Multivariate Regular Variation and its applications

by

Diouldé Habibatou Mariko

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the Master of Science in Mathematics

Department of Mathematics and Statistics
Faculty of Science
University of Ottawa

Abstract

In this thesis, we review the basic notions related to univariate regular variation and study some fundamental properties of regularly varying random variables. We then consider the notion of regular variation in the multivariate case. After collecting some results from multivariate regular variation for random vectors with values in $\mathbb{R}_+^d$, we discuss its properties and examine several examples of multivariate regularly varying random vectors such as independent and identically distributed random vectors, fully dependent random vectors and other models. We also present the elements of univariate and multivariate extreme value theory and emphasize the connection with multivariate regular variation. Some measures of extreme dependence such as the stable tail dependence function and the Pickands dependence function are presented. We end the study by conducting a data analysis using financial data. In the univariate case, graphical tools such as quantile-quantile plots, mean excess plots and Hill plots are used in order to determine the underlying distribution of the univariate data. In the multivariate case, non-parametric estimators of the stable tail dependence function and the Pickands dependence function are used to describe the dependence structure of the multivariate data.

Acknowledgements

I would like to express my deepest gratitude to my supervisors Professor Raluca Balan and Professor Rafal Kulik for their valuable guidance, their engagement, their precise organization and their patience throughout the learning process of this master's thesis. Without their supervision and constant help, this thesis would not have been possible. Working with them has been an absolute pleasure and an unforgettable experience.

No words can begin to express how grateful I am for my family. Thank you for the unfailing faith, support, and unconditional love you have provided throughout my entire studies. Thank you for inspiring me, encouraging me on a daily basis and keeping my morale up during the difficult periods of doubt. Even though we were miles away, it felt like you were right beside me, cheering me on.

I would also like to thank my friends and house mates who have always been there for me when I needed a break from writing. The memories we have shared the past few years have made this experience even more amazing.

I also place on record, my sense of gratitude to one and all, who directly or indirectly, have made this goal of mine possible.

Last but not least, I dedicate this thesis to my father Khalilou Mariko who passed away shortly before I started this master's degree. He always believed in me and motivated me to be a better individual. He taught me that with hard work, faith and dedication I could accomplish anything I wanted in life and for that I am forever grateful.

Contents

Chapter 1

Introduction

Heavy-tail analysis has played an increasingly important role in insurance, finance and risk management in the recent years and has been dealt with in various research papers and books such as [1] and [10]. The distribution of a random variable X is said to have heavy-tails if it has a large probability of getting very large values. The most common example is the Pareto distribution. The right tail of the distribution is generally the tail that is of interest. It is well known that financial returns are usually heavy-tailed and therefore it is crucial for those in charge of the corresponding risk management calculations to rely on heavy-tailed methods.

One of the most important classes of distributions in the heavy-tail analysis is the regular variation class, which includes all the distributions with a regularly varying tail. Recall that a function f is regularly varying with index α if

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

A classical estimator of the tail index α is the Hill estimator. To efficiently characterize the multivariate heavy-tailed phenomena, de Haan and Ferreira ([4]) as well as Resnick ([10]) extended the regularly varying distributions to the multivariate case, leading to the concept of multivariate regular variation. This concept has been widely used to analyse multivariate extremes since there is a natural connection between the limit theory of the component-wise maximum of independent and identically distributed random vectors and the theory of multivariate regular variation.

The goal of this thesis was to review and study some already known results from the literature on regular variation and extreme value theory and apply this theory to financial data. When the data display dependence among extreme values, inferences based on multivariate tail probabilities are needed. In order to describe the dependence structure of multivariate data, we will need extreme value theory. It is essential to quantify the strength of the dependence in the tails since it can heavily influence the risk assessment. Another way to investigate strength of dependence in the tails is through copulas, a copula being a multivariate probability distribution for which the marginal probability distribution of each variable is uniform. However, we will not address this topic in this thesis.

Extreme value theory is a branch of statistics dealing with the likelihood of an extremely unlikely event. The probability that a data point that deviates significantly from the mean will occur is of interest. Extreme value analysis is used in many disciplines, such as finance, hydrology, engineering, climate and environment (see [12]). For example, extreme value analysis might be used in the field of hydrology to estimate the probability of an unusually large flooding event. The link between regular variation and extreme value theory comes from the fact that a random vector with positive components is in the max-domain of attraction of a multivariate extreme value distribution with Fréchet marginals if and only if the vector has a distribution which is regularly varying. This fact is clearly demonstrated in Chapter 4. Hence, regular variation of the tail of a multivariate distribution function is an essential tool for describing the max-domain of attraction of a multivariate extreme value distribution.

The mathematical framework of multivariate regular variation comes into play by providing tools to compute tail probabilities and is used to give a semi-parametric class of distribution functions, where we know for instance, how the extremes behave according to the Fisher-Tippett-Gnedenko Theorem. The previously mentioned theorem stipulates that the component-wise maximum of a sample of independent and identically distributed random variables after proper normalization can only converge in distribution to one of three possible distributions: the Gumbel distribution, the Fréchet distribution or the Weibull distribution, which allows us to classify distributions into classes of max-domains of attraction for one of the three types of extreme value distributions. In this thesis we will only focus on the Fréchet max-domain of attraction. Once we know more about the dependence structure of the multivariate data, it is easier to construct the univariate marginals and to make assumptions about the underlying distribution of the univariate data.

This thesis is structured as follows. In Chapter 2, we review the basic notions related to univariate regular variation and summarize some fundamental properties of regularly varying random variables. In Chapter 3, we consider the notion of regular variation in the multivariate case. In this chapter we study the concept of multivariate regular variation for random vectors with values in $\mathbb{R}_+^d$ and we discuss its properties. We then examine several examples of multivariate regularly varying random vectors such as independent and identically distributed random vectors, fully dependent random vectors and other models. In Chapter 4, we present the elements of univariate and multivariate extreme value theory and highlight the connection with multivariate regular variation. We also present some measures of extreme dependence such as the stable tail dependence function and the Pickands dependence function. Finally, in Chapter 5, we conduct some data analysis. In the univariate case, we use graphical tools such as quantile-quantile plots, mean excess plots and Hill plots in order to determine the underlying distribution of the univariate data. In the multivariate case, we used non-parametric estimators of the stable tail dependence function and the Pickands dependence function to describe the dependence structure of the multivariate data. Some additional measure-theoretic probability results are offered in Appendix A. Appendix B contains the R codes used for the simulations in Chapter 5.

Chapter 2

Univariate regular variation

In this chapter, we present the basic definitions related to regular variation and collect some fundamental properties of regularly varying random variables.

2.1 Regular variation of real valued functions

In this section, we review some basic properties of regular varying functions.

2.1.1 Slowly varying functions

Definition 2.1.1. *Let ℓ:$(0,\infty) \to (0,\infty)$ be a real valued function. We say that ℓ is **slowly varying** if*

$$\lim_{x\to+\infty} \frac{\ell(tx)}{\ell(x)} = 1, \forall t > 0.$$

In this case,we write $\ell \in SV$.

Example 2.1.2. Let $\ell(x) = \ln(x)$.Then

$$\lim_{x\to+\infty} \frac{\ell(tx)}{\ell(x)} = \lim_{x\to+\infty} \frac{\ln(tx)}{\ln(x)} = \lim_{x\to+\infty} \frac{\ln(t)+\ln(x)}{\ln(x)} = \lim_{x\to+\infty} \frac{\ln(t)}{\ln(x)} + \lim_{x\to+\infty} \frac{\ln(x)}{\ln(x)} = 0+1 = 1.$$

Hence the function $\ell(x) = \ln(x) \in SV$.

Property 2.1.3.

1. *If $\ell \in SV$, then $\ell^\alpha \in SV$ for any $\alpha \in \mathbb{R}$.*
2. *If $\ell_1, \ell_2 \in SV$, then $\ell_1\ell_2 \in SV$.*
3. *If $\ell_1, \ell_2 \in SV$ and $\lim_{x\to+\infty} \ell_1(x)$, $\lim_{x\to+\infty} \ell_2(x)$ exist and are finite, then $\ell_1 + \ell_2 \in SV$.*

Proof. Here we will only prove properties 2 and 3. Property 1 has a trivial proof.

2. We have
$$\lim_{x\to+\infty}\frac{\ell_1(tx)\ell_2(tx)}{\ell_1(x)\ell_2(x)}=\lim_{x\to+\infty}\frac{\ell_1(tx)}{\ell_1(x)}\cdot\lim_{x\to+\infty}\frac{\ell_2(tx)}{\ell_2(x)}=1\cdot 1=1.$$

3. Denote $\lim\limits_{x\to+\infty}\ell_1(x)=c_1$ and $\lim\limits_{x\to+\infty}\ell_2(x)=c_2$. Then
$$\begin{aligned}\lim_{x\to+\infty}\frac{(\ell_1+\ell_2)(tx)}{(\ell_1+\ell_2)(x)}&=\lim_{x\to+\infty}\frac{\ell_1(tx)+\ell_2(tx)}{\ell_1(x)+\ell_2(x)}=\lim_{x\to+\infty}\frac{\ell_1(x)\frac{\ell_1(tx)}{\ell_1(x)}+\ell_2(x)\frac{\ell_2(tx)}{\ell_2(x)}}{\ell_1(x)+\ell_2(x)}\\&=\frac{c_1+c_2}{c_1+c_2}=1.\end{aligned}$$

□

One of the several key results (that give the theory power and utility) is the Karamata's Representation Theorem (see, e.g., Corollary 2.1 in [10]).

Theorem 2.1.4 (Karamata's Representation Theorem)**.** *For any $\ell\in SV$ we can write*
$$\ell(x)=c(x)\exp\left(\int_{x_0}^{x}\frac{\eta(t)}{t}dt\right),\ \textit{for all } x>x_0,$$
where $\lim\limits_{x\to+\infty}\eta(x)=0$ *and* $\lim\limits_{x\to+\infty}c(x)=c_0$*.*

One of the several useful properties of regularly varyong functions is know as Potter's Bounds (see, e.g., Proposition 2.6 (ii) of [10] (with $\rho=0$)).

Theorem 2.1.5 (Potter's Bounds)**.** *Let $\ell\in SV$ such that ℓ is locally bounded away from zero (i.e $\ell(x)\geq a>0$ for some $a>0$ for all $x\in\mathbb{R}$). For all $\varepsilon>0$ there exists a constant $t_0>0$ such that*
$$(1-\varepsilon)x^{-\varepsilon}\leq\frac{\ell(tx)}{\ell(t)}\leq(1+\varepsilon)x^{\varepsilon},\ \textit{for all } x\geq 1 \textit{ and } t\geq t_0.$$

2.1.2 Regularly varying functions

Definition 2.1.6. *Let f: $(0,\infty)\to(0,\infty)$ be a measurable function. We say that f is* ***regularly varying*** *with index $\alpha\in\mathbb{R}$ if*
$$\lim_{x\to+\infty}\frac{f(tx)}{f(x)}=t^{\alpha},\ \textit{for all } t>0.$$
In this case, we write $f\in RV_\alpha$.

Theorem 2.1.7 (Representation Theorem)**.** *Any function $f \in RV_\alpha$ with $\alpha \in \mathbb{R}$ can be written as $f(t) = t^\alpha \ell(t)$, where ℓ is a slowly varying function. In the case of a slowly varying function, $\alpha = 0$.*

Proof. By Definition 2.1.6, we know that $\lim\limits_{x\to+\infty} \frac{f(tx)}{t^\alpha f(x)} = 1$ for any $t > 0$. Hence the function ℓ defined by $\ell(t) = t^{-\alpha} f(t)$ is slowly varying since

$$\lim_{x\to+\infty} \frac{\ell(tx)}{\ell(x)} = \lim_{x\to+\infty} \frac{(tx)^{-\alpha} f(tx)}{x^{-\alpha} f(x)} = \lim_{x\to+\infty} \frac{t^{-\alpha} f(tx)}{f(x)} = 1.$$

Therefore, $f(t) = t^\alpha \ell(t)$. □

2.2 Regular variation of random variables

Definition 2.2.1. *Let X be an arbitrary random variable and $\overline{F}(x) = P(X > x)$ be its tail distribution function. We say that X is **regularly varying** if the following two conditions are satisfied:*

- $\lim\limits_{x\to+\infty} \overline{F}(tx)/\overline{F}(x) = t^{-\alpha}$, *i.e $\overline{F} \in RV_{-\alpha}$ for some $\alpha > 0$;*
- *there exists $p \in [0, 1]$ such that*

$$\lim_{x\to+\infty} \frac{P(X > x)}{P(|X| > x)} = p. \tag{2.1}$$

In this case,we write $X \in RV_{-\alpha}$.

Based on the Representation Theorem, the tail distribution function of a positive random variable X can be written as $\overline{F}(x) = x^{-\alpha}\ell(x)$, where ℓ is a slowly regularly function.

Lemma 2.2.2. *Let $X \in RV_{-\alpha}$ for some $\alpha > 0$. For any $\varepsilon > 0$, there exists $c > 0$ such that for any $x, y > 0$*

$$\frac{P(yX > x)}{P(X > x)} \leq c(1 \vee y)^{\alpha+\varepsilon},$$

where $\vee$ denotes the maximum.

Proof. The proof we present here is taken from [7]. Let $X \in RV_{-\alpha}$. Hence, $P(X > x) = x^{-\alpha}\ell(x)$, for some $\ell \in SV$.

- If $y \le 1$, then:

$$P(yX > x) \le P(X > x) \text{ and } \frac{P(yX > x)}{P(X > x)} \le 1 \le (1 \vee y)^{\alpha+\varepsilon}.$$

- If $y \ge 1$, then by Theorem 2.1.5, we have:

$$\begin{aligned} \frac{P(yX > x)}{P(X > x)} &= \frac{\overline{F}_X(\frac{x}{y})}{\overline{F}_X(x)} = \frac{(\frac{x}{y})^{-\alpha}\ell(\frac{x}{y})}{x^{-\alpha}\ell(x)} = \frac{y^\alpha \ell(\frac{x}{y})}{\ell(x)} \le y^\alpha c_\varepsilon \left(\frac{x}{yx}\right)^\varepsilon = c_\varepsilon y^{\alpha-\varepsilon} \le c_\varepsilon y^{\alpha+\varepsilon} \\ &= c_\varepsilon (1 \vee y)^{\alpha+\varepsilon}. \end{aligned}$$

□

Lemma 2.2.3 (Breiman's lemma)**.** *Let X and Y be two independent positive random variables such that $X \in RV_{-\alpha}$. If there exist $\varepsilon > 0$ such that $E(Y^{\alpha+\varepsilon}) < \infty$, then*

$$\frac{P(XY > x)}{P(X > x)} \to E(Y^\alpha), \text{ as } x \to \infty.$$

Proof. The proof is borrowed from [7]. Define

$$G_x(y) = \frac{P(yX > x)}{P(X > x)} = y^\alpha \frac{\ell(\frac{x}{y})}{\ell(x)},$$

where we used Representation theorem for the second equality. Since $\ell \in SV$, for any $y > 0$, $\lim\limits_{x\to+\infty} G_x(y) = y^\alpha$ (i.e. $G_x(y)$ converges pointwise to y^α when $x \to \infty$). Note that $G_x(y) = \frac{P(yX>x)}{P(X>x)} \le (1 \vee y)^{\alpha+\varepsilon}$ (see Lemma 2.2.2).

Now, we will show that $\lim\limits_{x\to+\infty} E(G_x(Y)) = E(Y^\alpha)$. We denote by F_Y the distribution function of Y. Then

$$\begin{aligned} \lim_{x\to+\infty} E(G_x(Y)) &= \lim_{x\to+\infty} \int_{\mathbb{R}} G_x(y) F_Y(dy) = \lim_{x\to+\infty} \int_{\mathbb{R}} y^\alpha \frac{\ell(\frac{x}{y})}{\ell(x)} F_Y(dy) \\ &= \int_{\mathbb{R}} y^\alpha \lim_{x\to+\infty} \frac{\ell(\frac{x}{y})}{\ell(x)} F_Y(dy) = \int_{\mathbb{R}} y^\alpha F_Y(dy) = E(Y^\alpha). \end{aligned}$$

The interchange of the limit with the integral is justified by the dominated convergence theorem (see Theorem A.1.5) using the bound obtained above for $G_x(y)$ and the fact that $E[(1 \vee Y)^{\alpha+\varepsilon}] = E(Y^{\alpha+\varepsilon} 1_{\{Y>1\}}) + E(1_{\{Y\le1\}}) < \infty$. □

Lemma 2.2.4. *Let X and Y be independent nonnegative random variables such that*

- $P(X > x) = x^{-\alpha} L(x)$, *for some* $\alpha > 0$*;*

- $\lim\limits_{x\to+\infty} \frac{P(Y>x)}{P(X>x)} = 0.$

Then

$$\lim_{x\to+\infty} \frac{P(X+Y>x)}{P(X>x)} = 1. \tag{2.2}$$

Proof. We use the argument of [12]. Let $\varepsilon \in (0,1)$ be arbitrary. Then

$$\begin{aligned}
\limsup_{x\to+\infty} \frac{P(X+Y>x)}{P(X>x)} &= \limsup_{x\to+\infty} \frac{P(X+Y>x, Y>\varepsilon x) + P(X+Y>x, Y\le \varepsilon x)}{P(X>x)} \\
&\le \limsup_{x\to+\infty} \frac{P(Y>\varepsilon x) + P(X+\varepsilon x > x)}{P(X>x)} \\
&= \limsup_{x\to+\infty} \frac{P(Y>\varepsilon x) + P(X > x(1-\varepsilon))}{P(X>x)} \\
&\le \limsup_{x\to+\infty} \frac{P(Y>\varepsilon x)}{P(X>\varepsilon x)} \frac{P(X>\varepsilon x)}{P(X>x)} + \limsup_{x\to+\infty} \frac{P(X>x(1-\varepsilon))}{P(X>x)} \\
&= 0\cdot \varepsilon^{-\alpha} + (1-\varepsilon)^{-\alpha} = (1-\varepsilon)^{-\alpha}.
\end{aligned}$$

Taking $\varepsilon \to 0$, we conclude that

$$\lim_{\varepsilon\to 0} \limsup_{x\to+\infty} \frac{P(X+Y>x)}{P(X>x)} \le 1. \tag{2.3}$$

For the other inequality, we note that since Y is nonnegative, we have

$$P(X+Y>x) \ge P(X>x)$$

Hence

$$\liminf_{x\to+\infty} \frac{P(X+Y>x)}{P(X>x)} \ge \liminf_{x\to+\infty} \frac{P(X>x)}{P(X>x)} = 1. \tag{2.4}$$

Relation (2.2) follows from (2.3) and (2.4). $\square$

2.3 Vague convergence

In this section, we review the concept of vague convergence, which is used for giving an alternative definition of regular variation.

Definition 2.3.1. *Let E be a LCCB space (i.e a locally compact space with a countable basis), and $\mathcal{E}$ its Borel σ-field. We say that a measure μ on $(E,\mathcal{E})$ is* ***Radon*** *if $\mu(A) < \infty$ for any compact set $A \subset E$. We denote by $M_+(E)$ the set of Radon measures on E.*

Definition 2.3.2. *Let E be a LCCB space and $\mathcal{E}$ its Borel σ-field. Let $\{\mu_n\}_{n\in\mathbb{N}}$ be a sequence of Radon measures defined on $(E, \mathcal{E})$. We say that $\{\mu_n\}_{n\in\mathbb{N}}$* ***converges vaguely*** *to μ if for any continuous function $f : E \to [0, +\infty)$ with compact support*

$$\int_\Omega f(x)\nu_n(dx) \to \int_\Omega f(x)\nu(dx), \text{ as } n \to \infty.$$

In this case we write $\mu_n \xrightarrow{v} \mu$.

Remark 2.3.3. We have $\mu_n \xrightarrow{v} \mu$ if and only if $\mu_n(B) \xrightarrow{v} \mu(B)$ for any relatively compact set in B such that $\mu(\partial B) = 0$ (see [6]). Here $\partial B = \overline{B} \cap B^\circ$, where $\overline{B}$ is the closure of B and B° is the interior of B.

The main result of this section is the following equivalence between the concept of regular variation and vague convergence (cf. Theorem 3.6 of [10]).

Theorem 2.3.4. *Suppose that X is a nonnegative random variable with distribution function $F(x)$. Set $\overline{F}(x) = 1 - F(x)$. The following statements are equivalent:*

(i) There exists $\alpha > 0$ such that $X \in RV_{-\alpha}$.

(ii) There exists $\alpha > 0$ and a sequence $\{a_n\}$ with $a_n \to \infty$ such that

$$\lim_{n\to+\infty} n\overline{F}(a_n x) = x^{-\alpha}, x > 0. \tag{2.5}$$

(iii) There exists $\alpha > 0$ and a sequence $\{a_n\}$ with $a_n \to \infty$ such that

$$\mu_n(\cdot) := n\mathbb{P}\left(\frac{X}{a_n} \in \cdot\right) \xrightarrow{v} \nu_\alpha(\cdot) \text{ in } M_+((0,\infty]),$$

where $\nu_\alpha((x,\infty]) = x^{-\alpha}$ for all $x > 0$. Here $M_+((0,\infty])$ is the space of Radon measures on $(0,\infty]$.

In order to prove Theorem 2.3.4, we consider first the following lemma (cf. Proposition 2.3 of [10]):

Lemma 2.3.5. *(i) A measurable function $U : (0,\infty) \to (0,\infty)$ is regularly varying (with index $\alpha \in \mathbb{R}$) if and only if there exists a function $h : (0,\infty) \to (0,\infty)$ such that*

$$\lim_{x\to+\infty} \frac{U(tx)}{U(t)} = h(x), \text{ for any } x > 0. \tag{2.6}$$

In this case, $h(x) = x^\alpha$, for all $x > 0$.

(ii) *A monotone function $U : (0, \infty) \to (0, \infty)$ is regularly varying if and only if there exist two sequences $\{a_n\}$ and $\{\lambda_n\}$ such that*

$$a_n \to \infty \quad \text{and} \quad \frac{\lambda_n}{\lambda_{n+1}} \to 1, \ \text{as } n \to \infty,$$

and for all $x > 0$,

$$\lim_{n\to\infty} \lambda_n U(a_n x) =: \chi(x) \ \text{exists, is positive and definite.} \tag{2.7}$$

Then $U \in RV_\alpha$ for some $\alpha \in \mathbb{R}$ and $\chi(x) = \chi(1)x^\alpha$ for all $x > 0$.

Proof. We only need to prove that (2.6) implies that $U \in RV$.

(i) The function h is measurable since it is a limit of a family of measurable functions. Then for $x > 0$ and $y > 0$,

$$\lim_{t\to+\infty} \frac{U(txy)}{U(t)} = \lim_{t\to+\infty} \left(\frac{U(txy)}{U(tx)} \cdot \frac{U(tx)}{U(x)} \right) = h(y)h(x).$$

Therefore, $h(xy) = h(x)h(y)$ for all $x, y > 0$. By a well-known criterion, it follows that $h(x) = x^\alpha$ for some $\alpha \in \mathbb{R}$.

(ii) We assume that U is non-decreasing. (The case when U is non-increasing is similar.) Since $a_n \to \infty$, for each t there exists a finite $n(t)$ defined by:

$$n(t) = \inf\{n : a_{n+1} > t\}$$

so that, $a_{n(t)} \le t \le a_{n(t)+1}$. Therefore, by monotonicity of U, for any $x > 0$

$$\lim_{t\to\infty} \left(\frac{\lambda_{n(t)+1}}{\lambda_{n(t)}} \right) \left(\frac{\lambda_{n(t)} U(a_{n(t)}x)}{\lambda_{n(t)+1} U(a_{n(t)+1})} \right) \le \lim_{t\to\infty} \frac{U(tx)}{U(t)} \le \lim_{t\to\infty} \left(\frac{\lambda_{n(t)}}{\lambda_{n(t)+1}} \right) \left(\frac{\lambda_{n(t)+1} U(a_{n(t)+1}x)}{\lambda_{n(t)} U(a_{n(t)})} \right).$$

Hence $\lim_{t\to\infty} \frac{U(tx)}{U(t)} = 1 \cdot \frac{\chi(x)}{\chi(1)} = h(x)$. By part (i), $U \in RV_\alpha$ for some $\alpha \in \mathbb{R}$.

□

Proof of Theorem 2.3.4. We follow the lines of the proof of Theorem 3.6 of [10].

$(i) \Rightarrow (ii)$ We choose a sequence $\{a_n\}$ with $a_n \to \infty$ such that $nP(X > a_n) \to 1$, e.g. $a_n = F^{-1}(1 - \frac{1}{n})$. Then

$$\begin{aligned}\lim_{n\to+\infty} nP(X > a_n x) &= \lim_{n\to+\infty} \left(\frac{P(X > a_n x)}{P(X > a_n)} \cdot nP(X > a_n) \right) \\ &= \lim_{n\to+\infty} \left(\frac{\overline{F}(a_n x)}{\overline{F}(a_n)} \cdot nP(X > a_n) \right) = x^{-\alpha} \cdot 1 = x^{-\alpha}.\end{aligned}$$

$(ii) \Rightarrow (i)$ We use Lemma 2.3.5 (ii) with $U(x) = \overline{F}(x)$ and $\lambda_n = n$. We know that $\lim_{n\to+\infty} n\bar{F}(a_n x) = x^{-\alpha} := \chi(x)$. It follows that $\overline{F}(x) \in RV_{-\alpha}$.

$(ii) \Rightarrow (iii)$ Let $f : (0, \infty] \to [0, \infty)$ be a continuous function with compact support. We want to show that

$$\mu_n(f) := nE\left[f\left(\frac{X_1}{a_n}\right)\right] = \int_0^\infty f(x) n\mathbb{P}\left[\frac{X_1}{a_n} \in dx\right] \to \nu_\alpha(f).$$

Since f has compact support, the support of f is contained in $(\delta, \infty]$ for some $\delta > 0$. By (2.5), we know that

$$\mu_n\left((x, \infty]\right) \to x^{-\alpha} = \nu_\alpha(x, \infty], \text{ for all } x > 0.$$

On the set $(\delta, \infty]$, we consider the probability measures:

$$P_n(\cdot) = \frac{\mu_n(\cdot)}{\mu_n(\delta, \infty]} \quad \text{and} \quad P(\cdot) = \frac{\nu_\alpha(\cdot)}{\delta^{-\alpha}},$$

Then for $y \in (\delta, \infty)$

$$P_n\left((y, \infty]\right) \to P\left((y, \infty]\right) = \frac{y^{-\alpha}}{\delta^{-\alpha}}.$$

Therefore, $P_n\left((\delta, y]\right) \to P\left((\delta, y]\right)$ for any $y \in (\delta, \infty)$. By Theorem A.1.7 (Appendix A), $\{P_n\}$ converges weakly to P. Since f is bounded and continuous on $(\delta, \infty]$, this implies that $P_n(f) \to P(f)$; that is,

$$\frac{\mu_n(f)}{\mu_n((\delta, \infty])} \to \frac{\nu_\alpha(f)}{\delta^{-\alpha}}.$$

Since $\mu_n((\delta, \infty]) \to \delta^{-\alpha}$, this implies that $\mu_n(f) \to \nu_\alpha(f)$. We conclude that $(\mu_n)_{n\geq 1}$ converges vaguely to ν_α.

$(iii) \Rightarrow (i)$ Since $\mu_n \xrightarrow{v} \nu_\alpha$, then by Remark 2.3.3, we have $\mu_n((x, \infty]) \to \nu_\alpha((x, \infty]) = x^{-\alpha}$, for all $x > 0$, since $(x, \infty]$ is relatively compact set and $\nu_\alpha(\partial(x, \infty]) = \nu_\alpha(\{x\}) = 0$. This shows that (2.7) holds for $\lambda_n = n$, $U = \overline{F}$ and $\chi(x) = x^{-\alpha}$ since $n\overline{F}(a_n x) = nP\left(\frac{X}{a_n} \in (x, \infty]\right)\Big) = \mu_n((x, \infty])$. By Lemma 2.3.5 (ii), it follows that $\overline{F} \in RV_{-\alpha}$.

$\square$

Chapter 3

Multivariate regular variation

In this chapter, we review the concept of multivariate regular variation for random vectors with values in $\mathbb{R}_+^d$ and we discuss its properties.

3.1 Multivariate regular variation of real valued functions

In this section, we introduce the concept of multivariate regular variation.

3.1.1 Multivariate regularly varying functions

Definition 3.1.1. *Let $h : \mathbb{R}^d \to (0, \infty)$ be a measurable function. We say that h is* ***multivariate regularly varying*** *with limit function $\lambda \neq 0$ if for any $\mathbf{x} \in \mathbb{R}^d$*

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

where $\mathbf{1} = (1, 1, \cdots, 1) \in \mathbb{R}^d$.

The next result gives a criterion for checking that a function is multivariate regularly varying.

Lemma 3.1.2. *(i) If h is multivariate regularly varying with limit function λ, then there exists a function $V \in RV_\rho$ such that*

$$\lim_{t \to +\infty} \frac{h(t\mathbf{x})}{V(t)} = \lambda(\mathbf{x}), \text{ for all } \mathbf{x} \in \mathbb{R}^d. \tag{3.1}$$

In this case, $V(t) = h(t\mathbf{1})$ and $\lambda(s\mathbf{x}) = s^\rho \lambda(\mathbf{x})$ for any $s > 0$ and $\mathbf{x} \in \mathbb{R}^d$.

(ii) If there exists a function $V \in RV_\rho$ for some $\rho \in \mathbb{R}$ such that (3.1) holds, then h is multivariate regularly varying with limit function $\lambda' = c\lambda$ where $c = \frac{1}{\lambda(\mathbf{1})}$.

Proof. We follow the argument on page 167 of [10].

(i) If h is multivariate regularly varying with limit function λ, then (3.1) holds with $V(t) = h(t\mathbf{1})$. We need to prove that $V \in RV_\rho$ for some $\rho \in \mathbb{R}$.
Fix $\mathbf{x} \in \mathbb{R}^d$. Define $U : (0, \infty)$ by $U(t) = h(t\mathbf{x})$. For any $s > 0$, we have

$$\lim_{t\to+\infty} \frac{U(ts)}{U(t)} = \lim_{t\to+\infty} \frac{h(ts\mathbf{x})}{h(t\mathbf{x})} = \lim_{t\to+\infty} \frac{h(ts\mathbf{x})}{h(t\mathbf{1})} \cdot \frac{h(t\mathbf{1})}{h(t\mathbf{x})} = \frac{\lambda(s\mathbf{x})}{\lambda(\mathbf{x})}. \tag{3.2}$$

By Lemma 2.3.5 (i), it follows that $U \in RV_{\rho(\mathbf{x})}$ for some $\rho(\mathbf{x}) \in \mathbb{R}$, and

$$\frac{\lambda(s\mathbf{x})}{\lambda(\mathbf{x})} = s^{\rho(\mathbf{x})}, \text{ for all } s > 0. \tag{3.3}$$

We have,

$$s^{\rho(\mathbf{y})} = \lim_{t\to+\infty} \frac{h(ts\mathbf{y})}{h(t\mathbf{y})} = \lim_{t\to+\infty} \frac{\frac{h(ts\mathbf{y})/h(ts\mathbf{1})}{h(ts\mathbf{x})/h(ts\mathbf{1})}}{\frac{h(t\mathbf{y})/h(t\mathbf{1})}{h(t\mathbf{x})/h(t\mathbf{1})}} \cdot \frac{h(ts\mathbf{x})}{h(t\mathbf{x})} = \frac{\frac{\lambda(\mathbf{y})}{\lambda(\mathbf{x})}}{\frac{\lambda(\mathbf{y})}{\lambda(\mathbf{x})}} \cdot s^{\rho(\mathbf{x})} = s^{\rho(\mathbf{x})}$$

and hence $\rho(\mathbf{x}) = \rho(\mathbf{y})$. Relation (3.3) implies that

$$\lambda(s\mathbf{x}) = s^\rho \lambda(\mathbf{x}) \text{ for any } s > 0,\ \mathbf{x} \in \mathbb{R}^d.$$

(ii) To prove the other implication, suppose that (3.1) holds for some function $V \in RV_\rho$ with $\rho \in \mathbb{R}$. Then for any $\mathbf{x} \in \mathbb{R}^d$,

$$\lim_{t\to+\infty} \frac{h(t\mathbf{x})}{h(t\mathbf{1})} = \lim_{t\to+\infty} \frac{h(t\mathbf{x})}{V(t)} \cdot \frac{V(t)}{h(t\mathbf{1})} = \frac{\lambda(\mathbf{x})}{\lambda(\mathbf{1})} =: \lambda'(\mathbf{x}).$$

This shows that h is multivariate regularly varying with limit function λ'.

□

3.1.2 The polar coordinate transformation

We recall that a *norm* on $\mathbb{R}^d$ is a mapping $\|\cdot\| : \mathbb{R}^d \to [0, \infty)$ which satisfies the following properties:

1. $\|\mathbf{x}\| > 0$ for all $\mathbf{x} \in \mathbb{R}^d$, and $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = 0$;

2. $\|c\mathbf{x}\| = |c|\|\mathbf{x}\|$ for all $\mathbf{x} \in \mathbb{R}^d$ and $c \in \mathbb{R}$;

3. For all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ we have $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

Examples of norms on $\mathbb{R}^d$ are:

i) The Euclidean norm:

$$\|\mathbf{x}\| = \sqrt{\sum_{i=1}^{d} x_i^2}.$$

ii) The L_p-norm:

$$\|\mathbf{x}\| = \left(\sum_{i=1}^{d} |x_i|^d\right), \text{ for all } p \geq 0.$$

iii) The maximum norm:

$$\|\mathbf{x}\| = \max\{|x_1|, \ldots, |x_d|\}.$$

Given a chosen norm $\|\cdot\|$, the unit sphere on $\mathbb{R}^d$ is defined by

$$\aleph_+ = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = 1\}.$$

Definition 3.1.3. *The* ***polar coordinate transformation*** $T : \mathbb{R}^d \setminus \{0\} \to (0, \infty) \times \aleph_+$ *is defined by:*

$$T(\mathbf{x}) = \left(\|\mathbf{x}\|, \frac{\mathbf{x}}{\|\mathbf{x}\|}\right) =: (r, a).$$

This has an inverse transformation $T^{-1} : (0, \infty) \times \aleph_+ \to \mathbb{R}^d \setminus \{0\}$ given by

$$T^{-1}(r, a) = ra.$$

3.2 Multivariate regular variation of random vectors

Recall that the distribution function of a d-dimensional random vector $\mathbf{X} = (X_1, \cdots, X_d)$ is defined by $F(\mathbf{x}) = P(\mathbf{X} \leq \mathbf{x})$ for any $\mathbf{x} \in \mathbb{R}^d$. Here we use the lexicographic order in $\mathbb{R}^d$: if $\mathbf{x} = (x_1, \cdots, x_d)$ and $\mathbf{y} = (y_1, \cdots, y_d)$ then $\mathbf{x} \leq \mathbf{y}$ if $x_i \leq y_i$ for all $i = 1, \cdots, d$.

Definition 3.2.1. *Let* $\mathbf{X}$ *be a* d*-dimensional random vector that takes values in* $[0, \infty)^d$*. Suppose that the distribution of* $\mathbf{X}$ *is* F*. We say that* X *is* ***multivariate regularly***

varying *if there exist a sequence* $\{a_n\}$ *with* $a_n \to \infty$ *and a Radon measure* ν *on* $\mathbb{E}_\infty = [0,\infty]^d \setminus \{0\}$ *such that*

$$nF(a_n \cdot) = nP\left(\frac{\mathbf{X}}{a_n} \in \cdot\right) \xrightarrow{v} \nu \text{ in } M_+(\mathbb{E}_\infty). \tag{3.4}$$

Here $M_+(\mathbb{E}_\infty)$ *is the space of Radon measures on* $\mathbb{E}_\infty$*. In this case, we say that* ν *is the limit measure of* X *and* $\{a_n\}_n$ *is the normalizing sequence.*

Remark 3.2.2. Let G be a distribution function on $\mathbb{R}_+^d$ defined by

$$G(\mathbf{x}) = e^{-\nu([\mathbf{0},\mathbf{x}]^c)}, x \in \mathbf{R}_+^d.$$

In this case, we say that ν is the *exponent measure* of G. The meaning of this construction will be explained in Proposition 4.2.14.

We include below several technical results which we will need for the proofs.

Lemma 3.2.3 (Lemma 6.1 of [10])**.** *Let* $\mu_n \in M_+(\mathbb{E}_\infty)$ *for all* $n \geq 0$*. Then*

$$\mu_n \xrightarrow{v} \mu_0 \text{ in } M_+(\mathbb{E}_\infty)$$

if and only if

$$\mu_n([\mathbf{0},\mathbf{x}]^c) \to \mu_0([\mathbf{0},\mathbf{x}]^c)$$

for all $\mathbf{x} \in [\mathbf{0},\infty)^d \setminus \{\mathbf{0}\}$*, which are continuity points of the limit* $\mu_0([\mathbf{0},\cdot]^c)$*.*

Proposition 3.2.4 (Proposition 5.5 of [10])**.** *Let* $\mathbb{E}_1$ *and* $\mathbb{E}_2$ *be LCCB spaces. Let* $\mathcal{K}(\mathbb{E}_i)$ *be the class of compact subsets of* $\mathbb{E}_i$*, for* $i = 1, 2$*. Suppose that* $T : \mathbb{E}_1 \to \mathbb{E}_2$ *is a continuous function such that*

$$T^{-1}(K_2) \in \mathcal{K}(\mathbb{E}_1) \text{ for all } K_2 \in \mathcal{K}(\mathbb{E}_2).$$

If $\mu_n \xrightarrow{v} \mu_0$ *in* $M_+(\mathbb{E}_1)$*, then*

$$\widehat{T}(\mu_n) = \mu_n \circ T^{-1} \xrightarrow{v} \mu_0 \circ T^{-1} = \widehat{T}(\mu_0)$$

in $M_+(\mathbb{E}_2)$*.*

We recall that ν_α is the measure on $(0,\infty]$ (see Theorem 2.3.4) given by

$$\nu_\alpha\left((x,\infty]\right) = x^{-\alpha} \text{ for any } x \in (0,\infty) \text{ and } \nu_\alpha(\{\infty\}) = 0.$$

The next theorem provides the equivalence between the multivariate regular variation of F, the vague convergence and the polar transformation. In this result, $\|\cdot\|$ denotes an arbitrary norm on $\mathbb{R}^d$.

Theorem 3.2.5 (Theorem 6.1 of [10]). *Suppose that* $\mathbf{X}$ *is a nonnegative d-dimensional random vector with distribution function F. The following statements are equivalent:*

(i) There exists a Radon measure ν' *on* $\mathbb{E}_\infty$ *with* $\nu'([\mathbf{0},\mathbf{1}]^c) = 1$ *such that*

$$\lim_{t\to+\infty} \frac{1-F(t\mathbf{x})}{1-F(t\mathbf{1})} = \lim_{t\to+\infty} \frac{P\left(\frac{\mathbf{X}}{t} \in [\mathbf{0},\mathbf{x}]^c\right)}{P\left(\frac{\mathbf{X}}{t} \in [\mathbf{0},\mathbf{1}]^c\right)} = \nu'([\mathbf{0},\mathbf{x}]^c)$$

for all points $\mathbf{x} \in \mathbb{E}_\infty$ *which are continuity points of the function* $\nu([0,\cdot]^c)$.

(ii) There exist a function $a(\cdot)$ *with* $\lim_{t\to+\infty} a(t) = \infty$ *and a Radon measure* ν *on* $\mathbb{E}_\infty$*, such that in* $M_+(\mathbb{E}_\infty)$*,*

$$tP\left(\frac{\mathbf{X}}{a(t)} \in \cdot\right) \xrightarrow{v} \nu, \text{ as } t \to \infty.$$

(iii) There exist a sequence $\{a_n\}$ *with* $a_n \to \infty$ *and a Radon measure* ν *on* $\mathbb{E}_\infty$ *such that in* $M_+(\mathbb{E}_\infty)$*,*

$$nP\left(\frac{\mathbf{X}}{a_n} \in \cdot\right) \xrightarrow{v} \nu, \text{ as } n \to \infty. \tag{3.5}$$

(iv) There exist a probability measure $S(\cdot)$ *on* $\aleph_+$ *(the unit sphere on* $\mathbb{R}^d$*)and a function* $a(\cdot)$ *with* $\lim_{t\to+\infty} a(t) = \infty$ *such that for* $(\mathbf{R},\mathbf{\Theta}) = (\|\mathbf{X}\|, \frac{\mathbf{X}}{\|\mathbf{X}\|})$*, we have*

$$tP\left[\left(\frac{\mathbf{R}}{a(t)}, \mathbf{\Theta}\right) \in \cdot\right] \xrightarrow{v} c\nu_\alpha \times S$$

in $M_+((0,\infty] \times \aleph_+)$ *for some* $c > 0$.

(v) There exist a probability measure $S(\cdot)$ *on* $\aleph_+$*, and a sequence* $\{a_n\}$ *with* $a_n \to \infty$ *such that for* $(\mathbf{R},\mathbf{\Theta}) = (\|\mathbf{X}\|, \frac{\mathbf{X}}{\|\mathbf{X}\|})$*, we have*

$$nP\left[\left(\frac{\mathbf{R}}{a_n}, \mathbf{\Theta}\right) \in \cdot\right] \xrightarrow{v} c\nu_\alpha \times S \tag{3.6}$$

in $M_+((0,\infty] \times \aleph_+)$ *for some* $c > 0$.

Remark 3.2.6. Note that (i) is equivalent to saying that the function $h(\mathbf{x}) = 1-F(\mathbf{x}) = P(\mathbf{X} \in [\mathbf{0},\mathbf{x}]^c)$ is multivariate regularly varying with limit function $\lambda(\mathbf{x}) = \nu'([\mathbf{0},\mathbf{x}]^c)$. The measure S given by (iv) is called the *angular measure* of $\mathbf{X}$. We will show that the measures ν and S are related via the relationship:

$$\nu \circ T^{-1} = c\nu_\alpha \times S \text{ on } (0,\infty) \times \aleph_+. \tag{3.7}$$

When either one of the conditions $(i)-(v)$ of Theorem 3.2.5 holds, $\mathbf{X}$ is multivariate regularly varying.

Proof of Theorem 3.2.5. We follow the lines of the proof of Theorem 6.1 of [10].

$(i) \Rightarrow (ii)$ We know from (i) that $\overline{F}(\mathbf{x}) = 1-F(\mathbf{x})$ is multivariate regularly varying with limit function $\lambda(x) = \nu([\mathbf{0},\mathbf{x}]^c)$. We denote $U(t) = \overline{F}(t\mathbf{1})$. Using the same argument as in the proof of Lemma 3.1.2 (see (3.2) and (3.3) with x=1) we obtain:

$$\lim_{t\to+\infty} \frac{U(ts)}{U(t)} = s^{-\alpha}.$$

This proves that U is RV with index α. We define $a(t)$ such that

$$\overline{F}(a(t)\mathbf{1}) \sim \frac{1}{t} \text{ as } t \to \infty, \text{ i.e.} \lim_{t\to+\infty} tP\left(\frac{\mathbf{X}}{a(t)} \in [\mathbf{0},\mathbf{1}]^c\right) = 1.$$

From Lemma 3.2.3 we know, $\mu_t \xrightarrow{v} \mu$ in $M_+(\mathbb{E}_\infty)$ if and only if $\mu_t([\mathbf{0},\mathbf{x}]^c) \to \mu([\mathbf{0},\mathbf{x}]^c)$. Let $\mu_t(\cdot) = tP\left(\frac{\mathbf{X}}{a(t)} \in \cdot\right)$ and $\mu(\cdot) = \nu(\cdot)$. For any $\mathbf{x}$ fixed,we have

$$\lim_{t\to+\infty} \frac{P\left(\frac{\mathbf{X}}{t} \in [0,\mathbf{x}]^c\right)}{P\left(\frac{\mathbf{X}}{t} \in [\mathbf{0},\mathbf{1}]^c\right)} = \nu([\mathbf{0},\mathbf{x}]^c) \text{ and hence } \lim_{t\to+\infty} \frac{tP\left(\frac{\mathbf{X}}{a(t)} \in [0,\mathbf{x}]^c\right)}{tP\left(\frac{\mathbf{X}}{a(t)} \in [\mathbf{0},\mathbf{1}]^c\right)} = \nu([\mathbf{0},\mathbf{x}]^c).$$

We know that $\lim_{t\to+\infty} tP\left(\frac{\mathbf{X}}{a(t)} \in [\mathbf{0},\mathbf{1}]^c\right) = 1$. Hence,

$$\lim_{t\to+\infty} tP\left(\frac{\mathbf{X}}{a(t)} \in [\mathbf{0},\mathbf{x}]^c\right) = \nu([\mathbf{0},\mathbf{x}]^c) \text{ i.e } \lim_{t\to+\infty} \mu_t([\mathbf{0},\mathbf{x}]^c) = \mu([\mathbf{0},\mathbf{x}]^c).$$

Therefore, we have $\mu_t \xrightarrow{v} \mu$. We conclude that

$$tP\left[\frac{\mathbf{X}}{a(t)} \in \cdot\right] \xrightarrow{v} \nu, \text{ as } t \to \infty.$$

$(ii) \Rightarrow (iii)$ This proof is obvious. Replace t by n.

$(iii) \Rightarrow (i)$ It can be shown that there exists a function $a(t) \in RV_{1/\alpha}$ such that $a(n) = a_n$. For any a continuity point $\mathbf{x}$ of $\nu([\mathbf{0},\cdot]^c)$, we have by Lemma 3.2.3 that

$$nP\left(\frac{\mathbf{X}}{a_n} \in [\mathbf{0},\mathbf{x}]^c\right) \to \nu([\mathbf{0},\mathbf{x}]^c).$$

Let $V(t) = \frac{1}{a^{-1}(t)}$, where $a^{-1}(t)$ is the generalized inverse of the function a defined by

$$a^{-1}(s) = \inf\{t > 0; a(t) \geq s\}.$$

The function a^{-1} is also regularly varying, and so is the function U. For any t, there exists an integer $n(t)$ such that

$$a_{n(t)} \leq t < a_{n(t)+1},$$

and so

$$\begin{aligned} a^{-1}(t)P\left(\frac{\mathbf{X}}{t} \in [\mathbf{0},\mathbf{x}]^c\right) &\leq a^{-1}\left(a(n(t)+1)\right)P\left(\frac{\mathbf{X}}{a(n(t))} \in [\mathbf{0},\mathbf{x}]^c\right) \\ &\sim n(t)P\left(\frac{\mathbf{X}}{a(n(t))} \in [\mathbf{0},\mathbf{x}]^c\right) \\ &\to \nu([\mathbf{0},\mathbf{x}]^c) \text{ as } t \to \infty. \end{aligned}$$

Hence $\limsup_{t\to+\infty} \frac{P\left(\frac{\mathbf{X}}{t}\in[0,\mathbf{x}]^c\right)}{V(t)} \leq \nu([\mathbf{0},\mathbf{x}]^c)$. Similarly,

$$\begin{aligned} a^{-1}(t)P\left(\frac{\mathbf{X}}{t} \in [\mathbf{0},\mathbf{x}]^c\right) &\geq a^{-1}\left(a(n(t))\right)P\left(\frac{\mathbf{X}}{a(n(t))} \in [\mathbf{0},\mathbf{x}]^c\right) \\ &\sim n(t)P\left(\frac{\mathbf{X}}{a(n(t))} \in [\mathbf{0},\mathbf{x}]^c\right) \\ &\to \nu([0,\mathbf{x}]^c) \text{ as } t \to \infty, \end{aligned}$$

and hence $\liminf_{t\to+\infty} \frac{P\left(\frac{\mathbf{X}}{t}\in[0,\mathbf{x}]^c\right)}{V(t)} \geq \nu([\mathbf{0},\mathbf{x}]^c)$. Therefore,

$$\lim_{t\to+\infty} \frac{P\left(\frac{\mathbf{X}}{t} \in [\mathbf{0},\mathbf{x}]^c\right)}{V(t)} = \nu([\mathbf{0},\mathbf{x}]^c) =: \lambda(\mathbf{x}).$$

Since $V(t)$ is regularly varying, by Lemma 3.1.2 we can say that $\overline{F}(\mathbf{x}) = P\left(\frac{\mathbf{X}}{t} \in [\mathbf{0},\mathbf{x}]^c\right)$ is multivariate regularly varying with limit function $\lambda' = c_0\lambda$, where $c_0 = \frac{1}{\lambda(\mathbf{1})} = \frac{1}{\nu([\mathbf{0},\mathbf{1}]^c)}$. Therefore,

$$\lim_{t\to+\infty} \frac{\overline{F}(t\mathbf{x})}{\overline{F}(t\mathbf{1})} = \lim_{t\to+\infty} \frac{P\left[\frac{\mathbf{X}}{t} \in [\mathbf{0},\mathbf{x}]^c\right]}{P\left[\frac{\mathbf{X}}{t} \in [\mathbf{0},\mathbf{1}]^c\right]} = \nu'([\mathbf{0},\mathbf{x}]^c),$$

where $\nu'([\mathbf{0},\mathbf{x}]^c) = \lambda'(\mathbf{x}) = c_0\lambda(\mathbf{x}) = c_0\nu([\mathbf{0},\mathbf{x}]^c)$.

(iii) $\Rightarrow$ *(v)* We assume that (3.5) holds in $M_+(\mathbb{E})$. We will prove that (3.6) holds in $M_+((0,\infty)\times\aleph_+)$. For the extension to $M_+(\mathbb{E}_\infty)$ and $M_+((0,\infty]\times\aleph_+)$ we refer to pages 176-178 of [10]. Let $\mu_n = nP\circ\left(\frac{\mathbf{X}}{a_n}\right)^{-1}$ and $\mu(\cdot)=\nu(\cdot)$. Then

$$\mu_n(A) = n\left(P\circ\left(\frac{\mathbf{X}}{a_n}\right)^{-1}\right)(A) = nP\left(\frac{\mathbf{X}}{a_n}\in A\right).$$

We apply Proposition 3.2.4 with $\mathbb{E}_1=(0,\infty)^d\setminus\{0\}$, $\mathbb{E}_2=(0,\infty)\times\aleph_+$ and $T:\mathbb{E}_1\to\mathbb{E}_2$ the polar coordinate transformation. We conclude that

$$\mu_n\circ T^{-1}\xrightarrow{v}\mu\circ T^{-1} \text{ in } \mathbb{E}_2.$$

Note that, $(\mu_n\circ T^{-1})(A) = nP\left(T\left(\frac{\mathbf{X}}{a_n}\right)\in A\right)$. It remains to prove that (3.7) holds. We have shown that *(iii)* implies *(i)*. This means that

$$\lim_{t\to+\infty}\frac{P\left(\frac{\mathbf{X}}{t}\in[\mathbf{0},\mathbf{x}]^c\right)}{P(\frac{\mathbf{X}}{t}\in[\mathbf{0},\mathbf{1}]^c)} = \nu'\left([\mathbf{0},\mathbf{x}]^c\right).$$

Hence $\overline{F}(x) = P(\mathbf{X}\in[\mathbf{0},\mathbf{x}]^c)$ is multivariate regularly varying with limit function $\lambda(x)=\nu'\left([\mathbf{0},\mathbf{x}]^c\right)$. We have seen that $\nu\left([\mathbf{0},\mathbf{x}]^c\right) = c_1\nu'\left([\mathbf{0},\mathbf{x}]^c\right)$, where $c_1=\nu\left([\mathbf{0},\mathbf{1}]^c\right)$. By Lemma 3.1.2 we know

$$\lambda(s\mathbf{x}) = s^{-\alpha}\lambda(\mathbf{x}), \text{ i.e.}$$

$$\nu\left([0,s\mathbf{x}]^c\right) = s^{-\alpha}\nu\left([0,\mathbf{x}]^c\right) \text{ for any } \mathbf{x}\in(0,\infty)^d \text{ and } s>0.$$

More generally, it can be proved that, for any set A of the form $[0,\mathbf{x}]^c$ we have

$$\nu(sA) = s^{-\alpha}\nu(A) \text{ for any } s>0. \tag{3.8}$$

For any set $\Lambda\subset\aleph_+$ and $t>0$, we consider the following "pizza slice shaped":

$$A_{\Lambda,t} = \{\mathbf{x}\in(0,\infty)^d; \|\mathbf{x}\|>t, \frac{\mathbf{x}}{\|\mathbf{x}\|}\in\Lambda\}.$$

We now verify the equality of measures given by (3.7) for sets of the form $(t,\infty)\times\Lambda$:

$$\begin{aligned}(\nu\circ T^{-1})\left((t,\infty)\times\Lambda\right) &= \nu\{\mathbf{x}; T(\mathbf{x})\in(t,\infty)\times\Lambda\}\\ &= \nu\{\mathbf{x}; \left(\|\mathbf{x}\|, \frac{\mathbf{x}}{\|\mathbf{x}\|}\right)\in(t,\infty)\times\Lambda\}\\ &= \nu\{\mathbf{x}; \|\mathbf{x}\|>t, \frac{\mathbf{x}}{\|\mathbf{x}\|}\in\Lambda\}\\ &= \nu\{t\cdot\frac{\mathbf{x}}{t}; \|\frac{\mathbf{x}}{t}\|>1, \frac{\mathbf{x}/t}{\|\mathbf{x}\|/t}\in\Lambda\}\\ &= \nu\{t\mathbf{y}; \|\mathbf{y}\|>1, \frac{\mathbf{y}}{\|\mathbf{y}\|}\in\Lambda\}\\ &= t^{-\alpha}\nu\{\mathbf{y}; \|\mathbf{y}\|>1, \frac{\mathbf{y}}{\|\mathbf{y}\|}\in\Lambda\},\ \text{using (3.8)}\\ &=: (c\nu_\alpha\times S)\left((t,\infty)\times\Lambda\right),\end{aligned}$$

where we define

$$S(\Lambda)=\frac{1}{c}\nu(\{\mathbf{y}; \|\mathbf{y}\|>1, \frac{\mathbf{y}}{\|\mathbf{y}\|}\in\Lambda\})\ \text{ and }\ c=\nu(\{\mathbf{y}; \|\mathbf{y}\|>1\}). \tag{3.9}$$

The proof is complete since using the previous relation, it can be proved that

$$(\nu\circ T^{-1})(A)=(c\nu_\alpha\times S)(A) \text{ for any set } A\subset(0,\infty)\times\aleph_+.$$

$(v)\Rightarrow(iii)$ The proof is very similar to $(iii)\Rightarrow(v)$.

$(iv)\Rightarrow(v)$ The proof is similar to $(ii)\Rightarrow(iii)$. We omit the details.

□

In the proof of Theorem 3.2.5 we represented the angular measure S in terms of the limiting measure ν. Conversely, the limiting measure ν in the definition of regular variation has the following representation in terms of the angular measure S.

Proposition 3.2.7 (Proposition 6.4 of [10])**.** *Let T be the polar coordinate transformation and ν be a Radon measure on $\mathbb{E}_\infty$ such that (3.7) holds with $v_\alpha=((x,\infty])=x^{-\alpha}$, for all $x>0$, and S a probability measure on $\aleph_+$. Then for any $\mathbf{x}\in[0,\infty)^d\setminus\{\mathbf{0}\}$, we have:*

$$\nu([\mathbf{0},\mathbf{x}]^c)=c\int_{\aleph_+}\bigvee_{i=1}^{d}\left(\frac{x_i}{a_i}\right)^{-\alpha}S(d\mathbf{a}), \tag{3.10}$$

where $\bigvee_{i=1}^{d} y_i=\max_{1\le i\le d} y_i$.

Proof. We follow the lines of the proof of Proposition 6.4 of [10]. We have

$$\begin{aligned} T([\mathbf{0},\mathbf{x}]^c) &= T(\{\mathbf{y}\in[\mathbf{0},\infty)^d\setminus\{\mathbf{0}\}; y_i > x_i \text{ for some } i=1,\cdots,d\}) \\ &= \{(r,a)\in(0,\infty)\times\aleph_+; ra_i > x_i \text{ for some } i=1,\cdots,d\} \\ &= \{(r,a)\in(0,\infty)\times\aleph_+; r > \frac{x_i}{a_i} \text{ for some } i=1,\cdots,d\} \\ &= \{(r,a)\in(0,\infty)\times\aleph_+; r > \bigwedge_{i=1}^d \frac{x_i}{a_i}\}, \end{aligned} \tag{3.11}$$

where $\bigwedge_{i=1}^d y_i = \min_{1\le i\le d} y_i$. Since T is bijective, using (3.11), we have

$$\begin{aligned} \nu([\mathbf{0},\mathbf{x}]^c) &= (\nu\circ T^{-1})(T([\mathbf{0},\mathbf{x}]^c)) \\ &= (\nu\circ T^{-1})\{(r,a)\in(0,\infty)\times\aleph_+; r > \bigwedge_{i=1}^d \frac{x_i}{a_i}\} \\ &= (c\nu_\alpha\times S)\{(r,a)\in(0,\infty)\times\aleph_+; r > \bigwedge_{i=1}^d \frac{x_i}{a_i}\} \\ &= c\int_{\aleph_+}\left[\int_{[r>\bigwedge_{i=1}^d \frac{x_i}{a_i}]}\nu_\alpha(dr)\right]S(d\mathbf{a}) \\ &= c\int_{\aleph_+}\left(\bigwedge_{i=1}^d \frac{x_i}{a_i}\right)^{-\alpha}S(d\mathbf{a}) \\ &= c\int_{\aleph_+}\bigvee_{i=1}^d\left(\frac{x_i}{a_i}\right)^{-\alpha}S(d\mathbf{a}), \end{aligned}$$

where we used (3.7) for the third equality above. $\square$

In the next lemma we connect the conditional and the tail probabilities to the limiting measure ν.

Lemma 3.2.8. *Let $\mathbf{X} = (X_1, X_2)$ be a bivariate regularly varying random variable on $\mathbb{R}_+^2$ of index α and $\{a_n\}_n$ a sequence with $a_n\to\infty$ such that $nP(X_1 > a_n)\to 1$. Then*

a)

$$\lim_{n\to+\infty} P(X_2 > a_n x_2 | X_1 > a_n x_1) = x_1^\alpha \nu((x_1,\infty)\times(x_2,\infty)). \tag{3.12}$$

b)

$$\lim_{n\to+\infty} \frac{P(\{X_1 > a_n x_1\}\cup\{X_2 > a_n x_2\})}{P(X_1 > a_n x_1)} = x_1^\alpha\nu([\mathbf{0},\mathbf{x}]^c). \tag{3.13}$$

Proof. a) We have:

$$\begin{aligned}
\lim_{n\to+\infty} P(X_2 > a_n x_2 | X_1 > a_n x_1) &= \lim_{n\to+\infty} \frac{P(\{X_1 > a_n x_1\} \cap \{X_2 > a_n x_2\})}{P(X_1 > a_n x_1)} \\
&= \lim_{n\to+\infty} \frac{nP\left(\left(\frac{X_1}{a_n}, \frac{X_2}{a_n}\right) \in (x_1, \infty) \times (x_2, \infty)\right)}{nP\left(\frac{X_1}{a_n} > x_1\right)} \\
&= x_1^{\alpha} \nu((x_1, \infty) \times (x_2, \infty)).
\end{aligned}$$

b) Similarly to part a), w have:

$$\begin{aligned}
\lim_{n\to+\infty} \frac{P(\{X_1 > a_n x_1\} \cup \{X_2 > a_n x_2\})}{P(X_1 > a_n x_1)} &= \lim_{n\to+\infty} \frac{nP\left(\{\frac{X_1}{a_n} > x_1\} \cup \{\frac{X_2}{a_n} > x_2\}\right)}{nP\left(\frac{X_1}{a_n} > x_1\right)} \\
&= \lim_{n\to+\infty} \frac{nP\left(\left(\frac{X_1}{a_n}, \frac{X_2}{a_n}\right) \in [\mathbf{0}, \mathbf{x}]^c\right)}{nP\left(\frac{X_1}{a_n} > x_1\right)} \\
&= x_1^{\alpha} \nu([\mathbf{0}, \mathbf{x}]^c).
\end{aligned}$$

□

3.3 Multivariate regular variation: Examples

In this section, we examine several examples of multivariate regularly varying random vectors.

3.3.1 Independence

Proposition 3.3.1. *Let* $\mathbf{X} = (X_1, \cdots, X_d)$ *be a non-negative random vector where* $(X_j)_{j=1,\cdots,d}$ *are independent and identically distributed random variables. We choose* $(a_n)_n$ *such that* $\lim_{n\to+\infty} nP(X_1 > a_n) = 1$, *e.g.*

$$a_n = \inf\{x \geq 0; F_1(x) \geq 1 - \frac{1}{n}\},$$

where $F_1(x) = P(X_1 \leq x)$. *Then*

(i) $\mathbf{X}$ *is multivariate regularly varying and (3.5) holds with the measure* ν *given by*

$$\nu(dx_1, \cdots, dx_d) = \sum_{i=1}^{d} \delta_0(dx_1) \times \cdots \times \nu_\alpha(dx_i) \times \cdots \times \delta_0(dx_d),$$

where δ_0 *denotes the Dirac measure at* 0*:* $\delta_0(A) = \mathbb{1}_A(0)$.

(ii) If we let $R = \|\mathbf{X}\|$ *and* $\Theta = \frac{\mathbf{X}}{\|\mathbf{X}\|}$ *then (3.6) holds with* $c = d$ *and* $S = \sum_{i=1}^{d} \frac{1}{d}\delta_{e_i}$*, with* $e_i = (0, \cdots, 1, \cdots, 0)$*, where* δ_{e_i} *denotes the Dirac measure at* e_i*:* $\delta_{e_i}(A) = \mathbb{1}_A(e_i)$.

Proof. (i) By Theorem 3.2.5 and Lemma 3.2.3 it suffices to show that for any $\mathbf{x} \in \mathbb{E}_\infty$

$$nP\left(\frac{\mathbf{X}}{a_n} \in [\mathbf{0}, \mathbf{x}]^c\right) \to \nu([\mathbf{0}, \mathbf{x}]^c).$$

Note that for the measure given by (i), $\nu([\mathbf{0}, \mathbf{x}]) = \sum_{i=1}^{d} {x_i}^{-\alpha}$.
Using the inclusion-exclusion principle we have

$$\begin{aligned} nP\left(\frac{\mathbf{X}}{a_n} \in [\mathbf{0}, \mathbf{x}]^c\right) &= nP\{\bigcup_{i=1}^{d}(X_i > a_n x_i)\} \\ &= \sum_{i=1}^{d} nP(X_i > a_n x_i) - \sum_{1 \le i < j \le d} nP(X_i > a_n x_i, X_j > a_n x_j) \\ &+ \sum_{1 \le i < j < k \le d} nP(X_i > a_n x_i, X_j > a_n x_j, X_k > a_n x_k) + \cdots \\ &+ (-1)^{d+1} nP(X_1 > a_n x_1, \cdots, X_d > a_n x_d). \end{aligned}$$

Except the first term, all the terms above converge to 0 as $n \to \infty$. To see this, we examine the second term (the other terms are similar). Using the independence between X_i and X_j we have

$$\begin{aligned} \lim_{n \to +\infty} nP(X_i > a_n x_i, X_j > a_n x_j) &= \lim_{n \to +\infty} nP(X_i > a_n x_i) P(X_j > a_n x_j) \\ &= \lim_{n \to +\infty} nP(X_i > a_n x_i) nP(X_j > a_n x_j) \cdot \frac{1}{n} \\ &= {x_i}^{-\alpha} \cdot {x_j}^{-\alpha} \cdot 0 \\ &= 0. \end{aligned}$$

For the first term we have

$$\begin{aligned} \lim_{n \to +\infty} n \sum_{i=1}^{d} P(X_i > a_n x_i) &= \sum_{i=1}^{d} \lim_{n \to +\infty} nP(X_i > a_n x_i) \\ &= \sum_{i=1}^{d} {x_i}^{-\alpha}. \end{aligned}$$

Here we used the fact that $X_1 \in RV_{-\alpha}$ and $nP(X_1 > a_n) \to 1$ implies that $nP(X_1 > a_n x) \to x^{-\alpha}$.

(ii) In the proof of Theorem 3.2.5 (the part $(iii) \Rightarrow (v)$), we showed that for any $\Lambda \subset \aleph_+$,

$$S(\Lambda) = \frac{1}{c}\nu(\{\mathbf{y} \in \mathbb{E}_\infty; \|\mathbf{y}\| > 1, \frac{\mathbf{y}}{\|\mathbf{y}\|} \in \Lambda\}),$$

where $c = \nu(\{\mathbf{y} \in \mathbb{E}_\infty; \|\mathbf{y}\| > 1\})$. In our case, we have shown in part (i) that the measure ν concentrates on the axes $A_1, A_2, \cdots, A_d$ of $\mathbb{E}_\infty$, where

$$A_i = \{te_i; t > 0\} \text{ is the } i\text{-th axis.}$$

Hence,

$$\begin{aligned}
S(\Lambda) &= \frac{1}{c}\sum_{i=1}^{d}\nu(\{\mathbf{y} \in \mathbb{E}_\infty; \|\mathbf{y}\| > 1, \frac{\mathbf{y}}{\|\mathbf{y}\|} \in \Lambda\} \cap \{te_i; t > 0\}) \\
&= \frac{1}{c}\sum_{i=1}^{d}\nu(\{te_i; t > 1, \frac{e_i}{\|e_i\|} \in \Lambda\}) \\
&= \frac{1}{c}\sum_{i=1}^{d}\nu(\{te_i; t > 1\})\mathbb{1}_\Lambda(e_i).
\end{aligned}$$

Note that

$$\begin{aligned}
\nu(\{te_i; t > 1\}) &= \nu(\{(0, \cdots, t, \cdots, 0); t > 1)\}) \\
&= \nu(\{0\} \times \cdots \times (1, \infty) \times \cdots \times \{0\}) \\
&= \nu_\alpha((1, \infty)) = 1^{-\alpha} = 1.
\end{aligned}$$

Hence, $S(\Lambda) = \frac{1}{c}\sum\limits_{i=1}^{d}\delta_{e_i}(\Lambda)$. Finally, note that

$$\begin{aligned}
c &= \nu(\{\mathbf{y} \in \mathbb{E}_\infty; \|\mathbf{y}\| > 1\}) \\
&= \nu((1, \infty) \times \{0\} \times \cdots \times \{0\}) + \cdots + \nu(\{0\} \times \cdots \times (1, \infty)) \\
&= \nu_\alpha((1, \infty)) + \cdots + \nu_\alpha((1, \infty)) \\
&= 1^{-\alpha} + \cdots + 1^{-\alpha} = d.
\end{aligned}$$

Therefore, we proved that $c = d$ and $S = \sum\limits_{i=1}^{d}\frac{1}{d}\delta_{e_i}(\Lambda)$.

$\square$

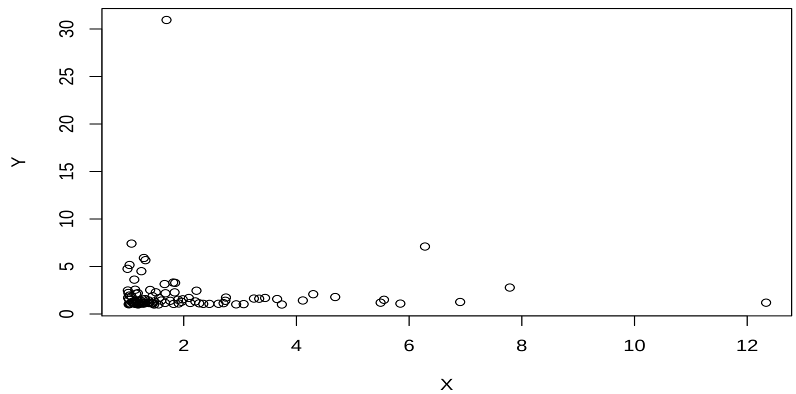

Figure 3.1: X and Y are independent Pareto with $\alpha = 2$.

3.3.2 Full dependence

Proposition 3.3.2. *Let* $\mathbf{X} = (X_1, \cdots, X_1)$ *be a non-negative random vector with* $X_1 \in RV_{-\alpha}$. *We choose* $(a_n)_n$ *such that* $\lim_{n\to+\infty} nP(X_1 > a_n) = 1$. *Then,*

(i) $\mathbf{X}$ *is multivariate regularly varying and (3.5) holds for a measure* ν *which concentrates on* $\Delta = \{t\mathbf{1}; t > 0\} = \{(t, \cdots, t); t > 0\}$ *and is given by* $\nu(\{t\mathbf{1}; t > x\}) = x^{-\alpha} = \nu_\alpha((x, \infty])$, *for any* $x > 0$.

(ii) If we let $R = \|\mathbf{X}\|$ *and* $\Theta = \frac{\mathbf{X}}{\|\mathbf{X}\|}$ *then (3.6) holds with* $c = (\frac{1}{\sqrt{d}})^{-\alpha}$ *and* $S = \delta_{\mathbf{1}/\|\mathbf{1}\|}$.

Proof. (i) Let f $\in {C_K}^+(\mathbb{E}_\infty)$ and $\mu_n(\cdot) = nP \circ \left(\frac{\mathbf{X}}{a_n}\right)^{-1}$. It suffices to show that $\mu_n(f) \to \nu(f)$. To show this we note that

$$\mu_n(f) = nE\left[f\left(\frac{\mathbf{X}}{a_n}\right)\right] = n\int_0^\infty f(x, \cdots, x)\left[P \circ \left(\frac{\mathbf{X}}{a_n}\right)^{-1}\right](dx)$$
$$\longrightarrow \int_0^\infty f(x, \cdots, x)\nu_\alpha(dx) = \int_{\mathbb{E}_\infty} f(y)\nu(dy) = \nu(f).$$

(ii) By Theorem 3.2.5 we know that (3.6) holds with (S, c) given by (3.9). By part (i), we know that and ν concentrates on $\Delta = \{t\mathbf{1}; t > 0\}$. This means that

$$\begin{aligned}
S(\Lambda) &= \frac{1}{c}\nu(\{\mathbf{y} \in \mathbb{E}_\infty; \|\mathbf{y}\| > 1, \frac{\mathbf{y}}{\|\mathbf{y}\|} \in \Lambda\} \cap \{t\mathbf{1}; t > 0\}) \\
&= \frac{1}{c}\nu(\{t\mathbf{1}; t > \frac{1}{\|\mathbf{1}\|}, \frac{\mathbf{1}}{\|\mathbf{1}\|} \in \Lambda\}) \\
&= \frac{1}{c}\nu(\{t\mathbf{1}; t > \frac{1}{\|\mathbf{1}\|}\})\mathbb{1}_\Lambda\left(\frac{\mathbf{1}}{\|\mathbf{1}\|}\right) \\
&= \mathbb{1}_\Lambda\left(\frac{\mathbf{1}}{\|\mathbf{1}\|}\right)
\end{aligned}$$

where for the last equality above we used the fact that

$$c = \nu(\{t\mathbf{1}; t > \frac{1}{\|\mathbf{1}\|}\}) = \nu_\alpha((\frac{1}{\sqrt{d}}, \infty)) = \left(\frac{1}{\sqrt{d}}\right)^{-\alpha}.$$

Therefore,we proved that $c = \left(\frac{1}{\sqrt{d}}\right)^{-\alpha}$ and $S = S_{\mathbf{1}/\|\mathbf{1}\|}(\Lambda)$.

$\square$

3.3.3 Other models

To introduce other models, we proceed with the following Proposition.

Proposition 3.3.3. *Let R be a non-negative random variable which is regularly varying with index $\alpha > 0$ and $(a_n)_{n\geq 1}$ a sequence of positive numbers such that*

$$nP(R > a_n x) \to x^{-\alpha}, \ \forall x > 0.$$

Let Θ be an arbitrary random variable on $\aleph_+$, with distribution S such that Θ is independent of R. Then the random vector $\mathbf{X} = R\Theta$ is multivariate regularly varying with limit measure ν given by (3.10) with $c = 1$.

Proof. We apply Theorem 3.2.5. Due to the independence between R and Θ, we have

$$nP\left(\left(\frac{R}{a_n}, \Theta\right) \in \cdot\right) = nP\left(\frac{R}{a_n} \in \cdot\right) P(\Theta \in \cdot) \xrightarrow{v} \nu_\alpha \times S(\cdot) \text{ in } M_+(\mathbb{E}_\infty).$$

$\square$

We consider now an example of a multivariate regularly varying random vector on $\mathbb{R}^2$. (The definition of the multivariate regular variation on $\mathbb{R}^2$ is similar to $\mathbb{R}^2_+$, except that we no longer have property (ii) in Theorem 3.2.5.)

Example 3.3.4 (Example 6.1 of [10])**.** Let $\mathbf{X} = (X_1, X_2)$ be a random vector on $\mathbb{R}^2$ with a bivariate Cauchy density

$$f(x,y) = \frac{1}{2\pi}(1 + x^2 + y^2)^{-3/2},\ (x,y) \in \mathbb{R}^2.$$

We denote by $R = \|\mathbf{X}\| = \sqrt{{X_1}^2 + {X_2}^2}$ the Euclidean norm of $\mathbf{X}$ and we let $\Theta = \frac{\mathbf{X}}{\|\mathbf{X}\|}$. Transforming to the polar coordinates,

$$r^2 = x^2 + y^2, \quad \theta = \arctan y/x,$$

we see that the density function of (R, Θ) is

$$\begin{aligned} f_{R,\Theta}(r,\theta) &= f(r\cos\theta, r\sin\theta) \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} \\ &= f(r\cos\theta, r\sin\theta) r \\ &= r(1+r^2)^{-3/2}\frac{1}{2\pi}. \end{aligned}$$

Therefore, Θ has a uniform distribution on $[0, 2\pi]$ and the density of R is $f_R(r) = r(1+r^2)^{-3/2}$. Hence

$$P(R > r) = \int_r^\infty v(1+v^2)^{-3/2} dv = (1+r^2)^{-1/2} \sim r^{-1},\ \text{as } r \to \infty,$$

It follows that R is regularly varying with index $\alpha = -1$. In addition, R and Θ are independent.

Next, we verify that relation (3.6) holds with $a_n = n$. For any measurable sets $A \subset (0,\infty)$ and $B \subset [0, 2\pi]$, we have

$$\begin{aligned} nP\left[\left(\frac{R}{n}, \Theta\right) \in A \times B\right] &= nP\left(\tfrac{R}{n} \in A\right) P(\Theta \in B) \\ &= nP\left(\tfrac{R}{n} \in A\right) \cdot \tfrac{1}{2\pi}\ell(B) \end{aligned} \tag{3.14}$$

with ℓ being the Lebesgue measure. Note that

$$nP\left(\frac{R}{n} > x\right) = n\int_{nx}^\infty r(1+r^2)^{-3/2} dr = \frac{n}{(1+n^2x^2)^{1/2}} \to x^{-1} = \nu_1((1,\infty)),\ \text{as } n \to \infty. \tag{3.15}$$

In general, it can be proved that

$$nP\left(\frac{R}{n} \in A\right) \to \nu_1(A)$$

for any measurable set $A \in (0, \infty)$. Using (3.14) and (3.15), we infer that

$$nP\left[\left(\frac{R}{n}, \Theta\right) \in \cdot\right] \xrightarrow{v} \nu_1 \times S, \tag{3.16}$$

where S is the uniform distribution on $[0, 2\pi]$ and $\nu_1\left((x, \infty]\right) = x^{-1}$. Hence, according to the analogue of Theorem 3.2.5 for $\mathbb{R}^2$, $\mathbf{X}$ is multivariate regularly varying with limiting measure ν.

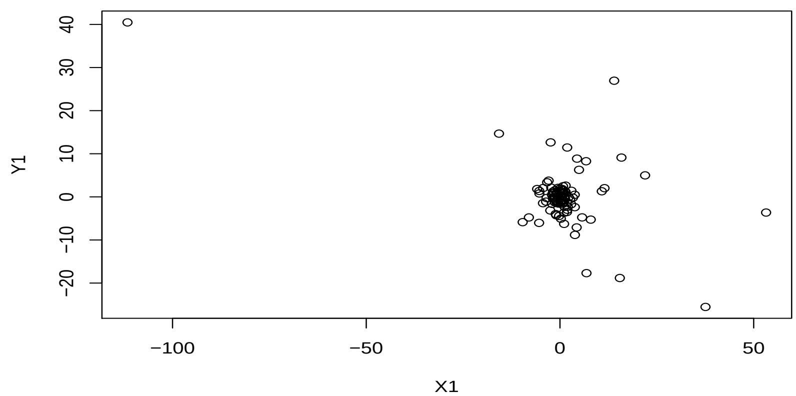

Figure 3.2: Bivariate Cauchy simulation.

Chapter 4

Extreme value theory

In this chapter, we introduce the basic elements of extreme value theory in the multivariate case, and we discuss the connections with multivariate regular variation.

4.1 Univariate extreme value theory

In this section, we review some known results from the univariate extreme value theory.

Definition 4.1.1. *Let G be a distribution on $\mathbb{R}$. We say that G is a* ***univariate extreme value*** *(EV) distribution if there exist a sequence $(X_i)_{i\geq 1}$ of independent and identically distributed random variables in $\mathbb{R}$ (with distribution F) and sequences $(a_n)_n \subset \mathbb{R}_+$ and $(b_n)_n \subset \mathbb{R}$ such that*

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} Y, \tag{4.1}$$

where Y is a random variable with distribution G and $M_n = \max_{i\leq n} X_i$. In this case, we say that F is in the ***max-domain of attraction*** *of G and we write $F \in MaxD(G)$.*

The distribution of the maximum M_n is given by:

$$P(M_n \leq x) = P(X_1 \leq x, \dots, X_n \leq x) = P(X_1 \leq x) \dots P(X_n \leq x) = F^n(x). \tag{4.2}$$

The following technical lemma will be needed below.

Lemma 4.1.2 (Proposition 2.2 of [2])**.** *Assume that $U : \mathbb{R}_+ \to \mathbb{R}$ is a monotone function which satisfies the following condition:*

$$\lim_{x\to+\infty} \frac{U(xu) - U(x)}{a(x)} = h(u), \quad \textit{for any} \ \ u > 0, \tag{4.3}$$

for some function $a : \mathbb{R}_+ \to \mathbb{R}_+$. Then there exists some $\gamma \in \mathbb{R}$ such that

$$h(u) = h_\gamma(u) := \int_1^u v^{\gamma-1} dv, \quad \text{for any} \quad u > 0. \tag{4.4}$$

That is $h_\gamma(u) = \frac{u^\gamma - 1}{\gamma}$ if $\gamma \neq 0$ and $h_0(u) = \ln(u)$.

In the next theorem, we will need the distribution G_γ, defined as follows:

- For $\gamma > 0$,

$$G_\gamma(x) = \begin{cases} \exp\{-(1+\gamma x)^{-1/\gamma}\}, & \text{if } x > -\frac{1}{\gamma} \\ 0, & \text{otherwise} \end{cases}. \tag{4.5}$$

- For $\gamma < 0$,

$$G_\gamma(x) = \begin{cases} \exp\{-(1+\gamma x)^{-1/\gamma}\}, & \text{if } x < -\frac{1}{\gamma} \\ 1, & \text{otherwise} \end{cases}. \tag{4.6}$$

- For $\gamma = 0$,

$$G_0(x) = \exp(-\exp(-x)), \quad \text{for any} \quad x \in \mathbb{R}. \tag{4.7}$$

Remark 4.1.3. Instead of (4.5),(4.6) or (4.7) we will use the following notation convention:

$$G_\gamma(x) = \exp\{-(1+\gamma x)^{-1/\gamma}\} \ \text{ if } 1 + \gamma x > 0.$$

Definition 4.1.4. *A distribution G on $\mathbb{R}$ is called **max-stable** if for any positive integer k, there exist $\alpha_k \in \mathbb{R}_+$ and $\beta_k \in \mathbb{R}$ such that*

$$G^k(x) = G(\alpha_k x + \beta_k), \ \textit{for all } x \in \mathbb{R}. \tag{4.8}$$

Lemma 4.1.5. *For any $\gamma \in \mathbb{R}, G_\gamma$ is a max-stable distribution.*

Proof. We need to show that G_γ satisfies relation (4.8) for some $\alpha_k \in \mathbb{R}_+$ and $\beta_k \in \mathbb{R}$, for any positive integer k. We consider only the case $\gamma > 0$. (The case $\gamma < 0$ is similar.) We need to find α_k and β_k such that

$$G_\gamma^k(x) = G_\gamma(\alpha_k x + \beta_k), \ \text{for all } x \in \mathbb{R}.$$

Consider the values $x \in \mathbb{R}$ for which $1 + \gamma x > 0$ and $1 + \gamma(\alpha_k x + \beta_k) > 0$. Using the formula of G_γ the previous equation becomes

$$\exp\{-k(1+\gamma x)^{-1/\gamma}\} = \exp\{-(1+\gamma(\alpha_k x + \beta_k))^{-1/\gamma}\},$$

or equivalently

$$k^{-\gamma}(1+\gamma x) = 1 + \gamma(\alpha_k x + \beta_k), \quad \text{i.e.}$$

$$\gamma x + 1 = \gamma \alpha_k k^{\gamma} x + k^{\gamma}(1 + \gamma \beta_k).$$

Hence, by equalizing the coefficients we get :

$$\alpha_k = k^{-\gamma} \;\text{ and }\; \beta_k = \frac{k^{-\gamma} - 1}{\gamma}.$$

We now consider the case $\gamma = 0$. In this case $G_0(x) = \exp(-\exp(-x))$. As in the previous case, we have to find α_k and β_k such that

$$G_0^k(x) = G_0(\alpha_k x + \beta_k), \text{ for all } x \in \mathbb{R}.$$

Using the formula of G_0 this equation becomes

$$\exp(-k \exp(-x)) = \exp(-\exp(-\alpha_k x - \beta_k)),$$

or equivalently

$$k \exp(-x) = \exp(-\alpha_k x - \beta_k), \quad \text{i.e.}$$

$$x - \log(k) = \alpha_k x + \beta_k.$$

Hence, by equalizing the coefficients we get :

$$\alpha_k = 1 \;\text{ and }\; \beta_k = -\log(k).$$

Therefore, there exist $\alpha_k \in \mathbb{R}_+$ and $\beta_k \in \mathbb{R}$ such that for any positive integer k relation (4.8) holds. $\square$

The following result will be proved in a more general form in Section 4.2 in the multivariate case (see Theorem 4.2.10 and Definition 4.2.7 of the concept of max-stable distributions). We omit the proof in the univariate case.

Theorem 4.1.6. *The class of univariate extreme value distributions coincides with the class of max-stable distributions.*

A crucial step in the proof of Theorem 4.1.6 is based on the following two lemmas:

Lemma 4.1.7 (Convergence of Types Lemma, Proposition 0.2 page 4 of [9])**.** *Let U and V be two non-degenerate distributions on $\mathbb{R}$.*

a) Suppose that for each $n \geq 1, F_n$ is a distribution on $\mathbb{R}, a_n \in \mathbb{R}_+, b_n \in \mathbb{R}, \alpha_n \in \mathbb{R}_+, \beta_n \in \mathbb{R}$ such that

$$F_n(a_n x + b_n) \to U(x) \tag{4.9}$$

for any continuity point x of U and

$$F_n(\alpha_n x + \beta_n) \to V(x) \tag{4.10}$$

for any continuity point x of V. Then $\alpha_n/a_n \to A > 0, (\beta_n - b_n)/a_n \to B \in \mathbb{R}$ and

$$V(x) = U(Ax + B) \text{ for all } x \in \mathbb{R}. \tag{4.11}$$

An equivalent formulation can be given in terms of random variables, as follows: if $(X_n)_{n\geq 1}$ is a sequence of random variables such that

$$\frac{X_n - b_n}{a_n} \xrightarrow{d} X,$$

where X has distribution U and

$$\frac{X_n - \beta_n}{\alpha_n} \xrightarrow{d} Y,$$

where Y has distribution V, then $\alpha_n/a_n \to A > 0, (\beta_n - b_n)/a_n \to B \in \mathbb{R}$ and

$$Y \stackrel{d}{=} \frac{X - B}{A}.$$

b) If $\alpha_n/a_n \to A > 0, (\beta_n - b_n)/a_n \to B \in \mathbb{R}$ and (4.9) holds for any continuity point x of U, then (4.10) holds for any continuity point x of V, with V given by (4.11).

Lemma 4.1.8. *Let F be a distribution on $\mathbb{R}, (a_n)_n \subset \mathbb{R}_+$ and $(b_n)_n \subset \mathbb{R}$ such that*

$$F^n(a_n x + b_n) \to G(x)$$

for any continuity point x of G, where G is a non-degenerate distribution. Then for any integer $k > 0$, there exist $\alpha_k \in \mathbb{R}_+$ and $\beta_k \in \mathbb{R}$ such that

$$\lim_{n\to+\infty} \frac{a_n}{a_{nk}} = \alpha_k, \quad \lim_{n\to+\infty} \frac{b_n - \beta_n}{a_{nk}} = \beta_k \quad \text{and} \quad G^k(x) = G(\alpha_k x + \beta_k).$$

Proof. We use the argument on page 9 of [9]. We apply the Convergence of Types Lemma with $F_n = F^{nk}$. For any $k > 0$ fixed, we have:

$$F^{nk}(a_{nk}x + b_{nk}) \to G(x) =: U(x)$$

and

$$F^{nk}(a_n x + b_n) = \{F^n(a_n x + b_n)\}^k \to G^k(x) =: V(x).$$

The conclusion follows from Lemma 4.1.7. $\square$

The next theorem is the main result of this section (see, e.g, Theorem 2.1 of [2]).

Theorem 4.1.9. *G is a univariate EV distribution if and only if $G = G_\gamma$ for some $\gamma \in \mathbb{R}$.*

Proof. a) We first show that if G is an EV distribution then $G = G_\gamma$ for some γ. We follow the lines of the proof given on p. 46-48 of [2]. Since G is an EV distribution, there exist a sequence $(X_i)_{i\geq 1}$ of independent and identically distributed random variables (with distribution F) and some sequences $(a_n)_n \in \mathbb{R}_+$ and $(b_n)_n \in \mathbb{R}$ such that (4.1) holds, where $M_n = \max_{i\leq n} X_i$. We assume for simplicity that F has density function f. In particular, F is continuous. Let F_n be the distribution function of M_n. Thus $F_n(x) = F^n(x)$ (see (4.2)) and $F_n'(x) = nF^{n-1}(x)f(x)$. We denote $Y_n = (M_n - b_n)/a_n$.

Let $z : \mathbb{R} \to \mathbb{R}$ be a continuous bounded function. By Theorem A.1.8 (Appendix A), we have:

$$E[z(Y_n)] \to E[z(Y)].$$

Note that

$$\begin{aligned} E[z(Y_n)] &= E\left[z\left(\frac{M_n - b_n}{a_n}\right)\right] = \int_{\mathbb{R}} z\left(\frac{x-b_n}{a_n}\right) dF_n(x) = \int_{\mathbb{R}} z\left(\frac{x-b_n}{a_n}\right) F_n'(x)dx \\ &= n\int_{\mathbb{R}} z\left(\frac{x-b_n}{a_n}\right) F^{n-1}(x)f(x)dx. \end{aligned}$$

We use the substitution $F(x) = 1 - \frac{1}{y}$, i.e. $x = F^{-1}\left(1-\frac{1}{y}\right) =: U(y)$. The Jacobian of this transformation is

$$dx = \frac{1}{F'(F^{-1}(1-1/y))}\frac{1}{y^2}dy = \frac{1}{f(U(y))}\frac{1}{y^2}dy.$$

Note that $F^{n-1}(x) = [F(U(y))]^{n-1} = (1-1/y)^{n-1}$ since $F(U(y)) = F(F^{-1}(1-1/y)) = 1 - 1/y$. We obtain:

$$\begin{aligned} E[z(Y_n)] &= = n\int_1^\infty z\left(\frac{U(y)-b_n}{a_n}\right)\left(1-\frac{1}{y}\right)^{n-1} f(U(y))\frac{1}{f(U(y))}\frac{1}{y^2}dy \\ &= n\int_1^\infty z\left(\frac{U(y)-b_n}{a_n}\right)\left(1-\frac{1}{y}\right)^{n-1}\frac{1}{y^2}dy. \end{aligned}$$

We now use the substitution $y = n/v$. We obtain:

$$E[z(Y_n)] = \int_0^n z\left(\frac{U(n/v)-b_n}{a_n}\right)\left(1-\frac{v}{n}\right)^{n-1} dv.$$

Assume that $b_n = U(n)$, $a_n = a(n)$ for certain functions $a : \mathbb{R}_+ \to \mathbb{R}_+$ and $U : \mathbb{R}_+ \to \mathbb{R}$, where U satisfies (4.3). By Lemma 4.1.2, there exists some $\gamma \in \mathbb{R}$ such that $h(u) = h_\gamma(u)$ for all $u > 0$. Then taking $u = 1/v$ in relation (4.3) we have

$$\lim_{n\to+\infty} \frac{U(n/v) - b_n}{a_n} = h_\gamma\left(\frac{1}{v}\right), \quad \text{for any} \quad v > 0.$$

Since $\lim\limits_{n\to+\infty}(1 - v/n)^{n-1} = e^{-v}$ by the Dominated Convergence Theorem, it follows that

$$E[z(Y)] = \lim_{n\to+\infty} E[z(Y_n)] = \int_0^\infty z\left(h_\gamma\left(\frac{1}{v}\right)\right) e^{-v} dv =: I.$$

To evaluate the integral I, we consider separately three cases:

(i) If $\gamma > 0$, we use the substitution $s = (v^{-\gamma} - 1)/\gamma$. The inverse transformation is $v = (1 + \gamma s)^{-1/\gamma}$. In this case,

$$\begin{aligned} I &= \int_0^\infty z\left(\frac{v^{-\gamma} - 1}{\gamma}\right) e^{-v} dv = \int_{-1/\gamma}^\infty z(s) \exp\{-(1+\gamma s)^{-1/\gamma}\} \left[-\frac{1}{\gamma}(1+\gamma s)^{-\frac{1}{\gamma}-1}\gamma\right] ds \\ &= \int_{-\infty}^\infty z(s) dG_\gamma(s), \end{aligned}$$

where G_γ is given by (4.5).

(ii) If $\gamma < 0$, we use the same substitution $s = (v^{-\gamma} - 1)/\gamma$, but in this case

$$\begin{aligned} I &= \int_0^\infty z\left(\frac{v^{-\gamma} - 1}{\gamma}\right) e^{-v} dv = \int_{-\infty}^{-1/\gamma} z(s) \exp\{-(1+\gamma s)^{-1/\gamma}\} \left[-\frac{1}{\gamma}(1+\gamma s)^{-\frac{1}{\gamma}-1}\gamma\right] ds \\ &= \int_{-\infty}^\infty z(s) dG_\gamma(s), \end{aligned}$$

where G_γ is given by (4.6).

(iii) If $\gamma = 0$, then $h_0(u) = \ln(u)$ and $h_0\left(\frac{1}{v}\right) = \ln\left(\frac{1}{v}\right) = -\ln(v)$. We use the substitution $s = -\ln(v)$. The inverse transformation is $v = e^{-s}$. We obtain:

$$I = \int_0^\infty z(-\ln(v))e^{-v} dv = \int_{-\infty}^\infty z(s)e^{-e^{-s}}(-e^{-s})ds = \int_{-\infty}^\infty z(s) dG_0(s),$$

where G_0 is given by (4.7).

b) Suppose now that $G = G_\gamma$ for some $\gamma \in \mathbb{R}$. By Lemma 4.1.5 we know that G_γ is a max-stable distribution. Hence, by Theorem 4.1.6 G_γ is an EV distribution. □

Lemma 4.1.10. *The class of EV distributions is a location-scale family i.e. if Y has EV distribution G then*

$$\widetilde{Y} = \mu + \sigma Y$$

has also an EV distribution, for any $\mu \in \mathbb{R}$ and $\sigma > 0$.

Proof. Since Y has an EV distribution, there exist a sequence $(X_i)_{i\geq 1}$ of independent and identically distributed random variables and some sequences $(a_n)_n \subset \mathbb{R}_+, (b_n)_n \subset \mathbb{R}$ such that (4.1) holds, where $M_n = \max_{i\leq n} X_i$. By Slutsky Theorem, it follows that

$$\mu + \frac{\max_{i\leq n}(\sigma X_i) - \sigma b_n}{a_n} = \mu + \frac{\sigma(M_n - b_n)}{a_n} \xrightarrow{d} \mu + \sigma Y = \widetilde{Y}.$$

This means that

$$\frac{\max_{i\leq n}(\sigma X_i) - (\sigma b_n - \mu a_n)}{a_n} \xrightarrow{d} \widetilde{Y}$$

So there exist a sequence $\{\widetilde{X_i} = \sigma X_i, i \geq 1\}$ of independent and identically distributed random variables and sequences $\widetilde{a_n} = a_n, \widetilde{b_n} = \sigma b_n - \mu a_n$ such that

$$\frac{\max_{i\leq n} \widetilde{X_i} - \widetilde{b_n}}{\widetilde{a_n}} \xrightarrow{d} \widetilde{Y}.$$

This proves that $\widetilde{Y}$ has an EV distribution. $\square$

Remark 4.1.11. Lemma 4.1.10 can be restated in terms of distributions as follows. If G is an EV distribution then

$$\widetilde{G}(x) = G\left(\frac{x-\mu}{\sigma}\right), \quad \text{for any } x \in \mathbb{R}$$

is also an EV distribution, for any $\mu \in \mathbb{R}$ and $\sigma > 0$. In particular, by Theorem 4.1.9, it follows that the class of EV distributions coincides with the following class:

$$\left\{\widetilde{G}_\gamma(x) = G_\gamma\left(\frac{x-\mu}{\sigma}\right); \gamma \in \mathbb{R}, \mu \in \mathbb{R}, \sigma > 0\right\}.$$

Remark 4.1.12. The class of EV distributions essentially involves three types of extreme value distributions which depend on the shape parameter γ. The sub-families defined by $\gamma = 0$, $\gamma > 0$ and $\gamma < 0$ correspond, respectively, to the Gumbel, Fréchet and Weibull-type families, whose cumulative distribution functions are defined below.

- Fréchet distribution ($\gamma = \alpha^{-1} > 0$)

$$G(x) = \begin{cases} \exp\left\{-\left(\frac{x-\mu}{\sigma}\right)^{-\alpha}\right\}, & \text{if } x > \mu \\ 0, & \text{otherwise} \end{cases}. \tag{4.12}$$

- Weibull-type distribution ($\gamma = -\alpha^{-1} < 0$)

$$G(x) = \begin{cases} \exp\left\{-\left(-\frac{x-\mu}{\sigma}\right)^{-\alpha}\right\}, & \text{if } x < \mu \\ 1, & \text{otherwise} \end{cases}. \tag{4.13}$$

- Gumbel distribution

$$G(x) = \exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right], \text{ for all } x \in \mathbb{R}. \tag{4.14}$$

Remark 4.1.13. The regularly varying distributions have a close relation with extreme value theory. Indeed, regularly varying distributions belong to the max-domain of attraction of the Fréchet distribution in a general multivariate case (see Proposition 4.2.13).

4.2 Multivariate extreme value theory

In this section, we introduce the class of multivariate extreme valued distributions and we discuss some of their properties. We begin by introducing some notation. If $\mathbf{a}, \mathbf{b}, \mathbf{x} \in \mathbb{R}^d$ then we denote

$$\mathbf{ax} + \mathbf{b} = (a_1 x_1 + b_1, \ldots, a_d x_d + b_d)$$

and we let

$$\frac{1}{\mathbf{a}} = \left(\frac{1}{a_1}, \ldots, \frac{1}{a_d}\right) \text{ if } a_j \neq 0 \text{ for all } j = 1, \ldots, d.$$

If $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are random vectors in $\mathbb{R}^d$ with $\mathbf{X}_i = (X_{i,j})_{1 \le j \le d}$, we denote by $\mathbf{M}_n = \max_{i \le n} \mathbf{X}_i$ the random vector defined by $\mathbf{M}_n = (M_{n,j})_{1 \le j \le d}$ with $M_{n,j} = \max_{i \le n} X_{i,j}$, for all $j = 1, \ldots, d$. If $(\mathbf{a}_n)_n$ and $(\mathbf{b}_n)_n$ are sequences in $\mathbb{R}^d$ with $\mathbf{a}_n = (a_{n,j})_{1 \le j \le d}$ and $\mathbf{b}_n = (b_{n,j})_{1 \le j \le d}$, we denote

$$\frac{\mathbf{M}_n - \mathbf{b}_n}{\mathbf{a}_n} = \left(\frac{M_{n,j} - b_{n,j}}{a_{n,j}}\right)_{1 \le j \le d}.$$

If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ then we write $\mathbf{x} \le \mathbf{y}$ if $x_j \le y_j$ for all $j = 1, \ldots, d$.

Definition 4.2.1. *Let G be a distribution on $\mathbb{R}^d$. We say that G is a **multivariate extreme value** (MEV) distribution if there exist a sequence $(\mathbf{X}_i)_{i \ge 1}$ of independent and identically distributed random vectors in $\mathbb{R}^d$ (with distribution F) and sequences $(\mathbf{a}_n)_n \subset \mathbb{R}^d_+$ and $(\mathbf{b}_n)_n \subset \mathbb{R}^d$ such that*

$$\frac{\mathbf{M}_n - \mathbf{b}_n}{\mathbf{a}_n} \xrightarrow{d} \mathbf{Y} \tag{4.15}$$

where $\mathbf{Y}$ *is a random vector with distribution* G *and* $\mathbf{M}_n = \max_{i \leq n} \mathbf{X}_i$*. In this case, we say that* F *is in the* ***max-domain of attraction*** *of* G *and we write* $F \in MaxD(G)$*.*

Remark 4.2.2. Note that (4.15) can also be written as:

$$F^n(\mathbf{a}_n\mathbf{x} + \mathbf{b}_n) \xrightarrow{w} G(\mathbf{x}), \tag{4.16}$$

where $\xrightarrow{w}$ denotes the weak convergence. To see this, note that (4.15) is equivalent to

$$P\left(\frac{\mathbf{M}_n - \mathbf{b}_n}{\mathbf{a}_n} \leq \mathbf{x}\right) \to P(\mathbf{Y} \leq \mathbf{x}) = G(\mathbf{x})$$

for all continuity points $\mathbf{x}$ of G, and

$$\begin{aligned} P\left(\frac{\mathbf{M}_n - \mathbf{b}_n}{\mathbf{a}_n} \leq \mathbf{x}\right) &= P(\mathbf{M}_n \leq \mathbf{a}_n\mathbf{x} + \mathbf{b}_n) \\ &= P\left(\bigcap_{i=1}^{n} \{\mathbf{X}_i \leq \mathbf{a}_n\mathbf{x} + \mathbf{b}_n\}\right) \\ &= \prod_{i=1}^{n} P(\mathbf{X}_i \leq \mathbf{a}_n\mathbf{x} + \mathbf{b}_n) \\ &= \{P(\mathbf{X}_1 \leq \mathbf{a}_n\mathbf{x} + \mathbf{b}_n)\}^n \\ &= F^n(\mathbf{a}_n\mathbf{x} + \mathbf{b}_n). \end{aligned}$$

The multivariate convergence in (4.15) implies the marginal convergence to a univariate EV distribution. To proe this, we need the following lemma:

Lemma 4.2.3. *Let* $(\mathbf{Y}_n)_{n\geq 1}$ *and* $\mathbf{Y}$ *be vectors in* $\mathbb{R}^d$ *such that* $\mathbf{Y}_n = (Y_{n,j})_{1\leq j\leq d}$ *and* $\mathbf{Y} = (Y_j)_{1\leq j\leq d}$*. If* $\mathbf{Y}_n \xrightarrow{d} \mathbf{Y}$ *then* $Y_{n,j} \xrightarrow{d} Y_j$*, for all* $j = 1, \ldots, d$*.*

Proof. Fix $j = 1, \ldots, d$. Let $f : \mathbb{R} \to \mathbb{R}$ be continuous and bounded function. Consider the function $h : \mathbb{R} \to \mathbb{R}$ defined by $h(x_1, \ldots, x_d) = f(x_j)$. Clearly, h is continuous and bounded. Hence,

$$E[f(Y_{n,j})] = E[h(Y_n)] \to E[h(Y)] = E[f(Y_j)].$$

□

Corollary 4.2.4. *If (4.15) holds then*

$$\frac{M_{n,j} - b_{n,j}}{a_{n,j}} \xrightarrow{d} Y_j, \quad \textit{for all } j = 1, \ldots d \tag{4.17}$$

where Y_j *has distribution* G_j*,* G_j *being the* j*-th marginal of* G*. This shows that if* G *is MEV and* $F \in MaxD(G)$ *then the* j*-th marginal of* G *is a EV distribution and* $F_j \in MaxD(G_j)$*, for all* $j = 1, \ldots, d$*.*

Proof. This is an immediate consequence of Lemma 4.2.3. □

Remark 4.2.5. Note that (4.17) can be written as

$$F_j{}^n(a_{n,j}x_j + b_{n,j}) \xrightarrow{w} G_j(x_j) \tag{4.18}$$

for any x_j which is a continuity point of G_j. (This follows by the same argument as in Remark 4.2.2).

Remark 4.2.6. Since G_j is an extreme value distribution, G_j is continuous. Therefore, G is continuous. (A distribution in $\mathbb{R}^d$ which has all marginal distributions continuous is continuous). Hence, if G is a MEV distribution, relation (4.16) holds for *all* $x \in \mathbb{R}^d$.

Now, we extend the univariate equivalence between max-stability and EV distributions given by Theorem
refThm 4.1.6 to the multivariate case. First, we introduce the definition of max-stable distributions.

Definition 4.2.7. *A distribution G on $\mathbb{R}^d$ is called **max-stable** if for any positive integer k, there exist $\alpha_k \in \mathbb{R}^d_+$ and $\beta_k \in \mathbb{R}^d$ such that*

$$G^k(\mathbf{x}) = G(\alpha_k \mathbf{x} + \beta_k), \text{ for all } \mathbf{x} \in \mathbb{R}^d. \tag{4.19}$$

Remark 4.2.8. Note that (4.19) is equivalent to saying that

$$\max_{i \le k} \mathbf{Y}_i \stackrel{d}{=} \frac{\mathbf{Y} - \beta_k}{\alpha_k} \tag{4.20}$$

where $(\mathbf{Y}_i)_{i \ge 1}$ are independent and identically distributed random vectors in $\mathbb{R}^d$ with distribution G. To see this, we note that

$$P(\max_{i \le k} \mathbf{Y}_i \le \mathbf{x}) = P\left(\bigcap_{i=1}^{k} \{\mathbf{Y_i} \le \mathbf{x}\}\right) = \{P(\mathbf{Y_1} \le \mathbf{x})\}^k = G^k(\mathbf{x})$$

and

$$P\left(\frac{\mathbf{Y} - \beta_k}{\alpha_k} \le \mathbf{x}\right) = P(\mathbf{Y} \le \alpha_k \mathbf{x} + \beta_k) = G(\alpha_k \mathbf{x} + \beta_k).$$

Remark 4.2.9. Relation (4.19) has an equivalent formulation which we explain below. In (4.19), we let $\mathbf{y} = \alpha_k \mathbf{x} + \beta_k$. Then $\mathbf{x} = \frac{\mathbf{y} - \beta_k}{\alpha_k}$ (recall that all these operations are done component-wise, i.e $x_j = \frac{y_j - \beta_{k,j}}{\alpha_{k,j}}$ for all $j = 1, \ldots, d$). So (4.19) becomes

$$G^k\left(\frac{\mathbf{y} - \beta_k}{\alpha_k}\right) = G(\mathbf{y}), \text{ for all } \mathbf{y} \in \mathbb{R}^d. \tag{4.21}$$

We denote ${\alpha_k}' = \frac{1}{\alpha_k}$ and ${\beta_k}' = \frac{-\beta_k}{\alpha_k}$. Hence, (4.21) is equivalent to

$$G^k({\alpha_k}'\mathbf{y} + {\beta_k}') = G(\mathbf{y}) \text{ for all } \mathbf{y} \in \mathbb{R}^d.$$

This coincides with relation (8.2) of [2].

The following theorem is the main result of this section (see page 255 of [2]).

Theorem 4.2.10. *The class of multivariate extreme value distributions coincides with the class of max-stable distribution.*

Proof. We follow the lines of the proof of Proposition 5.9 in [9] (see also page 255 of [2]). The proof is divided into two parts.

a) Suppose that G is a max-stable distribution i.e. (4.19) holds for any $k > 0$. We have to prove that there exist some independent and identically distributed random vectors $(\mathbf{X}_i)_{i\geq 1}$ in $\mathbb{R}^d$ (with distribution function F) and some sequences $(\mathbf{a}_n)_n \subset \mathbb{R}^d_+$ and $(\mathbf{b}_n)_n \subset \mathbb{R}^d$ such that (4.15) holds. We take $F = G$, $\mathbf{X}_i = \mathbf{Y}_i$, where $(\mathbf{Y}_i)_{i\geq 1}$ are independent and identically distributed random vectors with law G, $\mathbf{a}_n = \frac{1}{\alpha_n}$ and $\mathbf{b}_n = \frac{-\beta_n}{\alpha_n}$. Then, using relation (4.20) with $k = n$,

$$\frac{\mathbf{M}_n - \mathbf{b}_n}{\mathbf{a}_n} = \frac{\max_{i\leq n}\mathbf{Y}_i + \frac{\beta_n}{\alpha_n}}{1/\alpha_n} = \alpha_n \max_{i\leq n}\mathbf{Y}_i + \beta_n \stackrel{d}{=} \mathbf{Y}$$

for any n. This clearly implies that $\dfrac{\mathbf{M}_n - \mathbf{b}_n}{\mathbf{a}_n} \xrightarrow{d} \mathbf{Y}$.

b) Suppose that G is a MEV distribution i.e. there exist a sequence $(\mathbf{X}_i)_{i\geq 1}$ of independent and identically distributed random vectors and there exist some sequences $(\mathbf{a}_n)_n \subset \mathbb{R}^d_+$ and $(\mathbf{b}_n)_n \subset \mathbb{R}^d$ such that (4.15) holds, where $\mathbf{M}_n = \max_{i\leq n}\mathbf{X}_i$. We have to prove that G is a max-stable distribution, i.e. for any $k > 0$ there exist $\alpha_k \in \mathbb{R}^d_+$ and $\beta_k \in \mathbb{R}^d$ such that (4.19) holds. By Remark 4.2.9, it suffices to prove that (4.21) holds.

Recall that (4.15) implies the convergence in distribution of the marginals, i.e. relation (4.17) holds for any $j = 1, \ldots, d$. This means that relation (4.18) holds for any $j = 1, \ldots, d$. By Lemma 4.1.8, for any $k > 0$, there exist $\alpha_{k,j} \in \mathbb{R}_+$ and $\beta_{k,j} \in \mathbb{R}$ such that

$$\lim_{n\to+\infty} \frac{a_{n,j}}{a_{nk,j}} = \alpha_{k,j}, \quad \lim_{n\to+\infty} \frac{b_{n,j} - b_{nk,j}}{a_{nk,j}} = \beta_{k,j}, \tag{4.22}$$

and $G_j^k(x) = G_j(\alpha_{k,j}x + \beta_{k,j})$ for any $x \in \mathbb{R}$.

Let $\mathbf{Y}_k$ be a random vector in $\mathbb{R}^d$ with distribution G^k, for any $k > 0$ fixed and $\mathbf{Y} = \mathbf{Y}_1$. By (4.16),

$$P\left(\frac{\mathbf{M}_{nk} - \mathbf{b}_n}{\mathbf{a}_n} \leq \mathbf{x}\right) = P(\mathbf{M}_{nk} \leq \mathbf{a}_n\mathbf{x} + \mathbf{b}_n) = \{F^n(\mathbf{a}_n\mathbf{x} + \mathbf{b}_n)\}^k \to G^k(x).$$

This means that

$$\frac{\mathbf{M}_{nk}-\mathbf{b}_n}{\mathbf{a}_n} \xrightarrow{d} \mathbf{Y}_k. \tag{4.23}$$

In particular,

$$\frac{M_{nk,j}-b_{n,j}}{a_{n,j}} \xrightarrow{d} Y_{k,j} \quad \text{for any } j=1,\ldots,d.$$

Note that for any $k>0$ fixed, by (4.15), if we focus on the subsequence $\{nk\}_{n\geq 1}$,

$$\frac{\mathbf{M}_{nk}-\mathbf{b}_{nk}}{\mathbf{a}_{nk}} \xrightarrow{d} \mathbf{Y} \sim G. \tag{4.24}$$

We write

$$\begin{aligned}\frac{\mathbf{M}_{nk}-\mathbf{b}_{nk}}{\mathbf{a}_{nk}} &= \left(\frac{M_{nk,j}-b_{nk,j}}{a_{nk,j}}; 1\leq j\leq d\right)\\ &= \left(\frac{M_{nk,j}-b_{n,j}}{a_{n,j}}\cdot\frac{a_{n,j}}{a_{nk,j}}+\frac{b_{n,j}-b_{nk,j}}{a_{nk,j}}; 1\leq j\leq d\right).\end{aligned}$$

The term on the left-hand side of the previous relation converges in distribution to $\mathbf{Y}$ by (4.24). The term on the right-hand side converges to $(\alpha_{k,j}Y_{k,j}+\beta_{k,j}; 1\leq j\leq d) = \alpha_k\mathbf{Y}_k+\beta_k$ due to (4.22) and (4.23), by a multivariate version of Slutsky's theorem. Since the limit has to be the same, we infer that

$$\mathbf{Y} \stackrel{d}{=} \alpha_k\mathbf{Y}_k+\beta_k, \tag{4.25}$$

which means that the two random vectors have the same distribution. We obtain that for any $\mathbf{x}\in\mathbb{R}^d$

$$\begin{aligned}G(\mathbf{x}) = P(\mathbf{Y}\leq\mathbf{x}) &= P(\alpha_k\mathbf{Y}_k+\beta_k\leq\mathbf{x})\\ &= P\left(\mathbf{Y}_k\leq\frac{\mathbf{x}-\beta_k}{\alpha_k}\right) = G^k\left(\frac{\mathbf{x}-\beta_k}{\alpha_k}\right).\end{aligned}$$

This proves (4.21). We conclude that G is a max-stable distribution. $\square$

Recall that in the univariate case, we proved that any EV distribution has to be of the form $G_\gamma\left(\frac{x-\mu}{\sigma}\right)$ for some $\gamma\in\mathbb{R},\mu\in\mathbb{R}$ and $\sigma>0$ (see Remark 4.1.11). Since the marginals of a MEV distribution are EV distributions (see Corollary 4.2.4), we obtain the following result:

Theorem 4.2.11. *Let G be a MEV distribution, and G_j be its j-th marginal for $j=1,\ldots,d$. If G_j is given by*

$$G_j(x_j) = G_{\gamma_j}\left(\frac{x_j-\mu_j}{\sigma_j}\right) \text{ for all } x_j\in\mathbb{R}^d,$$

for some $\gamma_j \in \mathbb{R}, \mu_j \in \mathbb{R}$ *and* $\sigma_j \in \mathbb{R}_+$*, then :*

$$G_j^{-1}(e^{-1/z_j}) = \mu_j + h_{\gamma_j}(z_j) \text{ for all } z_j \in \mathbb{R}_+, \tag{4.26}$$

where h_{γ_j} *is given by (4.4).*

Proof. We consider only the case $\gamma_j > 0$. (The case $\gamma_j < 0$ is similar.) We solve for x_j in the following equation:

$$e^{-1/z_j} = G_j(x_j). \tag{4.27}$$

Using the formula of G_j and assuming that $1 + \gamma_j \frac{x_j - \mu_j}{\sigma_j} > 0$, this equation becomes:

$$e^{-1/z_j} = \exp\left\{-\left(1 + \gamma_j \frac{x_j - \mu_j}{\sigma_j}\right)^{-1/\gamma_j}\right\},$$

or equivalently

$$-\frac{1}{z_j} = -\left(1 + \gamma_j \frac{x_j - \mu_j}{\sigma_j}\right)^{-1/\gamma_j}, \text{ i.e. } z_j^{-\gamma_j} = 1 + \gamma_j \frac{x_j - \mu_j}{\sigma_j}.$$

Hence, $x_j = \mu_j + \sigma_j \frac{z_j^{-\gamma_j} - 1}{\gamma_j}$. Therefore, applying G_j^{-1} to both sides of equation (4.27) we get:

$$G_j^{-1}(e^{-1/z_j}) = x_j = \mu_j + \sigma_j h_{\gamma_j}(z_j). \tag{4.28}$$

We now consider the case $\gamma_j = 0$. In this case,

$$G_j(x_j) = G_0\left(\frac{x_j - \mu_j}{\sigma_j}\right) = \exp\{-\exp\left(-\frac{x_j - \mu_j}{\sigma_j}\right)\}.$$

As in the previous case, we solve for x_j in the equation (4.27) which becomes :

$$e^{-1/z_j} = \exp\left\{-\exp\left(-\frac{x_j - \mu_j}{\sigma_j}\right)\right\},$$

or equivalently

$$\frac{1}{z_j} = \exp\left(-\frac{x_j - \mu_j}{\sigma_j}\right).$$

Hence $-\frac{x_j - \mu_j}{\sigma_j} = \ln(\frac{1}{z_j}) = -\ln(z_j)$, i.e $x_j = \mu_j + \sigma_j \ln(z_j)$. Therefore, applying G_j^{-1} to both sides of equation (4.27), we obtain again (4.28). $\square$

We consider the transformation $T_G : \mathbb{R}_+^d \to \mathbb{R}_+^d$ defined by

$$T_G(x_1, \dots, x_d) = \left(-\frac{1}{\log G_1(x_1)}, \dots, -\frac{1}{\log G_d(x_d)} \right). \tag{4.29}$$

Note that T_G depends only on the marginals $G_1, \dots, G_d$ of G.

The next result shows how a multivariate distribution G can be transformed to a "standard" version (see page 256 of [2]).

Theorem 4.2.12. *Let $\mathbf{Y} = (Y_i)_{1 \le j \le d}$ be a random vector with distribution function G and G_j be the j-th marginal of G for $j = 1, \dots, d$. Let G_* be the distribution function of the random vector*

$$\mathbf{Y}_* = T_G(\mathbf{Y}) = \left(-\frac{1}{\log G_1(Y_1)}, \dots, -\frac{1}{\log G_d(Y_d)} \right).$$

Then

(a) The marginals of G_* are Fréchet distributions with parameter $\alpha = 1$, i.e. for any $j = 1, \dots, d$

$$P\left(-\frac{1}{\log G_j(Y_j)} \le z_j \right) = e^{-1/z_j}, \text{ for all } z_j \in \mathbb{R}_+.$$

(b) For any $\mathbf{z} = (z_1, \dots, z_d) \in \mathbb{R}_+^d$,

$$G_*(\mathbf{z}) = G\left(G_1^{-1}(e^{-1/z_1}), \dots, G_d^{-1}(e^{-1/z_d}) \right). \tag{4.30}$$

(c) For any $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}_+^d$,

$$G(\mathbf{x}) = G_*\left(-\frac{1}{\log G_1(x_1)}, \dots, -\frac{1}{\log G_d(x_d)} \right). \tag{4.31}$$

Proof. (a) The result follows by Lemma A.2.3 (Appendix A).

(b) Using the same calculations as in the proof of Lemma A.2.3, we see that

$$\begin{aligned} G_*(\mathbf{z}) &= P\left(-\frac{1}{\log G_j(Y_j)} \le z_j, j = 1, \dots, d \right) = P(Y_j \le G_j^{-1}(e^{-1/z_j}), j = 1, \dots, d) \\ &= G(G_1^{-1}(e^{-1/z_1}), \dots, G_d^{-1}(e^{-1/z_d})). \end{aligned}$$

(c) This is proved similarly to (b). We omit the details.

□

The next result deals with a special case of the limiting distribution G, that occurs in case of multivariate regularly varying vectors.

Proposition 4.2.13. *Let $(\mathbf{X}_i)_{i\geq 1}$ be a sequence of independent and identically distributed multivariate regularly varying random vectors in $\mathbb{R}_+^d$ of index $\alpha > 0$ with limit measure ν. Then*

$$\frac{1}{a_n}\max_{i\leq n}\mathbf{X}_i \xrightarrow{w} \mathbf{Y} \sim G \tag{4.32}$$

where

$$G(\mathbf{x}) = e^{-\nu([\mathbf{0},\mathbf{x}]^c)} \text{ for all } \mathbf{x} \in \mathbb{R}_+^d. \tag{4.33}$$

In particular, G is a MEV distribution and the marginal distributions of G are Fréchet with index α, i.e.

$$G_j(x_j) = e^{-c_j x_j^{-\alpha}} \quad \text{for all } x_j > 0,$$

where $c_j = \nu(\mathbb{R}_+ \times \ldots \times (1,\infty) \times \ldots \times \mathbb{R}_+)$.

Proof. Recall that, by the definition of multivariate regular variation, $nP(\mathbf{X}_1 > a_n\mathbf{x}) \to \nu([\mathbf{0},\mathbf{x}]^c)$ for all $\mathbf{x} \in \mathbb{R}_+^d$ (see Definition 3.2.1). Using the fact that

$$\left(1 + \frac{\mathbf{x}_n}{n}\right) \to e^{\mathbf{x}} \text{ if } \mathbf{x}_n \to \mathbf{x},$$

we obtain that for any $\mathbf{x} \in \mathbb{R}_+^d$,

$$\begin{aligned}
\lim_{n\to+\infty} P(\max_{i\leq n}\mathbf{X}_i \leq a_n\mathbf{x}) &= \lim_{n\to+\infty} P(\mathbf{X}_1 \leq a_n\mathbf{x})^n \\
&= \lim_{n\to+\infty} \left(1 - \frac{nP(\mathbf{X}_1 > a_n\mathbf{x})}{n}\right)^n \\
&= e^{-\nu([\mathbf{0},\mathbf{x}]^c)}.
\end{aligned}$$

To simplify the writing, we assume that $d = 2$. (The case of arbitrary $d \geq 1$ is similar.) We denote by ν_1 the first marginal of ν. Since $(\mathbf{X}_i)_{i\geq 1}$ are independent and identically distributed we have for all $x_1 > 0$

$$\begin{aligned}
G_1(x_1) &= \lim_{x_2\to+\infty} G(x_1,x_2) = \lim_{x_2\to+\infty} e^{-\nu([\mathbf{0},\mathbf{x}]^c)} \\
&= e^{-\nu((x_1,\infty)\times\mathbb{R}_+)} = e^{-\nu_1((x_1,\infty))} = e^{-x_1^{-\alpha}\nu_1((1,\infty))} = e^{-c_1 x_1^{-\alpha}},
\end{aligned}$$

where $c_1 = \nu_1((1,\infty))$. Here we used the following scaling property of the measure ν_1:

$$\nu_1(sA) = s^{-\alpha}\nu_1(A)$$

which is inherited from the scaling property (3.8) of the measure ν. Similarly it can be proved that $G_2(x_2) = e^{-c_2 x_2^{-\alpha}}$ for all $x_2 > 0$, where $c_2 = \nu_2((1,\infty))$ and ν_2 is the second marginal of ν. $\square$

Remark 4.2.14. Propostion 4.2.13 shows that in the case of a sequence $\{\mathbf{X}_i\}_{i\geq 1}$ of multivariate regularly varying random vectors, the exponent measure (given by Remark 3.2.2) of the limiting distribution G of the normalized maximum $\mathbf{M}_n = \max_{i\leq n} \mathbf{X}_i$ coincides with the limit measure of $\mathbf{X}_1$.

Remark 4.2.15. Let G be a distribution function on $\mathbb{R}_+$ given by

$$G(x) = e^{-cx^{-\alpha}}, \quad \text{for all } x > 0$$

for some $\alpha > 0$ and $c > 0$. Then for all $x > 0$ and $s > 0$,

$$G(sx) = e^{-cs^{-\alpha}x^{-\alpha}} = (e^{-cx^{-\alpha}})^{s^{-\alpha}} = [G(x)]^{s^{-\alpha}}$$

and hence

$$\log G(sx) = s^{-\alpha} \log G(x). \tag{4.34}$$

The next result shows that if the random vector $\mathbf{X}$ is multivariate regularly varying and lies in the max-domain of attraction of the vector $\mathbf{Y}$ with distribution G, then the transformed vector $T_G(\mathbf{X})$ is in the max-domain of attraction of $\mathbf{Y}_* = T_G(\mathbf{Y})$.

Proposition 4.2.16. *Let $(\mathbf{X}_i)_{i\geq 1}$ be independent and identically distributed random vectors on $\mathbb{R}_+^d$ such that $\mathbf{X}_1$ is multivariate regularly varying with index $\alpha > 0$, limiting measure ν and normalizing sequence $\{a_n\}_n$, i.e. (3.4) holds. Let G be the distribution function given by (4.33). We assume that G is continuous and strictly increasing. We define $(\mathbf{X}_i)_* = T_G(\mathbf{X_i})$. Then*

$$\frac{1}{a_n^{\alpha}} \max_{i\leq n} (\mathbf{X}_i)_* \xrightarrow{d} \mathbf{Y}_* = T_G(\mathbf{Y}) \tag{4.35}$$

where $\mathbf{Y}$ is a random vector on $\mathbb{R}_+^d$ with distribution G and $\mathbf{Y}_*$ has distribution function G_* given by (4.30).

Proof. We denote $\mathbf{X}_i = (X_{i,1}, \ldots, X_{i,d})$. By (4.31) and (4.32), we have

$$\begin{aligned} P\left(\frac{1}{a_n}\max_{i\leq n} X_{i,1} \leq x_1, \ldots, \frac{1}{a_n}\max_{i\leq n} X_{i,d} \leq x_d\right) &\to G(x_1, \ldots, x_d) \\ &= G_*\left(-\frac{1}{\log G_1(x_1)}, \ldots, -\frac{1}{\log G_d(x_d)}\right). \end{aligned}$$

We denote $z_j = -\frac{1}{\log G_j(x_j)}$ for all $j = 1, \ldots, d$. Hence $x_j = G_j^{-1}(e^{-1/z_j})$ for all $j = 1, \ldots, d$, and the previous relation becomes:

$$P\left(\frac{1}{a_n}\max_{i\leq n} X_{i,1} \leq G_1^{-1}(e^{-1/z_1}), \ldots, \frac{1}{a_n}\max_{i\leq n} X_{i,d} \leq G_d^{-1}(e^{-1/z_d})\right) \to G_*(z_1, \ldots, z_d) \tag{4.36}$$

for all $\mathbf{z} = (z_1, \ldots, z_d) \in \mathbb{R}_+^d$. We examine separately the event

$$\left\{ \frac{1}{a_n} \max_{i \le n} X_{i,j} \le G_j^{-1}(e^{-1/z_j}) \right\}$$

for any $j = 1, \ldots, d$ fixed. Note that $\frac{1}{a_n} \max_{i \le n} X_{i,j} \le G_j^{-1}(e^{-1/z_j})$ means that $\frac{X_{i,j}}{a_n} \le G_j^{-1}(e^{-1/z_j})$ for all $i \le n$, which is equivalent to

$$G_j \left(\frac{X_{i,j}}{a_n} \right) \le e^{-1/z_j} \quad \text{for all } i \le n, \text{ i.e.}$$

$$\log G_j \left(\frac{X_{i,j}}{a_n} \right) \le -\frac{1}{z_j} \quad \text{for all } i \le n.$$

Using (4.34) with $x = X_{i,j}$ and $s = a_n^{-1}$, this last relationship is equivalent to

$$a_n^{\alpha} \log G_j(X_{i,j}) \le -\frac{1}{z_j} \quad \text{for all } i \le n,$$

which can be written also as

$$\frac{1}{a_n^{\alpha}} \left(-\frac{1}{\log G_j(X_{i,j})} \right) \le z_j \quad \text{for all } i \le n.$$

This means that

$$\frac{1}{a_n^{\alpha}} \max_{i \le n} \left(-\frac{1}{\log G_j(X_{i,j})} \right) \le z_j.$$

To summarize, we have proved that for any $j = 1, \ldots, d$ fixed,

$$\left\{ \frac{1}{a_n} \max_{i \le n} X_{i,j} \le G_j^{-1}(e^{-1/z_j}) \right\} = \left\{ \frac{1}{a_n^{\alpha}} \max_{i \le n} \left(-\frac{1}{\log G_j(X_{i,j})} \right) \le z_j \right\}.$$

Taking the intersection of these d events, we obtain

$$\begin{aligned}
&\left\{ \frac{1}{a_n} \max_{i \le n} X_{i,1} \le G_1^{-1}(e^{-1/z_1}), \ldots, \frac{1}{a_n} \max_{i \le n} X_{i,d} \le G_d^{-1}(e^{-1/z_d}) \right\} \\
&= \left\{ \frac{1}{a_n^{\alpha}} \max_{i \le n} \left(-\frac{1}{\log G_1(X_{i,1})} \right) \le z_1, \ldots, \frac{1}{a_n^{\alpha}} \max_{i \le n} \left(-\frac{1}{\log G_d(X_{i,d})} \right) \le z_d \right\} \\
&= \left\{ \frac{1}{a_n^{\alpha}} \max_{i \le n} T_G(\mathbf{X}_i) \le \mathbf{z} \right\}, \qquad (4.37)
\end{aligned}$$

where the second equality is due to the fact that

$$\max_{i \le n} \mathbf{W}_i := (\max_{i \le n} W_{i,1}, \ldots, \max_{i \le n} W_{i,d})$$

with $\mathbf{W}_i = T_G(\mathbf{X}_i)$. Finally, from (4.36) and (4.37), we infer that

$$P\left(\frac{1}{a_n^{\alpha}} \max_{i \le n} T_G(\mathbf{X}_i) \le \mathbf{z} \right) \to G_*(\mathbf{z})$$

for all $\mathbf{z} \in \mathbb{R}_+^d$. This concludes the proof. $\square$

Remark 4.2.17. Let G be a distribution function on $\mathbb{R}_+^d$ and G_* be given by (4.30). Let ν, ν be the exponent measures of G, G_* respectively. (Recall the definition of exponent measure given in Remark 3.2.2.) The link between ν and ν_* is given by

$$\nu((-\infty, \mathbf{x}]^c) = -\log G(\mathbf{x}) = -\log G_*(\mathbf{z}) = \nu_*([\mathbf{0}, \mathbf{z}]^c), \quad \text{for all } \mathbf{z} \in \mathbb{R}_+^d$$

where $z_j = -\frac{1}{\log G_1(x_1)}$ for all $j = 1, \dots, d$. We set

$$V_*(\mathbf{z}) := -\log G_*(\mathbf{z}) = \nu_*([\mathbf{0}, \mathbf{z}]^c). \tag{4.38}$$

While ν contains information about marginals and the dependence structure of G, ν_* is a "standard" version that contains information about the dependence structure only. Hence, we can have two distributions G and $\widetilde{G}$ with different marginals and different measures ν and $\widetilde{\nu}$, but with the same ν_*.

4.2.1 Measures of extreme dependence

Definition 4.2.18. *Let G be a distribution function on $\mathbb{R}_+^d$, G_* be given by (4.30) and V_* be given by (4.38). Let ν_* be the exponent measure of G_*. The **stable tail dependence function** of G is defined by*

$$\ell(\mathbf{v}) = V_*\left(\frac{1}{v_1}, \dots, \frac{1}{v_d}\right) = \nu_*([\mathbf{0}, (1/v_1, \dots, 1/v_d)]^c), \quad \textit{for all } \mathbf{v} \in \mathbb{R}_+^d. \tag{4.39}$$

Remark 4.2.19. Using the scaling property of the measure ν_*, $\nu_*(sA) = s^{-1}\nu_*(A)$, it can be proved that:

$$\ell(s\mathbf{v}) = s\ell(\mathbf{v}) \text{ for all } s \in \mathbb{R}_+.$$

Lemma 4.2.20 (Equation (8.13) and (8.14) in [2])**.** *Let G be a distribution function on $\mathbb{R}_+^d$ and G_* be given by (4.30).*

(a) The function ℓ defined by (4.39) can be written in terms of G as follows:

$$\ell(\mathbf{v}) = -\log G(G_1^{-1}(e^{-v_1}), \dots, G_d^{-1}(e^{-v_d})), \quad \textit{for all } \mathbf{v} \in \mathbb{R}_+^d. \tag{4.40}$$

(b) Conversely, we can reconstruct a distribution G from its marginal distributions G_j and its stable tail dependence function ℓ as follows:

$$-\log G(\mathbf{x}) = \ell(-\log G_1(x_1), \dots, -\log G_d(x_d)), \quad \textit{for all } \mathbf{x} \in \mathbb{R}^d.$$

Proof. (a) Using (4.30) and (4.38) we have

$$-\log G(G_1^{-1}(e^{-v_1}),\ldots,G_d^{-1}(e^{-v_d})) = -\log G_*\left(\frac{1}{v_1},\ldots,\frac{1}{v_d}\right) = V_*\left(\frac{1}{v_1},\ldots,\frac{1}{v_d}\right) = \ell(\mathbf{v}).$$

(b) Letting $z_j = -\frac{1}{\log G_j(x_j)}$ for $j = 1,\ldots,d$, we obtain:

$$-\log G(\mathbf{x}) = -\log G_*(\mathbf{z}) = V_*(\mathbf{z}) = \ell\left(\frac{1}{\mathbf{z}}\right) = \ell(-\log G_1(x_1),\ldots,-\log G_d(x_d)).$$

□

We now consider several examples of stable tail dependence functions ℓ.

Example 4.2.21. Let $\mathbf{Y} = (Y_1,\ldots,Y_d)$ where $Y_1,\ldots,Y_d$ are independent. Let G be the distribution function of $\mathbf{Y}$. Then $G(\mathbf{x}) = G_1(x_1)\ldots G_d(x_d)$ where G_j is the i-th marginal distribution of G and hence $-\log G(\mathbf{x}) = \sum_{j=1}^d -\log G_j(x_j)$. We write this relation for $x_j = G_j^{-1}(e^{v_j})$. Using (4.40), we obtain:

$$\ell(\mathbf{v}) = v_1 + \ldots + v_d.$$

Example 4.2.22. Let $\mathbf{Y} = (Y_1,\ldots,Y_d)$ where $Y_1 = \ldots = Y_d$. Then $\ell(\mathbf{v}) = \max_{i=1,\ldots,d} v_i$. To see this, we assume for simplicity that $d = 2$. (The case of arbitrary $d \geq 1$ is similar.) Let G be the distribution function of $\mathbf{Y} = (Y_1, Y_2)$ and G_j be the j-th marginal of G. Then $G_1 = G_2$ and

$$G(\mathbf{x}) = P(Y_1 \leq x_1, Y_1 \leq x_2) = P(Y_1 \leq x_1 \wedge x_2) = G_1(x_1 \wedge x_2) = G_1(x_1) \wedge G_1(x_2).$$

Using (4.40), we get

$$\begin{aligned}
\ell(v_1, v_2) &= -\log G(G_1^{-1}(e^{-v_1}), G_2^{-1}(e^{-v_2})) \\
&= -\log G(G_1^{-1}(e^{-v_1}), G_1^{-1}(e^{-v_2})) \\
&= -\log\{G_1(G_1^{-1}(e^{-v_1})) \wedge G_1(G_1^{-1}(e^{-v_2}))\} \\
&= -\log(e^{-v_1} \wedge e^{-v_2}) = v_1 \vee v_2.
\end{aligned}$$

Definition 4.2.23. *Let G be a distribution function on $\mathbb{R}_+^2$ and ℓ be the stable tail dependence function. The* ***Pickands dependence function*** *is defined by*

$$A(t) = \ell(1-t, t), \quad \textit{for all } t \in [0,1]. \tag{4.41}$$

Lemma 4.2.24. *We have the following relation between the Pickands dependence function A and the stable tail dependence function ℓ:*

$$\ell(v_1, v_2) = (v_1 + v_2)A\left(\frac{v_2}{v_1 + v_2}\right), \text{ for all } v_1, v_2 \in \mathbb{R}_+^2. \tag{4.42}$$

Proof. Using (4.41) and Remark 4.2.19, we have

$$(v_1 + v_2)A\left(\frac{v_2}{v_1 + v_2}\right) = (v_1 + v_2)\ell\left(\frac{v_1}{v_1 + v_2}, \frac{v_2}{v_1 + v_2}\right) = \ell(v_1, v_2).$$

□

Example 4.2.25. Let $\mathbf{Y} = (Y_1, Y_2)$, where Y_1, Y_2 are independent. Then $A(t) = 1$.

Example 4.2.26. Let $\mathbf{Y} = (Y_1, Y_2)$, where $Y_1 = Y_2$. Then $A(t) = t \vee (1 - t)$.

Chapter 5

Exploratory data analysis

5.1 Univariate data

In this section we will focus on exploratory data analysis techniques such as quantile-quantile and mean excess plots which provide a fair amount of information about the tail of a distribution. These graphical tools are used to help us decide on a reasonable model.

Definition 5.1.1. *Let $\{X_i\}_{i\geq 1}$ be a sequence of independent random variables with distribution function F. Arranging the sample in ascending order generates a new family of observations, written as $X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$, called* ***order statistics*** *associated with the original sample. In particular, the random variable $X_{i,n}$ denotes the i-th order statistic.*

5.1.1 Quantile-quantile plot

The quantile-quantile plot is an exploratory graphical tool used when trying to answer the goodness of fit question: does a particular model provide a plausible fit to the distribution of the random variable at hand?
The basic idea is to compare the theoretical quantiles with sample quantiles. If the data indeed follow the assumed distribution, then the points on the quantile-quantile plot will fall approximately on a straight line.

Example 5.1.2 (Exponential distribution)**.** Let X be an exponential random variable with distribution function $F(x) = 1 - e^{-\lambda x},\ \lambda > 0$. The quantile function has the form

$Q_\lambda(p) = \frac{-1}{\lambda}\log(1-p)$, for $p \in (0,1)$. Using the observations, we estimate the unknown quantile function Q by the empirical approximation $\hat{Q}_n$ defined by

$$\hat{Q}_n(p) = X_{i,n}, \text{ for } \frac{i-1}{n} < p \leq \frac{i}{n}.$$

In this section we will use $p = p_{i,n} := \frac{i}{n+1}$. In an orthogonal coordinate system, the points with values

$$(-\log(1-p), \hat{Q}_n(p)) = (-\log(1-p_{i,n}), X_{i,n}) \tag{5.1}$$

are plotted. Note that the parameter λ can be considered as a nuisance parameter here since its value is not our main focus at this moment. When a straight line pattern is obtained, we can say that the exponential model provides a plausible fit for the given population and the slope of a fitted line can be used as an estimate of $\frac{1}{\lambda}$.

In order to discuss quantile-quantile plot for Pareto distribution, we state the following Lemma.

Lemma 5.1.3. *Let X be the standard Pareto random variable with index $\alpha > 0$. Then $Y = \log(X)$ follows an exponential distribution.*

Proof. We know that $\overline{F}(x) = P(X > x) = x^{-\alpha}$. Then

$$P(Y > y) = P(\log(X) > y) = P(X > e^y) = e^{-\alpha y}.$$

Hence $Y \sim Exp(\alpha)$. $\square$

Example 5.1.4 (Pareto distribution)**.** Using Lemma 5.1.3 and (5.1) we can conclude that the quantile-quantile plot coordinates for the Pareto distribution are of the form

$$(-\log(1-p_{i,n}), \log(X_{i,n})).$$

Theorem 5.1.5. *Let X be a Log-normal random variable with mean μ and variance σ^2 then $Y = \log(X)$ follows a normal distribution.*

Proof. We know that $F(x) = P(X \leq x) = \Phi(\frac{\log(x)-\mu}{\sigma})$.

$$P(Y \leq y) = P(\log(X) \leq y) = P(X \leq e^y) = \Phi\left(\frac{\log(e^y)-\mu}{\sigma}\right) = \Phi\left(\frac{y-\mu}{\sigma}\right).$$

Hence $Y \sim N(\mu, \sigma^2)$. $\square$

Example 5.1.6 (Log-normal distribution)**.** Using Theorem 5.1.5 and the fact that $\Phi^{-1}(p)$ is the quantile function of a normal random variable, it follows that the quantile-quantile plot coordinates for the log-normal distribution are of the form

$$(\Phi^{-1}(p_{i,n}), \log(X_{i,n})).$$

Example 5.1.7 (Weibull distribution)**.** Let X be a Weibull random variable with distribution function $F(x) = 1 - e^{-\lambda x^\tau}$, $\lambda > 0$ and $\tau > 0$. The quantile function has the form

$$Q_\lambda(p) = \left(\frac{-1}{\lambda}\log(1-p)\right)^{1/\tau}, \text{ for } p \in (0,1).$$

Taking the logarithm of the quantile function we have

$$\log(Q(p)) = \frac{1}{\tau}\log\left(\frac{-1}{\lambda}\log(1-p)\right) = \frac{1}{\tau}\log\left(\frac{1}{\lambda}\right) + \frac{1}{\tau}\log(-\log(1-p)).$$

Hence, the quantile-quantile plot coordinates for the Weibull distribution are of the form

$$(\log(-\log(1-p)), \log(X_{i,n})).$$

5.1.2 Excess plot

In this section, we will introduce the concept of mean excess plot which is a graphical tool often used to understand the right tail behaviour of an underlying distribution.

Theoretical mean excess function

Definition 5.1.8. Let X be a nonnegative random variable. The **mean excess function** of X is defined by

$$e(t) = E(X - t | X > t), \text{ for any } t > 0$$

assuming that $E(X) < \infty$.

Example 5.1.9 (Exponential distribution)**.** Let X be an exponential random variable with distribution function $F(x) = 1 - e^{-\lambda x},\ \lambda > 0$. Then

$$\begin{aligned}
e(t) &= \frac{E[(X-t)1\{X>t\}]}{P(X>t)} \\
&= \frac{\int_t^\infty (x-t)f(x)dx}{\overline{F}(t)} \\
&= e^{\lambda t}\left[\int_t^\infty x\lambda e^{-\lambda x}dx - t\overline{F}(t)\right] \\
&= \lambda e^{\lambda t}\left[\frac{-e^{-\lambda x}(\lambda x+1)}{\lambda^2}\right]_t^\infty - t \\
&= \frac{1}{\lambda}e^{\lambda t}(0 + e^{-\lambda t} + \lambda t e^{-\lambda}) - t = \frac{1}{\lambda} + t - t = \frac{1}{\lambda}.
\end{aligned}$$

Hence, the mean excess function of an exponential random variable is $e(t) = \frac{1}{\lambda}$.

Example 5.1.10 (Pareto distribution)**.** Let X be a Pareto random variable with distribution function $F(x) = 1 - x^{-\alpha}$. Then for $\alpha > 1$,

$$\begin{aligned}
e(t) &= \frac{\int_t^\infty xf(x)dx - t\overline{F}(t)}{\overline{F}(t)} \\
&= t^\alpha\left[\int_t^\infty \alpha x^\alpha dx - t^{-\alpha+1}\right] \\
&= \alpha t^\alpha\left[\frac{x^{-\alpha+1}}{-\alpha+1}\right]_t^\infty - t \\
&= \alpha t^\alpha\left(0 - \frac{t^{-\alpha+1}}{-\alpha+1}\right) - t = \frac{\alpha t}{\alpha-1} - t = \frac{t}{\alpha-1}.
\end{aligned}$$

Hence, the mean excess function of a Pareto random variable is $e(t) = \frac{t}{\alpha-1}$.

Example 5.1.11 (Log-normal distribution)**.** Let X be a log-normal random variable with distribution function $F(x) = \Phi(\frac{\log(x)-\mu}{\sigma})$. Then

$$e(t) = \frac{\int_t^\infty xf(x)dx - t\overline{F}(t)}{\overline{F}(t)} = \frac{\int_t^\infty xf(x)dx}{\overline{\Phi}(\frac{\log t-\mu}{\sigma})} - t,$$

where $\overline{\Phi}(u) = 1 - \Phi(u)$.

First of all, we need to compute $\int_t^\infty xf(x)dx$ in order to obtain $e(t)$,

$$\begin{aligned}
\int_t^\infty xf(x)dx &= \frac{1}{\sigma\sqrt{2\pi}}\int_t^\infty e^{-\frac{(\log x-\mu)^2}{2\sigma^2}}dx \\
&= \frac{1}{\sigma\sqrt{2\pi}}\int_{\frac{\log t-\mu}{\sigma}}^\infty e^{-\frac{y^2}{2}}\sigma e^{\sigma y+\mu}dy \\
&= \frac{e^\mu}{\sqrt{2\pi}}\int_{\frac{\log t-\mu}{\sigma}}^\infty e^{-\frac{y^2}{2}+\sigma y}dy \\
&= \frac{e^{\mu+\frac{\sigma^2}{2}}}{\sqrt{2\pi}}\int_{\frac{\log t-\mu}{\sigma}}^\infty e^{-\frac{1}{2}(y-\sigma)^2}dy \\
&= e^{\mu+\frac{\sigma^2}{2}}\int_{\frac{\log t-\mu-\sigma^2}{\sigma}}^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz \\
&= e^{\mu+\frac{\sigma^2}{2}}\overline{\Phi}(\frac{\log t-\mu-\sigma^2}{\sigma}).
\end{aligned}$$

It follows that,

$$e(t) = e^{\mu+\frac{\sigma^2}{2}}\frac{\overline{\Phi}(\frac{\log t-\mu-\sigma^2}{\sigma})}{\overline{\Phi}(\frac{\log t-\mu}{\sigma})} - t.$$

In addition, we know that when $t\to\infty$,

$$\overline{\Phi}(t) \sim \frac{\phi(t)}{t} = t\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}},$$

where $f(t)\sim h(t)$ means that $\lim\limits_{t\to+\infty}\frac{\phi(t)}{h(t)} = 1$. Therefore,

$$\begin{aligned}
\frac{\overline{\Phi}(\frac{\log t-\mu-\sigma^2}{\sigma})}{\overline{\Phi}(\frac{\log t-\mu}{\sigma})} &\sim \frac{\phi(\frac{\log t-\mu-\sigma^2}{\sigma})}{(\frac{\log t-\mu-\sigma^2}{\sigma})} \div \frac{\phi(\frac{\log t-\mu}{\sigma})}{(\frac{\log t-\mu}{\sigma})} \\
&= \frac{e^{-\frac{(\frac{\log t-\mu-\sigma^2}{\sigma})^2}{2}}}{e^{-\frac{(\frac{\log t-\mu}{\sigma})^2}{2}}}\cdot\frac{\log t-\mu}{\log t-\mu-\sigma^2} \\
&= e^{\log t-\mu-\frac{\sigma^2}{2}}\frac{\log t-\mu}{\log t-\mu-\sigma^2}.
\end{aligned}$$

We can conclude that,

$$\begin{aligned}
e(t) &\sim e^{\mu+\frac{\sigma^2}{2}}e^{\log t-\mu-\frac{\sigma^2}{2}}\frac{\log t-\mu}{\log t-\mu-\sigma^2} - t \\
&= t\cdot\frac{\log t-\mu}{\log t-\mu-\sigma^2} - t \\
&= \frac{\sigma^2}{\log t-\mu-\sigma^2}.
\end{aligned}$$

Empirical mean excess function

In practice, the mean excess function $e(t)$ is estimated by the empirical mean excess function $\hat{e}_n(t)$ defined by

$$\hat{e}_n(t) = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i \mathbb{1}_{\{X_i>t\}}}{\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i>t\}}} - t, \tag{5.2}$$

where

$$\mathbb{1}_{\{X_i>t\}} = \begin{cases} 1, & \text{if } \; X_i > t \\ 0, & \text{otherwise} \end{cases}.$$

The empirical mean excess function $\hat{e}_n$ is often plotted at the values $t = X_{n-k,n}$ for $k = 1, \cdots, n-1$, the $(k+1)$-largest observation. Hence, (5.2) becomes:

$$\begin{aligned} \hat{e}_n(t) = \hat{e}_n(X_{n-k,n}) &= \frac{\frac{1}{n}\sum_{i=1}^{n} X_i \mathbb{1}_{\{X_i>X_{n-k,n}\}}}{\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i>X_{n-k,n}\}}} - X_{n-k,n} \\ &= \frac{\frac{1}{n}\sum_{i=0}^{n-1} X_{n-i,n} \mathbb{1}_{\{X_{n-i,n}>X_{n-k,n}\}}}{\frac{k}{n}} - X_{n-k,n} \\ &= \frac{1}{k}\sum_{i=0}^{n-1} X_{n-i,n} \mathbb{1}_{\{X_{n-i,n}>X_{n-k,n}\}} \\ &= \frac{1}{k}\sum_{i=0}^{k-1} X_{n-i,n} - X_{n-k,n}. \end{aligned}$$

The estimates of the mean excesses are given by

$$e_{k,n} = \hat{e}_n(X_{n-k,n}) = \frac{1}{k}\sum_{i=0}^{k-1} X_{n-i,n} - X_{n-k,n}, \text{ where } k = 1, \ldots, n-1. \tag{5.3}$$

Remark 5.1.12. Plots of empirical mean excess values $e_{k,n}$ as introduced in (5.3) can be constructed in two alternative ways: $e_{k,n}$ versus k, or $e_{k,n}$ versus $X_{n-k,n}$.

The appropriate underlying distribution can therefore be selected by comparing the shapes of the empirical and theoretical mean excess functions.

5.1.3 Tail index estimation

In this section, we will focus on the case where the underlying distribution is Pareto-type with index $\alpha > 0$. Historically, one of the most important statistical issues related to distributions with regularly varying tail is the estimation of the tail index. We will outline a few parameter estimation methods.

Maximum likelihood estimator for Pareto

The maximum likelihood estimation is a method of estimating population characteristics from a sample by choosing the values of the parameters that will maximize the probability of getting the particular sample actually obtained from the population.

Definition 5.1.13. *Let $X_1, \ldots, X_n$ be a random sample from the Pareto distribution with density function*

$$f(x; m, \alpha) = \frac{\alpha m^\alpha}{x^{\alpha+1}}.$$

Then the likelihood function L, has the following form:

$$L(m, \alpha) = \prod_{i=1}^{n} \frac{\alpha m^\alpha}{x_i^{\alpha+1}}, \quad \textit{for } 0 < m \le \min\{x_i\}, \alpha > 0.$$

We know that m can be no larger than $\min\{x_i\}$. Therefore, it is clear to see that the value of m that maximizes L is:

$$\hat{m} = \min\{x_i\}.$$

In order to find the maximum likelihood estimate for α, we take the logarithm of L. We do this because it is easier to differentiate $\log L$ than L itself. Logarithms are bijective functions, so the value of α that maximizes L also maximizes $\log L$. We have

$$\log L(m, \alpha) = \sum_{i=1}^{n} \log\left(\frac{\alpha m^\alpha}{{x_i}^{\alpha+1}}\right) = n\log(\alpha) + \alpha n \log(m) - (\alpha+1)\sum_{i=1}^{n} \log(x_i).$$

Hence, we obtain

$$\frac{d}{d\alpha} L(m, \alpha) = \frac{n}{\alpha} + n\log(m) - \sum_{i=1}^{n} \log(x_i).$$

Setting the derivative equal to zero, we get

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} \log\left(\frac{x_i}{\hat{m}}\right)} = \frac{n}{\sum_{i=1}^{n} \log(x_i) - n\min\{x_i\}}.$$

To confirm that we are indeed maximizing L, we need to look at the second derivative.

$$\frac{d^2}{d\alpha} L(m, \alpha) = -\frac{n}{\alpha^2} < 0.$$

The second derivative is negative, hence it is the maximum.

Hill estimator

Definition 5.1.14. *Let X be a regularly varying random variable and $Y = \log(X)$ be the log-transformed variable. The mean excess value of Y is known as the Hill estimator defined as follows*

$$H_{k,n} = \frac{1}{k}\sum_{i=1}^{k} \log X_{n-i+1,n} - \log X_{n-k,n}.$$

The Hill estimator is one of the most popular methods for estimating the tail index α.

Property 5.1.15. 1. *The Hill estimator $H_{k,n}$ converges in probability to $\hat{\alpha}_H = 1/\alpha$ when $n, k \to \infty$ and $k/n \to 0$.*

2. *Under an additional condition, it can also be shown that $\sqrt{k}(H_{k,n} - 1/\alpha) \to N(0, 1/\alpha^2)$, i.e., $H_{k,n}$ is asymptotically normally distributed.*

3. *The confidence interval for $1/\alpha$ based on the normal approximation is :*

$$\left(H_{k,n}\left(1 + \frac{z_{p/2}}{\sqrt{k}}\right), H_{k,n}\left(1 - \frac{z_{p/2}}{\sqrt{k}}\right)\right),$$

where $z_{p/2}$ is the $(p/2)$th quantile of the standard normal distribution.

Remark 5.1.16. For every choice of k, we obtain another estimator for α. We usually plot the estimates $H_{k,n}$ against k. 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.

5.1.4 Simulations

Example 5.1.17 (Exponential case)**.** We generate $n = 100$ observations from an exponential distribution with $\lambda = 3$.
Clearly the histogram (Figure 5.1) does not look symmetric, therefore we can conclude that the underlying distribution of the data is not normal. Looking at the histogram of the log-transformed data (Figure 5.2), we reach the same conclusions.
We now have to consider a heavier distribution. We construct two empirical excess plots. Based on the two plots (Figure 5.3 and Figure 5.4), we can see that the mean excess plot is constant. Hence, we can assume that the underlying distribution of the data is exponential. To confirm our claim we construct the exponential qq-plot of the data.

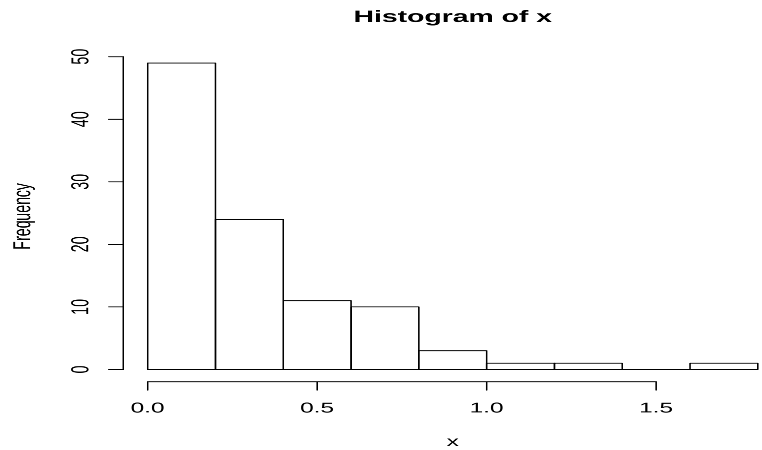

Figure 5.1: Histogram of X. (Example 5.1.17)

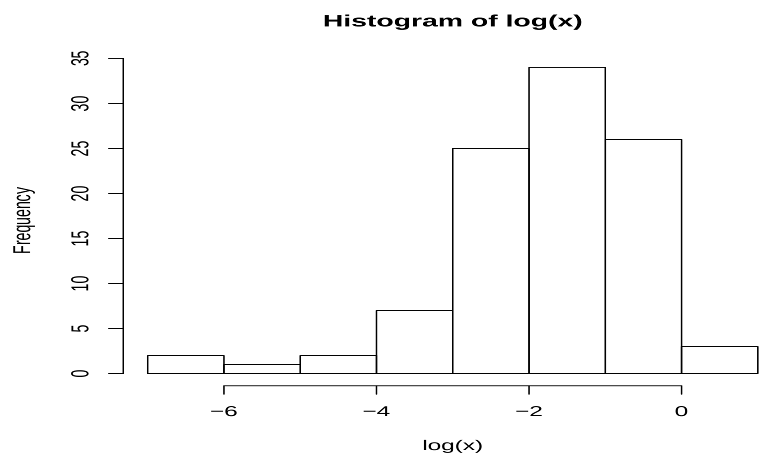

Figure 5.2: Histogram of $\log(X)$. (Example 5.1.17)

A straight line pattern is obtained (Figure 5.5), we can say that the exponential model provides a plausible fit for the given data and the slope of a fitted line can be used as an estimate of $\frac{1}{\lambda}$.

Example 5.1.18 (Pareto case)**.** We generate $n = 100$ observations from a Pareto distribution with $\alpha = 2$.
Clearly the histogram (Figure 5.6) does not look symmetric, therefore we can conclude that the underlying distribution of the data is not normal. Looking at the histogram of

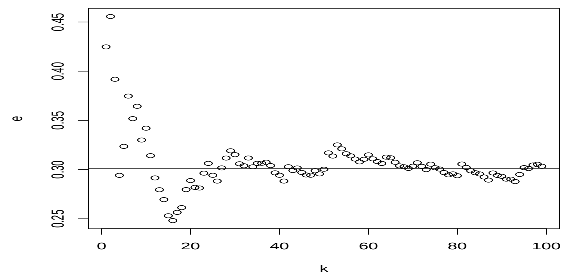

Figure 5.3: Mean excess plot(1). (Example 5.1.17)

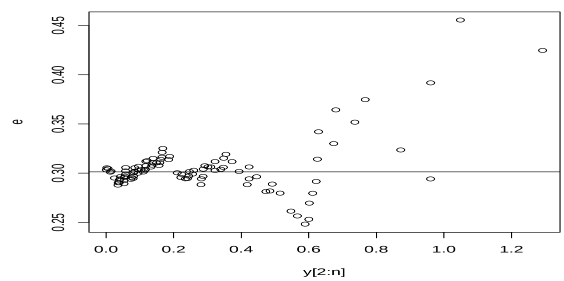

Figure 5.4: Mean excess plot(2). (Example 5.1.17)

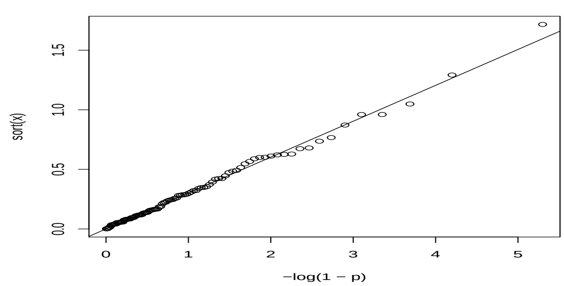

Figure 5.5: Exponential quantile-quantile plot. (Example 5.1.17)

the log-transformed data (Figure 5.7), we reach the same conclusion.
We now have to consider a heavier distribution. We construct the empirical excess plot: Based on the excess plot (Figure 5.8), we can see that the mean excess plot has a straight line pattern. Hence, we can assume that the underlying distribution of the data is Pareto.

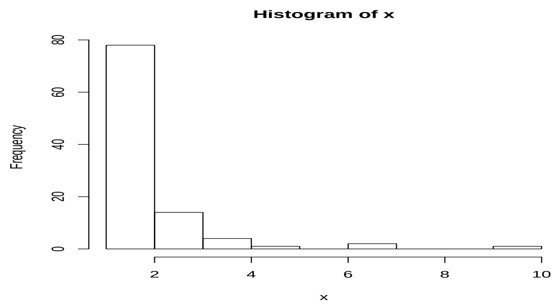

Figure 5.6: Histogram of X. (Example 5.1.18)

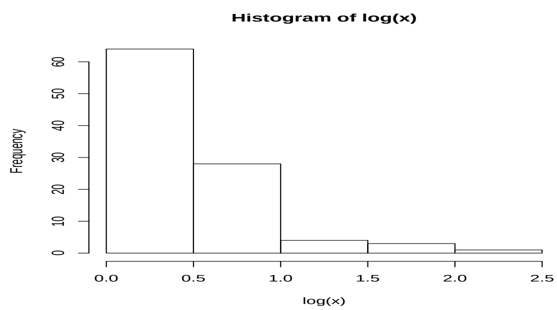

Figure 5.7: Histogram of $\log(X)$. (Example 5.1.18)

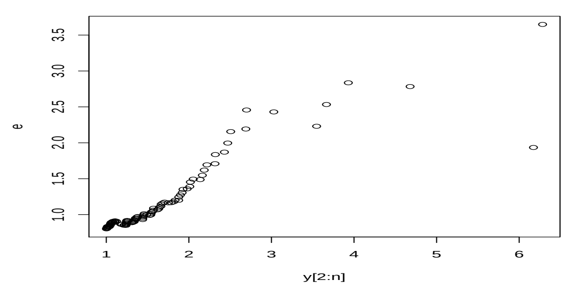

Figure 5.8: Mean excess plot of X. (Example 5.1.18)

To confirm our assumption we construct the Pareto qq-plot of the data.
A straight line pattern is obtained (Figure 5.9), we can say that the Pareto model provides a plausible fit for the given data. We will now, try to estimate the tail index α by constructing the Hill plot.

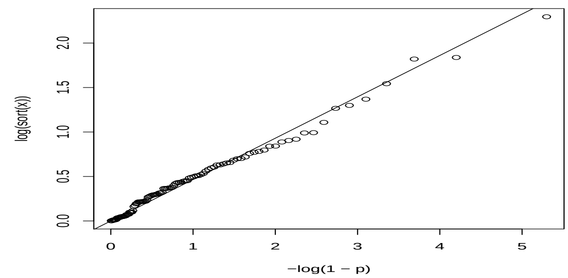

Figure 5.9: Pareto quantile-quantile plot. (Example 5.1.18)

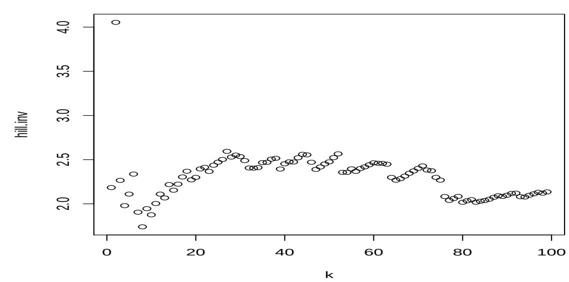

Figure 5.10: Hill plot of X. (Example 5.1.18)

The Hill plot (Figure 5.10) seems to stabilize around $\alpha = 2.5$.

Example 5.1.19 (Sum of a Normal and Pareto random variable)**.** We generate $n = 100$ observations from a Pareto distribution with $\alpha = 0.3$ and $n = 100$ observations from a normal distribution with $\mu = 3$ and $\sigma = 1$. We take the sum of the two random variables. Clearly the histogram (Figure 5.11) does not look symmetric, therefore we can conclude that the underlying distribution of the data is not normal. Looking at the histogram of the log-transformed data (Figure 5.12), we reach the same conclusion.
We now have to consider a heavier distribution. We construct the empirical excess plot: Based on the excess plot (Figure 5.13), we can see that the mean excess plot has a straight line pattern. Hence, we can assume that the underlying distribution of the data is Pareto. To confirm our assumption we will construct the Pareto qq-plot of the data.
A straight line pattern is obtained (Figure 5.14), we can say that the Pareto model provides a plausible fit for the given data. We will now, try to estimate the tail index α by constructing the Hill plot.

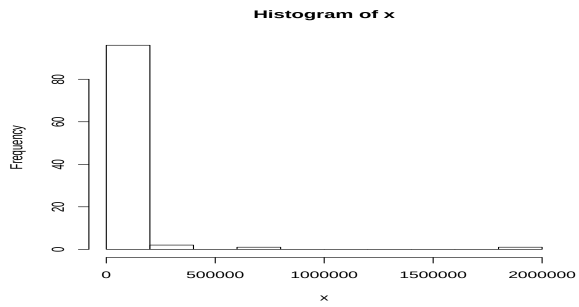

Figure 5.11: Histogram of X. (Example 5.1.19)

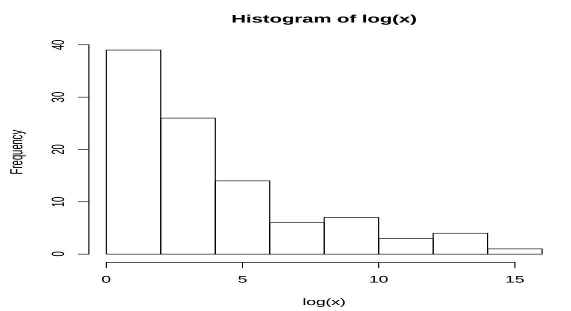

Figure 5.12: Histogram of $\log(X)$. (Example 5.1.19)

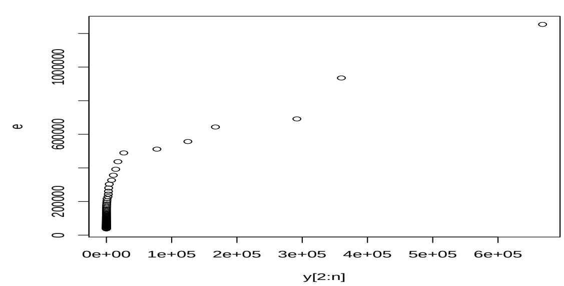

Figure 5.13: Mean excess plot of X. (Example 5.1.19)

The Hill plot (Figure 5.15) seems to stabilize around $\alpha = 0.05$.

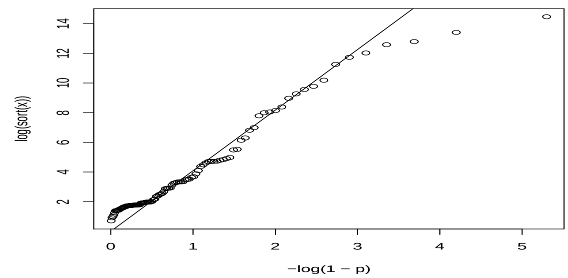

Figure 5.14: Pareto quantile-quantile plot. (Example 5.1.19)

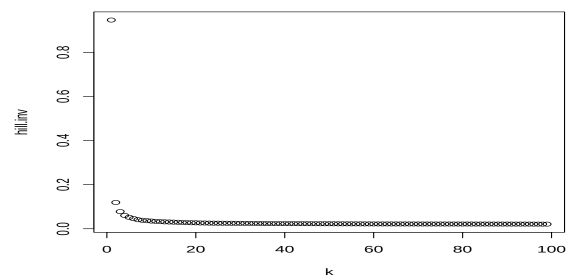

Figure 5.15: Hill plot of X. (Example 5.1.19)

5.2 Multivariate data

In this section, we will focus on the estimation of the stable tail dependence function and the Pickands dependence function. These estimators provide a fair amount of information about the dependence structure of the tails.

5.2.1 Non-parametric estimator of the stable tail dependence function

In this section, we will present two non-parametric estimators of the stable tail dependence function.

Estimator based on the order statistics

Result 5.2.1 (Introduction of [8]). *Let ℓ be the stable tail dependence function as defined in Chapter 4. To simplify the writing, we assume that $d = 2$. (The case of arbitrary $d \geq 1$ is similar.) The stable tail dependence function ℓ is estimated by the empirical stable tail dependence function $\hat{\ell}_n$ defined by*

$$\hat{\ell}_n(v_1, v_2) = \frac{1}{k} \sum_{i=1}^{n} \mathbb{1}_{\{v_1 X_{*i} \vee v_2 Y_{*i} > X_{*n-k,n}\}}, \text{ for all } v_1, v_2 \in \mathbb{R}_+^2, \tag{5.4}$$

where $X_{i,n}$ denotes the i-th order statistic.

Heuristic. We know that

$$\ell(v_1, v_2) = V_* \left(\frac{1}{v_1}, \frac{1}{v_2} \right) = \nu_*([\mathbf{0}, (1/v_1, 1/v_2)]^c), \text{ for all } \mathbf{v} \in \mathbb{R}_+^2,$$

and

$$\lim_{x \to +\infty} \frac{P(X_* > x\frac{1}{v_1} \vee Y_* > x\frac{1}{v_2})}{P(X_* > x)} \to \nu_*([\mathbf{0}, (1/v_1, 1/v_2)]^c),$$

where, $X_* = T_G(X)$ and $Y_* = T_G(Y)$. Taking x "big", we estimate the the probability on the left hand side, by

$$\hat{\ell}_n(v_1, v_2) = \frac{\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{v_1 X_{*i} > x \cup v_2 Y_{*i} > x\}}}{\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_{*i} > x\}}}, \tag{5.5}$$

where

$$\mathbb{1}_{\{X_{i*} > x\}} = \begin{cases} 1, & \text{if } X_{i*} > x \\ 0, & \text{otherwise} \end{cases}.$$

The empirical stable tail dependence function $\hat{\ell}_n$ is plotted at the values $x = X_{*n,n-k}$ for $k = 1, \cdots, n-1$, the $(k+1)$-largest observation. Hence, (5.5) becomes:

$$\hat{\ell}_n(v_1, v_2) = \frac{1}{k} \sum_{i=1}^{n} \mathbb{1}_{\{v_1 X_{*i} \vee v_2 Y_{*i} > X_{*n-k,n}\}}.$$

□

Rank-based estimator

Result 5.2.2. *(See Section 3 in [5].) Let R_i^X and R_i^Y be the rank of X_i among $X_1, \ldots, X_n$ and the rank of Y_i among $Y_1, \ldots, Y_n$ respectively, where $i = 1, \ldots, n$. Let $\tilde{\ell}_n$ be a non-parametric estimator of the stable tail dependence function ℓ defined by*

$$\tilde{\ell}_n(x, y) = \frac{1}{k} \sum_{i=1}^{n} \mathbb{1}_{\{R_i^X > n + \frac{1}{2} - kx \text{ or } R_i^Y > n + \frac{1}{2} - ky\}}, \text{ for all } x, y \in \mathbb{R}_+^2. \tag{5.6}$$

5.2.2 Non-parametric estimator of the Pickands dependence function

In this section, we will present two non-parametric estimators of the Pickands dependence function.

Estimator based on the order statistics

Result 5.2.3. *Let ℓ be the stable tail dependence function. The Pickands dependence function A is defined by*

$$A(t) = \ell(1-t, t), \quad \textit{for all } t \in [0,1].$$

Taking $v_1 = 1-t$ and $v_2 = t$, it follows from (5.4) that the Pickands dependence function A by can be estimated by the empirical function $\hat{A}_n$ defined by

$$\hat{A}_n(t) = \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}_{\{(1-t)X_{*i} \vee tY_{*i} > X_{n,n-k}\}}, \ \textit{for all } t \in [0,1].$$

Rank-based estimator

Result 5.2.4. *Let R_i^X and R_i^Y be the rank of X_i among $X_1, \ldots, X_n$ and the rank of Y_i among $Y_1, \ldots, Y_n$ respectively, where $i = 1, \ldots, n$. It follows from (5.6) that*

$$\tilde{A}_n(t) = \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}_{\{R_i^X > n+\frac{1}{2}-k(1-t) \cup R_i^Y > n+\frac{1}{2}-kt\}}, \ \textit{for all } t \in [0,1].$$

where $\tilde{A}_n$ is a non-parametric estimator of the Pickands dependence function A.

5.2.3 Simulations

Estimation of the stable tail dependence function

Example 5.2.5 (Independence)**.** We generate independently two random variables X and Y of size $n = 1000$ from a Pareto distribution with index $\alpha = 4$. We want to estimate the stable tail dependence function for four different values of k, $k = 50, 100, 200, 300$. We choose $v_1 = t[i]$ and $v_2 = t[j]$ such that $t[i] = \frac{i}{M}, t[j] = \frac{j}{M}$, where $M = 100$ and $i, j = 1, \ldots, n$. We will first use the estimator based on the order statistics, we obtain Figure 5.16.

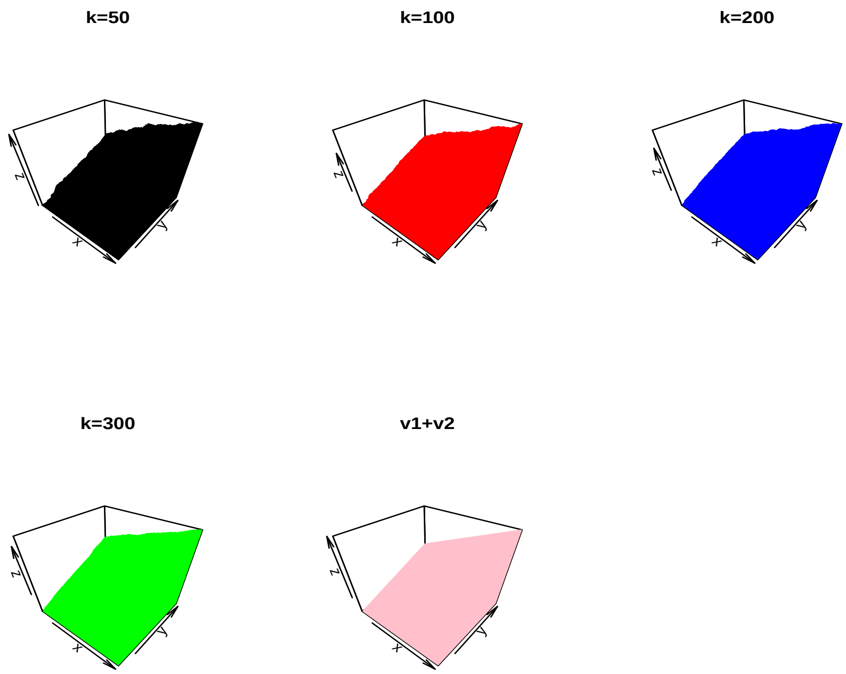

Figure 5.16: Stable tail dependence function using the order statistics (Example 5.2.5).

We know from Chapter 4 that theoretically if X and Y are independent then $\ell(v_1, v_2) = v_1 + v_2$. The fith figure (pink) represents $v_1 + v_2$, clearly, the plots for all the different values of k match the theoretical plot.

We now want to estimate the stable tail dependence function for different values of k using this time the rank-based estimator. We obtain Figure 5.17.

Again, the plots for the the different values of k match the theoretical plot. Therefore, X and Y are independent.

Example 5.2.6 (Complete dependence)**.** We generate two random variables X and Y of size $n = 1000$ from a Pareto distribution with index $\alpha = 4$ such that $X = Y$. We want to estimate the stable tail dependence function for four different values of k, $k = 50, 100, 200, 300$. We choose $v_1 = t[i]$ and $v_2 = t[j]$ such that $t[i] = \frac{i}{M}, t[j] = \frac{j}{M}$, where $M = 100$ and $i, j = 1, \ldots, n$. We will first use the estimator based on the order statistics, we obtain Figure 5.18.

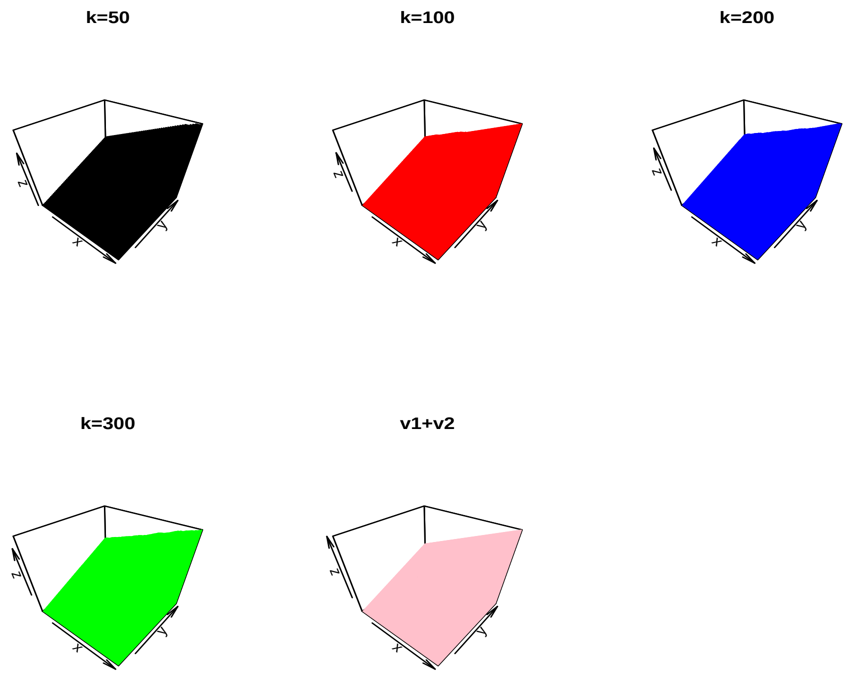

Figure 5.17: Stable tail dependence function using the ranks (Example 5.2.5).

We know from Chapter 4 that theoretically if $X = Y$ then $\ell(v_1, v_2) = v_1 \vee v_2$. The fith figure (pink) represents $v_1 \vee v_2$, clearly, the plots for all the different values of k match the theoretical plot.

We now want to estimate the stable tail dependence function for different values of k using this time the rank-based estimator. We obtain Figure 5.19.

Again, the plots for the the different values of k match the theoretical plot.

Estimation of the Pickands dependence function

Example 5.2.7 (Independence)**.** We generate independently two random variables X and Y of size $n = 1000$ from a Pareto distribution with index $\alpha = 4$. We want to estimate the Pickands dependence function for four different values of k, $k = 50, 100, 200, 300$. We choose $t = t[i]$ such that $t[i] = \frac{i}{M}$, where $M = 100$ and $i = 1, \ldots, n$. We will first use the estimator based on the order-statistic, we obtain Figure 5.20.

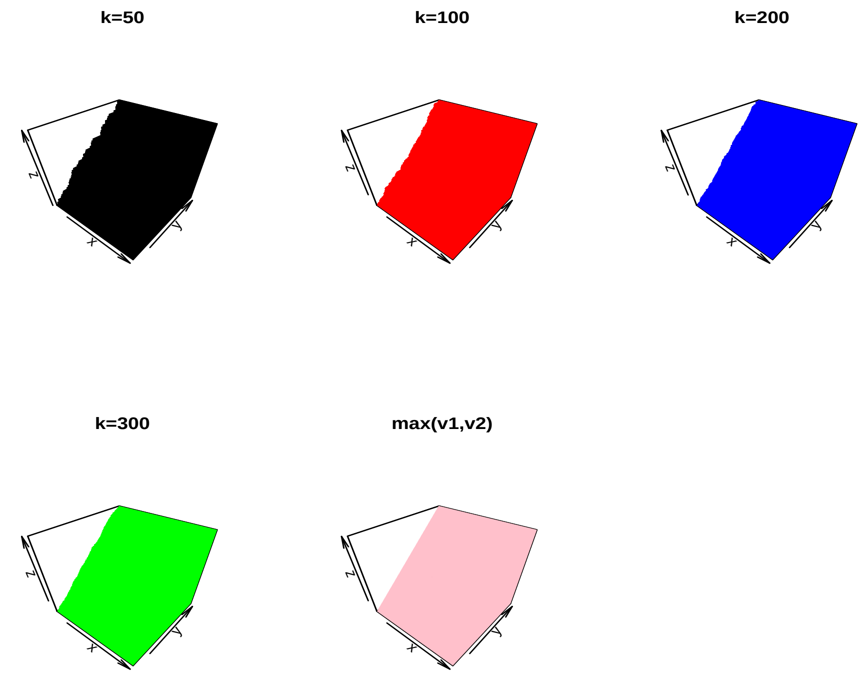

Figure 5.18: Stable tail dependence function using the order statistics (Example 5.2.6).

We know from Chapter 4 that theoretically if X and Y are independent then $A(t) = 1$. We can clearly see that the plots for all the different values of k stabilize around 1.

We now want to estimate the stable tail dependence function for different values of k using this time the rank-based estimator. We obtain Figure 5.21.

Again, the plots for the the different values of k stabilize around 1.

Example 5.2.8 (Complete dependence)**.** We generate two random variables X and Y of size $n = 1000$ from a Pareto distribution with index $\alpha = 4$ such that $X = Y$. We want to estimate the Pickands dependence function for four different values of k, $k = 50, 100, 200, 300$. We choose t such that $t[i] = \frac{i}{M}$, where $M = 100$. We will first use the estimator based on the order statistics, we obtain Figure 5.22.

We know from Chapter 4 that theoretically if $X = Y$ then $A(t) = (1-t) \vee t$. We can see that the plots for all the different values of k have the same shape as the theoretical plot of $(1-t) \vee t$.

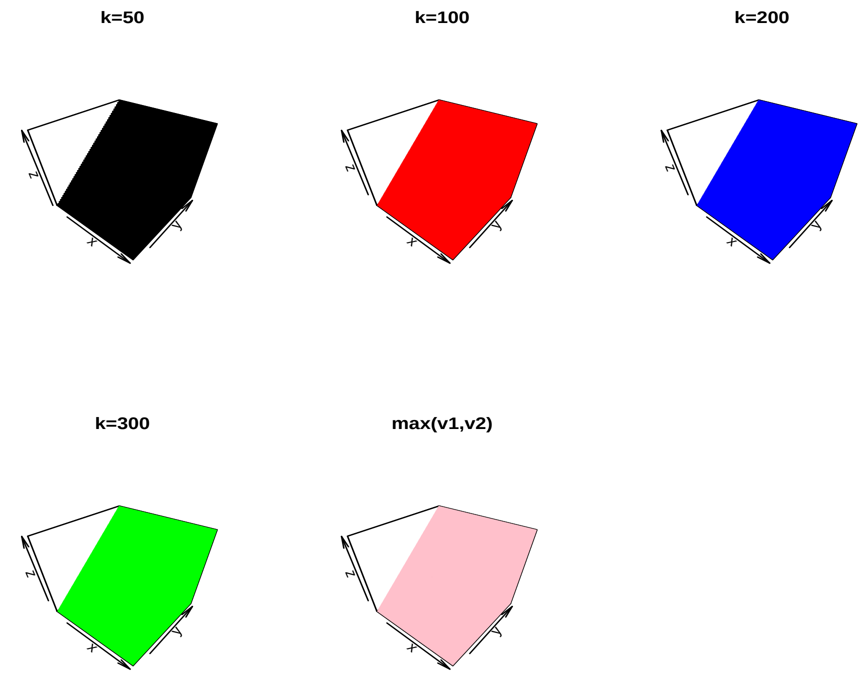

Figure 5.19: Stable tail dependence function using the ranks (Example 5.2.6).

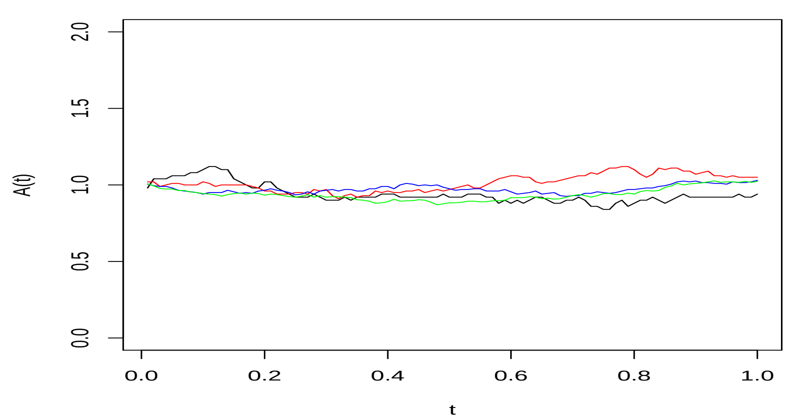

Figure 5.20: Pickands dependence function based on the order statistics (Example 5.2.7).

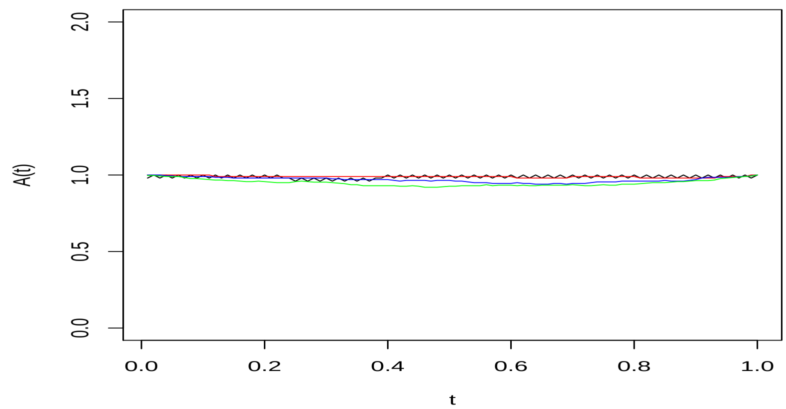

Figure 5.21: Pickands dependence function based on the ranks (Example 5.2.7).

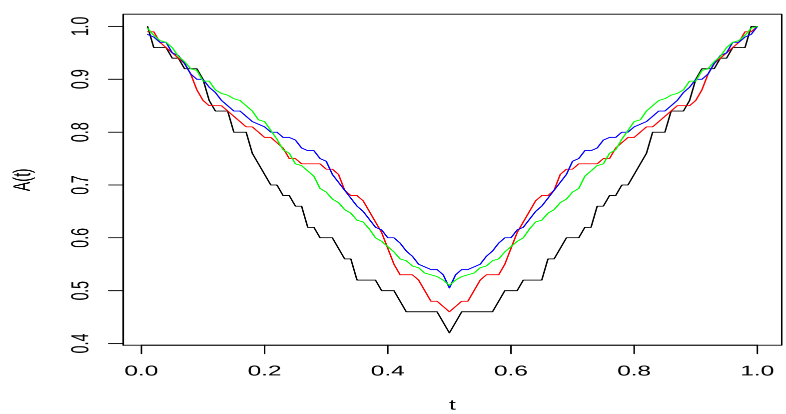

Figure 5.22: Pickands dependence function based on order statistics (Example 5.2.8).

We now want to estimate the Pickands dependence function for different values of k using this time the rank-based estimator. We obtain Figure 5.23.

Again, the plots for the the different values of k have the same shape as the theoretical plot of $(1-t) \vee t$.

Remark 5.2.9. Based on the simulations, the Pickand's dependence function seems to stabilize better when using the rank-based estimator, hence we can conclude that the estimator based on ranks behaves better than the estimator based on the order-statisitcs.

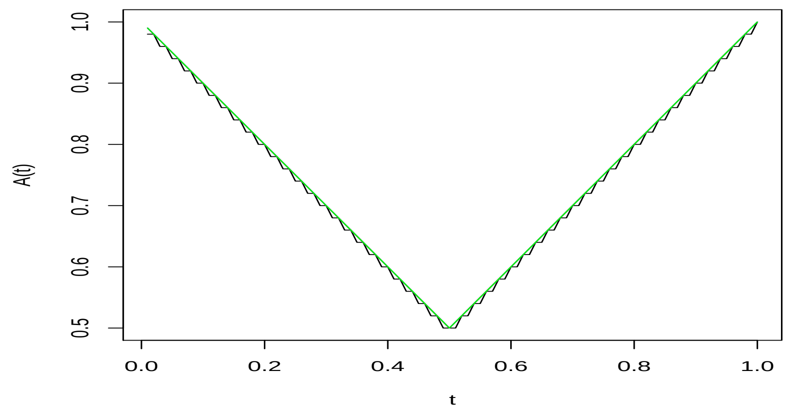

Figure 5.23: Pickands dependence function based on the ranks (Example 5.2.8).

5.3 Data analysis

In this section, we will conduct a data analysis using financial data.

Example 5.3.1. Our data analysis will be based on the NASDAQ stock prices and SP500 stock prices from 1972 to 2015. Before starting our analysis we have to compute the absolute value of the log-returns. We plot the Pickands dependence function using the rank method.

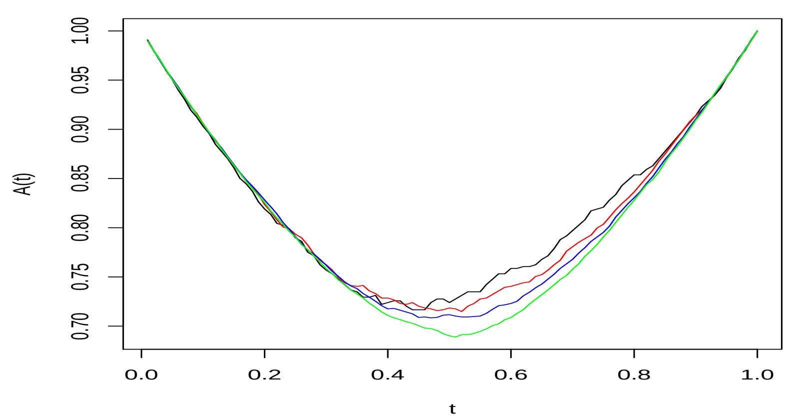

Figure 5.24: Pickands dependence function (Example 5.3.1).

We can see that the plot (Figure 5.24) has an almost "V" shape. This suggests that log-returns of the stock prices are dependent in the tails.

Example 5.3.2. For this example, our data analysis will be based on the MICROSOFT stock prices and FACEBOOK stock prices from 2012 to 2015 . Again we compute the absolute value of the log-returns before proceeding to the analysis and using the rank method we plot the Pickands dependence function.

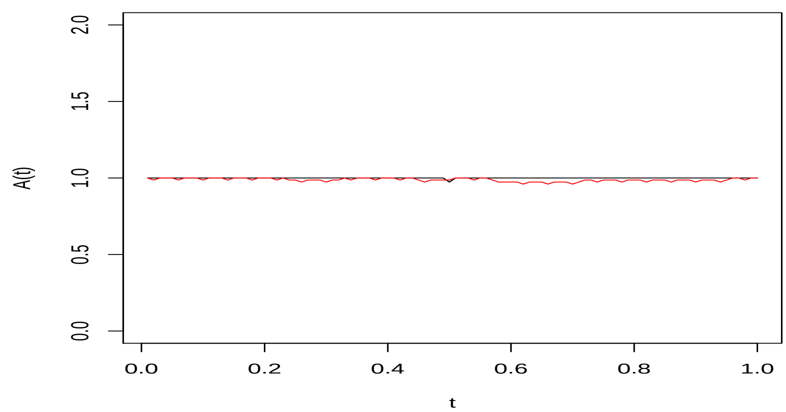

Figure 5.25: Pickands dependence function (Example 5.3.2).

Based on the plot (Figure 5.25) we can see that the Pickands dependence function stabilizes around 1 therefore we can say that the log-returns of the stock prices are independent in the tails.

Since the log-returns of the stock prices are independent in the tails, we will now study separately each stock prices to determine the respective underlying distributions.

Example 5.3.3 (MIROSOFT)**.** We first plot the histogram of the absolute value of the log-returns.

Clearly the histogram (Figure 5.26) does not look symmetric, hence we can say that the underlying distribution of the data is not normal. We now consider a heavier distribution. For that we will plot the mean excess plot.

The mean excess plot (Figure 5.27) seems to stabilize around 0.02 and looks like the mean excess plot of an exponential distribution. To confirm our assumption, we will construct an exponential qq-plot of the data.

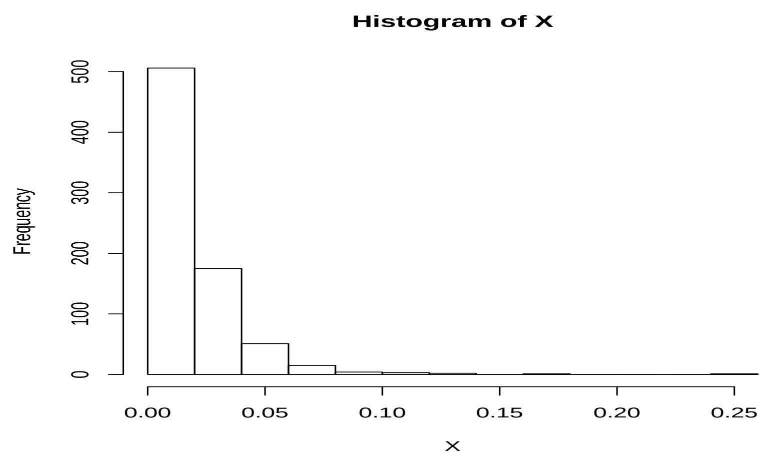

Figure 5.26: Histogram of the absolute value of the log-returns (Example 5.3.3).

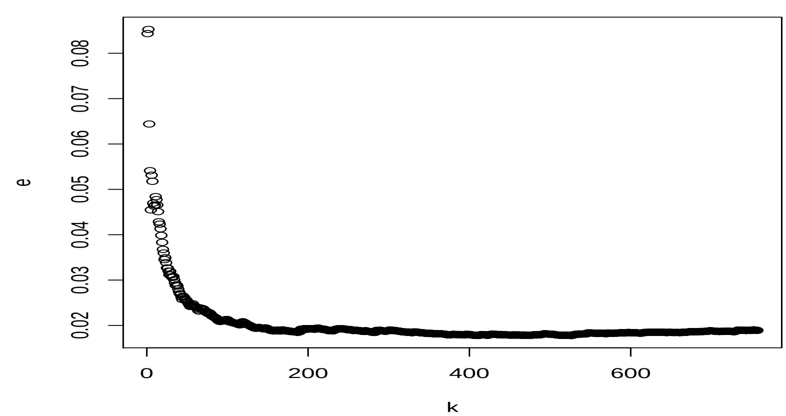

Figure 5.27: Mean excess plot of the absolute value of the log-returns (Example 5.3.3).

A straight line pattern is obtained (Figure 5.28). We conclude by saying that the exponential model provides a plausible fit for the given data and the slope of the fitted line can be used as an estimate of the inverse of the rate parameter. Using the maximum-likelihood estimator we obtain $\frac{1}{\hat{\lambda}} = 0.01888137$.

Example 5.3.4 (FACEBOOK)**.** We first plot the histogram of the absolute value of the log-returns.

Clearly the histogram (Figure 5.31) does not look symmetric, hence we can say that

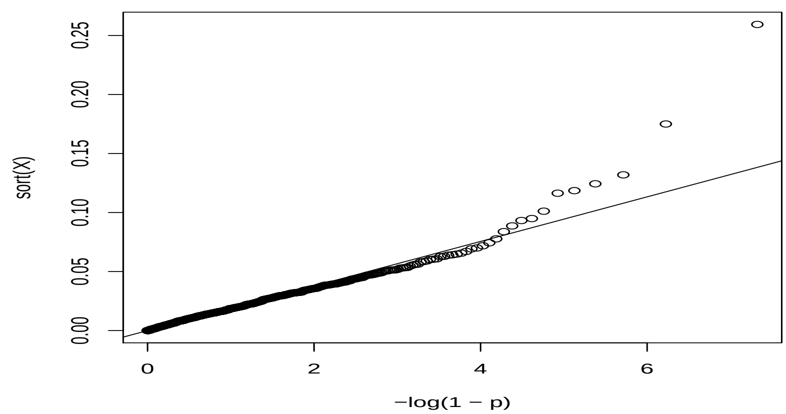

Figure 5.28: Exponential quantile-quantile plot of the absolute value of the log-returns (Example 5.3.3).

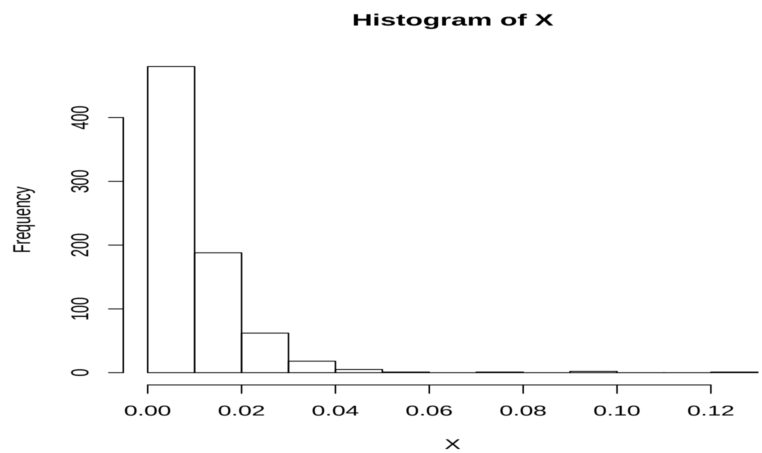

Figure 5.29: Histogram of the absolute value of the log-returns (Example 5.3.4).

the underlying distribution of the data is not normal. We now consider a heavier distribution. For that we will plot the mean excess plot.

The mean excess plot (Figure 5.30) seems to stabilize around 0.01 and looks like the mean excess plot of an exponential distribution. To confirm our assumption, we will construct an exponential qq-plot of the data.

A straight line pattern is obtained (Figure 5.31). We conclude by saying that the

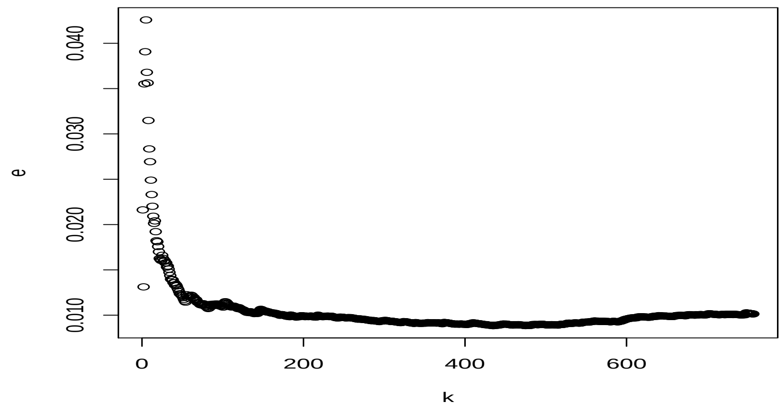

Figure 5.30: Mean excess plot of the absolute value of the log-returns (Example 5.3.4).

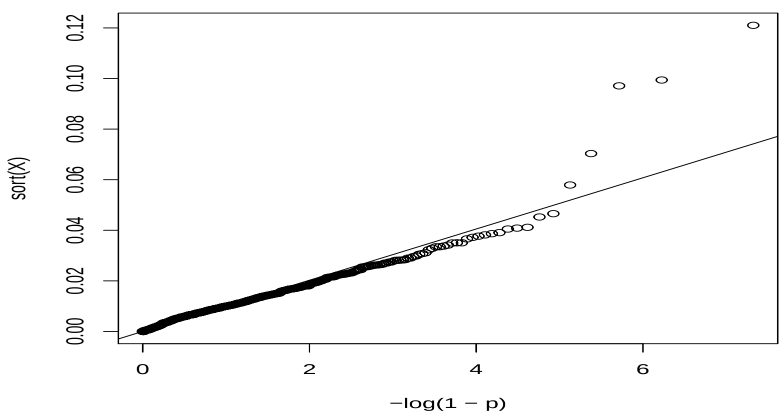

Figure 5.31: Exponential quantile-quantile plot of the absolute value of the log-returns (Example 5.3.4).

exponential model provides a plausible fit for the given data and the slope of the fitted line can be used as an estimate of the inverse of the rate parameter. Using the maximum-likelihood estimator we obtain $\frac{1}{\hat{\lambda}} = 0.01888137$.

Appendix A

Measure-theoretic probability results

A.1 Probability theory

In this section, we collect some results from measure-theoretic probability which are used in the thesis. The main reference for these results is [3].

Definition A.1.1. *Let Ω be an arbitrary set. We say that $\mathcal{F}$ is a **σ-algebra** on Ω if it satisfies the following properties:*

(i) $\emptyset \in \mathcal{F}$.

(ii) *if* $A \in \mathcal{F}$ *then* $A^c \in \mathcal{F}$.

(iii) *if* $(A_n)_n \in \mathcal{F}$, *for all* $n \geq 1$, *then* $\bigcup_{n\geq 1} A_n \in \mathcal{F}$.

*In this case, we say that $(\Omega, \mathcal{F})$ is a **measurable space**.*

Definition A.1.2. *Let $(\Omega, \mathcal{F})$ be a measurable space. A function $\mu : \mathcal{F} \to [0, +\infty]$ is called a **measure** on $(\Omega, \mathcal{F})$ if it satisfies the following properties:*

(i) $\mu(\emptyset) = 0$;

(ii) $\mu(\cup_n A_n) = \sum_n \mu(A_n)$ *for any disjoint sets* $(A_n)_n \in \mathcal{F}$.

*In this case, we say that $(\Omega, \mathcal{F}, \mu)$ is a **measure space**.*
*A **probability measure** is a measure P on $(\Omega, \mathcal{F})$ such that $P(\Omega) = 1$. In this case,we say that the space $(\Omega, \mathcal{F}, P)$ is a **probability space**.*

An important example is the case $\Omega = \mathbb{R}$ and $\mathcal{F} = \mathcal{B}$ where $\mathcal{B}$ is the class of the Borel sets in $\mathbb{R}$. (Recall that any interval in $\mathbb{R}$ is a Borel set.)

Definition A.1.3. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. A function $X : \Omega \to \mathbb{R}$ is called a* ***random variable*** *if $\{X \in A\} = \{\omega \in \Omega; X(\omega) \in A\} \in \mathcal{F}$ for any Borel set $A \in \mathcal{B}$. In this case we define the probability measure μ on $(\mathbb{R}, \mathcal{B})$ by $\mu(A) = P(X \in A)$ for any $A \in \mathcal{B}$. We say that μ is* ***the law (or distribution) of X****.*

Definition A.1.4. *Let $X, (X_n)_{n\geq 1}$ be random variables defined on a probability space $(\Omega, \mathcal{F}, P)$. Let F and F_n be the distribution functions of X and X_n, respectively, $n \geq 1$, i.e $F(x) = P(X \leq x), x \in \mathbb{R}$ and $F_n(x) = P(X_n \leq x), x \in R$. We say that $(X_n)_{n\geq 1}$* ***converges in distribution*** *to X if*

$$\lim_{x \to +\infty} F_n(x) = F(x)$$

for any $x \in \mathbb{R}$ which is a continuity point of F. In this case we write $X_n \xrightarrow{d} X$.

Theorem A.1.5 (Dominated convergence theorem, Theorem 16.4 of [3])**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measurable space and $X, \{X_n\}_{n\geq 1}$ be $\mathcal{F}$-measurable functions defined on Ω. Assume that $X_n(\omega) \to X(\omega)$ for any $\omega \in \Omega$. If there exists an $\mathbb{R}$-valued $\mathcal{F}$-measurable function $Y : \Omega \to \mathbb{R}_+$ such that*

$$\int_\Omega |Y| d\mu < \infty \quad \textit{and} \quad |X_n(\omega)| \leq Y(\omega) \quad \forall \omega \in \Omega, \quad \forall n \geq 1$$

then

$$\lim_{n \to +\infty} \int_\Omega X_n d\mu = \int_\Omega X d\mu.$$

Definition A.1.6. *A sequence of measures $\{\mu_n\}_{n\in\mathbb{N}}$ defined on a measurable space $(\Omega, \mathcal{F})$* ***converges weakly*** *to a measure μ defined on $(\Omega, \mathcal{F})$ if for any bounded continuous function $f : \Omega \to \mathbb{R}$*

$$\int_\Omega f(x)\mu_n(dx) \to \int_\Omega f(x)\mu(dx), \;\; \textit{as } n \to \infty.$$

In this case, we write $\mu_n \xrightarrow{w} \mu$. We denote $\mu(f) = \int_\Omega f(x)\mu(dx)$.

Theorem A.1.7 ((Theorem 25.8 of [3]))**.** *Let $X, (X_n)_{n\geq 1}$ be random variables defined on a probability space $(\Omega, \mathcal{F}, P)$. We denote by μ and μ_n the law of X and X_n respectively. Then $X_n \xrightarrow{d} X$ if and only if $\mu_n \xrightarrow{w} \mu$.*

Theorem A.1.8 (Helly-Bray Theorem, Theorem 25.12 of [3])**.** *Let Y_n and Y be random variables with respective distribution functions F_n and F. Then $Y_n \xrightarrow{d} Y$ if and only if for any bounded continuous function $z : \mathbb{R} \to \mathbb{R}$*

$$E[z(Y_n)] \to E[z(Y)].$$

Theorem A.1.9 (Change of Variable Theorem, Theorem 6.1 of [[11])**.** *]. Let $(\Omega, \mathcal{F}, P)$ be a probability space and X be a random variable on Ω with distribution μ. Then for any Borel-measurable function $f : \mathbb{R} \to \mathbb{R}$, we have*

$$\int_\Omega f(X(\omega))P(d\omega) = \int_{-\infty}^{\infty} f(t)\mu(dt),$$

i.e $E[f(X)] = \mu(f)$, provided that either side is well-defined. In other words, the expected value of the random variable $f(X)$ with respect to the probability measure P on Ω is equal to the expected value of the function f with respect to the measure μ on $\mathbb{R}$.

A.2 Transformations of random variables

In this section, we include some useful results about transformation of random variables.

Lemma A.2.1. *If X is a random variable with distribution function F then $F(X)$ has a uniform distribution on $[0, 1]$.*

Proof. The proof of this result is standard and we omit. □

Lemma A.2.2. *Let X be a random variable such that $X > 1$ and F be its distribution function. For any $x > 1$, we define*

$$U(x) = F^{-1}\left(1 - \frac{1}{x}\right).$$

Then $U^{-1}(X)$ has Pareto distribution with parameter $\alpha = 1$.

Proof. For any $x > 1$, we have:

$$\begin{aligned} P(U^{-1}(X) > x) &= P(X > U(x)) \\ &= P(X > F^{-1}(1 - 1/x)) \\ &= P(F(X) > 1 - 1/x) \\ &= 1 - (1 - 1/x) = x^{-1}, \end{aligned}$$

where for the second last equation, we used the fact that $F(X)$ has a uniform distribution on $[0, 1]$. Hence, $U^{-1}(X)$ has a Pareto distribution with parameter $\alpha = 1$. □

Lemma A.2.3. *Let Y be a random variable with distribution function G. Then*

$$Y_* = -\frac{1}{\log G(Y)}$$

has a standard Fréchet distribution with $d = 1$.

Proof. We have:

$$\begin{aligned}
P\left(-\frac{1}{\log G(Y)} \leq z\right) &= P\left(\frac{1}{\log G(Y)} \geq -z\right) \\
&= P\left(\log G(Y) \leq -\frac{1}{z}\right) \\
&= P(G(Y) \leq e^{-1/z}) \\
&= P(Y \leq G^{-1}(e^{-1/z})) \\
&= G(G^{-1}(e^{-1/z})) = e^{-1/z} = e^{-z^{-1}}.
\end{aligned}$$

Therefore, Y_* has a Frechét distribution with parameter $\alpha = 1$. □

A.3 Notation

LCCB space = locally compact space with a countable basis.

f^{-1} = the left-continuous inverse of a monotone function f defined by

$$f^{-1}(x) = \inf\{y : f(y) \geq x\}.$$

$\mathbb{E}_\infty = [0, \infty]^d \setminus \{\mathbf{0}\}$.

$M_+(\mathbb{E}_\infty)$ = the space of non-negative Radon measures on $\mathbb{E}_\infty$.

ν_α = a measure on $(0, \infty]$ given by $\nu(x, \infty] = x^{-\alpha},\ \alpha > 0,\ x > 0$.

$\mathcal{K}(S)$ = the compact subsets of a metric space S.

$a(t)$ = the quantile function of a distribution function $F(x)$, defined by

$$a(t) = F^{-1}(1 - \frac{1}{t}).$$

$\aleph_+$ = the unit sphere on $\mathbb{R}^d$ with respect to a chosen norm $\|\cdot\|$.

Appendix B

R codes

B.1 Simulations for the univariate data

In this section, we give the R codes used in the first section of Chapter 5.

Analysis of the simulated exponential distribution

```
n=100; lambda=3
x=rexp(n,lambda)
hist(x)
qqnorm(x)
hist(log(x))
qqnorm(log(x))
y=rev(sort(x))
e=cumsum(y[1:(n-1)])/(1:(n-1))-y[2:n]
plot(k,e)
plot(y[2:n],e)
p=ppoints(n)
plot(-log(1-p),sort(x))
mle=n/sum(x)
```

Analysis of the simulated Pareto distribution

```
n=100; a=2
u=runif(100)
```

```
x=(1-u)^(-1/a)
hist(x)
qqnorm(x)
hist(log(x))
qqnorm(log(x))
y=rev(sort(x))
e=cumsum(y[1:(n-1)])/(1:(n-1))-y[2:n]
plot(y[2:n],e)
minimum=min(x)
alpha.est=length(x)/(sum(log(x)-log(minimum)))
p=ppoints(n)
plot(-log(1-p),log(sort(x)))
abline(0,1/alpha.est)
y=rev(sort(x))
k=1:(n-1)
l=log(y)
hill=cumsum(l[1:n-1])/(1:(n-1))-l[2:n]
hill.inv=1/hill
plot(k,hill.inv)
```

B.2 Simulations for the multivariate data

In this section, we give the R codes used in the second section of Chapter 5.

Estimation of the stable tail dependence function using the order statistics

```
n=1000; M=100;a1=4; a2=4;
p1=0.05;p2=0.1;p3=0.2;p4=0.3;
k1=floor(p1*n); k2=floor(p2*n); k3=floor(p3*n); k4=floor(p4*n);
t=c(1:M)/M
L1=matrix(0,M,M); L2=matrix(0,M,M); L3=matrix(0,M,M); L4=matrix(0,M,M);
V=rep(NA,n)
U1=runif(n); U2=runif(n);
U1=runif(n); U2=U1;
X1=(1-U1)^{-1/a1}; X2=(1-U2)^{-1/a2}
```

```
Y1=X1^{a1}; Y2=X2^{a2}
Z1=rev(sort(Y1)); Z2=rev(sort(Y2))
for(i in (1:M))
        {
for(j in (1:M)){
               for(l in (1:n)){
                       V[l]=max((t[i])*Y1[l],t[j]*Y2[l])
                       }
                L1[i,j]=(k1)^{-1}*sum(V>Z1[k1+1]);
                L2[i,j]=(k2)^{-1}*sum(V>Z1[k2+1]);
                L3[i,j]=(k3)^{-1}*sum(V>Z1[k3+1]);
                L4[i,j]=(k4)^{-1}*sum(V>Z1[k4+1]);
               }
        }
L0=matrix(0,M,M);
for(i in (1:M))
        {
for(j in (1:M)){
               L0[i,j]=t[i]+t[j];
               }
        }
L0=matrix(0,M,M);
for(i in (1:M))
        {
for(j in (1:M)){
               L0[i,j]=max(t[i],t[j])
               }
        }
par(mfrow=c(2,3))
hist3D(t,t,L1,type="l",main="k=50",col="black");
hist3D(t,t,L2,type="l",main="k=100",col="red");
hist3D(t,t,L3,type="l",main="k=200",col="blue");
hist3D(t,t,L4,type="l",main="k=300",col="green");
hist3D(t,t,L0,type="l",main="v1+v2",col="pink")
hist3D(t,t,L0,type="l",main="max(v1,v2)",col=pink")
```

Rank-based estimation of the stable tail dependence function

```
n=1000; M=100;a1=4; a2=4;
p1=0.05;p2=0.1;p3=0.2;p4=0.3;
k1=floor(p1*n); k2=floor(p2*n); k3=floor(p3*n); k4=floor(p4*n);
t=c(1:M)/M
L1=matrix(NA,M,M); L2=matrix(NA,M,M); L3=matrix(NA,M,M); L4=matrix(NA,M,M);
V1=array(NA,dim=c(n,M,M));V2=array(NA,dim=c(n,M,M));V3=array(NA,dim=c(n,M,M));
V4=array(NA,dim=c(n,M,M));
U1=runif(n); U2=runif(n);
U1=runif(n); U2=U1;
X1=(1-U1)^{-1/a1}; X2=(1-U2)^{-1/a2}
R1=rank(X1);R2=rank(X2)
for(i in (1:M))
        {
for(j in (1:M)){
               for(l in (1:n)){
            V1[l,i,j]=R1[l]>n+0.5-k1*t[i]|R2[l]>n+0.5-k1*t[j];
            V2[l,i,j]=R1[l]>n+0.5-k2*t[i]|R2[l]>n+0.5-k2*t[j];
            V3[l,i,j]=R1[l]>n+0.5-k3*t[i]|R2[l]>n+0.5-k3*t[j];
            V4[l,i,j]=R1[l]>n+0.5-k4*t[i]|R2[l]>n+0.5-k4*t[j]}
                       }
}
for(i in (1:M))
        {
for(j in (1:M)){
                L1[i,j]=k1^{-1}*sum(V1[,i,j]);
                L2[i,j]=k2^{-1}*sum(V2[,i,j]);
                L3[i,j]=k3^{-1}*sum(V3[,i,j]);
                L4[i,j]=k4^{-1}*sum(V4[,i,j]);
                    }
}
L0=matrix(NA,M,M);
for(i in (1:M))
```

```
        {
for(j in (1:M)){
               L0[i,j]=t[i]+t[j];
               }
        }
L0=matrix(0,M,M);
for(i in (1:M))
        {
for(j in (1:M)){
               L0[i,j]=max(t[i],t[j])
               }
        }
par(mfrow=c(2,3))
hist3D(t,t,L1,type="l",main="k=50",col="black");
hist3D(t,t,L2,type="l",main="k=100",col="red");
hist3D(t,t,L3,type="l",main="k=200",col="blue");
hist3D(t,t,L4,type="l",main="k=300",col="green");
hist3D(t,t,L0,type="l",main="v1+v2",col="pink")
hist3D(t,t,L0,type="l",main="max(v1,v2)",col="pink")
```

Estimation of the Pickands dependence function using the order statistics

```
n=1000; M=100;a1=4; a2=4;
p1=0.05;p2=0.1;p3=0.2;p4=0.3;
k1=floor(p1*n); k2=floor(p2*n); k3=floor(p3*n); k4=floor(p4*n);
t=c(1:M)/M
A1=rep(NA,M); A2=rep(NA,M); A3=rep(NA,M); A4=rep(NA,M);
V=rep(NA,n)
U1=runif(n); U2=runif(n);
U1=runif(n); U2=U1;
X1=(1-U1)^{-1/a1}; X2=(1-U2)^{-1/a2}
Y1=X1^{a1}; Y2=X2^{a2}
Z1=rev(sort(Y1)); Z2=rev(sort(Y2))
for(j in (1:M)){
               for(l in (1:n)){
                       V[l]=max((1-t[j])*Y1[l],t[j]*Y2[l])
```

```
                         }
                A1[j]=(k1)^{-1}*sum(V>Z1[k1+1]);
                A2[j]=(k2)^{-1}*sum(V>Z1[k2+1]);
                A3[j]=(k3)^{-1}*sum(V>Z1[k3+1]);
                A4[j]=(k4)^{-1}*sum(V>Z1[k4+1]);
               }
maximum=max(A1,A2,A3,A4);
minimum=min(A1,A2,A3,A4);
plot(t,A1,type="l",xlab="t",ylab="A(t)",ylim=c(minimum,maximum),col="black");
plot(t,A1,type="l",xlab="t",ylab="A(t)",ylim=c(0,2),col="black");
lines(t,A2,type="l",col="red");
lines(t,A3,type="l",col="blue");
lines(t,A4,type="l",col="green");
```

Rank-based estimation of the Pickands dependence function

```
n=1000; M=100;a1=4; a2=4;
p1=0.05;p2=0.1;p3=0.2;p4=0.3;
k1=floor(p1*n); k2=floor(p2*n); k3=floor(p3*n); k4=floor(p4*n);
t=c(1:M)/M
A1=rep(NA,M); A2=rep(NA,M); A3=rep(NA,M); A4=rep(NA,M);
V1=matrix(NA,n,M);V2=matrix(NA,n,M);V3=matrix(NA,n,M);V4=matrix(NA,n,M);
U1=runif(n); U2=runif(n);
U1=runif(n); U2=U1;
X1=(1-U1)^{-1/a1}; X2=(1-U2)^{-1/a2}
R1=rank(X1);R2=rank(X2)
for(j in (1:M)){
               for(l in (1:n)){
            V1[l,j]=R1[l]>n+0.5-k1*(1-t[j])|R2[l]>n+0.5-k1*t[j];
            V2[l,j]=R1[l]>n+0.5-k2*(1-t[j])|R2[l]>n+0.5-k2*t[j];
            V3[l,j]=R1[l]>n+0.5-k3*(1-t[j])|R2[l]>n+0.5-k3*t[j];
            V4[l,j]=R1[l]>n+0.5-k4*(1-t[j])|R2[l]>n+0.5-k4*t[j]}
}
for (j in (1:M)){
    A1[j]=k1^{-1}*sum(V1[,j]);
    A2[j]=k2^{-1}*sum(V2[,j]);
```

```
    A3[j]=k3^{-1}*sum(V3[,j]);
    A4[j]=k4^{-1}*sum(V4[,j]);
}
maximum=max(A1,A2,A3,A4);
minimum=min(A1,A2,A3,A4);
plot(t,A1,type="l",xlab="t",ylab="A(t)",ylim=c(minimum,maximum),col="black");
plot(t,A1,type="l",xlab="t",ylab="A(t)",ylim=c(0,2),col="black");
lines(t,A2,type="l",col="red");
lines(t,A3,type="l",col="blue");
lines(t,A4,type="l",col="green");
```

Bibliography

[1] R. Adler, A. Feldman, and M. Taqqu. *A practical guive to Heavy Tails: Statistical Techniques and Applications*. Springer-Verlag, 1998.

[2] J. Beirlant, Y. Goegebeur, J. Segers, J Teugels, D. De Waal, and C Ferro. *Statistics of Extremes: Theory and Applications*. Wiley, 2004.

[3] P. Billingsley. *Probability and Measure*. Wiley-Interscience, third edition, 1995.

[4] L. de Haan and A. Ferreira. *Extreme Value-Theory: An Introduction*. Springer-Verlag, 2006.

[5] J. Einmahl, A. Krajina, and J. Segers. A method of moments estimator of tail dependence. Technical report, Tilberg University, 2008.

[6] 0. Kallenberg. *Random Measures*. Academic Press, third edition, 1983.

[7] R. Kulik. Notes on regularly varying random variables, 2014.

[8] L. Peng. A practial way for estimating tail dependence functions. *Statistica Sinica*, 20, 2010.

[9] S. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer, 1987.

[10] S. Resnick. *Heavy-Tail Phenomena:Probabilistic and Statistical Modelling*. Springer, 2009.

[11] J. Rosenthal. *A First Look at Rigorous Probability Theory*. World Scientific Publishing Co., second edition, 2007.

[12] G. Samorodnitsky and J. Szulga. An asymptotic evaluation of the tail of a multiple symmetric - stable integral. *The Annals of Probability*, 17, 1989.