CHAPTER 2

THEORETICAL BACKGROUND

2.1 Introduction

The scope of this chapter is to present and describe the main results of classical extreme value theory, since all the statistical techniques that have been developed for inference of extremes (and which are going to be discussed in more depth in subsequent chapters) are based on these results. More particularly, extreme value theory is concerned with the form of the limiting (non-degenerate) d.f.'s of the extremes (maxima) as well as the conditions under which such a limit exists. Furthermore, in the context of extreme value theory, characterization and other properties of families of d.f.'s with common limits have been developed.

Throughout this chapter X_1, X_2, ... is a sequence of i.i.d. non-degenerate r.v.'s with common d.f. F. In particular, $\{X_1,...,X_n\}$ denotes a random sample of size n from d.f. F. Our interest is focused on the 'behaviour' of the sample maxima

$$M_n = \max(X_1, ..., X_n), \qquad n \geq 2.$$

Of course, using basic probabilistic calculations, one can easily derive that the exact d.f. of M_n is

$$P(M_n \leq x) = [F(x)]^n, \qquad x \in \Re,\ n \in N$$

which leads to

$$M_n \xrightarrow{a.s.} x_F,$$

where $x_F = \sup\{x \in \Re : F(x) < 1\}$ is the right endpoint of F.

Still, this result does not provide much of insight about the extremes of a d.f. A larger amount of information about the order of magnitude of maxima is provided by weak convergence results. This is exactly the essence of extreme value theory.

The main analytic tool of extreme value theory is the theory of regularly varying functions, while the basic probabilistic tool is point process theory. So, before proceeding to the presentation of extreme value theory, we provide an introduction to concepts such as regularly varying functions, among others, which are commonly used in extreme value theory and are necessary for a better comprehension of the logic and of the results of extreme value theory.

In the sections to follow, apart from providing the main results of extreme value theory (sections 2.2 and 2.3), we describe the main distributions that are used in the context of extreme-value analysis, that is the Generalized Extreme-Value distribution (section 2.4) and the Generalized Pareto distribution (section 2.6). In section 2.5, some results are provided for the case that instead of maxima, we are interested in minima, while in the final section we present some characteristic examples of known d.f.'s whose maxima have a non-degenerate limiting d.f.

2.2 Theory of Regular Variation

The concept of regular variation is widely used in extreme value theory. Below, we summarize some of the main results (definitions, extensions, properties) of regular variation theory which are relevant to our scope. These are further elaborated, accompanied by proofs, in the encyclopaedic volume on the subject by Bingham et al. (1987).

Definition: *Regularly Varying Function* (Embrechts et al., 1997)

A positive, Lebesgue measurable function L on $(0,\infty)$ is regularly varying (at ∞) of index $\alpha \in \Re$ (and we write $L \in RV_\alpha$) if

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = t^{\alpha}, \quad \text{for all } t>0.$$

Any regularly varying function can be decomposed as $L(x) = x^{\alpha} l(x)$, where l is a so-called slowly-varying function.

Definition: ***Slowly Varying Function*** (Embrechts et al., 1997)

A positive, Lebesgue measurable function l on $(0,\infty)$ is slowly varying (at ∞) (and we write $l \in RV_0$) if

$$\lim_{x\to\infty} \frac{l(tx)}{l(x)} = 1, \text{ for all } t>0.$$

Notice that a slowly varying function is essentially a regularly varying function with index 0. A nice 'interpretation' of slowly varying functions is that of bifurcation functions describing how a particular function, mainly behaving as a power function, differs from that particular power function. Typical examples are positive constants or functions converging to a positive constant, logarithms and iterated logarithms.

Definition: ***Rapidly Varying Function*** (Embrechts et al., 1997)

A positive, Lebesgue measurable function h on $(0,\infty)$ is rapidly varying (at ∞) with index $-\infty$ (and we write $h \in RV_{-\infty}$) if

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

Though the above definitions give us an idea about the form of regularly (slowly or rapidly) varying functions, still a more useful insight into them is provided by the next representation theorems. These are the forms used in extreme value theory. The decomposition into two measurable functions with known properties helps the interpretation and, more importantly, the parameter estimation of regular (accordingly slowly or rapidly) varying functions.

Theorem: ***Representation Theorem for Regularly Varying Functions*** (Embrechts et al., 1997)

If $L \in RV_\alpha$ for some $\alpha \in \Re$, then

$$L(x) = c(x)\exp\left\{\int_z^x \frac{\delta(u)}{u}\,du\right\}, \qquad x \ge z,$$

for some z>0 where c and δ are measurable functions, such that $c(x) \to c_0 \in \Re$ and $\delta(x) \to \alpha$ as $x \to \infty$. The converse implication also holds.

Theorem: ***Representation Theorem for Slowly Varying Functions*** (Beirlant et al., 1996)

If $l \in RV_0$, then

$$l(x) = c(x)\exp\left\{\int_z^x \frac{\delta(u)}{u} du\right\}, x \geq z,$$

for some $z>0$ where c and δ are measurable functions, such that $c(x) \to c_0 \in \Re$ and $\delta(x) \to 0$ as $x \to \infty$. The converse also holds.

Theorem: ***Representation Theorem for Rapidly Varying Functions*** (Embrechts et. al, 1997)

If $h \in RV_{-\infty}$, then there exist functions c and δ, such that $c(x) \to c_0 \in \Re$, $\delta(x) \to -\infty$ as $x \to \infty$ and for some $z>0$

$$h(x) = c(x)\exp\left\{\int_z^x \frac{\delta(u)}{u} du\right\}, \qquad x \geq z,$$

The converse also holds.

The following result of Karamata is also very useful, since it is often used in proofs of theorems of extreme-value theory. It essentially says that integrals of regularly varying functions are again regularly varying functions, or more precisely, one can 'take the slowly varying function out of the integral'.

Theorem: ***Karamata's Theorem*** (Karamata, 1933)

Let l be a slowly varying function, bounded in $[x_0, \infty)$ for some $x_0 \geq 0$. Then

(a) for $\alpha > -1$,

$$\int_{x_0}^x t^\alpha l(t)dt \sim (\alpha+1)^{-1} x^{\alpha+1} l(x), \qquad \text{as } x \to \infty,$$

(b) for $\alpha < -1$,

$$\int_x^\infty t^\alpha l(t)dt \sim -(\alpha+1)^{-1} x^{\alpha+1} l(x), \qquad \text{as } x \to \infty.$$

2.3 Limit Laws for Maxima

Now, we return to the main topic of this chapter, which is the presentation of extreme value theory. The first and most fundamental component of extreme value theory is the determination of all the possible limiting d.f.'s of (properly normalized) maxima. This problem has already been resolved in 1928 by Fisher and Tippet, in their famous theorem which is regarded to be the foundation of classical extreme value theory.

Theorem: ***Fisher-Tippet Theorem, Limit Laws for Maxima*** (Fisher and Tippet, 1928)

Let (X_n) be a sequence of i.i.d. r.v.'s and $M_n = \max(X_1, ..., X_n)$. If there exist norming constants $c_n > 0$, $d_n \in \Re$ and some non-degenerate d.f. H such that

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

then H belongs to the type of one of the following three d.f.'s:

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

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

$$\text{Gumbel}: \ \Lambda(x) = \exp\left[-e^{-x}\right], \ x \in \Re.$$

A comprehensive sketch of the proof can be found in Embrechts et al. (1997). In simple words, this theorem states that ***if*** the maximum value of a d.f. tends (in distribution) to a non-degenerate d.f. then this limiting d.f. can ***only*** be one of the three forms given above. The d.f.'s Φ_α, Ψ_α, Λ are called *standard extreme value* d.f.'s, and the corresponding r.v.'s *standard extremal* r.v.'s, and are going to be described in more details in a subsequent section of this chapter.

Still, this theorem leaves some questions open. Two other issues that need to be addressed so as to have a complete description of the behaviour of extremes are :

1) Identification of the necessary conditions that a d.f. F should fulfil, so that its normalized maxima converge weakly to one of the *extreme value* d.f.'s.
2) Determination of the centring constant d_n and normalizing constant c_n (jointly called norming constants).

In the sequel, we deal with these two questions.

In order to explore the necessary conditions for the existence of a limiting d.f. H, it is useful to adopt a systematic approach towards the set of d.f.'s whose maxima have the same limiting d.f. So, we introduce the notion of *maximum domain of attraction*.

Definition : *Maximum Domain of Attraction* (Embrechts et al., 1997)

The r.v. X (the d.f. F of X, or the distribution of X) is said to belong to the maximum domain of attraction of the extreme value distribution H if there exist constants $c_n > 0$, $d_n \in \Re$ such that

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

holds. We write $X \in$ MDA(H) (or $F \in$ MDA(H)).

That is, the MDA of the extreme value distribution H is the family of distributions whose maxima tends, in distribution, to H.

A first result which stems from the effort to find the conditions under which a d.f. has limiting d.f. of maxima is the following characterization property.

Proposition : *Characterization of MDA(H)* (Embrechts et al., 1997)

The d.f. F belongs to the maximum domain of attraction of the extreme value distribution H with norming constants $c_n > 0$, $d_n \in \Re$ if and only if

$$\lim_{n\to\infty}\left\{n\overline{F}\left(c_n x + d_n\right)\right\} = -\ln\left[H(x)\right], \quad x \in \Re,\ \overline{F} = 1 - F .$$

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

This general characterization property can lead to necessary and sufficient conditions for a d.f. F to have one of the three specific limiting laws for maxima. This can be achieved using the concept of regular variation for the Fréchet and Weibull case (though in the last case a small modification is required), while for the Gumbel case the notion of rapidly varying functions is needed. In the sequel we present the characterization properties (their proofs can be found in Embrechts et al. ,1997).

Theorem : *Maximum Domain of Attraction of Φ_α* (Embrechts et al., 1997)

The d.f. 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*.

Every F $\in$ MDA(Φ_α) has an infinite right endpoint $x_F = \infty$. Essentially, MDA(Φ_α) embraces all the distributions with right tails regularly varying with index $-\alpha$, that is $1 - F(x) \to x^{-a}$, $x \to \infty$. These d.f.'s are called Pareto-type or heavy-tailed distributions.

Theorem : *Maximum Domain of Attraction of Ψ_α* (Embrechts et al., 1997)

The d.f. 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*.

All d.f.'s in MDA(Ψ_α) have a finite right endpoint x_F. This family of distributions has only 'rescaled' right tails tending to infinity in the form of x^{-a} (i.e. regularly varying), that is $1 - F(x_F - x^{-1}) \to x^{-a}$, $x \to \infty$.

Theorem : *Maximum Domain of Attraction of Λ* (Embrechts et al., 1997)

The d.f. F with right endpoint $x_F \le \infty$ belongs to the maximum domain of attraction of Λ if and only if there exists some $z < x_F$ such that F can be written as

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

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 $\lim_{x \to x_F^-} a'(x) = 0$.

Notice, that now there is no direct linkage with regular variation notion. However, MDA(Λ) covers a wider collection of d.f.'s with very different tails ranging from moderately heavily (such as the lognormal) to light (e.g. normal distribution). D.f.'s of this class can be either unbounded or bounded on the right.

Still, the above conditions are difficult to verify in practice. Von Mises (1936) has given some simpler sufficient conditions which can be used for practical purposes. These conditions for each one of the three limiting cases are given in the sequel.

Corollary : ***Von Mises Condition for MDA(Φ_α)*** (Embrechts et al., 1997)

Let F be an absolutely continuous d.f. with density f satisfying

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

then $F \in$ MDA(Φ_α).

Corollary : ***Von Mises Condition for MDA(Ψ_α)*** (Embrechts et al., 1997)

Let F be an absolutely continuous d.f. with density f which is positive on some finite interval (z, x_F). If

$$\lim_{x\to x_F^-}\frac{(x_F-x)f(x)}{\overline{F}(x)}=\alpha>0,$$

then $F \in$ MDA(Ψ_α).

Corollary : ***Von Mises Condition for MDA(Λ)*** (Embrechts et al., 1997)

Let F be a d.f. with right endpoint $x_F \leq \infty$, such that for $z < x_F$ F has the representation

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

where c is some positive constant, $a(x)$ is a positive, absolutely continuous function (with respect to Lebesgue measure) with density $a'(x)$ having $l\lim_{x\to x_F^-}a'(x)=0$.

Then $F \in$ MDA(Λ).

Note that the d.f.'s just described are known as 'von Mises functions'. A long list of references dealing with these theoretical issues can be found in Johnson et al. (1995).

Through the procedures used to derive the previous results, proper forms for the norming constants appear which are different for the three different types of limiting d.f.'s. These formulae, answering the second question we posed, are summarized in the following table.

Table 2.1. Norming Constants for the Maximum Domains of Attraction

Maximum Domain of Attraction	***Centring Constant d_n***	***Normalizing Constant c_n***
Fréchet	0	$F^{\leftarrow}(1-n^{-1})$
Weibull	x_F	x_F - $F^{\leftarrow}(1-n^{-1})$
Gumbel	$a(d_n)$	$F^{\leftarrow}(1-n^{-1})$

Note: $F^{\leftarrow}(u)=\inf\{x: F(x)\geq u\}$ is the generalized inverse function of the d.f. F, i.e. it is the well-known quantile function.

2.4 Generalized Extreme-Value Distribution

According to the fundamental theorem of Fisher-Tippet, the possible limiting d.f.'s of maxima are of three distinct types. In the previous section, we explored some of their properties and it is an immediate result that there are large analogues among the three extreme value d.f.'s. Actually, these three d.f.'s can be summarized into a single family of distributions, known as *Generalized Extreme-Value distribution*, using a different parametrization. The idea of this unification is attributed to von Mises (1936), though often is also attributed to Jenkinson (1955). The corresponding definition is:

Definition : ***Generalized Extreme-Value distribution (GEV)*** (see Embrechts et al., 1997)

The GEV d.f. H_γ is defined by the formula

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

where $1+\gamma x>0$, i.e. the support of H_γ, is

$x>-\gamma^{-1}$ for $\gamma>0$,

$x<-\gamma^{-1}$ for $\gamma<0$, and

$x\in\Re$ for $\gamma=0$.

The corresponding p.d.f. (for $1+\gamma x>0$) is given by the formula

$$h_\gamma(x) = \begin{cases} (1+\gamma x)^{-(1+1/\gamma)} \exp\left[-(1+\gamma x)^{-1/\gamma}\right] & if \quad \gamma \neq 0 \\ \exp\left[-\left(x + \exp(-x)\right)\right] & if \quad \gamma = 0. \end{cases}$$

One can easily derive the correspondence between the generalized extreme-value distribution and the three standard extreme value d.f.'s. Specifically,

for $\gamma > 0$, $H_\gamma \rightarrow \Phi_\alpha$ with $\alpha = \gamma^{-1}$,

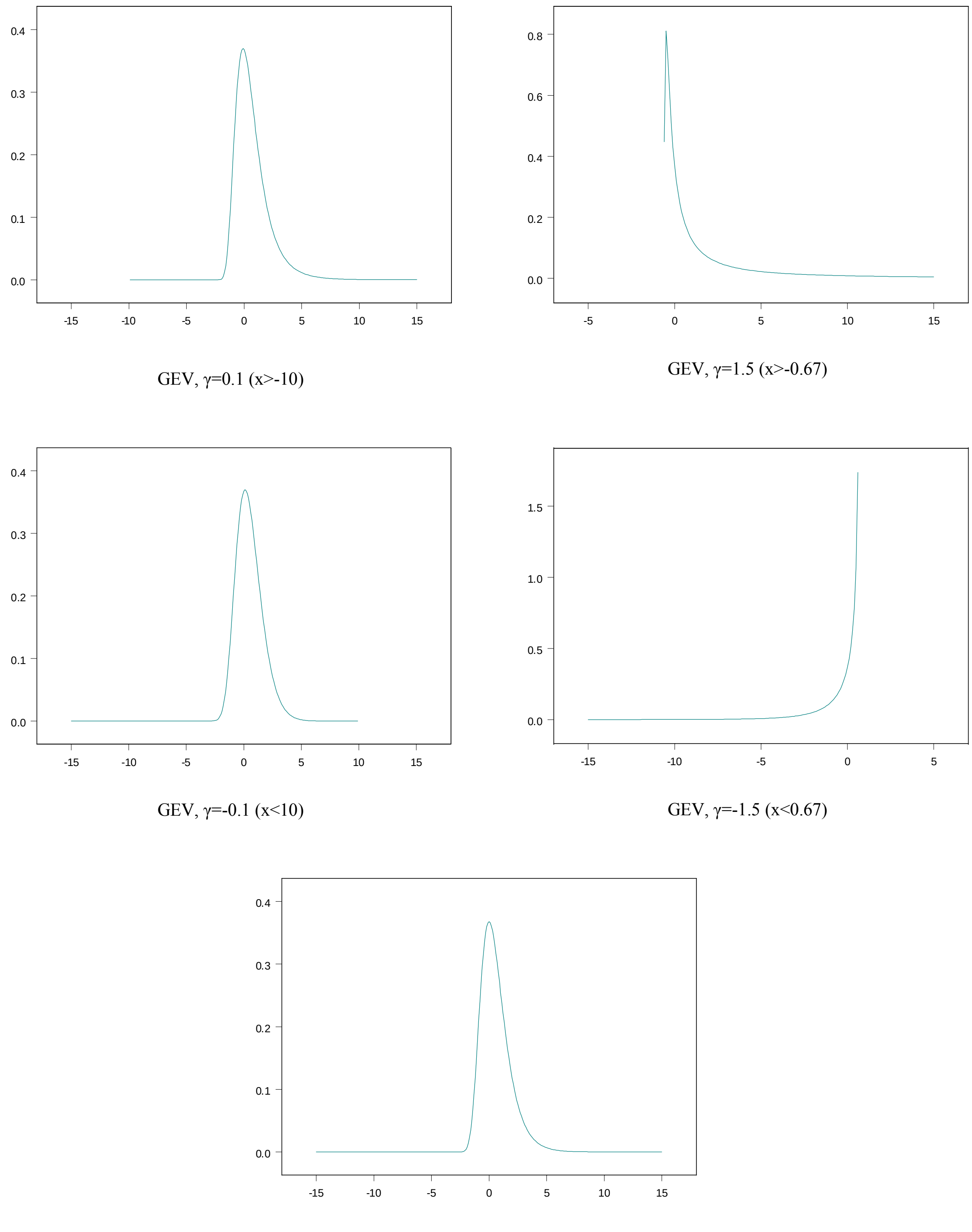
for $\gamma < 0$, $H_\gamma \rightarrow \Psi_\alpha$ with $\alpha = -\gamma^{-1}$, and

for $\gamma = 0$, $H_\gamma \rightarrow \Lambda$.

The parameter γ is called 'shape parameter', though it is often referred to as 'extreme-value index' or 'tail index'.

Such a one-parameter representation of the three standard cases in one family of d.f.'s will turn out to be particularly useful. Actually, its introduction was mainly motivated by statistical applications.

In the figure below, we give a visual inspection of the form of the limiting d.f. of normalized maxima, depending on the shape parameter γ. It is really interesting to note the extensive similarities that exist among the graphs of GEV p.d.f. with γ equal to +0.1, -0.1 and 0. The similarities are spotted in the middle part of the density, for γ in the interval [-2, +2]. The tails of the p.d.f.'s are the ones that differ. In any case for γ=0.1 x cannot be smaller than –10, for γ=-0.1 x cannot exceed 10, while for γ=0 x can take any value in the axis of real numbers.



GEV, γ=0 (x∈ℜ)

Figure 2.1. Probability density functions of the Generalized Extreme-Value distribution for shape-parameter values $\gamma = \pm 0.1,\ \pm 1.5,\ 0$.

In the previous section we have presented characterization properties separately for each standard extreme value d.f. After the unification, a 'global' characterization property can be developed, which includes all the three limit laws for maxima. Notice, also, that the maximum domain of attraction of H_γ is constructed in the same way as the maximum domain of attraction of the three standard extreme value d.f's. Actually, there are two forms to characterize GEV.

Theorem: ***Characterization I of Generalized Extreme-Value distribution*** (Embrechts et al., 1997)

The d.f. F with right endpoint $x_F \le \infty$ belongs to the maximum domain of attraction of H_γ ($F \in MDA(H_\gamma)$) if and only if, for $1+\gamma x>0$,

$$\lim_{u\to x_F^-}\frac{\overline{F}(u+xa(u))}{\overline{F}(u)}=\begin{cases}(1+\gamma x)^{-1/\gamma} & if \quad \gamma\neq 0\\ e^{-x} & if \quad \gamma=0\end{cases}$$

where $a(x)$ is a positive, measurable function.

An interesting probabilistic interpretation stems by observing that $\overline{F}(u+xa(u))/\overline{F}(u)=P((X-u)/a(u)>x|X>u)$. That is, the above characterization provides a distributional approximation for the scalled excess over the (high) threshold u. The appropriate scaling factor is $a(u)$.

Theorem: ***Characterization II of Generalized Extreme-value distribution*** (Embrechts et al., 1997)

The d.f. F with right endpoint $x_F \le \infty$ belongs to the maximum domain of attraction of H_γ ($F \in MDA(H_\gamma)$) if and only if, for x, y>0, y≠1,

$$\lim_{s\to\infty}\frac{U(sx)-U(s)}{U(sy)-U(s)}=\begin{cases}\dfrac{x^\gamma-1}{y^\gamma-1} & if \quad \gamma\neq 0\\ \dfrac{\ln x}{\ln y} & if \quad \gamma=0\end{cases}$$

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

A reformulation of this relation leads to an estimation procedure for quantiles outside the range of the data, while a special case of this formula is also used to motivate the Pickands estimator of γ. These issues will be further elaborated on chapter 4.

A sketch of the proof for both characterization properties is given in Embrechts, et al. (1997). This 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. In simple words, it provides the necessary and sufficient conditions for a d.f. F to have maxima that converge weakly to a non-degenerate d.f. If F does not satisfy the above conditions then its maxima do not converge in any distribution. Another point that we should mention is that the limit d.f.'s are unique only up to affine transformations, i.e. in each case the exact limiting d.f. is not necessarily the standard GEV distribution but it can be a d.f. of the more general location-scale GEV family, defined in the sequel. Still, this does not cause any identification problem, since in any case, by appropriate normalization of maxima, we can end up in the standard GEV. This is the reason for which, in most parts of the present thesis we are mainly concerned only with the standard GEV.

- ***Max-Stability of Extreme Value Distributions***

A characteristic property of extreme value distributions is that they are closed (up to affine transformations) for maxima (max-stability property). In general, the definition of max-stability is the following

Definition : *Max-Stable Distribution* (Embrechts et al., 1997)

A non-degenerate r.v. X (the corresponding distribution or df) is called max-stable if

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

holds for i.i.d. X, X_1, ..., X_n, appropriate constants $c_n > 0$, $d_n \in \Re$ and every $n \geq 2$.

From the definition alone we can see that every max-stable distribution is a limit distribution for maxima of i.i.d. r.v.'s. Moreover, max-stable distributions are the only limit laws for normalized maxima. This is indicated by the following theorem (whose proof is supplied by Embrechts et al., 1997).

Theorem : *Limit Property of Max-Stable Laws* (Embrechts et al., 1997)

The class of max-stable distributions coincides with the class of all possible (non-degenerate) limit laws for (properly normalized) maxima of i.i.d. r.v.'s.

Now, we can formally state and prove the max-stability property of extreme value distributions.

Proposition

The extreme value distributions (Fréchet, Weibull and Gumbel) are max-stable distributions and they are the only distributions with this property

Proof :

First, we are going to show that, indeed the extreme value distributions satisfy the defining property of max-stable distributions.

- Fréchet case

Let X, X_1, ..., X_n be i.i.d. from Fréchet(α), with common distribution function

$$F(x) = \begin{cases} 0, & x \le 0 \\ \exp\left[-x^{-\alpha}\right], & x > 0 \end{cases} \qquad (2.1),$$

M_n=max(X_1,..X_n) with distribution function $F_{M_n}(x) = \left(F(x)\right)^n$ (2.2)

Substituting (2.1) in (2.2), we get

For x>0

$$F_{M_n}(x) = \left(\exp\left[-x^{-\alpha}\right]\right)^n = \exp\left[-nx^{-a}\right] = \exp\left[-\left(n^{-1/\alpha}x\right)^{-\alpha}\right]$$

$$\Rightarrow F_{M_n}(x) = F\left(n^{-1/\alpha}x\right)$$

For x≤0, $F_{M_n}(x) = (0)^n = 0 = F\left(n^{-1/\alpha}x\right)$

That is,

$$\max(X_1, \ldots, X_n) \overset{d}{=} n^{1/\alpha} X$$

which proves the max-stability of the Fréchet d.f.

- Weibull case

Analogously to previous case, let X, X_1, ..., X_n be i.i.d. from Weibull(α), with common distribution function

$$F(x)=\begin{cases}\exp\left[-(-x)^{\alpha}\right], & x\le 0\\ 1, & x>0\end{cases} \qquad (2.3)$$

$M_n=\max(X_1,..X_n)$ with distribution function $F_{M_n}(x)=\left(F(x)\right)^n$ (2.4)

Substituting (2.3) in (2.4), we get

For x≤0

$$F_{M_n}(x)=\left(\exp\left[-(-x)^{\alpha}\right]\right)^n=\exp\left[-n(-x)^{a}\right]=\exp\left[-\left(-n^{1/\alpha}x\right)^{\alpha}\right]$$

$$\Rightarrow F_{M_n}(x)=F\left(n^{1/\alpha}x\right)$$

For x>0, $F_{M_n}(x)=(1)^n=1=F\left(n^{1/\alpha}x\right)$

That is,

$$\max(X_1,\ldots,X_n)\overset{d}{=}n^{-1/\alpha}X$$

which proves the max-stability of the Weibull d.f.

- <u>Gumbel case</u>

Let X, X_1, ..., X_n be i.i.d. from Gumbel, with common distribution function

$$\Lambda(x)=\exp\left[-e^{-x}\right], \quad x\in\Re \qquad (2.5)$$

$M_n=\max(X_1,..X_n)$ with distribution function $F_{M_n}(x)=\left(F(x)\right)^n$ (2.6)

Substituting (2.5) in (2.6), we get

$$F_{M_n}(x)=\left(\exp\left[-e^{-x}\right]\right)^n=\exp\left[-ne^{-x}\right]=\exp\left[-e^{\ln n}e^{-x}\right]=\exp\left[-e^{-(x-\ln n)}\right]$$

$$\Rightarrow F_{M_n}(x)=F(x-\ln n)$$

That is,

$$\max(X_1,\ldots,X_n)\overset{d}{=}X+\ln n$$

which proves the max-stability of the Weibull d.f.

Up to this point, we have showed that the extreme value distributions are max-stable. In order to show that they are the only max-stable distributions, we just have to combine the theorem of ‘Limit property of max-stable laws’ with the Fisher-Tippet Theorem. Indeed, according to Fisher-Tippet Theorem, extreme value d.f.s are the only possible limit laws of (properly normalized) maxima. On the other hand, the limit property of max-stable laws dictates that the class of max-stable distributions coincides with the class of all possible (non-degenerate) limit laws for (properly normalized) maxima of i.i.d.

Graphically we have that

$$\{\text{Extreme Value Distributions}\} \Leftrightarrow \{\text{Limit Laws of properly normalized Maxima}\}$$

and

$$\{\text{Max- Stable Distributions}\} \Leftrightarrow \{\text{Limit Laws of properly normalized Maxima}\}.$$

Consequently,

$$\{\text{Extreme Value Distributions}\} \Leftrightarrow \{\text{Max- Stable Distributions}\}.$$

Up to now, we have focused on the GEV as limiting d.f. of the extremes of other d.f.'s, i.e. we were within the context of extreme value theory. At this point, we are going to concentrate on the properties of Generalized Extreme-Value d.f. itself , i.e. we are going to investigate Generalized Extreme-Value d.f. from the point of view of 'distribution theory'.

❒

- ***Location – Scale Family of Generalized Extreme-Value distribution***

If a r.v. X has d.f. H_γ , then the r.v. $\mu + \sigma X$ has the d.f. $H_{\gamma,\mu,\sigma}(x) = H_\gamma\left(\frac{x-\mu}{\sigma}\right)$. $\mu \in \Re$ is the location parameter and $\sigma > 0$ is the scale parameter. The support of the d.f. is adjusted accordingly to the GEV.

Notice that the location parameter μ is the left endpoint in case $\gamma > 0$, and the right endpoint if $\gamma < 0$.

- ***Distributional Characteristics of Generalized Extreme-Value distribution***

- The mean exists only for $\gamma < 1$, and is given by $E\left(H_\gamma\right) = \frac{\Gamma(1-\gamma)-1}{\gamma}$, where $\Gamma(.)$ is the Gamma function.

- The variance exists only for $\gamma < 1/2$, and equals $Var(H_\gamma) = \frac{\Gamma(1-2\gamma) - \Gamma^2(1-\gamma)}{\gamma^2}$.

- The GEV densities are unimodal with mode $\text{mod}(H_\gamma) = \frac{(1+\gamma)^{-\gamma} - 1}{\gamma}$ for $\gamma > -1$, while they are J-shaped for $\gamma \le -1$.

The case $\gamma = 0$ is included in the latter formulas by considering limits with γ tending to zero. That is,

$E(H_0) = \lim_{\gamma \to 0} E(H_\gamma) = \int_0^\infty (-\ln x) e^{-x} dx = \lambda$, where λ=0.577216 is Euler's constant,

$Var(H_0) = \lim_{\gamma \to 0} Var(H_\gamma) = \pi^2/6$, and

$\text{mod}(H_0) = 0$.

2.5 Limit Laws for Minima

Up to now, we have been entirely concerned with the study of maxima, i.e. the largest values of a d.f. Still, not few are the cases and practical situations where we are primarily interested in the minima (smallest values). Fortunately, the study of minima is totally equivalent to the study of maxima, since it holds that

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

All the findings, mentioned above, concerning maxima have an analogous formula and hold for minima as well.

So, there is an one-to-one relationship between limiting d.f.'s of maxima and minima. More particularly,

if $P\left(\max_{i \le n}(-X_i) \le b_n + a_n x\right) \xrightarrow{n \to \infty} H(x)$

then $P\left(\min_{i \le n}(X_i) \le d_n + c_n x\right) \xrightarrow{n \to \infty} 1 - H(-x)$,

where $c_n = a_n$ and $d_n = -b_n$.

This yields that the possible limiting d.f.'s of minima can be only of three types (in the sense of Fisher-Tippet) :

Converse Fréchet : $\widetilde{\Phi}_\alpha(x) = 1 - \Phi_\alpha(-x) = \begin{cases} 1 - \exp\left[-(-x)^{-\alpha}\right], & x \le 0 \\ 1, & x > 0 \end{cases}$ α > 0.

Converse Weibull : $\widetilde{\Psi}_\alpha(x) = 1 - \Psi_\alpha(-x) = \begin{cases} 0, & x \le 0 \\ 1 - \exp\left[-(x)^{\alpha}\right], & x > 0 \end{cases}$ α > 0.

Converse Gumbel : $\widetilde{\Lambda}(x) = 1 - \Lambda(-x) = 1 - \exp\left[-e^{x}\right], \qquad x \in \Re.$

The limiting d.f.'s of minima can also be simply called converse extreme value d.f.'s. Some well-known special cases are:

- The converse Gumbel $\widetilde{\Lambda}(x)$ is also called Gompertz d.f. This is the d.f. that satisfies the famous Gompertz law

- $\widetilde{\Psi}_1(x)$ is the exponential d.f. on the positive half-line

- $\widetilde{\Psi}_2(x)$ is the Rayleigh d.f. that is also of interest in some particular statistical applications. A characteristic relationship is that if the areas of random circles are exponentially distributed, then the diameters have a Rayleigh d.f.

Under the unified representation (γ-reparametrization) the limiting d.f. of minima is the *Converse Generalized Extreme-Value distribution* :

$$\widetilde{H}_\gamma(x) = 1 - H_\gamma(-x) = \begin{cases} \exp\left[-(1-\gamma x)^{-1/\gamma}\right] & if \quad \gamma \ne 0 \\ \exp\left[-\exp(x)\right] & if \quad \gamma = 0 \end{cases}$$

where $1 - \gamma x > 0$, i.e. the support of $\widetilde{H}_\gamma$ is

$x < \gamma^{-1}$ for $\gamma > 0$
$x > \gamma^{-1}$ for $\gamma < 0$
$x \in \Re$ for $\gamma = 0$.

Another characteristic property of maxima which is 'transmitted' to minima is that of max-stability. More precisely, the limiting d.f.'s of minima are characterized by the min-stability. That is, the d.f's converse Fréchet, converse Weibull and converse Gumbel are the only distributions for which it holds that

$$P\left\{\min_{i\le n}\left(X_i\right)\le d_n + c_n x\right\} = 1-\left(1-F\left(d_n + c_n x\right)\right)^n = F(x)$$

for a certain choice of constants $c_n > 0$, $d_n \in \Re$, where (X_i) are iid random variables with common d.f. F ($\widetilde{\Phi}$, $\widetilde{\Psi}$, or $\widetilde{\Lambda}$).

The min-stability can also be expressed in terms of the survivor function $\overline{F}$, that is

$$P\left\{\min_{i\le n}\left(X_i\right) > d_n + c_n x\right\} = \overline{F}^{\,n}\left(d_n + c_n x\right) = \overline{F}(x).$$

2.6 Generalized Pareto Distribution

Another very useful distribution in the context of extreme value theory is the Generalized Pareto Distribution (GPD, in short). Apart from the theoretical connection to the extreme value theory, the GPD has been extensively used in practical applications for statistical inference in extremes. Many statistical procedures in the field we are working have been inspired by GPD. More details for this will be provided in following chapters. The formal definition of the GPD is as follows.

Definition : *The Generalized Pareto Distribution* (Embrechts et al., 1997)

The Generalized Pareto distribution G_γ is defined by the formula

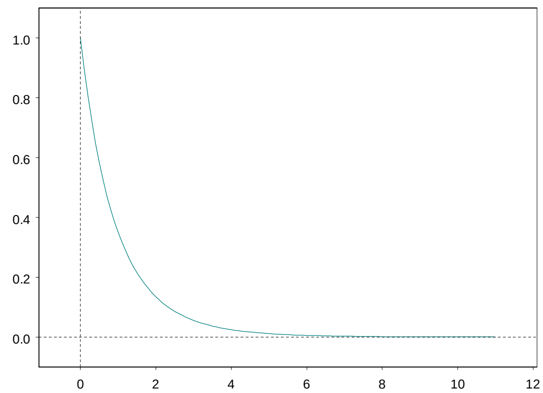
$$G_\gamma(x) = \begin{cases} 1-\left(1+\gamma x\right)^{-1/\gamma} & if \quad \gamma \ne 0 \\ 1-e^{-x} & if \quad \gamma = 0 \end{cases}$$

where the support of G_γ is
$x \ge 0$ if $\gamma \ge 0$, and
$0 \le x \le -1/\gamma$ if $\gamma < 0$.

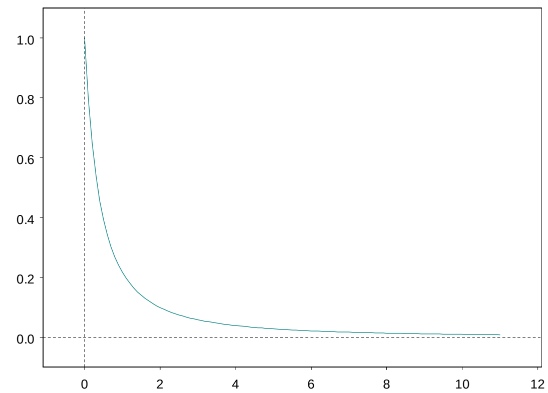
The corresponding p.d.f. is

$$g_\gamma(x) = \begin{cases} \left(1+\gamma x\right)^{-(1+1/\gamma)} & if \quad \gamma \ne 0 \\ e^{-x} & if \quad \gamma = 0. \end{cases}$$

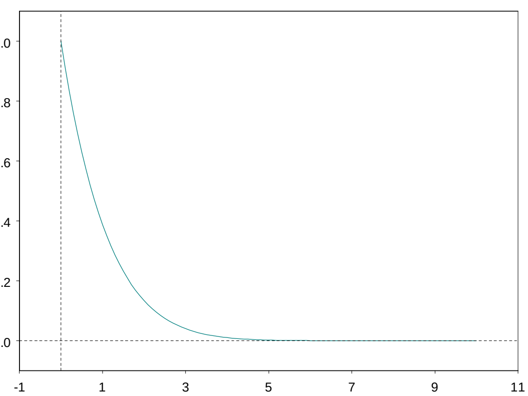
The following figure displays the generalized Pareto p.d.f. for several values of the parameter γ. As was the case in the GEV p.d.f., large similarities exist among the graphs of p.d.f. of the GPD for γ is equal to +0.1, -0.1 and 0. Their main difference is that in the case of γ= -0.1 x cannot exceed 10, while in the two other cases x can become infinitely large.



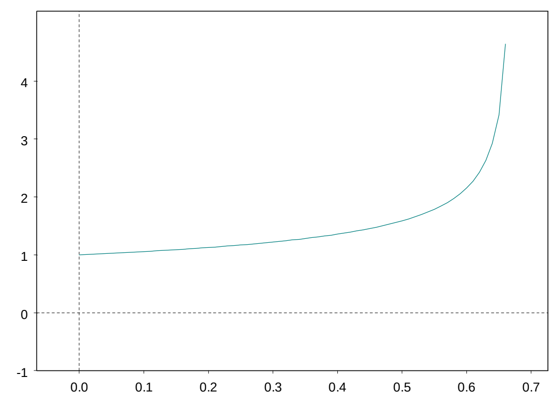
GPD, γ=0.1 (x≥0)



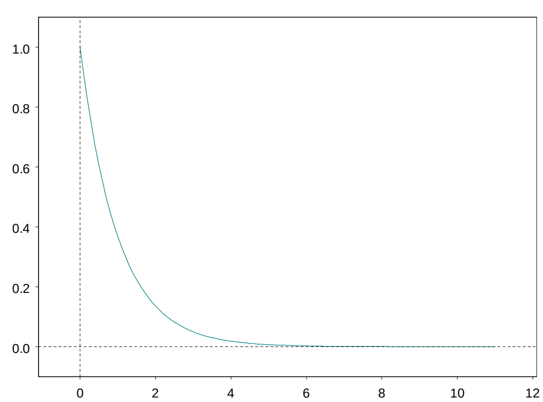
GPD, γ=1.5 (x≥0)



GPD, γ=-0.1 (0≤x≤10)



GPD, γ=-1.5 (0≤x≤0.67)



GPD, γ=0 (x≥0)

Figure 2.2. Probability density functions of the Generalized Pareto distribution for shape-parameter values $\gamma = \pm 0.1,\ \pm 1.5,\ 0$.

In order to obtain the full statistical family of GPD's, we have to add location and scale parameters $\nu \in \Re$, and $\beta > 0$ respectively. So, the related location-scale family $G_{\gamma;\nu,\beta}$ is

$G_{\gamma q \nu,\beta}(x) = G_\gamma\left(\frac{x-\nu}{\beta}\right)$, where the support is adjusted accordingly. The location parameter ν is always the left endpoint of the support of the distribution. Notice also that, for $\gamma > 0$ the GPD corresponds to Pareto d.f., for $\gamma < 0$ to Beta d.f., while for $\gamma = 0$ we get the standard exponential distribution.

The strong connection of Generalized Pareto d.f. to extreme value theory is revealed in the following considerations. Indeed, if we take a more thorough look at the first characterization property of the maximum domain of attraction of generalized extreme-value d.f. (MDA(H_γ)), we notice that the left-hand side of the relation can be rewritten as

$$\frac{\overline{F}\left(u + xa(u)\right)}{\overline{F}(u)} = P\left(X > u + xa(u) \middle| X > u\right) = P\left(\frac{X-u}{a(u)} > x \middle| X > u\right) = \overline{F}_u\left(a(u)x\right),$$

where F_u stands for the conditional distribution of X given that X>u.

That is, if X $\in$ MDA(H_γ), then it holds that, for $1 + \gamma x > 0$,

$$\lim_{u \to x_F^-} P\left(\frac{X-u}{a(u)} > x \middle| X > u\right) = \begin{cases} \left(1+\gamma x\right)^{-1/\gamma} & if \quad \gamma \neq 0 \\ e^{-x} & if \quad \gamma = 0 \end{cases}$$

which gives a distributional approximation for the scaled excesses over the (high) threshold u by the GPD. The appropriate scaling factor is $a(u)$. In this context, F_u is called excess d.f.

So, by making use of the previous definition, we can easily derive the following proposition, which reveals the connection of the Generalized Pareto distribution to extreme-value theory.

Proposition : ***Limit distribution of scaled excesses over high thresholds*** (Embrechts et al., 1997)

If the d.f. F with right endpoint $x_F \leq \infty$ belongs to the maximum domain of attraction of H_γ ($F \in MDA(H_\gamma)$) then, for $1+\gamma x>0$,

$$\lim_{u \to x_F^-} P\left(\frac{X-u}{a(u)} > x \middle| X>u \right) \stackrel{d}{=} \overline{G}_{\gamma, a(u)}(x).$$

Properties of GPD

Maybe, the most important property of the GPD is the one known as 'POT-stability', (POT stands for Peaks Over Threshold), i.e. the class of GPD's is closed with respect to changes of the threshold. Formally, we have the following proposition.

Proposition : ***POT-Stability of Generalized Pareto Distributions*** (Reiss and Thomas, 1997)

Generalized Pareto d.f.'s are the only continuous d.f.'s such that for a certain choice of constants b_u and a_u, it holds that

$$F^{[u]}(b_u + a_u x) = F(x),$$

where $F^{[u]}(x) = P(X \leq x | X>u) = \dfrac{F(x)-F(u)}{\overline{F}(u)}$ is the exceedance d.f. over the threshold u (truncation of F left at u).

That is, the truncated version of a GPD remains in the same family of d.f.'s.

The usefulness of this proposition lies on the fact that if we can assume that the d.f. of the scaled excesses over a high threshold v is well approximated by GPD, then the same holds for any higher threshold u>v. This remark will turn out to be of great practical interest in the case of estimation of the extreme-value index γ, as we will see in subsequent chapter.

Other Properties of Generalized Pareto Distributions

Let $G_{\gamma,\beta}$ be a GPD with shape parameter γ and scale parameter β (when written as β(u) it implies dependence of β on u).

- For every $\gamma \in \Re$, F ∈ MDA(H_γ) if and only if

$$\lim_{u \to x_F^-} \sup_{0<x<x_F-u} \left| F_u(x) - G_{\gamma,\beta(u)} \right| = 0,$$

where $F_u(x) = P\left(X - u \le x \middle| X > u\right)$, x≥0, is the excess d.f. over the threshold u.

- For every x_i, i=1,2 (to the appropriate domain of support),

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

This property is a reformulation of the POT-stability property.

- If $\gamma < 1$, then for $u < x_F$, it holds that

$$e(u) = E\left(X - u \middle| X > u\right) = \frac{\beta + \gamma u}{1-\gamma}, \qquad \beta + \gamma u > 0,$$

where e(u) is called mean-excess function.

This property shows that the mean-excess function is linear with respect to u. This remark is a key-point to many statistical techniques (estimation methods for γ) that we will present later.

- If $\gamma < 1$, then the followings hold:

E(X) < ∞

$$E\left[\left(1 + \frac{\gamma}{\beta} X\right)^{-r}\right] = \frac{1}{1+\gamma r}, \quad \text{for } r>-1/\gamma$$

$$E\left[\left(\ln\left(1 + \frac{\gamma}{\beta} X\right)\right)^{k}\right] = \gamma^k k!, \quad k \in N$$

$$E\left[X\left(\overline{G}_{\gamma,\beta}(X)\right)^{r}\right] = \frac{\beta}{(r+1-\gamma)(r+1)}, \quad (r+1)/|\gamma| > 0.$$

Moreover,

if $\gamma < 1/r$, $r \in N$, then $E\left(X^r\right) = \dfrac{\beta^r \Gamma\left(\gamma^{-1} - r\right)}{\gamma^{r+1}\Gamma\left(1+\gamma^{-1}\right)} r!$.

- Let N be a Poisson r.v. with parameter λ ($P(\lambda)$), independent of the i.i.d. sequence (X_n) with a $G_{\gamma,\beta}$ d.f., and $M_N = \max(X_1, ..., X_N)$. Then, it holds that

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

where $\mu = \beta\gamma^{-1}\left(\lambda^{\gamma} - 1\right)$ is the location parameter and $\sigma = \beta\lambda^{\gamma}$ is the scale parameter.

The essence of this property is that if the number of exceedances over a high threshold is exact Poisson and the excess d.f. is an exact GPD, then the maximum of this excesses has an *exact* GEV distribution.

The above properties *suggest* the following *approximate model* for the exceedance times and the excesses of an i.i.d. 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 (and searching for that point from which linearity seems to start).
- The distribution of the maximum of a Poisson number of i.i.d. excesses over a high threshold is a GEV.

2.7 Examples and Counter-Examples

In a previous section we have mentioned the necessary and sufficient conditions, for a d.f. to have maxima that converge weakly to a particular d.f. form. These conditions may seem particularly complicated and difficult to be fulfilled. Still, many, if not most, of the well-known distributions have been proven to fulfil these conditions, thus, leading to a formal study of the maxima of such distributions. In the table that follows we present a list of such d.f.'s, accompanied with the shape parameter γ of the MDA(H_γ) to which each one of them belongs.

Table 2.2. Distributions belonging to MDA(H_γ), $\gamma \in \Re$.

Name	Distribution Function F(x)	Shape parameter γ
Uniform (0,1)	x	-1
Beta	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^x u^{\alpha-1}(1-u)^{\beta-1}du$	-1/β
Reversed Burr	$1-\left(\frac{\beta}{\beta+(x_F-x)^{-\tau}}\right)^{\lambda}$	$-1/\tau\lambda$
Extreme Value	$\exp\left(-(1+\gamma x)^{-1/\gamma}\right)$	γ
Benktander II	$1-x^{-(1-\beta)}\exp\left(-\frac{\alpha}{\beta}x^{\beta}\right)$	0
Weibull	$1-\exp\left(-\lambda x^{\tau}\right)$	0
Exponential	$1-\exp(-\lambda x)$	0
Gamma	$\frac{\lambda^m}{\Gamma(m)}\int_0^x u^{m-1}e^{-\lambda u}du$	0
One-sided Logistic	$\frac{e^x-1}{e^x+1}$	0
Lognormal	$\int_0^x \frac{1}{\sqrt{2\pi\sigma^2}\,u}\exp\left(-\frac{1}{2\sigma^2}(\ln u-\mu)^2\right)du$	0
Benktander I	$1-x^{-(1+\alpha+\beta\ln x)}\left(1+2\frac{\beta}{\alpha}\ln x\right)$	$\frac{1}{1+\alpha}$ if β=0 0 if β>0
Pareto	$1-x^{-\alpha}$	1/α

Burr	$1-\left(\dfrac{\beta}{\beta+x^{\tau}}\right)^{\lambda}$	1/τλ
Generalized Pareto	$G_{\gamma}(x)=1-(1+\gamma x)^{-1/\gamma}$	γ
Loggamma	$\dfrac{\lambda^{m}}{\Gamma(m)}\int_{0}^{x}(\ln u)^{m-1}u^{-\lambda-1}du$	1/λ
Loghyperbolic	$D\int_{0}^{x}\exp\left(-\dfrac{\phi+\alpha}{2}\sqrt{\delta^{2}+(\ln u-\mu)^{2}}+\dfrac{\phi-\alpha}{2}(\ln u-\mu)\right)\dfrac{du}{u}$	1/α
Log-logistic	$\dfrac{\beta x^{\alpha}}{1+\beta x^{\alpha}}$	1/α
Fréchet	$\exp(-x^{-a})$	1/α
Cauchy	$\int_{-\infty}^{x}\dfrac{1}{\pi(1+u^{2})}du$	1

The above table presents only some of the cases of d.f.'s which belong to the maximum domain of attraction of the generalized extreme-value distribution (H_γ). A much longer list could have resulted if we had considered not only distributions with completely specified d.f.'s, but also d.f.'s for which only tails are specified. For example, Pareto-like d.f.'s belong to MDA(H_γ), with $\gamma>0$, while exponential-like d.f.'s belong to MDA(H_0).

Still, there are important distributions, commonly used in practice, which could not be included in the above list. Indeed, one needs certain continuity conditions on F at its right endpoint to include F at MDA(H_γ), which rules out many of the discrete distributions. Neither the normalized maxima of i.i.d. Poisson distributed r.v.'s, of Geometric r.v.'s nor of negative Binomial r.v.'s converge to a non-degenerate limit distribution. More particularly, it can be shown (see Embrechts et. al, 1997) that for any d.f. with a jump at its *finite right endpoint*, no limiting distribution for maxima exists, whatever the normalization. Additionally, in case of discrete d.f.'s with *infinite right endpoints*, jump heights which do not decay sufficiently fast (i.e. $\lim_{n\to\infty}\dfrac{\overline{F}(n)}{\overline{F}(n-1)}<1$) prohibit the existence of non-degenerate limit distributions for normalized maxima.