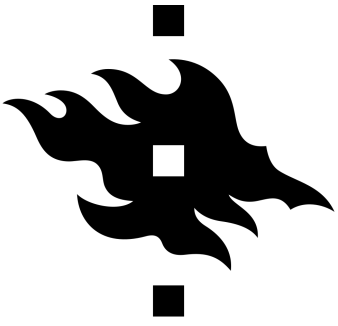
Multivariate Regular Variation

Master's Programme in Mathematics and Statistics

Department of Mathematics and Statistics

2020

Alexandre Bernardo



HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta/Osasto — Fakultet/Sektion — Faculty	Koulutusohjelma — Utbildningsprogram — Degree programme
Faculty of Science	MAST

Tekijä — Författare — Author
Alexandre Bernardo

Työn nimi — Arbetets titel — Title

Multivariate Regular Variation

Oppiaine — Läroämne — Subject
Mathematics

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Master's Thesis	September 2020	55 pages

Tiivistelmä — Referat — Abstract

In insurance and reinsurance, heavy-tail analysis is used to model insurance claim sizes and frequencies in order to quantify the risk to the insurance company and to set appropriate premium rates. One of the reasons for this application comes from the fact that excess claims covered by reinsurance companies are very large, and so a natural field for heavy-tail analysis.

In finance, the multivariate returns process often exhibits heavy-tail marginal distributions with little or no correlation between the components of the random vector (even though it is a highly correlated process when taking the square or the absolute values of the returns). The fact that vectors which are considered independent by conventional standards may still exhibit dependence of large realizations leads to the use of techniques from classical extreme-value theory, that contains heavy-tail analysis, in estimating an extreme quantile of the profit-and-loss density called value-at-risk (VaR).

The need of the industry to understand the dependence between random vectors for very large values, as exemplified above, makes the concept of multivariate regular variation a current topic of great interest. This thesis discusses multivariate regular variation, showing that, by having multiple equivalent characterizations and and by being quite easy to handle, it is an excellent tool to address the real-world issues raised previously.

The thesis is structured as follows. At first, some mathematical background is covered: the notions of regular variation of a tail distribution in one dimension is introduced, as well as different concepts of convergence of probability measures, namely vague convergence and M^*-convergence. The preference in using the latter over the former is briefly discussed.

The thesis then proceeds to the main definition of this work, that of multivariate regular variation, which involves a limit measure and a scaling function. It is shown that multivariate regular variation can be expressed in polar coordinates, by replacing the limit measure with a product of a one-dimensional measure with a tail index and a spectral measure.

Looking for a second source of regular variation leads to the concept of hidden regular variation, to which a new hidden limit measure is associated.

Estimation of the tail index, the spectral measure and the support of the limit measure are next considered. Some examples of risk vectors are next analyzed, such as risk vectors with independent components and risk vectors with repeated components.

The support estimator presented earlier is then computed in some examples with simulated data to display its efficiency. However, when the estimator is computed with real-life data (the value of stocks for different companies), it does not seem to suit the sample in an adequate way.

The conclusion is drawn that, although the mathematical background for the theory is quite solid, more research needs to be done when applying it to real-life data, namely having a reliable way to check whether the data stems from a multivariate regular distribution, as well as identifying the support of the limit measure.

Avainsanat — Nyckelord — Keywords
Multivariate regular variation, Hidden regular variation,

Säilytyspaikka — Förvaringsställe — Where deposited
Kumpulan tiedekirjasto

Muita tietoja — Övriga uppgifter — Additional information

Contents

1 Introduction **3**

2 Mathematical background **3**
2.1 Convergence of measures 4
2.2 The problem with vague convergence 7

3 Multivariate regular variation **8**
3.1 Polar coordinates and the spectral measure 13

4 Hidden regular variation **20**

5 Parameter estimation **23**
5.1 Support estimation 26
5.2 Estimator for the spectral measure 27

6 Simulated examples **32**
6.1 Independent components 32
6.2 Asymptotic independence 35
6.3 Identical components 35
6.4 Removal of independent components 37

7 Support estimation for simulated data **39**

8 A worked example: daily stock returns **42**

9 Final Remarks **45**

A R codes **49**
A.1 Section 6.1 49
A.2 Section 6.3 49
A.3 Chapter 7 50
A.3.1 Multivariate regular variation 50
A.3.2 Hidden regular variation 51
A.3.3 Regular variation with exponential noise 52

List of abbreviations and symbols

RV_α	Set of functions in $\mathbb{R}$ that have regular variation with parameter α.
$\overline{F}$	Tail function of a random variable X, such that $\overline{F}(x) = \mathbb{P}[X > x]$, for appropriate x.
$\overline{F}^{-1}$	Generalized inverse distribution function of a random variable X. We have $\overline{F}^{-1}(y) = \inf_x\{\overline{F}(x) \geqslant y\}$.
$f \sim g$	Approximation notation meaning that $\lim_{x\to\infty} f(x)/g(x) = 1$.
$\mathbb{1}_A$	Indicator function of a set A. We have that $\mathbb{1}_A(x) = 1$ if $x \in A$ and $\mathbb{1}_A(x) = 0$ if $x \notin A$.
$\mathbb{1}[\cdot]$	Indicator function for events in a probability space. If A is an event, then $\mathbb{1}[A] = 1$ if A is true and $\mathbb{1}[A] = 0$ if A is not true.
ϵ_X	Random indicator measure of a random variable X. We have that $\epsilon_X(A) = 1$ if $X \in A$ and $\epsilon_X(A) = 0$ if $X \notin A$.
$d(\cdot)$	Distance function.
$f \to a$	Notation meaning that the limit of a function f equals value a as $n \to \infty$ or $t \to \infty$.
$\mu_n \Rightarrow \mu$	Weak convergence of the sequence of measures $\{\mu_n\}$ to a measure μ.
$\mu_n \xrightarrow{\nu} \mu$	Vague convergence of the sequence of measures $\{\mu_n\}$ to a measure μ.
$\mu_n \xrightarrow{*} \mu$	$\mathbb{M}^*$-convergence of the sequence of measures $\{\mu_n\}$ to a measure μ.
$\forall$	Notation meaning "for all".
$\exists$	Notation meaning "there is" or "exists".
∂A	Boundary of a set A with respect to some topology.
$\overline{A}$	Closure of a set A with respect to some topology.
A^c	Complement of a set A (with respect to some space).
iff	Abbreviation of "if and only if".
wrt	Abbreviation of "with respect to".

1 Introduction

The study of regular variation in multiple dimensions is the main focus of this thesis. This concept can be considered integrated in heavy-tail analysis, which in turn is a branch of extreme-value theory.

Suppose that an insurance company handles multiple types of insurance, with several insured clients, for instance, having car, home and some umbrella insurance. If it happens that an extreme event, such as the passing of a hurricane or the spread of large wildfires, strikes the residence area of those clients, then it is very likely that there will be a great quantity of insurance claims pertaining to different lines of business being filed. How can the insurance company hedge against that risk? Is there a way to know what are the chances that such a situation would occur?

The theory of multivariate regular variation attempts to answer these and many other questions. The core idea is that the probability that a random vector could be very large, which translates the likelihood of an extreme event taking place, will, for an appropriate scaling function, tend to a certain measure called the limit measure. Knowing this limit measure would allow us to ascertain how likely it is for the value of the random vector to go above a certain threshold, as well as to determine how does that likelihood evolve for different thresholds.

As it is shown throughout this text, multivariate regular variation is not a simple extension of uni-dimensional regular variation, and it requires some additional mathematical background, namely that which relates to different kinds of convergence of probability measures. Certain tools from Topology, Probability Theory and Measure Theory are used, but these can generally be found in most textbooks of those fields.

2 Mathematical background

We will need some background information on convergence of measures and regular variation of distributions. Regular variation is the central concept of this thesis, and in particular in the multivariate case, which will be defined in the next chapter. We begin this chapter by presenting it in one dimension.

Definition 2.1. ***(Univariate Regular Variation)*** *Let $U : \mathbb{R}_+ \to \mathbb{R}_+$ be a function and $\alpha \in \mathbb{R}$. We say that U has* regular variation *with parameter α, or $U \in RV_\alpha$, if, for all $x > 0$,*

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

Example 2.1. *If X is a random variable that follows a Pareto distribution with parameters 1 and α (scale and shape, respectively), then the tail function $\overline{F}(x) = \mathbb{P}[X > x] = x^{-\alpha}$ has regular variation with parameter $-\alpha$.*

In order to discuss the different notions of convergence of measures, it is necessary first to introduce the concept of Radon measure.

Definition 2.2. *Let μ be a measure over a locally compact, Hausdorff space Y. Then μ is*

1. Locally finite *iff for all $K \subset Y$, K compact*: $\mu(K) < \infty$
2. Inner regular *iff for all $V \subset Y$, V open, measurable*:

$$\mu(V) = \sup_{\substack{K\, compact \\ K \subset Y}} \{\mu(K)\}$$

3. Outer regular *iff for all $A \subset Y$*: $\mu(A) = \inf\{\mu(V) | A \subset V,\ V \text{ open}\}$
4. Radon measure *iff it is locally finite, inner regular and outer regular*

We denote by $\mathcal{M}(Y)$ the space of locally finite measures on Y. For a locally compact, Hausdorff space Y, $\mathbb{M}_+(Y)$ denotes the space of positive Radon measures on Y. We say that a measure is positive if it is a function from measurable sets of a measurable space to the non-negative real numbers.

Example 2.2. *Consider the Dirac measure δ_x on a topological space X for a given $x \in X$. For any measurable set $A \subset X$, we have*

$$\delta_x(A) = \mathbb{1}_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}.$$

The Dirac measure on any topological space is a Radon measure.

2.1 Convergence of measures

Here we define the convergence of measures. In particular, we want to clarify the meaning of the convergence of a sequence of probability measures μ_n on a locally compact Hausdorff space Y.

Given a measure μ on a topological space X and a function $f : X \to X$, we denote by μf the L^1 norm of f with respect to μ, that is,

$$\mu f = \int f d\mu = \int_X f(x) d\mu(x)$$

Definition 2.3. ***(Weak Convergence)*** *Let $\{\mu_n\}$ be a sequence of positive probability measures on a locally compact Hausdorff space Y. We say that the sequence $\{\mu_n\}$* converges weakly *to a measure μ on Y, denoted by $\mu_n \xrightarrow{w} \mu$ or $\mu_n \Rightarrow \mu$, if and only if $\mu_n f \to \mu f$, for every bounded continuous function $f\colon Y \to Y$.*

Remark 2.1. *Note that the definition of weak convergence depends on the notion of continuity and, therefore, on the topology chosen for the space Y. It is possible for the set of continuous functions to change in such a way that a sequence of measures converges weakly to some measure in one topology but not on another. Examples* 2.3 *and* 2.4 *demonstrate that possibility.*

Like many notions of convergence of measures, the concept of weak convergence relies on having convergence of the form

$$\mu_n f \to \mu f$$

for some set of test functions f. The idea is that the average value of each of these test functions with respect to the sequence of measures $\{\mu_n\}$ should converge. The following proposition presents an alternative characterization of weak convergence that relies on test sets, rather than test functions.

Proposition 2.1. *Let $\{\mu_n\}$ be a sequence of positive probability measures on a locally compact Hausdorff space Y. Then, if μ is a probability measure on Y, we have that*

$$\mu_n \Rightarrow \mu \ \textit{iff} \ \lim_{n\to\infty} \mu_n(B) = \mu(B),$$

for all sets $B \subset Y$ such that $\mu(\partial B) = 0$.

Proof. See ([2], Section 2, p. 16). □

We now check some examples of sequences of measures exhibiting weak convergence in $\mathbb{R}$.

Example 2.3. *Taking $Y = \mathbb{R}$, with the usual topology, let μ_n be $\delta_{\frac{1}{n}}$, the Dirac measure at $1/n$. Then $\mu_n \Rightarrow \mu$, where μ is the Dirac measure at 0.*

Example 2.4. *As in Example* 2.3*, we consider $Y = \mathbb{R}$ and μ_n as the Dirac measure at $1/n$, $\delta_{\frac{1}{n}}$. However, we now equip $\mathbb{R}$ with the trivial topology $\{\varnothing, \mathbb{R}\}$. The function $f\colon \mathbb{R} \to \mathbb{R}$, defined by*

$$f(x) = \begin{cases} -x^2 - 2x, & x \leqslant 0 \\ x^2 + 1, & x > 0 \end{cases},$$

for $x \in \mathbb{R}$, is continuous on this topology. We have $\mu_n f \to 1 \neq 0 = \delta_0 f$, and so, in this topology, $\{\delta_{\frac{1}{n}}\}$ does not converge weakly to δ_0.

Another concept of convergence of measures, and the one which is normally used when working with the notion of regular variation, is that of vague convergence. The idea behind it is nearly identical to the one in weak convergence. The difference is that the set of test functions is now smaller.

Definition 2.4. ***(Vague Convergence)*** *Let $\{\mu_n\}$ be a sequence of locally finite, positive probability measures on a locally compact, Hausdorff space Y. We say that the sequence $\{\mu_n\}$* converges vaguely *to a measure μ on $\mathcal{M}(Y)$, denoted by $\mu_n \xrightarrow{\nu} \mu$, if and only if $\mu_n f \to \mu f$, for all positive continuous functions $f : Y \to Y$ with compact support.*

Remark 2.2. *Since all compact sets are bounded, we have that any continuous function with compact support is also bounded, and so weak convergence implies vague convergence.*

A third way of handling convergence of measures is the $\mathbb{M}^*$-convergence. For that we will need a few definitions and some notation. Two sets A and B are said to be *bounded away from each other* if $\overline{A} \cap \overline{B} = \varnothing$, where $\overline{S}$ denotes the closure of the set S with respect to some topology. Unless stated otherwise, we will consider the usual topology for the spaces $\mathbb{R}^d$, $d \geqslant 1$, and its subspaces. A set $\mathbb{C} \subset \mathbb{R}^d$ is called a *cone* if $x \in \mathbb{C}$ implies $tx \in \mathbb{C}$ for all $t > 0$. Note that this definition allows for cones which do not include the origin. The complement of a closed cone in $\mathbb{C}$ (a closed subset of $\mathbb{C}$ that is also a cone) is an open cone in $\mathbb{C}$ (it is an open subset of $\mathbb{C}$ that is a cone).

Definition 2.5. ***($\mathbb{M}^*$-Convergence)*** *Let $\mathbb{C}$ and $\mathbb{F}$ be two cones in $[0,\infty)^d$ such that $\{\mathbf{0}\} \subset \mathbb{F} \subset \mathbb{C}$. Let $\mathcal{C}$ be the Borel σ-algebra of $\mathbb{C}$ and the σ-algebra of $\mathbb{O} = \mathbb{C} \backslash \mathbb{F}$ be $\mathcal{O} = \{B \subset \mathbb{O} : B \in \mathcal{C}\}$. Denote by $\mathbb{M}^*(\mathbb{C}, \mathbb{O})$ the class of Borel measures on $\mathcal{O}$ that assign finite measure to all sets in $\mathcal{O}$ that are bounded away from $\mathbb{F}$.*

Let $\{\mu_n\}$ be a sequence of measures in $\mathbb{M}^*(\mathbb{C}, \mathbb{O})$ and μ a measure also in $\mathbb{M}^*(\mathbb{C}, \mathbb{O})$. We say that the sequence $\{\mu_n\}$ $\mathbb{M}^*$-converges *to a measure μ, denoted by $\mu_n \xrightarrow{*} \mu$, if and only if $\mu_n(B) \to \mu(B)$, for all $B \in \mathcal{O}$ such that $\mu(\partial B) = 0$ and B is bounded away from $\mathbb{F}$.*

Example 2.5. *Let $d = 2$, $\mathbb{C} = [0,\infty)^2$ and $\mathbb{F} = \{\mathbf{0}\}$. Here $\mathbb{M}^*(\mathbb{C}, \mathbb{O})$ is the class of finite Borel measures on $(0,\infty)^2$. Consider the sequence of measures $\{\alpha_n\}$, where, for each natural number n,*

$$\alpha_n(A) = \begin{cases} 1, & \exists (x^1, x^2) \in A \colon x^1 + x^2 \leqslant n \text{ or } A = \varnothing \\ 0, & \text{otherwise} \end{cases}.$$

For any set A in $\mathbb{C}$, we have $\alpha_n(A) \to 1$, and so $\alpha_n \xrightarrow{*} \mathbf{1}$, where $\mathbf{1}$ is the constant measure that gives value 1 to every set in $\mathbb{C}$.

We also define $\mathcal{C}_{\mathbb{F}}$ as the set of all bounded, continuous real-valued functions f on $\mathbb{C}$ for which there is a number $r > 0$ such that $f(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in \mathbb{C}$ satisfying $d(\boldsymbol{x}, \mathbb{F}) < r$, for some distance function d.

Similarly to other notions of convergence of measures, we also have the analogous Portmanteau Theorem for the $\mathbb{M}^*$-convergence.

Theorem 2.1. ***(Portmanteau Theorem)*** *Let $\{\mu_n\}$ be a sequence of measures in $\mathbb{M}^*(\mathbb{C}, \mathbb{O})$ and μ a measure in $\mathbb{M}^*(\mathbb{C}, \mathbb{O})$, as in Definition 2.5. The following are equivalent:*

1. *$\mu_n \xrightarrow{*} \mu$ in $\mathbb{M}^*(\mathbb{C}, \mathbb{O})$.*
2. *$\mu_n f \to \mu f$, for every $f \in \mathcal{C}_{\mathbb{F}}$.*
3. *$\liminf_{n\to\infty} \mu_n(G) \geqslant \mu(G)$ and $\limsup_{n\to\infty} \mu_n(A) \leqslant \mu(A)$, for all closed sets $A \in \mathcal{O}$ and open sets $G \in \mathcal{O}$ such that $\overline{G} \cap \mathbb{F} = \varnothing$.*

Proof. See ([11], Section 3, pp. 15-17). □

The following lemma is a consequence of the Portmanteau Theorem. The lemma states that, for a sequence of measures $\{\mu_n\}$ and possible limit measure μ, it is sufficient to check condition $\mu_n(B) \to \mu(B)$ for only a certain type of sets B, in order to determine whether $\mu_n \xrightarrow{*} \mu$.

Lemma 2.1. *Let $\{\mu_n\}$ be a sequence of measures in $\mathbb{M}^*(\mathbb{C}, \mathbb{O})$ and μ a measure in $\mathbb{M}^*(\mathbb{C}, \mathbb{O})$, as in Definition 2.5. Define $f\colon \mathbb{O} \to \mathbb{R}$ by $f(\boldsymbol{x}) = \mu([\mathbf{0}, \boldsymbol{x}]^c)$. Let $x \in \mathbb{O}$ be a continuity point of $\mu([\mathbf{0}, \cdot]^c)$. Then*

$$\mu_n \xrightarrow{*} \mu \text{ iff } \mu_n([\mathbf{0}, \boldsymbol{x}]^c) \to \mu([\mathbf{0}, \boldsymbol{x}]^c).$$

Proof. See ([14], Section 6, p. 174). □

2.2 The problem with vague convergence

The usual way to define Multivariate Regular Variation uses vague convergence. We will be working with a definition that uses the $\mathbb{M}^*$-convergence of measures. The reason for this switch is to avoid the issues created by vague convergence. These issues are discussed next.

In the traditional definition, vague convergence is considered for Radon measures on $[0, \infty]^d \backslash \{\mathbf{0}\}$, that is, the set $[0, \infty)^d$ is compactified and the origin is removed, so that sets bounded away from $\{\mathbf{0}\}$ are relatively compact. This is adequate because Radon measures put finite measure on relatively compact sets as in Definition 2.2 (to see that this is true, recall that a relatively

compact set is a set whose closure is compact and all compact subsets of a Hausdorff space are relatively compact). However, note that, if we consider hidden regular variation, it may be necessary to remove more than just a point, so that the cone $\mathbb{F}$ in Definition 2.5 is larger a set than just the origin.

One of the problems caused by vague convergence is that polar coordinates, which prove to be quite useful in this context, are defined on $[0,\infty)^d\backslash\{\mathbf{0}\}$, and so, if we use the space $[0,\infty]^d\backslash\{\mathbf{0}\}$, we need to show the equivalence between the two notions of multivariate regular variation (polar on the former and Cartesian on the later), which is generally not as straightforward as it might seem.

Another problem has to do with geometric properties. For instance, in two dimensions ($d = 2$), if we consider two parallel but separate lines on $(0,\infty)^2$ with the same positive and finite slope, these lines both converge to the point (∞,∞). Thus, the two lines are not bounded away from each other, which may cause complications as some sets whose boundaries are two parallel lines are therefore not relatively compact. $\mathbb{M}^*$-convergence is a solid alternative to vague convergence that is able to avoid the idea of compactification and the issues generated by the procedure.

3 Multivariate regular variation

We consider non-negative random vectors $\boldsymbol{X} = (X^1, X^2, \ldots, X^d)$ called risk vectors. We say that the distribution of $\boldsymbol{X}$ has *multivariate regular variation* if there exists an increasing function $b\colon \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\lim_{t\to\infty} b(t) = \infty$$

and a non-negative Radon measure μ on $\mathbb{R}_+^d$ such that

$$t\mathbb{P}\left[\frac{\boldsymbol{X}}{b(t)} \in \cdot\right], \xrightarrow{\nu} \mu(\cdot) \tag{3.1}$$

where $\xrightarrow{\nu}$ denotes vague convergence in $\mathbb{M}_+(\mathbb{E})$. We may extend this standard definition to $\mathbb{M}^*$-convergence as follows.

Definition 3.1. ***(Multivariate Regular Variation)*** *Let $\mathbb{C}$ and $\mathbb{F}$ be closed cones containing the origin such that $\mathbb{F} \subset \mathbb{C} \subset [0,\infty)^d$. The distribution of a random vector $\boldsymbol{X}$ has* regular variation *on $\mathbb{O} = \mathbb{C}\backslash\mathbb{F}$ if there exists an increasing function $b\colon \mathbb{R} \to \mathbb{R}$ with regular variation such that $\lim_{t\to\infty} b(t) = \infty$ and a positive measure μ on $\mathbb{M}^*(\mathbb{C},\mathbb{O})$ such that*

$$t\mathbb{P}\left[\frac{\boldsymbol{X}}{b(t)} \in \cdot\right] \xrightarrow{*} \mu(\cdot)$$

in $\mathbb{M}^*(\mathbb{C},\mathbb{O})$. *When the cone* $\mathbb{C}$ *is clear from the context, we may simply write* $\mathbb{M}^*(\mathbb{O})$ *instead of* $\mathbb{M}^*(\mathbb{C},\mathbb{O})$.

We now present an example of multivariate regular variation in two dimensions, that is, in the case $d = 2$.

Example 3.1. *Let* $\boldsymbol{X} = (X^1, X^2)$ *be a two-dimensional random vector such that both components* X^1 *and* X^2 *are independent and identically distributed unit Pareto random variables, that is,* X^1 *and* X^2 *follow a Pareto distribution with both parameters equal to 1. This means that they have a common tail function* $\overline{F}$ *defined, for* $x > 0$, *by*

$$\overline{F}(x) = \mathbb{P}[X^1 > x] = \mathbb{P}[X^2 > x] = \frac{1}{x}.$$

Let $\mathbb{C} = [0,\infty)^2$ *and* $\mathbb{F} = \{\mathbf{0}\}$, *and so* $\mathbb{O} = [0,\infty)^2\backslash\{\mathbf{0}\}$. *We want to check that* $\boldsymbol{X}$ *has regular variation on* $\mathbb{O}$ *with* $b : \mathbb{R} \to \mathbb{R}$ *given by* $b(t) = t$. *By Lemma 2.1, it is enough to verify the limit of* $t\mathbb{P}\left[\dfrac{\boldsymbol{X}}{b(t)} \in [\mathbf{0},\boldsymbol{x}]^c\right]$, *as* $t \to \infty$, *for* $\boldsymbol{x} \in \mathbb{O}$.

Let $\boldsymbol{x} = (x^1, x^2) \in \mathbb{O}$. *It is only relevant to consider the case where* $x^1 \times x^2 \neq 0$ *since, by Lemma 2.1, it is enough to look at the continuity points of* $\mu([\mathbf{0},\cdot]^c)$. *Then, because* X^1 *and* X^2 *are independent, we have*

$$\begin{aligned}
t\mathbb{P}\left[\frac{\boldsymbol{X}}{b(t)} \in [\mathbf{0},\boldsymbol{x}]^c\right] &= t\mathbb{P}\left[\frac{X^1}{b(t)} > x^1 \vee \frac{X^2}{b(t)} > x^2\right] \\
&= t\mathbb{P}\left[\frac{X^1}{b(t)} > x^1\right] + t\mathbb{P}\left[\frac{X^2}{b(t)} > x^2\right] - t\mathbb{P}\left[\frac{X^1}{b(t)} > x^1\right]\mathbb{P}\left[\frac{X^2}{b(t)} > x^2\right] \\
&= t\overline{F}(b(t)x^1) + t\overline{F}(b(t)x^2) - t\overline{F}(b(t)x^1)\overline{F}(b(t)x^2) \\
&= t\frac{1}{b(t)x^1} + t\frac{1}{b(t)x^2} - t\frac{1}{b(t)x^1}\frac{1}{b(t)x^2} \\
&= \frac{1}{x^1} + \frac{1}{x^2} - \frac{1}{tx^1x^2}.
\end{aligned}$$

We then compute the limit

$$\lim_{t\to\infty} t\mathbb{P}\left[\frac{\boldsymbol{X}}{b(t)} \in [\mathbf{0},\boldsymbol{x}]^c\right] = \frac{1}{x^1} + \frac{1}{x^2},$$

and we thus conclude that the random vector $\boldsymbol{X}$ *has multivariate regular variation on* $\mathbb{O} = [0,\infty)^2\backslash\{\mathbf{0}\}$ *with scaling function* $b(t) = t$ *and limit measure* μ *given by* $\mu([\mathbf{0},\boldsymbol{x}]^c) = \dfrac{1}{x^1} + \dfrac{1}{x^2}$, *for* $\boldsymbol{x} = (x^1, x^2) \in \mathbb{O}$ *such that* $x^1 \times x^2 \neq 0$.

Before delving into alternative characterizations of multivariate regular variation, we check that the previous notion of univariate regular variation given by 2.1 is equivalent to the statement of Definition 3.1 for one dimension.

Proposition 3.1. *Let X be a random variable with tail function $\overline{F}(x) = \mathbb{P}[X > x]$ for $x \in \mathbb{R}$. Then the distribution of X has regular variation as in Definition* 3.1 *if and only if $\overline{F}$ has regular variation as in Definition* 2.1*.*

Proof. Suppose $\overline{F} \in RV_\alpha$ for some $\alpha \in \mathbb{R}$. Take $\mathbb{C} = [0, \infty)$ and $\mathbb{F} = \{0\}$, such that $\mathbb{O} = (0, \infty)$. Let μ be the Pareto measure given by $\mu((x, \infty)) = x^{-\alpha}$, for $x > 0$. Define the function b such that, for positive t,

$$b(t) = \overline{F}^{-1}(1/t).$$

Then, b satisfies the requisites from Definition 3.1 and

$$\overline{F}(b(t)) \sim \frac{1}{t}. \tag{3.2}$$

We want to prove that

$$t\mathbb{P}\left[\frac{X}{b(t)} \in \cdot\right] \xrightarrow{*} \mu(\cdot),$$

which, by Lemma 2.1, is equivalent to

$$t\mathbb{P}\left[\frac{X}{b(t)} > x\right] \to \mu([0, x]^c), \text{ as } t \to \infty,$$

which is the same as

$$t\overline{F}(b(t)x) \to x^{-\alpha}, \text{ as } t \to \infty.$$

Using (3.2), we have

$$t\overline{F}(b(t)x) = t\overline{F}(b(t))\frac{\overline{F}(b(t)x)}{\overline{F}(b(t))} \to 1 \times x^{-\alpha} = x^{-\alpha}, \text{ as } t \to \infty.$$

Suppose now that the distribution of X has regular variation as in Definition 3.1. We want to prove that the tail function $\overline{F}$ satisfies

$$\frac{\overline{F}(tx)}{\overline{F}(t)} \to x^{-\alpha}, \text{ as } t \to \infty.$$

The function b has regular variation and so there is a real number α such that $b \in RV_{\frac{1}{\alpha}}$. Using this fact and the dominated convergence theorem, we have

$$\begin{aligned}
\lim_{t\to\infty} \frac{\overline{F}(tx)}{\overline{F}(t)} &= \lim_{t\to\infty} \frac{\overline{F}(b(t)x)}{\overline{F}(b(t))} \\
&= \lim_{t\to\infty} \frac{t\mathbb{P}\left[\frac{X}{b(t)} > x\right]}{t\mathbb{P}\left[\frac{X}{b(t)} > 1\right]} \\
&= \lim_{t\to\infty} \frac{t\mathbb{P}\left[\frac{X}{b(t)}\frac{b(tx^\alpha)}{b(tx^\alpha)} > x\right]}{t\mathbb{P}\left[\frac{X}{b(t)} > 1\right]} \\
&= \lim_{t\to\infty} \frac{tx^\alpha x^{-\alpha}\mathbb{P}\left[\frac{X}{b(tx^\alpha)}x > x\right]}{t\mathbb{P}\left[\frac{X}{b(t)} > 1\right]} \\
&= \lim_{t\to\infty} \frac{tx^\alpha\mathbb{P}\left[\frac{X}{b(tx^\alpha)} > 1\right]}{t\mathbb{P}\left[\frac{X}{b(t)} > 1\right]} x^{-\alpha} \\
&= 1 \times x^{-\alpha} \\
&= x^{-\alpha}.
\end{aligned}$$

□

Regular variation can also be defined in terms of a sequence b_n instead of a function b. The next theorem gives us the equivalence between these notions.

Theorem 3.1. *Let $\boldsymbol{X}$ be a non-negative random vector in $\mathbb{R}^d$, that is, all the vector components X^1, ..., X^d are non-negative. Then the following statements are equivalent:*

1. *$\boldsymbol{X}$ has multivariate regular variation according to Definition 3.1.*
2. *There exists a sequence b_n such that $\lim_{n\to\infty} b_n = \infty$ and a positive measure μ in $\mathbb{M}^*(\mathbb{C}, \mathbb{O})$ such that*
$$n\mathbb{P}\left[\frac{\boldsymbol{X}}{b_n} \in \cdot\right] \xrightarrow{*} \mu(\cdot).$$
3. *There exists a positive measure μ in $\mathbb{M}^*(\mathbb{C}, \mathbb{O})$ such that*
$$\lim_{t\to\infty} \frac{1 - F(t\boldsymbol{x})}{1 - F(t\boldsymbol{1})} = \lim_{t\to\infty} \frac{\mathbb{P}[\boldsymbol{X}/t \in [\boldsymbol{0}, \boldsymbol{x}]^c]}{\mathbb{P}[\boldsymbol{X}/t \in [\boldsymbol{0}, \boldsymbol{1}]^c]} = \mu([\boldsymbol{0}, \boldsymbol{x}]^c])$$

for all $\boldsymbol{x} \neq \mathbf{0}$ *which are continuity points of the function* $\mu([\mathbf{0}, \cdot]^c)$.

Proof. The direction $(1 \Rightarrow 2)$ is quite easy. We just need to replace $b(t)$ with b_n in Definition 3.1.

The direction $(2 \Rightarrow 3)$ is found in ([14], Section 6, pp. 175-176).

For the direction $(3 \Rightarrow 1)$, we note that the univariate function $t \mapsto \overline{F}(t\mathbf{1})$ is regularly varying, that is, there is some $\alpha > 0$ such that $\overline{F}(t\mathbf{1}) \in RV_{-\alpha}$.

Define the function b as

$$b(t) = \overline{F}^{-1}(\frac{1}{t}\mathbf{1}),$$

for positive t. Then, b satisfies the requisites from Definition 3.1 and

$$\overline{F}(b(t)\mathbf{1}) \sim \frac{1}{t}. \tag{3.3}$$

Thus, we have

$$t\mathbb{P}\left[\frac{\boldsymbol{X}}{b(t)} \in [\mathbf{0}, \boldsymbol{x}]^c\right] = \frac{\overline{F}(b(t)\boldsymbol{x})}{1/t} \sim \frac{\overline{F}(b(t)\boldsymbol{x})}{\overline{F}(b(t)\mathbf{1})} \sim \frac{\overline{F}(t\boldsymbol{x})}{\overline{F}(t\mathbf{1})} \sim \mu([\mathbf{0}, \boldsymbol{x}]^c]).$$

Then, by Lemma 2.1, we conclude that the distribution of $\boldsymbol{X}$ has regular variation as in Definition 3.1. □

The limit measure in Definition 3.1 satisfies an important scaling property.

Proposition 3.2. ***(Scaling Property)*** *Let* $\boldsymbol{X}$ *be a non-negative random vector in* $\mathbb{R}^d$ *with regularly varying distribution according to Definition 3.1. Then, there is a number* α *such that* $b \in RV_{\frac{1}{\alpha}}$ *and* μ *satisfies, for* $\theta > 0$,

$$\mu(\theta B) = \theta^{-\alpha}\mu(B),$$

for all sets $B \in \mathcal{O}$ *such that* $\mu(\partial B) = 0$ *and* $\overline{B} \cap \overline{\mathbb{F}} = \varnothing$..

Proof. Let $\theta > 0$ be fixed. From Definition 3.1 we have

$$\lim_{t\to\infty} t\mathbb{P}\left[\frac{\boldsymbol{X}}{b(t)} \in B\right] = \mu(B)$$

for all $B \in \mathcal{O}$ such that $\mu(\partial B) = 0$ and B is bounded away from $\mathbb{F}$.

Definition 3.1 also states that b is a regularly varying function for some $\rho \in \mathbb{R}$, $b \in RV_\lambda$. We can take $\alpha = 1/\lambda$ so that $b \in RV_{\frac{1}{\alpha}}$, which means that, for $x > 0$,

$$\lim_{t\to\infty} \frac{b(xt)}{b(t)} = x^{\frac{1}{\alpha}}.$$

Morevoer, we note that

$$\lim_{t\to\infty} t\mathbb{P}\left[\frac{\boldsymbol{X}}{b(t)} \in \theta B\right] = \mu(\theta B),$$

which is well-defined because $\mathbb{O}$ is a cone and $\mu(\partial(\theta B)) = 0$, if $\mu(\partial B) = 0$. Then, for $\theta > 0$, we have

$$\mu(\theta B) = \lim_{t\to\infty} t\mathbb{P}\left[\frac{\boldsymbol{X}}{b(t)} \in \theta B\right] = \frac{1}{\theta^\alpha} \lim_{t\to\infty}(\theta^\alpha t))\mathbb{P}\left[\frac{\boldsymbol{X}}{b(\theta^\alpha t)}\frac{b(\theta^\alpha t)}{b(t)} \in \theta B\right] =$$
$$\overset{1}{=} \frac{1}{\theta^\alpha} \lim_{(\theta^\alpha t)\to\infty}(\theta^\alpha t))\mathbb{P}\left[\frac{\boldsymbol{X}}{b(\theta^\alpha t)} \in B\right] = \frac{1}{\theta^\alpha}\mu(B) = \theta^{-\alpha}\mu(B),$$

where in $\overset{1}{=}$ we use the fact that $b \in RV_{\frac{1}{\alpha}}$ (setting $x = \theta^\alpha$) and the dominated convergence theorem. This ends the proof.

□

3.1 Polar coordinates and the spectral measure

We can make use of the polar coordinates transformation

$$x \mapsto \left(||x||, \frac{x}{||x||}\right)$$

to decompose the limit measure μ in Definition 3.1 into a product measure. For vague convergence, assuming $b \in RV_{\frac{1}{\alpha}}$, we have by the scaling property

$$t\mathbb{P}\left[\left(\frac{||X||}{b(t)}, \frac{X}{||X||}\right) \in \cdot\right] \xrightarrow{\nu} \alpha\nu_{(\alpha)} \times S^*(\cdot),$$

where $\nu_{(\alpha)}$ is the Pareto measure given by $\nu_{(\alpha)}((x,\infty)) = x^{-\alpha}$, for $x > 0$, and S^* is a finite measure on the unit sphere

$$\partial\aleph_0 = \{\boldsymbol{y} \in \mathbb{E} : d(\boldsymbol{y}, \boldsymbol{0}) = 1\}.$$

We can then define a probability measure

$$S(\cdot) = \frac{S^*(\cdot)}{S^*(\partial\aleph_0)},$$

which is called the *spectral* or *angular measure*.

For $\mathbb{M}^*$ convergence, we consider the cones $\mathbb{C}$, $\mathbb{F}$ and $\mathbb{O}$ as in Definition 2.5, as well as the sets $\aleph_{\mathbb{O}} = \{\boldsymbol{x} \in \mathbb{O} : d(x, \mathbb{F}) \geqslant 1\}$ and $\partial\aleph_{\mathbb{O}} = \{\boldsymbol{x} \in \mathbb{O} : d(x, \mathbb{F}) = 1\}$. Definition 3.1 implies $\mu(\aleph_{\mathbb{O}})$ is positive and finite and, by the scaling property, the function b can be chosen such that $\mu(\aleph_{\mathbb{O}}) = 1$.

Example 3.2. *Let $\mathbb{C} = [0,\infty)^2$ and $\mathbb{F} = \{\mathbf{0}\}$. Let the distance function be the L^2-norm $d(x,\mathbb{F}) = ||x||_2 = \sqrt{(x^1)^2 + (x^2)^2}$, for $x = (x^1, x^2) \in \mathbb{C}$. Then $\partial\aleph_{\mathbb{O}} = \{\boldsymbol{x} : ||x|| = 1\}$, which is the unit simplex for the L^2-norm.*

Before we consider the decomposition of the limit measure μ in a one-dimensional Pareto measure and a probability measure defined on $\partial\aleph_{\mathbb{O}}$, let us introduce a definition and some notation.

Given a set $A \subset [0,\infty)^d$, we define $(A)_1 = \{x^1 : x \in A\}$, that is, $(A)_1$ is the projection of set A onto the first coordinate axis. For our limit measure μ and a continuous bijection $h : \mathbb{O} \mapsto (0,\infty) \times \partial\aleph_{\mathbb{O}}$, we define the set

$$\mathcal{A}_h = \{A \subset (0,\infty)\times\partial\aleph_{\mathbb{O}} \text{ measurable wrt } \mathcal{O}\colon \mu(h^{-1}(\partial A)) = 0, \overline{(A)_1}\cap\{0\} = \varnothing\}.$$

Definition 3.2. *A continuous bijection $h : \mathbb{O} \mapsto (0,\infty) \times \partial\aleph_{\mathbb{O}}$ is called a *-bijection if*

1. *For every set $A \in \mathcal{A}_h$, $h^{-1}(A)$ is bounded away from $\mathbb{F}$ and*
2. *For every measurable set $B \subset \mathbb{O}$ such that B is bounded away from $\mathbb{F}$ we have $\overline{(h(B))_1} \cap \{0\} = \varnothing$.*

Remark 3.1. *The projection onto the first coordinate axis present in Definition 3.2 follows from the need to transform the limit measure into the product of two components, one of them being a one-dimensional Pareto measure. It is not necessary to consider the first axis for this one-dimensional component. It is done here due to convention and clarity of writing.*

The following topological fact will prove useful. It relates the boundary of the inverse image of a set by a continuous bijection with the inverse image of its boundary by the same function.

Lemma 3.1. *Let $f : X \mapsto Y$ be a continuous bijective function between topological spaces X and Y. Then, for all $S \subset Y$,*

$$\partial(f^{-1}(S)) \subset f^{-1}(\partial S).$$

Proof. Recall that, given a set $A \subset Y$ and a point $x \in Y$, we have

$$x \in \partial A \text{ iff } B \cap A \neq \varnothing \text{ and } B \cap Y\backslash A \neq \varnothing,$$

for all neighborhoods B of x, that is, for all open sets B containing x.

Let $x \in \partial(f^{-1}(S))$. We want to prove that $x \in f^{-1}(\partial S)$, which, since f is a bijection, is equivalent to $f(x) \in \partial S$.

Note that, because f is a continuous bijection, if A is a neighborhood of x, then the set $f(A) = \{f(a)\colon a \in A\}$ is a neighborhood of $f(x)$. Similarly,

if B is a neighborhood of $f(x)$, then the set $f^{-1}(B) = \{f^{-1}(b)\colon b \in B\}$ is a neighborhood of x.

Let B be a neighborhood of $f(x)$. Then $f^{-1}(B)$ is a neighborhood of x and so, because $x \in \partial(f^{-1}(S))$, there are z_1, $z_2 \in f^{-1}(B)$ such that

1. $z_1 \in f^{-1}(S)$ and
2. $z_2 \notin f^{-1}(S)$.

Then $f(z_1) \in B$ and $f(z_1) \in S$, which means that $B \cap S \neq \varnothing$. Likewise, $f(z_2) \in B$ but $f(z_2) \notin S$, and so $B \cap Y\backslash S \neq \varnothing$. In other words, we have proven that any neighborhood of $f(x)$ intersects both the set S and its complement $Y\backslash S$, which means $f(x) \in \partial S$. □

We can now start moving towards regular variation in polar coordinates. Recall that $\mathcal{O} = \{B \subset \mathbb{O} : B \in \mathcal{C}\}$, where $\mathcal{C}$ is the Borel σ-algebra of $\mathbb{C}$.

Lemma 3.2. *For a *-bijection h we have*

$$h^{-1}(\mathcal{A}_h) = \{B \in \mathcal{O} : \mu(\partial B) = 0, \overline{B} \cap \mathbb{F} = \varnothing\},$$

where $h^{-1}(\mathcal{A}_h) = \{h^{-1}(A)\colon A \in \mathcal{A}_h\}$.

Proof. $(\subset)$

Let $B \in h^{-1}(\mathcal{A}_h)$. Then there is $A \in \mathcal{A}_h$ such that $B = h^{-1}(A)$. By definition of h, we have $B \subset \mathbb{O}$ and, because A is measurable, $B \in \mathcal{C}$. So $B \in \mathcal{O}$. Due to the continuity of h, we will use Lemma 3.1 to get

$$\partial(h^{-1}(A)) \subset h^{-1}(\partial A).$$

Because $A \in \mathcal{A}_h$, we have

$$\mu(\partial B) = \mu(\partial(h^{-1}(A))) \leqslant \mu(h^{-1}(\partial A)) = 0,$$

and so $\mu(\partial B) = 0$. Since h is a *-bijection, we have $\overline{h^{-1}(A)} \cap \mathbb{F} = \varnothing$.

$(\supset)$

Let $B \in \mathcal{O}$ and $\mu(\partial B) = 0$. Then $A = h(B) \subset (0, \infty) \times \partial\aleph_{\mathbb{O}}$. Since B is measurable, it follows that A is also measurable.

We will again use Lemma 3.1 due to the continuity of h^{-1}, obtaining $\partial h(B) \subset h(\partial B)$, and thus

$$\mu(h^{-1}(\partial h(B))) \leqslant \mu(h^{-1}(h(\partial B))) \text{ if and only if } \mu(h^{-1}(\partial A)) \leqslant \mu(\partial B),$$

which yields $\mu(\partial B) = 0$. Thus we have $\mu(h^{-1}(\partial A)) = 0$ and, from the fact that h is a *-bijection we get $\overline{(h(B))_1} \cap \{0\} = \varnothing$. □

The previous lemma will be used to characterize multivariate regular variation in polar coordinates, but, in order to apply it, we must first show that the the polar transformation is indeed a *-bijection.

Lemma 3.3. *The function*

$$h : \boldsymbol{x} \mapsto \left(d(\boldsymbol{x}, \mathbb{F}), \frac{\boldsymbol{x}}{d(\boldsymbol{x}, \mathbb{F})} \right),$$

defined in $\mathbb{O}$*, is a* *-*bijection.*

Proof. We first prove that h is a continuous bijection from $\mathbb{O}$ to $(0, \infty) \times \partial\aleph_{\mathbb{O}}$.

By definition of distance, $\boldsymbol{x} \mapsto d(\boldsymbol{x}, \mathbb{F})$ is a continuous function and so is h. The definition also gives us $d(\boldsymbol{x}, \mathbb{F}) \in [0, \infty)$ and, since $\mathbb{F}$ is closed, we have, for all $\boldsymbol{x} \in \mathbb{O} = \mathbb{C}\backslash\mathbb{F}$, $d(\boldsymbol{x}, \mathbb{F}) > 0$. Thus, $d(\boldsymbol{x}, \mathbb{F}) \in (0, \infty)$.

On the other hand,

$$d\left(\frac{\boldsymbol{x}}{d(\boldsymbol{x}, \mathbb{F})}, \mathbb{F} \right) = \frac{d(\boldsymbol{x}, \mathbb{F})}{d(\boldsymbol{x}, \mathbb{F})} d\left(\frac{\boldsymbol{x}}{d(\boldsymbol{x}, \mathbb{F})}, \mathbb{F} \right) = \frac{1}{d(\boldsymbol{x}, \mathbb{F})} d\left(\boldsymbol{x}, d(\boldsymbol{x}, \mathbb{F}) \mathbb{F} \right) = 1,$$

where $d(\boldsymbol{x}, \mathbb{F})\mathbb{F} = \mathbb{F}$, since $\mathbb{F}$ is a cone. This means that $\dfrac{\boldsymbol{x}}{d(\boldsymbol{x}, \mathbb{F})} \in \partial\aleph_{\mathbb{O}}$.

To prove that h is a bijection, we begin by noting that, for $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{O}$,

$$\begin{aligned} h(\boldsymbol{x}_1) = h(\boldsymbol{x}_2) &\iff \left(d(\boldsymbol{x}_1, \mathbb{F}), \frac{\boldsymbol{x}_1}{d(\boldsymbol{x}_1, \mathbb{F})} \right) = \left(d(\boldsymbol{x}_2, \mathbb{F}), \frac{\boldsymbol{x}_2}{d(\boldsymbol{x}_2, \mathbb{F})} \right) \\ &\iff \boldsymbol{x}_1 = \boldsymbol{x}_2, \end{aligned}$$

which proves that h is injective.

For surjectivity, given $\boldsymbol{y} \in (0, \infty) \times \partial\aleph_{\mathbb{O}}$, we can consider $\boldsymbol{y} = (y_1, \overline{y})$, with $y_1 \in (0, \infty)$ and $\overline{y} \in \partial\aleph_{\mathbb{O}}$. The corresponding $\boldsymbol{x}$ is given by the product $y_1 * \overline{y}$. In fact,

$$d(\boldsymbol{x}, \mathbb{F}) = d(y_1\overline{y}, \mathbb{F}) = d(y_1\overline{y}, y_1\mathbb{F}) = y_1 d(\overline{y}, \mathbb{F}) = y_1,$$

where the last equality comes from the fact that $\overline{y} \in \partial\aleph_{\mathbb{O}}$. We also have

$$\frac{\boldsymbol{x}}{d(\boldsymbol{x}, \mathbb{F})} = \frac{y_1\overline{y}}{y_1} = \overline{y},$$

and so h is surjective. We conclude that h is a bijection.

Let $A \in \mathcal{A}_h$. We have $\overline{(A)_1} \cap \{0\} = \varnothing$, which means that there is $r > 0$ such that $y_1 > r$, for all $\boldsymbol{y} = (y_1, \overline{y}) \in A$.

Thus, given $\boldsymbol{x} \in h^{-1}(A)$, there is, since h is a bijection, $\boldsymbol{y} = (y_1, \overline{y}) \in A$ such that $h(\boldsymbol{x}) = \boldsymbol{y}$. And we have $d(\boldsymbol{x}, \mathbb{F}) = y_1 > r > 0$. We conclude that $h^{-1}(A)$ is bounded away from $\mathbb{F}$. In a similar way we prove the last property of Definition 3.2, finishing the proof. □

We are finally ready to present multivariate regular variation in polar coordinates, which is given by the following proposition.

Proposition 3.3. *Let $\boldsymbol{X}$ be a random vector. Regular variation of $\boldsymbol{X}$ in $\mathbb{O}$ as defined in Definition 3.1 is equivalent to*

$$t\mathbb{P}\left[\left(\frac{d(\boldsymbol{X},\mathbb{F})}{b(t)},\frac{\boldsymbol{X}}{d(\boldsymbol{X},\mathbb{F})}\right)\in A\right]\to c\mu_{(\alpha)}\times S_{\mathbb{O}}(A)$$

for all measurable sets $A\in\mathcal{A}_h$, where $c>0$ is a constant, $b\in RV_{\frac{1}{\alpha}}$,

$$h:\boldsymbol{x}\mapsto\left(d(\boldsymbol{x},\mathbb{F}),\frac{\boldsymbol{x}}{d(\boldsymbol{x},\mathbb{F})}\right),$$

$\mu_{(\alpha)}$ is the Pareto measure given by $\mu_{(\alpha)}((x,\infty))=x^{-\alpha}$, for $x>0$, and

$$S_{\mathbb{O}}(\Lambda)=\frac{\mu(\{\boldsymbol{x}\in\mathbb{O}:d(\boldsymbol{x},\mathbb{F})\geqslant 1,\frac{\boldsymbol{x}}{d(\boldsymbol{x},\mathbb{F})}\in\Lambda\})}{\mu(\{\boldsymbol{x}\in\mathbb{O}:d(\boldsymbol{x},\mathbb{F})\geqslant 1\})},$$

for a set Λ in $\partial\aleph_{\mathbb{O}}$.

Remark 3.2. *$S_{\mathbb{O}}$ is a probability measure on $\partial\aleph_{\mathbb{O}}$, called the* spectral *or* angular measure *on $\mathbb{O}$. Parameter α is called the* tail index *of $\boldsymbol{X}$.*

Proof. By definition of $\mathbb{M}^*$-convergence in Definition 2.5, we have that regular variation on $\mathbb{O}$ is equivalent to

$$t\mathbb{P}\left[\frac{\boldsymbol{X}}{b(t)}\in B\right]\to\mu(B)$$

for all $B\in\mathcal{O}$ such that $\mu(\partial B)=0$ and $\overline{B}\cap\mathbb{F}=\varnothing$.

Since $h:\boldsymbol{x}\mapsto\left(d(\boldsymbol{x},\mathbb{F}),\frac{\boldsymbol{x}}{d(\boldsymbol{x},\mathbb{F})}\right)$ is a *-bijection by Lemma 3.3, we have, by Lemma 3.2, that regular variation on $\mathbb{O}$ is equivalent to

$$t\mathbb{P}\left[\frac{\boldsymbol{X}}{b(t)}\in h^{-1}(A)\right]\to\mu(h^{-1}(A))$$

for all measurable sets $A\in\mathcal{A}_h$.

We notice that

$$h\left(\frac{\boldsymbol{X}}{b(t)}\right)=\left(d\left(\frac{\boldsymbol{X}}{b(t)},\mathbb{F}\right),\frac{\boldsymbol{X}}{d(\frac{\boldsymbol{X}}{b(t)},\mathbb{F})}\right)=\left(\frac{d(\boldsymbol{X},\mathbb{F})}{b(t)},\frac{\boldsymbol{X}}{d(\boldsymbol{X},\mathbb{F})}\right),$$

and so $\dfrac{\boldsymbol{X}}{b(t)} \in h^{-1}(A)$ is equivalent to $\left(\dfrac{d(\boldsymbol{X},\mathbb{F})}{b(t)}, \dfrac{\boldsymbol{X}}{d(\boldsymbol{X},\mathbb{F})}\right) \in A$.

It is now enough to show that $\mu(h^{-1}(A)) = c\mu_{(\alpha)} \times S_{\mathbb{O}}(A)$, for all measurable $A \in \mathcal{A}_h$. In fact, it suffices to show that the equality holds for any $A \in \mathcal{A}_h$ of the form $[t,\infty) \times \Lambda$, with $t > 0$ and $\Lambda \subset \partial\aleph_{\mathbb{O}}$.

The scaling property in Proposition 3.2 guarantees the existence of α such that $b \in RV_{\frac{1}{\alpha}}$ and $\mu(\theta B) = \theta^{-\alpha}\mu(B)$ for any suitable set B and positive θ.

For such an A we have

$$\begin{aligned}
\mu(h^{-1}(A)) &= \mu(\{\boldsymbol{X} | h(\boldsymbol{X}) \in A\}) \\
&= \mu(\{\boldsymbol{X} | d(\boldsymbol{X},\mathbb{F}) \geqslant t, \frac{\boldsymbol{X}}{d(\boldsymbol{X},\mathbb{F})} \in \Lambda\}) \\
&= \mu(\{\boldsymbol{X} | d(\boldsymbol{X}/t,\mathbb{F}) \geqslant 1, \frac{\boldsymbol{X}/t}{d(\boldsymbol{X}/t,\mathbb{F})} \in \Lambda\}) \\
&= \mu(\{t\boldsymbol{Y} | d(\boldsymbol{Y},\mathbb{F}) \geqslant 1, \frac{\boldsymbol{Y}}{d(\boldsymbol{Y},\mathbb{F})} \in \Lambda\}) \\
&= t^{-\alpha}\mu(\{\boldsymbol{Y} | d(\boldsymbol{Y},\mathbb{F}) \geqslant 1, \frac{\boldsymbol{Y}}{d(\boldsymbol{Y},\mathbb{F})} \in \Lambda\}) \\
&= c\mu_{(\alpha)}([t,\infty))S_{\mathbb{O}}(\Lambda) \\
&= (c\mu_{(\alpha)} \times S_{\mathbb{O}})(A).
\end{aligned}$$

□

The following corollary of Proposition 3.3 provides a characterization of multivariate regular variation involving the random variable $d(\boldsymbol{X},\mathbb{F})$.

Corollary 3.1. *Let $\boldsymbol{X}$ be a random vector. Regular variation of $\boldsymbol{X}$ in $\mathbb{O}$ as defined in Definition 3.1 is equivalent to the following two conditions:*

1. *The distribution of $d(\boldsymbol{X},\mathbb{F})$ is regularly varying in $(0,\infty)$ and*
2. *The conditional distribution of $\boldsymbol{X}/d(\boldsymbol{X},\mathbb{F})$ given $d(\boldsymbol{X},\mathbb{F}) > t$ converges to the spectral measure, that is,*

$$\mathbb{P}[\boldsymbol{X}/d(\boldsymbol{X},\mathbb{F}) \in \cdot \mid d(\boldsymbol{X},\mathbb{F}) > t] \Rightarrow S_{\mathbb{O}}(\cdot).$$

Proof. Regular variation of $\boldsymbol{X}$ in $\mathbb{O}$ is equivalent to

$$t\mathbb{P}\left[\left(\frac{d(\boldsymbol{X},\mathbb{F})}{b(t)}, \frac{\boldsymbol{X}}{d(\boldsymbol{X},\mathbb{F})}\right) \in A\right] \to c\mu_{(\alpha)} \times S_{\mathbb{O}}(A),$$

for all measurable sets $A \in \mathcal{A}_h$, as in Proposition 3.3.

We note that any set $A \in \mathcal{A}_h$ can be written as the cartesian product $B_1 \times B_2$, with $B_1 \subset (0,\infty)$ and $B_2 \subset \partial\aleph_{\mathbb{O}}$. This means that, for a set $A = B_1 \times B_2 \subset (0,\infty) \times \partial\aleph_{\mathbb{O}}$, we have

$$\begin{aligned} t\mathbb{P}\left[\left(\frac{d(\boldsymbol{X},\mathbb{F})}{b(t)}, \frac{\boldsymbol{X}}{d(\boldsymbol{X},\mathbb{F})}\right) \in A\right] &= t\mathbb{P}\left[\frac{d(\boldsymbol{X},\mathbb{F})}{b(t)} \in B_1, \frac{\boldsymbol{X}}{d(\boldsymbol{X},\mathbb{F})} \in B_2\right] \\ &= t\mathbb{P}\left[\frac{d(\boldsymbol{X},\mathbb{F})}{b(t)} \in B_1\right] \mathbb{P}\left[\frac{\boldsymbol{X}}{d(\boldsymbol{X},\mathbb{F})} \in B_2 \middle| \frac{d(\boldsymbol{X},\mathbb{F})}{b(t)} \in B_1\right] \end{aligned}$$

Suppose the distribution of X has regular variation as in Definition 3.1, with $b \in RV_{\frac{1}{\alpha}}$. Then, for a set $A = B_1 \times B_2 \subset (0,\infty) \times \partial\aleph_{\mathbb{O}}$, we have

$$t\mathbb{P}\left[\frac{d(\boldsymbol{X},\mathbb{F})}{b(t)} \in B_1\right] \mathbb{P}\left[\frac{\boldsymbol{X}}{d(\boldsymbol{X},\mathbb{F})} \in B_2 \middle| \frac{d(\boldsymbol{X},\mathbb{F})}{b(t)} \in B_1\right] \to c\mu_{(\alpha)}(B_1) S_{\mathbb{O}}(B_2),$$

as $t \to \infty$.

We now prove that the one-dimensional random variable $d(\boldsymbol{X},\mathbb{F})$ has regular variation. Let $x \in \mathbb{R}$. Then, for $t > 0$ and a set $B_2 \subset \partial\aleph_{\mathbb{O}}$, we have

$$\begin{aligned} &\frac{\mathbb{P}\left[\frac{d(\boldsymbol{X},\mathbb{F})}{b(t)} > x\right]}{\mathbb{P}\left[\frac{d(\boldsymbol{X},\mathbb{F})}{b(t)} > 1\right]} \\ &= \frac{t\mathbb{P}\left[\frac{d(\boldsymbol{X},\mathbb{F})}{b(t)} > x\right] \mathbb{P}\left[\frac{\boldsymbol{X}}{d(\boldsymbol{X},\mathbb{F})} \in B_2 \middle| \frac{d(\boldsymbol{X},\mathbb{F})}{b(t)} > x\right] \mathbb{P}\left[\frac{\boldsymbol{X}}{d(\boldsymbol{X},\mathbb{F})} \in B_2 \middle| \frac{d(\boldsymbol{X},\mathbb{F})}{b(t)} > 1\right]}{t\mathbb{P}\left[\frac{d(\boldsymbol{X},\mathbb{F})}{b(t)} > 1\right] \mathbb{P}\left[\frac{\boldsymbol{X}}{d(\boldsymbol{X},\mathbb{F})} \in B_2 \middle| \frac{d(\boldsymbol{X},\mathbb{F})}{b(t)} > 1\right] \mathbb{P}\left[\frac{\boldsymbol{X}}{d(\boldsymbol{X},\mathbb{F})} \in B_2 \middle| \frac{d(\boldsymbol{X},\mathbb{F})}{b(t)} > x\right]}. \end{aligned}$$

Note that

$$\frac{d(\boldsymbol{X},\mathbb{F})}{b(t)} > x \iff \frac{d(\boldsymbol{X},\mathbb{F})}{b(t)} \frac{b(tx^\alpha)}{b(tx^\alpha)} > x \iff \frac{d(\boldsymbol{X},\mathbb{F})}{b(tx^\alpha)} \frac{b(tx^\alpha)}{b(t)} > x.$$

Since $b \in RV_{\frac{1}{\alpha}}$, we have

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

and so

$$\frac{\mathbb{P}\left[\frac{\boldsymbol{X}}{d(\boldsymbol{X},\mathbb{F})} \in B_2 \middle| \frac{d(\boldsymbol{X},\mathbb{F})}{b(t)} > 1\right]}{\mathbb{P}\left[\frac{\boldsymbol{X}}{d(\boldsymbol{X},\mathbb{F})} \in B_2 \middle| \frac{d(\boldsymbol{X},\mathbb{F})}{b(t)} > x\right]} \to 1 \text{ as } t \to \infty.$$

Then, since $\boldsymbol{X}$ has regular variation, we have

$$\lim_{t\to\infty}\frac{\mathbb{P}[d(\boldsymbol{X},\mathbb{F})>tx]}{\mathbb{P}[d(\boldsymbol{X},\mathbb{F})>t]}=\lim_{t\to\infty}\frac{\mathbb{P}\left[\frac{d(\boldsymbol{X},\mathbb{F})}{b(t)}>x\right]}{\mathbb{P}\left[\frac{d(\boldsymbol{X},\mathbb{F})}{b(t)}>1\right]}=\frac{c\mu_{(\alpha)}((x,\infty))S_{\mathbb{O}}(B_2)}{c\mu_{(\alpha)}((1,\infty))S_{\mathbb{O}}(B_2)}=x^{-\alpha},$$

which proves that $d(\boldsymbol{X},\mathbb{F})$ has regular variation. In other words, we have

$$t\mathbb{P}\left[\frac{d(\boldsymbol{X},\mathbb{F})}{b(t)}\in\cdot\right]\to c\mu_{(\alpha)}(\cdot),\ \text{ as }\ t\to\infty,$$

and so, by Proposition 2.1, we conclude that

$$\mathbb{P}[\boldsymbol{X}/d(\boldsymbol{X},\mathbb{F})\in\cdot\mid d(\boldsymbol{X},\mathbb{F})>t]\Rightarrow S_{\mathbb{O}}(\cdot).$$

Suppose that the distribution of $d(\boldsymbol{X},\mathbb{F})$ is regularly varying in $(0,\infty)$, that is, there is $\alpha\in\mathbb{R}$ such that $x\mapsto\mathbb{P}[d(\boldsymbol{X},\mathbb{F})>x]\in RV_{-\alpha}$, and we have

$$\mathbb{P}[\boldsymbol{X}/d(\boldsymbol{X},\mathbb{F})\in\cdot\mid d(\boldsymbol{X},\mathbb{F})>t]\Rightarrow S_{\mathbb{O}}(\cdot).$$

Then, for a set $A=B_1\times B_2\subset(0,\infty)\times\partial\aleph_{\mathbb{O}}$, we have, as $t\to\infty$, that

$$t\mathbb{P}\left[\frac{d(\boldsymbol{X},\mathbb{F})}{b(t)}\in B_1\right]\mathbb{P}\left[\frac{\boldsymbol{X}}{d(\boldsymbol{X},\mathbb{F})}\in B_2\middle|\frac{d(\boldsymbol{X},\mathbb{F})}{b(t)}\in B_1\right]\to c\mu_{(\alpha)}(B_1)S_{\mathbb{O}}(B_2),$$

which is equivalent to $\boldsymbol{X}$ having regular variation in $\mathbb{O}$ by Proposition 3.3. □

4 Hidden regular variation

There are cases when looking solely for regular variation is not enough. It may happen that, in the limit, the multivariate distribution may concentrate on more than one cone, and, if we limit ourselves to the regular variation described in the previous chapter, we are ignoring the smaller cones, since Definition 3.1 only concerns the largest cone. These other cones, although negligible according to that definition, might still play a role in applications of the theory. For instance, in an insurance setting, the cone associated to the second highest tail index might correspond to the second most likely catastrophic scenario, which is of interest to an insurance company. Thus, it is important to look for regular variation on a second cone, which leads us to the concept of hidden regular variation. It corresponds to the second order asymptotic behavior of tail functions in the setting of multivariate regular variation.

The idea is similar to that of regular variation: before, we had a closed cone $\mathbb{F} \subset \mathbb{C}$ containing the origin and we defined regular variation on $\mathbb{O} = \mathbb{C}\backslash\mathbb{F}$; now, we take a second closed cone $\mathbb{F}_1 \subset \mathbb{C}$ and we define hidden regular variation on $\mathbb{O}_1 = \mathbb{C}\backslash(\mathbb{F} \cup \mathbb{F}_1)$. For these new cones $\mathbb{F} \cup \mathbb{F}_1$ and $\mathbb{O}_1$ we consider the corresponding sets and σ-algebras as we did before for the cones $\mathbb{F}$ and $\mathbb{O}$. For example, we define the σ-algebra of $\mathbb{O}_1$ as $\mathcal{O}_1 = \{B \subset \mathbb{O}_1 \colon B \in \mathcal{C}\}$.

Definition 4.1. ***(Hidden Regular Variation)*** *Let $\mathbb{C}$ and $\mathbb{F}$ be closed cones containing the origin such that $\mathbb{F} \subset \mathbb{C} \subset [0,\infty)^d$ and let $\mathbb{F}_1 \subset \mathbb{C}$ be another closed cone containing the origin and different than $\mathbb{F}$. Denote $\mathbb{O} = \mathbb{C}\backslash\mathbb{F}$ and $\mathbb{O}_1 = \mathbb{C}\backslash(\mathbb{F} \cup \mathbb{F}_1)$. Let $\boldsymbol{X}$ be a random vector in $\mathbb{R}^d$ whose distribution has regular variation on $\mathbb{O}$ as in Definition 3.1. The distribution of $\boldsymbol{X}$ has* hidden regular variation *on $\mathbb{O}_1$ if:*

1. *The distribution of $\boldsymbol{X}$ has regular variation on $\mathbb{O}_1$ as in Definition 3.1, with scaling function b_1 and limit measure μ_1, and*
2. *The functions b and b_1 satisfy $\lim_{t\to\infty} \dfrac{b(t)}{b_1(t)} = \infty$.*

As it was the case in the previous chapter, hidden regular variation also has a scaling property.

Proposition 4.1. ***(Scaling Property of Hidden Regular Variation)*** *Let $\boldsymbol{X}$ be a non-negative random vector in $\mathbb{R}^d$ whose distribution has hidden regular variation according to Definition 4.1. Then there is a number $\alpha_1 > \alpha$ such that $b_1 \in RV_{\frac{1}{\alpha}}$ and μ_1 satisfies, for $\theta > 0$,*

$$\mu_1(\theta B) = \theta^{-\alpha_1}\mu_1(B),$$

for all sets $B \in \mathcal{O}_1$ such that $\mu_1(\partial B) = 0$ and $\overline{B} \cap \overline{\mathbb{F} \cup \mathbb{F}_1} = \varnothing$.

Proof. Let $\theta > 0$ be fixed. From Definition 4.1, the distribution of $\boldsymbol{X}$ has regular variation on $\mathbb{O}_1$ as in Definition 3.1, with scaling function b_1 and limit measure μ_1. By taking α_1 such that $b_1 \in RV_{\frac{1}{\alpha}}$, we have, by Proposition 3.2,

$$\mu_1(\theta B) = \theta^{-\alpha_1}\mu_1(B),$$

for any set $B \in \mathcal{O}_1$ such that $\mu_1(\partial B) = 0$ and $\overline{B} \cap \overline{\mathbb{F} \cup \mathbb{F}_1} = \varnothing$. □

Example 4.1. *We return to Example 3.1, where we had considered a two-dimensional vector $\boldsymbol{X} = (X^1, X^2)$ such that X^1 and X^2 are independent and identically distributed unit Pareto random variables, that is, they have common tail function $\overline{F}$ given by $\overline{F}(x) = \dfrac{1}{x}$, for $x > 0$.*

Recall that $\boldsymbol{X}$ has multivariate regular variation on $\mathbb{O} = \mathbb{C}\backslash\mathbb{F}$, where $\mathbb{C} = [0,\infty)^2$ and $\mathbb{F} = \{\mathbf{0}\}$. Now we want to prove that $\boldsymbol{X}$ also has hidden regular variation on a different cone $\mathbb{O}_1$ with scaling function b_1.

Following the steps of Example 3.1, it suffices to consider the limit of $t\mathbb{P}\left[\dfrac{\boldsymbol{X}}{b_1(t)} \in [\mathbf{0},\boldsymbol{x}]^c\right]$, as $t \to \infty$, for $\boldsymbol{x} \in \mathbb{O}_1$. For the scaling function b_1 we take $b_1(t) = \sqrt{t}$, and so we have, for $\boldsymbol{x} = (x^1, x^2) \in \mathbb{O}_1$,

$$\begin{aligned} t\mathbb{P}\left[\frac{\boldsymbol{X}}{b_1(t)} \in [\mathbf{0},\boldsymbol{x}]^c\right] &= t\overline{F}(b_1(t)x^1) + t\overline{F}(b_1(t)x^2) - t\overline{F}(b_1(t)x^1)\overline{F}(b_1(t)x^2) \\ &= \frac{t}{b_1(t)x^1} + \frac{t}{b_1(t)x^2} - \frac{t}{b_1(t)x^1}\frac{1}{b_1(t)x^2} \\ &= \sqrt{t}\left(\frac{1}{x^1} + \frac{1}{x^2}\right) - \frac{1}{x^1x^2}. \end{aligned}$$

We notice that $\mu([\mathbf{0},\boldsymbol{x}]^c) = \dfrac{1}{x^1} + \dfrac{1}{x^2}$ and so we exclude that part to get $\mu_1([\mathbf{0},\boldsymbol{x}]^c) = \dfrac{1}{x^1x^2}$. We do that since the function b_1 is dominated by the function b (in the sense that $\lim\limits_{t\to\infty} b(t)/b_1(t) = \infty$) and so the hidden cone is dominated by the larger one. The reason the function $b_1(t) = \sqrt{t}$ was appropriate is because this function satisfies $t/b_1^2(t) = 1$, which was what we wanted in the expression of $t\mathbb{P}\left[\dfrac{\boldsymbol{X}}{b_1(t)} \in [\mathbf{0},\boldsymbol{x}]^c\right]$.

We thus conclude that the random vector $\boldsymbol{X}$ has hidden regular variation on $\mathbb{O}_1 = (0,\infty)^2$ with scaling function $b_1(t) = \sqrt{t}$ and limit measure μ_1 given by $\mu_1([\mathbf{0},\boldsymbol{x}]^c) = \dfrac{1}{x^1x^2}$, for $\boldsymbol{x} = (x^1, x^2) \in \mathbb{O}_1$. More discussion of this example and many others is to be found at ([4], Section 3, pp. 12-13, and Section 5, pp. 18-23).

There are not many results about where to look for hidden regular variation. We should note, however, the following observation.

Proposition 4.2. *Let $\boldsymbol{X}$ be a random vector whose distribution has regular variation as in Definition 3.1. Then, the support of the limit measure μ is a closed cone $\mathbb{F}_\mu \subset \mathbb{C}$ containing the origin.*

Proof. Let $supp(\mu)$ denote the support of the limit measure μ. By definition of support, $supp(\mu)$ is closed and so we only need to show that, if $\boldsymbol{x} \in supp(\mu)$, then $t\boldsymbol{x} \in supp(\mu)$, for $t > 0$.

For small $\delta > 0$, we have, due to the scaling property,

$$\mu\left((t\boldsymbol{x} - \delta\mathbf{1}, t\boldsymbol{x} + \delta\mathbf{1}) \cap \mathbb{C}\right) = t^{-\alpha}\mu\left(\left(\boldsymbol{x} - \frac{\delta}{t}\mathbf{1}, t\boldsymbol{x} + \frac{\delta}{t}\mathbf{1}\right) \cap \mathbb{C}\right) > 0$$

because $\boldsymbol{x} \in supp(\mu)$. $\square$

Note that Definition 4.1 can be extended by considering another closed cone $\mathbb{F}_2$ containing the origin and the open cone $\mathbb{O}_2 = \mathbb{C}\backslash(\mathbb{F} \cup \mathbb{F}_1 \cup \mathbb{F}_2)$. This introduces a second hidden regular variation, with a scaling function b_2 and a second hidden limit measure μ_2. We can do this recursively to define as many hidden measures as we want. However, practice shows that from the third hidden regular variation onward, the cones where one could find some hidden measure are negligible and add almost nothing to our knowledge of the tail behaviour of $\boldsymbol{X}$.

5 Parameter estimation

In the previous chapters we have defined multivariate regular variation on a cone $\mathbb{O} \subset [0,\infty)^d$ and hidden regular variation on a second cone $\mathbb{O}_1 \subset \mathbb{O}$. We now explore some ways to estimate parameters associated with multivariate regular variation and hidden regular variation. More precisely, we will find estimators for the tail index of a regularly varying distribution, as well as for the support of its spectral measure.

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be i.i.d. (independent and identically distributed) random vectors in $[0,\infty)^d$ whose common distribution function has regular variation as in Definition 3.1 with limit measure μ. Let $\boldsymbol{X}$ be a random vector with the same distribution as $\boldsymbol{X}_1$.

By Corollary 3.1, the random variable $d(\boldsymbol{X}, \mathbb{F})$ is regularly varying as in Definition 3.1 with some exponent $\alpha > 0$ and scaling function b, which we may take as $b(t) = \overline{F}_d^{-1}(1/t)$, where F_d is the distribution function of $d(\boldsymbol{X}, \mathbb{F})$ and

$$\mathbb{P}\left[\frac{\boldsymbol{X}}{d(\boldsymbol{X},\mathbb{F})} \in \cdot \,\middle|\, d(\boldsymbol{X},\mathbb{F}) > t\right] \Rightarrow S_{\mathbb{O}}(\cdot), \text{ as } t \to \infty \tag{5.1}$$

in the class of all probability measures on $\partial\aleph_{\mathbb{O}} = \{\boldsymbol{x} \in \mathbb{O} : d(\boldsymbol{x},\mathbb{F}) = 1\}$, which we denote by $P(\partial\aleph_{\mathbb{O}})$.

We estimate the limit measure μ by using the connection present in Proposition 3.3 which states that the limit measure can be written as a product measure. Thus we can estimate the tail index α and the spectral measure $S_{\mathbb{O}}$ separately.

Since $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ are i.i.d. random vectors, we may take $d_1, \ldots, d_n$ as i.i.d. samples from a regular varying distribution on $(0,\infty)$, with $d_i = d(\boldsymbol{X}_i, \mathbb{F})$, for $1 \leqslant i \leqslant n$.

In this thesis we will not discuss possible estimators for the tail index α. There are several ways to estimate that parameter, namely through the Hill, Pickands or QQ estimators, all of which are analyzed in ([14], Section 4).

The following proposition gives an estimator for the spectral measure $S_{\mathbb{O}}$.

Proposition 5.1. ***(Estimator of*** $S_{\mathbb{O}}$***)*** *Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be i.i.d. random vectors whose common distribution function is regularly varying as in Definition 3.1. Let $\{k(n)\}$ be a sequence of natural numbers such that $k = k(n) \to \infty$ and $n/k \to \infty$ as $n \to \infty$. Then, we have that, in $P(\partial\aleph_{\mathbb{O}})$,*

$$S_n(\cdot) = \frac{\sum_{i=1}^{n} \epsilon_{(\frac{d_i}{b(n/k)}, \frac{\boldsymbol{X}_i}{d_i})}((1,\infty) \times \cdot)}{\sum_{i=1}^{n} \epsilon_{\frac{d_i}{b(n/k)}}(1,\infty)} \Rightarrow S_{\mathbb{O}}(\cdot).$$

The proof of Proposition 5.1, as well as all the auxiliary results needed to obtain it, are presented at the end of this chapter.

The problem with Proposition 5.1 is that it depends on us knowing $b(t)$, which is not generally true. We can, however, estimate $b(n/k)$ by replacing it with an estimator which can be calculated from the first n values of the sequence $d_1, d_2, \ldots$. To do that, we order the values $d_1, \ldots, d_n$ in a decreasing sequence $d_{(1)} \geqslant \cdots \geqslant d_{(n)}$. We want to show that $d_{(k+1)}$ is a consistent estimator of $b(n/k)$ as $n \to \infty$, $k \to \infty$ and $n/k \to \infty$. We start by presenting the definition of a consistent estimator.

Definition 5.1. ***(Consistent Estimator)*** *Let $\theta \in \mathbb{R}\backslash\{0\}$ and $X(n)$ be a random variable with values in $\mathbb{R}$. We say that $X(n)$ is a* consistent estimator *of parameter θ if, for all $\epsilon > 0$,*

$$\mathbb{P}\left[\left|\frac{X(n)}{\theta} - 1\right| > \epsilon\right] \to 0, \tag{5.2}$$

as $n \to \infty$.

Remark 5.1. *If the parameter θ is treated as a random variable in Definition 5.1, then condition (5.2) is equivalent to say that the random variable $X(n)$ converges in probability to θ.*

We now show that $d_{(k+1)}$ is a consistent estimator of $b(n/k)$ according to Definition 5.1.

Proposition 5.2. *Under the assumptions of Proposition 5.1, $d_{(k+1)}$ is a consistent estimator of $b(n/k)$.*

Proof. Let $\epsilon > 0$. We have

$$\begin{aligned}\mathbb{P}\left[\left|\frac{d_{(k+1)}}{b(n/k)} - 1\right| > \epsilon\right] &= \mathbb{P}\left[\frac{d_{(k+1)}}{b(n/k)} > 1+\epsilon\right] + \mathbb{P}\left[\frac{d_{(k+1)}}{b(n/k)} < 1-\epsilon\right] \\ &\leqslant \mathbb{P}\left[\frac{1}{k}\sum_{i=1}^{n} \epsilon_{\frac{d_i}{b(\frac{n}{k})}}((1+\epsilon,\infty)) \geqslant 1\right] \\ &+ \mathbb{P}\left[\frac{1}{k}\sum_{i=1}^{n} \epsilon_{\frac{d_i}{b(\frac{n}{k})}}([1-\epsilon,\infty)) < 1\right].\end{aligned}$$

By Proposition 5.6, because $d(\boldsymbol{X}, \mathbb{F})$ is regularly varying, we have that

$$\frac{1}{k}\sum_{i=1}^{n} \epsilon_{\frac{d_i}{b(\frac{n}{k})}} \Rightarrow \mu_{(\alpha)},$$

where $\mu_{(\alpha)}$ is the Pareto measure given by $\mu_{(\alpha)}((x,\infty)) = x^{-\alpha}$, for $x > 0$.

By Proposition 2.1, we have

$$\lim_{n\to\infty} \frac{1}{k}\sum_{i=1}^{n} \epsilon_{\frac{d_i}{b(\frac{n}{k})}}((1+\epsilon,\infty)) = (1+\epsilon)^{-\alpha}$$

and

$$\lim_{n\to\infty} \frac{1}{k}\sum_{i=1}^{n} \epsilon_{\frac{d_i}{b(\frac{n}{k})}}([1-\epsilon,\infty)) = (1-\epsilon)^{-\alpha}.$$

We notice that, for $\epsilon > 0$, $(1+\epsilon)^{-\alpha} < 1$ and $(1-\epsilon)^{-\alpha} > 1$. This fact, coupled with the continuous mapping theorem, leads us to conclude that

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{1}{k}\sum_{i=1}^{n} \epsilon_{\frac{d_i}{b(\frac{n}{k})}}((1+\epsilon,\infty)) \geqslant 1\right] = 0$$

and

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{1}{k}\sum_{i=1}^{n} \epsilon_{\frac{d_i}{b(\frac{n}{k})}}([1-\epsilon,\infty)) < 1\right] = 0.$$

which ends the proof. □

Thus, $d_{(k+1)}$ a consistent estimator of $b(n/k)$ as $n \to \infty$ and so we can replace $b(n/k)$ by $d_{(k+1)}$ in S_n of Proposition 5.1 to consider $\hat{S}_n$, defined as

$$\hat{S}_n(\cdot) = \frac{\sum_{i=1}^{n} \epsilon_{(\frac{d_i}{d_{(k+1)}}, \frac{X_i}{d_i})}((1,\infty)\times\cdot)}{\sum_{i=1}^{n} \epsilon_{\frac{d_i}{d_{(k+1)}}}(1,\infty)} = \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}\left[d_i/d_{(k+1)} > 1\right] \epsilon_{X_i/d_i}(\cdot).$$

Proposition 5.3 states that $\hat{S}_n$ can be used as an estimator of the spectral measure $S_{\mathbb{O}}$, replacing S_n.

Proposition 5.3. *Under the assumptions of Proposition 5.1, we have that, in* $P(\partial\aleph_{\mathbb{O}})$*,*

$$\hat{S}_n(\cdot) \Rightarrow S_{\mathbb{O}}(\cdot).$$

Proof. By replacing $b(n/k)$ with $d_{(k+1)}$ in Proposition 5.1, we get the desired result.

In order to do that, we need to show that $d_{(k+1)}$ satisfies the condition of regular variation that is satisfied by $b(n/k)$, which is

$$\frac{n}{k}\mathbb{P}\left[\frac{d(\boldsymbol{X},\mathbb{F})}{b(n/k)} > x\right] \to x^{-\alpha},$$

for $x > 0$, as $n \to \infty$.

Thus, we need to show that, for positive x,

$$\frac{n}{k}\mathbb{P}\left[\frac{d(\boldsymbol{X},\mathbb{F})}{d_{(k+1)}} > x\right] \to x^{-\alpha}.$$

Let $x > 0$. Then

$$\begin{aligned}\frac{n}{k}\mathbb{P}\left[\frac{d(\boldsymbol{X},\mathbb{F})}{d_{(k+1)}} > x\right] =& \frac{n}{k}\mathbb{P}\left[\frac{d(\boldsymbol{X},\mathbb{F})}{d_{(k+1)}} > x, \frac{d_{(k+1)}}{b(n/k)} = 1\right] \\ &+ \frac{n}{k}\mathbb{P}\left[\frac{d(\boldsymbol{X},\mathbb{F})}{d_{(k+1)}} > x, \frac{d_{(k+1)}}{b(n/k)} \neq 1\right].\end{aligned}$$

By Proposition 5.2, we have that

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{d_{(k+1)}}{b(n/k)} = 1\right] = 1,$$

and so

$$\lim_{n\to\infty} \frac{n}{k}\mathbb{P}\left[\frac{d(\boldsymbol{X},\mathbb{F})}{d_{(k+1)}} > x\right] = \lim_{n\to\infty} \frac{n}{k}\mathbb{P}\left[\frac{d(\boldsymbol{X},\mathbb{F})}{b(n/k)} > x\right] = x^{-\alpha}.$$

□

5.1 Support estimation

In the same way as before, let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be i.i.d. random vectors in $[0,\infty)^d$ whose common distribution function has regular variation as in Definition 3.1 with scaling function b and limit measure μ. Let $\boldsymbol{X}$ be a random vector with the same distribution as $\boldsymbol{X}_1$.

If we cannot identify a cone on which to seek multivariate regular variation, it would be challenging to apply the previous results directly. Even though the estimator from Proposition 5.3 can be used to estimate the entire spectral measure and then be used again for the hidden spectral measures, it is required to know what are the cones $\mathbb{O}_1$, $\mathbb{F}_1$, $\mathbb{O}_2$, and so on. Instead, we turn to estimation of the support of the limit measure μ, denoted by $supp(\mu)$. Due to the fact that

$$supp(\mu) = \{t \cdot supp(S_{\mathbb{O}}), t \geqslant 0\},$$

it suffices to estimate the support of $S_{\mathbb{O}}$. The following Proposition allows us to find an estimator of $supp(\mu)$.

Proposition 5.4. *Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be i.i.d. random vectors whose common distribution function is regularly varying as in Definition 3.1 with scaling function b and limit measure μ. Let $\{k(n)\}$ be a sequence of natural numbers such that $k = k(n) \to \infty$ and $n/k \to \infty$ as $n \to \infty$. Then, we have*

$$supp_{k,n} = \left\{ \frac{\boldsymbol{X}_i}{d_i} : d_i > d_{(k+1)}, i = 1, \ldots, n \right\} \Rightarrow supp(S_{\mathbb{O}}).$$

Proof. See ([10], Section 2, pp. 8-19). □

Proposition 5.4 provides a way to estimate $supp(S_{\mathbb{O}})$ and hence $supp(\mu)$. The only information we need to have to use this estimator concerns the cones $\mathbb{F}$ and $\mathbb{O}$. By knowing the (estimated) support of the limit measure, we can, at least theoretically, remove it from $\mathbb{O}$ and look in the remaining set for hidden regular variation. If our sample data is very large, we could know the support of the limit measure with such a level of detail that we could have some good notion of where to check for further hidden regular variation.

5.2 Estimator for the spectral measure

In this section we prove Proposition 5.1, which presents an estimator for the spectral measure $S_{\mathbb{O}}$. The main step required to that proof is making the bridge between weak convergence and $\mathbb{M}^*$-convergence (recall the convergence definitions in chapter 2). In order to do that, we first establish two lemmas, that make it easy to compare these different kinds of convergence.

We recall that $\mathcal{C}_{\mathbb{F}}$ is, following from Definition 2.5, the set of all bounded, continuous real-valued functions f on $\mathcal{C}$ for which there is a $r > 0$ such that $f(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in \mathbb{C}$ satisfying $d(\boldsymbol{x}, \mathbb{F}) < r$.

Lemma 5.1. *Let $\{\mu_n\}$ be a sequence of measures on $\mathbb{M}^*(\mathbb{C},\mathbb{O})$ and μ a measure also on $\mathbb{M}^*(\mathbb{C},\mathbb{O})$, as in Definition 2.5. We have*

$$\mu_n \Rightarrow \mu \text{ iff } \mathbb{E}(e^{-\mu_n(f)}) \to \mathbb{E}(e^{-\mu f}),$$

for all $f \in \mathcal{C}_{\mathbb{F}}$.

Proof. ($\Rightarrow$) Suppose $\mu_n \Rightarrow \mu$ and let f be a function in $\mathcal{C}_{\mathbb{F}}$. The map $\nu \mapsto e^{-\nu(f)}$ defined in $\mathbb{M}^*(\mathbb{C},\mathbb{O}) \mapsto [0,\infty)$ is continuous, and so by the continuous mapping theorem we have $e^{-\mu_n(f)} \Rightarrow e^{-\mu f}$. By Lebesgue's Dominated Convergence Theorem, we have $\mathbb{E}(e^{-\mu_n(f)}) \to \mathbb{E}(e^{-\mu f})$.

($\Leftarrow$) See ([14], Section 5, pp. 137-138). □

Now we can connect vague convergence with $\mathbb{M}^*$-convergence, as illustrated by Proposition 5.5.

Proposition 5.5. *Suppose that, for each $n \geqslant 1$, we have that $\{Z_{n,j}, j = 1,\ldots,n\}$ is a sequence of i.i.d. random variables. Suppose also that $\{a_n, n \geqslant 1\}$ is a positive sequence such that $\lim_{n\to\infty} a_n = \infty$. Then, for a measure $\mu \in \mathbb{M}^*(\mathbb{C},\mathbb{O})$, we have*

$$\frac{1}{a_n}\sum_{j=1}^{n} \epsilon_{Z_{n,j}} \Rightarrow \mu \text{ iff } \frac{n}{a_n}\mathbb{P}[Z_{n,1} \in \cdot] \xrightarrow{*} \mu(\cdot).$$

Proof. Using Lemma 5.1, we have that

$$\frac{1}{a_n}\sum_{j=1}^{n} \epsilon_{Z_{n,j}} \Rightarrow \mu$$

is equivalent to

$$\mathbb{E}(e^{-\frac{1}{a_n}\sum_{j=1}^{n}\epsilon_{Z_{n,j}}(f)}) \to \mathbb{E}(e^{-\mu f}),$$

for all $f \in \mathcal{C}_{\mathbb{F}}$.

We then compute

$$\begin{aligned}\mathbb{E}(e^{-\frac{1}{a_n}\sum_{j=1}^{n}\epsilon_{Z_{n,j}}(f)}) &= (\mathbb{E}(e^{-\frac{1}{a_n}f(Z_{n,1})}))^n \\ &= \left(1 - \frac{\int_{\mathbb{E}}(1-e^{-\frac{1}{a_n}f(x)})n\mathbb{P}[Z_{n,1}\in dx]}{n}\right)^n,\end{aligned}$$

which converges to $e^{-\mu f}$ if and only if

$$\int_{\mathbb{E}}(1-e^{-\frac{1}{a_n}f(x)})n\mathbb{P}[Z_{n,1}\in dx] \to \mu f. \tag{5.3}$$

We now show that (5.3) being valid for all $f \in \mathcal{C}_{\mathbb{F}}$ is equivalent to

$$\frac{n}{a_n}\mathbb{P}[Z_{n,1} \in \cdot] \xrightarrow{*} \mu(\cdot), \tag{5.4}$$

which finishes the proof.

We first note that, for positive y, we have $y - \frac{y^2}{2} \leqslant 1 - e^{-y} \leqslant y$, and so, for all $f \in \mathcal{C}_{\mathbb{F}}$,

$$\frac{f}{a_n} - \frac{f^2}{2a_n^2} \leqslant 1 - e^{-f/a_n} \leqslant \frac{f}{a_n}.$$

Thus, we also have

$$\begin{aligned}
&\int_{\mathbb{E}} f(x)\frac{n}{a_n}\mathbb{P}[Z_{n,1} \in dx] - \int_{\mathbb{E}} \frac{f^2(x)}{2a_n}\frac{n}{a_n}\mathbb{P}[Z_{n,1} \in dx] \\
&\qquad \leqslant \int_{\mathbb{E}} (1 - e^{-\frac{1}{a_n}f(x)})n\mathbb{P}[Z_{n,1} \in dx] \\
&\qquad \leqslant \int_{\mathbb{E}} f(x)\frac{n}{a_n}\mathbb{P}[Z_{n,1} \in dx].
\end{aligned}$$

((5.4)⇒(5.3)) Suppose (5.4) holds. We have, by Theorem 2.1,

$$\int_{\mathbb{E}} f(x)\frac{n}{a_n}\mathbb{P}[Z_{n,1} \in dx] \to \mu f.$$

and, since $a_n \to \infty$,

$$\int_{\mathbb{E}} \frac{f^2(x)}{2a_n}\frac{n}{a_n}\mathbb{P}[Z_{n,1} \in dx] = \frac{1}{2a_n}\int_{\mathbb{E}} f^2(x)\frac{n}{a_n}\mathbb{P}[Z_{n,1} \in dx] \sim \frac{\mu(f^2)}{2a_n} \to 0.$$

We conclude that

$$\lim_{n\to\infty} \int_{\mathbb{E}} (1 - e^{-\frac{1}{a_n}f(x)})n\mathbb{P}[Z_{n,1} \in dx] = \mu f.$$

((5.3)⇒(5.4)) Suppose (5.3) holds and let $f \in \mathcal{C}_{\mathbb{F}}$, such that $f \leqslant 1$. We have

$$\int_{\mathbb{E}} (1 - e^{-\frac{1}{a_n}f(x)})n\mathbb{P}[Z_{n,1} \in dx] \leqslant \int_{\mathbb{E}} f(x)\frac{n}{a_n}\mathbb{P}[Z_{n,1} \in dx],$$

which, since (5.3) holds, leads to

$$\liminf_{n\to\infty} \int_{\mathbb{E}} f(x)\frac{n}{a_n}\mathbb{P}[Z_{n,1} \in dx] \geqslant \mu f.$$

On the other hand,

$$\int_{\mathbb{E}} f(x)\frac{n}{a_n}\mathbb{P}[Z_{n,1}\in dx] - \int_{\mathbb{E}} \frac{f^2(x)}{2a_n}\frac{n}{a_n}\mathbb{P}[Z_{n,1}\in dx] \leqslant \int_{\mathbb{E}} (1-e^{-\frac{1}{a_n}f(x)})n\mathbb{P}[Z_{n,1}\in dx],$$

leading to

$$\limsup_{n\to\infty}\int_{\mathbb{E}}(f(x)-\frac{f^2(x)}{2a_n})\frac{n}{a_n}\mathbb{P}[Z_{n,1}\in dx]\leqslant \mu f.$$

We get (5.4) by once again using the fact that

$$\int_{\mathbb{E}} \frac{f^2(x)}{2a_n}\frac{n}{a_n}\mathbb{P}[Z_{n,1}\in dx]\to 0.$$

□

We may now present an equivalent definition for multivariate regular variation.

Proposition 5.6. *Under the assumptions of Definition 3.1, we have*

$$t\mathbb{P}\left[\frac{\boldsymbol{X}}{b(t)}\in\cdot\right]\xrightarrow{*}\mu(\cdot) \text{ iff } \frac{1}{k}\sum_{i=1}^{n}\epsilon_{\frac{\boldsymbol{X}_i}{b(\frac{n}{k})}}\Rightarrow\mu,$$

where $\{\boldsymbol{X},\boldsymbol{X_1},\boldsymbol{X_2},\dots\}$ *are independent and identically distributed and* $k=k(n)$ *is a sequence such that* $k(n)\to\infty$, $n/k\to\infty$ *and* $k(n)\sim k(n+1)$.

Proof. Define $\boldsymbol{X}_{n,j}=\dfrac{\boldsymbol{X}_j}{b(\frac{n}{k(n)})}$. Then, we have

$$\begin{aligned}
t\mathbb{P}\left[\frac{\boldsymbol{X}}{b(t)}\in\cdot\right]\xrightarrow{*}\mu(\cdot) &\overset{1}{\iff} n\mathbb{P}\left[\frac{\boldsymbol{X}}{b_n}\in\cdot\right]\xrightarrow{*}\mu(\cdot)\\
&\overset{2}{\iff} \frac{n}{k(n)}\mathbb{P}\left[\frac{\boldsymbol{X}}{b(n/k(n))}\in\cdot\right]\xrightarrow{*}\mu(\cdot)\\
&\iff \frac{n}{k}\mathbb{P}\left[\boldsymbol{X}_{n,1}\in\cdot\right]\xrightarrow{*}\mu(\cdot)\\
&\overset{3}{\iff} \frac{1}{k}\sum_{i=1}^{n}\epsilon_{\boldsymbol{X}_{n,1}}\Rightarrow\mu\\
&\iff \frac{1}{k}\sum_{i=1}^{n}\epsilon_{\frac{\boldsymbol{X}_i}{b(\frac{n}{k})}}\Rightarrow\mu,
\end{aligned}$$

where

1. $\overset{1}{\iff}$ is obtained through Proposition 3.1 (equivalent definitions of multivariate regular variation),

2. $\overset{2}{\iff}$ is due to the fact that $n/k \to \infty$,

3. $\overset{3}{\iff}$ comes from application of Proposition 5.5.

□

We are now ready to prove Propositions 5.1.

Proof of Proposition 5.1. We begin by noticing that, since $\{\boldsymbol{X}, \boldsymbol{X_1}, \boldsymbol{X_2}, \dots\}$ are independent and identically distributed regularly varying random variables, so are the distances $d_1, d_2, \dots$ (recall that $d_i = d(\boldsymbol{X_i}, \mathbb{F})$).

These distances $d_1, d_2, \dots$ have common distribution F_d on $(0, \infty)$ with normalizing function b, which can be chosen to be $b(t) = \overline{F}_d^{-1}(1/t)$. Thus, by definition of regular variation, we have

$$t\mathbb{P}\left[\frac{d_i}{b(t)} > x\right] \to cx^{-\alpha}$$

for all $i = 1, 2, \dots$ and for some constant c.

By Proposition 5.6 (in one dimension), that is equivalent to:

$$\frac{1}{k}\sum_{i=1}^{n} \epsilon_{\frac{d_i}{b(t)}}(1, \infty) \Rightarrow c$$

for all $i = 1, 2, \dots$.

Since the denominator of the expression in Proposition 5.1 tends to a constant, it suffices to show that the numerator tends to $S_{\mathbb{O}}$. We can rewrite that condition as

$$\frac{1}{k}\sum_{i=1}^{n} \epsilon_{\frac{\boldsymbol{X_i}}{d_i}}(\cdot)\mathbb{1}\left[\frac{d_i}{b(\frac{n}{k})} > 1\right] \Rightarrow S_{\mathbb{O}}(\cdot).$$

The sum on the left counts the sets to which $\dfrac{\boldsymbol{X_i}}{d_i}$ belongs, such that $\dfrac{d_i}{b(\frac{n}{k})} > 1$, that is, given that $\dfrac{d_i}{b(\frac{n}{k})} > 1$. We can then use Proposition 5.6 again to express it as

$$t\mathbb{P}\left[\frac{\boldsymbol{X_i}}{d_i b(t)} \in \cdot \,\middle|\, \frac{d_i}{b(t)} > 1\right] \xrightarrow{*} S_{\mathbb{O}}(\cdot).$$

But we know that (5.1) holds for all $\boldsymbol{X_i}$, and also that $b(t) = \overline{F}_d^{-1}(1/t)$. From there we get the desired result. □

6 Simulated examples

In this chapter we analyze some examples of multivariate regular variation.

6.1 Independent components

We consider a risk vector $\boldsymbol{X} = (X^1, \ldots, X^d)$, where X^j are independent and identically distributed with regularly varying tail $\overline{F}$ with parameter $-\alpha$. That is,

$$\lim_{t\to\infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = \lim_{t\to\infty} \frac{\mathbb{P}[X^j > tx]}{\mathbb{P}[X^j > t]} = x^{-\alpha}$$

for all $j = 1, \ldots, d$ and $x > 0$.

Then, if we define the function b as

$$b(t) = \overline{F}^{-1}(1/t),$$

we have

$$t\mathbb{P}\left[\frac{X^j}{b(t)} \in (x, \infty)\right] \to x^{-\alpha} \tag{6.1}$$

for all positive x.

This is true, because, due to the fact that $\overline{F}$ has regular variation, we have

$$\begin{aligned}
\lim_{t\to\infty} t\mathbb{P}\left[\frac{X^j}{b(t)} > x\right] &= \lim_{t\to\infty} t\overline{F}(b(t)x) \\
&= \lim_{t\to\infty} t\frac{\overline{F}(b(t)x)}{\overline{F}(b(t))}\overline{F}(b(t)) \\
&= \lim_{t\to\infty} \frac{\overline{F}(b(t)x)}{\overline{F}(b(t))} \lim_{t\to\infty} t\overline{F}(b(t)) \\
&= x^{-\alpha}.
\end{aligned}$$

Then, for big enough t,

$$t\overline{F}(b(t)) = \frac{\overline{F}(b(t))}{1/t} \sim \frac{\overline{F}(b(t))}{1/b(t)} \sim \frac{\overline{F}(b(t))}{\overline{F}(t)} \sim 1.$$

We will show that the distribution of X has multivariate regular distribution with function b as before and limit measure μ given by

$$\mu([\mathbf{0}, \boldsymbol{x}]^c) = \sum_{i=1}^{d} (x^i)^{-\alpha} \tag{6.2}$$

for $\boldsymbol{x} = (x^1, \ldots, x^d) \in [\mathbf{0}, \infty) \backslash \{\mathbf{0}\}$.

By Lemma 2.1, it is sufficient to show that, for $\boldsymbol{x} > \mathbf{0}$, we have

$$t\mathbb{P}\left[\frac{\boldsymbol{X}}{b(t)} \in [\mathbf{0}, \boldsymbol{x}]^c\right] \to \mu([\mathbf{0}, \boldsymbol{x}]^c).$$

We have, by the Principle of Inclusion and Exclusion,

$$\begin{aligned} t\mathbb{P}\left[\frac{\boldsymbol{X}}{b(t)} \in [\mathbf{0}, \boldsymbol{x}]^c\right] &= t\mathbb{P}\left\{\bigcup_{i=1}^{d}\left[\frac{X^i}{b(t)} > x^i\right]\right\} \\ &= \sum_{i=1}^{d} t\mathbb{P}\left[\frac{X^i}{b(t)} > x^i\right] - \sum_{1 \leqslant i < j \leqslant d} t\mathbb{P}\left[\frac{X^i}{b(t)} > x^i, \frac{X^j}{b(t)} > x^j\right] \\ &+ \sum_{1 \leqslant i < j < k \leqslant d}^{d} t\mathbb{P}\left[\frac{X^i}{b(t)} > x^i, \frac{X^j}{b(t)} > x^j, \frac{X^k}{b(t)} > x^k\right] \\ &- \ldots (-1)^{d+1} t\mathbb{P}\left\{\bigcap_{i=1}^{d}\left[\frac{X^i}{b(t)} > x^i\right]\right\}. \end{aligned}$$

For the first sum, we have, by (6.1),

$$\sum_{i=1}^{d} t\mathbb{P}\left[\frac{X^i}{b(t)} > x^i\right] \to \sum_{i=1}^{d} (x^i)^{-\alpha}.$$

The other sums go to zero, because all terms there can be bounded by an expression of the form (for $i \neq j$)

$$t\mathbb{P}\left[\frac{X^i}{b(t)} > x^i, \frac{X^j}{b(t)} > x^j\right] = t\mathbb{P}\left[\frac{X^i}{b(t)} > x^i\right]\mathbb{P}\left[\frac{X^j}{b(t)} > x^j\right] \to (x^i)^{-\alpha} \times 0 = 0.$$

We thus conclude that

$$t\mathbb{P}\left[\frac{\boldsymbol{X}}{b(t)} \in \cdot\right] \xrightarrow{*} \mu(\cdot).$$

We note that the measure μ puts zero mass in the interior of the positive quadrant. We will show that when $d = 2$ (the case for higher dimensions is similar).

Given $\boldsymbol{x} = (x, x) > \mathbf{0}$, we show that $\mu((\boldsymbol{x}, \infty)) = 0$. By denoting $Q_1 = [0, x] \times (x, \infty)$, $Q_2 = (\boldsymbol{x}, \infty) = (x, \infty)^2$ and $Q_3 = (x, \infty) \times [0, x]$, we have that $Q_i \cap Q_j = \varnothing$, whenever $i \neq j$, and

$$[\mathbf{0}, \boldsymbol{x}]^c = ([0, x] \times [0, x])^c = \bigcup_{i=1}^{3} Q_i.$$

Thus

$$\sum_{i=1}^{3} \mu(Q_i) = \mu(\bigcup_{i=1}^{3} Q_i) = \mu([\mathbf{0}, \boldsymbol{x}]^c) = 2x^{-\alpha}.$$

On the other hand,

$$\mu(Q_1) + \mu(Q_2) = \mu(Q_1 \cup Q_2) = \mu(([0, \infty) \times [0, x])^c) = x^{-\alpha},$$

and similarly

$$\mu(Q_2) + \mu(Q_3) = x^{-\alpha}.$$

We conclude that

$$2x^{-\alpha} = \sum_{i=1}^{3} \mu(Q_i) = \mu(Q_1) + \mu(Q_2) + \mu(Q_2) + \mu(Q_3) - \mu(Q_2) = 2x^{\alpha} - \mu(Q_2),$$

and so $\mu(Q_2) = \mu((\boldsymbol{x}, \infty)) = 0$.

Therefore, independent components are associated with a limit measure with support concentrated on the axes. Figure 6.1 presents an example of a two-dimensional vector with independent components.

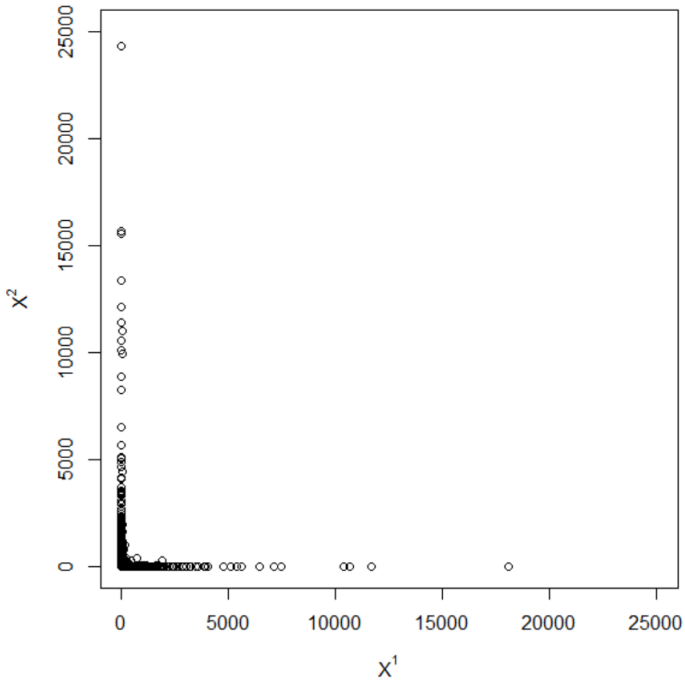


Figure 6.1: Simulated data ($N = 10^5$) for a two-dimensional random variable with independent Pareto marginal distributions.

6.2 Asymptotic independence

We consider again a risk vector $\boldsymbol{X} = (X^1, \ldots, X^d)$, but now we assume the distribution of $\boldsymbol{X}$ has multivariate regular variation with limit measure μ as defined in (6.2) and some increasing function b.

We say that $\boldsymbol{X}$ has asymptotic independence, even though we do not assume that each component of $\boldsymbol{X}$ is independent of each other.

We note that, in two dimensions ($d = 2$), we have

$$\begin{aligned}
\lim_{t\to\infty} \mathbb{P}[X^2 > t | X^1 > t] &= \lim_{t\to\infty} \frac{\mathbb{P}[X^2 > t, X^1 > t]}{\mathbb{P}[X^1 > t]} \\
&= \lim_{t\to\infty} \frac{\mathbb{P}[X^2 > b(t), X^1 > b(t)]}{\mathbb{P}[X^1 > b(t)]} \\
&= \lim_{t\to\infty} ct\mathbb{P}[X^2 > b(t), X^1 > b(t)] \\
&= \lim_{t\to\infty} ct\mathbb{P}[\frac{\boldsymbol{X}}{b(t)} > 1] \\
&= c\mu((\mathbf{1}, \infty)) \\
&= 0,
\end{aligned}$$

for some positive constant c and $\mu((\mathbf{1}, \infty)) = 0$ as we have shown before.

We can interpret this in the sense that, if one component of $\boldsymbol{X}$ is large, then there is a very small probability that the other component of $\boldsymbol{X}$ is also large.

6.3 Identical components

We consider a risk vector $\boldsymbol{X} = (X^1, \ldots, X^1)$, where each component is the same random variable X^1, which has a regularly varying tail distribution F with parameter $-\alpha$.

Then, if we define

$$b(t) = \overline{F}^{-1}(1/t),$$

we have, as previously,

$$t\mathbb{P}\left[\frac{X^j}{b(t)} \in (x, \infty)\right] \to x^{-\alpha} \tag{6.3}$$

for all positive x.

We can then compute

$$\begin{aligned} t\mathbb{P}\left[\frac{\boldsymbol{X}}{b(t)} \in [\mathbf{0}, \boldsymbol{x}]^c\right] &= t\mathbb{P}\left\{\bigcup_{i=1}^{d}\left[\frac{X^i}{b(t)} > x^i\right]\right\} \\ &= t\mathbb{P}\left[\frac{X^1}{b(t)} > \min_{i=1,\ldots,d} x^i\right] \\ &\to (\min_{i=1,\ldots,d} x^i)^{-\alpha}. \end{aligned}$$

Thus, for risk vectors that consist of only repeated components, we have multivariate regular variation with limit measure μ given by

$$\mu([\mathbf{0}, \boldsymbol{x}]^c) = (\min_{i=1,\ldots,d} x^i)^{-\alpha}. \tag{6.4}$$

This measure is concentrated on the line $\{(x^1, x^2, \ldots, x^d) : x^1 = x^2 = \cdots = x^d\}$, that is, the line where all components of a d-dimensional vector are the same.

To see that this is true, we consider the two-dimensional case (easily generalized for higher dimensions) with a vector $\boldsymbol{x} = (x, x)$, for some positive x, and the same partition of $[0, \boldsymbol{x}]^c$ as in 6.1., so that $Q_1 = [0, x] \times (x, \infty)$, $Q_2 = (\boldsymbol{x}, \infty) = (x, \infty)^2$ and $Q_3 = (x, \infty) \times [0, x]$.

We then have

$$\begin{aligned} \mu([\mathbf{0}, \boldsymbol{x}]^c) &= \mu(Q_1) + \mu(Q_2) + \mu(Q_3) \\ &= \mu(Q_1) + \mu(Q_2) + \mu(Q_3) + \mu(Q_2) - \mu(Q_2) \\ &= \mu(([0, \infty) \times [0, x])^c) + \mu(([0, x] \times [0, \infty))^c) - \mu(Q_2). \end{aligned}$$

By using the expression of the measure μ given by (6.4), we get

$$x^{-\alpha} = x^{-\alpha} + x^{-\alpha} - \mu(Q_2),$$

hence $\mu(Q_2) = x^{-\alpha}$.

But because $\mu([\mathbf{0}, \boldsymbol{x}]^c) = x^{-\alpha}$, we have that $\mu(Q_1) = \mu(Q_3) = 0$. Since this works with any positive x, we conclude that any set outside the line $y = x$ has measure zero. Figure 6.2 presents a graph of a two-dimensional random vector with identical components.

In the same way for the asymptotic independence, we say that a regularly varying distribution of a random vector $\boldsymbol{X}$ possesses asymptotic full dependence if the limit measure μ is given by (6.4), even if it not all vector components are equal.

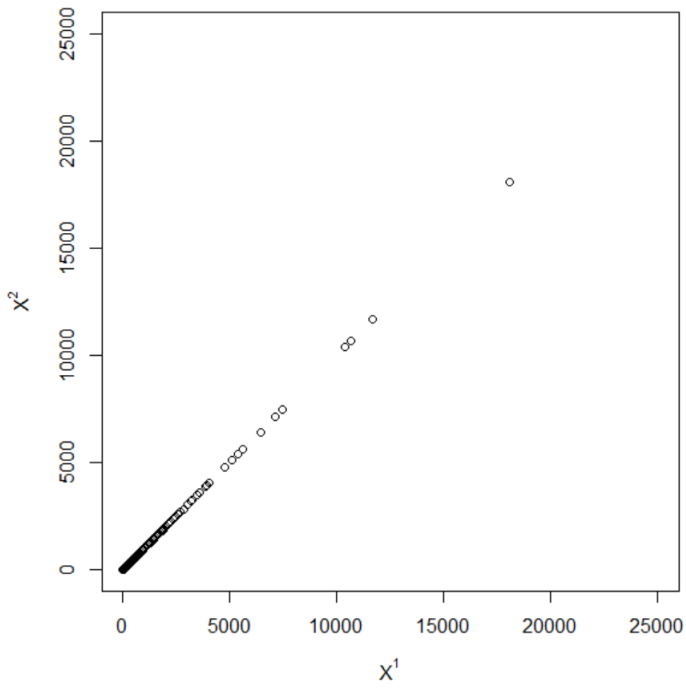


Figure 6.2: Simulated data ($N = 10^5$) for a two-dimensional random variable with a common Pareto marginal distribution.

6.4 Removal of independent components

Suppose a risk vector $\boldsymbol{X} = (X^1, \ldots, X^d)$, where we have one component that is independent from all the others. That is, there is $j \in \{1, \ldots, d\}$ such that X^j is independent from X^i, for all $i \neq j$.

Suppose further that all components of vector $\boldsymbol{X}$ have regularly varying tails with parameters α_1, α_2, ..., α_d, such that

$$\lim_{t\to\infty} \frac{\overline{F_i}(tx)}{\overline{F_i}(t)} = \lim_{t\to\infty} \frac{\mathbb{P}[X^i > tx]}{\mathbb{P}[X^i > t]} = x^{-\alpha_i}$$

for all $i = 1, \ldots, d$ and $x > 0$.

Let b be an increasing function, with $\lim_{t\to\infty} b(t) = \infty$, and such that

$$t\mathbb{P}\left[\frac{X^j}{b(t)} \in (x, \infty)\right] \to x^{-\alpha_j}$$

for all positive x. We have seen in section (6.1) that such function exists.

By the Principle of Inclusion and Exclusion, we have

$$\begin{aligned} t\mathbb{P}\left[\frac{\boldsymbol{X}}{b(t)} \in [\mathbf{0}, \boldsymbol{x}]^c\right] &= t\mathbb{P}\left\{\bigcup_{i=1}^{d}\left[\frac{X^i}{b(t)} > x^i\right]\right\} \\ &= \sum_{i=1}^{d} t\mathbb{P}\left[\frac{X^i}{b(t)} > x^i\right] - \sum_{1\leqslant i<k\leqslant d} t\mathbb{P}\left[\frac{X^i}{b(t)} > x^i, \frac{X^k}{b(t)} > x^k\right] \\ &+ \sum_{1\leqslant i<k<l\leqslant d}^{d} t\mathbb{P}\left[\frac{X^i}{b(t)} > x^i, \frac{X^k}{b(t)} > x^k, \frac{X^l}{b(t)} > x^l\right] \\ &- \ldots (-1)^{d+1} t\mathbb{P}\left\{\bigcap_{i=1}^{d}\left[\frac{X^i}{b(t)} > x^i\right]\right\}. \end{aligned}$$

For any term involving X^j, we have, due to independence,

$$\begin{aligned} &\lim_{t\to\infty} t\mathbb{P}\left[\frac{X^j}{b(t)} > x^j, \frac{X^{i_1}}{b(t)} > x^{i_1}, \ldots, \frac{X^{i_k}}{b(t)} > x^{i_k} \text{ for some } i_1, \ldots, i_k\right] \\ &\quad = \lim_{t\to\infty} t\mathbb{P}\left[\frac{X^j}{b(t)} > x^j\right] \mathbb{P}\left[\frac{X^{i_1}}{b(t)} > x^{i_1}, \ldots, \frac{X^{i_k}}{b(t)} > x^{i_k} \text{ for some } i_1, \ldots, i_k\right] \\ &\quad = (x^j)^{-\alpha_j} \times 0 = 0. \end{aligned}$$

So, in the limit, the only term of $t\mathbb{P}\left[\frac{\boldsymbol{X}}{b(t)} \in [\mathbf{0}, \boldsymbol{x}]^c\right]$ containing X^j that does not go to zero is $t\mathbb{P}\left[\frac{X^j}{b(t)} \in (x, \infty)\right]$, whose limit is $(x^j)^{-\alpha_j}$. The terms that remain (that do not contain X^j) are of the form $t\mathbb{P}\left[\frac{\boldsymbol{X^{-j}}}{b(t)} \in [\mathbf{0}, \boldsymbol{x^{-j}}]^c\right]$, where $\boldsymbol{X^{-j}}$ and $\boldsymbol{x^{-j}}$ are the (d-1)-dimensional vectors formed from $\boldsymbol{X}$ and $\boldsymbol{x}$, respectively, with the j^{th} component removed. Thus, we have

$$\begin{aligned} \lim_{t\to\infty} t\mathbb{P}\left[\frac{\boldsymbol{X}}{b(t)} \in [\mathbf{0}, \boldsymbol{x}]^c\right] &= \lim_{t\to\infty} t\mathbb{P}\left[\frac{\boldsymbol{X^{-j}}}{b(t)} \in [\mathbf{0}, \boldsymbol{x^{-j}}]^c\right] + t\mathbb{P}\left[\frac{X^j}{b(t)} \in (x, \infty)\right] \\ &= \lim_{t\to\infty} t\mathbb{P}\left[\frac{\boldsymbol{X^{-j}}}{b(t)} \in [\mathbf{0}, \boldsymbol{x^{-j}}]^c\right] + (x^j)^{-\alpha_j}. \end{aligned}$$

We can see that we can remove an independent component (any number of independent components, in fact, if we repeat this process) and study only the measure from the dependent components, since there is a simple relation between that measure and the global one.

7 Support estimation for simulated data

In this chapter we will compute the support estimator from chapter 5 using sets of simulated data and plot them in order to show how well the theoretical estimators work with known distributions. We will use the programming language R for generating and treating this data. The used programs can be checked in the appendices.

We will start with a simple example, where we generate a two-dimensional variable $\boldsymbol{X} = (X^1, X^2)$, from a random variable Y with Pareto distribution $Par(2, 2)$ and another random variable U, which is uniformly distributed on $[0.75, 0.95]$. We put $X^1 = YU$ and $X^2 = Y(1 - U)$.

Using this data, we compute the support estimator given by Proposition (5.4), $supp_{k,n} = \left\{ \frac{\boldsymbol{X}_i}{d_i} : d_i > d_{(k+1)}, i = 1, \ldots, n \right\}$, setting $k = n^{0.75}$, and obtain the following graph:

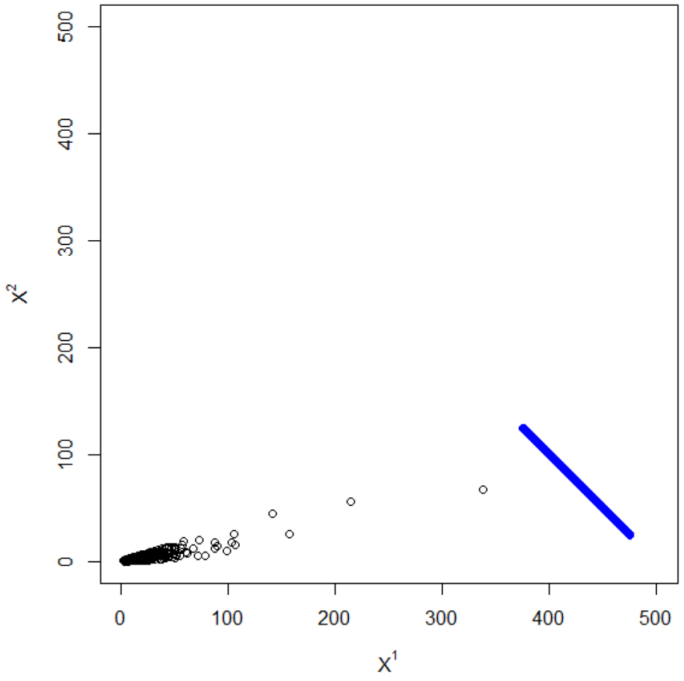


Figure 7.1: Simulated data ($N = 10^5$) from a Pareto distribution $Par(2, 2)$ mingled with an uniform distribution on $[0.75, 0.95]$. The blue line is the computed support estimator with L^1-norm. The value of k is $N^{0.75}$.

From graphical inspection, the support estimator seems to cover all the observations and the theoretical limit measure.

We consider a second example, where we generate two two-dimensional random variables: one with the same distribution from the previous example, and another with corresponding Pareto distribution $Par(2, 10)$ and uniform distribution on $[0.25, 0.45]$. We then sum the two variables, compute the support estimator of the resulting distribution and obtain Figure 7.2.

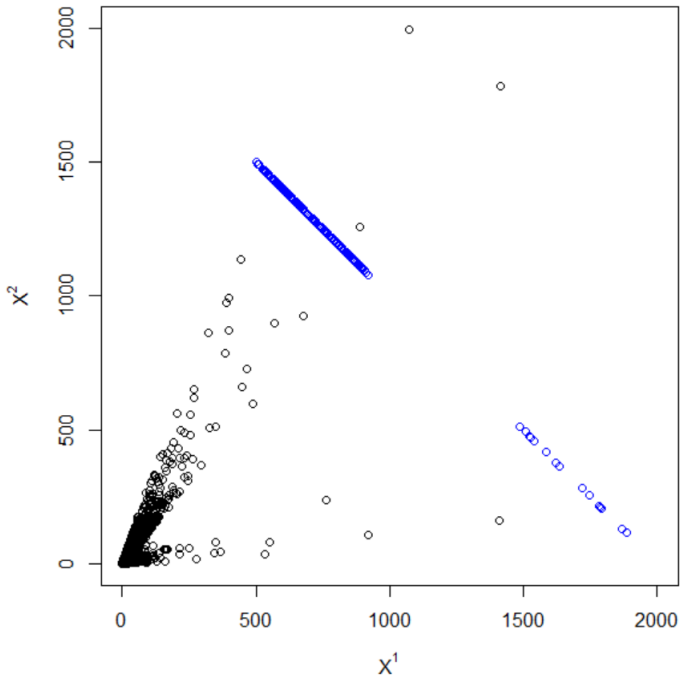


Figure 7.2: Simulated data ($N = 10^5$) from a sum of two random variables: the associated Pareto and Uniform parameters are $((2, 2); [0.75, 0.95])$ and $((2, 10); [0.25, 0.45])$. The blue line is the computed support estimator with L^1-norm. The value of k is $N^{0.485}$.

We now see that our support estimator holds well for hidden regular variation, where we have two cones with regular variation. The main regular variation is completely captured by the set of points of the estimator, and the hidden regular variation (the smaller cone) has a concentration of some points. Even though there are some outliers, it seems reasonable to assume one could get the correct limit measure from the support estimator.

The efficiency of the support estimator depends on the value of k: for larger values, we have more points in the estimator, and so we lose accuracy, since we will be considering points from the sample that do not match multivariate regular variation; on the other hand, for smaller values, we may not

consider enough points from the sample to cover the cone where the limit measure is. Thus a certain balance for k is required, and this usually means some trail-and-error for each sample.

In Figure 7.3, the same sample from Figure 7.2 is reproduced with two different values for k (dependent on the sample size N).

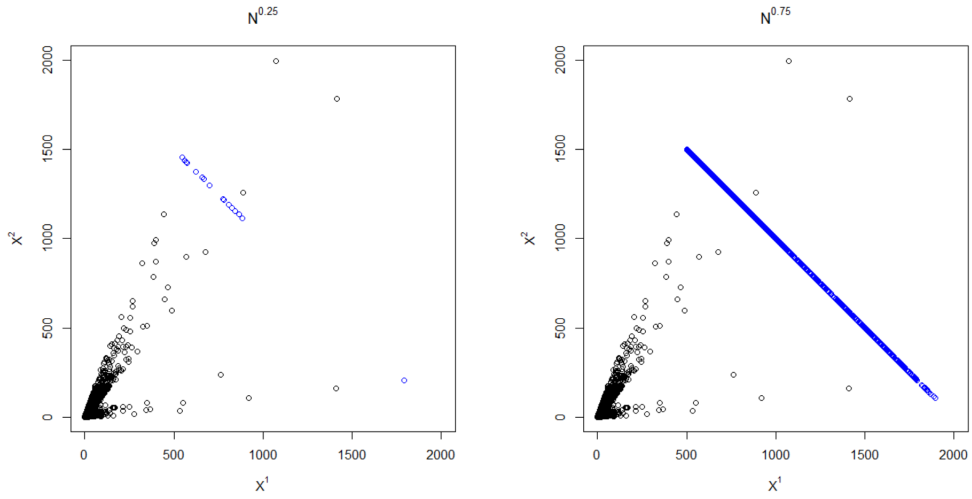


Figure 7.3: Simulated data from Figure 7.2 with two different values of k, based on the sample size $N = 10^5$.

These graphs illustrate the importance of choosing a good value for k. The sample distribution is identical to Figure 7.2, which corresponded to an adequate estimator, with $k = N^{0.485}$. On that graph there is a straight line associated with the first cone of regular variation and a small set of points linked to the hidden regular variation cone. The first graph of Figure 7.3 (on the left side), with $k = N^{0.25}$, presents the data with an estimator generated by a value to k that is evidently too small. The smaller the value of k is, the fewer points are selected for the estimator, making it harder to detect any regular variation. In this case, it detected something for the first cone (and thus we have a small set of points in that direction) and nothing for the second cone. On the other hand, the larger the value of k is, the more points form the estimator, and there is a risk of having too many points, leading the estimator to falsely identify regular variation cones. This happens on the second graph of Figure 7.3 (on the right side), where the value of $k = N^{0.75}$ is too large. The result is a straight line connecting the directions of the two cones, which is quite misleading.

In Figure 7.4, an exponential noise was added to both components of the random vector.

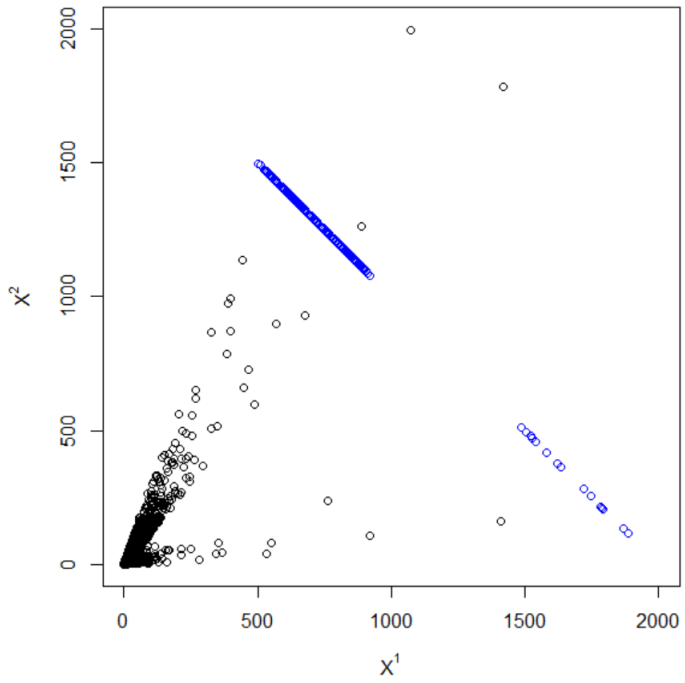


Figure 7.4: Simulated data ($N = 10^5$) from a sum of two random variables: the associated Pareto and Uniform parameters are $((2,2);[0.75,0.95])$ and $((2,10);[0.25,0.45])$. Exponential noise was added to both components of the random vector. The blue line is the computed support estimator with L^1-norm. The value of k is $N^{0.485}$.

The graphs presented here may lead us to conclude that the support estimator $supp_{k,n} = \left\{ \frac{\boldsymbol{X}_i}{d_i} : d_i > d_{(k+1)}, i = 1, \ldots, n \right\}$ does fairly well in detecting multivariate regular variation and hidden regular variation, as it suggests that adding a distribution whose tail is not regularly-varying does not cause a significant change in the support estimator. However, we must keep in mind that in this chapter we have worked with estimated data, which is much more well-behaved than real data.

8 A worked example: daily stock returns

Let us consider a real-life example: the stock returns of some companies indexed in the Standard & Poors (S&P) 500 Index. We have obtained this public data through Finance Yahoo! (see [16], [17]). We will work with the log-difference of the stock prices, that is, if S_t denotes the (opening) stock

price at time unit t (we generally consider daily stock prices), then we are interested in the return process R_t, defined, for any $t > 0$, by

$$R_t = \log \frac{S_t}{S_{t-1}}.$$

The preference of the return process over the price process S_t is mainly due to two factors. First, the return process is dimensionless in the sense that it is independent from the unit associated to the price process. It also does not depend on the initial capital or investment.

Another fact is that the return processes of different commodities are much more likely to provide us independent and identically distributed multidimensional sample points (at least for large enough observations), unlike the price processes, which do not satisfy this assumption.

In order to apply the Multivariate Regular Variation (MRV) model to our observations, we have to process the data in the following way.

1. Determine a certain threshold and remove from the sample all observations that are smaller than that threshold. This is done since MRV only concerns large enough values, and most observations (the smaller ones) are just noise that give no information about how fast the tail of the distribution is decaying. The threshold can be set by graphical inspection of the sample points.

2. Normalize the remaining sample vector. This still preserves the size relation between observations, but makes it easier to plot and check for MRV.

When the data is processed, we need to analyze the sample marginal distributions to check that their tails are regularly varying and, in particular, that they have the same parameter α. Two (complementary) ways of doing that are the Hill estimator and QQ plots. These and other methods are discussed in ([14], Section 4).

Even if each marginal distribution is similar to a regularly varying one-dimensional distribution, that does not mean that the joint distribution follows MRV. Furthermore, even if the joint distribution could come from a distribution with MRV, it is possible that we do not have sufficient data to infer it without much wishful thinking. We give here some examples.

We first consider the stock prices of Microsoft and Google, presented in Figure 8.1. We would expect some concentration around the line $y = x$, since these two companies work in the same line of business and thus should be highly dependent of one another. However, the graph does not show a

clear picture of MRV. The marginal distributions approximate Pareto distributions, but we would need more sample points for the largest observations.

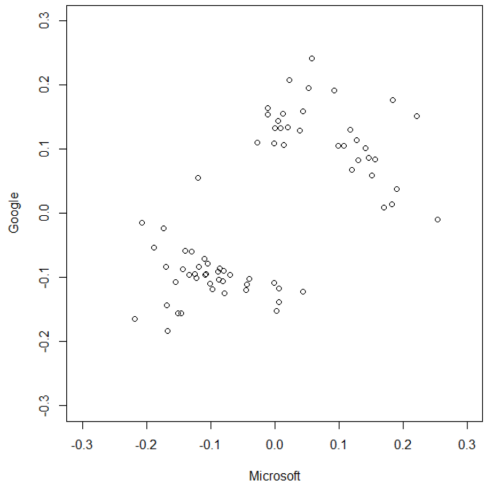


Figure 8.1: Daily return processes for Microsoft and Google stocks. The collected data corresponds to the period between 01/01/2005 and 01/01/2020 ($N = 5478$).

What happens when we look at stocks from different industries? If we consider stocks from Apple and General Motors, we obtain Figure 8.2.

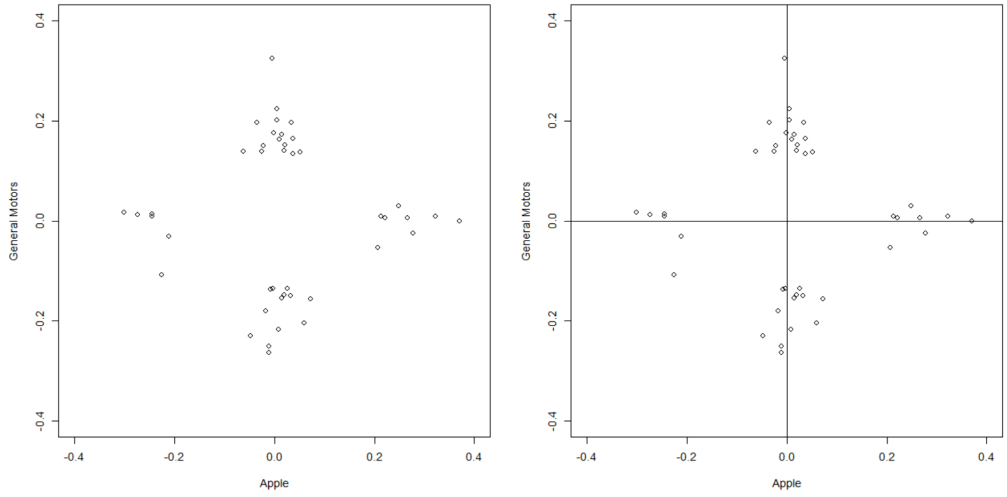


Figure 8.2: Daily return processes for Apple and General Motors stocks. The collected data corresponds to the period between 01/01/2005 and 01/01/2020 ($N = 5478$).

The graph does show some concentration of data points around the axis (as is the case in MRV with independent components), but the sample size is too small to necessarily conclude that there is MRV in this case.

9 Final Remarks

This entire work sprouts from its core concept: the definition of multivariate regular variation (Definition 3.1). Throughout the thesis several equivalent definitions were given. It makes sense to gather them here and present them in a single theorem.

Theorem 9.1. ***(Equivalent Definitions of Multivariate Regular Variation)*** *Let $\boldsymbol{X}$ be a random vector in $\mathbb{O} = \mathbb{C}\backslash\mathbb{F}$. Then, the following statements are equivalent:*

1. *There exist a increasing function $b \nearrow \infty$ and a non-zero measure μ on $\mathbb{M}^*(\mathbb{C}, \mathbb{O})$ such that*
$$t\mathbb{P}\left[\frac{\boldsymbol{X}}{b(t)} \in \cdot\right] \xrightarrow{*} \mu(\cdot).$$

2. *There exist a sequence b_n such that $\lim_{n\to\infty} b_n = \infty$ and a non-zero measure μ on $\mathbb{M}^*(\mathbb{C}, \mathbb{O})$ such that*
$$n\mathbb{P}\left[\frac{\boldsymbol{X}}{b_n} \in \cdot\right] \xrightarrow{*} \mu(\cdot).$$

3. *There exists a non-zero measure μ on $\mathbb{M}^*(\mathbb{C}, \mathbb{O})$ such that*
$$\lim_{t\to\infty} \frac{1 - F(t\boldsymbol{x})}{1 - F(t\boldsymbol{1})} = \lim_{t\to\infty} \frac{\mathbb{P}[\boldsymbol{X}/t \in [\boldsymbol{0}, \boldsymbol{x}]^c]}{\mathbb{P}[\boldsymbol{X}/t \in [\boldsymbol{0}, \boldsymbol{1}]^c]} = \mu([\boldsymbol{0}, \boldsymbol{x}]^c])$$
for all $\boldsymbol{x} \neq \boldsymbol{0}$ which are continuity points of the function $\mu([\boldsymbol{0}, \cdot]^c)$.

4. *There exist a function $b \in RV_{\frac{1}{\alpha}}$ and a constant $c > 0$ such that*
$$t\mathbb{P}\left[\left(\frac{d(\boldsymbol{X}, \mathbb{F})}{b(t)}, \frac{\boldsymbol{X}}{d(\boldsymbol{X}, \mathbb{F})}\right) \in A\right] \to c\mu_{(\alpha)} \times S_{\mathbb{O}}(A)$$
for all measurable sets $A \in \mathcal{A}_h$, where $h : \boldsymbol{x} \mapsto \left(d(\boldsymbol{x}, \mathbb{F}), \frac{\boldsymbol{x}}{d(\boldsymbol{x}, \mathbb{F})}\right)$, $\mu_{(\alpha)}$ is the Pareto measure given by $\mu_{(\alpha)}((x, \infty)) = x^{-\alpha}$, for $x > 0$, and
$$S_{\mathbb{O}}(\Lambda) = \frac{\mu(\{\boldsymbol{x} \in \mathbb{O} : d(\boldsymbol{x}, \mathbb{F}) \geqslant 1, \frac{\boldsymbol{x}}{d(\boldsymbol{x}, \mathbb{F})} \in \Lambda\})}{\mu(\{\boldsymbol{x} \in \mathbb{O} : d(\boldsymbol{x}, \mathbb{F}) \geqslant 1\})}.$$

5. *There exist a sequence* $k = k(n)$ *such that* $k(n) \to \infty$, $n/k \to \infty$ *and* $k(n) \sim k(n+1)$, *and a non-zero measure* μ *on* $\mathbb{M}^*(\mathbb{C}, \mathbb{O})$ *such that*

$$\frac{1}{k}\sum_{i=1}^{n} \epsilon_{\frac{\boldsymbol{X}_i}{b(\frac{n}{k})}} \Rightarrow \mu,$$

where $\{\boldsymbol{X}, \boldsymbol{X_1}, \boldsymbol{X_2}, \dots\}$ *are independent and identically distributed.*

Clearly multivariate regular variation theory has multiple applications and can be a powerful tool for describing and analyzing dependence structures of large sets of multivariate data. Although the theory, in this format, is relatively recent, there is a solid mathematical background that sustains it, namely heavy-tail analysis and extreme-value theory in general.

However, there is still a lot to be done and researched. One area where we can see the shortcomings of the current theory is real-life data, as shown in the previous chapter. It is quite hard to make sure that the data we are working with follows an actual regularly varying distribution, let alone finding the correct limit measure. The practical uses of the theory is therefore a field which could vastly benefit from additional research.

Despite these defects, the theory of multivariate regular variation is the best response to the problem, that is so often faced by several industries, of obtaining information about the dependence structure within the components of a multivariate random vector. The main advantages of this procedure are its straightforward implementation for practical purposes, its ease of handling from a mathematical point of view, as it assimilates rather easily in existing theory, and, related to that, the multiple equivalent ways it can be expressed. Competing approaches, such as copula theory, are either harder to manipulate theoretically or require knowing certain parameters, such as the marginal cumulative distributions of the vector components, which are not generally knowable in a real-world scenario. It would perhaps be beneficial for both fields of research if these various approaches could be somehow used together in a unified theory of heavy-tail dependence of random variables.

References

[1] P. Billingsley. *Probability and Measure*. Wiley Series in Probability and Statistics. Wiley, 1995.

[2] P. Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. Wiley, 2013.

[3] N.H. Bingham, C.M. Goldie, and J.L. Teugels. *Regular Variation*. Number no. 1 in Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1989.

[4] Bikramjit Das, Abhimanyu Mitra, and Sidney Resnick. Living on the multidimensional edge: seeking hidden risks using regular variation. *Adv. in Appl. Probab.*, 45(1):139–163, 2013.

[5] L. de Haan and Sidney I. Resnick. Estimating the limit distribution of multivariate extremes. *Communications in Statistics. Stochastic Models*, 9(2):275–309, 1993.

[6] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events: for Insurance and Finance*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg, 2013.

[7] H. Federer. *Geometric Measure Theory*. Classics in Mathematics. Springer Berlin Heidelberg, 2014.

[8] Henrik Hult and Filip Lindskog. Extremal behavior of stochastic integrals driven by regularly varying lévy processes. *The Annals of Probability*, 35(1):309–339, Jan 2007.

[9] O. Kallenberg. *Foundations of Modern Probability*. Probability and Its Applications. Springer New York, 2006.

[10] Jaakko Lehtomaa and Sidney Resnick. Asymptotic independence and support detection techniques for heavy-tailed multivariate data, 2019.

[11] Filip Lindskog, Sidney I. Resnick, and Joyjit Roy. Regularly varying measures on metric spaces: hidden regular variation and hidden jumps. *Probab. Surv.*, 11:270–314, 2014.

[12] I. Molchanov. *Theory of Random Sets*. Probability and Its Applications. Springer London, 2006.

[13] Sidney Resnick. The extremal dependence measure and asymptotic independence. *Stochastic Models*, 20(2):205–227, 2004.

[14] S.I. Resnick, T.V. Mikosch, and S.M. Robinson. *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Number v. 10 in Heavy-tail phenomena: probabilistic and statistical modeling. Springer, 2007.

[15] S.I. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer Series in Operations Research and Financial Engineering. Springer New York, 2013.

[16] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

[17] Jeffrey A. Ryan and Joshua M. Ulrich. *quantmod: Quantitative Financial Modelling Framework*, 2020. R package version 0.4.17.

A R codes

A.1 Section 6.1

```
library(extremefit)
# Fix seed that generates the random numbers
set.seed(1179)

# Number of samples
N<-10^5

# Sampling
p1<-rpareto(N,1,2)
p2<-rpareto(N,1,1)
q1<-p1
q2<-p2

# Plotting
par(mfrow=c(1,1),pty='s')
x1<-expression(X^1)
x2<-expression(X^2)
lim<-c(0,25000)
plot(q1,q1,xlim=lim,ylim=lim,xlab=x1,ylab=x2)
```

A.2 Section 6.3

```
library(extremefit)
# Fix seed that generates the random numbers
set.seed(1179)

# Number of samples
N<-10^5

# Sampling
p1<-rpareto(N,1,2)
q1<-p1

# Plotting
par(mfrow=c(1,1),pty='s')
x1<-expression(X^1)
x2<-expression(X^2)
lim<-c(0,25000)
plot(q1,q1,xlim=lim,ylim=lim,xlab=x1,ylab=x2)
```

A.3 Chapter 7

A.3.1 Multivariate regular variation

```
library(extremefit)
#Fix seed used in generating random numbers
set.seed(1143)

# Number of samples
N<-10^5

#Uniform distribution
U1<-runif(N,0.75,0.95)

#Pareto distribution
p1<-rpareto(N,2,2)

#Setting the two-dimensional distribution
q1<-p1*U1
q2<-p1*(1-U1)
x<-cbind(q1,q2)

#Estimating the support
normx<-x[,1]+x[,2]
sortnormx<-sort(normx,decreasing=TRUE)
summary(sortnormx)
dx<-x/normx

#Setting k
k<-N^(0.75)

#Getting only the points with the fitting norm
v1<-dx[,1]
v2<-dx[,2]
w1<-v1
w2<-v2
i<-1
j<-0
while (i<N) {
  if(normx[i]>sortnormx[k])
  {j<-j+1;v1[j]<-w1[i];v2[j]<-w2[i]}
  i<-i+1
}
```

```
#Plotting
par(pty="s",mfrow=c(1,1))
x1<-expression(X^1)
x2<-expression(X^2)
plot(q1,q2,xlim=c(0,500),ylim=c(0,500),xlab=x1,ylab=x2)
points(v1[1:j]*500,v2[1:j]*500, col="blue",lwd=.05)
```

A.3.2 Hidden regular variation

```
#Fix seed used in generating random numbers
set.seed(7698)

# Number of samples
N<-10^5

#Uniform distribution
U1<-runif(N,0.75,0.95)
U2<-runif(N,0.25,0.45)

#Pareto distribution
p1<-rpareto(N,2,2)
p2<-rpareto(N,2,10)

#Setting the two-dimensional distribution
q1<-p1*U1+p2*U2
q2<-p1*(1-U1)+p2*(1-U2)
x<-cbind(q1,q2)

#Estimating the support
normx<-x[,1]+x[,2]
sortnormx<-sort(normx,decreasing=TRUE)
dx<-x/normx

#Setting k
k<-N^(0.485)

#Getting only the points with the fitting norm
v1<-dx[,1]
v2<-dx[,2]
w1<-v1
w2<-v2
i<-1
j<-0
```

```
while (i<N) {
  if(normx[i]>sortnormx[k])
  {j<-j+1;v1[j]<-w1[i];v2[j]<-w2[i]}
  i<-i+1
}

#Plotting
par(pty="s",mfrow=c(1,1))
x1<-expression(X^1)
x2<-expression(X^2)
plot(q1,q2,xlim=c(0,2000),ylim=c(0,2000),xlab=x1,ylab=x2)
points(v1[1:j]*2000,v2[1:j]*2000, col="blue",lwd=.05)
```

A.3.3 Regular variation with exponential noise

```
#Fix seed that generates the random numbers
set.seed(7698)

# Number of samples
N<-10^5

# Sampling
U1<-runif(N,0.75,0.95)
U2<-runif(N,0.25,0.45)
p1<-rpareto(N,2,2)
p2<-rpareto(N,2,10)
exp1<-rexp(N,5)
qa1<-(p1)*U1
qa2<-(p1)*(1-U1)
qb1<-(p2)*U2
qb2<-(p2)*(1-U2)

#Setting the two-dimensional distribution
q1<-qa1+qb1+exp1
q2<-qa2+qb2+exp1
x<-cbind(q1,q2)

#Estimating the support
normx<-x[,1]+x[,2]
sortnormx<-sort(normx,decreasing=TRUE)
summary(sortnormx)
dx<-x/normx

#Setting k
```

```
k<-N^(0.485)

#Getting only the points with the fitting norm
v1<-dx[,1]
v2<-dx[,2]
w1<-v1
w2<-v2
i<-1
j<-0
while (i<N) {
  if(normx[i]>sortnormx[k])
  {j<-j+1;v1[j]<-w1[i];v2[j]<-w2[i]}
  i<-i+1
}

#Plotting
par(pty="s",mfrow=c(1,1))
x1<-expression(X^1)
x2<-expression(X^2)
plot(q1,q2,xlim=c(0,2000),ylim=c(0,2000),xlab=x1,ylab=x2)
points(v1[1:j]*2000,v2[1:j]*2000, col="blue",lwd=0.5)
```