we have $\widehat{\gamma}_n^P$ and for $u = v = q \in \mathbb{N}\backslash\{0, 1\}$, we obtain the Fraga Alves generalization introduced in [50].

2.3.5 Moment Estimator

Next we want to develop an estimator similar to the Hill estimator but one that can be used for general $\gamma \in \mathbb{R}$, not only for $\gamma > 0$. In order to introduce the estimator let us look at the behavior of the Hill estimator for general γ. We look at a slightly more general statistic.
An immediate problem with applying the Hill estimator for the case $\gamma < 0$ is that $\mathbb{U}(\infty) \leq 0$ is possible, in which case the logarithm of the observations is not defined. . In 1989, Dekkers, Einmahl and de Haan proposed in [39] an extension to any type of distributions, called moment estimator, is given by the following estimator:

$$\widehat{\gamma}_n^M := \mathbb{M}_n^{(1)} + 1 - \frac{1}{2}\left(1 - \left(\mathbb{M}_n^{(1)}\right)^2 / \mathbb{M}_n^{(2)}\right). \tag{2.20}$$

The asymptotic behavior of $\widehat{\gamma}_n^M$ is investigated as well.

Theorem 2.4 *Let $X_1, X_2, ..., X_n$ be a iid r.v. with cdf F. Suppose that $F \in \mathcal{D}(\mathcal{H}_\gamma)$ with $\gamma \in \mathbb{R}$. Then as $n \to \infty$, $k \to \infty$ and $k/n \to 0$*
(a) ***Weak Consistency:***

$$\widehat{\gamma}_n^M \xrightarrow{P} \gamma \text{ as } n \to \infty.$$

(b) ***Strong consistency:*** *If $k/(\log n)^\delta \to \infty$ as $n \to \infty$, for some $\delta > 0$, then*

$$\widehat{\gamma}_n^M \xrightarrow{a.s.} \gamma \text{ as } n \to \infty.$$

(c) ***Asymptotic normality:***

$$\sqrt{k}\left(\widehat{\gamma}_n^M - \gamma\right) \xrightarrow{d} \mathcal{N}\left(0, \eta^2\right) \text{ as } n \to \infty,$$

where

$$\eta^2 := \begin{cases} 1+\gamma^2, & \gamma \geq 0 \\ \left(1-\gamma^2\right)\left(1-2\gamma\right)\left(4 - 8\frac{1-2\gamma}{1-3\gamma} + \frac{(5-11\gamma)(1-2\gamma)}{(1-3\gamma)(1-4\gamma)}\right), & \gamma < 0 \end{cases}.$$

Remark 2.1

$$\widehat{\gamma}_- = \widehat{\gamma}_n^M - \mathbb{M}_n^{(1)} = 1 - \frac{1}{2}\left(1 - \left(\mathbb{M}_n^{(1)}\right)^2 / \mathbb{M}_n^{(2)}\right). \tag{2.21}$$

2.3.6 Kernel Type Estimators

A major drawback of the estimators above is the discrete character of their behavior in the sense that increasing k by 1, can change the actual value of the estimate considerably. In 1985, using a kernel function K, Csörgö, Deheuvels and Mason [25] proposed a smoother version of Hill's estimator (denoted by $\widehat{\gamma}_n^K$) and proved its consistency and asymptotic normality:

$$\widehat{\gamma}_n^K(h) := \sum_{i=1}^{n-1} \frac{i}{nh} \mathrm{K}\left(\frac{i}{nh}\right)\left(\log X_{n-i+1,n} - \log X_{n-i,n}\right) / \int_0^{1/h} \mathrm{K}(u)\,du. \tag{2.22}$$

where $h > 0$ is called bandwidth and K is a non-negative, non-increasing and right continuous function on $(0,1)$ such that $\int_0^\infty \mathrm{K}(u)\,du = 1$ and $\int_0^\infty \mathrm{K}^{-1/2}(u)\,du < \infty$. It is obvious that $\int_0^{1/h} \mathrm{K}(u)\,du$ can be replaced by $\left(\frac{1}{nh}\right)\sum_{i=1}^{n-1}\mathrm{K}\left(\frac{i}{nh}\right)$). Notice that, using the uniform kernel $\mathrm{K} = 1_{(0,1)}$ and $h = k/n$, we obtain the Hill estimator $\widehat{\gamma}^H$ as a special case.

The kernel type estimator depends in a continuous way on the bandwidth h representing the proportion of top order statistics used. Hence, plotting $\widehat{\gamma}_n^K$ as a function of h yields a smooth figure as opposed to the zigzag figure resulting from plotting any of the previous estimators as a function of k.
Unfortunately, $\widehat{\gamma}_n^K$ can, similarly to $\widehat{\gamma}_n^H$, only be used for the estimation of positive extreme value indices. For an investigation of the asymptotic properties of $\widehat{\gamma}_n^K$ with a discussion of the restrictions (on F and K) under which the asymptotic normality is established, we refer to [25] where it is also shown that it is possible to improve the (asymptotic) variance by choosing appropriate kernels. For a complete analysis of $\widehat{\gamma}_n^K$ and other kernel type estimators, one may e.g., consult [28] and [125]. The generalization of $\widehat{\gamma}_n^H$ to the estimation of any real-valued tail index is made by Groeneboom, Lopuhaä and de Wolf in [68] where they

introduce a new kernel type estimator (to be denoted by $\widehat{\gamma}_n^W$) that inherited the smooth behavior of $\widehat{\gamma}_n^K$ as well as the general applicability of $\widehat{\gamma}_n^M$.

$$\widehat{\gamma}_n^W := \widehat{\gamma}_n^{\mathrm{K}}(h) = \widehat{\gamma}_{pos} - 1 + \left(\widehat{q}_n^{(2)}\right) / \left(\widehat{q}_n^{(1)}\right), \tag{2.23}$$

where

$$\widehat{\gamma}_{pos} := \widehat{\gamma}_{pos}(h) = \sum_{i=1}^{n-1} \frac{i}{n} \mathrm{K}_h\left(\frac{i}{n}\right)(\log X_{n-i+1,n} - \log X_{n-i,n}),$$

$$\widehat{q}_n^{(1)} := \widehat{q}_n^{(1)}(h) = \sum_{i=1}^{n-1} \left(\frac{i}{n}\right)_h^{\alpha} \mathrm{K}\left(\frac{i}{n}\right)(\log X_{n-i+1,n} - \log X_{n-i,n}),$$

$$\widehat{q}_n^{(2)} := \widehat{q}_n^{(2)}(h) = \sum_{i=1}^{n-1} \frac{d}{du}\left[u_h^{\alpha+1}\mathrm{K}(u)\right]_{u=i/n}(\log X_{n-i+1,n} - \log X_{n-i,n}),$$

with $\mathrm{K}_h(u) = \mathrm{K}(u/h)/h$ and $\alpha > 0$. Here K is a kernel function null outside $(0,1)$ and satisfying K, K' and K'' bounded, $\mathrm{K}(1) = \mathrm{K}'(1) = 0$, $\int_0^1 \mathrm{K}(u)\,du = 1$ and $\int_0^1 u^{\alpha-1}\mathrm{K}(u)\,du \neq 0$. Notice that the first term in (**??**) is (almost) $\widehat{\gamma}^{\mathrm{K}}$. The basis of the construction of $\widehat{\gamma}^W$ is the von Mises condition

$$\lim_{t \to x^*} \frac{d}{dt}\left(F(t)/F'(t)\right) = \gamma.$$

The full description of the way $\widehat{\gamma}_n^W$ is derived, is given in [68] where consistency and asymptotic normality are established.

2.4 The choice of k in Hill's estimator

The choice of the optimal threshold or corresponding k is however a difficult problem and has been studied by many authors, as discussed in Beirlant et al. (2004), Markovich (2007) [93] and Meragni (2008) [100]. In this thesis we are interesting only with algorithm of Cheng and Peng (2001) for esteablished this fraction.

Note that, the asymptotic normality of Hill's estimator is used to construct (asymptotic) confidence intervals for the EVI γ of a cdf F belonging to Hall's class. Indeed, we have that (see, e.g. Hall, 1982)

$$\sqrt{k}\left(\widehat{\gamma}_n^H - \gamma\right) \xrightarrow{d} \mathcal{N}\left(0, \gamma^2\right), \quad as \quad n \to \infty,$$

iff $k = o\left(n^{-2\rho/(1-2\rho)}\right)$. Thus, for $0 < \alpha < 1$ the one-sided and two-sided intervals of confidence level $(1-\alpha)$ are respectively

$$I_1(\alpha) := \left(0, \widehat{\gamma}_n^H + z_\alpha \frac{\widehat{\gamma}_n^H}{\sqrt{k}}\right),$$

and

$$I_2(\alpha) := \left(\widehat{\gamma}_n - z_{\alpha/2}\frac{\widehat{\gamma}_n^H}{\sqrt{k}}, \widehat{\gamma}_n + z_{\alpha/2}\frac{\widehat{\gamma}_n^H}{\sqrt{k}}\right),$$

where z_ω $(0 < \omega < 1)$, is defined by $P(\mathcal{N}(0,1) \leq z_\omega) = 1-\omega$, i.e. z_ω is the $(1-\omega)$-quantile of the standard normal distribution. It is shown in Cheng and Peng (2001) that the corresponding coverage probabilities are

$$P(\gamma \in I_1(\alpha)) = 1-\alpha-\phi(z_\alpha)\left\{\frac{1+2z_\alpha^2}{3\sqrt{k}} - \frac{\rho d c^\rho}{(1-\rho)}\sqrt{k}\left(\frac{n}{k}\right)^\rho\right\}+o\left(\frac{1}{\sqrt{k}}+\sqrt{k}\left(\frac{n}{k}\right)^\rho\right)$$

and

$$P(\gamma \in I_2(\alpha)) = 1-\alpha+o\left(\frac{1}{\sqrt{k}}+\sqrt{k}\left(\frac{n}{k}\right)^\rho\right).$$

By minimizing the absolute coverage error for $I_1(\alpha)$, Cheng and Peng (2001) propose an optimal sample fraction

$$k^* := \begin{cases} \left(\dfrac{(1+2z_\alpha^2)(1-\rho)}{-3dc^\rho\rho(1-2\rho)}\right)^{1/(1-\rho)} n^{-\rho/(1-\rho)} & \text{if } d > 0, \\ \left(\dfrac{(1+2z_\alpha^2)(1-\rho)}{3dc^\rho\rho}\right)^{1/(1-\rho)} n^{-\rho/(1-\rho)} & \text{if } d < 0. \end{cases}$$

Notice that it is readily verified that $k^* = o\left(n^{-2\rho/(1-2\rho)}\right)$. Since k^* depends on quantities characterizing the unknown cdf F, Cheng and Peng (2001) introduce a plug-in estimate

$$\widehat{k}^* := \begin{cases} \left(\dfrac{(1+2z_\alpha^2)}{3\widehat{\delta}(1-2\widehat{\rho})}\right)^{1/(1-\widehat{\rho})} n^{-\widehat{\rho}/(1-\widehat{\rho})} & \text{if } \widehat{\delta} > 0, \\ \left(\dfrac{(1+2z_\alpha^2)}{-3\widehat{\delta}}\right)^{1/(1-\widehat{\rho})} n^{-\widehat{\rho}/(1-\widehat{\rho})} & \text{if } \widehat{\delta} < 0. \end{cases} \tag{2.24}$$

where

$$\widehat{\rho} := -\log\left(\left|\frac{\mathbb{M}_n^{(2)}\left(n/\left(2\sqrt{\log n}\right)\right) - 2\left\{\mathbb{M}_n^{(1)}\left(n/\left(2\sqrt{\log n}\right)\right)\right\}^2}{\mathbb{M}_n^{(2)}\left(n/\sqrt{\log n}\right) - 2\left\{\mathbb{M}_n^{(1)}\left(n/\sqrt{\log n}\right)\right\}^2}\right|\right)/\log 2,$$

and

$$\widehat{\delta} := (1-\widehat{\rho})(\log n)^{-\widehat{\rho}/2} \frac{\mathbb{M}_n^{(2)}\left(\frac{n}{\sqrt{\log n}}\right) - 2\left\{\mathbb{M}_n^{(1)}\left(\frac{n}{\sqrt{\log n}}\right)\right\}^2}{2\widehat{\rho}\left\{\mathbb{M}_n^{(1)}\left(\frac{n}{\sqrt{\log n}}\right)\right\}^2},$$

with $\mathbb{M}_n^{(r)}$ is given by equation (2.15).

2.5 POT Procedure

The POT is a statistical methodology very largely used in tail index and high quantile estimation. It is a central approach in the statistical analysis of extreme events and consists in using the GPD (defined by (1.15)) to approximate the distribution of excesses over a given (sufficiently high) threshold. Formally, this approximation is expressed by equation (1.33) of Theorem 1.18, which is a key result in the theory of extreme values. It makes a connection between the GEVD and the GPD and explains the importance of the latter distribution. This useful concept in the statistics of extremes schematically works as follows:

- **Step 1**: Select a high threshold u.
- **Step 2**: Fit a GPD to the excesses over u in order to get estimates for the shape and scale parameters γ and σ.

In the statistical literature, there exist several contributions on this modelling approach. For a detailed account, we refer the reader to the fundamental paper [**??**].

2.5.1 Fitting the GPD

Let $(X_1, ..., X_n)$ be a sample from a rv X with continuous cdf $F \in \mathcal{D}(H_\gamma)$.Fix a high threshold u and select, from the sample, only those observations $X_{i_1}, ..., X_{i_{N_u}}$ that exceed u. A GPD with parameters γ and $\sigma = \sigma(u)$ is expected to be a good approximation of the cdf F_u (defined by (1.12)) of the N_u excesses

$$Y_j := X_{i_j} - u > 0, j = 1, 2, ..., N_u.$$

Fitting a GPD to these excesses amounts to estimating parameters γ and σ based on $Y_1, ..., Y_{N_u}$.

2.5.2 Estimating GPD Parameters

ML estimation

As we applied two parameter estimation methods to GEV in Section 2.5, we use both two methods in the parameter estimation of GPD model. In case $\gamma \neq 0$, the likelihood function can be obtained directly by

$$L\left((\gamma,\sigma)\mid X\right) := \prod_{j=1}^{N_u}\left[\frac{1}{\sigma}\left(1+\frac{\gamma y_j}{\sigma}\right)^{-1/\gamma-1}\right].$$

The log-likelihood function is

$$l\left((\gamma,\sigma)\mid X\right) := -N_u \ln\sigma - \left(\frac{1}{\sigma}+1\right)\sum_{i=1}^{N_u}\ln\left(1+\frac{\gamma y_j}{\sigma}\right).$$

Taking partial derivatives of $l\left((\gamma,\sigma)\mid X\right)$ with respect to γ and σ, we obtain

$$\begin{cases} \frac{1}{\sigma^2}\sum_{j=1}^{N_u}\ln\left(1+\frac{\gamma y_j}{\sigma}\right) - \sum_{i=1}^{N_u}\frac{y_j}{\sigma+\gamma y_j} = 0 \\ -N_u + (1+\gamma)\sum_{i=1}^{N_u}\frac{y_j}{\sigma+\gamma y_j} = 0 \end{cases},$$

where $y_1,...,y_{N_u}$ is a realization of $Y_1,...,Y_{N_u}$. MLE $(\widehat{\gamma}_{N_u},\widehat{\sigma}_{N_u})$ of (γ,σ) as a solution of this system of equations. Notice that this system does not have explicit solutions and hence numerical methods are required to compute the estimate values $\widehat{\gamma}_{N_u}$ and $\widehat{\sigma}_{N_u}$. In Theorem 3.2 of [117], Smith shows the asymptotic normality of $(\widehat{\gamma}_{N_u},\widehat{\sigma}_{N_u})$ provided $\gamma > -1/2$. Specifically we have

$$\sqrt{N_u}\begin{pmatrix}\widehat{\gamma}_{N_u}-\gamma \\ \widehat{\sigma}_{N_u}-\sigma\end{pmatrix} \xrightarrow{d} \mathcal{N}_2\left(0,\mathbb{Q}^{-1}\right) \text{ as } N_u\to\infty, \tag{2.25}$$

where

$$\mathbb{Q}^{-1} := (1+\gamma)\begin{pmatrix}1+\gamma & -1 \\ -1 & 2\end{pmatrix}, \tag{2.26}$$

and $\mathcal{N}_2(\varepsilon,\omega)$ stands for the bivariate normal distribution with mean vector ε and covariance matrix ω. With this result, confidence intervals for parameter estimates are easily constructed.

PWM estimation

Similarly to the estimation of the GEVD parameters (see Subsection 2.3), Hosking and Wallis [78] proposed a PWM approach for the GPD parameters based on the following quantities:

$$w_r(\gamma,\sigma) = E\left[X\overline{\mathbb{G}}^r_{\gamma,\sigma}(x)\right] \text{ for } r = 1,2$$

where X has $\mathbb{G}_{\gamma,\sigma}$ as a cdf. Solving for $r = 0$ and 1, we immediately obtain

$$\gamma = \frac{w_0 - 4w_1}{w_0 - 2w_1}, \sigma = \frac{2w_0w_1}{w_0 - 2w_1}.$$

Replacing w_0 and w_1 by the respective empirical moments yields the PWM estimators $\widehat{\gamma}$ and $\widehat{\sigma}$ of the GPD parameters.

Estimating Distribution tails

By setting $x = u + y$ in relation (1.31), we may write the distribution tail as

$$\overline{F}(x) = \overline{F}_u(x-u)\,\overline{F}(u), \;\; u < x < x^*. \tag{2.27}$$

This means that estimates of the conditional tail $\overline{F}_u(x-u)$ and $\overline{F}(u)$ are needed in order to obtain an estimate for the (unconditional) tail $\overline{F}(x)$. After, we compute (by one of the above methods) estimates $\widehat{\gamma}_u$ and $\widehat{\sigma}_u$ for γ and σ respectively and by virtue of (1.33), we can estimate $\overline{F}_u(x-u)$ by

$$\widehat{\overline{F}}_u(x-u) := \overline{\mathbb{G}}_{\widehat{\gamma}_u,\widehat{\sigma}_u}(x-u) = \left(1 + \widehat{\gamma}_u \frac{x-u}{\widehat{\sigma}_u}\right)^{-1/\widehat{\gamma}_u}, u < x < x^*.$$

A natural estimate of $\overline{F}(u)$ is given by the sample edf or the empirical probability of exceedance

$$\widehat{\overline{F}}(u) := \widehat{\overline{F}}_n(u) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X>u\}} = \frac{N_u}{n}, u < x^*.$$

Putting all this together results in the following form for the distribution tail estimate

$$\widehat{\overline{F}}(x) = \frac{N_u}{n}\left(1 + \widehat{\gamma}_u \frac{x-u}{\widehat{\sigma}_u}\right)^{-1/\widehat{\gamma}_u}, u < x < x^*. \tag{2.28}$$

It is important to stress that this estimator is only valid for $x > u$.

2.5.3 Threshold Selection

The POT procedure supposes the existence of a suitable high threshold u for which the approximation of theorem (1.17) is good and above which we still have sufficient data to obtain accurate estimates for the shape and scale parameters. The choice of such a threshold is subject to a trade-off between bias and variance. Indeed, a value of u too high results in too few exceedances and consequently the estimators will have high variances and conversely if u is too small, then too many exceedances will make approximation (1.33) poor and the estimators will become biased. Therefore, the threshold has to be selected in an optimal manner. One tool of immediate use is available for this purpose. It is the emf-plot

$$\{(u, e_n(u)) : X_{1,n} \leq u \leq X_{n,n}\}.$$

where $e_n(u)$ is the empirical mef defined in (2.2). As we saw in Proposition 1.6, the emf of a GPD is linear in u. Therefore, we have to check the linearity of the plot above and choose u such that $e_n(x)$ is approximately linear for $x \geq u$. In other words, we select as threshold the value beyond which the emf-plot is (almost) linear. It is also common practice to fix threshold u at the $(k+1)$th largest observation $X_{n-k,n}$ and the problem becomes a matter of which value of k to take as an optimal choice.

2.6 Estimating High Quantiles

In the analysis of extremes, one is mainly concerned with the estimation of quantities related to rare events. In many areas of application, like for instance insurance, finance, hydrology and statistical quality control, a typical requirement is to find values, large enough, so that the chances of exceeding them are very small. We are then interested in estimating high quantiles.

Definition 2.7 *For the continuous cdf* $F(x)$ *the quantile* $x = x_p$ *of level* $(1-p)$, $p \in (0,1)$, *is the solution of the equation*

$$1 - F(x) = p.$$

The value x_p is the point that is exceeded with probability p. Using the functions introduced in definition 1.9, we define the $(1-p)$-quantile of F as

$$x_p := F^{\leftarrow}(1-p) = \mathbb{Q}(1-p) = \mathbb{U}(1/p). \tag{2.29}$$

2.6.1 Quantile Estimation

High quantile estimation plays an important role in the context of risk management where it is crucial to evaluate adequately the risk of a big loss that occurs very rarely. The main difficulty of this estimation is due to the fact that when p is very small, the point x_p is beyond the range of the sample $(X_1, ..., X_n)$ with drawn from an unknown cdf F.
As we use asymptotic theory, p must depend on the sample size n i.e., $p =: p_n$.
Two cases are possible for x_p, within and outside the sample.
If $p_n \to 0$ with $np_n \to c \in [1, \infty]$ as $n \to \infty$, the $(1 - p_n)$-quantile is within the sample, and
if $p_n \to 0$ with $np_n \to c \in [0, 1)$ as $n \to \infty$, the $(1 - p_n)$-quantile is outside the sample.
In other words, the within-sample estimation is possible up to the $(1/n)$-quantile whereas for $p < 1/n$, quantile estimates are beyond the range of the data. The latter case is the most relevant for purposes of real-life applications.
For the first situation, we have $\mathbb{Q}_n(s) = X_{n-i+1,n}$, then with $s = 1 - p = 1 - (i-1)/n$ for $i = 2, ..., n$, we get

$$\mathbb{Q}_n\left(1 - \frac{(i-1)}{n}\right) = X_{n-i+1,n}, i = 2, ..., n.$$

Hence, $X_{n-i+1,n}$ seems to be a natural estimator for the $\left(1 - \frac{(i-1)}{n}\right)$-quantile.
In the second case, we have to infer beyond the limits of the sample by extrapolating from intermediate quantiles. Obviously, this cannot be done without some kind of information on the tails of the distribution. An accurate modelling of the distribution tails is then needed. In other words, a good estimate of the tail index is essential to the process of extreme quantile estimation.
Since estimating high quantiles is directly linked to estimating the EVI, one would expect to find, in the literature, as many quantile estimators as there are tail index estimators. Moreover, confidence intervals for the quantile estimates are easily constructed since the proposed estimators are asymptotically normal.

Finally, endpoints (in case they are finite) are estimated as quantiles of order 1.

2.6.2 EVT-based Estimators

The GEVD $\mathcal{H}_\theta$, defined in section 2.2, is used to derive estimators for high quantiles based on a sample $(X_1, ..., X_n)$ with drawn from cdf F.

Case where F is exactly $\mathcal{H}_\theta$

The $(1-p)$-quantile $x_p = \mathcal{H}_\theta^{\leftarrow}(1-p)$ is naturally estimated by $\mathcal{H}_{\widehat{\theta}}^{\leftarrow}(1-p)$. That is

$$\widehat{x}_p := \begin{cases} \widehat{\mu} - \frac{\widehat{\sigma}}{\widehat{\gamma}}\left(1 - (-\log(1-p))^{-\widehat{\gamma}}\right) & \text{if } \gamma \neq 0 \\ \widehat{\mu} - \widehat{\sigma}\log\left(-\log(1-p)\right) & \text{if } \gamma = 0 \end{cases}, \tag{2.30}$$

where $\widehat{\theta} = (\widehat{\gamma}, \widehat{\mu}, \widehat{\sigma})$ is either obtained by the ML or the PWM methods discussed in section 2.2. When $\gamma < 0$, the endpoint is finite. It may be estimated by

$$\widehat{x}^* = \widehat{\mu} - \frac{\widehat{\sigma}}{\widehat{\gamma}}.$$

Case where $F \in DA(\mathcal{H}_\theta)$

Using relation (1.6) with large $u = a_n x + b_n$, we get a tail estimate of the form

$$\widehat{\overline{F}}(u) = \frac{1}{n}\left(1 + \widehat{\gamma}\frac{u - \widehat{b}_n}{\widehat{a}_n}\right)^{-1/\widehat{\gamma}},$$

where $\widehat{\gamma}$, $\widehat{b}_n$ and $\widehat{a}_n$ are appropriate estimates (based on the k upper order statistics $X_{n-k+1,n}, ..., X_{n,n}$) of the tail index γ and the norming constants a_n and b_n respectively. In case the $(1-p)$-quantile is within the range of the observations (i.e. $p \geq 1/n$), it readily can be estimated by

$$\widehat{x}_p := \widehat{a}_n \frac{(np)^{-\widehat{\gamma}} - 1}{\widehat{\gamma}} + \widehat{b}_n.$$

For the more typical case of outside the sample estimation (i.e. $p < 1/n$), we use a subsequence (n/k), where $k = k_n \to \infty$ and $k/n \to 0$ as $n \to \infty$.
Assuming, for notational convenience, that n/k is integer, we get

$$\widehat{x}_p := \widehat{a}_{n/k} \frac{(np/k)^{-\widehat{\gamma}} - 1}{\widehat{\gamma}} + \widehat{b}_{n/k}.$$

More details on this approach with the estimation of the norming constants can be found in Sections 6.4.1 and 6.4.3 of [46].
When $\gamma < 0$, the endpoint is finite. It may be estimated by

$$\widehat{x}^* = \widehat{b}_{n/k} - \frac{\widehat{a}_{n/k}}{\widehat{\gamma}}.$$

Next, we define large quantile estimators that are associated to the semiparametric estimators of the EVI γ introduced in section 2.3.

Weissman discussed in [123] the estimation of high quantiles for each one of the three standard extreme value distributions of Fisher-Tippett theorem 1.5 separately. For the Frechet class ($\gamma > 0$), the Weissman-type estimator of the $(1-p)$-quantile takes on the following form:

$$\widehat{x}_p^{(W)} := X_{n-k,n} \left(\frac{k}{np} \right)^{\widehat{\gamma}_n}, \tag{2.31}$$

where $\widehat{\gamma}_n$ is some consistent estimator of the tail index γ, often taken to be equal to Hill's estimator $\widehat{\gamma}_n^H$ resulting in the classical Weissman quantile estimator

$$\widehat{x}_p^{(W)} := X_{n-k,n} \left(\frac{k}{np} \right)^{\widehat{\gamma}_n^H}. \tag{2.32}$$

The asymptotic properties of this estimator are discussed and confidence intervals constructed under some conditions on F, k and p in, e.g. [90] and [95].
By observing that equation (2.31) may be rewritten into

$$\widehat{\mathbb{U}}(t) = \left(\frac{k}{n} \right)^{\widehat{\gamma}_n} X_{n-k,n} t^{\widehat{\gamma}_n}, t > n/k,$$

we readily obtain the following Weissman estimator for the distribution tail $\overline{F}(x)$:

$$\widehat{\overline{F}}(x) = \left(\frac{k}{n} \right) (X_{n-k,n})^{1/\widehat{\gamma}_n} x^{-\widehat{\gamma}_n}. \tag{2.33}$$

The estimator $\widehat{x}_p^{(RB)}$ of the $(1-p)$-quantile linked to reduced of bias estimator $\widehat{\gamma}_n^{(RB)}$ is derived in [65] (where a full analysis, with several asymptotic results and examples, is to be found) and is of the following form:

$$\widehat{x}_p^{(RB)} = \frac{X_{n-[k/2]:n} - X_{n-k:n}}{2^{\widehat{\gamma}^{RB}} - 1} \left(\frac{k}{np} \right)^{\widehat{\gamma}^{RB}} \times \left(1 - \frac{2^{\left(\widehat{\gamma}^{RB} + \widehat{\rho}\right)} - 1}{2^{\widehat{\gamma}^{RB}} - 1} \times \frac{\widehat{\gamma}^{RB} \widehat{\beta} \, (n/k)^{\widehat{\rho}}}{\widehat{\rho}} \right),$$

where $\widehat{\gamma}^{RB}, \widehat{\rho}, \widehat{\beta}$ are estimators of γ, ρ, β respectively, are given by formulas (2.13), (2.16) and (2.17) respectively.
The estimator $\widehat{x}_p^{(P)}$ of the $(1-p)$-quantile linked to Pickands estimator $\widehat{\gamma}_n^{(P)}$ is derived in [38] (where a full analysis, with several asymptotic results and examples, is to be found) and is of the following form:

$$\widehat{x}_p^{(P)} = X_{n-k+1,n} + \frac{(np/k)^{-\widehat{\gamma}_n^{(P)}} - 1}{1 - 2^{-\widehat{\gamma}_n^{(P)}}} \left(X_{n-k+1,n} - X_{n-2k+1,n} \right). \tag{2.34}$$

When $\gamma < 0$, the endpoint is finite. It may be estimated by

$$\widehat{x}_p^{(P)} = X_{n-k+1,n} + \frac{(X_{n-k+1,n} - X_{n-2k+1,n})}{2^{-\widehat{\gamma}_n^{(P)}} - 1}. \tag{2.35}$$

Similarly, the quantile of order $(1-p)$ is estimated in [37], on the basis the moment estimator $\widehat{\gamma}_n^{(M)}$, by

$$\widehat{x}_p^{(M)} = X_{n-k,n} + \frac{(np/k)^{-\widehat{\gamma}_n^{(M)}} - 1}{\widehat{\gamma}_n^{(M)}} \frac{X_{n-k,n}\mathbb{M}_n^{(1)}}{\varphi\left(\widehat{\gamma}_n^{(M)}\right)}, \tag{2.36}$$

where

$$\varphi(\gamma) = \begin{cases} 1, \gamma \geq 0 \\ 1/(1-\gamma), \gamma < 0 \end{cases}.$$

Actually, $\widehat{\gamma}_n^{(M)}$ could be replaced by any consistent estimator of γ in (2.36), leading to a more general estimator for x_p. Asymptotic results are established in [38] provided various conditions on F, k and p. This enables us to construct confidence intervals for x_p. When $\gamma < 0$, the endpoint is finite. It may be estimated by

$$\widehat{x}^{*(M)} = X_{n-k,n} + \left(1 - 1/\widehat{\gamma}_n^{(M)}\right) X_{n-k,n}\mathbb{M}_n^{(1)}.$$

2.6.3 POT-based Estimator

One of the advantages of the POT procedure is that it gives simple quantile estimators for different thresholds. Indeed, for a fixed threshold u, an estimator of quantile $x_p > u$ results immediately by inverting the tail estimate formula (2:20)

$$\widehat{x}_p := u + \frac{\widehat{\sigma}_u}{\widehat{\gamma}_u}\left(\left(\frac{N_u}{np}\right)^{\widehat{\gamma}_u} - 1\right).$$

Furthermore, the endpoint, in case it is finite $(\gamma < 0)$, is estimated by

$$\widehat{x}^* := u - \frac{\widehat{\sigma}_u}{\widehat{\gamma}_u}.$$

In practice, we usually fix u at the $(k+1)$th largest observation $X_{n-k,n}$ and a GPD is then fitted to the k excesses $(X_{n-k+1,n} - X_{n-k,n}), ..., (X_{n,n} - X_{n-k,n})$. The resulting estimators of parameters γ and σ are respectively denoted by $\widehat{\gamma}^{POT}$ and $\widehat{\sigma}^{POT}$. In this case, the quantile estimator is of the following form:

$$\widehat{x}_p^{POT} := X_{n-k,n} + \frac{\widehat{\sigma}^{POT}}{\widehat{\gamma}^{POT}}\left(\left(\frac{k}{np}\right)^{\widehat{\gamma}^{POT}} - 1\right), p < k/n.$$

The finite endpoint is then estimated by

$$\widehat{x}^{*POT} := X_{n-k,n} - \frac{\widehat{\sigma}^{POT}}{\widehat{\gamma}^{POT}}.$$

2.7 Estimating of the Location Parameter μ

Usually, the mean or the location paramater $\mu = EX_1$ is naturally estimated by the sample mean $\overline{X}_n$, which by the Central Limit Theorem (TCL) is asymptotically normal. Whereas for $\gamma \in (1/2, 1)$, X_1 has infinite variance and therfore the TCL is not valid anymore.

2.7.1 EVT Procedure: Estimator of Peng

In 2001, Peng [106] proposed an estimate of the mean as follows. For each $n \geq 1$, we have

$$\begin{aligned} \mu &= \int_0^1 \mathbb{Q}(s)\,ds \\ &= \int_0^{1-k/n} \mathbb{Q}(s)\,ds + \int_{1-k/n}^1 \mathbb{Q}(s)\,ds \\ &:= \mu_n^{(1)} + \mu_n^{(2)}, \end{aligned}$$

Then, Peng estimator of μ is defined as follows

$$\widehat{\mu}_n^P = \widehat{\mu}_n^{(1)} + \widehat{\mu}_n^{(2)}, \tag{2.37}$$

where $\widehat{\mu}_n^{(1)}$ is the *trimmed-mean* of the sample $X_1, ..., X_n$ is defined by

$$\widehat{\mu}_n^{(1)} = \frac{1}{n}\sum_{i=1}^{n-k} X_{i,n},$$

and

$$\widehat{\mu}_n^{(2)} := \frac{k}{n} X_{n-k+1,n} \frac{\widehat{\gamma}_n^H}{\widehat{\gamma}_n^H - 1},$$

is an estimator of $\mu_n^{(2)}$, where $\widehat{\gamma}_n^H$ is the Hill estimator of the index γ given by equation (2.11).

2.7.2 GPD Procedur: Estimator of Johansson

An alternative way to estimate μ may be made by GPD's approximation (see Johansson [58]).
Indeed, for each $n \geq 1$, we have

$$\begin{aligned}\mu &= \int_0^\infty x dF(x) \\ &= \int_0^{u_n} x dF(x) + \int_{u_n}^\infty x dF(x) \\ &= \int_0^{u_n} x dF(x) - \int_0^\infty (t+u_n) d\overline{F}((t+u_n)) \\ &= \mu_n^* + \tau_n,\end{aligned}$$

where

$$\mu_n^* = \int_0^{u_n} x dF(x) \text{ and } \tau_n = -\int_0^\infty (u_n+t) d\overline{F}(u_n+t).$$

Johansson's estimator of μ is given by

$$\widehat{\mu}_n^{(J)} = \int_0^{u_n} x dF_n(x) - \int_0^\infty (u_n+t) d\widehat{\overline{F}}(u_n+t),$$

where F_n is the empirical distribution function pertaining to the sample $X_1, X_2, ..., X_n$ and $\widehat{\overline{F}}(u_n + \cdot)$ is the estimator of distribution tailed $\overline{F}(u_n + \cdot)$. Integrating the last quation, we get

$$\widehat{\mu}_n^{(J)} = \widehat{\mu}_n^* + \widehat{\tau}_n, \tag{2.38}$$

where

$$\widehat{\mu}_n^* = \frac{1}{n}\sum_{i=1}^n X_i \mathbf{1}_{\{X_i \leq u_n\}} \text{ and } \widehat{\tau}_n = \widehat{p}_n \left(u_n + \frac{\widehat{\sigma}_n}{1-\widehat{\gamma}_n} \right),$$

with $\widehat{p}_n = \left(\sum_{i=1}^n 1_{\{X_i \geq u_n\}}\right)/n$ is an estimator of $p_n = P\left(X \geq u_n\right)$ and $\widehat{\gamma}_n, \widehat{\sigma}_n$ is estimators of γ, σ respectively.

Chapter 3

Financial Risk and Heavy Tails

3.1 Introduction

The measurement of financial risk has been one of the main preoccupations of actuaries and insurance practitioners for a very long time. Measures of financial risk manifest themselves explicitly in many different types of insurance problem, including the determination of reserves or capital, the setting of premiums and thresholds (e.g., for deductibles and reinsurance cedance levels), and the estimation of magnitudes such as expected claims, expected losses and probable maximum losses, they also manifest themselves implicitly in problems involving shortfall and ruin probabilities. In each of these cases, we are interested, explicitly or implicitly, in quantiles of some loss function or, more generally, in quantile-based risk measures.

Very many risk measures are proposed in the literature, the differences between them are the properties that they satisfy. Value at Risk (VaR) is one of most popular risk measures, due to its simplicity. VaR indicates the minimal loss incurred in the worse outcomes of the portfolios. But this risk measure is not always sub-additive, nor convex. To overcome this, Artzner, Delbaen, Elbner and Heath (1999) proposed the main properties that a risk measures must satisfy, thus establishing the notion of coherent risk measure.

After coherent risk measures and their properties were established, other classes of measures have been proposed, each with distinctive properties: convex (Föllmer and Shied, 2004), spectral (Acerbi, 2002) or deviation measures (Rockafellar et al. 2006).

Spectral risk measures are coherent risk measure that satisfies two additional

conditions. These measures have been applied to futures clearinghouse margin requirements in Cotter and Down (2006). Acerbi and Simonetti (2002) extend the results of Pflug and Rockafellar-Uryasev methodology to spectral risk measures.
A description of the axioms of risk pricing measures with many applications to insurance can be found in Wang, Young and Panjer (1997), and in the monograph by Kass, Goovaerts, Dhaene and Denuit (2001). The concept of distorted risk measures evolved from this line of work and ties in with the notion of capacity. Capacities are non-additive, monotone set functions which extend the notion of integral in a peculiar way. The evolution of this concept, from Choquet's work in the 1950's until the 1990's can be traced back from the review by Denneberg (1997).

This chapter provides an overview of the theory and estimation of these risks measures, and of their applications to insurance problems, focusing mainly on the VaR, coherent measures, spectral measures and distortion measures. In this chapter, we present estimates, based on EVT and POT method, for three major measures of market risk, namely the VaR, the CTE and the return level. Then, we use EVT results to derive an asymptotically normal estimator for the CTE for a loss distribution and we are establishing its confidence bounds (see [104]).

3.2 Types of financial risks

Financial institutions such as banks, hedge funds, and (re)insurance companies are exposed to several types of financial risks. Generally, they are classified into market risks, credit risks, liquidity risks, operational risks and legal risks. In a broader perspective, however, each of these corporations faces more general risks too, such as business risks and strategic risks. However, the daily business of financial institutions is concerned with managing an enormous number and variety of financial transactions and thus the financial risks are of key interest to the financial industry. The following description summarises the characteristics of the various financial risks.
Credit risk. This risk arises when a counterparty may fail or might be unwilling to meet its obligations and thus causes the asset holder to suffer financial loss. This class includes: downgrade risk, which refers to the risk that a counterparty might be downgraded by a rating agency; sovereign risk, which refers

to the default of a country; and settlement risk, which arises when there is non-simultaneous exchange of value (Bustany 1998).

Operational risk. This risk results from mistakes or failures in internal operations. It covers a wide area that can be divided into human/technology errors such as management failure, fraud, flawed system implementation, conducting business in an unethical or risky manner, and risks that are outside the control of the firm such as natural disasters and political or regulatory regime changes (Allen, Boudoukh, and Saunders 2004).

Legal risk. This risk is related to the legal uncertainties arising when a counterparty does not have the regulatory authority to enter financial transactions. It could also include activities that contravene government regulations, such as market manipulation and insider trading (Jorion 1997).

Liquidity risk. This risk consists of market/product liquidity risk and cash flow/funding liquidity risk. The latter relates to the inability to raise the necessary cash to roll over debt, or to meet the cash, margin, or collateral requirements of counterparties. Market/product liquidity risk is related to trading risk and arises when a financial institution is unable to execute a transaction in the prevailing market conditions. It may occur during market turmoil when liquidity dries out and the bid-ask spread increases dramatically. This risk is difficult to quantify and varies across market conditions (Crouhy, Galai, and Mark 2001).

Market risk. This risk arises from financial transactions and can be defined as the risk resulting from adverse movement in market prices. There are four major types of market risk (Basle Committee on Banking Supervision1996): Interest rate risk, Equity risk, Foreign exchange risk and Commodity price risk.

3.3 Risk Measure

Financial institutions have to regularly assess their capital adequacy to cover losses by reporting a number which reflects the minimum loss of their portfolio. In the past, people used to measure this loss by calculating the mean and variance. Unfortunately, these classical measures do not provide much information about extreme outcomes. Recently, in the early 1990's, new more reliable risk measures, like the VaR or the $\mathbb{CTE}$, have been introduced by a number of world financial institutions (J.P. Morgan, Bankers Trust,...).

The *Return Level* is another basic risk measure, mainly used to assess building

regulations for nuclear plants, dams, sea dykes, bridges,... Based on the largest observations of loss, all these measures are functions of extreme quantiles of loss distribution. Those whose are interested in full description of financial risk are referred to [16].
Risk measure can be formally described with the following definition.

Definition 3.1 *Let $(\Omega, \mathcal{A})$ a space mesurable where Ω is the outcome space and $\mathcal{A}$ is a $\sigma-$algebra defined on it. A risk is a random variable defined on $(\Omega, \mathcal{A})$, that is, $X : \Omega \to \mathbb{R}$ is a risk if $X^{-1}((-\infty, x]) \in \mathcal{A}$ for all $x \in \mathbb{R}$. A risk represents the final net loss of a position (contingency) currently held. When $X > 0$, we call it a loss, whereas when $X < 0$, we call it a gain.*

The class of all random variables on $(\Omega, \mathcal{A})$ is denoted by $\mathcal{X}$

Definition 3.2 *Any mapping $\rho : \mathcal{X} \to \mathbb{R} \cup \infty$ is called a risk measure. In case $\rho[X] = +\infty$, we say that the risk is unacceptable or non-insurable.*

3.4 Coherent Risk Measure

A risk measure is said to be coherent in the sense of Artzner et al. [5], [6] if it obeys the four properties or axioms that we now list.
Axiom T. Translation invariance: for all $X \in \mathcal{X}$ and all real numbers a, we have

$$\rho\left(X + a\right) = \rho\left(X\right) - a. \tag{3.1}$$

Translational invariance requires that the addition of a sure amount reduces pari passu the cash still needed to make our position acceptable, and its validity is obvious.

Remark 3.1 *Axiom **T** ensures that, for each X, $\rho\left(X + \rho\left(X\right)\right) = 0$. This equality has a natural interpretation in terms of the acceptance set associated to ρ.*

Axiom S. Subadditivity: for all $X_1, X_2 \in \mathcal{X}$

$$\rho\left(X_1 + X_2\right) \leq \rho\left(X_1\right) + \rho\left(X_2\right). \tag{3.2}$$

The subadditivity is the key property, this tells us that a portfolio made up of sub-portfolios will risk an amount which is no more than, and in some cases less than, the sum of the risks of the constituent subportfolios.

Axiom Ph. Positive homogeneity: for all $\lambda \geq 0$ and all $X \in \mathcal{X}$, $\rho(\lambda X) = \lambda\rho(X)$.
Positive homogeneity implies that the risk of a position is proportional to its scale or size, and makes sense if we are dealing with liquid positions in marketable instruments.
Axiom M. Monotonicity: for all X and $Y \in \mathcal{X}$ with $X \leq Y$, we have $\rho(X) \leq \rho(Y)$.
Monotonicity means that if Y has a greater value than X, then Y should have lower risk: this makes sense, because it means that less has to be added to Y than to X to make it acceptable, and the amount to be added is the risk measure.

Subadditivity axiom is the most important criterion we would expect a 'respectable' risk measure to satisfy. It reflects our expectation that aggregating individual risks should not increase overall risk, and this is a basic requirement of any 'respectable' risk measure, coherent or otherwise.

3.5 Representation of Coherent Risk Measures by Scenarios

These four axioms define the coherent measures of risks, which admit the following general representation:

Proposition 3.1 *Given the total return r on a reference investment, a risk measure ρ is coherent iff there exists a family $\mathcal{P}$ of probability measures on the set of states of nature, such that*

$$\rho(X) = \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[-X/r\right], \tag{3.3}$$

Thus, any coherent measure of risk appears as the expectation of the maximum loss over a given set of scenarios (the different probability measures $\mathbb{P} \in \mathcal{P}$). It is then obvious that the larger the set of scenarios, the larger the value of $\rho(X)$ and thus, the more conservative the risk measure.
The axioms of coherent risk measures have been very influential. These coherent risk measures can be used as capital requirements to regulate the risk assumed by market participants, traders, and insurance underwriters, as well as to allocate existing capitals. But we should realize that not all coherent risk measures are reasonable to use under certain practical situations.

Coherent risk measures were extended in general spaces by Delbaen [41].
Later were extended to convex risk measures, also called weakly coherent risk measures by relaxing the constraints of subadditivity and positive homogeneity, and instead requiring the following weaker condition:
Convexity (C):

$$\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y), \lambda \in [0,1].$$

Definition 3.3 *A map $\rho : \mathcal{X} \to \mathbb{R}$ will be called a convex risk measure if it satisfies the condition of convexity **(C)**, monotonicity **(M)**, and translation invariance **(T)**.*

Definition 3.4 *A map ρ: $\mathcal{X} \to \mathbb{R}$ satisfying the Subadditivity **(S)**, Positive homogeneity **(Ph)**, and the following two axioms*

(SH) Shift-invariance: $\rho\left(X+m\right) = \rho\left(X\right), \forall X \in \mathcal{X}, m \in \mathbb{R}$

(N) Nonegative: $\rho\left(X\right) \geq 0, \forall X \in \mathcal{X},$

are called deviation measures.

Standard deviation and semi-standard deviation are typical examples of this kind. Deviation measures and coherent risk measures are in fact mutually incompatible: there is no function can satisfy axioms **(C)** and **(SH)** at the same time.

3.6 Some Example of Risk Measures

Variance and standard deviation have been traditional risk measures in economics and finance since the pioneering work of Markowitz. The two risk measures exhibit a number of nice technical properties. For example, the variance of a portfolio return is the sum of the variance and covariance of the individual returns. Furthermore, variance can be used as a standard optimization function. Finally, there is a well established statistical methods to estimate variance and covariance. However, variance does not account for fat tails of the underlying distribution and therefore is inappropriate to describe the risk of low probability events, such as default risks. Secondly, variance penalizes ups and downs equally. Moreover, mean-variance decisions are usually not consistent with the expected

utility approach, unless returns are normally distributed or a quadratic utility function is chosen.
In the following, we shall consider some other familiar risk measures.

3.6.1 Value at Risk (VaR)

Value at Risk (VaR) was originally identified by the Group of Thirty (1993), a working group of academics, end-users, lawyers, dealers and financiers, whose major recommendation was to value positions on mark-to-market principles. It became popular in 1994 as the US investment bank J.P. Morgan made available to the public their own risk measurement system, called Risk Metrics (J.P. Morgan 1996).
Jorion (1997, p.19) gives the following definition of VaR.

Definition 3.5 *The risk measure VaR is summarises the expected maximum loss (or worst loss) over a target horizon within a given confidence interval.*

Using statistical language for the VaR.

Definition 3.6 (Quantile, VaR) *Let $t \in (0,1)$ be fixed and X be a real rv with df F in a probability space $(\Omega, \mathcal{A}, P)$. We then call*

$$VaR_t\left(x\right) = \mathbb{Q}_t\left(x\right) = F_X^{-1}\left(t\right), \tag{3.4}$$

the t-quantile of X. The VaR at confidence level t of X.

Remark 3.2 *The VaR is translation invariant, positive homogeneous and monotone. However, the subadditivity property (Axiom **S**) fails to hold for VaR in general, so VaR is not a coherent risk measure.*

Remark 3.3 *In fact, VaR is only subadditive in the restrictive case where the loss distribution is elliptically distributed, and this is of limited consolation because most real-world loss distributions are not elliptical ones. The failure of VaR to be subadditive is a fundamental problem because it means, in essence, that VaR has no claim to be regarded as a 'true' risk measure at all. The VaR is merely a quantile.*

In practice, $VaR(X)$ can be interpreted as the minimal amount of capital to be put back by an investor in order to preserve his solvency with a probability of at least t. However, standard deviation lacks the ability to describe the rare events and VaR is criticized because of its inability to aggregate the risk in a legal manner.
It is inappropriate to use VaR in practice because of its nonconvexity. It can have many local extremes, which will lead to unstable risk ranking. VaR is a model dependent risk measurement because, by definition, it depends on the initial reference probability.
At the latest in 1999, when coherent risk measures appeared, it became clear that VaR cannot be considered as an adequate risk measurement to allocate economic capital for financial institutions. In spite of this, as a compact representation of risk level, VaR has met the favor of regulatory agencies to measure downside risk and has been embraced by corporate risk managers as an important tool in overall risk management process.

3.6.2 Conditional Tail Expectation (CTE)

The CTE measure (also known as Tail-VaR or expected shortfall) is the conditional expectation of a loss variable X given that X exceeds a specified quantile (e.g., VaR_t). In other words, it measures the expected maximum loss in the 100% worst cases, over a given period.
The CTE because of these properties it has become a popular risk-measuring tool in insurance and finance industries. For example, use of the CTE for determining liabilities associated with variable life insurance and annuity products with guarantees is recommended in the United States (American Academy of Actuaries 2002) and required in Canada (Canadian Institute of Actuaries 2002).

Definition 3.7 *The Conditional Tail Expectation (CTE) at level t will be denoted by* $\mathbb{CTE}_t[X]$*, and is defined as*

$$\mathbb{CTE}_t[X] = E\left[X \mid X > VaR_t(X)\right], t \in (0,1)\,. \tag{3.5}$$

Equivalently,

$$\mathbb{CTE}_t[X] = VaR_t(X) + E\left[X - VaR_t(X)\right], t \in (0,1)\,.$$

This risk measure can be interpreted as the expected value of the shortfall in case the capital is set equal to $VaR_t(X) - t$.

This CTE risk measure is very familiar to actuaries, although it is usually known in actuarial circles as the Expected Tail Loss, Expected shortfall, Conditional VaR, Tail Conditional VaR, and Worst Conditional Expectation.

Proposition 3.2 *The Conditional Tail Expectation (CTE) is a coherent risk measure.*

Definition 3.8 *We assume that, the df F_X is continuous. Then, theThe Conditional Tail Expectation (CTE) is defined by*

$$\mathbb{CTE}_t[X] = \frac{1}{1-t}\int_t^1 VaR_t\,(x)\,dx. \tag{3.6}$$

Note that the $\mathbb{CTE}_t$ is always larger than the corresponding quantile.

3.6.3 Return Level

To define the return level, we use the block maxima observed over successive non overlapping time periods of equal length l (mostly consisting of one year), the distribution of the maxima is the (general) GEVD $\mathcal{H}_\gamma$.

Definition 3.9 *The return level is*

$$R_m = R_m(l) = \mathcal{H}_\theta^{\leftarrow}\left(1 - 1/m\right), m > 1. \tag{3.7}$$

is the expected level to be exceeded in one out of m periods of length l.

R_m is a conservative risk measure. It measures the maximum loss of a portfolio. If $l = 1$; a value of R_{20} equals 5% means that the maximum loss observed during a period of one year will exceed 5% once in twenty years on average.

3.7 Modeling Tails and Measures of Tail Risk

As we are particularly interested in the extreme risks faced by the clearinghouse, we model extreme returns using an Extreme Value (EV) approach. Perhaps the most suitable of these for our purposes is the Peaks over Threshold (POT) approach based on the Generalized Pareto distribution (GPD).

3.7.1 Tail probabilities and risk measures.

Assuming that u is sufficiently high, we observe that under equation (2.27), the distribution function for exceedances is given by: for $x \geq u$,

$$\overline{F}_u(x) = \overline{F}(u)\left(1+\gamma\frac{x-u}{\sigma}\right)^{-1/\gamma},$$

which, if we know $F(u)$, gives us a formula for tail probabilities. This formula may be inverted to obtain a high quantile of the underlying distribution, which we interpret as a VaR_t.
We will now derive analytical expressions for VaR_t and $\mathbb{CTE}_t$ defined respectively. For $t \geq F(u)$ we have that VaR is equal to

$$VaR_t = \mathbb{Q}_t(F) = u + \frac{\sigma}{\gamma}\left(\left(\frac{1-t}{\overline{F}(u)}\right)^{-\gamma} - 1\right) \tag{3.8}$$

Next, we may rewrite the The Conditional Tail Expectation as follows

$$\mathbb{CTE}_t[X] = VaR_t(X) + E\left[X - VaR_t(X) \mid X > VaR_t(X)\right], t \in (0,1), \tag{3.9}$$

where the second term on the right is the mean of excess distribution $F_{VaR_t}(y)$ over the threshold VaR_t. It is known that the mean excess function for the GPD with parameter $\gamma < 1$ is

$$e(z) = P(X - z \mid X > z) = \frac{\sigma+\gamma z}{1-\gamma}, \sigma+\gamma z > 0.$$

Similarly, the The Conditional Tail Expectation is given by

$$\mathbb{CTE}_t[X] = VaR_t(X) + \frac{\sigma+\gamma(VaR_t - u)}{1-\gamma} = \frac{VaR_t(X)}{1-\gamma} + \frac{\sigma-\gamma u}{1-\gamma}. \tag{3.10}$$

We now apply the block maxima method to our daily return data. The standard GEV is the limiting distribution of normalized extrema. Given that in practice we do not know the true distribution of the returns and, as a result, we do not have any idea about the norming constants a_n and b_n, we use the three parameter specification

$$\mathcal{H}_{\gamma,\sigma,\mu}(x) = \mathcal{H}_\gamma\left(\frac{x-\mu}{\sigma}\right), x \in D, D = \begin{cases}]-\infty, \mu-\sigma/\gamma[& \gamma < 0 \\ ]-\infty, +\infty[& \gamma = 0 \\ ]\mu-\sigma/\gamma, +\infty[& \gamma > 0 \end{cases},$$

of the GEV, which is the limiting distribution of the unnormalized maxima. The two additional parameters μ and σ are the location and the scale parameters representing the unknown norming constants.
The return level R_m is

$$R_m = \mathcal{H}^{\leftarrow}_{\gamma,\sigma,\mu}\left(1 - 1/m\right) = \begin{cases} \mu - \frac{\sigma}{\gamma}\left(1 - \left(-\log\left(1 - \frac{1}{m}\right)^{-\gamma}\right)\right) & \gamma \neq 0 \\ \mu - \sigma \log\left(-\log\left(1 - \frac{1}{m}\right)\right) & \gamma = 0 \end{cases}.$$

3.8 Spectral Risk Measures

Spectral risk measures are a generalization of the previous risk measures, in which the distribution function is pre-multiplied with a admissible risk aversion function which allow to introduce a subjective risk weight.
The spectral risk measures proposed by Acerbi (2002). Consider a risk measure M_ϕ defined by:

$$M_\phi = \int_0^1 \mathbb{Q}_p \phi\left(p\right) dp$$

where $\mathbb{Q}_p$ is the p loss quantile, $\phi\left(p\right)$ is a weighting function called the risk aversion function defined over p, and p is a range of cumulative probabilities $p \in [0, 1]$.
Following Acerbi (2004), the risk measure M_ϕ is coherent if and only if $\phi\left(p\right)$ satisfies the following properties:

1 $\phi\left(p\right) \geq 0$: weights are always non-negative.

2 $\int_0^1 \phi\left(p\right) dp = 1$: weights sum to one.

3 $\phi'\left(p\right) \geq 0$: higher losses have weights that are higher than or equal to those of smaller losses.

The CTE is a special case of M_ϕ obtained by setting $\phi\left(p\right)$ to the following:

$$\phi\left(p\right) = \frac{1}{1-t} 1_{\{p>t\}}.$$

3.9 Distortion Risk Measures

Distortion risk measures are a particular class of risk measures that have been extensively studied in actuarial literature in connection with the axiomatic theory of premium calculation, they were introduced by Denneberg (1990) and Wang (1996) and have been applied to a wide variety of insurance problems, most particularly distortion risk measures form an important class, they include Value at Risk, Conditional Tail Expectation and Wang's PH transform premium principle. Later in this section we review distortion functions and distortion risk measures. The objective of this section is to lay out the relationship between the characteristics of these risk measures and the criteria for coherence proposed by Artzner et al (1999), and we see that a distortion risk measure is coherent iff the associated distortion function is concave.

Definition 3.10 *A distortion risk measure is the expected loss under a transformation of the cumulative density function known as a distortion function, and the choice of distortion function determines the risk measure. More formally, if $F(x)$ is some cdf, the transformation $F^*(x) = g(F(x))$ is a distortion function if $g : [0,1] \to [0,1]$ is an increasing function with $g(0) = 0$ and $g(1) = 1$.*

The distortion risk measure is then the expectation of the random loss X using probabilities obtained from $F^*(x)$ rather than $F(x)$. Like coherent risk measures, distortion risk measures have the properties of monotonicity, positive homogeneity, and translational invariance.
A number of risk measures can be expressed as the expectation of the loss random variable under a change of measure accomplished using a distortion function. That is, for a risk measure Π, the price associated with a loss random variable $X \geq 0$ with distribution function F is

$$\Pi(x) = \int_0^{+\infty} g\left(\overline{F}(x)\right) dx. \tag{3.11}$$

Several risk measures of this form are discussed by Wirch and Hardy (1999). These authors note that, when g is concave, the risk measure is coherent.
We now present an alternative representation of risk measures originally expressed using distorted probabilities. This representation is convenient in developing empirical estimators of the risk measures.

Lemma 3.1 *Let X be a real-valued rv with df F, and let g be an increasing* *differentiable function with* $g(0) = 0$ *and* $g(1) = 1$*. Then*

$$\Pi(F) = \int_{-\infty}^{0} \left[g\left(\overline{F}(x)\right) - 1 \right] dx + \int_{0}^{+\infty} g\left(\overline{F}(x)\right) dx = \int_{0}^{1} F^{-1}(s)\, \Psi(s)\, ds, \tag{3.12}$$

where

$$\Psi(s) = g'(1-s).$$

3.9.1 Examples of distortion function

1. The VaR Measure: When the binary distortion function:

$$g(x) = \mathbf{1}_{\{x \geq t\}}, t \in (0,1).$$

2. The CTE Measure based on the distortion function:

$$\begin{cases} g(x) \geq \frac{x-t}{1-t} 1 \text{ if } x \geq t \\ g(x) = 0 \text{ if } x < t \end{cases}.$$

3. THE **PHT**-measure (Proportional Hasard Transform measure): When

$$g(s) = s^r, \text{with } 0 < r < 1.$$

4. The **WT**-Measure (Wang Tsansform measure): When

$$g(s) = \Phi\left(\Phi^{-1}(s) + \lambda\right),$$

where Φ is the standard normal df, and λ is a parameter that reflects the systematic risk of X.

5. Dual-power transform: The dual power transform is defined for parameter $r \geq 1$ with

$$g(s) = 1 - (1-s)^r.$$

6. Gini's principle:

$$g(s) = (1+r)x - rx^2.$$

7. Dennensberg's absolute deviation principle:

$$g(s) = \begin{cases} (1+r)x, \text{ for } 0 \leq s \leq 1/2 \\ r + (1-r)x, \text{ for } 1/2 \leq s \leq 1 \end{cases}.$$

8. Square-root transform:
$$g(s) = \frac{\sqrt{1-\ln(r)x}-1}{\sqrt{1-\ln(r)}-1}.$$

9. Exponential transform:
$$g(s) = \frac{1-r^x}{1-r}.$$

10. Logarithmic transform:
$$g(s) = \frac{\ln(1-\ln(r)x)}{\ln(1-\ln(r))}.$$

3.10 Estimations

3.10.1 Empirical Estimators for VaR and CTE.

By using the empirical quantile estimation we established empirical estimations of the VaR and $\mathbb{CTE}$ as follows
$$VaR_{t,n} = \mathbb{Q}_n(t) = X_{i,n}, t \in \left(\frac{i-1}{n}, \frac{i}{n}\right)$$
where $X_{i,n}$ is the order statistic of the iid rv's $X_1, ..., X_n$ with df F.
$$\mathbb{CTE}_{t,n} = \sum_{i=[nt]}^{n} X_{i,n}/\left(n-[nt]\right)$$

3.10.2 POT-Based estimator of VaR and CTE.

With equation (2.28), we have
$$\widehat{\overline{F}}(x) = \frac{N_u}{n}\left(1+\widehat{\gamma}\frac{x-u}{\widehat{\sigma}}\right)^{-1/\widehat{\gamma}}, u < x < x^*.$$
where n the sample size and N_u is the number of observations in excess of the u threshold and $\widehat{\gamma}$, $\widehat{\sigma}$ are estimates of γ and σ respectively, then we obtain an estimator of the VaR_t as follow
$$\widehat{VaR}_t = u + \frac{\widehat{\sigma}}{\widehat{\gamma}}\left(\left(\frac{n}{N_u}(1-t)\right)^{-\widehat{\gamma}} - 1\right). \quad (3.13)$$
Next, with equation 3.10, The Conditional Tail Expectation is estimated by
$$\widehat{\mathbb{CTE}}_t[X] = \frac{\widehat{VaR}}{1-\widehat{\gamma}} + \frac{\widehat{\sigma}-\widehat{\gamma}u}{1-\widehat{\gamma}} \quad (3.14)$$

3.10.3 The Hill Estimator Approach of the VaR

By using the Weissman estimator of the high quantile see equation (2.31), we have the $t-$quantile x_t

$$\widehat{x}_t^H = X_{n-k,n}\left(\frac{n}{k}(1-t)\right)^{-\widehat{\gamma}_n^H} = X_{n-k,n} + X_{n-k,n}\left[\left(\frac{n}{k}(1-t)\right)^{-\widehat{\gamma}_n^H} - 1\right],$$

where $\widehat{\gamma}_n^H$ is the Hill estimator (2.11) of the index γ.
Then, with the Value-at-Risk notation, we have

$$\widehat{VaR}_t^H = X_{n-k,n} + X_{n-k,n}\left[\left(\frac{n}{k}(1-t)\right)^{-\widehat{\gamma}_n^H} - 1\right]$$

3.10.4 Estimate of the return level

An estimator of R_m is defined as follow

$$\widehat{R}_m = \begin{cases} \widehat{\mu} - \frac{\widehat{\sigma}}{\widehat{\gamma}}\left(1-\left(-\log\left(1-\frac{1}{m}\right)^{-\widehat{\gamma}}\right)\right) & \gamma \neq 0 \\ \widehat{\mu} - \widehat{\sigma}\log\left(-\log\left(1-\frac{1}{m}\right)\right) & \gamma = 0 \end{cases}$$

Where $\widehat{\mu}, \widehat{\sigma}$ and $\widehat{\gamma}$ are estimators of μ, σ and γ respectively.

3.10.5 Non Parametric Estimation of the Measure of Distortion

Indeed, if in equation (3.12) we replace F by the empirical distribution function $\widehat{F}_n$, then the integral $\int_0^1 \widehat{F}_n^{-1}(s)\,\Psi(s)\,ds$, becomes, $\sum_{i=1}^n \left[\int_{(i-1)/n}^{i/n} \Psi(s)\,ds\right] X_{i,n}$, where $X_{1,n}, ..., X_{n,n}$ denote the ordered values of data $X_1, ..., X_n$. Hence, the empirical estimator of a risk measure $\Pi(F)$ is given by

$$\Pi(\widehat{F}_n) = \sum_{i=1}^n c_{in} X_{i,n}, \tag{3.15}$$

where $C_{in} = \int_{(i-1)/n}^{i/n} \Psi(s)\,ds$.
Jones and Zitikis (2003) employ asymptotic theory for L-statistics to prove that, for underlying distributions with a sufficient number of finite moments and under certain regularity conditions on function Ψ, the empirical estimator $\Pi(\widehat{F}_n)$ of a risk measure $\Pi(F)$ is strongly consistent and asymptotically normal with the mean $\Pi(F)$ and the variance $\sigma(\Psi, \Phi)/n$, where

$$\sigma^2(\Psi, \Phi) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left[F(\min(x,y)) - F(x)\,F(y)\right]\Psi(F(x))\,\Phi(F(y))\,dxdy. \tag{3.16}$$

3.11 Estimation of the Conditional Tail Expectation in the Case of Heavy-Tailed Losses

Naturally, the CTE is unknown since the cdf F is unknown. Hence, it is desirable to establish appropriate statistical inferential results such as confidence intervals for $\mathbb{CTE}(t)$ with specified confidence levels and margins of error. We shall next show how one can accomplish this task, initially assuming the classical moment assumption $\mathbf{E}[X^2] < \infty$. Namely, suppose that we have independent rv's X_1, X_2, $\ldots$, each with the cdf F, and let $X_{1:n} < \cdots < X_{n:n}$ denote the order statistics of $X_1, \ldots, X_n$. It is natural to define an empirical estimator of $\mathbb{CTE}(t)$ by the formula

$$\overline{\mathbb{CTE}}_n(t) = \frac{1}{1-t}\int_t^1 \mathbb{Q}_n(s)ds, \tag{3.17}$$

where $\mathbb{Q}_n(s)$ is the empirical quantile function, which is equal to the i^{th} order statistic $X_{i:n}$ for all $s \in ((i-1)/n, i/n]$, and for all $i = 1, \ldots, n$. The asymptotic behavior of the estimator $\overline{\mathbb{CTE}}_n(t)$ has been studied by Brazauskas *et al.* (2008), and we next formulate their most relevant for our paper result as a theorem.

Theorem 3.1 *Assume that* $\mathbf{E}[X^2] < \infty$*. Then for every* $t \in (0,1)$ *we have the asymptotic normality statement*

$$\sqrt{n}\big(\overline{\mathbb{CTE}}_n(t) - \mathbb{CTE}(t)\big)(1-t) \xrightarrow{d} \mathcal{N}(0, \sigma^2(t))$$

when $n \to \infty$*, where the asymptotic variance* $\sigma^2(t)$ *is given by the formula*

$$\sigma^2(t) = \int_t^1 \int_t^1 \left(\min(x,y) - xy\right) dQ(x)dQ(y).$$

The assumption $\mathbf{E}[X^2] < \infty$ is, however, quite restrictive as the following example shows. Suppose that F is the Pareto cdf with an index $\gamma > 0$, that is, $1 - F(x) = x^{-1/\gamma}$ for all $x \geq 1$. Let us focus on the case $\gamma < 1$, since when $\gamma \geq 1$, then $\mathbb{CTE}(t) = +\infty$ for every $t \in (0,1)$. Theorem 3.1 covers only the values $\gamma \in (0, 1/2)$ due the requirement $\mathbf{E}[X^2] < \infty$. When $\gamma \in [1/2, 1)$, we have $\mathbf{E}[X^2] = \infty$ but, nevertheless, $\mathbb{CTE}(t)$ is well defined and finite since $\mathbf{E}[X] < \infty$. Analogous remarks hold for other distributions with Pareto-like tails. We shall tackle this issue in the case of more general distributions than the just noted heavy-tailed Pareto.

3.11.1 Construction of a new CTE estimator

We have already noted that the 'old' estimator $\overline{\mathbb{CTE}}_n(t)$ does not yield asymptotic normality beyond the condition $\mathbf{E}[X^2] < \infty$. For this reason we next construct an alternative CTE estimator, which takes into account different asymptotic properties of moderate and high quantiles in the case of heavy-tailed distributions. Hence, from how on we assume that $\gamma \in (1/2, 1)$. Before indulging ourselves into construction details, we first formulate the new CTE estimator:

$$\widetilde{\mathbb{CTE}}_n(t) = \frac{1}{1-t} \int_t^{1-k/n} \mathbb{Q}_n(s)ds + \frac{kX_{n-k,n}}{n\,(1-t)\,(1-\widehat{\gamma})}, \tag{3.18}$$

where $\widehat{\gamma}$ is an estimator of the tail index $\gamma \in (1/2, 1)$, we use the Hill (1975) estimator (2.11).

We have based the construction of $\widetilde{\mathbb{CTE}}_n(t)$ on the recognition that one should estimate moderate and high quantiles differently when the underlying distribution is heavy-tailed. For this, we first recall that the high quantile $\mathbb{Q}_s$ is, by definition, equal to $\mathbb{Q}(1-s)$ for sufficiently small s. For an estimation theory of high quantiles in the case of heavy-tailed distributions we refer to, e.g., Weissman (1978), Dekkers and de Haan (1989), Matthys and Beirlant (2003), Gomes *et al.* (2005), and references therein. We shall use Weissman's estimator

$$\widetilde{\mathbb{Q}}_s = (k/n)^{\widehat{\gamma}} X_{n-k:n} s^{-\widehat{\gamma}}, \quad 0 < s < k/n,$$

of the high quantile $\mathbb{Q}_s$, where the integers $k = k_n \in \{1, \ldots, n\}$ are such that $k \to \infty$ and $k/n \to 0$ when $n \to \infty$. Then we write $\mathbb{CTE}(t)$ as the sum of $\mathbb{C}_{1,n}(t)$ and $\mathbb{C}_{2,n}(t)$, which are defined together with their empirical estimators as follows:

$$\mathbb{C}_{1,n}(t) = \frac{1}{1-t} \int_t^{1-k/n} \mathbb{Q}(s)ds \approx \frac{1}{1-t} \int_t^{1-k/n} \mathbb{Q}_n(s)ds = \widetilde{\mathbb{C}}_{1,n}(t)$$

and

$$\mathbb{C}_{2,n}(t) = \frac{1}{1-t} \int_{1-k/n}^{1} \mathbb{Q}(s)ds \approx \frac{1}{1-t} \int_{1-k/n}^{1} \widetilde{\mathbb{Q}}_{1-s}ds = \widetilde{\mathbb{C}}_{2,n}(t).$$

Simple integration gives the formula

$$\widetilde{\mathbb{C}}_{2,n}(t) = \frac{kX_{n-k,n}}{n\,(1-\widehat{\gamma})\,(1-t)}.$$

Consequently, the sum $\widetilde{\mathbb{C}}_{1,n}(t) + \widetilde{\mathbb{C}}_{2,n}(t)$ is an estimator of $\mathbb{CTE}(t)$, and this is exactly the estimator $\widetilde{\mathbb{CTE}}_n(t)$ introduced above. We shall investigate asymptotic normality of the new estimator in the next section, accompanied with an illustrative simulation study.

3.11.2 Main theorem and its practical implementation

The Next Theorem 3.2 below, establishes asymptotic normality of the new estimator $\widetilde{\mathbb{CTE}}_n(t)$.

Theorem 3.2 *Assume that the cdf F satisfies condition (1.39) with $\gamma \in (1/2, 1)$. Then for any sequence of integers $k = k_n \to \infty$ such that $k/n \to 0$ and $k^{1/2}A(n/k) \to 0$ when $n \to \infty$, we have that*

$$\frac{\sqrt{n}\big(\widetilde{\mathbb{CTE}}_n(t) - \mathbb{CTE}(t)\big)(1-t)}{\sqrt{k/n}X_{n-k:n}} \xrightarrow{d} \mathcal{N}\left(0, \sigma_\gamma^2\right) \tag{3.19}$$

for any fixed $t \in (0,1)$, where the asymptotic variance σ_γ^2 is given by the formula

$$\sigma_\gamma^2 = \frac{\gamma^4}{(1-\gamma)^4(2\gamma-1)}.$$

3.11.3 Simulation study

To discuss practical implementation of Theorem 3.2, we first fix a significance level $\varsigma \in (0,1)$ and use the classical notation $z_{\varsigma/2}$ for the $(1-\varsigma/2)$-quantile of the standard normal distribution $\mathcal{N}(0,1)$. Given a realization of the random variables $X_1, \ldots, X_n$ (e.g., claim amounts), which follow a cdf F satisfying the conditions of Theorem 3.2, we construct a level $1-\varsigma$ confidence interval for $\mathbb{CTE}(t)$ as follows. First, we need to choose an appropriate number k of extreme values. Since Hill's estimator has in general a substantial variance for small k and a considerable bias for large k, we search for a k that balances between the two shortcomings, which is indeed a well-known hurdle when estimating the tail index. To resolve this issue, several procedures have been suggested in the literature, and we refer to, e.g., Dekkers and de Haan (1993), Drees and Kaufmann (1998), Danielsson *et al.* (2001), Cheng and Peng (2001), Neves and Fraga Alves (2004), Gomes *et al.* (2009), and references therein. In our current study we employ the method of Cheng and Peng (2001) for deciding on an appropriate value k^* of k. Having computed Hill's estimator and consequently determined $X_{n-k^*:n}$, we then compute the corresponding values of $\widetilde{\mathbb{CTE}}_n(t)$ and $\sigma^2_{\widehat{\gamma}_n}$, and denote them by $\widetilde{\mathbb{CTE}}^*_n(t)$ and $\sigma^{2*}_{\widehat{\gamma}_n}$, respectively. Finally, using Theorem 3.2 we arrive at the following $(1-\varsigma)$-confidence interval for $\mathbb{CTE}(t)$:

$$\widetilde{\mathbb{CTE}}^*_n(t) \pm z_{\varsigma/2}\frac{\sqrt{k^*/n}X_{n-k^*:n}.\sigma^*_{\widehat{\gamma}_n}}{(1-t)\sqrt{n}}. \tag{3.20}$$

To illustrate the performance of this confidence interval, we have carried out a small-scale simulation study based on the Pareto cdf $F(x) = 1 - x^{-1/\gamma}, x \geq 1$, with the tail index $\gamma = 2/3$. The levels have been set to $t = 0.75$ and 0.90. We then generated 200 independent replicates of three samples of sizes $n = 1000$, 2000, and 5000. For every simulated sample, we obtained estimates $\widetilde{\mathbb{CTE}}_n(t)$. Then we calculated the arithmetic averages over the values from the 200 repetitions, with the absolute error (error) and root mean squared error (rmse) of the new estimator $\widetilde{\mathbb{CTE}}_n(t)$ reported in Tables (3.1), (3.2) for $\gamma = 2/3$ and Tables (3.3), (3.4) for $\gamma = 3/4$, where we also report 95%-confidence intervals (3.20) with their lower and upper bounds, coverage probabilities.
We note emphatically that the above coverage probabilities and lengths of confidence intervals can be improved by employing more precise but, naturally, considerably more complex estimators of the tail index. Many of such estimators are described in the monographs by Beirlant et al.(2004), Castillo et al. (2005), de Haan and Ferreira (2006) [70], Resnick (2007). Since the publication of these monographs, numerous journal articles have appeared on the topic.

$t = 0.75$			$C(t) = 7.005$				
n	k^*	$\widetilde{C}_n(t)$	error	rmse	lcb	ucb	cprob
1000	54	6.876	0.045	0.303	6.356	7.397	0.839
2000	100	6.831	0.025	0.231	6.463	7.199	0.882
5000	219	7.119	0.016	0.194	6.881	7.357	0.895

Table 3.1: Point estimates and 95%-confidence intervals for $\mathbb{CTE}(0.75)$ with tail index 2/3.

$t = 0.90$			$C(t) = 12.533$				
n	k*	$\widetilde{C}_n(t)$	error	rmse	lcb	ucb	cprob
1000	46	12.753	0.017	0.534	12.241	13.269	0.847
2000	97	12.487	0.003	0.294	12.137	12.838	0.841
5000	219	12.461	0.005	0.236	12.246	12.676	0.887

Table 3.2: Point estimates and 95%-confidence intervals for $\mathbb{CTE}(0.90)$ with tail index 2/3.

Figures 3.2 and 3.1 illustrates the performance and comparaison of the sample estimator (3.17) and the new estimator (3.18) of $\mathbb{CTE}(t)$ with respect to the

$t = 0.75$		$C(t) = 9.719$					
n	k*	$\widetilde{C}_n(t)$	error	rmse	lcb	ucb	cprob
1000	51	9.543	0.018	0.582	8.589	9.543	0.854
2000	104	9.808	0.009	0.466	9.150	10.466	0.888
5000	222	9.789	0.007	0.410	9.363	10.215	0.915

Table 3.3: Point estimates and 95%-confidence intervals for $\mathbb{CTE}(0.75)$ with tail index 3/4

$t = 0.90$		$C(t) = 18.494$					
n	k*	$\widetilde{C}_n(t)$	error	rmse	lcb	ucb	cprob
1000	48	18.199	0.015	0.989	17.437	18.960	0.874
2000	96	18.696	0.011	0.858	18.052	19.340	0.895
5000	229	18.541	0.002	0.798	18.092	18.990	0.925

Table 3.4: Point estimates and 95%-confidence intervals for $CTE(0.90)$ with tail index 3/4

sample size $n \geq 1$ with four different values of t is $t = 0.25$, $t = 0.50$, $t = 0.75$, and $t = 0.90$, such that the Figures 3.2 for the tail index $\gamma = 2/3$ and .Figures 3.1 is for the tail index $\gamma = 3/4$.

3.11.4 Proof of Theorem 3.2

We start the proof of Theorem 3.2 with the decomposition

$$\big(\widetilde{\mathbb{CTE}}_n(t) - \mathbb{CTE}(t)\big)(1-t) = A_{n,1}(t) + A_{n,2}, \tag{3.21}$$

where

$$A_{n,1}(t) = \int_t^{1-k/n} (\mathbb{Q}_n(s) - \mathbb{Q}(s))\, ds,$$

$$A_{n,2} = \frac{k/n}{1-\widehat{\gamma}_n} X_{n-k:n} - \int_{1-k/n}^{1} \mathbb{Q}(s) ds.$$

We shall show below that there are Brownian bridges $\mathbb{B}_n$ such that

$$\frac{\sqrt{n} A_{n,1}(t)}{\sqrt{k/n}\mathbb{Q}(1-k/n)} = -\frac{\int_0^{1-k/n} \mathbb{B}_n(s) d\mathbb{Q}(s)}{\sqrt{k/n}\mathbb{Q}(1-k/n)} + o_{\mathbf{P}}(1) \tag{3.22}$$

and

$$\frac{\sqrt{n} A_{n,2}}{\sqrt{k/n}\mathbb{Q}(1-k/n)} = \frac{\gamma^2}{(1-\gamma)^2}\sqrt{\frac{n}{k}}\left(\mathbb{B}_n(1-k/n) - \frac{1}{\gamma}\int_{1-k/n}^{1} \frac{\mathbb{B}_n(s)}{1-s} ds\right) + o_{\mathbf{P}}(1). \tag{3.23}$$

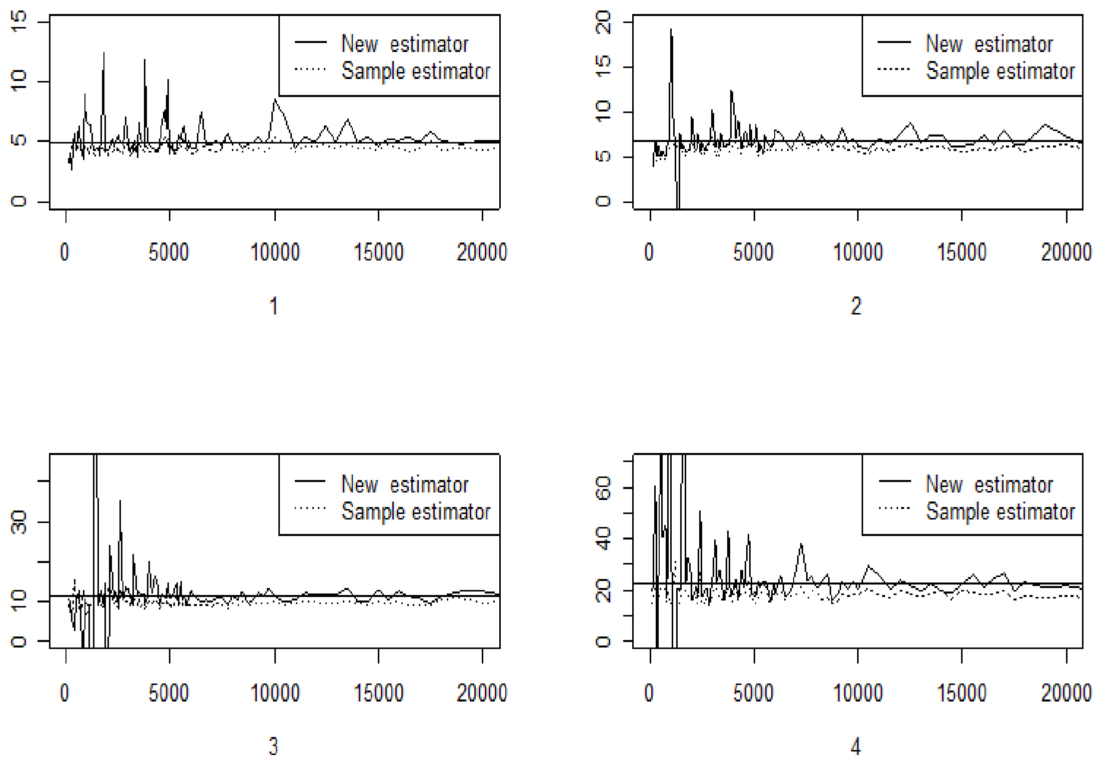

Figure 3.1: Values of the $\mathbb{CTE}$ estimator $\widetilde{\mathbb{CTE}}_n$ (vertical axis) versus sample sizes n (horizontal axis) evaluated at the levels $t = 0.25$, $t = 0.50$, $t = 0.75$, and $t = 0.90$ (panels 1–4, respectively) in the Pareto case with the tail index $\gamma = 3/4$.

Assuming for the time being that statements (3.22) and (3.23) hold, we complete the proof of Theorem 3.2. To simplify the presentation, we use the notation:

$$W_{1,n} = -\frac{\int_0^{1-k/n} \mathbb{B}_n(s)d\mathbb{Q}(s)}{\sqrt{k/n}\mathbb{Q}(1-k/n)},$$

$$W_{2,n} = \frac{\gamma^2}{(1-\gamma)^2}\sqrt{\frac{n}{k}}\,\mathbb{B}_n(1-k/n),$$

$$W_{3,n} = -\frac{\gamma}{(1-\gamma)^2}\sqrt{\frac{n}{k}}\int_{1-k/n}^1 \frac{\mathbb{B}_n(s)}{1-s}ds.$$

Hence, we have the asymptotic representation

$$\frac{\sqrt{n}\big(\widetilde{\mathbb{CTE}}_n(t) - \mathbb{CTE}(t)\big)(1-t)}{\sqrt{k/n}\mathbb{Q}(1-k/n)} = W_{1,n} + W_{2,n} + W_{3,n} + o_{\mathbf{P}}(1).$$

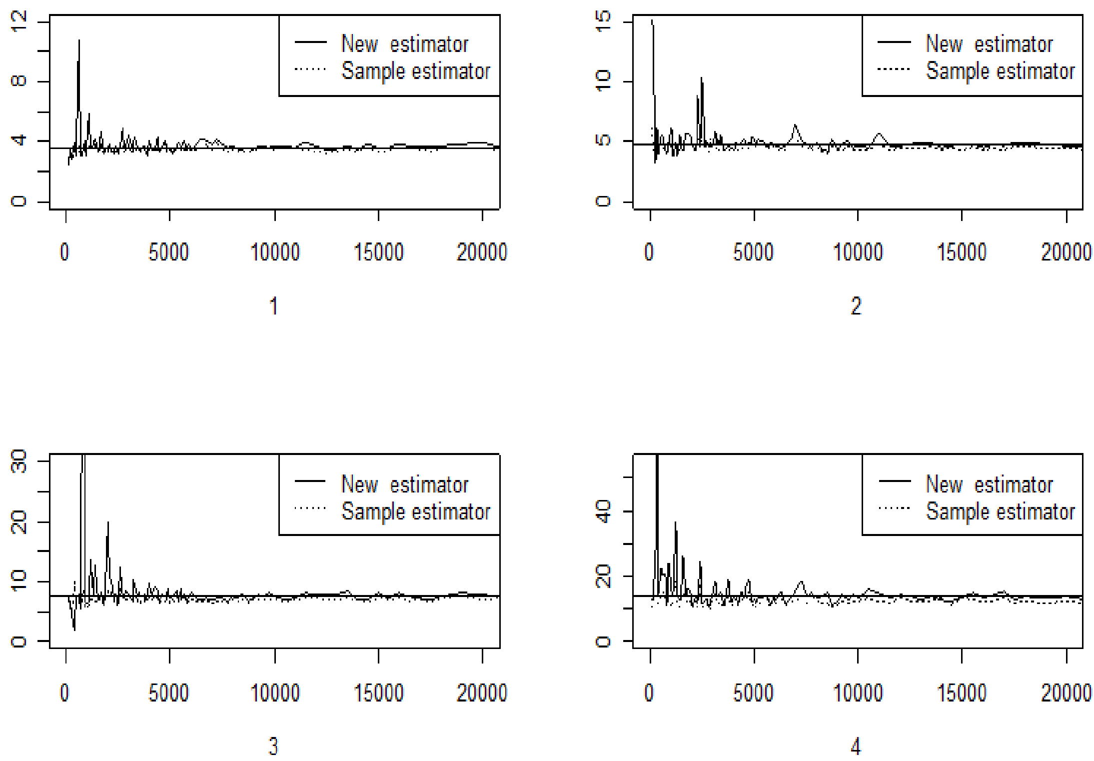

Figure 3.2: Values of the $\mathbb{CTE}$ estimator $\widetilde{\mathbb{CTE}}_n$ (vertical axis) versus sample sizes n (horizontal axis) evaluated at the levels $t = 0.25$, $t = 0.50$, $t = 0.75$, and $t = 0.90$ (panels 1–4, respectively) in the Pareto case with the tail index $\gamma = 2/3$.

The sum $W_{1,n} + W_{2,n} + W_{3,n}$ is a centered Gaussian random variable. To calculate its asymptotic variance, we establish the following limits:

$$\mathbf{E}[W_{1,n}^2] \to \frac{2\gamma}{2\gamma - 1}, \quad \mathbf{E}[W_{2,n}^2] \to \frac{\gamma^4}{(1-\gamma)^4}, \quad \mathbf{E}[W_{3,n}^2] \to \frac{2\gamma^2}{(1-\gamma)^4},$$
$$2\mathbf{E}[W_{1,n}W_{2,n}] \to \frac{-2\gamma^2}{(1-\gamma)^2}, \quad 2\mathbf{E}[W_{1,n}W_{3,n}] \to \frac{2\gamma}{(1-\gamma)^2},$$
$$2\mathbf{E}[W_{2,n}W_{3,n}] \to \frac{-2\gamma^3}{(1-\gamma)^4}.$$

Summing up the right-hand sides of the above six limits, we obtain σ_γ^2, whose expression in terms of the parameter γ is given in Theorem 3.2. Finally, since $X_{n-k:n}/\mathbb{Q}(1-k/n)$ converges in probability to 1 (see, e.g., the proof of Corollary in Necir and Meraghni (2009)), the classical Sultsky's lemma completes the proof of Theorem 3.2. Of course, we are still left to verify statements (3.22) and (3.23), which make the contents of the following two subsections.

Proof of statement (3.22).

If $\mathbb{Q}$ were continuously differentiable, then statement (3.22) would follow easily from the proof of Theorem 2 in Necir and Meraghni (2009). We do not assume differentiability of $\mathbb{Q}$ and thus a new proof is required, which is crucially based on the Vervaat process (see Zitikis (1998), Davydov and Zitikis (2003), Davydov and Zitikis (2004), and references therein)

$$V_n(t) = \int_0^t (\mathbb{Q}_n(s) - \mathbb{Q}(s))\, ds + \int_{-\infty}^{\mathbb{Q}(t)} (F_n(x) - F(x))\, dx.$$

Hence, for every t such that $0 < t < 1 - k/n$, which is satisfied for all sufficiently large n since t is fixed, we have that

$$\begin{aligned} A_{n,1}(t) &= \int_0^{1-k/n} (\mathbb{Q}_n(s) - \mathbb{Q}(s))\, ds - \int_0^t (\mathbb{Q}_n(s) - \mathbb{Q}(s))\, ds \\ &= -\int_{Q(t)}^{Q(1-k/n)} (F_n(x) - F(x))\, dx + V_n(1 - k/n) - V_n(t). \end{aligned} \tag{3.24}$$

It is well known (see Zitikis (1998), Davydov and Zitikis (2003), Davydov and Zitikis (2004)) that $V_n(t)$ is non-negative and does not exceed $-(F_n(\mathbb{Q}(t)) - t)(\mathbb{Q}_n(t) - \mathbb{Q}(t))$. Since the cdf F is continuous by assumption, we therefore have that

$$\sqrt{n}\, V_n(t) \leq |e_n(t)||\mathbb{Q}_n(t) - \mathbb{Q}(t)|, \tag{3.25}$$

where $e_n(t)$ is the the uniform empirical process $\sqrt{n}\,(F_n(\mathbb{Q}(t)) - F(\mathbb{Q}(t))$, which for large n looks like the Brownian bridge $\mathbb{B}_n(t)$. Note also that with the just introduced notation e_n, the integral on the right-hand side of equation (3.24) is equal to $\int_{\mathbb{Q}(t)}^{\mathbb{Q}(1-k/n)} e_n(F(x))dx$. Hence,

$$\begin{aligned} \frac{\sqrt{n}A_{n,1}(t)}{\sqrt{k/n}\mathbb{Q}(1-k/n)} &= -\frac{\int_{\mathbb{Q}(t)}^{\mathbb{Q}(1-k/n)} e_n(F(x))dx}{\sqrt{k/n}\mathbb{Q}(1-k/n)} \\ &\quad + O_{\mathbf{P}}(1)\frac{|e_n(1-k/n)||\mathbb{Q}_n(1-k/n) - \mathbb{Q}(1-k/n)|}{\sqrt{k/n}\mathbb{Q}(1-k/n)} \\ &\quad + O_{\mathbf{P}}(1)\frac{|e_n(t)||\mathbb{Q}_n(t) - \mathbb{Q}(t)|}{\sqrt{k/n}\mathbb{Q}(1-k/n)}. \end{aligned} \tag{3.26}$$

We shall next replace the empirical process e_n by an appropriate Brownian bridge $\mathbb{B}_n$ in the first integral on the right-hand side of equation (3.26) with an error term of magnitude $o_{\mathbf{P}}(1)$, and we shall also show that the second and third summands on the right-hand side of equation (3.26) are of the order $o_{\mathbf{P}}(1)$. The replacement of e_n by $\mathbb{B}_n$ can be accomplished using, for example, Corollary 2.1 on p. 48 of Csörgő *et al.* (1986), which states that on an appropriately constructed probability space and for any $0 \le \nu < 1/4$, we have that

$$\sup_{1/n \le s \le 1-1/n} \frac{|e_n(s) - \mathbb{B}_n(s)|}{s^{1/2-\nu}(1-s)^{1/2-\nu}} = O_{\mathbf{P}}(n^{-\nu}). \tag{3.27}$$

This result is applicable in the current situation since we can always place our original problem into the required probability space, because our main results are 'in probability'. Furthermore, since $\mathbb{Q}(t) \le x \le \mathbb{Q}(1-k/n)$, we have that $t \le F(x) \le 1 - k/n$. Hence, statement (3.27) implies that

$$-\frac{\int_{\mathbb{Q}(t)}^{\mathbb{Q}(1-k/n)} e_n(F(x))dx}{\sqrt{k/n}\mathbb{Q}(1-k/n)} = -\frac{\int_{\mathbb{Q}(t)}^{\mathbb{Q}(1-k/n)} \mathbb{B}_n(F(x))dx}{\sqrt{k/n}\mathbb{Q}(1-k/n)} + O_{\mathbf{P}}(1)\frac{\int_{\mathbb{Q}(t)}^{\mathbb{Q}(1-k/n)} (1-F(x))^{1/2-\nu}dx}{n^{\nu}\sqrt{k/n}\mathbb{Q}(1-k/n)}. \tag{3.28}$$

Changing the variables of integration and using the property $\sqrt{k/n}\mathbb{Q}(1-k/n) \to \infty$ when $n \to \infty$, we obtain that

$$-\frac{\int_{\mathbb{Q}(t)}^{\mathbb{Q}(1-k/n)} \mathbb{B}_n(F(x))dx}{\sqrt{k/n}\mathbb{Q}(1-k/n)} = -\frac{\int_0^{1-k/n} \mathbb{B}_n(s)d\mathbb{Q}(s)}{\sqrt{k/n}\mathbb{Q}(1-k/n)} + o_{\mathbf{P}}(1). \tag{3.29}$$

The main term on the right-hand side of equation (3.29) is $W_{1,n}$. We shall next show that the right-most summand of equation (3.28) converges to 0 when $n \to \infty$.

Changing the variable of integration and then integrating by parts, we obtain the bound

$$\frac{\int_{\mathbb{Q}(t)}^{\mathbb{Q}(1-k/n)} (1-F(x))^{1/2-\nu}dx}{n^{\nu}\sqrt{k/n}\mathbb{Q}(1-k/n)} \le \frac{(1-s)^{1/2-\nu}\mathbb{Q}(s)|_t^{1-k/n}}{n^{\nu}\sqrt{k/n}\mathbb{Q}(1-k/n)} + O(1)\frac{\int_t^{1-k/n} (1-s)^{-1/2-\nu}\mathbb{Q}(s)ds}{n^{\nu}\sqrt{k/n}\mathbb{Q}(1-k/n)}. \tag{3.30}$$

We want to show that the right-hand side of bound (3.30) converges to 0 when $n \to \infty$. For this, we first note that

$$\frac{(1-s)^{1/2-\nu}\mathbb{Q}(s)|_t^{1-k/n}}{n^{\nu}\sqrt{k/n}\mathbb{Q}(1-k/n)} = \frac{1}{k^{\nu}} - \frac{(1-t)^{1/2-\nu}\mathbb{Q}(t)}{n^{\nu}\sqrt{k/n}\mathbb{Q}(1-k/n)} \to 0. \tag{3.31}$$

Next, with the notation $\phi(u) = \mathbb{Q}(1-u)/u^{1/2+\nu}$ we have that

$$\frac{\int_t^{1-k/n}(1-s)^{-1/2-\nu}\mathbb{Q}(s)ds}{n^\nu\sqrt{k/n}\mathbb{Q}(1-k/n)} = \frac{1}{k^\nu}\frac{\int_{k/n}^{1-t}\phi(s)ds}{(k/n)\phi(k/n)} \to 0 \tag{3.32}$$

when $n \to \infty$, where the convergence to 0 follows from Result 1 in the Appendix of Necir and Meraghni (2009). Taking statements (3.30)–(3.32) together, we have that the right-most summand of equation (3.28) converges to 0 when $n \to \infty$. Consequently, in order to complete the proof of statement (3.22), we are left to show that the second and third summands on the right-hand side of equation (3.26) are of the order $o_{\mathbf{P}}(1)$. The third summand is of the order $o_{\mathbf{P}}(1)$ because $|e_n(t)||\mathbb{Q}_n(t) - \mathbb{Q}(t)| = O_{\mathbf{P}}(1)$ and $\sqrt{k/n}\mathbb{Q}(1-k/n) \to \infty$. Hence, we are only left to show that the second summand on the right-hand side of equation (3.26) is of the order $o_{\mathbf{P}}(1)$, for which we shall show that

$$\frac{|e_n(1-k/n)|}{\sqrt{k/n}}\left|\frac{\mathbb{Q}_n(1-k/n)}{\mathbb{Q}(1-k/n)} - 1\right| = o_{\mathbf{P}}(1). \tag{3.33}$$

To prove statement (3.33), we first note that

$$\frac{|e_n(1-k/n)|}{\sqrt{k/n}} \leq \frac{|e_n(1-k/n) - \mathbb{B}_n(1-k/n)|}{\sqrt{k/n}} + \frac{|\mathbb{B}_n(1-k/n)|}{\sqrt{k/n}}. \tag{3.34}$$

The first summand on the right-hand side of bound (3.34) is of the order $O_{\mathbf{P}}(1)$ due to statement (3.27) with $\nu = 0$. The second summand on the right-hand side of bound (3.34) is of the order $O_{\mathbf{P}}(1)$ due to a statement on p. 49 of Csörgő *et al.* (1986) (see the displayed bound just below statement (2.39) therein). Hence, to complete the proof of statement (3.33), we need to check that

$$\frac{\mathbb{Q}_n(1-k/n)}{\mathbb{Q}(1-k/n)} = 1 + o_{\mathbf{P}}(1). \tag{3.35}$$

Observe that, for each n, the distribution of $\mathbb{Q}_n(1-k/n)$ is same as that of $\mathbb{Q}\left(E_n^{-1}(1-k/n)\right)$, where E_n^{-1} is the uniform empirical quantile function. Furthermore, the processes $\{1 - E_n^{-1}(1-s),\ 0 \leq s \leq 1\}$ and $\{E_n^{-1}(s),\ 0 \leq s \leq 1\}$ are equal in distribution. Hence, statement (3.35) is equivalent to

$$\frac{\mathbb{Q}\left(1 - E_n^{-1}(k/n)\right)}{\mathbb{Q}(1-k/n)} = 1 + o_{\mathbf{P}}(1). \tag{3.36}$$

From the Glivenko-Cantelli theorem we have that $E_n^{-1}(k/n) - k/n \to 0$ almost surely, which also implies that $E_n^{-1}(k/n) \to 0$ since $k/n \to 0$ by our choice of k.

Moreover, we know from Theorem 0 and Remark 1 of Wellner (1978) that

$$\sup_{1/n \le s \le 1} s^{-1} \left| E_n^{-1}(s) - s \right| = o_{\mathbf{P}}(1), \tag{3.37}$$

from which we conclude that

$$nE_n^{-1}(k/n)/k = 1 + o_{\mathbf{P}}(1). \tag{3.38}$$

Since the function $s \mapsto \mathbb{Q}\,(1-s)$ is slowly varying at zero, using Potter's inequality (see the 5$^{\text{th}}$ assertion of Proposition B.1.9 on p. 367 of de Haan and Ferreira (2006)) [70] we obtain that

$$\frac{\mathbb{Q}\,(1 - E_n^{-1}(k/n))}{\mathbb{Q}(1-k/n)} = (1 + o_{\mathbf{P}}(1)) \left(nE_n^{-1}(k/n)/k \right)^{-\gamma \pm \theta} \tag{3.39}$$

for any $\theta \in (0, \gamma)$. In view of (3.38), the right-hand side of equation (3.39) is equal to $1 + o_{\mathbf{P}}(1)$, which implies statement (3.36) and thus finishes the proof of statement (3.22).

3.11.5 Proof of statement (3.23).

The proof of statement (3.23) is similar to that of Theorem 2 in Necir *et al.* (2007), though some adjustments are needed since we are now concerned with the CTE risk measure. We therefore present main blocks of the proof together with pinpointed references to Necir *et al.* (2007) for specific technical details.
We start the proof with the function $\mathbb{U}(z) = \mathbb{Q}(1 - 1/z)$ that was already used in the formulation of Theorem 3.2. Hence, if Y is a random variable with the distribution function $G(z) = 1 - 1/z$, $z \ge 1$, then $\mathbb{U}(Y) = \mathbb{Q}(G(Y)) \stackrel{d}{=} X$ since $G(Y)$ is a uniform on the interval $[0,1]$ random variable. Hence,

$$A_{n,2} = \frac{k/n}{1 - \widehat{\gamma}_n} \mathbb{U}(Y_{n-k:n}) - \int_0^{k/n} \mathbb{U}(1/s)ds,$$

and so we have

$$\begin{aligned} \frac{\sqrt{n} A_{n,2}}{(k/n)^{1/2} \mathbb{Q}(1-k/n)} &= \sqrt{k} \left(\frac{1}{1 - \widehat{\gamma}_n} \frac{\mathbb{U}(Y_{n-k:n})}{\mathbb{U}(n/k)} - \frac{(n/k) \int_0^{k/n} \mathbb{U}(1/s)ds}{\mathbb{U}(n/k)} \right) \\ &= \sqrt{k} \left(\frac{1}{1 - \widehat{\gamma}_n} \frac{\mathbb{U}(Y_{n-k:n})}{\mathbb{U}(n/k)} - \frac{1}{1-\gamma} \right) \\ &\quad + \sqrt{k} \left(\frac{1}{1-\gamma} + \frac{\int_1^\infty s^{-2} \mathbb{U}(ns/k)ds}{\mathbb{U}(n/k)} \right). \end{aligned} \tag{3.40}$$

We next show that the right-most term in equation (3.40) converges to 0 when $n \to \infty$. For this reason, we first rewrite the term as follows:

$$\sqrt{k}\left(\frac{1}{1-\gamma}+\frac{\int_1^\infty s^{-2}\mathbb{U}(ns/k)ds}{\mathbb{U}(n/k)}\right)=\sqrt{k}\int_1^\infty \frac{1}{s^2}\left(\frac{\mathbb{U}(ns/k)}{\mathbb{U}(n/k)}-s^\gamma\right)ds. \tag{3.41}$$

The right-hand side of equation (3.41) converges to 0 (see notes on p. 149 of Necir *et al.* (2007)) due to the second-order condition (1.39), which can equivalently be rewritten as

$$\lim_{z\to\infty}\frac{1}{A(z)}\left(\frac{\mathbb{U}(zs)}{\mathbb{U}(z)}-s^\gamma\right)=s^\gamma\frac{s^\rho-1}{\rho}$$

for every $s > 0$, where $A(z) = \gamma^2 a(\mathbb{U}(z))$. Note that $\sqrt{k}\,A(n/k) \to 0$ when $n \to \infty$. Hence, in order to complete the proof of statement (3.23), we need to check that

$$\begin{aligned}\sqrt{k}\left(\frac{1}{1-\widehat{\gamma}_n}\frac{\mathbb{U}(Y_{n-k:n})}{\mathbb{U}(n/k)}-\frac{1}{1-\gamma}\right) &= \frac{\gamma^2}{(1-\gamma)^2}\sqrt{\frac{n}{k}}\,\mathbb{B}_n(1-k/n)\\ &\quad -\frac{\gamma}{(1-\gamma)^2}\sqrt{\frac{n}{k}}\int_{1-k/n}^1\frac{\mathbb{B}_n(s)}{1-s}ds+o_{\mathbf{P}}(1).\end{aligned} \tag{3.42}$$

With Hill's estimator(2.11) is written in the form

$$\widehat{\gamma}_n=\frac{1}{k}\sum_{i=1}^k\log\left(\frac{\mathbb{U}(Y_{n-i+1:n})}{\mathbb{U}(Y_{n-k:n})}\right),$$

we proceed with the proof of statement (3.42) as follows:

$$\begin{aligned}\sqrt{k}\left(\frac{1-\gamma}{1-\widehat{\gamma}_n}\frac{\mathbb{U}(Y_{n-k:n})}{\mathbb{U}(n/k)}-1\right) &= \sqrt{k}\,\frac{1-\gamma}{1-\widehat{\gamma}_n}\left(\frac{\mathbb{U}(Y_{n-k:n})}{\mathbb{U}(n/k)}-\left(\frac{Y_{n-k:n}}{n/k}\right)^\gamma\right)\\ &\quad+\sqrt{k}\,\frac{1-\gamma}{1-\widehat{\gamma}_n}\left(\left(\frac{Y_{n-k:n}}{n/k}\right)^\gamma-1\right)+\sqrt{k}\,\frac{\widehat{\gamma}_n-\gamma}{1-\widehat{\gamma}_n}.\end{aligned} \tag{3.43}$$

Furthermore, we have that

$$\begin{aligned}\sqrt{k}\,\frac{\widehat{\gamma}_n-\gamma}{1-\widehat{\gamma}_n} &= \frac{1}{1-\widehat{\gamma}_n}\frac{1}{\sqrt{k}}\sum_{i=1}^k\left(\log\left(\frac{\mathbb{U}(Y_{n-i+1:n})}{\mathbb{U}(Y_{n-k:n})}\right)-\gamma\log\left(\frac{Y_{n-i+1:n}}{Y_{n-k:n}}\right)\right)\\ &\quad+\frac{\gamma}{1-\widehat{\gamma}_n}\frac{1}{\sqrt{k}}\sum_{i=1}^k\left(\log\left(\frac{Y_{n-i+1:n}}{Y_{n-k:n}}\right)-1\right).\end{aligned} \tag{3.44}$$

Arguments on p. 156 of Necir *et al.* (2007) imply that the first term on the right-hand side of equation (3.44) is of the order $O_{\mathbf{P}}(\sqrt{k}\, A(Y_{n-k:n}))$, and a note on p. 157 of Necir *et al.* (2007) says that $\sqrt{k}\, A(Y_{n-k:n}) = o_{\mathbf{P}}(1)$. Hence, the first term on the right-hand side of equation (3.44) is of the order $o_{\mathbf{P}}(1)$. Analogous considerations but now using bound (2.5) instead of (2.4) on p. 156 of Necir *et al.* (2007) imply that the first term on the right-hand side of equation (3.43) is of the order $o_{\mathbf{P}}(1)$. Hence, in summary, we have that

$$\sqrt{k}\left(\frac{1-\gamma}{1-\widehat{\gamma}_n}\frac{\mathbb{U}(Y_{n-k:n})}{\mathbb{U}(n/k)}-1\right) = \frac{1-\gamma}{1-\widehat{\gamma}_n}\sqrt{k}\left(\left(\frac{Y_{n-k:n}}{n/k}\right)^{\gamma}-1\right) + \frac{\gamma}{1-\widehat{\gamma}_n}\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\left(\log\left(\frac{Y_{n-i+1:n}}{Y_{n-k:n}}\right)-1\right)+o_{\mathbf{P}}(1). \tag{3.45}$$

We now need to connect the right-hand side of equation (3.45) with Brownian bridges $\mathbb{B}_n$. To this end, we first convert the Y-based order statistics into $\mathbb{U}$-based (i.e., uniform on $[0,1]$) order statistics. For this we recall that the cdf of Y is G, and thus Y is equal in distribution to $G^{-1}(\mathbb{U})$, which is $1/(1-\mathbb{U})$. Consequently,

$$\sqrt{k}\left(\frac{1-\gamma}{1-\widehat{\gamma}_n}\frac{\mathbb{U}(Y_{n-k:n})}{\mathbb{U}(n/k)}-1\right) = \frac{1-\gamma}{1-\widehat{\gamma}_n}\sqrt{k}\left(\left(\frac{1}{(n/k)(1-U_{n-k:n})}\right)^{\gamma}-1\right) + \frac{\gamma}{1-\widehat{\gamma}_n}\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\left(\log\left(\frac{(1-U_{n-k:n})}{(1-U_{n-i+1:n})}\right)-1\right)+o_{\mathbf{P}}(1). \tag{3.46}$$

Next we choose a sequence of Brownian bridges $\mathbb{B}_n$ (see pp. 158-159 in Necir *et al.* (2007) and references therein) such that the following two asymptotic representation hold:

$$\sqrt{k}\left(\left(\frac{1}{(n/k)(1-U_{n-k:n})}\right)^{\gamma}-1\right) = -\gamma\sqrt{\frac{n}{k}}\,\mathbb{B}_n(1-k/n)+o_{\mathbf{P}}(1)$$

and

$$\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\left(\log\left(\frac{(1-U_{n-k:n})}{(1-U_{n-i+1:n})}\right)-1\right) = \sqrt{\frac{n}{k}}\,\mathbb{B}_n(1-k/n) - \sqrt{\frac{n}{k}}\int_{1-k/n}^{1}\frac{\mathbb{B}_n(s)}{1-s}ds+o_{\mathbf{P}}(1).$$

Using these two statements on the right-hand side of equation (3.46) and also keeping in mind that $\widehat{\gamma}_n$ is a consistent estimator of γ (see Mason (1982)), we have that

$$\sqrt{k}\left(\frac{1-\gamma}{1-\widehat{\gamma}_n}\frac{\mathbb{U}(Y_{n-k:n})}{\mathbb{U}(n/k)}-1\right)=\frac{\gamma^2}{1-\gamma}\sqrt{\frac{n}{k}}\,\mathbb{B}_n(1-k/n)$$
$$-\frac{\gamma}{1-\gamma}\sqrt{\frac{n}{k}}\int_{1-k/n}^{1}\frac{\mathbb{B}_n(s)}{1-s}ds+o_{\mathbf{P}}(1). \qquad (3.47)$$

Dividing both sides of equation (3.47) by $1-\gamma$, we arrive at equation (3.42). This completes the proof of statement (3.23) and of Theorem 3.2 as well.

Chapter 4

Renewal theory and Estimation of The Renewal Function

4.1 Introduction

A renewal process is a generalization of the Poisson process. In essence, the Poisson process is a continuous-time Markov process on the positive integers (usually starting at zero) which has independent identically distributed holding times at each integer i (exponentially distributed) before advancing (with probability 1) to the next integer: $i + 1$. In the same informal spirit, we may define a renewal process to be the same thing, except that the holding times take on a more general distribution. (Note however that the iid property of the holding times is retained).

Renewal processes have a wide range of applications in the warranty control, in the reliability analysis of technical systems and, particularly, of telecommunication networks such as high-speed packet-switched networks like the Internet. Normally, measurement facilities count the events of interest, e.g., the number of requested and transferred Web pages, incoming or outgoing calls, frames, packets or cells in consecutive time intervals of fixed length. It is important for planning and control purposes to estimate the related traffic load in terms of the mean numbers of counted events and their variances in these intervals.

4.2 Main Definitions

Definition 4.1 *Let $X_1, X_2, ...$be iid non-negative rv's with cdf F. Let*

$$T_0 = 0, T_n = X_1 + X_2 + ... + X_n \text{ for all } n \geq 1,$$

and

$$N(t) = \max\{n : T_n \leq t\}. \tag{4.1}$$

Then we say $N(t)$ is a renewal process and that it has inter-arrival or lifetimes X_i.

The random variable $N(t)$ equals the number of renewals of these elements during time t.

Remark 4.1 *If the inter-arrival times $\{X_n\}$ are exponentially distributed with mean $1/\lambda$, then the renewal counting process $\{N(t), t > 0\}$ is a Poisson process with intensity λ.*

Definition 4.2 *The function*

$$H(t) = E(N(t)) = \sum_{n=1}^{\infty} P(T_n \leq t) = \sum_{n=1}^{\infty} F^{*n}(t), \tag{4.2}$$

is called the renewal function, for $t \geq 0$ where F^{*n} denotes the n-fold recursive Stieltjes convolution of F.

The function H satisfies the **renewal equation**

$$H(t) = F(t) + \int_0^t H(t-s)dF(s).$$

Note that, The renewal process $N(t)$ and the renewal function $H(t)$ are very important in the study of various theoretical and applied problems in queueing theory, reliability theory, storage theory, the theory of branching processes.

Definition 4.3 *The df F is said to be arithmetic if its support is on $\{0, \pm d, \pm 2d, ...\}$ for some constant d and otherwise is not arithmetic. The largest such d is called the period (or span) of the rv.*

4.3 Limit Theorems for Renewal Process

Theorem 4.1 *Let* $\{N(t), t > 0\}$ *be a renewal process corresponding to iid inter-arrival times* $\{X_n, n = I, 2, ...\}$
1) ***A symptotic expression for the mean*** $H(t) = E[N(t)]$*: if* $\mu = E[X_i] < \infty$*, then*

$$\lim_{t\to\infty} \frac{H(t)}{t} = \frac{1}{\mu}.$$

2) ***Asymptotic expression for the variance*** $Var[N(t)]$*, if* $\mu = E[X_i]$ *and* $\sigma^2 = Var[X_i]$ *are finite, then*

$$\lim_{t\to\infty} \frac{Var[N(t)]}{t} = \frac{\sigma^2}{\mu^3}.$$

3) ***Asymptotic normality of the renewal process****: if* $E[T^2] < \infty$*, then*

$$\frac{N(t) - t/\mu}{\sqrt{t\sigma^2/\mu^3}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Smith's (1954) key renewal theorem may be useful.

Theorem 4.2 (Key renewal theorem) *Let the cdf* $Q(t)$ *be continuous and* $Q(t) \geq 0$ *be a monotone non-increasing and integrable function on* $(0, \infty)$*. Then*

$$\lim_{t\to\infty} \int_0^t Q(t-s)\, H(ds) = \frac{1}{\mu} \int_0^\infty Q(x)\, dx, \tag{4.3}$$

where μ *denote the mean of inter-arrival time,* $\mu = E(X_i)$

Theorem 4.3 (Blackwell's theorem) *If* F *is not arithmetic*

$$H(t) - H(t-h) \to \frac{h}{\mu}, \text{ as } t \to \infty \tag{4.4}$$

for every $h > 0$*. If* F *is arithmetic the same is true when* h *is multiple of the span* d*.*

Theorem 4.4 *Let* $H(t)$ *be the mean value function of a renewal process corresponding to iid inter-arrival times with non-lattice df* $F(x)$ *and finite mean* μ*. Let* $g(t)$ *be a function satisfying the renewal equation*

$$g(t) = Q(t) + \int_0^t g(t-s) dF(s).$$

Then $g(t)$ *is given by*

$$g(t) = Q(t) + \int_0^t g(t-s)dH(s).$$

If $Q(t)$ *be continuous and* $Q(t) \geq 0$ *be a monotone non-increasing and integrable function on* $(0,\infty)$. *Then*

$$\lim_{t\to\infty} g(t) = \frac{1}{\mu}\int_0^\infty Q(x)\,dx, \tag{4.5}$$

Several rf-estimation methods have been developed for a known interarrival-time distribution. Unfortunately, explicit forms of the rf are obtained only in rare cases, for example, if the interarrival times have a uniform distribution, or for the wide class of matrix-exponential distributions (exponential and Erlang distributions belong to this class). Therefore, several attempts have been made to evaluate the rf computationally.
If the variance $\sigma^2 = var(X_i)$ of F is finite, then, applying Smith's theorem, the RF $H(t)$ for large t, may be approximated as follow

Theorem 4.5 *If* F *is not arithmetic with finite mean* μ *and finite variance* σ^2, *then*

$$H(t) \sim \frac{t}{\mu} + \frac{\sigma^2+\mu^2}{2\mu^2} \quad \text{as } t \to \infty. \tag{4.6}$$

4.4 Renewal Function when the first or the second moment is infinite

The classical renewal theorems do not tell much about the renewal function if the mean lifetime is infinite. To obtain more accurate results, when the lifetime distribution is regulary varying, $1-F(x) = x^{-\alpha}\mathbb{L}(x)$ and some $0<\alpha<1$, where $\mathbb{L}$ is slowly varying function, we have an analogue of the key renewal theorem by Feller (1971) and W.L. Smith proved a result of the two boundary case $\alpha = 0$ and $\alpha = 1$

Theorem 4.6 *If* F *is regulary varying distribution, then as* $t \to \infty$

$$H(t) \sim \begin{cases} 1/\mathbb{L}(x) & \text{if } \alpha = 0 \\ \frac{t^\alpha}{\mathbb{L}(t)}\frac{\sin\alpha\pi}{\alpha\pi} & 0<\alpha<1 \\ t/\int_0^t \overline{F}(x)dx & \text{if } \alpha = 1 \end{cases}.$$

Next, when the mean exist but the second moment is infinite Sgibnev[115] in 1981 shows the result.

Theorem 4.7 *If F is not arithmetic with finite mean μ and infinite variance σ^2, then when $t \to \infty$*

$$H(t) \sim \frac{t}{\mu} + \frac{1}{\mu^2} \int_0^t \left(\int_y^\infty (1 - F(x))\, dx \right) dy, \tag{4.7}$$

Note that the quantity $H(t) - \frac{t}{\mu}$ converges to a constant if and only if $E(X^2) < \infty$.

Theorem 4.8 *If F is regulary varying distribution, $1 - F(x) = x^{-\alpha}\mathbb{L}(x)$ and some $1 < \alpha < 2$, where $\mathbb{L}$ is slowly varying function, we have*

$$H(t) \sim \frac{t}{\mu} + \frac{t^2 (1 - F(t))}{\mu^2 (\alpha - 1)(\alpha - 2)}, t \to \infty. \tag{4.8}$$

This result (4.8) is derived by Teugels in 1968.

Remark 4.2 *This result has been extended to the case $1 < \alpha \le 2$, as follows*

$$H(t) - \frac{t}{\mu} \sim \frac{1}{\mu^2} \int_0^t \left(\mu - \int_0^y (1 - F(x))\, dx \right) dy, \ t \to \infty,$$

in Mohan 1976.

4.5 Non Parametric Estimation of the Renewal Function

There is a considerable body of literature on this topic, of which we shall next present an overview. Frees (52) in 1985, introduced and considered three non-parametric estimators of (4.2), The first is defined by replacing $F^{*k}(z)$ by its unbiased estimator:

$$F_n^k(z) := \begin{pmatrix} n \\ k \end{pmatrix}^{-1} \sum 1_{\{X_{i1}+X_{i2}+...+X_{ik} \le z\}}.$$

where the sum is taken over all k distinct indices $i_1, ..., i_k$ from the set $\{1, ..., n\}$. The sum in (4.2) is then truncated at some $m = m(n)$ dependent on n. This gives the estimator

$$\widehat{H}_n^1(z) := \sum_{k=0}^{m} F_n^k(z). \tag{4.9}$$

Frees[52] proves that $\widehat{H}_n(z)$ is a strongly consistent estimator of $H(z)$ and proves that this estimator is asymptotically normal distribution.
The second estimator suggested by Frees [52] is obtained by replacing F by the empirical distribution function F_n based on $X_1, ..., X_n$. The sum in (4.2) is then truncated at some $m = m(n)$ dependent on n. This gives the estimator

$$\widehat{H}_n^2(z) := \sum_{k=0}^{m} F_n^{*k}(z). \tag{4.10}$$

Frees[52] notes that the asymptotic results concerning estimator (4.9) hold for the estimator in (4.10) as well. However, just as with (4.9), choosing the parameter $m = m(n)$ in (4.10) remains a problem.
Frees[52] also mentions the possibility of using yet a third estimator, defined by replacing F by the empirical distribution function F_n, the sum in (4.2) is not truncated this time. This results in the estimator

$$\widehat{H}_n^3(z) := \sum_{k=0}^{\infty} F_n^{*k}(z). \tag{4.11}$$

Zhao and Subba Rao in 1997, estimated the renewal function by solving the renewal equation, incorporating a kernel estimate of the renewal density f .
Markovitch and Krieger[91] in 2006 used a histogram-type estimate, this estimator is without any information about the form of the underlying distribution and they using only an empirical sample.
From the literature overview above we can appreciate the complexities that arise in estimating the renewal function RF when $E(X^2) < \infty$, let alone when $E(X^2) = \infty$. Nevertheless, one can attempt to investigate the performance of this estimators when $E(X^2) = \infty$. This is certainly a natural and interesting avenue of research. Sgibnev [115] in 1981 shows that, when $t \to \infty$, the quantity $\mathbb{H}(t) - \frac{t}{\mu}$ converges to a constant iff $E(X^2) < \infty$. When $E(X^2) = \infty$, Sgibnev (1981) proves that

$$\mathbb{H}(t) - \frac{t}{\mu} \sim \frac{1}{\mu^2} \int_0^t \left(\int_y^\infty (1 - F(x))\, dx \right) dy,$$

Mark Bebbington et al, 2007[10], construct an empirical estimator for it as follows.

$$\widehat{\mathbb{H}}_n(t) \sim \frac{t}{\overline{X}} + \frac{1}{\overline{X}^2} \int_0^t \left(\int_y^\infty (1 - F_n(x))\, dx \right) dy$$

where F_n and $\overline{X}$ is the empirical distribution and sample mean respectively based on the sample $X_1, X_2, ..., X_n$. we write this estimatord as follows

$$\widehat{\mathbb{H}}_n(t) \sim \frac{t}{\overline{X}} + \frac{1}{2n\overline{X}^2} \sum_{i=1}^{n} (X_i \wedge t)^2 + \frac{t}{n\overline{X}^2} \sum_{i=1}^{n} (X_i - X_i \wedge t) . \tag{4.12}$$

Their main result says that whenever F belongs to the domain of attraction of a stable law S_α with $1/2 < \alpha < 1$ (see, e.g., Solotarev, 1986), the df of $\widehat{\mathbb{H}}_n(t)$ converges, for suitable normalizing constants, to S_α. This result provides confidence bounds for $\mathbb{H}(t)$ with respect to the quantiles of S_α.

4.6 Semi Parametric Estimation of the Renewal Function

In practice, it is a more realistic situation that the distribution is unknown or that just general information describing it is available. The restoration of the df, if the latter exists, may become complicated if the distributions of the rv's are heavy-tailed. Weibull distributions with a shape parameter less than one and Pareto distributions provide examples of such pdfs. Heavy-tailed distributions often arise in practice, for example, in insurance and queueing or in the characterization of World Wide Web (WWW) traffic (Ref.[92]).

Indeed, an important class of models having infinite second order moments is the set of heavy-tailed distributions (e.g., Pareto, Burr, Student, ...). A df F is said to be heavy-tailed with tail idex $\gamma > 0$ if

$$\overline{F}(x) = cx^{-1/\gamma} \left(1 + x^{-\delta}\mathbb{L}(x)\right), \text{ as } x \to \infty, \tag{4.13}$$

for $\gamma \in (0, 1)$, $\delta > 0$ and some real constant c, with $\mathbb{L}$ a slowly varying function at infinity.
Notice that when $\gamma \in (1/2, 1)$ we have $\mu < \infty$ and $E[X^2] = \infty$. In this case, an asymptotic approximation of the renewal function $\mathbb{H}(t)$ is given by (4.7).

Prior to Sgibnev (1981) [115], Teugels (1968) [120] obtained an approximation of $\mathbb{H}(t)$ when F is heavy-tailed with tail index $\gamma \in (1/2, 1)$:

$$\mathbb{H}(t) - \frac{t}{\mu} \sim \frac{\gamma^2 t^2 \overline{F}(t)}{\mu^2 (1-\gamma)(2\gamma - 1)}, \text{ as } t \to \infty. \tag{4.14}$$

Extreme value theory allows for an accurate modeling of the tails of any unknown distribution, making the (semi-parametric) statistical inference more performant for heavy-tailed distributions. Indeed, the semi-parametric approach permits to extrapolate beyond the largest value of a given sample while the non-parametric one does not since the empirical df vanishes outside the sample. This represents a big handicap for those dealing with heavy-tailed data.

Extreme value theory is applied with the POT method, it is based on Balkema-de Haan result which says that the distribution of the excesses over a fixed threshold is approximated by the generalized Pareto distribution (GPD) (see section (1.5.4). In our situation, we have a fixed threshold equal to the horizon $t = t_n$. Therefore, the POT method would be the appropriate choice to derive an estimator for $\mathbb{H}(t)$ by exploiting the heavy-tail property of df F used in approximation (4.7). The asymptotic normality of our estimator is established under suitable assumptions.

4.6.1 Estimating the renewal function in infinite time

The distribution of the excesses, over a "fixed" threshold t, pertaining to df F is defined by F_t as (1.31), with theorem (1.18) of Balkema and de Haan (1974) and Pickands (1975), that is F_t is approximated by a generalized Pareto distribution (GPD) function $\mathbb{G}_{\gamma,\sigma}$ with shape parameter $\gamma \in \mathbb{R}$ and scale parameter $\sigma = \sigma(t) > 0$, where $\mathbb{G}_{\gamma,\sigma}$ is define by equation (1.30).
Suppose now that $Y_1, ..., Y_N$ are drawn not from $\mathbb{G}_{\gamma,\sigma}$, but from F_t. In view of the asymptotic approximation (1.33), Smith (1985) [118] has proposed estimates for (γ, σ) via the Maximum Likelihood approach. The obtained estimators $(\widehat{\gamma}_N, \widehat{\sigma}_N)$ are given in equations (2.25) and (2.26).

Since we are interested in the renewal function in infinite time, we must assume that time t is large enough and for asymptotic considerations, we will assume that t depends on the sample size n. That is $t = t_n$, with $t_n \to \infty$ as $n \to \infty$. Relation (4.13) suggests that in order to construct an estimator of $\mathrm{H}(t_n)$, we need to estimate μ, γ and $\overline{F}(t_n)$. Let $n = n(t)$ be the number of $X_i's$, which are observed on horizon t_n and denote by

$$N_{t_n} := card\left(\{X_i > t_n : 1 \leq i \leq n\}\right),$$

the number of exceedances over t_n, with $card\,(K)$ being the cardinality of set K.

Notice that N_{t_n} is a binomial rv with parameters n and $p_n := \overline{F}(t_n)$ for which the natural estimator is $\widehat{p}_n := N_t/n$.

Select, from the sample $(X_1, ..., X_n)$, only those observations $X_{i_1}, ..., X_{i_{N_{t_n}}}$ that exceed t_n. The N_t excesses

$$E_{j:n} := X_{i_j} - t_n,\ j = 1, ..., N_{t_n},$$

are iid rv's with df F_{t_n}. The maximum likelihood estimators $(\widehat{\gamma}_n, \widehat{\sigma}_n)$ are solutions of the following system

$$\begin{cases} \dfrac{1}{v_n}\displaystyle\sum_{j1}^{v_n} \log\left(1+\gamma\frac{e_{j:n}}{\sigma}\right) = \gamma, \\ \dfrac{1}{v}\displaystyle\sum_{j=1}^{v_n} \frac{e_{j:n}/\sigma}{1+e_{j:n}/\sigma} = \dfrac{1}{1+\gamma}, \end{cases}$$

where v_n is an observation of N_{t_n} and the vector $(e_{1:n}, ..., e_{v_n:n})$ a realization of $(E_{1:n}, ..., E_{N_{t_n}:n})$. Regarding the distribution mean $\mu = E[X_1]$, we know that, for $\gamma \in (0, 1/2]$, X_1 has finite variance and therefore μ could naturally be estimated by the sample mean $\overline{X} := n^{-1}S_n$ which, by the Central Limit Theorem (CLT), is asymptotically normal. Whereas for $\gamma \in (1/2, 1)$, X_1 has infinite variance, in which case the CLT is no longer valid. This case is frequently met in real insurance data (see for instance, Beirlant *et al.*, 2001). Using the GPD approximation, Johansson (2003) in [58] has proposed an alternative estimator for μ given by equation (2.38) as follow:

$$\widehat{\mu}_n^{(J)} = \frac{1}{n}\sum_{i=1}^{n} X_i \mathbf{1}_{\{X_i \leq t_n\}} + \widehat{p}_n\left(t_n + \frac{\widehat{\sigma}_n}{1-\widehat{\gamma}_n}\right). \tag{4.15}$$

Here $\mathbf{1}_K$ denotes the indicator function of set K. Respectively substituting $\widehat{\mu}_n^{(J)}$, $\widehat{\gamma}_n$ and $\widehat{p}_n$ for μ, γ and $\overline{F}(t_n)$ in (4.14) yields the following estimator for the renewal function $\mathrm{H}(t_n)$

$$\widetilde{\mathbb{H}}_n(t_n) := \frac{t_n}{\widehat{\mu}_n^{(J)}} + \frac{\widehat{\gamma}_n^2 t_n^2 \widehat{p}_n}{\widehat{\mu}_n^{(J)2}\left(1-\widehat{\gamma}_n\right)\left(2\widehat{\gamma}_n - 1\right)}. \tag{4.16}$$

The asymptotic behavior of $\widetilde{\mathbb{H}}_n(t_n)$ is given by the following two theorems.

Theorem 4.9 *Let F be a df fulfilling* (4.13) *with* $\gamma \in (1/2, 1)$. *Suppose that $\mathbb{L}$ is locally bounded in $[x_0, +\infty)$ for $x_0 \geq 0$ and $x \mapsto x^{-\delta}\mathbb{L}(x)$ is non-increasing* *near infinity, for some* $\delta > 0$. *Then, for any* $t_n = O(n^{\alpha\gamma/4})$ *with* $\alpha \in (0,1)$, *we have*

$$\widetilde{\mathbb{H}}_n(t_n) - \mathbb{H}(t_n) = O_{\mathbb{P}}(n^{(\alpha/2)(\gamma-1/4)-1/2}), \text{ as } n \to \infty.$$

Theorem 4.10 *Let* F *be as in Theorem 4.9. Then for any* $t_n = O(n^{\alpha\gamma/4})$ *with* $\alpha \in (4/(1+2\gamma\delta), 1)$, *we have*

$$\frac{\sqrt{n}}{s_n t_n}\left(\widetilde{\mathbb{H}}_n(t_n) - \mathbb{H}(t_n)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \text{ as } n \to \infty,$$

where

$$\begin{aligned} s_n^2 &:= \gamma_n^2\theta_{1n}^2 + p_n(1-p_n)\left(\theta_{2n} + \theta_{1n}t_n + \frac{\theta_{1n}\sigma_n}{1-\gamma}\right)^2 \\ &+ \frac{(1+\gamma)^2}{p_n}\left(\theta_{3n} + \frac{\theta_{1n}p_n\sigma_n}{(1-\gamma)^2}\right)^2 + \frac{2(1+\gamma)\theta_{1n}^2 p_n^2\sigma_n^2}{p_n(1-\gamma)^2} \\ &- \frac{(1+\gamma)\theta_{1n}p_n\sigma_n}{p_n(1-\gamma)}\left(\theta_{3n} + \frac{\theta_{1n}p_n\sigma_n}{(1-\gamma)^2}\right), \end{aligned}$$

with

$$\begin{aligned} \theta_{1n} &:= -\frac{1}{\mu^2} - \frac{2\gamma^2 t_n p_n}{\mu^3(1-\gamma)(2\gamma-1)}, \\ \theta_{2n} &:= \frac{\gamma^2 t_n}{\mu^2(1-\gamma)(2\gamma-1)}, \\ \theta_{3n} &:= \frac{t_n p_n}{\mu^2(1-\gamma)(2\gamma-1)}\left(2\gamma + \frac{4\gamma^3 - 3\gamma^2}{(1-\gamma)(2\gamma-1)}\right), \end{aligned}$$

$\sigma_n := t_n\gamma$ *and* $\gamma_n^2 := Var\left(X_1\mathbf{1}_{\{X_1 \le t_n\}}\right)$. *Here* $\mathcal{N}(0,1)$ *stands for the standard normal rv.*

4.6.2 Simulation study

In this section, we carry out a simulation study (by means of the statistical software **R**, see Ihaka and Gentleman, 1996) to illustrate the performance of our estimation procedure, through its application to sets of samples taken from two distinct Pareto distributions $F(x) = 1 - x^{-1/\gamma}$, $x > 1$ (with tail indices $\gamma = 3/4$ and $\gamma = 2/3$). We fix the threshold at 4, which is a value above the intermediate statistic corresponding to the optimal fraction of upper order staistics in each sample. The latter is obtained by applying the algorithm of Cheng and Peng (2001). For each sample size, we generate 200 independent replicates. Our

True value $H = 2.222$						
	semi-parametric $\widehat{H}$			non-parametric $\widetilde{H}$		
sample size	mean	bias	rmse	mean	bias	rmse
1000	2.265	0.042	0.185	2.416	0.193	0.229
2000	2.247	0.024	0.157	2.054	−0.167	0.223
5000	2.223	0.001	0.129	2.073	−0.149	0.192

Table 4.1: Semi-parametric and non-parametric estimates of the renewal function of inter-occurence times of Pareto-distributed claims with shape parameter 2/3. Simulations are repeated 200 times for different sample

True value $H = 1.708$						
	semi-parametric $\widehat{H}$			non-parametric $\widetilde{H}$		
sample size	mean	bias	rmse	mean	bias	rmse
1000	1.696	−0.013	0.250	2.141	0.433	0.553
2000	1.719	0.011	0.183	1.908	0.199	0.288
5000	1.705	−0.003	0.119	1.686	−0.022	0.168

Table 4.2: Semi-parametric and non-parametric estimates of the renewal function of inter-occurence times of Pareto-distributed claims with shape parameter 3/4. Simulations are repeated 200 times for different sample

overall results are then taken as the empirical means of the values in the 200 repetitions. A comparison with the non-parametric estimator is done as well.

In the graphical illustration, we plot both estimators versus the sample size ranging from 1000 to 20000. Figure 4.1 clearly shows that the new estimator is consistent and that it is always better than the non-parametric one. For the numerical investigation, we take samples of sizes 1000, 2000 and 5000. In each case, we compute the semi-parametric estimate $\widetilde{\mathbb{H}}_n$ as well as the non-parametric estimate $\widehat{\mathbb{H}}_n$. We also provide the bias and the root mean squared error (rmse). The results are summarized in Tables (4.1) and (4.2) for γ= 3/4 and γ= 2/3 respectively. We notice that, regardless of the tail index value and the sample size, the semi-parametric estimation procedure is more accurate than the non-parametric one.

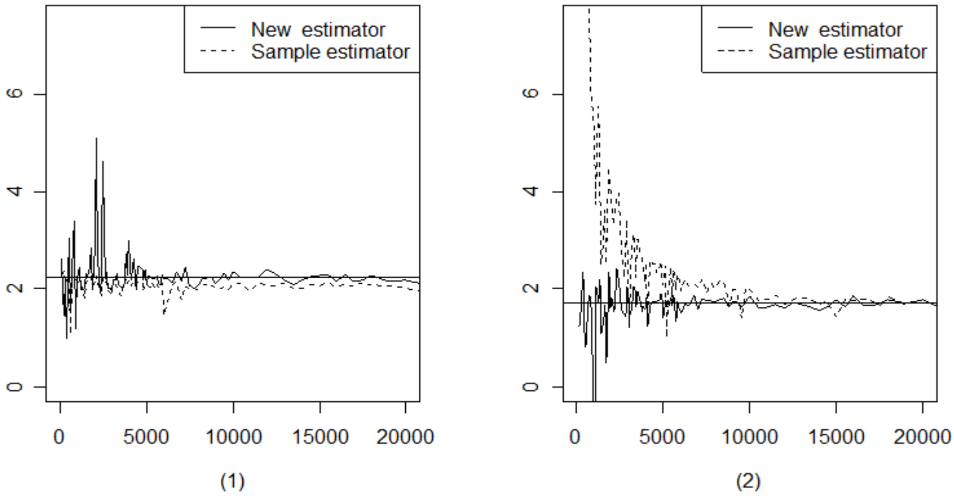

Figure 4.1: Plots of the new and sample estimators of the renewal function, of inter-occurence times of Pareto-distributed claims with tail indices 2/3 (panel 1) and 3/4 (panel 2), versus the sample size. The horizontal line represents the true value of the renewal function $\mathbb{R}(t)$ evaluated at $t = 4$.

4.7 Proofs

The following tools will be instrumental for our needs.

Proposition 4.1 *Let F be a cdf fulfilling* (4.13) *with* $\gamma \in (1/2, 1)$, $\delta > 0$ *and some real c. Suppose that $\mathbb{L}$ is locally bounded in $[x_0, +\infty)$ for $x_0 \geq 0$. Then for n large enough and for any $t_n = O\left(n^{\alpha\gamma/4}\right), \alpha \in (0,1)$, we have as $n \to \infty$*
$p_n = c(1 + o(1))n^{-\alpha/4}$, $\gamma_n^2 = O(n^{(\alpha/2)(\gamma-1/2)})$,
$s_n^2 = O(n^{(\alpha/2)(\gamma-1/2)})$ *and* $\sqrt{np_n} t_n^{-\delta} \mathbb{L}(t_n) = O\left(n^{-\alpha/8-\alpha\gamma\delta/4+1/2}\right)$.

Lemma 4.1 *Under the assumptions of Theorem (4.10), we have, for any real numbers u_1, u_2, u_3 and u_4,*

$$E\left[\exp\left\{iu_1 \frac{\sqrt{n}}{\gamma_n}\left(\widehat{\mu}_n^* - \mu_n^*\right) + i\sqrt{np_n}\left(u_2, u_3\right)\begin{pmatrix}\widehat{\sigma}_n/\sigma_n - 1\\ \widehat{\gamma}_n - \gamma\end{pmatrix} + iu_4 \frac{\sqrt{n}\left(\widehat{p}_n - p_n\right)}{\sqrt{p_n\left(1-p_n\right)}}\right\}\right]$$
$$\to \exp\left\{-\frac{u_1^2}{2} - \frac{1}{2}\left(u_2, u_3\right)\mathbb{Q}^{-1}\begin{pmatrix}u_2\\ u_3\end{pmatrix} - \frac{u_4^2}{2}\right\}, \text{ as } n \to \infty,$$

where $i^2 = -1$.

Proof of the Proposition. We will only prove the second result, the other ones are straightforward from (4.13). Let $x_0 > 0$ be such that $\overline{F}(x) = cx^{-1/\gamma}\left(1 + x^{-\delta}\mathbb{L}(x)\right)$, for $x > x_0$. Then for n large enough, we have

$$E\left[X_1 \mathbf{1}_{\{X_1 \leq t_n\}}\right] = \int_0^{t_n} xdF(x) = \int_0^{x_0} xdF(x) + \int_{x_0}^{t_n} xdF(x).$$

Recall that $\mu < \infty$, hence $\int_0^{x_0} xdF(x) < \infty$. Making use of the Proposition assumptions, we get for all large n, $E\left[X_1 \mathbf{1}_{\{X_1 \leq t_n\}}\right] = O(1)$ and $E\left[X_1^2 \mathbf{1}_{\{X_1 \leq t_n\}}\right] = O\left(t_n^{2-1/\gamma}\right)$ and therefore $\gamma_n^2 = O\left(n^{\alpha/2(\gamma/2-1)}\right)$. ■

Proof of the Lemma. See Johansson (2003)[58]. ■

Proof of Theorem 4.9. We readily check that for all large n,

$$\left(\widetilde{\mathbb{H}}_n(t_n) - \mathbb{H}(t_n)\right) t_n / \sim A_n + B_n + C_n,$$

where

$$A_n := \left(-\frac{1}{\widehat{\mu}_n^{(J)}\mu} - \frac{\gamma^2 t_n p_n \left(\widehat{\mu}_n^{(J)} + \mu\right)}{\widehat{\mu}_n^{(J)2}\mu^2 (1-\gamma)(2\gamma-1)} \right) \left(\widehat{\mu}_n^{(J)} - \mu\right),$$

$$B_n := \frac{\widehat{\gamma}_n^2 t_n}{\widehat{\mu}_n^{(J)2}(1-\widehat{\gamma}_n)(2\widehat{\gamma}_n - 1)} \left(\widehat{p}_n - p_n\right),$$

and

$$C_n := \frac{t_n p_n}{\widehat{\mu}_n^{(J)2}(1-\widehat{\gamma}_n)(2\widehat{\gamma}_n - 1)} \times \left(\widehat{\gamma}_n + \gamma + \frac{2\gamma^2(\widehat{\gamma}_n + \gamma) - 3\gamma^2}{(1-\gamma)(2\gamma-1)}\right)\left(\widehat{\gamma}_n - \gamma\right).$$

Johansson (2003)[58] proved that there exists a bounded sequence k_n such that

$$\widehat{\mu}_n^{(J)} - \mu = O_{\mathbb{P}}\left(\gamma_n \sqrt{k_n/n}\right), \text{ as } n \to \infty, \tag{4.17}$$

hence

$$\widehat{\mu}_n^{(J)} - \mu = O_{\mathbb{P}}\left(n^{(\alpha/4)(\gamma-1/2)-1/2}\right).$$

The first result of the Proposition yields that

$$t_n p_n\left(\widehat{\mu}_n^{(J)} - \mu\right) = O_{\mathbb{P}}\left(n^{(\alpha/4)(2\gamma-3/2)-1/2}\right).$$

Since $(\alpha/4)(2\gamma - 3/2) - 1/2 < 0$, then

$$t_n p_n\left(\widehat{\mu}_n^{(J)} - \mu\right) = o_{\mathbb{P}}(1).$$

On the other hand, by the CLT we have

$$\widehat{p}_n - p_n = O_{\mathbb{P}}\left(\sqrt{p_n/n}\right), \text{ as } n \to \infty, \tag{4.18}$$

then

$$t_n\left(\widehat{p}_n - p_n\right) = O_{\mathbb{P}}\left(n^{(\alpha/4)(\gamma-1/2)-1/2}\right) = o_{\mathbb{P}}(1).$$

On the other hand, from Smith (1987), yields

$$\widehat{\gamma}_n - \gamma = O_{\mathbb{P}}\left(t_n^{-\delta}\mathbb{L}(t_n)\right), \text{ as } n \to \infty, \tag{4.19}$$

it follows

$$\widehat{\gamma}_n^2 t_n\left(\widehat{p}_n - p_n\right) = O_{\mathbb{P}}\left(n^{(\alpha/4)(\gamma(1-2\delta)-1/2)-1/2}\right) = o_{\mathbb{P}}(1),$$

therefore

$$\frac{\widehat{\gamma}_n^2 t_n\left(\widehat{p}_n - p_n\right)}{\widehat{\mu}_n^{(J)2}\left(1-\widehat{\gamma}_n\right)\left(2\widehat{\gamma}_n - 1\right)} = o_{\mathbb{P}}(1), \text{ as } n \to \infty,$$

$$t_n p_n\left(\widehat{\gamma}_n - \gamma\right) = O_{\mathbb{P}}\left(n^{(\alpha/4)(\gamma(1-\delta)-1)}\right) = o_{\mathbb{P}}(1), \text{ as } n \to \infty,$$

$$\widehat{\gamma}_n t_n p_n\left(\widehat{\gamma}_n - \gamma\right) = O_{\mathbb{P}}\left(n^{(\alpha/4)(\gamma(1-2\delta)-1)}\right) = o_{\mathbb{P}}(1), \text{ as } n \to \infty,$$

and

$$p_n\left(\widehat{\gamma}_n - \gamma\right) = O_{\mathbb{P}}\left(n^{(-\alpha/4)((1+\gamma\delta))}\right) = o_{\mathbb{P}}(1), \text{ as } n \to \infty.$$

Thus

$$\frac{\left(\widehat{\gamma}_n + \gamma\right)}{\widehat{\mu}_n^{(J)2}\left(1-\widehat{\gamma}_n\right)\left(2\widehat{\gamma}_n - 1\right)} t_n p_n\left(\widehat{\gamma}_n - \gamma\right) \xrightarrow{\mathbb{P}} 0, \text{ as } n \to \infty,$$

and

$$\frac{t_n p_n\left(2\gamma^2\left(\widehat{\gamma}_n + \gamma\right) - 3\gamma^2\right)}{\widehat{\mu}_n^{(J)2}\left(1-\widehat{\gamma}_n\right)\left(2\widehat{\gamma}_n - 1\right)\left(1-\gamma\right)\left(2\gamma-1\right)}\left(\widehat{\gamma}_n - \gamma\right) \xrightarrow{\mathbb{P}} 0 \text{ as } n \to \infty.$$

Finally, we infer that $\widetilde{\mathbb{H}}_n(t_n) - \mathbb{H}(t_n) = O_{\mathbb{P}}(n^{(\alpha/2)(\gamma-1/4)-1/2})$ as $n \to \infty$. ∎

Proof of theorem 4.10. In the proof of Theorem 4.9, we have shown that $\widehat{\mu}_n^{(J)} = \mu + o_{\mathbb{P}}(1)$ as $n \to \infty$. In view of (4.19), it follows that, for all large n, $(\widetilde{\mathbb{H}}_n(t_n) - \mathbb{H}(t_n))/t_n$ may be rewritten into

$$\begin{aligned}\left(\widetilde{\mathbb{H}}_n(t_n) - \mathbb{H}(t_n)\right)/t_n &= \theta_1\left(1 + o_{\mathbb{P}}(1)\right)\left(\widehat{\mu}_n^{(J)} - \mu\right)\\ &+ \theta_2\left(1 + o_{\mathbb{P}}(1)\right)\left(\widehat{p}_n - p_n\right)\\ &+ \theta_3\left(1 + o_{\mathbb{P}}(1)\right)\left(\widehat{\gamma}_n - \gamma\right),\end{aligned}$$

where

$$\theta_1 = -\frac{1}{\mu^2} - \frac{2\gamma^2 t_n p_n}{\mu^3 (1-\gamma)(2\gamma-1)},$$
$$\theta_2 = \frac{\gamma^2 t_n}{\mu^2 (1-\gamma)(2\gamma-1)},$$

and

$$\theta_3 = \frac{t_n p_n}{\mu^2 (1-\gamma)(2\gamma-1)} \left(2\gamma + \frac{4\gamma^3 - 3\gamma^2}{(1-\gamma)(2\gamma-1)} \right).$$

Multiplying by $\sqrt{n}/\gamma_n$ and using the Proposition and the Lemma together with the continuous mapping theorem, we find that

$$\begin{aligned}\frac{\sqrt{n}}{\gamma_n t_n} \left(\widehat{\mathbb{H}}_n(t_n) - \mathbb{H}_n(t_n) \right) &= \theta_1 (1 + o_{\mathbb{P}}(1)) \frac{\sqrt{n}}{\gamma_n} \left(\widehat{\mu}_n^{(J)} - \mu \right) \\ &\quad + \theta_2 (1 + o_{\mathbb{P}}(1)) \frac{\sqrt{n}}{\gamma_n} (\widehat{p}_n - p_n) \\ &\quad + \theta_3 (1 + o_{\mathbb{P}}(1)) \frac{\sqrt{n}}{\gamma_n} (\widehat{\gamma}_n - \gamma).\end{aligned}$$

On the other hand, from Johansson (2003) [58], we have for all large n

$$\begin{aligned}\frac{\sqrt{n}}{\gamma_n} \left(\widehat{\mu}_n^{(J)} - \mu \right) &= \frac{\sqrt{n}}{\gamma_n} (\widehat{\mu}_n^* - \mu_n^*) + \left(t_n + \frac{\sigma_n}{1-\gamma_n} \right) \frac{\sqrt{n}}{\gamma_n} (\widehat{p}_n - p_n) \\ &\quad + \frac{p_n \sigma_n}{(1-\gamma)^2} \frac{\sqrt{n}}{\gamma_n} (\widehat{\gamma}_n - \gamma) + \frac{p_n}{1-\gamma} \frac{\sqrt{n}}{\gamma_n} (\widehat{\sigma}_n - \sigma_n) + o_{\mathbb{P}}(1).\end{aligned}$$

This enables us to rewrite $\frac{\sqrt{n}}{\gamma_n t_n} \left(\widehat{\mathbb{R}}_n(t_n) - \mathbb{R}_n(t_n) \right)$ into

$$\begin{aligned}&\theta_1 \frac{\sqrt{n}}{\gamma_n} (\widehat{\mu}_n^* - \mu_n^*) + \frac{\sqrt{p_n (1-p_n)}}{\gamma_n} \left(\theta_2 + \theta_1 \left(t_n + \frac{\sigma_n}{1-\gamma} \right) \right) \frac{\sqrt{n} (\widehat{p}_n - p_n)}{\sqrt{p_n (1-p_n)}} \\ &\quad + \theta_1 \frac{\sigma_n p_n}{\gamma_n (1-\gamma) \sqrt{p_n}} \sqrt{n p_n} (\widehat{\sigma}_n / \sigma_n - 1) \\ &\quad + \frac{1}{\gamma_n \sqrt{p_n}} \left(\theta_3 + \theta_1 \frac{p_n \sigma_n}{(1-\gamma)^2} \right) \sqrt{n p_n} (\widehat{\gamma}_n - \gamma) + o_{\mathbb{P}}(1).\end{aligned}$$

In view of Lemma 2, we infer that for all large n, the previous quantity

$$\begin{aligned}&\theta_1 \mathcal{W}_1 + \frac{\sqrt{p_n (1-p_n)}}{\gamma_n} \left(\theta_2 + \theta_1 \left(t_n + \frac{\sigma_n}{1-\gamma} \right) \right) \mathcal{W}_2 \\ &+ \frac{\sqrt{2(1+\gamma)} \theta_1 \sigma_n p_n}{\gamma_n (1-\gamma) \sqrt{p_n}} \mathcal{W}_3 + \frac{(1+\gamma)}{\gamma_n \sqrt{p_n}} \left(\theta_3 + \frac{\theta_1 p_n \sigma_n}{(1-\gamma)^2} \right) \mathcal{W}_4 + o_{\mathbb{P}}(1),\end{aligned}$$

where $(\mathcal{W}_i)_{i=1,4}$ are standard normal rv's with $E[W_iW_j] = 0$ for every $i, j = 1, ..., 4$ with $i \neq j$, except for

$$\begin{aligned}
E[\mathcal{W}_3\mathcal{W}_4] &= E\left[\frac{1}{\sqrt{2(1+\gamma)}}\sqrt{np_n}\left(\widehat{\sigma}_n/\sigma_n - 1\right)\frac{1}{(1+\gamma)}\sqrt{np_n}\left(\widehat{\gamma}_n - \gamma\right)\right] \\
&= \frac{1}{(1+\gamma)\sqrt{2(1+\gamma)}}E\left[\sqrt{np_n}\left(\widehat{\sigma}_n/\sigma_n - 1\right)\sqrt{np_n}\left(\widehat{\gamma}_n - \gamma\right)\right] \\
&= -\frac{1}{\sqrt{2(1+\gamma)}}.
\end{aligned}$$

Therefore the rv $\dfrac{\sqrt{n}}{\gamma_n t_n}\left(\widehat{\mathbb{H}}_n(t_n) - \mathbb{H}_n(t_n)\right)$ is Gaussian with mean zero with asymptotic variance

$$\begin{aligned}
K_n^2 &= \theta_1^2 + \frac{p_n(1-p_n)}{\gamma_n^2}\left(\theta_2 + \theta_1\left(t_n + \frac{\sigma_n}{1-\gamma}\right)\right)^2 \\
&+ \frac{2(1+\gamma)p_n\theta_1^2\sigma_n^2}{\gamma_n^2(1-\gamma)^2} + \frac{(1+\gamma)^2}{p_n\gamma_n^2}\left(\theta_3 + \theta_1\frac{p_n\sigma_n}{(1-\gamma)^2}\right)^2 \\
&- 2\frac{(1+\gamma)\sigma_n}{\gamma_n^2(1-\gamma)}\left(\theta_1\theta_3 + \theta_1^2\frac{p_n\sigma_n}{(1-\gamma)^2}\right) + o_{\mathbb{P}}(1) \\
&= s_n^2 + o_{\mathbb{P}}(1),
\end{aligned}$$

Observe now that $K_n^2 = s_n^2 + o_{\mathbb{P}}(1)$, where s_n^2 is that in theorem (4.10), this completes the proof of Theorem (4.10). ■

Conclusions and Future Research Directions

In this thesis, the use of Extreme Value Theory is advocated. This methodology provides a scientific approach for a difficult practical problem, namely that of estimating extreme risks when only little data is available. Due to its assumptions of heavy tails and the extreme value distributions that are used. For these reasons, EVT is important in the theory and practice of a category of problems that is normally hard to address. A presentation of the main methods of EVT has been made, with the support of graphical tools.

One of the most active areas in this field is extreme value theory applied to time dependent sequences and certain time series models. By adopting the general theory of point processes, there is a natural way explaining the independence between parameters.

Further research may extend the tail dependence estimations by applying multivariate distributions such as the copula methods. Another application of tail dependence is in the context of style analysis.

Out last approach is to apply the technique of proper transforms to extreme models using Bayesian inferences and/or Markov chains. A Bayesian analysis of extreme value data is desirable if we acquire other source of information through a prior distribution. It is specially helpful when the number of available observations is not large enough to apply large sample approximation in extreme value theory. The by product of Bayesian analysis is a more complete inference through the posterior distribution than the MLE method.

Bibliography

[1] Acerbi, C., Tasche, D., 2002. On the coherence of expected shortfall. J. Banking Finance 26, 1487–1503.

[2] Acerbi, C., Nordio, C., and Sirtori, C., Expected shortfall as a tool for financial risk management, Working paper, http://www.gloriamundi.org, 2001.

[3] Anderson, k, k., Athreya, K.B. (1987). A renewal theorem in the infinite mean case. Ann. Prob. 15 388-393. MR 88h:60154.

[4] Arnold, B.C. & Balakrishnan, N. and Nagaraja, H.N. A First Course in Order Statistics. Wiley, New York. (1992), 119,186,195.

[5] Artzner, P., 1999. Application of coherent risk measures to capital requirements in insurance. North Amer. Actuarial J. 3, 11–29.

[6] Artzner, Ph., Delbaen, F., Eber, J-M, Heath, D. 1999. Coherent measures of risk. Math. Finance 9, 203–228.

[7] Azzalini, A. Statistical Inference Based on the Likelihood. Chapman and Hall, London. (1996).

[8] Balkema, A.A. and de Haan L. (1974). Residual Lifetime at Great Age.Annals of Probability 2, 792-804.

[9] Bassi, F. & Embrechts, P. & Kafetzaki, M. "A survival kit on quantile estimation". Environmetrics 5 (1994), 221-239.

[10] Bebbington, M., Davydov, Y., Zitikis, R. (2007). Estimating the renewal function when the second moment is infinite. *Stoch. Models*, 23, 27–48.

[11] Beirlant, J., Teugels, J., 1987. Asymptotic normality of Hill's estimator. Extreme value theory (Oberwolfach, 1987), 148-155, Lecture Notes in Statistics, 51, Springer, New York, 1989.

[12] Beirlant, J., Goegebeur, Y., Teugels, J., Segers, J. (2004). *Statistics of Extremes: Theory and applications*. Chichester: Wiley.

[13] Beirlant, J., Matthys, G., Dierckx, G. (2001). Heavy-tailed distribution and rating. *ASTIN Bull.*, 31, 37–58.

[14] Beirlant, J., Teugels, J. (1989). Asymptotic normality of Hill's estimator. In J. Hüsler, R.-D. Reiss (Eds.) *Extreme Value Theory* (pp. 148–155). New York: Springer.

[15] Bingham, N., Goldie, C., and Teugels, J. (1987), Regular Variation, no. 27 in Encyclopedia of Mathematics and its Applications, Cambridge University Press.

[16] Bradley, B.O. and Taqqu, M.S. (2002). Financial Risk and Heavy Tails. \Heavy-tailed distributions in Finance", Svetlozar T. Rachev, editor, North Holland.

[17] Brazauskas, V., Jones, B. L., Puri, M. L., Zitikis, R. (2008). Estimating conditional tail expectation with actuarial applications in view. *J. Statist. Plann. Inference*, 138, 3590–3604.

[18] Cai, J., Li, H. (2005). Conditional tail expectations for multivariate phase-type distributions. *J. Appl. Probab.*, 42, 810–825.

[19] Castillo, E., Hadi, A. S., Balakrishnan, N., Sarabia, J. M. (2005). *Extreme Value and Related Models with Applications in Engineering and Science*. Hoboken, NJ: Wiley.

[20] Cheng, S., Peng, L. (2001). Confidence intervals for the tail index. *Bernoulli*, 7, 751–760.

[21] Coles, S. (2001), An Introduction to Statistical Modeling of Extreme Values, Springer-Verlag, London, Spring Series in Statistics.

[22] Crovella, M. E., Taqqu, M. S., and Bestavros, A. (1998), "Heavy-Tailed Probability Distributions in the World Wide Web," in A Practical Guide to Heavy Tails, eds.

[23] Csörgő, M., Mason, D. M. (1985). Central limit theorems for sums of extreme values. *Math. Proc. Camb. Phil. Soc.*, 98, 547–558.

[24] Csörgő, M., Csörgő, S., Horváth, L., Mason, D. M. (1986). Weighted empirical and quantile processes. *Ann. Probab.*, 14, 31–85.

[25] Csörgő, S., Deheuvels, P. and Mason, D. (1985). Kernel Estimates of the Tail Index of a Distribution. Annals of Statistics 13, 1050-1077.

[26] Csörgő, S. and Mason, D.M. (1985). Central Limit Theorems for Sums of Extreme Values. Mathematical Proceedings of the Cambridge Philosophical Society 98, 547-558.

[27] Csörgő, M., Csörgő, S., Horvarth, L. and Mason, D. (1986). Weighted Empirical and Quantile processes. Annals of Probability 14, 31-85.

[28] Csörgő, S. and Viharos, L. (1998). Estimating the Tail Index. Asymptotic Methods in Probability and Statistics. B. Szyszkowicz, ed., North Holland, 833-881.

[29] Danielsson, J. and de Vries, C.G. (1997). Tail Index and Quantile Estimation with Very High Frequency Data. Journal of Empirical Finance 4, 241-257.

[30] Danielsson, J., de Haan, L. Peng, L., de Vries, C. G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *J. Multivariate Anal.*, 76, 226–248.

[31] Davis, R. and Resnick, S.I. (1984). Tail Estimates Motivated by Extreme Value Theory. Annals of Statistics 12, 1467{1487.

[32] Davison, A.C. and Smith R.L. (1990). Models for Exceedances over High Thresholds (with discussion). Journal of the Royal Statistical Society, Series B 52, 393-442.

[33] Davydov, Y., Zitikis, R. (2003). Generalized Lorenz curves and convexifications of stochastic processes. *J. Appl. Probab.*, 40, 906–925.

[34] Davydov, Y., Zitikis, R. (2004). Convex rearrangements of random elements. In L. Horváth and B. Szyszkowicz (Eds.) *Asymptotic Methods in Stochastics* (pp. 141–171). Providence, RI: American Mathematical Society.

[35] Daykin, C.D., PentikÄainen, T. and Pesonen, M. (1994). Practical Risk Theory for Actuaries. Chapman and Hall, London.

[36] Deheuvels, P., Häusler, E. and Mason, D.M. (1988). Almost Sure Convergence of the Hill Estimator. Mathematical Proceedings of the Cambridge Philosophical Society 104, 371-381.

[37] Dekkers, A. L. M., Einmahl, J. H. J., de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *Ann. Statist.*, 17, 1833–1855.

[38] Dekkers, A. L. M., de Haan, L. (1989). On the estimation of the extreme-value index and large quantile estimation. *Ann. Statist.*, 17, 1795–1832.

[39] Dekkers, A. L. M., de Haan, L. (1993). Optimal choice of sample fraction in extreme-value estimation. *J. Multivariate Anal.*, 47, 173–195.

[40] Dekkers, A. L. M., Einmahl, J. H. J., de Haan, L., 1989. A moment estimator for the index of an extreme-value distribution. Ann. Statist. 17, 1833-1855.

[41] Delbaen, F., Coherent measures of risk on general probability spaces, Working papers, http://math.ethz.ch/ delbaen, 2000

[42] Denuit, M., Dhaene, J., Goovaerts, M., Kaas, R. (2005). *Actuarial Theory for Dependent Risks: Measures, Orders and Models.* Chichester: Wiley.

[43] Draisma, G., de Haan, L., Peng, L. and Perreira, T.T. (1999). A Bootstrap-Based Method to Achieve Optimality in Estimating the Extereme-Value Index. Extremes 2, 367-404.

[44] Drees, H.(1996). Refined Pickands Estimator of the Extreme Value Index. Annals of Statistics 23, 2059-2080.

[45] Drees, H., Kaufmann, E. (1998). Selection of the optimal sample fraction in univariate extreme value estimation. *Stochastic Process. Appl.*, 75, 149–195.

[46] Embrechts, P., Klüppelberg, C., Mikosch, T. (1997), Modelling Extremal Events for Insurance and Finance, Springer Verlag, Berlin.

[47] Erickson, K.B. (1970). Strong renewal theorem infinite mean. Trans. Amer. Math. Soc.151 263-291. MR 42:3873.

[48] Feller, W. (1971). An introduction to probability theory and its applications 2, 2nd ed. Wiley, New York. MR 42:5292.

[49] Fisher, R.A. & Tippet, L.H.C. "Limiting forms of the frequency distribution of the largest or smallest number of the sample". Proc. Cambridge Philos. Soc. 24 (1928), 180-190.

[50] Fraga Alves, M.I. (1995). Estimation of the Tail Parameter in the Domain of Attraction of an Extremal Distribution. Journal of Statistical Planning and inference 45, 143-173.

[51] Fraga Alves, M.I., Gomes, M.I., and de Haan, L. (2003). A New Class of Semi-Parametric Estimators of the Second Order Parameter. Portugaliae Mathematica 60, 194{213.

[52] Frees, E.W. Nonparametric renewal function estimation. Ann. Statist. 1986, 14, 1366–1378.

[53] Furman, E., Landsman, Z. (2006). Tail variance premium with applications for elliptical portfolio of risks. *ASTIN Bull.*, 36, 433–462.

[54] Furman, E., Zitikis, R. (2008). Weighted risk capital allocations. *Insurance Math. Econom.*, 43, 263–269.

[55] Geluk, J., de Haan, L., Resnick, S., Stărică, C. (1997). Second order regular variation, convolution and the central limit theorem. *Stochastic Process. Appl.*, 69, 139–135.

[56] Geluk, J. L. (1997) A renewal theorem with infinite-mean. Trans. Amer. Math. Soc. 125, Number 11 3407-3413.

[57] de Glyn A. Holton (2003). Value-At-Risk: Theory and Practice.

[58] Johansson, J. (2003) Estimating the mean of heavy-tailed distributions, Extremes 6, 91-131.

[59] Gnedenko, B.V. (1943). Sur la Distribution Limite du Terme Maximum d'une Série Aléatoire. Annales de Mathématiques 44, 423-453.

[60] Gnedenko, B.V. and Kolmogorov, A.N. (1954). Limit Distributions for Sums of Independent Random Variables. GITTL, Moscow; English Translation, Addison-Wiley.

[61] Gomes, M. I. and Figueiredo, F. (2006). Bias reduction in risk modelling: semi-parametric quantile estimation. Test 15:2, 375-396.

[62] Gomes, M.I., and Martins, M.J. (2002). "Asymptotically unbiased" estimators of the tail index based on external estimation of the second order parameter. Extremes 5:1, 5-31.

[63] Gomes, M. I. and Pestana, D. (2005). A sturdy reduced bias extreme quantile (VaR) estimator. J. Amer. Statist. Assoc., to appear.

[64] Gomes, M. I., Figueiredo, F., Mendonça, S. (2005). Asymptotically best linear unbiased tail estimators under a second-order regular variation condition. *J. Statist. Plann. Inference*, 134, 409–433.

[65] Gomes, M. I., Pestana, D. (2007). A simple second-order reduced bias' tail index estimator. *J. Stat. Comput. Simul.*, 77, 487–504.

[66] Gomes, M. I., Pestana, D., Caeiro, F. (2009) A note on the asymptotic variance at optimal levels of a bias-corrected Hill estimator. *Statist. Probab. Lett.*, 79, 295–303.

[67] Grandell, H. (1991). Aspects of Risk Theory. Springer, Berlin.Grandell, H. (1991). Aspects of Risk Theory. Springer, Berlin.

[68] Groeneboom, P., Lopuhaä, H.P.and de Wolf, P.P. (2003). Kernel Estimators for the Extreme Value Index. Annals of Statistics 31, 1956-1995.

[69] Gumbel, E.J. (1958). Statistics of Extremes. Columbia University Press, New York.

[70] de Haan, L., Ferreira, A., 2006. Extreme Value Theory: An introduction. Springer.

[71] de Haan, L., Stadtmüller, U., 1996. Generalized regular variation of second order. J. Austral. Math. Soc. Ser. A 61, 381-395.

[72] de Haan, L. and Resnick, S. (1996). Second Order Regular Variation and Rates of Convergence in Extreme Value Theory. Annals of Probability 24, 97-124.

[73] de Haan, L. and Peng, L. (1998). Comparison of Tail Index Estimators. Statistica Neerlandica 52, 60-70.

[74] Hall, P. (1982). On some simple estimates of an exponent of regular variation. *J. R. Statist. Soc.*, 44, 37–42.

[75] Hall, P., Welsh, A. H. (1985). Adaptative estimates of parameters of regular variation. *Ann. Statist.*, 13, 331–341.

[76] Hardy, M. R., Wirch, J. L. (2004). The iterated CTE: a dynamic risk measure. *N. Am. Actuar. J.*, 4, 62–75.

[77] Häusler, E. and Teugels, J.L. (1985). On Asymptotic Normality of Hill's estimator for the Exponent of Regular Variation. Annals of Statistics 13, 743-756.

[78] Hill, B. M. (1975). A simple approach to inference about the tail of a distribution. *Ann. Statist.*, 3, 1136–1174.

[79] Hosking, J.R.M.,Wallis, J.R. and Wood, E.F. (1985). Estimation of the Generalized Extreme Value Distribution by the Method of Probability Weighted Moments. Technometrics 27, 251-261.

[80] Hosking, J.R.M. and Wallis, J.R. (1987). Parameter and Quantile Estimation for the Generalized Pareto Distribution. Technometrics 29, 339-349.

[81] Ihaka, R., Gentleman, R., (1996). R: A language for data analysis and graphics. J. Comput. Graph. Statist. 5, 299-314.

[82] Landsman, Z. M., Valdez, E. A., 2003. Tail conditional expectations for elliptical distributions. North Amer. Actuarial J. 7, 55–71.

[83] Leadbetter, M.R., Lindgren, G. and Rootzèn, H. (1983). Extremes and Related Properties of Random Sequences and Processes. Springer, Berlin.

[84] Lehmann, E. L. and Casella, G., 1998. Theory of Point Estimation. Springer.

[85] Levy, J. B. and Taqqu, M. S. (2000), "Renewal Reward Processes with Heavy-Tailed Inter-Renewal Times and Heavy-Tailed Rewards," Bernoulli, 6, 23–44.

[86] Lévy, P. (1925). Calcul des probabilities. Paris, Gauthier-Villars.

[87] Kelly, F. P., Zachary, S., and Ziedins, I. (eds.) (1996), Stochastic Networks, Theory and Applications, vol. 4 of Royal Statistical Society Lecture Note Series, Oxford University Press.

[88] Kratz, M.F. and Resnick, S.I. (1996). The QQ-Estimator of the Index of Regular variation. Communications in Statistics - Stochastic Models 12, 699-724.

[89] Manistre, B. J., Hancock, G. H. (2005). Variance of the CTE estimator. *N. Am. Actuar. J.*, 9, 129–156.

[90] Markovich, N.M. (2005). High Quantile Estimation for Heavy-Tailed Distributions. Performance Evaluation 62, 178-192.

[91] Markovitch, N.M., Krieger, U.R. Non Parametric Estimation of the Renewal Function by Empirical Data. Stochastic Models, 22:175–199, 2006

[92] Markovitch, N.M., Krieger, U.R. Nonparametric estimation of long-tailed density functions and its application to the analysis of World Wide Web traffic. Performance Evaluation 2000, 42,205–222.

[93] Markovich, N.M. (2007). Nonparametric Analysis of Univariate Heavy-Tailed Data. John Wiley.

[94] Mason, D. M. (1982). Laws of the large numbers for sums of extreme values. *Ann. Probab.*, 10, 754–764.

[95] Matthys, G., Beirlant, J. (2003). Estimating the extreme value index and high quantiles with exponential regression models. *Statist. Sinica*, 13, 853–880.

[96] Mohan, N. R. (1976). Teugels renewal theorem and stable laws, Ann. Prob. 4 863-868. MR 54:6312.

[97] McNeil, A.J. (1997), "Estimating the tails of loss severity distributions using extreme value theory", Theory ASTIN Bulletin, Vol. 27 No.1, pp.1117-37.

[98] McNeil, A.J. (1998), "Calculating quantile risk measures for financial return series using extreme value theory", Department of Mathematics, ETH E-Collection, Swiss Federal Technical University, Zurich.

[99] McNeil, A.J. (1999), "Extreme value theory for risk managers", Internal Modeling CAD II, Risk Books, London, pp.93-113.

[100] Meraghni, D., (2008). Modeling Distribution Tails. Doctorat thesis, Med Khider University, Biskra-Algeria.

[101] Necir, A., Meraghni, D., Meddi, F. (2007). Statistical estimate of the proportional hazard premium of loss. *Scand. Actuarial J.*, 3, 147–161.

[102] Necir, A., Meraghni, D. (2009). Empirical estimation of the proportional hazard premium for heavy-tailed claim amounts. *Insurance Math. Econom.*, 45, 49–58.

[103] Necir, A., Rassoul, A. ,Meraghni, D. (2010). POT-based estimation of the renewal function of inter-occurrence times of heavy-tailed risks.Journal of Probability and Statistics. Volume 2010 (2010), Article ID 965672, doi: 10.1155/2010/965672

[104] Necir, A., Rassoul, A. , Zitikis, R. (2010). Estimating the Conditional Tail Expectation in the case Heavy-Tailed Losses. Journal of Probability and Statistics. Volume 2010 (2010), Article ID 596839, doi:10.1155/2010/596839

[105] Neves, C., Fraga Alves, M. I. (2004). Reiss and Thomas' automatic selection of the number of extremes. *Comput. Statist. Data Anal.*, 47, 689–704.

[106] Peng, L. (2001). Estimating the mean of a heavy-tailed distribution. *Statist. Probab. Lett.*, 52, 255–264.

[107] Peng, L., Qi, Y. (2004). Estimating the first and second-order parameters of a heavy-tailed distribution. *Aust. N. Z. J. Stat.*, 46, 305–312.

[108] Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. Annals of Statistics 3, 119-131.

[109] Pipiras, V. and Taqqu, M. S. (2000), "The Limit of a Renewal Reward Process with Heavy-Tailed Rewards is not a Linear Fractional Stable Motion," Bernoulli, 6, 607–614.

[110] Reiss, R.D. and Thomas, M. (2007). Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields. Birkhäuser, Basel.

[111] Resnick, S. (1987), Extreme Values, Regular Variation and Point Processes, Springer-Verlag.

[112] Resnick, S. I. (2007). *Heavy-tail Phenomena: Probabilistic and Statistical Modeling.* New York: Springer.

[113] Romano, J. P. and Wolf, M. (1999), "Subsampling inference for the mean in the heavy-tailed case," Metrika, 50, 55–69.

[114] Seneta, E. (1976), Regularly Varying Functions, no. 508 in Lecture Notes in Mathematics, Springer-Verlag.

[115] Sgibnev, M.S. Renewal theorem in the case of an infinite variance. Siberian Math. J. 1981, 22, 787–796.

[116] Sousa, B.C. (2002). A Contribution to the Estimation of the Tail Index of Heavy-Tailed Distributions. PhD. Thesis, University of Michigan.

[117] Smith, R. (1987), "Estimating Tails of Probability Distributions," The Annals of Statistics, 15, 1174–1207.

[118] Smith, R. L. (1985), "Maximum likelihood estimation in a class of nonregular cases,"Biometrika, 72, 67–90.

[119] Schneider, H.; Lin, B.S.; O'Cinneide, C. Comparison of nonparametric estimators for therenewal function. J. Royal Statist. Soc., Series C. 1990, 39, 55–61.

[120] Teugels, J.L. (1968). Renewal theorem when the first or the second moment is infinite. Ann. Math. Statist. 39 1210-1219. MR 37:5952.

[121] Von Mises, R. (1936). La Distribution de la plus Grande des n Valeurs. Selected papers, American Mathematical Society, 271-294.

[122] Wang, S.S. (1996). Premium Calculation by Transforming the Layer Premium Density. ASTIN Bulletin 26, 71-92.

[123] Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *J. Amer. Statist. Assoc.*, 73, 812–815.

[124] Wellner, J. A. (1978). Limit theorems for the ratio of the empirical distribution function to the true distribution function. *Z. Wahrsch. Verw. Gebiete*, 45, 73–88.

[125] de Wolf, P.P. (1999). Estimating the Extreme Value Index. Ph.D. Thesis, Delft University of Technology.

[126] Yun, S. (2002). On a Generalized Pickands Estimator of the Extreme Value Index. Journal of Statistical Planning and Inference 102, 389-409.

[127] Zitikis, R. (1998). The Vervaat process. In B. Szyszkowicz (Ed.) *Asymptotic Methods in Probability and Statistics* (pp. 667–694). Amsterdam: North-Holland.

[128] Zhongxian Han, B.S., M.A.S. (2003) Actuarial Modelling of Extremal Events using Transformed Generalized Extreme Value Distributions and Generalized Pareto Distribution. Ph.D. Thesis, The Ohio State University.

Résumé

La théorie des valeurs extrêmes (EVT) est apparue en 1928, dans le travail de Fisher et de Tippett décrivant le comportement du maximum d'une suite des variables aléatoires indépendantes et identiquement distribuées. Des diverses applications ont été mises en application avec succès dans beaucoup de domaines comme: science actuarielle, finances, sciences économiques, hydrologie, climatologie, télécommunications et sciences de la technologie. Dans cette thèse, nous présentons une vue d'ensemble sur la théorie des valeurs extrêmes et les différentes méthodes d'estimation d'index de la queue de distribution et des quantiles extrêmes.

Cette thèse comporte deux applications de la théorie de valeur extrême, en particulier, quand l'index de valeur extrême est positif, qui correspond à la classe des distributions à queue lourdes fréquemment utilisées, dans des modèles à des ensembles de données réels. Le premier consiste en une application dans le domaine actuariel, pour estimer une des mesures de risque plus usuelles, est appelé l'espérance conditionnelle de queue (CTE). La deuxième contribution est une application importante dans les domaines de la fiabilité des systèmes et dans la télécommunication, pour estimer la fonction de renouvellement.

ملخص

إن نظرية القيم القصوى التي ظهرت سنة 1928 في أعمال "فيشز و تيبت" وذلك في دراسة السلوك اللانهائي للقيمة العظمى لسلسلة متغيرات عشوائية موزعة بطريقة مستقلة و متشابهة. الكثير من التطبيقات أخذت في مختلف المجالات على غرار المالية, الاقتصاد, الهيدرولوجية, علم المناخ, التكنولوجيا وعلوم الاتصالات.

في هذه الأطروحة ، نقدم لمحة عامة حول نظرية القيم القصوى وأساليب مختلفة لتقدير الرقم القياسي من الذيل والتوزيعات الحدية القصوى.

هذه الرسالة تحتوي على نوعين من التطبيقات لنظرية القيمة القصوى ، وبصفة خاصة عندما يكون مؤشر القيمة القصوى موجب، والتي تتطابق مع فئة من التوزيعات الثقيلة الذيل التي يشيع استخدامها في نماذج لمجموعات ذات بيانات حقيقية. التطبيق الأول هو تطبيق في مجال المالية لتقدير واحد من أهم مقاييس المخاطر والذي يعد الأكثر شهرة ويطلق عليه اسم "توقع ذيل الشرطي". التطبيق الثاني هو تطبيق هام في مجال فعالية الأنظمة، وأنظمة الأمن وذلك لتقدير مايعرف بدالة التجديد.