Statistical assessment of extreme values of boron and arsenic in water of Arica and Parinacota

Ludy Mireya Núñez Soza

Supervisor: **Prof. Milan Stehlík**

Instituto de Estadística
Universidad de Valparaíso

This dissertation is submitted for the degree of
Doctor en Estadística

Executive Summary

The prediction of events that by their magnitude do not happen frequently and that have an impact on the environment is of great interest in various fields. Many researchers have focused on calculating with a high level of certainty the possible occurrence of these events, which at present have happened a little more often than expected. In many data analysis studies, "extreme events" are considered as outliers and therefore are ignored. Within the Theory of Probability and Statistics these unusual or rare events are known as extreme values, which, when plotted on a histogram are located in the tails of the distributions. At present it has been found that for a large number of data sets that these situations are heavier than classical distributions predict. Therefore, if we want to better model these heavy tails data series, we must have an appropriate mathematical method that explains the distribution with which such events occur. The answer to this problem is the **Extreme Value Theory**.

The extreme values theory is responsible for modeling the behavior of the maximum and/or minimum values of a data set, trying to find the shape of the limit distribution by which these values can be approximated. The limit distributions to which these values can converge are known as extreme values distributions. This asymptotic theory is similar to the Central Limit Theorem, because the extreme value theory focuses on the behavior of a sample of extremes.
According to Fréchet (1927), Nicolas Bernoulli was the first to introduce in 1709 the discussion of larger mean distances from the origin. Later in 1927, the French mathematician Maurice Fréchet obtained the first asymptotic distribution for maximum order statistics. Tippet and Fisher in 1928 established the theorem that perhaps is the most important in the extreme values theory. In 1936 Von Mises presented some simple and useful sufficient conditions for the convergence in distribution of the larger order statistic. The work done by these authors was the basis of the univariate theory of extreme values.
Later in 1943 Boris Gnedenko was in charge of classifying the asymptotic distributions of extreme values and of giving necessary and sufficient conditions under which such asymptotic distributions are valid. Subsequently De Haan's doctoral work in 1970 contributed to the development of this theory and the results of Pickans in 1975 were the first contribution to the multivariate extreme value theory. At present advances in the extreme values theory focus

mainly on the development of models and methods for extreme values of spatial phenomena and other more complex structures.

Historically, the theory of extreme values has been applied in the field of environmental sciences, in issues such as the flow of rivers, wind speed, extreme temperatures, etc. Also it has been used in the field of engineering, as for example, in security, structural reliability, etc. Nowadays the theory of extreme values is one of the main statistical theories applied in various sciences, for example in hydrology, finance, insurance, ecology, etc. In all above areas a matter of great interest is to predict these extreme events, which means to determine the probability distribution function associated to these extreme values and therefore the parameters of such distribution. One of the main problems in the prediction or estimation of these extreme values located in the tails of the distributions, is that standard densities estimation techniques fit well for data will higher density but may have important biases when estimating tails. Because of this, several authors have presented alternative methods of estimating. For example, Hosking et al. (1985) use the method of probability-weighted moments to derive estimators of the parameters and quantiles of the generalized extreme value distribution (GEV).

In 1970 Hill proposed a simple general approach to inference about the tail behavior of a distribution and recently Gomes and Stehlík (2014) provide recent developments on the Hill and related estimators of the extreme value index and discuss their properties, like mean square error, efficiency and robustness. On the other hand Stehlík et al. (2012) suggest to introduce the t-Hill estimator as a variant for the tail index of a Pareto distribution and Jordanova et al. (2016) analyze asymptotic characteristics of t-Hill and t-lgHill estimators. In this work we do a review of the estimators mentioned above, comparing them with the moment estimator mentioned in De Hann and Ferreira (2006) and the QQ estimator presented in Resnick (2007).

As we previously mentioned, the theory of extreme values is applied in different ambits of science, particularly in the analysis of environmental phenomena, such as atmospheric, soil and water pollution. One of the factors of water pollution is the presence of high concentrations of heavy metals in surface water and drinking water in some geographical areas. This contamination affects the quality of water in the northern zone of Chile, where precisely this resource is scarce and of poor quality, due to high salinity and high concentrations of heavy metals. In the *Región de Arica y Parinacota* the high levels of boron and arsenic recorded in the Lluta river and in the drinking water of the Parinacota province and the city of Arica require an analysis and statistical study from

the point of view of the theory of extreme values in order to determine robust estimates for the threshold of these pollutants.
Castillo and Venegas (2010) highlight that arsenic contamination is a problem affecting the northern area of Chile, so they recommend paying more attention to prevention and control. Tchernitchin et al. (2015) in their study on the presence of polymetals in drinking water of Arica and Parinacota determined that in some localities of the province of Parinacota the concentration of arsenic is higher than the Chilean norm. Cortes et al. (2011) report the results of a study where boron levels were measured in drinking water in Arica, where the concentration of boron in drinking water varied according to the sampling site (establishments, private homes and volunteer houses) between 0.22 and 11.30 mg/L, with a mean value of 4.18 mg/L, values that exceed what is recommended by the World Health Organization (WHO). This background led us to apply estimation methods from the theory of extreme values and to analyze an estimation model for the concentration of arsenic and boron in drinking water.
The aim of this thesis is to develop rigorous statistical methodology for measurement of pollutants in Chile, especially in Arica and Parinacota, in order to obtain more secure limits of toxic elements in water. Specifically, we put the focus on determining robust estimators of the tail index for the small sample case. We also propose a model to estimate arsenic and boron concentrations in drinking water.
The thesis consists of 5 chapters, in which various topics were studied according to the proposed objective. Chapter 1 presents the generalized extreme values distribution, which encompasses the three most known distributions in the theory of extreme values. We refer to the Fréchet, Gumbel and Reversed Weibull distributions. We also present other distributions for extreme values such as the Gamma distribution, the Beta distribution, the Weibull distribution and the generalized Pareto distribution. For each of the distributions mentioned above, the main characteristics are provided.
In Chapter 2 we review the maximum likelihood estimate applied in extreme value distributions. In addition, the method of moments is presented as an alternative in the estimation of the parameters of extreme value distributions.
The estimators for the shape parameter of the Pareto distribution are analyzed in Chapter 3, where important properties of these estimators are also mentioned. Furthermore, in this chapter alternative normality tests are presented to the commonly used tests, these tests have been applied for smaller data sets.

The necessary and sufficient conditions for a distribution of extreme value F to belong to the attraction domain of another distribution of extreme value G_γ for the $k-$dimensional (multivariate) case are analyzed in Chapter 4, together with a procedure for estimating a bivariate distribution.
Finally, in Chapter 5, we apply the tail index estimation methods, first to data from the *Dirección General de Aguas (DGA)* corresponding to measurements of arsenic and boron concentration in surface water of the Lluta river in the period $2000-2016$, which were obtained online on the DGA website. We then apply this methodology to estimate the shape parameter of the Pareto distribution associated with both the concentration of arsenic and the concentration of boron in drinking water. The arsenic concentration data was provided by the *Secretaria Regional Ministerial (SEREMI) de Salud de Arica y Parinacota* and corresponds to measurements made in 2014, in various monitoring sites of the Parinacota province. As for the data associated with the concentration of boron in drinking water, these correspond to measurements made between 2006 and 2008 by Cortes et al. (2011) at various monitoring sites of the city of Arica. This chapter also considers a model that allows predicting the concentration of arsenic and boron in drinking water, sometimes such model does not yield normal errors, therefore we provide a procedure to estimate the distribution function of these errors, together with the results of the normality tests of the RT class. Besides in Chapter 5, we show the dependence of arsenic observations measured in the Parinacota province, together with the bivariate joint distribution associated with these observations.

I dedicate this dissertation to Jesus, my inseparable friend, to my family and to my friends who always supported me throughout the process.

"Whoever drinks the water I shall give will never thirst; the water I shall give will become in him a spring of water welling up to eternal life"

John 4, 14

Acknowledgements

I would like to express my sincere gratitude to my advisor, Dr. Milan Stehlík. This research would not have been possible without relevant guidance and talent and without their motivation and constant support.
I thank the *Universidad de Tarapacá*, which granted the necessary permits and financial support, during this period. I also thank the *Universidad de Valparaíso* and the professors of the *Programa de Doctorado en Estadística* for providing me with the necessary tools to carry out this task.
My profound gratitude to God in the person of Jesus who is my strength and joy and light that illuminates my way, balsam of my life. A special thank you to my mother, brother and sisters, nephew and nieces, brothers-in-law who during this stage have been a fundamental support. Finally I thank my friends, Elizabeth, Cristián and Francisco for brightening my days and for their sincere friendship.

Abstract

Spanish Version

El problema de predecir los fenómenos extremos suscita gran interés de muchos investigadores. Diversas técnicas se han desarrollado con este propósito, las cuales están enmarcadas dentro de la teoría de valores extremos. En la zona norte de Chile, particularmente en la Región de Arica y Parinacota un evento extremo es la presencia de altas concentraciones de metales pesados en el agua potable. De aquí nace la necesidad de predecir con un alto grado de certeza la posibilidad de que ocurran estos eventos. En esta tesis consideramos estimadores robustos de los parámetros de la distribución asociada a estos eventos extremos. También proponemos un modelo para estimar las concentraciones de estos metales pesados y realizamos pruebas de normalidad para los errores de este modelo. Finalmente consideramos la posibilidad que las observaciones realizadas sean dependientes y posteriormente estimamos la distribución conjunta bivariada.

English Version

The problem of predicting extreme events is of great interest to many researchers. Various techniques has been developed for this purpose, which they are framed within the extreme value theory. In the northern zone of Chile, particularly in the *Región de Arica y Parinacota*, an extreme event is the presence of high concentrations of heavy metals in drinking water. From this arises the need to predict with a high degree of certainty the possibility of that these events occur. In this thesis we consider robust estimators of the parameters of the distribution associated with these extreme events. We also propose a model to estimate the concentrations of these heavy metals and perform normality tests for the errors of this model. Finally we consider the possibility that the observations are dependent and then estimate the bivariate joint distribution.

Table of contents

List of Figures

List of Tables

Chapter 1

Extreme Value Distributions

1.1 Background of Extreme Value Theory

Historically, the theory of extreme values has been applied in the field of environmental sciences, in issues such as the flow of rivers, wind speed, extreme temperatures, etc. Also it has been applied in the field of engineering, as for example, in security, structural reliability, etc. Nowadays the theory of extreme values is one of the main statistical theories applied in various sciences, for example in hydrology, finance, insurance, ecology, etc.
The works of Fréchet (1927), Fisher and Tippet (1928) and Von Mises (1936) led to the gradual development of the main theoretical result of the extreme values, which is the determination of the distribution function for maximum. It is also important to mention Boris Gnedenko who in 1943 showed such a distribution. In 1958 the book "Statistics of extremes" appeared, in which the statistical theory of extreme values began to be applied to problems arising from engineering. Until the beginning of the 70s only the theoretical development started, with the thesis of De Haan (1970) and the results of statistical inference developed by Pickand (1975); such results supposed the first contributions to the theory of multivariate extreme values and gave rise to the development of alternative models that are based on the exceedance of thresholds.

1.1.1 Generalized extreme value distribution

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent random variables, which have a common distribution function F. The maximum M_n is

$$M_n = \max\{X_1, X_2, \ldots, X_n\} \tag{1.1}$$

As $X_i,\ i = 1, \ldots, n$ is a sequence of i.i.d. random variables, the distribution function of M_n is obtained directly

$$\begin{aligned} P(M_n \leq z) &= P(X_1 \leq z, X_2 \leq z, \ldots X_n \leq z) \\ &= P(X_1 \leq z)P(X_2 \leq 2) \ldots P(X_n \leq z) \\ &= (F(z))^n \end{aligned}$$

The previously obtained distribution function converges to zero as $n \longrightarrow \infty$, for $z > z^*$ and to one for $z \leq z^*$, with $z^* := \sup\{z : F(z) < 1\}$. Therefore, in order to obtain nondegenerate distribution, a normalization is necessary which is based on the Central Limit Theorem. So, if we assume that there exists sequences of real constants $\{a_n > 0\}$ and $\{b_n\}$, $n \geq 1$, such that

$$M_n^* = \frac{M_n - b_n}{a_n} \tag{1.2}$$

have a nondegenerate distribution as $n \longrightarrow \infty$, i.e.,

$$\lim_{n \to \infty} F^n(a_n z + b_n) = G(z) \tag{1.3}$$

The entire range of possible limit distributions for M_n^* is given for the Fisher-Tippet-Gnedenko Theorem or Extreme Values Theorem (De Hann and Ferreira, 2006)

Theorem 1.1.1. (**Fisher and Tippet**(**1928**), **Gnedenko**(**1943**)) *If there exists sequences of real constants $\{a_n > 0\}$ and $\{b_n\}$, such that*

$$P\left(\frac{M_n - b_n}{a_n} \leq z\right) \longrightarrow G(z) \quad \text{as } n \longrightarrow \infty \tag{1.4}$$

where $G(z)$ is a nondegenerate distribution, then G corresponds to a Generalized Extreme Value distribution (GEV):

$$G(z) = \exp\left\{-\left[1 + \gamma\left(\frac{z-\mu}{\sigma}\right)\right]^{-\frac{1}{\gamma}}\right\}, \ \text{defined on } \left\{z : 1 + \gamma\left(\frac{z-\mu}{\sigma}\right) > 0\right\}, \tag{1.5}$$

where $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \gamma < \infty$, μ is the location parameter and σ the scale parameter.

The proof of this Theorem can be found in Leadbetter et al. (1983). This cumulative distribution function (cdf) limit $G(z)$ can be classified in three types acoording to the shape parameter γ, also called Extreme Value Index (EVI). We concidere the subclasses $\gamma > 0$, $\gamma = 0$ and $\gamma < 0$ separately

1. For $\gamma > 0$ and $\alpha = \frac{1}{\gamma}$, we get the class of distributions Fréchet.

$$G(z;\mu,\sigma,\gamma) = \begin{cases} 0, & y \leq 0 \\ \exp(-y^{-\alpha}), & y > 0 \end{cases} \tag{1.6}$$

   where $y = 1 + \gamma\left(\frac{z-\mu}{\sigma}\right)$.

2. For $\gamma = 0$, the distribution function is

$$G(z;\mu,\sigma,0) = \exp\left(-\mathrm{e}^{-\frac{z-\mu}{\sigma}}\right) \qquad z \in \mathbb{R} \tag{1.7}$$

   This distribution is called Gumbel distribution or Double Exponential distribution.

3. For $\gamma < 0$, $\alpha = -\frac{1}{\gamma} > 0$ and $y = -\left(1 + \gamma\left(\frac{z-\mu}{\sigma}\right)\right)$ we get

$$G(z;\mu,\sigma,\gamma) = \begin{cases} \exp(-(-y)^{\alpha}), & y < 0 \\ 1, & y \geq 0 \end{cases} \tag{1.8}$$

   This class is sometimes called Reversed Weibull distribution.

Example 1.1.1. *Figure 1.1 shows the densities obtained for different values of the parameters of the generalized extreme value (GEV) distribution.*

1. *Left figure: **Fréchet distribution** with $\mu = 1$, $\sigma = 0.5$ and $\gamma > 0$. In this graphic we observe that the distribution is more leptocurtic as $\frac{1}{\gamma}$ increases. We also note that the curves are bounded inferiorly.*

2. *Central figure **Gumbel distribution** with $\mu = 1$, $\gamma = 0$ and $\sigma > 0$. Here the distribution becomes flatter as $\sigma = \lambda$ increases and the curves are not bounded.*

3. *Right figure: **Reversed Weibull distribution**, with $\mu = 1$, $\sigma = 0.8$ and $\gamma < 0$, where the distribution is more leptocurtic as $\alpha = \frac{1}{\gamma}$ increases and the curves are bounded superiorly.*

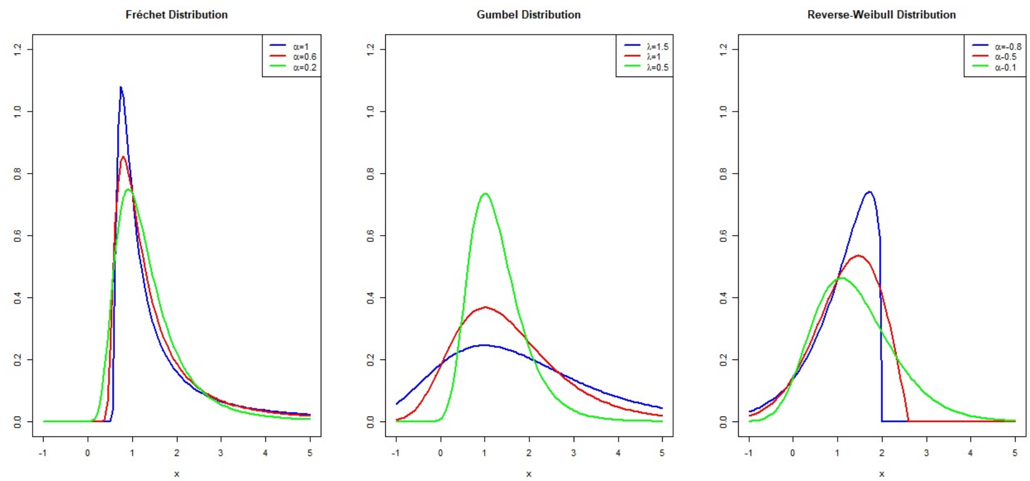

Figure 1.1: Examples of density function for various values of the parameters of the Generalized extreme value (GEV) distribution

1.1.2 Domain of Attraction

Before establishing necessary and sufficient conditions for a distribution function F to belong to the domain of attraction of G_γ, we will state the following theorem.

Theorem 1.1.2. *Let $a_n > 0$ and b_n real sequences of constans and G a non-degenerate distribution function. The following statements are equivalent:*

1.

$$\lim_{n \longrightarrow \infty} F^n(a_n X + b_n) = G(x), \tag{1.9}$$

for each continuity point x of G.

2.

$$\lim_{t \longrightarrow \infty} t\,(1 - F(a(t)x + b(t))) = -\log G(x), \tag{1.10}$$

for each continuity point x of G for which $0 < G(x) < 1$, $a(t) := a_{[t]}$, and $b(t) := b_{[t]}$ (with $[t]$ the integer part of t).

3.

$$\lim_{t \longrightarrow \infty} \frac{U(tx) - b(t)}{a(t)} = D(x), \tag{1.11}$$

for each $x > 0$ continuity point of $D(x) = G^{\leftarrow}(\mathrm{e}^{-1/x})$, $a(t) := a_{[t]}$, and $b(t) := b_{[t]}$.

We will determine the conditions of the attraction domain starting from condition (1.11) from Theorem 1.1.2, with $D(x) = \frac{x^\gamma - 1}{\gamma}$. That is,

$$\lim_{t \longrightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma} \tag{1.12}$$

for all $x > 0$, where γ is real parameter and a a suitable positive function. We prove the following results.

Theorem 1.1.3. *The distribution function F is in the domain of atraction of the extreme value distribution function G_γ if and only if*

1. *for $\gamma > 0$: $x^* = \sup\{x : F(x) < 1\}$ is infinite and*

$$\lim_{t \longrightarrow \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-1/\gamma}, \tag{1.13}$$

for all $x > 0$. This means that the function $1 - F$ is regularly varying at infinity with index $-\frac{1}{\gamma}$.

2. *for $\gamma < 0$: x^* is finite and*

$$\lim_{t \longrightarrow 0} \frac{1 - F(x^* - xt)}{1 - F(x^* - t)} = x^{-1/\gamma}, \tag{1.14}$$

for all $x > 0$.

3. *for $\gamma = 0$: x^* can be finite or infinite and*

$$\lim_{t \longrightarrow x^*} \frac{1 - F(t + xf(t))}{1 - F(t)} = e^{-x} \tag{1.15}$$

for all real x, where f is a suitable positive function.

If (1.15) holds for some f, then $\int_t^{x^*} (1 - F(s))ds < \infty$ for $t < x^*$ and (1.15) holds with

$$f(t) := \frac{\int_t^{x^*} (1 - F(s))ds}{1 - F(t)}. \tag{1.16}$$

Theorem 1.1.4. *The distribution function F is in the domain of attraction of the extreme value distribution G_γ if and only if*

1. *for $\gamma > 0$: $F(x) < 1$ for all x, $\int_1^\infty \frac{(1-F(x))}{x} dx < \infty$, and*

$$\lim_{t \longrightarrow \infty} \frac{\int_1^\infty (1-F(x)) \frac{dx}{x}}{1-F(t)} = \gamma; \tag{1.17}$$

2. *for $\gamma < 0$: there is $x^* < \infty$ such that $\int_{x^*-t}^{x^*} \frac{(1-F(x))}{(x^*-x)} dx < \infty$ and*

$$\lim_{t \longrightarrow 0} \frac{\int_{x^*-t}^{x^*} (1-F(x)) \frac{dx}{x^*-x}}{1-F(x^*-t)} = -\gamma; \tag{1.18}$$

3. *for $\gamma = 0$ (the right endpoint x^* may be finite or infinite):* $\int_x^{x^*} \int_t^{x^*} (1-F(s)) ds dt < \infty$ *and the function h defined by*

$$h(x) := \frac{(1-F(x)) \int_x^{x^*} \int_t^{x^*} (1-F(s)) ds dt}{\left(\int_x^{x^*} (1-F(s)) ds \right)^2} \tag{1.19}$$

satisfies

$$\lim_{t \longrightarrow x^*} h(t) = 1. \tag{1.20}$$

Remark 1.1.1. *Limit (1.17) is equivalent to*

$$\lim_{t \longrightarrow \infty} \mathrm{E}[(\log X - \log t) | X > t] = \gamma. \tag{1.21}$$

Remark 1.1.2. *Relation (1.17) will be the basis for the construction of the Hill estimator of γ.*

Remark 1.1.3. *Limit (1.18) can be interpreted as*

$$\lim_{t \longrightarrow 0} \mathrm{E}[(\log(x^* - X) - \log t) | X > x^* - t] = \gamma, \tag{1.22}$$

which will be the basis for the construction of negative Hill estimator.

Remark 1.1.4. *Relation (1.20) is equivalent to*

$$\lim_{t \longrightarrow x^*} \frac{\mathrm{E}[(X-t)^2 | X > t]}{\mathrm{E}^2[(X-t) | X > t]} = \lim_{t \longrightarrow x^*} 2h(t) = 2, \tag{1.23}$$

and this relations leads to the moment estimator of γ.

1.1.3 Life Distributions

The family of Exponential distributions is a family of the life distributions of one pamameter. It is one the most important due to the fact that several of the most commonly used families of life distributions are two or three-parameter extensions of the Exponential distributions. The Exponential distributions is a baseline for evaluating other families. Since they have only one parameter. Also they are quite simple to describe and are exceptionally susceptible to statitical analysis. Epstein and Sobel (1953) with their paper brought new attention to the uses of Exponential distribution. Marshall and Olkin (2007) provide some reasons why the family of exponential distributions plays a central role within the class of all life distributions.

Exponential Distribution

A continuous random variable X is said to have an Exponential distribution with parameter λ if its probability density function (pdf) is given by

$$f(x) = \begin{cases} \lambda \mathrm{e}^{-\lambda}, & x > 0 \\ 0, & \text{otherwise} \end{cases} \tag{1.24}$$

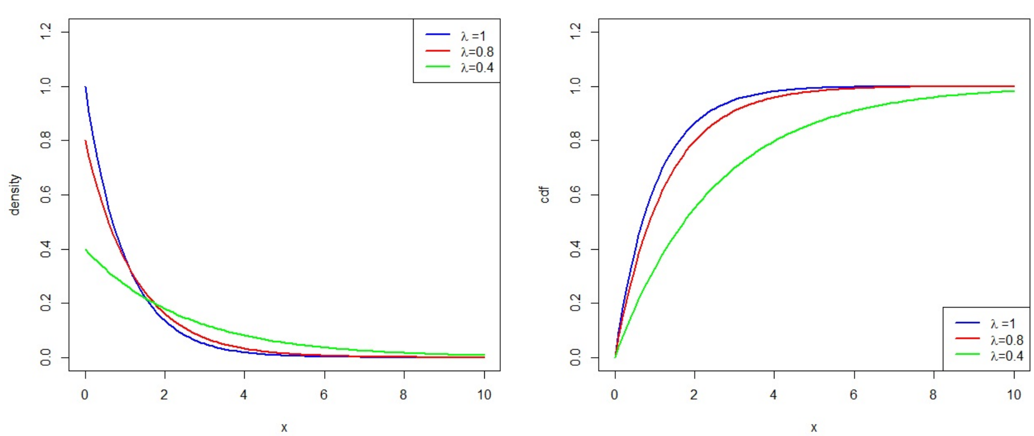

Figure 1.2: Exponential density function (*left chart*) and cumulative distribution function (*right chart*) for $\lambda = 1$, $\lambda = 0.8$ and $\lambda = 0.4$

For $X \sim \exp(\lambda)$, we have

$$F(x) = 1 - \mathrm{e}^{-\lambda x}, \qquad x \geq 0 \tag{1.25}$$

$$\mathrm{E}[X] = \frac{1}{\lambda}, \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2} \tag{1.26}$$

Remark 1.1.5. *As we have seen the Exponential distribution has a single parameter that serves as a scale parameter or as a frailty parameter. Furthermore, by introducing a parameter of age or parameter Laplace transform, the distribution is still exponential, only the parameter changes. The introduction of the parameters of moment and convolution leads to the gamma family.*

Remark 1.1.6. *The survival function $\overline{F}$ and hazard rate r of the Exponential distribution are*

$$\overline{F}(x) = \mathrm{e}^{-\lambda x}, \quad x \geq 0 \tag{1.27}$$

$$r(x) = \lambda, \quad x \geq 0 \tag{1.28}$$

Gamma Distribution

The Gamma distribution is a two-parameter family of continuous probability distributions. The exponential distribution, Erlang distribution, and chi-squared distribution are special cases of the Gamma distribution. Its pdf is given by

$$f(x;\theta,k) = \frac{x^{k-1}\mathrm{e}^{-\frac{x}{\theta}}}{\theta^k\Gamma(k)}, \quad x > 0, \quad \theta, k > 0 \tag{1.29}$$

where $\Gamma(\cdot)$ is the complete Gamma function and θ the scale parameter and k is the shape parameter. The cdf is

$$F(x;\theta,k) = \int_0^x f(u;\theta,k)du = \frac{\Upsilon(\frac{x}{\theta}), k}{\Gamma(k)}, \tag{1.30}$$

where $\Upsilon(\frac{x}{\theta}, k) = \int_0^{x/\theta} t^{k-1}\mathrm{e}^{-t}dt$ is the incomplete Gamma function.

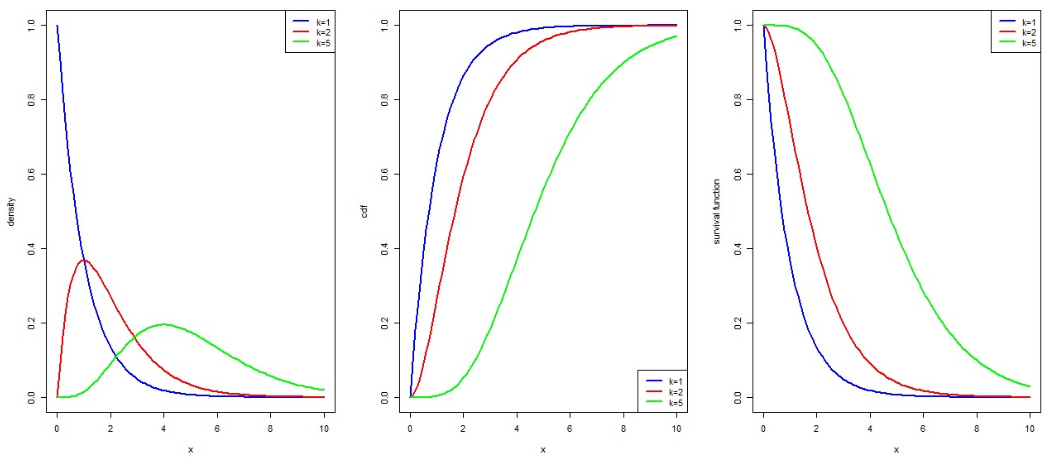

Figure 1.3: Gamma density function (*left chart*), cumulative distribution function (*central graphic*) and survival function (*right chart*) with $\theta = 1$ and $k = 1$, $k = 2$ and $k = 5$

Moments of the Gamma distribution

Proposition 1.1.1. *If F is a Gamma distribution with density (1.29), then*

$$\mathrm{E}[X^r] = \frac{\theta^r \Gamma(r+k)}{\Gamma(k)}, \quad r > -k \tag{1.31}$$

This result can be obtained directly from the definition of the Gamma function as an integral. From (1.31) follows that

1.

$$\begin{aligned} \mathrm{E}[X] &= \frac{\theta \Gamma(k+1)}{\Gamma(k)} \\ &= \frac{\theta k \Gamma(k)}{\Gamma(k)} \\ &= k\theta \end{aligned}$$

2.

$$\begin{aligned} \mathrm{E}[X^2] &= \frac{\theta^2\Gamma(k+2)}{\Gamma(k)} \\ &= \frac{\theta^2(k+1)\Gamma(k+1)}{\Gamma(k)} \\ &= \frac{\theta^2 k(k+1)\Gamma(k)}{\Gamma(k)} \\ &= k(k+1)\theta^2 \end{aligned}$$

Then

$$\begin{aligned} \mathrm{Var}(X) &= \mathrm{E}[X^2] - (\mathrm{E}[X])^2 \\ &= k(k+1)\theta^2 - (k\theta)^2 \\ &= (k^2 + k - k^2)\theta^2 \\ &= k\theta^2 \end{aligned}$$

Density Properties

Proposition 1.1.2. *The density (1.29) of the Gamma distribution is*

i. completely momotone, log convex, and decreasing, for $0 < k \leq 1$.

ii. log concave and unimodal, for $k \geq 1$, *with mode at the point*

$$x = (k-1)\theta$$

Distribution and Survival Function Properties

Proposition 1.1.3.

i. The Gamma distribution function F *is log concave, for all* k.

ii. The survival fuction $\overline{F} = 1 - F$ *corresponding to the Gamma density (1.29) is log concave, for* $k \geq 1$, *and log convex on* $[0, \infty)$, *for* $k \leq 1$.

Weibull Distribution

Considering the exponential distribution function with parameter $\frac{1}{\lambda}$, and introducing a power parameter k, the exponential survival function becomes in

$$\overline{F}(x;\lambda,k) = \exp\left\{-\left(\frac{x}{\lambda}\right)^k\right\}, \quad \lambda, k, x > 0; \tag{1.32}$$

which it is the survival function of the Weibull distibution.
From the expression (1.32) we can get directly the density of the Weibull distribution, which is given by

$$f(x;\lambda,k) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} \exp\left\{-\left(\frac{x}{\lambda}\right)^k\right\}, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{1.33}$$

Then, the Weibull distribution function with parameter of scale λ and parameter de shape k is

$$F(x;\lambda,k) = \begin{cases} 1 - \exp\left\{-\left(\frac{x}{\lambda}\right)^k\right\}, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{1.34}$$

Moments of the Weibull distribution
If X has a Weibull distribution,

$$\mathrm{E}[X^r] = \lambda^r \Gamma\left(\frac{r}{k} + 1\right), \quad r > -k \tag{1.35}$$

Thus, the first moment is

$$\mathrm{E}[X] = \lambda \Gamma\left(\frac{1}{k} + 1\right). \tag{1.36}$$

The variance the a Weibull distribution it can be explicitly expresed in term of Gamma functions using (1.35) in the definition $\mathrm{Var}(X) := \mathrm{E}[X^2] - (\mathrm{E}[X])^2$.

$$Var(X) = \lambda^2\left(\Gamma\left(\frac{2}{k} + 1\right) - \Gamma^2\left(\frac{1}{k} + 1\right)\right). \tag{1.37}$$

When the scale parameter $\lambda = 1$ and the parameter k takes large values, McEwen and Parresol (1991) have shown that

$$\mathrm{E}[X] \approx 1 - \frac{\gamma}{k} + \frac{\gamma^2 + \frac{\pi}{6}}{2k^2}, \qquad \mathrm{Var} \approx \frac{\pi^2}{6k^2},$$

where $\gamma \approx 0.577256649$ is Euler's constant. If k is small, Abramowitz and Stegun (1964, p. 257) show that

$$\mathrm{E}[X] \approx \sqrt{2\pi}\mathrm{e}^{-1/k} k^{-(1/k)-(1/2)};$$

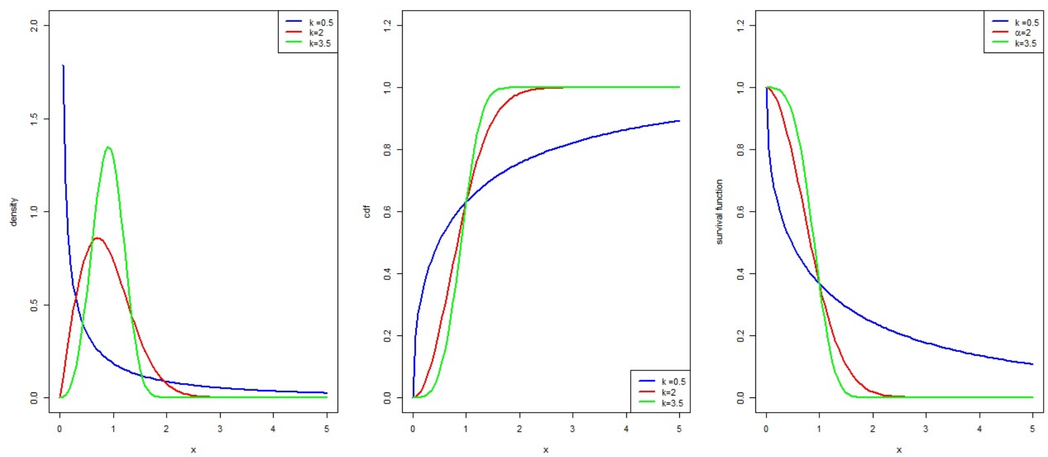

Figure 1.4: Weibull density function (*left chart*), cumulative distribution function (*central graphic*)and survival function (*right chart*) with $\lambda = 1$ and $k = 0.5$, $k = 2$ and $k = 3.5$

1.1.4 Beta Distribution

The family of Beta distributions is a two-parameters family with support $[0, 1]$. It has a number of desirable properties, and the density takes on a variety of shapes.
The two-parameters family of Beta distributions is defined in terms of the density

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha, \beta)}, \quad 0 \leq x \leq 1, \quad \alpha, \beta > 0, \tag{1.38}$$

where $\mathrm{B}(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.

The corresponding distribution function is

$$F(x; \alpha, \beta) = \frac{\mathrm{B}_x(\alpha, \beta)}{\mathrm{B}(\alpha, \beta)} = \mathrm{I}_x(\alpha, \beta), \tag{1.39}$$

where $\mathrm{B}_x(\alpha, \beta) = \int_o^x t^{\alpha-1}(1-t)^{\beta-1}dt$ is the incomplete Beta function and $\mathrm{I}_x(\alpha, \beta)$ is the regularized incomplete Beta function.

Special cases

a. $\alpha = 1$, (1.38) is the density of the distribution with survival function $\overline{F}(x;\beta) = (1-x)^{\beta}$.

b. $\beta = 1$, the density of the distribution $F(x;\alpha) = x^{\alpha}$ is obtained.

c. $\alpha = \beta = 1$, la Beta distribution is a Uniform distribution.

d. The dual of $f(x;\alpha,\beta)$ is

$$f(1-x;\alpha,\beta).$$

But in this case $f(x;\alpha,\beta) = f(1-x;\beta,\alpha)$, thus when $\alpha = \beta$, the density is symetric about $\frac{1}{2}$, consequently, it is self-dual.

e. α is an integer, the survival function $\overline{F}$ of f can be given explicity as a finite sum.

Shape of the Beta density

In terms of monotonicity and convexity, the shape of the Beta density (1.38) is determined using standard methods of calculus. Here we summarize the results of the aplication of this methods.
The shape of the Beta density (1.38) is identical to the shape of

$$\begin{aligned} g(x;\alpha,\beta) &= \mathrm{B}(\alpha,\beta) f(x;\alpha,\beta) \\ &= x^{\alpha-1}(1-x)^{\beta-1}, \qquad\qquad 0 \le x \le 1 \end{aligned}$$

To determine this shape, It is convenient to start by determining the first and second derivative of g.

$$\begin{aligned} g'(x;\alpha,\beta) &= (\alpha-1)x^{\alpha-2}(1-x)^{\beta-1} - (\beta-1)x^{\alpha-1}(1-x)^{\beta-2} \\ &= x^{\alpha-2}(1-x)^{\beta-2}\left[(\alpha-1)(1-x) - (\beta-1)x\right] \\ &= x^{\alpha-2}(1-x)^{\beta-2}\left[(\alpha-1) - (\alpha+\beta-2)x\right]. \end{aligned}$$

$$\begin{aligned} g''(x;\alpha,\beta) = &\ \left[(\alpha-2)x^{\alpha-3}(1-x)^{\beta-2} - (\beta-2)x^{\alpha-2}(1-x)^{\beta-3}\right]\left[(\alpha-1) - (\alpha+\beta-2)x\right] \\ &\ -x^{\alpha-2}(1-x)^{\beta-2}[\alpha+\beta-2] \\ = &\ x^{\alpha-3}(1-x)^{\beta-3}[(\alpha-1)(\alpha-2) - 2(\alpha-1)(\alpha+\beta-3)x \\ &\ +(\alpha+\beta-2)(\alpha+\beta-3)x^2]. \end{aligned}$$

Thus, $g'(x;\alpha,\beta)=0$ at the point

$$x_0 = \frac{\alpha-1}{\alpha+\beta-2}$$

wereas $g''(x;\alpha,\beta)=0$ at the points

$$x_- = \frac{(\alpha-1)(\alpha+\beta-3)-\sqrt{(\alpha-1)(\beta-1)(\alpha+\beta-3)}}{(\alpha+\beta-2)(\alpha+\beta-3)},$$

$$x_+ = \frac{(\alpha-1)(\alpha+\beta-3)+\sqrt{(\alpha-1)(\beta-1)(\alpha+\beta-3)}}{(\alpha+\beta-2)(\alpha+\beta-3)}.$$

Depending upon the values of α and β, x_0, x_-, and x_+ may or may not lie in the interval $(0,1)$. Additionall, x_- and x_+ may or may not be real.
As $g'(x;\alpha,\beta)=0$ for at most one point in $(0,1)$, it changes sign at most once in the interval $(0,1)$. On the other hand, because $g''(x;\alpha,\beta)=0$ at most two points in $(0,1)$, there are two points in $(0,1)$ where it changes sign.

a. $\alpha\le 1$, $\beta\le 1$. In this case, x_- and x_+ are not real so g'' is of one sign, which can easily be seen to be positive because it is posotive for x near 0. Moreover, $x_0\in(0,1)$. Thus, g is decreasing in $(0,x_0)$, increasing in $(x_0,1)$, and convex on $[0,1]$. In this case, $f(0;\alpha,\beta)=f(1;\alpha,\beta)=\infty$ (*top left graph* in Figure 1.5).

b. $\alpha\ge 1$, $\beta\ge 1$. In this case, it follows from above discussion that $f(x;\alpha,\beta)$ is unimodal with mode at x_0. Alternatively, when $\alpha\ge 1$, $\beta\ge 1$, it follows from the concavvity of the logarithm function that $\log f(x;\alpha,\beta)$ is concave. In this case, it i clear that $f(0;\alpha,\beta)=f(1;\alpha,\beta)=0$ (*top right graph* in Figure 1.5).

c. $\alpha\le 1$, $\beta\ge 1$. Using similar methods, it can be determinated that the density (1.38) is decreasing. If $1<\beta<2$, the density is first concave, then convex; if $\beta\ge 2$, then the density is convex (*bottom left graph* in Figure 1.5).

d. $\alpha\ge 1$, $\beta\le 1$. As in the previous case, it can be determined that the density (1.38) is increasing. If $1<\alpha<2$, then the density is first concave then convex; if $\alpha\ge 2$, the density is convex (*bottom right graph* in Figure 1.5).

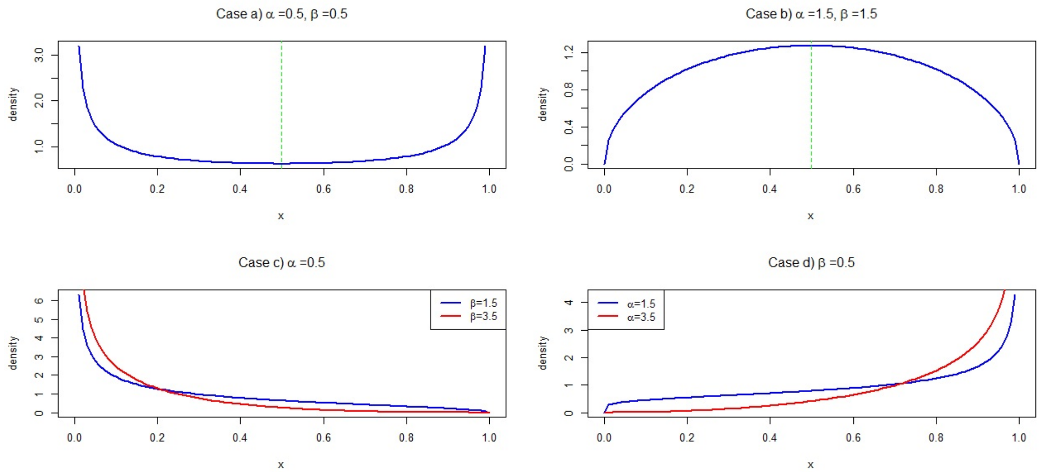

Figure 1.5: Densities of Beta distributions for various values of the parameters α and β

Moments of the Beta Distribution

Let X be a random variable with density $f(\cdot\,;\alpha,\beta)$. Because it is known that the density of the Beta distribution integrates to 1, it is easy to determine that $f(\cdot\,;\alpha,\beta)$ has rth moment

$$\mathrm{E}[X^r] = \mu_r = \frac{\mathrm{B}(\alpha+r,\beta)}{\mathrm{B}(\alpha,\beta)}.$$

Thus,

$$\mathrm{E}[X] = \mu_1 = \frac{\alpha}{\alpha+\beta}, \quad \mathrm{E}[X^2] = \mu_2 = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}, \quad \text{and}$$
$$\mathrm{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

By fixing $\mu_1 = \frac{\alpha}{\alpha+\beta}$ and letting $\alpha \longrightarrow 0$, $\beta \longrightarrow 0$ together, the upper bound for the variance is aproached; by letting $\alpha \longrightarrow \infty$, $\beta \longrightarrow \infty$, the lower bound for the variance is aproached. Thus, the Beta distribution allow for all pairs (μ_1, σ^2) possible under the constraint that $P\{0 \leq X \leq 1\} = 1$ apart from those values on the boundary which are achieved only by discrete distributions.

1.1.5 Pareto Distribution

The Pareto distribution was proposed by Vilfredo Pareto (1987). It is a continuous probability distribution with two parameters. It arises because Pareto was interested in describing income distributions. On the other hand in Cirilo (1979) we see that the distribution arises from the economic models. Thus the Pareto distribution has aplications in economy, sociology, geophysics, etc.
Arnold (1983) discussed in detail the several variations of the Pareto distribution. In Arnold's terminology, if X is a random variable with Pareto type I distribution, then the survival function is

$$\overline{F}(x;\lambda,\alpha) = \begin{cases} 1, & x \leq \lambda \\ \left(\frac{x}{\lambda}\right)^{-\alpha}, & x > \lambda, \end{cases} \tag{1.40}$$

where α is positive parameter. The Pareto type I distribution is characterized by a scale parameter (λ) and a shape parameter (α), which is known as the tail index.
From (1.40) the cdf of X random variable Pareto with parameters λ and α is

$$F(x;\lambda,\alpha) = \begin{cases} 0, & x \leq \lambda \\ 1-\left(\frac{x}{\lambda}\right)^{-\alpha}, & x > \lambda \end{cases} \tag{1.41}$$

By differentiation of (1.41) the density of the Pareto type I distribution is obtained.

$$f(x;\lambda,\alpha) = \begin{cases} 0, & x \leq \lambda \\ \frac{\alpha}{x}\left(\frac{x}{\lambda}\right)^{-\alpha}, & x > \lambda \end{cases} \tag{1.42}$$

Moments of Pareto type I distribution

Let X be a random variable whit Pareto type I distribution, for $r < \alpha$ it is defined the rth moment by

$$E[X^r] = \mu_r = \frac{\alpha\lambda^r}{\alpha - r},$$

for $r \geq \alpha$ the rth moment does not exist. Thus

$$\mathrm{E}[X] = \frac{\alpha\lambda}{\alpha-1} \qquad \mathrm{E}[X^2] = \frac{\alpha\lambda^2}{\alpha-2}$$

$$\mathrm{Var}(X) = \frac{\alpha}{\alpha-2}\left(\frac{\lambda}{\alpha-1}\right)^2$$

Various types of Pareto Distribution

In much of the literature the standard Pareto distribution is permitted to have an additional location parameter. We will call the resulting distribution the Pareto type II distribution. A convenient parameterization is provided by

$$\overline{F}(x;\mu,\sigma,\alpha)=\left[1+\left(\frac{x-\mu}{\sigma}\right)\right]^{-\alpha},\quad x>\mu, \tag{1.43}$$

where μ, the location parameter, is real, σ, the scale parameter is positive, and α the shape parameter is positive.
Truncated versions of the Pareto type II distribution have found application in simulation contexts, and have been proposed as plausible income distribution models (e.g., Arnold and Austin, 1987). Such distributions are of the form

$$F(x;\mu,\sigma,\alpha)=\begin{cases}0, & x\leq\mu\\ \frac{1-\left(1+\frac{x-\mu}{\sigma}\right)^{-\alpha}}{1-\left(1+\frac{\tau-\mu}{\sigma}\right)^{-\alpha}}, & \mu<x<\tau\\ 1, & x\geq\tau\end{cases}$$

where $-\infty<\mu<\tau<\infty$, $\sigma>0$, and $\alpha>0$.

An alternative variation on the Pareto theme, which provides tail behavior similar to (1.43), is provided by the Pareto type III family with survival function

$$\overline{F}(x;\mu,\sigma,\gamma)=\left[1+\left(\frac{x-\mu}{\sigma}\right)^{\frac{1}{\gamma}}\right]^{-1},\quad x>\mu, \tag{1.44}$$

where μ is real, σ is positive, and γ is positive. We will call γ the inequality parameter. If $\mu=0$ and $\gamma\leq 1$, then γ turns out to be precisely the Gini index of inequality.
We can generalize even more, introducing the shape parameter and the inequality parameter, finally obtaining the Pareto type IV distribution, whose survival function is

$$\overline{F}(x;\mu,\sigma,\gamma,\alpha)=\left[1+\left(\frac{x-\mu}{\sigma}\right)^{\frac{1}{\gamma}}\right]^{-\alpha},\quad x>\mu, \tag{1.45}$$

where μ (location) is real, σ (scale) is positive, γ (inequality) is positive, and α (shape) is positive. In this case γ is not identifiable with the Gini index.,

because this will only occur when $\alpha = 1$ and $\mu = 0$.
The family of Pareto type IV distributions provides a convenient vehicle for computing distributional results for the three more specialized families. They may be identified as special cases of the Pareto type IV family as follows:

$$
\begin{aligned}
P(I)(\sigma, \alpha) &= P(IV)(\sigma, \sigma, 1, \alpha) \\
P(II)(\mu, \sigma, \alpha) &= P(IV)(\mu, \sigma, 1, \alpha) \\
P(III)(\mu, \sigma, \gamma) &= P(IV)(\mu, \sigma, \gamma, 1),
\end{aligned}
$$

where, $P(I)$ is the Pareto type I distribution, $P(II)$ is the Pareto type II distribution, $P(III)$ is the Pareto type III distribution and $P(IV)$ is the Pareto type IV distribution.

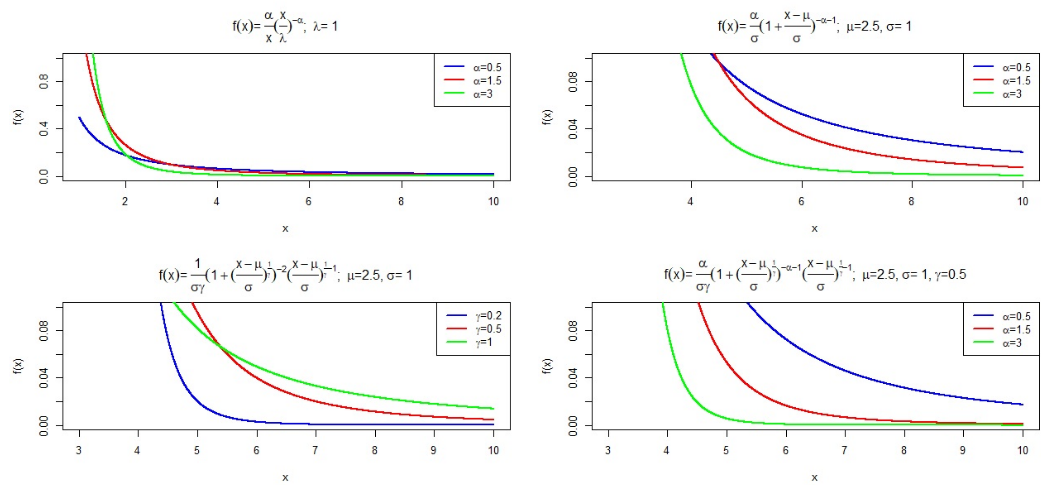

Figure 1.6: Densities of the various types of the Pareto distribution. Pareto type I distribution (*top left graph*), Pareto type II distribution (*top right graph*), Pareto type III distribution (*bottom left graph*) and Pareto type IV distribution (*bottom right graph*)

Feller (1971, p. 49) defined a Pareto distribution in a different way. Let Y be a random variable with Beta distribution whit parameters γ_1 and γ_2, i.e.,

$$f(y;\gamma_1,\gamma_2) = \frac{y^{\gamma_1-1}(1-y)^{\gamma_2-1}}{\mathrm{B}(\gamma_1,\gamma_2)}, \quad 0<y<1.$$

We define $U = Y^{-1} - 1$. Then U has what Feller called a Pareto distribution. By considering a linear form of a power of said random variable, we obtain a more general family, which we call the Feller-Pareto family. More precisely, if $Y \sim B(\gamma_1,\gamma_2)$ and if for μ real, $\sigma > 0$ and $\gamma > 0$, we define

$$W = \mu + \sigma\left(Y^{-1}-1\right)^{\gamma}, \tag{1.46}$$

then W has a Feller-Pareto distribution. We see how this family represents a generalization of the Pareto type IV family (and thus of the other Pareto families).
If $Y \sim B(\gamma_1,\gamma_2)$ and $U = Y^{-1} - 1$, then U has a Beta dsitribution of the second kind, equivalently a scaled F distribution with density

$$f(u;\gamma_1,\gamma_2) = \frac{u^{\gamma_2-1}(1+u)^{-\gamma_1-\gamma_2}}{\mathrm{B}(\gamma_1,\gamma_2)}, \quad u>0.$$

If W is as in (1.46), then

$$\begin{aligned} P(W>w) &= P(\mu+\sigma U^{\gamma} > w) \\ &= P\left(U > \left(\frac{w-\mu}{\sigma}\right)^{\frac{1}{\gamma}}\right) \end{aligned} \tag{1.47}$$

Thus we need a closed form for the survival function of U. This will be simplest when either γ_1 or γ_2 is 1. Let us consider the case when $\gamma_2 = 1$, so that

$$\overline{F}(u;\gamma_1) = (1+u)^{-\gamma_1}, \quad \mu > 0. \tag{1.48}$$

Combining (1.47) and (1.48), we see that, with $\gamma_2 = 1$,

$$P(W>w) = \left[1+\left(\frac{w-\mu}{\sigma}\right)^{\frac{1}{\gamma}}\right]^{-\gamma_1}$$

Referring to (1.45), we see that the Pareto type IV ($P(IV)$) distributions are identifiable with the Feller-Pareto (FP) distributions with $\gamma_2 = 1$, i.e.,

$$P(IV)(\mu,\sigma,\gamma,\alpha) = FP(\mu,\sigma,\gamma,\alpha,1).$$

In the case, when $\gamma_1 = 1$, we may verify that

$$P(W > w) = 1 - \left[1 + \left(\frac{w - \mu}{\sigma}\right)^{-\frac{1}{\gamma}}\right]^{-\gamma_2} \tag{1.49}$$

We may be seen that the Feller-Pareto family includes distributions (such as (1.49)) whose behavior is distinctly non-Paretian in the upper tail.

Chapter 2

Estimation of the parameters of Extreme Value Distributions

Numerous techniques have been proposed for the estimation of the parameters in an extreme value model. These include, for example: graphic techniques based on versions of probability diagrams, moment-based techniques, where the moments of the model are equated with their empirical equivalents; procedures in which the parameters are estimated as specific functions of order statistics and methods based on likelihood.

2.1 Maximum Likelihood Estimation

The method of maximun likelihood, by far, the most popular techique for deriving estimators of the parameters vector $\theta = (\theta_1, \ldots, \theta_p)^T$, which consists in choosing as estimator of θ the value $\hat{\theta}(X_1, \ldots, X_n)$ that maximizes the likelihood function, given by

$$L(\theta, \mathbf{X}) = \prod_{i=1}^{n} f(x_i, \theta_1, \ldots, \theta_p), \tag{2.1}$$

where x_i are the realizations of the i.i.d. random variables $X_1, \ldots, X_n$ with pdf common f. Generally maximize the likelihood function is often difficult. Since the likelihood is a positive function and the maximums of the likelihood are the same as those of the log-likelihood, as the logarithm function is increasing, we prefer to look the maximum likelihood estimators (MLE) using the log-likelihood given by

$$l(\theta, \mathbf{X}) = \sum_{i=1}^{n} \log(f(x_i, \theta_1, \ldots, \theta_p)). \tag{2.2}$$

The maximum likelihood estimators of the parameters $\theta_1, \ldots, \theta_p$ are obtained by solving the following system of equations.

$$\begin{aligned} \frac{\partial l(\theta_1, \ldots, \theta_p, \mathbf{X})}{\partial \theta_1} &= 0 \\ &\vdots \\ \frac{\partial l(\theta_1, \ldots, \theta_p, \mathbf{X})}{\partial \theta_p} &= 0 \end{aligned} \tag{2.3}$$

2.1.1 Properties of the maximum likelihood estimators

Under regularity conditions quite general of the density function $f(x, \theta)$, the maximum likelihood estimators have the following nice properties.

1. **Consistency**
$$P(\hat{\theta}_{MLE} = \theta) \longrightarrow 1, \quad \text{when} \quad n \longrightarrow \infty.$$

2. **Asymptotic Normality**
$$\sqrt{n}(\hat{\theta}_{MLE} - \theta) \ \xrightarrow{d} \ \mathrm{N}(0, \sigma^2_{MLE}).$$

3. **Asymptotic optimality (efficient)**.
Maximum likelihood estimator has the smallest variance.

4. **Invariance Property**.
Let $\tau = g(\theta)$ be a function of θ. Let $\hat{\theta}_n$ the maximum likelihood estimator of θ. Then $\hat{\tau}_n = g(\hat{\theta}_n)$ is the maximum likelihood of τ.

2.1.2 Maximum Likelihood Estimation for the GEV distribution

The family of GEV distributions does not meet the regularity conditions required by the usual asymptotic properties, so that the maximum likelihood estimator is valid. This is because the end-points of the GEV distribution are functions of the parameter values. Smith (1985) studied this problem in detail and obtained the following results in relation to parameter γ of the GEV distribution given in the Theorem 1.1.1.

1. When $\gamma > -0.5$, maximum likelihood estimators are regular.
2. When $-1 < \gamma < -0.5$, maximum likelihood estimators are generally obtainable, but do not have the standard asymptotic properties.
3. When $\gamma < -1$, maximum likelihood estimators are unlikely to be obtainable.

The case $\gamma \leq -0.5$ corresponds a distributions with a very short bounded upper tail. This situation is rarely encountered in applications of extreme value modeling, so the theoretical limitations of the maximum likelihood approach are usually no obstacle in practice.
Let $Z_1, Z_2, \ldots, Z_n$ be independent random variables with common distribution the GEV defined in Theorem 1.1.1. The log-likelihood for the GEV parameters when $\gamma \neq 0$ is

$$l(\mu,\sigma,\gamma) = -n\log\sigma - \left(1+\frac{1}{\gamma}\right)\sum_{i=1}^{n}\log\left(1+\gamma\left(\frac{z_i-\mu}{\sigma}\right)\right) - \sum_{i=1}^{n}\left(1+\gamma\left(\frac{z_i-\mu}{\sigma}\right)\right)^{-\frac{1}{\gamma}} \tag{2.4}$$

provided that

$$1+\gamma\left(\frac{z_i-\mu}{\sigma}\right) > 0, \quad \text{for} \quad i = 1,\ldots,n \tag{2.5}$$

The case when condition (2.5) is not met, corresponds to a configuration where at least one of the observed points falls beyond the distribution's end point, the likelihood is zero and the log-likelihood is equal to $-\infty$.
The case of $\gamma = 0$ requires the Gumbel limit of the GEV distribution, which leaves the log-likelihood

$$l(\mu,\sigma) = -n\log\sigma - \sum_{i=1}^{n}\left(\frac{z_i-\mu}{\sigma}\right) - \sum_{i=1}^{n}\exp\left\{-\left(\frac{z_i-\mu}{\sigma}\right)\right\} \tag{2.6}$$

Maximizing equations (2.4) and (2.6) with respect to the parameter vector (μ, σ, γ) leads to the maximum likelihood estimate with respect to the entire GEV family. There is no analytical solution, but for any given dataset, maximization is straightforward using standard numerical optimization algorithms. It must be guaranteed that these algorithms do not consider combinations of parameters that violate (2.5), and also the numerical difficulties that arise in

the evaluation of (2.4) in the neighborhood of $\gamma = 0$. The latter problem is solved easily by using (2.6) instead of (2.4) for the values that fall within a small neighborhood around zero.
Subject to the limitations on γ, the approximate distribution of $(\hat{\mu}, \hat{\sigma}, \hat{\gamma})$ is multivariate normal with mean (μ, σ, γ) and variance-covariance matrix equal to the inverse of the observed information matrix evaluated at the maximum likelihood estimator. Although this matrix can be obtained analytically, it is easier to use numerical differentiation techniques to evaluate the second derivatives, and standard numerical routines to determine the inverse.

2.1.3 Maximum likelihood for Weibull distribution

In the case of the Weibull model with parameters λ and k, it is not possible to calculate numerically by maximum likelihood the estimates of the parameters of the model, but we must resort to numerical methods to calculate these estimates.
The likelihood function for $X_1, \ldots, X_n$ i.i.d. random variables with Weibull common distribution is

$$L(\lambda), k) = \frac{k^n}{\lambda^{nk}} \prod_{i=1}^{n} x_i^{k-1} \exp\left\{ -\left(\frac{x_i}{\lambda}\right)^k \right\}, \tag{2.7}$$

then, the log-likelihood function is given by

$$l(\lambda, k) = n \log k - nk \log \lambda + (k-1) \sum_{i=1}^{n} \log x_i - \sum_{i=1}^{n} \left(\frac{x_i}{\lambda}\right)^k. \tag{2.8}$$

The maximum likelihood estimates are obtained by solving the equations resulting from equaling the two partial derivatives from $l(\lambda, k)$ to zero. As a result, the maximum likelihood estimator of k is obtained by solving the equation

$$\frac{\sum_{i=1}^{n} x_i^k \log x_i}{\sum_{i-1}^{n} x_i^k} - \frac{1}{k} - \frac{1}{n} \sum_{i=1}^{n} \log x_i = 0 \tag{2.9}$$

Although it is not possible to obtain an analytical solution, $\hat{k}$ can be calculated using numerical solutions, such as the Newton-Raphson method. Once the shape parameter is estimated, we obtain the estimator of the scale parameter λ by means of the expression

$$\hat{\lambda} = \left(\frac{1}{n} \sum_{i=1}^{n} x_i^{\hat{k}} \right)^{\frac{1}{\hat{k}}} \tag{2.10}$$

2.1.4 Maximun likelihood for Gamma distribution

In the case of tha Gamma distribution, we have that the likelihood function is given by

$$\mathrm{L}(\theta, k) = \frac{1}{\theta^{nk}(\Gamma(k))^n} \prod_{i=1}^{n} x_i^{k-1} \mathrm{e}^{-\frac{x_i}{\theta}}, \tag{2.11}$$

consequently the log-likelihood function is

$$\begin{aligned} l(\theta, k) &= -nk \log\theta - n\log(\Gamma(k)) + (k-1)\sum_{i=1}^{n} \log x_i - \frac{1}{\theta}\sum_{i=1}^{n} x_i \\ &= -nk\log\theta - n\log(\Gamma(k)) + n(k-1)\overline{\log x} - \frac{n}{\theta}\overline{x}, \end{aligned} \tag{2.12}$$

whose partial derivative with respect to θ is

$$\frac{\partial l}{\partial \theta} = -\frac{nk}{\theta} + \frac{n}{\theta^2}\overline{x}, \tag{2.13}$$

equaling this derivative to zero, it is easily obtained

$$\hat{\theta} = \frac{\overline{x}}{k}, \tag{2.14}$$

substituting this in (2.12) we get

$$l(\hat{\theta}, k) = -nk\log\overline{x} + nk\log k - n\log(\Gamma(k)) + n(k-1)\overline{\log x} - nk. \tag{2.15}$$

Minka (2002) describes two algorithms to maximize this function.

First algorithm.(Maximize iteratively to reduce the limit)
Because $k\log k$ is convex, we can use a linear lower limit

$$k\log k \geq (1+\log k_0)(k-k_0) + k_0\log k_0 \tag{2.16}$$

$$\begin{aligned} l(\hat{\theta}, k) \geq \quad & n(k-1)\overline{\log x} - n\log(\Gamma(k)) - nk\log\overline{x} \\ & + n(1+\log k_0)(k-k_0) + nk_0\log k_0 - nk \end{aligned} \tag{2.17}$$

The maximum is obtained by equating to zero the partial derivative with respect to k of the latter expression.

$$\begin{aligned} & n\overline{\log x} - n - n\log\overline{x} + n(1+\log k_0) - n \\ & \psi(\hat{k}) = \overline{\log x} - \log\overline{x} + \log k_0, \end{aligned} \tag{2.18}$$

where $\psi(k) = \frac{\Gamma'(k)}{\Gamma(k)}$ is the digamma function. The iteration process continues by configuring k in the current $\hat{k}$, then a new $\hat{\theta}$ is obtained by determining the inverse function of ψ. Because the log-likelihood is concave, this iteration must converge to the (unique) global maximum. Unfortunately, it can be quite slow.

Second algorithm("Generalized Newton")
This algorithm is much faster, and is obtained via "generalized Newton"(Minka, 2000). Using an approximation of the form,

$$l(\hat{\theta}, k) \approx c_0 + c_1 k + c_2 \log k, \tag{2.19}$$

the update is

$$\frac{1}{k^{new}} = \frac{1}{k} + \frac{\overline{\log x} - \log \overline{x} + \log k - \psi(k)}{k^2 \left(\frac{1}{k} - \psi'(k)\right)}, \tag{2.20}$$

which converges quickly.

2.1.5 Maximum likelihood for Beta distribution

The likelihood function for the Beta distribution is

$$\mathrm{L}(\alpha, \beta) = \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right)^n \prod_{i=1}^{n} x_i^{\alpha-1}(1-x_i)^{\beta-1}. \tag{2.21}$$

Thus, the log-likelihood function is

$$\begin{aligned} l(\alpha, \beta) = \quad & n\log(\Gamma(\alpha+\beta)) - n\log(\Gamma(\alpha)) - n\log(\Gamma(\beta)) \\ & +(\alpha-1)\textstyle\sum_{i=1}^{n} \log x_i + (\beta-1)\sum_{i=1}^{n} \log(1-x_i) \end{aligned} \tag{2.22}$$

Taking into account that the maximum likelihood estimators are determined equal to zero the partial derivatives of the log-likelihood with respect to each of the parameters, which implies solving the following system of equations.

$$\begin{aligned} \frac{\Gamma'(\alpha+\beta)}{\Gamma(\alpha+\beta)} - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + \frac{1}{n}\sum_{i=1}^{n} \log x_i = 0 \\ \frac{\Gamma'(\alpha+\beta)}{\Gamma(\alpha+\beta)} - \frac{\Gamma'(\beta)}{\Gamma(\beta)} + \frac{1}{n}\sum_{i=1}^{n} \log(1-x_i) = 0 \end{aligned} \tag{2.23}$$

But there is no closed solution for this system of equations. Owen (2008) proposes using the Newton-Raphson method to estimate $\theta = (\alpha, \beta)$ iteratively, through the following procedure.

$$\hat{\theta}_{i+1} = \hat{\theta}_i - \mathbf{G^{-1}g}, \tag{2.24}$$

where $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ and $\mathbf{g} = (g_1, g_2)^T$ is the vector of normal equations, with

$$\begin{aligned} g_1 &= \psi(\alpha) - \psi(\alpha+\beta) - \frac{1}{n}\sum_{i=1}^{n} \log x_i \\ g_2 &= \psi(\beta) - \psi(\alpha+\beta) - \frac{1}{n}\sum_{i=1}^{n} \log(1-x_i), \end{aligned} \tag{2.25}$$

and $\mathbf{G}$ is the matrix of second derivatives

$$\mathbf{G} = \begin{bmatrix} \frac{dg_1}{d\alpha} & \frac{dg_1}{d\beta} \\ \frac{dg_2}{d\alpha} & \frac{dg_2}{d\beta} \end{bmatrix}, \tag{2.26}$$

where

$$\begin{aligned} \frac{dg_1}{d\alpha} &= \psi'(\alpha) - \psi'(\alpha+\beta) \\ \frac{dg_1}{d\beta} &= \frac{dg_2}{d\alpha} = -\psi'(\alpha+\beta) \\ \frac{dg_2}{d\beta} &= \psi'(\beta) - \psi'(\alpha+\beta) \end{aligned} \tag{2.27}$$

and $\psi(\cdot)$ and $\psi'(\cdot)$ are the digamma and trigamma functions respectively, defined as

$$\psi(\cdot) = \frac{\Gamma'(\cdot)}{\Gamma(\cdot)}, \qquad \psi'(\cdot) = \frac{\Gamma''(\cdot)}{\Gamma(\cdot)} - \left(\frac{\Gamma'(\cdot)}{\Gamma(\cdot)}\right)^2 \tag{2.28}$$

The Newton-Raphson algorithm converges, as the estimates of α and β change by less than a tolerated amount with each successive iteration,to $\hat{\alpha}$ and $\hat{\beta}$.

2.1.6 Maximum likelihood for Pareto distribution

The simplest and most researched situation is where we habe a sample of size n from a classic Pareto distribution (i.e., Pareto type I distribution), whose survival distribution is given in equation (1.40) of the Chapter 1. Arnold (1983) denote by $X = (X_1, \dots, X_n)$ the sample and $X_{i:n}$, $(i = 1, \dots, n)$, the corresponding order statistics. The corresponding likelihood function is of the form

$$L(\lambda), \alpha) = \alpha^n \lambda^{n\alpha} \prod_{i=1}^{n} x_i^{-(\alpha+1)}, \quad \alpha > 0, \quad X_{1:n} > \lambda. \tag{2.29}$$

To obtain the maximum likelihood estimate $(\hat{\lambda}, \hat{\alpha})$ of (λ, α), we consider the log-likelihood function.

$$l(\lambda, \alpha) = n \log \alpha + n\alpha \log \lambda - (\alpha + 1) \sum_{i=1}^{n} \log x_i. \tag{2.30}$$

For a fixed α, the above expression is maximized when $\lambda = X_{1:n}$. Differentation with respect to α can be used to obtain the maximizing value of α. Thus we find

$$\hat{\lambda} = X_{1:n}, \tag{2.31}$$

$$\hat{\alpha} = \left[\frac{1}{n} \sum_{i=1}^{n} \log \left(\frac{x_i}{X_{1:n}} \right) \right]^{-1}. \tag{2.32}$$

The maximum likelihood estimates are consistent (in fact, strogly consistent) (cf. Muniruzzaman, 1957, and Quandt, 1966). This follows from evidently $X_{1:n} \xrightarrow{a.s.} \lambda$ and for the strong law of large numbers,

$$\frac{1}{n} \sum_{i=1}^{n} \log x_i \xrightarrow{a.s.} \mathrm{E}[\log X_1] = \log \lambda + \frac{1}{\alpha}.$$

Then, we can derive the exact form of the joint density of the maximum likehood estimates $(\hat{\lambda}, \hat{\alpha})$ (Muniruzzaman, 1957).
Considering that if $X_1 \sim P(IV)(\mu, \sigma, \lambda, \alpha) \Longrightarrow X_{1:n} \sim P(IV)(\mu, \sigma, \lambda, n\alpha)$ we have that $\hat{\sigma} \sim P(I)(\lambda, n\alpha)$. Thus it is required the conditional distribution of $\hat{\alpha}$ given $\hat{\lambda}$. Malik (1970) showed that in fact $\hat{\alpha}$ and $\hat{\lambda}$ are independent (a result implicit in Muniruzzaman, 1957; see also Baxter, 1980). Defining $Y_i = l \log X_i$, follow that $Y_i - \log \lambda \sim \Gamma(1, \alpha^{-1})$ (i.e., an exponencial distribution). The independence of $\hat{\lambda}$ y $\hat{\alpha}$ follows readily from the independence of exponential spacings affirmed in Theorem 2.1.1.

Theorem 2.1.1. *Let $X_1, \ldots, X_n$ be sample of size n from the exponential distribution ($\overline{F}(x) = \mathrm{e}^{-\mu x},\ x > 0$) with corresponding order statistics $X_{i:n},\ i = 1, \ldots, n$ and spacings $Y_{i:n} = X_{i:n} - X_{i-1:n},\ i = 1, \ldots, n$ ($X_{0:n} = 0$, by definition). The spacings are independent exponential random variables. In fact, the random variables $\{(n - i + 1)Y_{i:n},\ i = 1, \ldots, n\}$ are independent identically distributed random variables with common exponential survival function $\overline{F}(x) = \mathrm{e}^{-\mu x},\ x > 0$.*

For completeness we write the actual densities

$$f_{\hat{\lambda}}(u) = n\alpha\lambda^{n\alpha}u^{-(n\alpha+1)}, \quad u > \lambda, \tag{2.33}$$

$$f_{\hat{\alpha}}(v) = \frac{(\alpha n)^{n-1}}{\Gamma(n-1)v^n}\mathrm{e}^{-\frac{n\alpha}{v}}, \quad v > 0. \tag{2.34}$$

The reciprocal of $\hat{\alpha}$ has a Gamma distribution, i.e. $\hat{\alpha}^{-1} \sim \Gamma(n-1, (\alpha n)^{-1})$ or equivalently

$$\frac{2\alpha n}{\hat{\alpha}} \sim \chi^2_{2n-2} \tag{2.35}$$

Note also that

$$2\alpha n \log\left(\frac{\hat{\lambda}}{\lambda}\right) \sim \chi^2_2. \tag{2.36}$$

The above observations allow to easily determine the mean and the variance of maximum likelihood estimates:

$$\mathrm{E}[\hat{\lambda}] = \lambda\left(1 - \frac{1}{n\alpha}\right)^{-1}, \qquad \mathrm{Var}(\hat{\lambda}) = \lambda^2 n\alpha(n\alpha - 1)^{-2}(n\alpha - 2)^{-1}. \tag{2.37}$$

$$\mathrm{E}[\hat{\alpha}] = \alpha n(n-2)^{-1}, \qquad \mathrm{Var}(\hat{\alpha}) = (\alpha n)^2(n-2)^{-2}(n-3)^{-1}. \tag{2.38}$$

The biased character of the maximum likelihood estimates is evident from the above. The estimate $\hat{\lambda}$ always overestimates λ.

2.2 Method of Moments

According to Casella and Berger (2002). The method of moment is, perhaps, the oldest method of finding point estimators, dating back at least to Karl Pearson in the late 1800s. It has the virtue of being quite simple to use and almost always yield some sort of estimate. In many cases, unfortunately, this method yields estimators that may be improved upon. However, it is good

place to start when other methods prove to be intractable.
The method of moments consists of equaling a certain number of theoretical moments of the distribution of the population with the corresponding sampling moments. From here one or more equations are obtained which, when solved, allow estimating the unknown parameters of the population distribution.

Advantages

1. The method of moments is fairly simple and yields consistent estimators (under very weak assumptions).

2. The method of moments can assist in finding maximum likelihood estimates.

3. When estimating other structural parameters (e.g., parameters of a utility function, instead of parameters of a known probability distribution), appropriate probability distributions may not be known, and moment-based estimates may be preferred to maximum likelihood estimation.

Disadvantages

1. Estimates obtained by the moments method are often biased.

2. In some cases, infrequent with large samples but not so infrequent with small samples, the estimates given by the moments method are outside the parametric space.

3. Estimates by the moments method are not necessarily sufficient statistics.

Remark 2.2.1. *So far our analysis has focused on obtaining estimators considering original data that do not really need to have properties that ensure the existence of finite moments. For this reason, Chapter 3 will study finite moment estimators of such "complicated" distributions.*

Table 2.1: Estimators of moments for Gamma Distribution and Beta Distribution.

Distribution	Moments	Equations	Estimators
Gamma	$\mathrm{E}[X] = k\theta$	$k\theta = \overline{x}$	$\hat{\theta} = \frac{s^2}{\overline{x}}$
	$\mathrm{Var}(X) = k\theta^2$	$k\theta^2 = s^2$	$\hat{k} = \frac{\overline{x}^2}{s^2}$
Beta	$\mathrm{E}[X] = \frac{\alpha}{\alpha+\beta}$	$\frac{\alpha}{\alpha+\beta} = \overline{x}$	$\hat{\alpha} = \overline{x}\left(\frac{\overline{x}(1-\overline{x})}{s^2} - 1\right)$
	$\mathrm{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = s^2$	$\hat{\beta} = (1-\overline{x})\left(\frac{\overline{x}(1-\overline{x})}{s^2} - 1\right)$

Chapter 3

Estimators of the shape parameter of the Pareto Distribution

There is a considerable number of publications on estimators for the extreme value index (EVI). The various studies are inspired by diverse (equivalent) conditions that ensure the convergence of the distribution of the maximum sample $X_{n:n} = \max\{X_1, X_2, \ldots, X_n\}$ to a limit distribution of the extreme value type mentioned in Chapter 1. Requiring convergence makes sense because otherwise there would be no hope that we have something significant to say about the extreme quantiles in the limit or even beyond the range of the sample. At this point it is important to emphasize that the extreme value index determines the behavior of the extreme values, in this way the estimation of the EVI becomes an important problem to reliably estimate high quantiles. In the case of heavy-tailed, i.e. distributions with EVI greater than zero, Feuerverger and Hall (1999) and Beirlant et al. (1999) proposed an exponential regression model for log-spacings of order statistics based on the theory of slow variation with remainder. This model proved successful in EVI estimation and optimal sample fraction selection (see Matthys and Beirlant, 2000a) as well as in extreme quantile estimation (see Matthys and Beirlant, 2000b).

3.1 Hill estimator

Considering the Pareto-type distribution, whose survival function

$$\overline{F}(x) = x^{-\alpha}C(x),$$

where $\alpha > 0$ and C is a slowly varying function. The parameter $\gamma = \frac{1}{\alpha}$ is known as the extreme value index or tail index, which helps to indicate the size and frequency of extreme events under F. The Hill estimator enjoys a high degree of popularity because it has some good theoretical properties despite presenting some serious drawbacks. Beirlant et al. (2004) present four ways to introduce the Hill estimator, we will focus on the first form, that is, from the point of view of quantiles.
The quantile view, which comes from the quantile plots of the Pareto-type distributions.

a. Pareto-type distributions satisfy

$$\frac{\log Q(1-p)}{-\log p} \longrightarrow \gamma, \quad \text{when} \quad p \longrightarrow 0. \tag{3.1}$$

From this, it follows that a Pareto quantile plot is a exponential quantile plot based on the log-transformed data, it is finally linear with slope γ near the largest observations.

b. The slope of an ultimately linear exponential quantile plot can be estimated by the mean excess values of the type

$$\tilde{E}_{k,n} = \frac{\frac{1}{k}\sum_{j=1}^{k}\left(X_{n-j+1,n} - X_{n-k,n}\right)}{\frac{1}{k}\sum_{j=1}^{k}\left(\log\left(\frac{j}{n+1}\right) + \log\left(\frac{k+1}{n+1}\right)\right)}. \tag{3.2}$$

The combination of these two observations leads to the mean excess value of the log-transformed data, known as the Hill estimator (Hill, 1975):

$$\hat{\gamma}_k = H_{k,n} = \frac{1}{k}\sum_{j=1}^{k}\log\left(\frac{X_{n-j+1,n}}{X_{n-k,n}}\right), \tag{3.3}$$

where $X_{1,n}, X_{2,n}, \ldots, X_{n,n}$ are the order statistics and $X_{n-k,n}$ is the k-th threshold. The Hill estimator is based on a fact that for a sample $Y_1, Y_2, \ldots, Y_n$

from the strict Pareto distribution whit suport in $[1, \infty)$ and survival function $\overline{F}(x) = x^{-\alpha}$,

$$\frac{1}{\hat{\alpha}} = \frac{1}{n} \sum_{i=1}^{n} \log Y_i, \tag{3.4}$$

is the maximum likelihood estimator of $\frac{1}{\alpha}$.

3.1.1 Properties of the Hill estimator

1. Mason (1982) showed that $H_{k,n}$ is a **consistent estimator** for γ, as $k, n \longrightarrow \infty$, $\frac{k}{n} \longrightarrow 0$, whatever the slowly varying function l_F o l_U may be. This is true even for data that is weakly dependent (Hsing, 1991) or in the case of a linear process (Resnick and Stărică, 1995).

2. **Asymptotic normality** of $H_{k,n}$ was discussed among others in Hall (1982), Deheuvels et al. (1998) and references therein, and De Haan and Resnick (1998) and references therein.

3. In Drees (1998) and Beirlant et al. (2006), **variance and rate optimality** of the Hill estimator was derived for large submodels of the Pareto-type model.

3.1.2 Problems of the Hill estimator

1. We obtain an estimator different from γ in each election of k. The estimates $H_{k,n}$ against k are usually plotted, producing the Hill plot:

   $$\{(k, H_{k,n}) : 1 \leq k \leq n-1\}$$

   However, these plots typically are far from being constant, which makes it difficult to use the estimator in practice without further guideline on how to choose the value k. In the Figure 3.1, we illustrate this by means of a simulation from a Pareto type I distribution I and Pareto type II distribution. Resnick and Stărică (1997) (see also Drees et al., 2000), proposed a plot

   $$\{(\log k, H_{k,n}) : 1 \leq k \leq n-1\},$$

   which focuses the graphics in the appropriate area as we observed in the Figure 3.2.

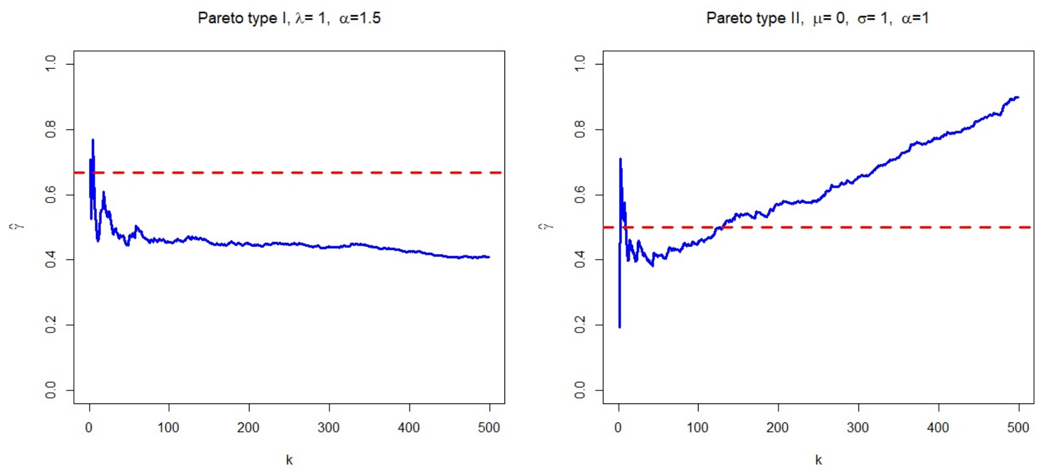

Figure 3.1: Plot of $H_{k,n} = \hat{\gamma} = \frac{1}{\hat{\alpha}}$ vs. k for simulated datasets of size $n = 500$

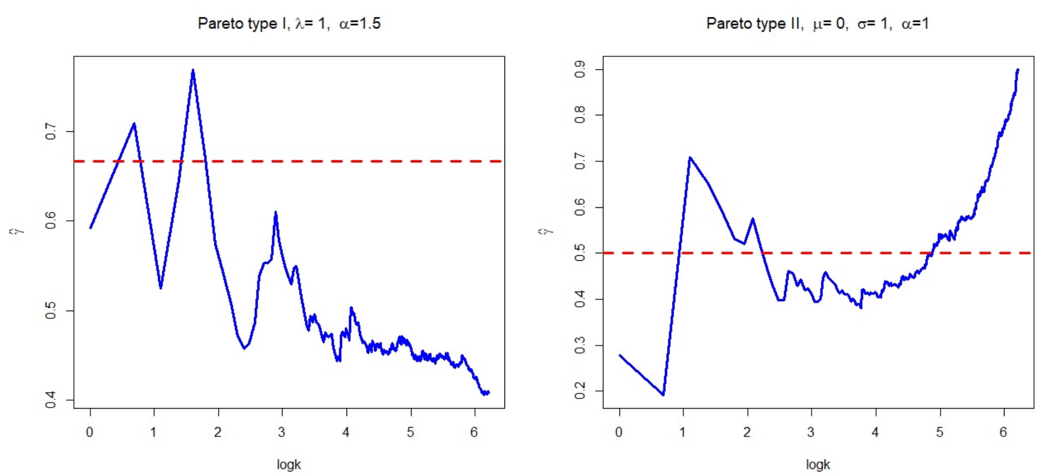

Figure 3.2: Plot of $H_{k,n} = \hat{\gamma} = \frac{1}{\hat{\alpha}}$ vs. $\log k$ for simulated datasets of size $n = 500$

2. In many instances, a severe bias can appear, which happens when the effect of the slowly varying part in the model disappears slowly in the Pareto quantile plot. That is, from the point of view of probability, by assuming that the relative excesses above the threshold follow a strict Pareto distribution is sometimes optimistic. A large bias leads to poor coverage probabilities of confidence intervals. In many practical cases, systematic overestimation or underestimation has to be avoided. The Figure 3.3 for a simulation experiment from a $|T4|$ distribution with $\gamma = \frac{1}{4}$ where only for the smallest values of k the median of the Hill estimator touches the correct value.

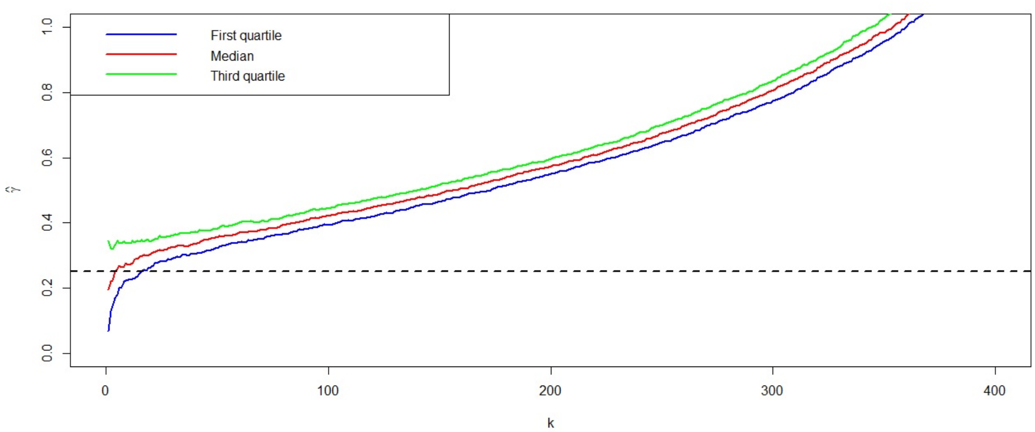

Figure 3.3: Median and quantiles of the Hill estimates $H_{k,n}$ as a function of k, obtained from 200 samples of size 500 from a $|T4|$ distribution

3. The Hill estimator, as well as many other common estimators based on logarithmically transformed data, is not invariant with respect to changes in the data. As mentioned by several authors, the inappropriate use of the Hill estimator along with a data change can also lead to systematic errors.

3.1.3 Hill plot of α

We previously mentioned the Hill plot $\{(k, H_{k,n}) : 1 \leq k \leq n-1\}$, now for the purpose of analyzing the use of the Hill estimator in the practice, we will make the Hill plot of α, mentioned in Resnick (2007).

$$\{(k, H_{k,n}^{-1}) : 1 \leq k \leq n-1\}, \tag{3.5}$$

and we hope that the graph looks stable for picking out a value of α. But, this happens wonderfully sometimes and sometimes the plots are not very revealing. For example, in Figure 3.4 we see one of many Hill Horror Plots. In this figure the left plot is for a simulation of size 20000 from an α-stable symmetric distribution with $\alpha = 1.5$ and the right plot is a simulated i.i.d. sample of size 20000, from the Gamma distribution with scale and shape parameters $\lambda = 1$ and $\alpha = 1$ respectively. Both plots exhibit extreme bias and do not lead to the correct answer. The problem is that the Hill estimator is designed for the Pareto distribution.

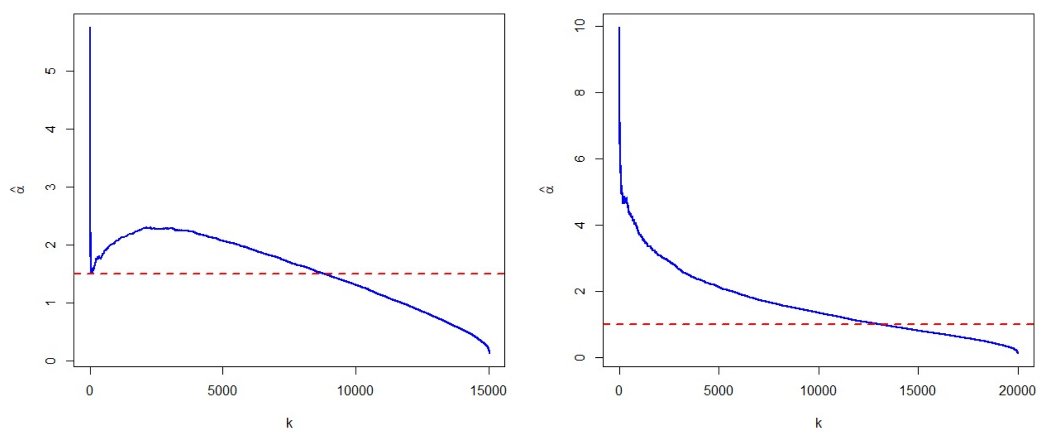

Figure 3.4: Hill Horror Plots of a α-stable symmetric distribution with $\alpha = 1.5$ (*left chart*) and a Gamma distribution with $\lambda = 1$ and $\alpha = 1$ (*right chart*)

According to the above, the use of the Hill estimator presents a series of difficulties which we can summarize as follows:

1. If we want to obtain a point estimate of α from a graph. What value of k should we use?

2. The graph can show considerable volatility and/or the true answer may be hidden in the graph.

3. The Hill estimator has optimality properties only when the underlying distribution is close to Pareto. In the case that the distribution is far from to Pareto there may be terrible error.

4. The Hill estimator is not location invariant. A shift in the location does not theoretically affect the tail index but may throw the Hill estimate way off. This means that the Hill estimator can be surprisingly sensitive to changes in location.

3.1.4 SmooHill plot

Resnick and Stărică (1997) depeloped a smoothing technique yielding. the smooHill plot. The technique involves selecting an integer r (usually $r = 2$ or $r = 3$), then to define

$$smooH_{k,n} = \frac{1}{(r-1)k} \sum_{j=k+1}^{rk} H_{j,n} \tag{3.6}$$

Resnick (2007) shows that the estimator defined above is also consistent.

Changing the scale
As an alternative to the Hill plot of α, sometimes is useful to diplay the information provided by the Hill or smooHill estimator as

$$\{(\theta, H_{[n^\theta],n}), 0 \leq \theta \leq \ 1\}, \tag{3.7}$$

where $[\bullet]$ is the smallest integer greater than or equal to $\bullet \geq 0$. Resnick (2007) called this graphic the *alternative Hill plot* (altHill). This alternative is sometimes revealing since the initial order statisctics are show more crearly and cover a bigger portion of the displayed space in a small neighborhood of α. The Figure 3.5 compares several Hill plot for 8000 observations from a α-stable distribution with $\alpha = 1.5$.

3.1.5 Last advances on the Hill estimator

Gomes and Stehlík (2014) provide developments on the Hill and related estimators of the extreme value index and discuss their properties, such as: mean square error, efficiency and robustness.

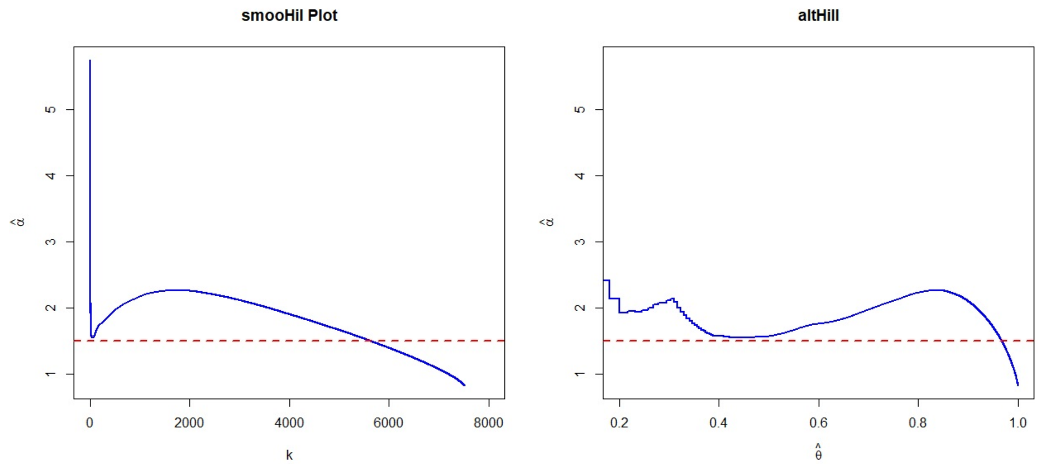

Figure 3.5: SmooHill plot (*left chart*) and altHill plot (*right chart*) of α-stable distribution with $\alpha = 1.5$ for a simulated dataset of size $n = 8000$

3.2 Moment estimator

De Haan and Ferreira (2006) depeloped an estimator similar to the Hill estimador, but that can be used for $\gamma \in \mathbb{R}$, not only for $\gamma > 0$.
Considering the function U as the left-continuous inverse of $\frac{1}{1-F}$, the authors assume that $U(\infty) > 0$, which can be achieved applying a change in the data, but it should be considered that this change influence in the behavior of the estimator.
The following Lemma corresponding to Lemma 3.5.1 of De Haan and Ferreira (2006), allows to introduce an estimator for a general γ.

Lemma 3.2.1. *Let $X_1, X_2, \ldots$ be i.i.d. random variables with distribution function F, assuming $F \in D(G_\gamma)$ ($D(G_\gamma)$ is the atraction domain of G_γ), $x^* = U(\infty) > 0$, i.e. for $x > 0$*

$$\lim_{t \longrightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}. \tag{3.8}$$

Define $j = 1, 2,$

$$M_n^{(j)} := \frac{1}{k} \sum_{i=0}^{k-1} (\log X_{n-i,n} - \log X_{n-k,n})^j. \tag{3.9}$$

Then for $k = k(n) \longrightarrow \infty$, $\frac{k}{n} \longrightarrow 0$, $n \longrightarrow \infty$,

$$\frac{M_n^{(j)}}{\left(\frac{a\left(\frac{n}{k}\right)}{U\left(\frac{n}{k}\right)}\right)^j} \xrightarrow{P} \prod_{i=1}^{j} \frac{i}{1 - i\gamma_-}, \tag{3.10}$$

with $\gamma_- = \min(0, \gamma)$.

From Lemma 3.2.1 it follows that the Hill estimator converges to zero for $\gamma \leq 0$; therefore this estimator is not informative in this range. However, under its conditions,

$$\frac{\left(M_n^{(1)}\right)^2}{M_n^{(2)}} \xrightarrow{P} \frac{1 - 2\gamma_-}{2(1 - \gamma_-)}. \tag{3.11}$$

Besides

$$\hat{\gamma}_H \xrightarrow{P} \gamma_+, \tag{3.12}$$

where $\gamma_+ := \max(0, \gamma)$.

This leads to the combination of the Hill estimator and the statistic in (3.11):

$$\hat{\gamma}_M := M_n^{(1)} + 1 - \frac{1}{2}\left(1 - \frac{\left(M_n^{(1)}\right)^2}{M_n^{(2)}}\right)^{-1}, \tag{3.13}$$

for which the authors proved the following theorem.

Theorem 3.2.1. *Let* $X_1, X_2, \dots$ *be i.i.d. random variables with distribution function* F. *Suppose* $F \in D(G_\gamma)$ *and* $x^* > 0$. *Then*

$$\hat{\gamma}_M \xrightarrow{P} \gamma, \tag{3.14}$$

for $\gamma \in \mathbb{R}$ *provided the sequence* k *is intermediate, i.e.,* $k = k(n) \longrightarrow \infty$, $\frac{k}{n} \longrightarrow 0$ *as* $n \longrightarrow \infty$, *i.e.,* $\hat{\gamma}_M$ *is consistent for* γ.

The estimator $\hat{\gamma}_M$ is called **moment estimator**, because the left-hand side of (3.10) converges to $\frac{\mathrm{E}[(Y^{\gamma_-}-1)^j]}{\gamma_-^j}$, $j = 1, 2$, which is the jth moment of the limiting generalized Pareto distribution.

3.3 QQ plot and QQ estimator

3.3.1 QQ plot

The QQ plot provides an infomal, but convenient way to test the hypothesis of goodness of fit.The method is empirical and requires to consider the order statistics notation indexed from smallest to largest and vice versa, i.e.,

$$X_{(1)} \geq X_{(2)} \geq \ldots, X_{(n)} \text{ order statistics indexing from largest to smallest,}$$
$$X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n} \text{ order statistics indexing fron smallest to largest,}$$

so that $X_{(i)} = X_{n-i+1}$.

For the empirical distribution of the sample $X_1, X_2, \ldots, X_n$ we have

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{[X_i \leq x]} = \frac{1}{n}\sum_{i=1}^{n} \epsilon_{X_i}((-\infty, x]),$$

where ϵ_{X_i} is the probabilite measure consisting of all mass at X_i.

The QQ plot consists in graphing the theoretical quantiles of F vs. the sample quantiles.

$$\left\{\left(F^{\leftarrow}\left(\frac{i}{n+1}\right), \hat{F}_n^{\leftarrow}\left(\frac{i}{n+1}\right)\right), 1 \leq i \leq n\right\} = \left\{\left(F^{\leftarrow}\left(\frac{i}{n+1}\right), X_{i,n}\right), 1 \leq i \leq n\right\}. \tag{3.15}$$

If the points should fall on or near the straight line $\{(x, x),\ x > 0\}$, then the null hypothesis is true, i.e. there is no evidence against the null hypothesis.

Remark 3.3.1. *Note that*

$$\hat{F}_n^{\leftarrow}\left(\frac{i}{n+1}\right) = X_{i:n}, \quad 1 \leq i \leq n.$$

Because, $\hat{F}_n^{\leftarrow}(q)$ *is left-continuous and*

$$\hat{F}_n^{\leftarrow}(q) = X_{i:n} \text{ for } \frac{i-1}{n} < q \leq \frac{i}{n}.$$

Since $\hat{F}_n^{\leftarrow}\left(\frac{i}{n}\right) = X_{i:n}$*, we need only to show that*

$$\frac{i}{n} - \frac{i}{n+1} = \frac{i}{(n+1)n} \quad \text{and} \quad \frac{i}{n+1} < 1,$$

then

$$\frac{i}{n} - \frac{i}{n+1} < \frac{1}{n}.$$

3.3.2 Adaptation to the heavy-tailed case

Suppose that the null hypothesis is

$$P(X > x) = \left(\frac{x}{x_l}\right)^{-\alpha}, \quad x > x_l, \tag{3.16}$$

with $x_l > 0$ and $\alpha > 0$. Assumption (3.16) means that $\frac{X}{x_l}$ is Pareto with left endpoint 1 and shape parameter α, and for $y > 0$,

$$P\left(\alpha \log \frac{X}{x_l} > y\right) = P\left(\frac{X}{x_l} > \mathrm{e}^{\frac{y}{\alpha}}\right) = \mathrm{e}^{-y}.$$

So $\alpha \log \frac{X}{x_l}$ is exponetial with parameter 1. Therefore,

$$\begin{aligned} P(\log X > y) &= P\left(\frac{\log X - \log x_l}{\alpha^{-1}} > \frac{y - \log x_l}{\alpha^{-1}}\right) \\ &= P\left(\alpha \log \frac{X}{x_l} > \frac{y - \log x_l}{\alpha^{-1}}\right) \\ &= \mathrm{e}^{-((y - \log x_l)/\alpha^{-1})}. \end{aligned}$$

If $W_1(x) = 1 - \mathrm{e}^{-x}$, $x > 0$, then

$$P(\log X > y) = \overline{W}_1\left(\frac{y - \log x_l}{\alpha^{-1}}\right),$$

which is a location-scale family with location parameter $\mu = \log x_l$ and scale parameter $\sigma = \alpha^{-1}$. The quantiles of this family are obtained in the following way

$$W_1(x) = 1 - \mathrm{e}^{-x} = q,$$

then

$$\begin{aligned} e^{-x} &= 1 - W_1(x) \\ -x &= \log(1 - W_1(x)) \\ x &= -\log(1-q). \end{aligned}$$

Therefore we must plot

$$\left\{ \left(-\log\left(1 - \frac{i}{n+1}\right), \log X_{i,n} \right), 1 \le i \le n \right\}, \tag{3.17}$$

If the null hypothesis (3.16) is correct, or at least aproximately correct, the plot should be roughly linear with intercept $\log x_l$ and slope α^{-1}.

3.3.3 QQ estimator

The regular variation assumption (3.16) means that we should not draw the line of least squares across all pairs of points in (3.17), but only through pairs corresponding to k upper-order statistics.
For n points in the plane $\{(x_i, y_i), 1 \le i \le n\}$, a standerd texbook indicates that the slope (SL) through these points is

$$\mathrm{SL}\left(\{(x_i, y_i), 1 \le i \le n\}\right) = \frac{\overline{S}_{xy} - \overline{x}\,\overline{y}}{\overline{S}_{xx} - \overline{x}^2},$$

where

$$\overline{S}_{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i, \qquad \overline{S}_{xx} = \frac{1}{n}\sum_{i=1}^{n} x_i^2.$$

If $X_1, X_2, \ldots, X_n$ is a random sample from the Pareto distribution

$$1 - F_\alpha(x) = x^{-\alpha}, \quad \alpha > 0, \quad x \ge 1,$$

then, the slope of least squares line through of the points

$$x_i = -\log\left(1 - \frac{i}{n+1}\right), \qquad y_i = \log X_{i,n},$$

gives an estimator of α^{-1}, called QQ estimator, whose expression is

$$\widehat{\alpha^{-1}} = \frac{\sum_{i=1}^{n}\left(-\log\left(\frac{1}{n+1}\right)\right)\left\{n \log X_{n-i+1,n} - \sum_{i=1}^{n} \log X_{n-j+1,n}\right\}}{n\sum_{i=1}^{n}\left(-\log\left(\frac{i}{n+1}\right)\right)^2 - \left[\sum_{i=1}^{n}\left(-\log\left(\frac{i}{n+1}\right)\right)\right]^2}. \tag{3.18}$$

With the regular variation assumption (3.16), we can modify (3.18) and to define the QQ estimator based in the k upper-order statistics.

$$\widehat{\alpha^{-1}} = \widehat{\alpha^{-1}}_{k,n} = \mathrm{SL}\left(\left\{\left(-\log\left(1-\frac{i}{k+1}\right), \log\left(\frac{X_{n-k+i,n}}{X_{n-k,n}}\right)\right), 1 \le i \le k\right\}\right). \tag{3.19}$$

Making some simplifications and considering the following property of the SL function.

$$\mathrm{SL}\left(\{(x_i, y_i), 1 \le i \le n\}\right) = \mathrm{SL}\left(\{(x_i + 1, y_i + b), 1 \le i \le n\}\right), \tag{3.20}$$

for any real numbers a, b. We get

$$\widehat{\alpha^{-1}} = \mathrm{SL}\left(\left\{\left(-\log\left(1-\frac{i}{k+1}\right), \log X_{n-k+i,n}\right), 1 \le i \le k\right\}\right). \tag{3.21}$$

Consistency of the QQ estimator

In view of (3.18), (3.20), and (3.3), we can rewrite the QQ estimator $\widehat{\alpha^{-1}}$ as

$$\begin{aligned}\widehat{\alpha^{-1}} &= \frac{\frac{1}{k}\sum_{i=1}^{k}\left(-\log\left(1-\frac{i}{k+1}\right)\right)\log\left(\frac{X_{n-k+i,n}}{X_{n-k,n}}\right) - \frac{1}{k}\sum_{i=1}^{k}\left(-\log\left(1-\frac{i}{k+1}\right)\right)H_{k,n}}{\frac{1}{k}\sum_{i=1}^{k}\left(-\log\left(1-\frac{i}{k+1}\right)\right)^2 - \left(\frac{1}{k}\sum_{i=1}^{k}\left(-\log\left(1-\frac{i}{k+1}\right)\right)\right)^2} \\ &= \frac{\frac{1}{k}\sum_{i=1}^{k}\left(-\log\left(\frac{i}{k+1}\right)\right)\log\left(\frac{X_{(i)}}{X_{(k+1)}}\right) - \frac{1}{k}\sum_{i=1}^{k}\left(-\log\left(\frac{i}{k+1}\right)\right)H_{k,n}}{\frac{1}{k}\sum_{i=1}^{k}\left(-\log\left(\frac{i}{k+1}\right)\right)^2 - \left(\frac{1}{k}\sum_{i=1}^{k}\left(-\log\left(\frac{i}{k+1}\right)\right)\right)^2}\end{aligned} \tag{3.22}$$

The following theorem corresponding to the Theorem 4.3 in Resnick (2007) ensures the weak consistency of the QQ estimator.

Theorem 3.3.1. *Suppose $X_1, \ldots, X_n$ are a random sample from F, a distribution with regularly varying tail satisfying*

$$P(X > x) = 1 - F(x) = \overline{F}(x) = x^{-\alpha}L(x), \quad x > 0. \tag{3.23}$$

Then the QQ estimator $\widehat{\alpha^{-1}}$ given in (3.22) is weakly consistent for $\frac{1}{\alpha}$:

$$\widehat{\alpha^{-1}} \xrightarrow{P} \alpha^{-1}$$

as $n \longrightarrow \infty$, $k = k(n) \longrightarrow \infty$ in such a way that $\frac{k}{n} \longrightarrow 0$.

3.4 t-Hill Estimator

3.4.1 Transformed score function

Pearson and Filon (1898), Edgeworth (1908a, and 1908b), and Fisher (1925) developed the score function. The score is the gradient with respect to parameter θ of the logarithm of the likelihood function $S(\theta, X) = \frac{\partial}{\partial\theta}\, logL(\theta, X)$. The variance of the score is the Fisher information $\mathcal{I}(\theta) = E_\theta[S^2]$.
Fabian (2001, 2007, 2010, and 2016) introduced the transformed score function (or t-score) of distribution, which reflect the main features of continuous probability distributions and enabling an introduction of their new relevant characteristics.

Definition 3.4.1. *Let $\eta : \mathcal{X} \longrightarrow \mathbb{R}$ be a strictly increasing smooth mapping. The* ***t-score*** *of distribution F with interval support $\mathcal{X} \subseteq \mathbb{R}$ and the (almost surely) two-times differentiable density $f(x)$ is*

$$T_F(x) = -\frac{1}{f(x)}\frac{d}{dx}\left[\frac{1}{\eta'(x)}f(x)\right], \tag{3.24}$$

where $\eta'(x) = \frac{d\eta(x)}{dx}$ is the Jacobian of the transformation. Supposing that the equation

$$T_F(x) = 0, \tag{3.25}$$

has unique solution x^*, then x^* is called the ***t-score mean*** *and the function*

$$S_F(x) = \eta'(x^*)T_F(x), \tag{3.26}$$

the ***generalized t-score function*** *of distribution F.*

Example 3.4.1. *Considering the Beta distribution*

$$f(x) = \frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}, \quad x \in (0,1), \quad \alpha, \beta > 0, \tag{3.27}$$

and using the Johnson mapping (see Fabián, 2016) with derivative

$$\eta'(x) = \frac{1}{x(1-x)}.$$

Then

$$
\begin{aligned}
T_F(x) &= -\frac{1}{\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}}\frac{d}{dx}\left[\frac{1}{\frac{1}{x(1-x)}}\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}\right] \\
&= -\frac{1}{x^{\alpha-1}(1-x)^{\beta-1}}\frac{d}{dx}\left[x^{\alpha}(1-x)^{\beta}\right] \\
&= -\frac{1}{x^{\alpha-1}(1-x)^{\beta-1}}\left[\alpha x^{\alpha-1}(1-x)^{\beta}-\beta x^{\alpha}(1-x)^{\beta-1}\right] \\
&= -\frac{1}{x^{\alpha-1}(1-x)^{\beta-1}}x^{\alpha-1}(1-x)^{\beta-1}[\alpha(1-x)-\beta x] \\
&= -\frac{1}{x^{\alpha-1}(1-x)^{\beta-1}}x^{\alpha-1}(1-x)^{\beta-1}[-(\alpha+\beta)x+\alpha] \\
&= (\alpha+\beta)x-\alpha, \qquad (3.28)
\end{aligned}
$$

and the ***t-score mean*** *is* $x^* = \frac{\alpha}{\alpha+\beta}$, *and the* ***generalized t-score function*** *is*

$$
\begin{aligned}
S_F(x) &= \frac{\alpha+\beta}{\alpha}\left[(\alpha+\beta)x-\alpha\right] \\
&= (\alpha+\beta)\left[\frac{\alpha+\beta}{\alpha}x-1\right] \\
&= (\alpha+\beta)\left[\frac{x}{x^*}-1\right]
\end{aligned}
$$

Example 3.4.2. *For the Pareto type I distribution, with* $\lambda = 1$, *we have that their function the density is*

$$
f(x) = \alpha x^{-(\alpha+1)}, \quad x > 1. \qquad (3.29)
$$

As this function does not have a visible Jacobian of any transformation, Jordanova et al. (2016) propose to use $\eta(x) = \log(\log x)$ *or* $\eta(x) = \log(x-1)$.

a. For $\eta(x) = \log(\log x) \Longrightarrow \eta'(x) = \frac{1}{x\log x}$*, then*

$$\begin{aligned}
T_F(x) &= -\frac{1}{\alpha x^{-(\alpha+1)}}\frac{d}{dx}\left[\frac{1}{\frac{1}{x\log x}}\alpha x^{-(\alpha+1)}\right] \\
&= -\frac{1}{x^{-(\alpha+1)}}\frac{d}{dx}\left[x^{-\alpha}\log x\right] \\
&= -\frac{1}{x^{-(\alpha+1)}}\left[-\alpha x^{-(\alpha+1)}\log x + x^{-(\alpha+1)}\right] \\
&= -\frac{1}{x^{-(\alpha+1)}}x^{-(\alpha+1)}(-\alpha\log x + 1) \\
&= \alpha\log x - 1, \quad (\textit{score of Fisher for } \alpha)
\end{aligned} \tag{3.30}$$

b. In the case that $\eta(x) = \log(x-1) \Longrightarrow \eta'(x) = \frac{1}{x-1}$*, consequently*

$$\begin{aligned}
T_F(x) &= -\frac{1}{\alpha x^{-(\alpha+1)}}\frac{d}{dx}\left[\frac{1}{\frac{1}{x-1}}\alpha x^{-(\alpha+1)}\right] \\
&= -\frac{1}{x^{-(\alpha+1)}}\frac{d}{dx}\left[x^{-\alpha} - x^{-(\alpha+1)}\right] \\
&= -\frac{1}{x^{-(\alpha+1)}}\left[-\alpha x^{-(\alpha+1)} + (\alpha+1)x^{-(\alpha+2)}\right] \\
&= -\frac{1}{x^{-(\alpha+1)}}x^{-(\alpha+1)}(-\alpha + (\alpha+1)x^{-1}) \\
&= \alpha - \frac{\alpha+1}{x},
\end{aligned} \tag{3.31}$$

Then, $x^* = \frac{\alpha+1}{\alpha}$ *and* $T_F(x) = \alpha\left(1 - \frac{x^*}{x}\right)$

3.4.2 Score moments

Fabián (2010) defines for any $k \in \mathbb{N}$ the **t-score moments**

$$\mathrm{E}T^k = \int_{\mathcal{X}} T_F^k(x) f(x,\theta)dx \tag{3.32}$$

Particularly $\mathrm{E}T = 0$ and $\mathrm{E}T^2$ is the Fisher information for x^* called t-score mean, which can be considered as a measure of central tendency of distributions (Fabián, 2004 , and Fabián, 2008). Jordanova et al. (2016) introduce for any $k \in \mathbb{N}$ the score moments

$$M_k(\theta) = \mathrm{E}S_F^k(\theta) = \int_{\mathcal{X}} S_F^k(x,\theta) f(x,\theta)dx, \tag{3.33}$$

which exists only if f satisfies the usual regularity requirements. It is easy to see that $M_1 = 0$ and the value $M_2 = \mathrm{E}S^2$ of distribution with (transformed) location parameter is the Fisher information for (transformed) location parameter. Accordingly, $\mathrm{E}S^2$ the Fisher information for the t-score mean, and the reciprocal value

$$\omega^2 = \frac{1}{\mathrm{E}S^2}, \tag{3.34}$$

the t-score variance, which seems to be a measure of variability (dispersion) of the distribution, even in cases where the usual variance does not exist (see Fabián, 2007).

Example 3.4.3. *From the example 3.4.2b) for the Pareto type I distribution, we have that* $T_F(x) = \alpha\left(1 - \frac{x^*}{x}\right)$ *with* $x^* = \frac{\alpha+1}{\alpha}$*, therefore*

1.

$$\begin{aligned}
\mathrm{E}T_F^k &= \int_1^\infty \alpha^2 \left(1 - \frac{x^*}{x}\right)^2 \alpha x^{-(\alpha+1)} dx \\
&= \alpha^3 \int_1^\infty \left(1 - 2\frac{x^*}{x} + \frac{(x^*)^2}{x^2}\right) x^{-(\alpha+1)} dx \\
&= \alpha^3 \int_1^\infty \left(x^{-(\alpha+1)} - 2x^* x^{-(\alpha+2)} + (x^*)^2 x^{-(\alpha+3)}\right) dx \\
&= \alpha^3 \left(-\frac{x^{-\alpha}}{\alpha} + 2x^* \frac{x^{-(\alpha+1)}}{\alpha+1} - (x^*)^2 \frac{x^{-(\alpha+2)}}{\alpha+2}\right)\Bigg|_1^\infty \\
&= \alpha^3 \left(-\frac{1}{\alpha} + \frac{(\alpha+1)^2}{\alpha^2(\alpha+2)}\right) \\
&= \alpha^3 \left(\frac{-\alpha(\alpha+2) + (\alpha+1)^2}{\alpha^2(\alpha+2)}\right) \\
&= \frac{\alpha}{\alpha+2}
\end{aligned} \tag{3.35}$$

2.

$$\begin{aligned}
\mathrm{E}S_F^2 &= \int_1^\infty \left(\eta'(x^*)\right)^2 T_F^2(x) f(x)dx \\
&= \left(\eta'(x^*)\right)^2 \int_1^\infty T_F^2(x) f(x)dx \\
&= \frac{1}{(x^*-1)^2}\mathrm{E}T_F^2 \\
&= \frac{1}{\left(\frac{\alpha+1}{\alpha}-1\right)^2}\frac{\alpha}{\alpha+2} \\
&= \frac{\alpha^3}{\alpha+2}
\end{aligned} \tag{3.36}$$

3.

$$\omega_F^2 = \frac{\alpha+2}{\alpha^3} \tag{3.37}$$

Definition 3.4.2. *(see Fabián, 2001, and Fabián, 2010) Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with common distribution F. Assuming that F belongs to the model family $\{F_\theta, \theta \in \Theta\}$, $\Theta \subseteq \mathbb{R}^m$. The parametric version the (3.33) yields the generalized moment estimation equations for θ,*

$$\hat{\theta}_{SM}: \; \frac{1}{n}\sum_{i=1}^{n} S_F^k(x_i\theta) = \mathrm{E}S_F^k(\theta), \quad k = 1, \ldots, m. \tag{3.38}$$

Remark 3.4.1. *$\hat{\theta}_{SM}$ is consistent and asymptotically normal.*

Remark 3.4.2. *The score moment estimators take the assumed form of the distribution. However, since data are involved in the estimation only through $S_F(x, \theta)$ and transformed scores for heavy-tailed distributions appeared to be bounded, the transformed score moment estimates of all parameters are protected against outliers in cases of heavy-tailed distributions.*

Remark 3.4.3. *Since, the t-scores and generailed t-score function differ only in a constant factor, we can use the t-scores instead of the generalized t-score function in the score moment equations (3.38).*

Example 3.4.4. *Considering the log-gamma distribution, whose density function is given by*

$$f(x) = \frac{\alpha^c}{\Gamma(c)}\frac{(\log x)^{c-1}}{x^{\alpha+1}}, \quad x > 1. \tag{3.39}$$

1. *Derive the t-score.*
   The density function (3.39) can be written as follows

$$f(x =)\frac{\alpha^c}{\Gamma(c)}\frac{(\log x)^c}{x^\alpha}\frac{1}{x\log x}, \tag{3.40}$$

then, the mapping is $\eta(x) = \log(\log x)$, *with* $\eta'(x) = \frac{1}{x\log x}$. *Replacing* $f(x)$ *and* $\eta'(x)$ *in the definition of the t-score, we have*

$$\begin{aligned} T_F(x) &= \frac{\Gamma(c)}{\alpha^c}\frac{x^{\alpha+1}}{(\log x)^{c-1}}\frac{d}{dx}\left[x\log x\frac{\alpha^c}{\Gamma(c)}\frac{(\log x)^c}{x^\alpha}\frac{1}{x\log x}\right] \\ &= \frac{x^{\alpha+1}}{(\log x)^{c-1}}\frac{d}{dx}\left[\frac{(\log x)^c}{x^\alpha}\right] \\ &= \frac{x^{\alpha+1}}{(\log x)^{c-1}}\left[c\frac{(\log x)^{c-1}}{x^{\alpha+1}} - \alpha\frac{(\log x)^c}{x^{\alpha+1}}\right] \\ &= c - \alpha\log x \end{aligned} \tag{3.41}$$

2. *Derive the t-score the moments for* $k = 2$

$$\begin{aligned} \mathrm{E}[T_F^2] &= \int_1^\infty (c-\alpha\log x)^2 f(x)dx \\ &= c^2\int_1^\infty f(x)dx - 2\alpha c\int_1^\infty (\log x)f(x)dx + \alpha^2\int_1^\infty (\log x)^2 f(x)dx \\ &= c^2 - 2c\int_1^\infty \frac{\alpha^{c+1}}{\Gamma(c)}\frac{(\log x)^c}{x^{\alpha+1}}dx + \int_1^\infty \frac{\alpha^{c+2}}{\Gamma(c)}\frac{(\log x)^{c+1}}{x^{\alpha+1}}dx \end{aligned}$$

On the other hand we have

$$\begin{aligned} \Gamma(c+1) &= c\Gamma(c) \Longrightarrow \Gamma(c) = \frac{\Gamma(c+1)}{c}, \\ \Gamma(c+2) &= \Gamma((c+1)+1) \\ &= (c+1)\Gamma(c+1) \\ &= (c+1)c\Gamma(c) \Longrightarrow \Gamma(c) = \frac{\Gamma(c+2)}{c^2+c}, \end{aligned}$$

then

$$\mathrm{E}[T_F^2] = c^2 - 2c^2\int_1^\infty \underbrace{\frac{\alpha^{c+1}}{\Gamma(c+1)}\frac{(\log x)^c}{x^{\alpha+1}}}_{(I)}dx + (c^2+c)\int_1^\infty \underbrace{\frac{\alpha^{c+2}}{\Gamma(c+2)}\frac{(\log x)^{c+1}}{x^{\alpha+1}}}_{(II)}dx \tag{3.42}$$

In equation (3.42), we have that I is the density function of a log-gamma distribution with parameters $c+1$ and α, and II is the density function of a log-gamma distribution with parameters $c+2$ and α. Thus,

$$\mathrm{E}[T_F^2] = c \tag{3.43}$$

3. *Solve the t-score moment equations*

$$\sum_{i=1}^{n}(c - \alpha \log x_i) = 0 \tag{3.44}$$

$$\frac{1}{n}\sum_{i=1}^{n}(c - \alpha \log x_i)^2 = c \tag{3.45}$$

By (3.44) we have that

$$c = \frac{\alpha}{n}\sum_{i=1}^{n}\log x_i, \tag{3.46}$$

after developing the square of the binomial involved in (3.45) and replacing (3.46), we obtain

$$\begin{aligned}\frac{\alpha^2}{n^2}\left(\sum_{i=1}^{n}\log x_i\right)^2 - 2\frac{\alpha^2}{n^2}\left(\sum_{i=1}^{n}\log x_i\right)^2 + \frac{\alpha^2}{n}\sum_{i=1}^{n}(\log x_i)^2 &= \frac{\alpha}{n}\sum_{i=1}^{n}\log x_i \\ -\frac{\alpha^2}{n^2}\left(\sum_{i=1}^{n}\log x_i\right)^2 + \frac{\alpha^2}{n}\sum_{i=1}^{n}(\log x_i)^2 &= \frac{\alpha}{n}\sum_{i=1}^{n}\log x_i.\end{aligned} \tag{3.47}$$

From this last equation we obtain $\hat{\alpha}$ and by consequence $\hat{c}$

$$\hat{\alpha} = \frac{s_1}{s_2 - s_1^2}, \qquad \hat{c} = \frac{s_1^2}{s_2 - s_1^2}, \tag{3.48}$$

where

$$s_1 = \frac{1}{n}\sum_{i=1}^{n}\log x_i, \qquad s_2 = \frac{1}{n}\sum_{i=1}^{n}(\log x_i)^2.$$

Example 3.4.5. *For the Pareto type I distribution with* $\lambda = 1$*, in the example 3.4.2b we determine that the t-score is* $T_F(x) = \alpha\left(1 - \frac{x^*}{x}\right)$ *where* $x^* = \frac{\alpha+1}{\alpha}$*. Also as* $\mathrm{E}T_F = 0$*, then*

$$\sum_{i=1}^{n} T(x_i, \alpha) = 0$$
$$\sum_{i=1}^{n} \left(\alpha - \frac{\alpha+1}{x_i}\right) = 0$$
$$\sum_{i=1}^{n} \left[\alpha\left(1 - \frac{1}{x_i}\right) - \frac{1}{x_i}\right] = 0$$
$$\alpha\left(n - \sum_{i=1}^{n} \frac{1}{x_i}\right) - \sum_{i=1}^{n} \frac{1}{x_i} = 0$$
$$\alpha\left(\frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} - 1\right) - 1 = 0$$
$$\hat{\alpha} = \frac{1}{\frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} - 1}$$

Stehlík et al. (2010) notice that the t-score estimator has the form

$$\hat{\alpha} = \frac{1}{\hat{x}^* - 1}, \tag{3.49}$$

where $\hat{x}^* = \overline{x}_H$*, with* $\overline{x}_H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$*.*

3.4.3 t-Hill estimator for Pareto distribution

From equation (3.49) Stehlík et al. (2012) suggest to introduce the **t-Hill estimator** as a variant of the Hill estimator, given by

$$\hat{\gamma}_k = \frac{1}{\hat{\alpha}_k} = H^*_{k,n} = \left(\frac{1}{k}\sum_{j=1}^{k} \frac{X_{n-k,n}}{X_{n-j+1,n}}\right)^{-1} - 1, \quad k = 1, 2, \ldots, n-1, \tag{3.50}$$

where $X_{1,n}, X_{2,n}, \ldots, X_{n,n}$ are the order estatists and $X_{n-k,n}$ is the kth threshold.

Remark 3.4.4. *Since, t-Hill estimator is based in the harmonic mean, which is a generalized t-Hill estimator, it is expected that to be resistant to large observations to a certain extent so that it could yield more realistic values than the ordinary Hill estimator.*

Remark 3.4.5. *Brilhante et al. (2013) and Beran et al. (2014) have made other generalizations of the Hill estimator. In particular, Beran et al. (2014) have studied the trade off between efficiency and robustness.*

In the following theorem Stehlík et al. (2012) provide the consistency of the t-Hill estimator for Pareto distribution.

Theorem 3.4.1. *t-Hill estimator for Pareto distribution is consistent.*

Jordanova et al. (2016) introduce the definition of the new estimator for $\gamma = \frac{1}{\alpha}$, considering the notation of order statistics and

$$M_n^{(j)} = \frac{1}{k}\sum_{i=1}^{k}\left(\log\left(\frac{X_{n-i+1,n}}{X_{n-k,n}}\right)\right)^j, \quad j = 1, 2, \dots. \tag{3.51}$$

Hence the **t-lgHill** estimator of $\gamma = \frac{1}{\alpha}$ is of the form

$$H_{k,n}^L = \frac{M_n^{(2)} - \left(M_n^{(1)}\right)^2}{M_n^{(1)}}. \tag{3.52}$$

3.4.4 Some asymptotic results

Jordanova et al. (2016) analyze some asymptotic characteristics of t-Hill estimator and t-lgHill estimator, which we present below.

t-Hill and t-lgHill for the case of moving average sequence

Suppose at least one of the real numbers c_j, $j = 0, 1, \dots$ is positive and there exists $\delta \in (0, 1)$, $\delta < \alpha$, such that

$$\sum_{j=0}^{\infty} |c_j|^{\delta} < \infty. \tag{3.53}$$

Consider the moving average sequence

$$X_n = \sum_{j=0}^{n} c_j Z_{n-j}, \quad -\infty < n < \infty, \tag{3.54}$$

where Z_i, $-\infty < i < \infty$ are non-negative i.i.d. innovations with distribution function G, such that $\overline{G} \in RV_{-\alpha}$, $\alpha > 0$. Considering the estimator t-Hill as previously introduced and the point measure

$$\mu_{X,k,n}(\cdot\,) := \frac{1}{k}\sum_{i=1}^{n} \varepsilon \left\{ \frac{X_i}{b\left(\frac{n}{k}\right)} \in (\cdot\,) \right\},$$

as a random element in the space $\mathbb{E}^+$ of positive Radon measures on $(0, \infty]$ endowed with the vague topology. Here

$$b(t) := F^{\leftarrow}\left(1 - \frac{1}{t}\right) = \left(\frac{1}{\overline{F}}\right)^{\leftarrow}(t), \tag{3.55}$$

is the tail function of F. Let

$$k \longrightarrow \infty, \quad \frac{k}{n} \longrightarrow 0. \tag{3.56}$$

By Proposition 3.3 of Resnick and Stărică (1995)

$$\frac{1}{k}\sum_{i=1}^{n} I_{\left[\frac{X_i}{b\left(\frac{n}{k}\right)} > 0\right]} \varepsilon \left\{ \frac{X_i}{b\left(\frac{n}{k}\right)} \in (\cdot\,) \right\} \Rightarrow \mu, \quad n \longrightarrow \infty,$$

in $\mathbb{E}^+$ where $\mu : \sigma((0, \infty]) \longrightarrow [0, \infty)$ and $\mu(x, \infty] = x^{-\alpha}$, $x > 0$.

However, for all $n \in \mathbb{N}$, $k = 1, 2, \ldots, n$ and $A \in \mathbb{E}^+$

$$\frac{1}{k}\sum_{i=1}^{n} I_{\left[\frac{X_i}{b\left(\frac{n}{k}\right)} > 0\right]} \varepsilon \left\{ \frac{X_i}{b\left(\frac{n}{k}\right)} \in (A) \right\} = \mu_{X,k,n}(A), \tag{3.57}$$

in distribution, therefore

$$\mu_{X,k,n} \Rightarrow \mu, \quad n \longrightarrow \infty, \tag{3.58}$$

By Proposition 2.1 in Resnick and Stărică (1995), having an intermediate sequence,

$$\frac{X_{(n-k,n)}}{b\left(\frac{n}{k}\right)} \xrightarrow{P} 1 \tag{3.59}$$

and by Proposition 2.2 in Resnick and Stărică (1995)

$$\hat{\mu}_{X,k,n}(\cdot) := \frac{1}{k}\sum_{i=1}^{n} \varepsilon\left\{\frac{X_i}{X_{(n-k,n)}} \in (\cdot)\right\} \xrightarrow{P} \mu(\cdot), \quad n \longrightarrow \infty, \tag{3.60}$$

in $\mathbb{E}^{+}$.

Using the following Potter's inequality for distribution functions with regularly varying tails. If a function H is regularly varying with exponent $\gamma \in \mathbb{R}$, then for $\varepsilon > 0$ there will exist $t_0(\varepsilon)$ such that for $t > t_0(\varepsilon)$ and $x \geq 1$,

$$(1-\varepsilon)x^{\gamma-\varepsilon} \leq \frac{H(tx)}{H(t)} \leq (1+\varepsilon)x^{\gamma+\varepsilon}. \tag{3.61}$$

Its proof can be found in De Haan L. (1970).

Lemma 3.4.1. *Let $X_1, X_2, \ldots$ be random variables with distribution function F such that*

$$\overline{F}(x) \in RV_{-\alpha} \tag{3.62}$$

Assume that $k \longrightarrow \infty$, $\frac{k}{n} \longrightarrow 0$ and $\mu_{X,k,n} \Rightarrow \mu$ as $n \longrightarrow \infty$. Define

$$\Phi_n = \frac{1}{k}\sum_{i=1}^{k} \varphi\left(\frac{X_{(n-i+1,n)}}{X_{(n-k,n)}}\right), \tag{3.63}$$

where φ is a differentiable function on $(1,+\infty)$ such that $x^{-\alpha}\varphi(x) \longrightarrow \infty$ as $x \longrightarrow \infty$ and $\int_1^{\infty} x^{-\alpha+\delta_1}|\varphi'(x)|dx < +\infty$, for all $t > 1$, $0 < \delta_1 < \alpha$. Then Φ_n is a weakly consistent estimator of $\varphi(1) + \int_1^{+\infty} x^{-\alpha}\varphi'(x)dx$.

As a consequence of Lemma 3.4.1, considering $\varphi(x) = \frac{1}{x}$ yields that $H^{*}_{k,n}$, defined in (3.50) is a weakly consistent estimator for $\gamma = \frac{1}{\alpha}$.

Proposition 3.4.1. *Suppose at least one of the real numbers c_j, $j = 0, 1, \ldots$ is positive and there exists $\delta < \alpha$ such that condition (3.53) is satisfied. Consider the moving average sequence (3.54) where Z_i, $-\infty < i < \infty$, are non-negative i.i.d. innovations with distribution function G, such that $\overline{G} \in \mathrm{RV}_{-\alpha}$, $\alpha > 0$. If (3.56), then $H^{*}_{k,n}$ is a weakly consistent estimator for $\frac{1}{\alpha}$.*

Proposition 3.4.2. *Suppose at least one of the real numbers $c_j, j = 0, 1, \ldots$ is positive and there exists $\delta \in (0,1)$, $\delta < \alpha$, such that the condition (3.53) is satisfied. Consider the moving average sequence (3.54), where $Z_i, -\infty < i < \infty$ are non-negative i.i.d. innovations with distribution function G, such that $\overline{G} \in RV_{-\alpha}$, $\alpha > 0$. If $k \longrightarrow \infty$, $\frac{k}{n} \longrightarrow 0$, then*

i. M_n^2 *is a weakly consistent estimator for* $\frac{2}{\alpha^2}$.

ii. *t-lgHill estimator* $H_{k,n}^L$ *is a weakly consistent estimator for* $\gamma = \frac{1}{\alpha}$.

Asymptotic normality of t-lgHill

Definition 3.4.3. *If the tail function (or survival function) of a non-negative random variable X is $\overline{F} := 1 - F$ and $\overline{F} : \mathbb{R} \longrightarrow [0,1]$ satisfies that $\overline{F} \in RV_{-\alpha}$ with $\alpha > 0$. Then $\overline{F}$ is said to be of second-order regular variation with parameter $\rho \leq 0$, if there exists a function $Q(t)$ that ultimately has a constant sign with $\lim\limits_{t \longrightarrow \infty} Q(t) = 0$ and a constant $c \neq 0$ such that*

$$\lim_{t \longrightarrow \infty} \frac{\frac{\overline{F}(tx)}{\overline{F}(t)} - x^{-\alpha}}{Q(t)} = H_{\alpha,\rho}(x) = cx^{-\alpha} \int_1^x u^{\rho-1} du, \quad x > 0. \tag{3.64}$$

Then it is written $\overline{F} \in 2RV_{\alpha,\rho}$ and $Q(t)$ is referred to as the auxiliary function of $\overline{F}$.

Remark 3.4.6. *From De Haan and Stadtmüller (1996) and Geluk et al. (1997) it is known that if $H_{\alpha,\rho}(x)$ is not a multiple of $x^{-\alpha}$ then $\rho < 0$ implies that there exists a $c \neq 0$ such that*

$$H_{\alpha,\rho}(x) = cx^{-\alpha} \frac{x^\rho - 1}{\rho},$$

and $|A| \in RV_\rho$ and no other choices of ρ are consistent with $Q(t) \longrightarrow 0$.

Remark 3.4.7. *There exists many distributions which satisfy the second order RV (Regular Variation) condition. These are (see e.g Drees et al., 2000): Cauchy $\gamma = 1$, $\rho = -2$, Fréchet(1) $\gamma = 1$, $\rho = -1$, Student t(4) $\gamma = \frac{1}{4}$, $\rho = -\frac{1}{2}$, Student t(10) $\gamma = \frac{1}{10}$, $\rho = -\frac{1}{5}$ or loggamma $\gamma = \frac{1}{3}$, $\rho = 0$.*

Remark 3.4.8. *The most seminal results on this topic could be found in De Haan and Ferreira (2006).*

Theorem 3.4.2. *Let $X_1, X_2, \ldots$ be i.i.d. randon variables. Suppose that the second order condition (3.64) holds with $\alpha > 0$ and auxiliary function Q. If $\sqrt{k}Q\left(\frac{n}{k}\right) \longrightarrow 0$, as $n \longrightarrow \infty$, then,*

$$\sqrt{k}\left(\alpha H^L_{k,n} - 1\right) \xrightarrow{d} N(0,8). \tag{3.65}$$

3.4.5 Comparisons

As t-Hill estimator is based on harmonic mean, it is resistant to large observations so that it may yield more realistic values for large k.
Stehlík et al. (2012) in order to show that the t-Hill estimator is robust to deviations. The experiment consists in generating samples from the Pareto contaminated distribution

$$F_s = (1-\varepsilon)F + \varepsilon H,$$

where F and H are Pareto type I distribution with scale parameter $\lambda = 1$ and different shape parameters and $0 < \varepsilon < 1$.
Jordanova et al. (2016) in order to compare the quality of the t-Hill and t-lgHill estimators with others estimators, they calculate the percentage relative bias (RB) and the relative root-mean-square error (RRMSE) define as

$$RB = \frac{\frac{1}{m}\sum_{i=1}^{m}(\hat{\gamma}_i - \gamma)}{\beta}100,$$
$$RRMSE = \frac{\sqrt{\frac{1}{m}\sum_{i=1}^{m}(\hat{\gamma}_i - \gamma)^2}}{\gamma}100,$$

where $\hat{\gamma}_i$ is the corresponding estimator of $\gamma = \frac{1}{\alpha}$ for the ith simulated sample, $i = 1, 2, \ldots, m$.
We compare the t-Hill and t-lgHill estimators with Hill, Moments and QQ estimators in orden to show the good features of t-Hill and t-lgHill, for this we carried out a simulation experiment in which we generated 100 samples of 1000 realizations of random variables Pareto type I with parameter $\alpha = 0.8$, $\alpha = 1.4$ and $\alpha = 2$. The results for RB and RRMSE are recorded in Table 3.1.
Figure 3.6, 3.7 and 3.8 show the asymptotic consistency of t-Hill and t-logHill compared with other estimators for a simulated datasest of size $n = 500$ of a Pareto type I distribution with $\alpha = 0.8$, $\alpha = 1.4$ and $\alpha = 2$.
When comparing both t-Hill and t-lgHill with Hill we observe in the Figure 3.6

that t-Hill estimator is closer to the true value of $\gamma = \frac{1}{\alpha}$. The same happens when we compare both t-Hill as t-lgHill with the moments estimator, which is shown in Figure 3.7. Finally, comparing both t-Hill and t-lgHill with the QQ estimator, we see in the graph central of Figure 3.8 that in the case of $\alpha = 1.4$ t-lgHill is closer to the true value of the parameter than the other estimators. On the other hand, in Table 3.1, we observe that t-Hill has a lower percentage of relative bias (RB) than the moments estimator and the QQ estimator in the case of the three values of α. For $\alpha = 1.4$, the percentage relative root-mean-square error (RRMSE) of t-Hill estimator is less that of the t-lgHill, Moments and QQ estimators.

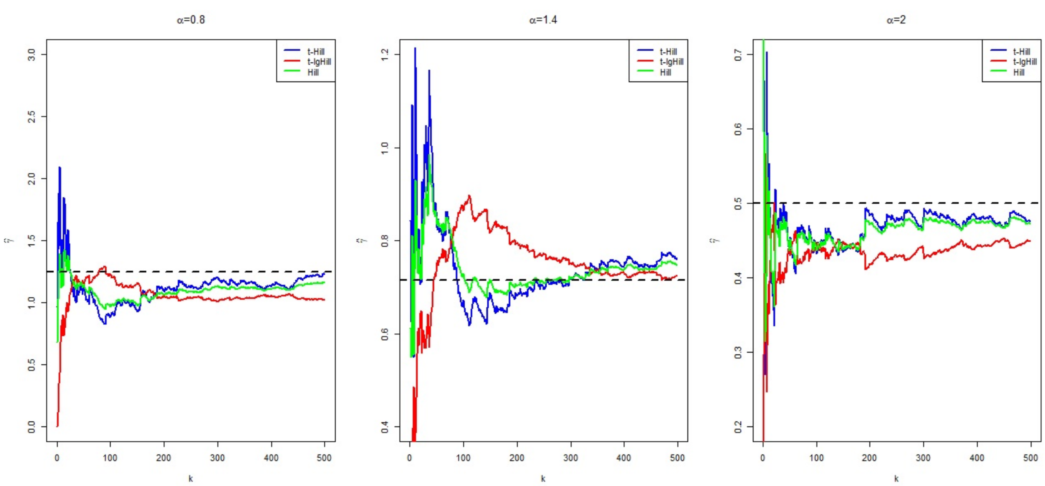

Figure 3.6: Comparation t-Hill and t-lgHill with Hill estimators for Extreme Values Index (EVI) $\gamma = \frac{1}{\alpha}$

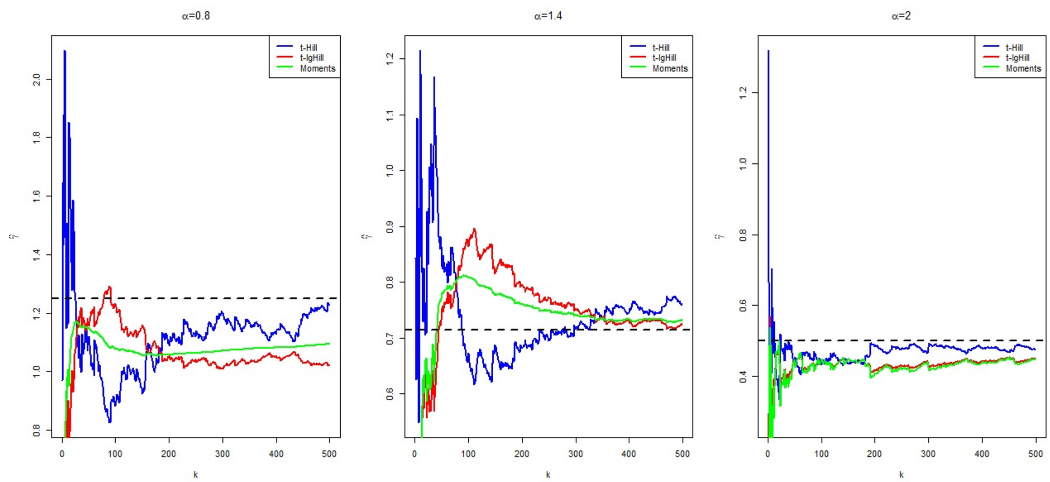

Figure 3.7: Comparation t-Hill and t-lgHill with Moments estimators for Extreme Values Index (EVI) $\gamma = \frac{1}{\alpha}$

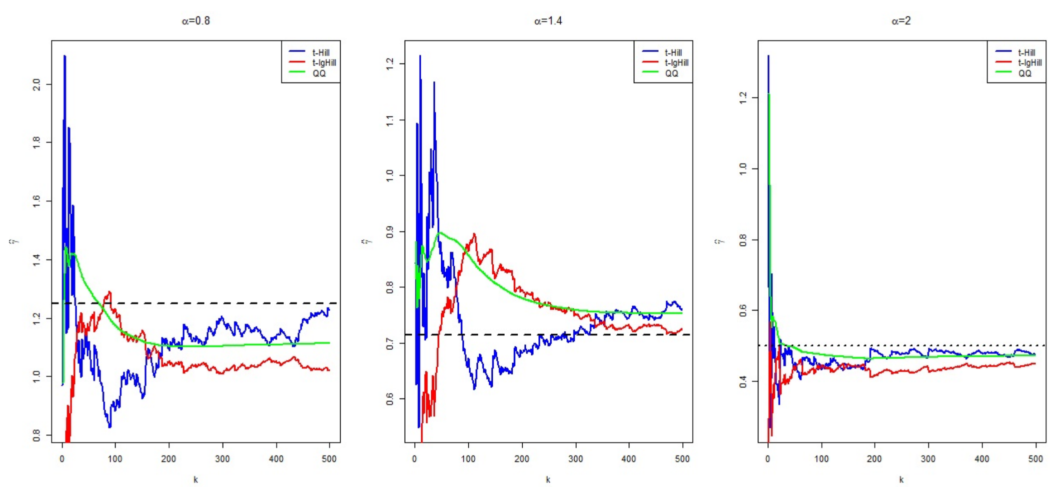

Figure 3.8: Comparation t-Hill and t-lgHill with QQ estimators for Extreme Values Index (EVI) $\gamma = \frac{1}{\alpha}$

Table 3.1: Results of the percentage relative bias (RB)and the percentage relative root-mean-square error (RRMSE) for the various estimators of the Pareto tail index $\gamma = \frac{1}{\alpha}$ for $n = 1000$ i.i.d. data with Pareto distribution $P(\alpha, 1)$

α	Estimator	RB	RRMSE
	Hill	-0.06434353	5.442569
	t-Hill	0.13160155	6.605862
0.8	t-lgHill	-0.23329831	10.455099
	Moments	-0.34940135	6.404770
	QQ	1.76659296	7.103728
	Hill	0.1845738	1.909956
	t-Hill	0.2890444	2.154207
1.4	t-lgHill	-0.3246375	3.827363
	Moments	-0.3405037	3.012380
	QQ	0.5620671	2.528399
	Hill	-0.05267297	0.7541040
	t-Hill	-0.06307524	0.7794487
2	t-lgHill	-0.03432581	1.8681507
	Moments	-0.14626960	1.8935889
	QQ	0.25929963	1.1944632

3.5 Robust test for normality

In simple linear regression, a general assumption is that errors follow a normal distribution. To verify this assumption, there are many tests. Thode (2002) carries out a complete discussion of several of these normality tests, among which the Shapiro-Wilk (SW) (Shapiro and Wilk, 1967) test, which is one of the most popular tests nowadays. In econometrics and related fields, the most widely adopted omnibus test for normality is the Jarque-Bera test (JB) (Jarque and Vera, 1980). Among the tests based on the empirical distribution function (EDF), the Lilliefors omnibus test (Kolmogorov-Smirnov) ($L(KS)$) (Lilliefors, 1967) is distinguished. However, the SW, JB, $L(KS)$ tests do not provide information about the nature of the deviation from normality, e.g. skewness, heavy tails or outliers.Therefore specialized tests directed to particular alternatives are needed.

Several authors affirm that the Jarque-Bera test behaves well in comparison with other tests of normality if the alternatives belong to the Pearson family. However this test behaves badly for the distributions of short tails and bimodal shape and sometimes it is biased.
Střelec and Stehlík (2009) introduce for the alternatives that are problematic

for the Jarque-Bera a general class of robust tests associated with these alternatives. In this sense, the authors present the general form of Jarque-Bera robust test and propose general guidelines for appropiate small samples testing for normality.

3.5.1 General class of robust tests

Střelec and Stehlík (2009) introduce a general class of robust test, the so called RT class. Particularly, the special cases of this class are the clasical JB test, the robust JB test introduced for Gel and Gastwirth (see Gel and Gastwirth, 2008), Geary's test (see Geary, 1935) and Urzúa's test (see Urzúa, 1996) and Uthoff's test (see Uthoff, 1973), whose statistics have asymptotically χ^2_2 distribution.
Střelec and Stehlík (2009) define the RT class of Jarque-Bera as

$$RT = \frac{k_1(n)}{C_1}\left(\frac{M^{\alpha_1}_{i_1,j_1}}{M^{\alpha_2}_{i_2,j_2}}\right)^2 + \frac{k_2(n)}{C_2}\left(\frac{M^{\alpha_3}_{i_3,j_3}}{M^{\alpha_4}_{i_4,j_4}} - 3\right)^2, \tag{3.66}$$

with

$$M_{i,j} = \frac{1}{n}\sum_{m=1}^{n} \varphi_j(X_m - M_{(i)}),$$

where $i \in \{0,1\}$ denotes either the arithmetic mean $M_{(0)} = \overline{x}$ or median $M_{(1)} = M_n$, $j \in \{0,1,2,3,4\}$ and φ_j is tractable and continuous function $\varphi_o = \sqrt{\frac{\pi}{2}}|x|$, $\varphi_1 = x$, $\varphi_2 = x^2$, $\varphi_3 = x^3$ and $\varphi_4 = x^4$.

3.5.2 Special cases of *RT* class

1. The classical **Jarque-Bera test** is a special case of RT test for $k_1(n) = n$, $k_2(n) = n$, $C_1 = 6$, $C_2 = 24$, $\alpha_1 = 1$, $\alpha_2 = \frac{3}{2}$, $\alpha_3 = 1$, $\alpha_4 = 2$, $i_1 = 0$, $i_2 = 0$, $i_3 = 0$, $i_4 = 0$, $j_1 = 3$, $j_2 = 2$, $j_3 = 4$, $j_4 = 2$. The Jarque-Bera test statistic JB is definided as

$$JB = \frac{n}{6}\left(\frac{\hat{\mu}_3}{\hat{\mu}_2^{\frac{3}{2}}}\right)^2 + \frac{n}{24}\left(\frac{\hat{\mu}_4}{\hat{\mu}_2^2} - 3\right)^2. \tag{3.67}$$

2. The **test of Urzúa** (see Urzúa, 1996) is special case of RT test for $k_1(n) = \frac{(n+1)(n+3)}{n-2}$, $k_2(n) = \frac{(n+1)^2(n+3)(n+5)}{n(n-2)(n-3)}$, $C_1 = 6$, $C_2 = 24$, $\alpha_1 = 1$,

$\alpha_2 = \frac{3}{2}$, $\alpha_3 = 1$, $\alpha_4 = 2$, $i_1 = 0$, $i_2 = 0$, $i_3 = 0$, $i_4 = 0$, $j_1 = 3$, $j_2 = 2$, $j_3 = 4$, $j_4 = 2$. Urzúa's Jarque-Bera test statistic JBU is defined as

$$JBU = \frac{\frac{(n+1)(n+3)}{n-2}}{6}\left(\frac{\hat{\mu}_3}{\hat{\mu}_2^{\frac{3}{2}}}\right)^2 + \frac{\frac{(n+1)^2(n+3)(n+5)}{n(n-2)(n-3)}}{24}\left(\frac{\hat{\mu}_4}{\hat{\mu}_2^2} - 3\right)^2. \tag{3.68}$$

3. The **robust Jarque-Bera test** (see Gel and Gastwirth, 2008) is a special case of RT test for $k_1(n) = n$, $k_2(n) = n$, $C_1 = 6$, $C_2 = 24$, $\alpha_1 = 1$, $\alpha_2 = 3$, $\alpha_3 = 1$, $\alpha_4 = 4$, $i_1 = 0$, $i_2 = 1$, $i_3 = 0$, $i_4 = 1$, $j_1 = 3$, $j_2 = 0$, $j_3 = 4$, $j_4 = 0$. The robust Jarque-Bera test statistic RJB is defined as

$$RJB = \frac{n}{6}\left(\frac{\hat{\mu}_3}{J_n^3}\right)^2 + \frac{n}{24}\left(\frac{\hat{\mu}_4}{J_n^4} - 3\right)^2, \tag{3.69}$$

where J_n is the average absolute deviation from the sample median (MAAD) (see Gastwith, 1982) defined by

$$J_n = \frac{\frac{\pi}{2}}{n}\sum_{i=1}^{n}|X_i - M_n|.$$

Chapter 4

Extreme Multivariate Values

In Chapters 1 to 3 we refer to the distribution functions of univariate extreme values and the estimations of their parameters. In particular, in Chapter 3 we studied various estimators for the extreme value index or tail index $\gamma = \frac{1}{\alpha}$, where α is the shape parameter of a Pareto type I distribution.
Now our analysis focuses on multivariate extreme values, considering that when studying two or more processes each process can be modeled using multivariate techniques, but there is strong evidence to study the interrelationships of extreme values.

The probability theory for extreme multivariate values has been well developed, there are similar results to the results obtained for the univariate case. Some of these results have been summarized by Galambos (1978). It is worth mentioning the work of Marshall and Olkin (1983), who obtain some multivariate analogs of Gnedenko's characterizations of attraction domain. More recently Li et al. (2012) discuss the parametric analysis of the bivariate failure time data.

4.1 Domain of Attraction

In the section 1.1.2 of the Chapter 1, we saw the necessary and sufficient conditions for that a distribution of extreme value F to belong to the attraction domain of other distribution of extreme value G_γ. Now we will analyze this for the $k-$dimensional case (multivariate) based on the work of Marshal and Olkin (1983). Previously we consider the following prerequisites.

Remark 4.1.1. *The univariate (max or min) extreme values distributions are the same type as* Φ_α, Ψ_α *or* Λ (Φ^*_α, Ψ^*_α *or* Λ^*), *where*

$$\begin{array}{ll}\Phi_\alpha(x) = \mathrm{e}^{-x^{-\alpha}}, \quad x > 0; & \Phi^*_\alpha(x) = 1 - \mathrm{e}^{-(-x)^{-\alpha}}, \quad x < 0 \quad (\alpha > 0),\\ \Psi_\alpha(x) = \mathrm{e}^{-(-x)^{\alpha}}, \quad x \le 0; & \Psi^*_\alpha(x) = 1 - \mathrm{e}^{-x^{\alpha}}, \quad x \ge 0 \quad (\alpha > 0),\\ \Lambda(x) = \mathrm{e}^{-\mathrm{e}^{-x}}, \quad -\infty < x < \infty; & \Lambda^*(x) = 1 - \mathrm{e}^{-\mathrm{e}^{x}}, \quad -\infty < x < \infty.\end{array}$$

Remark 4.1.2. *If* H *is the joint distribution of* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *then* H_i *denotes the marginal distribution of* $\mathbf{X}_i$, $i = 1, \ldots, k$.

4.1.1 Max domain of attraction

The following proposition is the k-dimensional version of (1.13) in the Chapter 1.

Proposition 4.1.1. *Let* G *be a* k*-dimensional max extreme value distribution such that* $G_i = \Phi_\alpha$, $i = 1, \ldots, k$ *and let* $\phi_i(t) = \overline{F}_i^{-1}\overline{F}_1(t)$, $i = 2, \ldots, k$, $-\infty < t < \infty$. F *is in the max domain of attraction of* G *if and only if*

$$\lim_{t \longrightarrow \infty} \frac{1 - F(tx_1, \phi_2(t)x_2, \ldots, \phi_k(t)x_k)}{1 - F_1(t)} = -\log G(\mathbf{x}), \quad \textit{for all } \mathbf{x}, \tag{4.1}$$

such that $G(\mathbf{x}) > 0$.

Proposition 4.1.2. *Let* G *be a* k*-dimensional max extreme value distribution such that* $G_i = \Psi_\alpha$, $i = 1, \ldots, k$. *Then* $F \in D_{\max}(G)$ *if and only if*

$$\textit{there exists } \mathbf{x}^0 \in \mathbb{R}^k \textit{ such that } F(\mathbf{x}^0) = 1 \textit{ and } F(\mathbf{x}) < 1 \textit{ if } \mathbf{x} \ne \mathbf{x}^0,\ x_i \le x_i^0,\ i = 1, \ldots, k, \tag{4.2}$$

and

$$\lim_{t \longrightarrow 0} \frac{1 - F((tx_1, \phi_2(t)x_2, \ldots, \phi_k(t)x_k) + \mathbf{x}^0)}{1 - F_1(x_1^0 - t)} = -\log G(\mathbf{x}), \tag{4.3}$$

for all $\mathbf{x}$ *such that* $G(\mathbf{x}) > 0$, *where*

$$\phi_i(t) = x_i^0 - \overline{F}_i^{-1}(\overline{F}_1(x_1^0 - t)), \quad i = 2, \ldots, k.$$

Proposition 4.1.3. *Let G be a k-dimensional max extreme value distribution such that $G_i = \Lambda$, $i = 1, \ldots, k$. Then $F \in D_{\max}(G)$ if and only if*

$$\lim_{t \longrightarrow x_1^0} \frac{1 - F(\mathbf{a}(t)\mathbf{x} + \mathbf{b}(t))}{1 - F_1(t)} = -\log G(\mathbf{x}), \quad \textit{for all } \mathbf{x} \textit{ such that } G(\mathbf{x}) > 0, \tag{4.4}$$

where

$$\begin{aligned} x_1^0 &= \sup\{t : F_1(t) < 1\}, \\ a_i(t) &= \overline{F}_i^{-1}(\mathrm{e}^{-1}\overline{F}_i(t)) - \overline{F}_i^{-1}\overline{F}_1(t), \qquad\qquad i = 1, \ldots, k. \\ b_i(t) &= \overline{F}_i^{-1}\overline{F}_1(t), \end{aligned}$$

Remark 4.1.3. *In the next proposition we assume that each F_i has density f_i and $r_i = \frac{f_i}{\overline{F}_i}$ is the corresponding hazard rate.*

Proposition 4.1.4. *Let G be a k-dimensional max extreme value distribution such that $G_i = \Lambda$, $i = 1, \ldots, k$.*
Let $\phi_i(t) = \overline{F}_i^{-1}\overline{F}_1(t)$, $i = 2, \ldots, k$, and let $x_1^0 = \sup\{x : F_1(x) < 1\}$. Then $F \in D_{\max}(G)$ if

$$\lim_{t \longrightarrow x_1^0} \frac{1 - F\left(\frac{x_1}{r_1(t)} + t, \frac{x_2}{r_2(\phi_2(t))} + \phi_2(t), \ldots, \frac{x_k}{r_k(\phi_k(t))} + \phi_k(t)\right)}{1 - F_1(t)} = -\log G(\mathbf{x}), \tag{4.5}$$

for all $\mathbf{x}$ such that $G(\mathbf{x}) > 0$.

4.1.2 Min domain of attraction

Proposition 4.1.5. *Let G be a k-dimensional min extreme value distribution such that $G_i = \Phi_\alpha^*$, $i = 1, \ldots, k$ and let $\phi_i(t) = \overline{F}_i^{-1}\overline{F}_1(t)$, $i = 2, \ldots, k$, $-\infty < t < \infty$. F is in the min domain of attraction of G if and only if*

$$\lim_{t \longrightarrow \infty} \frac{1 - \overline{F}(tx_1, \phi_2(t)x_2, \ldots, \phi_k(t)x_k)}{F_1(t)} = -\log \overline{G}(\mathbf{x}), \quad \textit{for all } \mathbf{x}, \tag{4.6}$$

such that $G(\mathbf{x}) > 0$.

Proposition 4.1.6. *Let G be a k-dimensional min extreme value distribution such that $G_i = \Psi^*_\alpha$, $i = 1, \ldots, k$. Then $F \in D_{\min}(G)$ if and only if there exists $\mathbf{x}_0 = (x_{01}, \ldots, x_{0k})$, such that*

$$F(\mathbf{x}_0) = 0 \text{ and } F(\mathbf{x}) > 0 \text{ if } \mathbf{x} \neq \mathbf{x}^0,\ x_i \geq x_{0i},\ i = 1, \ldots, k, \tag{4.7}$$

and

$$\lim_{t \longrightarrow 0} \frac{1 - \overline{F}((tx_1, \phi_2(t)x_2, \ldots, \phi_k(t)x_k) + \mathbf{x}_0)}{F_1(t + x_{01})} = -\log \overline{G}(\mathbf{x}), \tag{4.8}$$

for all $\mathbf{x}$ such that $\overline{G}(\mathbf{x}) > 0$, where

$$\phi_i(t) = F_i^{-1} F_1(x_{01} + t) - x_{0i}.$$

Proposition 4.1.7. *Let G be a k-dimensional min extreme value distribution such that $G_i = \Lambda^*$, $i = 1, \ldots, k$. Then $F \in D_{\min}(G)$ if and only if*

$$\lim_{t \longrightarrow x_{01}} \frac{1 - \overline{F}(\mathbf{a}(t)\mathbf{x} + \mathbf{b}(t))}{F_1(t)} = -\log \overline{G}(\mathbf{x}), \tag{4.9}$$

for all $\mathbf{x}$ such that $G(\mathbf{x}) > 0$, where

$$\begin{aligned} &x_0 = \inf\{t : F_1(t) > 0\}, \\ &a_i(t) = F_i^{-1}(\mathrm{e}^{-1} F_1(t)) - F_i^{-1} F_1(t), \qquad i = 1, \ldots, k. \\ &b_i(t) = F_i^{-1} F_1(t), \end{aligned}$$

Remark 4.1.4. *In the next proposition we assume that each F_i has density f_i and let $r_i^* = \frac{f_i}{F_i}$.*

Proposition 4.1.8. *Let G be a k-dimensional min extreme value distribution such that $G_i = \Lambda^*$, $i = 1, \ldots, k$.*
Let $\phi_i(t) = F_i^{-1} F_1(t)$, $i = 2, \ldots, k$, and let $x_{01} = \inf\{t : F_1(t) > 0\}$. Then $F \in D_{\min}(G)$ if

$$\lim_{t \longrightarrow x_{01}} \frac{1 - \overline{F}\left(\frac{x_1}{r_1^*(t)} + t, \frac{x_2}{r_2^*(\phi_2(t))} + \phi_2(t), \ldots, \frac{x_k}{r_k^*(\phi_k(t))} + \phi_k(t)\right)}{F_1(t)} = -\log \overline{G}(\mathbf{x}), \tag{4.10}$$

for all $\mathbf{x}$ such that $\overline{G}(\mathbf{x}) > 0$.

4.2 Association and independence

De Oliveira (1962/1963) noted that all multivariate extreme value distributions G of k variables $X_1, \ldots, X_k$ satisfy the condition

$$G(x_1, \ldots, x_k) \geq \prod_{i=1}^{k} G_i(x_i) \tag{4.11}$$

where G_i, $i = 1, \ldots, k$ are the marginal distributions.

Lehman (1966) called this property "positive quadrant dependence" in the case of $k = 2$. The positive quadrant dependence implies that all covariances are nonnegative.

Definition 4.2.1. *The random variables $X_1, \ldots, X_k$ are said to be associated if for every pair θ, Ψ nondecreasing functions defined on $\mathbb{R}^k$,*

$$\mathrm{Cov}(\theta(\mathbf{X}), \Psi(\mathbf{X})) \geq 0, \tag{4.12}$$

whenever the relevant expectations exist.

Esary et al. (1967) introduced this concept of positive dependence and showed that it is stronger than positive quadrant dependence.

Proposition 4.2.1. *If $X_1, \ldots, X_k$ have a multivariate extreme value distribution, then $X_1, \ldots, X_k$ are associated.*

Remark 4.2.1. *The above proposition shows that random variables with a multivariate extreme value distribution are always associated. Also the proposition 4.2.1 notes that pairwise independent random variables $X_1, \ldots, X_k$ having a multivariate extreme value distribution are mutually independent. This fact greatly simplifies the study of independence in multivariate extreme value distributions. This allows studies of asymptotic independence to be confined to the bivariate case.*

In the bivariate case for example: Geffroy (1958/1959) gives a necessary conditions for asymptotic independence of two maximums. Sibuya (1960), Berman (1961) and Galambos (1975) obtain other conditions.
Mikhailov(1974) and Galambos (1975) have obtained versions of the following proposition, which has a simple direct proof.

Proposition 4.2.2. *Suppose $k = 2$. Under the conditions accompanying the equations (4.1), (4.3) or (4.4), asymptotic independence occurs if and only if the limit in the respective equation is* 0 *when* $1-F$ *is replaced by* $\overline{F}$. *Under the conditions the equations (4.6), (4.8) or (4.9), asymptotic independence occurs if and only if the limit in the respective equation is* 0 *when* $1-\overline{F}$ *is replaced by* F.

4.2.1 Examples

Example 4.2.1. *Ali, Mikhail, Haq (1965)*

$$F(x_1, x_2) = \frac{F_1(x_1)F_2(x_2)}{1-\alpha\overline{F}_1(x_1)\overline{F}_2(x_2)}, \qquad -1 \le \alpha \le 1. \tag{4.13}$$

Since a bivariate extreme value distribution belongs to its own domain of attraction, it folows that the bivariate distribution (4.13) is a bivariariate extreme values distribution only in the case of independences, i.e. when $\alpha = 0$.

Example 4.2.2. *Suppose that*

$$\overline{F}(\mathbf{x}) = \overline{H}_1(x_1)\overline{H}_2(x_2)\overline{H}_3(\max(x_1, x_2)), \qquad -\infty < x_1, x_2 < \infty, \tag{4.14}$$

with

$$\overline{H}_1(z) = \mathrm{e}^{-\lambda_1 z}, \qquad \overline{H}_2(z) = \mathrm{e}^{-\lambda_2 z}, \qquad \overline{H}_3(z) = \mathrm{e}^{-\lambda_3 z},$$

$$z \ge 0 \text{ for some } \lambda_1, \lambda_2, \lambda_3 \ge 0,\ \lambda_1 + \lambda_3 > 0,\ \lambda_2 + \lambda_3 > 0.$$

Then F *is a bivariate exponential distribution (Marshal and Olkin, 1967). Here* $F_1, F_2 \in D_{\max}(\Lambda)$. *Then using the Proposition 4.1.4, it can be shown that* $F \in D_{\max}(G)$, *where*

$$G(x_1, x_2) = \begin{cases} \exp\{-\mathrm{e}^{-x_1} - \mathrm{e}^{-x_2}\}, & \max(\lambda_1, \lambda_2) > 0 \\ \exp\{-\mathrm{e}^{-x_1} - \mathrm{e}^{-x_2} + \mathrm{e}^{-\max(x_1, x_2)}\}, & \lambda_1 = \lambda_2 = 0 \end{cases} \tag{4.15}$$

In Example 5.2.2 of Galambos (1978), $F_1, F_2 \in D_{\min}(\Psi^*)$ *and from Proposition 4.1.6, it can be shown that* $F \in D_{\min}(G)$, *where*

$$\overline{G}(x_1, x_2) = \exp\left\{-\left[\lambda_1 x_1 + \lambda_2 \frac{\lambda_1+\lambda_3}{\lambda_2+\lambda_3} x_2 + \lambda_3 \max\left(x_1, \frac{\lambda_1+\lambda_3}{\lambda_2+\lambda_3} x_2\right)\right]\right\}, \quad x_1, x_2 \ge 0. \tag{4.16}$$

4.3 Estimation procedures

Marshall and Olkin (1967) introduced the so-called Marshall-Olkin bivariate exponential (MOBE) model, which assumes the joint survival function of T_1 and T_2 has the form

$$S(t_1, t_2) = \exp\{-\lambda_1 t_1 - \lambda_2 t_2 - \lambda_3 \max(t_1, t_2)\}, \tag{4.17}$$

where λ_1, λ_2 and λ_3 are some parameters.
Lu (1989) introduced two parameters β_1 and β_2 to model (4.17), getting

$$S(t_1, t_2) = \exp\{-\lambda_1 t_1^{\beta_1} - \lambda_2 t_2^{\beta_2} - \lambda_3 \max(t_1^{\beta_1}, t_2^{\beta_2})\}. \tag{4.18}$$

This model with $\beta_1 = \beta_2$ was referred to as the Marshall-Olkin bivariate Weibull (MOBW) model by Kundu and Dey (2009).
In the models (4.17) and (4.18), we can easily see that parameter λ_3 represents the correlation between T_1 and T_2, and $\lambda_3 = 0$ means that T_1 and T_2 are independent. Also we can show that under the model (4.17), both T_1 and T_2 follow exponential marginal distribution, i.e.

$$T_1 \sim \text{Exp}(\lambda_1 + \lambda_3), \qquad T_2 \sim \text{Exp}(\lambda_2 + \lambda_3).$$

Several authors have discussed the issue of fitting a model to bivariate data. For example

1. Bhattacharyya and Johnson (1971) considered the MOBE model and pointed out that the maximum likelihood estimate may not exist or be unique.

2. Kundu and Dey (2009) discussed the MOBW model for the case where $\beta_1 = \beta_2$ and when one observes exact failure time data. They developed an EM algorithm for the maximum likelihood estimate.

3. Nandi and Devan (2010) generalized the EM algorithm to the situation of right-censored failure time data. They used a pseudolikelihood function and assumed that $\beta_1 = \beta_2$.

4. Li et al. (2012) present and investigate two general estimation procedures to fit the bivariate model (4.18) to bivariate data, based on the Marshal-Olkin models. A graphical approach and marginal approach.

We consider the grafical estimation procedure.

4.3.1 Graphical Estimation Procedure

Assuming $(T_1, T_2) \sim MOBW(\beta_1, \beta_2, \lambda_1, \lambda_2, \lambda_3)$ the model given in (4.18), and the main goals is to estimate the unknown parameters β_1, β_2, λ_1, λ_2 and λ_3. Under this assumption is easy to show that the marginal distribution functions of T_1 and T_2 have the form,

$$F_1(t_1) = 1 - \exp\{-(\lambda_1 + \lambda_3)t_1^{\beta_1}\}, \tag{4.19}$$

$$F_2(t_2) = 1 - \exp\{-(\lambda_2 + \lambda_3)t_2^{\beta_2}\}. \tag{4.20}$$

That is, both T_1 and T_2 follow the univariate Weibull distribution. Let $S_1(t_1)$ and $S_2(t_2)$ be the marginal survival functions of T_1 and T_2 respectively. Then,

$$\log\left(-\log(S_1(t_1))\right) = \beta_1 \log(t_1) + \log(\lambda_1 + \lambda_3),$$
$$\log\left(-\log(S_2(t_2))\right) = \beta_2 \log(t_2) + \log(\lambda_2 + \lambda_3).$$

Therefore, if $S_1(t_1)$ and $S_2(t_2)$ are known, one can fit the linear models given above to estimate the parameters involved.

Let $\hat{S}_k$ be the Kaplan-Meier estimate of S_k based on the rigth-censored data $\{(t_{ki}, \delta_{ki}); i = 1, \ldots, n\}$ on T_k, $k = 1, 2$. Define

$$\mathbf{b}_k = \begin{bmatrix} \log(\lambda_k + \lambda_3) \\ \beta_k \end{bmatrix}, \quad \mathbf{y}_k = \begin{bmatrix} \log(-\log(\hat{S}_k(t_{k1}))) \\ \cdot \\ \cdot \\ \cdot \\ log(-\log(\hat{S}_k(t_{kn}))) \end{bmatrix}, \quad \mathbf{X}_k = \begin{bmatrix} 1 & \log(t_{k1}) \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & \log(t_{kn}) \end{bmatrix} \tag{4.21}$$

Then it is natural to fit the data $\left\{\left(\hat{S}_k(t_{ki}), t_{ki}\right); i = 1, \ldots, n\right\}$ to the linear model

$$\mathbf{y}_k = \mathbf{X}_k \mathbf{b}_k,$$

and estimate β_k and $\lambda_k + \lambda_3$ by the least-square estimates, $k = 1, 2$.

Let $\hat{\beta}_1$, $\hat{\beta}_2$, $\widehat{\lambda_1 + \lambda_3}$ and $\widehat{\lambda_2 + \lambda_3}$ be the estimates of the parameters given above. To estimate individual λ_1, λ_2 and λ_3, we assume that a consistent estimate, $\hat{S}(t_1, t_2)$, of the joint survival function $S(t_1, t_2)$ is availabre, considering that given $\hat{\beta}_1$ and $\hat{\beta}_2$ we can divide the observed data $\{(t_{1i}, t_{2i})\}$ in two parts,

$$I_1 = \{(t_{1i}, t_{2i}); t_{1i}^{\hat{\beta}_1} > t_{2i}^{\hat{\beta}_2}\}, \qquad I_2\{(t_{1i}, t_{2i}); t_{1i}^{\hat{\beta}_1} < t_{2i}^{\hat{\beta}_2}\}.$$

Under model (4.18), we have

- For the data points in I_1

$$\log\left\{-\log S(t_{1i},t_{2i})-(\lambda_1+\lambda_3)*t_{1i}^{\beta_1}\right\}=\beta_2*\log(t_{2i})+\log(\lambda_2) \tag{4.22}$$

- For the data points in I_2

$$\log\left\{-\log S(t_{1i},t_{2i})-(\lambda_2+\lambda_3)*t_{2i}^{\beta_2}\right\}=\beta_1*\log(t_{1i})+\log(\lambda_1) \tag{4.23}$$

Now, replacing β_1, β_2, $\lambda_1+\lambda_3$ and $S(t_1,t_2)$ by $\hat{\beta}_1$, $\hat{\beta}_2$, $\widehat{\lambda_1+\lambda_3}$ and $\hat{S}(t_1,t_2)$ in equation (4.22),

$$\hat{\lambda}_3=\widehat{\lambda_2+\lambda_3}-\hat{\lambda}_2, \qquad \text{and} \qquad \hat{\lambda}_1=\widehat{\lambda_1+\lambda_3}-\hat{\lambda}_3.$$

Besides, one can perform the goodness of fit tests by fitting the data I_2 to model (4.23) replacing β_1, β_2, λ_1, λ_2, λ_3 and $S(t_1,t_2)$ by $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\lambda}_1$, $\hat{\lambda}_2$, $\hat{\lambda}_3$ and $\hat{S}(t_1,t_2)$.
The above procedure requires to estimate the joint survival function $S(t_1,t_2)$, which one can estimate using the empirical joint survival function.

$$\hat{S}(t_1,t_2)=\frac{1}{n}\sum_{i=1}^{n}I(t_{1i}\geq t_1,t_{2i}\geq t_2). \tag{4.24}$$

In the case of right-censored data, one can use the estimate given in Dabrowska (1988) and Hougaard (2000), which is in the following steps.

1. Define

$$R(t_1,t_2)=\sum_i I(T_{1i}\geq t_1,T_{2i}\geq t_2)$$
$$K_{11}(t_1,t_2)=\sum_i I(T_{1i}=t_1,\delta_{1i}=1,T_{2i}=t_2,\delta_{2i}=1),$$

   which can be considered as the bivariate risk and failure numbers, respectevely.

2. Define

$$K_{10}(t_1,t_2) = \sum_i I(T_{1i}=t_1, \delta_{1i}=1, T_{2i} \geq t_2),$$
$$K_{01}(t_1,t_2) = \sum_i I(T_{2i}=t_2, \delta_{2i}=1, T_{1i} \geq t_1),$$
$$L_{11}(t_1,t_2) = \frac{K_{11}(t_1,t_2)}{R(t_1,t_2)},$$
$$L_{10}(t_1,t_2) = \frac{K_{10}(t_1,t_2)}{R(t_1,t_2)},$$
$$L_{01}(t_1,t_2) = \frac{K_{01}(t_1,t_2)}{R(t_1,t_2)}.$$

3. The Dabrowska bivariate survival estimate is given by

$$\hat{S}(t_1,t_2) = \hat{S}_1(t_1)\hat{S}_2(t_2) \prod_{\substack{0\leq u\leq t_1 \\ 0\leq v\leq t_2}} \{1 - H(u,v)\},$$

where

$$\hat{S}_1(t_1) = \prod_{u\leq t_1}\{1 - L_{10}(u,0)\}, \qquad \hat{S}_2(t_2) = \prod_{v\leq t_2}\{1 - L_{01}(0,v)\},$$

$$H(u,v) = \frac{L_{10}(u,v)L_{01}(u,v) - L_{11}(u,v)}{\{1 - L_{10}(u,v)\}\{1 - L_{01}(u,v)\}}.$$

Note that $\hat{S}$ jump only at the observed time points where $\delta_1 = 1$ or $\delta_2 = 1$.

Remark 4.3.1. *All parameters estimates $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\lambda}_1$, $\hat{\lambda}_2$ and $\hat{\lambda}_3$ are functions of the Kaplan-Meier estimates $\hat{S}_k(t_k)$ and $\hat{S}(t_1,t_2)$. One can show that they are asymptotically unbiased and consistent, and their distributions can be approximated by the normal distributions.*

To estimate the variance for $\hat{\beta}_k$ and $\widehat{\lambda_k + \lambda_3}$, it is natural to use

$$\mathrm{cov}\{\hat{\mathbf{b}}_k\} = (\mathbf{X}_k^T\mathbf{X}_k)^{-1}\mathbf{X}_k^T \mathrm{cov}(\mathbf{y}_k)\mathbf{X}_k(\mathbf{X}_k^T\mathbf{X}_k)^{-1}$$

where $\text{cov}(\mathbf{y}_k)$ is aproximated by the sandwich estimate $H\text{cov}(\hat{S}_k(\mathbf{t}_k))H^T$ and $\text{cov}(\hat{S}_k(\mathbf{t}_k))$ by the modified Greenwood's formula given in Kang and Koehler (1997), where

$$H = H(S_k) = \begin{bmatrix} \frac{1}{S_k(t_{k1})\log(S_k(t_{k1}))} & 0 & \cdots & 0 \\ 0 & \frac{1}{S_k(t_{k2})\log(S_k(t_{k2}))} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{S_k(t_{kn})\log(S_k(t_{kn}))} \end{bmatrix},$$

with S_k replaced by $\hat{S}_k$.

Remark 4.3.2. *In general, it is difficult to estimate the covariance matrix of $\hat{S}(t_1, t_2)$ (Dabrowska, 1989). Li et al (2012) applied the simple bootstrap procesure, where the simulation results show that these variance estimates seem to work well for practical situations.*

Chapter 5

Applications

5.1 Estimation of the extreme value index

We address a sensitive issue, the issue of water quality for irrigation in the northern zone of Chile, where precisely this resource is scarce and of poor quality, due to high salinity and high concentrations of heavy metals like boron and arsenic. The high levels of boron and arsenic recorded in the river Lluta located in the *Región de Arica y Parinacota* make necessary an analysis and statistical study from the point of view of the extreme value theory.
Considering that in the analysis of extreme values a central issue is the estimation of the extreme value index (EVI), which describes the behavior of the tail of a distribution. This real value parameter indicates the heaviness of the tail, i.e., how extreme and frequent extreme events can be under the given probability distribution. There is a substantial number of publications on estimators for this EVI, such as those mentioned in Chapter 3.
Núñez-Soza and Stehlík (2017) consider the Hill, t-Hill and t-lgHill estimators, taking into account that the t-Hill estimator is asymptotically consistent and more robust.

5.1.1 LLuta river Data

The *Dirección General de Aguas (DGA)* belonging to the *Ministerio de Obras Públicas (MOP)* of the Government of Chile periodically makes measurements of the physico-chemical parameters in the different basins of the rivers of Chile. In particular, in the Lluta river basin there exist 7 monitoring stations where these parameters are controlled. Lluta river waters have high concentrations of

arsenic and boron. For example MOP-DGA (2004) reports that boron values from the DGA monitoring campaign have values between 4 mg/L (monitoring station DGA in *Caracarani* - Spring Summer) to 27 mg/L (monitoring station DGA *Río Lluta en Panamerica* - Autumn). In this same report the arsenic values from the DGA monitoring campaign present values between 0.1 mg/L (monitoring station DGA *Río Lluta en Panamericana* - Winter) to 0.6 mg/L (monitoring station DGA *Río Lluta en Chapizca* - Winter). Tchernitchin et al. (2015) determine a boron concentration of 12.5 mg/L and an arsenic concentration 0.184 mg/L (in Lluta river near the bocatoma). The Chilean norm (Nch 1333, 1978) establishes a level of 0.75 mg/L for boron and 0.10 mg/L for arsenic.
The previous antecedents suggest a extreme value distribution for both the boron and the arsenic concentration.
We considered the boron and arsenic measurements obtained by the DGA in the 7 monitoring stations of Table 5.1, in the period from 2000 to 2016. Figure 5.1 shows the concentrations of boron and arsenic observed in the 7 monitoring stations during the period indicated. In addition, a statistical summary of measurements of boron and arsenic concentration is recorded in Table 5.2. Both in Figure 5.1 and in Table 5.2, high concentrations of boron and arsenic are observed, many of these values are above the recommended standard.

Table 5.1: Monitoring stations Lluta river basin with their geographical coordinates

Station	Latitude	Longitude	Altitude
Río Azufre antes Río Caracarani	8031772	424042	3970
Río Caracarani en Alcerreca	8011521	433472	3253
Río Colpitas en Alcerreca	8010570	433770	3251
Río Lluta en Tocontasi	7968987	404886	1850
Vertiente al Pichin	7964296	389973	654
Río Lluta en Panamericana	7965042	362677	10
Río Lluta en Chapisca	7968994	406412	0

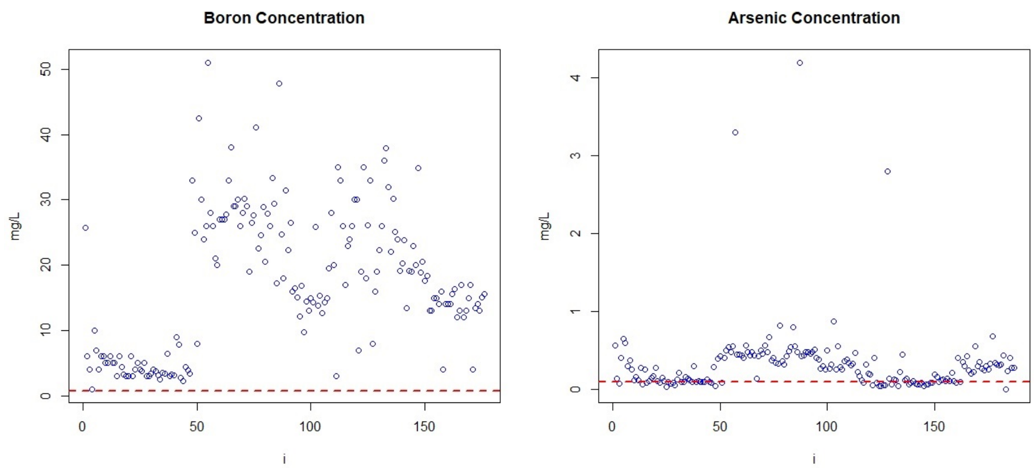

Figure 5.1: Data distribution of the boron concentration *(left chart)* and arsenic concentration *(right chart)* observed in the monitoring sites of the Lluta river

Table 5.2: Statistical summary of measurements of boron concentrarion and arsenic concentration 2000-2016.

Statistic	Boron	Arsenic
Minimum	1.000	0.0010
First quartile	6.319	0.1070
Median	16.399	0.2550
Third quartile	26.000	0.4200
Maximum	51.000	4.940
Mean	17.368	0.3214
Standar deviation	10.784	0.4431

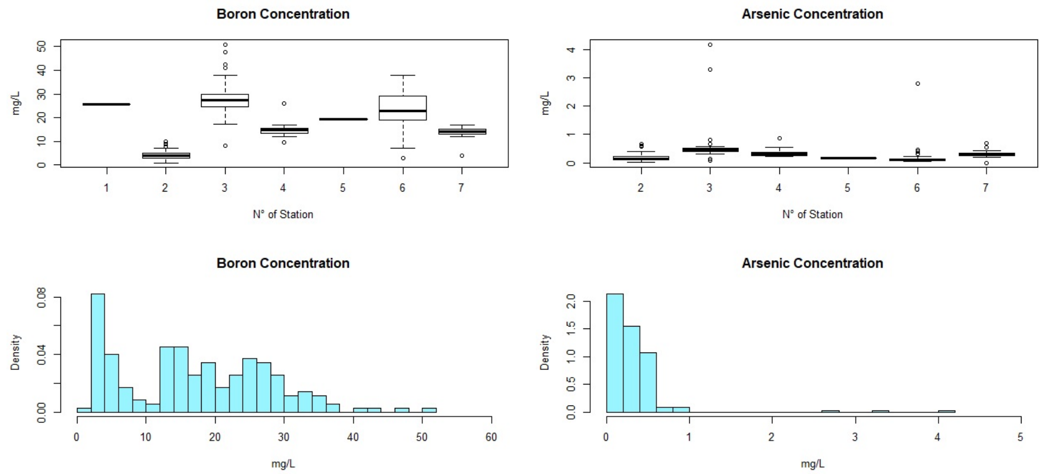

Figure 5.2: Boxplot and histogram of boron concentration and arsenic concentration observed in the monitoring sites of the Lluta river

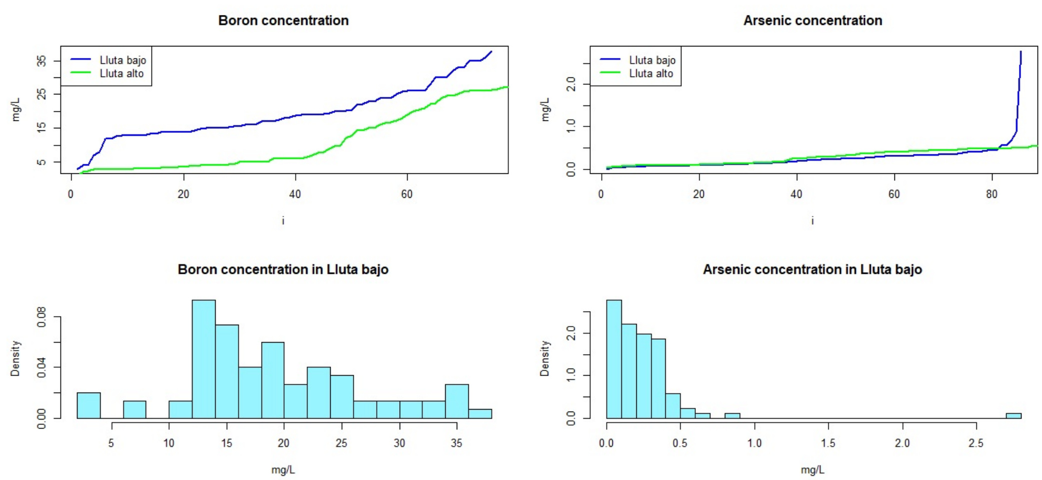

Figure 5.3: Plot and histogram of boron concentration and arsenic concentration observed in the monitoring sites of the sub-basin *Río Lluta Bajo*

In order to calculate the Hill, t-Hill and t-lgHill estimators of the extreme value index γ, we first determined the order statistics $X_{1,n}, \ldots, X_{n,n}$ for the boron and arsenic measurements, later we obtain the estimates mentioned above from the following expressions,

$$H_{k,n} = \frac{1}{k}\sum_{j=1}^{k} \log\left(\frac{X_{n-j+1,n}}{X_{n-k,n}}\right), \qquad \text{Hill estimator} \tag{5.1}$$

$$H^{*}_{k,n} = \left(\frac{1}{k}\sum_{j=1}^{n} \frac{X_{n-k,n}}{X_{n-j+1,n}}\right)^{-1} - 1, \qquad \text{t-Hill estimator} \tag{5.2}$$

$$H^{L}_{k,n} = \frac{\sum_{j=1}^{k}\left(\log\left(\frac{X_{n-j+1,n}}{X_{n-k,n}}\right)\right)^{2}}{\sum_{j=1}^{k}\log\left(\frac{X_{n-j+1,n}}{X_{n-k,n}}\right)} - \frac{1}{k}\sum_{j=1}^{k}\log\left(\frac{X_{n-j+1,n}}{X_{n-k,n}}\right), \quad \text{t-lgHill estimator} \tag{5.3}$$

where $k = 1, \ldots, n-1$.

The results obtained for the different values of k are shown in Figure 5.4 for data Lluta river basin and Figure 5.5 corresponds to the that estimate were conducted, considering the data of the sub-basin *Río Lluta bajo*.
Both in Figure 5.4, as well as in Figure 5.5, we observe that the t-lgHill estimator is more stable than the t-Hill and Hill estimators, which behave well for $k < 60$ in the case of the boron concentration and $k < 80$ in the case of arsenic concentration.

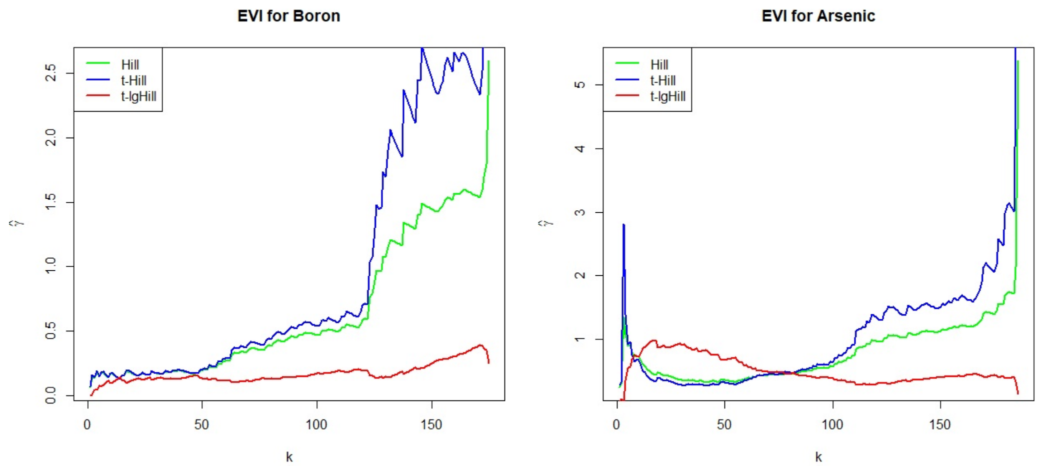

Figure 5.4: Extreme Value Index (EVI) estimators for a sample of 150+ data of boron and arsenic concentration observed in monitoring sites of the Lluta river

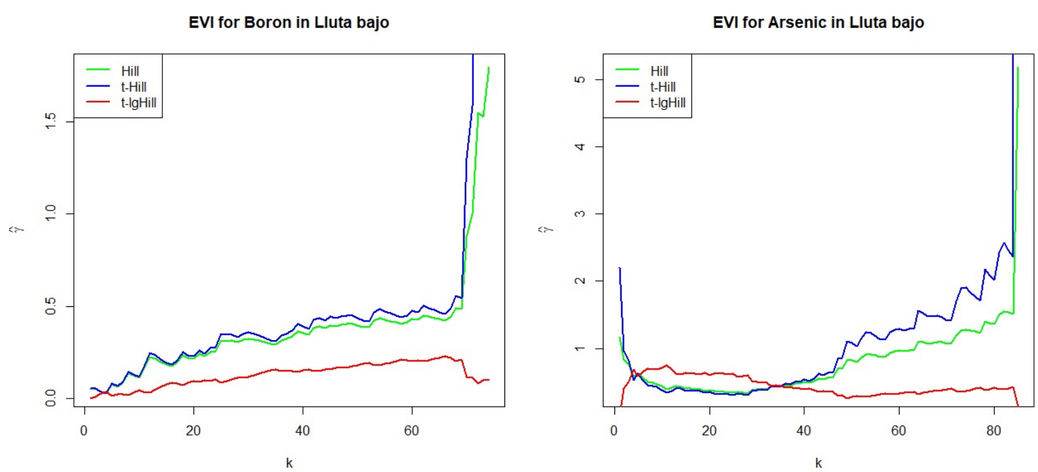

Figure 5.5: Extreme Values Index (EVI) estimators for a sample of 80+ data boron and arsenic concentration observed in monitoring sites of the sub-basin *Río LLuta Bajo*

5.1.2 Confidence bands

In order to build confidence bands for $\gamma = \frac{1}{\alpha}$ estimators, it is necessary to consider the following theorem

Theorem 5.1.1. *(Efron,1982) If $T(F) = F^{-1}(p)$ is the pth quantile, then jackknife variance estimate is inconsistent. For the median ($p = \frac{1}{2}$) we have that*

$$\frac{v_{jack}}{\sigma_n^2} \sim \left(\frac{\chi_2^2}{2}\right)^2,$$

where σ_n^2 is the asymptotic variance of the sample median.

Consequently, we can not easily use both the Delta nonparametric method and the the nonparametric Jacknife method to build confidence bands. Then we will resort to the Boostrapt method (Wassermann (2003)), which is a method to estimate the variance and the distribution of a statistic $T_n = g(X_1, \ldots, X_n)$. We can also use the bootstrap to construct confidence intervals. There are several ways to construct bootstrap confidence intervals. They vary in ease of calculation and accuracy. We consider Normal Interval given by

$$T_n \pm z_{\frac{\alpha}{2}} \widehat{se}_{boot},$$

with

$$\widehat{se}_{boot} = \sqrt{\frac{1}{B}\sum_{i=1}^{B}\left(T_i^* - \frac{1}{B}\sum_{i=1}^{B} T_i^*\right)^2},$$

where T_i^* is obtained from a sample where data have been replaced by others from the same sample, B is the number of booststrap samples considered and $z_{\frac{\alpha}{2}}$ is $\frac{\alpha}{2}$th quantile of Standard Normal distribution.
In this case $T_n = \hat{\gamma}$, where $\hat{\gamma}$ can be $H_{k,n}$, $H^*_{k,n}$ or $H^L_{k,n}$, which are functions of the order statistics $X_{1,n}, \ldots, X_{n,n}$.
Figure 5.6 shows the confidence bands obtained for the aforementioned statistics.

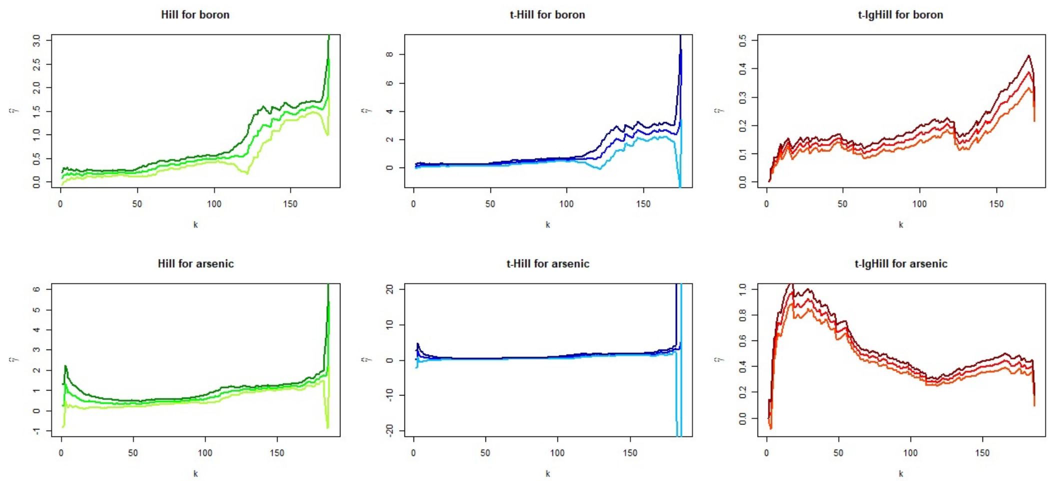

Figure 5.6: Confidence band for the extreme value index γ in the case of the boron concentration (*top graphics*) y of the arsenic concentration (*bottom graphics*)

5.2 Estimation of model coefficients

A current and very important problem nowadays is the quality of drinking water, due to its origin, to the continuous acts of contamination that occur in natural water sources and to the lack of clear and homogeneous regulation worldwide. Chile is not alien of this problem, especially in the northern zone where water is scarce and high levels of heavy metals are present, in addition to this we can not forget the contamination produced by the mining industry whose plants are often located near the sources of natural water. We should also consider that the Chilean norm for drinking water quality does not contemplate the same parameters that other standards consider, such as the recommendations of the World Health Organization (WHO).
In Chile, the problem of water quality becomes more delicate in areas where there is water shortage because drinking water is obtained from groundwater, which presents high levels of heavy metals, such as arsenic, boron, lead, etc. There exists a large number of research on the harmful effects on health produced by these heavy metals. Hence the importance of determining the level or concentration of these metals in drinking water. Sethlík et al. (2018) suggest a regression model with zero slope, both for boron concentration and arsenic

concentration, and conduct a study on the errors or residuals of this model.

5.2.1 Parinacota province Data

Arsenic is one of the few substances that have shown to cause cancer in humans by drinking water consumption. The World Health Organization (WHO, 2011) indicates that in various countries around the world, diseases caused by arsenic such as cancer, give rise to a significant problem for public health. Castillo and Venegas (2010) highlight that contamination by arsenic is a problem that affects north of Chile and hence they recomemmend to put more atention in terms of prevention and auditing process. In Chile, the entity in charge of the auditing process is the *Superintendencia de Servicios Sanitarios (SISS)*. A report of the *Controloria General de la República de Chile* (2015), about compliance of the normative for quality of the drinking water in several localities of Chile, shows infingements of the Chilean normative (NCh 409/1, 2005), because in this report, four over 15 samples of drinking water in Arica exceed the limit of 0.01 mg/L for the arsenic in the drinking water. Tchernitchin et al. (2015) report the presence of heavy metals in drinking water of *Arica y Parinacota*, e.g., in Chucuyo it has been registered 0.2738 mg/L of arsenic concentration and in the north zone of Arica city it hat been recorded arsenic levels of 0.0101 mg/L.
The Chilean governmental entity in charge of protection of public health and ensure the sanitary objetives is the *Subsecretaria de Salud Pública*. At the regional level there are units dependent on the *Subsecretaria de Salud Pública*, who work to improve the quality of life by carrying out monitoring, controls and inspection. In *Arica y Parinacota* is the *Secretaria Regional Ministerial (SEREMI) de Salud de Arica y Parinacota*, which regularly makes measurements of the quality of drinking water to ensure that quality standards are met.
In this context it is that in the year 2014 the *SEREMI de Salud de Arica y Parinacota* carried out an analysis of drinking water in the province of Parinacota. The analysis considered 28 monitoring sites located in this province. At each monitoring site a water sample was taken considering the date and hour. Subsequently, these water samples were analyzed by the *Instituto de Salud Pública (ISP) de Chile*. The analysis considered the measurement of the concentration of heavy metals present in drinking water particularly the arsenic concentration. We consider 24 monitoring sites because 4 of them were not able to acess for the monitoring information. Table 5.3 shows the loca-

tions where the monitoring sites are located. Figure 5.7 shows in the plot on the left the arsenic concentration observed in each site and the plot on the right corresponds to the histogram of this concentration and in Table 5.4 the statistical summary for arsenic concentration is recorded.

Table 5.3: *Comuna* and Locality of monitoring sites of the Parinacota province

Comuna	Locality	Site
Putre	Guallatire	Reten Guallatire
	Putre	Teniente del Campo 690
	Socoroma	Escuela San Francisco de Asis Casa particular
	Parinacota	Colegio Parinacota Patio
	Caquena	Escuela G-38 Vivienda S/N
	Belen	Consultorio
	Chucuyo	Reten de Carabineros Agua de Vertiente S/N
	Zapahuira	Restaurant Internacional Restaurant San Pedro Restaurant Zapahuira Casa S/N
	Chungara	Complejo Chungara
	Ticnamar	Colegio el Marquez
	Chapiquiña	Plaza
General Lagos	Alcerreca	Carabineros
	Visviri	Escuela G-35 S/N
	Ancolacane	Casa particular
	Tacora	S/N

Table 5.4: Statistical summary for arsenic concentration in drinking water of the Paricacota province 2014

Statistic	mg/L
Minimum	0.00300
First quartile	0.00500
Median	0.00500
Third quartile	0.02150
Maximum	0.60200
Mean	0.04433
Standar deviation	0.12237

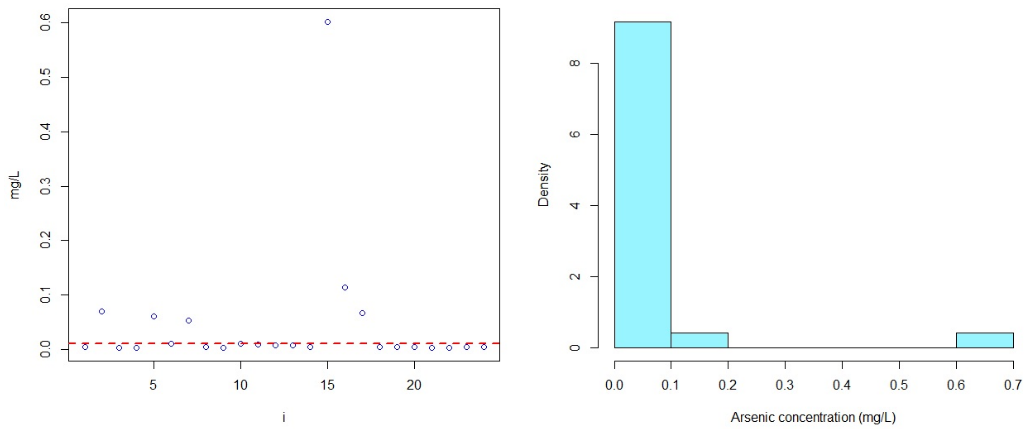

Figure 5.7: Arsenic concentration observed in monitoring sites of the Parinacota province

5.2.2 Arica city Data

According to the World Health Organization (WHO, 2011) the shortterm and longterm oral exposures to boric acid or borax in laboratory animals have demonstrated that the male reproductive tract is a consistent target of toxicity, causing testicular lesions in rats, mice and dogs who have received boric acid or borax in food or drinking water. Negative results in a large number of mutagenicity assays indicate that boric acid and borax are not genotoxic. Also the World Health Organization (WHO, 2009) indicates that the concentrations of boron found in drinking water from Chile, Germany, the United Kingdom and the USA ranged from 0.01 to 15.0 mg/L, with most values below 0.5 mg/L. These values are consistent with ranges and means observed for groundwater and surface water. This consistency is supported by two factors:

1) Boron concentrations in water are largely dependent on the leaching of boron from the surrounding geology and, to a decreasing extent, wastewater discharges.

2) Boron is not removed by conventional wastewater and drinking water treatment methods.

The recommended maximum limits for boron in waters for human consumption according to WHO is 2.4 mg/L ; for the European Economic Community (EEC) it is 1 mg/L; for Canada it is 5 mg/L and for Peru it is 1 mg/L. In Chile, boron is not regulated in the regulations applicable to water for human consumption, D.S. 735 and NCh 409. But, as we mentioned earlier, boron is regulated for irrigation water and other uses through the NCh 1333/78, modif. 1987, with a concentration of 0.75 mg/L. Several studies carried out in the *Región de Arica y Parinacota* show overwhelming results for the concentration of boron in drinking water. Tchernitchin et al. (2015) observed a maximum concentration of 14.45 mg/L in March 2014 in the locality of *Poconchile* of the *Región de Arica y Parinacota* and a maximum concentration of 4.585 mg/L in the northern sector of the city of Arica. Cortés et al. (2011) in a study carried out in the city of Arica at the request of the *Ministerio de Salud (MINSAL) de Chile*, between 2006 and 2008 they recorded boron concentrations from 0.22 to 11.3 mg/L with an average value of 4.18 mg/L. Cortes et al. (2011) evaluated exposition to boron using drinking water and urine samples in north of Chile. They took 75 samples of drinking water from different monitoring sites to measure then boron concentration levels. We considered the data provided by Sandra Cortés, Ph.D, which correspond to 13 urban radio monitoring sites, whose georeferential information is recorded in Table 5.5, in these sites a total of 56 measurements of boron concentration were made, which are shown in Figure 5.8 together with the corresponding histogram. In addition, Table 5.6 records the statistical summary corresponding to the boron concentration recorded in these 56 drinking water samples.

Table 5.5: Latitude and longitude (UTM) of monitoring sites and its corresponding pond

Street	Latitude	Longitude	Pond	Latitude	Longitude
Santiago Arata 1	7959586	363285	Cerro Chuño	7957149	364907
Santiago Arata 2	7961898	363260			
Diego Portales 1	7955916	363332			
Diego Portales 2	7956921	362418			
18 de Septiembre	7955298	363306			
21 de Mayo	7955097	362512			
Av. Manuel Castillo	7954427	364188	Saucache	7954322	363913
Tambo Quemado	7954215	364382			
Caupolicán	7955535	362080	Cerro La Cruz	7955496	361851
Maipú 1	7955937	361516			
Maipú 2	7956028	361419			
Chacabuco	7956385	361342			
General Velásquez	7957008	361254			

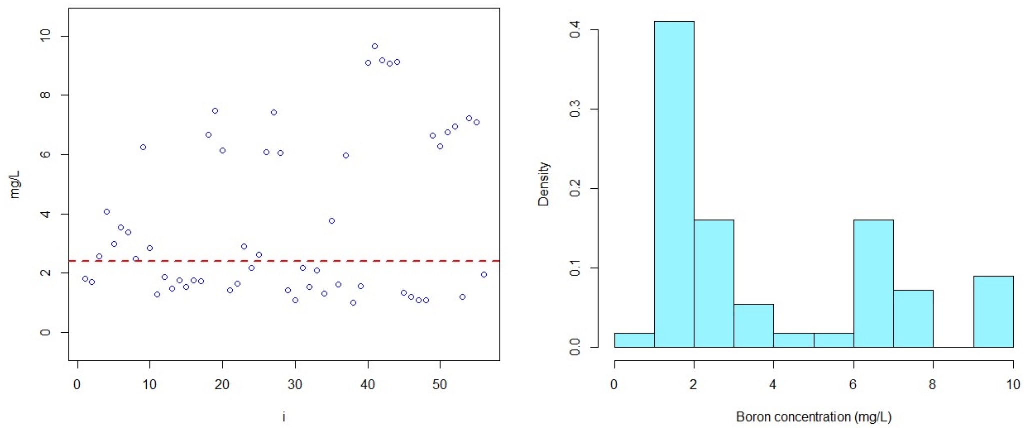

Figure 5.8: Boron concentration observed in monitoring sites of the Arica city

Table 5.6: Statistical summary for boron concentration in Arica 2006-2008

Statistic	mg/L
Minimum	0.990
First quartile	1.555
Median	2.515
Third quartile	6.255
Maximum	9.650
Mean	3.788
Standar deviation	2.737

At each monitoring site, an average of 4 water samples were taken and the boron concentration in each sample was recorded. We consider this information to determine the harmonic mean at each site, which we define by

$$H_i = \frac{x_{min,i} x_{max,i}}{\overline{x}_i} \quad i = 1, \ldots, 13, \tag{5.4}$$

where $x_{min,i}$ and $x_{max,i}$ is the minimun y maximum observed in the ith monitoring site respectively, and $\overline{x}_i$ is the mean concentration of the ith monitoring site.

5.2.3 Nonparametric and parametric bootstrap

In order to generate replicates of the measurements of arsenic and boron in each monitoring site we use nonparametric and parametric boostrap. In the case of parametric bootstrap, we assume that concentration levels have a Pareto type I distribution function, given by

$$F(x) = 1 - \left(\frac{x}{\lambda}\right)^{-\alpha}, \quad x > \lambda, \tag{5.5}$$

where $\alpha > 0$ es the shape parameter and λ is the scale parameter, such that

$$\lambda = \begin{cases} 0.0029, & x \text{ is the arsenic concentration} \\ 1, & x \text{ is the boron concentration} \end{cases} \tag{5.6}$$

The shape parameter that characterizes the tail distribution is estimated using the Hill, t-Hill and t-lgHill estimators mentioned in Chapter 3 and in the section 5.1 of this chapter.
In Table 5.7 we recorded for some values of k, the Hill, t-Hill and t-lgHill estimators obtained for the arsenic concentration and in the plot of Figure 5.9 we observed the behavior of these estimators for the different values of de k. This plot indicates that as k grows the estimated values are closer.
On the other hand, in the plot of Figure 5.10 we observe the values corresponding to the Hill, t-Hill and t-lgHill estimators for the boron concentration. Here, as k increases the values of the estimators are closer. Table 5.8 shows for some values of k, the value of the estimator of α.

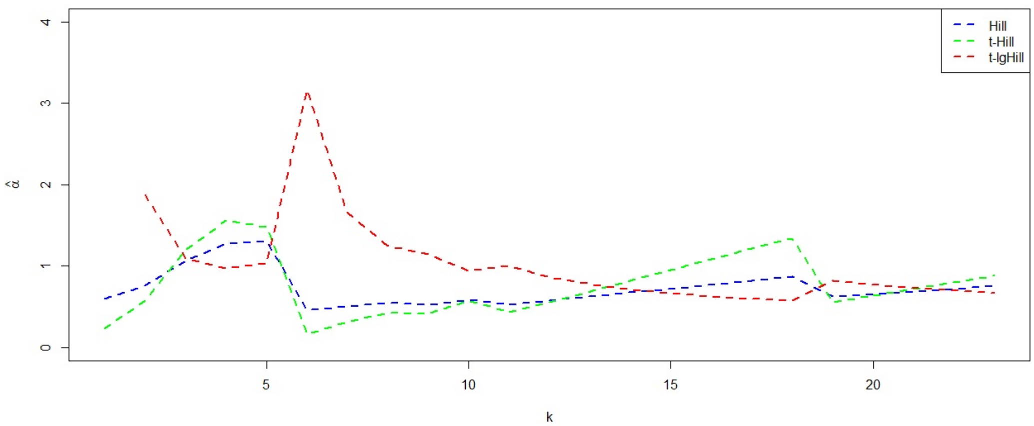

Figure 5.9: Estimators of shape parameter of the arsenic concentration distribution

Table 5.7: Shape parameters estimators of the arsenic concentration distribution for various values of k

k	Hill	t-Hill	t-lgHill
5	1.3071179	1.4827351	1.0347461
10	0.5871482	0.5724770	0.9416491
15	0.7234958	0.9517829	0.6656967
20	0.6570657	0.6378449	0.7683475
21	0.6899190	0.7197371	0.7278196
22	0.7227723	0.8016294	0.6945165
23	0.7556256	0.8835216	0.6666642

Table 5.8: Shape parameters estimators of the boron concentration distribution for various values of k

k	Hill	t-Hill	t-lgHill
3	2.571361	2.3149890	6.395571
6	1.944436	1.6746981	5.158933
9	1.423741	1.1653345	4.027709
10	1.246238	0.9718690	3.985237
11	1.339746	1.1272776	3.159342
12	1.449752	1.3037861	2.661926

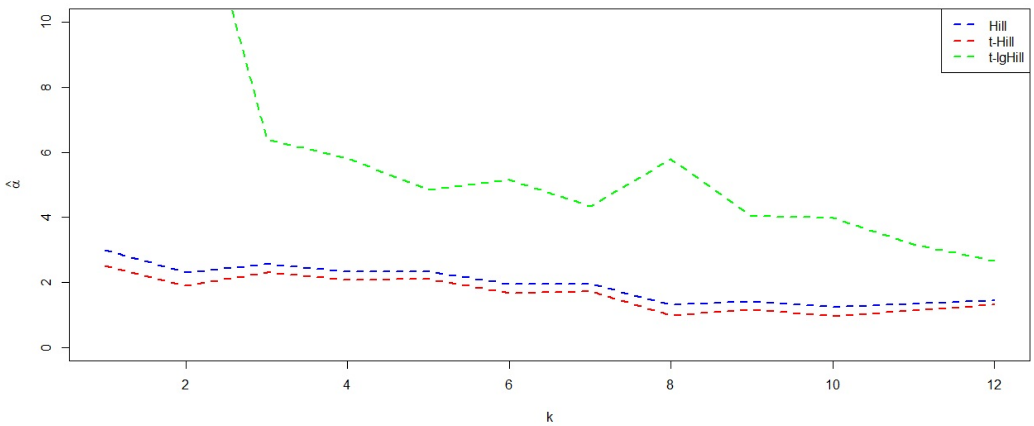

Figure 5.10: Estimators of the shape parameter of the boron concentration distribution (harmonic mean)

5.2.4 Linear regression and quantile regression

We use linear regression and quantile regression to model the levels of arsenic and boron concentration observed in Parinacota 2014 and Arica 2006-2008 respectively.
The simple linear regression model is given by

$$Y = a + bX + \varepsilon, \tag{5.7}$$

where Y is the response variable, X represents the predictor variable and ε is the error, which we assume to be Gaussian distributed.
On the other hand, the quantile regression (see Koenker and Basset, 1978) considers a linear functon of X, such that

$$Q_Y(\tau|X) = a(\tau) + b(\tau)X, \tag{5.8}$$

where the expression τ states that the parameters are for a specific τ quantile. This means that the parameters vary with τ due to the effect of the Tth quantile of the unknown error distribution ε.
For the case of Parinacota province data, we modeled the information obtained for the arsenic concentration and the time (in hours respect of the reference point). The predictor variable was the time (in thousands of hours) elapsed

from the reference point 00:00 hours of January 1, 2014 until the day and hour at which the original sample was taken.
In order to compute the parameter estimation of the simple linear regression and quantile regression with $\tau = 0.75$, we consider samples generated with nonparametric and parametric bootstrap, then we have the following cases for the response variable.

1. **Nonparametric bootstrap**. In each monitoring site a new sample with 24 realizations are generated randomly acording to nonparametric bootstrap. Then, we compute the harmonic mean according to the expression (5.4). The estimations for the intercept (a) and for the slope (b) of the models are shown in Table 5.9. Furthermore, each regression fit are shown in Figure 5.11.

2. **Parametric bootstrap**. It is considered the estimation of the shape parameter α associated with a specific value of k, to generate a new sample with 24 realizations from Pareto distribution (5.5) associated with the concentration of arsenic, which are generated randomly using the inverse sampling method. The estimations obtained for the intercept and the slope of the models, are based on the mean, obtained for the intercepts and for the slopes of the all replicates. These mean values are shows in Table 5.10 and the regression fit are shown in the Figure 5.12.

The results recorded in Tables 5.9 and 5.10 and the adjusted straight line of Figures 5.12 and 5.13 suggest that the slope of the linear regression model of the arsenic concentration is zero. To verify this in section 5.2.5, the results of the $t-$test are entered.

Table 5.9: Estimated coefficients of linear and quantile regression model for arsenic concentration in the case of nonparametric bootstrap

Model	Intercept	Slope
Linear regression	0.2277	0.0004184
Quantile regression with $\tau = 0.75$	0.03482	-0.00022

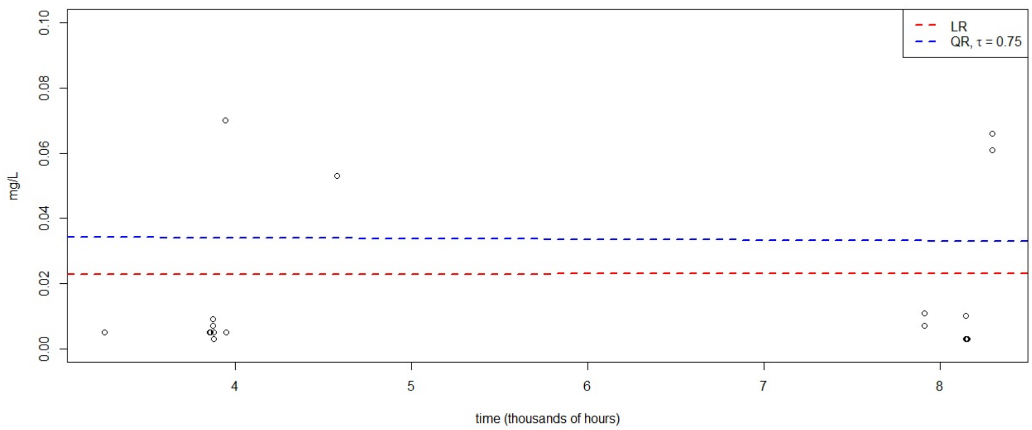

Figure 5.11: Scatterplot and linear and quantile regression fit of arsenic concentration vs. time

Table 5.10: Estimated coefficients of linear and quantile regression model for arsenic concentration in the case the parametric bootstrap

Model		Intercept	Slope
Linear regression	Hill	0.01450919	-0.0002223181
	t-Hill	0.01582100	-0.0004582841
	t-lgHill	0.02532391	-0.0001360984
Quantile regression with $\tau = 0.75$	Hill	0.01935514	-0.0006145886
	t-Hill	0.02075175	-0.0008088861
	t-lgHill	0.03961337	-0.0013331716

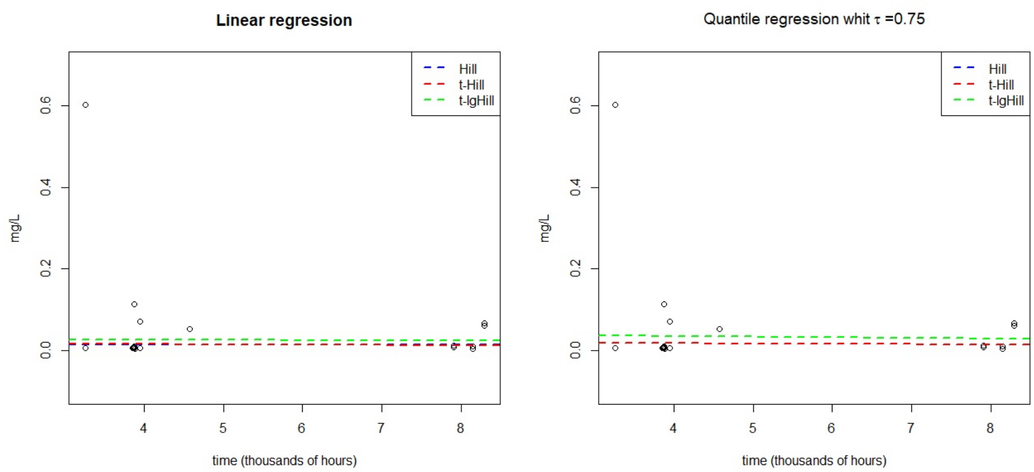

Figure 5.12: Linear and quantile regression fit of arsenic concentration vs. time

In the case of the Arica city data, we modeled the boron concentration in drinking water and the distance (in thousands of units) from the monitoring site to the pond fron where the drinking water is supplied. The parameters estimation of the simple linear regression model and the quantile model with $\tau = 0.75$, are estimated considering the two cases mentioned above.

1. **Nonparametric bootstrap**. A new sample with 13 realizations is generated in each monitoring site. The 13 realizations are generated randomly according to nonparametric bootstrap. The estimators for the intercept and slope of the models based on the harmonic mean obtained in each monitoring site and the distance. This estimates are recorded in the Table 5.11. The regression fits are shown in the Figure 5.13.

2. **Parametric bootstrap**. Considering the estimators of the shape parameter (see Table 5.8) for a particular value of the k, a new sample of 13 realizations of the random variable with Pareto distribution it is generated with the cdf given by the equation (5.5). associated with the concentration of boron, which are generated randomly using the inverse sampling method. The estimators of the intercept and the slope of the models, are based on the mean value of the intercept and the slope of all the replicates, this values are shown in Table 5.12. In addition, fitted regression models are displayed in Figure 5.14.

For the linear regression model of the boron concentration, the straight line adjusted in Figure 5.13 and the estimated value of the slope recorded in Table 5.11 suggest that the slope is zero. On the other hand, when we use the Hill, t-Hill and t-lgHill estimators of $\alpha = \frac{1}{\gamma}$, to generate measurements of the boron concentration, we obtain a slope close to zero, as shown in Table 5.12 and Figure 5.14. In section 5.2.5 we check whether the slope of the regression model is or is not zero.

Table 5.11: Estimated coefficients of linear and quantile regression model for boron concentration in the case of nonparametric bootstrap

Model	Intercept	Slope
Linear regression	3.03213806	0.02772321
Quantile regression with $\tau = 0.75$	3.5029599	0.1938831

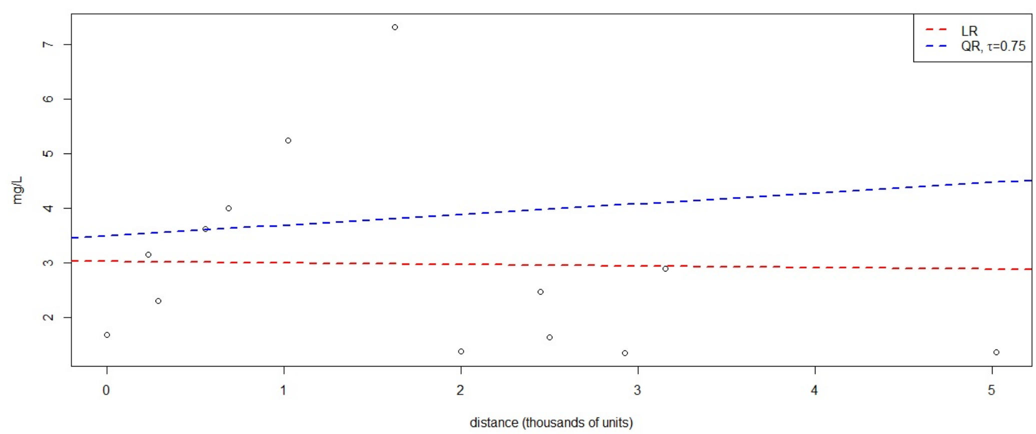

Figure 5.13: Scatterplot and linear and quantile regression fit of boron concentration vs. distance

Table 5.12: Estimated coefficients of linear and quantile regression model for boron concentration in the case of parametric bootstrap

Model		Intercept	Slope
Linear regression	Hill	2.6126198	-0.1777463
	t-Hill	2.9456053	-0.2209566
	t-lgHill	1.21025474	0.01059225
Quantile regression with $\tau = 0.75$	Hill	2.39498193	-0.01478547
	t-Hill	2.4205596	0.1279842
	t-lgHill	1.23102661	0.04812397

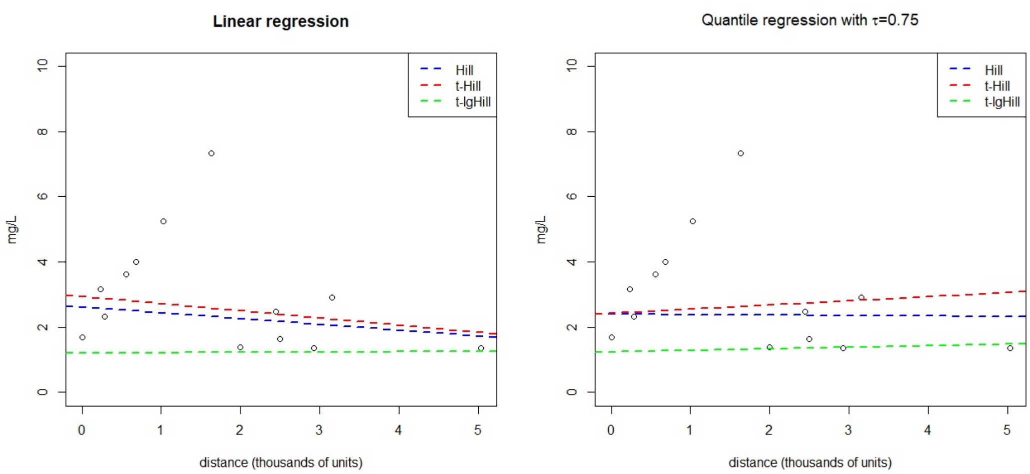

Figure 5.14: Linear and quantile regression of boron concentration vs. distance

5.2.5 t-test and confidence interval

In linear regression it is not common to reject the hypothesis of constant regression when applying the $t-$test. If we look at the data from the province of Parinacoa and the city of Arica then, to test the hypotheses that the regression slope is zero, the t statistic of the test has a value that does not reject the hypotheses of the slope of the regression equal to zero. Remember that the hypotheses of the test for this case are

$$H_0: \ b = 0 \qquad \text{v/s} \qquad H_1: \ b \neq 0, \tag{5.9}$$

where H_0 is rejected at level α if $|t| > t_{n-2,\frac{\alpha}{2}}$, with

$$t = \frac{\hat{b} - b}{\sqrt{\frac{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}}} \sim t_{n-2} \tag{5.10}$$

We can also construct a confidence interval for the slope of the linear regression model, in this case we use the nonparametric Delta method mentioned in Wasserman (2003), where the $(1-\alpha)100\%$ confidence interval is given by

$$T(\hat{F}_n) \pm z_{1-\frac{\alpha}{2}}\widehat{se}, \qquad \widehat{se} = \sqrt{\frac{\frac{1}{B}\sum_{i=1}^{B}\hat{L}^2(X_i)}{B}} \tag{5.11}$$

In this case $T(\hat{F}_n) = \hat{b}$ and $\hat{L}^2(X_i) = (\hat{b}_i - \overline{b})^2$, where B is the number of replicates, $\hat{b}_i$ is the estimated slope of the ith replicate and $\overline{b}$ the mean of all $\hat{b}_i$.

Results for Parinacota province data

In the Figure 5.12 it can be noticed that the fitted regression with simple linear regression and the quantile regression with $\tau = 0.75$ methods, had slopes close to zero, so tend to be constant functions.
When t-test is computed considering the realizations of the regression variable (X: time) and the dependend variable (Y: arsenic concentration) it is obtained $|t| = 1.023132$ and the upper quantile for $\alpha = 0.2$ and $n-2 = 22$ is 1.321, therefore the hypotesis that the slope of regression is zero is not rejected. The 95% of confidence interval obtained considering 20 replicates performed by the parametric bootstrap method are shown in Table 5.13, this intervals also indicate that the slope of regression is equal to zero. In addition in the Table 5.13 are shown the t statistics for the 20 replicates.

Table 5.13: t-values for the t-test and confidence limits for the arsenic concentration model

	$\hat{b}$	$\|t\|$	Lower limit	Upper limit
Hill	-0.0002223181	0.01756401	-0.001574587	0.00129950
t-Hill	-0.0004582841	0.03622943	-0.002186477	0.001269909
t-lgHill	-0.0001360984	0.0109571	-0.002882692	0.002610496

Results for Arica city data

According to Table 5.12 and Figure 5.14 in the t-lgHill estimation of the shape parameter α it can be noticed that fitted regression for simple linear regression and quantile regression whit $\tau = 0.75$ models, had slopes near to zero, meaning that the function could be constant.
The t-test for the slope of simple linear regression model with predictor variable the distance (X) and regression variable the boron concentration (Y) gives a value of $|t| = 1.803088$ and the upper quantile for $\alpha = 0.2$ and $n - 2 = 11$ is 1.36343. Then, this result indicate that the hypothesis of a zero slope in the regression is rejected. Table 5.14 shows the values of the t statistics for 20 replicates using parametric bootstrap, together with the limits of the 95% confidence interval. This interval indicate that the slope of the regression model can be zero.

Table 5.14: t-values for t-test and confidence limits for the boron concentration model

	$\hat{b}$	$\|t\|$	Lower limit	Upper limit
Hill	-0.1777463	0.427721	-0.5744093	0.2189166
t-Hill	-0.2209566	0.5417988	-0.5659523	0.1240391
t-lgHill	0.01059225	0.02046295	-0.0168825	0.0327431

Conclusion from $t-$test and from confidence intervals

According to the results obtained in the $t-$test and the confidence intervals determined, we can conclude that the regression slope is zero, i.e. we do not reject the hypothesis of constant regression.

5.2.6 Distribution of errors

So far we have focused on estimating for the coefficients of the simple linear regression model

$$Y = a + bX + \varepsilon \tag{5.12}$$

Theoretically, in this model it is assumed that errors are distributed according to a normal model. The statistical summary of Table 5.15 for the errors gives us little information, but when considering the histograms corresponding to the errors or residuals that occur in the estimation of the arsenic or boron models shown in Figure 5.15, we can easily see that this assumption is not totally true. Then, the question to answer is how are these errors distributed?

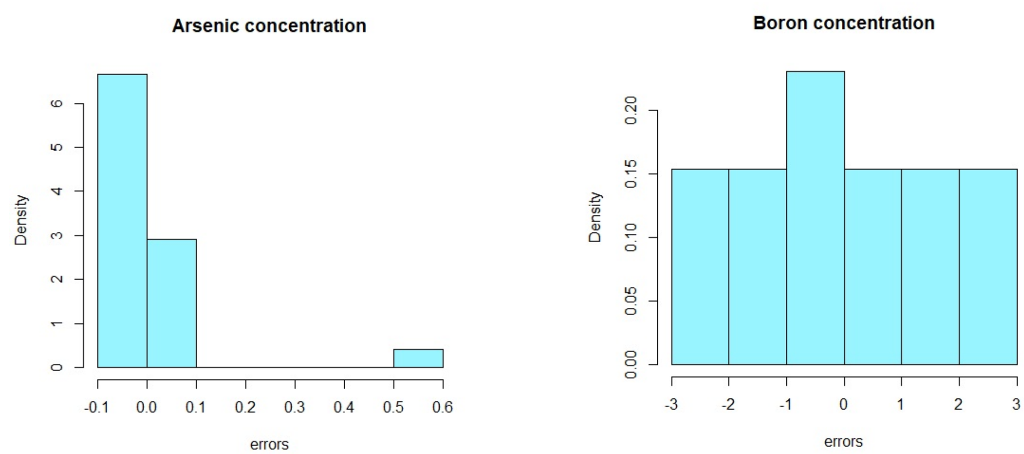

Figure 5.15: Histograms of errors for the models of arsenic concentration (*left chart*) and boron concentration (*right chart*)

Table 5.15: Statistical summary for the errors of the models of arsenic and boron concentration

Statistic	Arsenic	Boron
Minimum	-0.08062	-2.13386
First quantile	-0.05788	-1.16891
Median	-0.02133	-0.04958
Third quantile	0.01906	1.10325
Maximum	0.50156	2.46285
Mean	0.00000	0.00000
Standar deviation	0.1154028	1.549972

In order to answer the question, we consider as a hypothetical distribution the Beta distribution function with parameters $\alpha > 0$ and $\beta > 0$, whose density is given by

$$f(x, \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathbf{B}(\alpha, \beta)}, \tag{5.13}$$

where $\mathbf{B}(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ and $x \in (0, 1)$. For this variable, we have the moments

$$E[X] = \frac{\alpha}{\alpha+\beta} \qquad Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}, \tag{5.14}$$

this moments allow to estimate the parameters. Previously because $x \in (0, 1)$ is it necessary to perform a transformation of the errors such that

$$\text{error}^* = c + d \cdot \text{error}, \quad c, d \in \mathbb{R}.$$

Results for the ordinary least squares method

For the Parinacota province data and the simple linear regression model, where the independent variable is X, the time and the response variable is Y the arsenic concentration in water samples of the 24 monitoring sites, or which we consider the following cases:

1. Y: Measurements of arsenic concentration in the 24 monitoring sites for the Parinacota province data.

2. Y: Harmonic mean of arsenic concentrations generated in each monitoring site with nonparametric bootstrap.

The estimates obtained using the ordinary least squares method in case 1) are $\hat{Y} = 0.24658045 - 0.011155019X$. In the case 2) $\hat{a} = 0.0458957247$ and $\hat{b} = 0.0002014336$. Figure 5.16 corresponds to the histogram of the errors (error*) together with the curve representing the hypothetical pdf for cases 1) and 2). In addition, the corresponding estimates of the parameters of the beta distribution are recorded in Table 5.16.

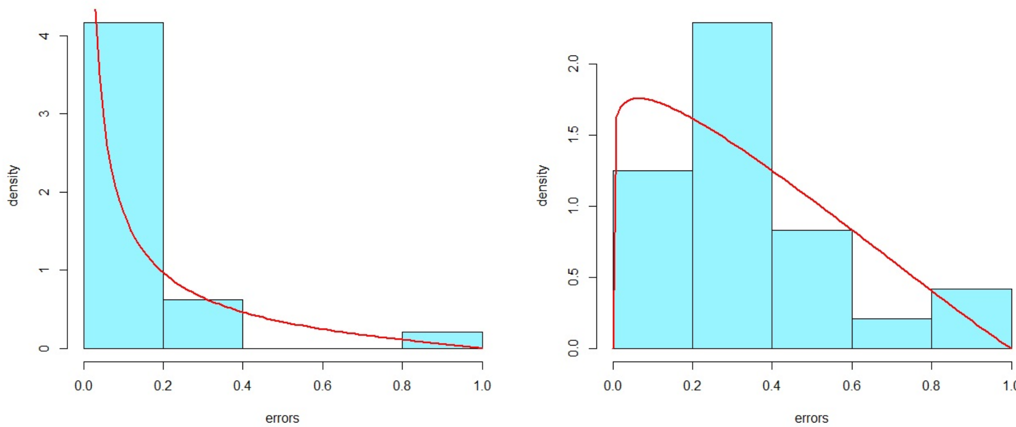

Figure 5.16: Histogram of the 24 errors for the arsenic concentration model for case 1 (*left chart*) and for case 2 (*right chart*)

Table 5.16: Estimates of the parameters of the Beta distribution of errors of the arsenic concentration model

Cases	$\hat{\alpha}$	$\hat{\beta}$
1	0.3002239	1.86787
2	1.073885	2.07299

In the case of Arica city data and the simple linear regression model, where the independent variable is X distance from the pond that provides the water to the monitoring site and the response variable Y the boron concentration recorded in the water samples of the 13 monitoring sites considered in the study. The response variable considers the following cases.

1. Y: Harmonic mean of the boron concentration obtained from concentrations of boron recorded in the water samples of each monitoring site for the Arica city data.

2. Y: Harmonic mean of boron concentrations generated in each monitoring site with nonparametric bootstrap.

The estimated model for these data considering Y of case 1) is given by $\hat{Y} = 1.9683675 + 0.5745863X$. For case 2) we have $\hat{a} = 3.38192356$ and $\hat{b} = -0.01412159$. The histogram of the errors (error*) is shown in Figure 5.17 and the estimates of the parameters of the hypothetical distribution are shown in Table 5.17.

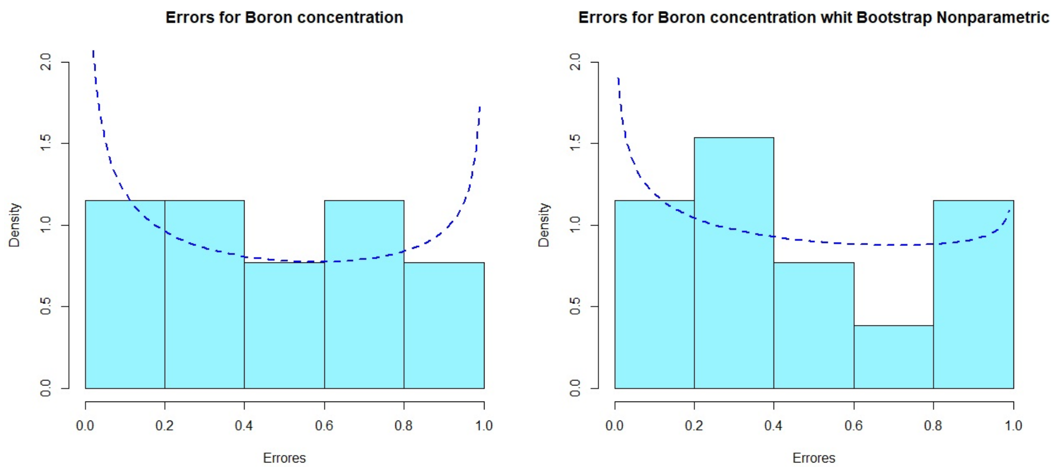

Figure 5.17: Histogram of 13 errors of the boron concentration model for case 1 (*left chart*) and case 2 (*right chart*)

Table 5.17: Estimates of the parameters of the Beta distribution of errors of the boron concentration model

	$\hat{\alpha}$	$\hat{\beta}$
1	0.6358964	0.7339391
2	0.79411	0.9154096

5.2.7 Score Regression

We consider the Score Regression, where we assume that the distribution F of random variable ξ is a member of the parametric family $\mathcal{F}_{\mathcal{X}} = \{\mathcal{F}_0 \in \Theta \subseteq \mathbb{R}\}$ whit support $\mathcal{X} = (0, \infty)$. The linear regression problem can be formulated as a task to find a y b in (5.7), with

$$\varepsilon = \xi - \xi^*, \tag{5.15}$$

where $\xi^* = \xi^*(\theta)$ is the score mean of the distribution function F. Hence, the elements ε_i of the vector ε can have both signs, whereas the elements ξ_i of the vector ξ are positive.
After setting initial values of a and b of the model (5.7), the solution iterative to the regression problem consist in two steps.

Step 1. Estimation of the score mean $\varepsilon^*(\theta)$ by estimating parameter θ of F_θ. In principle any "good"estimate $\hat{\theta}$ of θ can be used. In cases if heavy-tailed distributions we used the score moment estimates (Fabián, 2010 and Fabián, 2016) the solution of equations

$$\frac{1}{n}\sum_{i=1}^{n} S_F^k(x_i;\theta), \qquad k = 1, \ldots, m. \tag{5.16}$$

Score function of heavy-tailed distributions are bounded, so that the score moment estimates are slightly non-efficient, but robust with respect to outliers.

Remark 5.2.1. *This step, which to a certain extent corresponds to the estimation of the scale parameter in the method of least squares (LS method), provides* $\xi^* = \xi(\theta)$ *and residuals*

$$\varepsilon_i = \xi_i - \xi^* = y_i - a - bx_i \tag{5.17}$$

Step 2. The generalization of the LS method is the regression with parametric function *Fisher information* of random variable ξ. Then, we minimize

$$I_F(\hat{\theta}) = \sum_{i=1}^{n} S_F^2(\xi_i, \hat{\theta}), \tag{5.18}$$

with respect to a and b, where $\xi_i = y_i - (a + bx_i) + \hat{\xi}^*(\hat{\theta})$ and S_F is the score funtion distribution (sfd) of F_θ. The goal is to find the regression line with minimum information of residuals with respect to the score mean of ξ. By differentiating (5.18) with respect to a and b, we obtain a system of equations

$$\frac{\partial}{\partial a} I_F(\hat{\theta}) = \sum_{i=1}^{n} \psi_F(\xi_i) = 0 \tag{5.19}$$

$$\frac{\partial}{\partial b} I_F(\hat{\theta}) = \sum_{i=1}^{n} \psi_F(\xi_i) x_i = 0 \tag{5.20}$$

with inference function

$$\psi_F(\xi) = S_F(\xi; \hat{\theta}) S'_F(\xi; \hat{\theta}), \tag{5.21}$$

where S'_F is the derivative of S_F with respect to the corresponding variable.
Table 1 in Fabián (2016) shows functions $T_F(x)$, score means x^* and score regression inference function. For example, in the case of the Beta distribution with $x^* = \frac{\alpha}{\beta}$ the sfd is $S_F(x, \hat{x}^*) = \frac{\alpha^2}{\beta} \frac{x - \hat{x}^*}{x+1}$, such that $S'_F(x, \hat{x}^*) = \frac{\alpha^2}{\beta} \frac{\hat{x}^* + 1}{(x+1)^2}$ and the inference function is $\psi_F(x) = \frac{1}{(x+1)^2}$.

Iterative process and asymptotic distribution of estimates

Since the focus is to determine the distribution function of the ε_i, we perform an iterative process, which we detail in the following steps.

Step 1. Compute $\varepsilon_{ij} = y_i - (\hat{a} + \hat{b} x_i)$, and $\xi_{ij} = \varepsilon_{ij} + 1.005 * |\varepsilon_{ij}|$, with $j = 0$.

Step 2. Derive $\hat{\xi}_j^*$ from the equation of the first score moment

$$\sum_{i=1}^{n} \frac{\xi_{ij} - \hat{\xi}_j^*}{\xi_{ij} + 1} = 0,$$

in this case

$$\hat{\xi}_j^* = \frac{\sum_{i=1}^{n} \frac{\xi_{ij}}{\xi_{ij}+1}}{\sum_{i=1}^{n} \frac{1}{\xi_{ij}+1}} \tag{5.22}$$

Step 3. Solve the system of equations, whose unknowns are ξ_{ij},

$$\sum_{i=1}^{n} \frac{\xi_{ij} - \hat{\xi}_j^*}{(\xi_{ij}+1)^3} = 0$$
$$\sum_{i=1}^{n} \frac{\xi_{ij} - \hat{\xi}_j^*}{(\xi_{ij}+1)^3} x_i = 0 \tag{5.23}$$

Step 4. Considering the values of ξ_{ij} and the expression (5.22) we determine $\hat{\xi}_{j+1}^*$.

Step 5. Compute $|\hat{\xi}_{j+1}^* - \hat{\xi}_j^*|$, if this value is less than a small fixed constant, the process is terminated, otherwise steps 3, 4 and 5 are repeated successively.

Remark 5.2.2. *In the system of equations (5.23) there exist n unknowns, which means that the set of solutions is linearly dependent. To solve this problem, we will consider three unknowns,*

$$\begin{aligned} \xi_{ij} &= \xi_{1j}, \quad i = 1, \dots, k \\ \xi_{ij} &= \xi_{2j}, \quad i = k+1, \dots, m \\ \xi_{ij} &= \xi_{3j}, \quad i = m+1, \dots, n \end{aligned} \tag{5.24}$$

with $1 < k < m < n$*, then the system of equations (5.23) is as follows*

$$k\frac{\xi_{1j} - \hat{\xi}_j^*}{(\xi_{1j}+1)^3} + (m-k)\frac{\xi_{2j} - \hat{\xi}_j^*}{(\xi_{2j}+1)^3} + (n-m)\frac{\xi_{3j} - \hat{\xi}_j^*}{(\xi_{3j}+1)^3} = 0$$
$$k\frac{\xi_{ij} - \hat{\xi}_j^*}{(\xi_{ij}+1)^3} x_i + (m-k)\frac{\xi_{ij} - \hat{\xi}_j^*}{(\xi_{ij}+1)^3} x_i + (n-m)\frac{\xi_{ij} - \hat{\xi}_j^*}{(\xi_{ij}+1)^3} x_i = 0 \tag{5.25}$$

In order to make the solution of this system of equations more viable, we made the following substitution.

$$\frac{\xi_{1j} - \hat{\xi}_j^*}{(\xi_{1j} + 1)^3} = s \tag{5.26}$$

$$\frac{\xi_{2j} - \hat{\xi}_j^*}{(\xi_{2j} + 1)^3} = t \tag{5.27}$$

$$\frac{\xi_{3j} - \hat{\xi}_j^*}{(\xi_{3j} + 1)^3} = u \tag{5.28}$$

Consequently, the system of equations (5.25) is transformed into a linear system of equations, whose solutions are given by

$$s = \frac{(m-k)\sum_{i=m+1}^{n} x_i - (n-m)\sum_{i=k+1}^{m} x_i}{k\sum_{i=k+1}^{m} x_i - (m-k)\sum_{i=1}^{k} x_i} u := g_1(u) \tag{5.29}$$

$$t = \frac{k\sum_{i=m+1}^{n} x_i - (n-m)\sum_{i=1}^{k} x_i}{(m-k)\sum_{i=1}^{k} x_i - k\sum_{i=k+1}^{m} x_i} u := g_2(u) \tag{5.30}$$

By replacing s and t in equations (5.26) and (5.27) respectively, the following polynomials of degree 3 are obtained.

$$g_1(u)\xi_{1j}^3 + 3g_1(u)\xi_{1j}^2 + (3g_1(u) - 1)\xi_{1j} + g_1(u) + \hat{\xi}_j^* = 0 \tag{5.31}$$

$$g_2(u)\xi_{2j}^3 + 3g_2(u)\xi_{2j}^2 + (3g_2(u) - 1)\xi_{2j} + g_2(u) + \hat{\xi}_j^* = 0 \tag{5.32}$$

As theoretically a polynomial of degree 3 can have:

a. a real root of multiplicity 3.

b. A real root and two complex.

c. Three different real roots.

d. Two real roots, one simple and one double.

We consider the positive real root closest to zero.

Results for Parinacota province data

For the data of Parinacota we consider three groups of size 8 ($k = 8$, $m = 16$ and $n = 24$), for the partition of the ξ_{ij} and $\hat{\hat{\xi}}_j^* = \text{median}\{\xi_{i(j-1),\ i=1,\dots n}\}$, we obtain the distribution of the errors shown in the plot to the right of Figure 5.18 and the graph on the left of Figure 5.18 shows the curve representing the values of $\hat{\xi}_j^*$ obtained in each iteration until that $|\hat{\xi}_{j+1}^* - \hat{\xi}_j^*| < 7 * 10^-6$. In addition, the Table 5.18 shows the values of the estimates of the parameters of the Beta distribution function.

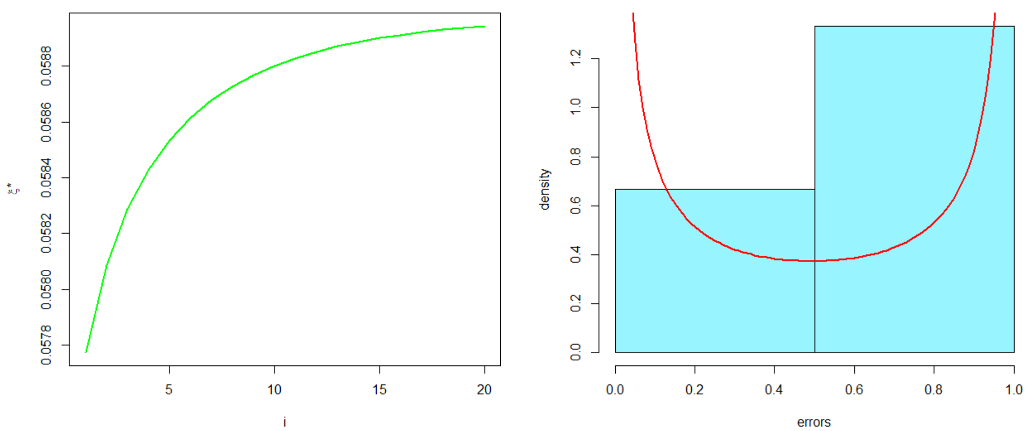

Figure 5.18: Distributions of errors for Parinacota province data

Table 5.18: Results of the estimation for Parinacota province data

Parameter	Estimator
α	0.3242644
β	0.2054242

Results for Arica city data

In the case of Arica city data we consider two groups of size 5 and a group of size 3 for the partition of the ξ_{ij}, then $k = 5$, $m = 10$, $n = 13$. Here

$\hat{\xi}_j^* = \max\{\xi_{i(j-1)}, i = 1, \ldots, n\}$ and the tolerance is $|\hat{\xi}_{j+1}^* - \hat{\xi}_j^*| < 8.75 * 10^{-5}$. Table 5.19 records the values of the estimates of the parameters of the Beta distribution function. On the other hand, Figure 5.19 shows in the graph on the left the curve representing the values of $\hat{\xi}_j^*$ in each iteration and in the plot on the right the corresponding distribution of the errors is described together with the curve representing the hypothetical distribution.

Table 5.19: Results of the estimation for Arica city data

Parameter	Estimator
α	0.271412
β	0.9673527

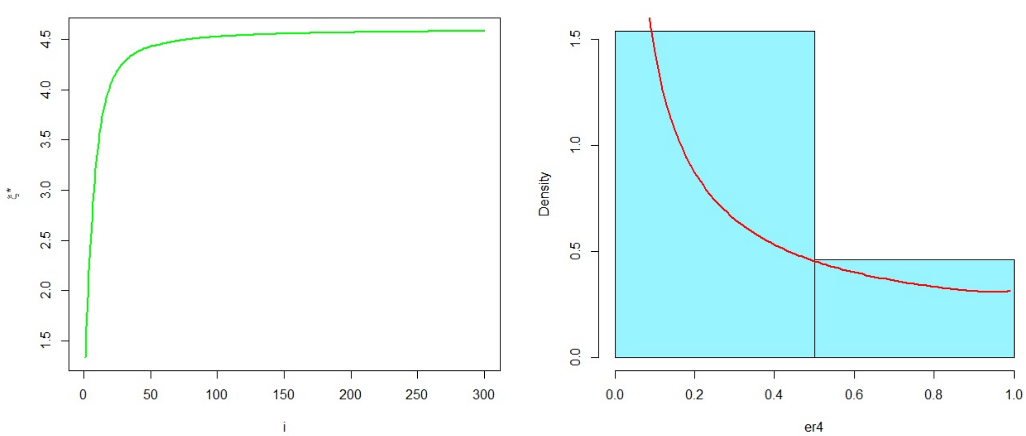

Figure 5.19: Distribution of errors for Arica city data

5.2.8 Comparison

With the purpose of comparing the score regression and ordinary least squares methods, we performed a simulation experiment, where the coefficients of model (5.7) are computed, under the assumption that $\varepsilon \sim \text{Beta}(\hat{\alpha}, \hat{\beta})$, the experiment was repeated 500 times with randomly generated errors. For Parinacota province data and Arica city data, the plot of the estimates of coeficients it is shown in the Figure 5.20 and Figure 5.21 respectively. A summary of this estimates can be observed in the Table 5.20 for Parinacota province data and Table 5.21 for Arica city data. In adition, the Figures 5.22 and 5.23 show the straight lines obtained in one of the repetitions of the simulation experiment carried out considering Parinacota province data and Arica city data respectively.

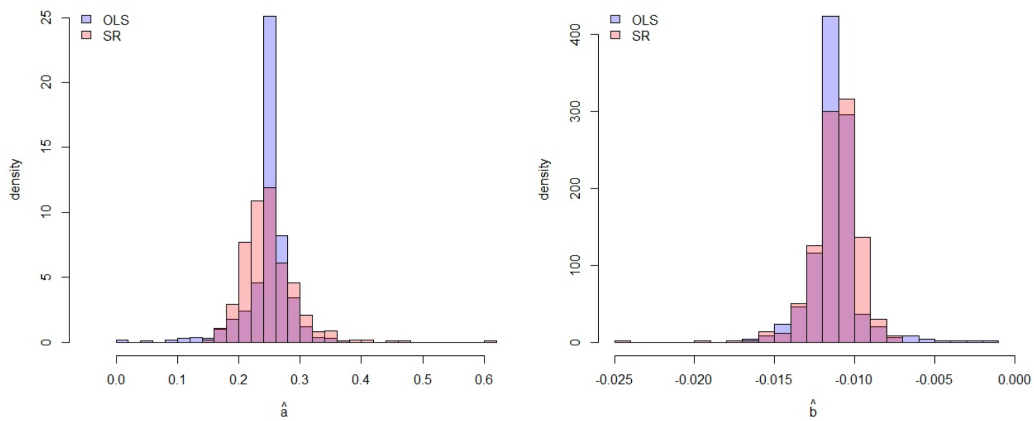

Figure 5.20: Histograms of the estimates of the intercept $\hat{a}$ (*left chart*) and of the slope $\hat{b}$ (*right chart*) in 500 repetitions of the experiment for Parinacota province dada

Table 5.20: Average estimates of regression coefficients and its corresponding Mean Square Errors (MSEs) for Parinacota province data

Method	$\hat{a}$	$\hat{b}$	MSE($\hat{a}$)	MSE($\hat{b}$)
OLS	0.2465546	-0.01121905	0.001437416	0.00000224898
SR	0.2488893	-0.01116552	0.001963972	0.000002261362

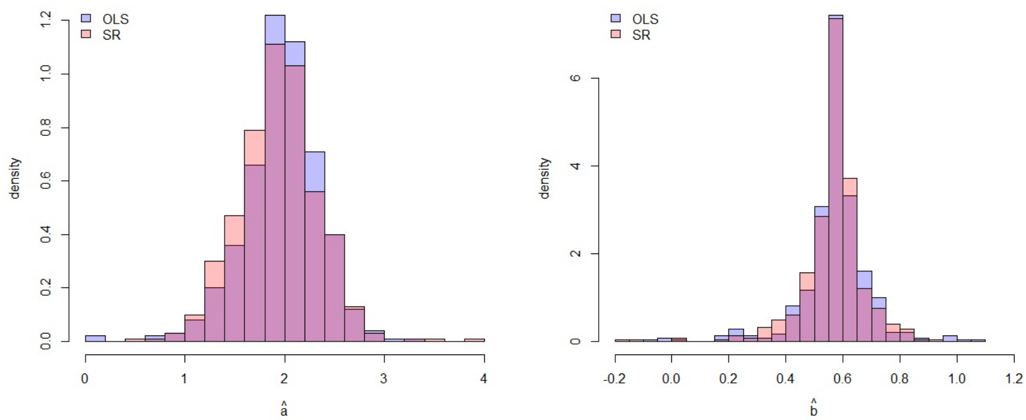

Figure 5.21: Histograms of the estimates of the intercept $\hat{a}$ (*left chart*) and of the slope $\hat{b}$ (*right chart*) in 500 repetitions of the experiment for Arica city data

Table 5.21: Average estimates of regression coefficients and its corresponding Mean Square Errors (MSEs) for Arica city data

Method	$\hat{a}$	$\hat{b}$	MSE($\hat{a}$)	MSE($\hat{b}$)
OLS	1.976411	0.573059	0.1487761	0.01392477
SR	1.935672	0.5669903	0.1590618	0.01218666

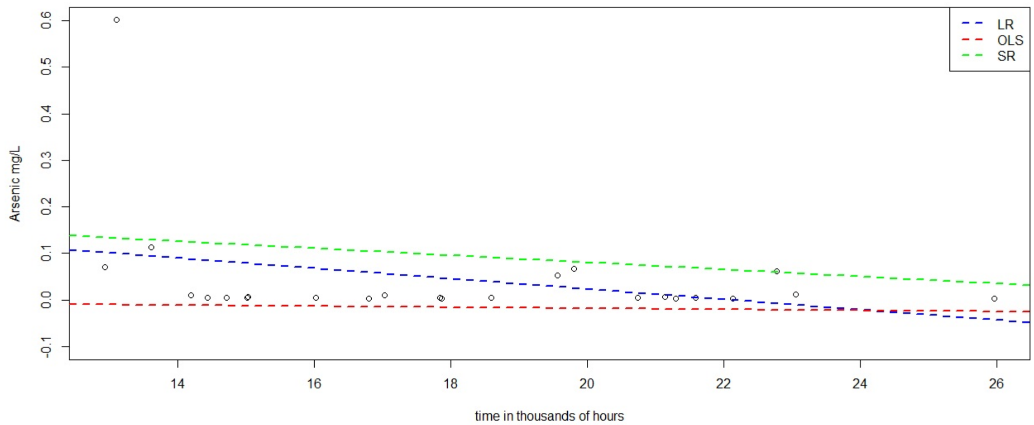

Figure 5.22: Scatterplot and OLS and SR for Parinacota province data

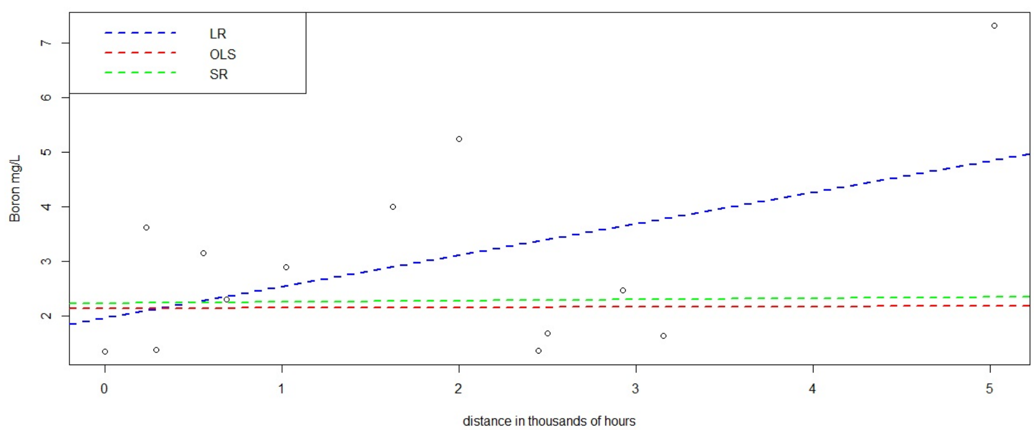

Figure 5.23: Scatterplot and OLS and SR for Arica city data

Conclusions from the score regression

Under the linear score regression that models the conditional score mean of a response Y (concentration of arsenic or boron) given a set of predictors as a linear function of X (time or distance). Assuming that the distribution of the error term belongs to a parametric family that includes skewed and heavy-talied distributions. We can conclude according to the adjusted straight line (green) of Figure 5.22 and of Figure 5.23 that the slope of model in (5.7) is zero. From Figure 5.20 and Figure 5.21 and from Table 5.20 and Table 5.21 we can infer that the regression score shown to be better than regression methods based on assumption of symmetric distribution for the error term.

5.2.9 Results of the RT class tests

In order to verify that the errors in the simple linear regression model do not follow a normal distribution function, we apply the normality tests JB, JBU and RJB mentioned in section 3.5 of Chapter 3. The results of these tests are shown in the Table 5.22 for the case of arsenic concentration and Table 5.23 for the case of boron concentration. In both tables we observe the results for the case of the errors when performing the estimation with the ordinary least square method (OLS) and when considering the score regression (SR).

Table 5.22: Test statistics of the errors in the model of arsenic concentration

Test	OLS		SR	
	Statistic	p-value	Statistic	p-value
JB	226.6455	0	2.368688	0.3059468
JBU	355.586	0	4.213574	0.1216281
RJB	3593.747	0	1.081866	0.5822048

Table 5.23: Test statistics of errors in the model of boron concentration

Test	OLS		SR	
	Statistic	p-value	Statistic	p-value
JB	0.8611352	0.643671	1.042992	0.5936318
JBU	2.54892	0.2795819	2.623266	0.2693887
RJB	0.6369571	0.7272547	0.533832	0.7657374

According to what is observed in Table 5.22, the normality hypotheses is rejected only in the case of performing the estimation with the ordinary least square method (OLS). For the score regression (SR), the p-values indicate that the normality assumption is not rejected. On the other hand, the p-values recorded in Table 5.23 are greater than a significance level $\alpha < 0.1$, Therefore we can not reject the hypotheses that the errors produced in the linear regression model of the boron concentration are distributed according to a normal model.

5.3 Dependence on random variables

We previously mentioned the study conducted by *SEREMI de Salud of Arica y Parinacota*, in which 24 monitoring sites were considered, where measurements of the level of arsenic concentration in samples of drinking water were made, which is provided by the *Sistema de Agua Potable Rural (APR)*, that supply the rural areas of Chile. The 24 monitoring sites are shown in Table 5.3.

In the previous section a model for arsenic concentration was addressed, under the assumption of independence, but the recorded measurements of the arsenic concentration X in mg/L at each monitoring site are not necessarily realizations of independent random variables . This problem was addressed by Filus et al. (2018), where the authors estimate the joint function of two random variables U_1 and U_2, whose realizations correspond to

$$u_{1,i} = x_i, \qquad i = 1, \cdots, n-1 \tag{5.33}$$

$$u_{2,j} = x_j \qquad j = 2, \cdots, n \tag{5.34}$$

To estimate the joint distribution function F, we followed example 4.2.2 of the Chapter 4, which is presented in Marshal and Olkin (1983), where the survival function $\overline{F} = 1 - F$, is gigen by

$$\overline{F}(\mathbf{u}) = \overline{H}_1(u_1)\overline{H}_2(u_2)\overline{H}_3(\max(u_1, u_2)), \qquad -\infty \leq u_1, u_2 \leq \infty, \tag{5.35}$$

with $\overline{H}_1(z) = \exp(-\lambda_1 z)$, $\overline{H}_2(z) = \exp(-\lambda_2 z)$, and $\overline{H}_3 = \exp(-\lambda_3 z)$ for $z > 0$ and for some λ_1, λ_2, $\lambda_3 \geq 0$ such that $\lambda_1 + \lambda_3 > 0$, $\lambda_2 + \lambda_3 > 0$. According to example 5.5.5 of Galambos (1978) the marginal distributions functions $F_1, F_2 \in D_{min}(\Psi^*)$, then by proposition 4.1.6 of the Chapter 4, corresponding to proposition 4.2 of Marshall and Olkin (1983) $F \in D_{min}(G)$, where

$$\overline{G}(u_1, u_2) = \exp\left\{-\left[\lambda_1 u_1 + \lambda_2 \frac{\lambda_+ \lambda_3}{\lambda_2 + \lambda_3} u_2 + \lambda_3 \max\left(u_1, \frac{\lambda_1 + \lambda_3}{\lambda_2 + \lambda_3} u_2\right)\right]\right\}, \quad u_1, u_2 \geq 0. \tag{5.36}$$

Thus U_1, U_2 have a bivariate extreme value distribution, then by proposition 4.2.1 of the Chapter 4, corresponding to proposition 5.1 of Marshall and Olkin (1983) U_1, U_2 are associated.
We estimate the parameters λ_1, λ_2 and λ_3 considering what was suggested by Li et al. (2011), who define for $k = 1, 2$

$$\mathbf{b}_k = \begin{bmatrix} \log(\lambda_k + \lambda_3) \\ \beta_k \end{bmatrix}, \quad \mathbf{y}_k = \begin{bmatrix} \log(-\log(\hat{S}_k(u_{k1}))) \\ \cdot \\ \cdot \\ \cdot \\ log(-\log(\hat{S}_k(u_{kn}))) \end{bmatrix}, \quad \mathbf{X}_k = \begin{bmatrix} 1 & \log(u_{k1}) \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & \log(u_{kn}) \end{bmatrix} \tag{5.37}$$

We estimate the survival function S using the empirical cumulative function,

$$S_k(u_{ik}) = 1 - S_n(u_{ik}). \tag{5.38}$$

The parameters involved in (5.35) and (5.36) are estimated by fitting the data to the linear model

$$\mathbf{y}_k = \mathbf{X}_k \mathbf{b}_k, \tag{5.39}$$

Then β_k and $\lambda_k + \lambda_3$ are estimated by the ordinary least squares method.

5.3.1 Data

Considering the 24 arsenic measurements made in the 24 monitoring sites (see Table 5.3), two groups of size 23 are formed, each group corresponds to the realizations of two variables U_1 and U_2. Table 5.24 shows the statistical summary for these variables.

Table 5.24: Statistical summary for the two random variables U_1 and U_2

Statistic	U_1	U_2
Minimun	0.003	0.003
First quartile	0.005	0.005
Median	0.005	0.005
Third quartile	0.0375	0.0375
Maximum	0.602	0.602
Mean	0.0423	0.0423
Standar deviation	0.1154096	0.1154096

5.3.2 Results

Table 5.25 shows the ordinary least squares estimators for the parameters involved in the model (5.36), and the Figure 5.24 shows the graph of the estimated joint distribution function and the corresponding survival function.

Table 5.25: Ordinary least squares estimators of the parameters of the joint bivariate distribution function

k	$\hat{\lambda}_k + \hat{\lambda}_3$	$\hat{\beta}_k$
1	15.91007	0.6399636
2	15.91007	0.6399636

In Table 5.25 we see that,

$$\hat{\lambda}_1 + \hat{\lambda}_3 = \hat{\lambda}_2 + \hat{\lambda}_3,$$

then $\hat{\lambda}_1 = \hat{\lambda}_2$. Considering $\hat{\lambda}_1 = \hat{\lambda}_2 = 0$, we have $\hat{\lambda}_3 = 15.91007$, therefore the estimated joint distribution $G(u_1, u_2)$ is obtained from the estimation of $\overline{G}(u_1, u_2) = \exp\left\{-\hat{\lambda}_3 \max(u_1, u_2)\right\}$.

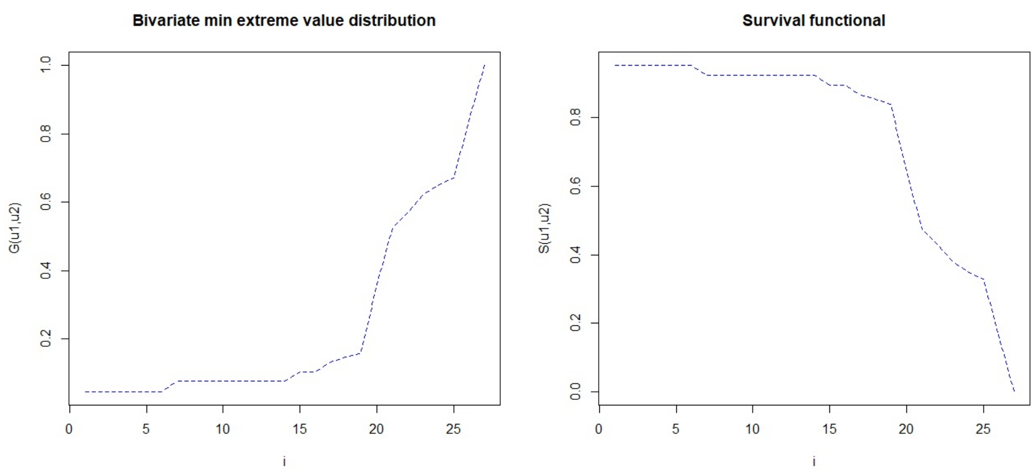

Figure 5.24: Graphs of bivariate distribution function $G(u_1, u_2)$ (*left chart*) and survival distribution $\overline{G}(u_1, U_2) = S(u_1, u_2)$ (*right chart*)

References

[1] M. Abramowitz, and I. A. Stegun. *Handbook of Mathematical Functions*. National Bureau of Standards, Applied Mathematics Series **55**, U.S. Government Printing Office, Washington D.C., 1964. (reprinted by Dover Publications, 1970).

[2] M. Ali, N. N. Mikhail, and M. S. Haq. A class of bivariate distributions including the bivariate logistic. *Journal of multivariate analysis, 8*(3), 405-412, 1978.

[3] B. C. Arnold. *Pareto Distributions*. Fairland, Maryland: International Cooperative Publishing House, 1983.

[4] B. C. Arnold, and K. Austin. Truncated Pareto distributions: Frexible, tractable and familiar. Thechnical Report 150. Department of Statistics, University of California, Riverside. 1987.

[5] M. A. Baxter. Minimum variance unbiesed estimation of the parameters of the Pareto distribution. *Metrica 27*(1), 133-138, 1980.

[6] J. Beirlant, G. Dierckx, Y. Goegebeur, and G. Matthys. Tail index estimation and exponential regression model. *Extremes, 2*(2), 177-200, 1999.

[7] J. Beirlant, Y. Goegebeur, J. Segers, and J. L. Teugels. *Statistics of extremes: Theory and applications*. John Wiley & Sons, 2006.

[8] J. Beran, D. Schell, and M. Stehlík. The harmonic moment tail index estimator: asymptotic distribution and robustness. *Annals of the Institute of Statistical Mathematics, 66*(1), 193-220, 2014.

[9] S. M. Berman. Convergence to bivariate limiting extreme value distributions. *Annals of the Institute of Statistical Mathematics, 13*(1), 217-223. 1961.

[10] G. K. Bhattacharyya, and R. A. Johnson. *Maximum Likelihood Estimation and Hypothesis Testing in the Bivariate Exponential Model of Marshall and Olkin* (No. UWIS-DS-71-276). WISCONSIN UNIV-MADISON DEPT OF STATISTICS, 1971.

[11] P. Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.

[12] M. Brilhante, M. Gomes, D. Pestana. A simple generalization of the Hill estimator. *Comput. Stat. Data Anal, 57*, 518-535, 2013.

[13] G. Casella, and R. L. Berger. *Statistical inference*. Pacific Grove, C.A: Duxbury, 2002.

[14] B. Castillo, and G. Venegas. Impacto y consecuencias del arsénico en la salud y el medio ambiente en el norte de Chile. *Revista Interamericana de Ambiente y Turismo-Riat, 6*, 53-60, 2010.

[15] R. Cirilo. *The Economics of Vilfredo Pareto*. Frank Cass, London, 1979.

[16] D. Cline. Infinite series of random variables with regularly varying tails. *Tech. Rep No 83-24. Institute of Applied Mathematics and Statistics, University of British Columbia*, 1983.

[17] Contraloria General de la República de Chile. Informe Final Superintendencia de Servicios Sanitarios, 2015. URL http://www.contraloria.cl/SicaProd/SICAv3-BIFAPortalCGR/faces/.

[18] S. Cortes, E. Reynaga-Delgado, A. M. Sancha, and C. Ferreccio. Boron exposure assessment using drinking water and urine in the North of Chile. *Science of the Total Environment, 410*, 96-101, 2011.

[19] D. M. Dabrowska. Kaplan-Meier estimate on the plane. *The Annals of Statistics, 16*(4), 1475-1489, 1988.

[20] D. M. Dabrowska. Kaplan-Meier estimate on the plane: weak convergence, LIL, and the bootstrap. *Journal of Multivariate analysis, 29*(2), 308-325, 1989.

[21] L. De Haan. *On Regular Variation and its Applications to the Weak Convergence of Sample Extremes*. PhD thesis, Mathematical Centre tract, 32, Amsterdann, 1970.

[22] L. De Haan, and U. Stadtmüller. Generalized regular variation of second order. *Journal of the Australian Mathematical Society, 61*(3), 381-395, 1996.

[23] L. De Haan, and S. Resnick. On asymptotic normality of the Hill estimator. *Sthochastic Models 14*(4), 849-866, 1998.

[24] L. De Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering, Springer, New York, 2006.

[25] T. De Oliveira. Structure theory of bivariate estremes extensions. *Estudos de Matemática, Estadistica e Econometria, 7*, 165-195, 1962/1663.

[26] P. Deheuvels, E. Haeusler, and D. M. Mason. Almost sure convergence of the Hill estimator. In *Mathematicsl Proccedings of the Cambridge Philosophical Society*. Cambridge University Press, p 371-38, 1998.

[27] A. L. Dekkers, and L. De Haan. On the estimation of the extreme-value index and large quantile estimation. *The Annals of Statistics, 17*(4), 1795-1832, 1989.

[28] H. Drees. Optimal rates convergence for estimates of the extreme value index. *Annals of Statistics, 26*, 434-448, 1998.

[29] H. Drees, L. De Haan, and S. Resnick. How to make a Hill plot. *Annals of Statistics*, 254-274, 2000.

[30] F. Y. Edgeworth. On the probable errors of frequency-constants. *Journal of the Royal Statistical Society, 71*(2), 381-397, 1908a.

[31] F. Y. Edgeworth, On the probable errors of frequency-constants (contd.). *Journal of the Royal Statistical Society, 71*(4), 651-678, 1908b.

[32] B. Efron. *The jackknife, the bootstrap, and other resampling plans* (Vol. 38). Siam, 1982.

[33] B. Epstein, and M. Sobel. Life testing. *Journal of the American Statistical Association 48*, 486-502, 1953.

[34] J. D. Esary, F. Proschan, and D. W. Walkup. Association of random variables, with applications. *The Annals of Mathematical Statistics*, 1466-1474, 1967.

[35] Z. Fabián. Induced cores and their use in robust parametric estimation. *Communications in Statistics-Theory and Methods, 30*(3), 537-555, 2001.

[36] Z. Fabián. Estimation of simple characteristics of samples from skewed and heavy-tailed distributions. In *Recent Advances in Stochastic Modeling and Data Analysis*, p. 43-50, 2007.

[37] Z. Fabián. Score Moment Estimators. In *Lechavallier, Y., Saporta, G. (eds.) Proceedings of Conference COMPSTAT'2010. Springer, Physica-Verlag*, 2010.

[38] Z. Fabián. Score function of distribution and revival of the moment method. *Communications in Statistics-Theory and Methods, 45*(4), 1118-1136, 2016

[39] W. Feller. *Introduction to Probability Theory and its Aplications, Vol 2*, (2nd ed.). Wiley, New York, 1971.

[40] A. Feuerverger, and P. Hall. Estimating a tail exponent by modelling departure from a Pareto distribution. *The Annals of Statistics, 27*(2), 760-781, 1999.

[41] J. Filus, L. Filus, Y. Lu, P. Jordanova, B. C. Arnold, L. Núñez-Soza, S. Stehlíková, and M. Stehlík. On parameter dependence and related topics: The impact of Jerzy Filus from genesis to recent developments (with discussion). Submitted to *CRC Book*, 2018.

[42] R. A. Fisher. *Statistical methods for research workers*. Genesis Publishing Pvt Ltd, 1925.

[43] R. Fisher, and L. H. C. Tippet. Limiting form of the frecuency distributions of the largest or smaller member of a sample. *Proceeding of the Cambridge Philosophical Society 24*, 180-190, 1928.

[44] M. Fréchet. Sur la loi de probabilité de l'écart maximum. *Ann. Soc. Math. Polon 6*, 93-116, 1927.

[45] J. Galambos. Order statistics of samples from multivariate distributions. *Journal of the American Statistical Association, 70*(351a), 674-680, 1975.

[46] J. Galambos. *The asymptotic theory of extreme order statistic*. Wiley, New York, 1978.

[47] J. L. Gastwirth. Statistical properties of a measure of tax assessment uniformity. *Journal of Statistical Planning and Inference, 6*(1), 1-12, 1982.

[48] R. C. Geary. The ratio of the mean deviation to the standard deviation as a test of normality. *Biometrika, 27*(3/4), 310-332, 1935.

[49] J. Geffroy. Contributions à la théorie des valeurs extrêmes. *Publ. Inst. Statist. Univ. Paris, 7*, 37-185, 1958/1959.

[50] Y. R. Gel, and J. L. Gastwirth. A robust modification of the Jarque-Bera test of normality. *Economics Letters, 99*(1), 30-32, 2008.

[51] J. Geluk, L. de Haan, S. Resnick, and C. Stărică. Second-order regular variation, convolution and the central limit theorem. *Stochastic Processes and their Applications, 69*(2), 139-159, 1997.

[52] B. Gnedenko. Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics*, 423-453, 1943.

[53] M. I. Gomes, and M. Stehlík. The Latest Advances on the Hill Estimator and Its Modifications. In *Topics in Nonparametric Statistics*. Sprimger, New york, NY. p. 323-33, 2014.

[54] P. Hall. On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society B 44*, 37-42, 1982.

[55] B. M. Hill. A simple general approach to inference about the tail of a distribution. *Annals of Statistics 3*, 1163-1174, 1975.

[56] J. R. Hosking, J. R. Wallis, and E. F. Wood, Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometric, 27*(3), 251-26, 1985.

[57] P. Hougaard. *Analysis of multivariate survival data*. Springer, New York, 2000.

[58] T. Hsing. On tail index estimation using dependent data. *The Annals of Statistics*, 1547-1569, 1991.

[59] C. M. Jarque, and A. K. Bera. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics letters, 6*(3), 255-259, 1980.

[60] P. Jordanova, Z. Fabián, P. Hermann, L. Střelec, A. Rivera, S. Girard, ... and M. Stehlík. Weak properties and robustness of t-Hill estimators. *Extremes, 19*(4), 591-626, 2016.

[61] R. Koenker, and G. Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society, 46*, 33-50, 1978

[62] D. Kundu, and A. K. Dey. Estimating the parameters of the Marshall-Olkin bivariate Weibull distribution by EM algorithm. *Computational Statistics and Data Analysis, 53*(4), 956-965, 2009.

[63] M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extreme and related properties of random sequences and processe*. Springer series in Statistics. Springer-Verlag, New York, 1983.

[64] E. L. Lehmann. Some concepts of dependence. *The Annals of Mathematical Statistics*, 1137-1153, 1966.

[65] Y. Li, J. Sun, and S. Song. Statistical analysis of bivariate failure time data with Marshall-Olkin Weibull models. *Computational statistics* & *data analysis, 56*(6), 2041-2050, 2012.

[66] H. W. Lilliefors. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American statistical Association, 62*(318), 399-402, 1967.

[67] J. C. Lu. Weibull extensions of the Freund and Marshall-Olkin bivariate exponential models. *IEEE Transactions on Reliability, 38*(5), 615-619, 1989.

[68] H. J. Malik. Estimation of the parameters of the Pareto distribution. *Metrica 15*, 125-132, 1970.

[69] A. W. Marshall, and I. Olkin. A multivariate exponential distribution. *Journal of the American Statistical Association, 62*(317), 30-44, 1967.

[70] A. W. Marshall, and I. Olkin. Domains of attraction of multivariate extreme value distributions. *The Annals of Probability*, 168-177, 1983.

[71] A. W. Marshall, and I. Olkin. *Life distributions*. Springer Series in Statistics. Springer, New York, 2007.

[72] D. M. Mason. Laws of large numbers for sums of extreme values. *The Annals of Probability*, 754-764, 1982.

[73] G. Matthys, and J. Beirlant. Adaptive threshold selection in tail index estimation. *Extremes and Integrated Risk Management*, 37-49, 2000a.

[74] G. Matthys, and J. Beirlant. Extreme cuantile estimation for heavy-tailed distributions. *Preprint, Universitair Centrum voor Statistick, Katholieke Universiteit Leuven*, 2000b.

[75] R. P. McEwen, and B. R. Parresol. Moment expressions and summary statistics for the complete and truncated Weibull distribution. *Communications in Statistics: theory and Methods 20*, 1361-1372, 1991.

[76] T.P. Minka. Estimating a Gamma distribution. *Microsoft Research, Cambridge, UK, Tech. Rep*, 2002

[77] T.P. Minka. Beyond Newton's method, 2000.

[78] V. G. Mikhailov. Asymptotic independence of vector components of multivariate extreme order statistics. *Theory of Probability & Its Applications, 19*(4), 817-821, 1975.

[79] MOP-DGA, CADE-IDEPE. Diagnóstico y clasificación de los cursos y cuerpos de agua según objetivo de calidad. *Cuenca del río Lluta*, Santiago, Chile, 2004.

[80] A. N. M. Muniruzzaman. On measures of location and dispersion and test of hypothesis in a Pareto population. *Calcutta Statistical Association Bulletin 7*,15-122, 1957.

[81] S. Nandi, and I. Dewan. An EM algorithm for estimating the parameters of bivariate Weibull distribution under random censoring. *Computational Statistics and Data Analysis, 54*(6), 1559-1569, 2010.

[82] N. C. Oficial. NCh 1333 Of. *Requisitos de calidad del agua para diferentes usos*, 1978.

[83] N. C, Oficial. NCh 409/1. Agua potable - parte 1: Requisitos. *Instituto Nacional de Normalización, INN. Santiago, 14* pp, 2005.

[84] L. Núñez-Soza, and M. Stehlík. Robust extreme value estimators for levels of boron and arsenic for Lluta river. *Proceedings of AIP (American Inst. of Physics)*, 2017.

[85] C. E. B. Owen. Parameter estimation for the beta distribution, 2008.

[86] V. Pareto. *Cours d'économie politique*. Vol II. F. Rouge, Lausanne, 1987.

[87] K. Pearson, and L. N. G. Filon. Mathematical contributions to the theory of evolution. IV. On the probable errors of frequency constants and on the influence of random selection on variation and correlation. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 191*, 229-311, 1898.

[88] J. Pickand III. Statistical inference using extreme order statistics.*Annals of Statistics 3*, 119-131, 1975.

[89] R. E. Quandt. Old and new methods of estimation and the Pareto distribution. *Metrica 10*, 55-82, 1966.

[90] S. Resnick, and C. Stărică. Consistency of Hill's estimator for dependent data. *Journal of Applied probability, 32*(1), 139-167, 1995.

[91] S. Resnick, and C. Stărică. Smoothing the Hill estimator. *Advances in Applied Probability, 29*(1), 271-293, 1997.

[92] S. Resnick. *Heavy-tail phenomena: probabilistic and statistical modeling.* Springer Science & Business Media, 2007.

[93] S. S. Shaphiro, and M. B. Wilk. An analysis of variance test for normality. *Biometrika, 52*(3), 591-611, 1965.

[94] M. Sibuya. Bivariate extreme statistics, I. *Annals of the Institute of Statistical Mathematics, 11*(3), 195-210, 1960.

[95] R. L. Smith. Maximum likelihood estimation in class of nonregular cases. *Biomethrika 72*, 67-90, 1985.

[96] M. Stehlík, R. Potocký, H. Waldl, and Z. Fabián. On the favorable estimation for fitting heavy tailed data. *Computational Statistics, 25*(3), 485-503, 2010.

[97] M. Stehlík, Z. Fabián, and L. Střelec. Small sample robust testing for normality against Pareto tails. *Communications in Statistics-Simulation and Computation, 41*(7), 1167-1194, 2012.

[98] M. Stehlík, L. Núñez-Soza, Z. Fabián, M. Jiřina, P. Jordanova, I. Arancibia, and J. Kisel'ák. On ecological aspects of dynamics for zero slope regression for water pollution in Chile. In revision for *Ecological Complexity*, 2018

[99] L. Střelec, and M. Stehlík. On robust testing for normality- Proceedings of 6th St. Petersburg Workshop on Sim., 755-759, 2009.

[100] C. M. Urzúa. On the correct use of omnibus tests for normality. *Economics Letters, 53*(3), 247-251, 1996

[101] V. A. Uthoff. The most powerful scale and location invariant test of the normal versus the double exponential. *The Annals of Statistics, 1*(1), 170-174, 1973.

[102] A. N. Tchernitchin, J. Ríos, I. Cortés, and L. Gaete. Polimetales en agua de Arica-Parinacota. Posibles orígenes y efectos en salud. XIV Congreso Geológico, La Serena, Chile, 2015.

[103] H. C. Thode. *Testing for normality*, Statistics: textbooks and monographs. Dekker, New York, 2002.

[104] R. Von Mises. La distribution de la plus grande de n valeurs. *Revue Mathématique de l'Union Interbalcanique 1*, 141-160, 1936.

[105] L. Wassermann. *All of nonparametric statistics*. New York, Springer, 2003.

[106] World Health Organization. Boron in drinking-water: Background document for development of WHO Guidelines for Drinking-water Quality, 2009.

[107] World Health Organization. *Guidelines for drinking water quality*, 4th edn. *World Health Organization*, Geneva, 2011.

Appendix A

A.1 Theorems

Theorem A.1.1. ***Drees (1998)*** *Suppose for a measurable function f and a positive function a we have*

$$\lim_{t\longrightarrow\infty}\frac{f(tx)-f(t)}{a(t)}=\frac{x^{\gamma}-1}{\gamma},$$

for all $x>0$, where γ is a real parameter, i.e., $f\in\mathbf{ERV}_{\gamma}$. Then for all $\varepsilon,\delta>0$ there is a $t_0=t_0(\varepsilon,\delta)$ such that for $t,tx\geq t_0$,

$$\left|\frac{f(tx)-f(t)}{a_0(t)}-\frac{x^{\gamma}-1}{\gamma}\right|\leq\varepsilon x^{\gamma}\max\{x^{\delta},x^{-\delta}\},$$

where

$$a_0(t):=\begin{cases}\gamma f(t), & \gamma>0,\\ -\gamma(f(\infty)-f(t)), & \gamma<0,\\ f(t)-t^{-1}\int_0^t f(s)ds, & \gamma=0.\end{cases}$$

A.2 Proof of Theorems, Lemmas and propositions

Proof of Lemma 3.2.1 Chapter 3

For $\gamma>0$ relation relation (3.8) simplifies to

$$\lim_{t\longrightarrow\infty}\frac{U(tx)}{U(t)}=x^{\gamma},\quad x>0,$$

i.e.,

$$\lim_{t\longrightarrow\infty} \frac{\log U(tx) - \log U(t)}{\gamma} = \log x.$$

For $\gamma \leq 0$ we have

$$\lim_{t\longrightarrow\infty} \frac{U(tx)}{U(t)} = 1 \quad \text{for all } x > 0.$$

Hence (3.8) is equivalent to

$$\lim_{t\longrightarrow\infty} \frac{\log U(tx) - \log U(t)}{\frac{a(t)}{U(t)}} = \frac{x^\gamma - 1}{\gamma}.$$

Summarizing we get that relation (3.8) is equivalent to

$$\lim_{t\longrightarrow\infty} \frac{\log U(tx) - \log U(t)}{\frac{a(t)}{U(t)}} = \frac{x^\gamma - 1}{\gamma_-}, \tag{A.1}$$

with $\gamma_- := \min\{0, \gamma\}$ and

$$\lim_{t\longrightarrow\infty} \frac{a(t)}{U(t)})\gamma_+, \tag{A.2}$$

with $\gamma_+ := \max\{0, \gamma\}$. Next we use the inequalities of previous Theorem: for each $\varepsilon > 0$ there exists t_0 such that for $t \geq t_0$, $x \geq 1$,

$$-\varepsilon x^{\gamma_- + \varepsilon} < \frac{\log U(tx) - \log U(t)}{q_0(t)} - \frac{x^{\gamma_-} - 1}{\gamma_-} < \varepsilon x^{\gamma_- + \varepsilon}, \tag{A.3}$$

where q_0 is a positive function satisfying $q_0(t) \sim \frac{a(t)}{U(t)}$, as $t \longrightarrow \infty$.

Let $Y_1, Y_2, \ldots, Y_n$ be independent and identically distributed with distribution function $1 - \frac{1}{x}$, $x > 1$. We apply these inequalities with $t := Y_{n-k,n}$ (tending to infinity a.s., as $n \longrightarrow \infty$) and $x := \frac{Y_{n-i,n}}{Y_{n-k,n}}$. We then have eventually, uniformly for $0 \leq i \leq k-1$,

$$\frac{\log U(Y_{n-i,n}) - \log U(Y_{n-k,n})}{q_0(Y_{n-k,n})} - \frac{\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right)^{\gamma_-} - 1}{\gamma_-} < \varepsilon \left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right)^{\gamma_- + \varepsilon}$$

It follows by adding the inequalities for $i = 0, 1, \ldots, k-1$ that

$$\begin{aligned}
&\frac{\frac{1}{k}\sum_{i=0}^{k-1}\left(\log U(Y_{n-i,n}) - \log U(Y_{n-k,n})\right)}{q_0(Y_{n-k,n})} \\
&< \frac{1}{k}\sum_{i=0}^{k-1}\frac{\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right)^{\gamma_-} - 1}{\gamma_-} + \frac{\varepsilon}{k}\sum_{i=0}^{k-1}\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right)^{\gamma_-+\varepsilon} \\
&\stackrel{d}{=} \frac{1}{k}\sum_{i=1}^{k}\frac{(Y_i^*)^{\gamma_-} - 1}{\gamma_-} + \frac{\varepsilon}{k}\sum_{i=0}^{k}(Y_i^*)^{\gamma_-+\varepsilon}
\end{aligned}$$

with $Y_1^*, Y_2^*, \ldots, Y_k^*$ independent and identically distributed with distribution function $1 - \frac{1}{x}$, $x > 1$, by the reasoning from the proof of Lemma 3.3.2 in De Haan and Ferreira (2006). By the law of large numbers the right-hand side converges in probability to the mean, i.e,

$$\mathbf{E}\frac{Y^{\gamma_-} - 1}{\gamma_-} + \varepsilon\mathbf{E}Y^{\gamma_-+\varepsilon} = \frac{1}{1-\gamma_-} + \frac{\varepsilon}{1-\gamma_--\varepsilon}. \tag{A.4}$$

A similar lower bound applies.
Next by squaring the expensions (A.3) we find that

$$\left(\frac{\log U(tx) - \log U(t)}{q_0(t)}\right)^2$$

is bounded above by

$$\left(\frac{x^{\gamma_-} - 1}{\gamma_-}\right)^2 + 2\varepsilon x^{\gamma_-+\varepsilon}\frac{x^{\gamma_-} - 1}{\gamma_-} + \varepsilon^2 x^{2\gamma_-+2\varepsilon},$$

and a similar lower bound applies. Starting from these inequalities, we follow the same reasoning as before. This leads to (3.10) for $j = 2$.

■

Proof of Theorem 3.3.1 Chapter 3
Write the denominator in (3.22) as

$$\frac{1}{k}S_{xx} - \left(\frac{1}{k}S_x\right)^2,$$

where as $n \longrightarrow \infty$,

$$\frac{1}{k}S_{xx} = \frac{1}{k}\sum_{i=1}^{k}\left(-\log\left(\frac{i}{k+1}\right)\right)^2 \sim \int_0^1 (-\log x)^2 dx \tag{A.5}$$

$$= \int_0^\infty y^2 \mathrm{e}^{-y} dy = 2 \tag{A.6}$$

and

$$\frac{1}{k}S_{x} = \frac{1}{k}\sum_{i=1}^{k}\left(-\log\left(\frac{i}{k+1}\right)\right) \sim \int_0^1 (-\log x) dx \tag{A.7}$$

$$= \int_0^\infty y \mathrm{e}^{-y} dy = 1. \tag{A.8}$$

Furthermore, as $n \longrightarrow \infty$, $k \longrightarrow \infty$, $\frac{n}{k} \longrightarrow \infty$,

$$\frac{1}{k}\sum_{i=1}^{k}\left(-\log\left(\frac{i}{k+1}\right)\right) H_{k,n} \sim \int_0^1 (-\log x) dx H_{k,n} \xrightarrow{P} \frac{1}{\alpha}$$

by the weak consistency of the Hill estimator. So far consistency of the QQ estimator, it suffices to show that

$$A_n := \frac{1}{k}\sum_{i=1}^{k}\left(-\log\left(\frac{i}{k+1}\right)\right) \log\left(\frac{X_{(i)}}{X_{(k+1)}}\right) \xrightarrow{P} \frac{2}{\alpha}. \tag{A.9}$$

Recall from (4.18) (p.82 in Resnick, 2007) that

$$\frac{X_{[kt]}}{b\left(\frac{n}{k}\right)} \Rightarrow t^{-\frac{1}{\alpha}} \tag{A.10}$$

in $D(0,\infty]$. Now write

$$A_n = \frac{k+1}{k}\int_0^1 \left(-\log\left(\frac{[(k+1)t]}{k+1}\right) \log\left(\frac{X_{[(k+1)t]}}{X_{(k+1)}}\right)\right) dt.$$

We claim this converges in probability to

$$\xrightarrow{P} \int_0^1 (-\log t)(\log t^{-\frac{1}{\alpha}}) dt$$

$$= \frac{1}{\alpha}\int_0^1 (-\log t)^2 dt = \frac{2}{\alpha}. \tag{A.11}$$

The convergence near zero in (A.11) is a problem since (A.10) holds in $D(0,\infty]$ and hence does not cover neighborhoods of zero. So we have to use a converging together argument based on Theorem 3.5 (p.56 in Resnick, 2007). Convergence of the integral in (A.11) over the region $(\delta,1)$ is guaranteed by the fact that (A.10) is tantamount to local uniform convergence *away from zero*, and hence on $(\delta,1)$ there is uniform convergence. So it suffices to show that

$$\lim_{\delta\longrightarrow 0}\limsup_{k\longrightarrow\infty} P\left[\int_0^{\delta}(-\log t)\log\left(\frac{X_{[(k+1)t]}}{X_{k+1}}\right)dt>\eta\right]=0. \tag{A.12}$$

We do this by using Potter's inequalities and Rényi's representation of order statistics (see Problem 4.1 p.114 in Resnick, 2007). Recall that Potter's inequalities (2.31) (Resnick, 2007) take the following form: Since $\frac{1}{1-F}$ is regularly varying with index α, the inverse $b=\left(\frac{1}{1-F}\right)^{\leftarrow}$ is regularly varying with index $\frac{1}{\alpha}$, and for $\epsilon>0$, there exists $t_0=t_0(\epsilon)$ such that if $y\geq 1$ and $t\geq t_0$,

$$(1-\epsilon)y^{\alpha^{-1}-\epsilon}\leq\frac{b(ty)}{b(t)}\leq(1+\epsilon)y^{\alpha^{-1}+\epsilon}. \tag{A.13}$$

We now replace this in terms of the function

$$R=-\log(1-F)=\log b^{\leftarrow}.$$

Then $b=R^{\leftarrow}\circ\log$; taking logarithms in (A.13) and then converting from a multiplicative to an additive form yields that

$$\log(1-\epsilon)+(\alpha^{-1}-\epsilon)y\leq\log R^{\leftarrow}(s+y)-\log R^{\leftarrow}(s)\leq\log(1+\epsilon)+(\alpha^{-1}+\epsilon)y \tag{A.14}$$

for $s\geq\log t_0$ and $y\geq 0$.

The reason for introducing the R function is that if $E_1,E_2,\dots,E_n$ are i.i.d. unit exponentially distributed random variables, then

$$(X_1,X_2,\dots,X_n)\stackrel{\mathrm{d}}{=}(R^{\leftarrow}(E_j);j=1,\dots,n).$$

The Rényi representation gives for the spacings of exponential order statistics,

$$\begin{aligned}(E_{1,n},E_{2,n}-E_{1,n},\dots,E_{n,n})&\stackrel{\mathrm{d}}{=}\left(\frac{E_n}{n},\frac{E_{n-1}}{n-1},\dots,\frac{E_n}{1}\right)\\(E_{(1)}-E_{(2)},E_{(2)}-E_{(3)},\dots,E_{(n-1)}-E_{(n)})&\stackrel{\mathrm{d}}{=}\left(\frac{E_1}{1},\frac{E_2}{2},\dots,\frac{E_n}{n}\right).\end{aligned} \tag{A.15}$$

Now we have

$$
\begin{aligned}
&P\left[\int_0^\delta (-\log t)\log\left(\frac{X_{[(k+1)t]}}{X_{(k+1)}}\right)dt > \eta\right] \\
&= P\left[\int_0^\delta (-\log t)\log\left(\frac{R^{\leftarrow}(E_{[(k+1)t]}) - E_{(k+1)} + E_{(k+1)})}{R^{\leftarrow}(E_{(k+1)})}\right)dt > \eta\right] \\
&\leq P\left[\int_0^\delta (-\log t)\log\left(\frac{R^{\leftarrow}(E_{[(k+1)t]}) - E_{(k+1)} + E_{(k+1)})}{R^{\leftarrow}(E_{(k+1)})}\right)dt > \eta,\ \mathrm{e}^{E_{(k+1)}} > t_0\right] + o_k(1),
\end{aligned}
$$

since $E_{(k)} \xrightarrow{P} \infty$ by Problem 4.3 (p.115 in Resnick, 2007). Ignore the term $o_k(1)$ since it goes to zero with k. Apply (A.15) and we get the upper bound

$$
\leq P\left[\int_0^\delta (-\log t)[(1+\epsilon) + (\alpha^{-1}+\epsilon)(E_{[(k+1)t]} - E_{(k+1)})]dt > \eta\right]
$$

and for small $\eta' > 0$ this is bounded by

$$
\leq P\left[\int_0^\delta (-\log t)(E_{[(k+1)t]} - E_{(k+1)})dt \ > \eta'\right].
$$

Apply Markov's inequality to get the upper bound

$$
\begin{aligned}
&\leq \frac{1}{\eta'}\int_0^\delta (-\log t)\mathbf{E}(E_{[(k+1)t]} - E_{(k+1)})dt \\
&= \frac{1}{\eta'}\int_0^\delta (-\log t)\sum_{l=[(k+1)t]}^{k+1}\frac{1}{l}dt;
\end{aligned}
$$

as $k \longrightarrow \infty$, this is asymptotic to $\frac{1}{\eta'}\int_0^\delta(-\log t)(-\log t)dt$, and as $\delta \longrightarrow 0$, this is asymptotic to $\frac{1}{\eta'}\delta(\log\delta)^2 \longrightarrow 0$.

■

Proof of Theorem 3.4.1 Chapter 3

Consistency of the tail empirical measure, defined as a random element of $M_+((0,\infty])$, the space of non negative Radon measures on $(0,\infty]$, implies the consistency of t-Hill estimator for $\frac{1}{\alpha+1}$. The proof proceeds by series of steps following the proof of classical Hill estimator in Resnick (2007).

Step 1. Consistency of empirical measure (given in 4.14 by Resnick, 2007) implies $\frac{X(k)}{b\left(\frac{n}{k}\right)} \xrightarrow{P} 1$ as $n \longrightarrow \infty$, $k \longrightarrow \infty$, $\frac{k}{n} \longrightarrow 0$. Here $b(t)$ is a quantile function and $X(k)$ is the kth largest order statistic.
This allow us to consider $X(k)$ as a consistent estimator of $b\left(\frac{n}{k}\right)$.

Step 2. In $M_+(0,\infty] : v_n \xrightarrow{P} v_\alpha$ as $n \longrightarrow \infty$, $k \longrightarrow \infty$, $\frac{k}{n} \longrightarrow 0$. This is proved by a scaling argument.
Define the operator $T : M_+((0,\infty]) \times (0,\infty) \longrightarrow M_+((0,\infty])$ by

$$T(\mu, x)(A) = \mu(xA).$$

From (4.14) in Resnick (2007) and Proposition 3.1 therein we get joint weak convergence $\left(v_n, \frac{X(k)}{b\left(\frac{n}{k}\right)}\right) \Rightarrow (v_\alpha, 1)$ in $M_+(0,\infty] \times (0,\infty)$.
Since $\hat{v}_n(\cdot\,) = T\left(v_n, \frac{X(k)}{b\left(\frac{n}{k}\right)}\right)$, the conclusion will follow by the continuous mapping theorem and continuity of the operator T at $(v_\alpha, 1)$.

Step 3. Integrate the tails of the mesures against $x^{-2}dx$. The integral functional is continuous on $[1, M]$ for any M and so it only on $[M, \infty]$ that care must be exercised. By the 2nd converging together Theorem (see Theorem 3.5 in Resnick, 2007), only we show that

$$\lim_{M\longrightarrow\infty} \limsup_{n\longrightarrow\infty} P\left[\int_M^\infty \hat{v}_n(x,\infty]x^{-2}dx > \delta\right] = 0$$

We have

$$P\left[\int_M^\infty \hat{v}_n(x,\infty]x^{-2}dx > \delta\right] \le I + II,$$

where

$$II \le P\left[\left|\frac{\hat{b}\left(\frac{n}{k}\right)}{b\left(\frac{n}{k}\right)} - 1\right| \ge \eta\right] \longrightarrow 0 \;\; \text{and}$$

$$I \le P\left[\int_M^\infty v_n((1-\eta)x,\infty]x^{-2}dx > \delta\right] = P\left[\int_{M(1-\eta)}^\infty v_n(x,\infty]x^{-2}dx > \delta\right]$$

and this probability has a bound from Markov's inequality

$$\begin{aligned}&\delta^{-1}\mathrm{E}\left[\int_{M(1-\eta)}^{\infty} v_n(x,\infty]x^{-2}dx\right]\\&= \delta^{-1}\int_{M(1-\eta)}^{\infty}\frac{n}{k}P\left[X_1 > b\left(\frac{n}{k}\right)x\right]x^{-2}dx \longrightarrow \delta^{-1}\int_{M(1-\eta)}^{\infty} x^{-2-\alpha}dx,\end{aligned}$$

for $n \leftarrow \infty$.

Finally,

$$\delta^{-1}\int_{M(1-\eta)}^{\infty} x^{-2-\alpha}dx = \frac{const}{M^{\alpha+1}} \longrightarrow 0 \quad \text{for} \quad M \longrightarrow \infty.$$

We have applied Karatamata's Theorem (see Theorem 2.1 in Resnick, 2007, p. 25.).

Step 4. We proved that

$$\int_1^{\infty} \hat{v}_n(x,\infty]x^{-2}dx \xrightarrow{P} \int_1^{\infty} v_\alpha(x,\infty]x^{-2}dx. \tag{A.16}$$

So $\int_1^\infty \hat{v}_n(x,\infty]x^{-2}dx$ is a consistent estimator of $\frac{1}{\alpha+1}$ and we just need to see that this is indeed the modified Hill estimator. This is done as follows (here $\varepsilon(x)A$ is a characteristic function of set A):

$$\begin{aligned}\int_1^{\infty} \hat{v}_n(x,\infty]x^{-2}dx &= \int_1^{\infty}\frac{1}{k}\sum_{i=1}^{n}\varepsilon\left(\frac{X_i}{\hat{b}\left(\frac{n}{k}\right)}\right)(x,\infty]x^{-2}dx\\&= \sum_{i=1}^{n}\int_1^{\frac{X_i}{\hat{b}\left(\frac{n}{k}\right)}\vee 1} x^{-2}dx\\&= 1 - \frac{1}{k}\sum_{i=1}^{n}\frac{1}{\frac{X(i)}{X(k)}}.\end{aligned}$$

■

Proof Lemma 3.4.1 Chapter 3
The result is proved analogously to Resnick and Stărică (1993). An integration by part entails

$$
\begin{aligned}
\Phi_n &= \int_1^\infty \varphi(x) \left\{ \frac{1}{k} \sum_{i=1}^n \varepsilon \left\{ \frac{X_i}{X_{(n-k,n)}} \in dx \right\} \right\} \\
&= \int_1^\infty \varphi(x) \hat{\mu}_{X,k,n}(dx) \\
&= - \int_1^\infty \varphi(x) d\hat{\mu}_{X,k,n}(x, \infty] \\
&= \lim_{x \longrightarrow 1} \varphi(x) \hat{\mu}_{X,k,n}(x, \infty] - \lim_{x \longrightarrow \infty} \varphi(x) \hat{\mu}_{X,k,n}(x, \infty] + \int_1^\infty \hat{\mu}_{X,k,n}(x, \infty] d\varphi(x).
\end{aligned}
$$

The positive measures $\hat{\mu}_{X,k,n}$ are Radon and the measure $\hat{\mu}_{X,k,n}$ converges to μ and $\alpha > 0$. Therefore, in view (3.60) and the definition of μ, the L'Hôpital's rule yields

$$
\begin{aligned}
\lim_{n \longrightarrow \infty} \lim_{x \longrightarrow \infty} \varphi(x) \hat{\mu}_{X,k,n}(x, \infty] &= \lim_{x \longrightarrow \infty} \lim_{n \longrightarrow \infty} \varphi(x) \hat{\mu}_{X,k,n}(x, \infty] \\
&= \lim_{x \longrightarrow \infty} \frac{\varphi(x)}{x^\alpha} \\
&= 0,
\end{aligned}
$$

and similarly,

$$
\begin{aligned}
\lim_{n \longrightarrow \infty} \lim_{x \longrightarrow 1} \varphi(x) \hat{\mu}_{X,k,n}(x, \infty] &= \lim_{x \longrightarrow 1} \lim_{n \longrightarrow \infty} \varphi(x) \hat{\mu}_{X,k,n}(x, \infty] \\
&= \varphi(1).
\end{aligned}
$$

Then, for all $t \geq 1$,

$$
\lim_{n \longrightarrow \infty} \Phi_n = \lim_{n \longrightarrow \infty} \left\{ \int_1^t \hat{\mu}_{X,k,n}(x, \infty] \varphi'(x) dx + \int_t^\infty \hat{\mu}_{X,k,n}(x, \infty] \varphi'(x) dx \right\}.
$$

Let us first remark that

$$
\begin{aligned}
\lim_{n \longrightarrow \infty} \int_1^t \hat{\mu}_{X,k,n}(x, \infty] \varphi'(x) dx &= \int_1^t \mu(x, \infty] \varphi'(x) dx \\
&= \int_1^t x^{-\alpha} \varphi'(x) dx
\end{aligned}
$$

and therefore

$$\lim_{t\longrightarrow\infty}\lim_{n\longrightarrow\infty}\int_1^t \hat{\mu}_{X,k,n}(x,\infty]\varphi'(x)dx = \int_1^{+\infty} x^{-\alpha}\varphi'(x)dx.$$

Second, in order to apply Theorem 4.2 of Billingsley (2013), we have to check that

$$\lim_{t\longrightarrow\infty}\lim_{n\longrightarrow\infty} P\left\{\int_t^\infty \hat{\mu}_{X,k,n}(x,\infty]|\varphi'(x)|dx > \varepsilon\right\} = 0. \tag{A.17}$$

Our aim is to replace $\hat{\mu}_{X,k,n}$ with $\mu_{X,k,n}$. To this aim, we choose $\delta > 0$ and consider

$$\begin{aligned}
&P\left\{\int_t^\infty \hat{\mu}_{X,k,n}(x,\infty]|\varphi'(x)|dx > \varepsilon\right\}\\
&= P\left\{\int_t^\infty \mu_{X,k,n}\left(\frac{X_{(n-k,n)}}{b\left(\frac{n}{k}\right)}x,\infty\right]|\varphi'(x)|dx > \varepsilon\right\}\\
&= P\left\{\int_t^\infty \mu_{X,k,n}\left(\frac{X_{(n-k,n)}}{b\left(\frac{n}{k}\right)}x,\infty\right]|\varphi'(x)|dx > \varepsilon,\ \left|\frac{X_{(n-k,n)}}{b\left(\frac{n}{k}\right)}-1\right| < \delta\right\}\\
&\quad + P\left\{\int_t^\infty \mu_{X,k,n}\left(\frac{X_{(n-k,n)}}{b\left(\frac{n}{k}\right)}x,\infty\right]|\varphi'(x)|dx > \varepsilon,\ \left|\frac{X_{(n-k,n)}}{b\left(\frac{n}{k}\right)}-1\right| \geq \delta\right\}.
\end{aligned}$$

Clearly,

$$\begin{aligned}
0 &\leq P\left\{\int_t^\infty \mu_{X,k,n}\left(\frac{X_{(n-k,n)}}{b\left(\frac{n}{k}\right)}x,\infty\right]|\varphi'(x)|dx > \varepsilon,\ \left|\frac{X_{(n-k,n)}}{b\left(\frac{n}{k}\right)}-1\right| \geq \delta\right\}\\
&\leq P\left\{\left|\frac{X_{(n-k,n)}}{b\left(\frac{n}{k}\right)}-1\right| \geq \delta\right\}
\end{aligned}$$

which, from (3.59), converges to zero when $n \longrightarrow \infty$. In view of (3.59), for any $0 < \delta_0 < 1+\alpha$ there exists n_{δ_0} large enough such that all $t > 1$ and $n > n_{\delta_0}$,

$$\begin{aligned}
&P\left\{\int_t^\infty \mu_{X,k,n}\left(\frac{X_{(n-k,n)}}{b\left(\frac{n}{k}\right)}x,\infty\right]|\varphi'(x)|dx > \varepsilon,\ \left|\frac{X_{(n-k,n)}}{b\left(\frac{n}{k}\right)}-1\right| < \delta\right\}\\
&\leq P\left\{\int_t^\infty \mu_{X,k,n}((1-\delta_0)x,\infty]|\varphi'(x)|dx > \varepsilon,\ \left|\frac{X_{(n-k,n)}}{b\left(\frac{n}{k}\right)}-1\right| < \delta\right\}\\
&\leq P\left\{\int_t^\infty \mu_{X,k,n}((1-\delta_0)x,\infty]|\varphi'(x)|dx > \varepsilon\right\}.
\end{aligned}$$

Chebishev's inequality guarantees that

$$P\left\{\int_t^{\infty} \mu_{X,k,n}((1-\delta_0)x,\infty]|\varphi'(x)|dx > \varepsilon\right\} \leq \frac{1}{\varepsilon}\mathrm{E}\int_t^{\infty} \mu_{X,k,n}((1-\delta_0)x,\infty]|\varphi'(x)|dx.$$

Since $X_1, X_2, \cdots, X_n$ are indentically distributed, it follow that

$$\begin{aligned}
\mathrm{E}\mu_{X,k,n}((1-\delta_0)x,\infty] &= \frac{1}{k}\sum_{i=1}^{n} P\left\{\frac{X_i}{b\left(\frac{n}{k}\right)} \in ((1-\delta_0)x,\infty]\right\} \\
&= \frac{n}{k}P\left\{\frac{X_1}{b\left(\frac{n}{k}\right)} \in ((1-\delta_0)x,\infty]\right\} \\
&= \frac{n}{k}\overline{F}((1-\delta_0)xb\left(\frac{n}{k}\right) \\
&= \frac{n}{k}\overline{F}\left(b\left(\frac{n}{k}\right)\right)\frac{\overline{F}\left((1-\delta_0)xb\left(\frac{n}{k}\right)\right)}{\overline{F}\left(b\left(\frac{n}{k}\right)\right)}.
\end{aligned}$$

From the definition of b, $\frac{n}{k} \longrightarrow \infty$ implies $b\left(\frac{n}{k}\right) \longrightarrow \infty$ and

$$\frac{n}{k}\overline{F}\left(b\left(\frac{n}{k}\right)\right) \longrightarrow 1$$

as $n \longrightarrow \infty$. Thus, in view of (3.62) and Potter's inequality for regularly varying functions, for any $0 < \delta_1 < \alpha$ there exist x_0 and n_{δ_1} large enough such that, for all $x > \max\{1, x_0\}$ and $n > n_{\delta_1}$,

$$\mathrm{E}\mu_{X,k,n}((1-\delta_0)x,\infty] \leq (1+\delta_1)((1-\delta_0)x)^{-\alpha+\delta_1},$$

and for all $t > 1$,

$$\begin{aligned}
&\int_t^{\infty} \mathrm{E}\mu_{X,k,n}((1-\delta_0)x,\infty]|\varphi'(x)|dx \\
&\leq (1+\delta_1)((1-\delta_0)x)^{-\alpha+\delta_1}\int_t^{\infty} x^{-\alpha+\delta_1}|\varphi'(x)|dx.
\end{aligned}$$

We apply Fubibi's Theorem and obtain that for all $t > 1$ and $n > n_{\delta}$,

$$
\begin{aligned}
&P\left\{\int_t^\infty \mu_{X,k,n}((1-\delta_0)x,\infty]|\varphi'(x)|dx > \varepsilon\right\} \\
&\le \frac{1}{\varepsilon}\int_t^\infty \mathrm{E}\mu_{X,k,n}((1-\delta_0)x,\infty]|\varphi'(x)|dx \\
&\le \frac{1}{\varepsilon}(1+\delta_1)(1-\delta_0)^{-\alpha+\delta_1}\int_t^\infty x^{-\alpha+\delta_1}|\varphi'(x)|dx.
\end{aligned}
$$

As a consequence,

$$
\lim_{t\longrightarrow\infty}\lim_{n\longrightarrow\infty} P\left\{\int_t^\infty \hat{\mu}_{X,k,n}\left(\frac{X_{(n-k,n)}}{b\left(\frac{n}{k}\right)}x,\infty\right]|\varphi'(x)|dx > \varepsilon,\ \left|\frac{X_{(n-k,n)}}{b\left(\frac{n}{k}\right)}-1\right| < \delta\right\} = 0
$$

and (A.17) is satisfied.

■

Proof Proposition 3.4.1 Chapter 3
The random variables X_n, $-\infty < n < \infty$, are identically distributed. Cline (1983) proves that under these setting

$$
\overline{F}(x) = P\left(\sum_{j=0}^{\infty} c_j Z_j\right) \sim \sum_{j:c_j>0}^{\infty} c_j^{\alpha}\overline{G}(x) \in \mathrm{RV}_{-\alpha}.
$$

Using (3.58), apply the above Lemma 3.4.1 and complete the proof.

■

Proof Proposition 3.4.2 Chapter 3

1. The random variables X_n, $-\infty < n < \infty$, are identically distributed. Cline (1983) proves that under these settings

   $$
   \overline{F}(x) = P\left(\sum_{j=0}^{\infty} c_j Z_j > r\right) \sim \sum_{j:c_j>0}^{\infty} c_j^{\alpha}\overline{G}(x) \in \mathrm{RV}_{-\alpha}.
   $$

   Using (3.58), apply Lemma 3.4.1 and complete the proof.

2. We use Slutsky's arguments and obtain weak consistency of the t-lgHill estimator in case of infinite moving avarage (MA) sequence.

∎

Proof Theorem 3.4.2 Chapter 3
Recall that t-lgHill estimator $H_{k,n}^L$ of $\gamma = \frac{1}{\alpha}$ can be written as

$$H_{k,n}^L = \frac{M_n^{(2)} - \left(M_n^{(1)}\right)^2}{M_n^{(1)}}, \tag{A.18}$$

while the Moment estimator is defined as

$$\hat{\gamma}_M = M_n^{(1)} + 1 - \frac{1}{2}\left(1 - \frac{\left(M_n^{(1)}\right)^2}{M_n^{(2)}}\right)^{-1}, \tag{A.19}$$

see (3.5.9) p.102, in De Haan and Ferreira (2006), also (1.7) in page 1834 of Dekkers and De Haan (1989). The difference between the two estimators appears clearly when comparing (A.18) and (A.19). In the following, using $\hat{\gamma}_- := \hat{\gamma}_M - M_n^{(1)}$, see Remark 3.5.7 in De Haan and Ferreira (2006), which is an estimator of $\gamma_- := \min\{0, \gamma\}$. It is thus clear that t-lgHill estimator (A.18) can be rewritten as

$$H_{k,n}^L = \frac{M_n^{(1)}}{1 - 2\hat{\gamma}_-}. \tag{A.20}$$

Corollary 3.5.6 of De Haan and Ferreira (2006) states that

$$\sqrt{k}\left\{1 - \frac{1}{2}\left(1 - \frac{\left(M_n^{(1)}\right)^2}{M_n^{(2)}}\right)^{-1} - \gamma_-\right\} \xrightarrow{d} (1-2\gamma_-)(1-\gamma_-)^2\left\{\left(\frac{1}{2} - \gamma_-\right)Q - 2P\right\},$$

where Q, P are Gaussian distributions given in Lemma 3.5.5 of De Hann and Ferreira (2006). Using the previous notation and remarking that $\gamma_- = 0$, this result can be simplified as $\xi_n := \sqrt{k}\hat{\gamma}_-$ converges in distribution to $\frac{1}{2}Q - 2P$. Similarly, let $\xi_n' := \sqrt{k}(M_n^{(1)} - \gamma_+)$ where $\gamma_+ := \max\{0, \gamma\} = \gamma = \frac{1}{\alpha}$. It is easily that ξ_n' can be expanded as $\xi_n' = \sqrt{k}(\hat{\gamma}_M - \gamma) - \sqrt{k}\hat{\gamma}_-$. The limit distribution of the first term is established in (3.5.24), page 109 of De Haan and Ferreira (2006):

$$\sqrt{k}((\hat{\gamma}_M - \gamma) \xrightarrow{d} \frac{1}{2}Q - (\gamma - 2)P.$$

The limit distribution of second term has already been established to be $\frac{1}{2}Q - 2P$ and therefore ξ_n' converges in distribution to γP. Plugging ξ_n and ξ_n' in (A.20) yields

$$\begin{aligned} H_{k,n}^L &= \frac{\gamma + k^{-\frac{1}{2}}\xi_n'}{1 - 2k^{-\frac{1}{2}}\xi_n} \\ &= (\gamma + k^{-\frac{1}{2}}\xi_n')(1 + 2k^{-\frac{1}{2}}\xi_n(1 + o_p(1))) \\ &= \gamma + k^{-\frac{1}{2}}(\xi_n' + 2\gamma\xi_n) + o_p(k^{-\frac{1}{2}}). \end{aligned}$$

Thus,

$$\sqrt{k}(H_{k,n}^L - \gamma) = \xi_n' + 2\gamma\xi_n + o_p(1) \xrightarrow{d} \gamma P + 2\gamma\left(\frac{1}{2}Q - 2P\right) = \gamma(Q - 2P).$$

According to Lemma 3.5.5 of De Haan and Ferreira (2006), (P, Q) is a bivariate centered Gaussian random vector with covariance matrix $\begin{bmatrix} 1 & 4 \\ 4 & 20 \end{bmatrix}$. The asymptotic variance in such case is

$$\mathrm{Var}(Q - 2P) = \mathrm{Var}(Q) + 4\mathrm{Var}(P) - 4\mathrm{Cov}(P, Q) = 8.$$

As a conclusion, we have

$$\sqrt{k}\left(\frac{H_{k,n}^L}{\gamma} - 1\right) \xrightarrow{d} N(0, 8).$$

■

Appendix B

B.1 Code examples for one-dimensional variables

B.1.1 Examples of extreme value distribution

```
##########  Section 1.1.1 ##########
## GEV
library(evd)
par(mfrow=c(1,3))
curve(dgev(x,loc=1,scale=0.5,shape=1,log=FALSE),
      xlim=c(-1,5),ylim=c(0,1.2),ylab=" ",
      col="blue",lwd=2,main="Fréchet Distribution")
curve(dgev(x,loc=1,scale=0.5,shape=0.6,log=FALSE),
      col="red",lwd=2,add=TRUE)
curve(dgev(x,loc=1,scale=0.5,shape=0.2,log=FALSE),
      col="green", lwd=2, add=TRUE)
legend("topright",legend=c(expression(paste(alpha,"=1")),
       expression(paste(alpha,"=0.6")),
       expression(paste(alpha,"=0.2"))),
       col=c("blue","red","green"),lty=c(1,1,1),
       lwd=c(2,2,2))
curve(dgev(x,loc=1,scale=1.5,shape=0,log=FALSE),
      xlim=c(-1,5),ylim=c(0,1.2),ylab=" ",
      col="blue",lwd=2,main="Gumbel Distribution")
curve(dgev(x,loc=1,scale=1,shape=0,log=FALSE),
      col="red",lwd=2,add=TRUE)
curve(dgev(x,loc=1,scale=0.5,shape=0,log=FALSE),
      col="green",lwd=2,add=TRUE)
legend("topright",legend=c(expression(paste(lambda,"=1.5")),
       expression(paste(lambda,"=1")),
       expression(paste(lambda,"=0.5"))),
```

```
        col=c("blue","red","green"),lty=c(1,1,1),
        lwd=c(2,2,2))
curve(dgev(x,loc=1,scale=0.8,shape=-0.8,log=FALSE),
      xlim=c(-1,5),ylim=c(0,1.2),ylab=" ",
      col="blue",lwd=2,main="Reverse-Weibull Distribution")
curve(dgev(x,loc=1,scale=0.8,shape=-0.5,log=FALSE),
      col="red",lwd=2, add=TRUE)
curve(dgev(x,loc=1, scale=0.8, shape=-0.1, log=FALSE),
      col="green",lwd=2,add=TRUE)
legend("topright",legend=c(expression(paste(alpha,"=-0.8")),
        expression(paste(alpha," -0.5")),
        expression(paste(alpha," -0.1"))),
        col=c("blue","red","green"),lty=c(1,1,1),
        lwd=c(2,2,2))

########## Section 1.1.3 ##########
## Exponential Distribution
par(mfrow=c(1,2))
curve(dexp(x,1), xlim=c(0,10), ylim=c(0,1.2), col="blue",
      lwd=2, ylab="density", main= " ")
curve(dexp(x,0.8), col="red", lwd=2, add=TRUE)
curve(dexp(x,0.4),col="green", lwd=2,add=TRUE)
legend("topright",legend=c(expression(paste(lambda," =1")),
        expression(paste(lambda,"=0.8")),
        expression(paste(lambda,"=0.4"))),
        col=c("blue","red","green"),lty=c(1,1,1),
        lwd=c(2,2,2))
curve(pexp(x,1),xlim=c(0,10),ylim=c(0,1.2),col="blue",
      lwd=2,ylab="cdf",main=" ")
curve(pexp(x,0.8),col="red",lwd=2,add=TRUE)
curve(pexp(x,0.4),col="green",lwd=2,add=TRUE)
legend("bottomright",legend=c(expression(paste(lambda," =1")),
        expression(paste(lambda,"=0.8")),
        expression(paste(lambda,"=0.4"))),
        col=c("blue","red","green"),lty=c(1,1,1),
        lwd=c(2,2,2))

## Gamma Distribution
par(mfrow=c(1,3))
curve(dgamma(x,1,1,log=FALSE),ylab="density",
      xlim=c(0,10),ylim=c(0,1),col="blue",lwd=2,main=" ")
curve(dgamma(x,2,1,log=FALSE),col="red",lwd=2,add=TRUE)
curve(dgamma(x,5,1,log=FALSE),col="green",lwd=2,add=TRUE)
legend("topright",legend=c("k=1","k=2","k=5"),
        col=c("blue","red","green"),lty=c(1,1,1),
        lwd=c(2,2,2))
```

```
curve(pgamma(x,1,1,log=FALSE),ylab="cdf",xlim=c(0,10),
      ylim=c(0,1),col="blue",lwd=2,main=" ")
curve(pgamma(x,2,1,log=FALSE),col="red",lwd=2,add=TRUE)
curve(pgamma(x,5,1,log=FALSE),col="green",lwd=2,add=TRUE)
legend("bottomright",legend=c("k=1","k=2","k=5"),
       col=c("blue","red","green"),lty=c(1,1,1),
       lwd=c(2,2,2))
curve(1-pgamma(x,1,1,log=FALSE),ylab="survival function",
      xlim=c(0,10),ylim=c(0,1),col="blue",lwd=2,main=" ")
curve(1-pgamma(x,2,1,log=FALSE),col="red",lwd=2,
      add=TRUE)
curve(1-pgamma(x,5,1,log=FALSE),col="green",lwd=2,
      add=TRUE)
legend("topright",legend=c("k=1","k=2","k=5"),
       col=c("blue","red","green"),lty=c(1,1,1),
       lwd=c(2,2,2))

## Weibull Distribution
par(mfrow=c(1,3))
curve(dweibull(x,shape=0.5,scale=1,log=FALSE),
      ylab="density",xlim=c(0,5),ylim=c(0,2),
      col="blue",lwd=2,main=" ")
curve(dweibull(x,shape=2,1,log=FALSE),col="red",lwd=2,
      add=TRUE)
curve(dweibull(x,shape=3.5,log=FALSE),col="green",lwd=2,
      add=TRUE)
legend("topright",legend=c(expression(paste(alpha,"=0.5")),
       expression(paste(alpha,"=2")),
       expression(paste(alpha, "=3.5"))),
       col=c("blue","red","green"),lty=c(1,1,1),
       lwd=c(2,2,2))
curve(pweibull(x,shape=0.5,scale=1,log=FALSE),
      ylab="cdf",xlim=c(0,5),ylim=c(0,1.2),
      col="blue",lwd=2,main=" ")
curve(pweibull(x,shape=2,scale=1,log=FALSE),col="red",
      lwd=2,add=TRUE)
curve(pweibull(x,shape=3.5,scale=1,log=FALSE),col="green",
      lwd=2,add=TRUE)
legend("bottomright",legend=c(expression(paste(alpha,"=0.5")),
       expression(paste(alpha, "=2")),
       expression(paste(alpha, "=3.5"))),
       col=c("blue","red","green"),lty=c(1,1,1),
       lwd=c(2,2,2))
curve(1-pweibull(x,shape=0.5,scale=1,log=FALSE),
      ylab="survival function",xlim=c(0,5),ylim=c(0,1.2),
      col="blue",lwd=2,main=" ")
```

```
curve(1-pweibull(x,shape=2,scale=1,log=FALSE),col="red",
      lwd=2,add=TRUE)
curve(1-pweibull(x,shape=3.5,scale=1,log=FALSE),col="green",
      lwd=2,add=TRUE)
legend("topright",legend=c(expression(paste(alpha,"=0.5")),
        expression(paste(alpha, "=2")),
        expression(paste(alpha, "=3.5"))),
        col=c("blue","red","green"),lty=c(1,1,1),
        lwd=c(2,2,2))

########## Section 1.1.4 ##########
## Beta Distribution
par(mfrow=c(2,2))
curve(dbeta(x,0.5,0.5,log=FALSE),ylab="density",col="blue",
      lwd=2,main=expression(paste("Case a)",alpha,"=0.5,",
      beta,"=0.5")))
abline(v=0.5,lty=2, col="green",lwd=1)
curve(dbeta(x,1.5,1.5,log=FALSE),ylab="density",col="blue",
      lwd=2,main=expression(paste("Case b)",alpha,"=1.5,",
      beta,"=1.5")))
abline(v=0.5,lty=2, col="green",lwd=1)
curve(dbeta(x,0.5,1.5,log=FALSE),ylab="density",col="blue",
      lwd=2,main=expression(paste("Case c)",alpha,"=0.5")))
curve(dbeta(x,0.5,3.5,log=FALSE),col="red",lwd=2,add=TRUE)
legend("topright",legend=c(expression(paste(beta,"=1.5")),
        expression(paste(beta,"=3.5"))),
        col=c("blue","red"),lty=c(1,1),lwd=c(2,2))
curve(dbeta(x,1.2,0.6,log=FALSE),ylab="density",col="blue",
      lwd=2,main=expression(paste("Case d)",beta,"=0.5")))
curve(dbeta(x,3,0.6,log=FALSE),col="red",lwd=2,add=TRUE)
legend("topleft",legend=c(expression(paste(alpha,"=1.5")),
        expression(paste(alpha,"=3.5"))),
        col=c("blue","red"),lty=c(1,1),lwd=c(2,2))

########## Section 1.1.5 ##########
## Pareto Distribution
a <- c(0.5,1.5,3 )
la <- 1
mu <- 2.5
ga <- c(0.2,0.5,1)
fI1 <- function(x) { (a[1]/x)*(x/la)^(-a[1])}
fI2 <- function(x) { (a[2]/x)*(x/la)^(-a[2])}
fI3 <- function(x) { (a[3]/x)*(x/la)^(-a[3])}
fII1 <- function(x) { (a[1]/la)*(1+(x-mu)/la)^{-a[1]-1}}
fII2 <- function(x) { (a[2]/la)*(1+(x-mu)/la)^{-a[2]-1}}
fII3 <- function(x) { (a[3]/la)*(1+(x-mu)/la)^{-a[3]-1}}
```

```
fIII1 <- function(x) {(1/(la*ga[1]))*
          (1+((x-mu)/la)^{1/ga[1]})^(-2)*
          (x-mu)/la)^(1/ga[1]-1)}
fIII2 <- function(x) {(1/(la*ga[2]))*
          (1+((x-mu)/la)^{1/ga[2]})^(-2)*
          ((x-mu)/la)^(1/ga[2]-1)}
fIII3 <- function(x) {(1/(la*ga[3]))*
          (1+((x-mu)/la)^{1/ga[3]})^(-2)*
          ((x-mu)/la)^(1/ga[3]-1)}
fIV1 <- function(x) {(a[1]/(la*ga[2]))*
         (1+((x-mu)/la)^{1/ga[2]})^(-a[1]-1)*
         ((x-mu)/la)^(1/ga[2]-1)}
fIV2 <- function(x) {(a[2]/(la*ga[2]))*
         (1+((x-mu)/la)^{1/ga[2]})^(-a[2]-1)*
         ((x-mu)/la)^(1/ga[2]-1)}
fIV3 <- function(x) {(a[3]/(la*ga[2]))*
         (1+((x-mu)/la)^{1/ga[2]})^(-a[3]-1)*
         ((x-mu)/la)^(1/ga[2]-1)}
par(mfrow=c(2,2))
curve(fI1,1,10,5000,ylab="f(x)",ylim=c(0,1),lty=1,
     lwd=2,col="blue",main=expression(paste(f(x),"=",
     frac(alpha,x)(frac(x,lambda))^-alpha,;  ",sep=" ",
     lambda,"=  1")))
curve(fI2,1,10,5000,lty=1,lwd=2,col="red",add=TRUE)
curve(fI3,1,10,5000,lty=1,lwd=2,col="green",add=TRUE)
legend("topright",legend=c(expression(paste(alpha,"=0.5")),
      expression(paste(alpha,"=1.5  ")),
      expression(paste(alpha, "=3  "))),
      col=c("blue","red","green"),lty=c(1,1,1),
      lwd=c(2,2,2))
curve(fII1,2.5,10,5000,ylab="f(x)",ylim=c(0,.1),lty=1,
     lwd=2,col="blue",main=expression(paste(f(x),"=",
     frac(alpha,sigma)(1+frac(x-mu,sigma))^{-alpha-1},";  ",
     sep=" ",mu,"=2.5,",sep=" ",sigma, "=  1")))
curve(fII2,2.5,10,5000,lty=1,lwd=2,col="red",add=TRUE)
curve(fII3,2.5,10,5000,lty=1,lwd=2,col="green",add=TRUE)
legend("topright",legend=c(expression(paste(alpha,"=0.5  ")),
      expression(paste(alpha,"=1.5  ")),
      expression(paste(alpha, "=3  "))),
      col=c("blue","red","green"),lty=c(1,1,1),
      lwd=c(2,2,2))
curve(fIII1,3,10,5000,ylab="f(x)",ylim=c(0,.1),lty=1,
     lwd=2,col="blue",main=expression(paste(f(x),"=",
     frac(1,sigma*gamma)(1+(frac(x-mu,sigma))^
     {frac(1,gamma)})^{-2},(frac(x-mu,sigma))^{frac(1,gamma)-1},
     ";  ",sep=" ",mu,"=2.5,",sep=" ",sigma,"=  1")))
```

```
curve(fIII2,3,10,5000,lty=1,lwd=2,col="red",add=TRUE)
curve(fIII3,3,10,5000,lty=1,lwd=2,col="green",add=TRUE)
legend("topright",legend=c(expression(paste(gamma,"=0.2")),
        expression(paste(gamma,"=0.5  ")),
        expression(paste(gamma,"=1  "))),
        col=c("blue","red","green"),lty=c(1,1,1),
        lwd=c(2,2,2))
curve(fIV1,3,10,5000,ylab="f(x)",ylim=c(0,.1),lty=1,
      lwd=2,col="blue", main=expression(paste(f(x),"= ",
      frac(alpha,sigma*gamma)(1+(frac(x-mu,sigma))^
      {frac(1,gamma)})^{-alpha-1},
      (frac(x-mu,sigma))^{frac(1,gamma) - 1},";",sep=" ",
      mu,"=2.5,",sep=" ",sigma,"= 1,",sep=" ",gamma,"=0.5")))
curve(fIV2,3,10,5000,lty=1,lwd=2,col="red",add=TRUE)
curve(fIV3,3,10,5000,lty=1,lwd=2,col="green",add=TRUE)
legend("topright",legend=c(expression(paste(alpha,"=0.5  ")),
        expression(paste(alpha,"=1.5  ")),
        expression(paste(alpha, "=3  "))),
        col=c("blue","red","green"),lty=c(1,1,1),
        lwd=c(2,2,2))
```

B.1.2 Simulation code to compare the Pareto tail index estimators

```
########## Section 3.4.5 ##########
#### Simulation to compare EVI estimators###
### Graphics
## Generate Data
n <- 5*10^2
m <- 3
a1 <- 0.8
a2 <- 1.4
a3 <- 2
a <- c(a1,a2,a3)
P <- matrix(0,n,m)
##
##  Pareto simulation ##
library(VGAM)
## Generate sample and determine estimators
for(j in 1:m){
   P[,j] <- sort(rpareto(n,1,a[j]))
}

##  Hill, t-Hill and t-lgHill estimators
```

```
H <- matrix(0,n-1,m)
TH <- matrix(0,n-1,m)
TLH <- matrix(0,n-1,m)
M <- matrix(0,n-1,m)
#
for(j in 1:m){
    for(k in 1:(n-1)){
        d <- numeric(k)
        for(i in 1:k){
            d[i] <- P[n-i+1,j]/P[n-k,j]
        }
        H[k,j] <- (1/k)*sum(log(d))
        TH[k,j] <- ((1/k)*sum(d^{-1}))^{-1}-1
        TLH[k,j] <- sum((log(d))^2)/sum(log(d))-(1/k)*
                    sum(log(d))
        M[k,j] <- H[k,j]+1-(1/2)*(1-(H[k,j]^2)/((1/k)*
                  sum((log(d))^2)))^(-1)
    }
}

## QQ Estimator
QQ <- matrix(0,n-1,m)
for(j in 1:m){
    for(k in 1:(n-1)){
        c <- numeric(k)
        r <- numeric(k)
        for(i in 1:k){
            c[i] <- 1 - i/(k+1)
            r[i] <- P[n-k+i,j]/P[n-k,j]
        }
        QQ[k,j] <- (sum((-log(c))*log(r))-sum((-log(c))*H[k,j]))/
                   (sum((-log(c))^2)-(1/k)*(sum(-log(c)))^2)
   }
}

## Graphics
par(mfrow=c(1,3))
plot(TH[,1],xlab="k",ylab=expression(paste(hat(gamma))),
     ylim=c(0,3),col="blue",type="l",lwd=2,
     main=expression(paste(alpha,"=0.8")))
points(TLH[,1],col="red",type="l",lwd=2)
points(H[,1],col="green",type="l",lwd=2)
abline(h=1/a[1],col="black", lty=2,lwd=2)
legend("topright",legend=c("t-Hill","t-lgHill","Hill"),
        col=c("blue","red","green"),lty=c(1,1,1),lwd=c(2,2,2))
plot(TH[,2],xlab="k",ylab=expression(paste(hat(gamma))),
```

```
      ylim=c(0.4,1.2),col="blue",type="l",lwd=2,
      main=expression(paste(alpha,"=1.4")))
points(TLH[,2],col="red",type="l",lwd=2)
points(H[,2],col="green",type="l",lwd=2)
abline(h=1/a[2],col="black", lty=2,lwd=2)
legend("topright",legend=c("t-Hill","t-lgHill","Hill"),
         col=c("blue","red","green"),lty=c(1,1,1),lwd=c(2,2,2))
plot(TH[,3],xlab="k",ylab=expression(paste(hat(gamma))),
      ylim=c(0.2,0.7),col="blue",type="l",lwd=2,
      main=expression(paste(alpha,"=2")))
points(TLH[,3],col="red",type="l",lwd=2)
points(H[,3],col="green",type="l",lwd=2)
abline(h=1/a[3],col="black", lty=2,lwd=2)
legend("topright",legend=c("t-Hill","t-lgHill","Hill"),
         col=c("blue","red","green"),lty=c(1,1,1),
         lwd=c(2,2,2))
par(mfrow=c(1,3))
plot(TH[,1],xlab="k",ylab=expression(paste(hat(gamma))),
      col="blue",type="l",lwd=2,
      main=expression(paste(alpha,"=0.8")))
points(TLH[,1],col="red",type="l",lwd=2)
points(M[,1],col="green",type="l",lwd=2)
abline(h=1/a[1],col="black",lty=2,lwd=2)
legend("topright",legend=c("t-Hill","t-lgHill","Moments"),
         col=c("blue","red","green"),lty=c(1,1,1),
         lwd=c(2,2,2))
plot(TH[,2],xlab="k",ylab=expression(paste(hat(gamma))),
      col="blue",type="l",lwd=2,
      main=expression(paste(alpha,"=1.4")))
points(TLH[,2],col="red",type="l",lwd=2)
points(M[,2],col="green",type="l",lwd=2)
abline(h=1/a[2],col="black",lty=2,lwd=2)
legend("topright",legend=c("t-Hill","t-lgHill","Moments"),
         col=c("blue","red","green"),lty=c(1,1,1),
         lwd=c(2,2,2))
plot(TH[,3],xlab="k",ylab=expression(paste(hat(gamma))),
      col="blue",type="l",lwd=2,
      main=expression(paste(alpha,"=2")))
points(TLH[,3],col="red",type="l",lwd=2)
points(M[,3],col="green",type="l",lwd=2)
abline(h=1/a[3],col="black",lty=2,lwd=2)
legend("topright",legend=c("t-Hill","t-lgHill","Moments"),
         col=c("blue","red","green"),lty=c(1,1,1),
         lwd=c(2,2,2))
par(mfrow=c(1,3))
plot(TH[,1],xlab="k",ylab=expression(paste(hat(gamma))),
```

```
     col="blue",type="l",lwd=2,
     main=expression(paste(alpha,"=0.8")))
points(TLH[,1],col="red",type="l",lwd=2)
points(QQ[,1],col="green",type="l",lwd=2)
abline(h=1/a[1],col="black",lty=2,lwd=2)
legend("topright",legend=c("t-Hill","t-lgHill","QQ"),
       col=c("blue","red","green"),lty=c(1,1,1),
       lwd=c(2,2,2))
plot(TH[,2],xlab="k",ylab=expression(paste(hat(gamma))),
     col="blue",type="l",lwd=2,
     main=expression(paste(alpha,"=1.4")))
points(TLH[,2],col="red",type="l",lwd=2)
points(QQ[,2],col="green",type="l",lwd=2)
abline(h=1/a[2],col="black",lty=2,lwd=2)
legend("topright",legend=c("t-Hill","t-lgHill","QQ"),
       col=c("blue","red","green"),lty=c(1,1,1),
       lwd=c(2,2,2))
plot(TH[,3],xlab="k",ylab=expression(paste(hat(gamma))),
     col="blue",type="l",lwd=2,
     main=expression(paste(alpha,"=2")))
points(TLH[,3],col="red",type="l",lwd=2)
points(QQ[,3],col="green",type="l",lwd=2)
abline(h=1/a[3],col="black",lty=3,lwd=2)
legend("topright",legend=c("t-Hill","t-lgHill","QQ"),
       col=c("blue","red","green"),lty=c(1,1,1),
       lwd=c(2,2,2))
##
### End Graphics

### Quadratic errors
## Generate Data
n <- 10^3
m <- 10^2
a1 <- 0.8
a2 <- 1.4
a3 <- 2
P1 <- matrix(0,n,m)
P2 <- matrix(0,n,m)
P3 <- matrix(0,n,m)
for(j in 1:m){
   P1[,j] <- sort(rpareto(n,1,a[1]))
   P2[,j] <- sort(rpareto(n,1,a[2]))
   P3[,j] <- sort(rpareto(n,1,a[3]))
}
##  Hill, t-Hill and t-lgHill Estimators
H1 <- matrix(0,n-1,m)
```

```
H2 <- matrix(0,n-1,m)
H3 <- matrix(0,n-1,m)
TH1 <- matrix(0,n-1,m)
TH2 <- matrix(0,n-1,m)
TH3 <- matrix(0,n-1,m)
TLH1 <- matrix(0,n-1,m)
TLH2 <- matrix(0,n-1,m)
TLH3 <- matrix(0,n-1,m)
M1 <- matrix(0,n-1,m)
M2 <- matrix(0,n-1,m)
M3 <- matrix(0,n-1,m)
for(j in 1:m){
   for(k in 1:(n-1)){
      d1 <- numeric(k)
      d2 <- numeric(k)
      d3 <- numeric(k)
      for(i in 1:k){
         d1[i] <- P1[n-i+1,j]/P1[n-k,j]
         d2[i] <- P2[n-i+1,j]/P2[n-k,j]
         d3[i] <- P3[n-i+1,j]/P3[n-k,j]
     }
     H1[k,j] <- (1/k)*sum(log(d1))
     H2[k,j] <- (1/k)*sum(log(d2))
     H3[k,j] <- (1/k)*sum(log(d3))
     TH1[k,j] <- ((1/k)*sum(d1^{-1}))^{-1}-1
     TH2[k,j] <- ((1/k)*sum(d2^{-1}))^{-1}-1
     TH3[k,j] <- ((1/k)*sum(d3^{-1}))^{-1}-1
     TLH1[k,j] <- sum((log(d1))^2)/sum(log(d1))-(1/k)*
                  sum(log(d1))
     TLH2[k,j] <- sum((log(d2))^2)/sum(log(d2))-(1/k)*
                  sum(log(d2))
     TLH3[k,j] <- sum((log(d3))^2)/sum(log(d3))-(1/k)*
                  sum(log(d3))
     M1[k,j] <- H1[k,j]+1-(1/2)*(1-(H1[k,j]^2)/((1/k)*
                sum((log(d1))^2)))^(-1)
     M2[k,j] <- H2[k,j]+1-(1/2)*(1-(H2[k,j]^2)/((1/k)*
                sum((log(d2))^2)))^(-1)
     M3[k,j] <- H3[k,j]+1-(1/2)*(1-(H3[k,j]^2)/((1/k)*
                sum((log(d3))^2)))^(-1)
   }
}
## QQ Estimator
QQ1 <- matrix(0,n-1,m)
QQ2 <- matrix(0,n-1,m)
QQ3 <- matrix(0,n-1,m)
for(j in 1:m){
```

```
    for(k in 1:(n-1)){
        c <- numeric(k)
        r1 <- numeric(k)
        r2 <- numeric(k)
        r3 <- numeric(k)
        for(i in 1:k){
            c[i] <- 1 - i/(k+1)
            r1[i] <- P1[n-k+i,j]/P1[n-k,j]
            r2[i] <- P2[n-k+i,j]/P2[n-k,j]
            r3[i] <- P3[n-k+i,j]/P3[n-k,j]
        }
        QQ1[k,j] <- (sum((-log(c))*log(r1))-
                    sum((-log(c))*H1[k,j]))/
                    (sum((-log(c))^2)-(1/k)*(sum(-log(c)))^2)
        QQ2[k,j] <- (sum((-log(c))*log(r2))-
                    sum((-log(c))*H2[k,j]))/
                    (sum((-log(c))^2)-(1/k)*(sum(-log(c)))^2)
        QQ3[k,j] <- (sum((-log(c))*log(r3))-
                    sum((-log(c))*H3[k,j]))/
                    (sum((-log(c))^2)-(1/k)*(sum(-log(c)))^2)
    }
}
## Graphics
par(mfrow=c(1,3))
plot(TH1[,1],xlab="k",ylab=" ",col="blue",type="l",
     lwd=2,main=expression(paste(alpha,"=0.8")))
points(TLH1[,1],col="red",type="l",lwd=2)
points(H1[,1],col="green",type="l",lwd=2)
legend("topright",legend=c("t-Hill","t-lgHill","Hill"),
        col=c("blue","red","green"),lty=c(1,1,1),
        lwd=c(2,2,2))
plot(TH2[,1],xlab="k",ylab=" ",col="blue",type="l",
     lwd=2,main=expression(paste(alpha,"=1.4")))
points(TLH2[,1],col="red",type="l",lwd=2)
points(H2[,1],col="green",type="l",lwd=2)
legend("topright",legend=c("t-Hill","t-lgHill","Hill"),
        col=c("blue","red","green"),lty=c(1,1,1),
        lwd=c(2,2,2))
plot(TH3[,1],xlab="k",ylab=" ",col="blue",type="l",
     lwd=2,main=expression(paste(alpha,"=2")))
points(TLH3[,1],col="red",type="l",lwd=2)
points(H3[,1],col="green",type="l",lwd=2)
legend("topright",legend=c("t-Hill","t-lgHill","Hill"),
        col=c("blue","red","green"),lty=c(1,1,1),
        lwd=c(2,2,2))
par(mfrow=c(1,3))
```

```
plot(TH1[,1],xlab="k",ylab=" ",col="blue",type="l",
     lwd=2,main=expression(paste(alpha,"=0.8")))
points(M1[,1],col="purple",type="l",lwd=2)
points(QQ1[,1],col="coral",type="l",lwd=2)
legend("topright",legend=c("t-Hill","Moments","QQ"),
        col=c("blue","purple","coral"),lty=c(1,1,1),
        lwd=c(2,2,2))
plot(TH2[,1],xlab="k",ylab=" ",col="blue",type="l",
     lwd=2,main=expression(paste(alpha,"=1.4")))
points(M2[,1],col="purple",type="l",lwd=2)
points(QQ2[,1],col="coral",type="l",lwd=2)
legend("topright",legend=c("t-Hill","Moments","QQ"),
        col=c("blue","purple","coral"),lty=c(1,1,1),
        lwd=c(2,2,2))
plot(TH3[,1],xlab="k",ylab=" ",col="blue",type="l",
     lwd=2,main=expression(paste(alpha,"=2")))
points(M3[,1],col="purple",type="l",lwd=2)
points(QQ3[,1],col="coral",type="l",lwd=2)
legend("topright",legend=c("t-Hill","Moments","QQ"),
        col=c("blue","purple","coral"),lty=c(1,1,1),
        lwd=c(2,2,2))
## Errors
RB <- matrix(0,5,3)
RRMSE <- matrix(0,5,3)
RB[1,1] <- (1/m)*sum(H1[n-1,]-(1/a[1]))*(1/a[1])*100
RB[2,1] <- (1/m)*sum(TH1[n-1,]-(1/a[1]))*(1/a[1])*100
RB[3,1] <- (1/m)*sum(TLH1[n-1,]-(1/a[1]))*(1/a[1])*100
RB[4,1] <- (1/m)*sum(M1[n-1,]-(1/a[1]))*(1/a[1])*100
RB[5,1] <- (1/m)*sum(QQ1[n-1,]-(1/a[1]))*(1/a[1])*100
RB[,1]
#
RB[1,2] <- (1/m)*sum(H2[n-1,]-(1/a[2]))*(1/a[2])*100
RB[2,2] <- (1/m)*sum(TH2[n-1,]-(1/a[2]))*(1/a[2])*100
RB[3,2] <- (1/m)*sum(TLH2[n-1,]-(1/a[2]))*(1/a[2])*100
RB[4,2] <- (1/m)*sum(M2[n-1,]-(1/a[2]))*(1/a[2])*100
RB[5,2] <- (1/m)*sum(QQ2[n-1,]-(1/a[2]))*(1/a[2])*100
RB[,2]
#
RB[1,3] <- (1/m)*sum(H3[n-1,]-(1/a[3]))*(1/a[3])*100
RB[2,3] <- (1/m)*sum(TH3[n-1,]-(1/a[3]))*(1/a[3])*100
RB[3,3] <- (1/m)*sum(TLH3[n-1,]-(1/a[3]))*(1/a[3])*100
RB[4,3] <- (1/m)*sum(M3[n-1,]-(1/a[3]))*(1/a[3])*100
RB[5,3] <- (1/m)*sum(QQ3[n-1,]-(1/a[3]))*(1/a[3])*100
RB[,3]
#
RRMSE[1,1] <- sqrt((1/m)*sum((H1[n-1,]-(1/a[1]))^2))*
```

```
                   (1/a[1])*100
RRMSE[2,1] <- sqrt((1/m)*sum((TH1[n-1,]-(1/a[1]))^2))*
                   (1/a[1])*100
RRMSE[3,1] <- sqrt((1/m)*sum((TLH1[n-1,]-(1/a[1]))^2))*
                   (1/a[1])*100
RRMSE[4,1] <- sqrt((1/m)*sum((M1[n-1,]-(1/a[1]))^2))*
                   (1/a[1])*100
RRMSE[5,1] <- sqrt((1/m)*sum((QQ1[n-1,]-(1/a[1]))^2))*
                   (1/a[1])*100
RRMSE[,1]
#
RRMSE[1,2] <- sqrt((1/m)*sum((H2[n-1,]-(1/a[2]))^2))*
                   (1/a[2])*100
RRMSE[2,2] <- sqrt((1/m)*sum((TH2[n-1,]-(1/a[2]))^2))*
                   (1/a[2])*100
RRMSE[3,2] <- sqrt((1/m)*sum((TLH2[n-1,]-(1/a[2]))^2))*
                   (1/a[2])*100
RRMSE[4,2] <- sqrt((1/m)*sum((M2[n-1,]-(1/a[2]))^2))*
                   (1/a[2])*100
RRMSE[5,2] <- sqrt((1/m)*sum((QQ2[n-1,]-(1/a[2]))^2))*
                   (1/a[2])*100
RRMSE[,2]
#
RRMSE[1,3] <- sqrt((1/m)*sum((H3[n-1,]-(1/a[3]))^2))*
                   (1/a[3])*100
RRMSE[2,3] <- sqrt((1/m)*sum((TH3[n-1,]-(1/a[3]))^2))*
                   (1/a[3])*100
RRMSE[3,3] <- sqrt((1/m)*sum((TLH3[n-1,]-(1/a[3]))^2))*
                   (1/a[3])*100
RRMSE[4,3] <- sqrt((1/m)*sum((M3[n-1,]-(1/a[3]))^2))*
                   (1/a[3])*100
RRMSE[5,3] <- sqrt((1/m)*sum((QQ3[n-1,]-(1/a[3]))^2))*
                   (1/a[3])*100
RRMSE[,3]
## Errors 1:Hill 2:t-Hill 3:t-lgHill 4:Moments 5:QQ
RB1 <- matrix(0,3,3)
RB2 <- matrix(0,3,3)
RB3 <- matrix(0,3,3)
RB4 <- matrix(0,3,3)
RB5 <- matrix(0,3,3)
RRMSE1 <- matrix(0,3,3)
RRMSE2 <- matrix(0,3,3)
RRMSE3 <- matrix(0,3,3)
RRMSE4 <- matrix(0,3,3)
RRMSE5 <- matrix(0,3,3)
for(i in 1:3){
```

```
    RB1[1,i] <- (1/(n-1)*sum(H1-(1/a1))*(1/a1)*100
    RB1[2,i] <- (1/m)*sum(H2[id[i],]-(1/a2))*(1/a2)*100
    RB1[3,i] <- (1/m)*sum(H3[id[i],]-(1/a3))*(1/a3)*100
    RB2[1,i] <- (1/m)*sum(TH1[id[i],]-(1/a1))*(1/a1)*100
    RB2[2,i] <- (1/m)*sum(TH2[id[i],]-(1/a2))*(1/a2)*100
    RB2[3,i] <- (1/m)*sum(TH3[id[i],]-(1/a3))*(1/a3)*100
    RB3[1,i] <- (1/m)*sum(TLH1[id[i],]-(1/a1))*(1/a1)*100
    RB3[2,i] <- (1/m)*sum(TLH2[id[i],]-(1/a2))*(1/a2)*100
    RB3[3,i] <- (1/m)*sum(TLH3[id[i],]-(1/a3))*(1/a3)*100
    RB4[1,i] <- (1/m)*sum(M1[id[i],]-(1/a1))*(1/a1)*100
    RB4[2,i] <- (1/m)*sum(M2[id[i],]-(1/a2))*(1/a2)*100
    RB4[3,i] <- (1/m)*sum(M3[id[i],]-(1/a3))*(1/a3)*100
    RB5[1,i] <- (1/m)*sum(QQ1[id[i],]-(1/a1))*(1/a1)*100
    RB5[2,i] <- (1/m)*sum(QQ2[id[i],]-(1/a2))*(1/a2)*100
    RB5[3,i] <- (1/m)*sum(QQ3[id[i],]-(1/a3))*(1/a3)*100
    RRMSE1[1,i] <- sqrt((1/m)*sum((H1[id[i],]-(1/a1))^2))*
                   (1/a1)*100
    RRMSE1[2,i] <- sqrt((1/m)*sum((H2[id[i],]-(1/a2))^2))*
                   (1/a2)*100
    RRMSE1[3,i] <- sqrt((1/m)*sum((H3[id[i],]-(1/a3))^2))*
                   (1/a3)*100
    RRMSE2[1,i] <- sqrt((1/m)*sum((TH1[id[i],]-(1/a1))^2))*
                   (1/a1)*100
    RRMSE2[2,i] <- sqrt((1/m)*sum((TH2[id[i],]-(1/a2))^2))*
                   (1/a2)*100
    RRMSE2[3,i] <- sqrt((1/m)*sum((TH3[id[i],]-(1/a3))^2))*
                   (1/a3)*100
    RRMSE3[1,i] <- sqrt((1/m)*sum((TLH1[id[i],]-(1/a1))^2))*
                   (1/a1)*100
    RRMSE3[2,i] <- sqrt((1/m)*sum((TLH2[id[i],]-(1/a2))^2))*
                   (1/a2)*100
    RRMSE3[3,i] <- sqrt((1/m)*sum((TLH3[id[i],]-(1/a3))^2))*
                   (1/a3)*100
    RRMSE4[1,i] <- sqrt((1/m)*sum((M1[id[i],]-(1/a1))^2))*
                   (1/a1)*100
    RRMSE4[2,i] <- sqrt((1/m)*sum((M2[id[i],]-(1/a2))^2))*
                   (1/a2)*100
    RRMSE4[3,i] <- sqrt((1/m)*sum((M3[id[i],]-(1/a3))^2))*
                   (1/a3)*100
    RRMSE5[1,i] <- sqrt((1/m)*sum((QQ1[id[i],]-(1/a1))^2))*
                   (1/a1)*100
    RRMSE5[2,i] <- sqrt((1/m)*sum((QQ2[id[i],]-(1/a2))^2))*
                   (1/a2)*100
    RRMSE5[3,i] <- sqrt((1/m)*sum((QQ3[id[i],]-(1/a3))^2))*
                   (1/a3)*100
}
```

B.2 Code example for two-dimensional variables

B.2.1 Code for dependence of random variables

```
########## Section 5.3.2 ##########
### Parinacota Data
## Read Data
dpa <- read.table("C:/Users/Ludy/Desktop/Tesis/
       Datosyscript/Parinacota2.txt", head=TRUE)
head(dpa)

X <- dpa$Arse
n <- length(X)
## Assign values to variables U1 and U2
U1 <- sort(X[1:(n-1)])
U2 <- sort(X[2:n])
summary(U1)
summary(U2)
sd(U1)
sd(U2)
E1 <- ecdf(U1)
E2 <- ecdf(U2)
S1 <- 1- E1(U1)
S2 <- 1- E2(U2)
ep <- 10^(-8)
S1 <- S1 + ep
S2 <- S2 + ep
Y1 <- log(-log(S1))
Y2 <- log(-log(S2))
V <- rep(1,n-1)
LT1 <- log(U1)
LT2 <- log(U2)
X1 <- cbind(V,LT1)
X2 <- cbind(V,LT2)
## Estimate the coefficients of the models
B1 <- (solve(t(X1)%*%X1))%*%(t(X1)%*%Y1)
B2 <- (solve(t(X2)%*%X2))%*%(t(X2)%*%Y2)
a1 <- exp(B1[1,1]); a1
a2 <- exp(B2[1,1]); a2
b1 <- B1[2,1]; b1
b2 <- B2[2,1]; b2
## Determine the bivariate joint function
G <- function(x,y) {1-exp(-a1*pmax(x,y))}
G(U1,U2)
```

```
S <- 1-G(U1,U2)
## Graph the bivariate joint function and survival function
par(mfrow=c(1,2))
plot(G(U1,U2), xlab="i",ylab="G(u1,u2)",col="blue",type="l",
     lty=2, main="Bivariate min extreme value distribution")
plot(S, xlab="i",ylab="S(u1,u2)",col="blue",type="l",
     lty=2, main="Survival functional")

open3d()
plot3d(U1,U2,G,col="blue", xlab="u1",ylab="u2",
       zlab="G(u1,u2)",type="s",lty=2)
```