An introduction to extreme-value theory

Thomas Opitz

BioSP, INRAE, Avignon

Short course at the Stochastic Geometry Days

Dijon, 12-13 June, 2023

Biostatistique & Processus Spatiaux

Plan for this course

Three sessions of two hours, with (very roughly) the following main topics:

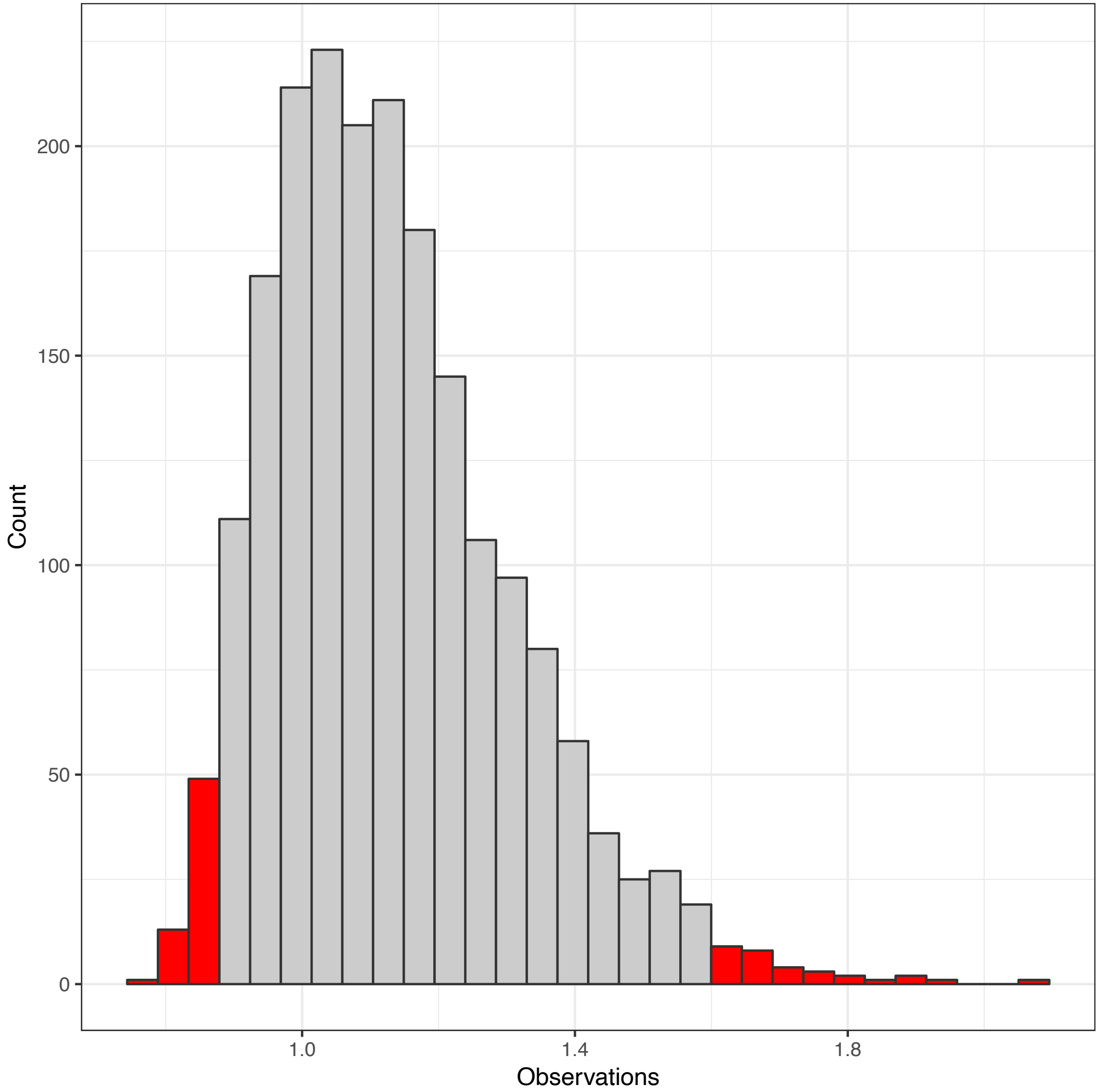
- Theory for univariate extremes
- Theory for dependent extremes based on maxima and point processes
- Theory for dependent extremes based on threshold exceedances

The origins of Extreme-Value Theory (EVT)

- A **probabilistic theory** with its origins in the first half of the 20th century:
  - Fréchet (1927). Sur la loi de probabilité de l'écart maximum. *Annales de la Société Polonaise de Mathématique*.
  - Fisher, Tippett (1928). Limiting forms of the frequency distribution of the largest and smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*.
  - von Mises (1936). La distribution de la plus grande de n valeurs. *Revue Mathématique de l'Union Interbalcanique*
  - Gnedenko (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*.
- Strong development of **multivariate and process theory** since the 1970s
- **Statistical methods and applications**
  - Often at the origin of theoretical developments
    (for example, Tippett's work for the cotton industry)
  - Seminal monograph *Statistics of Extremes* (1958) of Gumbel
  - Numerous applications since the 1980s
  - Today, strong use for finance/insurance and climate/environment
  - Typical goals:
    - Estimate and extrapolate extreme-event probabilities
    - Stochastically generate new extreme-event scenarios

Extreme events

Extreme events are located in the upper or lower **tail of the distribution**:



Without loss of generality, we focus on the extremes in the upper tail.

Classical asymptotic frameworks: Averages / Extremes

Consider independent and identically distributed (i.i.d.) random variables $X_1, X_2, \ldots$

Averages $\overline{S}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

Central Limit Theorem

$$\frac{\overline{S}_n - \mu}{\sigma_n} \to Z \sim \mathcal{N}(0,1)$$

Gaussian limit distribution
(Sum-stability)

Spatial extension:

Gaussian processes

Geostatistics

Extremes (maxima) $M_n = \max_{i=1}^{n} X_i$

Fisher–Tippett–Gnedenko Theorem

$$\frac{M_n - a_n}{b_n} \to Z \sim \mathrm{GEV}(\xi) \text{ (tail index } \xi \in \mathbb{R})$$

Extreme-value limit distribution
(Max-stability)

Spatial extension:

Max-stable processes

Spatial Extreme-Value Theory

The trinity of the three fundamental approaches

Three asymptotic approaches to study extreme events in an i.i.d. sample $\{X_i\}$:

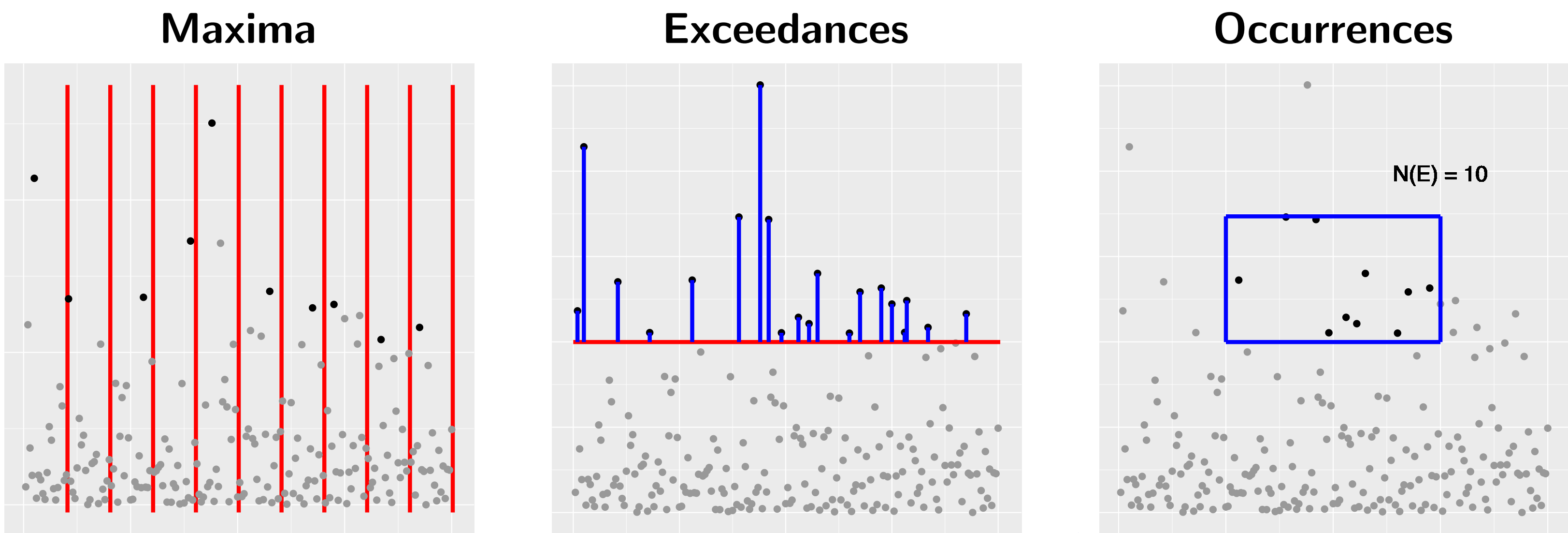
1. **Block maxima**: $M_n = \max_{i=1}^n X_i$ using blocks of size n
2. **Threshold exceedances** above a high threshold u: $(X_i - u) \mid X_i \geq u$
3. **Occurrence counts**: $N(E) = |\{X_i \in E,\ i = 1, \ldots, n\}|$ for extreme events E

Asymptotic theory

For

- increasing block size n,
- for increasing threshold u, and
- for more and more extreme event sets E,

we obtain **coherent theoretical representations across the three approaches**.



The maximum of a sample

For a series of **independent and identically distributed (iid) random variables**

$$X_i \sim F, \quad i = 1, 2, \ldots$$

we consider the **maximum**

$$M_n = \max_{i=1}^{n} X_i \sim F^n,$$

where

$$F^n(x) = (F(x))^n.$$

The fundamental extreme-value limit theorem

Fisher–Tippett–Gnedenko Theorem

Let X_i, $i = 1, 2, \ldots$ iid. If deterministic normalizing sequences a_n (location) and $b_n > 0$ (scale) exist such that

$$\frac{M_n - a_n}{b_n} \xrightarrow{d} Z \sim G, \quad n \to \infty, \quad (\star)$$

with a nondegenerate limit distribution G, then G is of one of the **three types of extreme-value distributions**:

- **(Reverse) Weibull**: $\tilde{G}(z) = \exp(-(-x)_+^{-\alpha})$ with $\alpha > 0$ (with support $(-\infty, 0)$)
- **Gumbel**: $\tilde{G}(z) = \exp(-\exp(-x))$ (with support $\mathbb{R}$)
- **Fréchet**: $\tilde{G}(z) = \exp(-x_+^{\alpha})$ with $\alpha > 0$ (with support $(0, \infty)$)

Remarks:

- Being of a certain type means being equal up to a location-scale transformation: $G(z) = \tilde{G}(a + bz)$ with some $b > 0$, $a \in \mathbb{R}$. We can always choose a_n, b_n such that $G = \tilde{G}$.
- If convergence $(\star)$ holds, we say that **F is in the maximum domain of attraction (MDA) of G**.
- Equivalently to $(\star)$, we have $F^n(a_n + b_n z) \to G(z)$, $n \to \infty$, $z \in \mathbb{R}$.

Sketch of the proof (1)

A key ingredient is the **Extremal-Types Theorem**, here copied from Embrechts, Klueppelberg, Mikosch (1996). An early proof is due to Gnedenko & Kolmogorov (1954).

Extremal-Types Theorem

Let $A, B, A_1, A_2, \ldots$ be random variables and $b_n > 0$, $\beta_n > 0$ and $a_n, \alpha_n \in \mathbb{R}$ be deterministic sequences. If the following convergence holds,

$$\frac{A_n - a_n}{b_n} \xrightarrow{d} A, \quad n \to \infty,$$

then the alternative convergence

$$\frac{A_n - \alpha_n}{\beta_n} \xrightarrow{d} B, \quad n \to \infty, \tag{1}$$

holds if and only if

$$\frac{b_n}{\beta_n} \to b \in [0, \infty), \quad \frac{a_n - \alpha_n}{\beta_n} \to a \in \mathbb{R}, \quad n \to \infty.$$

If (1) holds, then $B \stackrel{d}{=} bA + a$ with a, b being uniquely determined. Moreover, A is nondegenerate if and only if $b > 0$, and the A and B are said to belong to the same type.

Sketch of the proof (2)

In the following, all convergences are understood for $n \to \infty$.

1. If the convergence $F^n(a_n + b_n z) \to G(z)$ holds, then for any $t > 0$,

$$F^{\lfloor nt \rfloor}(a_{\lfloor nt \rfloor} + b_{\lfloor nt \rfloor} z) \to G(z), \quad z \in \mathbb{R}. \tag{2}$$

2. Observe that

$$F^{\lfloor nt \rfloor}(a_n + b_n z) = (F^n(a_n + b_n z))^{\lfloor nt \rfloor / n} \to G^t(z). \tag{3}$$

3. Using the Extremal-Types Theorem, there exist deterministic functions $\gamma(t) > 0$ and $\delta(t)$ such that

$$\frac{b_n}{b_{\lfloor nt \rfloor}} \to \gamma(t), \quad \frac{a_n - a_{\lfloor nt \rfloor}}{b_{\lfloor nt \rfloor}} \to \delta(t), \quad t > 0.$$

   By considering (2) and (3), we get

$$G^t(z) = G(\delta(t) + \gamma(t) z), \quad t > 0.$$

4. A consequence of the last equality is that for $s, t > 0$,

$$\gamma(st) = \gamma(s)\gamma(t), \quad \delta(st) = \gamma(t)\delta(s) + \delta(t).$$

5. The solutions of this functional equation are given by the three distribution functions of the reverse Weibull, Gumbel and Fréchet type.

Generalized Extreme-Value distribution (GEV)

The **Generalized Extreme-Value distributions (GEV)** uses threes parameter to jointly represent all possible limit distributions G:

$$G(z) = \mathrm{GEV}(z; \xi, \mu, \sigma) = \exp\left(- \left[1 + \xi \frac{z - \mu}{\sigma} \right]_+^{-1/\xi} \right) \quad (\star\star)$$

- **Shape parameter** (or **tail index**) $\xi \in \mathbb{R}$, determining the extremal type:
  - Reverse-Weibull MDA for $\xi < 0$
  - Gumbel MDA for $\xi = 0$
  - Fréchet MDA for $\xi > 0$
- Location parameter $\mu \in \mathbb{R}$
- Scale parameter $\sigma > 0$

For $\xi = 0$, $(\star\star)$ is the limit for $\xi \to 0$: $G(z) = \exp(-\exp(-(z - \mu)/\sigma))$, $z \in \mathbb{R}$.

The $(...)_+$-operator in $(\star\star)$ means that the distribution G has positive density dG/dz for values z satisfying $1 + \xi \frac{z-\mu}{\sigma} > 0$

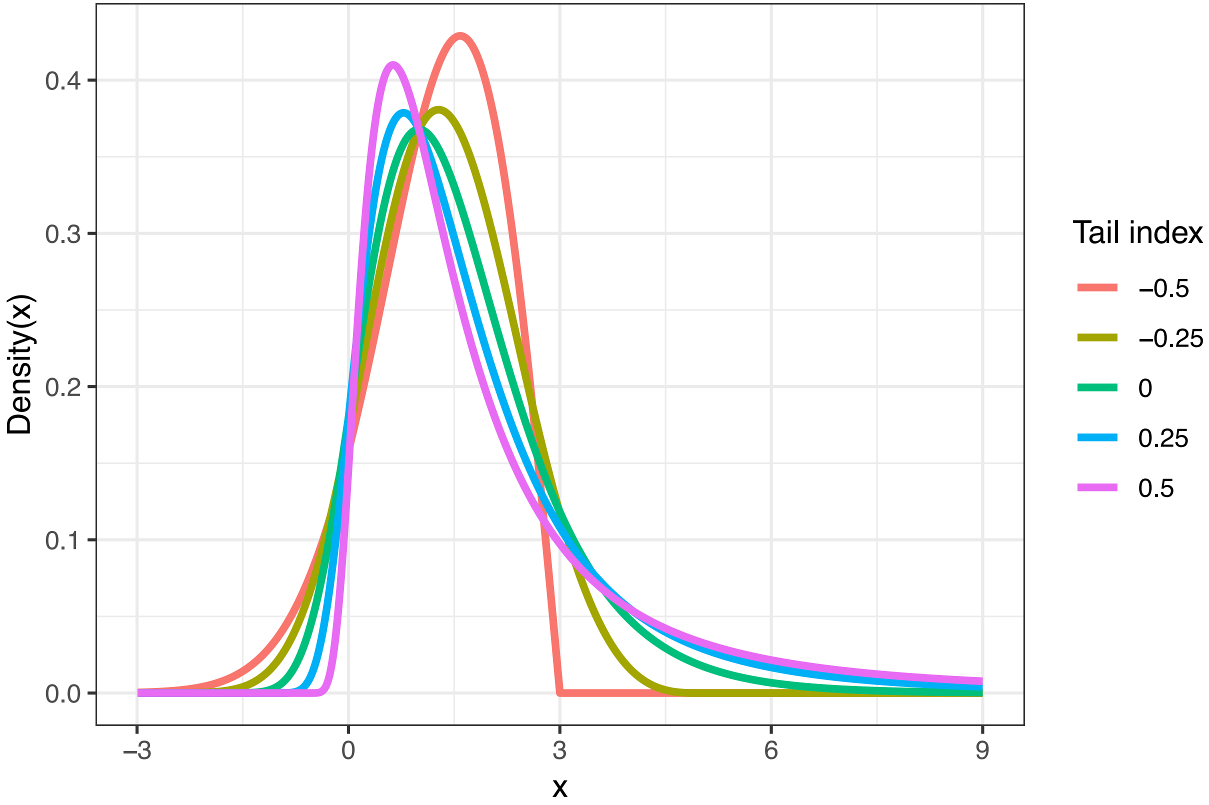
$\Rightarrow$ **Support of the GEV:** $A_{\xi,\sigma,\mu} = \begin{cases} (-\infty, \mu - \sigma/\xi), & \xi < 0, \\ (-\infty, \infty), & \xi = 0, \\ (\mu - \sigma/\xi, \infty), & \xi > 0. \end{cases}$

Illustration: GEV densities

In the MDA convergence ($\star$), we can always choose the normalizing sequences a_n, b_n such that $\mu = 0$, $\sigma = 1$, as for the probability densities shown below.

The three types have very different upper tail structure:

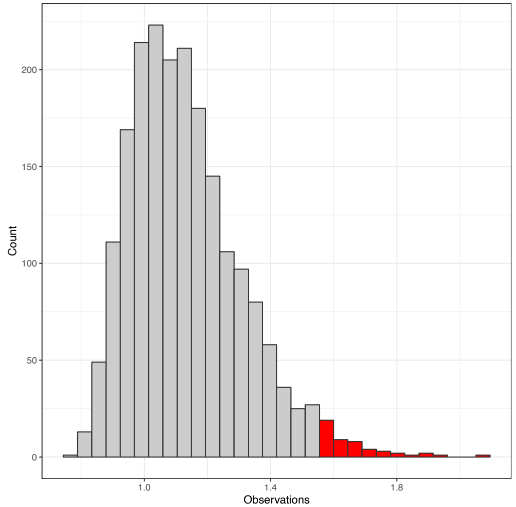
- Reverse-Weibull for $\xi < 0$: light tails with finite upper endpoint (GEV finite upper endpoint is $\mu - \sigma/\xi$)
- Gumbel for $\xi = 0$: exponential tail
- Fréchet for $\xi > 0$: power-law tails, i.e., heavy tails



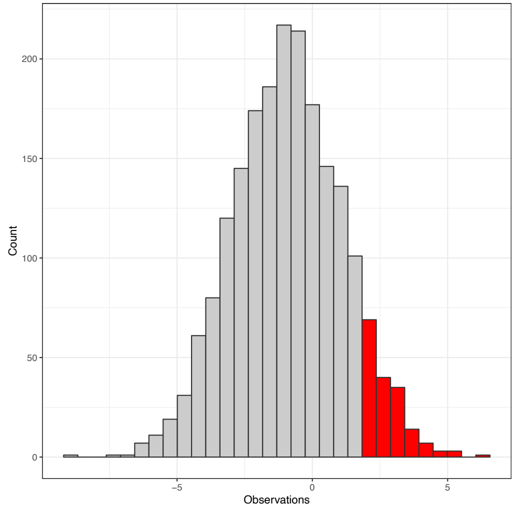
Empirical illustration

Histograms of i.i.d. samples X_i, $i = 1, 2, \ldots, n$, with different tail index ξ.

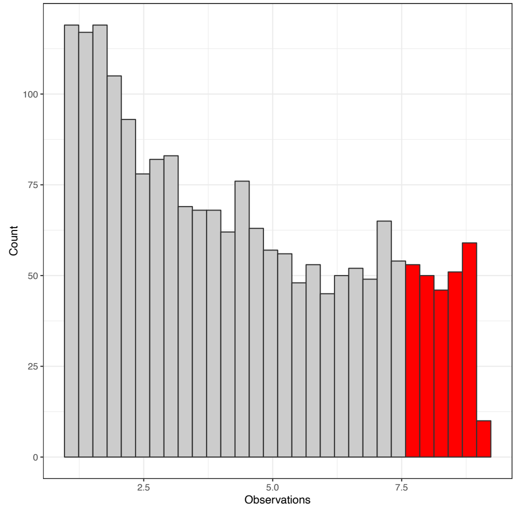
Heavy tail
$\xi > 0$



Exponential tail
$\xi = 0$



Bounded tail
$\xi < 0$



Examples of MDAs of common distributions:

- $\xi > 0$: Pareto ($\xi = 1/\mathrm{shape}$), student's t ($\xi = \mathrm{shape}$)
- $\xi = 0$: Normal, Exponential, Gamma, Lognormal
- $\xi < 0$: Uniform ($\xi = -1$), Beta

Example: GEV limit of the exponential distribution

Consider the standard exponential distribution with cdf $F(x) = 1 - \exp(-x), \; x > 0$.
The distribution F^n of the maximum $M_n = \max_{i=1}^n X_i$, where $X_i \overset{iid}{\sim} F$, $i = 1, \ldots, n$, is

$$F^n(x) = (1 - exp(-x))^n \,.$$

Can we find a_n and b_n such that $\lim_{n\to\infty} F^n(a_n + b_n x)$ exists and is nondegenerate?

For $x > -\log n$,

$$F^n(\log n + x) = (1 - \exp(-(\log n + x))^n = \left(1 - \frac{\exp(-x)}{n}\right)^n$$
$$\to \exp(-\exp(-x)), \quad n \to \infty$$

Conclusion:

- Using $a_n = \log(n)$ and $b_n = 1$, we obtain $\lim_{n\to\infty} F^n(a_n + b_n x) = \exp(-\exp(-x))$ for any $x \in \mathbb{R}$.
- The exponential distribution is in the **maximum domain of attraction of the standard Gumbel distribution**, i.e., the GEV with $\xi = 0$, $\mu = 0$, $\sigma = 1$.

A key theoretical characterisation of extreme-value limit distributions is as follows:

Class of extreme-value limit distributions G = Class of max-stable distributions

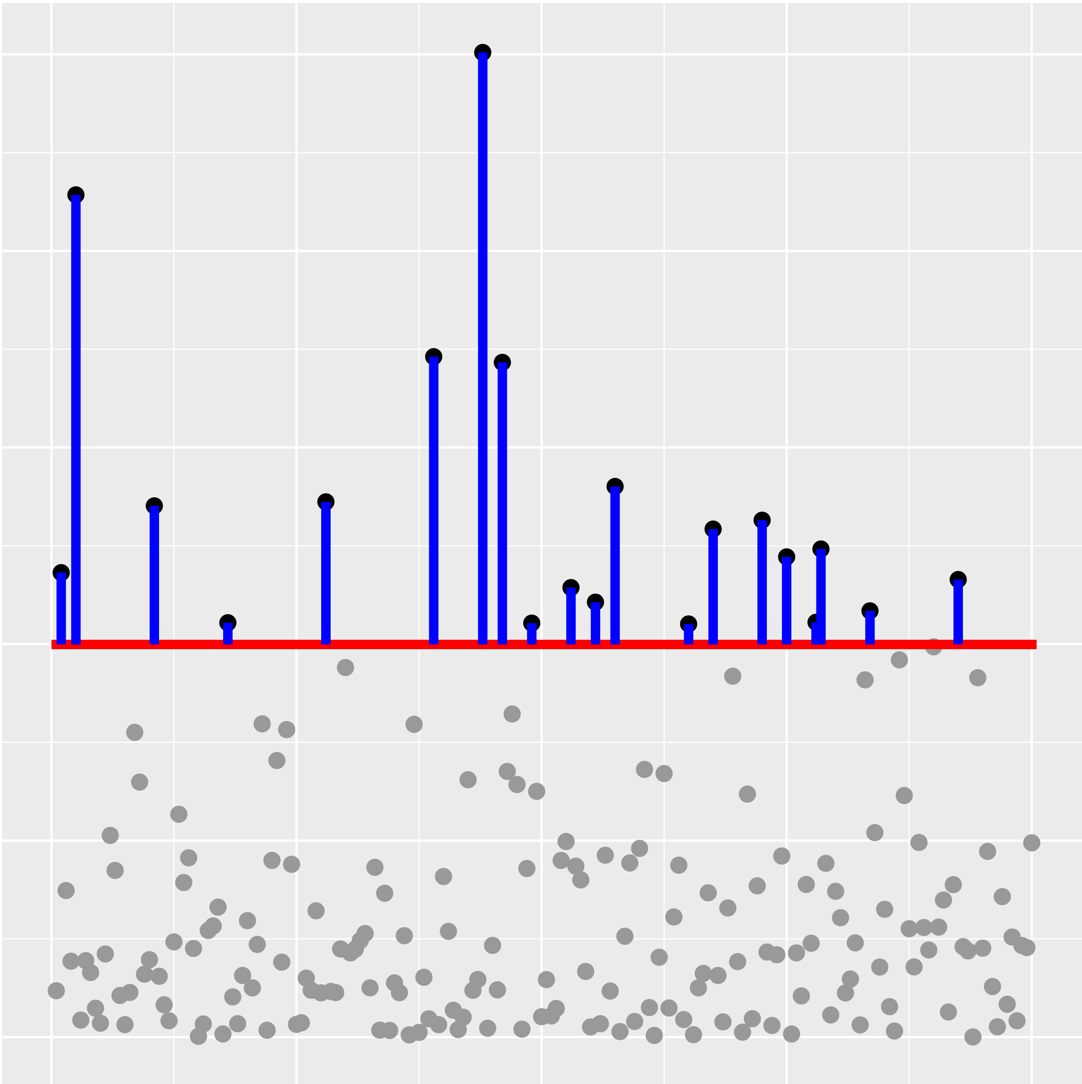
Max-stable distribution

A probability distribution G is called **max-stable** if for any $n \in \mathbb{N}$ there exist appropriate choices of deterministic normalizing sequences α_n and $\beta_n > 0$ such that

$$G^n(\alpha_n + \beta_n z) = G(z), \quad \text{for any} n \in \mathbb{N}.$$

This also means that the MDA limit ($\star$) is exact (and not asymptotic) if F is max-stable.

Threshold exceedances in a univariate sample



What are possible limits for threshold excesses

$$X - u \quad \text{given} \quad X > u \quad ?$$

Generalized Pareto limits for threshold exceedances

Consider iid $X, X_1, X_2, \ldots$ where $X \sim F$
with upper endpoint $x^\star = \sup\{x \in \mathbb{R} : F(x) < 1\} \in (-\infty, \infty]$.

Pickands–Balkema–de-Haan Theorem

Suppose that $M_n = \max(X_1, \ldots, X_n)$ converges to a $\mathrm{GEV}(\xi, \mu, \sigma)$ distribution according to the Fisher–Tippett–Gnedenko theorem.
Equivalently, there exists a scaling function $\sigma(u) > 0$ such that

$$(X - u)/\sigma(u) \mid (X > u) \quad \to \quad Y, \quad u \to x^\star,$$

and Y follows the **Generalized Pareto Distribution** $\mathrm{GPD}(\xi, \sigma_{GPD})$ given as

$$\mathrm{GPD}(y; \xi, \sigma_{GPD}) = \mathrm{Pr}(Y \leq y) = 1 - (1 + \xi y/\sigma_{GPD})_+^{-1/\xi} \quad y > 0,$$

with scale parameter $\sigma_{GPD} > 0$.

- This result dates back to the 1970s.
- As before, the case $\xi = 0$ is interpreted as the limit for $\xi \to 0$:

$$\mathrm{GPD}(y; 0, \sigma_{GPD}) = 1 - \exp(-y/\sigma_{GPD}), \quad y > 0$$

(= Exponential distribution).

Sketch of the proof

We here sketch the proof of "$\Rightarrow$"
(Convergence of maxima leads to convergence of threshold excesses).

1. Set $u_n = a_n + b_n\tilde{u}$ for $\tilde{u}$ chosen in the support of the $\mathrm{GEV}(\xi, \mu, \sigma)$. Then,

$$\Pr((X - u_n)/b_n > y \mid X > u_n) = \frac{1 - F(a_n + b_n(y + \tilde{u}))}{1 - F(a_n + b_n\tilde{u})}. \tag{4}$$

2. On the one hand, the MDA condition $F^n(a_n + b_n z) \to G(z)$ implies

$$\log F(a_n + b_n z) \approx \frac{1}{n} \log G(z), \quad \text{for large } n.$$

   On the other hand, since $F(a_n + b_n z) \approx 1$ as n increases, we can use the first-order approximation $\log(1 + x) \approx x$ for small $|x|$, such that

$$\log F(a_n + b_n z) \approx F(a_n + b_n z) - 1.$$

   Combining the two yields

$$1 - F(a_n + b_n z) \approx -\frac{1}{n} \log G(z). \tag{5}$$

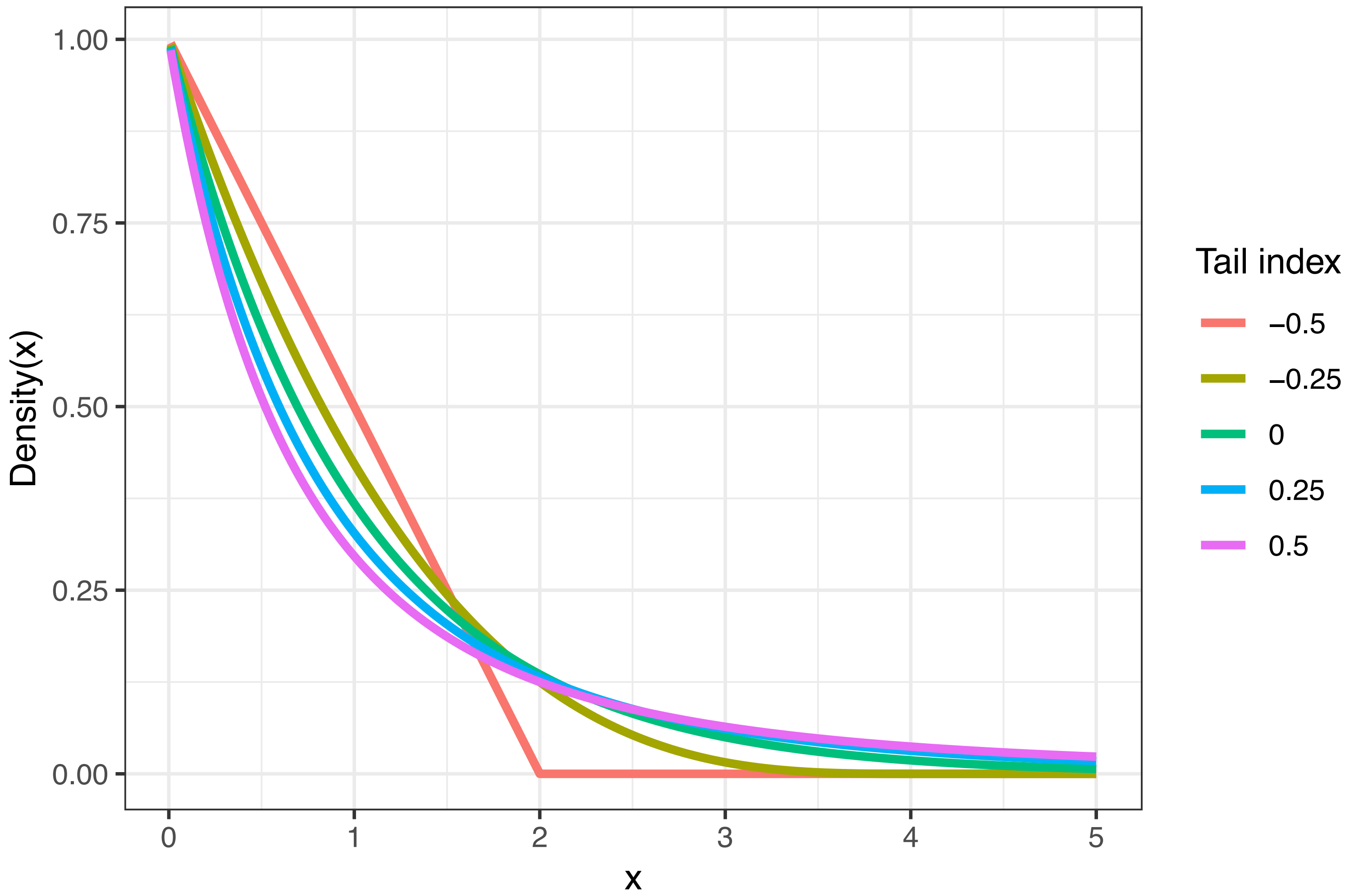
3. By using the approximation (5) for the numerator and denominator of (4), we get

$$\Pr((X - u_n)/b_n > y \mid X > u_n) \to \frac{\log G(\tilde{u} + y)}{\log G(\tilde{u})} = 1 - \mathrm{GPD}(y; \xi, \sigma_{GPD}), \quad n \to \infty;$$

   with $\sigma_{GPD} = \sigma + \xi(\tilde{u} - \mu) > 0$, and we can set $\sigma(u_n) = b_n$.

Illustration: GPD densities

The value of the tail index ξ characterizes the shape of the distribution.
Here, σ_{GPD} is fixed to 1.



Peaks-over-threshold stability

By analogy with **max-stability** of GEV limit distributions for maxima, we have **Peaks-Over-Threshold (POT) stability** for limit distributions of threshold exceedances.

Peaks-Over-Threshold stability of the GPD

Suppose that $Y \sim \mathrm{GPD}(\xi, \sigma_{GPD})$. Consider a new, higher threshold $\tilde{u} > 0$ such that $\mathrm{GPD}(\tilde{u}; \xi, \sigma_{GPD}) < 1$. Then

$$Y - \tilde{u} \mid (Y > \tilde{u}) \sim \mathrm{GPD}(\xi, \tilde{\sigma}_{GPD}), \quad \tilde{\sigma}_{GPD} = \sigma_{GPD} + \xi \tilde{u}.$$

Exercice: Prove this using pencil + paper by showing

$$\frac{1 - \mathrm{GPD}(\tilde{u} + y; \xi, \sigma_{GPD})}{1 - \mathrm{GPD}(\tilde{u}; \xi, \sigma_{GPD})} = 1 - \mathrm{GPD}(y; \xi, \tilde{\sigma}_{GPD})$$

$\Rightarrow$ Application of the POT approach to a GPD yields again a GPD!

For $\xi = 0$, where the GPD is the exponential distribution, the POT stability is also known as the **lack-of-memory property**.

Point-process convergence

The **trinity** of univariate extreme-value limits is completed by point patterns.

Theorem (Point-process convergence)

For i.i.d. copies $X_1, X_2, \dots$ of $X \sim F$, the following two statements are equivalent:

1. The distribution F is in the maximum domain of attraction of the max-stable distribution G with support $A_{\xi,\sigma,\mu}$ for the normalizing sequences $a_n \in \mathbb{R}$ and $b_n > 0$.
2. For the normalizing sequences $a_n \in \mathbb{R}$ and $b_n > 0$, we have the following point-process convergence with a locally finite Poisson-process process limit:

$$\left\{\left(\frac{i}{n}, \frac{X_i - a_n}{b_n}\right), \ i = 1, \dots, n\right\} \to \{(t_i, P_i), \ i \in \mathbb{N}\} \sim \mathrm{PPP}(\lambda_1 \times \Lambda), \quad n \to \infty,$$

with intensity measure $\lambda_1 \times \Lambda$ where λ_1 is the Lebesgue measure on $(0, 1)$.

If 1) and 2) hold, then $G(z) = \exp(-\Lambda[z; \infty))$, and the **exponent measure Λ** defined on $A_{\xi,\sigma,\mu}$ is characterized by its tail measure

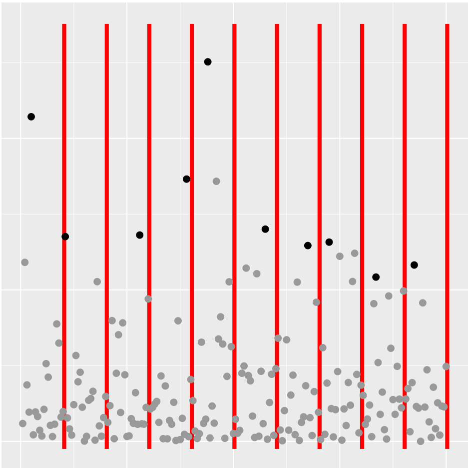
$$\Lambda[z, \infty) = -\log G(z) = \begin{cases} \left(1 + \xi \frac{z-\mu}{\sigma}\right)^{-1/\xi}, & \xi \neq 0 \\ \exp\left(\frac{z-\mu}{\sigma}\right), & \xi = 0 \end{cases}, \qquad \mu \in \mathbb{R}, \ \sigma > 0.$$

Remark: Λ is singular at $\inf A_{\xi,\sigma,\mu}$.

Summary: The extreme-value trinity

We allow for affine-linear rescaling $\tilde{X}_i = \frac{X_i - b_n}{a_n}$ of the iid sample X_i, $i = 1, \ldots, n$.

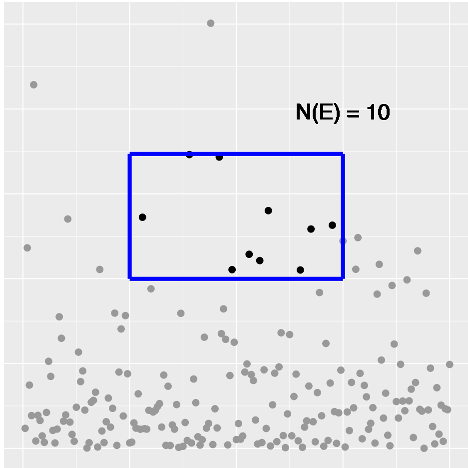
Maxima



$\Pr(\max_{i=1}^n \tilde{X}_i \le z)$
$\to \exp\left(-\Lambda[z, \infty)\right)$

Max-stable distr. (GEV)

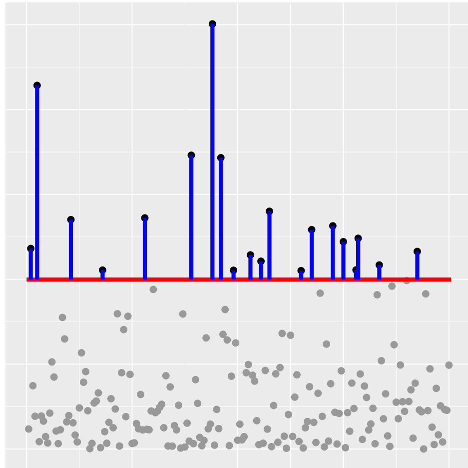
Occurrence counts



$\Pr(N(E) = k) \to$
$\exp(-(\lambda_1 \times \Lambda)(E)) \frac{(\lambda_1 \times \Lambda)(E)^k}{k!}$

Poisson process

Threshold exceedances



$\Pr(\tilde{X}_i - u > y \mid \tilde{X}_i > u)$
$\to \Lambda[y, \infty)/\Lambda[u, \infty)$

Gen. Pareto distr. (GPD)

Exponent measure Λ possessing asymptotic stability:
for any event E and $c > 0$, there are constants $\alpha(c) \in \mathbb{R}$, $\beta(c) > 0$ such that

$$c \times \Lambda(E) = \Lambda\left(\frac{E - \alpha(c)}{\beta(c)}\right)$$