Extremes

Robert L. Wolpert
Department of Statistical Science
Duke University, Durham, NC, USA

Vsn 13, 2014-05-20

1 Extreme Values

Most probability books do a fine job of covering the approximate probability distribution of *sums* (or averages) of independent random variables. If $\{X_j\}$ are independent and identically distributed (iid) with any distribution having a finite mean μ and variance σ^2, the sum and average

$$S_n := \sum_{j=1}^{n} \qquad \bar{X}_n := \frac{1}{n} S_n$$

are each asymptotically normally distributed in the sense that their standardized version

$$Z_n := \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

satisfies

$$\lim_{n\to\infty} \mathsf{P}[a < Z_n \le b] = \Phi(b) - \Phi(a)$$

uniformly in $-\infty < a < b < \infty$, where

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-z^2/2}\, dz$$

denotes the standard Normal cdf function. Some texts go further and discuss limits for sums of random variables X_j that do *not* have finite means or variances— in that case the α-Stable distribution emerges as another (in fact, the only other) possible limiting distribution for normalized sums of the form

$$\frac{S_n - b_n}{a_n}$$

for suitable non-random sequences $\{a_n\}$, $\{b_n\}$.

In light of recent concerns about economic crises and climate changes leading to catastrophes in storm and drought severity, temperature, hurricane intensity, and such, there is a new interest in looking not at the probability distributions of *averages* (like $\bar{X}_n$) but at those of *extremes*, like:

$$X_n^* := \max_{1\le j\le n} X_j.$$

The best tool for studying sums of iid random variables is the characteristic function $\chi(\omega) = \mathsf{E}e^{i\omega X_j}$, because the chf χ_n for the sum S_n has a simple expression: $\chi_n(\omega) = \chi(\omega)^n$. The best tool for studying *maxima* or minima of iid random variables is the CDF, for the same reason:

$$F_n(x) = \mathsf{P}[X_n^* \le x] = \mathsf{P}\left\{\cap_{i=1}^n [X_i \le x]\right\} = F(x)^n.$$

For $(X_n^* - b_n)/a_n$ to have a limiting distribution $G(z)$, we would need

$$\begin{aligned}
\mathsf{P}\left\{\frac{X_n^* - b_n}{a_n} \le z\right\} &= F_n(b_n + za_n)\\
&= F(b_n + za_n)^n\\
&= [1 - \bar{F}(b_n + za_n)]^n\\
&\to G(z).
\end{aligned}$$

We'll need $\bar{F}(b_n + za_n) \asymp \frac{1}{n}$, or $(b_n + za_n) \asymp F^{-1}\big(1 - \frac{1}{n}\big)$, so good starting places would be a_n or b_n to be about $F^{-1}(1 - \frac{1}{n})$. Let's look at examples.

1.1 Example 1: Exponential Distribution

Let $\{X_j\}$ have independent Exponential distributions $X_j \overset{\text{iid}}{\sim} \mathsf{Ex}(\lambda)$, and let X_n^* be the largest of the first n. Can we find non-random sequences $\{a_n\}$, $\{b_n\}$ and a limiting cdf $G(z)$ for which

$$\lim_{n\to\infty} \mathsf{P}\left[\frac{X_n^* - b_n}{a_n} \le z\right] = G(z)?$$

For any sequences $\{a_n\}$, $\{b_n\}$ the exact probabilities are

$$\begin{aligned}
\mathsf{P}\left[\frac{X_n^* - b_n}{a_n} \le z\right] &= \mathsf{P}[X_n^* \le a_n z + b_n]\\
&= \mathsf{P}\left\{\cap_{j=1}^n [X_j \le a_n z + b_n]\right\}\\
&= \{\mathsf{P}[X_1 \le a_n z + b_n]\}^n\\
&= \left\{1 - e^{-\lambda(a_n z + b_n)}\right\}^n
\end{aligned}$$

The goal is to find $\{a_n, b_n\}$ for which this converges as $n \to \infty$ to a DF. For this we need the term in braces be $1 - O(1/n)$, so we need $\log n - \lambda(a_n z + b_n)$ to converge. If we now choose $a_n := 1/\lambda$ and $b_n := (\log n)/\lambda$,

$$\begin{aligned}
\mathsf{P}\left[\frac{X_n^* - b_n}{a_n} \le z\right] &= \left\{1 - \frac{1}{n}e^{-z}\right\}^n\\
&\to G(z) := \exp\big(-e^{-z}\big),
\end{aligned} \tag{1}$$

the standard Gumbel Distribution. Its median is $m^* = -\log\log 2 \approx 0.366513$ (since $G(-\log\log 2) = \exp(-\log 2) = 1/2$) and its mean is $\mu^* = \gamma_e \approx 0.577216$, the Euler-Mascheroni constant, so the median m_n^* and mean μ_n^* for X_n^* are

$$m_n^* = \frac{\log n - \log\log 2}{\lambda} \qquad \mu_n^* = \frac{\log n + \gamma_e}{\lambda},$$

each growing with n at a logarithmic rate.

For example, if we imagine that sprinters' speed in m/s are given by the $\mathsf{Ex}(1)$ distribution, then the fastest speed of n independently-drawn sprinters would have approximately the re-scaled Gumbel Distribution with median $m_n^* = \log n - \log\log 2$; this has even odds of exceeding Usain Bolt's 2009 world-record 100m pace of 9.69s if

$$\begin{aligned}\log n - \log\log 2 &\ge \frac{100\text{m}}{9.69\text{s}}\\ &= 10.32\text{m/s}\\ \log n &\ge \log\log 2 + 10.32\\ n &\ge \exp(-0.37 + 10.32 = 9.95)\\ &= 21\,023.73,\end{aligned}$$

i.e., there's about an even chance that one of 21,024 independent $\mathsf{Ex}(1)$ random variables would exceed Bolt's pace.

For this example we can compute exactly the median for X_n^* or, if we prefer, the probability that X_n^* exceeds 9.95 for $n = 21024$; the latter, for example, is

$$\mathsf{P}[X_{21024}^* > 10.32] = \big[1 - \exp(-10.32)\big]^{21024} = 0.5000176,$$

so the approximation is quite good.

1.2 Example 2: Normal Distribution

Now let $\{X_j\}$ have independent standard Normal distributions $X_j \overset{\text{iid}}{\sim} \mathsf{No}(0,1)$, set $X_n^* := \max_{1\le j\le n} X_j$, and seek non-random $\{a_n\}$, $\{b_n\}$ and a limiting cdf $G(z)$ for $a_n^{-1}(X_n^* - b_n)$. First we need to note that, for $x > 0$,

$$\begin{aligned}\Phi(-x) &= \int_x^\infty \phi(z)\,dz\\ &\le \int_x^\infty \frac{z}{x}\phi(z)\,dz = \frac{1}{x\sqrt{2\pi}}\int_x^\infty z e^{-z^2/2}\,dz = \frac{1}{x}\phi(x);\end{aligned}$$

Gordon's Inequality improves this to the two-sided bound

$$1 \le \frac{\phi(x)}{x\Phi(-x)} \le 1 + \frac{1}{x^2}$$

for every $x > 0$. Now let $b_n := -\Phi^{-1}(1/n)$ be the $(1-1/n)$'th quantile (so $\Phi(-b_n) = 1/n$) and set $a_n := 1/b_n$; note that $b_n \asymp \sqrt{2\log n}$ grows as $n \to \infty$, while $a_n \to 0$. By Taylor's theorem and the evenness of $\phi(z)$, for fixed $z \in \mathbb{R}$,

$$\begin{aligned}\log\Phi(-a_n z - b_n) &= \log\Phi(-b_n) - a_n z\frac{\phi(-b_n)}{\Phi(-b_n)} + o(a_n z)\\ &= \log\frac{1}{n} - z\frac{\phi(b_n)}{b_n\,\Phi(-b_n)} + o(a_n z)\\ &= \log\frac{1}{n} - z + o(a_n z)\end{aligned}$$

so

$$\begin{aligned}
\mathsf{P}[X_1 \le a_n z + b_n] &= \Phi(a_n z + b_n) \\
&= 1 - \frac{1}{n} e^{-z + o(1/\sqrt{\log n})}, \text{ and} \\
\mathsf{P}[X_n^* \le a_n z + b_n] &\approx \left[1 - n^{-1} e^{-z}\right]^n \\
&\approx \exp(-e^{-z}) =: G(z),
\end{aligned}$$

again the Gumbel distribution. Similarly, if $\{X_i\} \overset{\text{iid}}{\sim} \mathsf{No}(\mu, \sigma^2)$ (now with arbitrary mean and variance) then we simply change the location and scale to find that with $b_n := \mu - \sigma \Phi^{-1}(1/n)$ and $a_n := -\sigma/\Phi^{-1}(1/n)$ we have

$$\mathsf{P}\left[\frac{X_n^* - b_n}{a_n} \le z\right] \to G(z) = e^{-e^{-z}},$$

with median

$$m_n^* = \mu - \sigma \Phi^{-1}(1/n) + (\log \log 2)\sigma/\Phi^{-1}(1/n)$$

growing like $\sigma\sqrt{2 \log n}$ as $n \to \infty$.

Typically unbounded distributions like the Exponential and Normal (as well as the Gamma, Log-normal, Weibull, *etc.*) whose tails fall off exponentially or faster will have this same Gumbel limiting distribution for the maxima, and will have medians (and other quantiles) that grow as $n \to \infty$ at the rate of (some power of) $\log n$.

1.3 Example 3: Pareto Distribution

Distributions with "fatter tails" (*i.e.*, those for which $\mathsf{P}[X > x]$ falls off no faster than a *power* of x) will have a different limit. For example, let $\{U_j\}$ be iid Uniform random variables and set $X_j = 1/U_j$; then X_j has the "unit Pareto distribution" determined by

$$\mathsf{P}[X_j > x] = 1/x, \qquad x \ge 1$$

and the maximum X_n^* of n iid unit Paretos will satisfy

$$\mathsf{P}[X_n^* \le a_n z + b_n] = \left(1 - [a_n z + b_n]^{-1}\right)^n \qquad\qquad a_n z + b_n \ge 1.$$

With $a_n := n$ and $b_n := 0$,

$$= \left(1 - \frac{1}{nz}\right)^n \to e^{-1/z} =: F(z), \qquad\qquad z > 0, \tag{2}$$

the "unit Fréchet Distribution". Similarly for $X_j = \epsilon U_j^{-1/\alpha}$ with the $\mathsf{Pa}(\alpha, \epsilon)$ distribution satisfying

$$\mathsf{P}[X_j > x] = \epsilon^\alpha / x^\alpha, \qquad x \ge \epsilon,$$

set $a_n := n^{1/\alpha}\epsilon$ and $b_n := 0$, then

$$\mathsf{P}[X_n^* \le a_n z + b_n] = \left(1 - \frac{1}{n} z^{-\alpha}\right)^n \to e^{-z^{-\alpha}} =: F(z \mid \alpha), \qquad z > 0,$$

the Fréchet distribution with shape parameter $\alpha > 0$. The Fréchet median is $(\log 2)^{-1/\alpha}$, so X_n^* has median

$$m_n^* = n^{1/\alpha}\epsilon(\log 2)^{-1/\alpha}$$

that grows like a power of n, while the mean is infinite. This limiting behavior is typical for heavy-tailed distributions such as the t, α-stable, and Pareto.

1.4 Example 4a: Beta Distribution Minimum

For $\alpha, \beta > 0$, the $1/\alpha$'th power of an exponential $\mathsf{Ex}(\beta)$ random variable has the Weibull $\mathsf{We}(\alpha, \beta)$ distribution, with Survival Function (SF) $\bar{F}(x) = \mathsf{P}[X > x] = \exp(-\beta x^\alpha)$ for $x \geq 0$. It follows that the *minimum* X_{*n} of n iid $\mathsf{We}(\alpha, \beta)$ random variables satisfies

$$\mathsf{P}[X_{*n} > x] = \left\{e^{-\beta x^\alpha}\right\}^n = e^{-n\beta x^\alpha},$$

again Weibull but now with the $X_* \sim \mathsf{We}(\alpha, n\beta)$ distribution. For $(X_{*n} - b_n)/a_n$ to have a limiting distribution we need

$$\mathsf{P}\left\{\frac{X_{*n} - b_n}{a_n} > z\right\} = e^{-n\beta(b_n + a_n z)^\alpha}$$

to converge as $n \to \infty$; evidently it will if $b_n := 0$ and $a_n := (n\beta)^{-1/\alpha}$:

$$= e^{-z^\alpha}, \qquad z > 0$$

the $\mathsf{We}(\alpha, 1)$. The minimum of n iid $\mathsf{Be}(\alpha, \beta)$ random variables has SF

$$\begin{aligned}\mathsf{P}\left\{X_{*n} > z\right\} &= \left(1 - c\int_0^z x^{\alpha-1}(1-x)^{\beta-1}\,dx\right)^n \\ &\approx (1 - (c/\alpha)z^\alpha)^n\end{aligned}$$

for $c = \Gamma(\alpha+\beta)/\Gamma(\alpha)\Gamma(\beta)$, so for convergence we will need $(b_n + a_n z)^\alpha \asymp 1/n$. The choice $b_n := 0$ and $a_n := (\alpha/nc)^{1/\alpha}$ leads again to the $\mathsf{We}(\alpha, 1)$ limiting distribution for the minimum.

1.4.1 Example 4b: Beta Distribution Maximum

Let $\{X_i\} \overset{\text{iid}}{\sim} \mathsf{Be}(\beta, \alpha)$ and set $Y_i := [1 - X_i]$. Then $\{Y_i\} \overset{\text{iid}}{\sim} \mathsf{Be}(\alpha, \beta)$ and $X_n^* = 1 - Y_{*n}$, so

$$\begin{aligned}\mathsf{P}\left\{\frac{X_n^* - b_n}{a_n} < z\right\} &= \mathsf{P}\left\{\frac{Y_{*n} - (1 - b_n)}{a_n} > -z\right\} \\ &\approx e^{-(-z)^\alpha}, \qquad z < 0\end{aligned}$$

for $b_n := 1$ and $a_n := (\alpha/nc)^{1/\alpha}$, with c as before, now for $z < 0$. This is called the *reversed* Weibull distribution, with cdf and pdf

$$\begin{aligned}G(z \mid \alpha) &= e^{-(-z)^\alpha} \qquad z < 0 \\ g(z \mid \alpha) &= (-z)^{\alpha-1}e^{-(-z)^\alpha}\mathbf{1}_{\{z<0\}},\end{aligned} \tag{3}$$

with median $m_n^* = -(n/\log 2)^{-1/\alpha}$ increasing to zero as $n \to \infty$.

Similarly the maximum X_n^* of n iid uniform random variables $X_j \sim \mathsf{Un}(L, R)$ on an arbitrary interval has limiting distribution:

$$\begin{aligned}
\mathsf{P}[a_n^{-1}[X_n^* - b_n] \le z] &= \mathsf{P}[X_n^* \le a_n z + b_n] \\
&= \left[1 - \frac{R - a_n z - b_n}{R - L}\right]^n && \text{if } L \le a_n z + b_n \le R \\
&= (1 + z/n)^n \to e^z && \text{if } -n \le z \le 0
\end{aligned}$$

for $a_n := (R - L)/n$ and $b_n := R$, the unit Reversed $\mathsf{We}(1)$ Weibull. Now the median for X_n^* is

$$m_n^* = R - (R - L)(\log 2)/n,$$

increasing at rate $1/n$ to an upper bound of R. The suitably standardized minimum and maximum of n independent $\mathsf{Be}(\alpha, \beta)$ random variables have asymptotic $\mathsf{We}(\alpha)$ and reverse $\mathsf{We}(\beta)$ distributions, respectively. These are typical of the maximal behavior for bounded random variables with continuous distributions.

1.5 The Three Types Theorem

Fisher and Tippett (1928) first proved that location-scale families of these three distributions— Gumbel (1), Fréchet (2), and reversed Weibull (3)— are the *only* possible limits for maxima of independent random variables. That is, if there exist nonrandom sequences $a_n > 0$ and $b_n \in \mathbb{R}$ and a nondegenerate distribution G such that the maximum $X_n^* := \max_{j \le n} X_j$ of iid random variables $\{X_j\}$ satisfies

$$\mathsf{P}\left\{\frac{X_n^* - b_n}{a_n} \le z\right\} \to G(z) \tag{4}$$

then G must be one of these three distributions: Gumbel, Fréchet, or reversed Weibull. Half a century later McFadden (1978) discovered that all three of these limiting distributions could be expressed in the same functional form as special cases of a single three-parameter "Generalized Extreme Value" (GEV) distribution, with cdf

$$G(x; \mu, \sigma, \xi) = \exp\left\{-\left[1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right]^{-1/\xi}\right\} \tag{5}$$

which reduces to the Fréchet with $\alpha = 1/\xi$ if $\xi > 0$, reversed Weibull with $\alpha = -1/\xi$ if $\xi < 0$, and Gumbel as $\xi \to 0$ (see Appendix A.5 on *p.* 17 for more details). In some ways I feel this was unfortunate, because now it is common for people to model and fit the GEV without thinking very clearly about the specific form of their data and distributions.

The key idea for the three-types theorem is to notice that any distribution G satisfying (4) must also have the property that for all n, the maximum of n independent random variables with cdf G must *also* (after suitable shift and scale changes) have cdf G— *i.e.*, that for any n there exist constants a_n and b_n such that for all $z \in \mathbb{R}$,

$$G(z)^n = G(a_n\, z + b_n).$$

It turns out that the only cdf that satisfies this equation is (5).

2 Threshold Exceedances

In this section we'll explore a different way of looking at the same limiting distributions of maxima, the "peaks over thresholds" or "PoT" approach.

As before let $\{X_j\}$ be iid for $1 \le j \le n$ and set[1] $T_j := \frac{j-1/2}{n} \in (0,1)$. Let a_n and b_n be real numbers and set $Y_j := a_n X_j + b_n$. The vector $N(R_i)$ of the numbers of points (T_j, Y_j) in disjoint rectangles $R_i := (s_i, t_i] \times (u_i, v_i]$ with $0 \le s_i < t_i \le 1$ and $u \le u_i < v_i \le \infty$ will have a multinomial distribution with parameters n and $\vec{p}$, where[2]

$$p_i \approx (t_i - s_i)\left[F(a_n v_i + b_n) - F(a_n u_i + b_n)\right].$$

For sufficiently large u and n, the $\{N(R_i)\}$ will be approximately independent Poisson random variables, with means

$$\lambda_i = n\, p_i.$$

Here we look for choices of a_n and b_n for which λ_i has a simple form, and then exploit it.

2.1 Example 1: Weibull Distribution

If $\mathsf{P}[X_j > x] = e^{-\beta x^\alpha}$ for $x > 0$, then for the choice $b_n := [\beta^{-1}\log n]^{1/\alpha}$ and $a_n := b_n/(\alpha \log n)$ we have for all large enough z,

$$\begin{aligned} n[1 - F(a_n z + b_n)] &= n \exp\big(-\beta(a_n z + b_n)^\alpha\big) \\ &= n \exp\big(-\log n (1 + z/\alpha \log n)^\alpha\big) \\ &= n \exp\big(-\log n (1 + z/\log n + o(1/\log n))\big) \\ &\approx e^{-z}, \end{aligned}$$

so $\{T_j, Y_j = (X_j - b_n)/a_n\}$ have approximately the Poisson distribution on $(0,1] \times \mathbb{R}$ with intensity measure $\nu(dt\,dy) = dt\, e^{-y} dy$ (illustrated in Figure (1)). A similar approach with suitable a_n, b_n works for any other distribution in the Gumbel domain.

The maximum $M_t := \max\{Y_j :\ T_j \le t\}$ is a non-decreasing stochastic process on the unit interval $0 < t \le 1$, with cdf

$$\begin{aligned} F_t(z) &= \mathsf{P}[M_t \le z] \\ &= \mathsf{P}[\text{No Poisson points in } (0,t] \times (z, \infty)] \\ &= e^{-te^{-z}}, \end{aligned}$$

the Gumbel distribution. The events $\{M_t \le z\}$ and $\left\{X^*_{\lfloor nt \rfloor} \le a_n z + b_n\right\}$ are identical.

2.1.1 Related Max-Stable Process

Let $\{(T_j, Y_j)\}$ be the points of a $\mathsf{Po}(dt\, e^{-y} dy)$ random field on all of $\mathbb{R}^d \times \mathbb{R}_+$, and let $f(t)$ be any positive function with finite Laplace transform. Define a random process by

$$Z(t) := \sup_j \{Y_j / f(T_j - t)\}.$$

[1]The following results would be identical if instead we took $\{T_j\} \overset{\text{iid}}{\sim} \mathsf{Un}(0,1)$.

[2]The approximation would be exact for $\{T_j\} \overset{\text{iid}}{\sim} \mathsf{Un}(0,1)$.

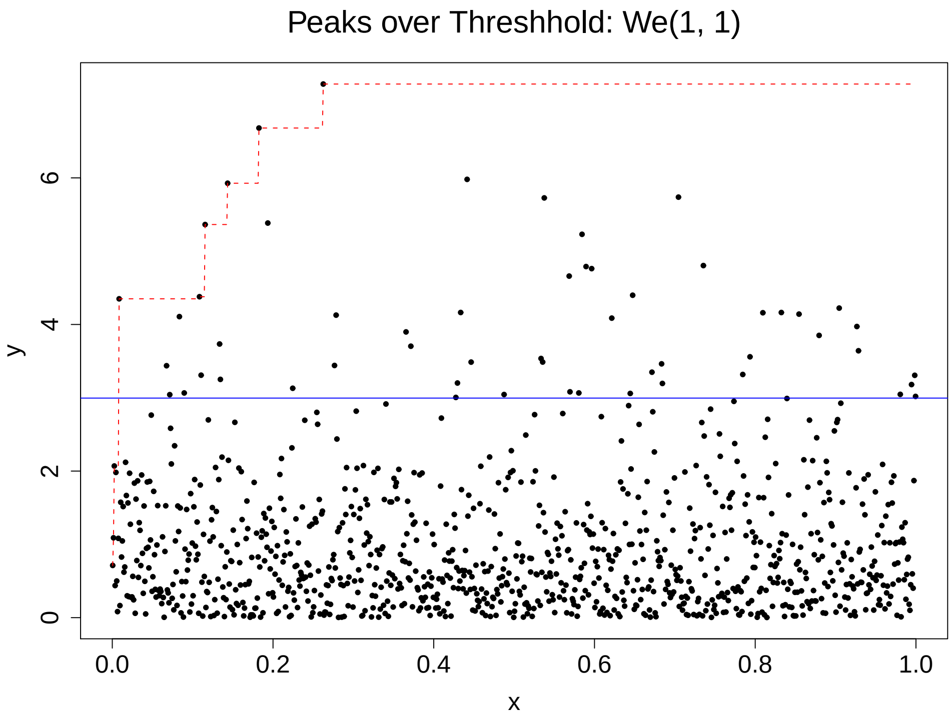

Figure 1: Simulation of 1000 scaled Weibull draws. Horizontal line is at 95% quantile. Cumulative maximum M_t is shown as dotted line.

If $f(t) = \sum a_i \mathbf{1}_{A_i}(t)$ is a simple function, then

$$\begin{aligned}
\mathsf{P}[Z(t) \le z] &= \prod_i \mathsf{P}\big[\sup_j \{Y_j/a_i \le z : \ T_j - t \in A_i\big] \\
&= \prod_i \mathsf{P}\big[\text{No Poisson pts in } (A_i + t) \times (a_i z, \infty)\big] \\
&= \prod_i \exp\big(-|A_i| e^{-b_i z}\big) \\
&= \exp\left(-\int e^{-z f(s)} ds\right),
\end{aligned}$$

so $Z(t)$ is a stationary process. For any (not necessarily simple) positive function $f(t)$ on $\mathbb{R}^d$, the same identity follows from LDCT.

2.2 Example 2: Pareto Distribution

If $\mathsf{P}[X_j > x] = \epsilon^\alpha x^{-\alpha}$ for $x > \epsilon$, then for the choice $a_n := \epsilon n^{1/\alpha}$ and $b_n := 0$ we have for all large enough z,

$$
\begin{aligned}
n[1 - F(a_n z + b_n)] &= n\big(\epsilon^\alpha (\epsilon n^{1/\alpha} z)^{-\alpha}\big) \\
&= z^{-\alpha},
\end{aligned}
$$

so $\{T_j, Y_j = (X_j - b_n)/a_n\}$ have approximately the Poisson distribution on $(0,1] \times \mathbb{R}_+$ with intensity measure $\nu(dt\,dy) = dt\,\alpha y^{-\alpha-1}\,dy$. A similar approach with suitable a_n, b_n works for any other distribution in the Fréchet domain.

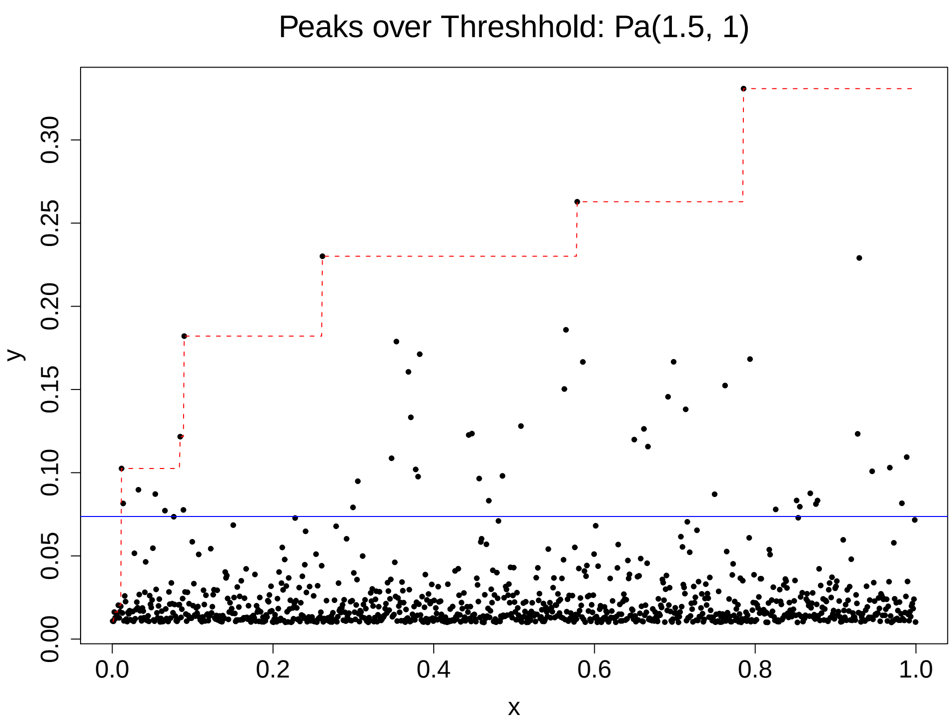

Figure 2: Simulation of 1000 scaled Pareto draws. Horizontal line is at 95% quantile. Cumulative maximum M_t is shown as dotted line.

The maximum $M_t := \max\{Y_j : \; T_j \le t\}$ is a non-decreasing stochastic process on the unit interval $0 < t \le 1$, with cdf

$$
\begin{aligned}
F_t(z) &= \mathsf{P}[M_t \le z] \\
&= \mathsf{P}[\text{No Poisson points in } (0,t] \times (z,\infty)] \\
&= e^{-tz^{-\alpha}},
\end{aligned}
$$

the Fréchet distribution. The events $\{M_t \le z\}$ and $\big\{X^*_{[nt]} \le a_n z + b_n\big\}$ are identical.

Note that the *sum* of the $\{Y_j : \; T_j \le t\}$ will be finite almost-surely if $\int_0^\infty (z \wedge 1)\alpha z^{-\alpha-1}\,dz < \infty$, *i.e.*, if $0 < \alpha < 1$; in that case the non-decreasing process

$$S_t := \sum \{Y_j : \ T_j \le t\}$$

is a fully-skewed α-Stable SII process with distribution

$$\sim \mathsf{St_A}\left(\alpha, \beta = 1, \gamma = t\Gamma(1-\alpha)\cos\tfrac{\pi\alpha}{2}, \delta = 0\right)$$

and the $\{Y_j\}$ are the "jumps" of S_t. A similar representation holds for $1 \le \alpha < 2$, but "compensation" is required (sort of like subtracting an infinite drift from S_t). There is no α-Stable process for $\alpha > 2$, although the connection between Fréchet distribution and the Poisson point process remains.

2.2.1 Related Max-Stable Process

Let $\{(T_j, Y_j)\}$ be the points of a $\mathsf{Po}(dt\,\alpha y^{-\alpha-1}\,dy)$ random field on all of $\mathbb{R}^d \times \mathbb{R}_+$, and let $0 \le f(t) \in L_\alpha(\mathbb{R}^d, dt)$. Define a random process by

$$Z(t) = \sup_j \{Y_j f(t - T_j)\}.$$

If $f(t) = \sum a_i \mathbf{1}_{A_i}(t)$ is a simple function, then

$$\begin{aligned}
\mathsf{P}[Z(t) \le z] &= \prod_i \mathsf{P}\big[\sup_j \{Y_j\, a_i \le z : \ t - T_j \in A_i\big] \\
&= \prod_i \mathsf{P}\big[\text{No Poisson pts in } (t - A_i) \times (z/a_i, \infty)\big] \\
&= \prod_i \exp\big(- |A_i| (z/a_i)^{-\alpha}\big) \\
&= \exp\left(-z^{-\alpha} \int f(s)^\alpha ds \right),
\end{aligned}$$

so $Z(t)$ is a stationary process with a Fréchet $\mathsf{Fr}\big(\alpha, \|f\|_\alpha^\alpha\big)$ distribution. For non-simple $0 \le f \in L_\alpha$, the same identity follows from LDCT.

2.3 Example 3: Beta Distribution

If $X_j \overset{\text{iid}}{\sim} \mathsf{Be}(\alpha, \beta)$ then for small ϵ, $x^{\alpha-1} \approx 1$ for $x > 1 - \epsilon$ and so

$$\begin{aligned}
\mathsf{P}[X_j > 1 - \epsilon] &\approx \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_{1-\epsilon}^1 (1-x)^{\beta-1}\, dx \\
&= \frac{\epsilon^\beta}{\beta B(\alpha,\beta)}, \qquad B(\alpha,\beta) := \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.
\end{aligned}$$

For $a_n := (\beta B(\alpha,\beta)/n)^{1/\beta}$ and $b_n := 1$, we have

$$\begin{aligned} n\mathsf{P}[X_j > a_n z + b_n] &\approx \frac{n}{\beta B(\alpha,\beta)}(1-a_n z - b_n)^\beta \\ &= (-z)^\beta, \qquad z<0 \end{aligned}$$

so $\{T_j, Y_j = (X_j - b_n)/a_n\}$ have approximately the Poisson distribution on $(0,1]\times\mathbb{R}_-$ with intensity measure $\nu(dt\,dy) = dt\,\beta(-y)^{\beta-1}\,dy$. A similar approach with suitable a_n, b_n works for any other distribution in the Reverse Weibull domain.

The maximum $M_t := \max\{Y_j :\ T_j \le t\}$ is a non-decreasing stochastic process on the unit interval $0 < t \le 1$, with cdf

$$\begin{aligned} F_t(z) &= \mathsf{P}[M_t \le z] \\ &= \mathsf{P}[\text{No Poisson points in } (0,t]\times(z,\infty)] \\ &= e^{-t(-z)^\beta}, \qquad z<0, \end{aligned}$$

the reversed Weibull distribution. The events $\{M_t \le z\}$ and $\left\{X^*_{\lfloor nt\rfloor} \le a_n z + b_n\right\}$ are identical.

The *minimum* of n iid $\mathsf{Be}(\alpha,\beta)$ random variables can be studied in the same way; for $a_n := (\alpha B(\alpha,\beta)/n)^{1/\alpha}$ and $b_n := 0$, the points $\{T_j, Y_j = (X_j - b_n)/a_n\}$ have approximately the Poisson distribution on $(0,1]\times\mathbb{R}_+$ with intensity measure $\nu(dt\,dy) = dt\,\alpha y^{\alpha-1}\,dy$, and the cumulative minimum $m_t = \min\{Y_j :\ T_j \le t\}$ is a non-increasing stochastic process satisfying $\mathsf{P}[m_t > z] = e^{-tz^\alpha}$ for $z \ge 0$, the usual (un-reversed) Weibull.

2.3.1 Related Max-Stable Process

Let $\{(T_j, Y_j)\}$ be the points of a $\mathsf{Po}(dt\,\alpha y^{\alpha-1}\,dy)$ random field on all of $\mathbb{R}^d\times\mathbb{R}_+$, and let $0 < f(t) \in L_\alpha(\mathbb{R}^d, dt)$. Define a random process by

$$Z(t) = \inf_j \{Y_j / f(t - T_j)\}.$$

If $f(t) = \sum a_i \mathbf{1}_{A_i}(t)$ is a simple function, then

$$\begin{aligned} \mathsf{P}[Z(t) > z] &= \prod_i \mathsf{P}\big[\sup_j \{Y_j/a_i > z :\ t - T_j \in A_i\big] \\ &= \prod_i \mathsf{P}\big[\text{No Poisson pts in } (t - A_i)\times(0, z\,a_i]\big] \\ &= \prod_i \exp\big(-|A_i|(z\,a_i)^\alpha\big) \\ &= \exp\left(-z^\alpha \int f(s)^\alpha ds\right), \end{aligned}$$

so $Z(t)$ is a stationary process with a Weibull $\mathsf{We}\big(\alpha, \|f\|_\alpha^\alpha\big)$ distribution. For non-simple $0 \le f \in L_\alpha$, the same identity follows from LDCT.

3 PoT Inference

Distribute points $\{y_j\}$ according to $\mathsf{Po}\big(\nu(dy)\big)$ and fix u in the support of ν. Let J be the number of points $Y_j > u$ (or $Y_j < u$ for the Weibull case), for $\nu(dy) = \alpha\gamma y^{-\alpha-1}\,dy$ (Fréchet) or $\nu(dy) = \alpha\gamma y^{\alpha-1}\,dy$ (Weibull) on $\mathbb{R}_+$, or $\nu(dy) = \gamma e^{-y}\,dy$ (Gumbel) on $\mathbb{R}$. Denoting the density of $\nu(dy)$ by $\nu(y)$, we can express the joint pdf for J and the J threshold exceedances $\{x_j\}$ as

$$L(\alpha,\gamma,a,b) = a^{-k} \prod_{j\le J} \left\{ \nu_\alpha\Big(\frac{x_j - b}{a}\Big)\right\} \quad \exp\left\{ -\nu_\alpha\Big(\frac{u-b}{a},\infty\Big)\right\}$$

and regard it as a likelihood function for α, γ, and the scale and location parameters a, b. It can probably be used to get MLEs and Fisher Information and maybe conjugate and Jeffreys' priors. The rate λ_u of exceedances of level u may also be interesting.

4 Multivariate EVT

In many application areas the problem arises of studying the extremes for random vectors. Examples include the daily prices or returns of multiple stocks, funds, indices, or other financial instruments; precipitation levels at multiple locations; the size and transmission speed of internet streams; or wind speeds and wave heights at vulnerable locations. Extreme value theory is much less well-developed for multivariate random vectors than it is for univariate quantities.

The customary approach to studying the distribution of extremes for random *vectors* begins by transforming each component of the vector to a standard EV distribution (often the "unit Fréchet" with cdf $G(x) = \exp(-1/x)$), then exploring dependence among the components. The initial transformation is most often performed parametrically by estimating the three parameters of the GEV separately for each dimension; then transforming to uniformity by the CDF for that GEV (usually ignoring uncertainty in the parameter estimation), then to unit Fréchet by the inverse CDF $G^{-1}(u) = -1/\log u$.

4.1 Asymptotic Dependence & Independence

Let (X,Y) be a two-dimensional random vector with unit Fréchet marginal distributions. The *extremal index*, denoted θ by some authors (such as Smith and Weissman, 1994), and χ by others (including Coles et al., 1999, whom we follow here), is

$$\begin{aligned} \chi &= \lim_{z\to\infty} \mathsf{P}[Y > z \mid X > z] \qquad (6) \\ &= \lim_{z\to\infty} \frac{\mathsf{P}[X > z,\ Y > z]}{1 - \exp(-1/z)} = \lim_{z\to\infty} \mathsf{P}[X > z \mid Y > z]. \end{aligned}$$

This expression is both symmetric in X and Y, and invariant under (identical) component-wise monotone transformations.

Evidently χ, when it exists, takes values between 0 and 1. The components X and Y are called *asymptotically independent* if $\chi = 0$. Surprisingly (for most of us, anyway), every nondegenerate bivariate normal distribution (even one with correlation $\rho = 0.9999$) is asymptotically independent.

If we take the monotone transformation to unit $\mathsf{No}(0,1)$ marginals with covariance $\rho < 1$, then

$$
\begin{aligned}
Y \mid X &\sim \mathsf{No}\big(\rho X,\ 1-\rho^2\big) \qquad \text{so} \\
\mathsf{P}[Y > z \mid X = x] &= \mathsf{P}\Big[\frac{Y-\rho x}{\sqrt{1-\rho^2}} > \frac{z-\rho x}{\sqrt{1-\rho^2}}\Big] = \Phi\Big(\frac{-z+\rho x}{\sqrt{1-\rho^2}}\Big) \\
\mathsf{P}[Y > z \mid X > z] &= \frac{1}{\Phi(-z)} \int_z^\infty \Phi\Big(\frac{\rho x - z}{\sqrt{1-\rho^2}}\Big)\ \varphi(x)\, dx \\
&\asymp \frac{1}{z}\Phi\Big(\frac{(\rho-1)z}{\sqrt{1-\rho^2}}\Big) \to 0 \text{ as } z \to \infty
\end{aligned}
$$

for any $\rho < 1$.

Any value of $\chi \in [0,1]$ is possible. To see this, take $0 \le \lambda \le 1$ and consider the "bivariate logistic model" with cdf

$$
G(x,y) = \exp\Big\{-\Big[x^{-1/\lambda} + y^{-1/\lambda}\Big]^\lambda\Big\}
$$

for $\lambda > 0$, and $G(x,y) = \exp\big(-1/\min(x,y)\big)$ (the limit) for $\lambda = 0$. Evidently X and Y each have unit Fréchet marginals (take the limits $x \to \infty$ and $y \to \infty$), and

$$
\begin{aligned}
\chi &= \lim_{z\to\infty} \frac{\mathsf{P}[X > z,\ Y > z]}{\mathsf{P}[X > z]} \\
&= \lim_{z\to\infty} \frac{1 - \mathsf{P}[X \le z] - \mathsf{P}[Y \le z] + \mathsf{P}[X \le z,\ Y \le z]}{1 - \mathsf{P}[X \le z]} \\
&= \lim_{z\to\infty} \frac{1 - 2G(z) + G(z,z)}{1 - G(z)} \\
&= 2 - \lim_{z\to\infty} \frac{1 - G(z,z)}{1 - G(z)} \\
&= 2 - \lim_{z\to\infty} \frac{1 - \exp(-2^\lambda/z)}{1 - \exp(-1/z)} = 2 - 2^\lambda
\end{aligned}
$$

by L'Hôpital's rule. This ranges from 0 to 1 as λ ranges from 1 to 0.

4.2 Multivariate EV Distributions

Let $\{(X_i, Y_i)\}$ be iid random vectors in $\mathbb{R}^2$ with Fréchet marginals and, for $n \in \mathbb{N}$, denote the component-wise maxima by:

$$
M_n := (X_n^*, Y_n^*), \qquad X_n^* := \max_{1\le i\le n} X_i, \qquad Y_n^* := \max_{1\le i\le n} Y_i.
$$

Then

$$
\mathsf{P}[X_n^*/n \le z] = \mathsf{P}[X_1 \le nz]^n = \big(e^{-1/nz}\big)^n = e^{-1/z}
$$

and similarly $\mathsf{P}[Y_n^*/n \le z] = e^{-1/z}$, so both marginals of M_n/n are unit Fréchet.

Theorem 1 *If there exists a non-degenerate bivariate distribution $G(x,y)$ such that $M_n/n \Rightarrow G(x,y)$ as $n \to \infty$,* i.e., *that*

$$
\mathsf{P}\left[X_n^* \le nx,\ Y_n^* \le ny\right] \to G(x,y),
$$

then

$$G(x,y) = e^{-V(x,y)} \tag{7}$$

for a nonnegative function $V : \mathbb{R}_+^2 \to \mathbb{R}_+$ *of the form*

$$V(x,y) = 2\int_{\Delta^1} \max\left(\frac{\sigma_1}{x}, \frac{\sigma_2}{y}\right) H(d\sigma) \tag{8}$$

for some probability measure $H(d\sigma)$ *on the unit simplex* $\Delta^1 \subset \mathbb{R}_+^2$ *with mean*

$$\int_{\Delta^1} \sigma\, H(d\sigma) = \left(\tfrac{1}{2}, \tfrac{1}{2}\right). \tag{9}$$

Every such "spectral measure" H gives rise to a bivariate extreme value distribution; below we'll motivate this by showing how H arises and where (8) comes from. From Eqns (7, 8), the marginal distribution function for X must be $G(x,\infty) = \exp\big(-V(x,\infty)\big) = \exp\big(-2\int_{\Delta^1} \sigma^1 H(d\sigma)/x$, so (9) is simply a standardization condition ensuring that X and Y have unit Fréchet marginal distributions. Meanwhile, let's set $G(x) := \exp(-1/x)$ and note that the extremal index of (6) can also be calculated as $\chi = \lim_{u\to 1} \chi(u)$ for

$$\begin{aligned}
\chi(u) :&= \frac{\mathsf{P}[G(X) > u,\ G(Y) > u]}{\mathsf{P}[G(X) > u]} \\
&= \frac{1 - \mathsf{P}[G(X) \le u] - \mathsf{P}[G(Y) \le u] + \mathsf{P}[G(X) \le u,\ G(Y) \le u]}{\mathsf{P}[G(X) > u]} \\
&= \frac{1 - 2u + \mathsf{P}[G(X) \le u,\ G(Y) \le u]}{1-u} \\
&= 2 - \frac{1 - \mathsf{P}[G(X) \le u,\ G(Y) \le u]}{1-u} \\
&= 2 - \frac{\log \mathsf{P}[G(X) \le u,\ G(Y) \le u]}{\log u} + O(1-u)
\end{aligned}$$

since $\log(1-\epsilon) = -\epsilon + O(\epsilon^2)$ for $\epsilon \approx 0$. With $u = G(z)$, or $z = -1/\log u$, we have $\mathsf{P}[G(X) \le u,\ G(Y) \le u] = \mathsf{P}[X \le z,\ Y \le z] = G(z,z) = \exp\big(-V(z,z)\big)$ so

$$\chi(u) \approx 2 - \frac{-V(z,z)}{-1/z} = 2 - zV(z,z)$$

and by (8) in the limit we have

$$\chi = 2 - 2\int_{\Delta^1} \max(\sigma_1, \sigma_2)\, H(d\sigma).$$

This will be zero if and only if $\max(\sigma_1, \sigma_2)$ is one on the support of H, *i.e.*, if and only if H is supported entirely on the boundary $\partial\Delta^1 = \{(0,1), (1,0)\}$.

4.3 Poisson Connection

Let H be a probability measure on Δ^1 satisfying (9), and consider a Poisson random measure $\mathcal{N}(dx\, dy)$ on the first quadrant whose intensity can be written $2H(d\sigma)r^{-2}dr$ in polar coordinates

$r = x + y$, $\sigma = (x, y)/r$. Let X^* and Y^* denote the maxima of the x and y coordinates of the mass points of $\mathcal{N}(dx\,dy)$, respectively. For $x, y > 0$ the event that $[X^* \le x,\ Y^* \le y]$ is just the event that $\mathcal{N}$ assigns zero points to $\big([0, x] \times [0, y]\big)^c$. We can compute this in polar coordinates as

$$\begin{aligned}
\mathsf{P}[X^* \le x,\ Y^* \le y] &= \exp\left\{ -\int_{\big([0,x]\times[0,y]\big)^c} 2H(d\sigma)\, r^{-2} dr \right\} \\
&= \exp\left\{ -\int_{(r\sigma_1 > x)\,\|\,(r\sigma_2 > y)} 2H(d\sigma)\, r^{-2} dr \right\} \\
&= \exp\left\{ -\int_{r > \min(x/\sigma_1, y/\sigma_2)} 2H(d\sigma)\, r^{-2} dr \right\} \\
&= \exp\left\{ -\int_{\Delta^1} \frac{2}{\min(x/\sigma_1, y/\sigma_2)}\, H(d\sigma) \right\} \\
&= \exp\left\{ -2 \int_{\Delta^1} \max\left(\frac{\sigma_1}{x}, \frac{\sigma_2}{y} \right) H(d\sigma) \right\},
\end{aligned}$$

exactly the same as $G(x, y)$ from (7). Thus for large n the extremes of the vectors $\{(X_j/n, Y_j/n)\}$, $1 \le j \le n$ behave like the extremes of a Poisson point cloud with intensity measure $2H(d\sigma)r^{-2}dr$. In $d \ge 2$ dimensions the same things work, of course, with Poisson intensity measure $dH(d\sigma)r^{-2}$ on $\mathbb{R}_+^d = \Delta^{d-1} \times \mathbb{R}_+$.

Coles et al. (1999) also define a second index

$$\bar{\chi} := \lim_{z \to \infty} \frac{2 \log \mathsf{P}[X > z]}{\log \mathsf{P}[X > z,\ Y > z]} - 1,$$

taking values in the interval $[-1, 1]$, which depends on the *minimum* of X, Y in the tails; they argue that it measures a degree of dependence for asymptotically independent variables (those for which $\chi = 0$). It vanishes for independent X, Y, and takes the value $+1$ for fully-dependent $X \equiv Y$.

A Appendix: A Few Less Familiar Distributions

Several distributions pop up when exploring extremes that are less studied than the usual suspects; here we collect a bit about them.

A.1 Pareto

If $U \sim \mathsf{Un}(0,1)$ and $\alpha > 0$, then $X = U^{-1/\alpha}$ has the *Pareto* distribution taking all values in $(1, \infty)$. The survival function (SF) and density function (pdf) are

$$\begin{aligned} \mathsf{P}[X > x] &= \mathsf{P}[U^{-1/\alpha} > x] \\ &= \mathsf{P}[U < x^{-\alpha}] \\ &= x^{-\alpha}, \qquad x > 1 \\ f(x) &= \alpha x^{-\alpha-1} \mathbf{1}_{\{x>1\}}. \end{aligned}$$

This is the prototype "heavy-tailed" distribution, whose SF and pdf fall off like powers of x (instead of the exponential fall-off typical of most commonly-studied distributions). The mean is infinite for $\alpha \le 1$, and $1/(\alpha - 1) < \infty$ for $\alpha > 1$; the variance infinite for $\alpha \le 2$.

It is frequently taken to be part of a two-parameter scale family ($Y := \epsilon X \sim \mathsf{Pa}(\alpha, \epsilon)$, taking all values in (ϵ, ∞)) and less commonly part of a three-parameter location/scale family.

A.2 Gumbel

If $Y \sim \mathsf{Ex}(1)$ is a standard exponential random variable, then $X = -\log Y$ has the standard *Gumbel* distribution taking all values in $\mathbb{R}$. The CDF and pdf are

$$\begin{aligned} \mathsf{P}[X \le x] &= \mathsf{P}[Y \ge e^{-x}] \\ &= e^{-e^{-x}} \\ f(x) &= e^{-x-e^{-x}} \end{aligned}$$

and the mean is $\mathsf{E}X = \gamma_e \approx 0.5772$, the Euler-McLaren constant. Since the mode is zero, the distribution is skewed to the right; the tail probabilities fall off exponentially as $x \to \infty$, but much faster as $x \to -\infty$. It is commonly taken to be part of a two-parameter location/scale family.

A.3 Fréchet

If $Y \sim \mathsf{Ex}(1)$ is a standard exponential random variable and $\alpha > 0$, then $X = Y^{-1/\alpha}$ has the standard *Fréchet* distribution taking all values in $\mathbb{R}_+$. The PDF and pdf are

$$\begin{aligned} \mathsf{P}[X \le x] &= \mathsf{P}[Y \ge x^{-\alpha}], \qquad x > 0 \\ &= e^{-x^{-\alpha}} \\ f(x) &= \alpha x^{-\alpha-1} e^{-x^{-\alpha}} \mathbf{1}_{\{x>0\}} \end{aligned}$$

and the mean is $\mathsf{E}X = \Gamma(1 - 1/\alpha)$ for $\alpha > 1$, or infinity for $\alpha \le 1$. The variance is infinite for $\alpha \le 2$. The mode $(1 + 1/\alpha)^{-1/\alpha}$ and median $(\log 2)^{-1/\alpha}$ are well-defined for all $\alpha > 0$.

This too is a heavy-tailed distribution, with SF and cdf falling off at the same rates as the $\mathsf{Pa}(\alpha)$. It is commonly taken to be part of a three-parameter location/scale family.

A.4 Weibull

If $Y \sim \mathsf{Ex}(1)$ is a standard exponential random variable and $\alpha > 0$ then $X = Y^{1/\alpha}$ has the *Weibull* distribution taking all values in $\mathbb{R}_+$. The SF and pdf are

$$\begin{aligned} \mathsf{P}[X > x] &= \mathsf{P}[Y > x^\alpha], \qquad x > 0 \\ &= e^{-x^\alpha} \\ f(x) &= \alpha x^{\alpha-1} e^{-x^\alpha} \mathbf{1}_{\{x>0\}} \end{aligned}$$

and the mean $\mathsf{E}X = \Gamma(1 + 1/\alpha)$ and variance are finite for all $\alpha > 0$.
It is commonly taken to be part of a two-parameter scale family, with SF $S(x) = \exp(-\beta x^\alpha)$ and hence hazard function

$$h(x) = f(x)/S(x) = \frac{\alpha\beta x^{\alpha-1}\exp\left(-\beta x^\alpha\right)}{\exp\left(-\beta x^\alpha\right)} = \alpha\beta x^{\alpha-1},$$

a monomial in x that can be either increasing (for $\alpha > 1$) or decreasing (for $\alpha < 1$) to model failure times for systems with either increasing or decreasing instantaneous hazard.
If $X \sim \mathsf{We}(\alpha)$ has the Weibull distribution then $Z := -X$ has the *reversed Weibull* distribution, with pdf

$$g(z) = \alpha(-z)^{\alpha-1} e^{-(-z)^\alpha} \mathbf{1}_{\{z<0\}}.$$

A.5 GEV

McFadden (1978) discovered that location/scale families built on the Gumbel, Fréchet, and reversed Weibull distribution were all special cases of the Generalized Extreme Value distribution, with conventional CDF parameterization given by:

$$G(x; \mu, \sigma, \xi) = \exp\left\{ -\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi} \right\} \tag{5}$$

for those x satisfying $1 + \xi(x - \mu)/\sigma > 0$, and pdf:

$$g(x; \mu, \sigma, \xi) = \frac{1}{\sigma}\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1-1/\xi} \exp\left\{ -\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi} \right\}.$$

Note the range of GEV depends on the sign of ξ: $X \in (\mu - \sigma/\xi, \infty)$ for $\xi > 0$, $X \in \mathbb{R}$ for $\xi = 0$, and $X \in (-\infty, \mu - \sigma/\xi)$ for $\xi < 0$. Evidently (5) is a location/scale family built on a standard GEV distribution ($\mu = 0$, $\sigma = 1$) with CDF and pdf:

$$\begin{aligned} G(x; \xi) &= \exp\left\{ -\left(1 + \xi x\right)^{-1/\xi} \right\} \\ g(x; \xi) &= [1 + \xi x]^{-1-1/\xi} \exp\left\{ -\left(1 + \xi x\right)^{-1/\xi} \right\}. \end{aligned}$$

The standard Gumbel, Fréchet, and reversed Weibull can each be expressed in terms of $G(x;\mu,\sigma,\xi)$ from (5):

$$\begin{array}{llll}
\text{Fréchet:} & \exp\left(-x^{-\alpha}\right) & = G(x;\, \mu=1, \quad \sigma=\frac{1}{\alpha}, \xi=\frac{1}{\alpha}), & \xi>0 \\
\text{Gumbel:} & \exp\left(-e^{-x}\right) & = G(x;\, \mu=0, \quad \sigma=1,\, \xi=0), & \xi=0 \\
\text{Rev Weibull:} & \exp\left(-(-x)^{\alpha}\right) & = G(x;\, \mu=-1, \sigma=\frac{1}{\alpha}, \xi=-\frac{1}{\alpha}), & \xi<0.
\end{array}$$

Note that if $\{X_i\} \overset{\text{iid}}{\sim} \mathsf{GEV}(\mu,\sigma,\xi)$ then $X_n^* := \max_{1\le i\le n}\{X_i\} \sim \mathsf{GEV}(\mu^*,\sigma^*,\xi)$ for $\mu^* := \mu + \sigma(n^\xi - 1)/\xi$ and $\sigma^* := \sigma n^\xi$, *i.e.*, the maximum of the first n also has the GEV distribution with the same shape parameter ξ, larger location parameter $\mu^* > \mu$, and scale σ^* that is larger (resp, smaller) than that of X_i if $\xi > 0$ (resp, $\xi < 0$). This property (that $G(x;\mu,\sigma,\xi)^n = G(x;\mu_n^*,\sigma_n^*,\xi)$ for some μ_n^* and σ_n^*, for each $n \in \mathbb{N}$) characterizes the GEV, and is the basis for the Three Types theorem.

A.6 GPD

If $X \sim \mathsf{GEV}(\mu,\sigma,\xi)$ for $\xi > 0$ with CDF

$$G(x;\mu,\sigma,\xi) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \tag{5}$$

then for $y > 0$ the *exceedances* $Y = [X-u]$ of a high level $u \gg \mu - \sigma/\xi$ satisfy

$$\begin{aligned}
\mathsf{P}[Y>y \mid Y>0] &= \mathsf{P}[X>y+u \mid X>u] \\
&= \frac{1-\exp\left\{-\left[1+\xi\left(\frac{y+u-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}}{1-\exp\left\{-\left[1+\xi\left(\frac{u-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}} \\
&\approx \frac{\left[1+\xi\left(\frac{y+u-\mu}{\sigma}\right)\right]^{-1/\xi}}{\left[1+\xi\left(\frac{u-\mu}{\sigma}\right)\right]^{-1/\xi}} \\
&= \left[\frac{\sigma+\xi\,(y+u-\mu)}{\sigma+\xi\,(\quad u-\mu)}\right]^{-1/\xi} \\
&= \left[1+\xi y/\hat\sigma\right]^{-1/\xi}, \qquad \text{w/ } \hat\sigma := \sigma + \xi\,(u-\mu)\,.
\end{aligned}$$

This is the *generalized Pareto distribution* $\mathsf{GPD}(\xi,\hat\sigma)$, with CDF $H(y) = 1 - [1+\xi y/\hat\sigma]_+^{-1/\xi}$ for $y > 0$ and mean

$$\mathsf{E}[Y] = \int_0^\infty \bar H(y)\,dy = \hat\sigma/(1-\xi), \qquad \xi < 1$$

(or infinity if $\xi \ge 1$), so for $0 < \xi < 1$ (*i.e.*, the Fréchet case with $\alpha > 1$),

$$\mathsf{E}[X-u \mid X>u] \approx \frac{\sigma-\xi\mu}{1-\xi} + \frac{\xi}{1-\xi}u = \frac{\sigma}{1-\xi} + \frac{\xi}{1-\xi}(u-\mu)$$

is linear in u with a slope that determines ξ. The variance of the GPD is also available in closed form, infinite for $\xi \ge \frac{1}{2}$ and, for $0 < \xi < \frac{1}{2}$,

$$\mathsf{V}[X \mid X>u] = \mathsf{V}[Y] = \frac{\hat\sigma^2}{(1-\xi)^2(1-2\xi)} = \frac{\mathsf{E}[Y]^2}{1-2\xi}.$$

For $0 < \xi < \infty$, when X has a Fréchet distribution, the GPD is a scaled (by $\xi/\hat\sigma$) and offset (to zero) version of the ordinary Pareto distribution. It has the interesting property that, for any $v > 0$ and $y > 0$,

$$\begin{aligned}\mathsf{P}[Y > y + v \mid Y > v] &= \frac{[1+\xi(y+v)/\hat\sigma]^{-1/\xi}}{[1+\xi(v)/\hat\sigma]^{-1/\xi}} = \left[\frac{\hat\sigma + \xi v + \xi y}{\hat\sigma + \xi v}\right]^{-1/\xi} \\ &= [1 + \xi y/\hat\sigma']^{-1/\xi}, \qquad \hat\sigma' := \hat\sigma + \xi v,\end{aligned}$$

i.e., the conditional distribution of $(Y - v)$ given $[Y > v]$ is again $\mathsf{GPD}(\xi, \hat\sigma')$.

This is the key to *estimating* the shape ξ and threshold u_0 above which extremes are modeled sufficiently well by the GPD, a black art. A plot of the empirical "Mean Residual Life" (MRL) $Y := (X - u)$, plotted against u, should be approximately linear above some threshold u_0. Unfortunately the variation around that line gets wider and wider with increasing u (because the MRL is estimated on the basis of fewer and fewer extreme events as u increases). The variance and mean calculations above should make it possible to generate error bars.

A common estimator of $\xi \equiv 1/\alpha$ in the Fréchet case is "Hill's Index" (Hill, 1975). Let $\{X_{(i)}\}$ be the order statistics (with $X_{(1)}$ the largest) for an iid sample of $n \in \mathbb{N}$ observations $\{X_j\}$ and, for each $1 \le k \le n$, set

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

This is just the MLE based on the observations that exceed an order statistic. Resnick and Stărică (1998) showed it to be consistent as $k \to \infty$ and $n/k \to \infty$, even for many dependent sequences.

References

Coles, S. G. (2001), *An Introduction to Statistical Modeling of Extreme Values*, New York, NY: Springer-Verlag.

Coles, S. G., Heffernan, J. E., and Tawn, J. A. (1999), “Dependence measures for extreme value analysis,” *Extremes*, 2, 339–365.

Fisher, R. A. and Tippett, L. H. C. (1928), “Limiting forms of the frequency distributions of the largest or smallest member of a sample,” *Proceedings of the Cambridge Philosophical Society*, 24, 180–190.

Hill, B. M. (1975), “A Simple General Approach to Inference About the Tail of a Distribution,” *Annals of Statistics*, 3, 1163–1174.

McFadden, D. (1978), “Modeling the Choice of Residential Location,” in *Spatial Interaction Theory and Planning Models*, eds. A. Karlqvist, L. Lundqvist, F. Snickars, and J. W. Weibull, Amsterdam, NL: North Holland, *Studies in Regional Science and Urban Economics*, volume 3, pp. 75–96, reprinted in ed. J. Quigley, London, UK: Edward Elgar, *The Economics of Housing, Volume I*, 1997, 531–552.

Resnick, S. I. (1987), *Extreme Values, Regular Variation, and Point Processes*, *Applied Probability*, volume 4, New York, NY: Springer-Verlag.

Resnick, S. I. and Stărică, C. (1998), “Tail index estimation for dependent data,” *Annals of Applied Probabability*, 8, 1156–1183, doi:10.1214/aoap/1028903376.

Smith, R. L. and Weissman, I. (1994), “Estimating the Extremal Index,” *Journal of the Royal Statistical Society, Ser.* B*: Statistical Methodology*, 56, 515–528.

LaTeX'd: November 24, 2014