

Chapter 3

Extreme Value Theory



From travel disruptions to natural disasters, extreme events have long captured the public’s imagination and attention. Due to their rarity and often associated calamity, they make waves in the news (Fig. 3.1) and stir discussion in the public realm: is it a freak event? Events of this sort may be shrouded in mystery for the general public, but a particular branch of probability theory, notably Extreme Value Theory (EVT), offers insight to their inherent scarcity and stark magnitude. EVT is a wonderfully rich and versatile theory which has already been adopted by a wide variety of disciplines in a plentiful way. From its humble beginnings in reliability engineering and hydrology, it has now expanded much further; it can be used to model the occurrences of records (say for example in athletic events) or quantify the probability of floods with magnitude greater than what has been observed in the past, i.e it allows us extrapolate beyond the range of available data!

In this book, we are interested in what EVT can tell us about electricity consumption of individual households. We already know a lot about what regions and countries do on average but not enough about what happens at the substation level or at least not with enough accuracy. We want to consider “worst” case scenario such as an area-wide blackout or the “very bad” case scenario such as a circuit fuse blowout or a low-voltage event. Distribution System Operators (DSO) may want to know how much electricity they will need to make available for the busiest time of day up to two weeks in advance. Local councils or policy makers may want to decide if a particular substation is equipped to meet the demands of the residents and if it needs an upgrade or maintenance. EVT can definitely help us to answer some of these questions and perhaps even more as we develop and adapt the theory and approaches further.

There are many ways to infer properties about a population based on various sample statistics. Depending on the statistic, a theory about how well it estimates

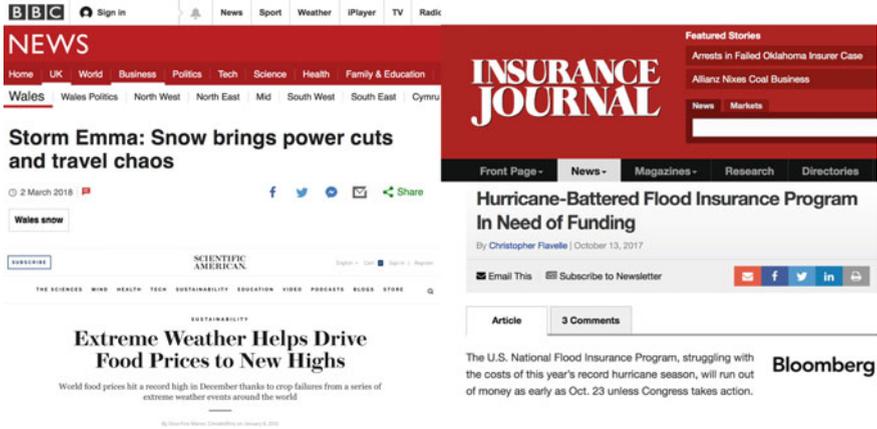


Fig. 3.1 News headlines reporting impacts of various extreme events

the parameter of interest can be developed. The sample average is one such, very common, statistic. Together with the law of large numbers and subsequently the central limit theorem (as well as others), a well known framework exists. However, this framework is lacking in various ways. Some of these limitations are linked to the assumptions of the normal distribution of finite mean and variance. But what if the underlying distribution does not have finite variance or indeed even a finite mean? Not only this, the processes involved in generating a “typical” event may be different to the processes involved in generating an extreme event, e.g. the difference between a windy day and a storm event. Or perhaps extreme events may come from different distributions: for example, internet protocols are the set of rules that set standards on data being transmitted using the internet (or another network). Faloutsos et al. [1] concluded that power-laws can help analyse the average behaviour network protocols whereas simulations from [2] showed exponential distribution in the tails.

EVT establishes the probabilistic and statistical theory of a different sample statistic: unsurprisingly, of extremes. Even though the study of EVT did not gain much traction before [3], some fundamental studies had already emerged in the earlier part of the twentieth century. While not the earliest analysis of extremes, the development of the general theory started with the work of [4]. The paper concerns itself with the distribution of the range of random samples drawn from the normal distribution. It was this work that officially introduced the concept of the distribution of the largest value. In the following years, [5, 6] evaluated the expected value and median of such distributions and the latter extended the question to non-normal distributions. The work of [7–9] gave the asymptotic distributions for the sample extremes. These works summatively give us the extreme value theorem and analogues of the central limit theory for partial or sample maxima. As this is one of the most fundamental results of EVT, we will explore it in more detail in Sect. 3.2.

In essence both the central limit theorem and the extreme value theorem are concerned with describing the same thing; an unusual event. The event may occur as a result from an accumulation of many events or from a single event which exceeds some critical threshold (or not), studied by [10]. Consider, a river whose water levels fluctuate seasonally. These fluctuations may erode its bank overtime or a single flash flooding may break the riverbank entirely. The first is a result of a cumulative effect with which the central limit theorem is concerned whereas the latter is the result of an event which exceeded what the riverbank could withstand, i.e. an extreme event, with which extreme value theory is concerned.

Analogous to measuring “normal” behaviour using a mean, median or a mode, “extreme” behaviour may also be defined in multiple ways. Each definition will give rise to specific results which may be linked to, but different from, results derived from other definitions. However, these results complement each other and allow application to different scenarios/disciplines depending on the nature of the data and the question posed. The subject of EVT is concerned with extrapolation beyond the range of the sample measurements. Hence, it is an asymptotic theory by nature, i.e. results tell us what happens when sample size tends to infinity.

The overarching goal of any asymptotic theory is two fold:

1. to provide the necessary and sufficient conditions to ensure that a specific distribution function (d.f.) occurs in the limit, which are rather qualitative conditions, and
2. to find all the possible distributions that may occur in the limit and derive a generalised form for those distributions.

The first goal is known as the *domains of attraction* problem whereas the second is known as the *limit problem*. Before we take a look at the asymptotic theory, it is valuable to review some of the statistical background and terminology that will be prevalent throughout this report.

The rest of this chapter is dedicated to an in depth exploration of extreme value theory, particularly the probabilistic foundations for the methods of inference presented in Chap. 4 and ensuing application in Chap. 5. In the following section, we will introduce and clarify some of the central terminology and nomenclature. In Sect. 3.2, we will explore the fundamental extreme value theorem which gives rise to the generalised form to the limiting distribution of sample maxima and other relevant results. We will then consider results for the Peaks over threshold (POT) approach in Sect. 3.3. In Sect. 3.4, some theory on regularly varying functions is discussed. The theory of regular variation is quite central to EVT though often it operates from the background. Familiarity with regular variation is highly recommended for the readers interested in a mathematical and theoretical treatment of EVT. Finally, in Sect. 4.4, we consider the case where condition of identically distributed rv’s can be relaxed. In Chaps. 3 and 4 as a whole, we aim to provide intuitive understanding of the theoretical results, to expound reasons for these being in great demand to many applications and to illustrate them with some examples of challenges arising in the energy sector.

3.1 Basic Definitions

As was mentioned earlier, various definitions of “extremes” give rise to different, but complementary, limiting distributions. In this section, we want to formalise what we mean by “extreme” as well as introduce the terminology that will be prevalent throughout the book.

Suppose $X_1, X_2, \dots, X_n, \dots$ is a sequence of independent random variables (r.v’s). Throughout this book and until said otherwise we shall assume that all these rv’s are generated by the same physical process and therefore it is reasonable to assume that any sample (X_1, X_2, \dots, X_n) , made out of (usually the first) n random variables in the sequence, is a random sample of independent and identically distributed (i.i.d) random variables with common distribution function (d.f.) $F(x) := \mathbb{P}(X \leq x)$, for all $x \in \mathbb{R}$. The non-stationary case, where the underlying d.f. function is allowed to vary with the time or location i of X_i has been worked through in the extreme values context by [11]. We shall refrain to delve into detail on this case within this chapter but in Chap.4 we shall refer to the statistical methodology for inference on extreme that has sprang from the domain of attraction characterisation furnished by [11]. We shall assume the underlying df is continuous with probability density function (PDF) f . We also often talk about the *support* of X ; this is the set of all values of x for which the pdf is strictly positive. The lower endpoint of the support of F or the *left endpoint* is denoted by x_F i.e.,

$$x_F := \inf\{x : F(x) > 0\}.$$

Equivalently, we define the *upper (or right) endpoint* of F by

$$x^F := \sup\{x : F(x) < 1\}. \tag{3.1}$$

which can be either finite or infinite. When each of these values exist finite, these are probabilistically speaking the smallest and largest values that can ever be observed, respectively. In reality we are not likely to observe such extremes or if we do observe extremes we do not know how close they are to the endpoints. This is only to be aggravated by what is generally known in many applications, such as financial or actuarial mathematics, that there might no such thing as a finite upper bound. The main broad aim of EVT is centred around this idea. Its purpose is to enable estimation of tale-telling features of extreme events, right up to the level of that unprecedented extreme event so unlikely to occur that we do not expect it to crop up in the data. Until they do... Now that we have established that EVT sets about teetering on the bring of the sample, aiming to extrapolation beyond the range of the observed data, the sample maximum inherently features as the relevant summary statistic we will be interested in characterising. Preferably with a fully-fledged and complete characterisation, but flexible enough to be taken up by wider applied sciences. A probabilistic result pertaining to the sample maximum such that, for its simplicity and mild assumptions could serve as a gateway for practitioners to find bespoke tail-related models that

were not previously accessible or easily interpreted; a theorem like this would be dramatically useful. It turns out that such a result, with resonance often likened to the Central Limit Theorem, already exists. This theorem, the Extreme Value or Extremal Types theorem is the centrepiece to the next section.

Before get underway to the asymptotic (or limit) theory of extremes, we need to familiarise ourselves with the following concepts of convergence for sequences:

- A sequence of random variables X_1, X_2, \dots is said to *converge in distribution* to a random variable X [notation: $X_n \xrightarrow{d} X$] if, for all $x \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} F_n(x) = F(x).$$

This is also known as weak convergence.

- The sequence *converges in probability* to X [notation: $X_n \xrightarrow{\mathbb{P}} X$] if, for any $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0.$$

- The sequence *converges almost surely* to X [notation: $X_n \xrightarrow{a.s.} X$] if,

$$P\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1.$$

Almost sure convergence is also referred to as strong convergence.

3.2 Maximum of a Random Sample

In the classical large sample (central limit) theory, we are interested in finding the limit distribution of linearly normalised partial sums, $S_n := X_1 + \dots + X_n$, where $X_1, X_2, \dots, X_n \dots$ are i.i.d. random variables. Whilst the focus here is on the aggregation or accumulation of many observable events, non of these being dominant, the Extreme Value theory shifts to edge of the sample where, for its huge magnitude and potentially catastrophic impact, one single event dominate the aggregation of the data. In the long run, the maximum might not be any less than the sum. Although this seems a bold claim, its probabilistic meaning will become more glaringly apparent later on when we introduce the concept of heavy tails.

The Central Limit theorem entails that the partial sum $S_n = X_1 + X_2 + \dots + X_n$ of linearly normalised random variables, with constants $a_n > 0$ and b_n , drawn from an iid sequence is asymptotically normal, i.e.

$$\frac{S_n - b_n}{a_n} \xrightarrow[n \rightarrow \infty]{d} N(0, 1).$$

reflecting Charlie Winsor's principle that "All actual distributions are Gaussian in the middle". The suitable choice of constant for the Central Limit theorem (CLT) to hold in its classical form are $b_n = nE(X_1)$ and $a_n = \sqrt{n\text{Var}(X_1)}$. Therefore, an important requirement is the existence of finite moment of second order, i.e. $E|X_1|^2 < \infty$. This renders the CLT inapplicable to a number of important distributions such as the Cauchy distribution. We refer to [12, Chap. 1] for further aspects on the class of sub-exponential distributions, which includes but is far from limited to the Cauchy distribution.

As it was originally developed, the Extreme Value theorem is concerned with partial maxima $X_{n,n} := \max(X_1, X_2, \dots, X_n)$ of an iid (or weakly dependent) sequence of rv's. Thus, the related *extremal limit problem* is to find out if there exist constants $a_n > 0$ and b_n such that the limit distribution to $a_n^{-1}(X_{n,n} - b_n)$ is non-degenerate. It is worth highlighting that the sample maximum itself converges *almost surely* to the upper endpoint of the underlying distribution F to the sampled data, for the df of the maximum is given by $P(X_{n,n} \leq x) = F^n(x) \rightarrow \mathbb{1}_{\{x \geq x^F\}}$, as $n \rightarrow \infty$, with $x^F \leq \infty$ and $\{X_{n,n}\}_{n \geq 1}$ recognisably a non-decreasing sequence. Here, the indicator function denoted by $\mathbb{1}_A$ is equal to 1 if A holds true and is 0 otherwise, meaning the probability distribution for the maximum distribution has mass confined to the upper endpoint. Hence, the matter now fundamentally lies in answering the following questions: (i) Is it possible to find $a_n > 0$ and b_n such that

$$\lim_{n \rightarrow \infty} P\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x), \quad (3.2)$$

for all x continuity points of a non-degenerate cdf G ? (ii) What kind of cdf G can be attained in the limit? (iii) How can G be specified in terms of F ? (iv) What are suitable choices for constants a_n and b_n question (i) is referring to? Each of these questions is addressed with great detail in the excellent book by [13].

The celebrated *Extreme Value theorem* gives us the only three possible distributions that G can be. The extreme value theorem (with contributions from [3, 8, 14]) and its counterpart for exceedances above a threshold [15] ascertain that inference about rare events can be drawn on the larger (or lower) observations in the sample. The precise statement is provided in the next theorem. Corresponding result for minima is readily accessible by using the device $X_{1,n} = -\max(-X_1, -X_2, \dots, -X_n)$.

Theorem 3.1 (Extremal Value Theorem). *Let $\{X_n\}_{n \geq 1}$ be a sequence of i.i.d. random variables with the same continuous df F . If there exist constants $a_n > 0$ and $b_n \in \mathbb{R}$, and some non-degenerate d.f. G such that Eq. (3.2) holds, then G must be only one of three types of d.f.'s:*

Fréchet

$$\Phi_\alpha(x) = \exp(-x^{-\alpha}), \quad x > 0 \quad (3.3)$$

Gumbel

$$\Lambda(x) = \exp(-e^{-x}), \quad x \in \mathbb{R}, \quad (3.4)$$

Weibull

$$\Psi_\alpha(x) = \exp(-(-x)^\alpha), x \leq 0 \quad (3.5)$$

The Fréchet, Gumbel and Weibull d.f.'s can be in turn nested in a one-parameter family of distribution through the *von Mises-Jenkinson parameterisation* [16, 17]. Notably, the Generalised Extreme Value (GEV) distribution with df given by

$$G_\gamma(x) = \begin{cases} \exp(-(1 + \gamma x)^{-1/\gamma}), & \gamma \neq 0, 1 + \gamma x > 0 \\ \exp(-e^{-x}), & \gamma = 0. \end{cases} \quad (3.6)$$

The parameter $\gamma \in \mathbb{R}$, the so-called *extreme value index* (EVI), governs the shape of the GEV distribution. In the literature, the EVI is also referred to as the shape parameter of the GEV. We will explore this notion more fully after establishing what it means for F to be in the maximum domain of attraction of a GEV distribution.

Definition 3.1. F said to be in the (maximum-) domain of attraction of G_γ [notation: $F \in \mathcal{D}(G_\gamma)$] if it is possible to redefine constants $a_n > 0$ and b_n provided in Eq. (3.2) in such a way that,

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G_\gamma(x), \quad (3.7)$$

with G_γ given by Eq. (3.6).

We now describe briefly the most salient features in a distribution belonging to each of the maximum domains of attraction:

1. If $\gamma > 0$, then the df F underlying the random sample (X_1, X_2, \dots, X_n) is said to be in the domain of attraction of a Fréchet distribution with d.f. given by $\Phi_{1/\gamma}$. This domain encloses all df's that are heavy-tailed. Qualitatively, this means that the probability of extreme events are ultimately non-negligible, and that upper endpoint x^F is infinite. Moreover, moments $\mathbb{E}(|X|^k)$ of order $k > 1/\gamma$ are not finite [18]. Formally, heavy-tailed distributions are those whose tail probability, $1 - F(x)$, is larger than that of an exponential distribution. Thus, noting that $1 - G_\gamma(x) \sim \gamma^{-1/\gamma} x^{-1/\gamma}$, as $x \rightarrow \infty$, we can then see that $G_{\gamma>0}$ is a heavy tailed distribution. Pareto, Generalised Pareto, and Inverse Gamma distributions are examples of distributions in the domain of attraction of the Fréchet distribution. Table 2.1 in [19] gives a longer list of distributions in the Fréchet domain of attraction and the corresponding expansion of the tail distribution as well as the EVI in terms of the parameters of the distribution.
2. At a first glance, the Gumbel domain of attraction would be the case of most simple inference as it is obvious that allowing to set $\gamma = 0$, renders any estimation of EVI unnecessary. However, the Gumbel domain attracts a plethora of distributions, sliding through a considerable range of tail-heavinesses, from the reversed Fréchet to the Log-Normal, and possessing either finite or infinite upper endpoint. Particular examples and a characterisation of the Gumbel domain can be found in [20].

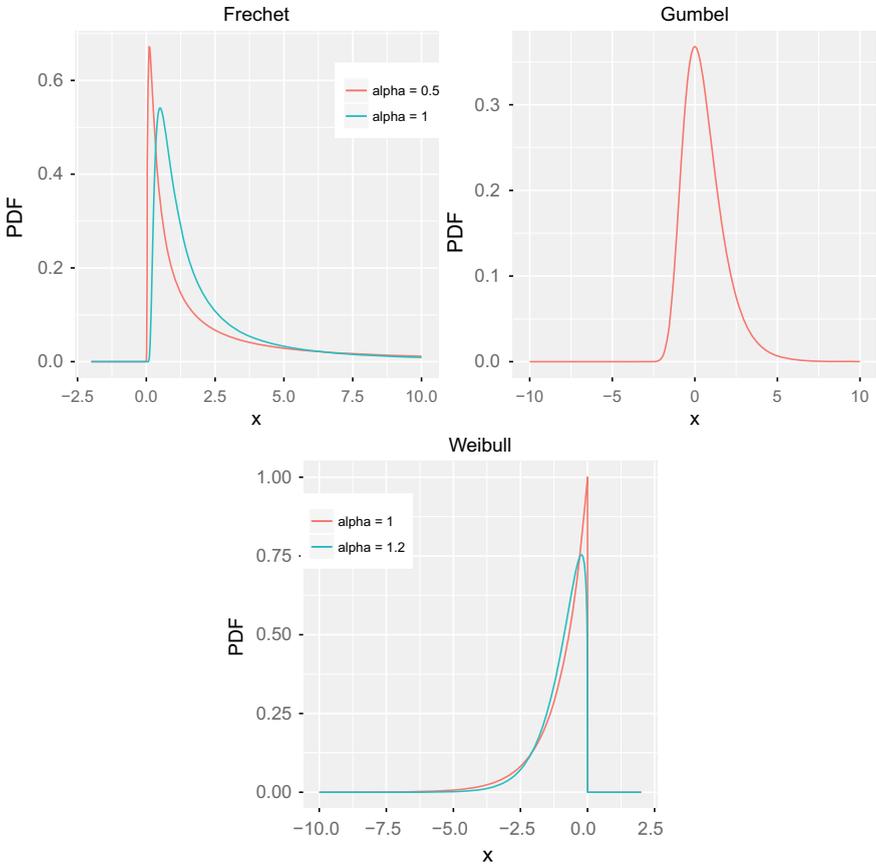


Fig. 3.2 Probability density function of several the extreme value distributions with varying $\alpha = 1/|\gamma|, \alpha > 0$

3. Lastly, for $\gamma < 0$, F is said to be in the domain of attraction of the Weibull distribution with d.f. $\Psi_{-1/\gamma}$. This domain contains short-tailed distributions with finite right endpoint. The case study presented in this report, that of electricity load of individual households, most likely belongs to the Weibull domain of attraction. This is because there is both a contractual obligation for customers to limit their electricity consumption and also physical limit to how much the fuse box can take before it short circuits.

Figure 3.2 displays the prototypical distributions to each domain of attraction, for selected values of α . We highlight the long tails, polynomial decaying tails exhibited by the chosen Fréchet distributions, contrasting with the short, upper bounded, tails ascribed to the Weibull domain.

Now that we have tackled the extremal limit problem, we can start to dissect the domains of attraction. There are enough results on this to publish multiple chapters however not all results are pertinent to our case study. Thus, we take a smaller set of results for the sake of brevity and comprehension. We present the first theorem (Theorem 3.2) which presents a set of equivalent conditions for F to belong to some domain of attraction condition. The proof for this (as for most results) are omitted but can be found in [18] (see Theorem 1.1.6). Theorem 3.2 allows us to make several important observations and connections. As noted before, we have two ways now to see that $F \in \mathcal{D}(G_\gamma)$, one in terms of the tail distribution function $1 - F$ and the other in terms of the tail quantile function U defined as

$$U(t) = \left(\frac{1}{1-F} \right)^{\leftarrow}(t) = F^{\leftarrow} \left(1 - \frac{1}{t} \right), \quad t \geq 1. \quad (3.8)$$

We note that the upper endpoint can thus be viewed as the ultimate quantile in the sense that $x^F = U(\infty) := \lim_{t \rightarrow \infty} U(t) \leq \infty$. Secondly, we have possible forms for b_n and some indication as to what a_n might be.

Theorem 3.2. *For $\gamma \in \mathbb{R}$, the following statements are equivalent:*

1. *There exists real constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that*

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G_\gamma(x) = \exp\left(- (1 + \gamma x)^{-1/\gamma}\right), \quad (3.9)$$

for all x with $1 + \gamma x > 0$.

2. *There is a positive function a such that for $x > 0$,*

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}, \quad (3.10)$$

where for $\gamma = 0$, the right hand side is interpreted at continuity, i.e. taking the limit as $\gamma \rightarrow 0$ giving rise to $\log x$. We use the notation $U \in ERV_\gamma$.

3. *There is a positive function a such that*

$$\lim_{t \rightarrow \infty} t[1 - F(a(t)x + U(t))] = (1 + \gamma x)^{-1/\gamma}, \quad (3.11)$$

for all x with $1 + \gamma x > 0$.

4. *There exists a positive function f such that*

$$\lim_{t \uparrow x^F} \frac{1 - F(t + xf(t))}{1 - F(t)} = (1 + \gamma x)^{-1/\gamma}, \quad (3.12)$$

for all x for which $1 + \gamma x > 0$.

Moreover, a_n and b_n in Eq. (3.9) holds with $a(n)$ and $U(n)$, respectively. Also, $f(t) = a(1/(1 - F(t)))$.

We see in the theorem above where theory of regular variation may come in with regard to U . Though we have deferred the discussion of regular variation to Sect. 3.4, it is useful to note that the tail quantile function U is of extended regular variation (cf. Definition 3.6) and only if F belongs to the domain of attraction of some extreme value distribution. Extreme value conditions of this quantitative nature have resonance from a rich theory and toolbox that we can borrow and apply readily, making asymptotic analysis much more elegant. We can see what the normalising constants might be and we see that we can prove that particular d.f. belongs to the domain of attraction of a generalised extreme value distribution either using F or using U . We may look at another theorem which links the index of regular variations with the extreme value index. Proceeding along these lines, the next theorem borrows terminology and insight from the theory of regular variation; though defined later (Definition 3.4), a slowly varying function ℓ satisfies $\lim_{t \rightarrow \infty} \ell(tx)/\ell(x) = 1$. Theorem 3.3 gives us the tail distribution of F , denoted by $\bar{F} = 1 - F$ in terms of a regularly varying function and the EVI. Noticing that \bar{F} is a regularly varying function means we can integrate it using Karamata's theorem (Theorem 3.13) which is useful for formulating functions f satisfying Eq. (3.12).

Theorem 3.3. *Let ℓ be some slowly varying function and $\bar{F}(x) := 1 - F(x)$ denote the survival function, where F is the d.f. association with the random variable X . Let x^F denote the upper endpoint of the df F .*

1. F is the Fréchet domain of attraction, i.e. $F \in \mathcal{D}(G_\gamma)$ for $\gamma > 0$, if and only if

$$\bar{F}(x) = x^{-1/\gamma} \ell(x) \iff \lim_{t \rightarrow \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-1/\gamma},$$

for all $x > 0$.

2. F is in the Weibull domain of attraction, i.e. $F \in \mathcal{D}(G_\gamma)$ for $\gamma < 0$, if and only if

$$\bar{F}(x^F - x^{-1}) = x^{-1/\gamma} \ell(x) \iff \lim_{t \downarrow 0} \frac{1 - F(x^F - tx)}{1 - F(x^F - t)} = x^{-1/\gamma},$$

for all $x > 0$.

3. F is in the domain of attraction of the Gumbel distribution, i.e. $\gamma = 0$ with $x^F \leq \infty$

$$\lim_{t \uparrow x^F} \frac{1 - F(t + xf(t))}{1 - F(t)} = e^{-x},$$

for all $x \in \mathbb{R}$, with f a suitable positive auxiliary function. If the above equation holds for some f , then it also holds with $f(t) := \left(\int_t^{x^F} (1 - F(s)) ds \right) / (1 - F(t))$ where the numerator of the integral exists finite for $t < x^F$.

Theorem 3.4. $F \in \mathcal{D}(G_\gamma)$ if and only if for some positive function f ,

$$\lim_{t \uparrow x^F} \frac{1 - F(t + xf(t))}{1 - F(t)} = (1 + \gamma x)^{-1/\gamma} \quad (3.13)$$

for all x with $1 + \gamma x > 0$. If the above holds for some f , then it also holds with

$$f(t) = \begin{cases} \gamma t, & \gamma > 0 \\ -\gamma(x^F - t), & \gamma < 0 \\ \frac{\int_t^{x^F} 1-F(x)dx}{1-F(t)}, & \gamma = 0. \end{cases}$$

Moreover, any f that satisfies Eq. (3.13) also satisfies

$$\begin{cases} \lim_{t \rightarrow \infty} \frac{f(t)}{t} = \gamma, & \gamma > 0, \\ \lim_{t \uparrow x^F} \frac{f(t)}{x^F - t} = -\gamma, & \gamma < 0, \\ f(t) \sim f_1(t), & \gamma = 0, \end{cases}$$

where $f_1(t)$ is some function for which $f_1'(t) \rightarrow 0$ as $t \uparrow x^F$.

In this section, we have thus far mentioned a suitable function f which plays various roles however it should not be interpreted as probability density function of F , unless explicitly stated as such. Theorem 3.4 gives us alternative forms for f and its limit relations.

Theorem 3.5. *If $F \in \mathcal{D}(G_\gamma)$, then*

1. for $\gamma > 0$:

$$\lim_{n \rightarrow \infty} F^n(a_n x) = \exp(-x^{-1/\gamma}) = \Phi_{-1/\gamma}(x)$$

holds for $x > 0$ with $a_n := U(n)$;

2. for $\gamma < 0$:

$$\lim_{n \rightarrow \infty} F^n(a_n x + x^F) = \exp(-(-x)^{-1/\gamma}) = \Psi_{-1/\gamma}(x)$$

for $x < 0$ holds with $a_n := x^F - U(n)$;

3. for $\gamma = 0$:

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = \exp(-e^{-x}) = \Lambda(x)$$

holds for all x with $a_n := f(U(n))$ and $b_n := U(n)$ where f is as defined in Theorem 3.3 (3).

We briefly consider maxima that have not been normalised and have been normalised for various sample sizes $n = 1, 7, 30, 365, 3650$ (Fig. 3.3). The left plot of Fig. 3.3 shows the distribution of the maxima where the lines represent, from left to right, each of the sample sizes. The right hand side of the same plot shows how quickly, the normalised maxima go to the Gumbel distribution; the exponential distribution belongs to the Gumbel domain of attraction. The appropriate normalising constants for F standard exponential are $a_n = 1$ and $b_n = \log(n)$. Deriving this is left as an exercise.

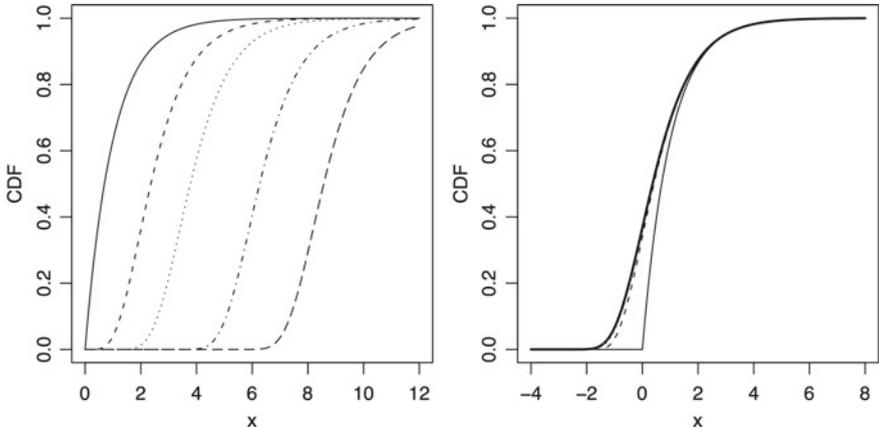


Fig. 3.3 Distributions of maxima (left) and normalised maxima (right) of $n = 1, 7, 30, 365, 3650$ standard exponential random variables, with the Gumbel d.f. (solid heavy line)

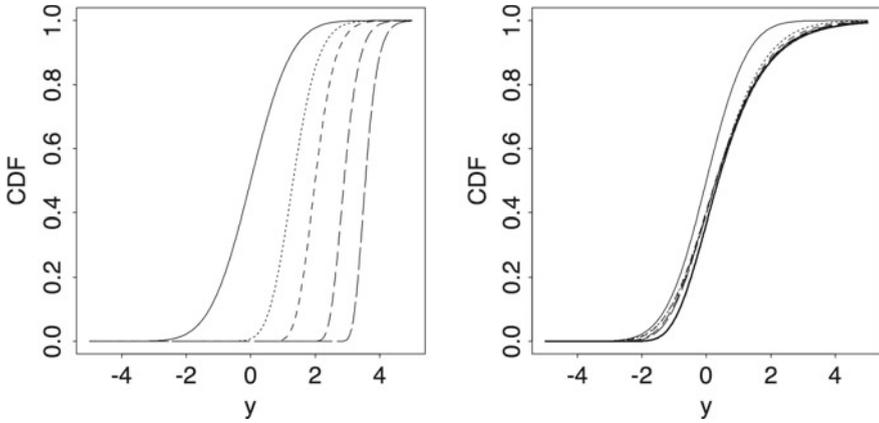


Fig. 3.4 Distributions of maxima (left) and normalised maxima (right) of $n = 1, 7, 30, 365, 3650$ standard normal random variables, with the Gumbel d.f. (solid heavy line)

Doing the same thing except now with F standard normal distribution, Fig. 3.4 shows that the convergence is slow. In this case, $a_n = (\log n)^{-1/2}$ and $b_n = (\log n)^{1/2} - 0.5 (\log n)^{-1/2} (\log \log n + \log 4\pi)$. As before, F standard normal belongs to the Gumbel domain of attraction.

Theorem 3.5 gives which sequences a_n and b_n should be used to normalise maxima in order to ensure that F is in the maximum domain of attraction of a specific G_γ . Note that the values of a_n and b_n changes with the sample size, n . If γ were known before hand, knowing the true value of the normalising constants may help with simulations or numerical experiments. However in practice, we do not know γ and it must be estimated. Thus, we can use the von Mises condition give us a work around.

Theorem 3.6 (von Mises Condition). *Let $r(x)$ be the hazard function defined by*

$$r(x) = \frac{f(x)}{1 - F(x)} \quad (3.14)$$

where $f(x)$ is the probability density function and F is the corresponding d.f.

1. *If $x^F = \infty$ and $\lim_{x \uparrow \infty} x r(x) = 1/\gamma > 0$, then $F \in \mathcal{D}(\Phi_{1/\gamma})$.*
2. *If $x^F < \infty$ and $\lim_{x \uparrow x^F} (x^F - x) r(x) = -1/\gamma > 0$, then $F \in \mathcal{D}(\Psi_{-1/\gamma})$.*
3. *If $r(x)$ is ultimately positive in the negative neighbourhood of x^F , is differentiable there and satisfies $\lim_{x \uparrow x^F} \frac{d}{dx} r(x) = 0$, then $F \in \mathcal{D}(\Lambda)$.*

The von Mises conditions given in Theorem 3.6 is particularly useful when one is interested in conducting simulations. We may sample from a known distribution F which readily gives us the probability density function, f . Thus, without knowledge of the appropriate normalising constants, the von Mises conditions allow us to find the domain of attraction of F .

We have discussed the asymptotic theory of the maximum from a sample. Earlier we mentioned that in practice, we divide the data into blocks of length n and take the maximum from each block and conduct inference on them. The results we have discussed in this section, tell us what happens as the block size becomes infinitely large. The approach of sampling maxima from blocks is, unsurprisingly known as the *Block Maxima* approach. As [21] pointed out, the block maxima model offers many practical advantages (over the Peaks Over Threshold, Sect. 3.3). The block maxima method is the appropriate statistical model when only the most extreme data are available; for example, historically temperature data was recorded in daily minimum, average and maximum. In cases where the time series may have periodicity, we can remove some dependence by dividing the blocks in such a way that dependence may exist within the block but not between blocks. We will now consider an alternative but equivalent method.

3.3 Exceedances and Order Statistics

When conducting inference on the tail of a distribution, it is wasteful to consider only the most extreme observation. We may be able to glean valuable information by utilising more than just the maximum. For such cases, we may study either exceedances over a (high) threshold (Sect. 3.3.1) or we may consider order statistics (Sect. 3.3.2). In each case, we get different limiting distributions. In what follows we will discuss what the limiting distributions are in each case and how they relate to the extreme value distributions and the results from Sect. 3.2.

3.3.1 Exceedances

In this instance, the idea is that statistical inference is based on observations that exceed a high threshold, u , i.e., either on X_i or on $X_i - u$ provided that $X_i > u$ for $i \leq n$. The exact conditions under which the POT model holds is justified by second order refinements (cf. Sect. 3.4) whereas typically it has been taken for granted that the block maxima method follows the extreme value distribution very well. We saw this from the discussion from Fig. 3.4. For large sample sizes, the performance of the block maxima method and the peaks over threshold method is comparable. However, when the sample is not large, there may be some estimations where the Peaks over threshold (POT) model is more efficient [21].

Since we have familiarised ourselves with the convergence of partial maxima, we now do the same for exceedance. We will show that appropriately normalised exceedances converge to the Generalised Pareto distribution. This is the POT model. The work on exceedances was independently initiated by [15, 22]. As before, we will start with definitions and then proceed to establishing the Generalised Pareto as the limiting distribution.

Definition 3.2. Let X be a random variable with d.f. F and right endpoint x^F . Suppose we have the threshold $u < x^F$. Then the d.f., F_u , of the random variable X over the threshold u is defined to be

$$F_u(x) = \mathbb{P}(X - u \leq x | X > u) = \frac{F(x + u) - F(u)}{1 - F(u)}, \quad 0 \leq x < x^F - u. \quad (3.15)$$

Definition 3.3. The *Generalised Pareto* distribution is defined as

$$W_\gamma(x) = \begin{cases} 1 - (1 + \gamma x)^{-1/\gamma}, & \gamma \neq 0, \\ 1 - e^{-x}, & \gamma = 0 \end{cases} \quad (3.16)$$

where

$$\begin{cases} x \geq 0, & \gamma \geq 0, \\ 0 \leq x \leq -1/\gamma, & \gamma < 0. \end{cases}$$

Note that, as for G_γ , the Generalised Pareto distribution also has scale and location parameters: $W_{\gamma; \nu, \beta}(x) := W_\gamma((x - \nu)/\beta)$. Further, for x with $1 + \gamma x > 0$,

$$1 + \log G_\gamma(x) = W_\gamma(x).$$

Now that we have looked at the d.f. of exceedance and defined the Generalised Pareto distribution, the latter can be established as the limiting distribution of exceedances.

Theorem 3.7 (Balkema-de Haan-Pickands Theorem). *One can find a positive, measurable function β such that*

$$\lim_{u \uparrow x^F} \sup_{0 \leq x \leq x^F - u} |F_u(x) - GP_{\gamma; 0, \beta(u)}(x)| = 0 \quad (3.17)$$

if and only if $F \in \mathcal{D}(G_\gamma)$.

Not only does the Balkema-de Haan-Pickands theorem allow us to use more than the maximum, it also connects the d.f. of the random variables to that of exceedances over a threshold; from knowing the limiting distribution of F_u , we also know about the domain of attraction of F and vice versa. The shape parameter in both cases is the same and thus their interpretation is the same as before, i.e., γ describes the tail-heaviness of F if Eq.(3.17) is satisfied. Holmes and Moriarty [23] used the Generalised Pareto distribution to model particular storms of interest for applications in wind engineering and [24] used the POT method to analyse financial risk.

3.3.2 Asymptotic Distribution of Certain Order Statistics

In the previous section, we talked about how the POT approach can use data more efficiently. The efficiency relies on choosing the threshold appropriately. If the threshold is too low, then the exceedances are no longer from the tail and the bias is dominant. On the other hand, if the threshold is too high, then very few data points exceed it and the variance is high and confidence in the results is low. We can consider this idea of balancing the effects of bias and variance by considering a certain kind of order statistics. This is the topic of this section.

Suppose $X_1, X_2, \dots, X_n, \dots$ are i.i.d. random variables with common d.f. F . If we take a finite sample X_1, \dots, X_n and order it from minimum to maximum, then we get the n th order statistics:

$$X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}. \quad (3.18)$$

Furthermore, we can define the k th upper order statistic, $X_{n-k,n}$, to be the k th largest value in the finite sample; the n th upper order statistic, i.e. $k = n$ is the maximum and the first upper order statistic, i.e. $k = 1$, is the minimum. Depending on k and its relation to n , the k th upper order statistic can be classified in at least three different ways which leads to different asymptotic distributions. Arnold et al. [13] classified $X_{n-k,n}$ to be one the following three order statistics:

1. *Central Order Statistics:* $X_{n-k,n}$ is considered to be a central order statistic if $k = [np] + 1$ where $0 < p < 1$ and $[\cdot]$ characterises the function which is the smallest integer larger than the argument.
2. *Extreme Order Statistics:* $X_{n-k,n}$ is an extreme order statistic when either k or $n - k$ is fixed and $n \rightarrow \infty$.

3. *Intermediate Order Statistics*: $X_{n-k,n}$ is an intermediate sequence if both k and $n - k$ approach infinity but $k/n \rightarrow 0$ or 1. In this book we present results for $k/n \rightarrow 0$ and we also assume that k varies with n i.e. $k = k(n)$.

Note that the conditions which ensure that $X_{n-k,n}$ is an intermediate order statistic has similar notions of balancing bias and variance; insisting that $k/n \rightarrow 0$ means that all data points larger than $X_{n-k,n}$ is a small part of the entire population and ensures analyses pertains to the tail of the distribution. However, for asymptotic results to hold, some of which we have seen in Sects. 3.2 and 3.3.1 and will see in this section, we require a large enough sample, i.e. k should go to infinity. As such identifying k appropriately is a crucial and a non-trivial part of extreme value analysis and also proves useful for the POT model as it allows us to chose u to be value which corresponds to the intermediate order statistics, $X_{n-k,n}$.

Since we use intermediate order statistics in our case study on electricity load in Chap. 5, it is of more immediate interest to us but for the sake of completeness and intuitive understanding we discuss the asymptotic distribution of all three order statistics. First, we consider the convergence of the k th upper order statistics.

Theorem 3.8. *Let F be a d.f. with a right (left) endpoint $x^F \leq \infty$ ($x_F \geq -\infty$) and $k = k(n)$ be a non-decreasing integer sequence such that*

$$\lim_{n \rightarrow \infty} \frac{k(n)}{n} = c \in [0, 1].$$

1. *Then $X_{n-k(n),n} \xrightarrow{a.s.} x^F(x_F)$ as $c = 0$ ($c = 1$).*
2. *If we instead assume that $c \in (0, 1)$ is such that there is unique solution $x(c)$ of the equation $\bar{F}(x) = c$. Then*

$$X_{n-k(n),n} \xrightarrow{a.s.} x(c).$$

Note that result 3.8 of Theorem 3.8 relates to intermediate order statistics whereas result 3.8 relates to central order statistics. The proof is simple and can be found in [12]. We now proceed to the discussion of the asymptotic distribution for each of the order statistics.

Theorem 3.9 (Asymptotic distribution of a central order statistic). *For $0 < p < 1$, let F be an absolutely continuous d.f. with density function f which is positive at $F^{\leftarrow}(p)$ and is continuous at that point. For $k = [np] + 1$, as $n \rightarrow \infty$,*

$$\sqrt{n} f(F^{\leftarrow}(p)) \frac{X_{n-k,n} - F^{\leftarrow}(p)}{\sqrt{p(1-p)}} \xrightarrow{d} \mathcal{N}(0, 1), \quad (3.19)$$

where $\mathcal{N}(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2 . Thus note that the central order statistics, when appropriately normalised, converges to the normal distribution. This property is known as *asymptotic normality* and is particu-

larly desirable for estimators as it allows for the construction of confidence intervals with relative ease. The proof of Theorem 3.9 can be found in [13].

We can consider the asymptotic distribution of the extreme order statistics (also known as the upper order statistic) which no longer exhibits asymptotic normality. Instead, in this case, we recover links to the Generalized Extreme Value distribution, G .

Theorem 3.10 (Asymptotic distribution of an extreme order statistic). *For any real x , $\mathbb{P}(X_{n-k+1,n} \leq a_n x + b_n) \rightarrow G(x)$ as $n \rightarrow \infty$ if and only for any fixed k ,*

$$\mathcal{P}(X_{n-k+1,n} \leq a_n x + b_n) \xrightarrow{n \rightarrow \infty} \sum_{j=0}^{k-1} G(x) \frac{(-\log G(x))^j}{j!}, \quad (3.20)$$

for all x .

The proof can be found in [13]. Note that F being in the domain of attraction of an extreme value distribution implies that Eq. (3.20) holds with the same a_n and b_n and thus establishes a strong link between the asymptotic behaviour of extreme order statistics and the sample maxima. However, when k is allowed to vary with n as for intermediate order statistics, we again acquire asymptotic normality.

Theorem 3.11 (Asymptotic distribution of an intermediate order statistic). *Suppose the von Mises condition given in Theorem 3.6 holds for some G_γ . Suppose further that $k \rightarrow \infty$ and $k/n \rightarrow 0$ as $n \rightarrow \infty$. Then*

$$\sqrt{k} \frac{X_{n-k,n} - U(\frac{n}{k})}{\frac{n}{k} U'(\frac{n}{k})} \xrightarrow{d} \mathcal{N}(0, 1). \quad (3.21)$$

A proof for Theorem 3.11 can be found in [25]. Thus we see that although intermediate order statistics is somewhere between central order statistics and extreme order statistics and intuitively closer to the latter, its asymptotic behaviour is more akin to that of the central order statistics. Theorem 3.11 also gives us the appropriate normalisation. We now consider an example as it will demonstrate how to use Theorems 3.9 and 3.11. It is also useful for numerical simulation.

Similarly, Theorems 3.9 and 3.10 can be used to choose appropriate normalisation for the relevant order statistics. Of course, in the above example, we have readily applied Theorem 3.11. In practice, we will need to check the von Mises condition or other relevant assumptions. This is taken for granted in the above example.

Order statistics are particularly useful as they are used to build various estimators for γ and x^F . The commonly used Hill estimator for $\gamma > 0$, is an example as is the more general Pickands estimator.

3.4 Extended Regular Variation

We have already alluded to the topics in this section however due to the technical complexity, it is given only at the end. The theory of regular variation provides us a tool box for understanding various functions that we have come across. Moreover, to set the theory that we have discussed within a wider framework, stronger conditions are necessary. These conditions follow readily if we are familiar with the theory of regular variation. The topics in this section may seem disjointed and irrelevant but in fact, it is instrumental to making extreme value theory as rich and robust as it is. We will start with the fundamentals.

Definition 3.4. Let ℓ be an eventually positive function on \mathbb{R}_+ . Then ℓ is said to be *slowly varying* if and only if

$$\lim_{u \rightarrow \infty} \frac{\ell(ux)}{\ell(x)} = 1.$$

Similarly, we can offer a more general version as follows.

Definition 3.5. Let f be an eventually positive function on \mathbb{R}_+ . Then f is said to be *regularly varying* with index ρ if and only if there exists a real constant ρ such that

$$\lim_{u \rightarrow \infty} \frac{f(ux)}{f(x)} = x^\rho. \quad (3.22)$$

ρ is called the *index of regular variation* [notation: $f \in RV_\rho$]. Note that if f satisfies Eq. (3.22) with $\rho = 0$, then f is slowly varying. Strictly speaking the above definitions require $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ to be Lebesgue measurable. We can readily assume this as most functions in our case are continuous and thus satisfy Lebesgue measurability. Note also that all regularly varying functions f can be written in terms of the a slowly varying function ℓ , i.e., if $f \in RV_\rho$, then $f(x) = x^\rho \ell(x)$ where $\ell \in RV_0$. Note then that in Theorem 3.3, the tail of F was regularly varying in both the Fréchet and Weibull cases.

We can make this even more general by considering functions that are of *extended regular variation* and/or belonging to a class of functions denoted by Π .

Definition 3.6. A measurable function $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ is said to be of extended regular variation if there exists a function $a : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that for some $\alpha \in \mathbb{R} \setminus \{0\}$ and for $x > 0$,

$$\lim_{u \rightarrow \infty} \frac{f(ux) - f(x)}{a(x)} = \frac{x^\alpha - 1}{\alpha}. \quad (3.23)$$

[Notation: $f \in ERV_\alpha$]. The function a is the *auxiliary function* for f . While we do not show this, $a \in RV_\alpha$. We can see now observe that $F \in \mathcal{D}(G_\gamma) \implies U \in ERV_\gamma$ with auxiliary function $a(t)$ (cf. Theorem 3.2). Not only this but we can link f to be regularly varying as follows.

Theorem 3.12. *Suppose f holds with Eq. (3.23), with $\alpha \neq 0$. Then*

1. *If $\alpha > 0$, then $\lim_{x \rightarrow \infty} f(x)/a(x) = 1/\alpha$ and hence $f \in RV_\alpha$.*
2. *If $\alpha < 0$, then $f(\infty) := \lim_{x \rightarrow \infty} f(x)$ exists, $\lim_{x \rightarrow \infty} (f(\infty) - f(x))/a(x) = -1/\alpha$ and hence $f(\infty) - f(x) \in RV_\alpha$.*

The proof can be found in Appendix B of [18]. Since we now have the relation between the normalising constants and EVI with the index of regular variation, it can be used to construct estimators for the EVI. It can also be used in simulations where the true value is known or can be calculated.

Definition 3.7. A measurable function $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ is said to belong to the class Π if there exist a function $a : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that, for $x > 0$,

$$\lim_{u \rightarrow \infty} \frac{f(ux) - f(x)}{a(x)} = \log x,$$

where a is again the auxiliary function for f [Notation: $f \in \Pi$ or $f \in \Pi(a)$]. In this case, a is measurable and slowly varying. Note that functions that belong to class Π are a special case of functions which are of extended regular variation, i.e. where the index is 0. Next we consider *Karamata's theorem* which gives us a way to integrate regularly varying function.

Theorem 3.13 (Karamata's Theorem). *Suppose $f \in RV_\alpha$. There exists $t_0 > 0$ such that $f(t)$ is positive and locally bounded for $t \geq t_0$. If $\alpha \geq -1$, then*

$$\lim_{t \rightarrow \infty} \frac{tf(t)}{\int_{t_0}^t f(s)ds} = \alpha + 1. \quad (3.24)$$

If $\alpha < -1$, or $\alpha = -1$ and $\int_0^\infty f(s)ds < \infty$, then

$$\lim_{t \rightarrow \infty} \frac{tf(t)}{\int_t^\infty f(s)ds} = -\alpha - 1. \quad (3.25)$$

Conversely, if Eq. (3.24) holds with $-1 < \alpha < \infty$, then $f \in RV_\alpha$. If Eq. (3.25) holds with $-\infty < \alpha < -1$, then $f \in RV_\alpha$.

Note that in Theorem 3.13, the converse for $\alpha = -1$ does not necessarily imply that f is regularly varying.

It is obvious how the definitions and theorems we have looked at so far are relevant; we have provided examples of functions that were used in the report that satisfy one or more definition. Recall that in Sect. 3.3, we made mentions of second order refinements. The next part, though glance rather terse at first glance, provides a good source of valuable information to the prediction of distinctive features in extreme data. We shall look further at extended regular variation of U in Eq. (3.10) (i.e., Eq. (3.6) specialised in U) to give thorough insight as to how the normalised spacings of quantiles attain the GPD tail quantile function in the limit. The second order refinement below seeks to address the order of convergence in Eq. (3.10).

Definition 3.8. The function U is said to satisfy the *second order refinement* if for some positive function a and some positive or negative function A with $\lim_{t \rightarrow \infty} A(t) = 0$,

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a(t)} - \frac{x^\gamma - 1}{\gamma}}{A(t)} =: H(x), \quad x > 0, \quad (3.26)$$

where H is some function that is not a multiple of $D_\gamma := (x^\gamma - 1)/\gamma$.

The non-multiplicity condition is merely to avoid trivial results. The functions a and A may be called the first-order and second-order auxiliary functions, respectively. As before, the function A controls the speed of convergence in Eq. (3.10). The next theorem establishes the form of H and gives some properties of the auxiliary functions.

Theorem 3.14. *Suppose the second order refinement Eq. (3.26) holds. Then there exist constants $c_1, c_2 \in \mathbb{R}$ and some parameter $\rho \leq 0$ such that*

$$H_{\gamma, \rho}(x) = c_1 \int_1^x s^{\gamma-1} \int_1^s u^{\rho-1} du ds + c_2 \int_1^x s^{\gamma+\rho-1} ds. \quad (3.27)$$

Moreover, for $x > 0$,

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{\frac{a(tx)}{a(t)} - x^\gamma}{A(t)} &= c_1 x^\gamma \frac{x^\rho - 1}{\rho}, \\ \lim_{t \rightarrow \infty} \frac{A(tx)}{A(t)} &= x^\rho. \end{aligned} \quad (3.28)$$

The results are understood in continuity i.e. taking the limit as ρ and/or γ goes to zero. This gives us that

$$H_{\gamma, \rho}(x) = \begin{cases} \frac{c_1}{\rho} (D_{\gamma+\rho}(x) - D_\gamma(x)) + c_2 D_{\gamma+\rho}(x), & \rho \neq 0, \gamma \neq 0 \\ \frac{c_1}{\gamma} (x^\gamma \log x - D_\gamma(x)) + c_2 D_\gamma(x), & \rho = 1, \gamma \neq 0 \\ \frac{c_1}{2} (\log x)^2 + c_2 \log x, & \rho = 0, \gamma = 0. \end{cases} \quad (3.29)$$

The case of $\gamma = 0$ and/or $\rho = 0$ is interpreted in the limiting sense as $\log x$. Without loss of generality the constants featuring in the above can set fixed at $c_1 = 1$ and $c_2 = 0$ (cf. Corollary 2.3.4 of [18]). The parameter ρ describes the speed of convergence in Eq. (3.26): ρ close to zero implies slow convergence whereas if $|\rho|$ large, then convergence is fast. The above theorem results from the work of [26]. Finally, we can provide the sufficient second order condition of von Mises type.

Theorem 3.15. *Suppose the tail quantile function U is twice differentiable. Define $A(t) := \frac{tU''(t)}{u'(t)} - \gamma + 1$. If A has constant sign for large t , $\lim_{t \rightarrow \infty} A(t) = 0$, and if $|A| \in RV_\rho$ for $\rho \leq 0$, then for $x > 0$,*

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{tU'(t)} - \frac{x^\gamma - 1}{\gamma}}{A(t)} = H_{\gamma, \rho}(x).$$

These definitions and results may seem unrelated or arbitrary but in fact some of the proofs of other results borrow understanding from the theory of regular variation, and functions such as the tail quantile function U as seen as of extended regular variation. Thus, regular variation theory allows us to extend the theory of extremes much further in a very natural way, it enables a full characterisation at high levels of the process generating the data by looking at the asymptotic behaviour of the exceedances above a sufficiently high threshold. It also allows us to prove asymptotic normality for various estimators. Thus, though quite involved, it is a very useful tool in extreme value analyses and is highly recommended for the enthusiastic or mathematically motivated reader.

In conclusion, extreme value theory gives us a broad and well grounded foundation to extrapolate beyond the range of available data. Using either sample maxima or exceedances over a threshold, valuable inferences about extremes can be made. These are made rigorous by the first order and second order conditioning, which are underpinned by the broader still theory of regular variation. Moreover, we have techniques to conduct these analyses even when conditions of independence and stationarity do not hold. These results have already been adapted to fields such as finance, flood forecasting, climate change. They are accessible to yet more fields, and in this book they will be adapted for electricity demand in low-voltage networks.

References

1. Faloutsos, M., Faloutsos, P., Faloutsos, C.: On power-law relationships of the internet topology. In: ACM SIGCOMM Computer Communication Review, vol. 29, pp. 251–262. ACM (1999)
2. Yook, S.-H., Jeong, H., Barabási, A.-L.: Modeling the internet's large-scale topology. Proc. Natl. Acad. Sci. **99**(21), 13382–13386 (2002)
3. Haande Haan, L.: On regular variation and its application to the weak convergence of sample extremes. Ph.D. thesis, Mathematisch Centrum Amsterdam (1970)
4. Bortkiewicz von Bortkiewicz, L.: Variationsbreite und Mittlerer Fehler. Berliner Mathematische Gesellschaft (1922)
5. Mises von Mises, L.: Über die Variationsbreite einer einer Beobachtungsreihe. Sitzungsberichte der Berliner Mathematischen Gesellschaft **22**, 3–8 (1923)
6. Dodd, E.L.: The greatest and the least variate under general laws of error. Trans. Am. Math. Soc. **25**(4), 525–539 (1923)
7. Fréchet, M.: Sur la loi de probabilité de l'écart maximum. Annales de la Société Polonaise de Mathématique 93–117 (1927)
8. Fisher, R.A., Tippett, L.H.C.: Limiting forms of the frequency distribution of the largest or smallest member of a sample. In: Mathematical Proceedings of the Cambridge Philosophical Society, vol. 24, pp. 180–190. Cambridge University Press (1928)
9. Gnedenko, B.V.: On a local limit theorem of the theory of probability. Uspekhi Matematicheskikh Nauk **3**(3), 187–194 (1948)
10. Gumbel, E.: Statistics of Extremes, p. 247. Columbia University Press, New York (1958)
11. Haande Haan, L.: Convergence of heteroscedastic extremes. Stat. Probab. Lett. **101**, 38–39 (2015)
12. Embrechts, P., Klüppelberg, C., Mikosch, T.: Modelling Extremal Events for Insurance and Finance. Springer, Berlin (1997)

13. Arnold, B.C., Balakrishnan, N., Nagaraja, H.N.: *A First Course in Order Statistics*. Wiley Series in Probability and Mathematical Statistics. Wiley (1992)
14. Gnedenko, B.: Sur la distribution limite du terme maximum d'une serie aleatoire. *Ann. Math.* 423–453 (1943)
15. Balkema, A.A., Haande Haan, L.: Residual life time at great age. *Ann. Probab.* 792–804 (1974)
16. Mises von Mises, R.: La distribution de la plus grande de n valeurs. *Rev. math. Union interbalcanique* **1**(1) (1936)
17. Jenkinson, A.F.: The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Q. J. R. Meteorol. Soc.* **81**(348), 158–171 (1955)
18. Haande Haan, L., Ferreira, A.: *Extreme Value Theory: An Introduction*. Springer (2006)
19. Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J.L.: *Statistics of Extremes: Theory and Applications*. Wiley (2004)
20. Fraga Alves, I., Neves, C.: Estimation of the finite right endpoint in the Gumbel domain. *Statistica Sinica* **24**, 1811–1835 (2014)
21. Ferreira, A., de Haan, L.: On the block maxima method in extreme value theory: PWM estimators. *Ann. Statist.* **43**(1), 276–298 (2015)
22. Pickands, J.: Statistical inference using extreme order statistics. *Ann. Stat.* 119–131 (1975)
23. Holmes, J., Moriarty, W.: Application of the generalized Pareto distribution to extreme value analysis in wind engineering. *J. Wind. Eng. Ind. Aerodyn.* **83**(1), 1–10 (1999)
24. Gilli, M., Këllezzi, E.: An application of extreme value theory for measuring financial risk. *Comput. Econ.* **27**(2–3), 207–228 (2006)
25. Falk, M.: Best attainable rate of joint convergence of extremes. In: *Extreme Value Theory*, pp. 1–9. Springer (1989)
26. Haande Haan, L., Stadtmüller, U.: Generalized regular variation of second order. *J. Aust. Math. Soc.* **61**(3), 381–395 (1996)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

