Chapter 2

Extreme Value Theory

In this chapter we summarize results in extreme value theory, which are primarily based on the condition that the upper tail of the underlying df is in the δ-neighborhood of a generalized Pareto distribution (GPD). This condition, which looks a bit restrictive at first sight (see Section 2.2), is however essentially equivalent to the condition that rates of convergence in extreme value theory are at least of algebraic order (see Theorem 2.2.5). The δ-neighborhood is therefore a natural candidate to be considered, if one is interested in reasonable rates of convergence of the functional laws of small numbers in extreme value theory (Theorem 2.3.2) as well as of parameter estimators (Theorems 2.4.4, 2.4.5 and 2.5.4).

2.1 Domains of Attraction, von Mises Conditions

Recall from Example 1.3.4 that a df F belongs to the domain of attraction of an extreme value df (EVD) $G_\beta(x) = \exp(-(1+\beta x)^{-1/\beta})$, $1+\beta x > 0$, denoted by $F \in \mathcal{D}(G_\beta)$, iff there exist constants $a_n > 0$, $b_n \in \mathbb{R}$ such that

$$\begin{aligned} &F^n(a_n x + b_n) \longrightarrow_{n\to\infty} G_\beta(x), \qquad x \in \mathbb{R} \\ \iff\ &P((Z_{n:n} - b_n)/a_n \le x) \longrightarrow_{n\to\infty} G_\beta(x), \qquad x \in \mathbb{R}, \end{aligned}$$

where $Z_{n:n}$ is the sample maximum of an iid sample $Z_1, \dots, Z_n$ with common df F. Moreover, $Z_{1:n} \le \dots \le Z_{n:n}$ denote the pertaining order statistics.

The Gnedenko-De Haan Theorem

The following famous result due to Gnedenko [176] (and partially due to de Haan [184]) provides necessary as well as sufficient conditions for $F \in \mathcal{D}(G_\beta)$.

M. Falk et al., *Laws of Small Numbers: Extremes and Rare Events*, 3rd ed.,
DOI 10.1007/978-3-0348-0009-9_2,

Theorem 2.1.1 (Gnedenko-de Haan). *Let G be a non-degenerate df. Suppose that F is a df with the property that for some constants $a_n > 0$, $b_n \in \mathbb{R}$,*

$$F^n(a_n x + b_n) \longrightarrow_{n\to\infty} G(x),$$

for any point of continuity x of G. Then G is up to a location and scale shift an EVD G_β, i.e., $F \in \mathcal{D}(G) = \mathcal{D}(G_\beta)$.

Put $\omega(F) := \sup\{x \in \mathbb{R} : F(x) < 1\}$. Then we have

(i) *$F \in \mathcal{D}(G_\beta)$ with $\beta > 0 \iff \omega(F) = \infty$ and*

$$\lim_{t\to\infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-1/\beta}, \qquad x > 0.$$

The normalizing constants can be chosen as $b_n = 0$ and $a_n = F^{-1}(1 - n^{-1})$, $n \in \mathbb{N}$, where $F^{-1}(q) := \inf\{t \in \mathbb{R} : F(t) \geq q\}$, $q \in (0, 1)$, denotes the quantile function or generalized inverse of F.

(ii) *$F \in \mathcal{D}(G_\beta)$ with $\beta < 0 \iff \omega(F) < \infty$ and*

$$\lim_{t\to\infty} \frac{1 - F(\omega(F) - \frac{1}{tx})}{1 - F(\omega(F) - \frac{1}{t})} = x^{1/\beta}, \qquad x > 0.$$

The normalizing constants can be chosen as $b_n = \omega(F)$ and $a_n = \omega(F) - F^{-1}(1 - n^{-1})$.

(iii) *$F \in \mathcal{D}(G_0) \iff$ there exists $t_0 < \omega(F)$ such that $\int_{t_0}^{\omega(F)} 1 - F(x)\, dx < \infty$ and*

$$\lim_{t\to\omega(F)} \frac{1 - F(t + xR(t))}{1 - F(t)} = \exp(-x), \qquad x \in \mathbb{R},$$

where $R(t) := \int_t^{\omega(F)} 1 - F(y)\, dy/(1 - F(t))$, $t < \omega(F)$. The norming constants can be chosen as $b_n = F^{-1}(1 - n^{-1})$ and $a_n = R(b_n)$.

It is actually sufficient to consider in part (iii) of the preceding Theorem 2.1.1 only $x \geq 0$, as shown by Worms [464]. In this case, the stated condition has a probability meaning in terms of conditional distributions, known as the *additive excess property*. We refer to Section 1.3 of Kotz and Nadarajah [293] for a further discussion.

Note that we have for any $\beta \in \mathbb{R}$,

$$F \in \mathcal{D}(G_\beta) \iff F(\cdot + \alpha) \in \mathcal{D}(G_\beta)$$

for any $\alpha \in \mathbb{R}$. Without loss of generality we will therefore assume in the following that $\omega(F) > 0$.

Von Mises Conditions

The following sufficient condition for $F \in \mathcal{D}(G_\beta)$ goes essentially back to von Mises [336].

Theorem 2.1.2 (Von Mises conditions). *Let F have a positive derivative f on $[x_0, \omega(F))$ for some $0 < x_0 < \omega(F)$.*

(i) *If there exist $\beta \in \mathbb{R}$ and $c > 0$ such that $\omega(F) = \omega(H_\beta)$ and*

$$\lim_{x \uparrow \omega(F)} \frac{(1+\beta x) f(x)}{1 - F(x)} = c, \tag{VM}$$

then $F \in \mathcal{D}(G_{\beta/c})$.

(ii) *Suppose in addition that f is differentiable. If*

$$\lim_{x \uparrow \omega(F)} \frac{d}{dx}\left(\frac{1 - F(x)}{f(x)}\right) = 0, \tag{VM$_0$}$$

then $F \in \mathcal{D}(G_0)$.

Condition (VM$_0$) is the original criterion due to von Mises [336, page 285] in case $\beta = 0$. Note that it is equivalent to the condition

$$\lim_{x \uparrow \omega(F)} \frac{1 - F(x)}{f(x)} \frac{f'(x)}{f(x)} = -1$$

and, thus, (VM) in case $\beta = 0$ and (VM$_0$) can be linked by l'Hôpital's rule. Condition (VM) will play a crucial role in what follows in connection with generalized Pareto distributions.

If F has ultimately a positive derivative, which is monotone in a left neighborhood of $\omega(F) = \omega(H_\beta)$ for some $\beta \neq 0$, and if $F \in \mathcal{D}(G_{\beta/c})$ for some $c > 0$, then F satisfies (VM) with β and c (see Theorems 2.7.1 (ii), 2.7.2 (ii) in de Haan [184]). Consequently, if F has ultimately a positive derivative f such that $\exp(-x) f(x)$ is non-increasing in a left neighborhood of $\omega(F) = \infty$, and if $F(\log(x)),\ x > 0$, is in $\mathcal{D}(G_{1/c})$ for some $c > 0$, then F satisfies (VM) with c and $\beta = 0$.

A df F is in $\mathcal{D}(G_0)$ iff there exists a df F^* with $\omega(F^*) = \omega(F)$, which satisfies (VM$_0$) and which is tail equivalent to F^*, i.e.,

$$\lim_{x \uparrow \omega(F)} \frac{1 - F(x)}{1 - F^*(x)} = 1,$$

see Balkema and de Haan [21].

Proof. We prove only the case $\beta = 0$ in condition (VM), the cases $\beta > 0$ and $\beta < 0$ can be shown in complete analogy (see Theorem 2.7.1 in Galambos [167] and Theorems 2.7.1 and 2.7.2 in de Haan [184]). Observe first that

$$\int_{t_0}^{\omega(F)} 1 - F(x)\, dx = \int_{t_0}^{\omega(F)} \frac{1 - F(x)}{f(x)} f(x)\, dx \leq 2/c$$

if t_0 is large. We have further by l'Hôpital's rule

$$\lim_{t\to\omega(F)} R(t) = \lim_{t\to\omega(F)} \frac{\int_t^{\omega(F)} 1-F(x)\,dx}{1-F(t)} = \lim_{t\to\omega(F)} \frac{1-F(t)}{f(t)} = 1/c.$$

Put now $g(t) := -f(t)/(1-F(t)) = (\log(1-F(t)))'$, $t \ge t_0$. Then we have the representation

$$1-F(t) = C\exp\Big(\int_{t_0}^t g(y)\,dy\Big), \qquad t\ge t_0,$$

with some constant $C>0$ and thus,

$$\frac{1-F(t+xR(t))}{1-F(t)} = \exp\Big(\int_t^{t+xR(t)} g(y)\,dy\Big) \to_{t\to\omega(F)} \exp(-x), \qquad x\in\mathbb{R},$$

since $\lim_{y\to\omega(F)} g(y) = -c$ and $\lim_{t\to\omega(F)} R(t) = 1/c$. The assertion now follows from Theorem 2.1.1 (iii). $\square$

Distributions F with differentiable upper tail of Gamma type that is, $\lim_{x\to\infty} F'(x)/((b^p/\Gamma(p))\, e^{-bx}x^{p-1}) = 1$ with $b,p>0$ satisfy (VM) with $\beta=0$. Condition (VM) with $\beta>0$ is, for example, satisfied for F with differentiable upper tail of Cauchy-type, whereas triangular type distributions satisfy (VM) with $\beta<0$. We have equality in (VM) with F being a GPD $H_\beta(x) = 1-(1+\beta x)^{-1/\beta}$, for all $x\ge 0$ such that $1+\beta x>0$.

The standard normal df Φ satisfies $\lim_{x\to\infty} x(1-\Phi(x))/\Phi'(x) = 1$ and does not satisfy, therefore, condition (VM) but (VM_0).

The following result states that we have equality in (VM) *only* for a GPD. It indicates therefore a particular relationship between df with GPD-like upper tail and the von Mises condition (VM), which we will reveal later. Its proof can easily be established (see also Corollary 1.2 in Falk and Marohn [143]).

Proposition 2.1.3. *We have ultimately equality in (VM) for a df F iff F is ultimately a GPD. Precisely, we have equality in (VM) for $x\in[x_0,\omega(F))$ iff there exist $a>0$, $b\in\mathbb{R}$ such that*

$$1-F(x) = 1-H_{\beta/c}(ax+b), \qquad x\in[x_0,\omega(F)),$$

where $b=(a-c)/\beta$ in case $\beta\ne 0$ and $a=c$ in case $\beta=0$.

Differentiable Tail Equivalence

Denote by h_β the density of the GPD H_β that is,

$$h_\beta(x) = \frac{G'_\beta(x)}{G_\beta(x)} = (1+\beta x)^{-1/\beta-1} \quad \text{for} \quad \begin{cases} x\ge 0 & \text{if} \quad \beta\ge 0 \\ 0\le x< -1/\beta & \text{if} \quad \beta<0. \end{cases}$$

Note that with $b = (a - c)/\beta$ if $\beta \neq 0$ and $a = c$ if $\beta = 0$ we have

$$\frac{1 + \beta x}{c} = \frac{1 - H_{\beta/c}(ax + b)}{ah_{\beta/c}(ax + b)}$$

for all x in a left neighborhood of $\omega(H_\beta) = \omega(H_{\beta/c}(ax + b))$. If F satisfies (VM), we can write therefore for any $a > 0$ and $b \in \mathbb{R}$ such that $b = (a - c)/\beta$ if $\beta \neq 0$ and $a = c$ if $\beta = 0$,

$$\begin{aligned} 1 &= \lim_{x \to \omega(F)} \frac{f(x)}{1 - F(x)} \frac{1 + \beta x}{c} \\ &= \lim_{x \to \omega(F)} \frac{f(x)}{ah_{\beta/c}(ax + b)} \frac{1 - H_{\beta/c}(ax + b)}{1 - F(x)}. \end{aligned} \tag{2.1}$$

As a consequence, we obtain that under (VM) a df F is *tail equivalent* to the GPD $H_{\beta/c}(ax + b)$, for some $a > 0,\ b \in \mathbb{R}$ with $b = (a - c)/\beta$ if $\beta \neq 0$ and $a = 1$ if $\beta = 0$, iff F and $H_{\beta/c}(ax + b)$ are *differentiable tail equivalent*. Precisely

$$\begin{aligned} &\lim_{x \to \omega(F)} \frac{1 - F(x)}{1 - H_{\beta/c}(ax + b)} \quad \text{exists in } [0, \infty] \\ \iff &\lim_{x \to \omega(F)} \frac{f(x)}{ah_{\beta/c}(ax + b)} \quad \text{exists in } [0, \infty] \end{aligned}$$

and in this case these limits coincide. Note that the "if" part of this conclusion follows from l'Hôpital's rule anyway.

Von Mises Condition With Remainder

The preceding considerations indicate that the condition (VM) is closely related to the assumption that the upper tail of F is close to that of a GPD. This idea can be made rigorous if we consider the rate at which the limit in (VM) is attained.

Suppose that F satisfies (VM) with $\beta \in \mathbb{R}$ and $c > 0$ and define by

$$\eta(x) := \frac{(1 + \beta x) f(x)}{1 - F(x)} - c, \qquad x \in [x_0, \omega(F)),$$

the remainder function in condition (VM). Then we can write for any $a > 0,\ b \in \mathbb{R}$ with $b = (a - c)/\beta$ if $\beta \neq 0$ and $a = c$ if $\beta = 0$,

$$\frac{f(x)}{ah_{\beta/c}(ax + b)} = \frac{1 - F(x_1)}{1 - H_{\beta/c}(ax_1 + b)} \exp\Big(- \int_{x_1}^{x} \frac{\eta(t)}{1 + \beta t}\, dt \Big) \Big(1 + \frac{\eta(x)}{c} \Big), \tag{2.2}$$

$x \in [x_1, \omega(F))$, where $x_1 \in [x_0, \omega(F))$ is chosen such that $ax_1 + b > 0$. Recall that for $\beta < 0$ we have $ax + b = ax + (a - c)/\beta \leq \omega(H_{\beta/c}) = -c/\beta \iff x \leq -1/\beta = \omega(H_\beta) = \omega(F)$. The following result is now immediate from the preceding representation (2.2) and equation (2.1).

Proposition 2.1.4. *Suppose that F satisfies (VM) with $\beta \in \mathbb{R}$ and $c > 0$. Then we have for any $a > 0$, $b \in \mathbb{R}$, with $b = (a-c)/\beta$ if $\beta \neq 0$ and $a = c$ if $\beta = 0$,*

$$\lim_{x \uparrow \omega(F)} \frac{1-F(x)}{1-H_{\beta/c}(ax+b)} = \lim_{x \uparrow \omega(F)} \frac{f(x)}{ah_{\beta/c}(ax+b)} = \begin{cases} 0 \\ \alpha \in (0,\infty) \\ \infty \end{cases}$$

$$\iff \int_{x_0}^{\omega(F)} \frac{\eta(t)}{1+\beta t}\, dt = \begin{cases} \infty \\ d \in \mathbb{R} \\ -\infty\,. \end{cases}$$

Observe that, for any $a, c, \alpha > 0$,

$$\alpha\Big(1 - H_{\beta/c}\Big(ax + \frac{a-c}{\beta}\Big)\Big) = 1 - H_{\beta/c}\Big(a\alpha^{-\beta/c}x + \frac{a\alpha^{-\beta/c} - c}{\beta}\Big) \tag{2.3}$$

if $\beta \neq 0$ and

$$\alpha(1 - H_0(ax+b)) = 1 - H_0(ax + b - \log(a)). \tag{2.4}$$

Consequently, we can find by Proposition 2.1.4 constants $a > 0$, $b \in \mathbb{R}$, with $b = (a-c)/\beta$ if $\beta \neq 0$ and $a = c$ if $\beta = 0$, such that

$$\lim_{x \uparrow \omega(F)} \frac{1-F(x)}{1-H_{\beta/c}(ax+b)} = \lim_{x \uparrow \omega(F)} \frac{f(x)}{ah_{\beta/c}(ax+b)} = 1$$

iff

$$-\infty < \int_{x_0}^{\omega(F)} \frac{\eta(t)}{1+\beta t}\, dt < \infty.$$

The preceding result reveals that a df F satisfying (VM) is tail equivalent (or, equivalently, differentiable tail equivalent) to a GPD iff the remainder function η converges to zero fast enough; precisely iff $\int_{x_0}^{\omega(F)} \eta(t)/(1+\beta t)\, dt \in \mathbb{R}$.

Observe now that the condition

$$\eta(x) = O((1 - H_\beta(x))^\delta) \text{ as } x \to \omega(F) = \omega(H_\beta)$$

for some $\delta > 0$ implies that $\int_{x_0}^{\omega(F)} \eta(t)/(1+\beta t)\, dt \in \mathbb{R}$ and

$$\int_x^{\omega(F)} \frac{\eta(t)}{1+\beta t}\, dt = O((1 - H_\beta(x))^\delta) \text{ as } x \to \omega(F).$$

The following result is therefore immediate from equation (2.2) and Taylor expansion of exp at zero.

Proposition 2.1.5. *Suppose that F satisfies (VM) with $\beta \in \mathbb{R}$ and $c > 0$ such that $\eta(x) = O((1 - H_\beta(x))^\delta)$ as $x \to \omega(F)$ for some $\delta > 0$. Then there exist $a > 0$, $b \in \mathbb{R}$, with $b = (a-c)/\beta$ if $\beta \neq 0$ and $a = c$ if $\beta = 0$, such that*

$$f(x) = ah_{\beta/c}(ax+b)\Big(1 + O((1 - H_\beta(x))^\delta)\Big)$$

for any x in a left neighborhood of $\omega(F)$.

It is worth mentioning that under suitable conditions also the reverse implication in Proposition 2.1.5 holds. For the proof of this result, which is Proposition 2.1.7 below, we need the following auxiliary result.

Lemma 2.1.6. *Suppose that F and G are df having positive derivatives f and g near $\omega(F) = \omega(G)$. If $\psi \geq 0$ is a decreasing function defined on a left neighborhood of $\omega(F)$ with $\lim_{x\to\omega(F)} \psi(x) = 0$ such that*

$$|f(x)/g(x) - 1| = O(\psi(x)),$$

then

$$|(1 - G(x))/(1 - F(x)) - 1| = O(\psi(x))$$

as $x \to \omega(F) = \omega(G)$.

Proof. The assertion is immediate from the inequalities

$$\begin{aligned}\Big|\frac{1 - G(x)}{1 - F(x)} - 1\Big| &\leq \int_x^{\omega(F)} \Big|\frac{f(t)}{g(t)} - 1\Big|\, dG(t)/(1 - F(x))\\ &\leq C\psi(x)(1 - G(x))/(1 - F(x)),\end{aligned}$$

where C is some positive constant. □

Proposition 2.1.7. *Suppose that F satisfies condition (VM) with $\beta \in \mathbb{R}$ and $c > 0$. We require further that, in a left neighborhood of $\omega(F)$,*

$$f(x) = a h_{\beta/c}(ax + b)\Big(1 + O((1 - H_\beta(x))^\delta)\Big)$$

for some $\delta > 0$, $a > 0$, $b \in \mathbb{R}$, where $b = (a - c)/\beta$ if $\beta \neq 0$ and $a = c$ if $\beta = 0$. Then the remainder function

$$\eta(x) = \frac{f(x)(1 + \beta x)}{1 - F(x)} - c$$

is also of order $(1 - H_\beta(x))^\delta$ that is,

$$\eta(x) = O((1 - H_\beta(x))^\delta) \text{ as } x \to \omega(F).$$

Proof. Write for x, in a left neighborhood of $\omega(F) = \omega(H_\beta)$,

$$\begin{aligned}\eta(x) &= c\frac{1 - H_{\beta/c}(ax + b)}{1 - F(x)}\ \frac{f(x)}{a h_{\beta/c}(ax + b)} - c\\ &= c\Big(\frac{1 - H_{\beta/c}(ax + b)}{1 - F(x)} - 1\Big)\ \frac{f(x)}{a h_{\beta/c}(ax + b)} + c\Big(\frac{f(x)}{a h_{\beta/c}(ax + b)} - 1\Big)\\ &= O((1 - H_\beta(x))^\delta)\end{aligned}$$

by Lemma 2.1.6. □

Rates of Convergence of Extremes

The growth condition $\eta(x) = O((1 - H_\beta(x))^\delta)$ is actually a fairly general one as revealed by the following result, which is taken from Falk and Marohn [143], Theorem 3.2. It roughly states that this growth condition is already satisfied, if $F^n(a_n x + b_n)$ approaches its limit G_β at a rate which is proportional to a power of n. For a multivariate version of this result we refer to Theorem 5.5.5.

Define the norming constants $c_n = c_n(\beta) > 0$ and $d_n = d_n(\beta) \in \mathbb{R}$ by

$$c_n := \begin{cases} n^\beta & \text{if } \beta \neq 0 \\ 1 & \text{if } \beta = 0 \end{cases}, \qquad d_n := \begin{cases} \frac{n^\beta - 1}{\beta} & \text{if } \beta \neq 0 \\ \log(n) & \text{if } \beta = 0. \end{cases}$$

With these norming constants we have, for any $\beta \in \mathbb{R}$,

$$H_\beta(c_n x + d_n) \longrightarrow_{n\to\infty} G_\beta(x), \qquad x \in \mathbb{R},$$

as is seen immediately.

Theorem 2.1.8. *Suppose that F satisfies (VM) with $\beta \in \mathbb{R}$ and $c > 0$ such that $\int_{x_0}^{\omega(F)} \eta(t)/(1+\beta t)\,dt \in \mathbb{R}$. Then we know from Proposition* 2.1.4 *and equations* (2.3), (2.4) *that*

$$\lim_{x\uparrow\omega(F)} \frac{1 - F(x)}{1 - H_{\beta/c}(ax+b)} = 1$$

for some $a > 0$, $b \in \mathbb{R}$, with $b = (a-c)/\beta$ if $\beta \neq 0$ and $a = c$ if $\beta = 0$. Consequently, we obtain with $a_n := c_n(\beta/c)/a$, $b_n := (d_n(\beta/c) - b)/a$ that

$$\sup_{x\in\mathbb{R}} |F^n(a_n x + b_n) - G_{\beta/c}(x)| \longrightarrow_{n\to\infty} 0.$$

If we require in addition that

$$\lim_{x\uparrow\omega(F)} \frac{\eta(x)}{r(x)} = 1$$

for some monotone function $r : (x_0, \omega(F)) \to \mathbb{R}$ and

$$\sup_{x\in\mathbb{R}} |F^n(a_n x + b_n) - G_{\beta/c}(x)| = O(n^{-\delta})$$

for some $\delta > 0$, then

$$\eta(x) = O((1 - H_\beta(x))^{c\delta})$$

as $x \to \omega(F) = \omega(H_\beta)$.

The following result is now immediate from Theorem 2.1.8 and Proposition 2.1.5.

Corollary 2.1.9. *Suppose that F satisfies (VM) with $\beta \in \mathbb{R}$, $c > 0$ such that $\int_{x_0}^{\omega(F)} \eta(t)/(1+\beta t)\, dt \in \mathbb{R}$ and*

$$\lim_{x \uparrow \omega(F)} \frac{\eta(x)}{r(x)} = 1$$

for some monotone function $r : (x_0, \omega(F)) \to \mathbb{R}$. If for some $\delta > 0$,

$$\sup_{x \in \mathbb{R}} |F^n(a_n x + b_n) - G_{\beta/c}(x)| = O(n^{-\delta}),$$

with $a_n > 0$, b_n as in Theorem 2.1.8, then there exist $a > 0$, $b \in \mathbb{R}$ with $b = (a-c)/\beta$ if $\beta \neq 0$ and $a = c$ if $\beta = 0$, such that

$$f(x) = a h_{\beta/c}(ax + b)\Big(1 + O((1 - H_\beta(x))^{c\delta})\Big)$$

for any x in a left neighborhood of $\omega(F) = \omega(H_\beta)$.

Our next result is a consequence of Corollary 5.5.5 in Reiss [385] and Proposition 2.1.5 (see also Theorems 2.2.4 and 2.2.5). By $\mathbb{B}^k$ we denote the Borel-σ-field in $\mathbb{R}^k$.

Theorem 2.1.10. *Suppose that F satisfies (VM) with $\beta \in \mathbb{R}$, $c > 0$ such that $\eta(x) = O((1 - H_\beta(x))^\delta)$ as $x \to \omega(F)$ for some $\delta > 0$. Then there exist $a_n > 0$, $b_n \in \mathbb{R}$ such that for $k \in \{1, \dots, n\}$ and $n \in \mathbb{N}$,*

$$\begin{aligned}
&\sup_{B \in \mathbb{B}^k} \Big| P(((Z_{n-i+1:n} - b_n)/a_n)_{i \le k} \in B) \\
&\qquad - \begin{cases} P((\beta(\sum_{j \le i} \xi_j)^{-\beta})_{i \le k} \in B)\Big| & \quad\textit{if}\quad \beta \neq 0 \\ P((-\log(\sum_{j \le i} \xi_j))_{i \le k} \in B)\Big| & \quad\phantom{\textit{if}}\quad \beta = 0 \end{cases} \\
&= O((k/n)^{\delta/c} k^{1/2} + k/n),
\end{aligned}$$

where $\xi_1, \xi_2, \dots$ are independent and standard exponential rv.

Best Attainable Rates of Convergence

One of the significant properties of GPD is the fact that these distributions yield the best rate of joint convergence of the upper extremes, equally standardized, if the underlying df F is ultimately continuous and strictly increasing in its upper tail. This is captured in the following result. By $G_\beta^{(k)}$ we denote the distribution of $(\beta(\sum_{j \le i} \xi_j)^{-\beta})_{i \le k}$ if $\beta \neq 0$ and of $(-\log(\sum_{j \le i} \xi_j))_{i \le k}$ if $\beta = 0$, where $\xi_1, \xi_2, \dots$ is again a sequence of independent and standard exponential rv and $k \in \mathbb{N}$. These distributions $G_\beta^{(k)}$ are the only possible classes of weak limits of the joint distribution of the k largest and equally standardized order statistics in an iid sample (see Theorem 2.2.2 and Remark 2.2.3).

Theorem 2.1.11. *Suppose that F is continuous and strictly increasing in a left neighborhood of $\omega(F)$. There exist norming constants $a_n > 0, b_n \in \mathbb{R}$ and a positive constant C such that, for any $k \in \{1, \ldots, n\}$, $n \in \mathbb{N}$,*

$$\sup_{B \in \mathbb{B}^k} \left| P\Big(((Z_{n-i+1:n} - b_n)/a_n))_{i \leq k} \in B \Big) - G_\beta^{(k)}(B) \right| \leq Ck/n$$

iff there exist $c > 0$, $d \in \mathbb{R}$ such that $F(x) = H_\beta(cx + d)$ for x near $\omega(F)$.

The *if*-part of this result is due to Reiss [383], Theorems 2.6 and 3.2, while the *only if*-part follows from Corollary 2.1.13 below.

The bound in Theorem 2.1.11 tends to zero as n tends to infinity for any sequence $k = k(n)$ such that $k/n \longrightarrow_{n\to\infty} 0$. The following result which is taken from Falk [129], reveals that this is a characteristic property of GPD that is, only df F, whose upper tails coincide with that of a GPD, entail approximation by $G_\beta^{(k)}$ for *any* such sequence k.

By $G_{\beta,(k)}$ we denote the k-th onedimensional marginal distribution of $G_\beta^{(k)}$ that is, $G_{\beta,(k)}$ is the distribution of $(\beta \sum_{j\leq k} \xi_j)^{-\beta}$ if $\beta \neq 0$, and of $-\log(\sum_{j\leq k} \xi_j)$ if $\beta = 0$. We suppose that F is ultimately continuous and strictly increasing in its upper tail.

Theorem 2.1.12. *If there exist $a_n > 0$, $b_n \in \mathbb{R}$ such that*

$$\sup_{t\in\mathbb{R}} \left| P((Z_{n-k+1:n} - b_n)/a_n \leq t) - G_{\beta,(k)}(t) \right| \longrightarrow_{n\to\infty} 0$$

for any sequence $k = k(n) \in \{1, \ldots, n\}$, $n \in \mathbb{N}$, with $k/n \longrightarrow_{n\to\infty} 0$, then there exist $c > 0$, $d \in \mathbb{R}$ such that $F(x) = H_\beta(cx + d)$ for x near $\omega(F)$.

The following consequence is obvious.

Corollary 2.1.13. *If there exist constants $a_n > 0$, $b_n \in \mathbb{R}$ such that for any $k \in \{1, \ldots, n\}$, $n \in \mathbb{N}$,*

$$\sup_{t\in\mathbb{R}} \left| P((Z_{n-k+1:n} - b_n)/a_n \leq t) - G_{\beta,(k)}(t) \right| \leq g(k/n),$$

where $g : [0, 1] \to \mathbb{R}$ satisfies $\lim_{x\to 0} g(x) = 0$, then the conclusion of Theorem 2.1.12 holds.

With the particular choice $g(x) = Cx$, $x \in [0, 1]$, the preceding result obviously yields the *only if*-part of Theorem 2.1.11. A multivariate extension of Theorem 2.1.12 and Corollary 2.1.13 will be established in Theorem 5.4.7 and Corollary 5.4.8.

2.2 The δ-Neighborhood of a GPD

Distribution functions F, which satisfy the von Mises condition (VM) from Theorem 2.1.2 with rapidly vanishing remainder term η, are members of certain δ-neighborhoods $Q_i(\delta)$, $i = 1, 2, 3$, of GPD defined below. These classes $Q_i(\delta)$ will be our semiparametric models, underlying the upper tail of F, for statistical inference about extreme quantities such as extreme quantiles of F outside the range of given iid data from F (see Section 2.4).

The Standard Form of GPD

Define for $\alpha > 0$ the following df,

$$W_{1,\alpha}(x) := 1 - x^{-\alpha}, \qquad x \geq 1,$$

which is the usual class of Pareto distributions,

$$W_{2,\alpha}(x) := 1 - (-x)^{\alpha}, \qquad -1 \leq x \leq 0,$$

which consist of certain beta distributions as, e.g., the uniform distribution on $(-1, 0)$ for $\alpha = 1$, and

$$W_3(x) := 1 - \exp(-x), \qquad x \geq 0,$$

the standard exponential distribution.

Notice that $W_i,\ i = 1, 2, 3$, corresponds to $H_\beta,\ \beta > 0,\ \beta < 0,\ \beta = 0$, and we call a df $W \in \{W_{1,\alpha}, W_{2,\alpha}, W_3 : \alpha > 0\}$ a GPD as well. While $H_\beta(x) = 1 + \log(G_\beta(x)),\ x \geq 0$, was derived in Example 1.3.4 from the *von Mises representation*

$$G_\beta(x) = \exp(-(1 + \beta x)^{-1/\beta}), \qquad 1 + \beta x > 0, \quad \beta \in \mathbb{R},$$

of an EVD G_β, the df W_i can equally be derived from an EVD G_i given in its standard form. Put for $i = 1, 2, 3$ and $\alpha > 0$,

$$G_{1,\alpha}(x) := \begin{cases} 0, & x \leq 0 \\ \exp(-x^{-\alpha}), & x > 0, \end{cases}$$

$$G_{2,\alpha}(x) := \begin{cases} \exp(-(-x)^{\alpha}), & x \leq 0 \\ 1, & x > 0, \end{cases}$$

$$G_3(x) := \exp(-e^{-x}), \qquad x \in \mathbb{R},$$

being the Fréchet, (reversed) Weibull and Gumbel distribution. Notice that the Fréchet and Weibull df can be regained from G_β by the equations

$$\begin{aligned} G_{1,1/\beta}(x) &= G_\beta((x-1)/\beta) \\ G_{2,-1/\beta}(x) &= G_\beta(-(x+1)/\beta) \end{aligned} \qquad \text{if} \qquad \begin{aligned} \beta &> 0 \\ \beta &< 0. \end{aligned}$$

Further we have for $G = G_{1,\alpha},\ G_{2,\alpha},\ G_3$ with $\alpha > 0$,

$$W(x) = 1 + \log(G(x)), \qquad \log(G(x)) > -1.$$

While we do explicitly distinguish in our notation between the classes of GPD H_β and W_i, we handle EVD G a bit laxly. But this should cause no confusion in the sequel.

δ-Neighborhoods

Suppose that the df F satisfies condition (VM) with $\beta \in \mathbb{R},\ c > 0$ such that for some $\delta > 0$ the remainder term η satisfies $\eta(x) = O((1 - H_\beta(x))^\delta)$ as $x \to \omega(F)$. Then we know from Proposition 2.1.5 that for some $a > 0,\ b \in \mathbb{R}$, with $b = (a-c)/\beta$ if $\beta \neq 0$ and $a = c$ if $\beta = 0$,

$$\begin{aligned} f(x) &= a h_{\beta/c}(ax+b)\Big(1 + O((1-H_\beta(x))^\delta)\Big) \\ &= \begin{cases} \tilde a w_{1,c/\beta}(\tilde a x)\Big(1+O((1-W_{1,c/\beta}(x))^{\tilde\delta}\Big), & \beta > 0 \\ \tilde a w_{2,-c/\beta}(\tilde a(x-\omega(F)))\Big(1+O((1-W_{2,-c/\beta}(x-\omega(F)))^{\tilde\delta})\Big), & \beta < 0 \\ \tilde a w_3(ax+b)\Big(1+O((1-W_3(ax))^{\delta/c})\Big), & \beta = 0, \end{cases} \end{aligned}$$

for some $\tilde a, \tilde\delta > 0$, where we denote by w the density of W. As a consequence, F is a member of one of the following semiparametric classes $Q_i(\delta),\ i = 1,2,3$ of df. In view of Corollary 2.1.9, theses classes $Q_i(\delta)$, which we call *δ-neighborhoods of GPD*, are therefore quite natural models for the upper tail of a df F. Such classes were first studied by Weiss [457]. Put for $\delta > 0$,

$Q_1(\delta) := \Big\{ F : \omega(F) = \infty$ and F has a density f on $[x_0, \infty)$ for some $x_0 > 0$ such that for some shape parameter $\alpha > 0$ and some scale parameter $a > 0$ on $[x_0, \infty)$,

$$f(x) = \frac{1}{a} w_{1,\alpha}\Big(\frac{x}{a}\Big)\Big(1 + O((1 - W_{1,\alpha}(x))^\delta)\Big)\Big\},$$

$Q_2(\delta) := \Big\{ F : \omega(F) < \infty$ and F has a density f on $[x_0, \omega(F))$ for some $x_0 < \omega(F)$ such that for some shape parameter $\alpha > 0$ and some scale parameter $a > 0$ on $[x_0, \omega(F))$,

$$f(x) = \frac{1}{a} w_{2,\alpha}\Big(\frac{x-\omega(F)}{a}\Big)\Big(1 + O\Big((1 - W_{2,\alpha}(x - \omega(F)))^\delta\Big)\Big)\Big\},$$

$Q_3(\delta) := \Big\{ F : \omega(F) = \infty$ and F has a density f on $[x_0, \infty)$ for some $x_0 > 0$ such that for some scale and location parameters $a > 0,\ b \in \mathbb{R}$ on $[x_0, \infty)$,

$$f(x) = \frac{1}{a} w_3\Big(\frac{x-b}{a}\Big)\Big(1 + O\Big(\Big(1 - W_3\Big(\frac{x}{a}\Big)\Big)^\delta\Big)\Big)\Big\}.$$

We will see that in case $F \in Q_i(\delta)$, $i = 1$ or 2, a suitable data transformation, which does not depend on the shape parameter $\alpha > 0$ and the scale parameter $a > 0$, transposes the underlying df F to $Q_3(\delta)$; this reduces for example the estimation of extreme quantiles of F to the estimation of the scale and location parameters a, b in the family $Q_3(\delta)$ see Section 2.4).

The EVD G_i lies in $Q_i(1)$, $i = 1, 2, 3$. The Cauchy distribution is in $Q_1(1)$, Student's t_n distribution with n degrees of freedom is in $Q_1(2/n)$, a triangular distribution lies in $Q_2(\delta)$ for any $\delta > 0$. Distributions F with upper Gamma tail that is, $f(x) = (c^p/\Gamma(p))e^{-cx}x^{p-1}$, $x \geq x_0 > 0$, with $c, p > 0$ and $p \neq 1$ do not belong to any class $Q_i(\delta)$.

A df F which belongs to one of the classes $Q_i(\delta)$ is obviously tail equivalent to the corresponding GPD $W_{i,\alpha}$ that is,

$$\lim_{x \to \omega(F)} \frac{1 - F(x)}{1 - W_{i,\alpha}((x-b)/a)} = 1 \tag{2.5}$$

for some $a > 0$, $b \in \mathbb{R}$, with $b = 0$ in case $i = 1$ and $b = \omega(F)$ in case $i = 2$. Interpret $W_{3,\alpha}$ simply as W_3, as in the case $i = 3$ there is no shape parameter α. Consequently, we obtain from (2.5)

$$\lim_{q \to 0} \frac{F^{-1}(1-q)}{W_{i,\alpha}((\cdot - b)/a)^{-1}(1-q)} = \lim_{q \to 0} \frac{F^{-1}(1-q)}{aW_{i,\alpha}^{-1}(1-q) + b} = 1, \tag{2.6}$$

and the estimation of large quantiles $F^{-1}(1-q)$ of F that is, for q near 0, then reduces within a certain error bound to the estimation of $aW_{i,\alpha}^{-1}(1-q) + b$.

The following result quantifies the error in (2.5) and (2.6) for a df F in a δ-neighborhood of a GPD.

Proposition 2.2.1. *Suppose that F lies in $Q_i(\delta)$ for some $\delta > 0$ that is, F is tail equivalent to some $W_{i,\alpha}((\cdot - b)/a)$, $i = 1, 2$ or 3, with $b = 0$ if $i = 1$ and $b = \omega(F)$ if $i = 2$. Then,*

(i) $$1 - F(x) = \left(1 - W_{i,\alpha}\left(\frac{x-b}{a}\right)\right)\left(1 + \psi_i(x)\right) \text{ as } x \to \omega(F),$$

where $\psi_i(x)$ decreases to zero at the order $O((1 - W_{i,\alpha}((x-b)/a))^\delta)$. We have in addition

(ii) $$F^{-1}(1-q) = \left(aW_{i,\alpha}^{-1}(1-q) + b\right)(1 + R_i(q)),$$

where

$$R_i(q) = \begin{cases} O(q^\delta) & \text{if } i = 1 \text{ or } 2 \\ O(q^\delta / \log(q)) & \text{if } i = 3 \end{cases}$$

as $q \to 0$. Recall our convention $W_{3,\alpha} = W_3$.

Proof. Part (i) follows from elementary computations. The proof of part (ii) requires a bit more effort. From (i) we deduce the existence of a positive constant K such that, for q near zero with $W_{a,b}(t) := W_{i,\alpha}((t-b)/a)$,

$$\begin{aligned}
F^{-1}(1-q) &= \inf\{t \geq x_q : q \geq 1 - F(t)\} \\
&= \inf\left\{t \geq x_q : q \geq \frac{1-F(t)}{1-W_{a,b}(t)}(1-W_{a,b}(t))\right\} \\
&\begin{cases} \leq \inf\{t \geq x_q : q \geq (1+K\cdot r(t))(1-W_{a,b}(t))\} \\ \geq \inf\{t \geq x_q : q \geq (1-K\cdot r(t))(1-W_{a,b}(t))\}, \end{cases}
\end{aligned}$$

where $r(x) = x^{-\alpha\delta}, |x-\omega(F)|^{\alpha\delta}, \exp(-(\delta/a)x)$ in case $i = 1, 2, 3$, and $x_q \to \omega(F)$ as $q \to 0$. Choose now

$$t_q^- := \begin{cases} aq^{-1/\alpha}(1-K_1q^\delta)^{-1/\alpha} & & i=1 \\ \omega(F) - aq^{1/\alpha}(1-K_1q^\delta)^{1/\alpha} & \text{in case} & i=2 \\ -a\log\{q(1-K_1q^\delta)\} + b & & i=3 \end{cases}$$

and

$$t_q^+ := \begin{cases} aq^{-1/\alpha}(1+K_1q^\delta)^{-1/\alpha} & & i=1 \\ \omega(F) - aq^{1/\alpha}(1+K_1q^\delta)^{1/\alpha} & \text{in case} & i=2 \\ -a\log\{q(1+K_1q^\delta)\} + b & & i=3, \end{cases}$$

for some large positive constant K_1. Then

$$(1+Kr(t_q^-))(1-W_{a,b}(t_q^-)) \leq q \quad \text{and} \quad (1-Kr(t_q^+))(1-W_{a,b}(t_q^+)) > q$$

for q near zero if K_1 is chosen large enough; recall that $b = 0$ in case $i = 1$. Consequently, we obtain for q near zero

$$\begin{aligned}
t_q^+ &\leq \inf\{t \geq x_q : q \geq (1-Kr(t))(1-W_{a,b}(t))\} \\
&\leq F^{-1}(1-q) \\
&\leq \inf\{t \geq x_q : q \geq (1+Kr(t))(1-W_{a,b}(t))\} \leq t_q^-.
\end{aligned}$$

The assertion now follows from the identity

$$W_{a,b}^{-1}(1-q) = aW_{i,\alpha}^{-1}(1-q) + b = \begin{cases} aq^{-1/\alpha} & & i=1 \\ \omega(F) - aq^{1/\alpha} & \text{in case} & i=2 \\ -a\log(q) + b & & i=3 \end{cases}$$

and elementary computations, which show that

$$t_q^+ = W_{a,b}^{-1}(1-q)(1+O(R(q))), \; t_q^- = W_{a,b}^{-1}(1-q)(1+O(R(q))). \qquad \square$$

The approximation of the upper tail $1-F(x)$ for large x by Pareto tails under von Mises conditions on F was discussed by Davis and Resnick [93]. New in the preceding result is the assumption that F lies in a δ-neighborhood of a GPD, which entails the handy error terms in the expansions of the tail and of large quantiles of F in terms of GPD ones. As we have explained above, this assumption $F \in Q_i(\delta)$ is actually a fairly general one.

Data Transformations

Suppose that F is in $Q_1(\delta)$. Then F has ultimately a density f such that, for some $\alpha, a > 0$,

$$f(x) = \frac{1}{a} w_{1,\alpha}\Big(\frac{x}{a}\Big)\Big(1 + O((1 - W_{1,\alpha}(x))^\delta)\Big)$$

as $x \to \infty$. In this case, the df with upper tail

$$F_1(x) := F(\exp(x)), \qquad x \geq x_0, \tag{2.7}$$

is in $Q_3(\delta)$. To be precise, F_1 has ultimately a density f_1 such that

$$\begin{aligned} f_1(x) &= \alpha w_3\Big(\frac{x - \log(a)}{1/\alpha}\Big)\Big(1 + O((1 - W_3(\alpha x))^\delta)\Big) \\ &= \frac{1}{a_0} w_3\Big(\frac{x - b_0}{a_0}\Big)\Big(1 + O\Big(\Big(1 - W_3\Big(\frac{x}{a_0}\Big)\Big)^\delta\Big)\Big), \qquad x \geq x_0, \end{aligned}$$

with $a_0 = 1/\alpha$ and $b_0 = \log(a)$.

If we suppose that F is in $Q_2(\delta)$ that is,

$$f(x) = \frac{1}{a} w_{2,\alpha}\Big(\frac{x - \omega(F)}{a}\Big)\Big(1 + O((1 - W_{2,\alpha}(x - \omega(F)))^\delta)\Big)$$

as $x \to \omega(F) < \infty$ for some $\alpha, a > 0$, then

$$F_2(x) := F(\omega(F) - \exp(-x)), \qquad x \in \mathbb{R}, \tag{2.8}$$

is in $Q_3(\delta)$. The df F_2 has ultimately a density f_2 such that

$$\begin{aligned} f_2(x) &= \alpha w_3\Big(\frac{x + \log(a)}{1/\alpha}\Big)\Big(1 + O((1 - W_3(\alpha x))^\delta)\Big) \\ &= \frac{1}{a_0} w_3\Big(\frac{x - b_0}{a_0}\Big)\Big(1 + O\Big(\Big(1 - W_3\Big(\frac{x}{a_0}\Big)\Big)^\delta\Big)\Big), \qquad x \geq x_0, \end{aligned}$$

with $a_0 = 1/\alpha$ and $b_0 = -\log(a)$.

The message of the preceding considerations can be summarized as follows. Suppose it is known that F is in $Q_1(\delta)$, $Q_2(\delta)$ or in $Q_3(\delta)$, but neither the particular shape parameter α nor the scale parameter a is known in case $F \in Q_i(\delta),\ i = 1, 2$. Then a suitable data transformation which does not depend on α and a results in an underlying df F_i which is in $Q_3(\delta)$. And in $Q_3(\delta)$ the estimation of large quantiles reduces to the estimation of a scale and location parameter for the exponential distribution; this in turn allows the application of standard techniques. Details will be given in the next section. A brief discussion of that case, where F is in $Q_2(\delta)$ but $\omega(F)$ is unknown, is given after Lemma 2.4.3.

If it is assumed that F lies in a δ-neighborhood $Q_i(\delta)$ of a GPD for some $i \in \{1, 2, 3\}$, but the index i is unknown, then an initial estimation of the class index i is necessary. A suggestion based on Pickands [371] estimator of the extreme value index α is discussed in Section 2.5.

Joint Asymptotic Distribution of Extremes

The following result describes the set of possible limiting distributions of the joint distribution of the k largest order statistics $Z_{n:n} \ge \cdots \ge Z_{n-k+1:n}$, equally standardized, in an iid sample $Z_1, \dots, Z_n$. By $\longrightarrow_{\mathcal{D}}$ we denote the usual weak convergence.

Theorem 2.2.2 (Dwass [117]). *Let $Z_1, Z_2, \dots$ be iid rv. Then we have for an EVD G and norming constants $a_n > 0$, $b_n \in \mathbb{R}$,*

$$\begin{aligned} &\frac{Z_{n:n} - b_n}{a_n} \longrightarrow_D G \\ \iff\ &\Big(\frac{Z_{n-i+1:n} - b_n}{a_n}\Big)_{i\le k} \longrightarrow_D G^{(k)} \qquad \text{for any} \quad k \in \mathbb{N}, \end{aligned}$$

where the distribution $G^{(k)}/\mathbb{B}^k$ has Lebesgue density $g^{(k)}(x_1, \dots, x_k) = G(x_k) \prod_{i\le k} G'(x_i)/G(x_i)$ for $x_1 > \cdots > x_k$ and zero elsewhere.

Remark 2.2.3. Let $\xi_1, \xi_2, \dots$ be a sequence of independent and standard exponential rv. Then $G^{(k)}_{1,\alpha}$ is the distribution of $((\sum_{j\le i} \xi_j)^{-1/\alpha})_{i\le k}$, $G^{(k)}_{2,\alpha}$ that of $(-(\sum_{j\le i} \xi_j)^{1/\alpha})_{i\le k}$ and $G^{(k)}_3$ that of $(-\log(\sum_{j\le i} \xi_j))_{i\le k}$. This representation was already utilized in Theorems 2.1.10-2.1.12.

Proof of Theorem 2.2.2. We have to show the only-if part of the assertion. Consider without loss of generality $Z_i = F^{-1}(U_i)$, where $U_1, U_2, \dots$ are independent and uniformly on (0,1) distributed rv, and where F denotes the df of Z_i. Then we have the representation $(Z_{i:n})_{i\le n} = (F^{-1}(U_{i:n}))_{i\le n}$, and by the equivalence

$$F^{-1}(q) \le t \iff q \le F(t), \qquad q \in (0,1), \quad t \in \mathbb{R},$$

we can write

$$\begin{aligned} &P\Big((Z_{n-i+1:n} - b_n)/a_n \le x_i,\ 1 \le i \le k\Big) \\ &= P\Big(F^{-1}(U_{n-i+1:n}) \le a_n x_i + b_n,\ 1 \le i \le k\Big) \\ &= P\Big(U_{n-i+1:n} \le F(a_n x_i + b_n),\ 1 \le i \le k\Big) \\ &= P\Big(n(U_{n-i+1:n} - 1) \le n(F(a_n x_i + b_n) - 1),\ 1 \le i \le k\Big). \end{aligned}$$

As the convergence $F^n(a_n x + b_n) \longrightarrow_{n\to\infty} G(x)$, $x \in \mathbb{R}$, is equivalent to $n(F(a_n x + b_n) - 1) \longrightarrow_{n\to\infty} \log(G(x))$, $0 < G(x) \le 1$, and, as is easy to see, $(n(U_{n-i+1:n} - 1))_{i\le k} \longrightarrow_D G^{(k)}_{2,1}$ with density $g^{(k)}_{2,1}(x_1, \dots, x_k) = \exp(x_k)$ if $0 > x_1 > \cdots > x_k$ and 0 elsewhere, we obtain

$$P\Big((Z_{n-i+1:n} - b_n)/a_n \le x_i,\ 1 \le i \le k\Big) \longrightarrow_{n\to\infty} G^{(k)}_{2,1}\Big((\log(G(x_i)))_{i\le k}\Big).$$

This implies the assertion. $\square$

For a proof of the following result, which provides a rate of convergence in the preceding theorem if the upper tail of the underlying distribution is in a δ-neighborhood of a GPD, we refer to Corollary 5.5.5 of Reiss [385] (cf. also Theorem 2.1.10).

Theorem 2.2.4. *Suppose that the df F is in a δ-neighborhood $Q_i(\delta)$ of a GPD $W_i = W_{1,\alpha}$, $i = 1, 2$ or 3. Then there obviously exist constants $a > 0$, $b \in \mathbb{R}$, with $b = 0$ if $i = 1$, $b = \omega(F)$ if $i = 2$, such that*

$$af(ax+b) = w_i(x)\Big(1 + O((1 - W_i(x))^{\delta})\Big) \tag{2.9}$$

for all x in a left neighborhood of $\omega(W_{i,\alpha})$. Consequently, we obtain from Corollary 5.5.5 in Reiss [385]

$$\sup_{B\in\mathbb{B}^k} \left| P\Big(\Big(\Big(\frac{Z_{n-j+1:n} - b}{a} - d_n\Big)/c_n\Big)_{j\le k} \in B\Big) - G^{(k)}(B)\right|$$

$$= O((k/n)^{\delta} k^{1/2} + k/n),$$

where $d_n = 0$ for $i = 1, 2$; $d_n = \log(n)$ for $i = 3$; $c_n = n^{1/\alpha}, n^{-1/\alpha}$, 1 for $i = 1, 2, 3$.

Notice that df F whose upper tails *coincide* with that of a GPD, are actually the only ones where the term $(k/n)^{\delta} k^{1/2}$ in the preceding bound can be dropped (cf. Theorem 2.1.11). This is indicated by Theorem 2.2.4, as δ can then and only then be chosen arbitrarily large.

Summarizing the Results

The following list of equivalences now follows from Proposition 2.1.4, 2.1.5 and Theorem 2.1.8, 2.2.4. They summarize our considerations of this section and the preceding one.

Theorem 2.2.5. *Suppose that F satisfies condition (VM) from the preceding section with $\beta \in \mathbb{R}$ and $c > 0$, such that the remainder function $\eta(x)$ is proportional to some monotone function as $x \to \omega(F) = \omega(H_\beta)$ and $\int_{x_0}^{\omega(F)} \eta(t)/(1 + \beta t)\, dt \in \mathbb{R}$. Then there exist $a > 0$, $b \in \mathbb{R}$ with $b = -1/\beta$ if $\beta \ne 0$, such that*

$$\lim_{x\uparrow\omega(W_i)} \frac{1 - F(ax+b)}{1 - W_i(x)} = \lim_{x\uparrow\omega(W_i)} \frac{af(ax+b)}{w_i(x)} = 1,$$

where $i = 1, 2, 3$ if $\beta > 0, < 0, = 0$ and $W_i = W_{1,c/\beta}, W_{2,c/\beta}, W_3$. Consequently, with c_n, d_n as in the preceding result

$$\sup_{x\in\mathbb{R}} \left| P\Big(\Big(\frac{Z_{n:n} - b}{a} - d_n\Big)/c_n \le x\Big) - G_i(x)\right| \longrightarrow_{n\to\infty} 0,$$

where $Z_1, \dots, Z_n$ are iid with common df F. Moreover, we have the following list of equivalences:

$$\begin{aligned}
&\sup_{x\in\mathbb{R}} \Big| P\Big(\Big(\frac{Z_{n:n}-b}{a} - d_n\Big)/c_n \le x\Big) - G_i(x)\Big| = O(n^{-\delta}) \text{ for some } \delta > 0\\
\iff &\text{ there exists } \delta > 0 \text{ such that for } x \to \omega(F)\\
&\qquad \eta(x) = O((1-H_\beta(x))^\delta)\\
\iff &\ F \text{ is in a } \delta\text{-neighborhood } Q_i(\delta) \text{ of the GPD } W_i\\
\iff &\text{ there exists } \delta > 0 \text{ such that, for } k \in \{1, \dots, n\},\ n \in \mathbb{N},\\
&\qquad \sup_{B\in\mathbb{B}^k} \Big| P\Big(\Big(\Big(\frac{Z_{n-j+1:n}-b}{a} - d_n\Big)/c_n\Big)_{j\le k} \in B\Big) - G^{(k)}(B)\Big|\\
&\qquad = O\Big((k/n)^\delta k^{1/2} + k/n\Big).
\end{aligned}$$

2.3 The Peaks-Over-Threshold Method

The following example seems to represent one of the first applications of the POT approach (de Haan [189]).

Example 2.3.1. After the disastrous flood of February 1st, 1953, in which the sea-dikes broke in several parts of the Netherlands and nearly two thousand people were killed, the Dutch government appointed a committee (so-called Delta-committee) to recommend an appropriate level for the dikes (called Delta-level since) since no specific statistical study had been done to fix a safer level for the sea-dikes before 1953. The Dutch government set as the standard for the sea-dikes that at any time in a given year the sea level exceeds the level of the dikes with probability 1/10,000. A statistical group from the Mathematical Centre in Amsterdam headed by D. van Dantzig showed that high tides occurring during certain dangerous windstorms (to ensure independence) within the dangerous winter months December, January and February (for homogeneity) follow closely an exponential distribution if the smaller high tides are neglected.

If we model the annual maximum flood by a rv Z, the Dutch government wanted to determine therefore the $(1-q)$-quantile

$$F^{-1}(1-q) = \inf\{t \in \mathbb{R} : F(t) \ge 1-q\}$$

of Z, where F denotes the df of Z and q has the value 10^{-4}.

The Point Process of Exceedances

From the past we have observations $Z_1, \dots, Z_n$ (annual maximum floods), which we assume to be independent replicates of Z. With these rv we define the truncated empirical point process

$$N_n^{(t)}(\cdot) := \sum_{j\le n} \varepsilon_{Z_j}(\cdot \cap (t,\infty))$$

that is, we consider only those observations which exceed the *threshold* t. The process $N_n^{(t)}$ is therefore called the point process of the exceedances.

From Theorem 1.3.1 we know that we can write

$$N_n^{(t)}(\cdot) = \sum_{j\le K_t(n)} \varepsilon_{V_j^{(t)}+t}(\cdot),$$

where the *excesses* $V_1^{(t)}, V_2^{(t)}, \dots$ are independent replicates of a rv $V^{(t)}$ with df $F^{(t)}(\cdot) := P(Z \le t + \cdot | Z \ge t)$, and these are independent of the sample size $K_t(n) := \sum_{i\le n} \varepsilon_{Z_i}((t,\infty))$.

Without specific assumptions, the problem to determine $F^{-1}(1-q)$ is a non-parametric one. If we require however that the underlying df F is in a δ-neighborhood of a GPD, then this non-parametric problem can be approximated within a reasonable error bound by a parametric one.

Approximation of Excess Distributions

Suppose therefore that the df F of Z is in a δ-neighborhood $Q_i(\delta)$ of a GPD W_i that is, there exist δ, $a > 0$, $b \in \mathbb{R}$, with $b = 0$ if $i = 1$ and $b = \omega(F)$ if $i = 2$, such that, for $x \to \omega(F)$,

$$f(x) = \frac{1}{a} w_i\Big(\frac{x-b}{a}\Big)\Big(1 + O\Big(\Big(1 - W_i\Big(\frac{x-b}{a}\Big)\Big)^{\delta}\Big)\Big),$$

where F has density f in a left neighborhood of $\omega(F)$.

In this case, the df $F^{(t)}(s)$, $s \ge 0$, of the excess $V^{(t)}$ has density $f^{(t)}$ for all t in a left neighborhood of $\omega(F)$, with the representation

$$\begin{aligned} f^{(t)}(s) &= \frac{f(t+s)}{1-F(t)} \\ &= \frac{\frac{1}{a}w_i(\frac{t+s-b}{a})}{1-W_i(\frac{t-b}{a})}\;\frac{1+O((1-W_i(\frac{t+s-b)}{a}))^{\delta})}{1+O((1-W_i(\frac{t-b)}{a}))^{\delta})} \\ &= \frac{\frac{1}{a}w_i(\frac{t+s-b}{a})}{1-W_i(\frac{t-b}{a})}\Big(1 + O\Big(\Big(1 - W_i\Big(\frac{t-b}{a}\Big)\Big)^{\delta}\Big)\Big), \quad s \ge 0. \end{aligned}$$

Note that $a^{-1}w_i((t+s-b)/a)/(1-W_i((t-b)/a)$, $s \ge 0$, with $0 < W_i((t-b)/a) < 1$ and $b = 0$ if $i = 1$, $b = \omega(F)$ if $i = 2$, is again the density of a GPD $W_i^{(t)}$, precisely of

$$W_i^{(t)}(s) = \begin{cases} W_1\Big(1+\frac{s}{t}\Big) & \qquad i=1 \\ W_2\Big(-1+\frac{s}{\omega(F)-t}\Big) & \text{if} \quad i=2, \quad s \ge 0, \\ W_3\Big(\frac{s}{a}\Big) & \qquad i=3. \end{cases}$$

We can consequently approximate the truncated empirical point process

$$\begin{aligned} N_n^{(t)}(\cdot) &= \sum_{j \le n} \varepsilon_{Z_j}(\cdot \cap (t,\infty)) \\ &= \sum_{j \le K_t(n)} \varepsilon_{V_j^{(t)}+t}(\cdot), \end{aligned}$$

pertaining to the $K_t(n)$ exceedances $V_1^{(t)}+t,\dots,V_{K_t(n)}^{(t)}+t$ over the threshold t, by the binomial point process

$$M_n^{(t)} = \sum_{j \le K_t(n)} \varepsilon_{c\xi_j+d+t},$$

where $c=t,\ d=-t$ in case $i=1,\ c=d=\omega(F)-t$ in case $i=2$ and $c=a,\ d=0$ in case $i=3$, and $\xi_1,\xi_2\dots$ are independent copies of a rv ξ having df W_i, and independent also from their random counting number $K_t(n)$.

Bounds for the Process Approximations

Choose the particular threshold

$$t = aW_i^{-1}\Big(1-\frac{r}{n}\Big)+b,$$

with r/n less than a suitable positive constant c_0 such that t is in a proper left neighborhood of $\omega(F)=a\,\omega(W_i)+b$. By Corollary 1.2.4 (iv) we obtain for the Hellinger distance $H(N_n^{(t)},M_n^{(t)})$ between $N_n^{(t)}$ and $M_n^{(t)}$ uniformly for $0<r/n<c_0$ the bound

$$\begin{aligned} H(N_n^{(t)},M_n^{(t)}) &\le H(V^{(t)},c\xi+d)\,(E(K_t(n)))^{1/2} \\ &= O((1-W_i((t-b)/a))^{\delta})\,(E(K_t(n)))^{1/2} \\ &= O((r/n)^{\delta}\,(n(1-F(t)))^{1/2}) \;=\; O((r/n)^{\delta}\,r^{1/2}). \end{aligned}$$

As the Hellinger distance is in general bounded by $2^{1/2}$, we can drop the assumption $r/n \le c_0$ and the preceding bound is therefore true uniformly for $0<r<n$.

The preceding inequality explains the exponential fit of the high tides by van Dantzig in Example 2.3.1, if the smaller high tides are neglected. This peaks-over-threshold method is not only widely used by hydrologists to model large floods (Smith [420], Davison and Smith [94]), but also in insurance mathematics for modeling large claims (Teugels [441], [442], Kremer [297], Reiss [384]). For thorough discussions of the peaks-over-threshold approach in the investigation of extreme values and further references we refer to Section 6.5 of Embrechts et al. [122], Section 4 of Coles [71] and to Reiss and Thomas [389].

Replacing $M_n^{(t)}$ by the Poisson process

$$N_n^{(t)**}(\cdot) := \sum_{j \le \tau_t(n)} \varepsilon_{c\xi_j+d+t}(\cdot),$$

with $\tau_t(n)$ being a Poisson rv with parameter $n(1-F(t))$, we obtain therefore by Theorem 1.2.5 the following bound for the second-order Poisson process approximation of $N_n^{(t)}$ by $N_n^{(t)**}$,

$$\begin{aligned} H(N_n^{(t)}, N_n^{(t)**}) &= O\Big(r^{1/2}(r/n)^\delta + (1-F(t))\Big) \\ &= O\Big(r^{1/2}(r/n)^\delta + r/n\Big), \end{aligned}$$

uniformly for $0 < r < n$ and $n \in \mathbb{N}$.

The preceding considerations are summarized in the following result providing bounds for functional laws of small numbers in an EVD model.

Theorem 2.3.2. *Suppose that F is in the δ-neighborhood $Q_i(\delta)$ of some GPD W_i, $i = 1, 2$ or 3. Then there exist $a > 0$, $b \in \mathbb{R}$, with $b = 0$ in case $i = 1$ and $b = \omega(F)$ in case $i = 2$ such that*

$$\lim_{x \uparrow \omega(W_i)} \frac{1 - F(ax+b)}{1 - W_i(x)} = 1.$$

Define for $r \in (0, n)$ the threshold

$$t := t(n) := aW_i^{-1}\Big(1 - \frac{r}{n}\Big) + b$$

and denote by

$$N_n^{(t)} = \sum_{j \le n} \varepsilon_{Z_j}(\cdot \cap (t, \infty)) = \sum_{j \le K_t(n)} \varepsilon_{V_j^{(t)}+t}(\cdot)$$

the point process of the exceedances among $Z_1, \dots, Z_n$ over t.

Define the binomial process

$$M_n^{(t)} := \sum_{j \le K_t(n)} \varepsilon_{c\xi_j+d+t}$$

and the Poisson processes

$$N_n^{(t)*} := \sum_{j \le \tau_t(n)} \varepsilon_{V_j^{(t)}+t}, \; N_n^{(t)**} := \sum_{j \le \tau_t(n)} \varepsilon_{c\xi_j+d+t},$$

where $c = t,\ d = -t,$ *if* $i = 1,\ c = d = \omega(t) - t$ *if* $i = 1,\ c = a,\ d = 0$ *if* $i = 3;\ \xi_1, \xi_2, \dots$ *are iid rv with common df* W_i *and* $\tau_t(n)$ *is a Poisson rv with parameter* $n(1 - F(t))$, *independent of the sequences* $\xi_1, \xi_2, \dots$ *and of* $V_1^{(t)}, V_2^{(t)}, \dots$ *Then we have the following bounds, uniformly for* $0 < r < n$ *and* $n \in \mathbb{N}$,

$$H(N_n^{(t)}, M_n^{(t)}) = O(r^{1/2}(r/n)^\delta)$$

for the POT method,

$$H(N_n^{(t)}, N_n^{(t)*}) = O(r/n)$$

for the first-order Poisson process approximation and

$$H(N_n^{(t)}, N_n^{(t)**}) = O(r/n + r^{1/2}(r/n)^\delta)$$

for the second-order Poisson process approximation.

A binomial process approximation with an error bound based on the remainder function of the von Mises condition (VM_0) in Theorem 2.1.2 was established by Kaufmann and Reiss [287] (cf. also [389], 2nd ed., Section 6.4).

2.4 Parameter Estimation in δ-Neighborhoods of GPD

Suppose that we are given an iid sample of size n from a df F, which lies in a δ-neighborhood $Q_i(\delta)$ of a GPD W_i. Then there exist $\alpha,\ a > 0,\ b \in \mathbb{R}$, with $b = 0$ if $i = 1$ and $b = \omega(F)$ if $i = 2$ such that $F(x)$ and $1 - W_{i,\alpha}((x - b)/a)$ are tail equivalent. Interpret again $W_{3,\alpha}$ as W_3. We assume that the class index $i = 1, 2, 3$ and $\omega(F)$ are known. As shown in (2.7) and (2.8) in Section 2.2, a suitable data transformation, which does not depend on α or a, transposes $F \in Q_i(\delta),\ i = 1$ or 2, to a df F_i which is in $Q_3(\delta)$. And in $Q_3(\delta)$ the estimation of upper tails reduces to the estimation of a scale and location parameter $a_0 > 0,\ b_0 \in \mathbb{R}$ for the exponential distribution, which in turn allows the application of standard techniques. A brief discussion of that case, where F is in $Q_2(\delta)$ but $\omega(F)$ is unknown, is given after Lemma 2.4.3.

If it is assumed that F lies in a δ-neighborhood $Q_i(\delta)$ of a GPD, but the class index i is unknown, then an initial estimation of the class index i is necessary. A suggestion based on Pickands [371] estimator of the extremal index α is discussed in the next section.

Our considerations are close in spirit to Weissman [458], who considers n iid observations with common df F being the EVD $G_3((x - b_0)/a_0)$ with unknown scale and location parameter $a_0 > 0,\ b_0 \in \mathbb{R}$. Based on the upper k order statistics in the sample, he defines maximum-likelihood and UMVU estimators of a_0 and b_0 and resulting estimators of extreme quantiles $F^{-1}(1-c/n)$. Equally, he proposes the data transformations (2.7) and (2.8) in case $F = G_{1,\alpha}$ or $G_{2,\alpha}$ but considers no asymptotics.

Viewing F as an element of $Q_i(\delta)$, we can establish asymptotics for UMVU estimators of a, b and of resulting estimators of extreme quantiles $F^{-1}(1-q_n)$ with $q_n \to 0$ and $k = k(n) \to \infty$ as $n \to \infty$. It follows in particular from Corollary 2.4.6 that the error of the resulting estimator of $F^{-1}(1-q_n)$ is of the order $O_p(q_n^{-\gamma(i)} k^{-1/2}(\log^2(nq_n/k)+1)^{1/2})$, where $\gamma(i) = 1/\alpha, -1/\alpha, 0$ if $F \in Q_i(\delta),\ i = 1, 2, 3$.

This demonstrates the superiority of the estimators to the ones proposed by Dekkers and de Haan [108] and Dekkers et al. [107], if nq_n is of smaller order than k. Note however that our estimators are based on the assumption that the class index i of the condition $F \in Q_i(\delta)$ is known, whereas those estimators proposed by Dekkers et al. are uniformly consistent.

The Basic Approximation Lemma

The following consequence of Theorem 2.2.4 is crucial.

Lemma 2.4.1. *Suppose that F is in $Q_3(\delta)$ for some $\delta > 0$. Then there exist $a_0 > 0,\ b_0 \in \mathbb{R}$ such that*

$$\begin{aligned}
\sup_{B \in \mathbb{B}^{k+1}} \Big| & P\Big(((Z_{n-j+1:n} - Z_{n-k:n})_{j=k}^{1},\ Z_{n-k:n}) \in B\Big) \\
& \qquad - P\Big(\Big((a_0 X_{j:k})_{j \le k},\ (a_0/k^{1/2})Y + a_0 \log(n/k) + b_0\Big) \in B\Big)\Big| \\
= {} & O(k/n + (k/n)^\delta k^{1/2} + k^{-1/2}),
\end{aligned}$$

where $Y, X_1, \dots, X_k$ are independent rv, Y is standard normal and $X_1, \dots, X_k$ are standard exponential distributed.

Proof. By Theorem 2.2.4 there exist $a_0 > 0$ and $b_0 \in \mathbb{R}$ such that

$$\begin{aligned}
\sup_{B \in \mathbb{B}^k} & \Big| P\Big(((Z_{n-j+1:n} - b_0)/a_0 - \log(n))_{j \le k} \in B\Big) - G_3^{(k+1)}(B)\Big| \\
& = O((k/n)^\delta k^{1/2} + k/n).
\end{aligned}$$

Recall that $G_3^{(k+1)}$ is the distribution of the vector $(-\log(\sum_{j \le r} \xi_j))_{r \le k+1}$, where $\xi_1, \xi_2, \dots$ are independent and standard exponential distributed rv (Remark 2.2.3). Within the preceding bound, the rv $((Z_{n-j+1:n} - Z_{n-k:n})/a_0)_{j=k}^{1},\ (Z_{n-k:n} - b_0)/a_0 - \log(n))$ behaves therefore like

$$\left(\left(-\log\Big(\sum_{j\le r}\xi_j\Big)+\log\Big(\sum_{j\le k+1}\xi_j\Big)\right)_{r=k}^{1},\ -\log\Big(\sum_{j\le k+1}\xi_j\Big)\right)$$
$$=\left(\left(-\log\Big(\sum_{j\le r}\xi_j/\sum_{j\le k+1}\xi_j\Big)\right)_{r=k}^{1},\ -\log\Big(\sum_{j\le k+1}\xi_j\Big)\right)$$
$$=_D\left((X_{r:k})_{r\le k},\ -\log\Big(\sum_{j\le k+1}\xi_j\Big)\right),$$

where $X_1, X_2, \dots, \xi_1, \xi_2, \dots$ are independent sets of independent standard exponential rv. By $=_D$ we denote equality of distributions. This follows from the facts that $(\sum_{j\le r}\xi_j/\sum_{j\le k+1}\xi_j)_{r\le k}$ and $\sum_{j\le k+1}\xi_j$ are independent (Lemma 1.6.6 in Reiss [385]), that $(\sum_{j\le r}\xi_j/\sum_{j\le k+1}\xi_j)_{r\le k} =_D (U_{r:k})_{r\le k}$, where $U_1, \dots, U_k$ are independent and uniformly on (0,1) distributed rv (Corollary 1.6.9 in Reiss [385]), and that $-\log(1-U)$ is a standard exponential rv if U is uniformly on (0,1) distributed. Finally it is straightforward to show that $-\log(\sum_{j\le k+1}\xi_j)$ is in variational distance within the bound $O(k^{-1/2})$ distributed like $Y/k^{1/2} - \log(k)$. □

The preceding result shows that within a certain error bound depending on δ, the k excesses $(Z_{n-j+1:n} - Z_{n-k:n})_{j=k}^1$ over the *random threshold* $Z_{n-k:n}$ can be handled in case $F \in Q_3(\delta)$ like a complete set $(a_0 X_{j:k})_{j\le k}$ of order statistics from an exponential distribution with unknown scale parameter $a_0 > 0$, whereas the random threshold $Z_{n-k:n}$ behaves like $a_0 k^{-1/2} Y + a_0\log(n/k) + b_0$, where Y is a standard normal rv being independent of $(X_{j:k})_{j\le k}$. Notice that no information from $(Z_{n-j+1:n})_{j\le k+1}$ is lost if we consider $((Z_{n-j+1:n} - Z_{n-k:n})_{j=k}^1, Z_{n-k:n})$ instead.

Efficient Estimators of a_0 and b_0

After the transition to the model $((a_0X_{j:k})_{j\le k}, (a_0/k^{1/2})Y + a_0\log(n/k) + b_0)$, we search for efficient estimators of a_0 and b_0 within this model.

Ad hoc estimators of the parameters $a_0 > 0$, $b_0 \in \mathbb{R}$ in the model

$$\{(V_{j:k})_{j\le k}, \xi\}$$
$$=\Big\{(a_0X_{j:k})_{j\le k},\ (a_0/k^{1/2})Y + a_0\log(n/k) + b_0 : a_0 > 0,\ b_0 \in \mathbb{R}\Big\}$$

are

$$\hat a_k := k^{-1}\sum_{j\le k} V_{j:k}$$

and

$$\hat b_{k,n} := \xi - \hat a_k \log(n/k).$$

The joint density f_{a_0,b_0} of $((V_{j:k})_{j\le k},\xi)$ is

$$f_{a_0,b_0}(\boldsymbol{x},y) = \frac{k!\,k^{1/2}}{a_0^{k+1}(2\pi)^{1/2}}\exp\Big(-a_0^{-1}\sum_{j\le k}x_j\Big)\\ \times\exp\Big(-\frac{(y-a_0\log(n/k)-b_0)^2}{2a_0^2/k}\Big),$$

for $\boldsymbol{x}=(x_1,\dots,x_k)\in\mathbb{R}^k$, if $0<x_1<\dots<x_k$, $y\in\mathbb{R}$, and zero elsewhere (Example 1.4.2 (i) in Reiss [385]). This representation implies with respect to the family $\mathcal{P}:=\{f_{a_0,b_0}:a_0>0,\ b_0\in\mathbb{R}\}$. It is straightforward to show that $\mathcal{P}$ is an exponential family, and by using standard arguments from the theory of such families (see e.g. Chapter 3 of the book by Witting [463]), it is elementary to prove that $(\hat a_k,\hat b_{k,n})$ is a *complete* statistic as well. Altogether we have the following result.

Proposition 2.4.2. *The estimators $\hat a_k,\hat b_{k,n}$ are UMVU (uniformly minimum variance unbiased) estimators of a_0,b_0 for the family $\mathcal{P}=\{f_{a_0,b_0}:a_0>0,\ b_0\in\mathbb{R}\}$.*

It is straightforward to show that $k^{-1/2}\sum_{i\le k}(X_i-1)$ approaches the standard normal distribution $N(0,1)$ within the error bound $O(k^{-1/2})$ in variational distance. The following auxiliary result is therefore obvious.

Lemma 2.4.3. *We have uniformly in $\mathcal{P}$ the bound*

$$\sup_{B\in\mathbb{B}^2}\Big|P\Big(((k^{1/2}(\hat a_k-a_0),\ (k^{1/2}/\log(n/k))(\hat b_{k,n}-b_0))\in B\Big)\\ -P\Big((a_0\xi_1,(a_0/\log(n/k))\xi_2-a_0\xi_1)\in B\Big)\Big|\\ =O(k^{-1/2}),$$

where ξ_1,ξ_2 are independent standard normal rv.

Hill's Estimator and Friends

If we plug our initial data $(Z_{n-j+1:n}-Z_{n-k:n})_{j=k}^1$, $Z_{n-k:n}$ into $\hat a_k$ and $\hat b_{k,n}$, we obtain the estimators

$$\hat a_{n,3}:=\hat a_k((Z_{n-j+1:n}-Z_{n-k:n})_{j=k}^1)\\ =k^{-1}\sum_{j\le k}Z_{n-j+1:n}-Z_{n-k:n},$$

and

$$\hat b_{n,3}:=Z_{n-k:n}-\log(n/k)\hat a_{n,3}$$

of a_0 and b_0 in case $F\in Q_3(\delta)$.

If we suppose that F is in $Q_1(\delta)$, then we know already that $F_1(x) = F(\exp(x))$ is in $Q_3(\delta)$, where the initial shape α parameter of F becomes the scale parameter $a_0 = 1/\alpha$ (cf. (2.7)).

We replace therefore $Z_{j:n}$ in this case by the log-transformed data $\log(Z_{j:n} \wedge 1) = \log(\max\{Z_{j:n}, 1\})$ and define the estimators

$$\begin{aligned}\hat{a}_{n,1} &:= \hat{a}_{n,3}\Big((\log(Z_{n-j+1:n} \wedge 1) - \log(Z_{n-k:n} \wedge 1))_{j=k}^{1}\Big)\\ &= k^{-1}\sum_{j\le k}\log(Z_{n-j+1:n} \wedge 1) - \log(Z_{n-k:n} \wedge 1),\end{aligned}$$

and

$$\begin{aligned}\hat{b}_{n,1} &:= \hat{b}_{n,3}\Big((\log(Z_{n-j+1:n} \wedge 1) - \log(Z_{n-k:n} \wedge 1))_{j=k}^{1},\ \log(Z_{n-k:n} \wedge 1)\Big)\\ &= \log(Z_{n-k:n} \wedge 1) - \log(n/k)\hat{a}_{n,1}\end{aligned}$$

of a_0 and b_0. The estimator $\hat{a}_{n,1}$ is known as the *Hill estimator* (Hill [217]). It actually estimates $1/\alpha$, the reciprocal of the initial shape parameter α of F. Note that the upper tail of the df of $Z \wedge 1$ and of Z coincide as $\omega(F) = \infty$. Asymptotic normality of $k^{1/2}(\hat{a}_{n,1} - 1/\alpha)$ with mean 0 and variance $1/\alpha^2$ under suitable conditions on F and the sequence $k = k(n)$ is well known (see, for example, Hall [202], Csörgő and Mason [84], Hall and Welsh [204], Häusler and Teugels [212]). For a thorough discussion of the Hill estimator we refer to Section 6.4 of Embrechts et al. [122].

If the underlying F is in $Q_2(\delta)$, the transformation $-\log(\omega(F) - Z_j)$ of our initial data Z_j leads us back to $Q_3(\delta)$ with particular scale and location parameters $a_0 > 0,\ b_0 \in \mathbb{R}$ (cf. (2.8)). The pertaining estimators are now

$$\begin{aligned}\hat{a}_{n,2} &:= \hat{a}_{n,3}\Big((-\log(\omega(F) - Z_{n-j+1:n}) + \log(\omega(F) - Z_{n-k:n}))_{j=k}^{1}\Big)\\ &= \log(\omega(F) - Z_{n-k:n}) - k^{-1}\sum_{j\le k}\log(\omega(F) - Z_{n-j+1:n})\end{aligned}$$

and

$$\hat{b}_{n,2} := -\log(\omega(F) - Z_{n-k:n}) - \log(n/k)\hat{a}_{n,2}.$$

If the endpoint $\omega(F)$ of F is finite but unknown, then we can replace the transformation $-\log(\omega(F) - Z_{j:n})$ of our initial data Z_j by the data-driven transformation $-\log(Z_{n:n} - Z_{j:n})$ and j running from 1 to $n-1$. This yields the modified versions

$$\hat{a}'_{n,2} := \log\Big(Z_{n:n} - Z_{n-k:n}\Big) - (k-1)^{-1}\sum_{2\le j\le k}\log(Z_{n:n} - Z_{n-j+1:n})$$

and

$$\hat{b}'_{n,2} := -\log\Big(Z_{n:n} - Z_{n-k:n}\Big) - \log(n/k)\hat{a}'_{n,2}$$

of the estimators $\hat{a}_{n,2}$ and $\hat{b}_{n,2}$.

One can show (cf. Falk [134]) that in case $0 < \alpha < 2$, the data-driven estimators $\hat{a}'_{n,2}, \hat{b}'_{n,2}$ perform asymptotically as good as their counterparts $\hat{a}_{n,2}, \hat{b}_{n,2}$ with known $\omega(F)$. Precisely, if $k = k(n)$ satisfies $k/n \to 0$, $\log(n)/k^{1/2} \to 0$ as n tends to infinity, we have

$$k^{1/2}|\hat{a}_{n,2} - \hat{a}'_{n,2}| = o_P(1)$$

and

$$(k^{1/2}/\log(n/k))|\hat{b}_{n,2} - \hat{b}'_{n,2}| = o_P(1).$$

As a consequence, the asymptotic normality of $(\hat{a}_{n,2},\, \hat{b}_{n,2})$, which follows from the next result if in addition $(k/n)^\delta k^{1/2} \to 0$ as n increases, carries over to $(\hat{a}'_{n,2},\, \hat{b}'_{n,2})$. If $\alpha \geq 2$, then maximum likelihood estimators of $\omega(F)$, a and $1/\alpha$ can be obtained, based on an increasing number of upper-order statistics. We refer to Hall [201] and, in case α known, to Csörgő and Mason [85]. For a discussion of maximum likelihood estimation of general EVD we refer to Section 6.3.1 of Embrechts et al. [122], Section 1.7.5 of Kotz and Nadarajah [293] and to Section 4.1 of Reiss and Thomas [389]. The following result summarizes the preceding considerations and Proposition 2.2.1.

Theorem 2.4.4. *Suppose that F is in $Q_i(\delta)$, $i = 1, 2$ or 3 for some $\delta > 0$ that is, F is in particular tail equivalent to a GPD $W_{i,\alpha}((x-b)/a)$, where $b = 0$ if $i = 1$ and $b = \omega(F)$ if $i = 2$. Then we have in case*

$$\left.\begin{array}{lll} i=1: & 1-F_1(x) & = 1-F(\exp(x)) \\ i=2: & 1-F_2(x) & = 1-F(\omega(F)-\exp(-x)) \\ i=3: & 1-F_3(x) & := 1-F(x) \end{array}\right\}$$

$$= (1 - W_3((x-b_0)/a_0))\Big(1 + O(\exp(-(\delta/a_0)x))\Big)$$

with $a_0 = 1/\alpha$, $b_0 = \log(a)$ if $i = 1$; $a_0 = 1/\alpha$, $b_0 = -\log(a)$ if $i = 2$ and $a_0 = a$, $b_0 = b$ if $i = 3$. Furthermore,

$$f_3(x) = a_0^{-1} w_3((x-b_0)/a_0)\Big(1 + O(\exp(-(\delta/a_0)x))\Big)$$

for $x \to \infty$ and

$$F_i^{-1}(1-q) = \Big((a_0 W_3^{-1}(1-q) + b_0\Big)\Big(1 + O(q^\delta/\log(q))\Big),\ i = 1, 2, 3$$

as $q \to 0$. Finally, we have for $i = 1, 2, 3$ the representations

$$\begin{aligned} \sup_{B \in \mathbb{B}^2} \Big| P\Big(\Big(k^{1/2}(\hat{a}_{n,i} - a_0),\ (k^{1/2}/\log(n/k))(\hat{b}_{n,i} - b_0)\Big) \in B\Big) \\ -P\Big(\Big(a_0\xi_1, (a_0/\log(n/k))\xi_2 - a_0\xi_1\Big) \in B\Big)\Big| \\ = O(k/n + (k/n)^\delta k^{1/2} + k^{-1/2}), \end{aligned}$$

where ξ_1, ξ_2 are independent standard normal rv.

Note that in cases $i = 1$ and 2, estimators of the initial scale parameter a in the model $F \in Q_i(\delta)$ are given by $\exp(\hat{b}_{n,1}) \sim \exp(b_0) = a$ and $\exp(-\hat{b}_{n,2}) \sim \exp(-b_0) = a$, respectively. Their asymptotic behavior can easily be deduced from the preceding theorem and Taylor expansion of the exponential function.

The Pareto Model with Known Scale Factor

Suppose that the df F underlying the iid sample $Z_1, \dots, Z_n$ is in $Q_1(\delta)$. Then F has a density f on (x_0, ∞) such that

$$f(x) = \frac{1}{a} w_{1,1/\alpha}\Big(\frac{x}{a}\Big)\Big(1 + O((1 - W_{1,1/\alpha}(x))^\delta\Big), \quad x > x_0, \tag{2.10}$$

for some $\alpha, \delta, a > 0$. Note that we have replaced α by $1/\alpha$. The preceding result states that for the Hill estimator

$$\hat{a}_{n,1} = k^{-1} \sum_{j \le k} \log(Z_{n-j+1:n} \wedge 1) - \log(Z_{n-k:n} \wedge 1)$$

of $a_0 = \alpha$ we have

$$\begin{aligned}\sup_{B \in \mathbb{B}} \Big| P\Big(\frac{k^{1/2}}{\alpha}(\hat{a}_{n,1} - \alpha) \in B\Big) - N(0,1)(B)\Big| \\ = O(k/n + (k/n)^\delta k^{1/2} + k^{-1/2}),\end{aligned}$$

where $N(0,1)$ denotes the standard normal distribution on $\mathbb{R}$. If the scale parameter a in (2.10) is however *known*, then the Hill estimator is outperformed by

$$\hat{\alpha}_n := \frac{\log(Z_{n-k:n} \wedge 1) - \log(a)}{\log(n/k)},$$

as by Lemma 2.4.1 and the transformation (2.7)

$$\begin{aligned}\sup_{B \in \mathbb{B}} \Big| P\Big(\frac{k^{1/2}\log(n/k)}{\alpha}(\hat{\alpha}_n - \alpha) \in B\Big) - N(0,1)(B)\Big| \\ = O(k/n + (k/n)^\delta k^{1/2} + k^{-1/2}),\end{aligned}$$

showing that $\hat{\alpha}_n$ is more concentrated around α than $\hat{a}_{n,1}$.

This result, which looks strange at first sight, is closely related to the fact that $Z_{n-k:n}$ is the central sequence generating *local asymptotic normality* (LAN) of the loglikelihood processes of $(Z_{n-j+1:n})_{j \le k+1}$, indexed by α. In this sense, $Z_{n-k:n}$ carries asymptotically the complete information about the underlying shape parameter α that is contained in $(Z_{n-j+1:n})_{j \le k}$ (see Theorems 1.3, 2.2 and 2.3 in Falk [133]).

The Extreme Quantile Estimation

Since $W_3^{-1}(1-q) = -\log(q)$, we obtain from the preceding theorem the following result on the estimation of extreme quantiles outside the range of our actual data. We adopt the notation of the preceding result.

Theorem 2.4.5. *Suppose that F is in $Q_i(\delta)$, $i = 1, 2$ or 3. Then we have, for $i = 1, 2, 3$,*

$$\sup_{B\in\mathbb{B}} \Big| P\Big(F_i^{-1}(1-q) - (\hat{a}_{n,i}W_3^{-1}(1-q) + \hat{b}_{n,i}) \in B\Big) \\ -P\Big(a_0\xi_1 \log(qn/k)/k^{1/2} + a_0\xi_2/k^{1/2} + O(q^\delta) \in B\Big)\Big| \\ = O(k/n + (k/n)^\delta k^{1/2} + k^{-1/2})$$

uniformly for $q \to 0$, where ξ_1, ξ_2 are independent and standard normal rv.

The preceding result entails in particular that $\hat{a}_{n,i}W_3^{-1}(1-q) + \hat{b}_{n,i} = \hat{b}_{n,i} -\hat{a}_{n,i}\log(q)$ is a consistent estimator of $F_i^{-1}(1-q)$ for any sequence $q = q_n \longrightarrow_{n\to\infty} 0$ such that $\log(qn)/k^{1/2} \to 0$ with $k = k(n) \to \infty$ satisfying $(k/n)^\delta k^{1/2} \to 0$.

The bound $O(k/n+(k/n)^\delta k^{1/2}+k^{-1/2})$ for the normal approximation in Theorem 2.4.5 suggests that an optimal choice of $k = k(n) \to \infty$ is of order $n^{2\delta/(2\delta+1)}$, in which case the error bound $(k/n)^\delta k^{1/2}$ does not vanish asymptotically.

Note that $F_1^{-1}(1-q) = \log(F^{-1}(1-q))$ and $F_2^{-1}(1-q) = -\log(\omega(F) - F^{-1}(1-q))$. We can apply therefore the transformation $T_1(x) = \exp(x)$ and $T_2(x) = \omega(F) - \exp(-x)$ in case $i = 1, 2$ to the estimators of $F_i^{-1}(1-q)$ in Theorem 2.4.5, and we can deduce from this theorem the asymptotic behavior of the resulting estimators of the initial extreme quantile $F^{-1}(1-q)$.

Theorem 2.4.5 implies the following result.

Corollary 2.4.6. *Suppose that F is in $Q_i(\delta)$, $i = 1, 2$ or 3. Then we have*

(i)
$$q_n^{\gamma(i)}(F^{-1}(1-q_n) - T_i(\hat{b}_{n,i} - \hat{a}_{n,i}\log(q_n))) \longrightarrow_{n\to\infty} 0$$

in probability, with $\gamma(i) = 1/\alpha, -1/\alpha, 0$ if $i = 1, 2, 3$, and $T_1(x) = \exp(x)$, $T_2(x) = \omega(F) - \exp(-x)$, $T_3(x) = x$, $x \in \mathbb{R}$, for any sequence $q_n \longrightarrow_{n\to\infty} 0$ such that $\log(q_n n)/k^{1/2} \longrightarrow_{n\to\infty} 0$, where $k = k(n)$ satisfies $(k/n)^\delta k^{1/2} \longrightarrow_{n\to\infty} 0$.

(ii)
$$\sup_{t\in\mathbb{R}} \Big| P\Big(\frac{q_n^{\gamma(i)}k^{1/2}}{a(i)(\log^2(q_n n/k)+1)^{1/2}}(F^{-1}(1-q_n) - T_i(\hat{b}_{n,i} - \hat{a}_{n,i}\log(q_n))) \le t\Big) \\ -\Phi(t)\Big| \longrightarrow_{n\to\infty} 0,$$

for any sequence $q_n \longrightarrow_{n\to\infty} 0$ such that $k^{1/2}q_n^\delta$ is bounded and $\log(q_n n)/k^{1/2} \longrightarrow_{n\to\infty} 0$, where $k \to \infty$ satisfies $(k/n)^\delta k^{1/2} \longrightarrow_{n\to\infty} 0$, $a(i) = a/\alpha, a/\alpha, a$ if $i = 1, 2, 3$ and Φ denotes the standard normal df.

Proof. Theorem 2.4.5 implies

(i) $F_i^{-1}(1-q_n) - (\hat{b}_{n,i} - \hat{a}_{n,i} W_3^{-1}(1-q_n)) \longrightarrow_{n\to\infty} 0$ in probability for any sequence $q_n \longrightarrow_{n\to\infty} 0$ such that $\log(q_n n)/k^{1/2} \longrightarrow_{n\to\infty} 0$, where $k = k(n) \longrightarrow_{n\to\infty} \infty$ satisfies $(k/n)^\delta k^{1/2} \longrightarrow_{n\to\infty} 0$.

(ii)
$$\sup_{t\in\mathbb{R}} \left| P\Big(\frac{k^{1/2}}{a(\log^2(q_n n/k)+1)^{1/2}} \Big(F_i^{-1}(1-q_n) - (\hat{b}_{n,i} - \hat{a}_{n,i} W_3^{-1}(1-q_n)) \Big) \le t \Big) - \Phi(t) \right| \longrightarrow_{n\to\infty} 0$$

for any sequence $q_n \longrightarrow_{n\to\infty} 0$ such that $k^{1/2} q_n^\delta$ is bounded, where $k \to \infty$ satisfies $(k/n)^\delta\ k^{1/2} \longrightarrow_{n\to\infty} 0$.

The assertion of Corollary 2.4.6 now follows from the equation $F^{-1}(1-q_n) = T_i(F_i^{-1}(1-q_n)),\ i = 1, 2, 3$, Taylor expansion of T_i and the equation $F^{-1}(1-q_n) = a q_n^{-1/\alpha}(1+o(1))$ if $i = 1$; $\omega(F) - F^{-1}(1-q_n) = a q_n^{1/\alpha}(1+o(1))$ if $i = 2$ (see Proposition 2.2.1). □

Confidence Intervals

Theorem 2.4.5 can immediately be utilized to define confidence intervals for the extreme quantile $F^{-1}(1-q_n)$. Put for $q_n \in (0,1)$,

$$\begin{aligned} \widehat{F}_n^{-1}(1-q_n) &:= \hat{a}_{n,i} W_3^{-1}(1-q_n) + \hat{b}_{n,i} \\ &= \hat{b}_{n,i} - \hat{a}_{n,i}\log(q_n), \end{aligned}$$

and define the interval

$$\begin{aligned} I_n := \Big[&\hat{F}_n^{-1}(1-q_n) - \hat{a}_{n,i}(\log^2(q_n n/k)+1)^{1/2} k^{-1/2} \Phi^{-1}(1-\beta_1), \\ &\hat{F}_n^{-1}(1-q_n) + \hat{a}_{n,i}(\log^2(q_n n/k)+1)^{1/2} k^{-1/2} \Phi^{-1}(1-\beta_2) \Big], \end{aligned}$$

where $0 < \beta_1, \beta_2 < 1/2$. For $F \in Q_i(\delta)$ we obtain that I_n is a confidence interval for $F_i^{-1}(1-q_n)$ of asymptotic level $1-(\beta_1 + \beta_2)$ that is, $\lim_{n\to\infty} P(F_i^{-1}(1-q_n) \in I_n) = 1-(\beta_1+\beta_2)$. Consequently, we obtain from the equation $T_i(F_i^{-1}(1-q_n)) = F^{-1}(1-q_n)$,

$$\lim_{n\to\infty} P(F^{-1}(1-q_n) \in T_i(I_n)) = 1-(\beta_1+\beta_2)$$

with $T_1(x) = \exp(x)$, $T_2(x) = \omega(F) - \exp(-x)$ and $T_3(x) = x,\ x \in \mathbb{R}$, for any sequence $q_n \longrightarrow_{n\to\infty} 0$ such that $k^{1/2} q_n^\delta$ is bounded, where $k \to \infty$ satisfies $(k/n)^\delta$ $k^{1/2} \longrightarrow_{n\to\infty} 0$. Note that T_i are strictly monotone and continuous functions. The confidence interval $T_i(I_n)$ can therefore be deduced from I_n immediately by just transforming its endpoints. Note further that the length of I_n is in probability proportional to $(\log^2(q_n n/k)+1)^{1/2} k^{-1/2}$, which is a convex function in q_n with the minimum value $k^{-1/2}$ at $q_n = k/n$.

2.5 Initial Estimation of the Class Index

If it is assumed that F is in $Q_i(\delta)$ but the index i is unknown, an initial estimation of the index $i \in \{1, 2, 3\}$ is necessary. Such a decision can be based on graphical methods as described in Castillo et al. [61] or on numerical estimates like the following Pickands [371] estimator (for a discussion we refer to Dekkers and de Haan [108]).

The Pickands Estimator

Choose $m \in \{1, \dots, n/4\}$ and define

$$\hat\beta_n(m) := (\log(2))^{-1} \ \log\Big(\frac{Z_{n-m+1:n} - Z_{n-2m+1:n}}{Z_{n-2m+1:n} - Z_{n-4m+1:n}}\Big).$$

This estimator is an asymptotically consistent estimator of $\beta := 1/\alpha, -1/\alpha, 0$ in case $F \in Q_i(\delta)$ with pertaining shape parameter α. A stable positive or negative value of $\hat\beta_n(m)$ indicates therefore that F is in $Q_1(\delta)$ or $Q_2(\delta)$, while $\hat\beta_n(m)$ near zero indicates that $F \in Q_3(\delta)$. By $N(\mu, \sigma^2)$ we denote the normal distribution on $\mathbb{R}$ with mean μ and variance σ^2.

Proposition 2.5.1. *Suppose that F is in $Q_i(\delta)$, $i = 1, 2, 3$. Then we have*

$$\begin{aligned}\sup_{t\in\mathbb{R}} \Big| P\Big(m^{1/2}(\hat\beta_n(m) - \beta) \le t\Big) - N(0, \sigma_\beta^2)((-\infty, t])\Big| \\ = O((m/n)^\delta\, m^{1/2} + m/n + m^{-1/2}),\end{aligned}$$

where

$$\sigma_\beta^2 := \frac{1 + 2^{-2\beta-1}}{2\log^2(2)}\Big(\frac{\beta}{1 - 2^{-\beta}}\Big)^2, \qquad \beta \in \mathbb{R}.$$

Interpret $\sigma_0^2 = \lim_{\beta\to 0}\sigma_\beta^2 = 3/(4\log(2)^4)$.

The estimator $\hat\beta_n(m)$ of β can easily be motivated as follows. One expects by Proposition 2.2.1,

$$\begin{aligned}\frac{Z_{n-m+1:n} - Z_{n-2m+1:n}}{Z_{n-2m+1:n} - Z_{n-4m+1:n}} &\sim \frac{F^{-1}(1 - \frac{m}{n+1}) - F^{-1}(1 - \frac{2m}{n+1})}{F^{-1}(1 - \frac{2m}{n+1}) - F^{-1}(1 - \frac{4m}{n+1})} \\ &\sim \frac{W_i^{-1}(1 - \frac{m}{n+1}) - W_i^{-1}(1 - \frac{2m}{n+1})}{W_i^{-1}(1 - \frac{2m}{n+1}) - W_i^{-1}(1 - \frac{4m}{n+1})},\end{aligned}$$

with $W_i \in \{W_1, W_2, W_3\}$ being the GPD pertaining to F. Since location and scale shifts are canceled by the definition of the estimator $\hat\beta_n(m)$, we can assume without loss of generality that W_i has standard form. Now

$$W_i^{-1}(1-q) = \begin{cases} q^{-1/\alpha} & \\ -q^{1/\alpha} & \text{in case} \quad i = 2, \\ -\log(q) & \end{cases} \quad \begin{matrix} i = 1, \\ \\ i = 3, \end{matrix}$$

$q \in (0,1)$ and, thus,

$$\frac{W_i^{-1}(1-\frac{m}{n+1}) - W_i^{-1}(1-\frac{2m}{n+1})}{W_i^{-1}(1-\frac{2m}{n+1}) - W_i^{-1}(1-\frac{4m}{n+1})} = \begin{cases} 2^{1/\alpha}, & i=1, \\ 2^{-1/\alpha}, & i=2, \\ 1, & i=3, \end{cases}$$

which implies the approximation

$$\begin{aligned} \hat{\beta}_n(m) &\sim (\log(2))^{-1} \log\left(\frac{W_i^{-1}(1-\frac{m}{n+1}) - W_i^{-1}(1-\frac{2m}{n+1})}{W_i^{-1}(1-\frac{2m}{n+1}) - W_i^{-1}(1-\frac{4m}{n+1})}\right) \\ &= \begin{cases} 1/\alpha, & i=1, \\ -1/\alpha, & i=2, \\ 0, & i=3, \end{cases} \quad = \beta. \end{aligned}$$

Weak consistency of $\hat{\beta}_n(m)$ actually holds under the sole condition that F is in the domain of attraction of an EVD and $m = m(n)$ satisfies $m \to \infty,\ m/n \to 0$ as $n \to \infty$ (see Theorem 2.1 of Dekkers and de Haan [108]). Asymptotic normality of $\hat{\beta}_n(m)$ however, requires additional conditions on F (see also Theorem 2.3 of Dekkers and de Haan [108]).

Convex Combinations of Pickands Estimators

The limiting variance of Pickands estimator $\hat{\beta}_n(m)$ can considerably be reduced by a simple convex combination. Choose $p \in [0,1]$ and define for $m \in \{1, \dots, n/4\}$,

$$\begin{aligned} \hat{\beta}_n(m,p) &:= p \cdot \hat{\beta}_n([m/2]) + (1-p) \cdot \hat{\beta}_n(m) \\ &= (\log(2))^{-1} \log\left\{\left(\frac{Z_{n-[m/2]+1:n} - Z_{n-2[m/2]+1:n}}{Z_{n-2[m/2]+1:n} - Z_{n-4[m/2]+1:n}}\right)^p \right. \\ &\qquad \left. \times \left(\frac{Z_{n-m+1:n} - Z_{n-2m+1:n}}{Z_{n-2m+1:n} - Z_{n-4m+1:n}}\right)^{1-p}\right\}, \end{aligned}$$

where $[x]$ denotes the integral part of $x \in \mathbb{R}$. If m is even, $[m/2]$ equals $m/2$, and the preceding notation simplifies.

We consider the particular convex combination $\hat{\beta}_n(m,p)$ to be a natural extension of Pickands estimator $\hat{\beta}_n(m)$: As $\hat{\beta}_n(m)$ is built upon powers of 2 that is, of $2^0 m, 2^1 m, 2^2 m$, it is only natural (and makes the computations a bit easier) to involve the next smaller integer powers $2^{-1}m, 2^0 m, 2^1 m$ of 2 and to combine $\hat{\beta}_n(m)$ with $\hat{\beta}_n([m/2])$. As a next step one could consider linear combinations $\sum_{i \le k} p_i \hat{\beta}_n([m/2^{i-1}])$, $\sum_{i \le k} p_i = 1$, of length k. But as one uses the $4m$ largest observations in a sample of size n in the definition of $\hat{\beta}_n(m)$, with $4m$ having to be relatively small to n anyway, it is clear that $m/2$ is already a rather small number for making asymptotics ($m \to \infty$). For moderate sample sizes n, the length m will therefore be limited to 2 in a natural way. Higher linear combinations nevertheless perform still better in an asymptotic model (cf. Drees [114]).

In the following result we establish asymptotic normality of $\hat{\beta}_n(m,p)$. With $p = 0$ it complements results on the asymptotic normality of $\hat{\beta}_n(m) = \hat{\beta}_n(m,0)$ (cf. also Dekkers and de Haan ([108], Theorems 2.3, 2.5)). Its proof is outlined at the end of this section. A careful inspection of this proof also implies Proposition 2.5.1.

Lemma 2.5.2. *Suppose that F is in $Q_i(\delta)$, $i = 1,2,3$ for some $\delta > 0$. Then we have, for $m \in \{1,\dots,n/4\}$ and $p \in [0,1]$,*

$$\sup_{B\in\mathbb{B}} \Big| P(m^{1/2}(\hat{\beta}_n(m,p) - \beta) \in B) - P(\sigma_\beta \nu_\beta(p) X + O_p(m^{-1/2}) \in B) \Big|$$
$$= O\Big((m/n)^\delta m^{1/2} + m/n + m^{-1/2}\Big),$$

where X is a standard normal rv and

$$\nu_\beta(p) := 1 + p^2\Big(3 + \frac{4\cdot 2^{-\beta}}{2^{-2\beta}+2}\Big) - p\Big(2 + \frac{4\cdot 2^{-\beta}}{2^{-2\beta}+2}\Big).$$

The Asymptotic Relative Efficiency

The following result is an immediate consequence of Lemma 2.5.2.

Corollary 2.5.3. *Under the conditions of the preceding lemma we have, for $m = m(n) \longrightarrow_{n\to\infty} \infty$ such that $(m/n)^\delta m^{1/2} \to 0$ as $n \to \infty$,*

$$m^{1/2}(\hat{\beta}_n(m,p) - \beta) \ \to_D \ N(0, \sigma_\beta^2 \nu_\beta^2(p)).$$

By $\to_D$ we denote again convergence in distribution. Recall that σ_β^2 is the variance of the limiting central normal distribution of the standardized Pickands estimator $\sqrt{m}\ (\hat{\beta}_n(m) - \beta)$. The factor $\nu_\beta^2(p)$ is now the *asymptotic relative efficiency (ARE)* of $\hat{\beta}_n(m)$ with respect to $\hat{\beta}_n(m,p)$, which we define by the ratio of the variances of the limiting central normal distributions of $m^{1/2}(\hat{\beta}_n(m,p) - \beta)$ and $m^{1/2}(\hat{\beta}_n(m) - \beta)$:

$$\text{ARE}\ (\hat{\beta}_n(m)|\hat{\beta}_n(m,p)) := \nu_\beta^2(p).$$

The Optimal Choice of p

The optimal choice of p minimizing $\nu_\beta^2(p)$ is

$$p_{\text{opt}}(\beta) := \frac{(2^{-2\beta}+2) + 2\cdot 2^{-\beta}}{3(2^{-2\beta}+2) + 4\cdot 2^{-\beta}},$$

in which case $\nu_\beta^2(p)$ becomes

$$\nu_\beta^2(p_{\mathrm{opt}}(\beta)) = 1 - p_{\mathrm{opt}}(\beta) \cdot \Big(1 + 2\frac{2^{-\beta}}{2^{-2\beta}+2}\Big).$$

As $p_{\mathrm{opt}}(\beta)$ is strictly between 0 and 1, we have $\nu_\beta^2(p_{\mathrm{opt}}(\beta)) < 1$ and the convex combination $\hat\beta_n(m, p_{\mathrm{opt}}(\beta))$ is clearly superior to the Pickands estimator $\hat\beta_n(m)$.

The following figure displays the ARE function $\beta \mapsto \nu_\beta^2(p_{\mathrm{opt}}(\beta))$. As $\nu_\beta^2(p_{\mathrm{opt}}(\beta)) =: g(2^{-\beta})$ depends upon β through the transformation $2^{-\beta}$, we have plotted the function $g(x)$, $x \in \mathbb{R}$, with $x = 2^{-\beta}$. Notice that both for $x \to 0$ (that is, $\beta \to \infty$) and $x \to \infty$ (that is, $\beta \to -\infty$) the pertaining ARE function converges to 2/3 being the least upper bound, while .34 is an approximate infimum.

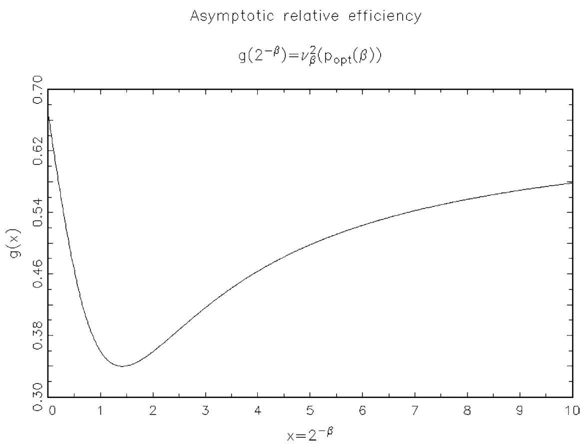

Figure 2.5.1. $g(x) = 1 - \frac{(x^2+2)+2x}{3(x^2+2)+4x}\Big(1 + \frac{2x}{x^2+2}\Big)$.

Data-Driven Optimal Estimators

The optimal p depends however on the unknown β and it is therefore reasonable to utilize the data-driven estimator

$$p_{\mathrm{opt}}(\tilde\beta_n) := \frac{(2^{-2\tilde\beta_n}+2) + 2\cdot 2^{-\tilde\beta_n}}{3(2^{-2\tilde\beta_n}+2) + 4\cdot 2^{-\tilde\beta_n}},$$

where $\tilde\beta_n$ is an initial estimator of β. If $\tilde\beta_n$ is asymptotically consistent, then, using Lemma 2.5.2, it is easy to see that the corresponding data-driven convex combination $\hat\beta_n(m, p_{\mathrm{opt}}(\tilde\beta_n))$ is asymptotically equivalent to the optimal convex combination $\hat\beta_n(m, p_{\mathrm{opt}}(\beta))$ with underlying β that is,

$$m^{1/2}\Big(\hat\beta_n(m, p_{\mathrm{opt}}(\tilde\beta_n)) - \hat\beta_n(m, p_{\mathrm{opt}}(\beta))\Big) = o_P(1),$$

so that their ARE is one.

A reasonable initial estimator of β is suggested by the fact that the particular parameter $\beta = 0$ is crucial as it is some kind of change point: If $\beta < 0$, then the right endpoint $\omega(F)$ of F is finite, while in case $\beta > 0$ the right endpoint of F is infinity. The question $\omega(F) = \infty$ or $\omega(F) < \infty$ is in case $\beta = 0$ numerically hard to decide, as an estimated value of β near 0 can always be misleading. In this case, graphical methods such as the one described in Castillo et al. [61] can be helpful. To safeguard oneself against this kind of a least favorable value β, it is therefore reasonable to utilize as an initial estimator $\tilde{\beta}_n$ that convex combination $\hat{\beta}_n(m, p)$, where p is chosen optimal for $\beta = 0$ that is, $p_{\text{opt}}(0) = 5/13$. We propose as an initial estimator therefore $\tilde{\beta}_n = \hat{\beta}_n(m, 5/13)$.

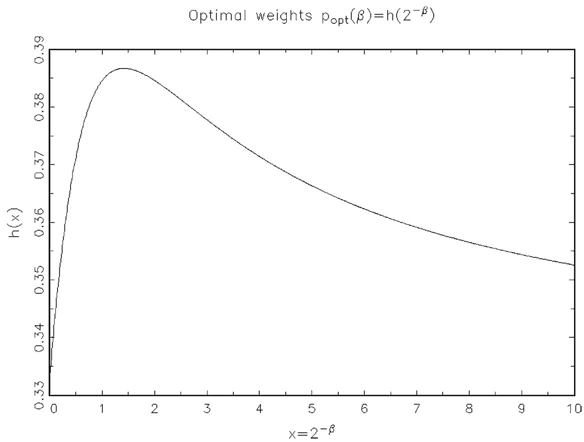

FIGURE 2.5.2. $h(x) = ((x^2 + 2) + 2x)/(3(x^2 + 2) + 4x)$.

Figure 2.5.2 displays the function of optimal weights $p_{\text{opt}}(\beta)$, again after the transformation $x = 2^{-\beta}$ that is, $p_{\text{opt}}(\beta) =: h(2^{-\beta})$. These weights do not widely spread out, they range between .33 and .39, roughly.

Note that the ARE of the Pickands estimator $\hat{\beta}_n(m)$ with respect to the optimal convex combination $\hat{\beta}_n(m, 5/13)$ in case $\beta = 0$ is 14/39. We summarize the preceding considerations in the following result.

Theorem 2.5.4. *Suppose that F is in $Q_i(\delta)$, $i = 1, 2$ or 3 for some $\delta > 0$. Then we have, for $m \to \infty$ such that $(m/n)^\delta m^{1/2} \to 0$ as $n \to \infty$,*

$$m^{1/2}\Big(\hat{\beta}_n(m, p_{opt}(\tilde{\beta}_n)) - \beta\Big) \to_D N\Big(0, \sigma_\beta^2\Big(1 - p_{opt}(\beta)\Big(1 + 2\frac{2^{-\beta}}{2^{-2\beta} + 2}\Big)\Big)\Big)$$

for any initial estimator sequence $\tilde{\beta}_n$ of β which is asymptotically consistent.

Dropping the δ-Neighborhood

The crucial step in the proof of Lemma 2.5.2 is the approximation

$$\begin{aligned}\Delta_{n,m} &:= \sup_{B\in\mathbb{B}^m} \Big| P(((Z_{n-j+1:n}-b_n)/a_n)_{j\le m}\in B) \\ &\qquad - \begin{cases} P\Big(\Big(\beta(\sum_{r\le j}\xi_r)^{-\beta}\Big)_{j\le m}\in B\Big)\Big| & \text{if} \quad \beta\ne 0,\\ P\Big(\Big(-\log(\sum_{r\le j}\xi_r)\Big)_{j\le m}\in B\Big)\Big| & \text{if} \quad \beta = 0,\end{cases}\\ &= O((m/n)^\delta m^{1/2}+m/n)\end{aligned}$$

of Theorem 2.2.4, where $\xi_1,\xi_2,\dots$ are independent and standard exponential rv and $a_n>0,\ b_n\in\mathbb{R}$ are suitable constants (see also Remark 2.2.3).

Lemma 2.5.2, Corollary 2.5.3 and Theorem 2.5.4 remain however true with $(m/n)^\delta m^{1/2}+m/n$ replaced by $\Delta_{n,m}$, if we drop the condition that F is in $Q_i(\delta)$ and require instead $\Delta_{n,m}\to 0$ for some sequence $m=m(n)\le n/4,\ m\to\infty$ as $n\to\infty$.

Then we can consider for example the case, where F is the standard normal df, which is not in any $Q_i(\delta)$; but in this case we have $\Delta_{n,m}=O(m^{1/2}\log^2(m+1)/\log(n))$ (cf. Example 2.33 in Falk [126]), allowing however only rather small sizes of $m=m(n)$ to ensure asymptotic normality.

Simulation Results

In this section we briefly report the results of extensive simulations which we have done for the comparison between $\hat\beta_n(m,\hat p_{\text{opt}})=\hat\beta_n(m,p_{\text{opt}}(\tilde\beta_n))$, based on the initial estimate $\tilde\beta_n=\hat\beta_n(m,5/13)$ and the Pickands estimator $\hat\beta_n(m)$.

These simulations with various choices of n,m and underlying df F support the theoretical advantage of using $\hat\beta_n(m,\hat p_{\text{opt}})$ in those cases, where F is in a δ-neighborhood of a GPD. Figures 2.5.3-2.5.6 exemplify the gain of relative performance which is typically obtained by using $\hat\beta_n(m,\hat p_{\text{opt}})$. In these cases we generated $n=50/100/200/400$ replicates $Z_1,\dots,Z_n$ of a (pseudo-) rv Z with different distribution F in each case. The estimators $\hat\beta_n(m,\hat p_{\text{opt}})$ and $\hat\beta_n(m)$ of the pertaining values of β with $m=6/8/12/40$ were computed, and we stored by $B:=|\hat\beta_n(m)-\beta|,\ C:=|\hat\beta_n(m,\hat p_{\text{opt}})-\beta|$ their corresponding absolute deviations. We generated $k=1000$ independent replicates $B_1,\dots,B_k$ and $C_1,\dots,C_k$ of B and C, with their sample quantile functions now visualizing the concentration of $\hat\beta_n(m)$ and $\hat\beta_n(m,\hat p_{\text{opt}})$ around β.

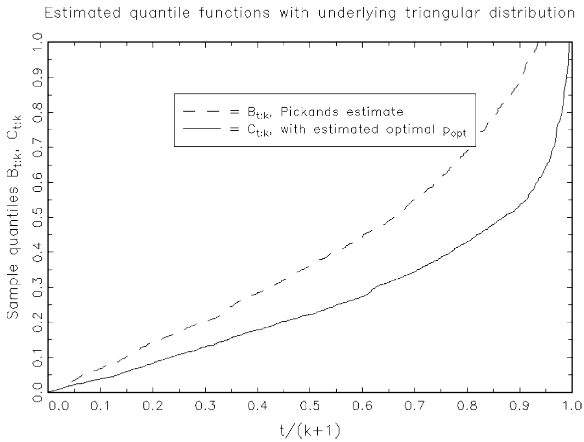

FIGURE 2.5.3. $\beta = -.5,\ n = 50,\ m = 6.$

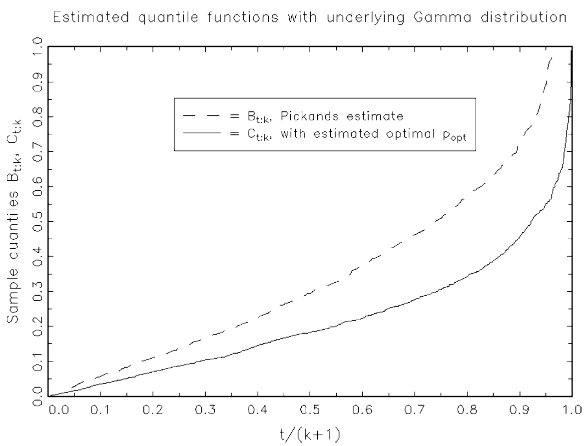

FIGURE 2.5.4. $\beta = 0,\ n = 100,\ m = 8.$

Figures 2.5.3-2.5.6 display the pertaining sample quantile functions $(t/(k+1),\ B_{t:k})$ and $(t/(k+1),\ C_{t:k})$, $t = 1, \dots, k = 1000$ for Z. By $B_{1:k} \le \cdots \le B_{k:k}$ and $C_{1:k} \le \cdots \le C_{k:k}$ we denote the ordered values of $B_1, \dots, B_k$ and $C_1, \dots, C_k$. In Figure 2.5.3, Z equals the sum of two independent and uniformly on (0,1) distributed rv ($\beta = -.5$); in Figure 2.5.4 the df F is a Gamma distribution ($Z = X_1 + X_2 + X_3$, X_1, X_2, X_3 independent and standard exponential, $\beta = 0$) and in Figure 2.5.5, F is a Cauchy distribution ($\beta = 1$). Elementary computations show that these distributions satisfy (VM) with rapidly decreasing remainder. The triangular distribution is in particular a GPD. In Figure 2.5.6, F is the normal distribution.

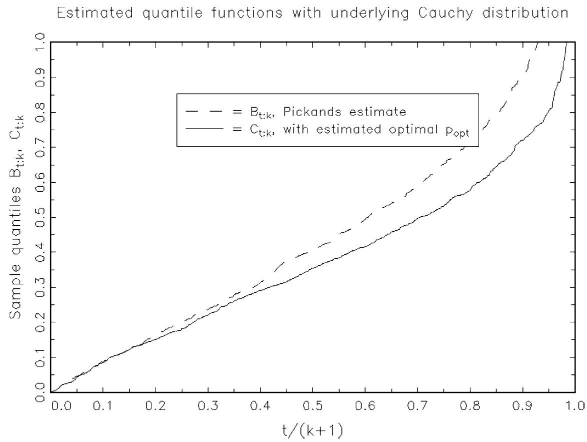

FIGURE 2.5.5. $\beta = 1,\ n = 200,\ m = 12.$

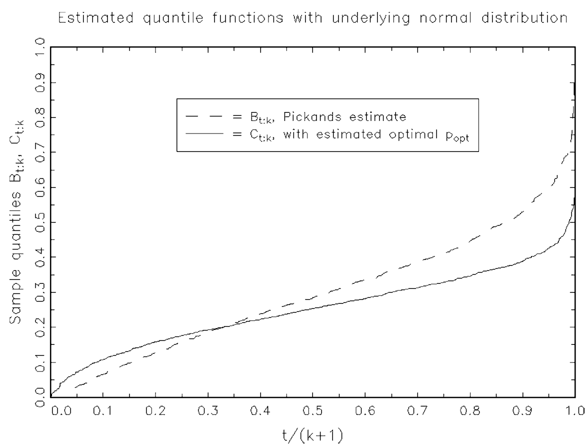

FIGURE 2.5.6. $\beta = 0,\ n = 400,\ m = 40.$

Except the normal distribution underlying Figure 2.5.6, these simulations are clearly in favor of the convex combination $\hat{\beta}_n(m, \hat{p}_{\text{opt}})$ as an estimator of β, even for moderate sample sizes n. In particular Figure 2.5.3 shows in this case with $\beta = -.5$, that $\hat{\beta}_n(m)$ would actually give a negative value between -1 and 0 with approximate probability .67, whereas the corresponding probability is approximately .87 for $\hat{\beta}_n(m, \hat{p}_{\text{opt}})$. Recall that a negative value of β implies that the underlying df has finite right endpoint. In Figure 2.5.4 with $\beta = 0$ the sample medians $B_{k/2:k}$ and $C_{k/2:k}$ are in particular interesting: While with approximate probability 1/2 the Pickands estimate $\hat{\beta}_n(m)$ has an absolute value less than .3, the combination $\hat{\beta}_n(m, \hat{p}_{\text{opt}})$ is less than .18 apart from $\beta = 0$ with approximate probability 1/2. Figure 2.5.6 is not in favor of $\hat{\beta}_n(m, \hat{p}_{\text{opt}})$. But this can be explained by the fact that the normal distribution does not belong to a δ-neighborhood of a GPD and the choice $m = 40$ is too large. This observation underlines the significance of δ-neighborhoods of GPD.

Our simulations showed that the relative performance of $\hat{\beta}_n(m, \hat{p}_{\text{opt}})$ is quite sensitive to the choice of m which corresponds to under- and oversmoothing in bandwidth selection problems in non-parametric curve estimation (see Marron [322] for a survey). Our simulations suggest as a rule of thumb the choice $m = (2/25)n$ for not too large sample size n that is, $n \leq 200$, roughly.

Notes on Competing Estimators

If one knows the sign of β, then one can use the estimators $\hat{a}_{n,i}$ of the shape parameter $1/\alpha = |\beta|$, which we have discussed in Section 2.4. Based on the $4m$ largest order statistics, $m^{1/2}(\hat{a}_{n,i} - |\beta|)$ is asymptotically normal distributed under appropriate conditions with mean 0 and variance $\beta^2/4$ in case $i = 1, 2$ and $\beta \neq 0$ (Theorem 2.4.4), and therefore outperforms $\hat{\beta}_n(m, \hat{p}_{\text{opt}})$ (see Theorem 2.5.4).

A competitor of $\hat{\beta}_n(m, p)$ is the moment estimator investigated by Dekkers et al. [107], which is outperformed by $\hat{\beta}_n(m, \hat{p}_{\text{opt}})$ if $\beta < 0$ is small enough. Alternatives such as the maximum-likelihood estimator or the method of probability-weighted moment (PWM) considered by Hosking and Wallis [224] require restrictions on β such as $\beta > -1$ (for the PWM method) and are therefore not universally applicable. A higher linear combination of Pickands estimators with estimated optimal scores was studied by Drees [114]. For thorough discussions of different approaches we refer to Section 9.6 of the monograph by Reiss [385], Sections 6.3 and 6.4 of Embrechts et al. [122], Section 5.1 of Reiss and Thomas [389], Section 1.7 of Kotz and Nadarajah [293] as well as to Beirlant et al. [32] and de Haan and Ferreira [190].

Outline of the Proof of Lemma 2.5.2: Put

$$A_n := \frac{Z_{n-[m/2]+1:n} - Z_{n-2[m/2]+1:n}}{Z_{n-2[m/2]+1:n} - Z_{n-4[m/2]+1:n}} - 2^\beta$$

and

$$B_n := \frac{Z_{n-m+1:n} - Z_{n-2m+1:n}}{Z_{n-2m+1:n} - Z_{n-4m+1:n}} - 2^\beta.$$

We will see below that A_n and B_n are both of order $O_P(m^{-1/2})$, so that we obtain, by the expansion $\log(1 + \varepsilon) = \varepsilon + O(\varepsilon^2)$ as $\varepsilon \to 0$,

$$\begin{aligned}\hat\beta_n(m,p)-\beta &= \frac{1}{\log(2)}\Big\{p\Big(\log(2^\beta+A_n)-\log(2^\beta)\Big)\\ &\qquad +(1-p)\Big(\log(2^\beta+B_n)-\log(2^\beta)\Big)\Big\}\\ &= \frac{1}{\log(2)}\Big\{p\log\Big(1+\frac{A_n}{2^\beta}\Big)+(1-p)\log\Big(1+\frac{B_n}{2^\beta}\Big)\Big\}\\ &= \frac{1}{2^\beta\log(2)}(pA_n+(1-p)B_n)+O_P(m^{-1}).\end{aligned}$$

From Theorem 2.2.4 we obtain within the error bound $O((m/n)^\delta m^{1/2}+m/n)$ in variational distance the representation

$$\begin{aligned}&pA_n+(1-p)B_n\\ &=\begin{cases} p\left(\dfrac{\Big(1+[m/2]^{-1}\sum_{j\le[m/2]}\eta_j\Big)^{-\beta}-\Big(2+[m/2]^{-1}\sum_{j\le 2[m/2]}\eta_j\Big)^{-\beta}}{\Big(2+[m/2]^{-1}\sum_{j\le 2[m/2]}\eta_j\Big)^{-\beta}-\Big(4+[m/2]^{-1}\sum_{j\le 4[m/2]}\eta_j\Big)^{-\beta}}-2^\beta\right)\\ +(1-p)\left(\dfrac{\Big(1+m^{-1}\sum_{j\le m}\eta_j\Big)^{-\beta}-\Big(2+m^{-1}\sum_{j\le 2m}\eta_j\Big)^{-\beta}}{\Big(2+m^{-1}\sum_{j\le 2m}\eta_j\Big)^{-\beta}-\Big(4+m^{-1}\sum_{j\le 4m}\eta_j\Big)^{-\beta}}-2^\beta\right), \\ \qquad \text{if}\quad \beta\ne 0\\ p\left(\dfrac{\log\left(\frac{2+[m/2]^{-1}\sum_{j\le 2[m/2]}\eta_j}{1+[m/2]^{-1}\sum_{j\le[m/2]}\eta_j}\right)}{\log\left(\frac{4+[m/2]^{-1}\sum_{j\le 4[m/2]}\eta_j}{2+[m/2]^{-1}\sum_{j\le 2[m/2]}\eta_j}\right)}-1\right)\\ +(1-p)\left(\dfrac{\log\left(\frac{2+m^{-1}\sum_{j\le 2m}\eta_j}{1+m^{-1}\sum_{j\le m}\eta_j}\right)}{\log\left(\frac{4+m^{-1}\sum_{j\le 4m}\eta_j}{2+m^{-1}\sum_{j\le 2m}\eta_j}\right)}-1\right), \qquad \text{if}\quad \beta=0,\end{cases}\end{aligned}$$

where $\eta_1+1,\eta_2+1,\dots$ are independent and standard exponential rv. Now elementary computations show that the distribution of $k^{-1/2}\sum_{j\le k}\eta_j$ approaches the standard normal distribution uniformly over all Borel sets at the rate $O(k^{-1/2})$ and thus, within the bound $O(m^{-1/2})$, we can replace the distribution of the right-hand side of the preceding equation by that of

$$p\left(\frac{\Big(1+\frac{X}{\sqrt{m/2}}\Big)^{-\beta}-\Big(2+\frac{X+Y}{\sqrt{m/2}}\Big)^{-\beta}}{\Big(2+\frac{X+Y}{\sqrt{m/2}}\Big)^{-\beta}-\Big(4+\frac{X+Y+\sqrt{2}Z}{\sqrt{m/2}}\Big)^{-\beta}}-2^\beta\right)$$

$$+(1-p)\left(\frac{\left(1+\frac{X+Y}{\sqrt{2m}}\right)^{-\beta}-\left(2+\frac{X+Y+\sqrt{2}Z}{\sqrt{2m}}\right)^{-\beta}}{\left(2+\frac{X+Y+\sqrt{2}Z}{\sqrt{2m}}\right)^{-\beta}-\left(4+\frac{X+Y+\sqrt{2}Z+2W}{\sqrt{2m}}\right)^{-\beta}}-2^{\beta}\right), \quad \text{if } \beta\neq 0,$$

and by

$$p\left(\frac{\log\left(\frac{2+\frac{X+Y}{\sqrt{m/2}}}{1+\frac{X}{\sqrt{m/2}}}\right)}{\log\left(\frac{4+\frac{X+Y+\sqrt{2}Z}{\sqrt{m/2}}}{2+\frac{X+Y}{\sqrt{m/2}}}\right)}-1\right)+(1-p)\left(\frac{\log\left(\frac{2+\frac{X+Y+\sqrt{2}Z}{\sqrt{2m}}}{1+\frac{X+Y}{\sqrt{2m}}}\right)}{\log\left(\frac{4+\frac{X+Y+\sqrt{2}Z+2W}{\sqrt{2m}}}{2+\frac{X+Y+\sqrt{2}Z}{\sqrt{2m}}}\right)}-1\right), \quad \text{if } \beta=0,$$

where X, Y, W, Z are independent standard normal rv. We have replaced $[m/2]$ in the preceding formula by $m/2$, which results in an error of order $O_P(1/m)$.

By Taylor's formula we have $(1+\varepsilon)^{-\beta} = 1-\beta\varepsilon+O(\varepsilon^2)$ for $\beta\neq 0$ and $\log(1+\varepsilon)=\varepsilon+O(\varepsilon^2)$ as $\varepsilon\to 0$. The assertion of Lemma 2.5.2 now follows from Taylor's formula and elementary computations. □

2.6 Power Normalization and p-Max Stable Laws

Let $Z_1,\dots Z_n$ be iid rv with common df F. In order to derive a more accurate approximation of the df of $Z_{n:n}$ by means of limiting df, Weinstein [456] and Pancheva [360] used a nonlinear normalization for $Z_{n:n}$. In particular, F is said to belong to the domain of attraction of a df H *under power normalization*, denoted by $F\in\mathcal{D}_p(H)$ if, for some $\alpha_n>0$, $\beta_n>0$,

$$F^n\big(\alpha_n|x|^{\beta_n}\operatorname{sign}(x)\big)\longrightarrow_\omega H(x), \qquad n\to\infty, \tag{2.11}$$

or in terms of rv,

$$(|Z_{n:n}|/\alpha_n)^{1/\beta_n}\operatorname{sign}(Z_{n:n})\longrightarrow_D Z, \qquad n\to\infty,$$

where Z is a rv with df H and $\operatorname{sign}(x)=-1$, 0 or 1 according as $x <, =$ or >0, respectively.

The Power-Max Stable Distributions

Pancheva [360] (see also Mohan and Ravi [339]) showed that a non-degenerate limit df H in (2.11) can up to a possible power transformation $H(\alpha|x|^\beta\operatorname{sign}(x))$, α, $\beta>0$, only be one of the following six different df H_i, $i=1,\dots,6$, where with $\gamma>0$,

$$H_1(x) = H_{1,\gamma}(x) = \begin{cases} 0 & \text{if } x \le 1, \\ \exp\big(-(\log(x))^{-\gamma}\big) & \text{if } x > 1, \end{cases}$$

$$H_2(x) = H_{2,\gamma}(x) = \begin{cases} 0 & \text{if } x \le 0, \\ \exp\big(-(-\log(x))^{\gamma}\big) & \text{if } 0 < x < 1, \\ 1 & \text{if } x \ge 1, \end{cases}$$

$$H_3(x) = \begin{cases} 0 & \text{if } x \le 0, \\ \exp(-1/x) & \text{if } x > 0, \end{cases}$$

$$H_4(x) = H_{4,\gamma}(x) = \begin{cases} 0 & \text{if } x \le -1, \\ \exp\big(-(-\log(-x))^{-\gamma}\big) & \text{if } -1 < x < 0, \\ 1 & \text{if } x \ge 0, \end{cases}$$

$$H_5(x) = H_{5,\gamma}(x) = \begin{cases} \exp\big(-(\log(-x))^{\gamma}\big) & \text{if } x < -1, \\ 1 & \text{if } x \ge -1, \end{cases}$$

$$H_6(x) = \begin{cases} \exp(x) & \text{if } x < 0, \\ 1 & \text{if } x \ge 0. \end{cases}$$

A df H is called *power-max stable* or *p-max stable* for short by Mohan and Ravi [339] if it satisfies the stability relation

$$H^n\big(\alpha_n |x|^{\beta_n} \operatorname{sign}(x)\big) = H(x), \qquad x \in \mathbb{R},\, n \in \mathbb{N},$$

for some $\alpha_n > 0$, $\beta_n > 0$. The df H_i, $i = 1, \dots, 6$, are p-max stable and, from a result of Pancheva [360], these are the only non-degenerate ones. Necessary and sufficient conditions for F to satisfy (2.11) were given by Mohan and Ravi [339], Mohan and Subramanya [340], and Subramanya [433]. In view of these considerations one might label the max stable EVD more precisely l-max stable, because they are max stable with respect to a linear normalization.

Max and Min Stable Distributions

We denote by $F \in \mathcal{D}_{\max}(G)$ the property that F is in the max-domain of attraction of an EVD G if

$$\frac{Z_{n:n} - b_n}{a_n} \longrightarrow_D G$$

for some norming constants $a_n > 0$, $b_n \in \mathbb{R}$. We denote by $F \in \mathcal{D}_{\min}(L)$ the property that F is in the min-domain of attraction of an EVD L if

$$\frac{Z_{1:n} - b_n}{a_n} \longrightarrow_D L.$$

From Theorem 2.1.1 we know that there are only three different types of possible non-degenerate limiting df G and L: G equals up to a possible linear transformation (i.e., location and scale shift) G_1, G_2 or G_3, where

$$G_1(x) = G_{1,\gamma}(x) = \begin{cases} 0 & \text{if } x \le 0, \\ \exp(-x^{-\gamma}) & \text{if } x > 0, \end{cases}$$

$$G_2(x) = G_{2,\gamma}(x) = \begin{cases} \exp(-(-x)^{\gamma}) & \text{if } x < 0, \\ 1 & \text{if } x \ge 0, \end{cases}$$

for some $\gamma > 0$, and $G_3(x) = \exp(-e^{-x})$, $x \in \mathbb{R}$. Note that $G_{1,1} = H_3$, $G_{2,1} = H_6$ and that G_3 is not a p-max stable law. The df L is up to a possible linear transformation equal to L_1, L_2 or L_3, where

$$L_1(x) = L_{1,\gamma}(x) = \begin{cases} 1 - \exp(-(-x)^{-\gamma}) & \text{if } x < 0, \\ 1 & \text{if } x \ge 0, \end{cases}$$

$$L_2(x) = L_{2,\gamma}(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 - \exp(-x^{\gamma}) & \text{if } x \ge 0, \end{cases}$$

for some $\gamma > 0$, and $L_3(x) = 1 - \exp(-e^{x})$, $x \in \mathbb{R}$.

The Characterization Theorem

The right endpoint $\omega(F) := \sup\{x : F(x) < 1\} \in (-\infty, \infty]$ of the df F plays a crucial role in the sequel.

The following result by Christoph and Falk [69] reveals that the upper *as well as* the lower tail behavior of F determine whether $F \in \mathcal{D}_p(H)$: The right endpoint $\omega(F) > 0$ yields the max stable distributions G, and $\omega(F) \le 0$ results in the min stable distributions L; this explains the number of six power types of p-max stable df.

Moreover, if $\omega(F) < \infty$ is *not* a point of continuity of F, i.e., if $P(Z_1 = \omega(F)) =: \rho > 0$, then $F \notin \mathcal{D}_p(H)$ for any non-degenerate df H. In this case we have

$$P(Z_{n:n} = \omega(F)) = 1 - P(Z_{n:n} < \omega(F)) = 1 - (1-\rho)^n \longrightarrow 1 \quad \text{as} \quad n \to \infty,$$

and

$$F^n\big(\alpha_n |x|^{\beta_n} \operatorname{sign}(x)\big) \begin{cases} \le (1-\rho)^n & \text{if} \quad \alpha_n |x|^{\beta_n} \operatorname{sign}(x) < \omega(F), \\ = 1 & \text{if} \quad \alpha_n |x|^{\beta_n} \operatorname{sign}(x) \ge \omega(F). \end{cases}$$

Hence, a limiting df H is necessarily degenerate in this case.

Theorem 2.6.1. *We have the following characterizations of the domains of attraction.*

(i) *Suppose that* $\omega(F) > 0$. *Put* $F^*(x) = 0$ *if* $x \le \min\{\log(\omega(F)/2), 0\}$ *and* $F^*(x) = F(\exp(x))$ *elsewhere. Then* F^* *is a df and*

$$F \in \mathcal{D}_p(H) \text{ for some non-degenerate } H \Longleftrightarrow F^* \in \mathcal{D}_{\max}(G). \tag{2.12}$$

In this case we have $H(x) = G((\log(x) - a)/b)$, $x > 0$, *for some* $b > 0$, $a \in \mathbb{R}$.

(ii) *Suppose that* $\omega(F) \le 0$ *and put* $F_*(x) := 1 - F(-\exp(x))$, $x \in \mathbb{R}$. *Then,*

$$F \in \mathcal{D}_p(H) \text{ for some non-degenerate } H \Longleftrightarrow F_* \in \mathcal{D}_{\min}(L). \tag{2.13}$$

In this case we have $H(x) = 1 - L((\log(-x) - a)/b)$, $x < 0$, *for some* $b > 0$, $a \in \mathbb{R}$.

As the domains of attraction of G and L as well as sequences of possible norming constants are precisely known (see Theorem 2.1.1), the preceding result together with its following proof characterizes the p-max stable distributions, their domains of attraction and the class of possible norming constants α_n, β_n in (2.11). In particular, we have $H_i(x) = G_i\big(\log(x)\big)$, $x > 0$, and $H_{i+3}(x) = 1 - L_i\big(\log(-x)\big)$, $x < 0$, $i = 1, 2, 3$. Subramanya [433], Remark 2.1, proved the special case $F \in \mathcal{D}_p(H_3)$ iff $F^* \in \mathcal{D}_{\max}(G_3)$.

Proof. (i) Suppose that $\omega(F) > 0$. In this case we have, for $x \le 0$ and any sequence $\alpha_n > 0$, $\beta_n > 0$,

$$\begin{aligned} F^n\big(\alpha_n |x|^{\beta_n} \operatorname{sign}(x)\big) &= P\big(Z_{n:n} \le \operatorname{sign}(x)\, \alpha_n |x|^{\beta_n}\big) \\ &\le P(Z_{n:n} \le 0) \longrightarrow 0 \quad \text{as} \quad n \to \infty. \end{aligned}$$

Let now $x > 0$ be a point of continuity of the limiting df H in (2.11) and put $c := \min\{\omega(F)/2, 1\}$. Then, as $n \to \infty$,

$$\begin{aligned} &P\big(Z_{n:n} \le \operatorname{sign}(x)\, \alpha_n |x|^{\beta_n}\big) \longrightarrow H(x) \\ &\Longleftrightarrow P\big(\log(Z_{n:n}) \le \beta_n \log(x) + \log(\alpha_n), Z_{n:n} \ge c\big) \longrightarrow H(x) \\ &\Longleftrightarrow P\big((Y_{n:n} - a_n)/b_n \le \log(x)\big) \longrightarrow H(x), \end{aligned}$$

where $b_n := \beta_n$, $a_n := \log(\alpha_n)$ and $Y_i := \log(Z_i) 1_{[c,\infty)}(Z_i) + \log(c) 1_{(-\infty,c)}(Z_i)$, $i = 1, \dots, n$ and $1_A(\cdot)$ is the indicator function of the set A. The rv $Y_1, \dots, Y_n$ are iid with common df satisfying

$$\begin{aligned} 1 - P(Y_i \le t) &= P(Y_i > t) \\ &= P(Z_i > \exp(t)) = 1 - F(\exp(t)), \qquad \exp(t) \ge c. \end{aligned}$$

Therefore, by the Gnedenko-de Haan Theorem 2.1.1 we obtain the equivalence (2.12) for some $G \in \{G_1, G_2, G_3\}$ with $H(x) = G_i(\log(x) - a)/b)$, $x > 0$, for some $b > 0$, $a \in \mathbb{R}$ and some $i \in \{1, 2, 3\}$. This completes the proof of part (i).

(ii) Suppose that $\omega(F) \leq 0$. Then we have for $x \geq 0$ and any $\alpha_n > 0$, $\beta_n > 0$

$$F^n\big(\alpha_n |x|^{\beta_n} \operatorname{sign}(x)\big) = P(Z_{n:n} \leq \alpha_n \, x^{\beta_n}) = 1.$$

Let now $x < 0$ be a point of continuity of a non-degenerate df H. Then, as $n \to \infty$,

$$\begin{aligned}
&P\big(Z_{n:n} \leq \operatorname{sign}(x)\alpha_n |x|^{\beta_n}\big) \longrightarrow H(x)\\
&\iff P(-Z_{n:n} \geq \alpha_n |x|^{\beta_n}) \longrightarrow H(x)\\
&\iff P\big(\log(-Z_{n:n}) \geq \beta_n \log(|x|) + \log(\alpha_n)\big) \longrightarrow H(x)\\
&\iff P\big((X_{1:n} - A_n)/B_n \geq \log(-x)\big) \longrightarrow H(x),
\end{aligned}$$

where $B_n := \beta_n$, $A_n := \log(\alpha_n)$ and $X_i := \log(-Z_i)$, $i = 1, \dots, n$ with df $F_*(x) = 1 - F(-\exp(x))$. Notice that the rv X_i is well defined, since $P(Z_i \geq 0) = 0$. In case $\omega(F) = 0$ and $\rho = P(Z_i = 0) > 0$, the limit H would necessarily be degenerate from (2.12). Hence, with $H_*(x) = 1 - H(-\exp(x))$ we find

$$F \in \mathcal{D}_p(H) \iff F_*(x) = 1 - F\big(\exp(x)\big) \in \mathcal{D}_{\min}(H_*),$$

and Theorem 2.1.1 leads to the representation

$$H(x) = 1 - L_i\big((\log(-x) - a)/b\big), \qquad x < 0,$$

for some $b > 0$, $a \in \mathbb{R}$ and some $i \in \{1, 2, 3\}$. This completes the proof. □

The characterization of p-max domains of attraction by the tail behavior of F and the sign of $\omega(F)$ as given in Theorems 3.1 - 3.3 by Mohan and Subramanya [340] (they are reproduced in Subramanya [433] as Theorems A, B, and C) now follows immediately from Theorem 2.6.1 and Theorem 2.1.1 for max domains under linear transformation. On the other hand, Theorems 2.2, 3.1, and 3.2 of Subramanya [433] are now a consequence of the above Theorem 2.6.1 and Theorems 10, 11 and 12 of de Haan [185] using only some substitutions.

Comparison of Max Domains of Attraction under Linear and Power Normalizations

Mohan and Ravi [339] compared the max domains of attraction under linear and power normalizations and proved the following result which shows that the class of max domains of attraction under linear normalization is a proper subset of the class of max domains of attraction under power normalization. This means that any df belonging to the max domain of attraction of some EVD limit law under linear normalization definitely belongs to the max domain of attraction of some p-max stable law under power normalization. Also, one can show that there are infinitely many df which belong to the max domain of attraction of a p-max stable law but do not belong to the max domain of attraction of any EVD limit law.

Theorem 2.6.2. *Let F be any df. Then*

$$\left.\begin{array}{ll}\text{(a)}\quad \text{(i)} & F \in \mathcal{D}_{\max}(G_{1,\gamma}) \\ \quad\quad\ \text{(ii)} & F \in \mathcal{D}_{\max}(G_3), \omega(F) = \infty\end{array}\right\} \Longrightarrow F \in \mathcal{D}_p(H_3),$$

(b) $F \in \mathcal{D}_{\max}(G_3), 0 < \omega(F) < \infty \Longleftrightarrow F \in \mathcal{D}_p(H_3), \omega(F) < \infty,$

(c) $F \in \mathcal{D}_{\max}(G_3), \omega(F) < 0 \Longleftrightarrow F \in \mathcal{D}_p(H_6), \omega(F) < 0,$

$$\left.\begin{array}{ll}\text{(d)}\quad \text{(i)} & F \in \mathcal{D}_{\max}(G_3), \omega(F) = 0 \\ \quad\quad\ \text{(ii)} & F \in \mathcal{D}_{\max}(G_{2,\gamma}), \omega(F) = 0\end{array}\right\} \Longrightarrow F \in \mathcal{D}_p(H_6),$$

(e) $F \in \mathcal{D}_{\max}(G_{2,\gamma}), \omega(F) > 0 \Longleftrightarrow F \in \mathcal{D}_p(H_{2,\gamma}),$

(f) $F \in \mathcal{D}_{\max}(G_{2,\gamma}), \omega(F) < 0 \Longleftrightarrow F \in \mathcal{D}_p(H_{4,\gamma}).$

Proof. Let $F \in \mathcal{D}_{\max}(G)$ for some $G \in \{G_1, G_2, G_3\}$. Then for some $a_n > 0, b_n \in \mathbb{R}$,

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G(x),\ x \in \mathbb{R}.$$

(a) (i) If $F \in \mathcal{D}_{\max}(G_{1,\gamma})$, then one can take $b_n = 0$ and hence setting $\alpha_n = a_n, \beta_n = 1/\gamma$,

$$\lambda^{(1)}(x) = \lambda_n^{(1)}(x) = \begin{cases} 0 & \text{if } x < 0, \\ x^{1/\alpha} & \text{if } 0 \le x, \end{cases}$$

we get

$$\lim_{n\to\infty} F^n\left(\alpha_n |x|^{\beta_n} \operatorname{sign}(x)\right) = \lim_{n\to\infty} F^n\left(a_n \lambda_n^{(1)}(x) + b_n\right) = G_{1,\gamma}(\lambda^{(1)}(x)) = H_3(x).$$

(ii) If $F \in \mathcal{D}_{\max}(G_3), \omega(F) = \infty$, then $b_n > 0$ for n large and $\lim_{n\to\infty} a_n/b_n = 0$. Setting $\alpha_n = b_n, \beta_n = a_n/b_n$,

$$\lambda_n^{(2)}(x) = \begin{cases} -1/\beta_n & \text{if } x \le 0, \\ \left(x^{\beta_n} - 1\right)/\beta_n & \text{if } 0 < x, \end{cases}$$

and

$$\lambda^{(2)}(x) = \begin{cases} -\infty & \text{if } x \le 0, \\ \log(x) & \text{if } 0 < x; \end{cases}$$

and proceeding as in the proof of (a) (i), we get $F \in \mathcal{D}_p(H_3)$ since $G_3(\lambda^{(2)}(x)) = H_3(x)$.

(b) If $F \in \mathcal{D}_{\max}(G_3), 0 < \omega(F) < \infty$, then the proof that $F \in \mathcal{D}_p(H_3)$ is the same as that in the case $\omega(F) = \infty$ above. So let $F \in \mathcal{D}_p(H_3)$ with $\omega(F) < \infty$. Then $\lim_{n\to\infty} \alpha_n = \omega(F), \lim_{n\to\infty} \beta_n = 0$. Setting $a_n = \alpha_n \beta_n, b_n = \alpha_n$,

$$u_n^{(1)}(x) = \begin{cases} 0 & \text{if } x \le -1/\beta_n, \\ (1 + \beta_n x)^{1/\beta_n} & \text{if } -1/\beta_n < x, \end{cases}$$

$$u^{(1)}(x) = \exp(x),$$

we get

$$\lim_{n\to\infty} F^n(a_n x + b_n) = \lim_{n\to\infty} F^n\left(\alpha_n |u_n^{(1)}(x)|^{\beta_n} \operatorname{sign}(x)\right) = H_3(u^{(1)}(x)) = G_3(x).$$

(c) If $F \in \mathcal{D}_{\max}(G_3), \omega(F) < 0$, then since $b_n < 0$ and $\lim_{n\to\infty} a_n = 0$, setting $\alpha_n = -b_n, \beta_n = -a_n/b_n$,

$$\lambda_n^{(3)}(x) = \begin{cases} (1 - |x|^{\beta_n})/\beta_n & \text{if} \quad x < 0, \\ 1/\beta_n & \text{if} \quad 0 \le x; \end{cases}$$

and

$$\lambda^{(3)}(x) = \begin{cases} -\log(-x) & \text{if} \quad x < 0, \\ \infty & \text{if} \quad 0 \le x; \end{cases}$$

we get $G_3(\lambda^{(3)} x) = H_6(x)$ and the claim follows as in the proof of (a) (i).
Now if $F \in \mathcal{D}_p(H_6)$, $\omega(F) < 0$, then $\lim_{n\to\infty} \alpha_n = -\omega(F), \lim_{n\to\infty} \beta_n = 0$. Setting $a_n = \alpha_n \beta_n, b_n = -\alpha_n$,

$$u_n^{(2)}(x) = \begin{cases} -(1 - \beta_n x)^{1/\beta_n} & \text{if} \quad x < 1/\beta_n, \\ 0 & \text{if} \quad 1/\beta_n \le x, \end{cases}$$

$$u^{(2)}(x) = -\exp(-x),$$

and proceeding as in the proof of (b), we get the result since $H_3(u^{(2)}(x)) = G_3(x)$.
(d) (i) Suppose $F \in \mathcal{D}_{\max}(G_3), \omega(F) = 0$. Then $\lim_{n\to\infty} b_n = 0$ and $\lim_{n\to\infty} a_n/b_n = 0$. Proceeding as in the proof of (c) we show that $F \in \mathcal{D}_p(H_6)$.
(ii) Now if $F \in \mathcal{D}_{\max}(G_{2,\gamma}), \omega(F) = 0$, then $b_n = 0$, and setting $\alpha_n = a_n, \beta_n = 1/\gamma$,

$$\lambda^{(4)}(x) = \lambda_n^{(4)}(x) = \begin{cases} -|x|^{1/\alpha} & \text{if} \quad x < 0, \\ 0 & \text{if} \quad 0 \le x, \end{cases}$$

we prove the claim as in (a)(i) using the fact that $G_{2,\gamma}(\lambda^{(4)})(x) = H_6(x)$.
(e) If $F \in \mathcal{D}_{\max}(G_{2,\gamma}), \omega(F) > 0$, then since $b_n = \omega(F)$ and $\lim_{n\to\infty} a_n = 0$, setting $\alpha_n = b_n, \beta_n = a_n/b_n$,

$$\lambda_n^{(5)}(x) = \begin{cases} -1/\beta_n & \text{if} \quad x \le 0, \\ (x^{\beta_n} - 1)/(\beta_n) & \text{if} \quad 0 < x \le 1, \\ 0 & \text{if} \quad 1 < x, \end{cases}$$

and

$$\lambda^{(5)}(x) = \begin{cases} -\infty & \text{if} \quad x \le 0, \\ \log(x) & \text{if} \quad 0 < x \le 1, \\ 0 & \text{if} \quad 1 < x; \end{cases},$$

we get $G_{2,\gamma}(\lambda^{(5)}(x)) = H_{2,\gamma}(x)$ and the claim follows as in the proof of (a) (i).
Now if $F \in \mathcal{D}_p(H_{2,\gamma})$, then $\lim_{n\to\infty} \alpha_n = \omega(F), \lim_{n\to\infty} \beta_n = 0$. Setting $a_n = \alpha_n \beta_n, b_n = \alpha_n$,

$$u_n^{(3)}(x) = \begin{cases} 0 & \text{if} \quad x \le -1/\beta_n, \\ (1 + \beta_n x)^{1/\beta_n} & \text{if} \quad 0 < x \le 1, \\ 1 & \text{if} \quad 0 < x, \end{cases}$$

$$u^{(3)}(x) = \begin{cases} \exp(x) & \text{if} \quad x < 0, \\ 1 & \text{if} \quad 0 \le x, \end{cases}$$

and proceeding as in the proof of (b), we get the result since $H_{2,\gamma}(u^{(3)}(x)) = G_{2,\gamma}(x)$.
(f) If $F \in \mathcal{D}_{\max}(G_{2,\gamma}), \omega(F) < 0$, then since $b_n = \omega(F)$ and $\lim_{n\to\infty} a_n = 0$, setting $\alpha_n = -b_n, \beta_n = -a_n/b_n$,

$$\lambda_n^{(6)}(x) = \begin{cases} (1 - |x|^{\beta_n})/\beta_n & \text{if} \quad x < -1, \\ 0 & \text{if} \quad -1 \le x, \end{cases}$$

and

$$\lambda^{(6)}(x) = \begin{cases} -\log(-x) & \text{if} \quad x < -1, \\ 0 & \text{if} \quad -1 \le x; \end{cases},$$

we get $G_{2,\gamma}(\lambda^{(6)}(x)) = H_{4,\gamma}(x)$ and the claim follows as in the proof of (a) (i).
Now if $F \in \mathcal{D}_p(H_{4,\gamma})$, then $\omega(F) < 0, \alpha_n = -\omega(F), \lim_{n\to\infty} \beta_n = 0$. Setting $a_n = \alpha_n\beta_n, b_n = -\alpha_n$,

$$u_n^{(4)}(x) = \begin{cases} -(1 - \beta_n x)^{\beta_n} & \text{if} \quad x < 0, \\ -1 & \text{if} \quad 0 \le x, \end{cases}$$

$$u^{(4)}(x) = \begin{cases} -\exp(-x) & \text{if} \quad x < 0, \\ -1 & \text{if} \quad 0 \le x, \end{cases}$$

and proceeding as in the proof of (b), we get the result since $H_{4,\gamma}(u^{(4)}(x)) = G_{2,\gamma}(x)$. The proof of the theorem is complete. □

Comparison of Max Domains of Attraction under Linear and Power Normalizations - The Multivariate Case

In this section we generalize Theorem 2.6.2 to the multivariate case. If $F \in \mathcal{D}_{\max}(G)$ for some max stable law G on $\mathbb{R}^d$ then we denote the normalizing constants by $a_n(i) > 0$ and $b_n(i)$, $i \le d$, so that

$$\lim_{n\to\infty} F^n\left(a_n(i)x_i + b_n(i), 1 \le i \le d\right) = G(\mathbf{x}), \qquad \mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d.$$

Similarly, if $F \in \mathcal{D}_p(H)$ for some p-max stable law H on $\mathbb{R}^d$ then we denote the normalizing constants by $\alpha_n(i) > 0$ and $\beta_n(i), 1 \le i \le d$, so that

$$\lim_{n\to\infty} F^n\left(\alpha_n(i)|x_i|^{\beta_n(i)}, 1 \le i \le d\right) = H(\mathbf{x}), \qquad \mathbf{x} \in \mathbb{R}^d.$$

For a df F on $\mathbb{R}^d$, let $F_{i(1)\ldots i(k)}$ denote the $(i(1)\ldots i(k))$-th $k-$variate marginal df, $1 \le i(1) < \cdots < i(k) \le d, 2 \le k \le d$.

Theorem 2.6.3. *Let $F \in \mathcal{D}_{\max}(G)$ for some max stable law G under linear normalization. Then there exists a p-max stable law H on $\mathbb{R}^d$ such that $F \in \mathcal{D}_p(H)$.*

Proof. Let $F \in \mathcal{D}_{\max}(G)$. Then for all $i \leq d, F_i \in \mathcal{D}_{\max}(G_i)$. Hence by Theorem 2.6.2, $F_i \in D_p(H_i)$, for some p-max stable law H_i which must be necessarily a p-type of one of the four p-max stable laws $H_{2,\gamma}, H_{4,\gamma}, H_3, H_6$. The normalization constants $\alpha_n(i), \beta_n(i)$ are determined by $a_n(i), b_n(i)$ as in the proof of Theorem 2.6.2. Further, it follows from the proof of Theorem 2.6.2 that there exists $\theta_n^{(i)}(x_i)$ such that

$$\begin{aligned}\lim_{n\to\infty} F_i^n\left(\alpha_n(i)|x_i|^{\beta_n(i)}\operatorname{sign}(x_i)\right) &= \lim_{n\to\infty} F_i^n\left(a_n(i)\theta_n^{(i)}(x_i)+b_n(i)\right)\\ &= G_i\left(\theta^{(i)}(x_i)\right),\end{aligned}$$

where $\theta_n^{(i)}$ is one of the $\lambda_n^{(j)}$, $j \leq 6$, defined in the proof of Theorem 2.6.2 depending upon which one of the conditions is satisfied by F_i and $\theta^{(i)} = \lim_{n\to\infty}\theta_n^{(i)}$. So, $H_i(x_i) = G_i\left(\theta^{(i)}(x_i)\right)$, $i \leq d$. Now fix $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$. If, for some $j \leq d$, $H_j(x_j) = 0$, then by Theorem 2.6.2, we have

$$F_i^n\left(\alpha_n(i)|x_i|^{\beta_n(i)}\operatorname{sign}(x_i),\, i \leq d\right) \leq F_j^n\left(\alpha_n(j)|x_j|^{\beta_n(j)}\operatorname{sign}(x_j)\right) \to 0,$$

as $n \to \infty$. Suppose now that for some integers $k, i(1), \ldots, i(k)$, we have $0 < H_{i(j)}\left(x_{i(j)}\right) < 1$, $j \leq k$, and $H_i(x_i) = 1$, $i \neq i(1), \ldots, i(k)$. Using uniform convergence, we have

$$\begin{aligned}&\liminf_{n\to\infty} F^n\left(\alpha_n(i)|x_i|^{\beta_n(i)}\operatorname{sign}(x_i),\, i \leq d\right)\\ &\geq \lim_{n\to\infty} F^n\left(a_n(i)\theta_n^{(i)}(x_i)+b_n(i),\, i \leq d\right)\\ &= G\left(\theta^{(i)}x_i,\, i \leq d\right)\\ &= G_{i(1)\ldots i(k)}\left(\theta^{(i(j))}(x_{i(j)}),\, j \leq k\right),\end{aligned}$$

since $H_i(x_i) = G_i\left(\theta^{(i)}(x_i)\right) = 1$, $i \neq i(1), \ldots, i(k)$. Again

$$\begin{aligned}&\limsup_{n\to\infty} F^n\left(\alpha_n(i)|x_i|^{\beta_n(i)}\operatorname{sign}(x_i),\, i \leq d\right)\\ &\leq \lim_{n\to\infty} F^n\left(a_n(i(j))\theta_n^{(i(j))}(x_{i(j)})+b_n(i(j)),\, j \leq k\right)\\ &= G_{i(1)\ldots i(k)}\left(\theta^{(i(j))}(x_{i(j)}),\, j \leq k\right).\end{aligned}$$

The claim now follows with

$$H(\mathbf{x}) = G\left(\theta^{(1)}(x_1), \ldots, \theta^{(d)}(x_d)\right). \qquad \square$$

In view of Theorems 2.6.2 and 2.6.3, it is clear that p-max stable laws collectively attract more distributions than do max stable laws under linear normalization collectively.

Examples

If F is the uniform distribution on $(-1, 0)$, then $\omega(F) = 0$ and (2.11) holds with $H = H_6$, $\alpha_n = 1/n$ and $\beta_n = 1$. Since $H_{2,1}$ is the uniform distribution on $(0, 1)$, it is p-max stable and (2.11) holds with $F = H = H_{2,1}$, $\alpha_n = 1$ and $\beta_n = 1$. For the uniform distribution F on $(-2, -1)$ we find $\omega(F) = -1$ and (2.11) holds with $H = H_{5,1}$, $\alpha_n = 1$ and $\beta_n = 1/n$.

If $F = F_\varepsilon$ is the uniform distribution on $(-1 + \varepsilon, \varepsilon)$ with $0 \leq \varepsilon < 1$, then $\omega(F_\varepsilon) = \varepsilon$. Here $F_\varepsilon \in \mathcal{D}_p(H_{2,1})$, and (2.11) holds with $\alpha_n = \varepsilon$ and $\beta_n = 1/(\varepsilon n)$ if $\varepsilon > 0$, whereas for $\varepsilon = 0$ we find (as mentioned above) $F_0 \in \mathcal{D}_p(H_6)$ with power-norming constants $\alpha_n = 1/n$ and $\beta_n = 1$. On the other hand for any fixed $n \geq 1$ we find $F_\varepsilon^n(\operatorname{sign}(x)\varepsilon|x|^{1/(\varepsilon n)}) \to 1_{(-\infty,-1]}(x)$ as $\varepsilon \to 0$. Here the limit distribution is degenerate.

The min stable df $L_{2,1}$ is an exponential law. On the other hand, $L_{2,1} \in \mathcal{D}_{\max}(G_3)$. If $F = L_{2,1}$, then $F(\exp(x)) = L_3(x) \in \mathcal{D}_{\max}(G_3)$. It follows from Theorem 2.6.1 that (2.11) holds with $H = H_3$, $\beta_n = 1/\log(n)$ and $\alpha_n = \log(n)$.

Let $F(x) = 1 - x^{-k}$ for $x \geq 1$, where $k > 0$. Then $F^n(n^{1/k}\, x) \to G_{1,\,k}(x)$ as $n \to \infty$, whereas by power normalization $F^n(n^{1/k}\, x^{1/k}) \to H_3(x)$. Note that $G_{1,\,1} = H_3$.

The df $G_{1,\gamma}$ for $\gamma > 0$ are under power transformation of type H_3, whereas the df $L_{1,\gamma}$ for $\gamma > 0$ are under power transformation of type H_6.

The max stable and min stable df are connected by the equation

$$L_i(x) = 1 - G_i(-x), \qquad x \in \mathbb{R},\ i = 1, 2, 3.$$

Under the six p-max stable df $H_1, \dots, H_6$ there are again three pairs. If the rv U has the df H_i, then the rv $V = -1/U$ has df H_{i+3}, $i = 1, 2$ or 3. The set of possible limit distributions of $Z_{1:n}$ under power normalization

$$(|Z_{1:n}|/\alpha_n^*)^{1/\beta_n^*} \operatorname{sign}\,(Z_{1:n}) \longrightarrow_D Z, \qquad n \to \infty,$$

for some suitable constants $\alpha_n^* > 0$ and $\beta_n^* > 0$ can be obtained from Theorem 2.6.1: The limit df equal up to a possible power transformation $1 - H_i(-x)$, $i = 1, \dots, 6$.

Put $F_1(x) = 1 - (\log(x))^{-1}$ for $x \geq e$. Then F_1 does not belong to any of $\mathcal{D}_{\max}$, but $F_1 \in \mathcal{D}_p(H_{1,\,1})$ with $\alpha_n = 1$ and $\beta_n = n$, see Galambos [167], Example 2.6.1, and Subramanya [433], Example 1. Taking now $F_2(x) = 1 - (\log\log(x))^{-1}$ for $x \geq \exp(e)$, then without any calculations it follows that F_2 does not belong to any of $\mathcal{D}_p$ since $F_2(\exp(x)) = F_1(x)$ does not belong to any of $\mathcal{D}_{\max}$.

If

$$F(x) = \begin{cases} 0 & \text{if } \ x < 1, \\ 1 - \exp(-(\log(x))^2) & \text{if } \ 1 \leq x, \end{cases}$$

then $F \in \mathcal{D}_p(H_3)$ with $\alpha_n = \exp(\sqrt{\log(n)})$, $\beta_n = 1/(2\sqrt{\log(n)})$. However, F does not belong to $\mathcal{D}_{\max}(G_{1,\gamma})$ or to $\mathcal{D}_{\max}(G_3)$.

If

$$F(x) = \begin{cases} 0 & \text{if} \quad x < -1, \\ 1 - \exp\left(-(\sqrt{-\log(-x)})\right) & \text{if} \quad 1 \le x < 0, \\ 1 & \text{if} \quad 0 \le x, \end{cases}$$

then $F \in \mathcal{D}_p(H_6)$ with $\alpha_n = \exp(-(\log(n))^2)$, $\beta_n = 2\log(n)$. Note that F does not belong to any $\mathcal{D}_{\max}(G_i)$ for any $i = 1, 2, 3$.

Note that df belonging to $\mathcal{D}_p(H_{1,\gamma})$ or $\mathcal{D}_p(H_{3,\gamma})$ do not belong by Theorem 2.6.2 to the max domain of attraction of any EVD limit law under linear normalization.

Applications of power-normalizations to the analysis of super-heavy tails are included in the subsequent section.

2.7 Heavy and Super-Heavy Tail Analysis

Distributions with light tails like the normal or the exponential distributions have been of central interest in classical statistics. Yet, to cover in particular risks in fields like flood frequency analysis, insurance and finance in an appropriate manner, it became necessary to include distributions which possess heavier tails. An early reference is the modeling of incomes by means of Pareto distributions.

One standard method to get distributions with heavier tails is the construction of log-distributions. Prominent examples are provided by log-normal and log-exponential distributions whereby the latter belongs to the above mentioned class of Pareto distributions. Normal and log-normal distributions possess an exponential decreasing upper tail and, as a consequence all moments of these distributions are finite. However, a log-normal distribution exhibits a higher kurtosis than the normal and, in this sense, its upper tail can be also considered heavier than that of the normal one. It is also a general rule that the mixture of distributions, as a model for heterogeneous populations, leads to heavier tails. For instance, log-Pareto df can be deduced as mixtures of Pareto df with respect to gamma df.

In contrast to normal, log-normal and exponential distributions, one can say that Pareto distributions are prototypes of distributions with heavy, upper tails. An important characteristic of this property is that not all moments are finite. Other prominent distributions of this type are, e.g., Student and sum-stable distributions with the exception of the normal one. All these distributions have power decreasing upper tails, a property shared by Pareto distributions.

Distributions with heavy tails have been systematically explored within the framework of extreme value theory with special emphasis laid on max-stable Fréchet and Pareto distributions where the latter ones possess a peaks-over-threshold (POT) stability. More precisely, one may speak of linearly max-stable or linearly POT-stable (l-max or l-POT) distributions in view of our explanations on page 66 and the remainder of this section. Related to this is the property that a distribution belongs to the l-max or l-POT domain of attraction of a Fréchet or

Pareto distribution if, and only if, the upper tail is regularly varying with negative index. In contrast to this, slowly varying upper tails will be of central importance in the subsequent context.

This means that we are out of the *"power-law-world"*, as Taleb's book, *"The Black Swan: the Impact of the Highly Improbable"* [439] entitles the class of distributions possessing a regularly varying upper tail or, equivalently, with polynomially decreasing upper tails. The designation of super-heavy concerns right tails decreasing to zero at a slower rate, as logarithmic, for instance. This also means that the classical *bible* for inferring about rare events, the Extreme Value Theory, is no longer applicable, since we are in presence of distributions with slowly varying tails.

We give a short overview of the peaks-over-threshold approach which is the recent common tool for statistical inference of heavy tailed distributions. Later on, we present extensions to the super-heavy tailed case.

Heavy Tail Analysis

We shortly address the peaks-over-threshold approach as already described at the end of Chapter 1 but take a slightly different point of view. We do not start with the assumption that a df F is in the max domain of attraction of some EVD G but we consider limiting distributions of exceedances in their own right. Recently, this has been the most commonly used statistical approach for heavy tailed distributions.

Recall that (1.22) indicates that, if $F \in \mathcal{D}(G)$ then GPD are the only possible limiting distributions of the linear normalized df $F^{[u_n]}$ of exceedances over thresholds u_n tending to $\omega(F)$. Hereby, $\omega(F) = \sup\{x : F(x) < 1\}$ is the right endpoint of the support of the df F and

$$F^{[u]}(x) = \frac{F(x) - F(u)}{1 - F(u)}, \quad x \geq u,$$

is the common df of exceedances above the threshold u. If X is a rv with df F then the exceedance df may be written as $F^{[u]}(x) = P(X \leq x \mid X > u)$ for $x \geq u$.

In what follows we assume that there exist real-valued functions $a(\cdot)$ and $b(\cdot) > 0$ such that

$$F^{[u]}(a(u) + b(u)x) \longrightarrow_{u \to \omega(F)} L(x) \tag{2.14}$$

for some non-degenerate df L. If (2.14) holds for df F and L we say that F is in the POT domain of attraction of L denoted by $F \in \mathcal{D}_{\mathrm{POT}}(L)$.

Notice that (2.14) can be formulated in terms of the survivor function $1 - F$ as

$$\frac{1 - F(a(u) + b(u)x)}{1 - F(u)} \longrightarrow_{u \to \omega(F)} 1 - L(x),$$

which corresponds to formula (1.21).

Due to results in Balkema and de Haan [22], stated as Theorem 2.7.1 below, we know that the limiting df L is POT-stable. Hereby, a df F is POT-stable if there exists constants $a(u)$ and $b(u) > 0$ such that

$$F^{[u]}(a(u) + b(u)x) = F(x) \tag{2.15}$$

for all points of continuity u in the interior of the support of F and $F(x) > 0$. The class of continuous POT-stable df, and, therefore, that of continuous limiting df of exceedances in (2.14) is provided by the family of generalized Pareto df (GPD). This result is stated in the following theorem which can be regarded as an extension of Theorem 1.3.5.

Theorem 2.7.1. *Let F be an df and L be a non-degenerate df. Suppose there exists real-valued functions $a(\cdot)$ and $b(\cdot) > 0$ such that*

$$F^{[u]}(a(u) + b(u)x) \longrightarrow_{u\to\omega(F)} L(x)$$

for all points of continuity x of F in the interior of its support. Then,

(i) *the limiting df L is POT-stable,*

(ii) *if L is continuous, then L is up to a location and scale shift a GPD W_γ,*

(iii) *the POT domain of attraction of a GPD W_γ coincides with the max domain of attraction of the corresponding EVD G_γ, thus $\mathcal{D}_{\text{POT}}(W_\gamma) = \mathcal{D}(G_\gamma)$.*

It is evident that all POT-stable df L appear as limiting df in Theorem 2.7.1 by choosing $F = L$. Therefore, GPD are the only continuous, POT-stable df.

For statistical applications, e.g., high quantile estimation, these results are of particular importance. Note that high q-quantiles $F^{-1}(q)$ of a df F with $q > F(u)$ for some threshold u only depend on the tail of F, thus $F(x)$ for $x > u$. Notice that for a df F and $x > u$,

$$\begin{aligned} F(x) &= F(u) + (1 - F(u))F^{[u]}(x) \\ &\approx F(u) + (1 - F(u))W_{\gamma,u,\sigma}(x) \end{aligned} \tag{2.16}$$

where the approximation is valid if $F \in \mathcal{D}_{\text{POT}}(W_\gamma)$ and u is sufficiently large. Now (2.16) provides a certain parametric model for the tail of F where a non-parametric part $F(u)$ can be replaced by an empirical counterpart. A similar "piecing-together approach" can also be utilized in the multivariate framework, cf. Chapter 5.

In what follows we concentrate on the heavy tail analysis, that is, on df F belonging to $\mathcal{D}_{\text{POT}}(W_\gamma)$ for some $\gamma > 0$, for which case $\omega(F) = \infty$. The model (2.16) offers the possibility to carry out statistical inference for such df. These df have the special property that their pertaining survivor function $1 - F$ is of *regular variation at infinity*. We include some remarks concerning the theory of regular varying functions and point out relations to the concept of POT-stability.

Regular and Slow Variation

Consider the Pareto distribution $W_{1,\alpha,0,\sigma}(x) = 1 - (x/\sigma)^{-\alpha}$, $x > \sigma > 0$, in the α-parametrization, with $\alpha > 0$. Recall that for any df F the survivor function of the pertaining exceedance df $F^{[u]}$ satisfies

$$\overline{F^{[u]}}(x) = \bar{F}(x)/\bar{F}(u), \quad x > u.$$

For $F = W_{1,\alpha,0,\sigma}$ and replacing x by ux one gets

$$\bar{F}(ux)/\bar{F}(u) = \overline{F^{[u]}}(ux) = x^{-\alpha}$$

which is the POT-stability of $W_{1,\alpha,0,\sigma}$. If F is an arbitrary Pareto df $W_{1,\gamma,\mu,\sigma}$ with additional location parameter μ this relation holds in the limit. We have

$$\bar{F}(ux)/\bar{F}(u) = \overline{F^{[u]}}(ux) = \left(\frac{x - \mu/u}{1 - \mu/u}\right)^{-\alpha} \longrightarrow_{u\to\infty} x^{-\alpha}, \qquad x \geq 1.$$

This implies that $\bar{F}$ is regularly varying at infinity according to the following definition: A measurable function $R : (0,\infty) \to (0,\infty)$ is called regularly varying at infinity with index (exponent of variation) ρ, denoted by $R \in RV_\rho$, if

$$R(tx)/R(t) \longrightarrow_{t\to\infty} x^\rho, \quad x > 0. \tag{2.17}$$

A comprehensive treatment of the theory of regular variation may, e.g., be found in Bingham et al. [46]. If $\rho = 0$ we have

$$R(tx)/R(t) \longrightarrow_{t\to\infty} 1;$$

in this case, R is said to be of slow variation at infinity ($R \in RV_0$). Recall from Theorem 2.1.1 together with Theorem 2.7.1, part (iii), that a df F is in the POT domain of attraction of some GPD W_γ, $\gamma > 0$, if $\bar{F} \in RV_{-1/\gamma}$. For any $R \in RV_\rho$ we have the representation

$$R(x) = x^\rho U(x), \quad x > a,$$

for some $a > 0$ sufficiently large and $U \in RV_0$. If a df F is in the POT domain of attraction of some GPD W_γ for $\gamma > 0$ (thus, $\bar{F} \in RV_{-1/\gamma}$) we call F heavy tailed.

The existence of finite β-power moments is restricted to the range $\beta < 1/\gamma$. Although there is no unified agreement on terminology, in literature the term *very heavy tailed* case has been attached to a degree of tail heaviness given by $0 < 1/\gamma < 1$.

Super-Heavy Tails and Slow Variation

The use of heavy tailed distributions constitutes a fundamental tool in the study of rare events and have been extensively used to model phenomena for which extreme values occur with a relatively high probability. Here, emphasis lies on the modelling of extreme events, i.e., events with a low frequency, but mostly with a high and often disastrous impact. For such situations it has became reasonable to consider an underlying distribution function F with polynomially decaying right tail, i.e., with tail distribution function

$$\bar{F} := 1 - F \in RV_{-\alpha}, \quad \alpha > 0. \tag{2.18}$$

Generalizing this heavy tail framework, it is also possible to consider the so-called *super-heavy tailed* case, for which $\alpha = 0$, i.e., $1 - F$ is a slowly varying function, decaying to zero at a logarithmic rate, for instance. We will consider two special classes of such super-heavy tailed df.

Class A. Notice that if X has a df F such that $\bar{F} \in RV_{-\alpha}$, for some positive α, then $Z := e^X$ has the df H with

$$\bar{H}(x) \sim (\log(x))^{-\alpha} U(\log(x)) \tag{2.19}$$

as $x \to \infty$, with $U \in RV_0$, meaning that the tail decays to zero at a logarithmic rate raised to some power. Although this transformation leads to a super-heavy tailed df it does not exhaust all possible slowly varying tail types. On the other hand, for the super-heavy tailed case there is no possible linear normalization of the maxima such that F belongs to any max-domain of attraction. Consider the case $U \equiv 1$ in (2.19). This gives the super-heavy tailed df $F(x) = 1 - \log(x)^{-\alpha}$, $x \geq e$. The pertaining survivor function satisfies

$$\bar{F}\left(x^{\log(u)}\right) / \bar{F}(u) = \bar{F}(x).$$

Subsequently, this property will be called the power-POT (p-POT) stability of F, it characterizes the class of limiting df of exceedances under power-normalization, cf. Theorem 2.7.2. Corresponding to the case of heavy tailed df an asymptotic version of this relation will be identified as an necessary and sufficient condition of a df to belong to certain p-POT domains of attraction, cf. Theorem 2.7.5.

Class B. The df F satisfies (2.18) if, and only if, there exists a positive function a such that

$$\lim_{t \to \infty} \frac{F(tx) - F(t)}{a(t)} = \frac{1 - x^{-\alpha}}{\alpha}, \quad x > 0. \tag{2.20}$$

For the latter it is enough to consider the auxiliary function $a = \alpha \bar{F}$ and thus $a \in RV_{-\alpha}$, $\alpha > 0$. A sub-class of slowly varying df is deduced from (2.20) by extension to the case of $\alpha = 0$, through the limit of the right-hand side of (2.20), as $\alpha \to 0$:

$$\lim_{t \to \infty} \frac{F(tx) - F(t)}{a(t)} = \log(x). \tag{2.21}$$

The above relation identifies the well-known class Π (cf., e.g., de Haan and Ferreira [190]). The class of super-heavy tailed distributions is characterized by (2.21).

More details about all distributions satisfying (2.20) with $\alpha \geq 0$ will be provided at the end of this section, together with testing procedures for super-heavy tails (see Theorems 2.7.12 and 2.7.13).

For the time being notice that the df given by $1 - 1/\log(x)$, $x > e$ belongs to both Classes A and B. Moreover, according to Proposition 2.7.10, any df H in Class A and resulting from composition with a df F such that the density $F' =: f$ exists, also belongs to the Class B. However, the reverse is not always true: for instance, the df $H(x) = 1 - 1/\log(\log(x))$, for $x > e^e$, belongs to B but not to A. Note that H is obtained by iterated log-transforms upon a Pareto df. Distributions of this type are investigated in Cormann [77] and Cormann and Reiss [76].

In the remainder of this section we study two special aspects of the statistical analysis of super-heavy tailed df. First we deal with asymptotic models for certain super-heavy tailed df related to df given by Class A. The second part concerns testing the presence of a certain form of super-heavy tails, namely Π-varying tailed distributions given by Class B.

Super-Heavy Tails in the Literature

We first give a short outline of the statistical literature dealing with super-heavy tails. Although there is no uniform agreement on terminology, the term *super-heavy tailed* has been attached, in the literature, to a degree of tail heaviness associated with slow variation. Examples of models with slowly varying tail are the log-Pareto, log-Fréchet and log-Cauchy distributions. We say that the rv X is a log-Pareto rv if $\log(X)$ is a Pareto rv.

In Galambos [166], Examples 1.3.3 and 2.6.1, the log-Pareto df

$$F(x) = 1 - 1/\log(x), \quad x > e, \tag{2.22}$$

serves as a df under which maxima possess "shocking" large values, not attained under the usual linear normalization pertaining to the domain of attraction of an EVD. Some theoretical results for super-heavy tailed distributions can be found in Resnick [392], Section 5, which is devoted to fill some "interesting gaps in classical limit theorems, which require the assumption that tails are even fatter than regularly varying tails". Two cases are considered in some probabilistic descriptions of *"fat"* tails, under the context of point process convergence results: slowly varying tails and its subclass of Π-varying distribution functions.

Another early reference to df with slowly varying tails, in conjunction with extreme value analysis, can be found in the book by Reiss and Thomas [388], Section 5.4, where log-Pareto distributions are regarded as mixtures of Pareto df with respect to gamma df. The authors have coined all log-Pareto df with the term "super-heavy" because the log-transformation leads to a df with a heavier tail than the tail heaviness of Pareto type.

Log-Pareto df within a generalized exponential power model are studied by Desgagné and Angers [100]. In a technical report, see Diebolt et al. [110], associated to Diebolt et al. [111], the authors mention another mixture distribution, different from the log-Pareto one, with super-heavy tails.

Moreover, Zeevi and Glynn [470] have studied properties of autoregressive processes with super-heavy tailed innovations, specifically, the case where the innovations are log-Pareto distributed. Their main objective was to illustrate the range of behavior that AR processes can exhibit in this super-heavy tailed setting. That paper studies recurrence properties of autoregressive (AR) processes with "super-heavy tailed" innovations. Specifically, they study the case where the innovations are distributed, roughly speaking, as log-Pareto rvs (i.e., the tail decay is essentially a logarithm raised to some power).

In Neves and Fraga Alves [352] and in Fraga Alves et al. [159] the tail index α is allowed to be 0, so as to embrace the class of super-heavy tailed distributions. Statistical tests then are developed in order to distinguish between heavy and super-heavy tailed probability distributions. This is done in a semi-parametric way, i.e., without specifying the exact df underlying the data in the sense of composite hypothesis testing. Therein, the authors present some simulation results concerning estimated power and type I error of the test. Application to data sets in teletraffic and seismology fields is also given.

Cormann and Reiss [76] introduced a full-fledged statistical model of log-Pareto distributions parametrized with two shape parameters and a scale parameter and show that these distributions constitute an appropriate model for super-heavy tailed phenomena. Log-Pareto distributions appear as limiting distributions of exceedances under power-normalization. Therein it is shown, that the well-known Pareto model is included in the proposed log-Pareto model for varying shape-parameters whence the log-Pareto model can be regarded as an extension of the Pareto model. This article also explores an hybrid maximum likelihood estimator for the log-Pareto model. The testing of the Pareto model against the log-Pareto model is considered in Villaseñor-Alva et al. [450].

The need of distributions with heavier tails than the Pareto type has also been claimed in Earth Sciences research. A usual statistical data analysis in seismology is done through the scalar seismic moment M, which is related to the earthquake moment magnitude m according to: $M = 10^{3(m+6)/2}$ (notice the power transformation with consequences on the distribution tail weight). Zaliapin et al. [469] presents an illustration of the distribution of seismic moment for Californian seismicity ($m \geq 5.5$), during the last two centuries, using an earthquake catalog and converting its magnitudes into seismic moments. They observed that

> *. . . with such a data set one does not observe fewer earthquakes of large seismic moment than expected according to the Pareto law. Indeed, . . . may even suggest that the Pareto distribution underestimates the frequency of earthquakes in this seismic moment range.*

In fact, these authors called the attention to the practitioners that:

Statistical data analysis, a significant part of modern Earth Sciences research, is led by the intuition of researchers traditionally trained to think in terms of "averages", "means", and "standard deviations". Curiously, an essential part of relevant natural processes does not allow such an interpretation, and appropriate statistical models do not have finite values of these characteristics.

The same data set has been analyzed by Neves and Fraga Alves [352] in the context of detecting super-heavy tails.

The P-Pot Stable Distributions

Recall the log-Pareto df $F(x) = 1 - 1/\log(x)$, $x \geq e$ mentioned above as an important example of a super-heavy tailed df. Such distributions cannot be studied within the POT-framework outlined in Sections 2.1 to 2.4 because they possess slowly varying tails. Nevertheless, p-max domains of attraction in Section 2.6 contain certain super-heavy tailed distributions.

We have noted in the previous section that the distribution of the largest order statistic $Z_{n:n}$ out of an iid sample $Z_1, \dots, Z_n$ can be approximated by certain p-max stables laws even if the common df belongs to a certain subclass of distributions with slowly varying distribution tails. In the present section we derive an asymptotic model for the upper tail of such a df F. Similarly to the linear normalization we consider the asymptotic relation

$$F^{[u]}\left(\operatorname{sign}(x)\alpha(u)|x|^{\beta(u)}\right) \longrightarrow_{u\to\omega(F)} L(x) \tag{2.23}$$

for all points of continuity x of L, where L is a non-degenerate df and $\alpha(\cdot)$, $\beta(\cdot) > 0$. Notice that (2.23) is equivalent to

$$\frac{1 - F\left(\operatorname{sign}(x)\alpha(u)|x|^{\beta(u)}\right)}{1 - F(u)} \longrightarrow_{u\to\omega(F)} 1 - L(x). \tag{2.24}$$

Recall that limiting distributions of exceedances under linear normalization are POT-stable. A similar results holds for limiting distributions under power-normalization. These distributions satisfy the p-POT stability property. A df F is p-POT stable if there are positive constants $\beta(u)$ and $\alpha(u)$ such that

$$F^{[u]}(\operatorname{sign}(x)\alpha(u)|x|^{\beta(u)}) = F(x) \tag{2.25}$$

for all x with $F(x) > 0$ and all continuity points u of F with $0 < F(u) < 1$.

Due to Theorem 2.7.1 we know that GPD are the only continuous POT-stable distributions. According to Theorem 1 in Cormann and Reiss [76], stated below as Theorem 2.7.2, we know that for every p-POT stable df L there is a POT-stable df W such that $L(x) = W(\log(x))$ if $\omega(L) > 0$, or $L(x) = W(-\log(-x))$ if $\omega(L) \leq 0$. As in Mohan and Ravi [339] we call a df F_1 a p-type of F_2, if $F_1(x) = F_2\left(\operatorname{sign}(x)\alpha|x|^{\beta}\right)$ for positive constants α and β.

Given a df F we define auxiliary df F^{**} and F_{**} by

$$F^{**}(x) = \frac{F(\exp(x)) - F(0)}{1 - F(0)}, \quad x \in \mathbb{R}, \tag{2.26}$$

if $\omega(F) > 0$, and

$$F_{**}(x) = F(-\exp(-x)), \quad x \in \mathbb{R}, \tag{2.27}$$

if $\omega(F) \leq 0$. These auxiliary df play a similar role for limiting df of exceedances as do the df F^* and F_* in Theorem 2.6.1 in the context of limiting df of maxima.

Theorem 2.7.2. *Let F be a df which is p-POT stable, cf.* (2.25). *Then,*

$$F(x) = W\left(\log(x)\right),$$

or

$$F(x) = W\left(-\log(-x)\right),$$

where W denotes a POT-stable df.

Proof. Let $0 < F(u) < 1$. First note that (2.25) is equivalent to

$$\frac{1 - F(\mathrm{sign}(x)\alpha(u)|x|^{\beta(u)})}{1 - F(u)} = 1 - F(x).$$

Let $F(0) > 0$. Then

$$\frac{1 - F(0)}{1 - F(u)} = 1 - F(0)$$

and $F(0) = 1$ because $0 < F(u) < 1$. Thus, we have $F(0) = 0$ or $F(0) = 1$ and, consequently, F has all the mass either on the positive or negative half-line.

(a) Let $F(0) = 0$ and, therefore, $F(x) = 0$ for all $x < 0$. It suffices to consider $x, u > 0$. Let $x > 0$, $F(x) > 0$ and $0 < F(u) < 1$. Then, (2.25) yields

$$\frac{1 - F\left(\alpha(u)x^{\beta(u)}\right)}{1 - F(u)} = 1 - F(x).$$

It follows that

$$\frac{1 - F\left(\alpha(\exp(u))\exp(x)^{\beta(\exp(u))}\right)}{1 - F(\exp(u))} = 1 - F(\exp(x))$$

for all x and continuity points u of $F(\exp(\cdot))$ with $F\left(\exp(x)\right) > 0$ and $0 < F\left(\exp(u)\right) < 1$. Furthermore,

$$\begin{aligned} &\frac{1 - F\left(\alpha(\exp(u))\exp(x)^{\beta(\exp(u))}\right)}{1 - F(\exp(u))} = 1 - F(\exp(x)) \\ \iff &\frac{1 - F\left(\exp\left(\log(\alpha(\exp(u))) + \beta(\exp(u))x\right)\right)}{1 - F\left(\exp(u)\right)} = 1 - F\left(\exp(x)\right). \end{aligned}$$

Observe that $F^{**} := F(\exp(\cdot))$ since $F(0) = 0$. The above computations yield

$$\frac{1 - F^{**}\left(\tilde{\alpha}(u) + \tilde{\beta}(u)x\right)}{1 - F^{**}(u)} = 1 - F^{**}(x)$$

with $\tilde{\alpha}(u) = \log\left(\alpha(\exp(u))\right)$ and $\tilde{\beta}(u) = \beta(\exp(u))$. Consequently, $F^{**} = W$ for some POT-stable df W and $F(\cdot) = W(\log(\cdot))$.

(b) Next assume that $F(0) = 1$. Let (2.25) hold for $x < 0$, $F(x) > 0$ and all continuity points u of F with $0 < F(u) < 1$. Then, similar arguments as in part (a) yield that (2.25) is equivalent to

$$\frac{1 - F_{**}(\tilde{\alpha}(u) + \tilde{\beta}(u)x)}{1 - F_{**}(u)} = 1 - F_{**}(x)$$

with $F_{**}(x) := F(-\exp(-x))$ where $\tilde{\alpha}(u)$ can be chosen as $\alpha(-\exp(-u))$ and $\tilde{\beta}(u) = \beta(-\exp(-u))$. Thus, F_{**} is a POT-stable df W and $F(x) = W\left(-\log(-x)\right)$, $x \le 0$. □

Due to the foregoing theorem all continuous p-POT stable df are p-types of the df

$$L_\gamma(x) = 1 - (1 + \gamma \log(x))^{-1/\gamma}, \quad x > 0,\ \gamma \in \mathbb{R} \tag{2.28}$$

which is a generalized log-Pareto distribution (GLPD), or

$$V_\gamma(x) := 1 - (1 - \gamma \log(-x))^{-1/\gamma}, \quad x < 0,\ \gamma \in \mathbb{R} \tag{2.29}$$

which may be addressed as negative generalized log-Pareto df. The case $\gamma = 0$ is again taken as the limit $\gamma \to 0$. Notice that only the p-POT stable law L_γ, $\gamma > 0$ is super-heavy tailed, while L_γ, $\gamma < 0$ and V_γ possess finite right endpoints. The df L_0 is a Pareto df and, therefore, heavy tailed.

Relations to p-max Stable Laws

We start with a representation of log-Pareto df by means of p-max stable df. Recall that a df F is p-max stable if there exist sequences α_n, $\beta_n > 0$ such that

$$F^n(\operatorname{sign}(x)\alpha_n |x|^{\beta_n}) = F(x), \quad x \in \mathbb{R}$$

and all positive integers n, cf. Section 2.6.

For the special p-max stable df

$$H_{1,\gamma}(x) = \exp(-(\log(x))^{-\gamma}), \quad x \ge 1,$$

with $\gamma > 0$, define

$$\begin{aligned} F_\gamma(x) &= 1 + \log\left(H_{1,\gamma}(x)\right) \\ &= 1 - (\log(x))^{-\gamma}, \quad x \ge \exp(1), \end{aligned} \tag{2.30}$$

which is a log-Pareto df with shape parameter $1/\gamma$.

In analogy to (2.30), the whole family of GLPDs in (2.28) can be deduced from p-max stable laws $H_{i,\gamma,\beta,\sigma}(x) = H_{i,\gamma}((x/\sigma)^\beta)$, $i = 1, 2, 3$. This relationship makes the theory of p-max df applicable to log-Pareto df to some extent.

Limiting Distributions of Exceedances

In the subsequent lines we present some unpublished material. We identify the limiting distributions of exceedances under power normalization in (2.23) as the class of p-POT stable df. We start with a technical lemma concerning F^{**} and F_{**}, cf. (2.26) and (2.27).

Lemma 2.7.3. *Let L be a non-degenerate limiting df in* (2.23) *for some df F. Then, for each point of continuity x in the interior of the support of L^{**} or L_{**}, respectively,*

(i) *there are functions $a(\cdot)$ and $b(\cdot) > 0$ such that*

$$\bar{F}^{**}(a(u) + b(u)x)/\bar{F}^{**}(u) \longrightarrow_{u\to\omega(F^{**})} \bar{L}^{**}(x)$$

if $\omega(F) > 0$, and

$$\bar{F}_{**}(a(u) + b(u)x)/\bar{F}_{**}(u) \longrightarrow_{u\to\omega(F_{**})} \bar{L}_{**}(x)$$

if $\omega(F) \le 0$.

(ii) *L^{**} and, respectively, L_{**} are POT-stable df.*

Proof. We outline the proof for both assertions merely for $\omega(F) > 0$. The case of $\omega(F) \le 0$ follows by similar arguments. Under (2.23) we first prove that the total mass of L is concentrated on the positive half-line and, therefore,

$$L(\exp(x)) = L^{**}(x), \quad x \in \mathbb{R}, \tag{2.31}$$

if $\omega(L) > 0$.

If $x < 0$, we have

$$F^{[u]}\left(\mathrm{sign}(x)\alpha(u)\,|x|^{\beta(u)}\right) \le F^{[u]}(0) \longrightarrow_{u\to\omega(F)} 0 \tag{2.32}$$

because $\omega(F) > 0$. This implies $L(x) = 0$ for all $x \le 0$.

Next consider $x > 0$. From (2.24) one gets

$$\bar{F}\left(\alpha(u)x^{\beta(u)}\right)/\bar{F}(u) \longrightarrow_{u\to\omega(F)} \bar{L}(x).$$

By straightforward computations,

$$\frac{\bar{F}\left(\exp\left(a(u) + b(u)x\right)\right)}{\bar{F}(\exp(u))} \longrightarrow_{\exp(u)\to\omega(F)} \bar{L}(\exp(x)) \tag{2.33}$$

for all $x \in \mathbb{R}$ with $a(u) = \log(\alpha(\exp(u))) \in \mathbb{R}$ and $b(u) = \beta(\exp(u)) > 0$. Therefore,

$$\frac{\bar{F}^{**}(a(u)+b(u)x)}{\bar{F}^{**}(u)} \longrightarrow_{u\to\omega(F^{**})} \bar{L}(\exp(x)) = \bar{L}^{**}(x), \tag{2.34}$$

and assertion (i) is verified. This also implies (ii) because limiting df under the linear normalization are necessarily POT-stable, cf. Theorem 2.7.1. □

Lemma 2.7.3 now offers the prerequisites to prove the the announced result concerning limiting df of exceedances under power-normalizations.

Theorem 2.7.4. *Every non-degenerate limiting df L in* (2.23) *is p-POT stable.*

Proof. Again, we merely prove the case $\omega(F) > 0$. From Lemma 2.7.3 (ii) we know that L^{**} is POT-stable. Thus, there are $a(u) \in \mathbb{R}$ and $b(u) > 0$ such that

$$\bar{L}^{**}(a(u)+b(u)x)/\bar{L}^{**}(u) = \bar{L}^{**}(x),$$

for each point of continuity u of L^{**} with $0 < L^{**}(u) < 1$ and $L^{**}(x) > 0$. This yields for $x, u > 0$,

$$\bar{L}^{**}(a(u)+b(u)\log(x))/\bar{L}^{**}(\log(u)) = \bar{L}^{**}(\log(x)).$$

Choosing $\alpha(u)$ and $\beta(u)$ as in the proof of Lemma 2.7.3 one gets from the equation $\bar{L}^{**}(a(u)+b(u)\log(x)) = \bar{L}^{**}\left(\log\left(\alpha(u)x^{\beta(u)}\right)\right)$ that

$$\bar{L}\left(\alpha(u)x^{\beta(u)}\right)/\bar{L}(u) = \bar{L}(x)$$

for all $x, u > 0$ with $0 < L^{**}(\log(u)) < 1$ and $L^{**}(\log(x)) > 0$. Notice that $L(x) = L^{**}(\log(x))$ if $x > 0$, and $L(x) = 0$ if $x \le 0$. This yields the p-POT stability of L according to the preceding equation. □

It is evident that the converse implication is also true, that is, every p-POT stable df L is a limiting df in (2.23) by choosing $F = L$. Summarizing the previous results we get that L is a limiting df of exceedances pertaining to a df F under power-normalization, if and only if, L^{**} (if $\omega(F) > 0$) or L_{**} (if $\omega(F) \le 0$) are POT-stable.

Domains of Attraction

Recall that within the linear framework, a df F belongs to the POT domain of attraction of a df W, denoted by $F \in \mathcal{D}_{\mathrm{POT}}(W)$, if there are functions $a(\cdot)$ and $b(\cdot) > 0$ such that

$$F^{[u]}(a(u)+b(u)x) \longrightarrow_{u\to\omega(F)} W(x). \tag{2.35}$$

Correspondingly, if relation (2.23) holds for df F and L, then F belongs to the p-POT domain of attraction of L, denoted by $F \in \mathcal{D}_{p\text{-}\mathrm{POT}}(L)$.

We characterize p-POT domains of attraction of a p-POT stable df L by means of POT domains of attraction of L^{**} or L_{**} which are POT-stable according to Theorem 2.7.4. As a direct consequence of Lemma 2.7.3(i) one gets Theorem 2.7.5.

Theorem 2.7.5. *For the p-POT domain of attraction $\mathcal{D}_{p\text{-}POT}(L)$ of a p-POT stable law L we have*

$$\mathcal{D}_{p\text{-}POT}(L) = \{F : F^{**} \in \mathcal{D}_{POT}(L^{**})\},$$

if $\omega(L) > 0$, and

$$\mathcal{D}_{p\text{-}POT}(L) = \{F : F_{**} \in \mathcal{D}_{POT}(L_{**})\}$$

if $\omega(L) \leq 0$.

P-POT domains of attraction of continuous p-POT stable laws can be deduced from p-max domains of attractions due to the identity of POT- and max-domains of attraction in the linear setup. The domains of attraction of the discrete p-POT stable laws have no counterpart in the framework of max-stable df. Their domains of attraction can be derived from the above theorem and Section 3 of Balkema and de Haan [22].

In the framework of super-heavy tail analysis we are merely interested in the super-heavy tailed p-POT stable laws, thus log-Pareto df. We also make use of a parametrization of log-Pareto df which is different from that in (2.28). Let

$$\tilde{L}_\gamma(x) = 1 - (\log(x))^{-1/\gamma}, \quad \gamma > 0,\, x \geq e. \tag{2.36}$$

It is apparent that $\tilde{L}_\gamma$ is a p-type of L_γ. Such df can be regarded as prototypes of p-POT stable df with slowly varying tails.

Corollary 2.7.6. *We have $F \in \mathcal{D}_{p\text{-}POT}(\tilde{L}_\gamma)$ if, and only if, there is a slowly varying function U and some $c > 0$ such that*

$$F(x) = 1 - (\log(x))^{-1/\gamma} U(\log(x)), \quad x > c. \tag{2.37}$$

Proof. This is a direct consequence of Theorem 2.7.5. We have for $x > 0$ that $\bar{F}(x) = \bar{F}(0)F^{**}(\log(x))$ for the df F^{**} which is in the POT domain of attraction of a Pareto df and, therefore, $\bar{F}^{**}$ is regularly varying at infinity. □

The p-POT domain of attraction of a log-Pareto df $\tilde{L}_\gamma$ can as well be characterized by a property with is deduced from regular variation, which characterizes the POT domain of attraction of Pareto df under linear transformation. Observe that

$$\bar{\tilde{L}}_\gamma\left(x^{\log(u)}\right) / \bar{\tilde{L}}_\gamma(u) = (\log(x))^{-1/\gamma},$$

which is the p-POT stability of $\tilde{L}_\gamma$. For the domain of attraction this relations holds in the limit and, furthermore, this yields a characterization of the domain attraction.

Corollary 2.7.7. *We have $F \in \mathcal{D}_{p\text{-POT}}(\tilde{L}_\gamma)$ if, and only if,*

$$\bar{F}\left(x^{\log(u)}\right)/\bar{F}(u) \longrightarrow_{u\to\infty} (\log(x))^{-1/\gamma}, \quad x > 1. \tag{2.38}$$

Proof. If $F \in \mathcal{D}_{p\text{-POT}}(\tilde{L}_\gamma)$ we have

$$\bar{F}(x) = (\log(x))^{-1/\gamma} U(\log(x)), \quad x > c$$

for some slowly varying function U and some $c > 0$. Therefore,

$$\begin{aligned}\frac{\bar{F}\left(x^{\log(u)}\right)}{\bar{F}(u)} &= \frac{\left(\log\left(x^{\log(u)}\right)\right)^{-1/\gamma} U\left(\log\left(x^{\log(u)}\right)\right)}{(\log(u))^{-1/\gamma} U(\log(u))} \\ &= (\log(x))^{-1/\gamma} \frac{U(\log(u)\log(x))}{U(\log(u))} \\ &\to (\log(x))^{-1/\gamma} \qquad \text{for } u \to \infty.\end{aligned}$$

Conversely, let

$$\lim_{u\to\infty} \bar{F}(x^{\log(u)})/\bar{F}(u) = (\log(x))^{-1/\gamma}$$

for $x > 1$. It follows that

$$\lim_{u\to\infty} \bar{F}(\exp(uy))/\bar{F}(\exp(u)) = y^{-1/\gamma}.$$

for all $y > 0$. Thus, $F^{**} \in \mathcal{D}_{\text{POT}}(W_\gamma)$ and, consequently, $F \in \mathcal{D}_{p\text{-POT}}\left(\tilde{L}_\gamma\right)$. □

We include a result about the invariance of $\mathcal{D}_{p\text{-POT}}\left(\tilde{L}_\gamma\right)$ under shift and power transformations.

Corollary 2.7.8. *The following equivalences hold true for $\mu \in \mathbb{R}$ and γ, β, $\sigma > 0$:*

$$F(\cdot) \in \mathcal{D}_{p\text{-POT}}\left(\tilde{L}_\gamma\right) \Leftrightarrow F((\cdot - \mu)) \in \mathcal{D}_{p\text{-POT}}\left(\tilde{L}_\gamma\right),$$

and

$$F(\cdot) \in \mathcal{D}_{p\text{-POT}}\left(\tilde{L}_\gamma\right) \Leftrightarrow F\left(\sigma(\cdot)^\beta\right) \in \mathcal{D}_{p\text{-POT}}\left(\tilde{L}_\gamma\right).$$

Proof. We only prove the first assertion because the second one is straightforward. Putting $F_\mu(x) = F(x - \mu)$ for $F \in \mathcal{D}_{p\text{-POT}}\left(\tilde{L}_\gamma\right)$, we get

$$\begin{aligned}\frac{1 - (F_\mu)^*(tx)}{1 - (F_\mu)^*(t)} &= \frac{\bar{F}(\exp(tx) - \mu)}{\bar{F}(\exp(t) - \mu)} \\ &= \frac{\bar{F}\left(\frac{\exp(tx)-\mu}{\exp(tx)}\exp(tx)\right)}{\bar{F}\left(\frac{\exp(t)-\mu}{\exp(t)}\exp(t)\right)}\end{aligned}$$

$$= \frac{\bar{F}\left(\exp\left(tx + \log\left(\frac{\exp(tx)-\mu}{\exp(tx)}\right)\right)\right)}{\bar{F}\left(\exp\left(t + \log\left(\frac{\exp(t)-\mu}{\exp(t)}\right)\right)\right)}$$
$$= \frac{F^{**}\left(tx + a_t\right)}{F^{**}(t + b_t)}$$

with

$$a_t = \log\left(\frac{\exp(tx) - \mu}{\exp(tx)}\right) \quad \text{and} \quad b_t = \log\left(\frac{\exp(t) - \mu}{\exp(t)}\right).$$

Obviously $a_t \to 0$ and $b_t \to 0$, hence using uniform convergence

$$\frac{\bar{F}^{**}\left(tx + a_t\right)}{\bar{F}^{**}(t + b_t)} \longrightarrow_{t\to\infty} x^{-1/\gamma}$$

and, thus, $F(\cdot - \mu) \in \mathcal{D}_{p\text{-POT}}$. $\square$

The previous result yields that

$$\mathcal{D}_{p\text{-POT}}(L) = \mathcal{D}_{p\text{-POT}}(\tilde{L}_\gamma) \tag{2.39}$$

for all p-types L of $\tilde{L}_\gamma$. It is easily seen that this result is valid for a p-POT domain of attraction of an arbitrary p-POT stable law. The result concerning location shifts cannot be extended to p-POT stable laws with finite right endpoints.

Mixtures of Regularly Varying DFs

We also deal with super-heavy tailed df given as mixtures of regularly varying df. We start with a result in Reiss and Thomas [389] concerning a relation of log-Pareto df,

$$\tilde{L}_\gamma(x) = 1 - (\log(x))^{-1/\gamma}, \quad x > e,\, \gamma > 0$$

and Pareto df,

$$\widetilde{W}_{\gamma,\sigma}(x) = 1 - (x/\sigma)^{-1/\gamma}, \quad x > \sigma,\, \gamma > 0.$$

Log-Pareto df can be represented as mixtures of certain Pareto df with respect to gamma densities. We have

$$\tilde{L}_\gamma(x) = \int_0^\infty \widetilde{W}_{1/z,e}(x) h_{1/\gamma}(z) dz \tag{2.40}$$

where h_α is the gamma density

$$h_\alpha(x) = \frac{1}{\Gamma(\alpha)} \exp(-x) x^{\alpha-1}. \tag{2.41}$$

We prove that this result can be extended to df in the domains of attraction of log-Pareto and Pareto df under power and, respectively, linear normalization. Assertion (ii) of the subsequent theorem is a modification and extension of Lemma 1 in Meerschaert and Scheffler [326], cf. also Cormann [77].

Theorem 2.7.9. *The following properties hold for the p-POT domain of attraction of a log-Pareto df* $\tilde{L}_\gamma$*:*

(i) *Let* $F \in \mathcal{D}_{p\text{-}POT}(\tilde{L}_\gamma)$ *for some* $\gamma > 0$*. Then there is a family of df* G_z*, with* $G_z \in \mathcal{D}_{POT}(\widetilde{W}_{1/z})$*, such that*

$$F(x) = \int_0^\infty G_z(x)p(z)dz,$$

where p *is a density which is ultimately monotone (monotone on* $[x_0, \infty)$ *for some* $x_0 > 0$*) and regularly varying at zero with index* $1/\gamma - 1$*.*

(ii) *Let* G_z *be a family of df with* $G_z \in \mathcal{D}(W_{1/z})$ *with representation*

$$G_z(x) = 1 - x^{-z}U(\log(x)), \quad x > a_1,$$

for some slowly varying function U *and some* $a_1 > 0$*. Then the mixture*

$$F(x) := \int_0^\infty G_z(x)p(z)dz,$$

where p *is a density as in* (i)*, has the representation*

$$F(x) = 1 - (\log(x))^{-1/\gamma} V(\log(x)), \quad x > a_2,$$

for some slowly varying function V *and some* $a_2 > 0$ *and, thus,* $F \in \mathcal{D}_{p\text{-}POT}(\tilde{L}_\gamma)$*.*

Proof. To prove (i) observe that the gamma density $h_{1/\gamma}$ in (2.41) satisfies the conditions imposed on p. Therefore, (i) is a direct consequence of (2.40) and Corollary 2.7.6. Therefore the statement is still true with p replaced by $h_{1/\gamma}$.

To prove (ii) notice that

$$1 - F(x) = \int_0^\infty e^{-z\log(x)}p(z)dzU(\log(x)).$$

The integral is now a function $\hat{p}(\log(\cdot))$ where $\hat{p}$ denotes the Laplace transform of p. Since p is assumed to be ultimately monotone and regularly varying at zero with index $1/\gamma - 1$ one can apply Theorem 4 on page 446 of Feller [156] getting

$$\int_0^\infty e^{-z\log(x)}p(z)dz = \log(x)^{-1/\gamma}\widetilde{V}(\log(x)), \quad x > a_3,$$

for some slowly varying function $\widetilde{V}$ and $a_3 > 0$. Now $V(x) := U(x)\widetilde{V}(x)$ is again slowly varying which completes the proof. $\square$

Testing for Super-Heavy Tails

In the subsequent lines the focus will be on statistical inference for distributions in Class B, namely on testing procedures for detecting the presence of a df F with Π-varying tail (2.21) underlying the sampled data. The main concern is to discriminate between a super-heavy tailed distribution and a distribution with a regularly varying tail. Since the non-negative parameter α is regarded as a gauge of tail heaviness, it can well serve the purpose of providing a straightforward distinction between super-heavy ($\alpha = 0$) and heavy tails ($\alpha > 0$). Moreover, note that if X is a rv with absolutely continuous df F in the Fréchet domain of attraction, i.e., satisfying (2.20), then e^X has a df H such that (2.21) holds. This is verified by the following proposition.

Proposition 2.7.10. *Let X be a rv with df F such that* (2.20) *holds and denote by $f := F'$ the corresponding density function. Define $Z := e^X$ with the df H. Then* (2.21) *holds with auxiliary function $a(t) := f(\log t)$, i.e., $H \in \Pi(a)$.*

Proof. The df of rv Z is related to the df of rv X through

$$H(x) = F(\log(x)) = (F \circ \log)(x).$$

Now notice that f is regularly varying with index $-\alpha-1 > -1$. Following the steps in the proof of Proposition B.2.15 (1) of de Haan and Ferreira [190], the following statements hold for the composition $F \circ \log$, since $\log \in \Pi$ and $F \in RV_{-\alpha}$:
for some $\theta = \theta(x,t) \in (0,1)$,

$$\begin{aligned}\frac{H(tx) - H(t)}{f(\log(t))} &= \frac{F(\log(tx)) - F(\log(t))}{f(\log(t))}\\ &= (\log(tx) - \log(t))\,\frac{f\left(\log(t) + \theta\{\log(tx) - \log(t)\}\right)}{f(\log(t))}\\ &= (\log(x))\,\frac{f\left(\log(t) + \theta\log(x)\right)}{f(\log(t))}\\ &= (\log(x))\,\frac{f\left(\log(t)\{1 + \theta\frac{\log(x)}{\log(t)}\}\right)}{f(\log(t))}\\ &\to_{t\to\infty} \log(x)\end{aligned}$$

by uniform convergence. □

Although the transformation via exponentiation projects a Pareto tailed distribution (2.20) into (2.21) as stated in Proposition 2.7.10, it is also possible to obtain a super-heavy tailed distribution in the sense of (2.21) via a similar transformation upon exponentially tailed distributions, i.e., departing from a df in the Gumbel max-domain of attraction. This is illustrated in Example 2.7.11 where a log-Weibull(β), $\beta \in (0,1)$, distribution is considered.

For the purpose of statistical inference, let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of rvs with common df F and let $X_{1,n} \le X_{2,n} \le \cdots \le X_{n,n}$ be their ascending order statistics. Furthermore, assume that F is a continuous and strictly increasing function.

In this context, in Fraga Alves et al. [159] and Neves and Fraga Alves [352] two test statistics have been developed to distinguish between heavy and super-heavy tailed probability distributions, i.e., for testing

$$H_0 : \alpha = 0 \ \text{[super-heavy]} \qquad vs. \qquad H_1 : \alpha > 0 \ \text{[heavy]} \tag{2.42}$$

in the framework carried out by the Class B of distribution functions (see equations (2.21) and (2.20)).

TEST 1. In Fraga Alves et al. [159], the asymptotic normality of the proposed statistic for testing (2.42) is proven under suitable and reasonable conditions. In particular, we need to require second-order refinements of (2.20) and (2.21) in order to specify the inherent rate of convergence. Hence, suppose there exists a positive or negative function A with $A(t) \to_{t\to\infty} 0$ and a second-order parameter $\rho \le 0$ such that

$$\lim_{t\to\infty} \frac{\frac{F(tx)-F(t)}{a(t)} - \frac{1-x^{-\alpha}}{\alpha}}{A(t)} = \frac{1}{\rho}\left(\frac{x^{-\alpha+\rho}-1}{-\alpha+\rho} - \frac{1-x^{-\alpha}}{\alpha}\right) =: H_{\alpha,\rho}(x), \tag{2.43}$$

for all $x > 0$ and some $\alpha \ge 0$. Appendix B of de Haan and Ferreira [190] offers a thorough catalog of second-order conditions, where it is also shown that necessarily $|A(t)| \in RV_\rho$.

Example 2.7.11 (log-Weibull distribution). *Let W be a random variable with min-stable Weibull(β) df, for $0 < \beta < 1$,*

$$F_W(x) = 1 - \exp(-x^\beta), \qquad x \ge 0.$$

Then the rv $X := e^W$ is log-Weibull distributed with df

$$F(x) = 1 - \exp(-(\log(x))^\beta), \qquad x \ge 1,\ 0 < \beta < 1.$$

From Taylor expansion of $F(tx) - F(t)$ one concludes that condition (2.43) holds with $\alpha = \rho = 0$, auxiliary functions

$$a(t) = \beta(\log(t))^{\beta-1}\exp(-(\log(t))^\beta)$$

and $A(t) = (\beta-1)/\log t$, $0 < \beta < 1$. Hence F belongs to the subclass defined by condition (2.43) *is, consequently, in Class B. However, F is not in Class A since $\log X$ has df F_W with an exponentially decaying tail.*

The test statistic introduced in Fraga Alves et al. [159] for discerning between super-heavy and heavy tailed distributions, as postulated in (2.42), is closely related to a new estimator for $\alpha \geq 0$. Both estimator and testing procedure evolve from the limiting relation below (with $j > 0$) which follows in turn from condition (2.20):

$$\lim_{t\to\infty} \int_1^\infty \frac{F(tx) - F(t)}{a(t)} \frac{dx}{x^{j+1}} = \int_1^\infty \frac{1 - x^{-\alpha}}{\alpha} \frac{dx}{x^{j+1}} = \frac{1}{j\,(j+\alpha)}.$$

The above equation entails that

$$\lim_{t\to\infty} \frac{\int_1^\infty \big(F(tx) - F(t)\big)\, dx/x^3}{\int_1^\infty \big(F(tx) - F(t)\big)\, dx/x^2} = \frac{1+\alpha}{2\,(2+\alpha)} \tag{2.44}$$

for $0 \leq \alpha < \infty$. Equation (2.44) can, furthermore, be rephrased as

$$\frac{\int_t^\infty \big(t/u\big)^2\, dF(u)}{\int_t^\infty (t/u)\, dF(u)} \to_{t\to\infty} \frac{1+\alpha}{2+\alpha} =: \psi(\alpha). \tag{2.45}$$

Replacing F by the empirical df F_n and t by the intermediate order statistic $X_{n-k,n}$, with $k = k_n$ a sequence of intermediate integers such that

$$k = k_n \to \infty, \quad k/n \to 0 \quad \text{as} \;\; n \to \infty, \tag{2.46}$$

the left-hand side of (2.45) becomes $\hat{\psi}_n(k)$, defined as

$$\hat{\psi}_n(k) := \frac{\sum_{i=0}^{k-1} \big(X_{n-k,n}/X_{n-i,n}\big)^2}{\sum_{i=0}^{k-1} X_{n-k,n}/X_{n-i,n}}. \tag{2.47}$$

On the other hand, the limiting function $\psi(\alpha)$ in (2.45) is a monotone continuous function. Hence, by simple inversion, we obtain the following estimator of $\alpha \geq 0$:

$$\hat{\alpha}_n(k) := \frac{2 \sum_{i=0}^{k-1} \big(X_{n-k,n}/X_{n-i,n}\big)^2 - \sum_{i=0}^{k-1} (X_{n-k,n}/X_{n-i,n})}{\sum_{i=0}^{k-1} (X_{n-k,n}/X_{n-i,n}) - \sum_{i=0}^{k-1} \big(X_{n-k,n}/X_{n-i,n}\big)^2}. \tag{2.48}$$

In the next theorem we establish without proof the asymptotic normality of the statistic $\hat{\psi}_n(k)$ introduced in (2.45), albeit under a mild second-order condition involving the intermediate sequence $k = k_n$. The result is akin to Theorem 2.4 in Fraga Alves et al. [159].

Theorem 2.7.12. *Let $k = k_n$ be a sequence of intermediate integers as in* (2.46) *and such that*

$$\left(n/\sqrt{k}\right) a(U(n/k)) \to \infty \tag{2.49}$$

as $n \to \infty$, where the function a is given in (2.43) *and U denotes the generalized inverse $U(t) := \left(\frac{1}{1-F}\right)^{\leftarrow}(t) = \inf\left\{x : F(x) \geq 1 - \frac{1}{t}\right\}$, for $t > 1$. If the second-order condition* (2.43) *holds with $\alpha \geq 0$ and*

$$(na(U(n/k)))^{1/2} A\left(U(n/k)\right) \to_{n\to\infty} \lambda \in \mathbb{R}, \tag{2.50}$$

then

$$\left(\sum_{i=0}^{k-1} \frac{X_{n-k:n}}{X_{n-i:n}}\right)^{1/2} \left(\hat{\psi}_n(k) - \frac{1+\alpha}{2+\alpha}\right) \to_D N(b, \sigma^2) \tag{2.51}$$

as $n \to \infty$, where

$$b := \frac{-\lambda\sqrt{1+\alpha}}{(2+\alpha)(1+\alpha-\rho)(2+\alpha-\rho)},$$
$$\sigma^2 := \frac{(1+\alpha)(4+3\alpha+\alpha^2)}{(2+\alpha)^3(3+\alpha)(4+\alpha)}.$$

An alternative formulation of (2.51) *is*

$$\left(n\, a\big(U(n/k)\big)\right)^{1/2}\left(\hat{\psi}_n(k) - \frac{1+\alpha}{2+\alpha}\right) \to_D N(b^\star, \sigma^{\star 2}),$$

as $n \to \infty$, where

$$b^\star := \frac{-\lambda\,(1+\alpha)}{(2+\alpha)(1+\alpha-\rho)(2+\alpha-\rho)},$$
$$\sigma^{\star 2} := \frac{(1+\alpha)^2(4+3\alpha+\alpha^2)}{(2+\alpha)^3(3+\alpha)(4+\alpha)}.$$

Theorem 2.7.12 has just provided a way to assess the presence of an underlying super-heavy tailed distribution. Taking k upper-order statistics from a sample of size n such that k accounts only for a small top sample fraction, in order to attain (2.46), we now define the test statistic

$$S_n(k) := \sqrt{24}\left(\sum_{i=0}^{k-1} \frac{X_{n-k,n}}{X_{n-i,n}}\right)^{1/2}\left(\hat{\psi}_n(k) - \frac{1}{2}\right). \tag{2.52}$$

The critical region for the one-sided test (2.42) at the nominal size $\bar{\alpha}$ is given by

$$\mathcal{R} := \left\{S_n(k) > z_{1-\bar{\alpha}}\right\},$$

where z_ε denotes the ε-quantile of the standard normal distribution.

It is worthwhile to mention that our null hypothesis is not only that the distribution F is in Class B defined in (2.21), but also F satisfies the second-order condition (2.43). Moreover, we should perform the test with a sequence k_n such that (2.50) holds with $\lambda = 0$. Condition (2.50) imposes an upper bound on the sequence k_n. For $\alpha = \rho = 0$, it seems difficult to prove that conditions (2.49) and (2.50) are never contradictory. However, if we replace (2.50) by the somewhat stricter condition $\sqrt{k_n}\, A(U(n/k_n)) \to_{n\to\infty} \lambda_1 \in \mathbb{R}$, we never hinder (2.49) from being valid. So, for any $\alpha \geq 0$, if $\sqrt{k_n}\, A(U(n/k_n)) \to_{n\to\infty} \lambda_1$ holds, then (2.50) holds with $\lambda = \sqrt{\alpha}\, \lambda_1$. The estimator of $\alpha \geq 0$ introduced in (2.48) is regarded as a way of testing for super-heavy tails. As an estimator for $\alpha > 0$ only, the present one is not really competitive.

TEST 2. The second proposal for testing (2.42) comes from Neves and Fraga Alves [352]; therein the test statistic $T_n(k)$, consisting of the ratio of maximum to the sum of log-excesses:

$$T_n(k) := \frac{\log(X_{n,n}) - \log(X_{n-k,n})}{\frac{1}{\log(k)} \sum_{i=0}^{k-1} \big(\log(X_{n-i,n}) - \log(X_{n-k,n})\big)} \tag{2.53}$$

proves to attain a standard Fréchet limit, as long as $k = k_n$ remains an intermediate sequence, under the simple null-hypothesis of condition (2.21) being fulfilled.

Theorem 2.7.13 below encloses a general result for heavy and super-heavy distributions belonging to the Class B (see (2.20) and (2.21)) thus suggesting a possible normalization for the test statistic $T_n(k)$ to attain a non-degenerate limit as n goes to infinity. Furthermore, results (i) and (ii) of Corollary 2.7.14 expound eventual differences in the stochastic behavior between super-heavy and heavy tailed distributions, accounting for power and consistency of the test, respectively.

First note that an equivalent characterization of the Class B can be formulated in terms of the tail quantile-type function U:

$$\lim_{t\to\infty} \frac{U\big(t + x\, q(t)\big)}{U(t)} = (1 + \alpha\, x)^{1/\alpha} \tag{2.54}$$

for all $1 + \alpha x > 0$, $\alpha \geq 0$, with a positive measurable function q such that

$$\lim_{t\to\infty} \frac{q\big(t + x\, q(t)\big)}{q(t)} = 1 + \alpha x \tag{2.55}$$

(cf. Lemma 2.7.15 below). This function q is called an auxiliary function for U.

If $\alpha = 0$, the right-hand side of (2.54) should be understood in the limiting sense as e^x while q becomes a self-neglecting function. This corresponds to an equivalent characterization of Class B as defined by (2.21). According to de Haan [184], Definition 1.5.1, we then say that the tail quantile function U belongs to the class Γ of functions of rapid variation (notation: $U \in \Gamma$).

Theorem 2.7.13. *Suppose the function U is such that condition* (2.54) *holds for some $\alpha \geq 0$. Let $k = k_n$ be a sequence of intermediate integers as in* (2.46)*. Then*

$$T_n(k) = O_p\left(\frac{1}{\log(k)}\right),$$

with $T_n(k)$ as defined in (2.53).

Corollary 2.7.14. *Under the conditions of Theorem* 2.7.13*,*

(i) *if $\alpha = 0$,*

$$\log(k)T_n(k) \to_D T^*, \tag{2.56}$$

where the limiting random variable T^* has a Fréchet df $\Phi(x) = \exp(-x^{-1})$, $x \geq 0$;

(ii) *if $\alpha > 0$,*

$$\log(k)T_n(k) \to_P 0. \tag{2.57}$$

Corollary 2.7.14 suffices to determine the critical region for assessing an underlying super-heavy tailed distribution. Considering the k upper-order statistics from a sample of size n such that k satisfies (2.46), we obtain the critical region for the one-sided test (2.42) at a nominal size $\bar{\alpha}$:

$$\mathcal{R} := \left\{\log(k)T_n(k) < \Phi^{-1}(\bar{\alpha})\right\},$$

where Φ^{-1} denotes the inverse of the standard Fréchet df Φ.

For the proof of Theorem 2.7.13 two auxiliary results are needed.

Lemma 2.7.15. *Suppose the function U is such that relation* (2.54) *holds with some $\alpha \geq 0$. Then, the auxiliary function q satisfies*

$$\lim_{t\to\infty} \frac{q(t)}{t} = \alpha \tag{2.58}$$

and

- *if $\alpha > 0$, then $U(\infty) := \lim_{t\to\infty} U(t) = \infty$ and U is of regular variation near infinity with index $1/\alpha$, i.e., $U \in RV_{1/\alpha}$;*
- *if $\alpha = 0$, then $U(\infty) = \infty$ and U is ∞-varying at infinity.*

Furthermore, for $\alpha = 0$,

$$\lim_{t\to\infty}\Big(\log(U(t + xq(t))) - \log(U(t))\Big) = x \quad \text{for every } x \in \mathbb{R}. \tag{2.59}$$

Lemma 2.7.15 coupled with condition (2.54) imposes the limit (2.55) on the auxiliary function $q(t)$.

Proof. In case $\alpha > 0$, the first part of the lemma follows directly from (2.54), whereas in case $\alpha = 0$ it is ensured by Lemma 1.5.1 and Theorem 1.5.1 of de Haan [184]. Relation (2.59) follows immediately from (2.54) with respect to $\alpha = 0$. □

Proposition 2.7.16. *Suppose condition* (2.54) *holds for some* $\alpha \geq 0$.

(i) *If* $\alpha > 0$, *then for any* $\varepsilon > 0$ *there exists* $t_0 = t_0(\varepsilon)$ *such that for* $t \geq t_0$, $x \geq 0$,

$$(1-\varepsilon)\,(1+\alpha\,x)^{\frac{1}{\alpha}-\varepsilon} \leq \frac{U\big(t+x\,q(t)\big)}{U(t)} \leq (1+\varepsilon)\,(1+\alpha\,x)^{\frac{1}{\alpha}+\varepsilon}. \tag{2.60}$$

(ii) *If* (2.54) *holds with* $\alpha = 0$ *then, for any* $\varepsilon > 0$, *there exists* $t_0 = t_0(\varepsilon)$ *such that for* $t \geq t_0$, *for all* $x \in \mathbb{R}$,

$$\frac{U(t+x\,q(t))}{U(t)} \leq (1+\varepsilon)\,\exp\big(x(1+\varepsilon)\big). \tag{2.61}$$

Proof. Inequalities in (2.60) follow immediately from Proposition 1.7 in Geluk and de Haan [170] when we settle $q(t) = \alpha t$ (see also (2.58) in Lemma 2.7.15) while (2.61) was extracted from Beirlant and Teugels [34], p.153. □

Lemma 2.7.17.

(i) *If* $U \in RV_{1/\alpha}$, $\alpha > 0$, *then, for any* $\varepsilon > 0$, *there exists* $t_0 = t_0(\varepsilon)$ *such that for* $t \geq t_0$ *and* $x \geq 1$,

$$(1-\varepsilon)\frac{1}{\alpha}\log(x) \leq \log(U(tx)) - \log(U(t)) \leq (1+\varepsilon)\frac{1}{\alpha}\log(x). \tag{2.62}$$

(ii) *If* $U \in \Gamma$ *then, for any* $\varepsilon > 0$, *there exists* $t_0 = t_0(\varepsilon)$ *such that for* $t \geq t_0$ *and for all* $x \in \mathbb{R}$,

$$\log(U(t+xq(t))) - \log(U(t)) \leq \varepsilon + x(1+\varepsilon). \tag{2.63}$$

Proof. Notice that once we apply the logarithmic transformation to relation (2.60) for large enough t, it becomes

$$\begin{aligned}(1-\varepsilon)\log\left((1+\alpha x)^{1/\alpha}\right) &\leq \log(U(t+xq(t))) - \log(U(t)) \\ &\leq (1+\varepsilon)\log\left((1+\alpha x)^{1/\alpha}\right).\end{aligned}$$

As before, the precise result is obtained by taking $q(t) = \alpha t$ with the concomitant translation of (2.54) for $\alpha > 0$ into the regularly varying property of U (cf. Lemma 2.7.15 again). The proof for (2.63) is similar and therefore omitted. □

Proof of Theorem 2.7.13. Let $(Y_{i,n})_{i=1}^n$ be the order statistics corresponding to the iid rv $(Y_i)_{i=1}^n$ with standard Pareto df $1 - y^{-1}$, for all $y \geq 1$. Taking into account the equality in distribution

$$(X_{i,n})_{i=1}^n =_D (U(Y_{i,n}))_{i=1}^n\,, \tag{2.64}$$

and defining

$$Q_n^{(i)} := \frac{Y_{n-i,n} - Y_{n-k,n}}{q(Y_{n-k,n})}, \quad i = 0, 1, \ldots, k-1, \tag{2.65}$$

as well as

$$M_n^{(1)} := \frac{1}{k}\sum_{i=0}^{k-1} \log(U(Y_{n-i,n})) - \log(U(Y_{n-k,n})), \tag{2.66}$$

we get in turn

$$\begin{aligned} T_n(k) &=_D \frac{\log(U(Y_{n,n})) - \log(U(Y_{n-k,n}))}{k\, M_n^{(1)}} \qquad (2.67)\\ &= \frac{\log(U(Y_{n,n})) - \log(U(Y_{n-k,n}))}{\sum_{i=0}^{k-1}\big(\log(U(Y_{n-i,n})) - \log(U(Y_{n-k,n}))\big)} \\ &= \frac{\log\big(U\big(Y_{n-k,n} + Q_{k,n}^{(0)}\, q(Y_{n-k,n})\big)\big) - \log(U(Y_{n-k,n}))}{\sum_{i=0}^{k-1}\Big(\log\big(U\big(Y_{n-k,n} + Q_{k,n}^{(i)}\, q(Y_{n-k,n})\big)\big) - \log(U(Y_{n-k,n}))\Big)}. \qquad (2.68) \end{aligned}$$

Bearing on the fact that the almost sure convergence $Y_{n-k,n} \to \infty$ holds with an intermediate sequence $k = k_n$ (cf. Embrechts *et al.* [122], Proposition 4.1.14), we can henceforth make use of condition (2.54). For ease of exposition, we shall consider the cases $\alpha > 0$ and $\alpha = 0$ separately.
Case $\alpha > 0$: As announced, the core of this part of the proof lies at relation (2.54). Added (2.62) from Lemma 2.7.17, we obtain the following inequality for any $\varepsilon > 0$ and sufficiently large n:

$$\begin{aligned} M_n^{(1)} &= \frac{1}{k}\sum_{i=0}^{k-1} \log\left(U\Big(\frac{Y_{n-i,n}}{Y_{n-k,n}}\, Y_{n-k,n}\Big)\right) - \log(U(Y_{n-k,n})) \\ &\leq (1+\varepsilon)\frac{1}{k}\sum_{i=0}^{k-1}\frac{1}{\alpha}\big(\log(Y_{n-i,n}) - \log(Y_{n-k,n})\big). \end{aligned}$$

Owing to Rényi's important representation for exponential spacings,

$$E_{k-i,k} =_D E_{n-i,n} - E_{n-k,n} \;=\; \log(Y_{n-i,n}) - \log(Y_{n-k,n}), \tag{2.69}$$

where $E_{n-i,n}$, $i = 0, 1, \ldots, k-1$, are the order statistics pertaining to independent standard exponential rv $E_i = \log(Y_i)$, we thus obtain

$$M_n^{(1)} = \frac{1}{k}\sum_{i=0}^{k-1} \log(U(Y_{n-i,n})) - \log(U(Y_{n-k,n}))$$
$$\leq \frac{1}{\alpha}\,(1+\varepsilon)\frac{1}{k}\sum_{i=0}^{k-1} \log(Y_{k-i,k}). \tag{2.70}$$

We can also establish a similar lower bound. The law of large numbers ensures the convergence in probability of the term on the right-hand side of (2.70) since, for an intermediate sequence $k = k_n$, as $n \to \infty$,

$$\frac{1}{k}\sum_{i=0}^{k-1} \log(Y_i) \to_P \int_1^\infty \frac{\log(y)}{y^2}\,dy = 1.$$

In conjunction with (2.58), the latter entails

$$L_n(k) := \frac{q(Y_{n-k,n})}{Y_{n-k,n}}\,M_n^{(1)} = 1 + o_p(1) \tag{2.71}$$

as $n \to \infty$. Hence, using (2.62) followed by (2.69) upon (2.67), we obtain, as $n \to \infty$,

$$T_n(k) =_D \frac{1}{k}\frac{q(Y_{n-k,n})}{Y_{n-k,n}}\frac{\log(U(Y_{n,n})) - \log(U(Y_{n-k,n}))}{L_n(k)}$$
$$= \frac{1}{k}(E_{k,k} - \log(k))\big(1 + o_p(1)\big) + \frac{\log(k)}{k}\big(1 + o_p(1)\big). \tag{2.72}$$

Finally, by noting that $E_{k,k} - \log(k) \to_D \Lambda$, as $k \to \infty$, where Λ is denoting a Gumbel rv, we obtain a slightly stronger result than the one stated in the present theorem. More specifically, we get from (2.72) that $T_n(k) = o_p(k^{-1/2})$, for any intermediate sequence $k = k_n$.
Case $\alpha = 0$: The proof concerning this case of super-heavy tailed distributions, virtually mimics the steps followed in the heavy tailed case ($\alpha > 0$). We get from (2.68) that $M_n^{(1)}$ as defined in (2.66) can be written as

$$M_n^{(1)} = \frac{1}{k}\sum_{i=0}^{k-1} \log\Big(U\big(Y_{n-k,n} + Q_{k,n}^{(i)}\,q(Y_{n-k,n})\big)\Big) - \log(U(Y_{n-k,n})).$$

Giving heed to the fact that, for each $i = 0, 1, \ldots, k-1$,

$$Q_{k,n}^{(i)} = \frac{Y_{n-k,n}}{q(Y_{n-k,n})}\Big(\frac{Y_{n-i,n}}{Y_{n-k,n}} - 1\Big)$$

is stochastically bounded away from zero (see Lemma 2.7.15), we can thus apply relation (2.63) from Lemma 2.7.17 in order to obtain, for any intermediate sequence $k = k_n$,

$$\frac{1}{k}\sum_{i=0}^{k-1} \log\Big(U\big(Y_{n-k,n} + Q_{k,n}^{(i)} q(Y_{n-k,n})\big)\Big) - \log(U(Y_{n-k,n})) \leq (1+\varepsilon)\,\frac{1}{k}\sum_{i=0}^{k-1} Q_{k,n}^{(i)},$$

as $n \to \infty$. Using Rényi's representation (2.69), we get

$$\frac{q(Y_{n-k,n})}{Y_{n-k,n}}\, M_n^{(1)} \leq (1+\varepsilon)\,\frac{1}{k}\sum_{i=0}^{k-1}\big(Y_{k-i,k} - 1\big). \tag{2.73}$$

It is worth noticing at this point that with constants $a_k^* > 0$, $b_k^* \in \mathbb{R}$ such that $a_k^* \sim k\pi/2$ and $b_k^* a_k^*/k \sim \log(k)$ as $k \to \infty$, this new random variable S_k^* defined by

$$S_k^* := \frac{1}{a_k^*}\sum_{i=0}^{k-1}\big(Y_i - 1\big) - b_k^*, \tag{2.74}$$

converges in distribution to a sum-stable law (cf. Geluk and de Haan [171]). Embedding S_k^* defined above in the right-hand side of (2.73), we ensure that $L_n(k)$ as introduced in (2.71) satisfies $L_n(k) = O_p(\log(k))$. Therefore, in view of (2.68), the proof is concluded by showing that it is possible to normalize the maximum of the log-spacings in such a way as to exhibit a non-degenerate behavior eventually. Since $U \in \Gamma$ we get in a similar way as before, for large enough n,

$$\begin{aligned}
&\frac{q(Y_{n-k,n})}{k\,Y_{n-k,n}}\Big(\log\big(U\big(Y_{n-k,n} + Q_{k,n}^{(0)} q(Y_{n-k,n})\big)\big) - \log(U(Y_{n-k,n}))\Big)\\
&= k^{-1}\Big(\frac{Y_{n,n}}{Y_{n-k,n}} - 1\Big)\big(1 + o_p(1)\big)\\
&= k^{-1}\big(Y_{k,k} - 1\big)\big(1 + o_p(1)\big) \;= O_p(1).
\end{aligned}$$

□

Proof of Corollary 2.7.14. *(i)* For $\alpha = 0$, the last part of the proof of Theorem 2.7.13 emphasizes that, as $n \to \infty$,

$$\begin{aligned}
\log(k)T_n(k) &=_D \frac{k^{-1}\big(\log(U(Y_{n,n})) - \log(U(Y_{n-k,n}))\big)}{L_n(k)/\log(k)}\,\frac{q(Y_{n-k,n})}{Y_{n-k,n}}\\
&= \big(T^* + o_p(1)\big)/\big(1 + o_p(1)\big)\\
&= T^*\big(1 + o_p(1)\big)
\end{aligned}$$

because, after suitable normalization by $a_k = k^{-1}$, the maximum of a sample of size k with standard Pareto parent distribution is attracted to a Fréchet law.
(ii) The precise result follows from (2.72) by straightforward calculations. □

Neves and Fraga Alves [352] present a finite (large) sample simulation study which seems to indicate that the conservative extent of Test 2 opens a window of opportunity for its applicability as a complement to Test 1. This is particularly true for the less heavy distributions lying in the class of super-heavy tails since in this case the number of wrong rejections is likely to rise high above the nominal level of the test based on (2.52). Moreover, the asymptotics pertaining to the test statistic $S_n(k)$ in (2.52) (cf. Theorem 2.7.12) require a second-order refinement of (2.21) (as in (2.43)), while the asymptotic behavior of the test statistic $T_n(k)$ only relies on the first-order conditions on the tail of F, meaning that we are actually testing $F \in$ Class B.

http://www.springer.com/978-3-0348-0008-2

Laws of Small Numbers: Extremes and Rare Events
Falk, M.; Hüsler, J.; Reiss, R.-D.
2011, XVI, 509 p., Softcover
ISBN: 978-3-0348-0008-2
A product of Birkhäuser Basel