Extreme Value Analysis : an Introduction

Myriam Garrido, Pascal Lezaud

SOCIETE FRANÇAISE DE STATISTIQUE

Journal de la Société Française de Statistique
Vol. 154 No. 2 (2013)

Extreme Value Analysis: an Introduction

Titre: Introduction à l'analyse des valeurs extrêmes

Myriam Charras-Garrido[1] and Pascal Lezaud[2]

Abstract: We provide an overview of the probability and statistical tools underlying the extreme value theory, which aims to predict occurrence of rare events. Firstly, we explain that the asymptotic distribution of extreme values belongs, in some sense, to the family of the generalised extreme value distributions which depend on a real parameter, called the extreme value index. Secondly, we discuss statistical tail estimation methods based on estimators of the extreme value index.

Résumé : Nous donnons un aperçu des résultats probabilistes et statistiques utilisés dans la théorie des valeurs extrêmes, dont l'objectif est de prédire l'occurrence d'événements rares. Dans la première partie de l'article, nous expliquons que la distribution asymptotique des valeurs extrêmes appartient, dans un certain sens, à la famille des distributions des valeurs extrêmes généralisées qui dépendent d'un paramètre réel, appelé l'indice de valeur extrême. Dans la seconde partie, nous discutons des méthodes d'évaluation statistiques des queues basées sur l'estimation de l'indice des valeurs extrêmes

Keywords: extreme value theory, max stable distributions, extreme value index, distribution tail estimation
Mots-clés : théorie des valeurs extrêmes, lois max-stables, indice des valeurs extrêmes, estimation en queue de distribution
AMS 2000 subject classifications: 60E07, 60G70, 62G32, 62E20

1. Introduction

The consideration of the major risks in our technological society has become vital because of the economic, environmental and human impacts of industrial disasters. One of the standard approaches to studying risks uses the extreme value theory; a branch of statistics dealing with the extreme deviations from the median of probability distributions. Of course, this approach is based on the language of probability theory and thus the first question to ask is whether a probability approach applies to the studied risk. For instance, can we use probabilities in order to study the disappearance of dinosaurs? More recently, the Fukushima disaster, only 25 years after that of Chernobyl, raises the question of the appropriateness of the probability methods used. Moreover, as explained in Bouleau (1991), the extreme value theory aims to predict occurrence of rare events (e.g. earthquakes of large magnitude), outside the range of available data (e.g. earthquakes of magnitude less than 2). So, its use requires some precautions, and in Bouleau (1991) the author concludes that

> The approach attributing a precise numerical value for the probability of a rare phenomenon is suspect, unless the laws of nature governing the phenomenon are explicitly and exhaustively known [...] This does

[1] INRA, UR346, F-63122 Saint-Genes-Champanelle, France.
E-mail: myriam.charras-garrido@clermont.inra.fr
[2] ENAC, MAIAA, F-31055 Toulouse, France.
E-mail: lezaud@recherche.enac.fr

not mean that the use of probability or probability concepts should be rejected.

Nevertheless, the extreme value theory remains a well suited technique capable of predicting extreme events. Although the application of this theory in the real world always needs to be viewed with a critical eye, we suggest, in this article, an overview of the mathematical and statistical theories underlying it.

As already said before, the main objective of the extreme value theory is to know or predict the statistical probabilities of events that have never (or rarely) been observed. Firstly, the statistical analysis of extreme values has been developed in order to study flood levels. Nowadays, the domains of application include other meteorological events (such as precipitation or wind speed), industry (for example important malfunctions), finance (*e.g.* financial crises), insurance (for very large claims due to catastrophic events), environmental sciences (like concentration of ozone in the air), etc.

Formally, we consider the sample $X_1,\dots,X_n$ of n independent and identically distributed (iid) random variables with common cumulative distribution function (cdf) F. We define the ordered sample by $X_{1,n} \le X_{2,n} \le \dots \le X_{n,n} = M_n$. We are interested in two related problems. The first one consists in estimating the tail of the survivor function $\bar{F} = 1-F$: given $h_n > M_n$, we want to estimate $p = \bar{F}(h_n)$. This corresponds to the estimation of the risk to get out a zone, for example the probability to exceed the level of a dyke for a flood application. The second problem consists in estimating extreme quantiles: given $p_n < 1/n$, we want to estimate $h = \bar{F}^{-1}(p_n)$. This corresponds to estimating the limit of a critical zone, as the level of a dyke for a flood application, to be exceeded with probability p_n. Note that since we are interested to extrapolate outside the range of available observations, we have to assume that the quantile probability depends on n and that $\lim_{n\to\infty} p_n = 0$.

In both problems, the same difficulty arises: the cdf F is unknown and difficult to estimate beyond observed data. We want to get over the maximal observation M_n, that means to extrapolate outside the range of the available observations. Both parametric and non parametric usual estimation methods fail in this case. For the parametric method, models considered to give similar results in the sample range can diverge in the tail. This is illustrated in Figure 1 that presents the relative difference between quantiles from a Gaussian and a Student distribution. For the non parametric method, $1-\widehat{F}_n(x) = 0$ if $x > M_n$, where $\widehat{F}$ denotes the empirical distribution function, *i.e.* it is estimated that outside the sample range nothing is likely to be observed. As we are interested in extreme values, an intuitive solution is to use only extreme values of the sample that may contain more information than the other observations on the tail behaviour. Formally, this solution leads to a semi-parametric approach that will be detailed later.

Before starting with the description of the estimation procedures, we need to introduce the probability background which is based on the elegant theory of max-stable distribution functions, the counterpart of the (alpha) stable distributions, see Feller (1971). The stable distributions are concerned with the limit behaviour of the partial sum $S_n = X_1 + X_2 + \cdots + X_n$, as $n \to \infty$, whereas the theory of sample extremes is related to the limit behaviour of M_n. The main result is the Fisher-Tippett-Gnedenko Theorem 2.3 which claims that M_n, after proper normalisation, converges in distribution to one of three possible distributions, the Gumbel distribution, the Fréchet distribution, or the Reversed Weibull distribution. In fact, it is possible to combine these three distributions together in a single family of continuous cdfs, known as the generalized extreme value (GEV)

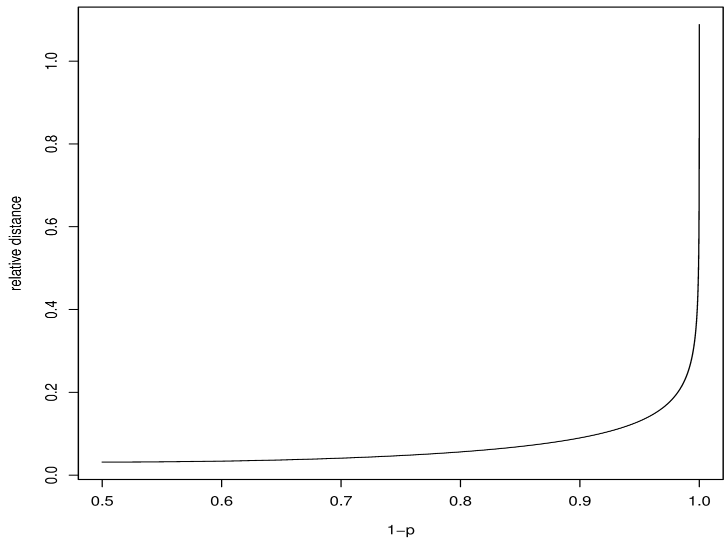

FIGURE 1. *Relative distance between quantiles of order p computed with $\mathcal{N}(0,1)$ and Student(4) models.*

distributions. A GEV is characterized by a real parameter γ, the extreme value index, as a stable distribution is it by a characteristic exponent $\alpha \in]0,2]$. Let us mention the similarity with the Gaussian Law, a stable distribution with $\alpha = 2$, and the Central Limit Theorem. Next we have to find some conditions to determine for a given cdf F the limiting distribution of M_n. The best tools suited to address that are the tail quantile function (cf. (3) for the definition) and the slowly varying functions. Finally, these results will be widened to some stationary time series.

The paper is articulated in two main Sections. In Section 2, we will set up the context in order to state the Fisher-Tippett-Gnedenko Theorem in Subsection 2.1. In this paper, we will follow closely the approach presented in Beirlant et al. (2004b), which transfers the convergence in distribution to the convergence of expectations for the class of real, bounded and continuous functions. Other recent texts include Embrechts et al. (2003) and Reiss and Thomas (1997). In Subsection 2.2, some equivalent conditions in terms of F will be given, since it is not easy to compute the tail quantile function. Finally, in Subsection 2.3 the condition about the independence between the X_i will be relaxed in order to adapt the previous result for stationary time series satisfying a weak dependence condition. The main result of this part is Theorem 2.12.

Section 3 addresses the statistical point of view. Subsection 3.1 gives asymptotic properties of extreme order statistics and related quantities and explains how they are used for this extrapolation to the distribution tail. Subsection 3.2 presents tail and quantile estimations using these extrapolations. In Subsection 3.3, different optimal control procedures on the quality of the estimates are explored, including graphical procedures, tests and confidence intervals.

2. The Probability theory of Extreme Values

Let us consider the sample $X_1,\dots,X_n$ of n iid random variables with common cdf F. We define the ordered sample by $X_{1,n} \le X_{2,n} \le \dots \le X_{n,n} = M_n$, and we are interested in the asymptotic distribution of the maxima M_n as $n \to \infty$. The distribution of M_n is easy to write down, since

$$\mathbb{P}(M_n \le x) = \mathbb{P}(X_1 \le x,\dots,X_n \le x) = F^n(x).$$

Intuitively extremes, which correspond to events with very small probability, happen near the upper end of the support of F, hence the asymptotic behaviour of M_n must be related to the right tail of the distribution near the right endpoint. We denote by $\omega(F) = \inf\{x \in \mathbb{R} : F(x) \ge 1\}$, the right endpoint of F and by $\bar{F}(x) = 1 - F(x) = \mathbb{P}(X > x)$ the survivor function of F. We obtain that for all $x < \omega(F)$, $\mathbb{P}(M_n \le x) = F^n(x) \to 0$ as $n \to \infty$, whereas for all $x \ge \omega(F)$ $\mathbb{P}(M_n \le x) = F^n(x) = 1$.

Thus M_n converges in probability to $\omega(F)$ as $n \to \infty$, and since the sequence M_n is increasing, M_n converge almost surely to $\omega(F)$. Of course, this information is not very useful, so we want to investigate the fluctuations of M_n in the similar way the Central Limit Theorem (CLT) is derived for the sum $S_n = \sum_i X_i$. More precisely, we look after conditions on F which ensure that there exists a sequence of numbers $\{b_n, n \ge 1\}$ and a sequence of positive numbers $\{a_n, n \ge 1\}$ such that for all real values x

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \le x\right) = F^n(a_n x + b_n) \to G(x) \tag{1}$$

as $n \to \infty$, where G is a non-degenerate distribution (i.e. without Dirac mass). If (1) holds, F is said to belong to the *domain of attraction* of G and we will write $F \in \mathscr{D}(G)$. The problem is twofold: (i) find all possible (non-degenerate) distributions G that can appear as a limit in (1), (ii) characterize the distributions F for which there exists sequences (a_n) and (b_n) such that (1) holds.

Introducing the threshold $u_n = u_n(x) := a_n x + b_n$ gives the more understanding interpretation of our problem, since

$$\mathbb{P}(M_n \le u_n) = F^n(u_n) = \left(1 - \frac{n\bar{F}(u_n)}{n}\right)^n.$$

Hence, we need rather conditions on the tail $\bar{F}$ to ensure that $\mathbb{P}(M_n \le u_n)$ converges to a non-trivial limit. The first result you obtain is the following:

Proposition 2.1. *For a given $\tau \in [0,\infty]$ and a sequence (u_n) of real numbers the two assertions (i) $n\bar{F}(u_n) \to \tau$, and (ii) $\mathbb{P}(M_n \le u_n) \to e^{-\tau}$ are equivalent.*

Clearly, Poisson's limit Theorem is the key behind this Proposition. Indeed, we assume for simplicity that $0 < \tau < \infty$ and we let $K_n(u_n) = \sum_{i=1}^n I_{\{X_i > u_n\}}$; it is the number of excesses over the threshold u_n in the sample $X_1,\dots,X_n$. This quantity has a binomial distribution with parameters n and $p = \bar{F}(u_n)$;

$$\mathbb{P}(K_n(u_n) = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

The Poisson's limit Theorem yields that $K_n(u_n)$ converges in law to a Poisson distribution with parameter τ if and only if $\mathbb{E}K_n(u_n) \to \tau$; this is nothing but Proposition 2.1.

Now, let us assume that $X_1 > u_n$ and consider the discrete time $T(u_n)$ such that $X_{1+T(u_n)} > u_n$ and $X_i \leq u_n$ for all $1 < i \leq T(u_n)$, i.e. $T(u_n) = \min\{i \geq 1 : X_{i+1} > u_n\}$. In order to hope for a limit distribution, we will have to normalize $T(u_n)$ by the factor n (so $T(u_n)/n \in (0,1]$); then

$$\mathbb{P}\left(n^{-1}T(u_n) > k/n\right) = \mathbb{P}(X_2 \leq u_n, \cdots, X_{k+1} \leq u_n | X_1 > u_n) = F(u_n)^{n(k/n)}.$$

Let $x > 0$, then for $k = \lfloor nx \rfloor$

$$\mathbb{P}(n^{-1}T(u_n) > x) = \mathbb{P}(n^{-1}T(u_n) > k/n) = \left(1 - \bar{F}(u_n)\right)^{n(k/n)},$$

hence if $n\bar{F}(u_n) \to \tau$ as $n \to \infty$, we have $\mathbb{P}(n^{-1}T(u_n) > x) \to e^{-\tau x}$, that means the excess times are asymptotically distributed according to an exponential law with parameter τ. The precise approach of this result requires the introduction of the point process of exceedances (N_n) defined by:

$$N_n(B) = \sum_{i=1}^{n} \delta_{i/n}(B) I_{\{X_i > u_n\}} = \sharp\{i/n \in B : X_i > u_n\},$$

where B is a Borel set on $(0,1]$ and $\delta_{i/n}(B) = 1$ if $i/n \in B$ and 0 else. Then we have the following result (see Resnick (1987)):

Proposition 2.2. *Let $(u_n)_{n\in\mathbb{N}}$ be threshold values tending to $\omega(F)$ as $n \to \infty$. Then, we have $\lim_{n\to\infty} n\bar{F}(u_n) = \tau \in (0,\infty)$, if and only if (N_n) converges in distribution to a Poisson process N with parameter τ as $n \to \infty$.*

2.1. The possible limits

Hereafter, we work under the assumption that the underlying cdf F is continuous and strictly increasing. What are the possible non-degenerate limit laws for the maxima M_n? Firstly, the limit law of a sequence of random variables is uniquely determined up to changes of location and scale (see Resnick (1987)), that means if there exists sequences (a_n) and (b_n) such that

$$\mathbb{P}\left(\frac{X_n - b_n}{a_n} \leq x\right) \to G(x),$$

then the relation

$$\mathbb{P}\left(\frac{X_n - \beta_n}{\alpha_n} \leq x\right) \to H(x),$$

holds for the sequences (β_n) and (α_n) if and only if

$$\lim_{n\to\infty} a_n/\alpha_n = \sigma \in [0,\infty), \quad \lim_{n\to\infty} (b_n - \beta_n)/\alpha_n = \mu \in \mathbb{R}.$$

In that case, $H(x) = G((x-\mu)/\sigma)$ and we say that H and G are of the *same type*. Thus, a cdf F cannot be in the domain of attraction of more than one type of cdf.

Furthermore, the question turns out to be closely related to the following property, identified by Fisher and Tippett (1928). Assume that the properly normalized and centred maxima M_n converges in distribution to G and let $n = mr$, with $m,n,r \in \mathbb{N}$. Hence, as $n \to \infty$, we have

$$F^n(a_m x + b_m) = [F^m(a_m x + b_m)]^r \to G^r(x).$$

From the previous discussion, it follows that there exist $a_r > 0$ and b_r such that $G^r(x) = G(a_r x + b_r)$; we say that the cdf G is *max-stable*.

To emphasize the role played by the tail function, we define an equivalence relation between cdfs in this way. Two cdfs F and H are called *tail-equivalent* if they have the same right end-point, i.e. if $\omega(F) = \omega(H) = x_0$, and

$$\lim_{x \uparrow x_0} \frac{1 - F(x)}{1 - H(x)} = A ,$$

for some constant A. Using the previous discussion, it can be shown (see Resnick (1987)) that $F \in \mathscr{D}(G)$ if and only if $H \in \mathscr{D}(G)$; moreover, we can take the same norming constants.

The main result of this Section is the Theorem of Fisher, Tippet and Gnedenko which characterizes the max-stable distribution functions.

Theorem 2.3 (Fisher-Tippett-Gnedenko Theorem)**.** *Let (X_n) be a sequence of iid random variables. If there exist norming constants $a_n > 0$, $b_n \in \mathbb{R}$ and some non degenerate cdf G such that $a_n^{-1}(M_n - b_n)$ converges in distribution to G, then G belongs to the type of one of the following three cdfs:*

$$\begin{aligned}
\textit{Gumbel:} \quad & G_0(x) = \exp(-e^{-x}), && x \in \mathbb{R}, \\
\textit{Fréchet:} \quad & G_{1,\alpha}(x) = \exp(-x^{-\alpha}), && x \geq 0,\ \alpha > 0, \\
\textit{Reversed Weibull:} \quad & G_{2,\alpha}(x) = \exp(-(-x)^{-\alpha}), && x \leq 0,\ \alpha < 0.
\end{aligned}$$

Figure 2 shows the convergence of $(M_n - b_n)/a_n$ to its extreme value limit in case of a uniform distribution $U[0,1]$.

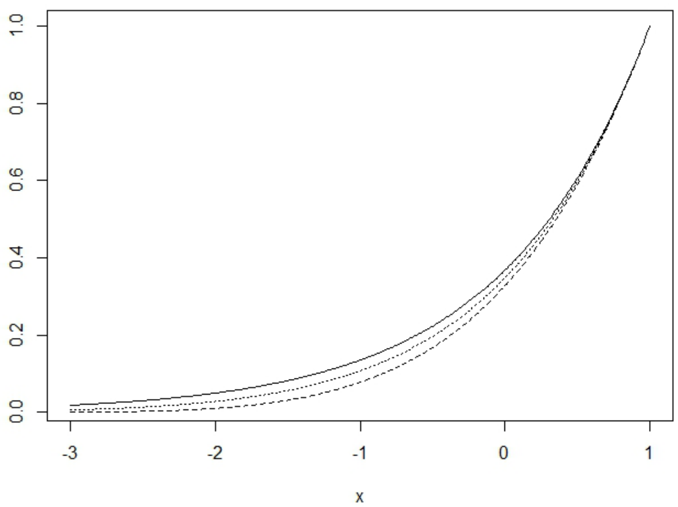

FIGURE 2. *Plot of* $\mathbb{P}((X_{n,n} - b_n)/a_n \leq x) = (1 + (x-1)/n)^n$ *for* $n = 5$ *(dashed line) and* $n = 10$ *(dotted line) and its limit* $exp(-(1-x))$ *(solid line) as* $n \to \infty$ *for a* $U[0,1]$ *distribution* F.

The three types of cdfs given in Theorem (2.3) can be thought of as members of a single family of cdfs. For that, let us introduce the new parameter $\gamma = 1/\alpha$ and the cdf

$$G_\gamma(x) = \exp(-(1+\gamma x)^{-1/\gamma}), \quad 1+\gamma x > 0. \tag{2}$$

The limiting case $\gamma \to 0$ corresponds to the Gumbel distribution. The cdf $G_\gamma(x)$ is known as the *generalized extreme value* or as the *extreme value* cdf in the von Mises form, and the parameter γ is called the *extreme value index*. Figure 3 gives examples of Gumbel, Fréchet and Reversed Weibull distributions.

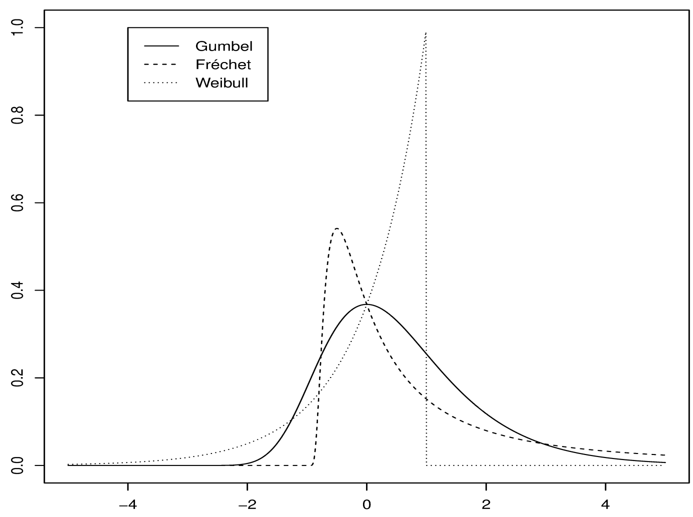

FIGURE 3. *Examples of Gumbel ($\gamma = 0$ in solid line), Fréchet (for $\gamma = 1$ in dashed line) and Reversed Weibull (for $\gamma = -1$ in dotted line) cdfs.*

Now, we will present the sketch of the Theorem's proof, following the approach of Beirlant et al. (2004b) which transfers the convergence in distribution to the convergence of expectations for the class of real, bounded and continuous functions (see Helly-Bray Theorem in Billingsley (1995)).

Let us introduce the *tail quantile function*

$$U(t) := \inf\{x \,:\, F(x) \geq 1 - 1/t\}, \tag{3}$$

which is non-decreasing over the interval $[1,\infty)$. Then, for any real, bounded and continuous functions f,

$$\begin{aligned}\mathbb{E}\left[f\left(a_n^{-1}(M_n - b_n)\right)\right] &= n\int_{-\infty}^{\infty} f\left(\frac{x-b_n}{a_n}\right) F^{n-1}(x)dF(x), \\ &= \int_0^n f\left(\frac{U(n/v)-b_n}{a_n}\right)\left(1-\frac{v}{n}\right)^{n-1} dv.\end{aligned}$$

Now observe that $(1-v/n)^{n-1} \to e^{-v}$, as $n \to \infty$, while the interval of integration extends to $[0,\infty)$. To obtain a limit for the left-hand term, we can make $a_n^{-1}(U(n/v)-b_n)$ convergent for all positive v. Considering the case $v=1$ suggests that $b_n = U(n)$ is an appropriate choice. Thereby, the natural condition to be imposed is that for some positive function a and any $u>0$

$$\lim_{x\to\infty} \frac{U(xu)-U(x)}{a(x)} = h(u) \text{ exists}, \quad (\mathscr{C})$$

with the limit function h not identically equal to zero. We have the following Proposition (Proposition 2.2 in Section 2.1 in Beirlant et al. (2004b))

Proposition 2.4. *The possible limits in* $(\mathscr{C})$ *are given by*

$$\begin{cases} h_\gamma(u) & = c\frac{u^\gamma-1}{\gamma} \qquad \gamma\neq 0 \\ h_0(u) & = c\log u \end{cases},$$

where $c \geq 0$ *and* γ *is real.*

The case $c=0$ has to be excluded since it leads to a degenerate limit, and the case $c>0$ can be reduced to the case $c=1$ by incorporating c in the function a. Hence, we replace the condition $(\mathscr{C})$ by

$$\lim_{x\to\infty} \frac{U(xu)-U(x)}{a(x)} = h_\gamma(u) \text{ exists}, \quad (\mathscr{C}_\gamma). \tag{4}$$

The above result entails that under $(\mathscr{C}_\gamma)$, we find that with $b_n = U(n)$ and $a_n = a(n)$

$$\mathbb{E}\left[f\left(a_n^{-1}(M_n-b_n)\right)\right] \to \int_0^\infty f\left(h_\gamma(1/v)\right)e^{-v}dv := \int_{-\infty}^\infty f(u)dG_\gamma(u),$$

as $n\to\infty$, with G_γ given by (2).

If we write $a(x) = x^\gamma \ell(x)$, then the limiting condition $a(xu)/a(x) \to u^\gamma$ leads to $\ell(xu)/\ell(x)\to 1$. This kind of condition refers to the notion of regular variation.

Definition 2.5. A positive measurable function ℓ on $(0,\infty)$ which satisfies

$$\lim_{x\to\infty} \frac{\ell(xu)}{\ell(x)} = 1, \qquad u>0,$$

is called *slowly varying* at ∞ (we write $\ell \in \mathscr{R}_0$).

A positive measurable function h on $(0,\infty)$ is *regularly varying* at ∞ of index $\gamma\in\mathbb{R}$ (we write $h\in\mathscr{R}_\gamma$) if

$$\lim_{x\to\infty} \frac{h(xu)}{h(x)} = x^\gamma, \qquad u>0.$$

The slowly varying functions play a fundamental role in probability theory, good references are the books of Feller (1971), Bingham et al. (1989) and Korevaar (2004). In particular, we have the following result due to Karamata (1933): $\ell\in\mathscr{R}_0$ if and only if it can be represented in the form

$$\ell(x) = c(x)\exp\left\{\int_1^x \frac{\varepsilon(u)}{u}du\right\},$$

where $c(x) \to c \in (0,\infty)$ and $\varepsilon(x) \to 0$ as $x \to \infty$. Typical examples are $\ell(x) = (\log x)^\beta$ for arbitrary β and $\ell(x) = \exp\{(\log x)^\beta\}$, where $\beta < 1$. Furthermore, if $h \in \mathscr{R}_\gamma$ with $\gamma > 0$, then $h(x) \to \infty$, while for $\gamma < 0$, $h(x) \to 0$, as $x \uparrow \infty$.

Because of their intrinsic importance, we distinguish between the three cases where $\gamma > 0$, $\gamma < 0$ and the intermediate case where $\gamma = 0$. We have the following result (see Theorem 2.3 in Section 2.6 in Beirlant et al. (2004b))

Theorem 2.6. *Let $(\mathscr{C}_\gamma)$ hold*

(i) Fréchet case: $\gamma > 0$. Here $\omega(F) = \infty$, the ratio $a(x)/U(x) \to \gamma$ as $x \to \infty$ and U is of the same regular variation as the auxiliary function a: moreover, $(\mathscr{C}_\gamma)$ is equivalent with the existence of a slowly varying function ℓ_U for which $U(x) = x^\gamma \ell_U(x)$.

(ii) Gumbel case: $\gamma = 0$. The ratios $a(x)/U(x) \to 0$ and $a(x)/\{\omega(F) - U(x)\} \to 0$ when $\omega(F)$ is finite.

(iii) Reversed Weibull case: $\gamma < 0$. Here $\omega(F)$ is finite, the ratio $a(x)/\{\omega(F) - U(x)\} \to -\gamma$ and $\{\omega(F) - U(x)\}$ is of the same regular variation as the auxiliary function a: moreover, $(\mathscr{C}_\gamma)$ is equivalent with the existence of a slowly varying function ℓ_U for which $\omega(F) - U(x) = x^\gamma \ell_U(x)$.

2.2. Equivalent conditions in terms of F

Until now, only necessary and sufficient conditions on U have been given in such a way that $F \in \mathscr{D}(G_\gamma)$. Nevertheless, it is not always easy to compute the tail quantile function of a cdf F. So, it could be preferable to express the relation between $(\mathscr{C}_\gamma)$ to the underlying distribution F.

The link between the tail of F and its tail quantile function U depends on the concept of the *de Bruyn conjugate* (see Proposition 2.5 in Section 2.9.3 in Beirlant et al. (2004b)).

Proposition 2.7. *If $\ell \in \mathscr{R}_0$, then there exists $\ell^* \in \mathscr{R}_0$, the de Bruyn conjugate of ℓ, such that*

$$\ell(x)\ell^*(x\ell(x)) \to 1, \qquad x \uparrow \infty.$$

Moreover, ℓ^* is asymptotically unique in the sense that if also $\widehat{\ell}$ is slowly varying and $\ell(x)\widehat{\ell}(x\ell(x)) \to 1$, then $\ell^* \sim \widehat{\ell}$. Furthermore, $(\ell^*)^* \sim \ell$.

This yields the full equivalence between the statements

$$1 - F(x) = x^{-1/\gamma}\ell_F(x), \quad \text{and} \quad U(x) = x^\gamma \ell_U(x),$$

where the two slowly varying functions ℓ_F and ℓ_U are linked together via the de Bruyn conjugation. So, according to Theorem 2.6 (i) and (iii) we get that

Theorem 2.8. *Referring to the notation of Theorem 2.6, we have:*

(i) Fréchet case: $\gamma > 0$. $F \in \mathscr{D}(G_\gamma)$ if and only if there exists a slowly varying function ℓ_F for which $\bar{F}(x) = x^{-1/\gamma}\ell_F(x)$. Moreover, the two slowly varying functions ℓ_U and ℓ_F are linked together via the de Bruyn conjugation.

(ii) Reversed Weibull case: $\gamma < 0$. $F \in \mathscr{D}(G_\gamma)$ if and only if there exists a slowly varying function ℓ_F for which $\bar{F}\left(\omega(F) - x^{-1}\right) \sim x^{1/\gamma}\ell_F(x), x \uparrow \infty$. Moreover, the two slowly varying functions ℓ_U and ℓ_F are linked together via the de Bruyn conjugation.

When the cdf F has a density f, it is possible to derive sufficient conditions in terms of the hazard function $r(x) = f(x)/(1-F(x))$. These conditions, which are due to von Mises (1975), are known as the von Mises conditions. In particular, the calculations involved on checking the attraction condition to G_0 are often tedious, in this respect, the von Mises criterion can be particularly useful.

Proposition 2.9 (von Mises' Theorem)**.** *Sufficient conditions on the density of a distribution for it belongs to $\mathscr{D}(G_\gamma)$ are the following:*

(i) Fréchet case: $\gamma > 0$. If $\omega(F) = \infty$ and $\lim_{x\uparrow\infty} xr(x) = 1/\gamma$, then $F \in \mathscr{D}(G_\gamma)$,

(ii) Gumbel case: $\gamma = 0$. $r(x)$ is ultimately positive in the neighbourhood of $\omega(F)$, is differentiable there and satisfies $\lim_{x\uparrow\omega(F)} \frac{dr(x)}{dx} = 0$, then $F \in \mathscr{D}(G_0)$

(iii) Reversed Weibull case: $\gamma < 0$. $\omega(F) < \infty$ and $\lim_{x\uparrow\omega(F)}(\omega(F)-x)r(x) = 1/\gamma$, then $F \in \mathscr{D}(G_\gamma)$.

Some examples of distributions which belong to the Fréchet, the Reversed Weibull and the Gumbel domain are given in respectively Table 1, Table 2 and Table 3. For more details about the norming constants a_n and b_n, see Embrechts et al. (2003). We also recall that the choice of these constants is not unique, for example we can choose α_n instead of a_n if $\lim_{n\to\infty} a_n/\alpha_n = 1$ (see the beginning of the Section 2.1).

TABLE 1. *A list of distributions in the Fréchet domain*

Distribution	$1-F(x)$	Extreme value index
Pareto	$\sim Kx^{-\alpha}$, $K,\alpha > 0$	$\frac{1}{\alpha}$
$F(m,n)$	$\int_x^\infty \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \omega^{\frac{m}{2}-1}\left(1+\frac{m}{n}\omega\right)^{-\frac{m+n}{2}} d\omega$ $x > 0;\ m,n > 0$	$\frac{2}{n}$
Fréchet	$1-\exp(-x^{-\alpha})$ $x > 0; \alpha > 0$	$\frac{1}{\alpha}$
T_n	$\int_x^\infty \frac{2\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{\omega^2}{n}\right)^{-\frac{n+1}{2}} d\omega$ $x > 0; m,n > 0$	$\frac{1}{n}$

TABLE 2. *A list of distributions in the Reversed Weibull domain*

Distribution	$1-F\left(\omega(F)-\frac{1}{x}\right)$	Extreme value index
Uniform	$\frac{1}{x}$ $x > 1$	-1
Beta(p,q)	$\int_{1-\frac{1}{x}}^1 \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} u^{p-1}(1-u)^{q-1} du$ $x > 1;\ p,q > 0$	$-\frac{1}{q}$
Reversed Weibull	$1-\exp(-x^{-\alpha})$ $x > 0;\ \alpha > 0$	$-\frac{1}{\alpha}$

Finally, we give an alternative condition for $(\mathscr{C}_\gamma)$ (Proposition 2.1 in Section 2.6 in Beirlant et al. (2004b)). It constitutes the basis for numerous statistical techniques to be discussed in Section 3.

TABLE 3. *A list of distributions in the Gumbel domain*

Distribution	$1-F(x)$
Weibull	$\exp(-\lambda x^\tau), \quad x>0;\ \lambda,\tau>0$
Exponential	$\exp(-\lambda x), \quad x>0;\ \lambda>0$
Gamma	$\frac{\lambda^m}{\Gamma(m)}\int_x^\infty u^{m-1}\exp(-\lambda u)du, \quad x>0;\ \alpha,m>0$
Logistic	$1/(1+\exp(x)), \quad x\in\mathbb{R}$
Normal	$\int_x^\infty \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x\in\mathbb{R};\ \sigma>0,\mu\in\mathbb{R}$
Log-normal	$\int_x^\infty \frac{1}{\sqrt{2\pi\sigma^2}u}\exp\left(-\frac{1}{2\sigma^2}(\log u-\mu)^2\right)du, \quad x>0;\ \mu\in\mathbb{R},\sigma>0$

Proposition 2.10. *The distribution F belongs to $\mathscr{D}(G_\gamma)$ if and only if for some auxiliary function b and $1+\gamma v>0$*

$$\frac{1-F(y+b(y)v)}{1-F(y)} \to (1+\gamma v)^{-1/\gamma}, \qquad (\mathscr{C}_\gamma^*)$$

as $y\to\omega(F)$. Then

$$\frac{b(y+vb(y))}{b(y)} \to 1+\gamma v.$$

Condition $(\mathscr{C}_\gamma^*)$ has an interesting probabilistic interpretation. Indeed, $(\mathscr{C}_\gamma^*)$ reformulates as

$$\lim_{x\uparrow\omega(F)} \mathbb{P}\left(\frac{X-v}{b(v)} > x \middle| X>v\right) = (1+\gamma v)^{-1/\gamma}.$$

Hence, the condition $(\mathscr{C}_\gamma^*)$ gives a distributional approximation for the scaled excesses over the high threshold v, and the appropriate scaling factor is $b(v)$. This motivates the following definitions.

Let X be a random variable with cdf F and right endpoint $\omega(F)$. For a fixed $u<\omega(F)$,

$$F_u(x) = \mathbb{P}(X-u\le x|X>u), \quad x\ge 0 \tag{5}$$

is the *excess* cdf of the random variable X over the threshold u. The function

$$e(u) = \mathbb{E}(X-u|X>u)$$

is called the *mean excess* function of X. The function e uniquely determines F. Indeed, whenever F is continuous, we have

$$1-F(x) = \frac{e(0)}{e(x)}\exp\left(-\int_0^x \frac{1}{e(u)}du\right), \quad x>0.$$

Define the cdf H_γ by

$$H_\gamma(x) = \begin{cases} 1-(1+\gamma x)^{-1/\gamma}, & \text{if } \gamma\neq 0,\\ 1-e^{-x}, & \text{if } \gamma=0,\end{cases}$$

where $x\ge 0$ if $\gamma\ge 0$ and $0\le x\le -1/\gamma$ if $\gamma<0$. H_γ is called a standard generalised Pareto distribution (GPD). In order to take into account a scale factor σ, we will denote

$$H_{\gamma,\sigma}(x) = \begin{cases} 1-(1+\gamma(x/\sigma))^{-1/\gamma}, & \text{if } \gamma\neq 0,\\ 1-e^{-x/\sigma}, & \text{if } \gamma=0,\end{cases} \tag{6}$$

which is defined for $x \in I\!R^+$ if $\gamma \geq 0$ and $x \in [0, -\sigma/\gamma[$ if $\gamma < 0$. Then, condition $(\mathscr{C}^*_\gamma)$ above suggests a GPD as appropriate approximation of the excess cdf F_u for large u. This result is often formulated as follows in Pickands (1975): for some function σ to be estimated from the data

$$\bar{F}_u(x) \approx \overline{H}_{\gamma,\,\sigma(u)}(x).$$

2.3. Extremes of Stationary Time Series

Beforehand, we restricted ourselves to iid random variables. However, in reality extremal events often tend to occur in clusters caused by local dependence. This requires a modification of standard methods for analysing extremes. We say that the sequence of random variable (X_i) is *strictly stationary* if for any integer $h \geq 0$ and $n \geq 1$, the distribution of the random vector $(X_{h+1}, \ldots, X_{h+n})$ does not depend on h. We seek the limiting distribution of $(M_n - b_n)/a_n$ for some choice of normalizing constants $a_n > 0$ and b_n. However, the limit distribution needs not to be the same as for the maximum $\widetilde{M}_n$ of the associated independent sequence $(\widetilde{X}_i)_{1 \leq i \leq n}$ with the same marginal distribution as (X_i). For instance, starting with an iid sequence $(Y_i, 1 \leq i \leq n+1)$ of random variables with common cdf H, we define a new sequence of random variables $(X_i, 1 \leq i \leq n)$ by $X_i = \max(Y_i, Y_{i+1})$. We see that the dependence causes large values to occur in pairs. Indeed, the random variables X_i are distributed according to the cdf $F = H^2$; so if F satisfies the equivalent conditions in Proposition 2.1, we conclude that $n\overline{H}(u_n) \to \tau/2$. Consequently, the maximum $M_n = X_{n,n}$ satisfies

$$\lim_{n \to \infty} \mathbb{P}(M_n \leq u_n) = e^{-\tau/2}.$$

To hope for the existence of a limiting distribution of $(M_n - b_n)/a_n$, the long-range dependence at extreme levels needs to be suitably restricted. To measure the long-range dependence, Leadbetter (1974) introduced a weak dependence condition known as the $D(u_n)$ condition. Before setting out this condition, let us introduce some notations as in Beirlant et al. (2004b). For a set J of positive integers, let $M(J) = \max_{i \in J} X_i$ (with $M(\emptyset) = -\infty$)). If $I = \{i_1, \ldots, i_p\}$, $J = \{j_1, \ldots, j_q\}$, we write that $I \prec J$ if and only if

$$1 \leq i_1 < \cdots < i_p < j_1 \cdots < j_q \leq n,$$

and the distance $d(I,J)$ between I and J is given by $d(I,J) = j_1 - i_p$.

Condition 2.11 ($D(u_n)$)**.** For any two disjoint subsets I, J of $\{1, \ldots, n\}$ such that $I \prec J$ and $d(I,J) \geq l_n$ we have

$$\Big| \mathbb{P}\left(\{M(I) \leq u_n\} \cap \{M(J) \leq u_n\}\right) - \mathbb{P}\left(M(I) \leq u_n\right) \mathbb{P}\left(M(J) \leq u_n\right) \Big| \leq \alpha_{n,\,l_n}$$

and $\alpha_{n,\,l_n} \to 0$ as $n \to \infty$ for some positive integer sequence l_n such that $l_n = o(n)$.

The $D(u_n)$ condition says that any two events of the form $\{M(I) \leq u_n\}$ and $\{M(J) \leq u_n\}$ become asymptotically independent as n increases when the index sets I and J are separated by a relatively short distance $l_n = o(n)$. This condition is much weaker than the standard forms of mixing condition (such as strong mixing).

Now, we partition the integers $\{1, \ldots, n\}$ into k_n disjoint blocks $I_j = \{(j-1)r_n + 1, \ldots, jr_n\}$ of size $r_n = o(n)$ with $k_n = [n/r_n]$ and, in case $k_n r_n < n$, a remainder block, $I_{k_n+1} = \{k_n r_n + 1, \ldots, n\}$.

A crucial point is that the events $\{X_i > u_n\}$ are sufficiently rare for the probability of an exceedance occurring near the ends of the blocks I_j to be negligible. Therefore, if we drop out the remainder block and the terminal sub-blocks $I'_j = \{jr_n - l_n + 1, \dots, jr_n\}$ of size l_n, we can consider only the sub-block $I^*_j = \{(j-1)r_n + 1, \dots, jr_n - l_n\}$ which are approximatively independent. Thus we get

$$\mathbb{P}(M_n \le u_n) = \mathbb{P}\left(\bigcap_{j=1}^{k_n} \{M(I^*_j) \le u_n\}\right) + o(1).$$

Finally, using condition $D(u_n)$ with $k_n \alpha_{n,l_n} \to 0$, we obtain

$$\left| \mathbb{P}\left(\bigcap_{j=1}^{k_n} \{M(I^*_j) \le u_n\}\right) - \mathbb{P}^{k_n}(\{M(I^*_1) \le u_n\}) \right| \le k_n \alpha_{n,l_n} \to 0,$$

as $n \to \infty$. Now, we observe that if thresholds u_n increase at a rate such that $\limsup n\overline{F}(u_n) > \infty$, then

$$\begin{aligned} \left|\mathbb{P}^{k_n}(M(I^*_1) \le u_n) - \mathbb{P}^{k_n}(M_{r_n} \le u_n)\right| &\le k_n \left|\mathbb{P}(M(I^*_1) \le u_n) - \mathbb{P}(M_{r_n} \le u_n)\right| \\ &= k_n \mathbb{P}(M(I^*_1) \le u_n < M(I'_1)) \\ &\le k_n l_n \mathbb{P}(X_1 > u_n) \to 0. \end{aligned}$$

So, under the $D(u_n)$, we obtain the appropriate condition

$$\mathbb{P}(M_n \le u_n) - \mathbb{P}^{k_n}(M_{r_n} \le u_n) \to 0 \tag{7}$$

from which the following fundamental results were derived, see Leadbetter (1974, 1983)

Theorem 2.12. *Let (X_n) be a stationary sequence for which there exist sequences of constants $a_n > 0$ and b_n and a non-degenerate distribution function G such that*

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \le x\right) \to G(x), \quad n \to \infty.$$

If $D(u_n)$ holds with $u_n = a_n x + b_n$ for each x such that $G(x) > 0$, then G is an extreme value distribution function.

Theorem 2.13. *If there exist sequences of constants $a_n > 0$ and b_n and a non-degenerate distribution function $\widetilde{G}$ such that*

$$\mathbb{P}\left(\frac{\widetilde{M}_n - b_n}{a_n} \le x\right) \to \widetilde{G}(x), \quad n \to \infty,$$

if $D(a_n x + b_n)$ holds for each x such that $\widetilde{G}(x) > 0$ and if $\mathbb{P}[(M_n - b_n)/a_n \le x]$ converges for some x, then we have

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \le x\right) \to G(x) := \widetilde{G}^{\theta}(x), \quad n \to \infty,$$

for some constant $\theta \in [0, 1]$.

Theorem 2.12 shows that the possible limiting distributions for maxima of stationary sequences satisfying the $D(u_n)$ condition are the same as those for maxima of independent sequences. Nevertheless Theorem 2.12 does not mean that the relations $M_n \in \mathscr{D}(G)$ and $\widetilde{M}_n \in \mathscr{D}(\widetilde{G})$ hold with $\widetilde{G} = G$. In fact, G is often of the form $\widetilde{G}^\theta$ for some $\theta \in [0,1]$ (see for instance the introductory example). This is precisely what Theorem 2.13 claims.

The constant θ is called *extremal index* and always belongs to the interval $[0,1]$. For instance, if we consider the max-autoregressive process of order one defined by the recursion

$$X_i = \max\{\alpha X_{i-1}, (1-\alpha)Z_i\}$$

where $0 \leq \alpha < 1$ and where the Z_i are independent Fréchet random variables. Then it can be proved that (cf. Beirlant et al. (2004b) Section 10.2.1)

$$\mathbb{P}(M_n \leq x) = \mathbb{P}(X_1 \leq x)[\mathbb{P}(Z_1 \leq x/(1-\alpha)] \to \exp[-(1-\alpha)/x] := G(x)$$

Whereas, $\widetilde{G}(x) = \exp(-1/x)$, so $\theta = 1-\alpha$. This example shows that any number in $(0,1]$ can be an extremal index. The case $\theta = 0$ is pathological, it entails that sample maxima M_n of the process are of smaller order than sample maxima $\widetilde{M}_n$. We refer to Leadbetter et al. (1983) and Denzel and O'Brien (1975) for some examples. Moreover, $\theta > 1$ is impossible; this follows from the following argument (see Embrechts et al. (2003) Section 8.1.1):

$$\mathbb{P}(M_n \leq u_n) = 1 - \mathbb{P}\left(\bigcup_{i=1}^{n}\{X_i > u_n\}\right) \geq 1 - n\overline{F}(u_n).$$

The left-hand side converges to $e^{-\theta\tau}$ whereas the right-hand side has limit $1-\tau$, hence $e^{-\theta\tau} \geq 1-\tau$ for all $\tau > 0$ which is possible only if $\theta \leq 1$. A case in which there is no extremal index is given in O'Brien (1974). In this article, each X_n is uniform over $[0,1]$, $X_1, X_3, \ldots$ being independent and X_{2n} a certain function of X_{2n-1} for each n. Finally, a case where $D(u_n)$ does not hold but the extremal index exists is given by the following example of Davis (1982). Let $Y_1, Y_2, \ldots$, be iid, and define the sequence

$$(X_1, X_2, X_3, \ldots) = (Y_1, Y_2, Y_2, Y_3, Y_3, \ldots) \text{ or } (Y_1, Y_1, Y_2, Y_2, \ldots)$$

each with probability $1/2$. It follows from Davis (1982) that the sequence (X_n) has extremal index $1/2$. However $D(u_n)$ does not hold: for example, if $X_1 = X_2$ then $X_n = X_{n+1}$ if n is odd and $X_n \neq X_{n+1}$ if n is even. For more details, we refer to Leadbetter (1983).

To sum up, unless θ is equal to one, the limiting distributions for the independent and stationary sequences are not the same. Moreover, if $\theta > 0$ then G is an extreme value distribution, but with different parameters than $\widetilde{G}$. Thus if

$$G(x) = \exp\left(-\left(1+\gamma\frac{x-\mu}{\sigma}\right)^{-1/\gamma}\right),$$

then we have

$$\widetilde{G}(x) = \exp\left(-\left(1+\gamma\frac{x-\widetilde{\mu}}{\widetilde{\sigma}}\right)^{-1/\gamma}\right),$$

with $\mu = \widetilde{\mu} - \widetilde{\sigma}(1-\theta^\gamma)/\gamma$ and $\sigma = \widetilde{\sigma}\theta^\gamma$ (if $\gamma = 0$, $\sigma = \widetilde{\sigma}$ and $\mu = \widetilde{\mu} + \sigma \log\theta$).

Under some regularity assumptions the limiting expected number of exceedances over u_n in a block containing at least one such exceedance is equal to $1/\theta$ (if $\theta > 0$). In fact, using the notations previously introduced, we obtain (see Beirlant et al. (2004b) Section 10.2.3)

$$\frac{1}{\theta} = \lim_{n\to\infty} \frac{r_n\bar{F}(u_n)}{\mathbb{P}(M_{r_n} > u_n)} = \mathbb{E}\left[\sum_{i=1}^{r_n} \mathbf{1}(X_i > u_n)\middle| M_{r_n} > u_n\right].$$

We can have an insight into this result with the following approach: let us assume that u_n is a threshold sequence such that $n\overline{F}(u_n) \to \tau$ and $\mathbb{P}(M_n \le u_n) \to \exp(-\theta\tau)$, then from (7) (with $k_n = \lfloor n/r_n \rfloor$) we get

$$\frac{n}{r_n}\mathbb{P}(M_{r_n} > u_n) \to \theta\tau,$$

and conclude that

$$\theta = \lim_{n\to\infty} \frac{\mathbb{P}(M_{r_n} > u_n)}{r_n\overline{F}(u_n)}.$$

Another interpretation of extremal event, due to O'Brien (1987), is that under some assumptions θ represents the limiting probability that an exceedance is followed by a run of observations below the threshold

$$\theta = \lim_{n\to\infty} \mathbb{P}\left(\max\{X_2, X_3, \ldots, X_{r_n}\} \le u_n | X_1 > u_n\right).$$

So, both interpretations identify $\theta = 1$ with exceedances occurring singly in the limit, unlike $\theta < 1$ which implies that exceedances tend to occur in clusters.

The case $\theta = 1$ can be checked by using the following sufficient condition $D'(u_n)$ introduced by Leadbetter (1974), when allied with $D(u_n)$,

Condition 2.14 ($D'(u_n)$)**.**

$$\lim_{k\to\infty} \limsup_{n\to\infty} n \sum_{j=2}^{\lfloor n/k \rfloor} P(X_1 > u_n, X_j > u_n) = 0.$$

Notice that $D'(u_n)$ implies

$$\mathbb{E}\left[\sum_{1\le i<j\le\lfloor n/k\rfloor} \mathbf{1}(X_i > u_n, X_j > u_n)\right] \le \lfloor n/k \rfloor \sum_{j=2}^{\lfloor n/k \rfloor} \mathbb{E}\left[\mathbf{1}(X_1 > u_n, X_j > u_n) \to 0\right],$$

so that, in the mean, joint exceedances of u_n by pairs (X_i, X_j) become unlikely for large n.

Verifying the conditions $D(u_n)$ and $D'(u_n)$ is, in general, tedious, except in the case of a Gaussian stationary sequence. Indeed, let $r(n) = \mathrm{cov}(X_0, X_n)$ be the auto-covariance function, then the so called Berman's condition $r(n)\log n \to \infty$ allied with $\limsup_{n\to\infty} n\overline{\Phi}(u_n) < \infty$, where Φ is the normal distribution, are sufficient to imply the both conditions $D(u_n)$ and $D'(u_n)$ (see Leadbetter et al. (1983)). Let recall that the normal distribution Φ is in the Gumbel maximum domain of attraction.

3. The Statistical point of view of Extreme Values Theory

As mentioned in the Introduction, the cdf F is unknown and difficult to estimate beyond observed data, so we need to extrapolate outside the range of the available observations. In this Section, using the properties developed in Section 2, we will introduce and discuss different procedures capable of carrying out this extrapolation.

3.1. Extrapolation to the distribution tail

Firstly, we can use the properties of the maximum M_n given in Section 2 for this extrapolation, as presented in Subsection 3.1.1. We can also base our extrapolation to the distribution tail on the excesses or peaks over a threshold as presented in Subsection 3.1.2. Both extrapolation procedures are derived from asymptotic procedures that correspond to a first order approximation of the distribution tail. Second order conditions as presented in Subsection 3.1.3 may help to improve this approximation.

3.1.1. Using maxima

Theorem 2.3 gives the asymptotic distribution of the maximum M_n. Then we use the approximation of the distribution of M_n by the generalized extreme value (GEV) cdf (2) to write

$$F(x) = P(M_n \leq x)^{1/n} \sim G_\gamma^{1/n}\left(\frac{x-b_n}{a_n}\right), \quad x \to \omega(F).$$

This gives a semi-parametric approximation of the tail of the cdf F. This approximation is illustrated in Figure 4 for a uniform distribution and different values of n, using the theoretical values of a_n, b_n and γ. Let recall that the uniform distribution is in the Reversed Weibull maximum domain of attraction, and that in this case $\gamma = -1$ (cf. Table 2).

We can equivalently approximate an extreme quantile by

$$\bar{F}^{-1}(p_n) = U(1/p_n) \sim b_n + \frac{a_n}{\gamma}\left((-\ln(1-p_n)^n)^{-\gamma} - 1\right), \quad p_n \to 0 \text{ when } n \to \infty.$$

In these two approximations, appear three quantities a_n, b_n and γ whose theoretical values are only known when the cdf F is known. In practice, these quantities are unknown. a_n corresponds to a shape parameter, b_n to a scale parameter, and γ is the extreme value index. These parameters would be estimated in Subsection 3.2 to produce semi-parametric estimations of the distribution tail. In this case, this estimation would be performed using a block maxima sample.

3.1.2. Using Peaks Over a Threshold: The POT method

Modelling block maxima is a wasteful approach to extreme value analysis if other data on extremes are available. A natural alternative is to regard observations that exceed some high threshold u, smaller than the right endpoint $\omega(F)$ of F as extreme events.

Excesses occur conditioned on the event that an observation is larger than a threshold u. They are denoted by $(Y_1, \ldots)$ and represented in Figure 5. The excess cdf F_u defined in (5) expresses

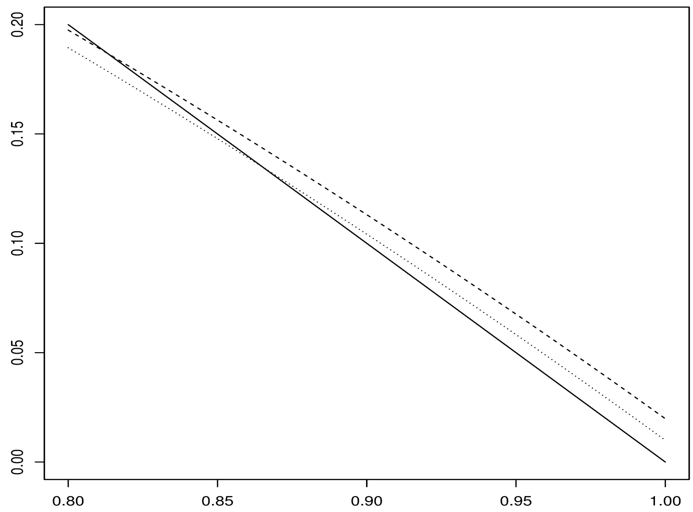

FIGURE 4. *Comparing* $\bar{F}(x)$ *(solid line) and* $1-G_{-1}^{1/n}((x-b_n)/a_n)$ *with* $a_n = n^{-1}$ *and* $b_n = 1$ *for* $n = 50$ *(dashed line) and* $n = 100$ *(dotted line) for a uniform distribution* F.

also as

$$F_u(y) = P(X \leq u+y|X \geq u) = 1 - \frac{\overline{F}(u+y)}{\overline{F}(u)}, \quad y > 0.$$

Pickands' Theorem (Pickands (1975)) implies that F_u can be approximated by a generalized Pareto distribution (GPD) function given by (6). Parameter γ is the extreme value index, and $\sigma = a_n + \gamma(u - b_n)$. In Section 3.1.1, approximating the distribution of the maximum by an EVD leads to semi-parametric estimations of the tail of the cdf F and an extreme quantile. Equivalently, approximating the distribution of the excesses over a threshold u may lead to the following semi-parametric approximations. For the tail of the cdf F, we have the semi-parametric approximation $F(x) \sim 1 - \overline{F}(u)\overline{H}_{\gamma,\sigma}(x-u)$, $x \to \omega(F)$. And for an extreme quantile, we obtain the semi-parametric approximation

$$\bar{F}^{-1}(p_n) \sim u + \frac{\sigma}{\gamma}\left[\left(\frac{p}{\overline{F}(u)}\right)^{-\gamma} - 1\right], \quad p_n \to 0 \text{ when } n \to \infty.$$

Again, we have three unknown parameters γ, σ and u to be estimated (see Subsection 3.2). Note that in practice, $u < M_n$ corresponds to a quantile inside the sample range that can be easily estimated by an observation (a quantile of the empirical distribution function). In practice, we choose $\widehat{u} = X_{n-k+1,n}$, where k is the number of excesses. However, this does not avoid the estimation of a parameter since k has to be accurately chosen. This choice is detailed in Subsection 3.3.1.

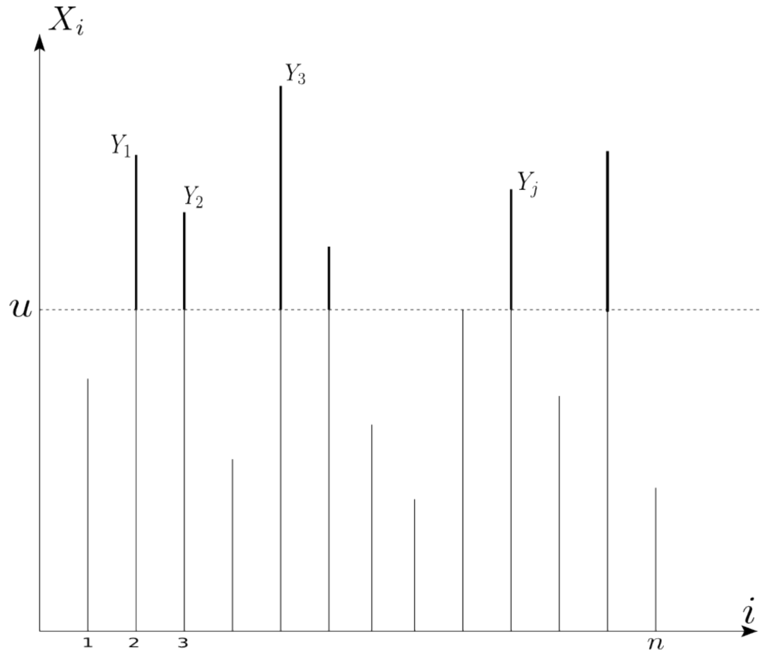

FIGURE 5. *Excesses* $(Y_1, \ldots)$ *over a threshold u.*

3.1.3. Second order conditions

The first order condition (C_γ), or equivalently (C^*_γ), relies to the convergence in distribution of the maximum M_n. We are now interested in the convergence rate for the distribution of the maximum M_n to the extreme value distribution. It corresponds to derive a remainder (see for example de Haan and Ferreira (2006) Section 2.3 or Beirlant et al. (2004a) Section 3.3) of the limit expressed by the first order condition (C_γ).

The function U (or the corresponding probability distribution) is said to satisfy the *second order condition* if for some positive function a and some positive or negative function A with $\lim_{t\to\infty} A(t) = 0$

$$\lim_{t\to\infty} \frac{\frac{U(tx)-U(t)}{a(t)} - \frac{x^\gamma - 1}{\gamma}}{A(t)} = \Psi(x),\ x > 0, \tag{8}$$

where Ψ is some function that is not a multiple of the function $(x^\gamma - 1)/\gamma$. Functions a and A are sometimes referred to as respectively first order and second order auxiliary functions. However, note that for A identically one, we obtain the first order condition (C_γ) with Ψ identically zero. The second order condition has been used to prove the asymptotic normality of different estimators and to define some of the estimators detailed in the following Section.

The following result (see de Haan and Ferreira (2006) Section 2.3) gives more insights on the functions a, A and Ψ .

Theorem 3.1. *Suppose that the second order condition (8) holds. Then there exists constants* c_1,

$c_2 \in I\!R$ and some parameter $\rho \leq 0$ such that

$$\Psi(x) = c_1 \int_1^x s^{\gamma-1} \int_1^s u^{\rho-1} du ds + c_2 \int_1^x s^{\gamma+\rho-1}. \tag{9}$$

Moreover, for $x > 0$,

$$\lim_{t\to\infty} \frac{\frac{a(tx)}{a(t)} - x^\gamma}{A(t)} = c_1 x^\gamma \frac{x^\gamma - 1}{\gamma} \tag{10}$$

and

$$\lim_{t\to\infty} \frac{A(tx)}{A(t)} = x^\rho. \tag{11}$$

Equation (11) means that function A is regularly varying with index ρ, while equation (10) gives a link between functions a and A. For $\rho \neq 0$, the limiting function Ψ can be expressed as

$$\Psi(X) = \frac{c_1}{\rho}\left(\frac{x^{\gamma+\rho}-1}{\gamma+\rho} - \frac{x^\gamma - 1}{\gamma}\right) + c_2 \frac{x^{\gamma+\rho}-1}{\gamma+\rho}.$$

If $\rho = 0$ and $\gamma \neq 0$, Ψ can be written as

$$\Psi(X) = \frac{c_1}{\gamma}\left(x^\gamma \log(x) - \frac{x^\gamma - 1}{\gamma}\right) + c_2 \frac{x^\gamma - 1}{\gamma}.$$

Finally, for $\rho = 0$ and $\gamma = 0$, Ψ can be written as

$$\Psi(X) = \frac{c_1}{2}(\log(x))^2 + c_2 \log(x).$$

There are several equivalent expressions for these quantities that can be found *e.g.* in de Haan and Ferreira (2006) Section 2.3 or Beirlant et al. (2004a) Section 3.3.

3.2. Estimation

We present the estimation procedure both for the block maxima and peak over threshold methods. Thus, in order to be general, we express the estimates from the original sample $(X_1, \ldots, X_n)$. We detail different estimates including maximum likelihood, moment, Pickands, Hill, regression and Bayesian estimates. In all cases, we focus on estimating the extreme value index γ. Other parameters can be deduced and are not detailed.

3.2.1. Maximum likelihood estimates

Maximum likelihood is usually one of the most natural estimates, largely used owing to its good properties and simple computation. However, in the case of extreme estimates, the support of the EVD (or the GPD) depends on the unknown parameter values. Then, as detailed by Smith (1985), the usual regularity conditions underlying the asymptotic properties of maximum likelihood estimators are not satisfied. In case $\gamma > -1/2$, the usual properties of consistency, asymptotic efficiency and asymptotic normality hold. But there is no analytic expression for the maximum likelihood estimates. Then, maximization of the log-likelihood may be performed by standard numerical optimization algorithms, see *e.g.* Prescott and Walden (1980, 1983), Hosking (2013) or Macleod (1989). An iterative formula is also available and presented in Castillo et al. (2004).

Moreover, remark that standard convergence properties are valuable for estimating using a sample issued from an EVD (or a GPD). Nevertheless, Fisher-Tippet-Gnedenko Theorem 2.3 (or Pickands' Theorem in Pickands (1975)) only guarantees that the maximum M_n (or the peaks over threshold) is approximately EVD (or GPD). Their accuracy in the context of extremes is more difficult to assess. However, asymptotic normality has been first proved for $\gamma > -1/2$, see *e.g.* de Haan and Ferreira (2006) Section 3.4. More recently, Zhou (2009, 2010) proves the asymptotic normality for $\gamma > -1$ and the non-consistency for $\gamma < -1$. Confidence intervals follow immediately from this approximate normality of the estimator. But these properties are limited to the range $\gamma > -1$ concerning the quantity to be estimated. In practice, the potential range of value of the parameter is unknown and thus the accuracy of the estimation cannot be assessed. Then, alternative estimates have been proposed.

3.2.2. *Moment and probability weighted moment estimates*

The probability weighted moments of a random variable X with cdf F, introduced by Greenwood et al. (1979), are the quantities $M_{p,r,s} = E\left(X^p F^r(X)(1-F(X))^s\right)$, for real p, r and s. The standard moments are obtained for $r = s = 0$. Moments and probability weighted moments do not exist for $\gamma \geq 1$. For $\gamma < 1$, we obtain for the EVD, setting $p = 1$ and $s = 0$,

$$M_{1,r,0} = \frac{1}{r+1}\left(b - \frac{a}{\gamma}[1-(r+1)^{\gamma}\Gamma(1-\gamma)]\right),$$

and for the GPD, setting $p = 1$ and $r = 0$,

$$M_{1,0,s} = \frac{\sigma}{(s+1)(s+1-\gamma)}.$$

By estimating these moments from a sample of block maxima or excesses over a threshold, we obtain estimates of the parameters a, b, σ, γ. Note that for block maxima and EVD, there is no analytic expression for the estimate of γ that has to be computed numerically. Conversely, for peaks over threshold and GPD, we have the following analytic expression given in Hosking and Wallis (1987)

$$\widehat{\gamma}_{PWM}(k) = 2 - \frac{\widehat{M}_{1,0,0}}{\widehat{M}_{1,0,0} - 2\widehat{M}_{1,0,1}} \text{ with } \widehat{M}_{1,0,s} = \frac{1}{k}\sum_{i=1}^{k}\left(1 - \frac{i}{k+1}\right)^{s} Y_{i,n}.$$

Its conceptual simplicity, its easy implementation and its good performance for small samples make this approach still very popular. However, this does not apply for strong heavy tails and in this case again, the range limitation $\gamma < 1$ concerns the quantity to be estimated. Moreover, the asymptotic normality is only valid for $\gamma \in]-1, 1/2[$, see Hosking and Wallis (1987) or de Haan and Ferreira (2006) Section 3.6.

To overcome these drawbacks, generalized probability weighted moment estimates have been proposed by Diebolt et al. (2007) for the parameters of the GDP distribution that exist for $\gamma < 2$

and are asymptotically normal for $\gamma \in]-1, 3/2[$. Diebolt et al. (2008) also proposed generalized probability weighted moment estimates for the parameters of the EVD distribution that exist for $\gamma < b+1$ and are asymptotically normal for $\gamma < 1/2 + b$, for some $b > 0$. However, since these are usual estimates, as the maximum likelihood estimate, they were not been specifically designed for extreme modelling. Conversely, the following estimates have been proposed in the context of extreme values.

3.2.3. *Hill and moment estimates*

Let $\gamma > 0$, *i.e.* we place in the Fréchet domain of attraction. From (i) in Theorem 2.8, we have

$$\lim_{t\to\infty} \overline{F}(tx)/\overline{F}(t) = x^{-1/\gamma}, \text{ for } x > 1.$$

This means that the distribution of the relative excesses X_i/t over a high threshold t conditionally on $X_i > t$ is approximately Pareto: $P(X/t > x | X > t) \simeq x^{-1/\gamma}$ for t large and $x > 1$. The likelihood equation for this Pareto distribution leads to the Hill's estimator (Hill (1975))

$$\hat{\gamma}_H(k) = \frac{1}{k}\sum_{i=1}^{k}(\log X_{n-i+1,n} - \log X_{n-k,n}).$$

We can also remark that, for $\gamma > 0$, an exponential quantile plot based on log-transformed data (also called generalized quantile plot) is ultimately linear with slope γ near the largest observations. This regression point of view also leads to the Hill estimate. This estimator can also be expressed as a simple average of scaled log-spacings

$$Z_i = i(\log X_{n-i+1,n} - \log X_{n-i,n}), \; j = 1, \dots, k. \tag{12}$$

The Hill estimate is designed from the extreme value theory and is consistent, see Mason (1982). It is also asymptotically normally distributed with mean γ and variance γ^2/k, see *e.g.* Beirlant et al. (2004a) Sections 4.2 and 4.3. Confidence intervals immediately follow from this approximate normality. But the definition of the Hill estimates and its properties are again limited to some ranges of γ, *i.e.* $\gamma > 0$. Moreover, in many instances a severe bias can appear related to the slowly varying part in the Pareto approximation. Furthermore, as many estimators based on log-transformed data, the Hill estimator is not invariant to shifts of the data. And as for all estimates of γ, for every choice of k, we obtain a different estimator, that can be very different in the case of the Hill estimator (see Figure 6).

The moment estimator has been introduced by Dekkers et al. (1989) as a direct generalization of the Hill estimator:

$$\hat{\gamma}_M(k) = \hat{\gamma}_H(k) + 1 - \frac{1}{2}\left(1 - \frac{\hat{\gamma}_H(k)}{H_k^{(2)}}\right),$$

with

$$H_k^{(2)} = \frac{1}{k}\sum_{i=1}^{k}(\log X_{n-i+1,n} - \log X_{n-k,n})^2.$$

This estimate is defined for $\gamma \in \mathbb{R}$ and is consistent. But it converges in probability to γ only for $\gamma \geq 0$, see Beirlant et al. (2004a) Section 5.2. Under appropriate conditions including the second-order condition, the asymptotic normality is established in Dekkers et al. (1989) and recalled for example in de Haan and Ferreira (2006) Section 3.5. It can be noted that the moment estimator is a biased estimator of γ.

3.2.4. *Other regression estimates*

The problem of non smoothness of the Hill estimate as a function of k can be solved with the partial least-squares regression procedure that minimizes with respect to δ and γ

$$\sum_{i=1}^{k}\left(\log X_{n-i+1,n}-\left(\delta+\gamma\log\frac{n+1}{i}\right)\right)^2.$$

This leads to the Zipf estimate, see *e.g.* Beirlant et al. (2004a) Section 4.3:

$$\hat{\gamma}_Z^+(k)=\frac{\frac{1}{k}\sum_{i=1}^{k}\left(\log\frac{k+1}{i}-\frac{1}{k}\sum_{j=1}^{k}\frac{k+1}{j}\right)\log X_{n-i+1,n}}{\frac{1}{k}\sum_{i=1}^{k}\log^2\frac{k+1}{i}-\left(\frac{1}{k}\sum_{i=1}^{k}\log\frac{k+1}{i}\right)^2}.$$

The asymptotic properties of this estimator are given *e.g.* in Csorgo and Viharos (1998).

Other refinements make use of the Hill estimate through $UH_{i,n}=X_{n-i,n}\hat{\gamma}_H(i)$ to reduce bias and to increase smoothness as a function of k. Using these UH statistics instead of the ordered statistics, the slope in the generalized quantile plot is estimated by (see Beirlant et al. (2004a) Section 5.2)

$$\hat{\gamma}^H(k)=\frac{1}{k}\sum_{i=1}^{k}(\log UH_{i,n}-\log UH_{k+1,n}),$$

and another Zipf estimate based on unconstrained least square regression (see Beirlant et al. (2002, 2004a), Section 5.2)

$$\hat{\gamma}^Z(k)=\frac{\frac{1}{k}\sum_{i=1}^{k}\left(\log\frac{k+1}{i+1}-\frac{1}{k}\sum_{j=1}^{k}\frac{k+1}{j+1}\right)\log UH_{i,n}}{\frac{1}{k}\sum_{i=1}^{k}\log^2\frac{k+1}{i+1}-\left(\frac{1}{k}\sum_{i=1}^{k}\log\frac{k+1}{i+1}\right)^2}.$$

One of the main interests of this last estimator is its smoothness as a function of k, which in some sense reduces the difficult problem of choosing k (detailed in Section 3.3.1).

Concerning the shift problems of the Hill estimate, a location-invariant variant is proposed in Fraga Alves (2002) using a secondary k-value denoted by k_0 $(< k)$

$$\hat{\gamma}^{(H)}(k_0,k)=\frac{1}{k_0}\sum_{i=1}^{k_0}\log\frac{X_{n-i+1,n}-X_{n-k,n}}{X_{n-k_0,n}-X_{n-k,n}}.$$

This estimator is consistent and asymptotically normal with mean γ and variance γ^2/k_0. Thus, its variance is not increased drastically compared to the Hill estimator.

3.2.5. Pickands Estimator

Condition $(\mathscr{C}_\gamma)$ given in equation (4) leads to

$$\frac{1}{\log 2}\log\left(\frac{U(4y)-U(2y)}{U(2y)-U(y)}\right)\sim\gamma,\ \text{for large } y.$$

Taking $y=(n+1)/k$ and replacing $U(x)$ by its empirical version $\widehat{U}_n(x)=X_{n-\lceil n/x\rceil+1,n}$ yields the Pickands estimator in Pickands (1975):

$$\hat{\gamma}_P(k)=\frac{1}{\log 2}\log\left(\frac{X_{n-\lceil k/4\rceil+1,n}-X_{n-\lceil k/2\rceil+1,n}}{X_{n-\lceil k/2\rceil+1,n}-X_{n-k+1,n}}\right).$$

The Pickands estimator is very simple but has a rather large asymptotic variance, see Dekkers and de Haan (1989). Moreover, as the Hill estimate, its is amply varying as a function of k. This is a problem as it makes crucial the choice of the fraction sample k to use for extreme estimation. Different variants have been proposed, see *e.g.* Segers (2005).

3.2.6. Bayesian estimates

An alternative to frequentist estimation, as presented until now, is to proceed to a Bayesian estimation. Some Bayesian estimates have been proposed in the literature and a review can be find *e.g.* in Coles and Powell (1996) or Coles (2001), Section 9.1. These estimators are also still under study: more recent articles present new Bayesian estimates for extreme values. For example, Stephenson and Tawn (2004) propose to estimate the parameters of the GPD distribution given the domain of attraction *i.e.* with constraints on parameter γ. Diebolt et al. (2005) propose quasi-conjugate Bayesian estimates for the parameters of the GPD distribution in the context of heavy tails *i.e.* for $\gamma>0$. do Nascimento et al. (2011) are concerned with extreme value density estimation using POT method and GPD distributions.

In our context of extreme values analysis, data are often scarce since we have to take into account only extreme data, *i.e.* a small fraction k of the original sample. One of the main reasons to use Bayesian estimation is the facility to include other sources of information through the chosen prior distribution. This can be particularly important in the context of extremes given the lack of information and the uncertainty in extrapolation. Moreover, the output of a Bayesian analysis, the posterior distribution, directly gives a measure of parameter uncertainty that allows to quantify the uncertainty in prediction. However, a Bayesian estimation implies the choice of a prior distribution that can greatly influence the result. Thus, this adds another choice to the determination of an adequate sample fraction k (detailed in Section 3.3.1).

3.2.7. Reducing bias

Classical extreme value index estimators are known to be quite sensitive to the number k of top order statistics used in the estimation. The recently developed second order reduced-bias estimators show much less sensitivity to changes in k, making the choice of k less crutial and allowing to use more data for extreme estimation. These estimators are based on the second order condition presented in Section 3.1.3. Many of them use an exponential representation including second order parameters.

Beirlant et al. (2004a)[Section 4.4], details that, for $\gamma > 0$ the scaled log-spacings Z_j, defined in equation (12), are approximately exponentially distributed with mean $\gamma + (k/j)^{\rho} b_{n,k}$. This implies that estimating γ from the log-spacing Z_j, as done with the Hill estimate, leads to a bias that is controlled by $b_{n,k}$. In the general case, it can be shown, as presented in Beirlant et al. (2004a)[Section 5.4], that the log-ratio spacings

$$j \log \frac{X_{n-j+1,n} - X_{n-k,n}}{X_{n-j,n} - X_{n-k,n}}, \quad j = 1, \ldots, k-1$$

are approximately exponentially distributed with mean $\gamma/(1-(j-(k+1))^{\gamma})$. A joint estimate of γ, $b_{n,k}$ and ρ computed from these properties, or variations of it, produces estimates of γ with reduced bias for heavy tail distributions or in the general case. Different proposals are presented in Beirlant et al. (2004a) Sections 4.5 and 5.7. In particular, Beirlant et al. (1999) perform a joint maximum likelihood for these three parameters at the same level k.

Another exponential approximation is firstly used in Feuerverger and Hall (1999). They consider that for $\gamma > 0$, the scaled log-spacings Z_j, defined in equation (12), are approximately exponentially distributed with mean $\gamma \exp(\beta (n/i)^{\rho})$ (with $\beta \neq 0$). They also proceed to the joint maximum likelihood estimation of the three unknown parameters at the same level k. Considering the same exponential approximation Gomes and Martins (2002) proposed a so-called external estimation of the second order parameter ρ, *i.e.* its estimation at a level k_1 higher than the level k used to estimate γ, together with a first order approximation for the maximum likelihood estimator of β. They then obtain quasi-maximum likelihood explicit estimators of γ and β, both computed at the same level k, and through that external estimation of ρ. This reduces the asymptotic variance of the γ estimator comparatively to the asymptotic variance of the γ estimator in Feuerverger and Hall (1999), where γ, β and ρ are estimated at the same level k. Gomes et al. (2007) build on this approach and propose an external estimation of both β and ρ by maximum likelihood both using a sample fraction k_1 larger than the sample fraction k used to estimate γ, also by maximum likelihood. This reduces the bias without increasing the asymptotic variance, which is kept at the value γ^2/k, the asymptotic variance of Hill's estimator. These estimators are thus better than the Hill estimator for all k.

3.3. Control procedures

Extreme value theory and estimation in the distribution tail are greatly influenced by several quantities. Firstly, we have to choose the tail sample fraction used for estimation. In this case, procedures for optimal choice of this tail fraction are presented in Section 3.3.1. We can also use graphical methods as presented in Section 3.3.2 to help to choose this tail fraction. Secondly, as detailed in Section 2, the tail behaviour is very different depending on the value of the parameter γ. Moreover, most of the estimates are not defined for any $\gamma \in \mathbb{R}$ but only for a smaller range of γ values. Some graphical procedures presented in Section 3.3.2 and the tests and confidence intervals presented in Section 3.3.3 can be used to assess the value of γ, the domain of attraction and the tail behaviour.

3.3.1. *Optimal choice of the tail sample fraction*

Practical application of the extreme value theory requires to select the tail sample fraction, *i.e.* the extreme values of the sample that may contain most information on the tail behaviour. Indeed, as illustrated in Figure 6 for the Hill estimator, for a small tail sample fraction k, the γ estimate strongly differs when changing the value of k. Moreover, this estimation also greatly varies when changing the sample for the same value of k, indicating a large variance of the estimate for small values of k. Conversely, for large values of k, the γ estimate presents a large bias, since the model assumption may be strongly violated, but a smaller variance. Indeed, we observe in Figure 6 that for large values of k, the γ estimates are close for the three simulated data sets.

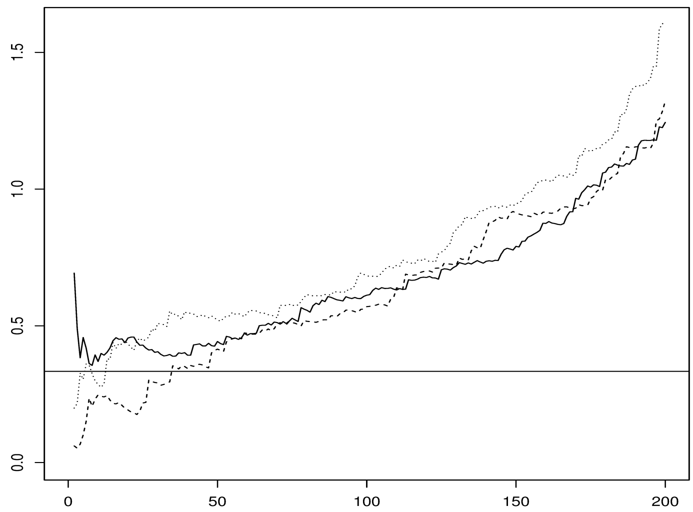

FIGURE 6. *Hill estimate of the extreme value index γ against different values of k and three data sets of size $n = 500$ simulated from a Student distribution of parameter 3 (with a true $\gamma = 1/3$).*

As noticed in Section 3.2.7, the bias of the estimates is controlled by the second order parameters, including parameter ρ. These additional parameters have been used to propose estimators with smaller bias and much less sensitive to changes in k. In the general case, the optimal k-value depends on γ and the parameters describing the second-order tail behaviour. Replacing these second order parameters by their joint estimates yields an estimate for the optimal value of k. For example, Guillou and Hall (2001) or Beirlant et al. (2004a) propose to choose the smallest value of k satisfying a given criterion which they defined.

When the asymptotic mean and variance of the estimates are known, an important alternative is to minimize the asymptotic mean squared error (AMSE) of the estimate of γ, of a tail probability or of a tail quantile, see *e.g.* Beirlant et al. (2004a). As detailed in the following Section, a mean squared error plot representing the AMSE depending on the value of k can also be useful.

3.3.2. Graphical procedures

As noticed in Section 3.3.1, we need to select the tail sample fraction, *e.g.* the number of upper extremes k, in order to apply extreme value theory for estimation purpose. Such a choice can be supported visually by a diagram. To this aim estimates of γ (see Figure 6), or other estimates, can be plotted against different values of k. For small values of k, the variance of the estimator is large and the bias is small, while for large values of k, the variance of the estimator is small and the bias is large. In between, there is a balance between the variance and the bias and we observe a plateau, where a suitable value of k may be chosen. Quite recent estimators, see *e.g.* Section 3.2.7, have the interesting property to present a relatively large plateau, that makes the choice of an appropriate value of k, less critical. To explore this balance between the variance and the bias, another option consists in plotting against the value of k, a mean square error computed, either from the true value when studying an estimate with simulated data sets or from an estimation obtained from real data sets.

As noticed above, the estimates of the extreme value index γ, and consequently the tail estimation, can be very different depending on the selected tail sample fraction. In particular, for large values of k, the model assumption may be strongly violated. It is then important to check the validity of the model. Thus, we present some graphical assessments for the validity of extreme value extrapolation. Firstly, we can use a probability plot (or PP-plot) which is a comparison of the empirical and fitted distribution functions, that may be equivalent if the model is convenient. For example, considering the ordered block maximum data $Z_{(1)} \geq \ldots \geq Z_{(m)}$, the PP-plot will consist of the points

$$\left(\frac{i}{m+1}, G_{\hat{\gamma}_m,\hat{\mu}_m,\hat{\sigma}_m}(Z_{(i)}) = \exp\left(-\left(1+\hat{\gamma}_m\frac{Z_{(i)}-\hat{\mu}_m}{\hat{\sigma}_m}\right)^{-1/\hat{\gamma}_m}\right)\right) \text{ for } i=1,\ldots,m.$$

We can also draw the PP-plot with the original sample. For example, in the POT case, we can represent the points (see Figure 7) as follows

$$\left(\frac{i}{k_n}, 1-\frac{k_n}{n}\overline{H}_{\hat{\gamma}_{k_n},\hat{\sigma}_{k_n}}(X_{n-k_n+i+1,n}-X_{n-k_n+1,n})\right) \text{ for } i=1,\ldots,k_n.$$

Secondly, we can use a quantile plot (or QQ-plot) which is a comparison between the empirical and model estimated quantiles, that may also be equivalent if the model is convenient. For example, the ordered block maximum data lead to plot the points

$$\left(G^{-1}_{\hat{\gamma}_m,\hat{\mu}_m,\hat{\sigma}_m}\left(\frac{i}{m+1}\right) = \hat{\mu}_m + \frac{\hat{\sigma}_m}{\hat{\gamma}_m}\left(1-\left(-\log\frac{i}{m+1}\right)\right)^{-\hat{\gamma}_m}\right) \text{ for } i=1,\ldots,m.$$

Again, we can also draw the QQ-plot with the original sample. For example, in the POT case, we can represent the points (see Figure 8)

$$\left(X_{n-k_n+i+1,n}, \overline{H}^{-1}_{\hat{\gamma}_{k_n},\hat{\sigma}_{k_n}}\left(\frac{k_n}{n}\left(1-\frac{i}{n}\right)\right)+X_{n-k_n+1,n}\right) \text{ for } i=1,\ldots,k_n.$$

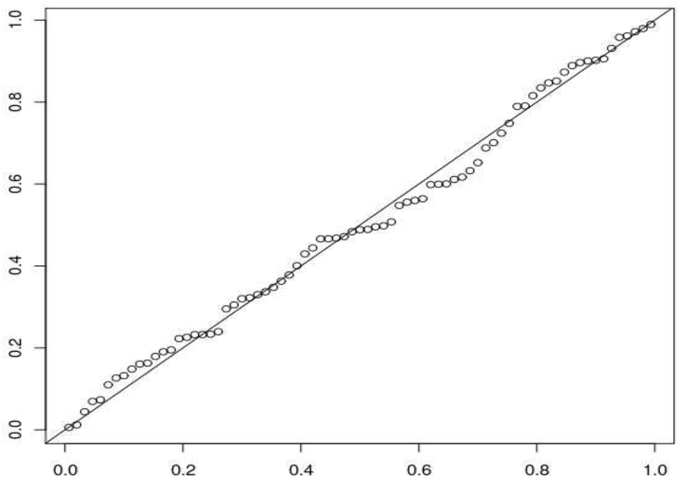

FIGURE 7. *Example of PP-plot for the original sample.*

In all the above mentioned PP or QQ-plots, the points should lie close to the unit diagonal. Substantial departures from linearity lead to suspect that either the parameter estimation method or the selected model (related for example to the chosen tail sample fraction) is inaccurate. A weakness of the PP-plot is that there is an over-smoothing, particularly in the upper and the lower tails of the distribution. Especially, the both coordinates are bounded to 1 for the largest data, *i.e.* the one of greatest interest for extreme values. Then, the probability plot provides the least information in the region of most interest. In consequence, Reiss and Thomas (2007) recommend to use the PP-plot principally to justify an hypothesis visually. They suggest to use other tools, including QQ-plot, whenever a critical attitude towards modelling is adopted. Indeed, a QQ-plot achieves a better compromise between the reduction of random data fluctuations and exhibition of special features and clues contained in the data.

There exists several other graphical tools including return level plots whose principles are analogous to those of the PP and QQ-plots. The density plot compares the density estimated by the model to a non-parametric estimation, *e.g.* histogram or kernel estimate. They are mainly of interest when the goal is to produce an estimation of the distribution tail and are not used when the goal is to estimate the extreme value index γ. Different variants of PP-plot or QQ-plot include a log-transform of the coordinates of the points. For example, the Hill and Zipf estimates (see Sections 3.2.3 and 3.2.4) are based on a generalized quantile plot. We will now focus in particular on the Gumbel plot. It is based on the fact that in the Gumbel maximal domain of attraction the excesses are exponentially distributed with parameter 1. The Gumbel plot consists in plotting the quantiles $-\log(i/k)$ against the ordered excesses $X_{n-k+i,n} - X_{n-k,n}$ as in Figure 9. In the Gumbel domain of attraction, see left panel of Figure 9, the points should lie close to the unit diagonal and the slope of the graph will give an estimate of the shape parameter, *e.g.* σ for the GPD. In the Fréchet domain of attraction an upward curvature may appear (see central panel of Figure 9), while a downward curvature may indicate a Reversed Weibull domain of attraction (see left panel of Figure 9). Outliers may also be detected using this plot. This last plot is mainly used to graphically assess the domain of attration of a data set.

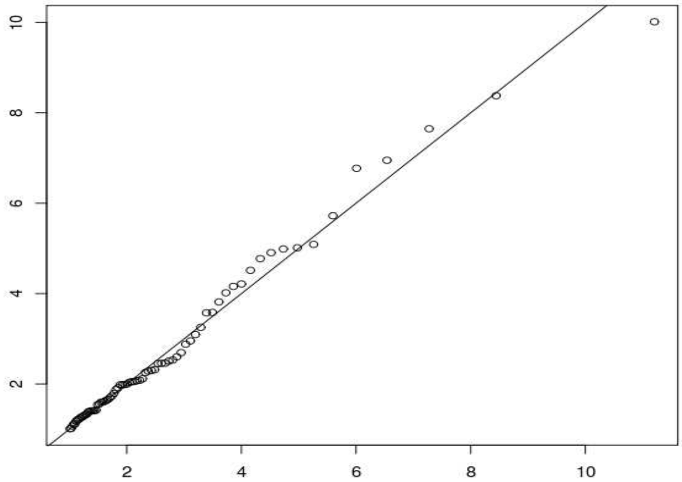

FIGURE 8. *Example of QQ-plot for the original sample.*

3.3.3. *Tests and Confidence intervals*

For many estimates, *e.g.* maximum likelihood or probability weighted moments, approximate normality is established, and confidence intervals for the GEV (or GPD) parameters follow as detailed for example in Castillo et al. (2004) Section 9.2. Direct application of the delta method yields approximate normality for the quantile corresponding estimates, and confidence intervals for the quantile can be deduced, as presented *e.g.* in Castillo et al. (2004). In other cases, the variance of the estimates may not be analytically readily available. In such cases, an estimate of the variance can be obtained using sampling methods such as jackknife and bootstrap methods presented in Efron (1979), with a preference for parametric bootstrap. In this simulation context, confidence intervals are obtained selecting empirical quantiles from the estimates (of parameters or quantiles) computed on a large number of simulated samples.

GEV has three special cases that have very different tail behaviours. For example, a distribution with a finite endpoint ($\omega(F)$) cannot be in the Fréchet domain of attraction, and conversely an unlimited distribution cannot be in the Reversed Weibull domain of attraction. Moreover, many estimates are limited to some ranges of the extreme value index γ. Model selection then focuses on deciding which one of these GEV particular case best fits the data. In particular, we wish to test $H_0 : \gamma = 0$ (Gumbel) versus $H_1 : \gamma \neq 0$ (Fréchet or Weibull), or $H_1 : \gamma < 0$ (Weilbull), or $H_1 : \gamma > 0$ (Fréchet). To this end, we can estimate γ for the GEV (or GPD) model using the maximum likelihood and perform a likelihood ratio test as detailed for example in Castillo et al. (2004) Sections 6.2 and 9.6. We can also use a confidence interval for γ, then check if it contains the value 0 and finally decide accordingly.

4. Conclusion

In this article, we presented the probability framework and the statistical analysis of extreme values. The probability framework starts with the famous Fisher-Tippett-Gnedenko Theorem 2.3

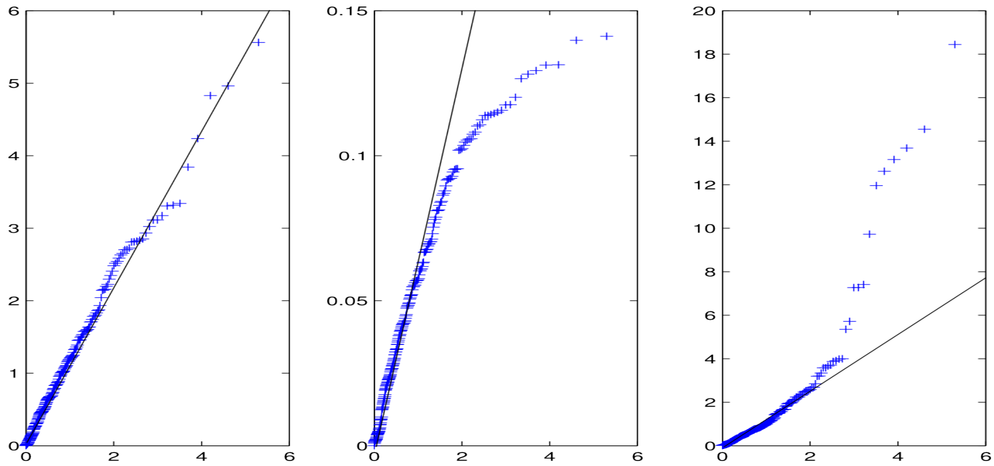

FIGURE 9. *Gumbel plot for the Gumbel (left panel), Fréchet (central panel) and Reversed Weibull (right panel) domains of attraction.*

which characterizes the three types of max-stable distributions. It remains to find the necessary and sufficient conditions to determine the domain of attraction of a specific distribution. The main tool to address this question is the notion of regular variation which plays an essential role in many limit theorems. Moreover, the Fisher-Tippett-Gnedenko Theorem restricts itself to iid random variables, so the necessity to modify the standard approach for analysing the extremes of stationary time series for instance. The results mainly obtained by M. R. Leadbetter ends the probability part of this article. We deliberately limited our presentation to the bases of the theory, so the point process approach has just been alluded and we omitted the multivariate extremes (see chapter 8 in Beirlant et al. (2004b)). The exceedances of a stochastic process, i.e. the study of $\mathbb{P}(\max_{0\leq s\leq t} X_s \geq b)$ for a stochastic process (X_t) are addressed in Aldous (1989), Berman (1992) and Falk et al. (2011). In addition, Adler (2000) and Azaïs and Wschebor (2009) are mainly dedicated to level sets and extrema of Gaussian random fields. At the heart of the Adler's approach stands the use of the Euler characteristic of level sets, whereas the book of Azaïs and Wschebor relies on Rice formula, a general tool to get results on the moments of the number of crossings of a given level by the process.

Estimating the distribution tail is a difficult problem since it implies an extrapolation. As a sign of this difficulty, numerous estimators have been proposed, some of them very recently, and none of them have made consensus. According to the application (and so the expected value of γ), the customs and practices of the applied field, the quantities of interest (estimation of γ, of the distribution tail, or of an extreme quantile) or the expected properties (low sensitivity to changes in k, low bias, low variance) different estimators can be chosen. The choice of an estimator can also be driven by practical considerations, since only some of the estimates proposed in the literature are available in classical softwares. A recent list can be find in Gilleland et al. (2013) and can help to choose estimates that are already implemented and then easy to apply. Extreme value modelling is still an active field. Topics like threshold or tail sample fraction selection, trends and change points in the tail behaviour, clustering, rates of convergence or penultimate approximations, among others, are still challenging. More details on open research topics concerning univariate extremes are given by Beirlant et al. (2012). Other challenges concern spatial extremes or non iid observations. Bases on spatial extremes can be found *e.g.* in Falk et al. (2011) or Castillo et al. (2004). Elements on extreme analysis of non iid observations are presented in Falk et al. (2011).

The statistical analysis of extreme values needs a long observation time because of the very low probability of the events considered. In many applications, such as complex systems with many interactions, collecting data is difficult, if not impossible. An alternative approach consists in the modelling of the process leading to the feared event. To achieve this, the first step requires that the considered system is formalized and only then, some estimate can be obtained by using simulation tools. Nevertheless, obtaining accurate estimates of rare event probabilities using traditional Monte Carlo techniques requires a huge amount of computing time. Many techniques for reducing the number of trials in Monte Carlo simulation have been proposed, the more promising is based on importance sampling. But to use importance sampling, we need to have a deep knowledge of the studied system and, even in such a case, importance sampling may not provide any speed-up. An alternative way to increase the relative number of visits to the rare event is to use trajectory splitting, based on the idea that there exist some well identifiable intermediate system states that are visited much more often than the target states themselves and behave as gateway states to reach the target states. For more details of the simulation of rare events, we suggest consulting Doucet et al. (2001), Bucklew (2011) and Rubino and Tuffin (2009).

Acknowledgement

We sincerely thank the Associate Editor and the referees for their careful reading, constructive comments, and relevant remarks.

References

Adler, R. J. (2000). On excursion sets, tube formulas and maxima of random fields. *Ann. Appl. Probab.*, 10(1):1–74.
Aldous, D. (1989). *Probability Approximations via the Poisson Clumping Heuristic*. Springer Verlag.
Azaïs, J. M. and Wschebor, M. (2009). *Level sets and extrema of random processes and fields*. John Wiley & Sons.
Beirlant, J., Caeiro, F., and Gomes, M. (2012). An overview and open research topics in statistics of univariate extremes. *REVSTAT - Statistical Journal*, 10(1):1–31.
Beirlant, J., Dierckx, G., Goegebeur, Y., and Matthys, G. (1999). Tail index estimation and an exponential regression model. *Extremes*, 2(2):177–200.
Beirlant, J., Dierckx, G., Guillou, A., and Staǎricaǎ, C. (2002). On exponential representations of log-spacings of extreme order statistics. *Extremes*, 5(2):157–180.
Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004a). *Statistics of Extremes*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd.
Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004b). *Statistics of Extremes: Theory and Applications*. Probability and Statistics. Wiley.
Berman, S. M. (1992). *Sojourns and extremes of stochastic processes*. Wadsworth and Brooks.
Billingsley, P. (1995). *Probability and Measure*. Wiley, 3ème edition.
Bingham, N., Goldie, C. M., and Teugels, J. L. (1989). *Regular Variation*. Cambridge University Press.
Bouleau, N. (1991). Splendeurs et misères des lois de valeurs extrêmes. *Extremes*. `http://halshs.archives-ouvertes.fr/docs/00/05/65/72/PDF/c15.pdf`.
Bucklew, J. (2011). *Introduction to rare event simulation*. Springer Series in Statistics. Springer-Verlag.
Castillo, E., Hadi, A., Balakrishnan, N., and Sarabia, J. (2004). *Extreme value and related models with applications in engineering and science*. Wiley.
Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer.

Coles, S. and Powell, E. (1996). Bayesian methods in extreme value modelling: A review and new developments. *International Statistical Review*, 64(1):119–136.

Csorgo, S. and Viharos, L. (1998). *Estimating the tail index*, pages 833–881.

Davis, R. (1982). Limit laws for the maximum and minimum of stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Werwandtl Gebiete*, **61**:31–42.

de Haan, L. and Ferreira, A. (2006). *Extreme value theory: An introduction*. Springer.

Dekkers, A. and de Haan, L. (1989). On the estimation of the extreme-value index and large quantile estimation. *Annals of Statistics*, 17(4):1795–1832.

Dekkers, A., Einmahl, J., and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *Annals of Statistics*, 17(4):1833–1855.

Denzel, G. and O'Brien, G. (1975). Limit theorems for extreme values of chain-dependent processes. *Annals of Probability*, 3:773–779.

Diebolt, J., El-Aroui, M., Garrido, M., and Girard, S. (2005). Quasi-conjugate bayes estimates for gpd parameters and application to heavy tails modelling. *Extremes*, 8(1-2):57–78.

Diebolt, J., Guillou, A., Naveau, P., and Ribereau, P. (2008). Improving probability-weighted moment methods for the generalized extreme value distribution. *REVSTAT - Statistical Journal*, 6(1):33–50.

Diebolt, J., Guillou, A., and Rached, I. (2007). Approximation of the distribution of excesses through a generalized probability-weighted moments method. *Journal of Statistical Planning and Inference*, 137(3):841–857.

do Nascimento, F., Gamerman, D., and Lopes, H. (2011). A semiparametric bayesian approach to extreme value estimation. *Statistics and Computing*, 22(2):661–675.

Doucet, A., de Freitas, N., and Gordon, N. (2001). *An Introduction to Sequential Monte Carlo Methods*. Springer Verlag.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The annals of Statistics*, 7(7):1–26.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (2003). *Modelling Extremal Events for Insurance and Finance*, volume 33 of *Applications of Mathematics*. Springer.

Falk, M., Hüsler, J., and Reiss, R. D. (2011). *Laws of small numbers: extremes and rare events*. Birkhäuser/Springer.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications II*. Wiley & Sons, 2ème edition.

Feuerverger, A. and Hall, P. (1999). Estimating a tail exponent by modelling departure from a pareto distribution. *Annals of Statistics*, 27(2):760–781.

Fisher, R. and Tippett, L. (1928). Limiting forms of the frequency distribution of the largest and smallest member of a sample. *Proc. Cambridge Phil. Soc.*, 24:180–190.

Fraga Alves, M. (2002). A location invariant hill-type estimator. *Extremes*, 4(3):199–217.

Gilleland, E., Ribatet, M., and Stephenson, A. (2013). A software review for extreme value analysis. *Extremes*, 16(1):103–119.

Gomes, M. and Martins, M. (2002). "asymptotically unbiased" estimators of the tail index based on external estimation of the second order parameter. *Extremes*, 5(1):5–31.

Gomes, M., Martins, M., and Neves, M. (2007). Improving second order reduced bias extreme value index estimation. *REVSTAT - Statistical Journal*, 5(2):177–207.

Greenwood, J., Landwehr, J., Matalas, N., and Wallis, J. (1979). Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form. *Water Ressources Research*, 15(5):1049–1054.

Guillou, A. and Hall, P. (2001). A diagnostic for selecting the threshold in extreme value analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):293–305. `http://doi.wiley.com/10.1111/1467-9868.00286`.

Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *The annals of statistics*, 3(5):1163–1174.

Hosking, J. (2013). Algorithm as 215: Maximum-likelihood estimation of the parameters of the generalized extreme-value distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 34(3):301–310.

Hosking, J. and Wallis, J. (1987). Parameter and quantile estimation for the generalized pareto distribution. *Technometrics*, 29(3):339–349.

Karamata, J. (1933). Sur un mode de croissance régulière. Théorèmes fondamentaux. *Bull. Soc. Math. France*, 61:55–62.

Korevaar, J. (2004). *Tauberian Theory: a century of developments*, volume 329 of *A Series of Comprehensive Studies in Mathematics*. Springer.

Leadbetter, M. (1974). On extreme values in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und*

Werwandtl Gebiete, **28**:289–303.

Leadbetter, M. (1983). Extremes and local dependence in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Werwandtl Gebiete*, **65**:291–306.

Leadbetter, M., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag.

Macleod, A. (1989). Algorithm as 245: A robust and reliable algorithm for the logarithm of the gamma function. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 38(2):397–402.

Mason, D. (1982). Laws of large numbers for sums of extreme values. *The Annals of Probability*, 10(3):754–764.

O'Brien, G. (1974). The maximum term of uniformly mixing stationary processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Werwandtl Gebiete*, **30**:57–63.

O'Brien, G. (1987). Extreme values for stationary and markov sequences. *Annals of Probability*, **15**:281–291.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, 3(1):119–131.

Prescott, P. and Walden, A. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika*, 67(3):723–724.

Prescott, P. and Walden, A. (1983). Maximum likelihood estimation of the parameters of the three-parameter generalized extreme-value distribution from censored samples. *Journal of Statistical Computation and Simulation*, 16(3-4):241–250.

Reiss, R. D. and Thomas, M. (1997). *Statistical Analysis of Extreme Values*. Birkhäuser.

Reiss, R.-D. and Thomas, M. (2007). *Statistical analysis of extreme values: with applications to insurance, finance, hydrology and other fields*. Birkhäuser, third edition.

Resnick, S. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer Verlag.

Rubino, G. and Tuffin, B., editors (2009). *Rare event simulation using Monte Carlo methods*. John Wiley & Sons.

Segers, J. (2005). Generalized pickands estimators for the extreme value index. *Journal of Statistical Planning and Inference*, 128(2):381–396.

Smith, R. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90.

Stephenson, A. and Tawn, J. (2004). Bayesian inference for extremes: Accounting for the three extremal types. *Extremes*, 7(4):291–307.

von Mises, R. (1975). La distribution de la plus grandes de *n* valeurs. *Reprinted(1954) in Selected Papers II: Amer. Math. Soc., Providence, R.I.*, pages 271–294.

Zhou, C. (2009). Existence and consistency of the maximum likelihood estimator for the extreme value index. *Journal of Multivariate Analysis*, 100(4):794–815.

Zhou, C. (2010). The extent of the maximum likelihood estimator for the extreme value index. *Journal of Multivariate Analysis*, 101(4):971–983.