Extremes - Lecture Notes (Draft)

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

A. Sabourin[1]

February 6, 2016

[1]LTCI, Telecom ParisTech, univeristé Paris-Saclay.

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 $X_t, 1 \le t \le n$. Some questions of interest for risk management are

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

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,

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

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

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

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 $\widehat{Q}_n(p) = X_{(n)}$. 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.

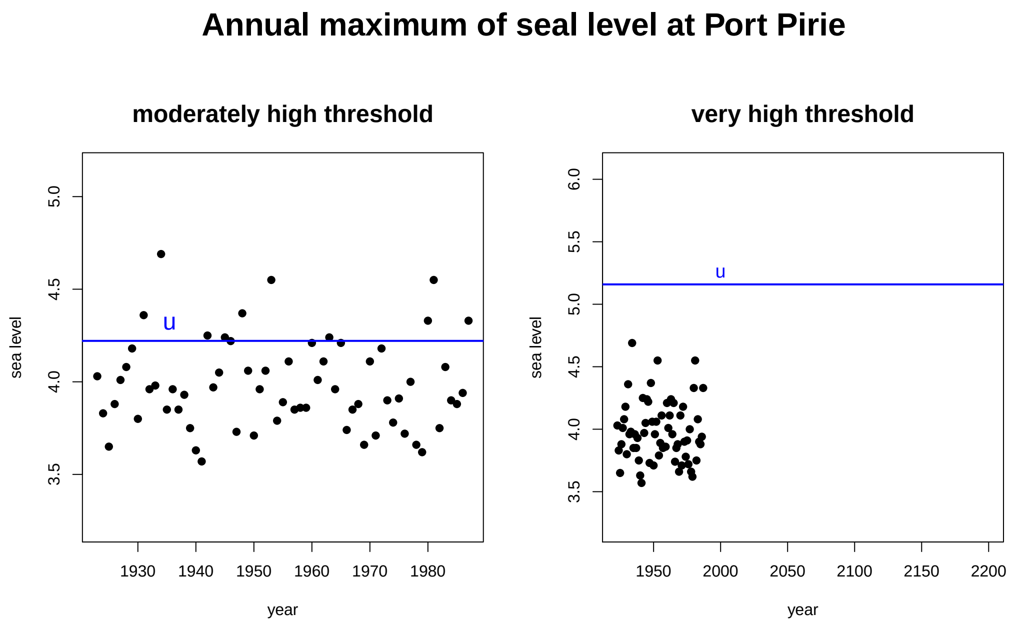

Figure 1.1: Why the empirical measure is not always useful as it is

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 with 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

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

where $\xrightarrow{d}$ 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

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

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.

1.1.4 Monotone functions and weak convergence

It is 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 convergence to the whole class of monotone functions, as follows.

Definition 1.1.3 (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

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

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 Intermediate results

1.2.1 Weak convergence of the inverse

Definition 1.2.1 (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 is $\inf\mathbb{R} = -\infty$ *and* $\inf\emptyset = +\infty$.

Remark 1.2.2. *It is left as an exercise to verify that $H^{\leftarrow}$ is indeed continuous from the left.*

Lemma 1.2.3 (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 1.2.4 (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 $n \in \mathbb{N}$, $\quad \inf_{\mathbb{R}} H_n \geq \inf_{\mathbb{R}} H$,

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

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 1.2.4 may seem intricated, but they are indeed satisfied in the particular case where we need it (*i.e.* in the proof of Theorem 1.4.1).

1.2.2 Convergence to types

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

Definition 1.2.5 (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 1.2.6 (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 1.2.4, 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 $y_1 < y_2$ such that $U^{\leftarrow}(y_1) < U^{\leftarrow}(y_2)$, and $\frac{F_n^{\leftarrow}(y_i)-b_n}{a_n} \xrightarrow{w} U^{\leftarrow}(y_i)$, $i = 1, 2$. 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 may find $z_1 < z_2$ such that $V^{\leftarrow}(z_2) - V^{\leftarrow}(z_1) > 0$,

$$\frac{F_n^{\leftarrow}(z_2) - F_n^{\leftarrow}(z_1)}{\tilde{a}_n} \xrightarrow{w} V^{\leftarrow}(z_2) - V^{\leftarrow}(z_1).$$

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

$$\frac{\tilde{a}_n}{a_n} \xrightarrow[n\to\infty]{} A := \frac{U^{\leftarrow}(z_2) - U^{\leftarrow}(z_1)}{V^{\leftarrow}(z_2) - V^{\leftarrow}(z_1)} > 0.$$

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.

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

$$\begin{aligned}
V^{\leftarrow}(y) &= \lim_{n\to\infty} \frac{F_n^{\leftarrow}(y) - \tilde{b}_n}{\tilde{a}_n} \\
&= \lim_{n\to\infty} \frac{a_n}{\tilde{a}_n}\left(\frac{F_n^{\leftarrow}(y) - b_n}{a_n} - \frac{(\tilde{b}_n - b_n)}{a_n}\right) \\
&= A^{-1}(U^{\leftarrow}(y) - B).
\end{aligned} \tag{1.5}$$

Since $V^{\leftarrow}$ and $U^{\leftarrow}$ are continuous from the left, the above equality holds for *all* $y \in \mathbb{R}$, whence, the two functions $V^{\leftarrow}$ and $A^{-1}(U^{\leftarrow} - B)$ are identical. Now, if Y is a standard uniform random variable, if we let $X = V^{\leftarrow}(Y)$ and $X' = A^{-1}(U^{\leftarrow}(Y) - B)$, then the *c.d.f.*'s of X and X' are respectively V and $U(A\cdot + B)$. But (1.5) shows that $X = X'$ almost surely. Thus X and X' have the same distribution functions, which means that $V(x) = U(Ax + B)$ for $x \in \mathbb{R}$ and (1.4) is true.

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 $y \in \mathcal{C}(U^{\leftarrow}) = \mathcal{C}(V^{\leftarrow})$,

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

We have shown that $F_n(\tilde{a}_n \bullet + \tilde{b}_n)^{\leftarrow} \xrightarrow{w} V^{\leftarrow}$, which implies, by Lemma 1.2.4, 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 $n \in \mathbb{N}$, then $\exists A_n, B_n$: $\frac{\sum_1^n X_i - B_n}{A_n} \overset{\mathrm{d}}{=} X_1$. Here and thereafter, '$\overset{\mathrm{d}}{=}$' 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 $\alpha_n = \alpha(n), \beta_n = \beta(n)$,

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

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 1.2.6, 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 (1.10) 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. ■

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

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

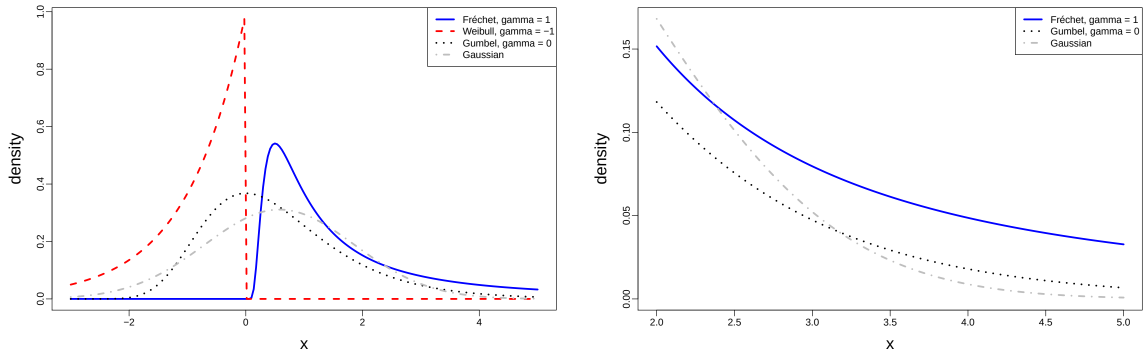

Figure 1.2: Density plot for the three extremal types, respectively ($\gamma = 1, \mu = 1, \sigma = 1$), ($\gamma = -1, \mu = -1, \sigma = 1$), ($\gamma = 0, \mu = 0, \sigma = 1$); compared with the Gaussian density with same mean and variance as the Gumbel one. The right panel is a zoom on the tail.

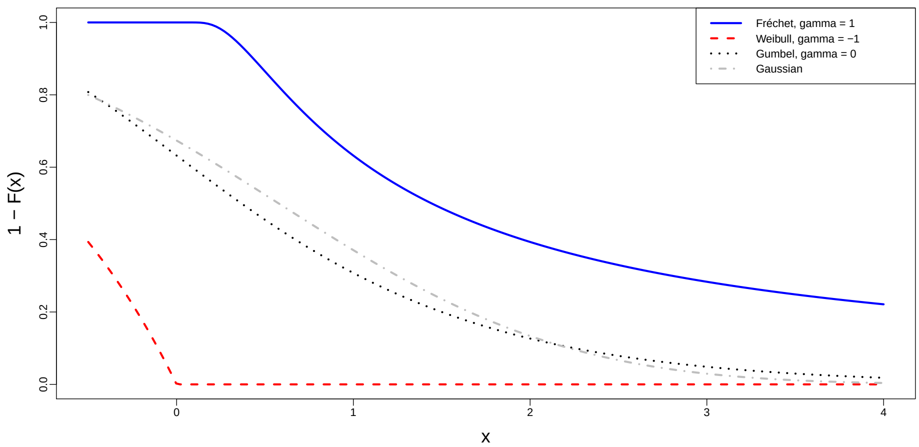

Figure 1.3: Survival function $1 - F(x)$ for the three extremal types, respectively ($\gamma = 1, \mu = 1, \sigma = 1$), ($\gamma = -1, \mu = -1, \sigma = 1$), ($\gamma = 0, \mu = 0, \sigma = 1$); compared with the Gaussian survival function with same mean and variance as the Gumbel one.

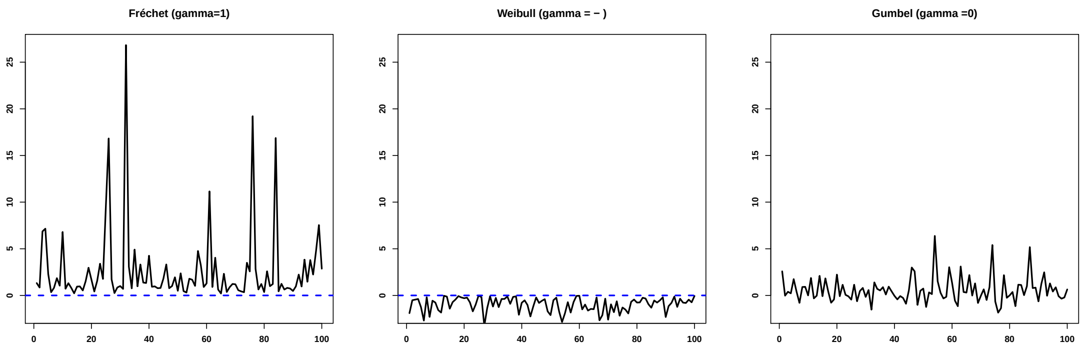

Figure 1.4: Series of i.i.d. random variables of the three extremal types, respectively ($\gamma = 1, \mu = 1, \sigma = 1$), ($\gamma = -1, \mu = -1, \sigma = 1$), ($\gamma = 0, \mu = 0, \sigma = 1$)

Exemple 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 may use heuristic calculus, 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

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

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 $x \in \mathbb{R}$,

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

Exercice 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 ?

Exercice 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

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

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

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

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 $t \in \mathbb{N}$, then by inversion also for $t \in \mathbb{Q}$, 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 $\beta(ts) = \alpha(t) + \beta(s), \quad s, t > 0$. 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 $s, t > 0$,

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

Pick $s \neq 1$. Then for $t > 0$,

$$\beta(t) = \underbrace{\frac{s^\gamma - 1}{\beta(s)}}_{C}(t^\gamma - 1) = C(t^\gamma - 1),$$

where $C \in \mathbb{R}$ is constant w.r.t. t. 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 $t \mapsto G^{1/t}(x)$ is strictly increasing. However, $G^{1/t}(x) = G(t^\gamma x)$, 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*

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

where $G_{0,1,\gamma}(x) = \exp\left(-(1 + \gamma x)_+^{-1/\gamma}\right)$. Indeed, one may choose $a'_n = \sigma a_n, b'_n = b_n + \mu a_n$, and use the convergence to type lemma 1.2.6.

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* $x \in \mathbb{R}$, $F^n(a_n x + b_n) \underset{n\to\infty}{\longrightarrow} e^{-(1+\gamma x)_+^{-1/\gamma}}$

(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

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

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 1.2.4.

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

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 $t > t_0$, $(\lfloor t \rfloor + 1)y < ty'$. Thus for large t, $\frac{U((\lfloor t \rfloor+1)y)-b(t)}{a(t)} \leq \frac{U(ty')-b(t)}{a(t)} \to \Gamma^{\leftarrow}(y') \leq \Gamma^{\leftarrow}(y) + \epsilon$. 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 (1.18); 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 1.2.4 and Remark 1.4.4, it is thus enough to show that

$$\forall y > 0, \quad \frac{U\left(\frac{y}{1-F(t)}\right) - t}{\sigma(t)} \xrightarrow[t \nearrow x_\star]{} \frac{y^\gamma - 1}{\gamma} := \Gamma^{\leftarrow}(y). \tag{1.24}$$

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 $y > 0, 0 < t < x_\star$,

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

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

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

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

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

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).

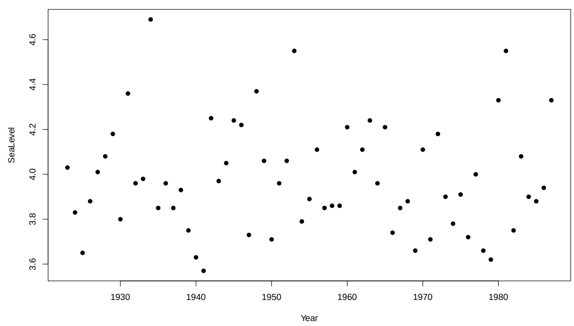

Figure 1.5: `portpirie` data in package `evd`: Annual maxima of the sea level at Port Pirie, 1923-1987

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

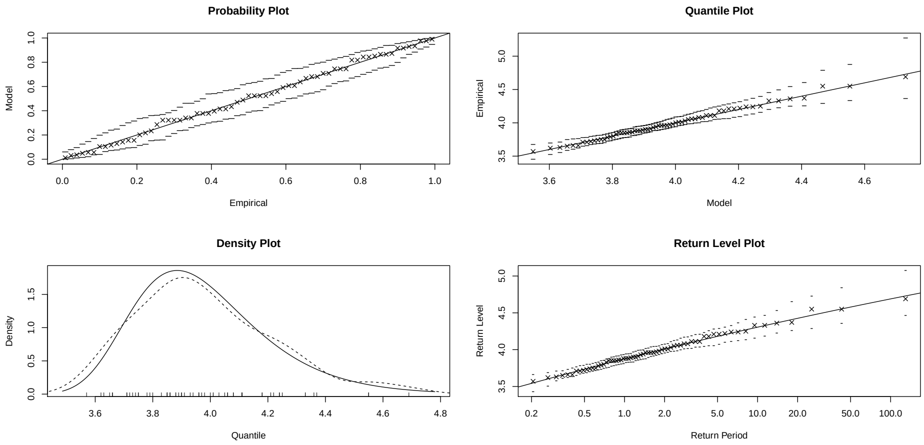

Figure 1.7: Graphical diagnostic plot for the GEV model fit on the Port Pirie dataset, as provided by `R` package `evd`.

plugin method,

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

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

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

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 submodel ($\gamma = 0$) 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 $y = i/n$ $(i = 1, \dots, n)$ is

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

(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.

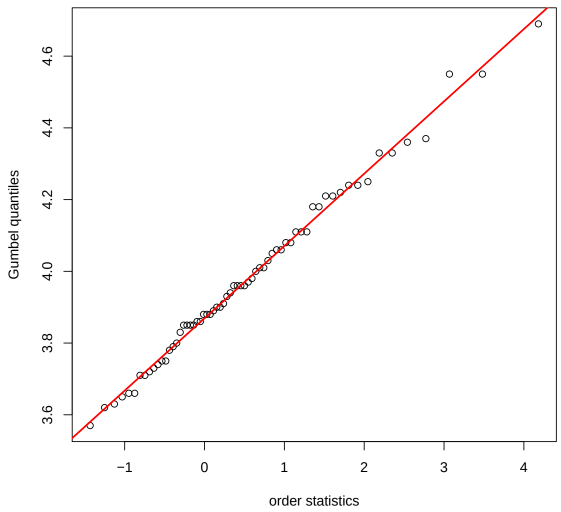

Figure 1.8: Gumbel plot for the Port Pirie dataset.

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.

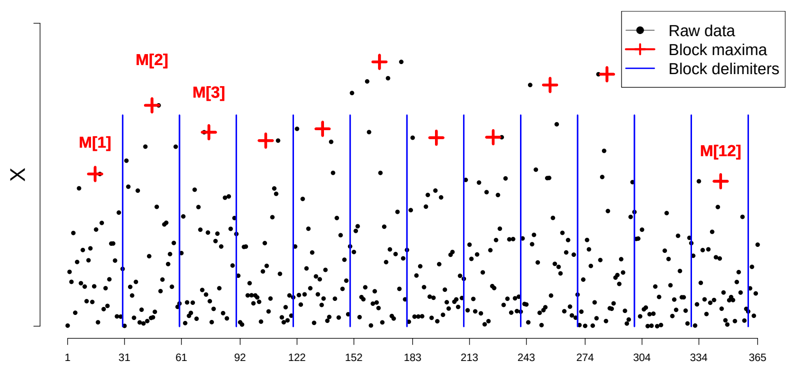

Figure 1.9: Work-flow for the block-maxima method.

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 .

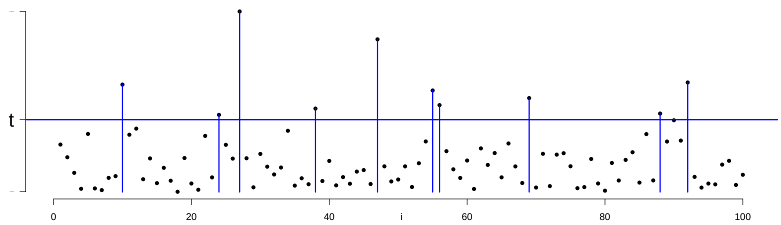

Figure 1.10: Work-flow for the POT procedure above a high threshold t: the raw data are the black dots, the 'excess' data $X_{i(r)}$ used for inference correspond to the blue lines.

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

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

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 $(X_i(r), r = 1, \ldots m)$ 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 the Poisson process

2.1 Regular variation of a real function

2.2 Karamata theorem and consequences

2.2.1 Karamata representation

2.2.2 von Mises conditions for Regular Variation

2.2.3 Fréchet domain of attraction

2.3 Regular variation and vague convergence

2.3.1 Vague convergence and limiting Poisson processes

2.4 Hill estimator

Chapter 3

Multivariate extremes

3.1 Multivariate extreme value distributions

stating the MDA conditions; MEV distributions are the max-stable distributions

3.2 Standardization

3.3 Max infinite divisibility

3.4 Angular measure

3.5 Regularly varying measures and Poisson representation

vague convergence / connection to functional regular variation/ multivariate regular variation of a random vector

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 1.2.4 (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 1.2.3, $H(x) < y$. 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 1.2.3 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 1.2.3 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 1.2.3 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 1.2.4 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 1.2.3, 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 1.2.3 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.