3

Fluctuations of Maxima

This chapter is concerned with classical extreme value theory and consequently it is fundamental for many results in this book. The central result is the Fisher–Tippett theorem which specifies the form of the limit distribution for centred and normalised maxima. The three families of possible limit laws are known as extreme value distributions. In Section 3.3 we describe their maximum domains of attraction and derive centring and normalising constants. A short summary is provided in Tables 3.4.2–3.4.4 where numerous examples are to be found.

The basic tool for studying rare events related to the extremes of a sample is the Poisson approximation: a first glimpse is given in Section 3.1. Poisson approximation is also the key to the weak limit theory of upper order statistics (see Section 4.2) and for the weak convergence of point processes (see Chapter 5).

The asymptotic theories for maxima and sums complement and contrast each other. Corresponding results exist for affinely transformed sums and maxima: stable distributions correspond to max–stable distributions, domains of attraction to maximum domains of attraction; see Chapter 2. Limit theorems for maxima and sums require appropriate normalising and centring constants. For maxima the latter are chosen as some quantile (or a closely related quantity) of the underlying marginal distribution. Empirical quantiles open the way for tail estimation. Chapter 6 is devoted to this important statistical problem. In Section 3.4 the mean excess function is introduced.

P. Embrechts et al., *Modelling Extremal Events*, Stochastic Modelling and Applied Probability 33, DOI 10.1007/978-3-642-33483-2_4,

It will prove to be a useful tool for distinguishing dfs in their right tail and plays an important role in tail estimation; see Chapter 6.

As in Chapters 1 and 2, regular variation continues to play a fundamental role. The maximum domain of attraction of an extreme value distribution can be characterised via regular variation and its extensions; see Section 3.3. We also study the relationship between subexponentiality and maximum domains of attraction. This will have consequences in Section 8.3, where the path of a risk process leading to ruin is characterised.

The theory of Section 3.4 allows us to present various results of the previous sections in a compact way. The key is the generalised extreme value distribution which also leads to the generalised Pareto distribution. These are two crucial notions which turn out to be very important for the statistics of rare events treated in Chapter 6.

The almost sure behaviour of maxima is considered in Section 3.5. These results find applications in Section 8.5, where we study the longest success–run in a random walk.

3.1 Limit Probabilities for Maxima

Throughout this chapter $X, X_1, X_2, \ldots$ is a sequence of iid non–degenerate rvs with common df F. Whereas in Chapter 2 we focussed on cumulative sums, in this chapter we investigate the fluctuations of the *sample maxima*

$$M_1 = X_1 , \quad M_n = \max\left(X_1, \ldots, X_n\right) , \quad n \geq 2 .$$

Corresponding results for minima can easily be obtained from those for maxima by using the identity

$$\min\left(X_1, \ldots, X_n\right) = -\max\left(-X_1, \ldots, -X_n\right) .$$

In Chapter 4 we continue with the analysis of the *upper order statistics* of the sample $X_1, \ldots, X_n$.

There is of course no difficulty in writing down the exact df of the maximum M_n:

$$P\left(M_n \leq x\right) = P\left(X_1 \leq x, \ldots, X_n \leq x\right) = F^n(x) , \quad x \in \mathbb{R} , \quad n \in \mathbb{N} .$$

Extremes happen "near" the upper end of the support of the distribution, hence intuitively the asymptotic behaviour of M_n must be related to the df F in its right tail near the right endpoint. We denote by

$$x_F = \sup\{x \in \mathbb{R} : \ F(x) < 1\}$$

the right endpoint of F. We immediately obtain, for all $x < x_F$,

$$P(M_n \le x) = F^n(x) \to 0\,, \quad n \to \infty\,,$$

and, in the case $x_F < \infty$, we have for $x \ge x_F$ that

$$P(M_n \le x) = F^n(x) = 1\,.$$

Thus $M_n \xrightarrow{P} x_F$ as $n \to \infty$, where $x_F \le \infty$. Since the sequence (M_n) is non–decreasing in n, it converges a.s., and hence we conclude that

$$M_n \xrightarrow{\text{a.s.}} x_F\,, \quad n \to \infty\,. \tag{3.1}$$

This fact does not provide a lot of information. More insight into the order of magnitude of maxima is given by weak convergence results for centred and normalised maxima. This is one of the main topics in classical extreme value theory. For instance, the fundamental Fisher–Tippett theorem (Theorem 3.2.3) has the following content: if there exist constants $c_n > 0$ and $d_n \in \mathbb{R}$ such that

$$c_n^{-1}(M_n - d_n) \xrightarrow{d} H\,, \quad n \to \infty\,, \tag{3.2}$$

for some non–degenerate distribution H, then H must be of the type of one of the three so-called standard *extreme value distributions*. This is similar to the CLT, where the stable distributions are the only possible non–degenerate limit laws. Consequently, one has to consider probabilities of the form

$$P\left(c_n^{-1}(M_n - d_n) \le x\right)\,,$$

which can be rewritten as

$$P(M_n \le u_n)\,, \tag{3.3}$$

where $u_n = u_n(x) = c_n x + d_n$. We first investigate (3.3) for general sequences (u_n), and afterwards come back to affine transformations as in (3.2). We ask:

Which conditions on F ensure that the limit of $P(M_n \le u_n)$ for $n \to \infty$ exists for appropriate constants u_n?

It turns out that one needs certain continuity conditions on F at its right endpoint. This rules out many important distributions. For instance, if F has a Poisson distribution, then $P(M_n \le u_n)$ never has a limit in $(0,1)$, whatever the sequence (u_n). This implies that the normalised maxima of iid Poisson distributed rvs do not have a non–degenerate limit distribution. This remark might be slightly disappointing, but it shows the crucial difference between sums and maxima. In the former case, the CLT yields the normal distribution as limit under the very general *moment condition* $EX^2 < \infty$. If $EX^2 = \infty$ the relatively small class of α–stable limit distributions enters. Only in that very heavy–tailed case do *conditions on the tail* $\overline{F} = 1 - F$ guarantee the existence of a limit distribution. In contrast to sums, *we always need rather delicate conditions on the tail* $\overline{F}$ *to ensure that* $P(M_n \le u_n)$ *converges to a non–trivial limit*, i.e. a number in $(0,1)$.

In what follows we answer the question above. We commence with an elementary result which is crucial for the understanding of the weak limit theory of sample maxima. It will become a standard tool throughout this book.

Proposition 3.1.1 (Poisson approximation)
For given $\tau \in [0, \infty]$ *and a sequence* (u_n) *of real numbers the following are equivalent*

$$n\overline{F}(u_n) \to \tau\,, \tag{3.4}$$

$$P(M_n \le u_n) \to e^{-\tau}\,. \tag{3.5}$$

Proof. Consider first $0 \le \tau < \infty$. If (3.4) holds, then

$$P(M_n \le u_n) = F^n(u_n) = \left(1 - \overline{F}(u_n)\right)^n = \left(1 - \frac{\tau}{n} + o\left(\frac{1}{n}\right)\right)^n\,,$$

which implies (3.5). Conversely, if (3.5) holds, then $\overline{F}(u_n) \to 0$. (Otherwise, $\overline{F}(u_{n_k})$ would be bounded away from 0 for some subsequence (n_k). Then $P(M_{n_k} \le u_{n_k}) = (1 - \overline{F}(u_{n_k}))^{n_k}$ would imply $P(M_{n_k} \le u_{n_k}) \to 0$.) Taking logarithms in (3.5) we have

$$-n \ln\left(1 - \overline{F}(u_n)\right) \to \tau\,.$$

Since $-\ln(1-x) \sim x$ for $x \to 0$ this implies that $n\overline{F}(u_n) = \tau + o(1)$, giving (3.4).

If $\tau = \infty$ and (3.4) holds, but (3.5) does not, there must be a subsequence (n_k) such that $P(M_{n_k} \le u_{n_k}) \to \exp\{-\tau'\}$ as $k \to \infty$ for some $\tau' < \infty$. But then (3.5) implies (3.4), so that $n_k\overline{F}(u_{n_k}) \to \tau' < \infty$, contradicting (3.4) with $\tau = \infty$. Similarly, (3.5) implies (3.4) for $\tau = \infty$. □

Remarks. 1) Clearly, Poisson's limit theorem is the key behind the above proof. Indeed, assume for simplicity $0 < \tau < \infty$ and define $B_n = \sum_{i=1}^n I_{\{X_i > u_n\}}$. This quantity has a binomial distribution with parameters $(n, \overline{F}(u_n))$. An application of Poisson's limit theorem yields $B_n \xrightarrow{d} Poi(\tau)$ if and only if $EB_n = n\overline{F}(u_n) \to \tau$ which is nothing but (3.4). Also notice that $P(M_n \le u_n) = P(B_n = 0) \to \exp\{-\tau\}$. This explains why (3.5) is sometimes referred to as *Poisson approximation* to the probability $P(M_n \le u_n)$.

2) Evidently, if there exists a sequence $(u_n^{(\tau)})$ satisfying (3.4) for some fixed $\tau > 0$, then we can find such a sequence for any $\tau > 0$. For instance, if $(u_n^{(1)})$ satisfies (3.4) with $\tau = 1$, $u_n^{(\tau)} = u_{[n/\tau]}^{(1)}$ obeys (3.4) for an arbitrary $\tau > 0$. □

By (3.1), (M_n) converges a.s. to the right endpoint x_F of the df F, hence

$$P(M_n \le x) \to \begin{cases} 0 & \text{if} \quad x < x_F\,, \\ 1 & \text{if} \quad x > x_F\,. \end{cases}$$

The following result extends this kind of 0–1 behaviour.

Corollary 3.1.2 *Suppose that $x_F < \infty$ and*

$$\overline{F}(x_F-) = F(x_F) - F(x_F-) > 0\,.$$

Then for every sequence (u_n) such that

$$P(M_n \le u_n) \to \rho\,,$$

either $\rho = 0$ or $\rho = 1$.

Proof. Since $0 \le \rho \le 1$, we may write $\rho = \exp\{-\tau\}$ with $0 \le \tau \le \infty$. By Proposition 3.1.1 we have $n\overline{F}(u_n) \to \tau$ as $n \to \infty$. If $u_n < x_F$ for infinitely many n we have $\overline{F}(u_n) \ge \overline{F}(x_{F-}) > 0$ for those n and hence $\tau = \infty$. The other possibility is that $u_n \ge x_F$ for all sufficiently large n, giving $n\overline{F}(u_n) = 0$, and hence $\tau = 0$. Thus $\tau = \infty$ or 0, giving $\rho = 0$ or 1. □

This result shows in particular that *for a df with a jump at its finite right endpoint no non–degenerate limit distribution for M_n exists*, whatever the normalisation.

A similar result is true for certain distributions with infinite right endpoint as we see from the following characterisation, given in Leadbetter, Lindgren and Rootzén [418], Theorem 1.7.13.

Theorem 3.1.3 *Let F be a df with right endpoint $x_F \le \infty$ and let $\tau \in (0, \infty)$. There exists a sequence (u_n) satisfying $n\overline{F}(u_n) \to \tau$ if and only if*

$$\lim_{x \uparrow x_F} \frac{\overline{F}(x)}{\overline{F}(x-)} = 1 \tag{3.6}$$

and $F(x_F-) = 1$. □

The result applies in particular to discrete distributions with infinite right endpoint. If the jump heights of the df decay sufficiently slowly, then a non–degenerate limit distribution for maxima does not exist. For instance, if X is integer–valued and $x_F = \infty$, then (3.6) translates into $\overline{F}(n)/\overline{F}(n-1) \to 1$ as $n \to \infty$.

These considerations show that some intricate asymptotic behaviour of (M_n) exists. The discreteness of a distribution can prevent the maxima from converging and instead forces "oscillatory" behaviour. Nonetheless, in this situation it is often possible to find a sequence (c_n) of integers such that $(M_n - c_n)$ is *tight*; i.e. every subsequence of $(M_n - c_n)$ contains a weakly convergent subsequence. This is true for the examples to follow; see Aldous [7], Section C2, Leadbetter et al. [418], Section 1.7.

Example 3.1.4 (Poisson distribution)

$$P(X = k) = e^{-\lambda}\, \lambda^k / k!\,, \quad k \in \mathbb{N}_0\,, \quad \lambda > 0\,.$$

Then

$$\begin{aligned} \frac{\overline{F}(k)}{\overline{F}(k-1)} &= 1 - \frac{F(k) - F(k-1)}{\overline{F}(k-1)} \\ &= 1 - \frac{\lambda^k}{k!} \left(\sum_{r=k}^{\infty} \frac{\lambda^r}{r!} \right)^{-1} \\ &= 1 - \left(1 + \sum_{r=k+1}^{\infty} \frac{k!}{r!}\, \lambda^{r-k} \right)^{-1} . \end{aligned}$$

The latter sum can be estimated as

$$\sum_{s=1}^{\infty} \frac{\lambda^s}{(k+1)(k+2)\cdots(k+s)} \le \sum_{s=1}^{\infty} \left(\frac{\lambda}{k} \right)^s = \frac{\lambda/k}{1 - \lambda/k}\,, \quad k > \lambda\,,$$

which tends to 0 as $k \to \infty$, so that $\overline{F}(k)/\overline{F}(k-1) \to 0$. Theorem 3.1.3 shows that no non–degenerate limit distribution for maxima exists and, furthermore, that no limit of the form $P(M_n \le u_n) \to \rho \in (0,1)$ exists, whatever the sequence of constants (u_n). □

Example 3.1.5 (Geometric distribution)

$$P(X = k) = p(1-p)^k\,, \quad k \in \mathbb{N}_0\,, \quad 0 < p < 1\,.$$

In this case, since $\overline{F}(k) = (1-p)^{k+1}$, we have that

$$\frac{\overline{F}(k)}{\overline{F}(k-1)} = 1 - p \in (0,1)\,.$$

Hence again no limit $P(M_n \le u_n) \to \rho$ exists except for $\rho = 0$ or 1.

Maxima of iid geometrically distributed rvs play a prominent role in the study of the length of the longest success–run in a random walk. We refer to Section 8.5, in particular to Theorem 8.5.13. □

Example 3.1.6 (Negative binomial distribution)

$$P(X = k) = \binom{v+k-1}{k} p^v (1-p)^k , \quad k \in \mathbb{N}_0 , \quad 0 < p < 1 , v > 0 .$$

For $v \in \mathbb{N}$ the negative binomial distribution generalises the geometric distribution in the following sense: the geometric distribution models the waiting time for the first success in a sequence of independent trials, whereas the negative binomial distribution models the waiting time for the vth success.

Using Stirling's formula we obtain

$$\lim_{k \to \infty} \frac{\overline{F}(k)}{\overline{F}(k-1)} = 1 - p \in (0,1) ;$$

i.e. no limit $P(M_n \le u_n) \to \rho$ exists except for $\rho = 0$ or 1. □

Notes and Comments

Extreme value theory is a classical topic in probability theory and mathematical statistics. Its origins go back to Fisher and Tippett [240]. Since then a large number of books and articles on extreme value theory has appeared. The interested reader may, for instance, consult the following textbooks: Adler [4], Aldous [7], Beirlant, Teugels and Vynckier [57], Berman [62], Falk, Hüsler and Reiss [225], Galambos [249], Gumbel [290], Leadbetter, Lindgren and Rootzén [418], Pfeifer [497], Reiss [526] and Resnick [530].

Some historical notes concerning the development of extreme value theory starting with Nicolas Bernoulli (1709) can be found in Reiss [526].

Our presentation is close in spirit to Leadbetter, Lindgren and Rootzén [418] and Resnick [530]. The latter book is primarily concerned with extreme value theory of iid observations. Two subjects are central: the main analytic tool of extreme value theory is the theory of regularly varying functions (see Appendix A3.1), and the basic probabilistic tool is point process theory (see Chapter 5). After a brief summary of results for iid observations, Leadbetter et al. [418] focus on extremes of stationary sequences and processes; see Sections 4.4, 5.3 and 5.5. Galambos [249] studies the weak and strong limit theory for extremes of iid observations. Moreover, Galambos [249] and also Resnick [530] include results on multivariate extremes. Beirlant et al. [57], Gumbel [290], Pfeifer [497] and Reiss [526] concentrate more on the statistical aspects; see Chapter 6 for more detailed information concerning statistical methods based on extreme value theory.

Extreme value theory for discrete distributions is treated, for instance, in Anderson [11, 12], Arnold, Balakrishnan and Nagaraja [20] and Gordon, Schilling and Waterman [280].

Adler [4], Berman [62] and Leadbetter et al. [418] study extremes of continuous–time (in particular Gaussian) processes.

3.2 Weak Convergence of Maxima Under Affine Transformations

We come back to the main topic of this chapter, namely to the characterisation of the possible limit laws for the maxima M_n of the iid sequence (X_n) under positive affine transformations; see (3.2). This extreme value problem can be considered as an analogue to the central limit problem. Consequently, the main parts of Sections 3.2 and 3.3 bear some resemblance to Section 2.2 and it is instructive to compare and contrast the corresponding results.

In this section we answer the question:

What are the possible (non–degenerate) limit laws for the maxima M_n when properly normalised and centred?

This question turns out to be closely related to the following:

Which distributions satisfy for all $n \geq 2$ the identity in law

$$\max(X_1, \ldots, X_n) \stackrel{d}{=} c_n X + d_n \tag{3.7}$$

for appropriate constants $c_n > 0$ and $d_n \in \mathbb{R}$?

The question is, in other words, which classes of distributions F are closed (up to affine transformations) for maxima. Relation (3.7) reminds us of the defining properties of a stable distribution; see (2.9) in Chapter 2. Those distributions are the only possible limit laws for sums of normalised and centred iid rvs. A similar notion exists for maxima.

Definition 3.2.1 (Max–stable distribution)
A non–degenerate rv X (the corresponding distribution or df) is called max–stable *if it satisfies* (3.7) *for iid $X, X_1, \ldots, X_n$, appropriate constants $c_n > 0$, $d_n \in \mathbb{R}$ and every $n \geq 2$.* □

Remark. 1) *From now on we refer to the centring constants d_n and the normalising constants c_n jointly as* norming constants. □

Assume for the moment that (X_n) is a sequence of iid max–stable rvs. Then (3.7) may be rewritten as follows

$$c_n^{-1}(M_n - d_n) \stackrel{d}{=} X . \tag{3.8}$$

We conclude that every max–stable distribution is a limit distribution for maxima of iid rvs. Moreover, max–stable distributions are the only limit laws for normalised maxima.

Theorem 3.2.2 (Limit property of max–stable laws)
The class of max–stable distributions coincides with the class of all possible (non–degenerate) limit laws for (properly normalised) maxima of iid rvs.

Proof. It remains to prove that the limit distribution of affinely transformed maxima is max–stable. Assume that for appropriate norming constants,

$$\lim_{n\to\infty} F^n\left(c_n x + d_n\right) = H(x)\,, \quad x \in \mathbb{R}\,,$$

for some non–degenerate df H. We anticipate here (and indeed state precisely in Theorem 3.2.3) that the possible limit dfs H are continuous functions on the whole of $\mathbb{R}$.

Then for every $k \in \mathbb{N}$

$$\lim_{n\to\infty} F^{nk}\left(c_n x + d_n\right) = \left(\lim_{n\to\infty} F^n\left(c_n x + d_n\right)\right)^k = H^k(x)\,, \quad x \in \mathbb{R}\,.$$

Furthermore,

$$\lim_{n\to\infty} F^{nk}\left(c_{nk} x + d_{nk}\right) = H(x)\,, \quad x \in \mathbb{R}\,.$$

By the convergence to types theorem (Theorem A1.5) there exist constants $\widetilde{c}_k > 0$ and $\widetilde{d}_k \in \mathbb{R}$ such that

$$\lim_{n\to\infty} \frac{c_{nk}}{c_n} = \widetilde{c}_k \quad \text{and} \quad \lim_{n\to\infty} \frac{d_{nk} - d_n}{c_n} = \widetilde{d}_k\,,$$

and for iid rvs $Y_1, \ldots, Y_k$ with df H,

$$\max\left(Y_1, \ldots, Y_k\right) \stackrel{d}{=} \widetilde{c}_k\, Y_1 + \widetilde{d}_k\,. \qquad \square$$

The following result is *the* basis of classical extreme value theory.

Theorem 3.2.3 (Fisher–Tippett theorem, limit laws for maxima)
Let (X_n) be a sequence of iid rvs. If there exist norming constants $c_n > 0$, $d_n \in \mathbb{R}$ and some non–degenerate df H such that

$$c_n^{-1}\left(M_n - d_n\right) \xrightarrow{d} H\,, \tag{3.9}$$

then H belongs to the type of one of the following three dfs:

$$\textit{Fréchet:} \qquad \Phi_\alpha(x) = \begin{cases} 0\,, & x \le 0 \\ \exp\left\{-x^{-\alpha}\right\}\,, & x > 0 \end{cases} \qquad \alpha > 0.$$

$$\textit{Weibull:} \qquad \Psi_\alpha(x) = \begin{cases} \exp\left\{-(-x)^{\alpha}\right\}\,, & x \le 0 \\ 1\,, & x > 0 \end{cases} \qquad \alpha > 0.$$

$$\textit{Gumbel:} \qquad \Lambda(x) = \exp\left\{-e^{-x}\right\}\,, \quad x \in \mathbb{R}\,.$$

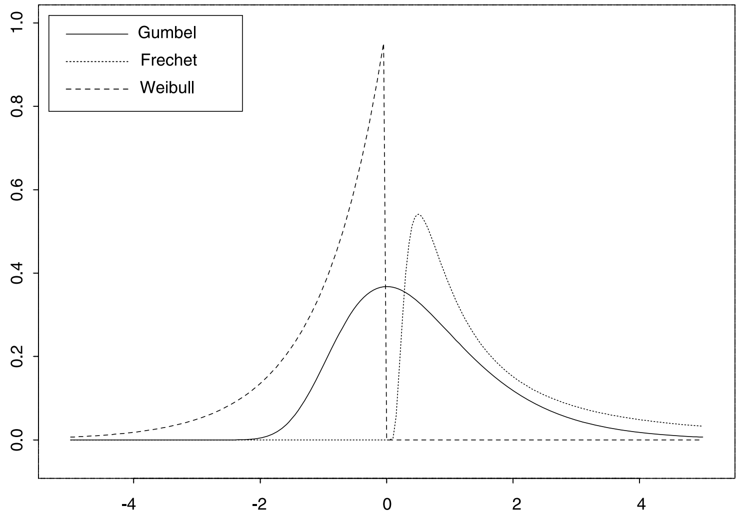


Figure 3.2.4 *Densities of the standard extreme value distributions. We chose* $\alpha = 1$ *for the Fréchet and the Weibull distribution.*

Sketch of the proof. Though a full proof is rather technical, we would like to show how the three limit–types appear; the main ingredient is again the convergence to types theorem, Theorem A1.5. Indeed, (3.9) implies that for all $t > 0$,

$$F^{[nt]}\left(c_{[nt]}x + d_{[nt]}\right) \to H(x)\,, \quad x \in \mathbb{R}\,,$$

where $[\cdot]$ denotes the integer part. However,

$$F^{[nt]}\left(c_n x + d_n\right) = \left(F^n\left(c_n x + d_n\right)\right)^{[nt]/n} \to H^t(x)\,,$$

so that by Theorem A1.5 there exist functions $\gamma(t) > 0$, $\delta(t) \in \mathbb{R}$ satisfying

$$\lim_{n\to\infty} \frac{c_n}{c_{[nt]}} = \gamma(t)\,, \quad \lim_{n\to\infty} \frac{d_n - d_{[nt]}}{c_{[nt]}} = \delta(t)\,, \quad t > 0\,,$$

and

$$H^t(x) = H(\gamma(t)x + \delta(t))\,. \tag{3.10}$$

It is not difficult to deduce from (3.10) that for $s, t > 0$

$$\gamma(st) = \gamma(s)\,\gamma(t)\,, \quad \delta(st) = \gamma(t)\,\delta(s) + \delta(t)\,. \tag{3.11}$$

The solution of the functional equations (3.10) and (3.11) leads to the three types Λ, Φ_α, Ψ_α. Details of the proof are for instance to be found in Resnick [530], Proposition 0.3. □

Remarks. 2) The limit law in (3.9) is unique only up to affine transformations. If the limit appears as $H(cx+d)$, i.e.

$$\lim_{n\to\infty} P\left(c_n^{-1}\left(M_n - d_n\right) \le x\right) = H(cx+d)\,,$$

then $H(x)$ is also a limit under a simple change of norming constants:

$$\lim_{n\to\infty} P\left(\tilde{c}_n^{-1}\left(M_n - \tilde{d}_n\right) \le x\right) = H(x)$$

with $\tilde{c}_n = c_n/c$ and $\tilde{d}_n = d_n - dc_n/c$. The convergence to types theorem shows precisely how affine transformations, weak convergence and types are related.

3) In Tables 1.2.5 and 1.2.6 we defined a Weibull df for $c, \alpha > 0$. For $c = 1$ it is given by

$$F_\alpha(x) = 1 - e^{-x^\alpha}\,, \quad x \ge 0\,,$$

which is the df of a positive rv. The Weibull distribution Ψ_α, as a limit distribution for maxima, is concentrated on $(-\infty, 0)$:

$$\Psi_\alpha(x) = 1 - F_\alpha(-x)\,, \quad x < 0\,.$$

In the context of extreme value theory we follow the convention and refer to Ψ_α as the Weibull distribution. We hope to avoid any confusion by a clear distinction between the two distributions whose extremal behaviour is completely different. Example 3.3.20 and Proposition 3.3.25 below show that F_α belongs to the maximum domain of attraction of the Gumbel distribution Λ.

4) The proof of Theorem 3.2.3 uses similar techniques as the proof of Theorems 2.2.2 and 2.2.3. Indeed, in the case $S_n = X_1 + \cdots + X_n$ we use the characteristic function $\phi_{S_n}(t) = (\phi_X(t))^n$, whereas for partial maxima we directly work with the df $F_{M_n}(x) = (F_X(x))^n$. So not suprisingly do we obtain functions like $\exp\{-c|t|^\alpha\}$ as possible limit chfs in the partial sum case, whereas such functions appear as limits for the dfs of normalised maxima.

5) Though, for modelling purposes, the types of Λ, Φ_α and Ψ_α are very different, from a mathematical point of view they are closely linked. Indeed, one immediately verifies the following properties. Suppose $X > 0$, then

$$X \text{ has df } \Phi_\alpha \iff \ln X^\alpha \text{ has df } \Lambda \iff -X^{-1} \text{ has df } \Psi_\alpha\,.$$

These relationships will appear again and again in various disguises throughout the book. $\square$

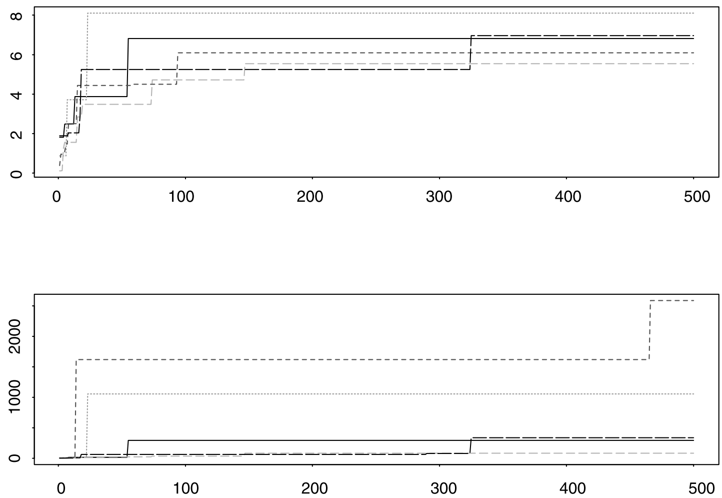


Figure 3.2.5 *Evolution of the maxima M_n of standard exponential (top) and Cauchy (bottom) samples. A sample path of (M_n) has a jump whenever $X_n > M_{n-1}$ (we say that M_n is a record). The graph seems to suggest that there occur more records for the exponential than for the Cauchy rvs. However, the distribution of the number of record times is approximately the same in both cases; see Theorem 5.4.7. The qualitative differences in the two graphs are due to a few large jumps for Cauchy distributed variables. Compared with those the smaller jumps are so tiny that they "disappear" from the computer graph; notice the difference between the vertical scales.*

Definition 3.2.6 (Extreme value distribution and extremal rv)
The dfs Φ_α, Ψ_α and Λ as presented in Theorem 3.2.3 *are called* standard extreme value distributions*, the corresponding rvs* standard extremal rvs*. Dfs of the types of Φ_α, Ψ_α and Λ are* extreme value distributions*; the corresponding rvs* extremal rvs. □

By Theorem 3.2.2, the extreme value distributions are precisely the max–stable distributions. Hence if X is an extremal rv it satisfies (3.8). In particular, the three cases in Theorem 3.2.3 correspond to

Fréchet:	$M_n \stackrel{d}{=} n^{1/\alpha}\, X$
Weibull:	$M_n \stackrel{d}{=} n^{-1/\alpha}\, X$
Gumbel:	$M_n \stackrel{d}{=} X + \ln n\,.$

Example 3.2.7 (Maxima of exponential rvs)
See also Figures 3.2.5 and 3.2.9. Let (X_i) be a sequence of iid standard exponential rvs. Then

$$\begin{aligned} P\left(M_n - \ln n \le x\right) &= \left(P\left(X \le x + \ln n\right)\right)^n \\ &= \left(1 - n^{-1}\, e^{-x}\right)^n \\ &\to \exp\left\{-e^{-x}\right\} = \Lambda(x)\,, \quad x \in \mathbb{R}\,. \end{aligned}$$

For comparison recall that for iid Gumbel rvs X_i,

$$P(M_n - \ln n \le x) = \Lambda(x)\,, \quad x \in \mathbb{R}\,. \qquad \square$$

Example 3.2.8 (Maxima of Cauchy rvs)
See also Figures 3.2.5 and 3.2.10. Let (X_i) be a sequence of iid standard Cauchy rvs. The standard Cauchy distribution is absolutely continuous with density

$$f(x) = \left(\pi\left(1 + x^2\right)\right)^{-1}, \quad x \in \mathbb{R}\,.$$

By l'Hospital's rule we obtain

$$\lim_{x\to\infty} \frac{\overline{F}(x)}{(\pi x)^{-1}} = \lim_{x\to\infty} \frac{f(x)}{\pi^{-1}x^{-2}} = \lim_{x\to\infty} \frac{\pi x^2}{\pi\left(1 + x^2\right)} = 1\,,$$

giving $\overline{F}(x) \sim (\pi x)^{-1}$. This implies

$$\begin{aligned} P\left(M_n \le \frac{nx}{\pi}\right) &= \left(1 - \overline{F}\left(\frac{nx}{\pi}\right)\right)^n \\ &= \left(1 - \frac{1}{n}\left(\frac{1}{x} + o(1)\right)\right)^n \\ &\to \exp\left\{-x^{-1}\right\} = \Phi_1(x)\,, \quad x > 0\,. \end{aligned} \qquad \square$$

Notes and Comments

Theorem 3.2.3 marked the beginning of extreme value theory as one of the central topics in probability theory and statistics. The limit laws for maxima were derived by Fisher and Tippett [240]. A first rigorous proof is due to Gnedenko [266]. De Haan [292] subsequently applied regular variation as an analytical tool. His work has been of great importance for the development of modern extreme value theory. Weissman [636] provided a simpler version of de Haan's proof, variations of which are now given in most textbooks on extreme value theory.

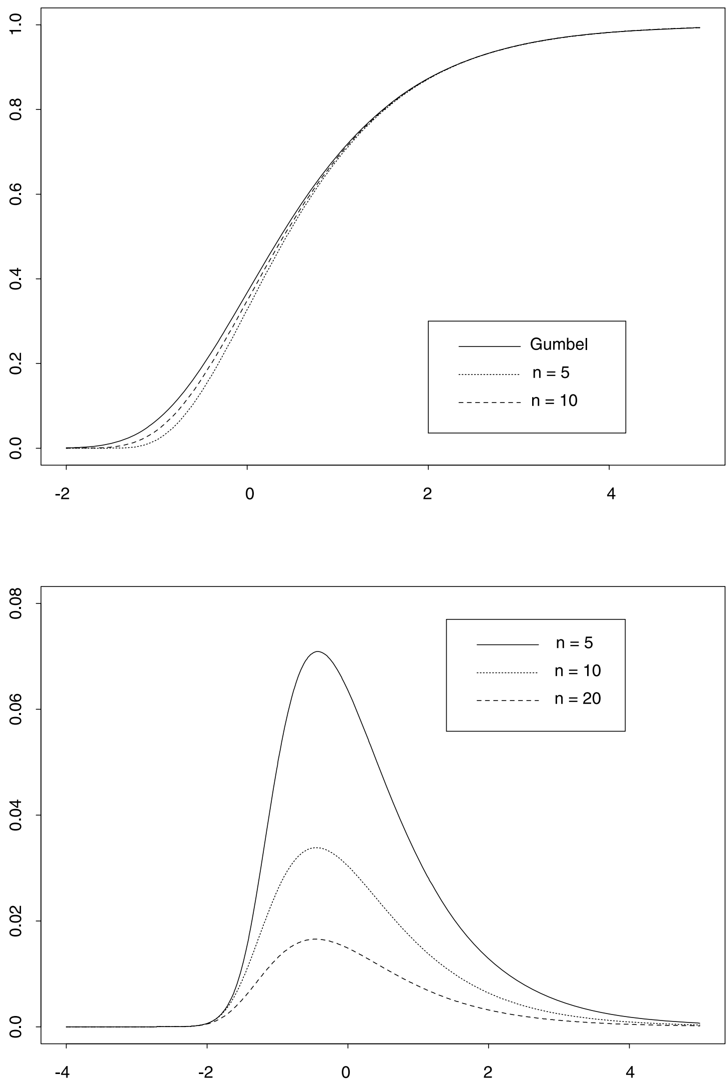


Figure 3.2.9 *The df $P(M_n - \ln n \le x)$ for n iid standard exponential rvs and the Gumbel df (top). In the bottom figure the relative error $(P(M_n - \ln n > x)/\overline{\Lambda}(x)) - 1$ of this approximation is illustrated.*

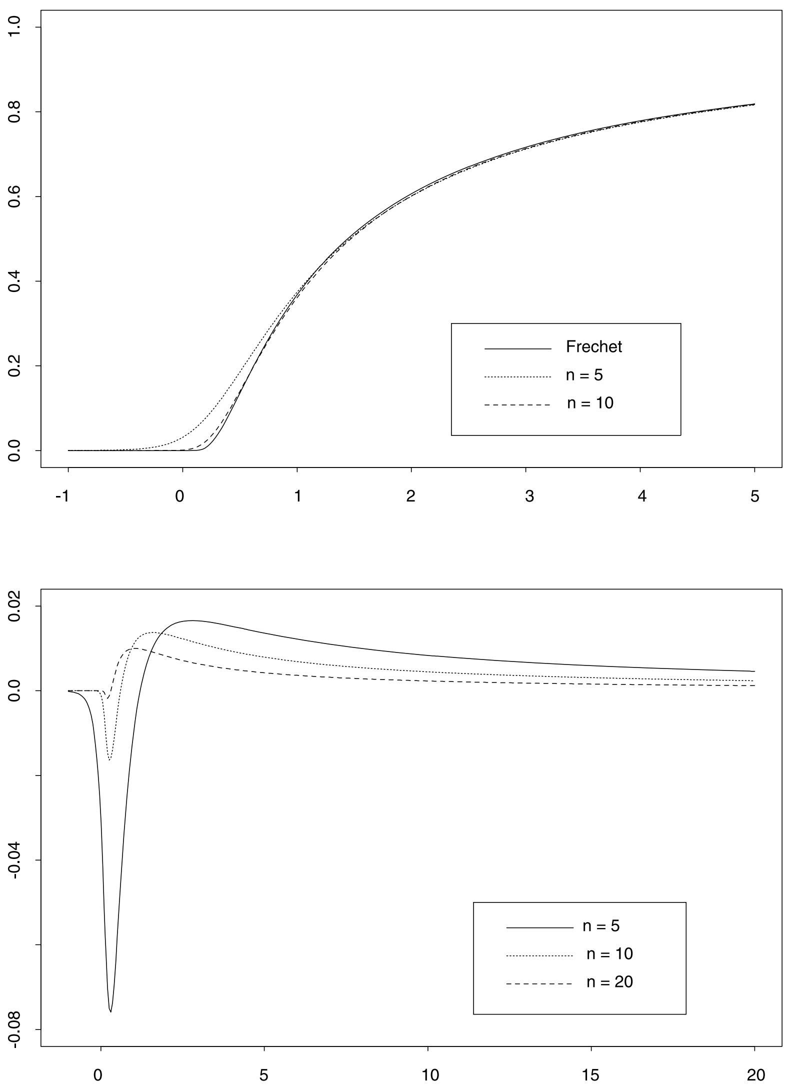


Figure 3.2.10 *The df* $P(\pi n^{-1} M_n \le x)$ *of* n *iid standard Cauchy rvs and the Fréchet df* Φ_1 *(top). In the bottom figure the relative error* $(P(\pi n^{-1} M_n > x)/\overline{\Phi}_1(x)) - 1$ *of this approximation is illustrated.*

3.3 Maximum Domains of Attraction and Norming Constants

In the preceding section we identified the extreme value distributions as the limit laws for normalised maxima of iid rvs; see Theorem 3.2.3. This section is devoted to the question:

Given an extreme value distribution H, what conditions on the df F imply that the normalised maxima M_n converge weakly to H?

Closely related to this question is the following:

How may we choose the norming constants $c_n > 0$ and $d_n \in \mathbb{R}$ such that

$$c_n^{-1}\left(M_n - d_n\right) \xrightarrow{d} H\,? \tag{3.12}$$

Can it happen that different norming constants imply convergence to different limit laws?

The last question can be answered immediately: the convergence to types theorem (Theorem A1.5) ensures that the limit law is uniquely determined up to affine transformations.

Before we answer the other questions recall from Section 2.2 how we proceeded with the sums $S_n = X_1 + \cdots + X_n$ of iid rvs: we collected all those dfs F in a common class for which the normalised sums S_n had the same stable limit distribution. Such a class is then called a domain of attraction (Definition 2.2.7). For maxima we proceed analogously.

Definition 3.3.1 (Maximum domain of attraction)
We say that the rv X (the df F of X, the distribution of X) belongs to the maximum domain of attraction of the extreme value distribution H *if there exist constants $c_n > 0$, $d_n \in \mathbb{R}$ such that* (3.12) *holds. We write* $X \in \mathrm{MDA}(H)$ $(F \in \mathrm{MDA}(H))$. □

Remark. Notice that the extreme value dfs are continuous on $\mathbb{R}$, hence $c_n^{-1}(M_n - d_n) \xrightarrow{d} H$ is equivalent to

$$\lim_{n\to\infty} P\left(M_n \le c_n x + d_n\right) = \lim_{n\to\infty} F^n\left(c_n x + d_n\right) = H(x)\,, \quad x \in \mathbb{R}\,. \quad \square$$

The following result is an immediate consequence of Proposition 3.1.1 and will be used throughout the following sections.

Proposition 3.3.2 (Characterisation of $\mathrm{MDA}(H)$)
The df F belongs to the maximum domain of attraction of the extreme value distribution H with norming constants $c_n > 0$, $d_n \in \mathbb{R}$ if and only if

$$\lim_{n\to\infty} n\overline{F}\left(c_n x + d_n\right) = -\ln H(x)\,, \quad x \in \mathbb{R}\,.$$

When $H(x) = 0$ the limit is interpreted as ∞. □

For every standard extreme value distribution we characterise its maximum domain of attraction. Using the concept of regular variation this is not too difficult for the Fréchet distribution Φ_α and the Weibull distribution Ψ_α; see Sections 3.3.1 and 3.3.2. Recall that a distribution tail $\overline{F}$ is regularly varying with index $-\alpha$ for some $\alpha \geq 0$, we write $\overline{F} \in \mathcal{R}_{-\alpha}$, if

$$\lim_{x\to\infty} \frac{\overline{F}(xt)}{\overline{F}(x)} = t^{-\alpha}, \quad t > 0.$$

The definition of regularly varying functions and those of their properties most important for our purposes can be found in Appendix A3.1. The interested reader may consult the monograph by Bingham, Goldie and Teugels [72] for an encyclopaedic treatment of regular variation. The maximum domain of attraction of the Gumbel distribution Λ is not so easily characterised; it consists of dfs whose right tail decreases to zero faster than any power function. This will be made precise in Section 3.3.3. If F has a density, simple sufficient conditions for F to be in the maximum domain of attraction of some extreme value distribution are due to von Mises. We present them below for practical (and historical) reasons.

The following concept defines an equivalence relation on the set of all dfs.

Definition 3.3.3 (Tail–equivalence)
Two dfs F and G are called tail–equivalent *if they have the same right endpoint, i.e. if $x_F = x_G$, and*

$$\lim_{x\uparrow x_F} \overline{F}(x)/\overline{G}(x) = c$$

for some constant $0 < c < \infty$. □

We show that every maximum domain of attraction is closed with respect to tail–equivalence, i.e. for tail–equivalent F and G, $F \in \mathrm{MDA}(H)$ if and only if $G \in \mathrm{MDA}(H)$. Moreover, for any two tail–equivalent distributions one can take the same norming constants. This will prove to be of great help for calculating norming constants which, in general, can become a rather tedious procedure.

Theorem 3.2.2 identifies the max–stable distributions as limit laws for affinely transformed maxima of iid rvs. The corresponding Theorem 2.2.2 for sums identifies the stable distributions as limit laws for centred and normalised sums. Sums are centred by their medians or by truncated means; see Proposition 2.2.14. The sample maximum M_n is the empirical version of the $(1-n^{-1})$–quantile of the underlying df F. Therefore the latter is an appropriate centring constant. Quantiles correspond to the "inverse" of a df, which is not always well–defined (dfs are not necessarily strictly increasing). In the following definition we fix upon a left–continuous version.

Definition 3.3.4 (Generalised inverse of a monotone function)
Suppose h is a non–decreasing function on $\mathbb{R}$. The generalised inverse of h *is defined as*

$$h^{\leftarrow}(t) = \inf\{x \in \mathbb{R} : h(x) \geq t\}\,.$$

(*We use the convention that the infimum of an empty set is ∞.*) □

Definition 3.3.5 (Quantile function)
The generalised inverse of the df F

$$F^{\leftarrow}(t) = \inf\{x \in \mathbb{R} :\ F(x) \geq t\}\,, \quad 0 < t < 1\,,$$

is called the quantile function *of the df F. The quantity $x_t = F^{\leftarrow}(t)$ defines the t–*quantile *of F.* □

We have summarised some properties of generalised inverse functions in Appendix A1.6.

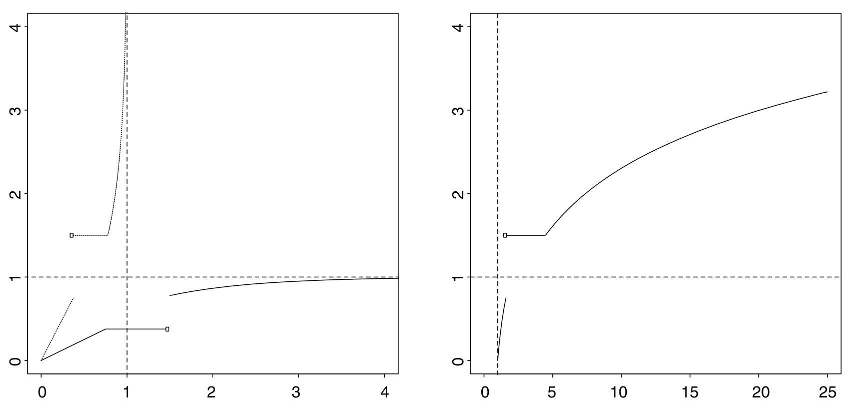


Figure 3.3.6 *An "interesting" df F, its quantile function $F^{\leftarrow}$ (left) and the corresponding function $F^{\leftarrow}(1 - x^{-1})$ (right).*

3.3.1 The Maximum Domain of Attraction of the Fréchet Distribution $\Phi_\alpha(x) = \exp\left\{-x^{-\alpha}\right\}$

In this section we characterise the maximum domain of attraction of Φ_α for $\alpha > 0$. By Taylor expansion,

$$1 - \Phi_\alpha(x) = 1 - \exp\left\{-x^{-\alpha}\right\} \sim x^{-\alpha}, \quad x \to \infty,$$

hence the tail of Φ_α decreases like a power law. We ask:

How far away can we move from a power tail
and still remain in $\mathrm{MDA}(\Phi_\alpha)$*?*

We show that the maximum domain of attraction of Φ_α consists of dfs F whose right tail is regularly varying with index $-\alpha$. For $F \in \mathrm{MDA}(\Phi_\alpha)$ the constants d_n can be chosen as 0 (centring is not necessary) and the c_n by means of the quantile function, more precisely by

$$\begin{aligned} c_n = F^{\leftarrow}(1-n^{-1}) &= \inf\left\{x \in \mathbb{R} : F(x) \geq 1 - n^{-1}\right\} \\ &= \inf\left\{x \in \mathbb{R} : \left(1/\overline{F}\right)(x) \geq n\right\} \\ &= \left(1/\overline{F}\right)^{\leftarrow}(n). \end{aligned} \tag{3.13}$$

Theorem 3.3.7 (Maximum domain of attraction of Φ_α)
The df F belongs to the maximum domain of attraction of Φ_α, $\alpha > 0$, if and only if $\overline{F}(x) = x^{-\alpha} L(x)$ for some slowly varying function L.
If $F \in \mathrm{MDA}(\Phi_\alpha)$, then

$$c_n^{-1} M_n \xrightarrow{d} \Phi_\alpha, \tag{3.14}$$

where the norming constants c_n can be chosen according to (3.13).

Notice that this result implies in particular that every $F \in \mathrm{MDA}(\Phi_\alpha)$ has an infinite right endpoint $x_F = \infty$. Furthermore, the norming constants c_n form a regularly varying sequence, more precisely, $c_n = n^{1/\alpha} L_1(n)$ for some slowly varying function L_1.

Proof. Let $\overline{F} \in \mathcal{R}_{-\alpha}$ for $\alpha > 0$. By the choice of c_n and regular variation,

$$\overline{F}(c_n) \sim n^{-1}, \qquad n \to \infty, \tag{3.15}$$

and hence $\overline{F}(c_n) \to 0$ giving $c_n \to \infty$. For $x > 0$,

$$n\overline{F}(c_n x) \sim \frac{\overline{F}(c_n x)}{\overline{F}(c_n)} \to x^{-\alpha}, \quad n \to \infty.$$

For $x < 0$, immediately $F^n(c_n x) \leq F^n(0) \to 0$, since regular variation requires $F(0) < 1$. By Proposition 3.3.2, $F \in \mathrm{MDA}(\Phi_\alpha)$.

Conversely, assume that $\lim_{n\to\infty} F^n(c_n x + d_n) = \Phi_\alpha(x)$ for all $x > 0$ and appropriate $c_n > 0$, $d_n \in \mathbb{R}$. This leads to

$$\lim_{n\to\infty} F^n(c_{[ns]}x + d_{[ns]}) = \Phi_\alpha^{1/s}(x) = \Phi_\alpha(s^{1/\alpha}x), \quad s > 0, x > 0.$$

By the convergence to types theorem (Theorem A1.5)

$$c_{[ns]}/c_n \to s^{1/\alpha} \quad \text{and} \quad \left(d_{[ns]} - d_n\right)/c_n \to 0\,.$$

Hence (c_n) is a regularly varying sequence in the sense of Definition A3.13, in particular $c_n \to \infty$. Assume first that $d_n = 0$, then $n\overline{F}(c_n x) \to x^{-\alpha}$ so that $\overline{F} \in \mathcal{R}_{-\alpha}$ because of Proposition A3.8(a). The case $d_n \neq 0$ is more involved, indeed one has to show that $d_n/c_n \to 0$. If the latter holds one can repeat the above argument by replacing d_n by 0. For details on this, see Bingham et al. [72], Theorem 8.13.2, or de Haan [292], Theorem 2.3.1. Resnick [530], Proposition 1.11 contains an alternative argument. □

We have found the answer to the above question:

$$\boxed{F \in \mathrm{MDA}(\Phi_\alpha) \quad \Longleftrightarrow \quad \overline{F} \in \mathcal{R}_{-\alpha}}$$

Thus we have a simple characterisation of $\mathrm{MDA}(\Phi_\alpha)$. Notice that this class of dfs contains "very heavy–tailed distributions" in the sense that $E(X^+)^\delta = \infty$ for $\delta > \alpha$. Thus they may be appropriate distributions for modelling large insurance claims and large fluctuations of prices, log–returns etc.

Von Mises found some easily verifiable conditions on the density of a distribution for it to belong to some maximum domain of attraction. The following is a consequence of Proposition A3.8(b).

Corollary 3.3.8 (Von Mises condition)
Let F be an absolutely continuous df with density f satisfying

$$\lim_{x\to\infty} \frac{x\,f(x)}{\overline{F}(x)} = \alpha > 0\,, \tag{3.16}$$

then $F \in \mathrm{MDA}(\Phi_\alpha)$. □

The class of dfs F with regularly varying tail $\overline{F}$ is obviously closed with respect to tail–equivalence (Definition 3.3.3). The following result gives us some insight into the structure of $\mathrm{MDA}(\Phi_\alpha)$. Besides this theoretical aspect, it will turn out to be a useful tool for calculating norming constants.

Proposition 3.3.9 (Closure property of $\mathrm{MDA}(\Phi_\alpha)$)
Let F and G be dfs and assume that $F \in \mathrm{MDA}(\Phi_\alpha)$ with norming constants $c_n > 0$, i.e.

$$\lim_{n\to\infty} F^n(c_n x) = \Phi_\alpha(x)\,, \quad x > 0\,. \tag{3.17}$$

Then

$$\lim_{n\to\infty} G^n(c_n x) = \Phi_\alpha(cx)\,, \quad x > 0\,,$$

for some $c > 0$ if and only if F and G are tail–equivalent with

$$\lim_{x\to\infty} \overline{F}(x)/\overline{G}(x) = c^\alpha\,.$$

Proof of the sufficiency. For the necessity part see Resnick [530], Proposition 1.19. Suppose that $\overline{F}(x) \sim q\,\overline{G}(x)$ as $x \to \infty$ for some $q > 0$. By Proposition 3.3.2 the limit relation (3.17) is equivalent to

$$\lim_{n\to\infty} n\overline{F}\,(c_n x) = x^{-\alpha}$$

for all $x > 0$. For such x, $c_n x \to \infty$ as $n \to \infty$ and hence, by tail–equivalence,

$$n\overline{G}\,(c_n x) \sim nq^{-1}\,\overline{F}\,(c_n x) \to q^{-1}\,x^{-\alpha}\,,$$

i.e. again by Proposition 3.3.2 ,

$$\lim_{n\to\infty} G^n\,(c_n x) = \exp\left\{-\left(q^{1/\alpha}x\right)^{-\alpha}\right\} = \Phi_\alpha\left(q^{1/\alpha}x\right)\,.$$

Now set $c = q^{1/\alpha}$. □

By Theorem 3.3.7, $F \in \mathrm{MDA}(\Phi_\alpha)$ if and only if $\overline{F} \in \mathcal{R}_{-\alpha}$. The representation theorem for regularly varying functions (Theorem A3.3) implies that every $F \in \mathrm{MDA}(\Phi_\alpha)$ is tail–equivalent to an absolutely continuous df satisfying (3.16). We can summarize this as follows:

MDA (Φ_α) consists of dfs satisfying the von Mises condition (3.16) and their tail–equivalent dfs.

We conclude this section with some examples.

Example 3.3.10 (Pareto–like distributions)
- Pareto
- Cauchy
- Burr
- Stable with exponent $\alpha < 2$.

The respective densities or dfs are given in Table 1.2.6; for stable distributions see Definition 2.2.1. All these distributions are Pareto–like in the sense that their right tails are of the form

$$\overline{F}(x) \sim Kx^{-\alpha}\,, \quad x \to \infty\,,$$

for some K, $\alpha > 0$. Obviously $\overline{F} \in \mathcal{R}_{-\alpha}$ which implies that $F \in \mathrm{MDA}(\Phi_\alpha)$ and as norming constants we can choose $c_n = (Kn)^{1/\alpha}$; see Theorem 3.3.7. Then

$$(Kn)^{-1/\alpha}\,M_n \xrightarrow{d} \Phi_\alpha\,.$$

The Cauchy distribution was treated in detail in Example 3.2.8. □

Example 3.3.11 (Loggamma distribution)
The loggamma distribution has tail

$$\overline{F}(x) \sim \frac{\alpha^{\beta-1}}{\Gamma(\beta)}(\ln x)^{\beta-1}x^{-\alpha}\,, \quad x \to \infty\,, \quad \alpha,\beta > 0\,. \tag{3.18}$$

Hence $\overline{F} \in \mathcal{R}_{-\alpha}$ which is equivalent to $F \in \mathrm{MDA}(\Phi_\alpha)$. According to Proposition 3.3.9 we choose c_n by means of the tail–equivalent right–hand side of (3.18). On applying (3.13) and taking logarithms we find we have to solve

$$\alpha \ln c_n - (\beta-1)\ln\ln c_n - \ln(\alpha^{\beta-1}/\Gamma(\beta)) = \ln n\,. \tag{3.19}$$

The solution satisfies

$$\ln c_n = \alpha^{-1}\left(\ln n + \ln r_n\right)\,,$$

where $\ln r_n = o(\ln n)$ as $n \to \infty$. We substitute this into equation (3.19) and obtain

$$\ln r_n = (\beta-1)\ln(\alpha^{-1}\ln n(1+o(1))) + \ln\left(\alpha^{\beta-1}/\Gamma(\beta)\right)\,.$$

This gives the norming constants

$$c_n \sim \left((\Gamma(\beta))^{-1}(\ln n)^{\beta-1}n\right)^{1/\alpha}\,.$$

Hence

$$\left((\Gamma(\beta))^{-1}(\ln n)^{\beta-1}n\right)^{-1/\alpha} M_n \quad \xrightarrow{d} \quad \Phi_\alpha\,. \qquad \square$$

3.3.2 The Maximum Domain of Attraction of the Weibull Distribution $\Psi_\alpha(x) = \exp\{-(-x)^\alpha\}$

In this section we characterise the maximum domain of attraction of Ψ_α for $\alpha > 0$. An important, though not at all obvious fact is that all dfs F in $\mathrm{MDA}(\Psi_\alpha)$ have finite right endpoint x_F. As was already indicated in Remark 5 of Section 3.2, Ψ_α and Φ_α are closely related, indeed

$$\Psi_\alpha\left(-x^{-1}\right) = \Phi_\alpha(x)\,, \quad x > 0\,.$$

Therefore we may expect that also $\mathrm{MDA}(\Psi_\alpha)$ and $\mathrm{MDA}(\Phi_\alpha)$ will be closely related. The following theorem confirms this.

Theorem 3.3.12 (Maximum domain of attraction of Ψ_α)
The df F belongs to the maximum domain of attraction of Ψ_α, $\alpha > 0$, if and only if $x_F < \infty$ and $\overline{F}(x_F - x^{-1}) = x^{-\alpha} L(x)$ for some slowly varying function L.

If $F \in \mathrm{MDA}(\Psi_\alpha)$, then

$$c_n^{-1}(M_n - x_F) \xrightarrow{d} \Psi_\alpha \,, \tag{3.20}$$

where the norming constants c_n can be chosen as $c_n = x_F - F^{\leftarrow}(1 - n^{-1})$ and $d_n = x_F$.

Sketch of the proof. The necessity part is difficult; see Resnick [530], Proposition 1.13. Sufficiency can be shown easily by exploiting the link between Φ_α and Ψ_α; see Remark 5 in Section 3.2. So suppose $x_F < \infty$ and $\overline{F}(x_F - x^{-1}) = x^{-\alpha} L(x)$ and define

$$F_*(x) = F\left(x_F - x^{-1}\right) \,, \quad x > 0 \,, \tag{3.21}$$

then $\overline{F}_* \in \mathcal{R}_{-\alpha}$ so that by Theorem 3.3.7, $F_* \in \mathrm{MDA}(\Phi_\alpha)$ with norming constants $c_n^* = F_*^{\leftarrow}(1 - n^{-1})$ and $d_n^* = 0$. The remaining part of the proof of sufficiency is now straightforward. Indeed, $F_* \in \mathrm{MDA}(\Phi_\alpha)$ implies that for $x > 0$,

$$F_*^n(c_n^* x) \to \Phi_\alpha(x) \,,$$

i.e.

$$F^n\left(x_F - (c_n^* x)^{-1}\right) \to \exp\left\{-x^{-\alpha}\right\} \,.$$

Substitute $x = -y^{-1}$, then

$$F^n(x_F + y/c_n^*) \to \exp\{-(-y)^\alpha\} \,, \quad y < 0 \,. \tag{3.22}$$

Finally,

$$\begin{aligned}
c_n^* &= F_*^{\leftarrow}\left(1 - n^{-1}\right) \\
&= \inf\left\{x \in \mathbb{R} : F(x_F - x^{-1}) \geq 1 - n^{-1}\right\} \\
&= \inf\left\{(x_F - u)^{-1} : F(u) \geq 1 - n^{-1}\right\} \\
&= \left(x_F - \inf\left\{u : F(u) \geq 1 - n^{-1}\right\}\right)^{-1} \\
&= \left(x_F - F^{\leftarrow}(1 - n^{-1})\right)^{-1} \,,
\end{aligned}$$

completing the proof because of (3.22). □

Consequently,

$$\boxed{F \in \mathrm{MDA}\,(\Psi_\alpha) \quad \Longleftrightarrow \quad x_F < \infty\,, \quad \overline{F}\left(x_F - x^{-1}\right) \in \mathcal{R}_{-\alpha}\,.}$$

Thus $\mathrm{MDA}(\Psi_\alpha)$ consists of dfs F with support bounded to the right. They may not be the best choice for modelling extremal events in insurance and finance, precisely because $x_F < \infty$. Though clearly in all circumstances in practice there is a (perhaps ridiculously high) upper limit, we may not want to incorporate this extra parameter x_F in our model. Often distributions with $x_F = \infty$ should be preferred since they allow for arbitrarily large values in a sample. Such distributions typically belong to $\mathrm{MDA}(\Phi_\alpha)$ or $\mathrm{MDA}(\Lambda)$. In Chapter 6 we shall discuss various such examples.

In the previous section we found it convenient to characterise membership in $\mathrm{MDA}(\Phi_\alpha)$ via the density of a df; see Corollary 3.3.8. Having in mind the transformation (3.21), Corollary 3.3.8 can be translated for $F \in \mathrm{MDA}(\Psi_\alpha)$.

Corollary 3.3.13 (Von Mises condition)
Let F be an absolutely continuous df with density f which is positive on some finite interval (z, x_F). If

$$\lim_{x \uparrow x_F} \frac{(x_F - x)\ f(x)}{\overline{F}(x)} = \alpha > 0\,, \tag{3.23}$$

then $F \in \mathrm{MDA}(\Psi_\alpha)$. □

Applying the transformation (3.21), Proposition 3.3.9 can be reformulated as follows.

Proposition 3.3.14 (Closure property of $\mathrm{MDA}\,(\Psi_\alpha)$)
Let F and G be dfs with right endpoints $x_F = x_G < \infty$ and assume that $F \in \mathrm{MDA}(\Psi_\alpha)$ with norming constants $c_n > 0$; i.e.

$$\lim_{n\to\infty} F^n\left(c_n x + x_F\right) = \Psi_\alpha(x)\,, \quad x < 0\,.$$

Then

$$\lim_{n\to\infty} G^n\left(c_n x + x_F\right) = \Psi_\alpha(cx)\,, \quad x < 0\,,$$

for some $c > 0$ if and only if F and G are tail–equivalent with

$$\lim_{x \uparrow x_F} \overline{F}(x)/\overline{G}(x) = c^{-\alpha}\,.$$ □

Notice that the representation theorem for regularly varying functions (Theorem A3.3) implies that every $F \in \mathrm{MDA}(\Psi_\alpha)$ is tail–equivalent to an absolutely continuous df satisfying (3.23). We summarize this as follows:

MDA(Ψ_α) consists of dfs satisfying the von Mises condition (3.23) and their tail–equivalent dfs.

We conclude this section with some examples of prominent MDA(Ψ_α)–members.

Example 3.3.15 (Uniform distribution on $(0,1)$)
Obviously, $x_F = 1$ and $\overline{F}(1-x^{-1}) = x^{-1} \in \mathcal{R}_{-1}$. Then by Theorem 3.3.12 we obtain $F \in \mathrm{MDA}(\Psi_1)$. Since $\overline{F}(1-n^{-1}) = n^{-1}$, we choose $c_n = n^{-1}$. This implies in particular

$$n\,(M_n - 1) \xrightarrow{d} \Psi_1 \,. \qquad \square$$

Example 3.3.16 (Power law behaviour at the finite right endpoint)
Let F be a df with finite right endpoint x_F and distribution tail

$$\overline{F}(x) = K\,(x_F - x)^\alpha\,, \quad x_F - K^{-1/\alpha} \le x \le x_F\,, \quad K, \alpha > 0\,.$$

By Theorem 3.3.12 this ensures that $F \in \mathrm{MDA}(\Psi_\alpha)$. The norming constants c_n can be chosen such that $\overline{F}(x_F - c_n) = n^{-1}$, i.e. $c_n = (nK)^{-1/\alpha}$ and, in particular,

$$(nK)^{1/\alpha}\,(M_n - x_F) \xrightarrow{d} \Psi_\alpha\,. \qquad \square$$

Example 3.3.17 (Beta distribution)
The beta distribution is absolutely continuous with density

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\, x^{a-1}(1-x)^{b-1}\,, \quad 0 < x < 1\,, \quad a\,, b > 0\,.$$

Notice that $f(1-x^{-1})$ is regularly varying with index $-(b-1)$ and hence, by Karamata's theorem (Theorem A3.6),

$$\overline{F}(1-x^{-1}) = \int_{1-x^{-1}}^{1} f(y)\,dy = \int_x^\infty f(1-y^{-1})y^{-2}\,dy \sim x^{-1} f(1-x^{-1})\,.$$

Hence $\overline{F}(1-x^{-1})$ is regularly varying with index $-b$ and

$$\overline{F}(x) \sim \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b+1)}\,(1-x)^b\,, \quad x \uparrow 1\,.$$

Thus the beta df is tail–equivalent to a df with power law behaviour at $x_F = 1$. By Proposition 3.3.14 the norming constants can be determined by this power law tail which fits into the framework of Example 3.3.16 above. $\square$

3.3.3 The Maximum Domain of Attraction of the Gumbel Distribution $\Lambda(x) = \exp\{-\exp\{-x\}\}$

Von Mises Functions

The maximum domain of attraction of the Gumbel distribution Λ covers a wide range of dfs F. Although there is no direct link with regular variation as in the maximum domains of attraction of the Fréchet and Weibull distribution, we will find extensions of regular variation which allow for a complete characterisation of MDA(Λ).

A Taylor expansion argument yields

$$1 - \Lambda(x) \sim e^{-x}\,, \quad x \to \infty\,,$$

hence $\overline{\Lambda}(x)$ decreases to zero at an exponential rate. Again the following question naturally arises:

How far away can we move from an exponential tail
and still remain in MDA(Λ)*?*

We will see in the present and the next section that MDA(Λ) contains dfs with very different tails, ranging from *moderately heavy* (such as the lognormal distribution) to *light* (such as the normal distribution). Also both cases $x_F < \infty$ and $x_F = \infty$ are possible. Before we give a general answer to the above question, we restrict ourselves to some absolutely continuous $F \in$ MDA(Λ) which have a simple representation, proposed by von Mises. These distributions provide an important building block of this maximum domain of attraction, and therefore we study them in detail. We will see later (Theorem 3.3.26 and Remark 4) that one only has to consider a slight modification of the von Mises functions in order to characterise MDA(Λ) completely.

Definition 3.3.18 (Von Mises function)
Let F be a df which is continuous at the right endpoint $x_F \le \infty$. Suppose there exists some $z < x_F$ such that F has representation

$$\overline{F}(x) = c \exp\left\{-\int_z^x \frac{1}{a(t)}\,dt\right\}\,, \quad z < x < x_F\,, \tag{3.24}$$

where c is some positive constant, $a(\cdot)$ is a positive and absolutely continuous function (*with respect to Lebesgue measure*) *with density a' and* $\lim_{x \uparrow x_F} a'(x) = 0$.

Then F is called a von Mises function*, the function $a(\cdot)$ the* auxiliary function *of F.* □

Remark. 1) Relation (3.24) should be compared with the Karamata representation of a regularly varying function; see Theorem A3.3. Substituting into (3.24) the function $a(x) = x/\delta(x)$ such that $\delta(x) \to \alpha \in [0,\infty)$ as $x \to \infty$, (3.24) becomes a regularly varying tail with index $-\alpha$. We will see later (see Remark 2 below) that the auxiliary function of a von Mises function with $x_F = \infty$ satisfies $a(x)/x \to 0$. It immediately follows that $\overline{F}(x)$ decreases to zero much faster than any power law $x^{-\alpha}$. □

We give some examples of von Mises functions.

Example 3.3.19 (Exponential distribution)

$$\overline{F}(x) = e^{-\lambda x}\,, \quad x \ge 0\,, \quad \lambda > 0\,.$$

F is a von Mises function with auxiliary function $a(x) = \lambda^{-1}$. □

Example 3.3.20 (Weibull distribution)

$$\overline{F}(x) = \exp\{-c\,x^{\tau}\}\,, \quad x \ge 0\,, \quad c\,,\tau > 0\,.$$

F is a von Mises function with auxiliary function

$$a(x) = c^{-1}\tau^{-1}x^{1-\tau}\,, \quad x > 0\,.$$ □

Example 3.3.21 (Erlang distribution)

$$\overline{F}(x) = e^{-\lambda x}\sum_{k=0}^{n-1}\frac{(\lambda x)^k}{k!}\,, \quad x \ge 0\,, \quad \lambda > 0\,, n \in \mathbb{N}\,.$$

F is a von Mises function with auxiliary function

$$a(x) = \sum_{k=0}^{n-1}\frac{(n-1)!}{(n-k-1)!}\,\lambda^{-(k+1)}\,x^{-k}\,, \quad x > 0\,.$$

Notice that F is the $\Gamma(n,\lambda)$ df. □

Example 3.3.22 (Exponential behaviour at the finite right endpoint)
Let F be a df with finite right endpoint x_F and distribution tail

$$\overline{F}(x) = K\exp\left\{-\frac{\alpha}{x_F - x}\right\}\,, \quad x < x_F\,, \quad \alpha\,, K > 0\,.$$

F is a von Mises function with auxiliary function

$$a(x) = \frac{(x_F - x)^2}{\alpha}\,, \quad x < x_F\,.$$

For $x_F = 1$, $\alpha = 1$ and $K = e$ we obtain for example

$$\overline{F}(x) = \exp\left\{-\frac{x}{1-x}\right\}\,, \quad 0 \le x < 1\,.$$ □

Example 3.3.23 (Differentiability at the right endpoint)
Let F be a df with right endpoint $x_F \le \infty$ and assume there exists some $z < x_F$ such that F is twice differentiable on (z, x_F) with positive density $f = F'$ and $F''(x) < 0$ for $z < x < x_F$. Then it is not difficult to see that F is a von Mises function with auxiliary function $a = \overline{F}/f$ if and only if

$$\lim_{x \uparrow x_F} \overline{F}(x)\, F''(x)/f^2(x) = -1\,. \qquad (3.25)$$

Indeed, let $z < x < x_F$ and set $Q(x) = -\ln \overline{F}(x)$ and $a(x) = 1/Q'(x) = \overline{F}(x)/f(x) > 0$. Hence F has representation (3.24). Furthermore,

$$a'(x) = -\frac{\overline{F}(x)\, F''(x)}{f^2(x)} - 1$$

and (3.25) is equivalent to $a'(x) \to 0$ as $x \uparrow x_F$.

Condition (3.25) applies to many distributions of interest, including the normal distribution; see Example 3.3.29. □

In Remark 1 above we gained some indication that regular variation does not seem to be the right tool for describing von Mises functions. Recall the notion of *rapidly varying function* from Definition A3.11. In particular, $\overline{F} \in \mathcal{R}_{-\infty}$ means that

$$\lim_{x\to\infty} \frac{\overline{F}(xt)}{\overline{F}(x)} = \begin{cases} 0 & \text{if} \quad t > 1\,, \\ \infty & \text{if} \quad 0 < t < 1\,. \end{cases}$$

It is mentioned in Appendix A3 that some of the important results for regularly varying functions can be extended to $\mathcal{R}_{-\infty}$ in a natural way; see Theorem A3.12.

Proposition 3.3.24 (Properties of von Mises functions)
Every von Mises function F is absolutely continuous on (z, x_F) with positive densitiy f. The auxiliary function can be chosen as $a(x) = \overline{F}(x)/f(x)$. Moreover, the following properties hold.

(a) *If $x_F = \infty$, then $\overline{F} \in \mathcal{R}_{-\infty}$ and*

$$\lim_{x\to\infty} \frac{x f(x)}{\overline{F}(x)} = \infty\,. \qquad (3.26)$$

(b) *If $x_F < \infty$, then $\overline{F}(x_F - x^{-1}) \in \mathcal{R}_{-\infty}$ and*

$$\lim_{x \uparrow x_F} \frac{(x_F - x) f(x)}{\overline{F}(x)} = \infty\,. \qquad (3.27)$$

Remarks. 2) It follows from (3.26) that $\lim_{x\to\infty} x^{-1}a(x) = 0$, and from (3.27) that $a(x) = o(x_F - x) = o(1)$ as $x \uparrow x_F$.

3) Note that $a^{-1}(x) = f(x)/\overline{F}(x)$ is the *hazard rate* of F. □

Proof. From representation (3.24) we obtain

$$\frac{d}{dx}\left(-\ln \overline{F}(x)\right) = \frac{f(x)}{\overline{F}(x)} = \frac{1}{a(x)}\,, \quad z < x < x_F\,.$$

(a) Since $a'(x) \to 0$ as $x \to \infty$ the Cesàro mean of a' also converges:

$$\lim_{x\to\infty} \frac{a(x)}{x} = \lim_{x\to\infty} \frac{1}{x}\int_z^x a'(t)\,dt = 0\,. \tag{3.28}$$

This implies (3.26). $\overline{F} \in \mathcal{R}_{-\infty}$ follows from an application of Theorem A3.12(b).

(b) We have

$$\begin{aligned}\lim_{x\uparrow x_F} \frac{a(x)}{x_F - x} &= \lim_{x\uparrow x_F} -\int_x^{x_F} \frac{a'(t)}{x_F - x}\,dt \\ &= \lim_{s\downarrow 0} \frac{1}{s}\int_0^s a'\left(x_F - t\right)\,dt\end{aligned}$$

by change of variables. Since $a'(x_F - t) \to 0$ as $t \downarrow 0$, the last limit tends to 0. This implies (3.27). $\overline{F}(x_F - x^{-1}) \in \mathcal{R}_{-\infty}$ follows as above. □

Now we can show that von Mises functions belong to the maximum domain of attraction of the Gumbel distribution. Moreover, the specific form of $\overline{F}$ allows to calculate the norming constants c_n from the auxiliary function.

Proposition 3.3.25 (Von Mises functions and MDA(Λ))
Suppose the df F is a von Mises function. Then $F \in \mathrm{MDA}(\Lambda)$. A possible choice of norming constants is

$$d_n = F^{\leftarrow}(1 - n^{-1}) \quad \textit{and} \quad c_n = a(d_n)\,, \tag{3.29}$$

where a is the auxiliary function of F.

Proof. Representation (3.24) implies for $t \in \mathbb{R}$ and x sufficiently close to x_F that

$$\frac{\overline{F}(x + t\,a(x))}{\overline{F}(x)} = \exp\left\{-\int_x^{x+t\,a(x)} \frac{1}{a(u)}\,du\right\}.$$

We set $v = (u - x)/a(x)$ and obtain

$$\frac{\overline{F}(x + t\,a(x))}{\overline{F}(x)} = \exp\left\{-\int_0^t \frac{a(x)}{a(x + v\,a(x))}\,dv\right\}. \tag{3.30}$$

We show that the integrand converges locally uniformly to 1. For given $\varepsilon > 0$ and $x \geq x_0(\varepsilon)$,

$$|a(x+va(x)) - a(x)| = \left| \int_x^{x+va(x)} a'(s)\, ds \right| \leq \varepsilon |v| a(x) \leq \varepsilon |t| a(x) \,,$$

where we used $a'(x) \to 0$ as $x \uparrow x_F$. This implies for $x \geq x_0(\varepsilon)$ that

$$\left| \frac{a(x+va(x))}{a(x)} - 1 \right| \leq \varepsilon |t| \,.$$

The right–hand side can be made arbitrarily small, hence

$$\lim_{x \uparrow x_F} \frac{a(x)}{a(x + v\, a(x))} = 1 \,, \tag{3.31}$$

uniformly on bounded v–intervals. This together with (3.30) yields

$$\lim_{x \uparrow x_F} \frac{\overline{F}(x + t\, a(x))}{\overline{F}(x)} = e^{-t} \tag{3.32}$$

uniformly on bounded t–intervals. Now choose the norming constants $d_n = (1/\overline{F})^{\leftarrow}(n)$ and $c_n = a(d_n)$. Then (3.32) implies

$$\lim_{n \to \infty} n\overline{F}\,(d_n + tc_n) = e^{-t} = -\ln \Lambda(t) \,, \quad t \in \mathbb{R} \,.$$

An application of Proposition 3.3.2 shows that $F \in \mathrm{MDA}(\Lambda)$. □

This result finishes our study of von Mises functions.

Characterisations of MDA(Λ)

Von Mises functions do not completely characterise the maximum domain of attraction of Λ. However, a slight modification of the defining relation (3.24) of a von Mises function yields a complete characterisation of MDA(Λ).

For a proof of the following result we refer to Resnick [530], Corollary 1.7 and Proposition 1.9.

Theorem 3.3.26 (Characterisation I of MDA(Λ))
The df F with right endpoint $x_F \leq \infty$ belongs to the maximum domain of attraction of Λ if and only if there exists some $z < x_F$ such that F has representation

$$\overline{F}(x) = c(x) \exp\left\{ -\int_z^x \frac{g(t)}{a(t)}\, dt \right\} , \quad z < x < x_F \,, \tag{3.33}$$

where c and g are measurable functions satisfying $c(x) \to c > 0$, $g(x) \to 1$ as $x \uparrow x_F$, and $a(x)$ is a positive, absolutely continuous function (with respect to

Lebesgue measure) with density $a'(x)$ *having* $\lim_{x \uparrow x_F} a'(x) = 0$.

For F *with representation* (3.33) *we can choose*

$$d_n = F^{\leftarrow}(1 - n^{-1}) \quad \textit{and} \quad c_n = a(d_n)$$

as norming constants.

A possible choice for the function a *is*

$$a(x) = \int_x^{x_F} \frac{\overline{F}(t)}{\overline{F}(x)}\, dt\,, \qquad x < x_F\,. \tag{3.34}$$

□

Motivated by von Mises functions, we call the function a in (3.33) an *auxiliary function* for F.

Remarks. 4) Representation (3.33) is not unique, there being some trade–off possible between the functions c and g. The following representation can be employed alternatively; see Resnick [530], Proposition 1.4:

$$\overline{F}(x) = c(x) \exp\left\{ - \int_z^x \frac{1}{a(t)}\, dt \right\}, \quad z < x < x_F\,, \tag{3.35}$$

for functions c and a with properties as in Theorem 3.3.26.

5) For a rv X the function $a(x)$ as defined in (3.34) is nothing but the *mean excess function*

$$a(x) = E(X - x \mid X > x)\,, \quad x < x_F\,;$$

see also Section 3.4 for a discussion on the use of this function. In Chapter 6 the mean excess function will turn out to be an important tool for statistical fitting of extremal event data. □

Another characterisation of MDA(Λ) was suggested in the proof of Proposition 3.3.25. There it was shown that every von Mises function satisfies (3.32), i.e. there exists a positive function $\widetilde{a}$ such that

$$\lim_{x \uparrow x_F} \frac{\overline{F}(x + t\widetilde{a}(x))}{\overline{F}(x)} = e^{-t}\,, \quad t \in \mathbb{R}\,. \tag{3.36}$$

Theorem 3.3.27 (Characterisation II of MDA(Λ))
The df F *belongs to the maximum domain of attraction of* Λ *if and only if there exists some positive function* $\widetilde{a}$ *such that* (3.36) *holds. A possible choice is* $\widetilde{a} = a$ *as given in* (3.34). □

The proof of this result is for instance to be found in de Haan [292], Theorem 2.5.1.

Now recall the notion of tail–equivalence (Definition 3.3.3). Similarly to the maximum domains of attraction of the Weibull and Fréchet distribution, tail–equivalence is an auxiliary tool to decide whether a particular distribution belongs to the maximum domain of attraction of Λ and to calculate the norming constants. In MDA(Λ) it is even more important because of the large variety of tails $\overline{F}$.

Proposition 3.3.28 (Closure property of MDA(Λ) under tail–equivalence)
Let F and G be dfs with the same right endpoint $x_F = x_G$ and assume that $F \in \mathrm{MDA}(\Lambda)$ with norming constants $c_n > 0$ and $d_n \in \mathbb{R}$; i.e.

$$\lim_{n\to\infty} F^n\left(c_n x + d_n\right) = \Lambda(x)\,, \quad x \in \mathbb{R}\,. \tag{3.37}$$

Then

$$\lim_{n\to\infty} G^n\left(c_n x + d_n\right) = \Lambda(x+b)\,, \quad x \in \mathbb{R}\,,$$

if and only if F and G are tail–equivalent with

$$\lim_{x\uparrow x_F} \overline{F}(x)/\overline{G}(x) = e^b\,.$$

Proof of the sufficiency. For a proof of the necessity see Resnick [530], Proposition 1.19. Suppose that $\overline{F}(x) \sim c\,\overline{G}(x)$ as $x \uparrow x_F$ for some $c > 0$. By Proposition 3.3.2 the limit relation (3.37) is equivalent to

$$\lim_{n\to\infty} n\overline{F}\left(c_n x + d_n\right) = e^{-x}\,, \quad x \in \mathbb{R}\,.$$

For such x, $c_n x + d_n \to x_F$ and hence, by tail–equivalence,

$$n\,\overline{G}\left(c_n x + d_n\right) \sim n c^{-1}\overline{F}\left(c_n x + d_n\right) \to c^{-1} e^{-x}\,, \quad x \in \mathbb{R}\,.$$

Therefore by Proposition 3.3.2,

$$\lim_{n\to\infty} G^n\left(c_n x + d_n\right) = \exp\left\{-e^{-(x+\ln c)}\right\} = \Lambda(x + \ln c)\,, \quad x \in \mathbb{R}\,.$$

Now set $\ln c = b$. $\square$

The results of this section yield a further complete characterisation of MDA(Λ).

MDA(Λ) consists of von Mises functions and their tail–equivalent dfs.

This statement and the examples discussed throughout this section show that MDA(Λ) consists of a large variety of distributions whose tails can be very different. Tails may range from moderately heavy (lognormal, heavy–tailed Weibull) to very light (exponential, dfs with support bounded to the right). Because of this, MDA(Λ) is perhaps the most interesting among all maximum domains of attraction. As a natural consequence of the variety of tails in MDA(Λ), the norming constants also vary considerably. Whereas in MDA(Φ_α) and MDA(Ψ_α) the norming constants are calculated by straightforward application of regular variation theory, more advanced results are needed for MDA(Λ). A complete theory has been developed by de Haan involving certain subclasses of $\mathcal{R}_{-\infty}$ and $\mathcal{R}_0$; see de Haan [292] or Bingham et al. [72], Chapter 3. Various examples below will illustrate the usefulness of results like Proposition 3.3.28.

Example 3.3.29 (Normal distribution)
See also Figure 3.3.30. Denote by Φ the df and by φ the density of the standard normal distribution. We first show that Φ is a von Mises function and check condition (3.25). An application of l'Hospital's rule to $\overline{\Phi}(x)/(x^{-1}\varphi(x))$ yields *Mill's ratio*, $\overline{\Phi}(x) \sim \varphi(x)/x$. Furthermore $\varphi'(x) = -x\varphi(x) < 0$ and

$$\lim_{x\to\infty} \frac{\overline{\Phi}(x)\,\varphi'(x)}{\varphi^2(x)} = -1\,.$$

Thus $\Phi \in$ MDA(Λ) by Example 3.3.23 and Proposition 3.3.25. We now calculate the norming constants. Use Mill's ratio again:

$$\overline{\Phi}(x) \sim \frac{\varphi(x)}{x} = \frac{1}{\sqrt{2\pi}\,x}\,e^{-x^2/2}\,, \quad x \to \infty\,, \tag{3.38}$$

and interpret the right–hand side as the tail of some df G. Then by Proposition 3.3.28, Φ and G have the same norming constants c_n and d_n. According to (3.29), $d_n = G^{\leftarrow}(1 - n^{-1})$. Hence look for a solution of $-\ln \overline{G}(d_n) = \ln n$; i.e.

$$\frac{1}{2}\,d_n^2 + \ln d_n + \frac{1}{2}\ln 2\pi = \ln n\,. \tag{3.39}$$

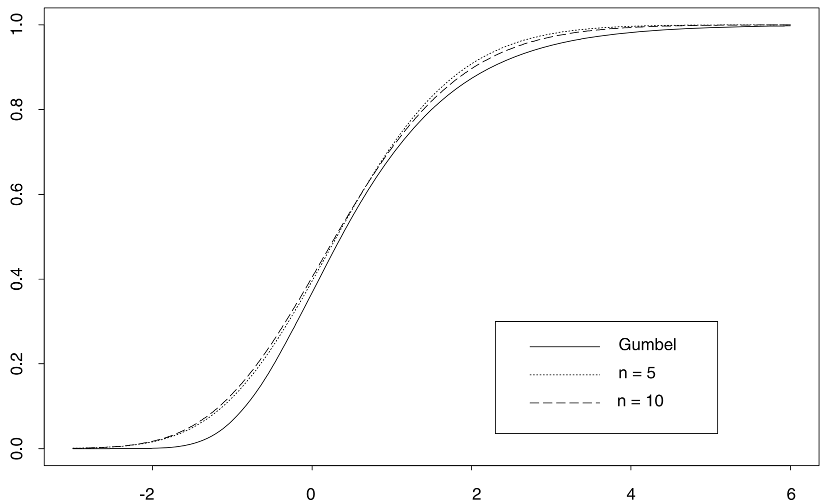
Then a Taylor expansion in (3.39) yields

$$d_n = (2\ln n)^{1/2} - \frac{\ln\ln n + \ln 4\pi}{2(2\ln n)^{1/2}} + o\left((\ln n)^{-1/2}\right)$$

as a possible choice for d_n. Since we can take $a(x) = \overline{\Phi}(x)/\varphi(x)$ we have that $a(x) \sim x^{-1}$ and therefore

$$c_n = a\,(d_n) \sim (2\ln n)^{-1/2}\,.$$

As the c_n are unique up to asymptotic equivalence, we choose



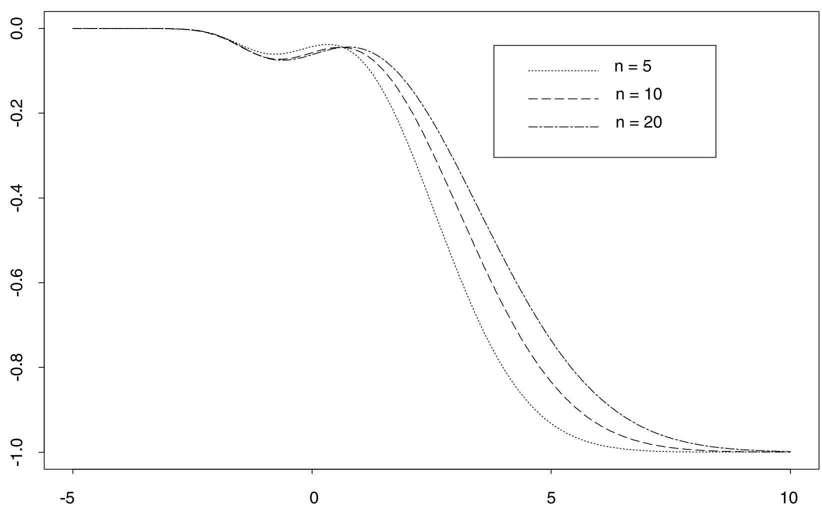


Figure 3.3.30 *Dfs of the normalised maxima of n standard normal rvs and the Gumbel df (top). In the bottom figure the relative error of this approximation for the tail is illustrated. The rate of convergence appears to be very slow.*

$$c_n = (2\ln n)^{-1/2}\,.$$

We conclude that

$$\sqrt{2\ln n}\,\left(M_n - \sqrt{2\ln n} + \frac{\ln\ln n + \ln 4\pi}{2(2\ln n)^{1/2}}\right) \xrightarrow{d} \Lambda\,. \tag{3.40}$$

Note that $c_n \to 0$, i.e. the distribution of M_n becomes less spread around d_n as n increases. □

Similarly, it can be proved that the gamma distribution also belongs to MDA(Λ). The norming constants are given in Table 3.4.4.

Another useful trick to calculate the norming constants is via monotone transformations. If g is an increasing function and $\widetilde{x} = g(x)$, then obviously

$$\widetilde{M}_n = \max\left(\widetilde{X}_1, \ldots, \widetilde{X}_n\right) = g\left(M_n\right)\,.$$

If $X \in \text{MDA}(\Lambda)$ with

$$\lim_{n\to\infty} P\left(M_n \le c_n x + d_n\right) = \Lambda(x)\,, \quad x \in \mathbb{R}\,,$$

then

$$\lim_{n\to\infty} P\left(\widetilde{M}_n \le g\left(c_n x + d_n\right)\right) = \Lambda(x)\,, \quad x \in \mathbb{R}\,.$$

In some cases, g may be expanded in a Taylor series about d_n and just linear terms suffice to give the limit law for $\widetilde{M}_n$, with changed constants $\tilde{c}_n = c_n g'(d_n)$ and $\tilde{d}_n = g(d_n)$. We apply this method to the lognormal distribution.

Example 3.3.31 (Lognormal distribution)
Let X be a standard normal rv and $g(x) = e^{\mu+\sigma x}$, $\mu \in \mathbb{R}$, $\sigma > 0$. Then

$$\widetilde{X} = g(X) = e^{\mu+\sigma X}$$

defines a lognormal rv. Since $X \in \text{MDA}(\Lambda)$ we obtain

$$\lim_{n\to\infty} P\left(\widetilde{M}_n \le e^{\mu+\sigma(c_n x + d_n)}\right) = \Lambda(x)\,, \quad x \in \mathbb{R}\,,$$

where c_n, d_n are the norming constants of the standard normal distribution as calculated in Example 3.3.29. This implies

$$\lim_{n\to\infty} P\left(e^{-\mu-\sigma d_n}\,\widetilde{M}_n \le 1 + \sigma c_n x + o\left(c_n\right)\right) = \Lambda(x)\,, \quad x \in \mathbb{R}\,.$$

Since $c_n \to 0$ it follows that

$$\frac{e^{-\mu-\sigma d_n}}{\sigma c_n}\,\left(\widetilde{M}_n - e^{\mu+\sigma d_n}\right) \xrightarrow{d} \Lambda\,,$$

so that $\widetilde{X} \in \text{MDA}(\Lambda)$ with norming constants

$$\widetilde{c}_n = \sigma c_n e^{\mu+\sigma d_n}\,, \qquad \widetilde{d}_n = e^{\mu+\sigma d_n}\,.$$

Explicit expressions for the norming constants of the lognormal distribution can be found in Table 3.4.4. □

Further Properties of Distributions in MDA(Λ)

In the remainder of this section we collect some further useful facts about distributions in MDA(Λ).

Corollary 3.3.32 (Existence of moments)
Assume that the rv X has df $F \in \mathrm{MDA}(\Lambda)$ with infinite right endpoint. Then $\overline{F} \in \mathcal{R}_{-\infty}$. In particular, $E(X^+)^\alpha < \infty$ for every $\alpha > 0$, where $X^+ = \max(0, X)$.

Proof. Every $F \in \mathrm{MDA}(\Lambda)$ is tail–equivalent to a von Mises function. If $x_F = \infty$, the latter have rapidly varying tails; see Proposition 3.3.24(a), which also implies the statement about the moments; see Theorem A3.12(a). $\square$

In Section 3.3.2 we showed that the maximum domains of attraction of Ψ_α and Φ_α are linked in a natural way. Now we show that MDA(Φ_α) can be embedded in MDA(Λ).

Example 3.3.33 (Embedding MDA(Φ_α) in MDA(Λ))
Let X have df $F \in \mathrm{MDA}(\Phi_\alpha)$ with norming constants c_n. Define

$$X^* = \ln(1 \vee X)$$

with df F^*. By Proposition 3.3.2 and Theorem 3.3.7, $F \in \mathrm{MDA}(\Phi_\alpha)$ if and only if

$$\lim_{n\to\infty} n\overline{F}(c_n x) = \lim_{n\to\infty} \frac{\overline{F}\,(c_n x)}{\overline{F}\,(c_n)} = x^{-\alpha}\,, \quad x > 0\,.$$

This implies that

$$\lim_{n\to\infty} \frac{\overline{F^*}\left(\alpha^{-1}x + \ln c_n\right)}{\overline{F^*}\,(\ln c_n)} = \lim_{n\to\infty} \frac{\overline{F}\left(c_n \exp\left\{\alpha^{-1}x\right\}\right)}{\overline{F}\,(c_n)} = e^{-x}\,, \quad x \in \mathbb{R}\,.$$

Hence $F^* \in \mathrm{MDA}(\Lambda)$ with norming constants $c_n^* = \alpha^{-1}$ and $d_n^* = \ln c_n$. As auxiliary function one can take

$$a^*(x) = \int_x^\infty \frac{\overline{F^*}(y)}{\overline{F^*}(x)}\, dy \qquad \square$$

Example 3.3.34 (Closure of MDA(Λ) under logarithmic transformations)
Let X have df $F \in \mathrm{MDA}(\Lambda)$ with $x_F = \infty$ and norming constants c_n, d_n, chosen according to Theorem 3.3.26. Define X^* and F^* as above. We intend to show that $F^* \in \mathrm{MDA}(\Lambda)$ with norming constants $d_n^* = \ln d_n$ and $c_n^* = c_n/d_n$. Since $a'(x) \to 0$, (3.28) holds, and since $d_n = F^{\leftarrow}(1 - n^{-1}) \to \infty$, it follows that

$$\frac{c_n}{d_n} = \frac{a(d_n)}{d_n} \to 0 .$$

Moreover,

$$\begin{aligned}
\overline{F^*}\,(c_n^* x + d_n^*) &= \overline{F}\left(\exp\left\{\frac{c_n}{d_n}x\right\} d_n\right)\\
&= \overline{F}\left(d_n\left(1+\frac{c_n}{d_n}x + o\left(\frac{c_n}{d_n}\right)\right)\right)\\
&= \overline{F}\,(c_n x + d_n + o\,(c_n))\\
&\sim \overline{F}\,(c_n x + d_n) \quad \sim \quad n^{-1}e^{-x}\,, \quad n \to \infty\,,
\end{aligned}$$

where we applied the uniformity of weak convergence to a continuous limit. The result follows from Proposition 3.3.2. □

Example 3.3.35 (Subexponential distributions and MDA(Λ))
Goldie and Resnick [276] characterise the dfs F that are both subexponential (we write $F \in \mathcal{S}$; see Definition 1.3.3) and in MDA(Λ). Starting from the representation (3.33) for $F \in$ MDA(Λ), they give necessary and sufficient conditions for $F \in \mathcal{S}$. In particular, $\lim_{x\to\infty} a(x) = \infty$ is necessary but not sufficient for $F \in \mathrm{MDA}(\Lambda)\cap\mathcal{S}$. A simple sufficient condition for $F \in \mathrm{MDA}(\Lambda) \cap \mathcal{S}$ is that a is eventually non–decreasing and that there exists some $t > 1$ such that

$$\liminf_{x\to\infty} \frac{a(tx)}{a(x)} > 1\,. \tag{3.41}$$

This condition is easily checked for the following distributions which are all von Mises functions and hence in MDA(Λ):

– *Benktander–type–I*

$$\overline{F}(x) = (1+2(\beta/\alpha)\ln x)\exp\{-(\beta(\ln x)^2+(\alpha+1)\ln x)\}\,, \quad x \ge 1\,, \quad \alpha\,,\beta > 0\,.$$

Here one can choose

$$a(x) = \frac{x}{\alpha + 2\beta \ln x}\,, \quad x \ge 1\,.$$

– *Benktander–type–II*

$$\overline{F}(x) = e^{\alpha/\beta}\, x^{-(1-\beta)} \exp\left\{-\frac{\alpha}{\beta}x^{\beta}\right\}\,, \quad x \ge 1\,, \quad \alpha > 0\,, 0 < \beta < 1\,,$$

with auxiliary function

$$a(x) = \frac{x^{1-\beta}}{\alpha}\,, \quad x \ge 1\,.$$

– *Weibull*

$$\overline{F}(x) = e^{-c\,x^{\tau}}\,, \quad x \geq 0\,, \quad 0 < \tau < 1\,, c > 0\,,$$

with auxiliary function

$$a(x) = c^{-1}\tau^{-1}x^{1-\tau}\,, \quad x \geq 0\,.$$

– *Lognormal*
with auxiliary function

$$a(x) = \frac{\overline{\Phi}(\sigma^{-1}(\ln x - \mu))\sigma x}{\varphi(\sigma^{-1}(\ln x - \mu))} \sim \frac{\sigma^2 x}{\ln x - \mu}\,, \quad x \to \infty\,.$$

The critical cases occur when F is in the tail close to an exponential distribution. For example, let

$$\overline{F}(x) \sim \exp\left\{-x(\ln x)^{\alpha}\right\}\,, \quad x \to \infty\,.$$

For $\alpha < 0$ we have $F \in \mathrm{MDA}(\Lambda) \cap \mathcal{S}$ in view of Theorem 2.7 in Goldie and Resnick [276], whereas for $\alpha \geq 0$, $F \in \mathrm{MDA}(\Lambda)$ but $F \notin \mathcal{S}$; see Example 1.4.3. □

Notes and Comments

There exist many results on the quality of convergence in extreme value limit theory. Topics include the convergence of moments, local limit theory and the convergence of densities, large deviations and uniform rates of convergence. We refer to Chapter 2 of Resnick [530] for a collection of such results.

Statistical methods based on extreme value theory are discussed in detail in Chapter 6. Various estimation methods will depend on an application of the Fisher–Tippett theorem and related results. The quality of those approximations will be crucial.

Figure 3.2.9 suggests a fast rate of convergence in the case of the exponential distribution: already for $n = 5$ the distribution of the normalised maximum is quite close to Λ, while for $n = 50$ they are almost indistinguishable. Indeed, it has been shown by Hall and Wellner [313] that for $F(x) = 1 - e^{-x}$, $x \geq 0$,

$$\sup_{x\in\mathbb{R}} \left| P\left(M_n - \ln n \leq x\right) - \exp\left\{-e^{-x}\right\}\right| \leq n^{-1}\left(2 + n^{-1}\right)\, e^{-2}\,.$$

In contrast to this rapid rate of convergence, the distribution of the normalised maximum of a sample of normal rvs converges extremely slowly to its limit distribution Λ; see Figure 3.3.30. This slow rate of convergence also depends on the particular choice of c_n and d_n. Hall [307] obtained an optimal rate by choosing c_n and d_n as solutions to

$$nc_n\varphi(c_n^{-1}) = 1 \quad \text{and} \quad d_n = c_n^{-1}\,,$$

where φ denotes the standard normal density. Then there exist constants $0 < c < C \leq 3$ such that

$$\frac{c}{\ln n} \leq \sup_{x\in\mathbb{R}} \left|P(M_n \leq c_n x + d_n) - \exp\{-e^{-x}\}\right| \leq \frac{C}{\ln n}\,, \quad n \geq 2\,.$$

Leadbetter, Lindgren and Rootzén [418] and Resnick [530] derive various rates for $F \in \mathrm{MDA}(\Phi_\alpha)$ and $F \in \mathrm{MDA}(\Lambda)$. They also give numerical values for some explicit examples. See also Balkema and de Haan [38], Beirlant and Willekens [58], de Haan and Resnick [301], Goldie and Smith [278], Smith [587] and references therein.

In order to discuss the next point, we introduce a parametric family $(H_\xi)_{\xi\in\mathbb{R}}$ of dfs containing the standard extreme value distributions, namely

$$H_\xi = \begin{cases} \Phi_{1/\xi} & \text{if} \quad \xi > 0, \\ \Lambda & \text{if} \quad \xi = 0, \\ \Psi_{-1/\xi} & \text{if} \quad \xi < 0. \end{cases}$$

The df H_ξ above is referred to as the *generalised extreme value distribution* with parameter ξ; a detailed discussion is given in Section 3.4. The condition $F \in \mathrm{MDA}(H_\xi)$ then yields the so-called *ultimate approximation*

$$F^n\left(c_n x + d_n\right) \approx H_\xi(x)$$

for appropriate norming constants $c_n > 0$ and $d_n \in \mathbb{R}$. One method for improving the rate of convergence in the latter limit relation was already discussed in the classical Fisher–Tippett paper [240]. The basic idea is the following: the parameter ξ can be obtained as a limit. For instance in the Gumbel case $\xi = 0$, F has representation (3.33) with $a'(x) \to 0$ as $x \to \infty$. The *penultimate approximation* now consists of replacing ξ by $\xi_n = a'(d_n)$ leading to the relation

$$F^n\left(c_n x + d_n\right) \approx H_{\xi_n}(x)\,.$$

Typically $\xi_n \neq 0$ so that in the Gumbel case, the penultimate approximation is based on a suitable Weibull ($\xi_n < 0$) or Fréchet ($\xi_n > 0$) approximation. The optimal rate of convergence $O\left((\ln n)^{-1}\right)$ for maxima of iid normal rvs in the ultimate approximation is improved to $O\left((\ln n)^{-2}\right)$ in the penultimate case, as shown in Cohen [128]. In special cases, further improvements can be given. For instance, using expansions of the df of normal maxima, Cohen [129] suggests

$$\Phi^n\left(\frac{x}{b_n}+b_n-\frac{1}{b_n}\right)\approx\Lambda(x)\left(1+e^{-x}\,\frac{x^2}{4\ln n}\right),$$

where $b_n^2\sim\ln n$. Further information is to be found in Joe [357] and Reiss [526].

3.4 The Generalised Extreme Value Distribution and the Generalised Pareto Distribution

In Section 3.2 we have shown that the standard extreme value distributions, together with their types, provide the only non-degenerate limit laws for affinely transformed maxima of iid rvs. As already mentioned in the Notes and Comments of the previous section, a one–parameter representation of the three standard cases in one family of dfs will turn out to be useful. They can be represented by introducing a parameter ξ so that

$\xi=\alpha^{-1}>0$	corresponds to the Fréchet distribution Φ_α ,
$\xi=0$	corresponds to the Gumbel distribution Λ ,
$\xi=-\alpha^{-1}<0$	corresponds to the Weibull distribution Ψ_α .

The following choice is by now widely accepted as *the* standard representation.

Definition 3.4.1 (Jenkinson–von Mises representation of the extreme value distributions: the generalised extreme value distribution (GEV))
Define the df H_ξ by

$$H_\xi(x)=\begin{cases}\exp\left\{-\left(1+\xi\,x\right)^{-1/\xi}\right\} & \text{if}\quad \xi\neq 0\,,\\ \exp\left\{-\exp\{-x\}\right\} & \text{if}\quad \xi=0\,,\end{cases}$$

where $1+\xi\,x>0$.

Table 3.4.2 *Maximum domain of attraction of the Fréchet distribution.*

Fréchet	$\Phi_\alpha(x) = \exp\{-x^{-\alpha}\}\,, \quad x > 0\,, \alpha > 0\,.$
MDA (Φ_α)	$x_F = \infty\,, \qquad \overline{F}(x) = x^{-\alpha}L(x)\,, \quad L \in \mathcal{R}_0\,.$
Norming constants	$c_n = F^{\leftarrow}(1 - n^{-1}) = n^{1/\alpha}L_1(n)\,, \quad L_1 \in \mathcal{R}_0\,, \quad d_n = 0\,.$
Limit result	$c_n^{-1}M_n \xrightarrow{d} \Phi_\alpha$
Examples	
Cauchy	$f(x) = (\pi(1+x^2))^{-1}, \quad x \in \mathbb{R}\,.$ $c_n = n/\pi$
Pareto Burr stable with index $\alpha < 2$	$\overline{F}(x) \sim Kx^{-\alpha}\,, \quad K\,, \alpha > 0\,.$ $c_n = (Kn)^{1/\alpha}$
Loggamma	$f(x) = \dfrac{\alpha^\beta}{\Gamma(\beta)}(\ln x)^{\beta-1}x^{-\alpha-1}\,, \quad x > 1\,, \quad \alpha, \beta > 0\,.$ $c_n = \left((\Gamma(\beta))^{-1}(\ln n)^{\beta-1}n\right)^{1/\alpha}$

Table 3.4.3 *Maximum domain of attraction of the Weibull distribution.*

Weibull	$\Psi_\alpha(x) = \exp\{-(-x)^\alpha\}\,, \quad x < 0\,, \quad \alpha > 0\,.$
MDA (Ψ_α)	$x_F < \infty\,, \quad \overline{F}\left(x_F - x^{-1}\right) = x^{-\alpha} L(x)\,, \quad L \in \mathcal{R}_0\,.$
Norming constants	$c_n = x_F - F^{\leftarrow}(1 - n^{-1}) = n^{-1/\alpha} L_1(n)\,, \quad L_1 \in \mathcal{R}_0\,, \quad d_n = x_F\,.$
Limit result	$c_n^{-1}(M_n - x_F) \xrightarrow{d} \Psi_\alpha$
Examples	
Uniform	$f(x) = 1\,, \quad x \in (0,1)\,.$ $c_n = n^{-1}\,, \quad d_n = 1$
Power law behaviour at x_F	$\overline{F}(x) = K(x_F - x)^\alpha\,, \quad x_F - K^{-1/\alpha} \le x \le x_F\,, \quad K\,,\alpha > 0\,.$ $c_n = (K\,n)^{-1/\alpha}\,, \quad d_n = x_F$
Beta	$f(x) = \dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\, x^{a-1}\,(1-x)^{b-1}\,, \quad 0 < x < 1\,, \quad a\,,b > 0\,.$ $c_n = \left(n\,\dfrac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b+1)}\right)^{-1/b}\,, \quad d_n = 1$

Table 3.4.4 *Maximum domain of attraction of the Gumbel distribution.*

Gumbel	$\Lambda(x) = \exp\left\{-e^{-x}\right\}, \quad x \in \mathbb{R}.$
MDA(Λ)	$x_F \le \infty, \quad \overline{F}(x) = c(x) \exp\left\{-\int_{x_0}^{x} \frac{g(t)}{a(t)}\, dt\right\}, \quad x_0 < x < x_F,$ where $c(x) \to c > 0, \quad g(x) \to 1, \quad a'(x) \to 0 \quad$ as $x \uparrow x_F$.
Norming constants	$d_n = F^{\leftarrow}(1 - n^{-1}), \quad c_n = a(d_n).$
Limit result	$c_n^{-1}(M_n - d_n) \xrightarrow{d} \Lambda$
Examples	
Exponential–like	$\overline{F}(x) \sim K e^{-\lambda x}, \quad K, \lambda > 0.$ $c_n = \lambda^{-1}$ $d_n = \lambda^{-1} \ln(Kn)$
Weibull–like	$\overline{F}(x) \sim K x^{\alpha} \exp\{-c x^{\tau}\}, \quad K, c, \tau > 0, \alpha \in \mathbb{R}.$ $c_n = (c\tau)^{-1} \left(c^{-1} \ln n\right)^{1/\tau - 1}$ $d_n = \left(c^{-1} \ln n\right)^{1/\tau} + \frac{1}{\tau} \left(c^{-1} \ln n\right)^{1/\tau - 1} \left\{\frac{\alpha}{c\tau} \ln(c^{-1} \ln n) + \frac{\ln K}{c}\right\}$

Gamma	$f(x) = \dfrac{\beta^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1}\, e^{-\beta x}\,, \quad x > 0\,, \quad \alpha\,,\beta > 0\,.$ $c_n = \beta^{-1}$ $d_n = \beta^{-1}(\ln n + (\alpha - 1)\ln\ln n - \ln \Gamma(\alpha))$
Normal	$\varphi(x) = \dfrac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\,, \quad x \in \mathbb{R}\,.$ $c_n = (2\ln n)^{-1/2}$ $d_n = \sqrt{2\ln n} - \dfrac{\ln(4\pi) + \ln\ln n}{2(2\ln n)^{1/2}}$
Lognormal	$f(x) = \dfrac{1}{\sqrt{2\pi}\sigma x}\, e^{-(\ln x - \mu)^2/2\sigma^2}\,, \quad x > 0\,, \quad \mu \in \mathbb{R}\,, \sigma > 0\,.$ $c_n = \sigma(2\ln n)^{-1/2} d_n$ $d_n = \exp\left\{\mu + \sigma\left(\sqrt{2\ln n} - \dfrac{\ln(4\pi) + \ln\ln n}{2(2\ln n)^{1/2}}\right)\right\}$

Exponential behaviour at $x_F < \infty$	$\overline{F}(x) = K \exp\left\{-\frac{\alpha}{x_F - x}\right\}, \quad x < x_F, \quad \alpha, K > 0.$ $c_n = \frac{\alpha}{(\ln(Kn))^2}$ $d_n = x_F - \frac{\alpha}{\ln(Kn)}$
Benktander–type–I	$\overline{F}(x) = (1 + 2(\beta/\alpha)\ln x)\exp\{-(\beta(\ln x)^2 + (\alpha+1)\ln x)\}, x > 1, \ \alpha, \beta > 0$ $c_n = \frac{1}{2\sqrt{\beta \ln n}} \exp\left\{-\frac{\alpha+1}{2\beta} + \sqrt{\frac{\ln n}{\beta}}\right\}$ $d_n = \exp\left\{-\frac{\alpha+1}{2\beta} + \sqrt{\frac{\ln n}{\beta}} + \frac{\ln\ln n + \ln(4\beta/\alpha^2) + (\alpha+1)^2/(2\beta)}{4\sqrt{\beta\ln n}}\right\}$
Benktander–type–II	$\overline{F}(x) = x^{-(1-\beta)} \exp\{-\frac{\alpha}{\beta}(x^\beta - 1)\}, \quad x > 1, \quad \alpha > 0, 0 < \beta < 1$ $c_n = \frac{1}{\alpha}\left(\frac{\beta}{\alpha}\ln n\right)^{1/\beta - 1}$ $d_n = \left(\frac{\beta}{\alpha}\ln n\right)^{1/\beta} + \frac{1}{\alpha\beta}\left(\frac{\beta}{\alpha}\ln n\right)^{1/\beta-1}\left\{\alpha - (1-\beta)\left(\ln\ln n + \ln\frac{\beta}{\alpha}\right)\right\}$

Hence the support of H_ξ *corresponds to*

$$\begin{aligned} x &> -\xi^{-1} \quad \textit{for} \quad \xi > 0\,, \\ x &< -\xi^{-1} \quad \textit{for} \quad \xi < 0\,, \\ x &\in \mathbb{R} \qquad \textit{for} \quad \xi = 0\,. \end{aligned}$$

H_ξ *is called a* standard generalised extreme value distribution (GEV). *One can introduce the related location–scale family* $H_{\xi;\mu,\psi}$ *by replacing the argument* x *above by* $(x-\mu)/\psi$ *for* $\mu \in \mathbb{R}\,, \psi > 0$. *The support has to be adjusted accordingly. We also refer to* $H_{\xi;\mu,\psi}$ *as GEV.* □

We consider the df H_0 as the limit of H_ξ for $\xi \to 0$. With this interpretation

$$H_\xi(x) = \exp\left\{ -\,(1+\xi\,x)^{-1/\xi} \right\}\,, \quad 1+\xi\,x > 0\,,$$

serves as a representation for all $\xi \in \mathbb{R}$. The densities of the standard GEV for $\xi = -1, 0, 1$ are shown in Figure 3.2.4.

The GEV provides a convenient unifying representation of the three extreme value types Gumbel, Fréchet and Weibull. Its introduction is mainly motivated by statistical applications; we refer to Chapter 6 where this will become transparent. There GEV fitting will turn out to be one of the fundamental concepts.

The following theorem is one of the basic results in extreme value theory. In a concise analytical way, it gives the essential information collected in the previous section on maximum domains of attraction. Moreover, it constitutes the basis for numerous statistical techniques to be discussed in Chapter 6. First recall the notion of the quantile function $F^{\leftarrow}$ of a df F and define

$$U(t) = F^{\leftarrow}(1-t^{-1})\,, \quad t > 0\,.$$

Theorem 3.4.5 (Characterisation of MDA(H_ξ))
For $\xi \in \mathbb{R}$ *the following assertions are equivalent:*

(*a*) $F \in \mathrm{MDA}(H_\xi)$.

(*b*) *There exists a positive, measurable function* $a(\cdot)$ *such that for* $1+\xi x > 0$,

$$\lim_{u \uparrow x_F} \frac{\overline{F}(u + x a(u))}{\overline{F}(u)} = \begin{cases} (1+\xi x)^{-1/\xi} & \textit{if} \quad \xi \neq 0\,, \\ e^{-x} & \textit{if} \quad \xi = 0\,. \end{cases} \tag{3.42}$$

(*c*) *For* $x, y > 0$, $y \neq 1$,

$$\lim_{s \to \infty} \frac{U(sx) - U(s)}{U(sy) - U(s)} = \begin{cases} \dfrac{x^\xi - 1}{y^\xi - 1} & \textit{if} \quad \xi \neq 0\,, \\[2ex] \dfrac{\ln x}{\ln y} & \textit{if} \quad \xi = 0\,. \end{cases} \tag{3.43}$$

Sketch of the proof. Below we give only the main ideas in order to show that the various conditions enter very naturally. Further details are to be found in the literature, for instance in de Haan [293].

(a)$\Leftrightarrow$(b) For $\xi = 0$ this is Theorem 3.3.27.

For $\xi > 0$ we have $H_\xi(x) = \Phi_\alpha(\alpha^{-1}(x+\alpha))$ for $\alpha = 1/\xi$. By Theorem 3.3.7, (a) is then equivalent to $\overline{F} \in \mathcal{R}_{-\alpha}$. By the representation theorem for regularly varying functions (Theorem A3.3), for some $z > 0$,

$$\overline{F}(x) = c(x) \exp\left\{ - \int_z^x \frac{1}{a(t)}\, dt \right\}, \quad z < x < \infty,$$

where $c(x) \to c > 0$ and $a(x)/x \to \alpha^{-1}$ as $x \to \infty$ locally uniformly. Hence

$$\lim_{u\to\infty} \frac{\overline{F}(u + xa(u))}{\overline{F}(u)} = \left(1 + \frac{x}{\alpha}\right)^{-\alpha},$$

which is (3.42). If (b) holds, choose $d_n = (1/\overline{F})^{\leftarrow}(n) = U(n)$, then

$$1/\overline{F}\,(d_n) \sim n,$$

and with $u = d_n$ in (3.42),

$$\left(1 + \frac{x}{\alpha}\right)^{-\alpha} = \lim_{n\to\infty} \frac{\overline{F}\,(d_n + xa\,(d_n))}{\overline{F}(d_n)} = \lim_{n\to\infty} n\overline{F}\,(d_n + x\,a(d_n)),$$

whence by Proposition 3.3.2, $F \in \mathrm{MDA}(H_\xi)$ for $\xi = \alpha^{-1}$.

The case $\xi < 0$ can be treated similarly.

(b)$\Leftrightarrow$(c) We restrict ourselves to the case $\xi \neq 0$, the proof for $\xi = 0$ being analogous. For simplicity, we assume that F is continuous and increasing on $(-\infty, x_F)$. Set $s = 1/\overline{F}(u)$, then (3.42) translates into

$$A_s(x) = \left(s\overline{F}(U(s) + xa(U(s)))\right)^{-1} \to (1 + \xi x)^{1/\xi}, \quad s \to \infty.$$

Now for every $s > 0$, $A_s(x)$ is decreasing and for $s \to \infty$ converges to a continuous function. Then because of Proposition A1.7, also $A_s^{\leftarrow}(t)$ converges pointwise to the inverse of the corresponding limit function, i.e.

$$\lim_{s\to\infty} \frac{U(st) - U(s)}{a(U(s))} = \frac{t^\xi - 1}{\xi}. \tag{3.44}$$

Now (3.43) is obtained by using (3.44) for $t = x$ and $t = y$ and taking the quotient. The proof of the converse can be given along the same lines. $\square$

Remarks. 1) For $x > 0$, condition (3.42) has an interesting probabilistic interpretation. Indeed, let X be a rv with df $F \in \mathrm{MDA}(H_\xi)$, then (3.42) reformulates as

$$\lim_{u \uparrow x_F} P\left(\frac{X-u}{a(u)} > x \,\middle|\, X > u\right) = \begin{cases} (1+\xi x)^{-1/\xi} & \text{if} \quad \xi \neq 0\,, \\ e^{-x} & \text{if} \quad \xi = 0\,. \end{cases} \tag{3.45}$$

Hence (3.45) gives a distributional approximation for the scaled excesses over the (high) threshold u. The appropriate scaling factor is $a(u)$. This interpretation will turn out to be crucial in many applications; see for instance Section 6.5.

2) In Section 6.4.4 we show how a reformulation of relation (3.43) immediately leads to an estimation procedure for quantiles outside the range of the data. A special case of (3.43) is also used to motivate the Pickands estimator of ξ; see Section 6.4.2.

3) In the proof of Theorem 3.4.5 there is implicitly given the result that $F \in \mathrm{MDA}(H_\xi)$ for some $\xi \in \mathbb{R}$ if and only if there exists some positive function $a_1(\cdot)$ such that

$$\lim_{s\to\infty} \frac{U(st)-U(s)}{a_1(s)} = \frac{t^\xi - 1}{\xi}\,, \quad t > 0\,. \tag{3.46}$$

If $\xi = 0$, the rhs of (3.46) has to be interpreted as $\ln t$. Moreover, for $\xi > 0$, (3.46) is equivalent to

$$\lim_{s\to\infty} \frac{U(st)}{U(s)} = t^\xi\,, \quad t > 0\,,$$

i.e. $U \in \mathcal{R}_\xi$ and hence $a_1(s) \sim \xi U(s)$, $s \to \infty$. For $\xi < 0$, F has finite right endpoint x_F, hence $U(\infty) = x_F < \infty$, and (3.46) is equivalent to

$$U(\infty) - U \in \mathcal{R}_\xi\,.$$

In this case, $a_1(s) \sim -\xi(U(\infty) - U(s))$, $s \to \infty$. The above formulations are for instance to be found as Lemmas 2.1 and 2.2 in Dekkers and de Haan [174]; see de Haan [293] for proofs. The case $\xi = 0$ in (3.46) gives rise to the so–called class of *Π–varying functions*, a strict subclass of $\mathcal{R}_0$. The recognition of the importance of Π–variation for the description of $\mathrm{MDA}(H_0)$ was one of the fundamental contributions to extreme value theory by de Haan [293]. □

In Remark 1 above we used the notion of excess. The following definition makes this precise.

Definition 3.4.6 (Excess distribution function, mean excess function)
Let X be a rv with df F and right endpoint x_F. For a fixed $u < x_F$,

$$F_u(x) = P(X - u \le x \mid X > u)\,, \quad x \ge 0, \tag{3.47}$$

is the excess df of the rv X (of the df F) over the threshold u. *The function*

Pareto	$\dfrac{\kappa+u}{\alpha-1}\,, \quad \alpha>1$
Burr	$\dfrac{u}{\alpha\tau-1}\,\left(1+o(1)\right)\,, \quad \alpha\tau>1$
Loggamma	$\dfrac{u}{\alpha-1}\,\left(1+o(1)\right)\,, \quad \alpha>1$
Lognormal	$\dfrac{\sigma^2 u}{\ln u-\mu}\,(1+o(1))$
Benktander–type–I	$\dfrac{u}{\alpha+2\beta\ln u}$
Benktander–type–II	$\dfrac{u^{1-\beta}}{\alpha}$
Weibull	$\dfrac{u^{1-\tau}}{c\tau}\,(1+o(1))$
Exponential	λ^{-1}
Gamma	$\beta^{-1}\left(1+\dfrac{\alpha-1}{\beta u}+o\left(\dfrac{1}{u}\right)\right)$
Truncated normal	$u^{-1}\,(1+o(1))$

Table 3.4.7 *Mean excess functions for some standard distributions. The parametrisation is taken from Tables 1.2.5 and 1.2.6. The asymptotic relations are to be understood for* $u \to \infty$.

$$e(u) = E(X-u \mid X>u)$$

is called the mean excess function of X. $\square$

Excesses over thresholds play a fundamental role in many fields. Different names arise from specific applications. For instance, F_u is known as the *excess–life* or *residual lifetime df* in reliability theory and medical statistics. In an insurance context, F_u is usually referred to as the *excess–of–loss df*. For a detailed discusssion of some basic properties and statistical applications of the mean excess function and the excess df, we refer to Sections 6.2.2 and 6.5.

Example 3.4.8 (Calculation of the mean excess function)
Using the definition of $e(u)$ and partial integration, the following formulae

are easily checked. They are useful for calculating the mean excess function in special cases. Suppose for ease of presentation that X is a positive rv with df F and finite expectation; trivial changes allow for support (x_0, ∞) for some $x_0 > 0$. Then

$$\begin{aligned} e(u) &= \int_u^{x_F} (x-u)dF(x)/\overline{F}(u) \\ &= \frac{1}{\overline{F}(u)} \int_u^{x_F} \overline{F}(x)\, dx\,, \quad 0 < u < x_F\,. \end{aligned} \tag{3.48}$$

Whenever F is continuous,

$$\overline{F}(x) = \frac{e(0)}{e(x)} \exp\left\{ - \int_0^x \frac{1}{e(u)}\, du \right\}, \quad x > 0\,. \tag{3.49}$$

It immediately follows from (3.49) that a continuous df is uniquely determined by its mean excess function. If, as in many cases of practical interest, $\overline{F} \in \mathcal{R}_{-\alpha}$ for some $\alpha > 1$, then an immediate application of Karamata's theorem (Theorem A3.6) yields $e(u) \sim u/(\alpha - 1)$ as $u \to \infty$. In Table 3.4.7 the mean excess functions of some standard distributions are summarised. □

The appearance of the rhs limit in (3.45) motivates the following definition.

Definition 3.4.9 (The generalised Pareto distribution (GPD))
Define the df G_ξ by

$$G_\xi(x) = \begin{cases} 1 - (1+\xi x)^{-1/\xi} & if \quad \xi \neq 0\,, \\ 1 - e^{-x} & if \quad \xi = 0\,, \end{cases} \tag{3.50}$$

where

$$\begin{aligned} x \geq 0 & \qquad if \quad \xi \geq 0\,, \\ 0 \leq x \leq -1/\xi & \qquad if \quad \xi < 0\,. \end{aligned}$$

G_ξ *is called a* standard generalised Pareto distribution (GPD). *One can introduce the related location–scale family $G_{\xi;\nu,\beta}$ by replacing the argument x above by $(x-\nu)/\beta$ for $\nu \in \mathbb{R}$, $\beta > 0$. The support has to be adjusted accordingly. We also refer to $G_{\xi;\nu,\beta}$ as GPD.* □

As in the case of H_0, G_0 can also be interpreted as the limit of G_ξ as $\xi \to 0$. The df $G_{\xi;0,\beta}$ plays an important role in Section 6.5. To economise on notation, we will denote

$$G_{\xi,\beta}(x) = 1 - \left(1 + \xi \frac{x}{\beta}\right)^{-1/\xi}, \quad x \in D(\xi, \beta)\,, \tag{3.51}$$

where

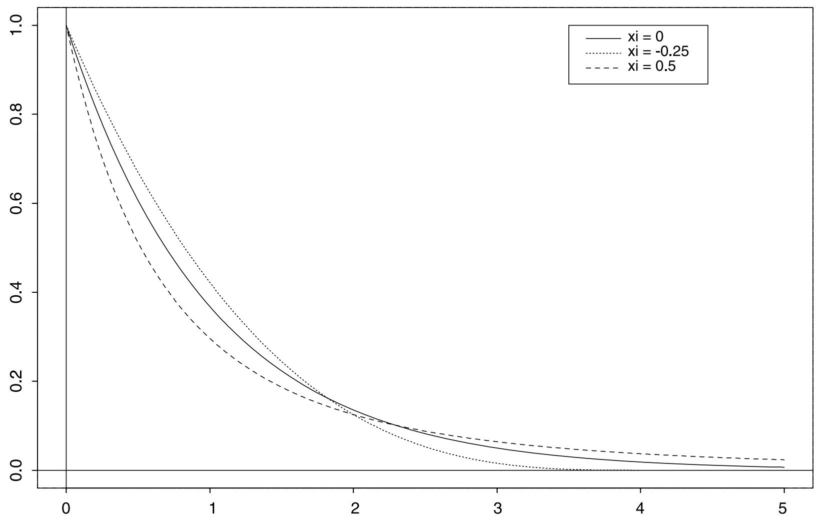


Figure 3.4.10 *Densities of the GPD for different parameters ξ and $\beta = 1$.*

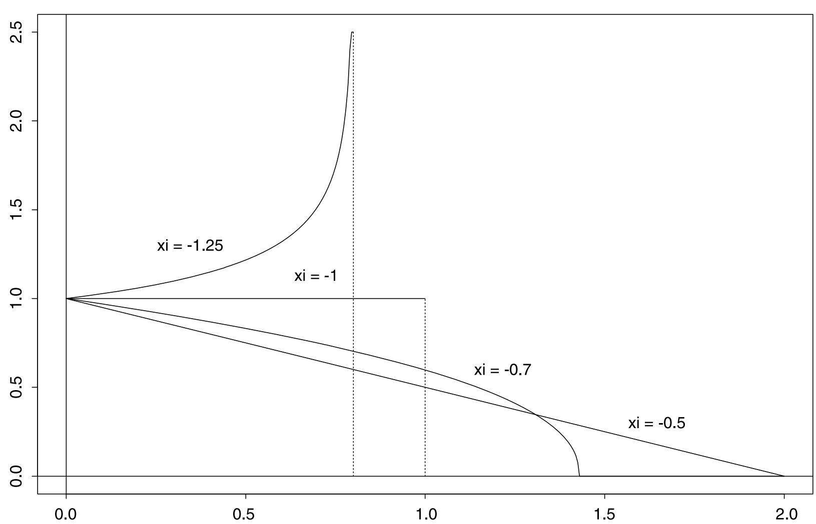


Figure 3.4.11 *Densities of the GPD for $\xi < 0$ and $\beta = 1$. Recall that they have compact support $[0, -1/\xi]$.*

$$x \in D(\xi,\beta) = \begin{cases} [0,\infty) & if \quad \xi \ge 0\,, \\ [0,-\beta/\xi] & if \quad \xi < 0\,. \end{cases}$$

Whenever we say that X has a GPD with parameters ξ and β, it is understood that X has df $G_{\xi,\beta}$.

Time to summarise:

The GEV

$$H_\xi\,, \quad \xi \in \mathbb{R}\,,$$

describes the limit distributions of normalised maxima.

The GPD

$$G_{\xi,\beta}\,, \quad \xi \in \mathbb{R}\,,\, \beta > 0\,,$$

appears as the limit distribution of scaled excesses over high thresholds.

GPD fitting is one of the most useful concepts in the statistics of extremal events; see Section 6.5. Here we collect some basic probabilistic properties of the GPD.

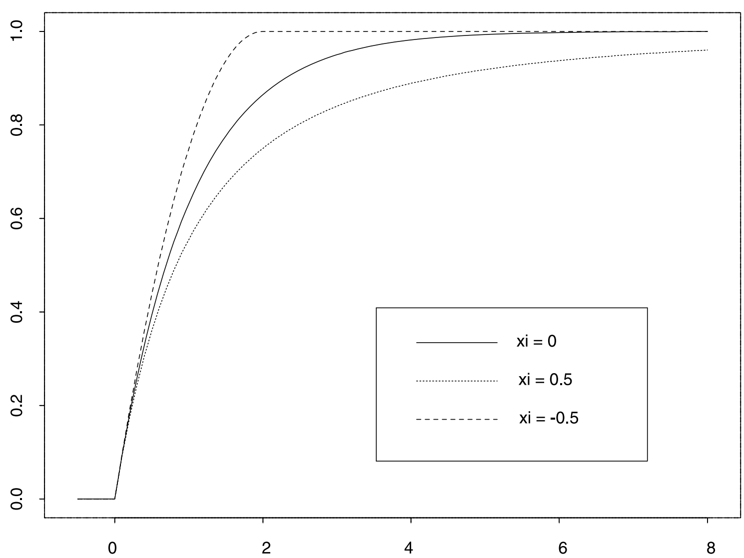


Figure 3.4.12 *GPD for different parameters ξ and $\beta = 1$.*

Theorem 3.4.13 (Properties of GPD)

(*a*) *Suppose X has a GPD with parameters ξ and β. Then $EX < \infty$ if and only if $\xi < 1$. In the latter case*

$$\begin{aligned} E\left(1+\frac{\xi}{\beta}X\right)^{-r} &= \frac{1}{1+\xi r}, \quad r > -1/\xi, \\ E\left(\ln\left(1+\frac{\xi}{\beta}X\right)\right)^{k} &= \xi^k\, k!, \quad k \in \mathbb{N}, \\ EX\left(\overline{G}_{\xi,\beta}(X)\right)^{r} &= \frac{\beta}{(r+1-\xi)(r+1)}, \quad (r+1)/|\xi| > 0. \end{aligned}$$

If $\xi < 1/r$ with $r \in \mathbb{N}$, then

$$EX^r = \frac{\beta^r}{\xi^{r+1}} \frac{\Gamma(\xi^{-1}-r)}{\Gamma(1+\xi^{-1})}\, r!\,.$$

(*b*) *For every $\xi \in \mathbb{R}$, $F \in \mathrm{MDA}(H_\xi)$ if and only if*

$$\lim_{u\uparrow x_F} \sup_{0<x<x_F-u} |F_u(x) - G_{\xi,\beta(u)}(x)| = 0 \tag{3.52}$$

for some positive function β.

(*c*) *Suppose $x_i \in D(\xi,\beta), i = 1, 2,$ then*

$$\frac{\overline{G}_{\xi,\beta}(x_1+x_2)}{\overline{G}_{\xi,\beta}(x_1)} = \overline{G}_{\xi,\beta+\xi x_1}(x_2)\,. \tag{3.53}$$

(*d*) *Assume that N is Poi(λ), independent of the iid sequence (X_n) with a GPD with parameters ξ and β. Write $M_N = \max(X_1,\dots,X_N)$. Then*

$$P\left(M_N \le x\right) = \exp\left\{-\lambda\left(1+\xi\frac{x}{\beta}\right)^{-1/\xi}\right\} = H_{\xi;\mu,\psi}(x)\,,$$

where $\mu = \beta\xi^{-1}(\lambda^{\xi}-1)$ and $\psi = \beta\lambda^{\xi}$.

(*e*) *Suppose X has GPD with parameters $\xi < 1$ and β. Then for $u < x_F$,*

$$e(u) = E(X-u \,|\, X > u) = \frac{\beta+\xi u}{1-\xi}, \quad \beta + u\xi > 0\,.$$

Proof. (a) and (c) follow by direct verification.

(b) In Theorem 3.4.5 (see Remark 1) we have already proved that $F \in \mathrm{MDA}(H_\xi)$ implies $F_u(\beta(u)x) \to G_{\xi,1}(x)$ for $x \in D(\xi,1)$. Because the limiting GPD is continuous, the uniformity of the convergence follows (see Appendix A1.1) which yields relation (3.52). Similarly, (3.52) yields (3.42) for $x \in D(\xi,1)$. The argument in Balkema and de Haan [37], p. 799, shows that

(3.42) can be extended to x satisfying $1+\xi x>0$. We conclude from Theorem 3.4.5 that $F \in \mathrm{MDA}(H_\xi)$.

(d) One immediately obtains

$$\begin{aligned} P(M_N \le x) &= \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} G_{\xi,\beta}^n(x) \\ &= \exp\{-\lambda \overline{G}_{\xi,\beta}(x)\} \\ &= \exp\left\{-\lambda\left(1+\xi \frac{x}{\beta}\right)^{-1/\xi}\right\} \\ &= \exp\left\{-\left(1+\xi \frac{x-\xi^{-1}\beta(\lambda^\xi-1)}{\beta\lambda^\xi}\right)^{-1/\xi}\right\}, \quad \xi \ne 0\,. \end{aligned}$$

The case $\xi = 0$ reduces to

$$P(M_N \le x) = \exp\left\{-e^{-(x-\beta\ln\lambda)/\beta}\right\}.$$

(e) This result immediately follows from the representation (3.48). □

Remarks. 4) Theorem 3.4.13 summarises various properties which are essential for the special role of the GPD in the statistical analysis of extremes. This will be made clear in Section 6.5.

5) The property (c) above is often reformulated as follows: the class of GPDs is closed with respect to changes of the threshold. Indeed the lhs in (3.53) is the conditional probability that, given our underlying rv exceeds x_1, it also exceeds the threshold $x_1 + x_2$. The rhs states that this probability is again of the generalised Pareto type. This closure property is important in reinsurance, where the GPDs are basic when treating excess–of–loss contracts; see for instance Conti [132]. In combination with property (d) it is also crucial for stop–loss treaties. For a discussion on different types of reinsurance treaties, see Section 8.7.

6) Property (b) above suggests a GPD as appropriate approximation of the excess df F_u for large u. This result goes back to Pickands [498] and is often formulated as follows. For some function β to be estimated from the data,

$$\overline{F}_u(x) = P(X-u>x \mid X>u) \approx \overline{G}_{\xi,\beta(u)}(x)\,, \quad x>0\,.$$

Alternatively one considers for $x>u$,

$$P(X>x \mid X>u) \approx \overline{G}_{\xi;u,\beta(u)}(x)\,.$$

In both cases u has to be taken sufficiently large. See Section 6.5 for more details. Together (b) and (e) give us a nice graphical technique for choosing the threshold u so high that an approximation of the excess df F_u by a GPD is justified: given an iid sample $X_1, \ldots, X_n$, construct the empirical mean excess function $e_n(u)$ as sample version of the mean excess function $e(u)$. From (e) we have that the mean excess function of a GPD is linear, hence check for a u–region where the graph of $e_n(u)$ becomes roughly linear. For such u an approximation of F_u by a GPD seems reasonable. In Section 6.5 we will use this approach for fitting excesses over high thresholds.

7) From Proposition 3.1.1 (see also the succeeding Remark 1) we have learnt that the number of the exceedances of a high threshold is roughly Poisson. From Remark 6 we conclude that an approximation of the excess df F_u by a GPD may be justified. Moreover, it can be shown that the number of exceedances and the excesses are independent in an asymptotic sense; see Leadbetter [417].

8) Property (d) now says that in a model, where the number of exceedances is exactly Poisson and the excess df is an exact GDP, the maximum of these excesses has an exact GEV. □

The above remarks suggest the following approximate model for the exceedance times and the excesses of an iid sample:

– The number of exceedances of a high threshold follows a Poisson process.
– Excesses over high thresholds can be modelled by a GPD.
– An appropriate value of the high threshold can be found by plotting the empirical mean excess function.
– The distribution of the maximum of a Poisson number of iid excesses over a high threshold is a GEV.

In interpreting the above summary, do look at the precise formulation of the underlying theorems. If at this point you want to see some of the above in action; see for instance Smith [595], p. 461. Alternatively, consult the examples in Section 6.5

Notes and Comments

In this section we summarised some of the probabilistic properties of the GEV and the GPD. They are crucial for the statistical analysis of extremal events as provided in Chapter 6. The GEV will be used for statistical inference of data which occur as iid maxima of certain time series, for instance annual maxima of rainfall, windspeed etc. Theorem 3.4.5 opens the way to tail and high quantile estimation. Part (b) of this theorem leads immediately to the definition of the GPD, which goes back to Pickands [498]. An approximation of the excess df by the GPD has been suggested by hydrologists under the name *peaks over threshold* method; see Section 6.5 for a detailed discussion. Weak limit theory for excess dfs originates from Balkema and de Haan [37]. Richard Smith and his collaborators have further developed the theory and applications of the GPD in various fields. Basic properties of the GPD can for instance be found in Smith [591]. Detailed discussions on the use of the mean excess function in insurance are to be found in Beirlant et al. [57] and Hogg and Klugman [330].

3.5 Almost Sure Behaviour of Maxima

In this section we study the a.s. behaviour of the maxima

$$M_1 = X_1\,,\quad M_n = \max\left(X_1,\ldots,X_n\right)\,,\quad n\geq 2\,,$$

for an iid sequence $X, X_1, X_2, \ldots$ with common non–degenerate df F.

At the beginning we might ask:

What kind of results can we expect?
Is there, for example, a general theorem like the SLLN for iid sums?

The answer to the latter question is, unfortunately, negative. We have already found in the previous sections that the weak limit theory for (M_n) is very sensitive with respect to the tails $\overline{F}(x) = P(X > x)$. The same applies to the a.s. behaviour.

We first study the probabilities (i.o. stands for "infinitely often")

$$P\left(M_n > u_n \quad \text{i.o.}\right) \quad \text{and} \quad P\left(M_n \leq u_n \quad \text{i.o.}\right)$$

for a non–decreasing sequence (u_n) of real numbers. We will fully characterise these probabilities in terms of the tails $\overline{F}(u_n)$.

We start with $P(M_n > u_n \text{ i.o.})$ which is not difficult to handle with the classical Borel–Cantelli lemmas. Recall from any textbook on probability theory that, for a sequence of events (A_n), $\{A_n \text{ i.o.}\}$ stands for

$$\limsup_{n\to\infty} A_n = \bigcap_{k=1}^{\infty} \bigcup_{n\geq k} A_n\,.$$

The standard Borel–Cantelli lemma states that

$$\sum_{n=1}^{\infty} P(A_n) < \infty \quad \text{implies} \quad P(A_n \quad \text{i.o.}) = 0\,.$$

Its partial converse for independent A_n tells us that

$$\sum_{n=1}^{\infty} P(A_n) = \infty \quad \text{implies} \quad P(A_n \quad \text{i.o.}) = 1\,.$$

A version of the following result for general independent rvs can be found in Galambos [249], Theorem 4.2.1.

Theorem 3.5.1 (Characterisation of the maximal a.s. growth of partial maxima)
Suppose that (u_n) is non–decreasing. Then

$$P(M_n > u_n \quad \text{i.o.}) = P(X_n > u_n \quad \text{i.o.})\,. \tag{3.54}$$

In particular,

$$P(M_n > u_n \quad \text{i.o.}) = 0 \quad \textit{or} \quad = 1$$

according as

$$\sum_{n=1}^{\infty} P(X > u_n) < \infty \quad \textit{or} \quad = \infty\,. \tag{3.55}$$

Notice that the second statement of the theorem is an immediate consequence of (3.54). Indeed, by the Borel–Cantelli lemma and its partial converse for independent events, $P(X_n > u_n \text{ i.o.}) = 0$ or $= 1$ according as (3.55) holds.

Proof. It suffices to show that (3.54) holds. Since $M_n \geq X_n$ for all n we need only to show that

$$P(M_n > u_n \quad \text{i.o.}) \leq P(X_n > u_n \quad \text{i.o.})\,. \tag{3.56}$$

Let x_F denote the right endpoint of the distribution F and suppose that $u_n \geq x_F$ for some n. Then

$$P(M_n > u_n) = P(X_n > u_n) = 0$$

for all large n, hence (3.56) is satisfied. Therefore assume that $u_n < x_F$ for all n. Then

$$F(u_n) < 1 \quad \text{for all} \quad n\,.$$

If $u_n \uparrow x_F$ and $M_n > u_n$ for infinitely many n then it is not difficult to see that there exist infinitely many n with the property $X_n > u_n$.
Now suppose that $u_n \uparrow b < x_F$. But then

$$\overline{F}(u_n) \geq \overline{F}(b) > 0\,.$$

By the converse Borel–Cantelli lemma, $1 = P(X_n > u_n \text{ i.o.})$, and then (3.54) necessarily holds. $\square$

The determination of the probabilities $P(M_n \le u_n \text{ i.o.})$ is quite tricky. The following is a final improvement, due to Klass [383, 384], of a result of Barndorff–Nielsen [44, 45]. For the history of this problem see Galambos [249] and Klass [383].

Theorem 3.5.2 (Characterisation of the minimal a.s. growth of partial maxima)
Suppose that (u_n) is non–decreasing and that the following conditions hold:

$$\overline{F}(u_n) \quad \to \quad 0\,, \tag{3.57}$$

$$n\overline{F}(u_n) \quad \to \quad \infty\,. \tag{3.58}$$

Then

$$P\left(M_n \le u_n \quad \text{i.o.}\right) = 0 \quad \text{or} \quad = 1$$

according as

$$\sum_{n=1}^{\infty} \overline{F}(u_n) \exp\left\{-n\overline{F}(u_n)\right\} < \infty \quad \text{or} \quad = \infty\,. \tag{3.59}$$

Moreover, if

$$\overline{F}(u_n) \to c > 0\,, \quad \text{then} \quad P\left(M_n \le u_n \quad \text{i.o.}\right) = 0\,,$$

while if

$$\liminf_{n\to\infty} n\overline{F}(u_n) < \infty\,, \qquad \text{then} \quad P(M_n \le u_n \quad \text{i.o.}) = 1\,.$$

Remarks. 1) Conditions (3.57) and (3.58) are natural; (3.57) just means that one of the following conditions holds: $u_n \uparrow \infty$ or $u_n \uparrow x_F$ for a df F continuous at its right endpoint x_F or $u_n \uparrow b > x_F$. From Proposition 3.1.1 we know that (3.58) is equivalent to $P(M_n \le u_n) \to 0$.

2) Condition (3.59) is, at a first glance, a little bit mysterious, but from the proof below its meaning becomes more transparent. But already notice that $\overline{F}(u_n)\exp\{-n\overline{F}(u_n)\}$ is close to the probability

$$P(M_n \le u_n, X_{n+1} > u_n) = P(M_n \le u_n, M_{n+1} > u_n)\,. \qquad \square$$

Sketch of the proof. We first deal with the case that $\overline{F}(u_n) \to c > 0$. Then

$$
\begin{aligned}
P\left(M_n \le u_n \quad \text{i.o.}\right) &= P\left(\bigcap_{i=1}^{\infty} \bigcup_{n \ge i} \{M_n \le u_n\}\right) \\
&= \lim_{i\to\infty} P\left(\bigcup_{n\ge i} \{M_n \le u_n\}\right) \\
&\le \lim_{i\to\infty} \sum_{n\ge i} P\left(M_n \le u_n\right) \\
&= \lim_{i\to\infty} \sum_{n\ge i} F^n\left(u_n\right) \\
&= 0\,,
\end{aligned}
$$

since $F(u_n) < 1 - c + \epsilon < 1$ for a small ϵ and sufficiently large n .

Next suppose that $\liminf_{n\to\infty} n\overline{F}(u_n) < \infty$. Then

$$
\begin{aligned}
P\left(M_n \le u_n \quad \text{i.o.}\right) &= \lim_{i\to\infty} P\left(\bigcup_{n\ge i} \{M_n \le u_n\}\right) \\
&\ge \limsup_{i\to\infty} P\left(M_i \le u_i\right) \\
&= \limsup_{i\to\infty} \ \exp\left\{i \ln\left(1 - \overline{F}\left(u_i\right)\right)\right\} \\
&= \limsup_{i\to\infty} \ \exp\left\{-i\overline{F}\left(u_i\right)\left(1 + o(1)\right)\right\} \\
&= \exp\left\{-\liminf_{i\to\infty} i\overline{F}\left(u_i\right)\right\} \\
&> 0\,.
\end{aligned}
$$

This and an application of the Hewitt–Savage 0–1 law prove that $P(M_n \le u_n \text{ i.o.}) = 1$.

Now we come to the main part of the theorem. We restrict ourselves to showing that the condition

$$
\sum_{n=1}^{\infty} \overline{F}\left(u_n\right) \exp\left\{-n\overline{F}\left(u_n\right)\right\} < \infty \tag{3.60}
$$

implies that

$$
P\left(M_n \le u_n \quad \text{i.o.}\right) = 0\,.
$$

Suppose that (3.60), (3.57) and (3.58) hold. We use a standard blocking technique for proving a.s. convergence results. Define the subsequence (n_k) as follows

$$n_1 = 1\,, \quad n_{k+1} = \min\left\{ j > n_k :\ (j - n_k)\,\overline{F}\,(u_{n_k}) \geq 1 \right\}\,.$$

This implies in particular that

$$(n_{k+1} - n_k)\,\overline{F}\,(u_{n_k}) \to 1\,. \tag{3.61}$$

Moreover, by (3.61), $n_{k+1}/n_k \to 1$. Hence there exists k_0 such that

$$n_j \overline{F}(u_{n_{j+1}}) \geq 1 \quad \text{for} \quad j \geq k_0\,.$$

Note also that the function $f(y) = y \exp\{-jy\}$ decreases for $y \geq j^{-1}$. Hence, for all $k \geq k_0$,

$$\begin{aligned}
&\sum_{n_k \leq j < n_{k+1}} \overline{F}\,(u_j) \exp\left\{ -j\overline{F}\,(u_j) \right\} \\
&\geq \sum_{n_k \leq j < n_{k+1}} \overline{F}\,(u_{n_k}) \exp\left\{ -j\overline{F}\,(u_{n_k}) \right\} \\
&\geq e^{-1}\,(n_{k+1} - n_k)\,\overline{F}\,(u_{n_k}) \exp\left\{ -n_k\overline{F}\,(u_{n_k}) \right\} \\
&\geq e^{-1} \exp\left\{ -n_k\overline{F}\,(u_{n_k}) \right\}\,.
\end{aligned}$$

Thus (3.60) implies that

$$\sum_{k=1}^{\infty} \exp\left\{ -n_k\overline{F}\,(u_{n_k}) \right\} < \infty\,. \tag{3.62}$$

(It can as it happens be shown that (3.60) is equivalent to (3.62).) Notice that

$$\begin{aligned}
P\,(M_n \leq u_n \quad \text{i.o.}) &= P\left(\bigcap_{k=1}^{\infty} \bigcup_{n \geq k} \{M_n \leq u_n\} \right) \\
&= \lim_{k \to \infty} P\left(\bigcup_{n \geq k} \{M_n \leq u_n\} \right) \\
&= \lim_{l \to \infty} P\left(\bigcup_{k \geq l} \bigcup_{n_k \leq j < n_{k+1}} \{M_j \leq u_j\} \right)
\end{aligned}$$

$$\leq \lim_{l\to\infty}\sum_{k\geq l} P\left(\bigcup_{n_k\leq j<n_{k+1}}\{M_j\leq u_j\}\right)$$

$$\leq \lim_{l\to\infty}\sum_{k\geq l} P\left(M_{n_k}\leq u_{n_{k+1}}\right)$$

$$\leq \lim_{l\to\infty}\sum_{k\geq l} \frac{P\left(M_{n_{k+1}}\leq u_{n_{k+1}}\right)}{P\left(M_{n_{k+1}-n_k}\leq u_{n_{k+1}}\right)}\,. \tag{3.63}$$

By construction of the n_k and by property (3.61),

$$\begin{aligned} P\left(M_{n_{k+1}-n_k}\leq u_{n_{k+1}}\right) &= \exp\left\{-(n_{k+1}-n_k)\overline{F}\left(u_{n_{k+1}}\right)(1+o(1))\right\} \\ &\geq \exp\left\{-(n_{k+1}-n_k)\overline{F}\left(u_{n_k}\right)(1+o(1))\right\} \\ &\to e^{-1}\,. \end{aligned}$$

This together with (3.63) and (3.62) yields $P(M_n\leq u_n \text{ i.o.})=0$.

The proof of the remaining part of the theorem is very technical. We refer to Klass [383, 384] for details. □

Recall (e.g. from Petrov [495, 496]) that the relation

$$\limsup_{n\to\infty} c_n^{-1}\left(M_n-d_n\right)=1 \quad \text{a.s.}$$

for $c_n>0$ and $d_n\in\mathbb{R}$ just means that

$$P\left(M_n>c_n(1+\epsilon)+d_n \quad \text{i.o.}\right)=0 \quad \text{or} \quad =1$$

according as $\epsilon>0$ or $\epsilon<0$ for small $|\epsilon|$. Similarly,

$$\liminf_{n\to\infty} c_n^{-1}\left(M_n-d_n\right)=1 \quad \text{a.s.}$$

holds if and only if

$$P\left(M_n\leq c_n(1+\epsilon)+d_n \quad \text{i.o.}\right)=0 \quad \text{or} \quad =1$$

according as $\epsilon<0$ or $\epsilon>0$ for small $|\epsilon|$. Then the following is immediate from Theorems 3.5.1 and 3.5.2.

Corollary 3.5.3 (Characterisation of the upper and lower limits of maxima)

(a) *Assume that the sequences $u_n(\epsilon)=c_n(1+\epsilon)+d_n$, $n\in\mathbb{N}$, are non–-decreasing for every $\epsilon\in(-\epsilon_0,\epsilon_0)$. Then the relation*

$$\sum_{n=1}^{\infty}\overline{F}\left(u_n(\epsilon)\right)<\infty \quad \textit{or} \quad =\infty$$

according as $\epsilon \in (0, \epsilon_0)$ *or* $\epsilon \in (-\epsilon_0, 0)$ *implies that*

$$\limsup_{n\to\infty} c_n^{-1}(M_n - d_n) = 1 \quad \text{a.s.}$$

(*b*) *Assume that the sequences* $u_n(\epsilon) = c_n(1+\epsilon) + d_n$, $n \in \mathbb{N}$, *are non–decreasing and satisfy* (3.57), (3.58) *for every* $\epsilon \in (-\epsilon_0, \epsilon_0)$. *Then the relation*

$$\sum_{n=1}^{\infty} \overline{F}(u_n(\epsilon)) \exp\left\{-n\overline{F}(u_n(\epsilon))\right\} < \infty \quad \textit{or} \quad = \infty$$

according as $\epsilon \in (-\epsilon_0, 0)$ *or* $\epsilon \in (0, \epsilon_0)$ *implies that*

$$\liminf_{n\to\infty} c_n^{-1}(M_n - d_n) = 1 \quad \text{a.s.} \qquad \square$$

We continue with several examples in order to illustrate the different options for the a.s. behaviour of the maxima M_n. Throughout we will use the following notation

$$\ln_0 x = x, \quad \ln_1 x = \max(0, \ln x), \quad \ln_k x = \max(0, \ln_{k-1} x), \quad k \geq 2,$$

i.e. $\ln_k x$ is the *kth iterated logarithm of* x.

Example 3.5.4 (Normal distribution, continuation of Example 3.3.29) Assume that $F = \Phi$ is the standard normal distribution. Then

$$\overline{\Phi}(x) \sim \frac{1}{\sqrt{2\pi}x} \exp\{-x^2/2\}. \tag{3.64}$$

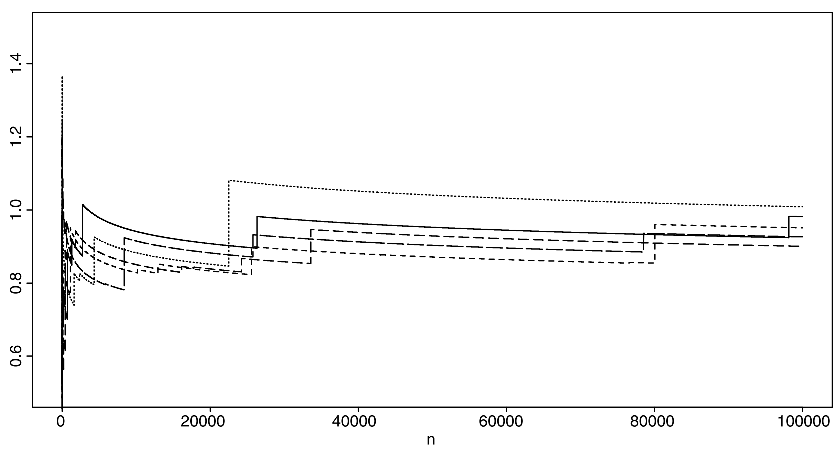


Figure 3.5.5 *Five sample paths of* $(M_n/\sqrt{2\ln n})$ *for* 100 000 *realisations of iid standard normal rvs. The rate of a.s. convergence to* 1 *appears to be very slow.*

From (3.40) we conclude

$$\frac{M_n}{\sqrt{2\ln n}} \xrightarrow{P} 1\,, \quad n \to \infty\,. \tag{3.65}$$

We are interested in a.s. refinements of this result.
Choose

$$u_n(\epsilon) = \sqrt{2\ln\left(\frac{(\ln_0 n \cdots \ln_r n)\ln_r^{\epsilon} n}{\sqrt{\ln n}}\right)}\,, \quad r \ge 0\,.$$

An application of Theorem 3.5.1 together with (3.64) yields

$$P\left(M_n > u_n(\epsilon) \quad \text{i.o.}\right) = 0 \quad \text{or} \quad = 1$$

according as $\epsilon > 0$ or $\epsilon < 0$ for small $|\epsilon|$ and hence, by Corollary 3.5.3,

$$\limsup_{n\to\infty} \frac{M_n}{\sqrt{2\ln n}} = 1 \quad \text{a.s.} \tag{3.66}$$

This result can further be refined. For example, notice that

$$\begin{aligned}
& P\left(M_n > \sqrt{2\ln\left(\frac{n\ln^{1+\epsilon} n}{\sqrt{\ln n}}\right)} \quad \text{i.o.}\right) \\
= \quad & P\left(M_n - \sqrt{2\ln\left(\frac{n}{\sqrt{\ln n}}\right)}\right. \\
& \qquad \left. > \sqrt{2\ln\left(\frac{n\ln^{1+\epsilon} n}{\sqrt{\ln n}}\right)} - \sqrt{2\ln\left(\frac{n}{\sqrt{\ln n}}\right)} \quad \text{i.o.}\right) \\
= \quad & 0 \quad \text{or} \quad = 1
\end{aligned}$$

according as $\epsilon > 0$ or $\epsilon \le 0$. By the mean value theorem, for small $|\epsilon|$ and certain $\theta_n \in (0,1)$,

$$\begin{aligned}
& P\left(M_n - \sqrt{2\ln\left(\frac{n}{\sqrt{\ln n}}\right)} > \frac{1}{2}\,\frac{2(1+\epsilon)\ln_2 n}{\left(2\ln(n/\sqrt{\ln n}) + \theta_n 2(1+\epsilon)\ln_2 n\right)^{1/2}} \quad \text{i.o.}\right) \\
& = 0 \quad \text{or} \quad = 1
\end{aligned}$$

according as $\epsilon > 0$ or $\epsilon \le 0$. In view of Corollary 3.5.3 this just means that

$$\limsup_{n\to\infty} \frac{\sqrt{2\ln n}}{\ln_2 n}\left(M_n - \sqrt{2\ln\left(\frac{n}{\sqrt{\ln n}}\right)}\right) = 1 \quad \text{a.s.}$$

By the same arguments,

$$\limsup_{n\to\infty} \frac{\sqrt{2\ln n}}{\ln_{r+1} n}\left(M_n - \sqrt{2\ln\left(\frac{\ln_0 n \cdots \ln_{r-1} n}{\sqrt{\ln n}}\right)}\right) = 1 \quad \text{a.s.}, \quad r \ge 1\,.$$

Now choose

$$u_n'(\epsilon) = \sqrt{2\ln\left(\frac{n}{\sqrt{4\pi \ln n}\,\ln((\ln_1 n \cdots \ln_r n)\ln_r^{\epsilon} n)}\right)}, \quad r \ge 1\,.$$

An application of Theorem 3.5.2 yields that

$$P\left(M_n \le u_n'(\epsilon) \quad \text{i.o.}\right) = 0 \quad \text{or} \quad = 1$$

according as $\epsilon > 0$ or $\epsilon < 0$ for small $|\epsilon|$. In particular, we may conclude that

$$\liminf_{n\to\infty} \frac{M_n}{\sqrt{2\ln n}} = 1 \quad \text{a.s.} \tag{3.67}$$

which, together with (3.66), yields the a.s. analogue to (3.65):

$$\lim_{n\to\infty} \frac{M_n}{\sqrt{2\ln n}} = 1 \quad \text{a.s.}$$

Refinements of relation (3.67) are possible in the same way as for lim sup. □

Example 3.5.6 (Exponential distribution, continuation of Example 3.2.7)
Let X be standard exponential with tail $\overline{F}(x) = e^{-x}$, $x \ge 0$. From Example 3.2.7 we conclude that

$$\frac{M_n}{\ln n} \xrightarrow{P} 1\,, \quad n \to \infty\,. \tag{3.68}$$

We are interested in a.s. refinements of this result.
Choose

$$u_n(\epsilon) = \ln\left((\ln_0 n \ln_1 n \cdots \ln_r n)\ln_r^{\epsilon} n\right)\,, \quad r \ge 0\,.$$

Then, for large n and small $|\epsilon|$,

$$\overline{F}\left(u_n(\epsilon)\right) = \frac{1}{(\ln_0 n \cdots \ln_r n)\ln_r^{\epsilon} n}\,.$$

An application of Theorem 3.5.1 yields that

$$P\left(M_n > u_n(\epsilon) \quad \text{i.o.}\right) = 0 \quad \text{or} \quad = 1$$

according as $\epsilon > 0$ or $\epsilon < 0$ for small $|\epsilon|$ and hence

$$\limsup_{n\to\infty} \frac{M_n}{\ln n} = 1 \quad \text{a.s.} \tag{3.69}$$

Moreover, we conclude from Theorem 3.5.2 that

$$P(M_n \le (1+\epsilon)\ln n \quad \text{i.o.}) = 0 \quad \text{or} \quad 1$$

according as $\epsilon < 0$ or $\epsilon > 0$ for small $|\epsilon|$ and hence

$$\liminf_{n\to\infty} \frac{M_n}{\ln n} = 1 \quad \text{a.s.} \tag{3.70}$$

which, together with (3.69), yields an a.s. analogue of (3.68):

$$\lim_{n\to\infty} \frac{M_n}{\ln n} = 1 \quad \text{a.s.}$$

Now choose

$$u_n'(\epsilon) = \ln\left(\frac{n}{\ln(\ln_2 n\,(\ln_1 n \cdots \ln_r n)\ln_r^{\epsilon} n)}\right), \quad r \ge 1\,.$$

Then, by Theorem 3.5.2,

$$P\left(M_n \le u_n'(\epsilon) \quad \text{i.o.}\right) = 0 \quad \text{or} \quad = 1 \tag{3.71}$$

according as $\epsilon > 0$ or $\epsilon < 0$ for small $|\epsilon|$. Refinements of (3.70) are possible. For example, for fixed $r \ge 1$,

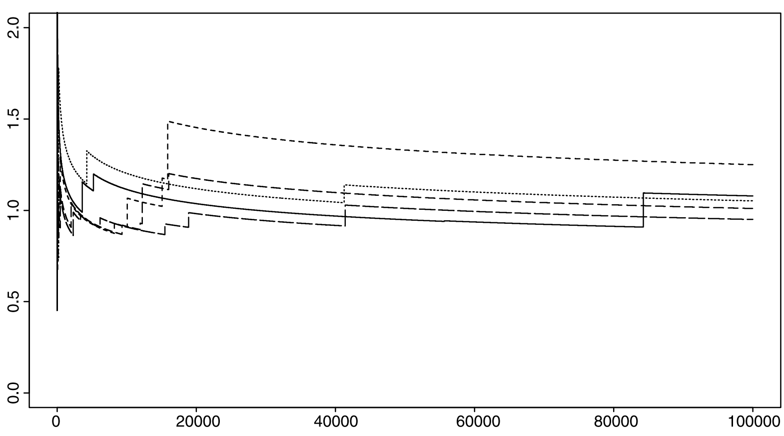


Figure 3.5.7 *Five sample paths of* $(M_n/\ln n)$ *for* 100 000 *realisations of iid standard exponential rvs. The rate of a.s. convergence to* 1 *appears to be very slow.*

$$\begin{aligned}
&P(M_n \le u_n'(\epsilon) \quad \text{i.o.})\\
&= P\Bigg(M_n - \Big(\ln n - \ln(\ln_3 n + (\ln_2 n + \cdots + \ln_{r+1} n))\Big)\\
&\qquad \le -\ln\left(\frac{\ln_3 n + (\ln_2 n + \cdots + \ln_{r+1} n + (1+\epsilon)\ln_{r+2} n)}{\ln_3 n + (\ln_2 n + \cdots + \ln_{r+1} n)}\right) \text{ i.o.}\Bigg)\\
&= P\Bigg(M_n - \Big(\ln n - \ln(\ln_3 n + (\ln_2 n + \cdots + \ln_{r+1} n))\Big)\\
&\qquad \le -\ln\left(1 + (1+\epsilon)\frac{\ln_{r+2} n}{\ln_3 n + (\ln_2 n + \cdots + \ln_{r+1} n)}\right) \quad \text{i.o.}\Bigg).
\end{aligned}$$

This together with (3.71), a mean value theorem argument and Theorem 3.5.2 imply that, for $r \ge 1$,

$$\liminf_{n\to\infty} \frac{\ln_2 n}{\ln_{r+2} n}\left(M_n - \Big(\ln n - \ln(\ln_3 n + (\ln_2 n + \cdots + \ln_{r+1} n))\Big)\right) = -1 \quad \text{a.s.}$$

□

Example 3.5.8 (Uniform distribution, continuation of Example 3.3.15)
Let F be uniform on $(0,1)$. From Example 3.3.15 we know that

$$M_n \stackrel{\text{a.s.}}{\to} 1\,.$$

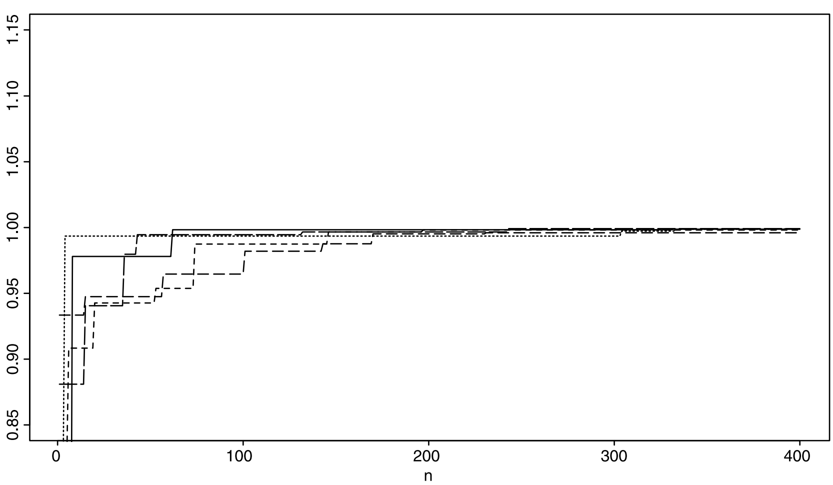


Figure 3.5.9 *Five sample paths of M_n for 400 realisations of iid $U(0,1)$ rvs. The rate of a.s. convergence to 1 is very fast.*

We derive some a.s. refinements of this limit result.
Choose

$$u_n(\epsilon) = 1 - \frac{1}{(\ln_0 n \ln_1 n \cdots \ln_r n) \ln_r^{\epsilon} n}\,, \quad r \geq 0\,.$$

Then, by Theorem 3.5.1,

$$P\left(M_n > u_n(\epsilon) \quad \text{i.o.}\right) = 0 \quad \text{or} \quad = 1$$

according as $\epsilon > 0$ or $\epsilon < 0$ for small $|\epsilon|$.
Now choose

$$u_n'(\epsilon) = 1 - \frac{1}{n} \ln\left(\ln_2 n \left(\ln_1 n \cdots \ln_r n\right) \ln_r^{\epsilon} n\right)\,, \quad r \geq 1\,.$$

Then, by Theorem 3.5.2,

$$P\left(M_n \leq u_n'(\epsilon) \quad \text{i.o.}\right) = 0 \quad \text{or} \quad = 1$$

according as $\epsilon > 0$ or $\epsilon < 0$ for small $|\epsilon|$. □

Notes and Comments

There does not exist very much literature on the a.s. behaviour of maxima of iid rvs. An extensive account can be found in Galambos [249]. The treatment of results about the normalised lim inf of maxima started with work by Barndorff–Nielsen [44, 45] who proved necessary and sufficient conditions under certain restrictions. A result in the same spirit was obtained by Robbins and Siegmund [544]. Klass [383, 384] finally proved the criterion of Theorem 3.5.2 under minimal restrictions on the df F and on the behaviour of (u_n). Goldie and Maller [275] use point process techniques to derive a.s. convergence results for order statistics and records.