Extremes - Lecture Notes

Master 2 'Mathématiques et applications', parcours 'Mathématiques de l'aléatoire', Orsay.

A. Sabourin[1]

January 2021

[1]LTCI, Telecom Paris, Institut polytechnique de Paris.

Contents

Chapter 1

Introduction to univariate extreme value theory

It seems that the rivers know the theory. It only remains to convince the engineers of the validity of this analysis.

Emil Julius Gumbel

- Course material for this chapter: Resnick (1987), chapters 0.1–0.3 (very concise); Leadbetter et al. (2012) (very detailed and easy to read), De Haan and Ferreira (2007), chapter 1 (additional results, more advanced).
- Other readings : Beirlant et al. (2004) (chapter 1) or Resnick (2007) Chapter 1 : examples of case studies and exploratory data analysis ; Coles (2001), chapters 3, 4: classical statistical methods.

1.1 Extreme Value Theory: what and why ?

1.1.1 Context, motivations

Extreme value theory (EVT) relies on elegant probability theory and finds natural statistical applications in many fields related to risk management (insurance, finance, telecommunication, climate, environmental sciences…).

To fix ideas (see Figure 1.1.1), call X our quantity of interest (X is a real valued random variable), which may be *e.g.* the water level on a coastal point, temperature, insurance claims …say we observe *i.i.d.* realizations $X_t, 1 \leq t \leq n$. Some questions of interest for risk management are

- Given a high threshold u, find $p = \mathbb{P}(X \geq p)$
- Given p (e.g. $p = 10^{-4}$), find u such that $\mathbb{P}(X > u) \leq p$.
- Given a long duration T (e.g. 10^4), and a high threshold u, find $p = \mathbb{P}(\max_{t \leq T} X_t \leq u)$.

In probabilistic terms, this is about estimating high quantiles or small probabilities. Unfortunately, it may happen that the sample size is too small for the naive empirical estimators to be of interest. As an example, if u is outside the range of observed data,

$$\hat{p}_n = \frac{1}{n} \sum_t \mathbb{1}_{x_t > u} = 0.$$

Another example about quantiles: we adopt throughout this course the following definition of the *quantile function*:

$$Q(p) = \inf\{x \in \mathbb{R} : F(x) \geq p\} \tag{1.1}$$

where F is the distribution function of the r.v. X under consideration.

An empirical counterpart of (1.1) based on n *i.i.d.* data is

$$\begin{aligned} \hat{Q}_n(p) &= \inf\{x \in \mathbb{R} : \hat{F}_n(x) \geq p\} = \inf\{x : \sum \mathbb{1}_{X_i \leq x} \geq np\} \\ &= X_{(\lceil np \rceil)}, \end{aligned}$$

where $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ are the order statistics of the sample $(X_t)_{1 \leq t \leq n}$. If one is interested in a very high quantile $Q(p)$ (*i.e.* p close to one) such that $1 > p > 1 - 1/n$, then $\lceil np \rceil = n$ and $\hat{Q}_n(p) = X_{(n)}$. There is no hope to estimate correctly such a quantile in a purely empirical manner.

To estimate such extremal quantities based on moderate sample sizes, one need additional assumption to be able to *extrapolate*, *e.g.* from what is observed above moderately high thresholds (left panel of Figure 1.1.1). As we shall see later on, it turns out that the answer to those questions depends on the (asymptotic) distribution of the *maximum* of n *i.i.d.* realizations of X, when n is large.

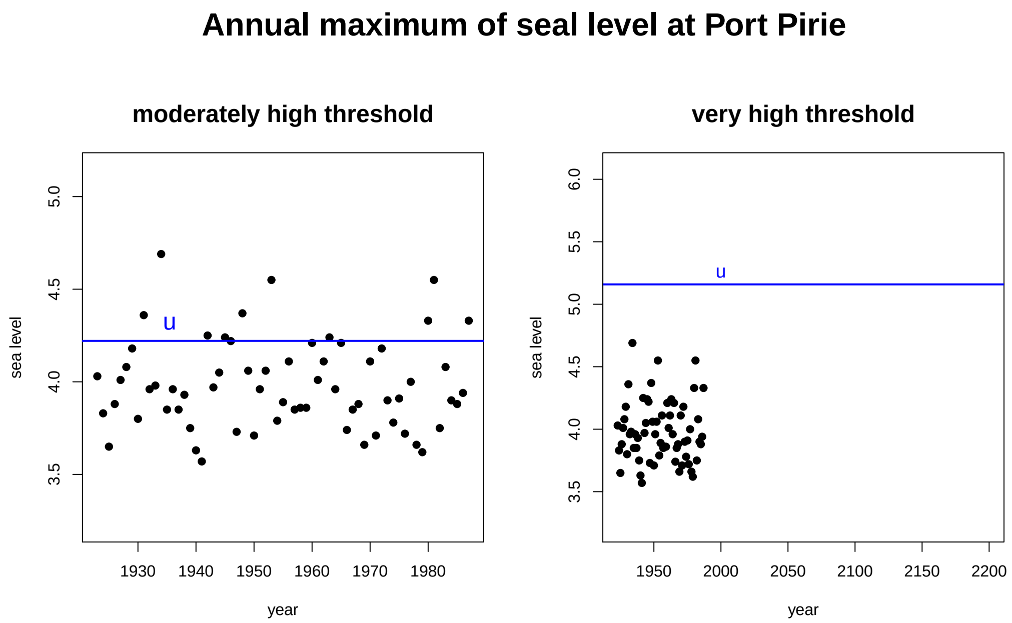

Figure 1.1: Why the empirical measure is not always useful as it is

Notations the maximum operator is denoted by $\bigvee$, so that for real numbers $x_i, 1 \leq i \leq n$, $\bigvee_{i=1}^n x_i = \max_{i=1}^n x_i$. Similarly, $\bigwedge$ is the minimum operator. In the multivariate case, these operators are understood componentwise, *i.e.* if $x_i = (x_{i,1}, \dots, x_{i,d})$,

$$\bigvee_{i=1}^{n} x_i = (\vee_{i=1}^n x_{i,1}, \dots, \vee_{i=1}^n . x_{i,d})$$

1.1.2 Rationale behind Extreme value theory (EVT)

The general purpose of EVT is to find statistical models for "extremes" (defined as maxima or excesses above large thresholds), supported by the theory (together with estimation tools). Consider i.i.d. copies X_i ($i \in \mathbb{N}$) of a random variable / vector / process X. Let us denote by $[X \mid \|X\| > u]$ the conditional distribution of X on the event $\{\|X\| > u\}$. Then under minimal assumptions,

$$\bigvee_{i=1}^{n} X_i \qquad \text{and} \qquad [X \mid \|X\| > u]$$

both converge to a certain class (as $n \to \infty$ or $u \to \infty$), up to a suitable normalization. Convergence of maxima is understood in the weak sense (convergence in distribution). Of course, convergence of the conditional distribution $[X \,\|\|X\| > u]$ is also a convergence in distribution. Interestingly enough, convergence of the maxima is *equivalent* to convergence of the conditional distribution of excesses. The main idea of extreme value analysis is to use the class of possible limits as a model for the law of the maximum over a long period of interest (the duration of a contract, the next 100 years for a dam, the next 1000 years for a nuclear plant, ...) or for the distribution of 'large' values (above a sufficiently high threshold). Inference in the appropriate model will be performed using the few (say k) largest data from a dataset of size n. Convergence of the various estimators is generally obtained under the assumption that $n \to \infty$, $k = k(n) \to \infty$, but k is a 'small proportion' of n, *i.e.* $k = o(n)$.

Why do we need renormalization ? If F is the cumulative distribution function (*c.d.f.*) of X and if the X_i's are i.i.d., then the *c.d.f.* for $M_n := \vee_{i=1}^n X_i$ is

$$F_n(x) = \mathbb{P}(M_n \leq x) = \mathbb{P}(\forall i \leq n, X_i \leq x) = F(x)^n.$$

Thus, if we do not 'normalize' the maximum, its distribution F_n is such that $F_n(x) \to 0$ as soon as $F(x) < 1$ and the limit distribution function (if there is one) is degenerate. Similarly, the distribution of X, given that $\|X\| \geq u$ 'escapes' to infinity.

1.1.3 A CLT for maxima ?

Recall hat the Central Limit Theorem states that, if X has finite second moment, we have

$$\frac{\sum_{i=1}^n X_i - b_n}{a_n} \xrightarrow{w} Z$$

where $\xrightarrow{w}$ stands for convergence in distribution, with Z a centered Gaussian distribution, $b_n = n\mathbb{E}(X)$ and $a_n = \sqrt{n}$.

In extreme value theory, the focus is on the maximum rather than the mean. The working hypothesis is the so-called *maximum-domain of attraction condition* (MDA):

There exist two sequences $(a_n)_{n\geq 0}, (b_n)_{n\geq 0}$ *of real numbers, with* $a_n > 0\ \forall n$, *and a non-degenerate random variable* Z, *such that*

$$\frac{\bigvee_{i=1}^{n} X_i - b_n}{a_n} \xrightarrow{w} Z \tag{MDA}$$

where $(X_i)_i$ *are i.i.d. random variables distributed as* X.

Remark 1.1.1 ('non-degenerate')**.** *A random variable is called 'non-degenerate' if its distribution is not concentrated at a single point. In other terms, it means that its* c.d.f. F *is such that* $\exists x < y \in \mathbb{R} : F(x) < F(y) < 1$.

In terms of distribution functions, the (MDA) is equivalent to the existence of a non-degenerate distribution function G such that

$$F^n(a_n x + b_n) \xrightarrow[n\to\infty]{} G(x) \tag{MDA'}$$

a each point x which is a continuity point of the limit G. If (MDA) (or alternatively (MDA')) holds, X (or F) is said to belong to the *maximum domain of attraction* of Z (or G).

Definition 1.1.2 (Extreme Value Distribution)**.** *A non-degenerate distribution function* G *is called an* extreme value distribution *if* (MDA') *is satisfied for some distribution function* F *and some sequences* $a_n > 0, b_n$.

Some natural questions are

- Under which conditions on F do we have (MDA') for some *c.d.f.* G ?
- What are the possible forms of the limit G ?
- How can we choose the sequences a_n, b_n ?
- What is the relation between (MDA) and the convergence of the conditional distribution of excesses above large thresholds ?

The aim of this chapter (and the next one) is to bring some answers, in the case where X is a real-valued random variable. The multivariate case will be the subject of the last chapter. This course does not cover the case where X is a (continuous) stochastic process.

1.2 Intermediate results

1.2.1 Monotone functions and weak convergence

It is a standard fact that convergence in distribution for random variables is the same as convergence of the associated distribution functions at each continuity point of the limit. In EVT, it is useful to extend this type of convergence to the whole class of monotone functions, as follows.

Definition 1.2.1 (Weak convergence of monotone functions)**.** *Let $(H_n)_{n\in\mathbb{N}}$ be a family of monotone functions $\mathbb{R} \to [-\infty, \infty]$. The functions H_n are said to* converge weakly*, and we write $H_n \xrightarrow{w} H$, if there exists a monotone function $H : \mathbb{R} \to [-\infty, \infty]$ such that*

$$\forall x \in \mathcal{C}(H), \quad H_n(x) \xrightarrow[n\to\infty]{} H(x).$$

where $\mathcal{C}(H)$ is the set of continuity points of H, that is

$$\mathcal{C}(H) = \{x \in \mathbb{R} : \ H(x) \in \mathbb{R} \text{ and } H \text{ is continuous at } x. \ \}$$

With this definition, if $(X_n)_{n\geq 0}$ and X are random variables with associated distribution functions $(F_n)_{n\geq 0}$ and F, then we indeed have

$$X_n \xrightarrow{w} X \iff F_n \xrightarrow{w} F \qquad \text{as } n \to \infty.$$

Notations When compositions of functions with affine scalings (or with other simple transformations) are involved, *e.g.* if we consider functions of the kind $x \mapsto F(ax + b)$, we will usually use the notation '$F(a \bullet + b)$' instead. As an example,

$$F_n(a_n \bullet + b_n) \xrightarrow{w} G, \qquad (n \to \infty)$$

means

$$\{x \mapsto F_n(a_n x + b_n)\} \xrightarrow{w} G \qquad (n \to \infty).$$

1.2.2 Weak convergence of the inverse

Definition 1.2.2 (Left-continuous inverse)**.** *Let H be a non-decreasing, right continuous function $\mathbb{R} \to [-\infty, \infty]$. The left-continuous inverse of H is the function*

$$\begin{aligned} H^{\leftarrow} &: \mathbb{R} \to [-\infty, \infty] \\ & y \mapsto H^{\leftarrow}(y) = \inf\{x \in \mathbb{R} : H(x) \geq y\}, \end{aligned}$$

with the convention that $\inf \mathbb{R} = -\infty$ *and* $\inf \emptyset = +\infty$.

Remark 1.2.3. *It is left as an exercise to verify that $H^{\leftarrow}$ is indeed continuous from the left.*

Lemma 1.2.4 (Order relations)
Let $H : \mathbb{R} \to [-\infty, \infty]$ be a non-decreasing, right-continuous function, and let $x \in \mathbb{R}$ and $y \in \mathbb{R}$. Define $A_y = \{t \in \mathbb{R} : H(t) \geq y\}$.
Then A_y is a closed set, and

$$H(x) \geq y \iff x \geq H^{\leftarrow}(y). \tag{1.2}$$

Proof. Notice first that A_y must be either the empty set, or $\mathbb{R}$, or of the form (u, ∞) or $[u, \infty)$, for some $u \in \mathbb{R}$.

1. If $A_y = \mathbb{R}$, A_y is obviously closed, and $H^{\leftarrow}(y) = -\infty$. Then both sides of (1.2) hold for any $x \in \mathbb{R}$.

2. if $A(y) = \emptyset$, A_y is closed again, and $H^{\leftarrow}(y) = +\infty$. Also, $\forall t \in \mathbb{R}, H(t) < y$, thus neither side of (1.2) can be true.

3. Otherwise, consider a sequence $u_n \downarrow u$. Each u_n belongs to A_y, thus $H(u_n) \geq y$. Since H is continuous from the right, $H(u) \geq y$ too, whence $u \in A_y$. Thus $A_y = [u, \infty)$ is closed in $\mathbb{R}$. By definition of $H^{\leftarrow}$, we have $H^{\leftarrow}(y) = \inf A_y = u$. Finally, (1.2) is obtained by noticing that

$$H(x) \geq y \iff x \in A_y \iff x \in [H^{\leftarrow}(y), \infty) \iff x \geq H^{\leftarrow}(y)$$

■

Lemma 1.2.5 (Weak convergence of the inverse)
Let $(H_n)_{n\in\mathbb{N}}$ and H be monotone functions $\mathbb{R} \to [-\infty, \infty]$. If

$$H_n \xrightarrow{w} H \quad \text{as} \quad n \to \infty,$$

then also

$$H_n^{\leftarrow} \xrightarrow{w} H^{\leftarrow} \quad \text{as} \quad n \to \infty.$$

Conversely, if we assume in addition that

(i) For all $n \in \mathbb{N}$, $\inf_{\mathbb{R}} H_n \geq \inf_{\mathbb{R}} H$,

(ii) For all $x \in \mathbb{R}$ such that $H(x) < \infty$, $H(x) < \sup_{t:H(t)<\infty} H(t)$,

Then weak convergence of $H_n^{\leftarrow}$ to $H^{\leftarrow}$ implies weak convergence of H_n to H.

The proof is deferred to Appendix A.2 Notice that the two conditions (i) and (ii) for the converse satement of Lemma 1.2.5 may seem intricate, but they are indeed satisfied in the particular case where we need it (*i.e.* in the proof of Theorem 1.4.1).

1.2.3 Convergence to types

The limiting form in (MDA') will be obtained 'up to rescaling', in the sense defined below.

Definition 1.2.6 (Functions of the same type). *to functions $U, V : \mathbb{R} \to [-\infty, \infty]$ are* of the same type *if $\exists A, B \in \mathbb{R}$, $A > 0$, such that*

$$\forall x \in \mathbb{R}, \quad V(x) = U(Ax + B).$$

The interesting fact about equality in type is that, if (MDA) or (MDA') holds for two different sequences, then the limits must be of the same type and the tails of the two sequences must be linked in the same way, as made precise below.

Lemma 1.2.7 (Convergence to types, Khintchine)
Let $(F_n)_n, U$ be cumulative distribution functions, U being non-degenerate. Let $a_n > 0$ and b_n $(n \in \mathbb{N})$ be two sequences of real numbers, such that

$$F_n(\, a_n \,\bullet\, + \, b_n\,) \xrightarrow{w} U. \tag{1.3}$$

Let $\tilde{a}_n > 0, \tilde{b}_n (n \in \mathbb{N})$ be two other sequences. Then, the following are equivalent:

(i) There exists another non-degenerate c.d.f. *V such that*

$$F_n(\tilde{a}_n \,\bullet\, + \tilde{b}_n) \xrightarrow{w} V$$

(ii) $\exists A > 0, B \in \mathbb{R}$ *such that*

$$\frac{\tilde{a}_n}{a_n} \xrightarrow[n\to\infty]{} A \quad ; \quad \frac{\tilde{b}_n - b_n}{a_n} \xrightarrow[n\to\infty]{} B.$$

Also, if (i) *or* (ii) *hold, then* U *and* V *are of the same type, namely*

$$V(\,\bullet\,) = U(A\,\bullet\, + b) \qquad (x \in \mathbb{R}). \tag{1.4}$$

Proof.
1. (i) $\Rightarrow$ (ii) **and** (1.4)**:**
Assume that (i) holds. Using Lemma 1.2.5, weak convergences in (i) and (1.3) may be inverted, so that

$$\frac{F_n^{\leftarrow} - b_n}{a_n} \xrightarrow{w} U^{\leftarrow} \quad \text{and} \quad \frac{F_n^{\leftarrow} - \tilde{b}_n}{\tilde{a}_n} \xrightarrow{w} V^{\leftarrow}$$

Non-degeneracy allows to pick $y_1 < y_2 \in \mathcal{C}(U^{\leftarrow}) \cap \mathcal{C}(V^{\leftarrow})$ such that $U^{\leftarrow}(y_1) < U^{\leftarrow}(y_2)$ and $V^{\leftarrow}(y_1) < V^{\leftarrow}(y_2)$. Thus $\frac{F_n^{\leftarrow}(y_i) - b_n}{a_n} \xrightarrow{w} U^{\leftarrow}(y_i)$, $i = 1, 2$. By substraction,

$$\frac{F_n^{\leftarrow}(y_2) - F_n^{\leftarrow}(y_1)}{a_n} \xrightarrow{w} U^{\leftarrow}(y_2) - U^{\leftarrow}(y_1).$$

In the same way, we have

$$\frac{F_n^{\leftarrow}(y_2) - F_n^{\leftarrow}(y_1)}{\tilde{a}_n} \xrightarrow{w} V^{\leftarrow}(y_2) - V^{\leftarrow}(y_1).$$

Dividing the two (which is possible since the limits are nonzero) yields

$$\frac{\tilde{a}_n}{a_n} \xrightarrow[n\to\infty]{} A := \frac{U^{\leftarrow}(y_2) - U^{\leftarrow}(y_1)}{V^{\leftarrow}(y_2) - V^{\leftarrow}(y_1)} > 0.$$

Also for $y \in \mathcal{C}(U^{\leftarrow}) \cap \mathcal{C}(V^{\leftarrow})$,

$$\frac{F_n^{\leftarrow}(y) - b_n}{a_n} - A\frac{F_n^{\leftarrow}(y) - \tilde{b}_n}{\tilde{a}_n} \xrightarrow{w} U^{\leftarrow}(y) - AV^{\leftarrow}(y).$$

However,

$$\begin{aligned}
\frac{F_n^{\leftarrow}(y) - b_n}{a_n} - A\frac{F_n^{\leftarrow}(y) - \tilde{b}_n}{\tilde{a}_n} &= \frac{F_n^{\leftarrow}(y) - b_n}{a_n} - A\frac{a_n}{\tilde{a}_n}\frac{F_n^{\leftarrow}(y) - \tilde{b}_n}{a_n} \\
&\sim_{n\to\infty} \frac{F_n^{\leftarrow}(y) - b_n}{a_n} - \frac{F_n^{\leftarrow}(y) - \tilde{b}_n}{a_n} \\
&= \frac{\tilde{b}_n - b_n}{a_n}
\end{aligned}$$

whence $\frac{\tilde{b}_n - b_n}{a_n} \to B := U^{\leftarrow}(y) - AV^{\leftarrow}(y)$; and (ii) is proved.

Another consequence is that the function $y \mapsto U^{\leftarrow}(y) - AV^{\leftarrow}(y)$ is identically equal to B on $\mathcal{C}(V^{\leftarrow}) \cap \mathcal{C}(U^{\leftarrow})$, so that $V^{\leftarrow}(y) = \frac{U^{\leftarrow}(y) - B}{A} = U[A\,\bullet\, + B]^{\leftarrow}$. By continuity from the left this identity holds everywhere on $\mathbb{R}$. We obtain

$$V^{\leftarrow} = [U(A\,\bullet\, + B)]^{\leftarrow}. \tag{1.5}$$

In order to conclude that $V = U(A\,\cdot\,+B)$, we need to show that, for two non decreasing functions G_1, G_2 such that $G_1^{\leftarrow} = G_2^{\leftarrow}$, it holds that $G_1 = G_2$. To see this, write for $x \in \mathbb{R}$,

$$\begin{aligned} G_1(x) &= sup\{y : \; G_1(x) \geq y\} \\ &= \sup\{y : \; x \geq G_1^{\leftarrow}(y)\} \qquad \text{from Lemma 1.2.4} \\ &= \sup\{y : \; x \geq G_2^{\leftarrow}(y)\} \\ &= G_2(x). \end{aligned}$$

2. (ii)$\Rightarrow$ (i) **and** (1.4)**:** Put $V(x) = U(Ax+B)$, $x \in \mathbb{R}$. Then $V^{\leftarrow}(y) = A^{-1}(U^{\leftarrow}(y) - B), y \in \mathbb{R}$. Reversing the argument leading to (1.5), we obtain, for $y \in \mathcal{C}(U^{\leftarrow}) = \mathcal{C}(V^{\leftarrow})$,

$$\begin{aligned} F_n(\tilde{a}_n \bullet + \tilde{b}_n)^{\leftarrow}(y) = \frac{F_n^{\leftarrow}(y) - \tilde{b}_n}{\tilde{a}_n} &= \frac{a_n}{\tilde{a}_n}\left(\frac{F_n^{\leftarrow}(y) - b_n}{a_n} - \frac{(\tilde{b}_n - b_n)}{a_n}\right) \\ &\xrightarrow[n\to\infty]{} A^{-1}(U^{\leftarrow}(y) - B) = V^{\leftarrow}(y). \end{aligned} \tag{1.6}$$

We have shown that $F_n(\tilde{a}_n \bullet + \tilde{b}_n)^{\leftarrow} \xrightarrow{w} V^{\leftarrow}$, which implies, by Lemma 1.2.5, that $F_n(\tilde{a}_n \bullet + \tilde{b}_n) \xrightarrow{w} V$, which is (i).

Since we have already proved that (i) forces $V(x) = U(Ax+B)$, the proof is complete. ∎

1.3 'Fundamental theorem' of EVT: Limit laws for maxima

1.3.1 Max-stable distributions

Getting back to our analogy with the CLT, remind that the limiting distribution $\mathcal{N}$ of rescaled sums (a Gaussian distribution) is *stable*, that is, if $X_i \overset{i.i.d.}{\sim} \mathcal{N}$, then for $n \in \mathbb{N}$, $\exists A_n, B_n$: $\frac{\sum_1^n X_i - B_n}{A_n} \overset{\text{d}}{=} X_1$. Here and thereafter, '$\overset{\text{d}}{=}$' means equality *in distribution.*

Replacing the sum-operator by the max-operator, one may reasonably expect an analogous property for extreme value distributions (*i.e.* the limit distributions G in (MDA')). It is indeed the case, if one consider *max-stability* instead of *stability*, as defined below.

Definition 1.3.1 (Max-stable distribution)**.** *A* c.d.f. G *is called* max-stable *if there exist functions* $\alpha(t) > 0, \beta(t)$ $(t > 0)$ *such that*

$$\forall t > 0, \forall x \in \mathbb{R}, \qquad G^t(\alpha(t)x + \beta(t)) = G(x).$$

In particular, if $(Z_i)_{i=1,\dots,n} \overset{i.i.d.}{\sim} G$, then $\bigvee_{i=1}^n Z_i \sim G^n$, so that, letting $\alpha_n = \alpha(n), \beta_n = \beta(n)$,

$$\frac{\bigvee_{i=1:n} Z_i - \beta_n}{\alpha_n} \overset{\text{d}}{=} Z_1.$$

Proposition 1.3.2 (Max-stable and extreme value distributions are the same)
Let G be a non-degenerate cumulative distribution function. Then G is an extreme value distribution if and only if it is max-stable.

Proof. If G is max-stable, it is obviously an extreme value distribution: (MDA') holds with $a_n = \alpha(n), b_n = \beta(n)$.

Conversely, assume that (MDA') holds for some F and sequences $a_n > 0, b_n$. Fix $t > 0$. On the one hand, for $x \in \mathcal{C}(G)$,

$$F^{\lfloor nt \rfloor}(a_{\lfloor nt \rfloor}x + b_{\lfloor nt \rfloor}) \xrightarrow[n\to\infty]{} G(x). \tag{1.7}$$

Also,

$$F^{\lfloor nt \rfloor}(a_n x + b_n) = (F^n(a_n x + b_n))^{\frac{\lfloor nt \rfloor}{n}} \xrightarrow[n\to\infty]{} G^t(x). \tag{1.8}$$

Using Khintchine Lemma 1.2.7, with $F_n = F^{nt}, U = G^t, V = G$ (note that G^t is necessarily a non-degenerate *c.d.f.* if G is so), there exist two real numbers $\alpha(t) > 0, \beta(t)$, such that

$$\frac{a_{\lfloor nt \rfloor}}{a_n} \xrightarrow[n\to\infty]{} \alpha(t)\ ; \quad \frac{b_{\lfloor nt \rfloor} - b_n}{a_n} \xrightarrow[n\to\infty]{} \beta(t) \tag{1.9}$$

and

$$G(x) = G^t(\alpha(t)x + \beta(t)), \quad x \in \mathbb{R}.$$

■

Lemma 1.3.3
The functions $t \mapsto \alpha(t) > 0$ and $t \mapsto \beta(t)$ in the definition of a max-stable distribution are uniquely determined by G, and are Borel-measurable.

Proof.
(1) To show that α and β are unique, it is enough to show that if a non degenerate *c.d.f.* G satisfies

$$G(x) = G(ax + b), \quad x \in \mathbb{R},$$

for some $a > 0, b \in \mathbb{R}$, then necessarily $a = 1, b = 0$. Define $T : x \mapsto ax + b$. The assumption rewrites $G = G \circ T$. Thus, $G = G \circ T^n$, for $n \in \mathbb{N}$. Thus, for $x \in \mathbb{R}$, $G(x) = \lim_n G(T^n x)$. It is then easy to see (exercise 1.1) that if $a \neq 1$, G must be degenerate, and then that b must be null.
(2) The argument leading to (1.9) in the proof of Proposition 1.3.2, with F replaced with G, shows that for $t > 0$,

$$\frac{\alpha(\lfloor nt \rfloor)}{\alpha(n)} \xrightarrow[n\to\infty]{} \alpha(t)\ ; \quad \frac{\beta(\lfloor nt \rfloor) - \beta(n)}{\alpha(n)} \xrightarrow[n\to\infty]{} \beta(t) \tag{1.10}$$

Now the functions $t \mapsto \frac{\alpha(n)}{\alpha(\lfloor nt \rfloor)}$ and $t \mapsto \frac{\beta(n) - \beta(\lfloor nt \rfloor)}{\alpha(\lfloor nt \rfloor)}$ are certainly measurable (they are piecewise constant). Since the pointwise limits of measurable functions are measurable, α and β are measurable. ■

Exercise 1.1:
Complete the proof of Lemma 1.3.3 (1): use the fact that for $x \in \mathbb{R}$, the sequence $(T^n x)_n$ is arithmetico-geometric, so that if $a \neq 1$, $\lim_n T^n x$ is either infinite, or a fixed point.

The next paragraph is the core of this chapter

1.3.2 Characterizing max-stable distribution

Before stating the result, notice that characterizing max-stable distributions is the same as characterizing extreme value distributions (the possible limits in (MDA')), according to Proposition 1.3.2.

Theorem 1.3.4 (Extreme value theorem (Fisher & Tipett 1928, Gnedenko 1943))
If G is a max-stable distribution, G is of one of the three types

(i) Fréchet : $\Phi_\alpha(x) = \begin{cases} e^{-x^{(-\alpha)}} & (x > 0) \\ 0 & (x \leq 0). \end{cases}$,
With $\alpha > 0$ *;*

(ii) Weibull : $\Psi_\alpha(x) = \begin{cases} e^{-(-x)^{(-\alpha)}} & (x < 0) \\ 1 & (x \geq 0). \end{cases}$,
With $\alpha < 0$*;*

(iii) Gumbel : $\Lambda(x) = e^{-e^{-x}}$*,* $x \in \mathbb{R}$*.*

It is convenient to use a common parametrization for the three types, as in the following statement (the verification is left to the reader):

Corollary 1.3.5
If G is a max-stable distribution, then $\exists \mu \in \mathbb{R}, \sigma > 0, \gamma \in \mathbb{R}$*, such that*

$$G(x) = G_{\mu,\sigma,\gamma}(x) := \exp\left[-\left(1 + \gamma\left(\frac{x-\mu}{\sigma}\right)\right)_+^{-1/\gamma}\right], \tag{1.11}$$

where $y_+ = \max(y, 0)$*, and where the above expression for* $\gamma = 0$ *is understood as its limit as* $\gamma \to 0$*, that is*

$$G(x) = \exp\left[-e^{-\frac{x-\mu}{\sigma}}\right].$$

Also,

- $\gamma = 0$ *if and only if G is of Gumbel type,*
- $\gamma > 0$ *if and only if G is of Fréchet type* Φ_α *with* $\alpha = 1/\gamma$*,*
- $\gamma < 0$ *if and only if G is of Weibull type* Ψ_α *with* $\alpha = 1/\gamma$*.*

Before some examples and the proof, Figures 1.2, 1.3 and 1.4 illustrate the three types. The first two figures explain why the distribution functions in the Fréchet domain of attraction are9+ usually referred to as *heavy tailed*, whereas those in the Gumbel domain are called *light tailed* (There is no agreement about the Weibull domain. Some authors use 'light tails', some others use 'bounded tails'). Also, Figure 1.4 indicates that a series of i.i.d. observations of heavy tailed variables is likely to contain more 'extreme' events than a series of light tailed variables: The Fréchet type corresponds to situation where extreme events occur 'quite often'. Typical examples include river discharge data, rainfall (in some cases), financial return times series, insurance claims.

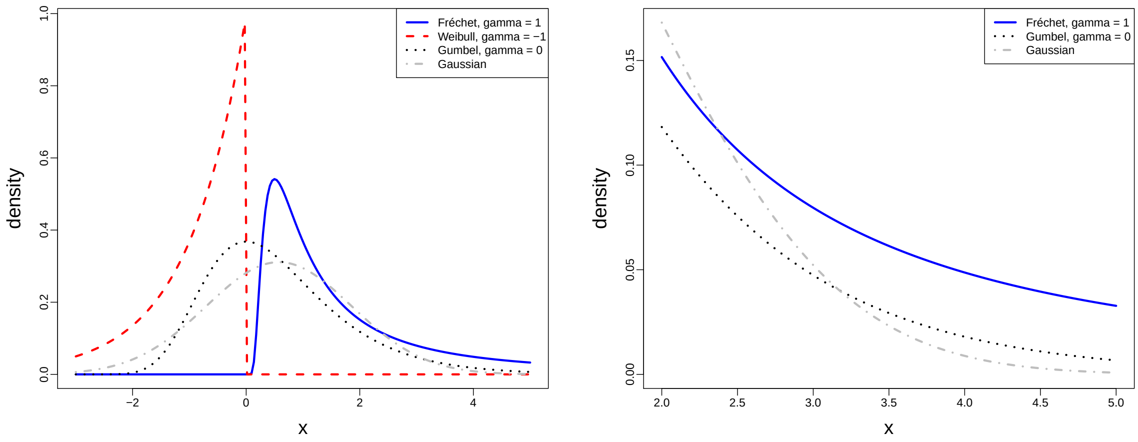

Figure 1.2: Density plot for the three extremal types, respectively $(\gamma = 1, \mu = 1, \sigma = 1)$, $(\gamma = -1, \mu = -1, \sigma = 1)$, $(\gamma = 0, \mu = 0, \sigma = 1)$; compared with the Gaussian density with same mean and variance as the Gumbel one. The right panel is a zoom on the tail.

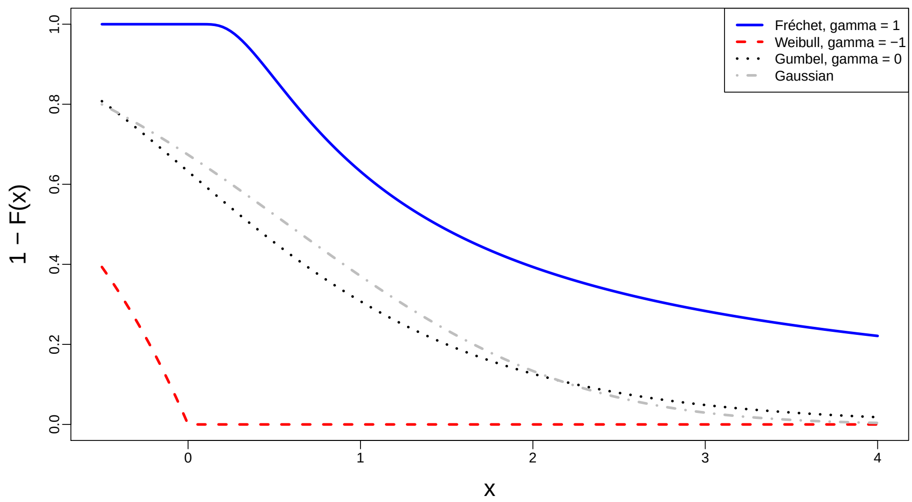

Figure 1.3: Survival function $1 - F(x)$ for the three extremal types, respectively $(\gamma = 1, \mu = 1, \sigma = 1)$, $(\gamma = -1, \mu = -1, \sigma = 1)$, $(\gamma = 0, \mu = 0, \sigma = 1)$; compared with the Gaussian survival function with same mean and variance as the Gumbel one.

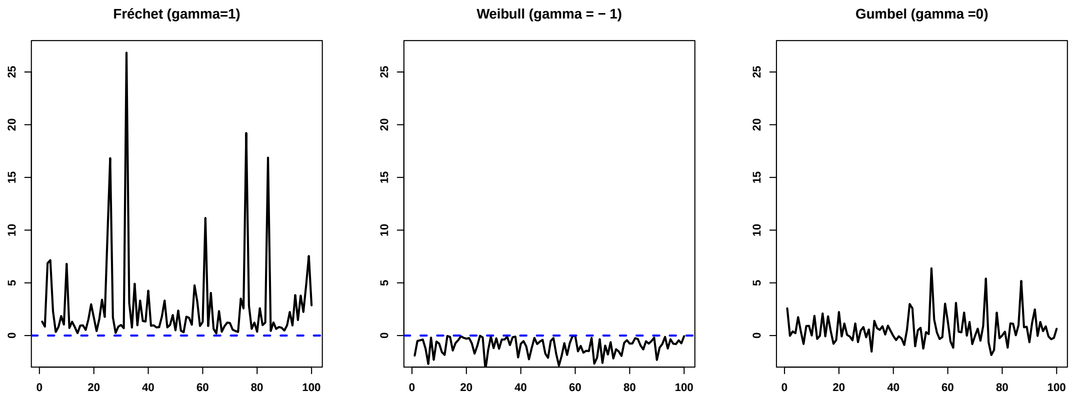

Figure 1.4: Series of i.i.d. random variables of the three extremal types, respectively $(\gamma = 1, \mu = 1, \sigma = 1)$, $(\gamma = -1, \mu = -1, \sigma = 1)$, $(\gamma = 0, \mu = 0, \sigma = 1)$

Example 1.1 (Exponential variable, Gumbel domain)**:**
Let F be an exponential distribution,

$$F(x) = \mathbb{1}_{x>0}\left(1 - e^{-\lambda x}\right).$$

In order to 'guess' possible norming constant, we shall proceed with heuristic computations, and prove in a second step that the sequences are indeed suitable. We may assume that for $x > 0$, $a_n x + b_n \xrightarrow[n\to\infty]{} \infty$ (otherwise,, $F^n(a_n x + b_n) \xrightarrow[n\to\infty]{} 0$. Thus

$$\begin{aligned} F^n(a_n x + b_n) &= \mathbb{1}_{a_n x + b_n > 0}\ (1 - e^{-\lambda(a_n x + b_n)})^n \\ &= \mathbb{1}_{a_n x + b_n > 0}\ \exp\left(n \log(1 - e^{-\lambda a_n x + \lambda b_n})\right) \\ &\approx \mathbb{1}_{a_n x + b_n > 0}\ \exp\left(n e^{-\lambda a_n x - \lambda b_n}\right) \\ &= \mathbb{1}_{a_n x + b_n > 0}\ \exp\left(e^{-\lambda a_n x + \log n - \lambda b_n}\right) \end{aligned}$$

If we set $a_n = 1, b_n = \lambda^{-1}\log n$, the latter expression does converge to $G(x) = \exp\left(-e^{-\lambda x}\right)$, which is of Gumbel type. Now we only need to check that the (MDA') condition is indeed satisfied with these sequences: for $x \in \mathbb{R}$,

$$\begin{aligned} F^n(a_n x + b_n) &= \mathbb{1}_{x + \log n/\lambda > 0}(1 - e^{-\lambda(x + \log(n)/\lambda)})^n \\ &= \exp\left(n \log(1 - e^{-\lambda x - \log(n))})\right) \qquad (n \geq e^{-\lambda x}) \\ &= \exp\left(n \log(1 - \frac{e^{-\lambda x}}{n})\right) \\ &= \exp\left(n\left(-\frac{e^{-\lambda x}}{n} + o(1/n)\right)\right) \\ &\xrightarrow[n\to\infty]{} \exp\left(-e^{-\lambda x}\right). \end{aligned}$$

Exercise 1.2 (Weibull domain)**:**
Check that the standard uniform variable is in the Weibull domain of attraction.
Hint: consider $a_n = 1/n$, $b_n = 1 - 1/n$. What is the corresponding limit distribution G ?

Exercise 1.3 (Fréchet domain of attraction)**:**
Show that the Pareto distribution $F(x) = \mathbb{1}_{x>u}\left(1 - \frac{u^\alpha}{x^\alpha}\right)$, where $\alpha > 0$, $u > 0$, belongs to the Fréchet max-domain of attraction. Exhibit suitable sequences a_n and b_n and explicit the limit G.

Proof of Theorem 1.3.4. Our proof mainly follows Resnick (1987).

Let $\alpha > 0$, β the norming functions such that for $t > 0$, $x \in \mathbb{R}$,

$$G(x) = G^t(\alpha(t)x + \beta(t)). \tag{1.12}$$

The key to the proof is to show that α and β satisfy a particular functional equation (the Hamel equation, see below), which solutions are known, and then to obtain the expression of G using (1.12) again.

First, for $t, s > 0$, $x \in \mathbb{R}$,

$$G^{1/(ts)}(x) = G(\alpha(ts)x + \beta(ts)).$$

but also

$$G^{1/(ts)}(x) = [G^{1/s}(x)]^{1/t} = G^{1/t}(\alpha(s)x + \beta(s)) = G(\alpha(t)\alpha(s)x + \alpha(t)\beta(s) + \beta(t)).$$

By Lemma 1.3.3, this implies that for $t, s > 0$,

$$\alpha(ts) = \alpha(t)\alpha(s) \tag{1.13}$$
$$\beta(ts) = \alpha(t)\beta(s) + \beta(t) = \alpha(s)\beta(t) + \beta(s), \tag{1.14}$$

where the last equality follows by interchanging the roles of s and t.

One recognizes in (1.13) the Hamel equation. It is easy to prove that the only continuous solutions of this equation are of the form $f(t) = t^\gamma$, for some $\gamma \in \mathbb{R}$ (this is obvious for $\log t \in \mathbb{N}$, then by inversion also for $\log t \in \mathbb{Q}$, and continuity achieves the proof for $t \in \mathbb{R}$.) In fact, it may be shown (See Hahn and Rosenthal (1948), pp. 116-118) that the only *measurable* solutions are also of this kind. Now, by 1.3.3, we know that α and β are measurable. Whence, $\exists \gamma \in \mathbb{R}$:

$$\forall t > 0, \quad \alpha(t) = t^\gamma$$

We distinguish three cases according to the sign of γ (to wit, γ will be the extreme value index appearing in (1.11))

case 1: $\gamma = 0$ In this case $\alpha \equiv 1$. Thus (1.14) yields $\beta(ts) = \beta(t) + \beta(s), \quad s, t > 0$. This is again the Hamel equation (up to log-scaling, that is: e^β satisfies (1.13)). Consequently, for some $\sigma \in \mathbb{R}$, $e^{\beta(t)} = t^\sigma$ (to wit, σ will be the scale parameter in (1.11)), that is

$$\beta(t) = \sigma \log t, \quad s, t > 0. \tag{1.15}$$

Going back to (1.12), we have

$$G^{1/t}(x) = G(x + \sigma \log t), \quad x \in \mathbb{R}, t > 0.$$

For x such that $0 < G(x) < 1$ (which exists by non-degeneracy of G), the function $t \mapsto G^{1/t}(x)$ is strictly increasing on $]1, \infty[$, thus $t \mapsto \sigma \log t$ must be strictly increasing , which means $\sigma > 0$. Then (1.12) with $x = 0$ yields $\forall t > 0, G(\sigma \log t) = G(0)^{1/t}$, *i.e.*, with $u = \sigma \log t$,

$$\forall u \in \mathbb{R}, G(u) = (G(0))^{e^{-u/\sigma}}$$

necessarily , $0 < G(0) < 1$, otherwise G would be constant on $\mathbb{R}$. Thus

$$\forall u \in \mathbb{R}, \quad G(u) = \exp\left[-e^{-u/\sigma}(-\log G(0))\right] = \exp\left[-e^{-(u-\mu)/\sigma}\right]$$

where μ is chosen so that $e^{\mu/\sigma} = -\log G(0)$, *i.e.* $\mu = \sigma \log(-\log G(0))$. Thus G is of Gumbel type ($G(x) = \Lambda((x - \mu)/\sigma)$.

case 2 : $\gamma \neq 0$ In this case, identity (1.14) implies, for $s, t > 0$,

$$\beta(t)(s^\gamma - 1) = \beta(s)(t^\gamma - 1).$$

Thus $t \mapsto \beta(t)/(t^\gamma - 1)$ (for $t \neq 1$) is constant, *i.e.*

$$\exists C \in \mathbb{R} : \quad \beta(t) = C(t^\gamma - 1) \qquad (t \neq 1).$$

Going back to (1.12), we obtain, for $x \in \mathbb{R}$,

$$G^{1/t}(x) = G(t^\gamma x + C(t^\gamma - 1)) = G[t^\gamma(x + C) - C],$$

so that $G^{1/t}(x - C) = G[t^\gamma x - C]$. Whence, putting $\Gamma(x) = G(x + C)$,

$$\Gamma^{1/t}(x) = \Gamma(t^\gamma x), \quad t > 0, x \in \mathbb{R}. \tag{1.16}$$

Note that this implies (setting $x = 0$ in the above equation) that $\Gamma(0) \in \{0, 1\}$. Also, $\Gamma(1) > 0$, otherwise Γ would be identically equal to 0. To conclude, it is enough to show that Γ if of one of the two first types (Fréchet or Gumbel). This will depend on the sign of γ.

1. **Case** $\gamma > 0$ Let us prove that we must have $\Gamma(1) < 1$: otherwise we would have for $t > 0$, $1 = \Gamma(t^\gamma)$, so that $\Gamma(0) = \lim_{t\to 0} \Gamma(t^\gamma) = 1$. But then $\exists x < 0$ such that $0 < \Gamma(x) < 1$, and for such an x the function $t \mapsto \Gamma^{1/t}(x)$ is strictly increasing. However, $\Gamma^{1/t}(x) = \Gamma(t^\gamma x)$, which is a non increasing function of t, a contradiction. Thus $0 < \Gamma(1) < 1$.

   We may thus rewrite (1.16) as (with $u = t^\gamma$, and $x = 1$)

   $$\Gamma(u) = \Gamma(1)^{u^{-1/\gamma}} = \exp\left[-u^{-1/\gamma}(-\log \Gamma(1))\right] = \exp\left[-(u/\sigma)^{-1/\gamma}\right], \quad u > 0. \tag{1.17}$$

   with $\sigma = (-\log \Gamma(1))^\gamma$. Thus $G(x) = \Gamma(x + C) = \Phi_{1/\gamma}((x + C)/\sigma$ (Fréchet type).

2. **Case** $\gamma < 0$**:** with a similar argument, one obtains that G is of the Weibull type.

■

Remark 1.3.6 (Choice of norming sequences and parameters of the limit)**.** *If F satisfies a MDA condition for some sequences (a_n, b_n) and if the limit is of the form $G_{\mu,\sigma,\gamma}(x)$ as in* (1.11)*, then it is always possible to choose other sequences a'_n, b'_n such that*

$$F^n(a'_n x + b'_n) \xrightarrow{w} G_{0,1,\gamma}$$

where $G_{0,1,\gamma}(x) = \exp\Big(-(1 + \gamma x)_+^{-1/\gamma}\Big)$. Indeed, one may choose $a'_n = \sigma a_n, b'_n = b_n + \mu a_n$, and use the convergence to type lemma 1.2.7.

Remark 1.3.7 (Continuity set of the limit)**.** *Since the max-stable distributions are continuous on $\mathbb{R}$ (this is obvious from their parametric form* (1.11)*), if F is in the domain of attraction of G, then convergence must occur for all $x \in \mathbb{R}$. In other words, in this case, weak convergence is the same as pointwise convergence on $\mathbb{R}$.*

1.4 Equivalent formulations in terms of excesses above thresholds

Our goal is to show that the condition (MDA') is equivalent to the convergence of the conditional distribution of excesses above t, in the following sense

Theorem 1.4.1 (Balkema, de Haan, 1974)
The following statements are equivalent

(i) $\exists a_n > 0, b_n$*: for all* $x \in \mathbb{R}$, $F^n(a_n x + b_n) \xrightarrow[n\to\infty]{} e^{-(1+\gamma x)_+^{-1/\gamma}}$

(ii) $\exists \sigma : (0,\infty) \to (0,\infty)$ *such that, for each* x *such that* $1+\gamma x > 0$,

$$\mathbb{P}\left(\frac{X-t}{\sigma(t)} > x \mid X > t\right) \underset{t\to x_\star}{\longrightarrow} -\log G(x) = (1+\gamma x)^{-1/\gamma}, \tag{1.18}$$

where $x_\star = F^{\leftarrow}(1)$ *is the right end-point of the support of* F*; which means in terms of distribution functions that*

$$\frac{1-F(t+\sigma(t)x)}{1-F(t)} \underset{t\nearrow x_\star}{\longrightarrow} (1+\gamma x)_+^{-1/\gamma}. \tag{1.19}$$

In such a case, σ *may be chosen as* $\sigma(t) = a\left(\frac{1}{1-F(t)}\right)$.

For the proof, we will use a series of equivalent characterization of the (MDA') condition in terms of survival functions $1-F$ and inverse functions.

Lemma 1.4.2 (Convergence of survival functions)
The (MDA') *condition is satisfied if and only if*

$$n(1-F(a_n \bullet + b_n)) \xrightarrow{w} -\log G.$$

Proof. By continuity of the logarithm function and its inverse,

$$(\text{MDA'}) \iff n \log F(a_n \bullet + b_n) \xrightarrow{w} \log G.$$

Now on both sides, for x such that $\log G(x)$ is finite, $F(a_n x + b_n)$ must converge to 1, thus

$$\log F(a_n x + b_n) = \log(1-(1-F(a_n x + b_n))) \sim_{n\to\infty} -[1-F(a_n x + b_n)],$$

whence the result. ∎

An immediate consequence is that (MDA') is equivalent to

$$\frac{1}{n(1-F(a_n \bullet + b_n))} \xrightarrow{w} \frac{-1}{\log G}. \tag{1.20}$$

Let $U = \left(\frac{1}{1-F}\right)^{\leftarrow}$ (*i.e.* $U(y) = F^{\leftarrow}(1-1/y), y>0$) and $\Gamma = \frac{-1}{\log G}$. From Lemma 1.2.5.

$$(1.20) \iff \frac{U(n\bullet) - b_n}{a_n} \xrightarrow{w} \Gamma^{\leftarrow} \quad \text{as } n\to\infty \tag{1.21}$$

Define $a(t) = a_{\lfloor t \rfloor}, b(t) = b_{\lfloor t \rfloor}$, $t>0$. The next lemma extends the above equality to all $t>0$.

Lemma 1.4.3
The (MDA') *condition is satisfied if and only if*

$$\frac{U(t\bullet) - b(t)}{a(t)} \xrightarrow{w} \Gamma^{\leftarrow}, \quad \text{as } t\to\infty \tag{1.22}$$

Proof. We only need to show that (MDA') implies (1.22). Indeed, the converse is immediate from (1.21).

Let $y \in \mathcal{C}(\Gamma^{\leftarrow})$. By monotonicity of U,

$$\frac{U(\lfloor t \rfloor y) - b(t)}{a(t)} \leq \frac{U(ty) - b(t)}{a(t)} \leq \frac{U((\lfloor t \rfloor + 1)y) - b(t)}{a(t)} \tag{1.23}$$

Fix $\epsilon >$, and choose $y' > y$ such that $\Gamma^{\leftarrow}(y') < \Gamma^{\leftarrow}(y) + \epsilon$. Then for some t_0 large enough and $t > t_0$, $(\lfloor t \rfloor + 1)y < \lfloor t \rfloor y'$. Thus for large t, $\frac{U((\lfloor t \rfloor+1)y)-b(t)}{a(t)} \leq \frac{U(\lfloor t \rfloor y')-b(t)}{a(t)} \to \Gamma^{\leftarrow}(y') \leq \Gamma^{\leftarrow}(y) + \epsilon$. Since the limit of the left-hand side of (1.23) is $\Gamma^{\leftarrow}(y)$, and since ϵ is arbitrary, the proof is complete. ■

Remark 1.4.4 (Continuity points of $\Gamma^{\leftarrow}$)**.** *Notice that weak convergence in* (1.21) *and* (1.22) *is equivalent to pointwise convergence for* $y > 0$. *Indeed,* $\Gamma = -1/\log G$ *induces a bijection (it is strictly increasing and continuous) from the interior of its support onto* $(0, \infty)$. *Thus, its left inverse is a real inverse and is also continuous on* $(0, \infty)$.

We may now proceed with the proof of the main result of this section.

Proof of Theorem 1.4.1. We prove that (MDA') implies (1.19); the proof of the converse is similar and is left as an exercise. Put $\sigma(t) = a(\frac{1}{1-F(t)})$ It is easily verified that the left-continuous inverse of the function

$$x \mapsto \frac{1 - F(t)}{1 - F(t + x\sigma(t))}$$

is

$$y \mapsto \frac{U(\frac{y}{1-F(t)}) - t}{\sigma(t)}.$$

Using Lemma 1.2.5 and Remark 1.4.4, it is thus enough to show that

$$\forall y \geq 1, \quad \frac{U\left(\frac{y}{1-F(t)}\right) - t}{\sigma(t)} \xrightarrow[t \nearrow x_\star]{} \frac{y^\gamma - 1}{\gamma} := \Gamma^{\leftarrow}(y). \tag{1.24}$$

However, using (1.22) from Lemma 1.4.3 for $y = 1$, we have

$$\frac{U(T) - b(T)}{a(T)} \xrightarrow[T \to \infty]{} \Gamma^{\leftarrow}(1) = \frac{1^\gamma - 1}{\gamma} = 0$$

But also for $y > 0$,

$$\frac{U(Ty) - b(T)}{a(T)} \xrightarrow[T \to \infty]{} \frac{y^\gamma - 1}{\gamma}.$$

By substraction,

$$\frac{U(Ty) - U(T)}{a(T)} \xrightarrow[T \to \infty]{} \frac{y^\gamma - 1}{\gamma}. \tag{1.25}$$

N.B: If we could replace T with $1/(1 - F(t))$, and t with $U(T)$ in (1.25), we would obtain (1.24) and the proof would be complete. This is the idea behind the remainder of the proof.

It is easy to show that if f is a right-continuous, non decreasing function, for $\epsilon > 0$, we have $f^{\leftarrow}(f(t)) \leq t \leq f^{\leftarrow}(f(t)+\epsilon)$. Thus, for $y > 0, 0 < t < x_\star$,

$$\begin{aligned}
0 \leq \frac{t - U\left(\frac{1}{1-F(t)}\right)}{a\left(\frac{1}{1-F(t)}\right)} &\leq \frac{U\left(\frac{1}{1-F(t)} + \epsilon\right) - U\left(\frac{1}{1-F(t)}\right)}{a\left(\frac{1}{1-F(t)}\right)} \\
&\leq \frac{U\left(\frac{1}{1-F(t)}(1+\epsilon)\right) - U\left(\frac{1}{1-F(t)}\right)}{a\left(\frac{1}{1-F(t)}\right)} \\
&\xrightarrow[t \nearrow x_\star]{} \Gamma^{\leftarrow}(1+\epsilon) = \frac{(1+\epsilon)^\gamma - 1}{\gamma}, \qquad (1.26)
\end{aligned}$$

where the last limit is obtained from (1.25) and the fact that $1/(1-F(t)) \xrightarrow[t \nearrow x_\star]{} +\infty$ (indeed, in case $x_\star < \infty$, F cannot have a jump at $x_\star$, see *e.g.* Leadbetter et al. (2012), Corollary 1.5.2). Since ϵ is arbitrary small, we conclude that

$$\frac{t - U\left(\frac{1}{1-F(t)}\right)}{a\left(\frac{1}{1-F(t)}\right)} \xrightarrow[t \nearrow x_\star]{} 0. \qquad (1.27)$$

As a consequence, for $y > 0$,

$$\begin{aligned}
\frac{U\left(\frac{y}{1-F(t)}\right) - t}{\sigma(t)} &\underset{t \nearrow x_\star}{\sim} \frac{U\left(\frac{y}{1-F(t)}\right) - U\left(\frac{1}{1-F(t)}\right)}{\sigma(t)} \\
&\xrightarrow[t \nearrow x_\star]{} \frac{y^\gamma - 1}{\gamma},
\end{aligned}$$

where the last limit is obtained from (1.25) as in (1.26). This shows (1.24) and completes the proof. ■

1.5 Case studies

The common idea between most statistical applications is to use the limits in the different convergence results presented in the above sections as *models* for the extremal data, where 'extremal' can be understood either as 'a maximum over a long period' or 'an excess above a high threshold'. A variant of the 'excess' view is to consider the *point process* of excesses above thresholds and use a Poisson approximation. This will be treated in the next chapter.

1.5.1 Annual maximum of the sea level

In order to fix a reasonable premium for real estate insurance, an insurance company is interested to potential damage induced from floods in a city close to the sea level. A dike does protect the city as long as the sea level is below some fixed level u_0. The question is : what is the probability of a flood occurring during a given year ? It may be shown that under weak temporal dependence (with mixing conditions), the extreme value theorem still holds. Thus, one may use the approximation for the annual maximum M_n $(n = 365)$:

$$\frac{M_n - b_n}{a_n} \overset{d}{\approx} Z$$

where $Z \sim G$ is a standard EV distribution, $G(x) = e^{-(1+\gamma x)^{-1/\gamma}}$, and a_n, b_n are unknown parameters. In other words, dropping the index n (which is fixed to 365), and setting $\mu = b_n$, $\sigma = a_n$ the assumption is

$$\mathbb{P}(M \leq x) = \mathbb{P}((M-\mu)/\sigma \leq (x-\mu)/\sigma) \simeq G((x-\mu)/\sigma) = \exp\left[-\left(1+\gamma\frac{x-\mu}{\sigma}\right)_+^{-1/\gamma}\right].$$

Thus, we assume that $M \sim G_{\mu,\sigma,\gamma}$ for some unknown (μ, σ, γ); in other words the statistical model for M is the parametric model

$$\mathcal{P} = \{G_{\mu,\sigma,\gamma} : \quad \mu \in \mathbb{R}, \sigma > 0, \gamma \in \mathbb{R}.\}$$

A widely used approach for inference of the is the maximum likelihood approach. It is implemented in numerous `R` models such as `ismev, extRemes, evd, fExtremes, EVIM, Xtremes, HYFRAN, EXTREMES` ... In our examples, we mainly use `evd` and `ismev`. Notice that it is also possible to resort to probability weighted moment methods. The dataset `portpirie` is part of these two packages. It contains annual maxima of the sea level at Port Pirie (Australia) (Figure 1.5), where a disastrous flood occur ed in 1934 (Figure 1.6).

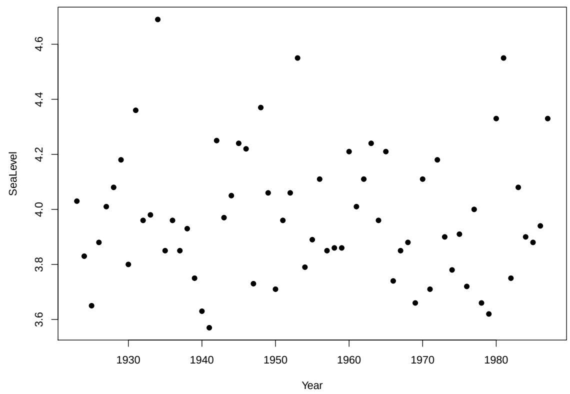

Figure 1.5: `portpirie` data in package `evd`: Annual maxima of the sea level at Port Pirie, 1923-1987

Figure 1.6: 1934 flood at Port Pirie, Australia

The next few lines of code show how to proceed with MLE estimation and obtain diagnostic plots (Figure 1.5.1).

```
> library(evd)
> fitgevpirie <- fgev(portpirie)
> fitgevpirie


Call: fgev(x = portpirie)
Deviance: -8.678117

Estimates
     loc     scale     shape
 3.87475   0.19805  -0.05012

Standard Errors
    loc    scale    shape
0.02793 0.02025 0.09826

Optimization Information
  Convergence: successful
  Function Evaluations: 30
  Gradient Evaluations: 8

> plot(fitgevpirie)
```

The probability of an excess of any threshold u may now reasonably be estimated by a

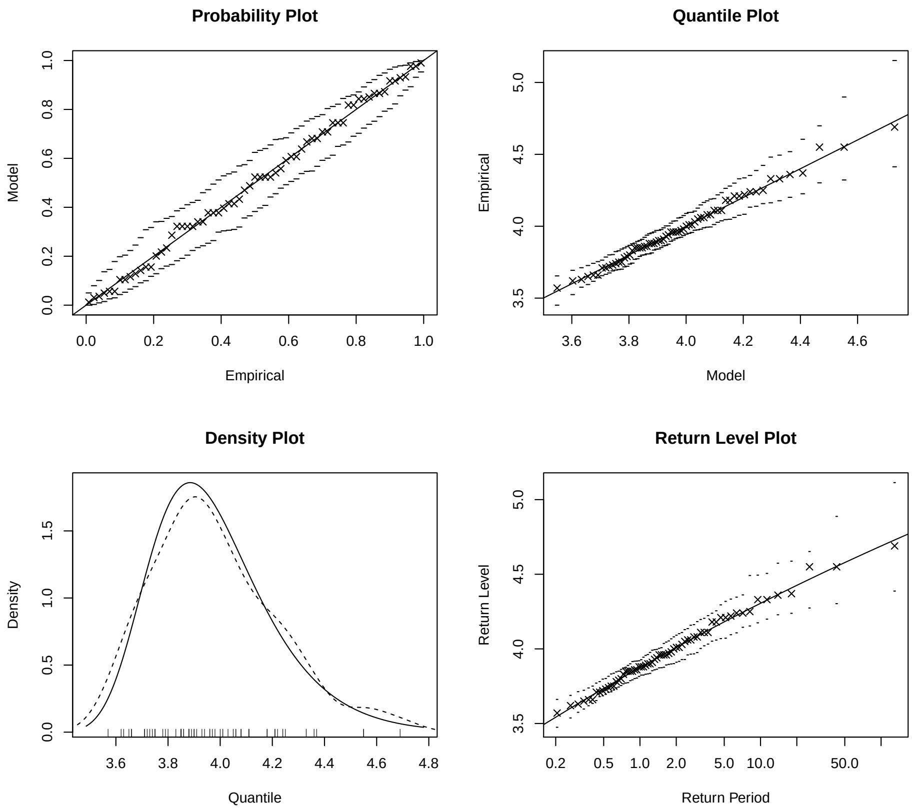

Figure 1.7: Graphical diagnostic plot for the GEV model fit on the Port Pirie dataset, as provided by R package evd.

plugin method,

$$\hat{p} = 1 - G_{\hat{\mu},\hat{\sigma},\hat{\gamma}}(u).$$

If the goal was to estimate a high quantile, say $z_p = F_n^{\leftarrow}(1-p)$, where F_n is the distribution of the annual maximum, one could again use plugin estimates and set

$$\hat{z}_p \;=\; G^{\leftarrow}_{\hat{\mu},\hat{\sigma},\hat{\gamma}}(1-p) = \begin{cases} \hat{\mu} + \dfrac{\hat{\sigma}}{\hat{\gamma}}\left[\left(\dfrac{-1}{\log(1-p)}\right)^{\gamma} - 1\right] & (\gamma \neq 0), \\ \mu + \sigma \log\left(\dfrac{-1}{\log(1-p)}\right) & (\gamma = 0). \end{cases}$$

In this introductory course, will not get into details about the consistency of these estimators. However, one may notice that, on this example, the maximum likelihood estimate is close to 0, compared to its estimated standard deviation. One may thus wonder if the Gumbel sub-model ($\gamma = 0$) provides a reasonable fit (this will impact in particular high quantile estimates, since the Gumbel distribution has unbounded support, contrary to the Weibull).

A simple visual diagnostic for this hypothesis is the following: The inverse of $G(x) = e^{-e^{-\frac{x-\mu}{\sigma}}}$ is

$$G^{\leftarrow}(y) = \sigma\left[-\log(-\log(y))\right] + \mu.$$

On the other hand, the empirical quantile of order $y = i/n$ $(i = 1, \ldots, n)$ is

$$\hat{G}^{\leftarrow}(i/n) = X_{(i)}$$

(the i^{th} order statistic)

If the Gumbel model is appropriate, we should have

$$X_{(i)} \approx \sigma - \log(-\log(\frac{i}{n+1})) + \mu,$$

for some $\sigma > 0$ and some $\mu \in \mathbb{R}$. Thus the graph of the points $(-\log(-\log(\frac{i}{n+1})); X_{(i)})$ (the so-called *Gumbel plot*) should be approximately affine. The graph obtained with the Port Pirie data is shown in Figure 1.5.1. It 'confirms' the null hypothesis of a Gumbel type distribution.

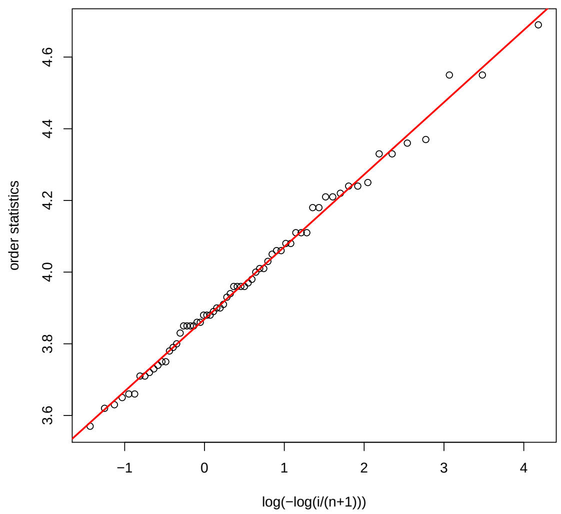

Figure 1.8: Gumbel plot for the Port Pirie dataset.

1.5.2 Method of block maxima

This is just a generalization of the above analysis. Given a series of n independent (or 'weakly' dependent), say daily, data $X_i, i \leq n$, the analyst may divide the data set into m block of size $k = n/m$ each (say $k = 30$ to work with monthly maxima), and assume that the maximum over each block

$$M_i = \bigvee_{r=ki+1}^{n} X_r, \quad i = 1, \ldots, m$$

approximately follows a GEV distribution, which parameters remain to be estimated. The rest follows the line of the Port Pirie example. Figure 1.5.2 illustrates this procedure.

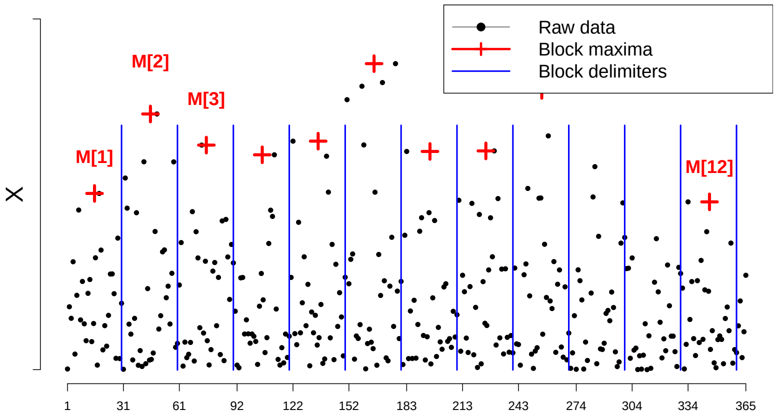

Figure 1.9: Work-flow for the block-maxima method.

1.5.3 Peaks-Over-Threshold

The 'Peaks-Over-Threshold' (POT) methods consider excesses over a fixed, relatively high threshold, instead of maxima. Consider the equivalent condition of (MDA) in terms of excesses above thresholds (1.18) (Theorem 1.4.1),

$$\frac{1 - F(t + \sigma(t) \bullet)}{1 - F(t)} \xrightarrow[t \nearrow x_\star]{} (1 + \gamma \bullet)_+^{-1/\gamma}.$$

For fixed, large enough t (but not too large, in order to observe 'some' data above t), we may use the approximation

$$\frac{1 - F(t + \sigma x)}{1 - F(t)} \approx (1 + \gamma x)_+^{-1/\gamma}.$$

In other terms, if $X \sim F$,

$$\mathbb{P}\left((X - t)/\sigma > x \mid X > t\right) \approx (1 + \gamma x)_+^{-1/\gamma}.$$

or, by a change of variables

$$\mathbb{P}\left(X - t > x \mid X > t\right) \approx \left(1 + \gamma \frac{x}{\sigma}\right)^{-1/\gamma}, \qquad x > 0$$

for some unknown parameters (σ, γ).

Consider an i.i.d. sample $X_i, i = 1, \dots n \sim F$. Estimation of the parameters (σ, γ) may be done using the *excesses* above t,

$$\{X_i : \quad X_i > t, i = 1, \dots, n\},$$

as illustrated in Figure 1.5.3 .

Let $(i(1), \dots, i(m))$ be the indices corresponding to an excess. Now the assumption for further inference is

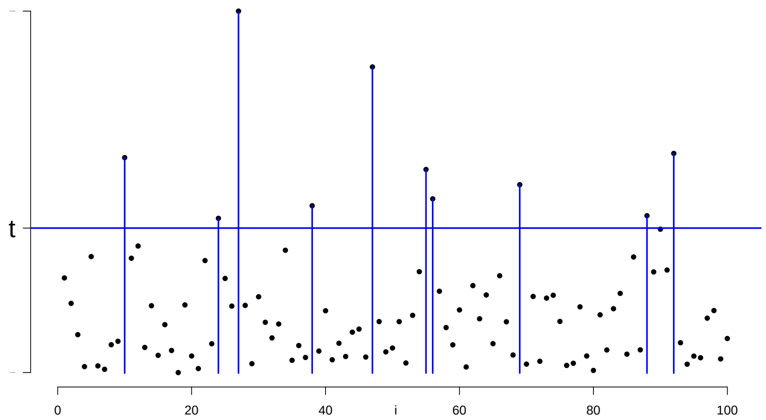

Figure 1.10: Work-flow for the POT procedure above a high threshold t: the raw data are the black dots, the 'excess' data $X_{i(r)}$ used for inference correspond to the blue lines.

$$\mathbb{P}\left(X_{i(r)} > x\right) \approx \left(1 + \gamma \frac{x-t}{\sigma}\right)_+^{-1/\gamma}, \qquad x > t. \tag{1.28}$$

i.e. $X_{i(r)} \sim H_{t,\sigma,\xi}(y)$, where $H_{\mu,\sigma,\gamma}$ is the *Generalized Pareto distribution* (GPD) with parameters $\mu \in \mathbb{R}, \sigma > 0, \gamma \in \mathbb{R}$,

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

Notice that in (1.28), the location parameter is automatically $\mu = t$. Also, the above quantity for $\gamma = 0$ should be interpreted as its limit as $\gamma \to 0$,

$$H_{\mu,\sigma,0}(x) = 1 - e^{-\frac{x-\mu}{\sigma}} \qquad (x > \mu).$$

The GPD model for the excesses $(X_i(r), r = 1, \dots m)$ is thus

$$\mathcal{P} = \{H_{t,\sigma,\gamma} \ : \qquad \sigma > 0, \gamma \in \mathbb{R}\}.$$

Again, the packages mentioned in the Port Pirie example provide routines for maximum likelihood estimation in the GPD model. In practice, one may use the estimated parameters in a plugin method in order to predict the probability of an excess above a high threshold $t' > t$ (even though no data has ever been observed above t').

Chapter 2

Regular variation and tail measures

2.1 Regular variation of a real function

Definition 2.1.1 (Regular variation). *A function $U : \mathbb{R}^+ \to \mathbb{R}^+$ is* regularly varying *(RV) if $\exists \rho \in \mathbb{R}$ such that*

$$\forall x > 0, \lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\rho.$$

The parameter ρ is called the regular variation index. *We write '$U \in RV(\rho)$', meaning U is RV with regular variation index ρ. If $\rho = 0$, U is called* slowly varying.

Example 2.1 (Fréchet survival function)**:**
$U(x) = 1 - \Phi_\alpha(x) = 1 - e^{-x^{-\alpha}}$ is $RV(-\alpha)$.

Example 2.2 (Generalized Pareto)**:**
$U(x) = (1 + \gamma x)^{-1/\gamma}$ is $RV(-1/\gamma)$.

Example 2.3 (Canonical: Pareto tail)**:**
$U(x) = x^{-\alpha}$, $x > 1$, is $RV(-\alpha)$.

Example 2.4 (slow variation)**:**
$U(x) = \log(1 + x)$ is slowly varying. If $\lim_{t\to\infty} f(t) = \ell \in \mathbb{R}$, then f is slowly varying, the converse is false.

Remark 2.1.2. *Remind from last chapter that the max-domain of attraction condition* (MDA) *is equivalent to condition* (1.19) *concerning the tail regularity, which is*

$$\frac{1 - F(t + \sigma(t)x)}{1 - F(t)} \xrightarrow[t \nearrow x_\star]{} (1 + \gamma x)_+^{-1/\gamma}.$$

This 'resembles' a RV condition. It will be shown that it is *equivalent to regular variation of $U = 1 - F$ in the case $\gamma > 0$.*

Remark 2.1.3 (Equivalent formulation of RV). *U is $RV(\rho)$ $\iff$ $\exists L$ a slowly varying function such that $U(x) = x^\rho L(x)$.*
(Proof: exercise)

Proposition 2.1.4 (A sufficient condition for RV)
If $\exists h : \mathbb{R}^{+*} \to \mathbb{R}^{+*}$, measurable such that $\forall x > 0, \lim_{t\to\infty} \frac{U(tx)}{U(t)} = h(x)$ then U is RV.

Proof. (sketch of): Show that such h satisfies the Hamel equation $h(xy) = h(x)h(y)$. ■

Proposition 2.1.5 (Another sufficient condition)
If U is monotone and if $\exists (a_n)_{n\geq 0} \in \mathbb{R}$ s.t. $a_n \to +\infty$, and a function $h : \mathbb{R}^{+*} \to \mathbb{R}^{+*}$ such that $\forall x > 0, \lim_n nU(a_n x) = h(x)$ then U is RV.

Proof. Assume that U is non decreasing. Put $n(t) = \inf\{n \geq 0 : a_n \geq t\}$. Then

$$\frac{U(a_{n(t)-1}x)}{U(a_{n(t)})} \leq \frac{U(tx)}{U(t)} \leq \frac{U(a_{n(t)}x)}{U(a_{n(t)-1})},$$

and both sides of the sandwich converge to $h(x)/h(1)$. Using Proposition 2.1.4 concludes.

The proof for the non increasing case is the same, up to switching the upper and lower bounds.

■

Exercise 2.1 (Reciprocal for Proposition 2.1.5)**:**
Let F be a cdf and asssume that $(1-F) \in RV(-\alpha)$, for some $\alpha < 0$. Define $U(t) = 1/(1-F)(t)$ and let $a_n = U^{\leftarrow}(n)$. Show that

$$n(1 - F(a_n x)) \to x^{-\alpha}, \text{ for } x > 0, \text{ as } t \to \infty. \tag{2.1}$$

hint: consider the ratio $\frac{1-F(a_n x)}{1-F(a_n)}$, and derive the limit of $U(U^{\leftarrow}(n))/n$.

Exercise 2.2 (regular variation and Fréchet domain of attraction)**:**
Let F be a *c.d.f.* The goal is to show the following: '$(1-F)$ is regularly varying with index $-\alpha < 0$ if and only if

$$\exists (a_n)_{n\geq 0} > 0 : F^n(a_n \bullet) \to \Phi_\alpha, \qquad \text{where } \Phi_\alpha(x) = e^{-x^{-\alpha}}, x > 0, \tag{2.2}$$

and to characterize the possible sequences a_n, up to tail equivalence.

1. Show that (2.2) $\Rightarrow \forall x > 0, F(x) < 1$, $F(a_n x) \to 1$, and $a_n \to \infty$.
2. Prove that (2.2)$\Rightarrow 1 - F$ is $RV(-\alpha)$.
3. Switching to the inverse function, show that (2.2)$\Rightarrow a_n \sim \left(\frac{1}{1-F}\right)^{\leftarrow}(n)$ as $n \to \infty$.
4. Check that if $(1 - F)$ $n(1 - F(a_n x)) \to x^{-\alpha}$ for some sequence a_n, then (2.2) holds true. Check that convergence also holds for any sequence $\tilde{a}_n \sim a_n$. Conclude.

2.2 Karamata theorem and consequences

Idea: For integration purposes (of the kind $\int_x^\infty U(t)\mathrm{d}t$ or $\int_0^x U(t)\mathrm{d}t$), If U is $RV(\rho)$ then it behaves as $t \mapsto t^\rho$ would, as $x \to \infty$. More precisely,

- if $U(t) = t^\rho$ and $\rho < -1$, $\int_x^\infty U(t)\mathrm{d}t = -(\rho + 1)^{-1}x^{\rho+1} = -(\rho + 1)^{-1}xU(x)$.
- if $U(t) = t^\rho$ and $\rho > -1$, $\int_0^x U(t)\mathrm{d}t = (\rho + 1)^{-1}x^{\rho+1} = (\rho + 1)^{-1}xU(x)$.

Karamata's theorem says that the same is true as $x \to \infty$ when $U \in RV(\rho)$.

Theorem 2.2.1 (Karamata)
Let $U : \mathbb{R}^+ \mapsto \mathbb{R}^+$ be a $RV(\rho)$ function, s.t. $\int_0^x U < \infty \forall x > 0$.

1. *If $\rho \geq -1$ then $x \mapsto \int_0^x U$ is $RV(\rho+1)$ and*
$$\lim_{x\to\infty} \frac{xU(x)}{\int_0^x U} = \rho + 1. \tag{2.3}$$
Conversely if (2.3) then $U \in RV(\rho)$.
2. *If $\rho < -1$ of if $\rho = 1$ and $\int_1^\infty U < \infty$ then $x \mapsto \int_x^\infty U$ is $RV(\rho+1)$ and*
$$\lim_{x\to\infty} \frac{xU(x)}{\int_x^\infty U} = -\rho - 1. \tag{2.4}$$
Conversely if (2.4) then $U \in RV(\rho)$.

Proof. See Resnick (1987), p. 17 or Resnick (2007), p. 25. ■

Corollary 2.2.2 (Karamata representation)
A function $L : \mathbb{R}^+ \to \mathbb{R}^+$ is slowly varying if and only if

- *$\exists c : \mathbb{R}^+ \to \mathbb{R}^+$ such that $\lim_\infty c(x) = c \in (0, \infty)$, and*
- *$\exists \epsilon : \mathbb{R}^+ \to \mathbb{R}^+$ such that $\lim_\infty \epsilon(x) = 0$,*

such that
$$L(x) = c(x) \exp\left(\int_1^x \frac{\epsilon(t)}{t} \mathrm{d}t\right). \tag{2.5}$$

Proof. The proof of the sufficiency of (2.5) is an easy exercise. For the converse, let $L \in RV(0)$. From Karamata theorem, we have
$$b(x) := \frac{xL(x)}{\int_0^x L} \to 1 \quad \text{as } x \to \infty.$$
By definition of b we may write
$$L(x) \quad = \quad \frac{b(x)}{x} \int_0^x L \quad = \quad b(x) \exp\{\log \frac{\int_0^x L}{x}\} \tag{2.6}$$
But also
$$\begin{aligned} \log \frac{\int_0^x L}{x} &= \int_1^x \frac{\mathrm{d}}{\mathrm{d}t}\Big[\log\Big(\int_0^t L\Big) - \log t\Big]\mathrm{d}t + D \quad (D : \text{a constant}) \\ &= \int_1^x \left(\frac{L(t)}{\int_0^t L} - \frac{1}{t}\right)\mathrm{d}t + D \quad (D : \text{a constant}) \\ &= \int_1^x \frac{b(t)-1}{t}\mathrm{d}t + D \end{aligned}$$
Setting $\epsilon(t) = b(t) - 1$, we have $\epsilon(t) \to 0$ and the latter display combined with (2.6) yields
$$L(x) = \underbrace{b(x)e^D}_{:=c(x)\to e^D>0} \exp\Big\{\int_1^x \frac{\epsilon(t)}{t}\mathrm{d}t\Big\},$$
which concludes the proof.

■

Corollary 2.2.3 (Karamata representation of RV functions)
$U \in RV(\rho) \iff U(x) = c(x)\exp \int_1^x \alpha(t)/t\mathrm{d}t$*, for some functions* $c(x) \to c > 0$ *and* $\alpha(t) \to \rho$*.*

Proof.

$$\begin{aligned} U \in RV(\rho) &\iff U(x) = L(x)x^\rho \\ &\iff U(x) = c(x)\exp\Big(\int_1^x \epsilon(t)/t\mathrm{d}t\Big)\, exp(\rho \log x) \quad \text{(Corollary 2.2.2)} \\ &\iff U(x) = c(x)\exp\Big(\int_1^x [\epsilon(t) + \rho]/t\mathrm{d}t\Big). \end{aligned}$$

■

The Karamata representation will prove useful at the end of this chapter, for proving the consistency of the Hill estimator (an estimator for the regular variation index).

2.3 Vague convergence of Radon measures

Most of the material of this section is borrowed from Resnick (1987), Chapter 3, which contains detailed proofs.

2.3.1 The space of Radon measures

In this course, the 'extreme events' will take place in a 'nice' space such as $(0,\infty)$ or $(0,\infty]$. Later on, for multivariate extremes, a very convenient space will be $\mathbf{E} = [0,\infty]^d \setminus \{\mathbf{0}\}$ where $\mathbf{0} = (0,\dots,0)$. The reason why we include $+\infty$ (via the Alexandroff's compactification) is that it makes the intervals $[x,\infty]$, for $x > 0$ compact.

Remark 2.3.1 (Alexandroff's space). *The space* $[0,\infty]$ *is defined as* $[0,\infty) \cup \{+\infty\}$*, where* $+\infty$ *is an arbitrary element which is greater than any element of* $[0,\infty)$*. The order* $\le$ *on* $[0,\infty)$ *is thus extended to* $[0,\infty]$*. The topology on* $[0,\infty]$*, i.e. the family of open sets then consists of*

- *All sets* $V \subset [0,\infty)$ *which are open sets for the usual topology (Euclidean) on* $\mathbb{R}$*.*
- *All sets* $V \subset [0,\infty]$ *such that* $+\infty \in V$ *and* V^c *is compact in* $[0,\infty)$*.*

After having compactified $[0,\infty)$ *at infinity, it is convenient to 'uncompactify' it by removing* 0*. We obtain the space* $\mathrm{E} = (0,\infty]$*. The idea behind is that we want* $tx \to \infty$ *as* $t \to \infty$ *for all* $x \in \mathrm{E}$*.*

Exercise 2.3:
Prove that the sets $[a,\infty]$, for $a > 0$ are compact in $\mathrm{E} = (0,\infty]$.

More generally, in the remainder of this course, we consider a space $\mathbf{E}$ which is locally compact, second countable, Hausdorff (LCSCH). Locally compact means that each point in $\mathbf{E}$ has a compact neighborhood. Second countable means that the topology on $\mathbf{E}$ has a countable base. Finally, Hausdorff means that for any pair $x \ne y \in \mathbf{E}$, there exists disjoint open sets U, V such that $x \in U$ and $y \in V$. In the sequel E is endowed with the Borel σ-field $\mathcal{E}$.

Definition 2.3.2 (Radon measures).

- *A measure* $\mu : (\mathrm{E}, \mathcal{E}) \to [0, \infty]$ *is called a* Radon measure *if for all compact set* $K \subset \mathrm{E}$, $\mu(K) < \infty$.
- *We denote* $\mathrm{M}(\mathrm{E})$ *the set of all Radon measures on* E.
- *In particular,* $\mathrm{M}(\mathrm{E})$ *contains* $\mathrm{M}_p(\mathrm{E})$ *the set of* Radon point measures, i.e. *measures of the kind* $\mu = \sum_{i \in D} \delta_{x_i}$*; where* D *is countable, and* $(x_i)_{i \in D} \in \mathrm{E}$ *has no accumulation point.*

2.3.2 Vague topology on $\mathrm{M}(\mathrm{E})$

Let $(\mu_n)_{n\in\mathbb{N}}, \mu \in \mathrm{M}(\mathrm{E})$. The sequence (μ_n) *converges vaguely* to μ if for all function $f \in \mathcal{C}_K$ (continuous with compact support), $\int_{\mathrm{E}} f \mathrm{d}\mu_n \to \int_{\mathrm{E}} f \mathrm{d}\mu$. We denote $\mu_n \overset{v}{\to} \mu$. In the sequel we denote $\mu(f) := \int_{\mathrm{E}} f \mathrm{d}\mu$. The topology associated to this notion of convergence is called the *vague topology* on $\mathrm{M}(\mathrm{E})$, denoted by $\mathcal{V}$. It is the topology generated by the evaluation maps $T_f : \mu \mapsto \mu(f)$, for $f \in \mathcal{C}_K$. A basis for $\mathcal{V}$ is the family of open sets

$$\Big\{V = \{\mu \in \mathrm{M}(\mathrm{E}) : a_i < \mu(f_i) < b_i, \quad \forall 1 \le i \le k\}, k \in \mathbb{N}, a_i < b_i \in \mathbb{R}, f_i \in \mathcal{C}_K\Big\}.$$

It can be shown that $(\mathrm{M}(\mathrm{E}), \mathcal{V})$ is a Polish space (separable, completely metrizable). Separable means that it contains a dense sequence; completely metrizable means that one can construct a distance on $\mathrm{M}(\mathrm{E})$ which is compatible with the topology, and for which $\mathrm{M}(\mathrm{E})$ becomes a complete space.

Similarly to the case of weak convergence, we have a 'Portmanteau theorem'

Theorem 2.3.3
The following are equivalent:

(i) $\mu_n \overset{v}{\to} \mu$.

(ii) $\mu_n(B) \to \mu(B)$ *for all set* B *such that* $\bar{B}$ *is compact and* $\mu(\partial B) = 0$.

(iii) For all compact $K \subset \mathrm{E}$, $\limsup \mu_n(K) \le \mu(K)$ *and for all open set* $G \subset \mathrm{E}$, $\liminf \mu_n(G) \ge \mu(G)$.

2.3.3 Regular variation and vague convergence of tail measures

In this section $\mathrm{E} = (0, \infty]$.

Theorem 2.3.4
Let F *be a* c.d.f. *and* $X \sim F$. *The following are equivalent*

(i) F *belongs to the max-domain of attraction of* Φ_α *(Fréchet distribution)*

(ii) $1 - F \in RV(-\alpha)$

(iii) $\exists (a_n)_{n\ge 0} : n(1 - F(a_n x)) \xrightarrow[n\to\infty]{} x^{-\alpha}$.

(iv) $\mu_n(\,\bullet\,) := n\mathbb{P}(\frac{X}{a_n} \in (\,\bullet\,)) \overset{v}{\to} \nu_\alpha(\,\bullet\,)$*, where* $\nu_\alpha[x, \infty) = x^{-\alpha}, x > 0$.

Proof. $(ii) \iff (iii)$ and $(ii) \Rightarrow (i)$ have been proven in Exercise 2.2. The fact that $(i) \Rightarrow (ii)$ is shown in Resnick (1987), proposition 1.11 p. 54. The proof relies on Karamata's theorem. It remains to see why $(iii) \iff (iv)$. Assume (iv). Then for $x > 0$,

$$\begin{aligned} n(1 - F(a_n x)) &= n\mathbb{P}(X/a_n \in (x, \infty)) \\ &= \mu_n(x, \infty) \\ &\to \nu_\alpha(x, \infty) = x^{-\alpha} \quad \text{by Theorem 2.3.3 (ii)}, \end{aligned}$$

which proves (iii). Now assume (iii). On order to show that (iv) holds, we need to show that for $f \in \mathcal{C}_K$, $\mu_n(f) \to \nu_\alpha(f)$. Let $f \in \mathcal{C}_K$. Let $S = supp(f) = cl\{x > 0 : f(x) > 0\}$, where $cl(A) = \bar{A}$ denotes the closure of a set A. Necessarily $0 \notin S$ otherwise S would not be closed in $(0, \infty]$. Thus $S \subset [\delta, \infty]$ for some $\delta > 0$. Introduce the probability measures P_n on $[\delta, \infty]$ defined by

$$P_n(A) = \mu_n(A)/\mu_n[\delta, \infty], \qquad A \subset [\delta, \infty],$$

(which is well defined because $[\delta, \infty]$ is compact, thus $\mu_n[\delta, \infty] < \infty$).

Using (iii), for all $x > \delta$, $P_n[x, \infty] \to (x/\delta)^{-\alpha}$. Thus, using the Portmanteau theorem for probability measures, P_n converges weakly to $P = \delta^\alpha \nu(\,\cdot\,)$. Now, f has compact support in $[\delta, \infty]$ implies that f is continuous and bounded on $[\delta, \infty]$. Thus, $P_n(f) \to P(f)$, which yields $\mu_n(f) \to \mu(f)$. ■

2.3.4 Exercises

The following exercises are borrowed from Resnick (1987), chapter 3.4

Exercise 2.4:
Show that the following transformations are continuous:

1.
$$\begin{aligned} T_1 : \mathrm{M}(E) \times \mathrm{M}(E) &\to \mathrm{M}(E) \\ (\mu_1, \mu_2) &\to \mu_1 + \mu_2. \end{aligned}$$

2.
$$\begin{aligned} T_1 : \mathrm{M}(E) \times (0, \infty) &\to \mathrm{M}(E) \\ (\mu, \lambda) &\to \lambda\mu. \end{aligned}$$

Exercise 2.5:
Let $(x_n)_{n \in \mathbb{N}}, x$ in E and $c_n \leq 0, c > 0$. Show that in $\mathrm{M}(E)$,

$$\mu_n := c_n \delta_{x_n} \xrightarrow{v} c\delta_x$$

if and only if $x_n \xrightarrow[n\to\infty]{} x$ and $c_n \xrightarrow[n\to\infty]{} c$.

Exercise 2.6:
Let $m_n = \sum_{i \in \mathbb{N}^*} n^{-1} \delta_{(\frac{i}{n})}$ and let m be the lebesgue measure on $(0, \infty)$. Show that $m_n \xrightarrow{v} m$.

2.4 Weak convergence of tail empirical measures

2.4.1 Random measures

Recall $(\mathrm{M}(\mathrm{E}), \mathcal{V})$ is a topological space. Thus it has a Borel σ-field $\mathcal{M}(\mathrm{E})$. It can be shown by monotone class arguments that $\mathcal{M}(\mathrm{E})$ is generated by the evaluation maps $T_f : \mu \mapsto \mu(f)$, for $f \in \mathcal{C}_K(\mathrm{E})$, or by the $T_F : \mu \mapsto \mu(F)$, for $F \subset E$ closed. Thus

$$\mathcal{M}(\mathrm{E}) = \sigma\Big\{T_f\,, f \in \mathcal{C}_K\Big\} \;=\; \sigma\Big\{T_F, F \subset \mathrm{E},\ \text{closed}\Big\}.$$

Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, a random measure ξ is thus a measurable mapping $(\Omega, \mathcal{A}) \to (\mathrm{M}(\mathrm{E}), \mathcal{M}(\mathrm{E}))$. A point process is a special case of such mapping, taking its value in $(\mathrm{M}_p(\mathrm{E}), \mathcal{M}_p(\mathrm{E}))$,where $\mathcal{M}_p(\mathrm{E})$ is the trace σ-field of $\mathcal{M}$ on M_p. The distribution of a random measure ξ is entirely determined by the 'finite dimensional distributions', *i.e.* by the laws of the random vectors $(\xi(f_1), \dots, \xi(f_k))$, where $k \in \mathbb{N}$ and $f_i \in \mathcal{C}_K, i \le k$.

A convenient tool for characterizing the law of random measures and their convergence in distribution is the Laplace transform, defined next.

Definition 2.4.1 (Lapace transorm of a random measure)**.** *The Laplace transform of a random measure ξ is the functional*

$$\begin{aligned} \mathcal{L}_\xi : \quad & \mathcal{C}_K \to \mathbb{R} \\ & f \mapsto \mathcal{L}_\xi(f) = \mathbb{E}\left(e^{-\xi(f)}\right) \quad = \int_\Omega e^{-\int_{\mathrm{E}} f(x)\xi(\omega, \mathrm{d}x)}\, \mathrm{d}\mathbb{P}(\omega) \end{aligned}$$

Since the law of a random vector $X \in \mathbb{R}^k$ is determined by its (usual) Laplace transform $t \mapsto \mathbb{E}\left(e^{-\langle t, X\rangle}\right)$, it is easy to see that the Laplace transform of a random measure also determines uniquely its distribution. In fact more is true: pointwise convergence of Laplace transforms of a sequence (ξ_n) determines weak convergence, as stated next.

2.4.2 Weak convergence in $\mathrm{M}(\mathrm{E})$

Proposition 2.4.2 (Characterization of weak convergence)
Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of random measures on E*. The following statements are equivalent*

(i) $\xi_n \xrightarrow{w} \xi$, i.e.$\forall \phi$ *bounded continuous* $\mathrm{M}(\mathrm{E}) \to \mathbb{R}$, $\mathbb{E}(\varphi(\xi_n)) \to \mathbb{E}(\varphi(\xi))$.

(ii) $\forall k \in \mathbb{N}, \forall (f_1, \dots, f_k) \in \mathcal{C}_K$, $(\xi_n(f_1), \dots, \xi_n(f_k)) \xrightarrow{w} (\xi(f_1), \dots, \xi(f_k))$.

(iii) $\forall f \in \mathcal{C}_K$, $\mathcal{L}_{\xi_n}(f) \to \mathcal{L}_\xi(f)$ *(pointwise convergence of the Laplace transforms)*

Proof.

- $(ii) \iff (iii)$ comes from standard properties of weak convergence in $\mathbb{R}^k$. Assume (ii) and fix $f \in \mathcal{C}_K$. Then letting $X_n = \xi_n(f)$ and $X = \xi(f)$, we have

  $$\mathcal{L}_{\xi_n}(f) = \mathbb{E}\left(e^{-\xi_n(f)}\right) = \mathcal{L}_{X_n}(1) \to \mathcal{L}_X(1),$$

  where the latter convergence comes from the fact that pointwise convergence of the Laplace transform of random variables is equivalent to their weak convergence. This proves (iii).

Conversely, assume (iii) and notice that the Laplace transform of the random vector $X_n = (\xi_n(f_1), \dots, \xi_n(f_k))$ is, for $t \in \mathbb{R}^k$,

$$\mathcal{L}_{X_n}(t) = \mathbb{E}\left(e^{-\langle t, X\rangle}\right) = \mathbb{E}\left(e^{-\sum_i t_i \xi_n(f_i)}\right) = \mathbb{E}\left(e^{-\xi_n(\sum_i t_i f_i)}\right) = \mathcal{L}_{\xi_n}\left(\sum_i t_i f_i\right).$$

Since $\sum_i t_i f_i \in \mathcal{C}_k$, the right-hand-side converges to $\mathcal{L}_\xi(\sum t_i f_i) = \mathcal{L}_X(t)$, where $X = (\xi(f_1), \dots, \xi(f_k))$ and the proof of $(iii) \Rightarrow (ii)$ is complete.

- $(i) \Rightarrow (ii)$ is a direct application of the continuous mapping theorem applied to the mapping $T : \mu \mapsto (\mu(f_1), \dots, \mu(f_k))$, which is continuous by definition of the vague topology.

- $(ii) \Rightarrow (i)$:
  Assume (ii). We need to show that $(\xi_n)_{n\in\mathbb{N}}$ is (a) relatively compact (*i.e.*that its closure is compact for the weak topology of weak convergence in $\mathrm{M}(E)$), and (b) the limits of any two converging subsequence coincide in distribution.

  (b) is an easy exercise: it is enough to show that for two possible limits ξ^1, ξ^2,

  $$\mathbb{P}\left(a_i < \xi^1(f_i) < b_i, 1 \le i \le k, a_i < b_i\right) = \mathbb{P}\left(a_i < \xi^2(f_i) < b_i, 1 \le i \le k, a_i < b_i\right).$$

  (a) requires more care. Since M(E) is a separable, metric space, the Prohorov's theorem applies (tightness implies relative compactness). It is thus enough to show that (ξ_n) is tight. To do this, use Lemma 3.20 p.153 in Resnick (1987): a sufficient condition is that $\xi_n(f)_{n\in\mathbb{N}}$ be tight, for all fixed $f \in \mathcal{C}_K$. Now the latter condition is satisfied because $\xi_n(f)$ converges weakly in $\mathbb{R}$.

■

2.4.3 Tail measure and tail empirical measure

In this section $\mathrm{E} = (0, \infty]$. Recall from Theorem 2.3.4 that for a *c.d.f.* F and $X \sim F$, the following equivalence:

- $1 - F$ is $RV(-\alpha)$, (*i.e.* $\exists (a_n)_{n\ge 0} : n(1 - F(a_n x) \to x^{-\alpha}$)
  $\Longleftrightarrow$
- $\mu_n \xrightarrow{v} \nu_\alpha$, where $\mu_n(A) = n\mathbb{P}(X/a_n \in A), A \subset (0, \infty]$ and $\nu_\alpha[x, \infty] = x^{-\alpha}, x > 0$.

We now define the empirical version of μ_n, and we shall see that this empirical version (a random measure) converges in distribution to ν_α as well, under the same assumptions.

Definition 2.4.3 (tail empirical measure)**.** *Let F be a* c.d.f. *on $\mathbb{R}^+$ and $X, (X_i)_{i\in\mathbb{N}} \overset{i.i.d.}{\sim} F$. Let $(a_n)_{n\in\mathbb{N}} > 0$ a sequence of positive numbers.*

Consider a sequence of integers $k(n)_{n\in\mathbb{N}} \in \mathbb{N}$, such that $k(n) \xrightarrow[n\to\infty]{} \infty$ and $\frac{k(n)}{n} \xrightarrow[n\to\infty]{} 0$. Write k instead of $k(n)$ for convenience. The tail empirical measure *associated to F and the sequence (a_n) is the random point measure*

$$\nu_{n,k} = \frac{1}{k}\sum_{i=1}^{k} \delta_{\left\{\frac{X_i}{a_{\lfloor n/k \rfloor}}\right\}}$$

Proposition 2.4.4 (weak CV of the tail empirical measure)
If $F \in RV(-\alpha)$ and $(a_n)_{n\geq 0} > 0$ is such that $\mu_n \overset{v}{\to} \nu_\alpha$, then the tail empirical measures converge weakly in $\mathrm{M}_+(\mathrm{E})$,

$$\nu_{n,k} \overset{w}{\to} \nu_\alpha.$$

Proposition 2.4.4 means that the tail empirical measure is a consistent estimator for the tail measure.

Proof. According to Proposition 2.4.2, we need to show that $\mathcal{L}_{\nu_{n,k}}(f) \to \mathcal{L}_{\delta_{\nu_\alpha}}(f) = e^{-\nu_\alpha(f)}$, for $f \in \mathcal{C}_K$.
Now, writing for a_s instead of $a_{\lfloor s \rfloor}$,

$$\begin{aligned}
\mathcal{L}_{\nu_{n,k}}(f) &= \mathbb{E}\left(\exp\Big\{-\int f \mathrm{d}\nu_{n,k}\Big\}\right) \\
&= \mathbb{E}\left(\exp\Big\{-\frac{1}{k}\sum_{i\leq n} f(\frac{X_i}{a_{n/k}})\Big\}\right) \\
&= \Big[\mathbb{E}\Big(\exp\{-\frac{1}{k}f(\frac{X_1}{a_{n/k}})\}\Big)\Big]^n \\
&= \Big[1 - \underbrace{\mathbb{E}\Big(1-\exp\{-\frac{1}{k}f(\frac{X_1}{a_{n/k}})\}\Big)}_{\mathcal{E}_{n,k}}\Big]^n
\end{aligned}$$

Now,

$$\mathcal{E}_{n,k} = \int_0^\infty (1 - e^{-\frac{1}{k}f(y)})\,\mathrm{d}P_{n,k}(y)$$
$$(\text{where } P_{n,k} \text{ is the law of } \frac{X_i}{a_{n/k}})$$

Since f is bounded we may use a Taylor expansion of the exponential function, uniformly over x, as $k \to \infty$,

$$\sup_{x>0}\Big|1 - e^{-\frac{f(x)}{k}} - \frac{f(x)}{k}\Big| = o(1/k).$$

Also $f \in \mathcal{C}_K(0,\infty]$ so that f vanishes at a neighborhood of 0, say on $(0,\delta]$ and for $0 \leq x \leq \delta$, $1 - e^{-f(x)/k} = 0$. We obtain

$$\begin{aligned}
\mathcal{E}_{n,k} &= \int_\delta^\infty f(x)/k + o(1/k)\mathrm{d}P_{n,k}(x) \\
&= n^{-1}\int_\delta^\infty (f(x) + o(1))\,\frac{n}{k}\mathrm{d}P_{n,k}(x).
\end{aligned}$$

Now for $I \subset (0,\infty]$ measurable, $\frac{n}{k}P_{n,k}(I) = \frac{n}{k}\mathbb{P}\left(X_1 \in a_{n/k}I\right) = \mu_{n/k}(I)$. Vague convergence of μ_t towards ν_α as $t \to \infty$ entails

$$n\mathcal{E}_{n,k} \to \int_\delta^\infty f\mathrm{d}\nu_\alpha = \int_0^\infty f\mathrm{d}\nu_\alpha.$$

Finally as $n \to \infty, k \to \infty$ and $n/k \to \infty$, we obtain

$$\mathcal{L}_{\nu_{n,k}}(f) = (1 - \frac{n\mathcal{E}_{n,k}}{n})^n \sim (1 - \nu_\alpha(f))^n \to e^{-\nu_\alpha(f)}.$$

■

Exercise 2.7:
Let $\{X_{k,n}, 1 \leq k \leq n, n \geq 1\}$ be random elements of E such that for each n, the $(X_{k,n}, k \leq n)$ are i.i.d. Let $(a_n)_{n\geq 0} > 0$ be a sequence such that $a_n \xrightarrow[n\to\infty]{} \infty$ and let $\mu \in \mathrm{M}(E)$. Define $\xi_n = \frac{1}{a_n}\sum_{k=1}^n \delta_{X_{k,n}}$ (a random measure) and $\mu_n = \frac{n}{a_n}\mathbb{P}\left(X_{1,n} \in \bullet\right)$. Show that

$$\mu_n \xrightarrow{v} \mu \iff \xi_n \xrightarrow{w} \mu \text{ in } \mathrm{M}(E).$$

2.4.4 Statistical application: Hill estimator

The Hill estimator is a classical estimator of the tail index α. Many other estimators exist (Pickand's estimator, CFG estimator, ...). In this course we limit ourselves to studying the consistency of the Hill estimator. Notice that sharper results exist such as asymptotic normality or concentration inequalities under additional regularity assumptions on the tails.

The idea behind the estimator is the following: Notice first that

$$\int_1^\infty \nu_\alpha[x,\infty]x^{-1}\mathrm{d}x = 1/\alpha.$$

The Hill estimator aims at approaching the quantity $1/\alpha$. The heuristic is to successively replace ν_α with μ_n, then a_n by a quantile, then μ_n with its empirical version $\nu_{k,n}$, as follows

$$\begin{aligned}\nu_\alpha(x,\infty] &\approx n(1-F(a_n x)) = n\mathbb{P}\left(X/a_n > x\right)\\ &\approx \nu_{k,n}[x,\infty] = \frac{1}{k}\sum_{i=1}^n \mathbb{1}\left\{\frac{X_i}{a_{\lfloor n/k\rfloor}} > x\right\}\end{aligned}$$

Now take $a_n = (1/(1-F))^{\leftarrow}(n) = F^{\leftarrow}(1-1/n)$ (see exercise 2.2 for the reason of this choice), and replace $F^{\leftarrow}(1-k/n)$ with its empirical version, which is the the k^{th} largest order statistic $X_{(k)}$, so that $a_{\lfloor n/k\rfloor} \approx X_{(n-k)}$. We get

$$\begin{aligned}\frac{1}{\alpha} &= \int_1^\infty x^{-1}\nu_\alpha(x,\infty]\\ &\approx \int_1^\infty x^{-1}\frac{1}{k}\sum_{i=1}^n \mathbb{1}\left\{\frac{X_i}{X_{(k)}} > x\right\}\mathrm{d}x\\ &= \frac{1}{k}\sum_{i=1}^n \int_1^\infty x^{-1}\mathbb{1}\left\{x < \frac{X_{(k)}}{X_i}\right\}\mathrm{d}x\\ &= \frac{1}{k}\sum_{i=1}^k \log\frac{X_{(i)}}{X_{(k)}}\end{aligned}$$

N.B Here the order statistics are ranked in decreasing order, $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(n)}$.

Proposition 2.4.5 (Hill estimator)
Let $X, (X_i) \overset{i.i.d.}{\sim} F$*, where* $1 - F \in RV(-\alpha)$*, for some* $\alpha > 0$*. Let* $k = k(n) \xrightarrow[n\to\infty]{} \infty$ *such that* $k/n \xrightarrow[n\to\infty]{} 0$*. The Hill estimator, defined by*

$$\widehat{1/\alpha}_n = \frac{1}{k}\sum_{i=1}^{k} \log \frac{X_{(i)}}{X_{(k)}}$$

is a consistent estimator of $1/\alpha$, i.e. *it converges in probability to* $1/\alpha$.

Proof. The proof follows the lines from Resnick (2007). Remind that $a(t) = a_{\lfloor t \rfloor}$. To alleviate notations, we denote

$$\nu_n = \frac{1}{k}\sum_{i=1}^{k} \delta_{\left\{\frac{X_i}{a(n/k)}\right\}} \quad (= \nu_{n,k})$$
$$\hat{\nu}_n = \frac{1}{k}\sum_{i=1}^{k} \delta_{\left\{\frac{X_i}{X_{(k)}}\right\}}$$

According to the arguments leading to the statement, $\widehat{1/\alpha}_n = \int_1^\infty x^{-1}\hat{\nu}_n[x,\infty]\mathrm{d}x$ and $1/\alpha = \int_1^\infty x^{-1}\nu_\alpha[x,\infty]\mathrm{d}x$. We need to show that

$$\int_1^\infty x^{-1}\hat{\nu}_n[x,\infty]\mathrm{d}x \xrightarrow[n\to\infty]{P} \int_1^\infty x^{-1}\nu_\alpha[x,\infty]\mathrm{d}x. \tag{2.7}$$

1. Behavior of the order statistics
We show that

$$\frac{X_{(k)}}{a(n/k)} \xrightarrow{P} 1. \tag{2.8}$$

Indeed for $\epsilon > 0$,

$$\begin{aligned}\mathbb{P}\left(\left|\frac{X_{(k)}}{a(n/k)} - 1\right| > \epsilon\right) &= \mathbb{P}\left(\frac{X_{(k)}}{a(n/k)} > 1+\epsilon\right) + \mathbb{P}\left(\frac{X_{(k)}}{a(n/k)} < 1-\epsilon\right) \\ &= \mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{n}\delta_{\frac{X_i}{a(n/k)}}(1+\epsilon,\infty\right) > 1] + \mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{n}\delta_{\frac{X_i}{a(n/k)}}(1-\epsilon,\infty\right) < 1] \\ &= \mathbb{P}(\nu_n(1+\epsilon,\infty] > 1) + \mathbb{P}\left(\nu_n(1-\epsilon,\infty) < 1\right]\end{aligned}$$

Now, Proposition 2.4.4 implies that $\nu_n(1+\epsilon,\infty] \xrightarrow{P} (1+\epsilon)^{-\alpha} < 1$ and $\nu_n(1-\epsilon,\infty] \xrightarrow{P} (1-\epsilon)^{-\alpha} < 1$. Whence, the latter display converges to zero and (2.8) is proved.

2. Convergence of $\hat{\nu}_n$ **in probability in** $\mathrm{M}_+(0,\infty]$
Notice first that

$$\hat{\nu}_n(\,\cdot\,) = \nu_n\left(\frac{X_{(k)}}{a(n/k)}\,\cdot\right).$$

Consider the operator

$$\begin{aligned}T : \mathrm{M}(0,\infty] \times \mathbb{R}_+^* &\to \mathrm{M}(0,\infty] \\ (\mu, x) &\mapsto \mu(x\,\,\cdot\,).\end{aligned}$$

Then $\hat{\nu}_n = T\Big(\nu_n, \frac{X_{(k)}}{a(n/k)}\Big)$. It can be shown (SeeResnick (2007), p. 83) that T is continuous at (ν_α, x) for $x > 0$ (see Resnick (2007) p. 84). Then (2.8) combined with the continuous mapping theorem yields

$$\hat{\nu}_n \xrightarrow{P} \nu_\alpha \qquad \text{in } \mathrm{M}(0,\infty]. \tag{2.9}$$

3. Convergence of $\int_1^\infty x^{-1}\hat{\nu}_n[x,\infty]\mathrm{d}x$ We are ready to prove (2.7). For $M > 0$, (2.7) is equivalent to

$$\underbrace{\int_1^M x^{-1}\hat{\nu}_n[x,\infty]\mathrm{d}x}_{A_{M,n}} + \underbrace{\int_M^\infty x^{-1}\hat{\nu}_n[x,\infty]\mathrm{d}x}_{B_{M,n}} \xrightarrow{P} \underbrace{\int_1^M x^{-1}\nu_\alpha[x,\infty]\mathrm{d}x}_{A_M} + \underbrace{\int_M^\infty x^{-1}\nu_\alpha[x,\infty]\mathrm{d}x}_{B_M} \tag{2.10}$$

- For any fixed $M > 0$, the mapping $\mu \mapsto \int_1^M x^{-1}\mu[x,\infty]\mathrm{d}x$ is continuous on $\mathrm{M}(0,\infty]$. To see this, notice that the integrand is a decreasing function of x, so that the integral can be framed between to Riemann sums. In addition, $\mu_n \xrightarrow{v} \mu$ implies that for fixed $x > 0$ which is not an atom of μ, $\mu_n[x,\infty] \to \mu[x,\infty]$.
- The continuous mapping theorem combined with (2.9) thus implies that $A_{M,n} \xrightarrow{P} A_M$, for any fixed M.
- Since $\lim_{M\to\infty} B_M = 0$, it is enough to show that for any $\epsilon > 0, \exists M_0 > 1$ such that $\forall M \geq M_0$,

$$\lim_n \mathbb{P}(B_{M,n} > \epsilon) \leq \delta. \tag{2.11}$$

Let $M > 1$ and $\eta > 0$. We have

$$\mathbb{P}(B_{M,n} > \epsilon) = \underbrace{\mathbb{P}\left(B_{M,n} > \epsilon, \left|\frac{X_{(k)}}{a(n/k)} - 1\right| > \eta\right)}_{p^1_{n,M}} + \underbrace{\mathbb{P}\left(B_{M,n} > \epsilon, \left|\frac{X_{(k)}}{a(n/k)} - 1\right| \leq \eta\right)}_{p^2_{n,M}}.$$

From (2.8), $p^1_{n,M} \leq \mathbb{P}\left(\left|\frac{X_{(k)}}{a(n/k)} - 1\right| > \eta\right) \to 0$ as $n \to \infty$. Also ,

$$\begin{aligned}
p^2_{n,M} &= \mathbb{P}\left(\int_M^\infty x^{-1}\nu_n\left[\frac{X_{(k)}}{a(n/k)}x,\infty\right]\mathrm{d}x > \epsilon\,,\ \left|\frac{X_{(k)}}{a(n/k)} - 1\right| \leq \eta\right) \\
&\leq \mathbb{P}\left(\int_M^\infty x^{-1}\nu_n\left[(1-\eta)x,\infty\right]\mathrm{d}x > \epsilon\right) \\
&= \mathbb{P}\left(\int_{M(1-\eta)}^\infty y^{-1}\nu_n\left[y,\infty\right]\mathrm{d}y > \epsilon\right) \\
&\overset{\text{Markov}}{\leq} \frac{1}{\epsilon}\,\mathbb{E}\left(\int_{M(1-\eta)}^\infty x^{-1}\nu_n\left[x,\infty\right]\mathrm{d}x\right) \\
&= \frac{1}{\epsilon}\int_{M(1-\eta)}^\infty x^{-1}\frac{n}{k}(1-F)\left(a(n/k)x\right)\mathrm{d}x \\
&= \frac{1}{\epsilon}\,\frac{n}{k}\int_{M(1-\eta)a(n/k)}^\infty \underbrace{x^{-1}(1-F)\left(x\right)}_{U(x)}\mathrm{d}x
\end{aligned}$$

The function U in the latter integrand is $RV(-\alpha-1)$, so Karamata theorem implies that $\int_T^\infty U \sim TU(T)/\alpha$, with $TU(T) = (1-F)(T)$, *i.e.*

$$\begin{aligned} \frac{1}{\epsilon}\,\frac{n}{k}\int_{M(1-\eta)a(n/k)}^{\infty} x^{-1}(1-F)\,(x)\,\mathrm{d}x \sim_{n\to\infty} & \;\frac{1}{\epsilon}\,\frac{n}{k}(1-F)\Big(M(1-\eta)a(n/k)\Big) \\ \xrightarrow[n\to\infty]{} & \;\frac{1}{\epsilon}\,\nu_\alpha\Big[M(1-\eta),\infty\Big] \\ = & \;\frac{1}{\epsilon}\,(M(1-\eta))^{-\alpha} \end{aligned}$$

Choosing M_0 large enough so that the latter quantity is less than $\delta/2$ for $M = M_0$ shows (2.11) and concludes the proof.

■

Chapter 3

Multivariate extremes

In this chapter, the multivariate extensions of ideas from univariate EVT are exposed. In particular the possible limiting distribution for maxima of multivariate samples are derived, and equivalent formulations in terms of limiting distributions of excesses are stated. Multivariate regular variation plays a central role when considering such multivariate 'excesses above large thresholds'.

Notations The usual order relation $\leq$ on $\mathbb{R}$ is extended to a partial order $\preceq$ on $\mathbb{R}^d$: $a \preceq b$ means $\forall j \in \{1, \dots, d\}, a_j \leq b_j$. Similarly $a \prec b$ means $a_j \prec b_j$ for all $j \leq d$. For $a \preceq b$, the 'rectangle' $[a, b]$ is the product $\prod_{i=1}^d [a_i, b_i]$. Binary operations are understood componentwise, *e.g.* $\vee(a, b) = (\vee(a_1, b_1), \vee(a_2, b_2), \dots, \vee(a_d, b_d))$. If $(X_n)_{n \geq 0}$ is an i.i.d. sample $\Omega \to \mathbb{R}^d$ we denote $X_i^{(j)}$ the j^{th} component of X_i.

3.1 Limit distributions of maxima

Let $X, X_1, \dots, X_n \overset{i.i.d.}{\sim} F$, where F is a multivariate *c.d.f.* $\mathbb{R}^d \to \mathbb{R}^+$ and $X = (X^{(1)}, \dots, X^{(d)})$ is a d-variate random vector. The basic assumption in multivariate EVT (MEVT) is that $\exists (a_n = (a_{1,n}, \dots, a_{d,n}))_{n \in \mathbb{N}} \succ 0, \exists (b_n = (b_{1,n}, \dots, b_{d,n}))_{n \in \mathbb{N}} \in \mathbb{R}^d, \exists Z$ a non-degenerate r.v., such that

$$\frac{\bigvee_{i=1}^n X_i - b_n}{a_n} \xrightarrow{w} Z. \tag{3.1}$$

An equivalent statement in terms of distribution is that $\exists G$ an non-degenerate *c.d.f.* on $\mathbb{R}^d$, such that

$$F^n(a_n \cdot + b_n) \xrightarrow{w} G(\cdot). \tag{3.2}$$

Any such limit G is called a **Multivariate extreme value distribution** (MEVD).

3.1.1 Max-stability

Definition 3.1.1. *A* c.d.f. $G : \mathbb{R}^d \to [0, 1]$ *is called* ***max-stable*** *if* $\forall t > 0, \exists \alpha(t) \succ 0, \exists \beta(t) \in \mathbb{R}^d$:

$$G(x) = G^t(\alpha(t)x + \beta(t)). \tag{3.3}$$

A non-degenerate random variable Z is called max-stable *if its distribution function is.*

Remark 3.1.2. *If* (3.3) *holds true for* $t \in \mathbb{N}$*, then it also holds for rational* t*'s, thus also for every positive real* t*. Thus, a non-degenerate random variable* Z *is max-stable if and only if for an i.i.d. sample* $(Z_n)_{n\geq 1} \sim Z$*,* $\quad \forall n \in \mathbb{N}^*,\ \exists \alpha(n) \succ 0,\ \exists \beta(n) \in \mathbb{R}^d$:

$$\frac{\bigvee_{i=1}^n Z_i - \beta(n)}{\alpha(n)} \stackrel{d}{=} Z \tag{3.4}$$

Proposition 3.1.3
A non-degenerate multivariate c.d.f. *is a MEVD if and only if it is max-stable.*

Proof. The fact that a max-stable distribution is a MEVD one is immediate: take $a_n = \alpha(n), b_n = \beta(n)$, then (3.2) since the right-hand side and the left-hand side are equal for all n. For the converse statement, notice first that multivariate weak convergence (3.2) entails weak convergence of the margins (to see this, apply (3.2) to $\mathbf{x} = (\infty, \infty, \ldots, x, \infty, \ldots)$). The arguments in the proof of Proposition 1.3.2 show that for all $1 \leq j \leq d$ and $t > 0$,

$$\exists \alpha_j(t) > 0, \exists \beta_j(t)\ :\ \frac{a_{j,\lfloor nt \rfloor}}{a_{j,n}} \xrightarrow[n\to\infty]{} \alpha_j(t)\ ;\ \frac{b_{j,\lfloor nt \rfloor} - b_{j,n}}{a_{j,n}} \xrightarrow[n\to\infty]{} \beta_j(t).$$

Define $\alpha(t) = (\alpha_1(t), \ldots, \alpha_d(t))$ and $\beta(t)$ similarly. On the one hand, for all $x \in \mathbb{R}^d$,we have

$$F^{\lfloor nt \rfloor}(a_{\lfloor nt \rfloor} x + b_{\lfloor nt \rfloor}) \xrightarrow[n\to\infty]{} G(x). \tag{3.5}$$

On the other hand, write

$$a_n x + b_n = a_{\lfloor nt \rfloor} \underbrace{\left(\frac{a_n}{a_{\lfloor nt \rfloor}} \left(x - \frac{(b_{\lfloor nt \rfloor} - b_n)}{a_n} \right) \right)}_{= y_n} + b_{\lfloor nt \rfloor}.$$

Notice that $y_n \xrightarrow[n\to\infty]{} \frac{x-\beta(t)}{\alpha(t)} := y$. Thus

$$F^{\lfloor nt \rfloor}(a_n x + b_n) = F^{\lfloor nt \rfloor}(a_{\lfloor nt \rfloor} y_n + b_{\lfloor nt \rfloor}) \xrightarrow[n\to\infty]{} G(y)$$

where the latter convergence is easily proved using $y_n \to y$ and the continuity of the limit G at y. The latter display combined with (3.5) shows that $G^t(x) = G(\frac{x-\beta(t)}{\alpha(t)})$. ■

3.1.2 Max infinite divisibility

Definition 3.1.4. *A* c.d.f. F *on* $\mathbb{R}^d$ *is called* ***max-infinitely divisible*** *(max-id) if* $\forall t > 0$*,* F^t *is a distribution function on* $\mathbb{R}^d$*.*

Remark 3.1.5. *1. A function* $H : \mathbb{R}^d \to [0, 1]$ *'is a* c.d.f.*' if it is the distribution function of a probability measure on* $\mathbb{R}^d$*. This is the case if and only if the following conditions are satisfied*

(a) $\lim_{x_i \to -\infty} H(x_1, \ldots, x_i, \ldots, x_d) = 0, \quad \forall i \leq d,\ \forall (x_j)_{j\neq i} \in \mathbb{R}^{d-1}$.

(b) $\lim_{x_1,\ldots,x_d \to +\infty} H(x_1, \ldots, x_d) = 1$

(c) $\forall j \leq d,\ x_j \mapsto H(x_1, \ldots, x_j, \ldots, x_d)$ *is right-continuous.*

(d) *For $a \preceq b$, denoting for $\beta \subset \{1,\dots d\}$, $x_{\beta,j} \begin{cases} a_j & \text{if } j \in \beta \\ b_j & \text{if } j \in \{1,\dots,d\} \setminus \beta \end{cases}$, it holds that*

$$H\{(a,b]\} := \sum_{\beta \subset \{1,\dots,d\}} (-1)^{|\beta|} H(x_\beta) \geq 0.$$

Thus, if F is a multivariate c.d.f. *then F^t is a* c.d.f. *if and only if condition (d) is satisfied for F^t.*

2. *If $d = 1$ and F is a* c.d.f. *then F^t is a* c.d.f. *for all $t > 0$. Thus every univariate* c.d.f. *is max-id.*

3. *Every max-stable distribution is max-id*

Exercise 3.1:
Are the *c.d.f.*'s F associated with the following probability distributions P max-id ?

1. in $\mathbb{R}^d$, $P = \delta_a$ for some $a \in \mathbb{R}^d$.
2. in $\mathbb{R}^2$, $P = \frac{1}{2}\left(\delta_{(1,0)} + \delta_{(0,1)}\right)$
3. in $\mathbb{R}^2$, $P = \frac{1}{2}\left(\delta_{(0,0)} + \delta_{(1,1)}\right)$

3.1.3 Characterizing Max-ID distributions

To begin with, we give an example of construction of max-id distribution. We shall prove later on that every max-id distribution can be represented this way.

Canonical example Let $\ell \in \mathbb{R}^d$ and let $E = [\ell, \infty] \setminus \{\ell\}$ the compactified orthant deprived from its 'origin' ℓ. In the sequel we need the following definition. Recall that point processes are special cases of random measures taking value in $\mathrm{M}_p(E)$.

Definition 3.1.6 (Poisson process). *A point process $N = \sum_{i \geq 1} \delta_{X_i}$ on a LCSCH space E is a* Poisson process *if and only if there is a Radon measure λ on E such that for all disjoint, measurable sets $(C_1, \dots, C_k)$ in E,*

1. *$(N(C_1), \dots, N(C_k))$ is a random vector*
2. *the random variables $N(C_i)_{i \leq k}$ are independent*
3. *$N(C_i) \sim \mathrm{Poisson}(\lambda(C_i))$*

The measure λ is called the intensity measure of N and we write $N \sim PP(\lambda)$ (PP stands for 'Poisson Process').

Let μ be a Radon measure on $(E, \mathcal{E})$. Consider the product space $E' = \mathbb{R}^+ \times E$. Define a Radon measure on E' as the product measure Lebesgue $\otimes \mu$, that is

$$\lambda((t_1, t_2) \times A) = (t_2 - t_1)\mu(A) \ , \qquad A \in \mathcal{E}, t_1 \leq t_2.$$

Let $N \sim PP(\lambda)$. Then one may write $N = \sum_{n\geq 0} \delta_{(t_n, Z_n)}$, where $t_n \in \mathbb{R}^+$ and $Z_n \in E$. For $t > 0$ consider the random variable

$$Y(t) = \left(\bigvee_{\{k \,:\, t_k < t\}} Z_k \right) \vee \ell.$$

Let F_t denote the distribution function of $Y(t)$. We shall show that $F_t = F_1^t$, thus F_1 will be proved to be a max-id distribution on $\mathbb{R}^d$, provided $Y(1)$ is indeed real valued with probability 1. By definition of F_t, we have for $x \not\succeq \ell$, $F_t(x) = 0$ and for $x \succeq \ell$,

$$\begin{aligned} F_t(x) &= \mathbb{P}\left(\bigvee_{k:t_k \leq t} Z_k \preceq x \right) \\ &= \mathbb{P}\left(N\left([0,t] \times [\ell, x]^c\right) = 0 \right) \\ &= \exp\{-\lambda([0,t] \times [\ell,x]^c)\} \\ &= e^{-t\mu([\ell,x]^c)}, \end{aligned}$$

where $[\ell, x]^c = E \setminus [\ell, x] \subset E$.

In particular $F_1(x) = e^{-\mu[\ell,x]^c}$ and we indeed have $F_t = F_1^t$. To make sure that $Y(1) \in \mathbb{R}^d$ with probability 1, we need to ensure that

1. $F_1(x) \xrightarrow[x \to +\infty]{} 1$, *i.e.* $\mu[\ell, x]^c \to 0$ as $x \to \infty$ in $\mathbb{R}^d$,
2. $F_1(x) \xrightarrow[x_j \to -\infty]{} 0$ for all $j \leq d$,

This is the case if and only if μ is such that

$$\begin{cases} \mu\Big[\cup_{j=1}^d \{x_j = +\infty\} \Big] = 0 & \text{(`}\mu\text{ puts no mass on lines at infinity')} \quad (3.6a) \\ \ell \succ \infty \text{ or } \mu[\ell, x]^c \xrightarrow[x \searrow \ell]{} +\infty & \text{(explosion at the origin)} \quad (3.6b) \end{cases}$$

Every Radon measure μ satisfying (3.6a) and (3.6b) thus gives rise to a max-id distribution function $F(x) = e^{-\mu[\ell,x]^c}$ on $\mathbb{R}^d$. In fact, every max-id distribution is of this kind, as shown next.

Proposition 3.1.7 (characterization of max-id distributions)
Let F be a non degenerate c.d.f. *on $\mathbb{R}^d$ and let $Y \sim F$. Then F is max-id if and only if $\exists \ell \in [-\infty, \infty]^d \setminus \{\infty\}$, $\exists \mu$ a Radon measure on $E = [\ell, \infty] \setminus \ell$, satisfying 3.6a and 3.6b, such that*

$$F(x) = \begin{cases} 0 & \text{if } \ell \not\preceq x \\ e^{-\mu[\ell,x]^c} & \text{if } \ell \preceq x. \end{cases} \tag{3.7}$$

In such a case, $\exists N = \sum_{k \geq 0} \delta_{(t_k, Z_k)}$ a Poisson Process on $E' = \mathbb{R}^+ \times E$ with intensity measure $\lambda =$ Lebesgue $\otimes \mu$, such that

$$Y \stackrel{d}{=} \left(\max_{k:t_k \leq 1} Z_k \right) \bigvee \ell,$$

with the convention that $\bigvee_{k \in \varnothing} Z_k = -\infty$.

sketch of proof. The sufficiency has been shown in the argument before the statement. Conversely, let F be max-id. We need to show that

1. The set $R = \{x \in [-\infty,\infty]^d : F(x) > 0\}$ is a rectangle of the kind $\prod_{j=1}^d R_j$ with $R_j = (\ell_j,\infty]$ or $R_j = [\ell_j,\infty]$ for some $\ell_j \in [-\infty,\infty)$.

2. There exists a Radon measure μ on $[-\infty,\infty]^d$ satisfying 3.6a, 3.6b such that

$$\begin{cases} \mu[\ell,x]^c = -\log F(x) & (x \succeq \ell) \\ \mu[-\infty,\infty]^d \setminus [\ell,\infty)^d = 0. \end{cases} \tag{3.8}$$

1. Define $R_j = \{x \in [-\infty,\infty] : F_j(x) > 0\}$. We want to show that $R = \prod_{j=1}^d R_j$. First, if $x \in R$ then for all j $F_j(x_j) > 0$ thus $x \in \prod R_j$. To prove the converse inclusion, let $x \in \prod_j R_j$. We need to show that $x \in R$. For all $i \le d$ it holds that $F_i(x_i) > 0$, whence $\exists \mathbf{y}_i \in R$ such that $x_i = \pi_i(\mathbf{y}_i)$. Define $\mathbf{y} = \wedge_{j=1}^d \mathbf{y}_j$. By construction, $\pi_i(\mathbf{y}) \le \pi_i(\mathbf{y}_i) = x_i$ for all i, thus $\mathbf{y} \preceq x$. It remains to show that (i) $\mathbf{y} \in R$, and that (ii) ($\mathbf{y} \in R, \mathbf{y} \preceq x \Rightarrow x \in R$). Claim (ii) derives immediately from the fact that F is non-decreasing along each coordinate. As for claim (i), since $\mathbf{y}_i \in R$ for all i, it is enough to show that for $y,z \in R$, we have $y \wedge z \in R$. To do so, we use the fact that for any distribution function H, by a union bound on the event $Z \not\preceq (y \wedge z)$ where $Z \sim H$, we have
$$1 - H(y \wedge z) \le (1 - H(y)) + (1 - H(y)).$$
Since F is max-id, $F^{1/n}$ is a distribution function for all n, so that the previous display yields
$$n(1 - F^{1/n}(y \wedge z) \le n(1 - F^{1/n}(y)) + n(1 - F^{1/n}(y)).$$
Taking the limit as $n \to \infty$ we obtain
$$-\log F(y \wedge z) \le -\log F(y) - \log F(z)$$
thus
$$F(y \wedge z) \ge F(y)F(z).$$
Finally, if $y,z \in R$ then $F(y) > 0$ and $F(z) > 0$. The previous display implies $F(y \wedge z) > 0$, *i.e.* $y \wedge z \in R$, which concludes the proof of (1).

2. We shall obtain μ as a vague limit of rescaled versions of the probability distributions P_t associated with F^t. Define $\mu_n = nP_{1/n}$. We show that μ_n is relatively compact and that the limits of two subsequences must coincide.

   - for $x \succ \ell$, denoting $[\ell,x]^c = [\ell,\infty] \setminus [\ell,x]$, we have

$$\begin{aligned} \mu_n[\ell,x]^c &= n(1 - F(x)^{1/n}) \\ &= n(1 - e^{1/n \log F(x)}) \quad \sim_{n\to\infty} n(-1/n \log F(x)) \\ &\xrightarrow[n\to\infty]{} -\log F(x) \end{aligned}$$

   Thus two limits of any two subsequences must coincide on the sets $[\ell,\infty]^c$, which generate the Borel σ-field $\mathcal{E}$ on E, so that thy coincide everywhere.

- As for sequential compactness, it is enough to show that for all compact set K in $E = [\ell, \infty] \setminus \{\ell\}$, $\sup_n \mu_n(K) < \infty$. But for such K, we have $K \subset [\ell, \infty] \setminus [\ell, \delta]$ for some $\delta \succ \ell$. Thus $\sup_n \mu_n(K) \leq \sup_n \mu_n[\ell, \infty] \setminus [\ell, \delta] < \infty$ since $\mu_n[\ell, \infty] \setminus [\ell, \delta] \to -\log F(\delta) < \infty$.

■

Exercise 3.2 (Resnick (1987), chap. 5: dependence structures)**:**

1. Let F be max-id on $\mathbb{R}^d$ with exponent measure μ. Show that F is a product $F(x) = \prod_{j=1}^{d} F_j(x_j) \iff \mu$ concentrates on the translated axes $\mathcal{C}_j = \{\ell_j + t\mathbf{e}_j, t \geq 0\}$, where $\mathbf{e}_j$ is the j^{th} canonical basis vector.
2. Let Y with max-id distribution in $\mathbb{R}^4$ with exponent measure μ. Give necessary and sufficient conditions on μ for (Y_1, Y_2) to be independent from (Y_3, Y_4).
3. Give an example of exponent measure in $\mathbb{R}^3$ such that $\mu(\ell, \infty]^3 = 0$ but no Y_i is independent from the complementary pair (for $Y \sim F$ with exponent measure μ).

3.2 Characterization of simple max-stable distributions

3.2.1 Reduction to the standard case

Since max-stable distributions are in particular max-id, we already know that any max-stable distribution G writes $G(x) = e^{-\mu[\ell,x]^c} \mathbb{1}_{x \succeq \ell}$ for some $\ell \in [-\infty, \infty)^d$ and some Radon measure μ satisfying 3.6a and 3.6b. More structure can be obtained when G has 'standard' margins, *i.e.*unit Fréchet margins

Definition 3.2.1 (Simple max-stable vector/distribution)**.** *$Z : \Omega \to \mathbb{R}^d$ is 'simple max-stable' if*

1. *For all $j \in \{1, \ldots, d\}$: $Z^{(j)} \sim \Phi_1$, i.e. $\mathbb{P}(Z^{(j)} \leq x) = e^{-1/x}, x > 0$, and*
2. *$\frac{1}{n} \bigvee_{i=1}^{n} Z_i \overset{d}{=} Z_1$, for $(Z_i)_{i \in \mathbb{N}} \overset{i.i.d.}{\sim} Z$.*

Proposition 3.2.2 (Standardizing max-stable distribution)
Let Z be a non-degenerate random vector. Then Z is a max-stable vector with normalizing sequences (a_n, b_n) such that $\frac{\bigvee Z_i - b_n}{a_n} \overset{d}{=} Z$ for $Z_i \overset{i.i.d.}{\sim} Z$, if and only if

1. *$\forall j$, $Z^{(j)}$ is a max-stable variable with norming constants $(a_{n,j}, b_{n,j})$*
2. *$Z^* := (-1/\log G_1(Z^{(1)}), \ldots, -1/\log G_d(Z^{(d)}))$ is a simple max-stable vector.*

The proof is left as an exercise.
Hint: For the direct implication, show that for $x \succ 0$, $\mathbb{P}\left(\vee_{i=1}^{n} Z_i^*/n \preceq x\right) = \mathbb{P}\left(Z_1^* \preceq x\right)$. The converse is similar.

A consequence of Proposition 3.2.2 is that it is enough to characterize simple max-stable vectors.

3.2.2 Angular Measure

If G is simple max-stable, we have $G^t(tx) = G(x)$ and the support of G is $[0,\infty]^d$. We thus take $E = [0,\infty]^d \setminus \{0\}$, and we have

$$G(x) = \begin{cases} e^{-\mu[0,x]^c} & \text{if } x \succeq 0 \\ 0 & \text{otherwise} . \end{cases}$$

From the homogeneity property of G we deduce that for $t > 0$,

$$\mu([0,tx]^c) = t^{-1}\mu[0,x]^c,$$

so that for any $A \in E$, measurable, $\mu(tA) = t^{-1}\mu(A)$.

Choose any norm $\|\cdot\|$ on $\mathbb{R}^d$ and denote S_{d-1} the corresponding positive orthant of the unit sphere. Define a (finite) measure Φ on S_{d-1} :

$$\Phi(A) = \mu\{tA \ : t \geq 1\}.$$

Exercise 3.3:
Show that Φ is a finite measure.

Let T denote the polar coordinate transformation, that is

$$T : [0,\infty]^d \setminus \{0\} \to (0,\infty] \times S_{d-1}$$
$$x \mapsto (\|x\|, \frac{1}{\|x\|}x).$$

Now the image measure of μ satisfies, for $r > 0$ and $A \subset S_{d-1}$,

$$\mu \circ T^{-1}[r,\infty] \times A r^{-1}\Phi(A),$$

it is thus a product measure. In other words, if (R, W) is a random pair following the probability measure $\frac{\mu\circ T^{-1}}{\mu\circ T^{-1}[1,\infty]\times S_{d-1}}$, Then $R \perp\!\!\!\perp W$, $W \sim \mathit{Pareto}(1)$ and $W \sim \Phi(\cdot)/\Phi(S_{d-1})$. Notice that R is a 'radius' and W is an 'angle'.

From a statistical point of view, this means that only the angular component of μ need to be estimated.

Recovering G from Φ Recall $G(x) = \exp\{-\mu[0,x]^c\}$. How to write G as a function of Φ ?

A change of variables and a call to Fubini show that

$$\mu[0,x]^c = \int_{S_{d-1}} \max_{j=1,\dots,d} \frac{w_j}{x_j}\, \mathrm{d}\Phi(w).$$

Exercise 3.4:
Prove the latter display.

Thus, for $x \succeq 0$,

$$G(x) = e^{-\int_{S_{d-1}} \max_{j=1,\dots,d} \frac{w_j}{x_j}\, \mathrm{d}\Phi(w)} \tag{3.9}$$

The choice of unit Fréchet margins implies that $G_j(1) = e^{-1}$. Writing (3.9) at $x = (0,\dots,0,1,0,\dots)$, one obtains 'first moment constraints' of Φ,

$$\forall j \leq d, \quad \int_{S_{d-1}} w_j \mathrm{d}\Phi(w) = 1. \tag{3.10}$$

Exercise 3.5:
Show that conversely, any distribution function of the kind $G(x) = \exp\{-\mu[0,x]^c\}$ with $\mathrm{d}\mu \circ T^{-1} = \frac{\mathrm{d}r}{r^2}\mathrm{d}\Phi$ where Φ is finite and satisfies (3.10) is simple max-stable.

We summarize the discussion:

Proposition 3.2.3 (Characterization of simple max-stable distributions)
Let G be a c.d.f. *on $\mathbb{R}^d$, and let Y be a random vector distributed according to G. Define $E = [0,\infty]^d \setminus \{0\}$. The following are equivalent.*

1. *G is simple max-stable.*

2. *$\exists \mu$ a Radon measure on E verifying*

$$\forall t > 0, \forall A \subset E \text{ measurable}, \mu(tA) = t^{-1}\mu(A) \tag{3.11}$$
$$\mu\{x \in E : x_j \geq 1\} = 1, \quad \forall j \in \{1,\dots,d\}\,. \tag{3.12}$$

   such that

$$G(x) = \begin{cases} 0 & \text{if } x \notin E \\ \exp\Big\{ -\mu[0,x]^c & \text{if } x \in E \end{cases}$$

3. *$\exists \Phi$ a finite measure on S_{d-1} satisfying the moment constraints (3.10) such that for $x \in \mathbb{R}^d$,*

$$G(x) = \begin{cases} 0 & \text{if } x \notin E \\ \exp\Big\{ -\int_{S_{d-1}} \max_{j=1}^d \frac{w_j}{x_j}\,\mathrm{d}\Phi(w)\Big\} & \text{if } x \in E \end{cases}$$

4. *$\exists$ a Radon measure μ satisfying (3.11) and (3.12) and $\exists$ a point process $N = \sum_{k\geq 1}\delta_{t_k,Z_k}$ with intensity measure $\mathrm{d}\lambda = \frac{\mathrm{d}t}{t^2}\otimes\mathrm{d}\mu$ such that*

$$Y = \bigvee_{k:t_k\leq 1} Z_k \bigvee 0.$$

5. *$\exists \Phi$ a finite measure on S_d verifying (3.10) and $\exists \Gamma = \sum_{k\geq 0}\delta_{R_k,W_k}$ a Poisson process with intensity measure $\frac{\mathrm{d}r}{r^2}\otimes\mathrm{d}\Phi$ such that $Y \stackrel{d}{=} \bigvee_{k\geq 1} R_k W_k$.*

Proof. The equivalence between (1), (2), (3) and (4) follows from the argument preceding the statement. The equivalence with statement (5) remains to be shown. We prove (3) $\iff$ (5).
... ∎

We conclude this section with a series of exercises taken from Resnick (1987), chapter 5.4.

Exercise 3.6:
Give an example of one of each of the following

- A *c.d.f.* G which is max-id but not max-stable
- A max-stable distribution which has one marginal distribution degenerate
- A max-stable distribution which is not absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$.

(take $d = 2$ for simplicity)

Exercise 3.7:
On $\mathbb{R}^3_+$, define $G(x, y, z) = \exp\{-\frac{1}{2}(x^{-1} \vee y^{-1} + x^{-1} \vee z^{-1} + y^{-1} \vee z^{-1})\}$. Check that G is simple max-stable. Pick a norm, and characterize the angular measure Φ associated with μ, where $G(x) = e^{-\mu[0,x]^c}$.

Exercise 3.8:
Same as Exercise 3.7 with $G(x, y, z) = \exp\{ -\frac{1}{2}(\sqrt{x^{-2} + y^{-2}} + \sqrt{x^{-2} + z^{-2}} + \sqrt{y^{-2} + z^{-2}})\}$.

Exercise 3.9:
Consider $d = 2$ and the angular measure for the L_2 norm, $\Phi = \sqrt{2}\delta_{(1/\sqrt{2},1/\sqrt{2}))}$. Give the expression of the simple max-stable *c.d.f.* G associated with Φ.

3.3 Maximum domain of attraction and Peaks-Over-Threshold: the multivariate case

As in the 1D case, the MDA condition may be reformulated in terms of Peaks-Over-Threshold conditions. A preliminary step is standardization. Then the 1D definition of regular variation can be extended and an interpretation in terms of point process convergence is possible.

3.3.1 Standardization in a max-domain of attraction

Idea: the multivariate MDA condition for a random vector X is equivalent to

1. MDA conditions on every component $X^{(j)}$
2. A standard MDA condition on the standardized variable

$$X^* = (-1/\log F_1(X^{(1)}), \dots, -1/\log F_d(X^{(d)})),$$

which writes $\frac{1}{n}\bigvee_1^n X_i^* \xrightarrow{w} Z^*$, where Z^* is a simple max-stable random vector.

Proposition 3.3.1
Let $X, X_i, i = 1, \dots, n \overset{i.i.d.}{\sim} F$. Assume that the marginal distributions F_j are continuous. Define for $i \geq 1$,

$$X_i^* = \Big(\frac{1}{1 - F_1(X_{i,1})}, \dots, \frac{1}{1 - F_d(X_{i,d}})\Big) .$$

Let G be a non-degenerate max-stable d.f. and let $Y \sim G$. The following statements are equivalent.

1. $\exists (a_n)_{n\geq 0} \in (\mathbb{R}_+^*)^d, \exists (b_n)_{n\geq 0} \in \mathbb{R}^d$: $\frac{\bigvee_{i=1}^n X_i - b_n}{a_n} \xrightarrow{w} Y$
2. *Marginal and joint weak convergence both occur:*

$$\begin{cases} \forall j \in \{1, \dots, d\}, \frac{\bigvee_{i=1}^n X_{i,j} - b_{n,j}}{a_{n,j}} \xrightarrow{w} Y_j \sim G_j, \\ \frac{1}{n}\bigvee_{i=1}^n X_i^* \xrightarrow{w} Y^* := (\frac{-1}{\log G_1(Y_1)}, \dots, \frac{-1}{\log G_d(Y_d)}) \end{cases}$$

Proof. We show that 1. $\Rightarrow$ 2., the proof of the converse statement is similar.
Assume 1. Then marginal convergence is immediate (joint weak convergence implies marginal weak convergence). We need to prove the weak convergence of the standardized maxima. It is enough to show that

$$\mathbb{P}^n(X^* \prec nx) \xrightarrow{w} \mathbb{P}(Y^* \prec x).$$

For $x \in \mathcal{C}(Y^*)$, we have, denoting $U_j(p) = F_j^{\leftarrow}(1 - 1/p)$.

$$\begin{aligned}\mathbb{P}^n(X^* \prec nx) &= \mathbb{P}\left(F_j(X_j) < 1 - \frac{1}{nx_j},\ j = 1, \dots, d\right)^n \\ &= \mathbb{P}\left(X_j < F_j^{\leftarrow}(1 - \frac{1}{nx_j}),\ j = 1, \dots, d\right)^n \\ &= \mathbb{P}\left(\frac{X_j - b_{n,j}}{a_{n,j}} < \frac{U_j(nx_j) - b_{n,j}}{a_{n,j}},\ j = 1, \dots, d\right)^n\end{aligned}$$

Now, the function $x \mapsto \frac{U_j(nx_j)-b_{n,j}}{a_{n,j}}$ is the generalized inverse of $x \mapsto 1/(n(1-F_j(a_{n,j}x+b_{n,j})))$. The latter expression is equivalent, for fixed x as $n \to \infty$, to $-1/\log(F_j^n(a_{n,j}x + b_{n,j}))$ which converges weakly to $-1/\log G_j(x_j)$ by continuous mapping. Lemma 1.2.4 (weak convergence of the inverse) thus implies that $\frac{U_j(nx_j)-b_{n,j}}{a_{n,j}} \xrightarrow{w} (-1/\log(G_j(x)))^{\leftarrow} = G_j^{\leftarrow}(e^{-1/x})$. By continuity, we obtain that

$$\begin{aligned}\mathbb{P}^n(X^* \prec nx) &\sim_{n\to\infty} \mathbb{P}\left(\frac{X_j - b_{n,j}}{a_{n,j}} < G_j^{\leftarrow}(e^{-1/x_j}),\ j = 1, \dots, d\right)^n \\ &\xrightarrow[n\to\infty]{} G(G_1^{\leftarrow}(e^{-1/x_1}), \dots, G_1^{\leftarrow}(e^{-1/x_1})) \\ &= \mathbb{P}(G_j(Y_j) < e^{-1/x_j},\ j = 1, \dots, d) \\ &= \mathbb{P}(Y^* \prec x).\end{aligned}$$

■

Proposition 3.3.1 tells us that we may restrict ourselves to the max-domain of attraction of simple max-stable vectors, *i.e.* consider random vectors $X^* \in \mathbb{R}^d$ such that

$$\frac{1}{n}\bigvee_{i=1}^{n} X_i^* \xrightarrow{w} Y^* \tag{3.13}$$

where Y^* is simple max-stable with distribution $G^*(x) = e^{-\mu[0,x]^c}$, $x \succ 0$. We shall see in the next paragraphs that this condition is equivalent to the weak convergence of properly normalized excesses above large multivariate thresholds, and also to a condition on the cdf of X^* which is a multivariate generalization of the Regular variation condition (2.1) introduced in Chapter 2.

3.3.2 Convergence of multivariate Peaks-over-Threshold

Let us rephrase the standard MDA condition (3.13): we have

$$\begin{aligned}
(3.13) \iff & F^*(nx) \xrightarrow[n\to\infty]{} G^*(x), \quad x \succ 0 \\
\iff & n\log[1-(1-F^*(nx))] \xrightarrow[n\to\infty]{} \log G^*(x), \quad x \succ 0 \\
\iff & n(1-F^*(nx)) \xrightarrow[n\to\infty]{} -\log G^*(x) = \mu[0,x]^c, \quad x \succ 0 \\
\iff & n\mathbb{P}\left(n^{-1}X^* \in [0,x]^c\right) \xrightarrow[n\to\infty]{} \mu[0,x]^c \\
\iff & t\mathbb{P}\left(t^{-1}X^* \in [0,x]^c\right) \xrightarrow[t\to\infty, t\in\mathbb{R}]{} \mu[0,x]^c
\end{aligned}$$

where the last equivalence is obtained by monotonicity of F^*. The same argument as in the proof of Proposition 3.1.7 shows that the above condition implies that the measures $\mu_t(\,\cdot\,) = t\mathbb{P}(t^{-1}X^* \in \,\cdot\,)$ converge vaguely to μ, and the converse implication is immediate. Another consequence is that μ characterizes the distribution of X 'far from the origin'. Indeed consider the conditional distribution $P_t(A) = \mathbb{P}\left(X^* \in tA \mid \|X^*\| > t\right)$ defined on $\Omega^c = \{x : \|x\| \geq 1\}$. Under the standard MDA condition, we have

$$\begin{aligned}
P_t(A) &= \frac{\mathbb{P}\left(X^* \in tA\right)}{\mathbb{P}(\|X^*\| > t)} \\
&= \frac{t\mathbb{P}\left(X^* \in tA\right)}{t\mathbb{P}(\|X^*\| > t)} \\
&\xrightarrow[t\to\infty]{\frac{\mu(A)}{\mu(\Omega^c)}} = \frac{1}{Z}\mu(A)
\end{aligned}$$

with $Z = \mu(\Omega^c)$ a normalizing constant. We summarize the discussion and leave the proofs of the remaining equivalences to the reader.

Proposition 3.3.2

Let $X = (X_1, \dots, X_d)$ be a random vector with distribution F with marginal distributions $F_j, j \leq d$, and let Y be a random vector with simple max-stable cdf $G^*(x) = e^{-\mu[0,x]^c}$ on $E = [0,\infty]^d] \setminus \{0\}$ and angular measure $\Phi(B) = \mu\{tW, t \geq 1, w \in B\}$, for $B \subset S_{d-1}$, measurable. The following conditions are equivalent

1. $F^n(nx) \xrightarrow[n\to\infty]{} G^*(x), \quad x \succ 0.$
2. $\frac{1}{n}\bigvee_{i\leq n} X_i \xrightarrow{w} Y.$
3. $\mu_t(\,\cdot\,) = t\mathbb{P}\left(t^{-1}X \in \,\cdot\,\right) \xrightarrow{v} \mu$ *on* E.
4. $F_1^n(nx_1) \xrightarrow[n\to\infty]{} e^{-1/x_1}$, $x_1 > 0$ *and on the space* E,
$$\mathbb{P}\left(t^{-1}X \in \cdot \;\middle|\; \|X\| > t\right) \xrightarrow{v} \frac{\mu(\,\cdot\,)}{\mu(\Omega^c)}\,.$$
5. $F_1^n(nx_1) \xrightarrow[n\to\infty]{} e^{-1/x_1}$, $x_1 > 0$ *and letting* $R = \|X\|$ *and* $W = \|X\|^{-1}X$, *for* $B \subset S_{d-1}$ *and* $r \geq 1$,
$$\mathbb{P}\left(W \in B, R > tr \mid R > t\right) \xrightarrow[t\to\infty]{} \frac{1}{r}\,\frac{\Phi(B)}{\Phi(S_{d-1})}\,.$$

Exercise 3.10:
Show that (4.) implies (3.) in Proposition 3.3.2.

3.3.3 Multivariate regular variation

Definition 3.3.3 (Multivariate RV of functions)**.** *Let* $E = [0, \infty]^d \setminus \{0\}$ *and* $U : E \to \mathbb{R}_+$ *a real valued function of* d *variables.* U *is called multivariate regularly varying if denoting by* $\mathbf{1}$ *the constant vector* $(1, \ldots, 1) \in \mathbb{R}^d$*, there exists a function* $\lambda : E \to \mathbb{R}^+$ *such that for all* $x \in E$*,*

$$\frac{U(tx)}{Ut\mathbf{1}} \xrightarrow[t\to\infty]{} \lambda(x)$$

Proposition 3.3.4
Let λ *be a limit function in the setting of Definition 3.3.3. Then* $\exists \rho \in \mathbb{R}$ *such that* λ *is* ρ*-homogeneous,* i.e. *for* $x \in E$ *and* $s > 0$*,*

$$\lambda(sx) = s^{\rho}\lambda(x).$$

Exercise 3.11:
Prove Proposition 3.3.4.
hint: prove the statement for fixed x then show that the RV index $\rho(x)$ does not depend on x.

If U is a multivariate RV function which limit function λ has homogeneity index ρ, then U is said to be regularly varying with index ρ and we denote $U \in RV(\rho)$.

Proposition 3.3.5 (Equivalent characterization of $RV(\rho)$)
A function $U : E \to \mathbb{R}_+$ *is* $RV(\rho)$ *for some* $\rho \in \mathbb{R}$ *iff. There exists a function* $V : \mathbb{R}_+^* \to \mathbb{R}_+^* \in RV(\rho)$ *(in the univariate sense) and a limit function* $\tilde{\lambda} : E \to \mathbb{R}_+$ *such that for* $x \in E$*,*

$$\frac{1}{V(t)}\, U(tx) \xrightarrow[t\to\infty]{} \tilde{\lambda}(x)$$

Exercise 3.12:
Prove Proposition 3.3.5.

In a probabilistic context, the considered RV function are the tail distributions of random variables.

Definition 3.3.6 (multivariate RV tails)**.** *A* c.d.f. F *in* $\mathbb{R}_+^d$ *has multivariate regularly varying tail if* $1 - F$ *is a multivariate RV function in the sense of Definition 3.3.3.*

The relationships between multivariate POT, vague convergence of tail measures and tail regular variation is summarized in the next proposition (See Resnick (2007), Th. 6.1)

Proposition 3.3.7 (Tail multivariate regular variation and vague convergence of measures)
Let $E = [0, \infty]^d \setminus \{0\}$ *and let* $X \sim F$ *a random vector valued in* $\mathbb{R}_+^d$*. Let* $\alpha > 0$*. The following are equivalent*

1. $1 - F \in RV(-\alpha)$.

2. $\exists (a_n)_n \to \infty$ *and* $\exists\ \nu \in \mathrm{M}(E)$ *such that*

$$n\mathbb{P}\left(a_n^{-1}X \in \cdot\right) \xrightarrow{v} \nu$$

where ν *is a homogeneous measure with index* $-\alpha$, i.e. $\nu(tA) = t^{-\alpha}\nu(A)$, *for* $t > 0$ *and* $A \subset E$, *measurable.*

3. There exist H *a probability measure on the sphere* S_{d-1}, *a constant* $c > 0$, *and a sequence* $a_n \to \infty$ *such that letting* $R = \|X\|$ *and* $W = X/R$,

$$n\mathbb{P}\left((R/a_n, W) \in \cdot\right) \xrightarrow{v} c\nu_\alpha \otimes H$$

in $\mathrm{M}((0, \infty] \times S_{d-1})$, *where* $\nu_\alpha[x, \infty] = x^{-\alpha}, x > 0$.

3.4 Tail regular variation and Poisson limits

Recall Definition 3.1.6 for a Poisson process on a general space. In the sequel we denote $PRM(\lambda)$ (Poisson Random Measure) a Poisson point process with intensity λ. The main result of this section stated below is that tail regular variation is equivalent to convergence of the normalized marked point process $\sum_i \delta_{(i/n, X_i/a_n)}$ towards a Poisson point process which intensity measure is closely linked to the limit measure associated to the normalizing sequence a_n.

Theorem 3.4.1 (Multivariate extremes and Poisson Process)
Let $(X_i)_{i\geq 1} \overset{i.i.d.}{\sim} X$ *be random variables valued in* $\mathbb{R}_+^d$, *let* a_n *be a normalizing sequence and let* μ *be a radon measure on* $E = [0, \infty]^d \setminus \{0\}$. *The following are equivalent.*

1. $n\mathbb{P}\left(a_n^{-1}X \in \cdot\right) \xrightarrow{v} \nu$
2. *The point process* $N_n = \sum_{i=1}^n \delta_{(i/n, X_i/a_n)} \xrightarrow{w} N = PRM(Lebesgue \otimes \mu)$ *on* $[0, 1] \times E$.

To prove this result we need some facts about Poisson processes.

Proposition 3.4.2 (Poisson process: characterization and construction)
Definition 3.1.6 is non empty, i.e. *for any Radon measure* λ *on a LCSCH space* E *there exists a random measure* $N \sim PRM(\lambda)$ *satisfying properties 1.,2., 3. therein. Also, properties 2. and 3. from the definition entirely characterize the law of the considered random measure, which Laplace transform is given by*

$$\mathcal{L}_N(f) = \exp\{-\int_E 1 - e^{-f(x)} \mathrm{d}\lambda(x)\},$$

for f *a non-negative, measurable function.*

Proof.
1. uniqueness, Laplace transform. We first prove the second assertion. Let λ be a Radon measure and let N be a point process satisfying the conditions of the definition. Since the Laplace transform entirely characterizes the law of N it is enough to show that $\mathcal{L}_N$ is entirely determined and that it writes as in the statement.

(a) For f a weighted indicator function $f = c\mathbb{1}_A$ for some $c > 0$ and $A \subset E$, measurable,

$$\begin{aligned}
\mathcal{L}_N(f) &= \mathbb{E}\left(e^{-N(f)}\right) = \mathbb{E}\left(e^{-cN(A)}\right) \\
&= \sum_{k\geq 0} e^{-kc}\frac{\lambda(A)^k}{k!}e^{-\lambda(A)} \\
&= \underbrace{\left(\sum_{k\geq 0}\frac{(\lambda(A)e^{-c})^k}{k!}e^{-\lambda(A)e^{-c}}\right)}_{=1} e^{-\lambda(A)(1-e^{-c})} \\
&= e^{-\int(1-e^{-f})\mathrm{d}\lambda}
\end{aligned}$$

which proves the statement for $f = c\mathbb{1}_A$.

(b) For $f = \sum_{i=1}^p c_i\mathbb{1}_{A_i}$ with $c_i > 0$ and where $A_i, i \leq p$ are disjoint measurable sets,

$$\begin{aligned}
\mathcal{L}_N(f) &= \mathbb{E}\left(e^{-\sum_{i\leq p} c_i N(A_i)}\right) \\
&= \mathbb{E}\left(\prod_{i\leq p} e^{-c_i N(A_i)}\right) \\
&= \prod_{i\leq p}\mathbb{E}\left(e^{-c_i N(A_i)}\right) \quad \text{(independence of the } N(A_i)'s) \\
&= \prod_{i\leq p}\exp\Big\{-\int(1-e^{-c_i\mathbb{1}_{A_i}})\mathrm{d}\lambda\Big\} \quad \text{(from (a))} \\
&= \exp\Big\{-\sum_{i\leq p}\int(1-e^{-c_i\mathbb{1}_{A_i}})\mathrm{d}\lambda\Big\} \\
&= \exp\Big\{-\int(1-e^{-\sum_{i\leq p}c_i\mathbb{1}_{A_i}})\mathrm{d}\lambda\Big\} \quad (A_i's \text{ are disjoint}) \\
&= e^{-\int(1-e^{-f})\mathrm{d}\lambda},
\end{aligned}$$

where the penultimate equality comes from the fact that if $x \notin \cup_i A_i$, then $\sum_i(1 - e^{-c_i\mathbb{1}_{A_i}(x)}) = 0$, otherwise if $x \in A_i$ for some i, $\sum_i(1 - e^{-c_i\mathbb{1}_{A_i}(x)}) = 1 - e^{-c_i}$.

(c) For $f \geq 0$, measurable, there exists a pointwise non decreasing sequence $(f_n)_{n\geq 0}$ of simple positive functions of the kind $f_n = \sum_{i\leq p_n} c_{i,n}\mathbb{1}_{A_{i,n}}$ where the $A_{i,n}$'s are disjoint for fixed n, such that $f_n \to f$ pointwise. By the monotone convergence theorem, almost surely, $N(f) = \int f\mathrm{d}N = \lim_n N(f_n)$. Also for all n, $e^{-N(f_n)} \leq 1$. We may thus apply the dominated convergence theorem to the random variables $e^{-N(f_n)}$, which yields

$$\begin{aligned}
\mathcal{L}_N(f) &= \lim_n \mathcal{L}_N(f_n) \\
&= \lim_n \exp\Big\{\int(1-e^{-f_n})\mathrm{d}\lambda\Big\} \\
&= \exp\Big\{\int(1-e^{-f})\mathrm{d}\lambda\Big\}
\end{aligned}$$

By monotone convergence again, which concludes the proof of the second assertion.

2. Existence: proof by construction. We now give a constructive proof of the existence of a random measure N satisfying the conditions of the definition. We proceed in two steps, first assuming that λ is finite on E, then relaxing this assumption.

2.(a): finite measure. Assume first that λ is a finite measure on E. Let M be a random variable following a Poisson distribution with parameter $\tilde{\lambda} = \lambda(E)$. Let $X_i, i \geq 1$ be an i.i.d. sequence independent from M following the probability distribution $\nu(\,\cdot\,) = \lambda(\,\cdot\,)/\tilde{\lambda}$ on E. Set $N = \sum_{i=1}^{M} \delta_{X_i}$. Then condition 1. of Definition 3.1.6 is obviously satisfied. Also for $A \subset E$ measurable, since conditionally to $M = m$, the number of X_i's belonging to A follows a binomial distribution $(m, p = \nu(A))$,

$$\begin{aligned}
\mathbb{P}\left(N(A) = k\right) &= \sum_{m=0}^{\infty} \mathbb{P}\left(M = m, \sum_{i=1}^{m} \mathbb{1}_A(X_i) = k\right) \\
&= \sum_{m=k}^{+\infty} \frac{\tilde{\lambda}^m}{m!} e^{-\tilde{\lambda}} \times \binom{m}{k} \nu(A)^k (1 - \nu(A))^{m-k} \\
&= \sum_{m=k}^{+\infty} \frac{\tilde{\lambda}^m}{m!} e^{-\tilde{\lambda}} \times \binom{m}{k} \lambda(A)^k (\tilde{\lambda} - \lambda(A))^{m-k} \tilde{\lambda}^{-m} \\
&= \sum_{m=k}^{+\infty} \frac{\lambda(A)^k}{k!} \frac{(\tilde{\lambda} - \lambda(A))^{m-k}}{(m-k)!} e^{-\tilde{\lambda}} \\
&= \frac{\lambda(A)^k}{k!} e^{-\lambda(A)} \underbrace{\sum_{m=k}^{\infty} \frac{(\tilde{\lambda} - \lambda(A))^{m-k}}{(m-k)!} e^{-(\tilde{\lambda} - \lambda(A))}}_{=1}
\end{aligned}$$

which proves that $N(A) \sim \mathcal{P}oiss(\lambda(A))$.
Turning to the independence property, let $A_1, \dots, A_p$ be disjoint measurable sets in E and let $A_0 = E \setminus \cup_i A_i$. For $k_1, \dots, k_p \in \mathbb{N}^p$, set $\pi = \mathbb{P}(N(A_i) = k_i, i = 1, \dots, p)$ and $K = \sum_{i=1}^{n} k_i$. We need to prove that $\pi = \prod_{i=1}^{p} \mathbb{P}(N(A_i) = k_i)$. Conditionally to $M = m$, the random vector $(N(A_0), N(A_1), \dots, N(A_p))$ follows a multinomial distribution with parameters $(m, (\nu(A_0), \dots, \nu(A_p))$. Whence

$$\begin{aligned}
\pi &= \sum_{m=K}^{\infty} \mathbb{P}\left(S = m\right) \mathbb{P}\left(N(A_0) = m - K, N(A_i) = k_i, i = 1, \dots, p\right) \\
&= \sum_{m=K}^{\infty} \frac{\tilde{\lambda}^m}{m!} e^{-\tilde{\lambda}} \times \frac{m!}{(m-K)! k_1! \cdots k_p!} \nu(A_0)^{m-K} \prod_{i=1}^{p} \nu(A_i)^{k_i} \\
&= \sum_{m=K}^{\infty} e^{-\lambda(E)} \frac{\lambda(A_0)^{m-K} \prod_{i=1}^{p} \lambda(A_i)^{k_i}}{(m-K)! k_1! \cdots k_p!} \\
&= \prod_{i=1}^{p} \frac{\lambda(A_i)^{k_i}}{k_i!} e^{-\lambda(A_i)} \times \underbrace{\sum_{m=K}^{\infty} e^{-\lambda(A_0)} \frac{\lambda(A_0)^{m-K}}{(m-K)!}}_{=1} \\
&= \prod_{i=1}^{p} \mathbb{P}\left(N(A_i) = k_i\right).
\end{aligned}$$

2.(b): Infinite measure Assume now that $\lambda(E) = +\infty$. Since E is locally compact and second countable, and since the measure λ of any compact set is finite (Radon measure), we may write E as a disjoint union $E = \sqcup_{k=1}^{\infty} E_k$ with $\lambda(E_k) < \infty$. Set $\lambda_k = \lambda(\cdot \cap E_k)$. Each λ_k is finite and we have $\lambda = \sum_k \lambda_k$. Construct on each E_k a Poisson Point process $N_k \sim PRM(\lambda_k)$ as in **2.(a)**, independently from each other and let $N = \sum_{k \geq 1} N_k$. It is now easy to show that N is a Poisson point process with intensity λ, by computing the Laplace transform of N. ■

Exercise 3.13:
Finish part **2.(b)** of the proof of Proposition 3.4.2, *i.e.* show that the process N constructed in the proof is indeed a Poisson process.

The following proposition allows to construct variants of a Poisson point process by transformation and marking

Proposition 3.4.3 (transformation and marking)
In the following statements, I is an index set which maybe $\mathbb{N}$ or a random integer interval of the kind $\{i = 1, \dots, M\}$ where M is a random integer.

(a) Let E, F be LCSCH spaces and let $T : E \to F$ be a continuous mapping. If $N \sim PRM(\lambda)$ on E then the image (random) measure $\tilde{N} = N \circ T^{-1}$ on F follows a $PRM(\lambda \circ T^{-1})$. Also is N admits the representation $N = \sum_{i \in I} \delta_{X_i}$, then

$$\tilde{N} \stackrel{d}{=} \sum_{i \in I} \delta_{T(X_i)}.$$

(b) If $N = \sum_{i \in I} \delta_{X_i} \sim PRM(\lambda)$ on E and if $(Y_i)_{i \geq 1}$ are i.i.d. according to some distribution P, independent from N, then

$$\tilde{N} = \sum_{i \in I} \delta_{(X_i, Y_i)} \sim PRM(\lambda \otimes P) \qquad \text{on } E \times F$$

Such a process $\tilde{N}$ is called a marked Poisson process.

Consider now an array $X_{i,n}, i \geq 1, n \geq 1$, where for each n, the $X_{i,n}$ are i.i.d. according to some distribution P_n on a LCSCH space E (In our context $X_{i,n} = X_i / a_n$). Define a finite measure on E , $\mu_n = n\mathbb{P}(X_{1,n} \in \cdot)$. In the above section we have shown that if $a_n \to \infty$ and $X_{i,n} = X_i / a_n$, then vague convergence of μ_n is equivalent to tail regular variation of X_1. We now show that this condition is in turn equivalent to the convergence of a marked point process based on the $X_{i,n}$'s towards a Poisson point process. Theorem 3.4.1 is thus an immediate consequence of the following proposition.

Proposition 3.4.4 (Vague convergence and Poisson process limit)
Let $X_{i,n}, i \geq 1, n \geq 1$ be a random array such that for each n, the $X_{i,n}$'s are i.i.d. according to some distribution P_n on a LCSCH space E. Define $\mu_n = nP_n$ and let μ be a Radon measure on E. Define a sequence of marked point processes $N_n = \sum_{i \geq 1} \delta_{(i/n, X_{i,n})}$ on $\mathbb{R} \times E$. The following statements are equivalent

1. *$\mu_n \xrightarrow{v} \mu$ on E*
2. *$N_n \xrightarrow{w} N \sim PRM(Lebesgue \otimes \mu)$ on $\mathbb{R}_+ \times E$.*

Proof.

1. $\Rightarrow$ **2.** Assume condition 1. from the statement. We need to show that for any non negative function $f \in \mathcal{C}_K(\mathbb{R}_+ \times E)$

$$\mathcal{L}_{N_n}(f) \to \mathcal{L}_N(f) = \exp\Big[- \int_{\mathbb{R}_+} \int_E (1 - e^{-f(t,x)}) \mathrm{d}t \mathrm{d}\mu(x) \Big].$$

$$\begin{aligned} \mathcal{L}_{N_n}(f) &= \mathbb{E}\left(\exp\Big\{ - \sum_{i=1}^{\infty} f(i/n, X_{i,n}) \Big\} \right) \\ &= \prod_{i=1}^{\infty} \mathbb{E}\left(\exp[-f(i/n, X_{1,n}))\right] \\ &= \prod_{i=1}^{\infty} \left(1 - \int_E 1 - e^{-f(i/n,x)} \mathrm{d}P_n(x) \right) \end{aligned}$$

Taking the negative logarithm,

$$-\log \mathcal{L}_{N_n}(f) = -\sum_{i=1}^{\infty} \log \left(1 - \int_E 1 - e^{-f(i/n,x)} \mathrm{d}P_n(x) \right)$$

We now show that $-\log\left(1 - \int_E 1 - e^{-f(i/n,x)}\mathrm{d}P_n(x)\right)$ may be replaced with $c_{i,n} := \int_E 1 - e^{-f(i/n,x)}\mathrm{d}P_n(x)$ in the above expression. If K is the compact support of f, note that $1 - e^{-f(t,x)} \leq \mathbb{1}_K(t,x)$. Also there exists compact sets $A \subset E$ and $I \subset \mathbb{R}_+$, such that $K \subset I \times A$. Thus

$$\sup_{i \geq 1} c_{i,n} \leq \mathbb{P}\left(X_{i,n} \in A\right) = \frac{1}{n}\mu_n(A).$$

Using the fact that $|\log(1+z) - z| \leq z^2$ for $|z| \leq 1/2$ we obtain for n sufficiently large,

$$\begin{aligned} &\left| -\log \mathcal{L}_{N_n}(f) - \sum_{i=1}^{\infty} \int_E 1 - e^{-f(i/n,x)} \mathrm{d}P_n(x) \right| \\ &= \left| \sum_{i=1}^{\infty} \log(1 - c_{i,n}) + c_{i,n} \right| \\ &\leq \sum_{i=1}^{\infty} c_{i,n}^2 \\ &\leq \sup_{i \geq 1} c_{i,n} \times \underbrace{\sum_{i=1}^{\infty} c_{i,n}}_{:= S_n} \\ &\leq \frac{\mu_n(A)}{n} \times S_n. \end{aligned} \tag{3.14}$$

In addition, the sum S_n can be seen as an integral on the product space $\mathbb{R}_+ \times E$ with respect to the measure $\lambda_n = (\sum_{i\geq 1} \delta_{i/n}) \otimes P_n$,

$$S_n = \sum_{i\geq 1} \int_E 1 - e^{-f(i/n,x)} \mathrm{d}P_n(x)$$
$$= \int_{\mathbb{R}_+\times E} 1 - e^{-f(t,x)} \mathrm{d}\lambda_n(t,x).$$

Now it is easy to see that $\mu_n \overset{v}{\to} \mu$ if and only if $\lambda_n \overset{v}{\to}$ Lebesque $\otimes\, \mu$ on $\mathbb{R}_+ \times E$. The function $(t,x) \mapsto 1 - e^{-f(t,x)}$ is continuous and its support is the same as the support K of f, which is compact in $\mathbb{R}_+ \times E$ by assumption. Thus vague convergence of λ_n yields

$$S_n \xrightarrow[n\to\infty]{} \int_{\mathbb{R}_+\times E} 1 - e^{-f(t,x)} \mathrm{d}t \otimes \mathrm{d}\mu(x) < \infty. \tag{3.15}$$

Combining (3.14) and (3.15) shows that

$$\left| -\log \mathcal{L}_{N_n}(f) - \underbrace{\sum_{i=1}^{\infty} \int_E 1 - e^{-f(i/n,x)} \mathrm{d}P_n(x)}_{S_n} \right| \xrightarrow[n\to\infty]{} 0. \tag{3.16}$$

Finally, (3.15) and (3.16) imply that

$$-\log \mathcal{L}_{N_n}(f) \xrightarrow[n\to\infty]{} \int_{\mathbb{R}_+\times E} 1 - e^{-f(t,x)} \mathrm{d}t \otimes \mathrm{d}\mu(x)$$

which concludes the proof.

2. $\Rightarrow$ **1.** Assume that $N_n \overset{w}{\to} N$ as in the statement.We use the fact that to prove vague convergence $\mu_n \overset{v}{\to} \mu$ it is enough to show that for any function $g \in \mathcal{C}_K^+(E)$,

$$\int_E 1 - e^{-g} \mathrm{d}\mu_n \to \int_E 1 - e^{-g} \mathrm{d}\mu. \tag{3.17}$$

Indeed in such a case, for any set A with compact closure such that $\mu(\partial A) = 0$, one may approach $\mathbb{1}_A$ by functions $h_M = 1 - e^{-M\mathbb{1}_A}$ and then $M\mathbb{1}_A$ by continuous functions $g_{M,p}$ in $\mathcal{C}_K^+$. The condition $\mu(\partial A) = 0$ implies that $\mu(A) = \lim_{M,p\to\infty} \int_E 1 - e^{-g_{M,p}} \mathrm{d}\mu$ whence the sufficiency of (3.17) for vague convergence of μ_n towards μ.

Let $g \in \mathcal{C}_K^+$. We need to show (3.17). Set $f(t,x) = \mathbb{1}_{[0,1]}(t)g(x)$. f is not continuous but may be approached by continuous functions $h_m \in \mathcal{C}_K^+(\mathbb{R}_+ \times E)$, so that weak convergence $N_n \overset{w}{\to} N$ implies convergence of the Laplace transforms $\mathcal{L}_{N_n}(f) \to \mathcal{L}_N(f)$. Also , we have

$$N_n(f) = \sum_{i\geq 1} \mathbb{1}_{[0,1]}(i/n) g(X_{i,n}) = \sum_{i=1}^{n} g(X_{i,n}),$$

thus

$$\mathcal{L}_{N_n}(f) = \mathbb{E}\left(e^{-\sum_{i=1}^n g(X_{i,n})}\right)$$
$$\mathbb{E}\left(e^{-g(X_{1,n})}\right)^n$$
$$= \left(1 - \frac{\int_E 1 - e^{-g} \mathrm{d}\mu_n}{n}\right)^n \tag{3.18}$$

On the other hand

$$\begin{aligned}\mathcal{L}_N(f) &= \int_{\mathbb{R}_+\times E} 1 - e^{-\mathbb{1}_{[0,1]}(t)g(x)}\mathrm{d}t\mathrm{d}\mu(x) \\ &= \int_E 1 - e^{-g(x)}\mathrm{d}\mu(x) \end{aligned} \tag{3.19}$$

In view of (3.18) and (3.19), $\mathcal{L}_{N_n}(f) \to \mathcal{L}_N(f)$ implies that $\int_E 1-e^{-g}\mathrm{d}\mu_n \to \int_E 1-e^{-g}\mathrm{d}\mu$. ■

Appendix A

Technicalities for Chapter 1

A.1 Monotone functions: additional results

Lemma A.1.1 (local uniform convergence of monotone functions)
Let $(H_n)_{n\in\mathbb{N}}$ and H be monotone functions $\mathbb{R} \to [-\infty, \infty]$, such that $H_n \xrightarrow{w} H$. If H is continuous on an interval $I \subset \mathbb{R}$ (in particular H has to be finite on I), then the convergence is locally uniform on I, i.e. *for $a < b \in I$,*

$$\sup_{x\in[a,b]} |H_n(x) - H(x)| \xrightarrow[n\to\infty]{} 0.$$

Sketch of proof. Since H is uniformly continuous on $[a, b]$; For $\epsilon > 0$, there is a subdivision $a = x_0 < x_1 < \cdots < x_K = b$; such that the variations of H are less than ϵ on each $[x_i, x_{i+1}]$. Use pointwise convergence on the finite set $(x_0, \ldots, x_k)$ and monotonicity to conclude. ■

A.2 Proof of Lemma 1.2.5 (Weak convergence of the inverse)

Weak convergence implies weak convergence of the inverse

We assume that $H_n \xrightarrow{w} H$, and we show that $H_n^{\leftarrow} \xrightarrow{w} H^{\leftarrow}$. Let $y \in \mathcal{C}(H^{\leftarrow})$. In particular $H^{\leftarrow}(y)$ is finite. Let $\epsilon > 0$. Since the discontinuity points of a monotone functions are at most countable, there exists $x \in \mathcal{C}(H)$ such that $H^{\leftarrow}(y) - \epsilon < x < H^{\leftarrow}(y)$. Then, from Lemma 1.2.4, $H(x) < y$. Since for such an x, $H_n(x) \xrightarrow[n\to\infty]{} H(x)$, we have for n large enough, $H_n(x) < y$ as well, so that, from Lemma 1.2.4 again, $x < H_n^{\leftarrow}(y)$. Thus, $\exists n_0$ such that for $n \geq n_0$, $H^{\leftarrow}(y) - \epsilon < x < H_n^{\leftarrow}(y)$. Since ϵ is arbitrary,

$$\liminf H_n^{\leftarrow}(y) \geq H^{\leftarrow}(y).$$

An upper bound on $\limsup H_n^{\leftarrow}(y)$ is obtained similarly: Since $y \in \mathcal{C}(H^{\leftarrow})$, we may choose $t > y$ such that $H^{\leftarrow}(t) \leq H^{\leftarrow}(y) + \epsilon$. Also, we may pick x' in $(H^{\leftarrow}(t), H^{\leftarrow}(t) + \epsilon) \cap \mathcal{C}(H)$. For such x', Lemma 1.2.4 implies

$$H(x') \geq t > y$$

Thus, for some n_1 and for all $n \geq n_1$, $H_n(x') \geq y$ as well, and using from Lemma 1.2.4 again, for such n,

$$H_n^{\leftarrow}(y) \leq x' \leq H_n^{\leftarrow}(y) + 2\epsilon.$$

Thus

$$\limsup H_n^{\leftarrow}(y) \leq H^{\leftarrow}(y),$$

and the proof is complete.

Converse statement

Let us assume that

$$H_n^{\leftarrow} \xrightarrow{w} H^{\leftarrow} \quad \text{as} \quad n \to \infty.$$

and that conditions (i) and (ii) from Lemma 1.2.5 are satisfied. Let $x \in \mathcal{C}(H)$ (in particular, $H(x)$ is finite) and $\epsilon > 0$. We need to show that for n large enough (say $n \geq n_0$),

$$H_n(x) \leq H(x) + \epsilon, \tag{A.1}$$

$$\text{and } H_n(x) \geq H(x) - \epsilon \tag{A.2}$$

We first show (A.1). By hypothesis (ii), $\exists x' > x : H(x) < H(x') < \sup_{t:H(t)<\infty} H(t)$. Thus $H^{\leftarrow}$ is finite on the open interval $(H(x), H(x'))$. The number of discontinuity points of $H^{\leftarrow}$ on this interval is at most countable, thus $\exists y \in \mathcal{C}(H^{\leftarrow}) : H(x) < y < H(x) + \epsilon$. Using Lemma 1.2.4, we obtain $x < H^{\leftarrow}(y)$. Weak convergence of $H_n^{\leftarrow}$ then implies that for n large enough, $x < H_n^{\leftarrow}(y)$ as well. Thus $H_n(x) < y < H(x) + \epsilon$, which proves (A.1).

For the proof of (A.2), we need to distinguish between the cases $H(x) > \inf_{\mathbb{R}} H$ and $H(x) = \inf_{\mathbb{R}} H$.

Case 1: $H(x) > \inf_{\mathbb{R}} H$**.** By continuity of H at x, we may choose $t < x$ such that $H(t) > \max(H(x) - \epsilon/2, \inf_{\mathbb{R}} H)$. Then $H^{\leftarrow}$ is finite on $(\inf_{\mathbb{R}} H, H(t))$, and again, admits only a countable number of discontinuity on this interval. Let then $y' \in \mathcal{C}(H^{\leftarrow})$ such that $H(t) - \epsilon/2 < y' < H(t)$. Lemma 1.2.4 again ensures that $H^{\leftarrow}(y') \leq t < x$, so that for n large enough, $H_n^{\leftarrow}(y') \leq x$ as well, whence $H_n(x) \geq y' > H(t) - \epsilon/2 > H(x) - \epsilon$ and (A.2) is true.

Case 2: $H(x) = \inf_{\mathbb{R}} H$**.** Since $x \in \mathcal{C}(H)$, necessarily $H(x) = \inf_{\mathbb{R}} H$ is finite, and hypothesis (i) ensures that for all $n \in \mathbb{N}$, $H_n(x) \geq H(x)$ so that (A.2) is immediate.

Bibliography

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of extremes: Theory and applications*. John Wiley & Sons: New York.

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer Verlag.

De Haan, L. and Ferreira, A. (2007). *Extreme value theory: an introduction*. Springer Science & Business Media.

Hahn, H. and Rosenthal, A. (1948). *Set functions*. University of New Mexico Press Albuquerque.

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (2012). *Extremes and related properties of random sequences and processes*. Springer Science & Business Media.

Resnick, S. (1987). *Extreme values, regular variation, and point processes, volume 4 of Applied Probability. A Series of the Applied Probability Trust*. Springer-Verlag, New York.

Resnick, S. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer Series in Operations Research and Financial Engineering.