Estimation of extreme risk, application to environmental data

ED-MSTII

Stéphane Girard
Inria Grenoble Rhône-Alpes
http://mistis.inrialpes.fr/people/girard/

2026

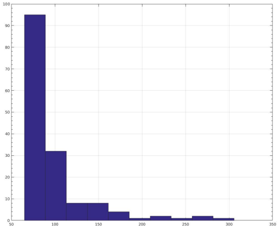
Outline

Three courses (September 28th, October 5th & 12th)

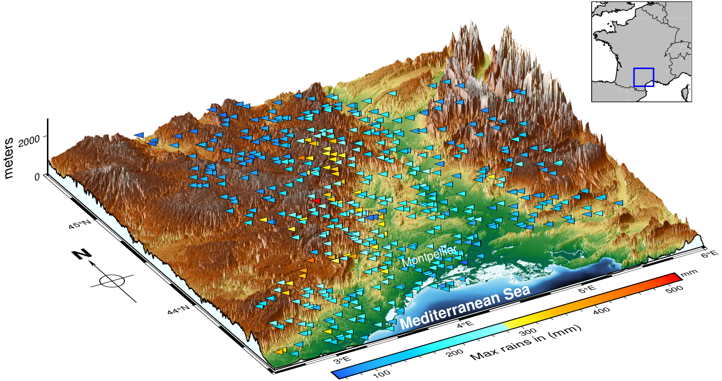
- Chapter 1: Motivations for extreme-value analysis
- Chapter 2: Introduction to extreme-value theory (probabilistic tools)
- Chapter 3: Application to the estimation of risk metrics (statistical aspects)
- Chapter 4: Measure of dependence with copulas

First (toy) example: Nidd river (England, North Yorkshire)

The flows of the Nidd river have been measured every three months during 38.5 years. This results in $n = 154$ measures, the maximum observed flow being about $300 m^3/s$.



Daily rainfall observations (in millimeters) from 1958 to 2000 in the Cévennes-Vivarais region (southern part of France, $\simeq 256 \times 283$ km^2) at 524 raingauge stations, leading to a dataset of size $n \simeq 524 \times 365 \times 43$ (missing data, NaN).



Pedagogical resources

Available on my webpage:
`https://mistis.inrialpes.fr/people/girard/dissemination`

- Slides and Python illustrations: `slides.pdf`, `illustration-chapter-1.ipynb` `illustration-chapter-2.ipynb`, `illustration-chapter-3.ipynb`, `illustration-chapter-4.ipynb`
- Data files: `nidd.d`, `cevennes_rains.csv`, `cevennes_stations.csv`, `temperature-normandie.csv`

References

- A nice introduction to EVT, (easy):
  Coles, S. (2001). *An introduction to statistical modeling of extreme values*, Springer.

- A good overview on EVT, (medium):
  Embrechts, P., Klüppelberg, C. and Mikosch, T. (1987). *Modelling Extremal Events for Insurance and Finance*, Springer.

- Full probabilistic and statistical aspects of EVT, (difficult):
  de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, Springer-Verlag.

- A lot of informations on copulas, (easy):
  Nelsen, R. (2006). *An Introduction to Copulas*, Springer.

- Links between EVT and point processes theory, (difficult):
  Resnick, S. (2008). *Extreme values, regular variation, and point processes*, Springer.

Chapter 1: Motivations for extreme-value analysis

Estimation of extreme risk, application to environmental data

Stephane Girard
Inria Grenoble Rhône-Alpes
http://mistis.inrialpes.fr/people/girard/

2026

Extreme events ...

... are rare but may have a high impact.

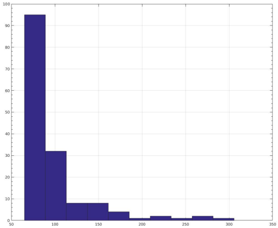
Left: Montpellier, France, 2014. Three hours of rainfall = 50% of mean annual rainfall (source: Dailymail).

Right: Texas, USA, 2025. Flow of the Guadalupe river increased from 2.7 to $4700 m^3/s$ (source: Radio Canada).

Goal: Compute risk measures associated with such extreme events.

Toy example: Nidd river (England, North Yorkshire)

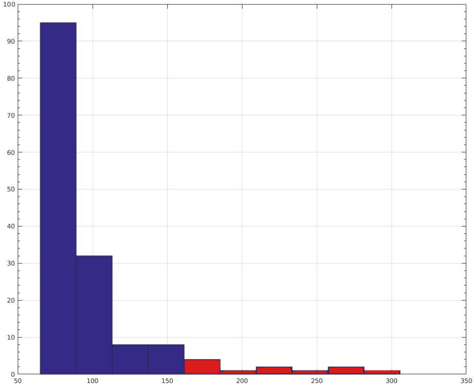
The flows of the Nidd river have been measured every three months during 38.5 years. This results in $n = 154$ measures, the maximum observed flow being about $300 m^3/s$.



Toy example: Nidd river (England, North Yorkshire)

From these data, it is possible to compute some exceedance probabilities such as:

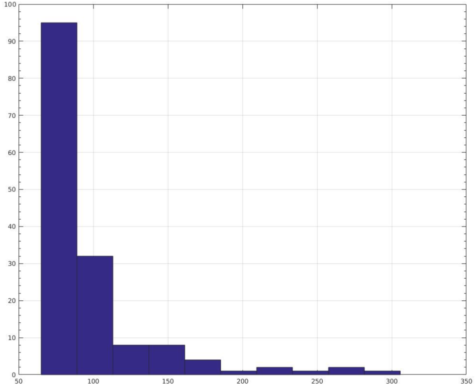
$$\mathbb{P}(X \geq 160) \simeq nb(X_i \geq 160)/n = 11/154 = 1/14$$



Toy example: Nidd river (England, North Yorkshire)

From these data, it is possible to compute some exceedance probabilities ... but not extreme ones!

$$\mathbb{P}(X > 500) \simeq nb(X_i > 500)/n = 0$$



The same problem occurs to estimate the 100-year river flow *i.e.* the flow t which is reached once every 100 years: $\mathbb{P}(X \geq t) = 1/400$ (quantile – return level – Value at Risk).

Statistical problem

Definition (Survival function)

$\bar{F}(x) = \mathbb{P}(X \geq x) = 1 - F(x)$ where F is the cumulative distribution function (cdf) associated with X.

Two problems: From $\{X_1, \ldots, X_n\}$ iid with cdf F,

1. **Estimation of small tail probabilites.**
   Given $x_n > X_{n,n} = \max(X_1, \ldots, X_n)$, estimate $p_n = \bar{F}(x_n)$.
2. **Estimation of extreme quantiles – return level – Value at Risk.**
   Given p_n such that $np_n \to c < \infty$ as $n \to \infty$, estimate $q(p_n)$ such that $p_n = \bar{F}(q(p_n))$.

Common difficulty: The survival function $\bar{F}(x)$ is unknown and difficult to estimate for $x > X_{n,n}$.

Remark. If $np_n \to c < \infty$ then

$$\mathbb{P}(q(p_n) > X_{n,n}) = \mathbb{P}^n(X_1 < q(p_n)) = F^n(q(p_n)) = (1 - p_n)^n$$
$$= \exp(n \log(1 - p_n)) \to e^{-c}.$$

Nonparametric approach

Empirical survival function: $\mathbb{P}(X \geq x)$ is estimated by the proportion of observations above x:

$$\hat{\bar{F}}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\{X_i \geq x\}$$

Problem: As illustrated on the Nidd river: $\hat{\bar{F}}_n(x) = 0$ if $x > X_{n,n}$.
From an asymptotic point of view: If $x_n \to \infty$, $\mathbb{E}(\hat{\bar{F}}_n(x_n)) = \bar{F}(x_n)$ and thus $\mathbb{E}(\hat{\bar{F}}_n(x_n)/\bar{F}(x_n)) = 1$ the estimator is still unbiased but

$$var(\hat{\bar{F}}_n(x_n)/\bar{F}(x_n)) = \frac{\bar{F}(x_n)F(x_n)}{n\bar{F}^2(x_n)} \sim \frac{1}{n\bar{F}(x_n)}.$$

The estimator is not convergent when $n\bar{F}(x_n) \to c > 0$.

Parametric approach

- Suppose *a priori* a parametric model for the survival function: $\bar{F} \in \{\bar{F}_\theta,\ \theta \in \Theta\}$.
- Estimate θ by $\hat{\theta}_n$.
- Use $\bar{F}_{\hat{\theta}_n}$ to estimate the risk measure.

Problem: A good fit on the sample does not necessarily lead to a good modelling above the maximum.

$\rightsquigarrow$ notebook1

Theorem of the Central Limit. Let $X_1, \dots, X_n$ be iid random variables with expectation μ and finite variance σ^2. Then,

$$\frac{\sqrt{n}}{\sigma}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right) \xrightarrow{d} \mathcal{N}(0,1).$$

Assume we have a similar result for the maxima, *i.e.*

$$\frac{1}{a_n}(\max(X_1, \dots, X_n) - b_n) \xrightarrow{d} Z,$$

where Z is a random variable with cdf H.

It would follow

$$\mathbb{P}\left(\frac{1}{a_n}(\max(X_1,\dots,X_n)-b_n)\leq x\right)\simeq\mathbb{P}(Z\leq x),$$

or equivalently,

$$\mathbb{P}\left(\max(X_1,\dots,X_n)\leq a_n x+b_n\right)\simeq\mathbb{P}(Z\leq x).$$

Letting $t=a_n x+b_n$ we obtain

$$\mathbb{P}\left(\max(X_1,\dots,X_n)\leq t\right)\simeq\mathbb{P}\left(Z\leq(t-b_n)/a_n\right).$$

Since $\mathbb{P}(\max(X_1,\dots,X_n)\leq t)=\mathbb{P}^n(X\leq t)$, it follows that

$$\mathbb{P}\left(X\leq t\right)\simeq\mathbb{P}^{1/n}\left(Z\leq(t-b_n)/a_n\right),$$

which can be rewritten as

$$1-\mathbb{P}\left(X\geq t\right)\simeq H^{1/n}\left((t-b_n)/a_n\right).$$

Taking the logarithm yields

$$\log(1 - \mathbb{P}(X \geq t)) \simeq \frac{1}{n} \log H\left((t - b_n)/a_n\right).$$

Since t is large, $\mathbb{P}(X \geq t)$ is small, a first order Taylor expansion of $\log(1 + u)$ thus yields

$$\mathbb{P}(X \geq t) \simeq -\frac{1}{n} \log H\left((t - b_n)/a_n\right)$$

for n large. Such an approximation would yield estimators for small tail probabilites and for extreme quantiles:

$$q(p_n) \simeq b_n + a_n H^{-1}(\exp(-np_n)).$$

The goal of the next chapter is to identify the limit distribution H and the normalizing constants a_n and b_n.

"The key message is that EVT cannot do magic – but it can do a whole lot better than empirical curve-fitting and guesswork. My answer to the sceptics is that if people aren't given well-founded methods like EVT, they'll just use dubious ones instead." – Jonathan Tawn

Chapter 2: Introduction to extreme-value theory

Estimation of extreme risk, application to environmental data

Stéphane Girard
Inria Grenoble Rhône-Alpes
http://mistis.inrialpes.fr/people/girard/

2026

1. Convergence of the maximum
2. Caracterisation of Maximum Domains of Attractions
3. Convergence of excesses

Let $X_1, \ldots, X_n$ be n i.i.d. random variables with cumulative distribution function (cdf) F. The maximum is denoted by $X_{n,n} = \max(X_1, \ldots, X_n)$. The cdf of $X_{n,n}$ is given by $\mathbb{P}(X_{n,n} \leq x) = F^n(x)$.

Definition 1

F belongs to the **maximum domain of attraction** of a cdf H if and only if there exist two sequences $a_n > 0$ and b_n such that

$$F^n(a_n x + b_n) \to H(x) \text{ as } n \to \infty \text{ for all } x \in \mathbb{R}.$$

Equivalently, $(X_{n,n} - b_n)/a_n \xrightarrow{d} Y$ where Y has cdf H.

This property is denoted for short by $F \in \text{MDA}(H)$. Except some pathological cases, every cdf F belongs to a maximum domain of attraction.

Type

Definition 2

Let G and H be two cdf. G and H are of **same type** if and only if there exist $a > 0$ and $b \in \mathbb{R}$ such that $H(x) = G(ax + b)$ for all $x \in \mathbb{R}$.

In there words, G and H are equal up to location and scale parameters.
For instance, all Gaussian distributions are of the same type.
The next proposition shows that the maximum domain of attraction is unique up to position and scale parameters.

Proposition 1

- *If $F \in MDA(G)$ and $F \in MDA(H)$ then, necessarily, G and H are of the same type.*
- *More precisely, if $a_n > 0$, $u_n > 0$, b_n and v_n are such that $F^n(a_n x + b_n) \to G(x)$ and $F^n(u_n x + v_n) \to H(x)$ as $n \to \infty$ for all $x \in \mathbb{R}$, then, necessarily, $u_n/a_n \to a$ and $(v_n - b_n)/a_n \to b$ as $n \to \infty$.*

Extreme-Value Theorem

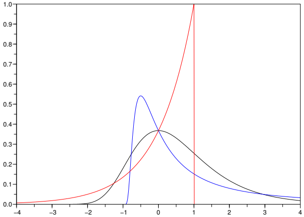
Theorem 1 (Fisher–Tippett–Gnedenko, 1928)

If $F \in MDA(G)$ then, necessarily, G is of the same type as the cdf

$$H_\gamma(x) = \begin{cases} \exp[-(1+\gamma x)_+^{-1/\gamma}] & \text{if } \gamma \neq 0, \\ \exp(-e^{-x}) \text{ if } \gamma = 0, \end{cases}$$

where $y_+ = \max(0, y)$.

H_γ is referred to as the cdf of the Extreme Value Distribution (EVD).
$\gamma \in \mathbb{R}$ is called the extreme-value index.



Example of densities associated with H_γ (black: $\gamma = 0$, blue: $\gamma = 1$ and red: $\gamma = -1$) $\rightsquigarrow$ notebook2#1

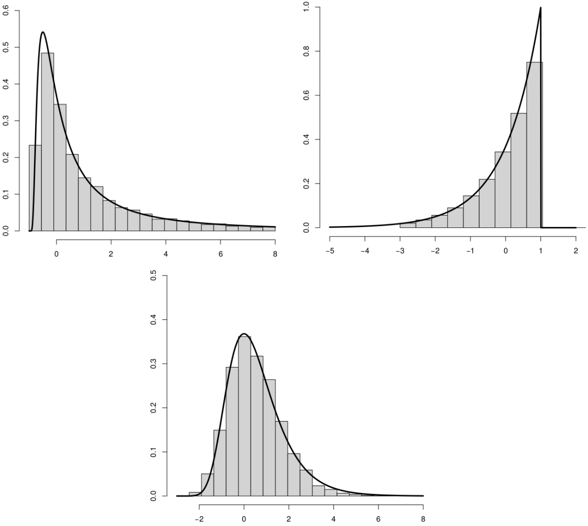
Following the sign of γ, three maximum domains of attraction are defined:

- If $\gamma > 0$, F is said to belong to the Fréchet maximum domain of attraction, $F \in \text{MDA(Fréchet)}$. It will be shown later that this domain of attraction includes distributions with heavy tails, *i.e.* their survival distribution functions decrease as a power function.
- If $\gamma = 0$, F is said to belong to the Gumbel maximum domain of attraction, $F \in \text{MDA(Gumbel)}$. This domain of attraction includes distributions with light tails, *i.e.* their survival distribution functions decrease as an exponential rate.
- If $\gamma < 0$, F is said to belong to the Weibull maximum of attraction, $F \in \text{MDA(Weibull)}$. This domain of attraction includes distributions with short tails, *i.e.* they have a finite endpoint $x_F = \inf\{x,\ F(x) \geq 1\}$.

A classification of numerous distributions by maximum domain of attraction is available in [3, Tables 3.4.2-3.4.4].

Maximum Domain of attraction	**Gumbel** $\gamma = 0$	**Fréchet** $\gamma > 0$	**Weibull** $\gamma < 0$
Distribution	Gaussian Exponential Lognormal Gamma Weibull	Cauchy Pareto Student Burr Fisher	Uniform Beta

Maximum domains of attraction associated with usual distributions.



Grey: Histogram of normalized maxima computed on 10^4 samples of size $n = 100$. Black: density of the EVD.

Left: Cauchy and $\gamma = 1$. Right: Beta$(1,1)$ and $\gamma = -1$. Bottom: standard Gaussian and $\gamma = 0$ $\rightsquigarrow$ notebook2#2

Proof of the Extreme-Value Theorem. Since $F \in MDA(G)$, there exist two sequences $a_n > 0$ and b_n such that

$$F^{[nt]}(a_{[nt]}x + b_{[nt]}) \to G(x) \tag{1}$$

as $n \to \infty$ for all $t > 0$ and $x \in \mathbb{R}$. Moreover,

$$F^{[nt]}(a_n x + b_n) = (F^n(a_n x + b_n))^{[nt]/n} \to G^t(x) \tag{2}$$

since $[nt]/n \to t$ as $n \to \infty$. In view of (1) and (2), Proposition 1 implies that

$$\frac{a_n}{a_{[nt]}} \to \alpha(t) > 0, \quad \frac{b_n - b_{[nt]}}{a_{[nt]}} \to \beta(t)$$

and that G and G^t are of the same type:

$$G^t(x) = G(\alpha(t)x + \beta(t)). \tag{3}$$

The proof consists in solving equation (3) with respect to G.

For all $t > 0$ and $s > 0$, the quantity $G^{st}(x)$ can be rewritten in three different ways:

$$G^{st}(x) = G(\alpha(st)x + \beta(st)) \tag{4}$$
$$= G^{s}(\alpha(t)x + \beta(t)) = G(\alpha(s)\alpha(t)x + \alpha(s)\beta(t) + \beta(s)) \tag{5}$$
$$= G^{t}(\alpha(s)x + \beta(s)) = G(\alpha(s)\alpha(t)x + \alpha(t)\beta(s) + \beta(t)). \tag{6}$$

Lemma 1

Let F be a non-degenerated cdf. If there exist $a > 0$, $c > 0$, $b \in \mathbb{R}$ and $d \in \mathbb{R}$ such that $F(ax + b) = F(cx + d)$ for all $x \in \mathbb{R}$ then, necessarily $a = c$ and $b = d$.

From (4), (5), (6) and in view of the above lemma, it follows that

$$\alpha(st) = \alpha(s)\alpha(t) \tag{7}$$
$$\beta(st) = \alpha(s)\beta(t) + \beta(s) \tag{8}$$
$$\beta(st) = \alpha(t)\beta(s) + \beta(t), \tag{9}$$

for all $s > 0$ and $t > 0$. It can be shown that the unique positive solution of equation (7) is $\alpha(t) = t^A$, for all $t > 0$ where $A \in \mathbb{R}$. Three cases appear.

1. If $A = 0$ then $\alpha(t) = 1$ for all $t > 0$ and equations (8) and (9) can be rewritten as

$$\beta(st) = \beta(s) + \beta(t). \tag{10}$$

It can be shown that the unique solution of this equation is $\beta(t) = \beta(e) \log t$, for all $t > 0$. Replacing in equation (3) yields the new equation on G:

$$G^t(x) = G(x + \beta(e) \log t),$$

for all $x \in \mathbb{R}$ and $t > 0$. Besides, since G is non-degenerated, there exists $x_0 \in \mathbb{R}$ such that $0 < G(x_0) < 1$. Thus, for all $t > 1$,

$$G(x_0) > G^t(x_0) = G(x_0 + \beta(e) \log t).$$

This implies that $\beta(e) < 0$. In the following, we thus note $\sigma = -\beta(e) > 0$.

1. *(continued).* Let us now prove that $0 < G(x) < 1$ for all $x \in \mathbb{R}$. To this end, assume there exists $x_0 \in \mathbb{R}$ such that $G(x_0) = 1$. It follows that $G^t(x_0) = 1$ for all $t > 0$ and therefore $G(x_0 - \sigma \log t) = 1$ for all $t > 0$. This implies that G is constant equal to 1, which is not possible. As a consequence, $G(x) < 1$ for all $x \in \mathbb{R}$ and a similar proof can be done for $G(x) > 0$ for all $x > 0$. Let us thus introduce $\theta = -\log G(0) > 0$ and recall that

$$G^t(x) = G(x - \sigma \log t)$$

for all $x \in \mathbb{R}$ and $t > 0$. In particular, for $x = 0$, we have

$$\exp(-\theta t) = G^t(0) = G(-\sigma \log t).$$

The change of variable $u = -\sigma \log t$ thus yields

$$G(u) = \exp(-\theta \exp(-u/\sigma))$$

for all $u \in \mathbb{R}$. This shows that G is of the same type as H_0 *i.e.*

$$G(x) = H_0(x/\sigma + \log \theta).$$

2. If $A < 0$, let us introduce $\gamma = -A > 0$ leading to $\alpha(t) = t^{-\gamma}$ for all $t > 0$. Equations (8) and (9) imply that

$$\alpha(s)\beta(t) + \beta(s) = \alpha(t)\beta(s) + \beta(t)$$

for all $s > 0$ and $t > 0$ or equivalently,

$$(1 - \alpha(s))\beta(t) = (1 - \alpha(t))\beta(s).$$

Excluding the value $t = s = 1$, we thus have

$$\frac{\beta(t)}{1 - \alpha(t)} = \frac{\beta(s)}{1 - \alpha(s)} = c.$$

It follows that $\beta(t) = c(1 - \alpha(t)) = c(1 - t^{-\gamma})$ where $c \in \mathbb{R}$ and $t > 0$. Replacing in equation (3) yields the new equation

$$G^t(x) = G(t^{-\gamma}(x - c) + c)$$

for all $x \in \mathbb{R}$ and $t > 0$.

2. *(continued)*. Letting $y = x - c$ and G_* the translated cdf defined by $G_*(y) = G(y + c)$, we end up with the equation

$$G_*^t(y) = G_*(t^{-\gamma} y)$$

for all $t > 0$ and $y \in \mathbb{R}$. Moreover, since G_* and G are of the same type, it is sufficient to solve the above equation. Let $y_0 \in \mathbb{R}$ such that $G_*(y_0) < 1$. Then, letting $t \to \infty$ in the previous equation yields $G_*(0) = 0$. It follows that $G_*(y) = 0$ for all $y \leq 0$. A proof similar to the case $A = 0$ can be done to prove that $0 < G_*(1) < 1$ and we thus introduce $\sigma = -\log G_*(1) > 0$. Considering $y = 1$ in the functional equation yields

$$G_*(t^{-\gamma}) = G_*^t(1) = \exp(-\sigma t)$$

for all $t > 0$. As a consequence, the change of variable $u = t^{-\gamma}$ entails

$$G_*(u) = \exp(-\sigma u^{-1/\gamma})$$

for all $u > 0$. It is easily seen that G_* is of the same type as H_γ *i.e.*

$$G_*(x) = H_\gamma(x\gamma^{-1}\sigma^{-\gamma} - \gamma^{-1})$$

for all $x \in \mathbb{R}$.

3. The case $A > 0$ is similar. ∎

Exercise 1 (Standard exponential distribution)

$F(x) = 1 - \exp(-x)$, $x > 0$. *Extreme-value index* $\gamma = 0$ *(Gumbel Maximum Domain of Attraction), sequences* $a_n = 1$ *and* $b_n = \log n$.

Exercise 2 (Pareto distribution)

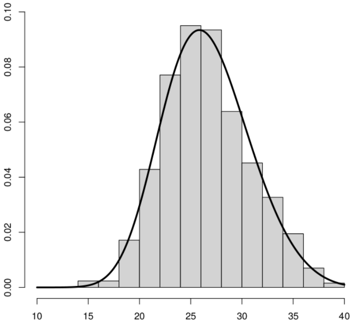
$F(x) = 1 - (x/a)^{-\alpha}$, $x > a > 0$ *and* $\alpha > 0$. *Extreme-value index* $\gamma = 1/\alpha > 0$ *(Fréchet Maximum Domain of Attraction), sequences* $a_n = a\alpha^{-1}n^{1/\alpha}$ *and* $b_n = an^{1/\alpha}$.

Exercise 3 (Standard uniform distribution)

$F(x) = x$ *if* $x \in [0,1]$, $F(x) = 1$ *if* $x > 1$, $F(x) = 0$ *otherwise. Extreme-value index* $\gamma = -1 < 0$ *(Weibull Maximum Domain of Attraction), sequences* $a_n = 1/n$ *and* $b_n = 1 - 1/n$.

In practice ...

The EVD is a natural model for phenomena whose observations can be interpreted as maxima. As an example: daily maximum temperatures observed in Normandie (France) during the summer months of 2016–2022. The EVD density (fitted by maximum likelihood, see next chapter) is compared with the histogram of maximum temperatures.
$\rightsquigarrow$ notebook2#3



Regularly-varying functions

The caracterisation of maximum domains of attractions relies on the theory of regularly-varying functions [1].

Definition 3

An eventually positive function U is **regularly-varying** with index $\delta \in \mathbb{R}$ at infinity if

$$\lim_{x \to \infty} \frac{U(\lambda x)}{U(x)} = \lambda^{\delta},$$

for all $\lambda > 0$. This property is denoted by $U \in \mathcal{RV}_{\delta}$. If $\delta = 0$, the function U is said to be **slowly-varying**.

$U \in \mathcal{RV}_{\delta}$ if and only if there exists $L \in \mathcal{RV}_0$ such that $U(x) = x^{\delta} L(x)$.

Example 4

All functions tending to a strictly positive limit are slowly-varying. Every power function of the logarithm is also slowly-varying.

As a consequence, every function in $\mathcal{RV}_{\delta}$ qualitatively behaves as $x \to x^{\delta}$ when x is large.

Fréchet Maximum Domain of Attraction

Theorem 2

- *F belongs to MDA(Fréchet) if and only if $\bar{F} := 1 - F$ is regularly-varying with index $-1/\gamma$. The associated extreme-value index is γ.*
- *Moreover, a possible choice for the normalizing sequences is $a_n = F^{-1}(1 - 1/n) = q(1/n)$ and $b_n = 0$.*

– Necessarily the endpoint of F is infinite.
– Heavy-tailed distributions.

Weibull Maximum Domain of Attraction

– Finite endpoint.
– One to one relationship with Fréchet Maximum Domain of Attraction.

For all cdf F with finite endpoint x_F, we denote by $F_\star$ the cdf defined by $F_\star(x) = F(x_F - 1/x)$ if $x > 0$ and $F_\star(x) = 0$ otherwise.

Theorem 3

- *F belongs to MDA(Weibull) if and only if x_F is finite and $F_\star$ belongs to MDA(Fréchet). Letting $\gamma > 0$ the extreme-value index associated with $F_\star$, the extreme-value index associated with F is $-\gamma$.*
- *Besides, a possible choice of normalizing sequences is $a_n = x_F - F^{-1}(1 - 1/n) = x_F - q(1/n)$ and $b_n = x_F$.*

Gumbel Maximum Domain of Attraction

– No simple characterisation result.

Theorem 4

F belongs to MDA(Gumbel) if and only if there exists $x_0 < x_F \leq \infty$ such that

$$\bar{F}(x) = c(x) \exp\left(-\int_{x_0}^{x} \frac{g(t)}{a(t)} dt\right),$$

where a, c and g are three functions verifying $a'(x) \to 0$, $c(x) \to c > 0$ and $g(x) \to 1$ as $x \to x_F$.

Exercise 4 (Standard Gaussian distribution)

The density is given by

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right),$$

and the survival function verifies $\bar{F}(x) \sim f(x)/x$ as $x \to \infty$. Check that $\mathcal{N}(0,1) \in$ MDA(Gumbel).

Solution. Letting,

- $x_0 > 0$,
- $\lambda(x) = x\bar{F}(x)/f(x) \to 1$ as $x \to \infty$,
- $k(x_0) = x_0 \exp(x_0^2/2)$,
- $g(x) = 1$,
- $a(x) = x/(x^2+1)$,
- $c(x) = \lambda(x)/(\sqrt{2\pi}k(x_0))$,

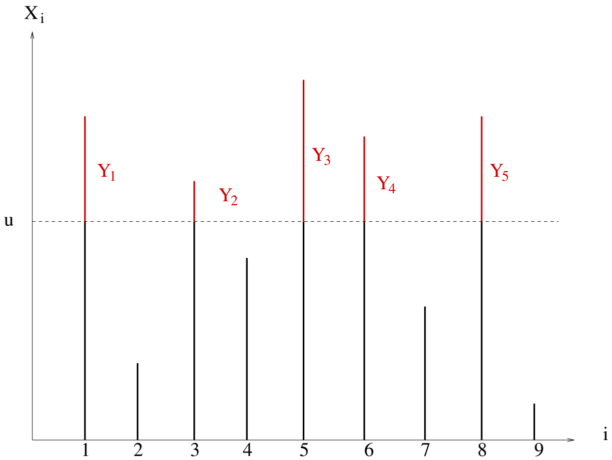
we end up with

$$\bar{F}(x) = c(x) \exp\left(-\int_{x_0}^{x} \frac{g(t)}{a(t)} dt\right).$$

Excesses

Definition 5

let X be random variable with cdf F and endpoint x_F. For all $u < x_F$ and $x \geq 0$, the cdf F_u defined by $F_u(x) = \mathbb{P}(X - u \leq x | X > u)$ is called the cdf of the excess of X above u.



Generalized Pareto Distribution

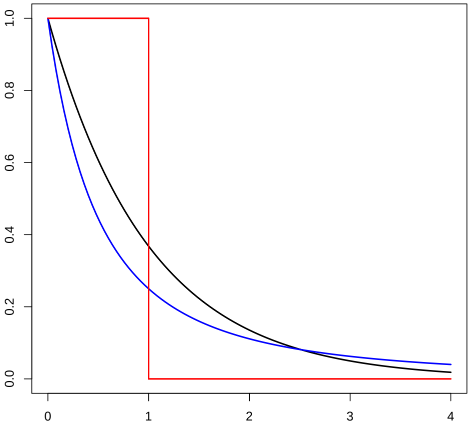
Definition 6

The Generalized Pareto Distribution (GPD) with parameters $\gamma \in \mathbb{R}$ and $\sigma > 0$ is defined by the cdf

$$
\begin{aligned}
G_{\gamma,\sigma}(y) &= 1 - \left(1 + \gamma \frac{y}{\sigma}\right)_+^{-1/\gamma} \quad \text{if } \gamma \neq 0, \\
&= 1 - \exp\left(-\frac{y}{\sigma}\right) \quad \text{otherwise.}
\end{aligned}
$$

Three particular cases:

- If $\gamma = 0$, then $G_{0,\sigma}$ is the cdf of $\mathrm{Exp}(\sigma)$,
- If $\gamma = -1$, then $G_{-1,\sigma}$ is the cdf of $\mathrm{Unif}[0, \sigma]$,
- If $\gamma > 0$, then $G_{\gamma,\sigma}$ is the cdf of a shifted Pareto distribution.



Example of densities associated with the Generalized Pareto Distribution
(black: $\gamma = 0$, blue: $\gamma = 1$ and red: $\gamma = -1$).

The distribution of the maximum can be approximated by an EVD if and only if the distribution of the excess can be approximated by a GPD. The shape parameters are the same.

Theorem 5 (Pickands theorem)

$F \in MDA(H_\gamma)$ if and only if there exists a positive function σ such that

$$\lim_{u \to x_F} \sup_{y \in [0, x_F - u]} \left| \bar{F}_u(y) - \bar{G}_{\gamma, \sigma(u)}(y) \right| = 0.$$

Exercise 5 (Standard exponential distribution)

$F(x) = 1 - \exp(-x)$, $x > 0$. *Extreme-value index* $\gamma = 0$ *(Gumbel Maximum Domain of Attraction) and* $\sigma(u) = 1$.

Exercise 6 (Pareto distribution)

$F(x) = 1 - (x/a)^{-\alpha}$, $x > a > 0$ *and* $\alpha > 0$. *Extreme-value index* $\gamma = 1/\alpha > 0$ *(Fréchet Maximum Domain of Attraction) and* $\sigma(u) = u/\alpha$.

Exercise 7 (Standard uniform distribution)

$F(x) = x$ *if* $x \in [0, 1]$, $F(x) = 1$ *if* $x > 1$, $F(x) = 0$ *otherwise. Extreme-value index* $\gamma = -1 < 0$ *(Weibull Maximum Domain of Attraction) and* $\sigma(u) = 1 - u$.

References

- Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation*, Cambridge University Press.
- de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, Springer-Verlag, New York.
- Embrechts, P., Klüppelberg, C. and Mikosch, T. (1987). *Modelling Extremal Events for Insurance and Finance*, Springer.
- Resnick, S. (2008). *Extreme values, regular variation, and point processes*, Springer.

Chapter 3: Application of EVT to the estimation of risk metrics

Estimation of extreme risk, application to environmental data

Stéphane Girard
Inria Grenoble Rhône-Alpes
http://mistis.inrialpes.fr/people/girard/

2026

As a consequence of the approximation derived in Chapter 1, Slide 11 and the extreme-value theorem (Theorem 1, Chapter 2), one has an approximation of **small tail probabilities**:

$$\begin{aligned}\mathbb{P}(X > x_n) = \bar{F}(x_n) &\simeq -\frac{1}{n}\log H_\gamma\left((x_n - b_n)/a_n\right)\\ &= \frac{1}{n}\left[1+\gamma\left(\frac{x_n - b_n}{a_n}\right)\right]^{-1/\gamma} \quad \text{if} \quad \gamma \neq 0\end{aligned}$$

when $x_n \to \infty$ as $n \to \infty$ and **extreme quantiles – VaR – return levels**:

$$\begin{aligned}q(p_n) = \bar{F}^{-1}(p_n) &\simeq b_n + a_n H_\gamma^{-1}(\exp(-np_n))\\ &= b_n + \frac{a_n}{\gamma}\left[(np_n)^{-\gamma} - 1\right] \quad \text{if} \quad \gamma \neq 0.\end{aligned}$$

when $p_n \to 0$ as $n \to \infty$.
(n is the size of the sample from which is taken the maximum)

Approximation of the Expected Shortfall

The **Expected Shortfall** of level p_n is defined as

$$\mathrm{ES}(p_n) = \mathbb{E}(X \mid X > q(p_n)) = \frac{1}{p_n}\int_0^{p_n} q(t)\mathrm{d}t.$$

Using the previous approximation of $q(t)$ as $t \to 0$, one gets

$$\mathrm{ES}(p_n) \quad \simeq \quad b_n + \frac{a_n}{\gamma}\left[\frac{(np_n)^{-\gamma}}{1-\gamma} - 1\right] \ \text{ if } \ \gamma \neq 0 \text{ and } \gamma < 1.$$

when $p_n \to 0$ as $n \to \infty$.

Inference (with the EVD approach)

In practice, a_n**,** b_n **and** γ **are unknown** (since F is unknown) and have to be estimated. The cdf of interest is

$$H_{\gamma,a,b}(x) \stackrel{def}{=} H_\gamma\left(\frac{x-b}{a}\right) = \exp\left\{-\left[1+\gamma\left(\frac{x-b}{a}\right)\right]_+^{-1/\gamma}\right\}.$$

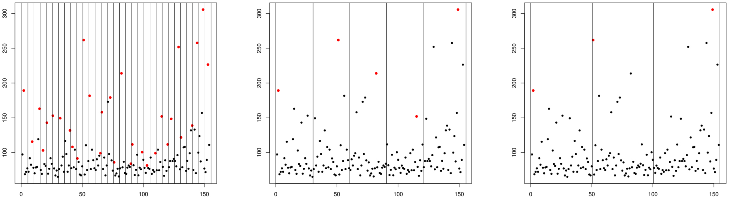
In the following, two estimators are considered:

- Maximum Likelihood Estimators (MLE),
- Probability Weighted Moments (PWM).

Let $\{Y_1,\ldots,Y_k\}$ be a sample of k iid random variables from $H_{\gamma,a,b}$. This starting point may be a problem since, usually, one does not directly observe a sample of maxima. They have to be extracted from the initial sample $\{X_1,\ldots,X_n\}$. The common practice is to divide the data into k blocks and to extract the maxima of each block (**block maxima method**). Two drawbacks:

- Loss of information (n data $\longrightarrow$ k data),
- The maxima are not exactly distributed from an EVD distribution.

Illustration on Nidd data



Block Maxima selection. Three block sizes k are considered. The associated block-maxima are depicted in red.

Inference (with the EVD approach)

– Estimation of **small tail probabilities**:

$$\hat{\bar{F}}(x_n) = \frac{k}{n}\left[1 + \hat{\gamma}\left(\frac{x_n - \hat{b}}{\hat{a}}\right)\right]^{-1/\hat{\gamma}}.$$

– Estimation of **extreme quantiles – VaR – return levels**:

$$\hat{q}(p_n) = \hat{b} + \frac{\hat{a}}{\hat{\gamma}}\left[\left(\frac{np_n}{k}\right)^{-\hat{\gamma}} - 1\right].$$

– Estimation of **Expected Shortfall**:

$$\widehat{\mathrm{ES}}(p_n) = \hat{b} + \frac{\hat{a}}{\hat{\gamma}}\left[\frac{1}{1-\hat{\gamma}}\left(\frac{np_n}{k}\right)^{-\hat{\gamma}} - 1\right].$$

(keeping in mind that maxima are extracted from samples of size n/k.)

The negative log-likelihood is given by

$$\ell(a,b,\gamma) = k\log a + \left(1+\frac{1}{\gamma}\right)\sum_{i=1}^{k}\log\left(1+\gamma\frac{Y_i-b}{a}\right) + \sum_{i=1}^{k}\left(1+\gamma\frac{Y_i-b}{a}\right)^{-1/\gamma}$$

if (a,b,γ) is such that $1+\gamma(Y_i-b)/a > 0$ for all $i=1,\ldots,k$ and $\ell(a,b,\gamma) = +\infty$ otherwise.

- The MLE is not explicit, its computation requires a numerical optimization procedure which may fail on small datasets.
- Since the support of the GEV depends on its parameters, the asymptotic normality of the MLE does not hold in the general case but a least for $\gamma > -1/2$ [18]. In this case, the Fisher information matrix is provided in [17] which allows to build asymptotic confidence intervals. Recently, [5] proved the consistency under the weaker assumption $F \in \mathrm{MDA}(H_\gamma)$, $\gamma > -1$.

In some context, one may assume that $\gamma = 0$ and fit a Gumbel distribution. The negative log-likelihood is obtained by letting $\gamma \to 0$ in the previous equation:

$$\ell(a, b, 0) = k \log a + \sum_{i=1}^{k} \left(\frac{Y_i - b}{a} \right) + \sum_{i=1}^{k} \exp \left[- \left(\frac{Y_i - b}{a} \right) \right].$$

- The MLE is still not explicit.
- Since the support of the Gumbel distribution does not depend on its parameters, the asymptotic normality automatically holds.

$\rightsquigarrow$ notebook3#1

Recall that $\{Y_1, \ldots, Y_k\}$ is a set of k iid random variables with cdf $H_{\gamma,a,b}$. The Probability Weighted Moment of order $r \geq 0$ is defined by [14]:

$$\mu_r = \mathbb{E}\left[Y H^r_{\gamma,a,b}(Y) \right].$$

Lemma 1

The PWM of order $r \geq 0$ exists provided that $\gamma < 1$ and is given by

$$\mu_r = \frac{1}{r+1}\left[b - \frac{a}{\gamma}\left\{ 1 - (r+1)^{\gamma}\Gamma(1-\gamma) \right\} \right],$$

where Γ is the Gamma function:

$$\Gamma(t) = \int_0^{+\infty} x^{t-1}\exp(-x)dx, \ t > 0.$$

To obtain a, b and γ, three PWMs are sufficient:

$$
\begin{aligned}
\mu_0 &= b - \frac{a}{\gamma}\left\{1 - \Gamma(1-\gamma)\right\} &(1)\\
2\mu_1 - \mu_0 &= -\frac{a}{\gamma}(1-2^\gamma)\Gamma(1-\gamma) &(2)\\
\frac{3\mu_2 - \mu_0}{2\mu_1 - \mu_0} &= \frac{1-3^\gamma}{1-2^\gamma}. &(3)
\end{aligned}
$$

From (3), it is possible to obtain γ with a numerical procedure. Then, a can be deduced from (2) and b can be deduced from (1). It remains to estimate μ_0, μ_1 and μ_2 to obtain estimators for a, b and γ.

To estimate μ_r, the idea is as follows. First, the mathematical expectation is replaced by the empirical mean:

$$\mu_r \simeq \frac{1}{k}\sum_{i=1}^{k} Y_i H^r_{\gamma,a,b}(Y_i) = \frac{1}{k}\sum_{i=1}^{k} Y_{i,k} H^r_{\gamma,a,b}(Y_{i,k}).$$

Second, $H_{\gamma,a,b}$ is replaced by the empirical cdf:

$$\mu_r \simeq \frac{1}{k}\sum_{i=1}^{k} Y_{i,k} \hat{F}^r_k(Y_{i,k}) = \frac{1}{k}\sum_{i=1}^{k} Y_{i,k}\left(\frac{i-1}{k}\right)^r.$$

One thus obtain

$$\hat{\mu}_r = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i-1}{k}\right)^r Y_{i,k}.$$

One may also use the alternative estimator:

$$\tilde{\mu}_r = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{(i-1)\cdots(i-r)}{(k-1)\cdots(k-r)}\right) Y_{i,k}.$$

The joint asymptotic normality of $(\hat{a},\hat{b},\hat{\gamma})$ is established in [9] under the assumption $F \in \mathrm{MDA}(H_\gamma)$, $\gamma < 1/2$.

1. Application of the Extreme-Value Theorem
2. Application of Pickands theorem
3. Semi-parametric approach in MDA(Fréchet)
4. Semi-parametric estimation in MDA(Weibull)
5. Semi-parametric estimation in MDA(Gumbel)

From Pickands theorem (Theorem 5, Chapter 2), one has, for all $y \geq 0$,

$$\bar{F}_u(y) = \frac{\bar{F}(u+y)}{\bar{F}(u)} \simeq \bar{G}_{\gamma,\sigma}(y),$$

where $\bar{G}_{\gamma,\sigma}$ is the survival function of the GPD. The change of variable $x = u + y$ yields for all $x \geq u$:

$$\bar{F}(x) \simeq \bar{F}(u)\bar{G}_{\gamma,\sigma}(x-u).$$

Finally, introducing $\alpha = \mathbb{P}(X > u) = \bar{F}(u)$, the survival function can be approximated when $x \to \infty$ by

$$\bar{F}(x) \simeq \alpha \bar{G}_{\gamma,\sigma}(x - \bar{F}^{-1}(\alpha)) = \alpha \bar{G}_{\gamma,\sigma}(x - q(\alpha)),$$

and its inverse is approximated when $p \to 0$ by

$$\bar{F}^{-1}(p) \simeq q(\alpha) + \bar{G}^{-1}_{\gamma,\sigma}(p/\alpha).$$

Approximation of risk measures

One thus obtains an approximation of **small tail probabilities**:

$$\mathbb{P}(X > x_n) \simeq \alpha \left[1 + \gamma \left(\frac{x_n - q(\alpha)}{\sigma}\right)\right]^{-1/\gamma} \quad \text{if} \quad \gamma \neq 0,$$

an approximation of **extreme quantiles – VaR – return levels**:

$$q(p_n) \simeq q(\alpha) + \frac{\sigma}{\gamma}\left[\left(\frac{p_n}{\alpha}\right)^{-\gamma} - 1\right] \quad \text{if} \quad \gamma \neq 0,$$

as well as an approximation of the **Expected Shortfall**:

$$\text{ES}(p_n) \simeq q(\alpha) + \frac{\sigma}{\gamma}\left[\frac{(p_n/\alpha)^{-\gamma}}{1-\gamma} - 1\right] \quad \text{if} \quad \gamma \neq 0 \text{ and } \gamma < 1.$$

Comparison between EVD and GPD approaches

For both approaches, the approximations are the same. For instance, in the case of extreme quantiles:

$$
\begin{aligned}
q(p_n) &\simeq q(\alpha) + \frac{\sigma}{\gamma}\left[\left(\frac{p_n}{\alpha}\right)^{-\gamma} - 1\right] && \text{(GPD)} \\
q(p_n) &\simeq b + \frac{a}{\gamma}\left[\left(\frac{np_n}{k}\right)^{-\gamma} - 1\right] && \text{(EVD)}
\end{aligned}
$$

There are three parameters to be estimated:

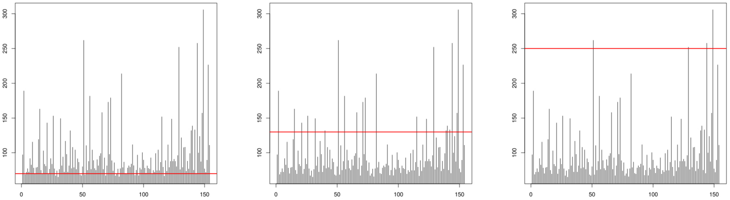
- the extreme-value index γ (shape parameter),
- a scale parameter σ (GPD) or a (EVD),
- a position parameter $q(\alpha)$ (GPD) or b (EVD).

Inference (with the GPD approach)

Let $\{Y_1, \ldots, Y_k\}$ be a sample of k iid random variables from $G_{\gamma,\sigma}$. Similarly to EVD case, this starting point may be a problem since, usually, one does not observe direclty a sample of excesses. They have to be extracted from an initial sample $\{X_1, \ldots, X_n\}$. The common practice is to choose a number of k excesses (*i.e.* the threshold is $X_{n-k,n}$) and select $\{Y_1, \ldots, Y_k\} := \{X_{n-k+1,n} - X_{n-k,n}, \ldots, X_{n,n} - X_{n-k,n}\}$: This is the **Peaks Over Threshold** (POT) method.

- Loss of information (n data $\longrightarrow$ k data),
- The excesses are not exactly distributed from a GPD distribution,
- Here, $\alpha = k/n$, the position parameter $q(k/n)$ is estimated by the empirical quantile $X_{n-k,n}$. It only remains to estimate γ and σ by MLE or PWM.

Illustration on Nidd data



Excesses selection. Three numbers of excesses k are considered. The associated thresholds are depicted in red.

Inference (with the GPD approach)

– Estimation of **small tail probabilities**:

$$\hat{\bar{F}}(x_n) = \frac{k}{n}\left[1 + \hat{\gamma}\left(\frac{x_n - X_{n-k,n}}{\hat{\sigma}}\right)\right]^{-1/\hat{\gamma}}.$$

– Estimation of **extreme quantiles – VaR – return levels**:

$$\hat{q}(p_n) = X_{n-k,n} + \frac{\hat{\sigma}}{\hat{\gamma}}\left[\left(\frac{np_n}{k}\right)^{-\hat{\gamma}} - 1\right].$$

– Estimation of **Expected Shortfall**:

$$\widehat{\mathrm{ES}}(p_n) = X_{n-k,n} + \frac{\hat{\sigma}}{\hat{\gamma}}\left[\frac{1}{1-\hat{\gamma}}\left(\frac{np_n}{k}\right)^{-\hat{\gamma}} - 1\right].$$

MLE

The negative log-likelihood is given by

$$\ell(\sigma,\gamma) = k \log \sigma + \left(1+\frac{1}{\gamma}\right) \sum_{i=1}^{k} \log\left(1+\gamma \frac{Y_i}{\sigma}\right)$$

if (σ,γ) is such that $1+\gamma Y_i/\sigma > 0$ for all $i = 1,\dots,k$ and $\ell(\sigma,\gamma) = +\infty$ otherwise. The reparametrization $(\tau,\gamma) := (\gamma/\sigma,\gamma)$ yields:

$$\tilde{\ell}(\tau,\gamma) = k \log(\gamma/\tau) + \left(1+\frac{1}{\gamma}\right) \sum_{i=1}^{k} \log\left(1+\tau Y_i\right) \tag{4}$$

and thus the equation

$$\frac{\partial \tilde{\ell}}{\partial \gamma} = \frac{k}{\gamma} - \frac{1}{\gamma^2} \sum_{i=1}^{k} \log\left(1+\tau Y_i\right) = 0$$

shows that $\hat{\gamma}$ can be expressed as a function of τ [12]:

$$\hat{\gamma}(\tau) = \frac{1}{k} \sum_{i=1}^{k} \log\left(1+\tau Y_i\right).$$

Replacing in (4), the MLE can be computed thanks to a one-parameter numerical optimization of $\tilde{\ell}(\tau,\hat{\gamma}(\tau))$.

- Since the support of the GPD depends on its parameters, the asymptotic normality of the MLE does not hold in the general case but only for $F \in \mathrm{MDA}(H_\gamma)$ with $\gamma > -1$ [20]. In this case, the Fisher information matrix is provided in [3] which allows to build asymptotic confidence intervals. [20] has also proved the non-consistency for $\gamma < -1$.
- In some context, one may assume $\gamma = 0$. In such a case, the GPD reduces to an exponential distribution and the associated estimator for extreme quantiles is referred to as the ET (exponential tail) estimator.

$\rightsquigarrow$ notebook3#2

Recall that $\{Y_1, \ldots, Y_k\}$ is a set of iid random variables with cdf $G_{\gamma,\sigma}$. In this case, the PWM of order $s \geq 0$ is defined by [15]:

$$\nu_s = \mathbb{E}\left[Y \bar{G}^s_{\gamma,\sigma}(Y)\right].$$

Lemma 2

The PWM of order $s \geq 0$ exists provided that $\gamma < 1$ and is given by

$$\nu_s = \frac{\sigma}{(s+1)(s+1-\gamma)}.$$

To obtain γ and σ two moments are enough:

$$\gamma = \frac{4\nu_1 - \nu_0}{2\nu_1 - \nu_0} \text{ and } \sigma = \frac{2\nu_1\nu_0}{\nu_0 - 2\nu_1},$$

where ν_0 and ν_1 can be easily estimated:

$$\hat{\nu}_s = \frac{1}{k}\sum_{i=1}^{k}\left(1 - \frac{i}{k}\right)^s Y_{i,k}.$$

Note that here, all estimators are explicit. The asymptotic normality of $(\hat{\gamma}, \hat{\sigma})$ is established in [2], for $F \in \mathrm{MDA}(H_\gamma)$ with $\gamma < 1/2$.

We focus on the Fréchet Maximum Domain of Attraction for which a simple characterisation of the survival function is available (Theorem 2, Chapter 2):

$$\bar{F}(x) = x^{-1/\gamma}\ell(x),$$

where ℓ is a slowly-varying function and $\gamma > 0$ (heavy-tailed distribution).

This can be interpreted as a semi-parametric model including:

- a parametric part $x^{-1/\gamma}$ which only depends on a parameter $\gamma > 0$,
- a non-parametric part $\ell(x)$, the only information being

$$\lim_{u\to\infty} \frac{\ell(tu)}{\ell(u)} = 1,$$

for all $t > 0$.

Application to the approximation of risk measures

For all $t > 0$,

$$\lim_{u \to \infty} \frac{\bar{F}(tu)}{\bar{F}(u)} = t^{-1/\gamma} \left(\lim_{u \to \infty} \frac{\ell(tu)}{\ell(u)} \right) = t^{-1/\gamma},$$

and one thus has the approximation:

$$\bar{F}(tu) \simeq \bar{F}(u) t^{-1/\gamma}.$$

Letting $x = tu$ and $\alpha = \bar{F}(u)$, approximations can be derived for small tail probabilities, extreme quantiles – VaR – return levels and Expected Shortfall:

$$\begin{aligned} \mathbb{P}(X > x_n) = \bar{F}(x_n) \quad &\simeq \quad \alpha \left(\frac{x_n}{q(\alpha)} \right)^{-1/\gamma}, \\ q(p_n) = \bar{F}^{-1}(p_n) \quad &\simeq \quad q(\alpha) \left(\frac{p_n}{\alpha} \right)^{-\gamma}, \\ \mathrm{ES}(p_n) \quad &\simeq \quad \frac{q(\alpha)}{1-\gamma} \left(\frac{p_n}{\alpha} \right)^{-\gamma} \text{ if } \gamma < 1. \end{aligned}$$

Comparison with EVD and GPD approaches

For all three approaches, the approximations are the same. For instance, in the case of small tail probabilities and $\gamma > 0$:

$$\bar{F}(x_n) \simeq \alpha \left[1 + \gamma \left(\frac{x_n - q(\alpha)}{\sigma}\right)\right]^{-1/\gamma} \qquad \text{(GPD)}$$

$$\bar{F}(x_n) \simeq \frac{k}{n} \left[1 + \gamma \left(\frac{x_n - a}{b}\right)\right]^{-1/\gamma} \qquad \text{(EVD)}$$

$$\bar{F}(x_n) \simeq \alpha \left(\frac{x_n}{q(\alpha)}\right)^{-1/\gamma} \qquad \text{(semi-parametric, Fréchet)}$$

The third approach is a particular case of the GPD method with $\sigma = \gamma q(\alpha)$. Thus, there are only two parameters to be estimated.

- As already seen, letting $\alpha = k/n$, the position parameter $q(\alpha)$ can be estimated by the ordered statistics $X_{n-k,n}$.
- It only remains to estimate the extreme-value index γ.

1) Semi-parametric estimation of γ: Hill estimator

The idea is to make use of the approximation

$$q(p) \simeq q(\alpha) \left(\frac{p}{\alpha}\right)^{-\gamma}.$$

Taking the logarithm yields

$$\log q(p) - \log q(\alpha) \simeq \gamma \log(\alpha/p).$$

Letting $\alpha = k/n$ and considering several values of $p = i/n$, $i = 1, \ldots, k-1$, one has

$$\log q(i/n) - \log q(k/n) \simeq \gamma \log(k/i).$$

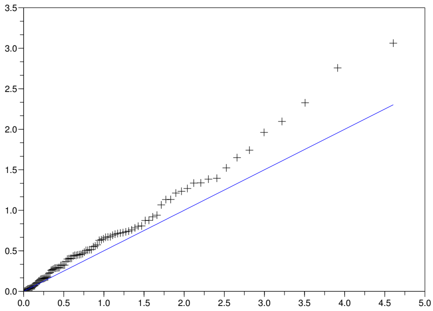
Now, the quantiles are estimated by the associated order statistics, leading to

$$\log X_{n-i+1,n} - \log X_{n-k+1,n} \simeq \gamma \log(k/i),$$

for $i = 1, \ldots, k-1$. The quality of this approximation can be assessed graphically.

Log quantile-quantile plot

Simulation of a sample of size $n = 500$ from a t_2 distribution ($\gamma = 1/2$). Here, $k = 100$.



Horizontally: $\log(k/i)$. Vertically: $y = x/2$ and $\log X_{n-i+1,n} - \log X_{n-k+1,n}$ for $i = 1, \ldots, k-1$.

Summing up the previous approximations for $i = 1, \ldots, k-1$, we get

$$\gamma \simeq \frac{\displaystyle\sum_{i=1}^{k-1} \log X_{n-i+1,n} - \log X_{n-k+1,n}}{\displaystyle\sum_{i=1}^{k-1} \log(k/i)}.$$

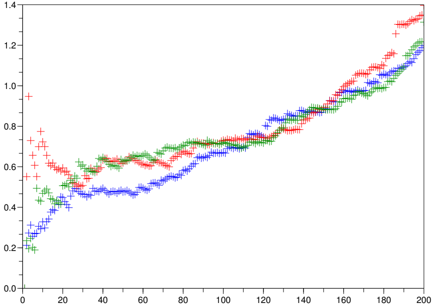
The lower part can be rewritten as $\log[k^{k-1}/(k-1)!)$. Thanks to Stirling's formula, one may show that is is asymptotically equivalent to $k-1$ when $k \to \infty$. We thus obtain (with $k \mapsto k+1$):

Definition 1 (Hill estimator [13])

$$\hat{\gamma}(k) = \frac{1}{k} \sum_{i=0}^{k-1} (\log X_{n-i,n} - \log X_{n-k,n}).$$

This estimator is scale invariant and asymptotically Gaussian [4, Theorem 3.2.5].

Estimation of γ on three different samples of size $n = 500$ from a t_2 distribution ($\gamma = 1/2$).

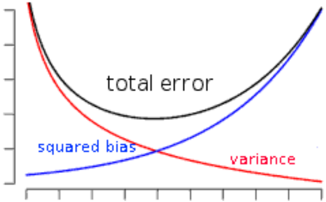


Horizontally: k. Vertically: $\hat{\gamma}(k)$ for $k = 1, \ldots, 200$.

In practice ...

The choice of k is difficult:

- If k is small, $\hat{\gamma}(k)$ is based on few observations, it has therefore a **large variance**.
- If k is large, then $X_{n-k,n}$ is no longer in the distribution tail and the approximation of the survival function by a power function is no longer true, $\hat{\gamma}(k)$ has a **large bias**.



Trade-off between bias and variance.

Inference (with the semi-approach in Fréchet MDA)

– Estimation of **small tail probabilities**:

$$\hat{\bar{F}}(x_n) = \frac{k}{n}\left(\frac{x_n}{X_{n-k,n}}\right)^{-1/\hat{\gamma}}.$$

– Estimation of **extreme quantiles – VaR – return levels** (Weissman estimator [19]):

$$\hat{q}(p_n) = X_{n-k,n}\left(\frac{np_n}{k}\right)^{-\hat{\gamma}}.$$

– Estimation of **Expected Shortfall**:

$$\widehat{\mathrm{ES}}(p_n) = \frac{X_{n-k,n}}{1-\hat{\gamma}}\left(\frac{np_n}{k}\right)^{-\hat{\gamma}}.$$

$\leadsto$ notebook3#3

To go further: asymptotic analysis

- Recall that, $F \in \text{DA}(\text{Fréchet})$ is equivalent to $\bar{F} \in \mathcal{RV}_{-1/\gamma}$. Introducing the **tail quantile function** $U(\cdot) = \bar{F}^{-1}(1/\cdot)$, the above properties are also equivalent to $U \in \mathcal{RV}_\gamma$, *i.e.* $U(tx)/U(t) \to x^\gamma$ for all $x > 0$ as $t \to \infty$. This property is sometimes referred to as a first order condition in the extreme-value literature. It is a sufficient condition to prove the consistency of extreme-value estimators.
- To establish the asymptotic distribution of an estimator, a second order condition is usually introduced: There exist $\gamma > 0$, $\rho \leq 0$ and a positive or negative function A with $A(t) \to 0$ as $t \to \infty$ such that

$$\frac{1}{A(t)}\left(\frac{U(tx)}{U(t)} - x^\gamma\right) \to x^\gamma \frac{x^\rho - 1}{\rho} \tag{5}$$

for all $x > 0$ as $t \to \infty$. The limit should be understood as $x^\gamma \log x$ when $\rho = 0$. Since $A(t) \to 0$ as $t \to \infty$, the limit (5) can be equivalently rewritten as

$$\frac{1}{A(t)}\left(\log U(tx) - \log U(t) - \gamma \log x\right) \to \frac{x^\rho - 1}{\rho} \tag{6}$$

for all $x > 0$ as $t \to \infty$. Moreover, it can be shown that $|A|$ is regularly varying with index ρ.

Theorem 1 (Consistency)

Let $X_1, \ldots, X_n$ be i.i.d. random variables with cdf $F \in MDA(H_\gamma)$, $\gamma > 0$. If $k \to \infty$ with $k/n \to 0$ as $n \to \infty$ then $\hat{\gamma}(k) \xrightarrow{\mathrm{P}} \gamma$.

Theorem 2 (Asymptotic normality)

Suppose the assumptions of Theorem 1 hold with the second order condition (6). If, moreover, $\sqrt{k}A(n/k) \to \lambda < \infty$, then

$$\sqrt{k}(\hat{\gamma}(k) - \gamma) \xrightarrow{d} \mathcal{N}(\lambda/(1-\rho), \gamma^2).$$

Theorem 2 shows that the asymptotic variance of Hill estimator is γ^2/k (a decreasing function of k) and that the asymptotic bias is $A(n/k)/(1-\rho)$ (an increasing function of k).

The asymptotic mean-squared error is thus given by

$$AMSE(k) = \frac{\gamma^2}{k} + \frac{A^2(n/k)}{(1-\rho)^2}.$$

This quantity is difficult to compute in practice since γ, ρ and $A(\cdot)$ are unknown. Nevertheless, its possible to derive the theoretical optimal value of k in the simplified case where $A(t) = \beta t^{\rho}$:

$$k(n) = \left(\frac{\gamma^2(1-\rho)^2}{-2\rho\beta^2}\right)^{1/(1-2\rho)} n^{-2\rho/(1-2\rho)}.$$

The optimal rate of convergence of Hill estimator is of order $n^{-\rho/(1-2\rho)}$. It approaches the parametric rate $n^{1/2}$ when $\rho \to -\infty$.

Both proofs of Theorem 1 and Theorem 2 rely on some common tools. First, let us note that, for all $i = 1, \dots, n$, $X_i \stackrel{d}{=} U(Y_i)$ where $Y_1, \dots, Y_n$ are i.i.d. from a standard Pareto distribution *i.e.* with cdf $1 - 1/y$, $y \geq 1$. It follows that

$$\hat{\gamma}(k) \stackrel{d}{=} \frac{1}{k} \sum_{i=0}^{k-1} \log U(Y_{n-i,n}) - \log U(Y_{n-k,n}).$$

Moreover, remarking that, for all $i = 1, \dots, n$, $\log Y_i$ follows a standard exponential distribution, the following lemma will reveal useful.

Lemma 3

Let $E_1, \dots, E_n$ be i.i.d. random variables with standard exponential distribution. Let $E_{1,n} \leq \dots \leq E_{n,n}$ be the associated order statistics and introduce $k \to \infty$ such that $k/n \to 0$ as $n \to \infty$.

- $E_{n-k,n} \xrightarrow{\mathrm{P}} \infty$,
- *If f is a function such that $var(f(E_1)) < \infty$, then*

$$D_n := \sqrt{k} \left(\frac{1}{k} \sum_{i=0}^{k-1} f(E_{n-i,n} - E_{n-k,n}) - \mathbb{E}(f(E_1)) \right)$$

is independent of $E_{n-k,n}$ and is asymptotically $\mathcal{N}(0, var(f(E_1)))$.

Theorem 3 (Rényi representation)

Let s $E_1, cdots, E_n$ be i.i.d standard exponential random variables. Then,

$$\left(E_{j,n},\ j=1,\dots,n\right) \stackrel{d}{=} \left(\sum_{r=1}^{j} \frac{F_r}{n-r+1},\ j=1,\dots,n\right),$$

where $F_1,\dots,F_n$ are i.i.d standard exponential random variables.

Proof of Lemma 3. The independence and $E_{n-k,n} \xrightarrow{\mathrm{P}} \infty$ are direct consequences of Rényi representation that also implies

$$\{E_{n-i,n} - E_{n-k,n}\}_{i=0,\dots,k-1} \stackrel{d}{=} \{E_{k-i,k}\}_{i=0,\dots,k-1}.$$

It follows that

$$D_n \stackrel{d}{=} \sqrt{k}\left(\frac{1}{k}\sum_{i=0}^{k-1} f(E_{k-i,k}) - \mathbb{E}(f(E_1))\right) = \sqrt{k}\left(\frac{1}{k}\sum_{i=0}^{k-1} f(E_i) - \mathbb{E}(f(E_1))\right),$$

and the central limit theorem concludes the proof. ■

Proof of Theorem 1. From the first order condition and Potter bounds, we have for all $\varepsilon > 0$, $\varepsilon' > 0$, $x \geq 1$ and $t \geq t_0$,

$$(1-\varepsilon)x^{\gamma-\varepsilon'} < \frac{U(tx)}{U(t)} < (1+\varepsilon)x^{\gamma+\varepsilon'},$$

or equivalently

$$\log(1-\varepsilon)+(\gamma-\varepsilon')\log x < \log U(tx) - \log U(t) < \log(1+\varepsilon)+(\gamma+\varepsilon')\log x.$$

Applying this inequality for $t = Y_{n-k,n} \xrightarrow{\mathrm{P}} \infty$ (Lemma 3) and $x = Y_{n-i,n}/Y_{n-k,n} \geq 1$ yields

$$\begin{aligned}
\log(1-\varepsilon)+(\gamma-\varepsilon')\log\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right) \quad &< \quad \log\left(\frac{U(Y_{n-i,n})}{U(Y_{n-k,n})}\right) \\
&< \quad \log(1+\varepsilon)+(\gamma+\varepsilon')\log\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right),
\end{aligned}$$

for all $i = 0, \ldots, k-1$.

As a consequence, it follows that

$$\log(1-\varepsilon)+(\gamma-\varepsilon')S_n<\hat{\gamma}_n<\log(1+\varepsilon)+(\gamma+\varepsilon')S_n$$

where we have defined

$$S_n := \frac{1}{k}\sum_{i=0}^{k-1}\log\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right).$$

Remarking that $\log Y \stackrel{d}{=} E$ yields

$$\begin{aligned} S_n &= \frac{1}{k}\sum_{i=0}^{k-1}(\log Y_{n-i,n}-\log Y_{n-k,n}) \stackrel{d}{=} \frac{1}{k}\sum_{i=0}^{k-1} E_{n-i,n}-E_{n-k,n} \\ &\stackrel{d}{=} \frac{1}{k}\sum_{i=0}^{k-1} E_{k-i,k} = \frac{1}{k}\sum_{i=0}^{k-1} E_i. \end{aligned}$$

The law of large number entails $S_n \xrightarrow{\mathrm{P}} 1$ and the result is proved. ■

Proof of Theorem 2. From (3.2.7) in [4], the second order condition (6) implies

$$\left|\frac{\log U(tx) - \log U(t) - \gamma \log x}{A_0(t)} - \frac{x^\rho - 1}{\rho}\right| \le \varepsilon x^{\rho+\varepsilon}, \qquad (7)$$

where $A_0(t) \sim A(t)$ as $t \to \infty$, $\varepsilon > 0$, $t \ge t_0$, $x \ge 1$. Applying this inequality for $t = Y_{n-k,n} \xrightarrow{\mathrm{P}} \infty$ (Lemma 3) and $x = Y_{n-i,n}/Y_{n-k,n} \ge 1$ yields

$$\hat{\gamma}(k) \stackrel{d}{=} \gamma S_n + A_0(Y_{n-k,n}) \frac{1}{k} \sum_{i=0}^{k-1} \frac{1}{\rho} \left(\left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right)^\rho - 1 \right) + R_n,$$

where S_n is defined in the proof of Theorem 1 and

$$|R_n| \le \varepsilon |A_0(Y_{n-k,n})| \frac{1}{k} \sum_{i=0}^{k-1} \left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right)^{\rho+\varepsilon}.$$

Introducing for all $a < 1$,

$$T_n(a) = \frac{1}{k} \sum_{i=0}^{k-1} \left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right)^a,$$

it follows

$$\sqrt{k}(\hat{\gamma}(k) - \gamma) \stackrel{d}{=} \gamma \sqrt{k}(S_n - 1) + \sqrt{k} A_0(Y_{n-k,n})(T_n(\rho) - 1)/\rho + \sqrt{k} R_n,$$

with $|R_n| \leq \varepsilon |A_0(Y_{n-k,n})| T_n(\rho + \varepsilon)$. Lemma 3 shows that $\sqrt{k}(S_n - 1)$ is asymptotically standard Gaussian. Besides,

$$\begin{aligned}
T_n(a) &= \frac{1}{k} \sum_{i=0}^{k-1} \exp\left(a(\log Y_{n-i,n} - \log Y_{n-k,n})\right) \\
&\stackrel{d}{=} \frac{1}{k} \sum_{i=0}^{k-1} \exp\left(a(E_{n-i,n} - E_{n-k,n})\right) \\
&\stackrel{d}{=} \frac{1}{k} \sum_{i=0}^{k-1} \exp\left(a E_{k-i,k}\right) = \frac{1}{k} \sum_{i=0}^{k-1} \exp\left(a E_i\right) \\
&\xrightarrow{\mathrm{P}} \mathbb{E}(\exp(a E_1)) = 1/(1-a),
\end{aligned}$$

from the law of large numbers.

Finally, properties of regularly-varying functions and $kY_{n-k,n}/n \xrightarrow{\mathrm{P}} 1$ show that

$$|A_0(Y_{n-k,n})| \sim |A(Y_{n-k,n})| = |A(n/k)|(1 + o_P(1)),$$

as $n \to \infty$. As a consequence,

$$\sqrt{k} A_0(Y_{n-k,n})(T_n(\rho) - 1)/\rho \xrightarrow{\mathrm{P}} \lambda/(1-\rho)$$

and

$$\sqrt{k}|R_n| \leq \frac{\lambda\varepsilon}{1-\rho-\varepsilon}(1 + o_P(1))$$

and the conclusion follows. ∎

2) Semi-parametric estimation of γ: Pickands estimator

The idea is to use twice the same approximation:

$$U(x) \simeq U(y)\left(\frac{x}{y}\right)^{\gamma} \text{ and } U(z) \simeq U(y)\left(\frac{z}{y}\right)^{\gamma}$$

recalling that U is regularly varying with index γ. It follows that

$$\frac{U(x)-U(y)}{U(y)-U(z)} \simeq \frac{(x/y)^{\gamma}-1}{1-(z/y)^{\gamma}}.$$

Considering $x/y = 2$ and $z/y = 1/2$ leads to

$$\frac{U(2y)-U(y)}{U(y)-U(y/2)} \simeq \frac{2^{\gamma}-1}{1-2^{-\gamma}} = 2^{\gamma}.$$

Taking the logarithm yields

$$\gamma \simeq \frac{1}{\log 2}\log\left(\frac{U(2y)-U(y)}{U(y)-U(y/2)}\right).$$

Letting $y = n/(2k)$, with $k \in \{1, \ldots, n\}$, we have

$$\gamma \simeq \frac{1}{\log 2} \log \left(\frac{U(n/k) - U(n/(2k))}{U(n/(2k)) - U(n/(4k))} \right).$$

Now, recall that $U(y) = \bar{F}^{-1}(1/y)$, estimating the unknown cdf F by the empirical one gives rise to:

Definition 2 (Pickands estimator [16])

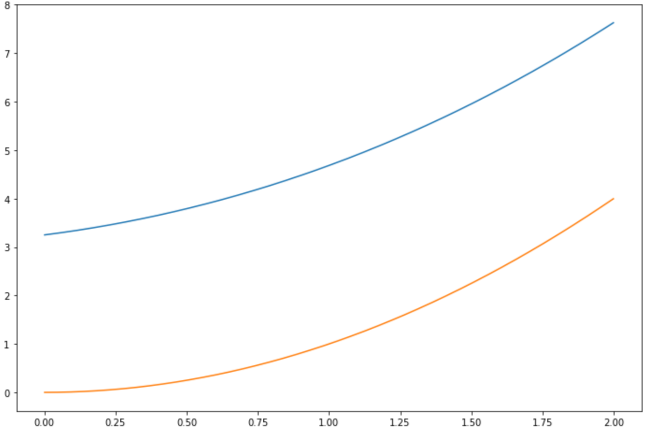
$$\tilde{\gamma}(k) = \frac{1}{\log 2} \log \left(\frac{X_{n-k,n} - X_{n-2k,n}}{X_{n-2k,n} - X_{n-4k,n}} \right).$$

- This estimator is location and scale invariant.
- It is consistent for all $\gamma \in \mathbb{R}$!
- An asymptotic normality result also holds for all $\gamma \in \mathbb{R}$, see [9], Theorem 3.3.5. The variance of the limiting Gaussian distribution is

$$v_P(\gamma) = \frac{\gamma^2 (2^{2\gamma+1} + 1)}{4(\log 2)^2 (2^\gamma - 1)^2},$$

to be compared to the Hill case where $v_H(\gamma) = \gamma^2$.

Comparison of variances



Variances as functions of $\gamma \in (0, 2]$. Orange: Hill, Blue: Pickands.

Here, we limit ourselves to proving the following consistency result ...

Theorem 4 (Consistency)

Let $X_1, \dots, X_n$ be i.i.d. random variables with cdf $F \in MDA(H_\gamma)$, $\gamma \in \mathbb{R}$. If $k \to \infty$ with $k/n \to 0$ as $n \to \infty$ then $\tilde{\gamma}(k) \xrightarrow{\mathrm{P}} \gamma$.

... in the particular case where $\gamma > 0$.

Proof. Assume $\gamma > 0$. Similarly to the proofs of Theorem 1 and Theorem 2 (dedicated to Hill estimator), the proof relies on that fact that, for all $i = 1, \dots, n$, $X_i \stackrel{d}{=} U(Y_i)$ where $Y_1, \dots, Y_n$ are i.i.d. from a standard Pareto distribution *i.e.* with cdf $1 - 1/y$, $y \leq 1$. It follows that

$$\tilde{\gamma}(k) \stackrel{d}{=} \frac{1}{\log 2} \log \left(\frac{U(Y_{n-k,n}) - U(Y_{n-2k,n})}{U(Y_{n-2k,n}) - U(Y_{n-4k,n})} \right).$$

The proof then follows that same lines as the heuristic used for introducing Pickands estimator:

$$\tilde{\gamma}(k) \stackrel{d}{=} \frac{1}{\log 2} \log \left(\frac{\frac{U(Y_{n-k,n})}{U(Y_{n-2k,n})} - 1}{1 - \frac{U(Y_{n-4k,n})}{U(Y_{n-2k,n})}} \right). \tag{8}$$

Clearly,

$$\frac{U(Y_{n-k,n})}{U(Y_{n-2k,n})} = \frac{U\left(Y_{n-2k,n} \frac{Y_{n-k,n}}{Y_{n-2k,n}}\right)}{U(Y_{n-2k,n})},$$

and recall that, from Potter bounds (see the proof of Theorem 1), for all $\varepsilon > 0$, $\varepsilon' > 0$, $x \geq 1$ and $t \geq t_0$,

$$(1-\varepsilon)x^{\gamma-\varepsilon'} < \frac{U(tx)}{U(t)} < (1+\varepsilon)x^{\gamma+\varepsilon'},$$

The idea is then to consider $t = Y_{n-2k,n}$ and $x = Y_{n-k,n}/Y_{n-2k,n} \geq 1$ in the above inequality, since $Y_{n-2k,n} \xrightarrow{\mathrm{P}} \infty$ as $n \to \infty$, see the proof of Theorem 2.

This entails:

$$(1-\varepsilon)\left(\frac{Y_{n-k,n}}{Y_{n-2k,n}}\right)^{\gamma-\varepsilon'} < \frac{U(Y_{n-k,n})}{U(Y_{n-2k,n})} < (1+\varepsilon)\left(\frac{Y_{n-k,n}}{Y_{n-2k,n}}\right)^{\gamma+\varepsilon'},$$

Besides,

$$\log\left(\frac{Y_{n-k,n}}{Y_{n-2k,n}}\right) \stackrel{d}{=} E_{n-k,n} - E_{n-2k,n} \stackrel{d}{=} E_{k,2k},$$

in view of Rényi representation. Note that $E_{k,2k}$ is the empirical median of the standard exponential distribution. It converges in probability to the theoretical median which is $\log 2$. Consequently, $Y_{n-k,n}/Y_{n-2k,n} \xrightarrow{\mathrm{P}} 2$ and therefore

$$U(Y_{n-k,n})/U(Y_{n-2k,n}) \xrightarrow{\mathrm{P}} 2^{\gamma}.$$

Similarly (replacing k by $2k$), one also has

$$U(Y_{n-2k,n})/U(Y_{n-4k,n}) \xrightarrow{\mathrm{P}} 2^{\gamma}.$$

The result follows from (8). ∎

3) Semi-parametric estimation of extreme quantiles

Basing on the approximation

$$q(p_n) := U(1/p_n) \simeq U(1/\alpha_n) \left(\frac{\alpha_n}{p_n} \right)^{\gamma},$$

and letting $\alpha_n = k_n/n$, the following estimator has been introduced:

Definition 3 (Weissman estimator [19])

$$\hat{q}(p_n) = X_{n-k_n,n} \left(\frac{k_n}{np_n} \right)^{\hat{\gamma}(k_n)},$$

where $\hat{\gamma}(k_n)$ is the Hill estimator.

Theorem 5 (Asymptotic normality)

Suppose the assumptions of Theorem 2 hold. If moreover $np_n/k_n \to 0$ as $n \to \infty$ then

$$\frac{\sqrt{k_n}\,(\log \hat{q}(p_n) - \log q(p_n))}{\log(k_n/(np_n))} \xrightarrow{d} \mathcal{N}(\lambda/(1-\rho), \gamma^2).$$

Proof. Recalling that $q(p_n) = U(1/p_n)$, $X_{n-k_n,n} \overset{d}{=} U(Y_{n-k_n,n})$ and $U \in \mathcal{RV}_\gamma$ yields

$$
\begin{aligned}
\log \hat{q}(p_n) &= \gamma \log Y_{n-k_n,n} + \log \ell(Y_{n-k_n,n}) + \hat{\gamma}_n \log(k_n/(np_n)), \\
\log q(p_n) &= \gamma \log(n/k_n) + \log \ell(1/p_n) + \gamma \log(k_n/(np_n)),
\end{aligned}
$$

and therefore

$$
\begin{aligned}
\frac{\sqrt{k_n}\,(\log \hat{q}(p_n) - \log q(p_n))}{\log(k_n/(np_n))} &\overset{d}{=} \gamma \frac{\sqrt{k_n}\,(\log Y_{n-k_n,n} - \log(n/k_n))}{\log(k_n/(np_n))} \\
&+ \frac{\sqrt{k_n}\,(\log \ell(Y_{n-k_n,n}) - \log \ell(1/p_n))}{\log(k_n/(np_n))} \\
&+ \sqrt{k_n}(\hat{\gamma}_n - \gamma) \\
&= T_{1,n} + T_{2,n} + T_{3,n}.
\end{aligned}
$$

The three terms are studied separately.

- From [4], Corollary 2.2.2, $\sqrt{k_n}(k_n Y_{n-k_n,n}/n - 1) \xrightarrow{d} \mathcal{N}(0,1)$. The delta-method yields that $\sqrt{k_n}(\log Y_{n-k_n,n} - \log(n/k_n))$ has the same asymptotic distribution and thus $T_{1,n} \xrightarrow{\mathrm{P}} 0$.
- The consequence of the second-order condition (7) can be written as

$$\left| \frac{\log \ell(tx) - \log \ell(t)}{A_0(t)} - \frac{x^\rho - 1}{\rho} \right| \leq \varepsilon x^{\rho+\varepsilon}.$$

Letting $t = Y_{n-k_n,n}$ and $x = 1/(p_n Y_{n-k_n,n})$, it follows that $x = (k_n/(np_n))(1 + o_P(1)) \xrightarrow{\mathrm{P}} \infty$ and consequently,

$$\begin{aligned} \log \ell(Y_{n-k_n,n}) - \log \ell(1/p_n) &= \frac{1}{\rho} A_0(Y_{n-k_n,n})(1 + o_P(1)) \\ &= \frac{1}{\rho} A(n/k_n)(1 + o_P(1)). \end{aligned}$$

This entails that

$$T_{2,n} = \frac{\lambda}{\rho} \frac{1}{\log(k_n/(np_n))}(1 + o_P(1)) \xrightarrow{\mathrm{P}} 0.$$

- Finally, $T_{3,n} \xrightarrow{d} \mathcal{N}(\lambda/(1-\rho), \gamma^2)$ from Theorem 2. ■

Let us now consider the case where $F \in \text{MDA(Weibull)}$. From Theorem 3 in Chapter 2 this implies that the endpoint is finite: $x_F < \infty$. Moreover, the extreme-value index is negative: $\gamma < 0$.

Lemma 4

Let $X_1, \dots, X_n$ be i.i.d. from a c.d.f. $F \in MDA(H_\gamma)$ with $\gamma < 0$ and introduce for all $i = 1, \dots, n$, $X_i^* = 1/(x_F - X_i)$. Then, $X_1^*, \dots, X_n^*$ are i.i.d. from the cdf $x \geq 0 \mapsto F_*(x) = F(x_F - 1/x)$ with $F_* \in MDA(H_{-\gamma})$.

Proof. Clearly, $X_1^*, \dots, X_n^*$ are i.i.d. since $X_1, \dots, X_n$ are i.i.d. Besides, for all $x \geq 0$,

$$\begin{aligned} F_*(x) &= \mathbb{P}(X_i^* \leq x) = \mathbb{P}(1/(x_F - X_i) \leq x) \\ &= \mathbb{P}(x_F - X_i \geq 1/x) = \mathbb{P}(X_i \leq x_F - 1/x) \\ &= F(x_F - 1/x), \end{aligned}$$

and Theorem 3 in Chapter 2 states that $F_* \in MDA(H_{-\gamma})$. ∎

As a consequence, each inference procedure dedicated to MDA(Fréchet) can be adapted to MDA(Weibull) by considering the transformed sample.

For instance, the extreme-value index of F_* can be estimated using Hill estimator computed on $X_1^*, \ldots, X_n^*$:

$$\begin{aligned}
-\tilde{\gamma}(k) &= \frac{1}{k}\sum_{i=0}^{k-1}(\log X_{n-i,n}^* - \log X_{n-k,n}^*) \\
&= -\frac{1}{k}\sum_{i=0}^{k-1}(\log(x_F - X_{n-i,n}) - \log(x_F - X_{n-k,n})),
\end{aligned}$$

leading to (after simplification):

$$\tilde{\gamma}(k) = \frac{1}{k}\sum_{i=0}^{k-1}(\log(x_F - X_{n-i,n}) - \log(x_F - X_{n-k,n})).$$

This is not, however, a proper estimate since it depends on x_F which is unknown in practice.

The natural idea, is then to estimate the endpoint by the maximum of the sample: $\hat{x}_F = X_{n,n}$. This gives rise to the negative Hill estimator:

Definition 5 (negative Hill estimator [8])

$$\hat{\gamma}(k) = \frac{1}{k}\sum_{i=1}^{k-1}(\log(X_{n,n} - X_{n-i,n}) - \log(X_{n,n} - X_{n-k,n})).$$

- This estimator is both location and scale invariant (better than the standard Hill estimator).
- It is consistent if $\gamma < -1/2$,...
- ... and asymptotically Gaussian if $-1 < \gamma < -1/2$, see [4], Theorem 3.6.4.

Estimation of the endpoint

As already seen, one can estimate the endpoint by $\hat{x}_F = X_{n,n}$. The asymptotic behaviour of this estimator is a direct consequence of the extreme-value theorem:

Theorem 6

Let $F \in MDA(H_\gamma)$ with $\gamma < 0$. Then, as $n \to \infty$,

$$\frac{\hat{x}_F - x_F}{x_F - F^{-1}(1 - 1/n)} \xrightarrow{d} Y,$$

where $Y \sim H_\gamma$.

Proof. Combine Theorem 1 and Theorem 3 in Chapter 2. ∎

In particular, we obtain that $\hat{x}_F \xrightarrow{\mathrm{P}} x_F$ with rate of convergence $1/(x_F - F^{-1}(1 - 1/n)) \to \infty$ since $F^{-1}(1 - 1/n) \to x_F$ as $n \to \infty$.

Another option is to use the previous approximations of extreme quantiles (for $\gamma < 0$):

$$\begin{aligned}\bar{F}^{-1}(p_n) &\simeq b_n + \frac{a_n}{\gamma}\left[\left(\frac{np_n}{k}\right)^{-\gamma} - 1\right] \text{ (EVD approach)},\\ \bar{F}^{-1}(p_n) &\simeq \bar{F}^{-1}(\alpha) + \frac{\sigma}{\gamma}\left[\left(\frac{p_n}{\alpha}\right)^{-\gamma} - 1\right] \text{ (GPD approach)},\end{aligned}$$

with $p_n = 0$ as probability level, leading to:

$$\begin{aligned}x_F &\simeq b_n - \frac{a_n}{\gamma} \text{ (EVD approach)},\\ x_F &\simeq \bar{F}^{-1}(\alpha) - \frac{\sigma}{\gamma} \text{ (GPD approach)},\end{aligned}$$

and then estimate the unknown parameters with an adapted inference procedure (MLE, PWM, ...).

Finally, let us consider the case where $F \in \mathrm{MDA}(\mathrm{Gumbel})$. The characterisation provided by Theorem 4 in Chapter 2 is not useful to perform semi-parametric inference.
One can then focus on a subclass of MDA(Gumbel):

Definition 7 (Weibull-tail distributions)

A distribution is Weibull-tailed if there exist $\theta > 0$ and a slowly-varying function ℓ such that

$$1 - F(x) = \exp(-H(x)) \text{ with } H^{-1}(t) = t^{\theta}\ell(t).$$

- H is called the **hazard function**,
- θ is refered to as the **Weibull-tail coefficient**.

As a consequence, $H^{-1} \in \mathcal{R}V_{\theta}$ and $H \in \mathcal{R}V_{1/\theta}$. Moreover, the endpoint is infinite $x_F = \infty$.

Proposition 1

If a cdf F is Weilbull-tailed with a twice differentiable hazard function H, then $F \in$MDA(Gumbel).

Proof: It relies on Theorem 4 in Chapter 2. Let $a(x) = 1/H'(x)$. Then,

$$x \mapsto |a'(x)| = \left| \frac{H''(x)}{(H'(x))^2} \right|$$

is regularly-varying with index $(1/\theta - 2) - 2(1/\theta - 1) = -1/\theta < 0$. As a consequence, $a'(x) \to 0$ as $\to \infty$.
Letting moreover, $g(x) = 1$ and for all x_0, $c(x) = \exp(-H(x_0))$, it follows that

$$\bar{F}(x) = c(x) \exp\left(- \int_{x_0}^{x} \frac{g(t)}{a(t)} dt \right),$$

and the result follows the characterisation of MDA(Gumbel). ∎

Examples

	θ	$\ell(x)$
Gaussian $\mathcal{N}(\mu, \sigma^2)$	$1/2$	$2^{1/2}\sigma - \frac{\sigma}{2^{3/2}} \frac{\log x}{x} + O(1/x)$
Gamma $\Gamma(\alpha, \beta)$	1	$\frac{1}{\beta} + \frac{\alpha - 1}{\beta} \frac{\log x}{x} + O(1/x)$
Weibull $\mathcal{W}(\alpha, \lambda)$	$1/\alpha$	λ

Remark: The Lognormal distribution belongs to MDA(Gumbel) but is not Weibull-tailed.

Estimation of the Weibull-tail coefficient

Consider the tail quantile function:

$$U(t) = \bar{F}^{-1}(1/t) = H^{-1}(\log t) = (\log t)^{\theta} \ell(\log t).$$

When both t and s are large:

$$\begin{aligned} \log U(t) - \log U(s) &= \theta \left(\log \log t - \log \log s\right) + \log \left(\frac{\ell(\log t)}{\ell(\log s)} \right) \\ &\simeq \theta \left(\log \log t - \log \log s\right), \end{aligned}$$

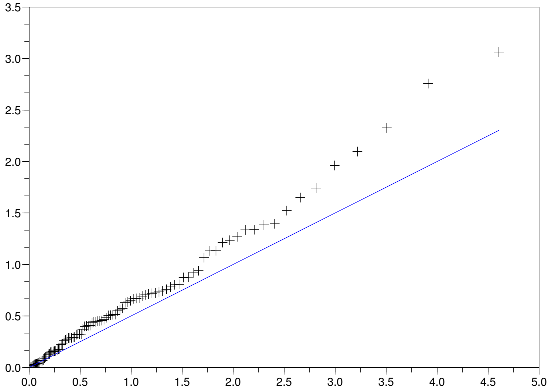
since ℓ is slowly-varying. Letting $t = n/i$, $s = n/k$ and estimating the quantiles by their empirical counterpart yield:

$$\log X_{n-i+1,n} - \log X_{n-k+1,n} \simeq \theta \left(\log \log(n/i) - \log \log(n/k)\right),$$

for all $i = 1, \ldots, k-1$. Similar to the approximation used to build Hill estimator, except the $\log \log$ terms.

Quantile-quantile plot

Simulation of a sample of size $n = 500$ from a Gaussian distribution ($\theta = 1/2$). Here, $k = 100$.



Horizontally: $\log\log(n/i) - \log\log(n/k)$. Vertically: $y = x/2$ and $\log X_{n-i+1,n} - \log X_{n-k+1,n}$ for $i = 1, \ldots, k-1$.

Summing up the previous approximations for $i = 1, \ldots, k-1$, and letting $k \mapsto k+1$, we get the so-called Hill-type estimator [11]:

$$\hat{\theta}(k) = \frac{\displaystyle\sum_{i=0}^{k-1} (\log X_{n-i,n} - \log X_{n-k,n})}{\displaystyle\sum_{i=0}^{k-1} (\log\log(n/(i+1)) - \log\log(n/(k+1)))}.$$

- Very similar to Hill estimator!
- Scale invariant,
- Consistency and asymptotic normality are established in [11].

Basing on the approximation

$$\frac{q(p_n)}{U(1/\alpha_n)} := \frac{U(1/p_n)}{U(1/\alpha_n)} \simeq \left(\frac{\log(1/p_n)}{\log(1/\alpha_n)}\right)^{\theta},$$

and letting $\alpha_n = k_n/n$, the following estimator has been introduced:

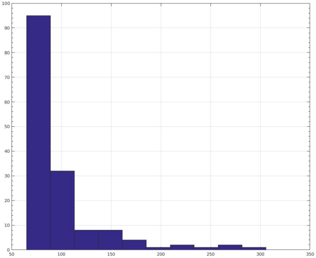
Definition 8 (Weissman type estimator [10])

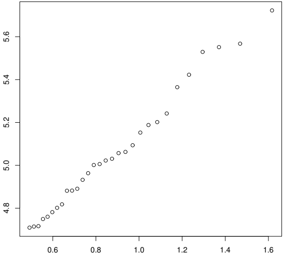
$$\hat{q}(p_n) = X_{n-k_n,n}\left(\frac{\log(1/p_n)}{\log(n/k_n)}\right)^{\hat{\theta}(k_n)},$$

where $\hat{\theta}(k_n)$ is an estimator of the Weibull-tail coefficient.

Asymptotic normality has been established in [10].

Application to the Nidd river dataset





$k = 29$ *yields an approximately linear quantile-quantile plot,* $\hat{\theta}(k) \simeq 0.89$.
One hundred year return level: $\hat{q}(p_n) \simeq 366\, m^3 s^{-1}$, recall that $X_{n,n} \simeq 305\, m^3 s^{-1}$.

References I

Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation*, Cambridge University Press.

Cai, J.-J., de Haan, L. and Zhou, C. (2013). Bias correction in extreme value statistics with index around zero. *Extremes*, **16**, 173–201.

Davison, A. C., and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society*. Series B, **52**, 393–442.

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer-Verlag, New York.

Dombry, C. (2015). Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. *Bernoulli*, **21**, 420–436.

Drees, H. (1998). On smooth statistical tail functionals. *Scandinavian Journal of Statistics*, **25**, 187–210.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1987). *Modelling Extremal Events for Insurance and Finance*, Springer.

Falk, M. (1995). Some best parameter estimates for distributions with finite endpoint. *Statistics*, **27**, 115–125.

Ferreira, A. and de Haan, L. (2015). On the block maxima method in extreme value theory: PWM estimators. *The Annals of Statistics*, **43**, 276–298.

Gardes, L. and Girard, S. (2005). Estimating extreme quantiles of Weibull tail-distributions, *Communication in Statistics - Theory and Methods*, **34**, 1065-1080.

Girard, S. (2004). A Hill type estimate of the Weibull tail-coefficient, *Communication in Statistics - Theory and Methods*, **33**, 205–234.

Grimshaw, S.D. (1993). Computing maximum likelihood estimates for the generalized Pareto distribution. *Technometrics*, **35**(2), 185–191.

References II

Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, **3**, 1163–1174.

Hosking, J.R.M., Wallis, J.R., and Wood, E.F. (1985). Estimation of the Generalized Extreme-Value distribution by the method of probability-weighted moments. *Technometrics*, **27**, 251–261.

Hosking, J.R.M., and Wallis, J.R. (1987). Parameter and quantile estimation for the Generalized Pareto Distribution. *Technometrics*, **29**, 1339–1349.

Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, **3**, 119–131.

Prescott, P., and Walden, A. T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika*, **67**, 723–724.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, **72**, 67–90.

Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, **73**, 812–815.

Zhou, C. (2010). The extent of the maximum likelihood estimator for the extreme value index. *Journal of Multivariate Analysis*, **101**, 971–983.

Chapter 4: Measure of dependence with copulas

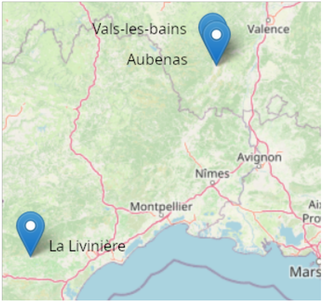
Estimation of extreme risk, application to environmental data

Stephane Girard
Inria Grenoble Rhône-Alpes
http://mistis.inrialpes.fr/people/girard/

2026

Motivation

- **Goal**: To model the dependence between rainfalls of two stations.



- **Multivariate distributions.** The catalog is poor:
  - *Elliptically contoured distributions* (margins = all the same *e.g.* Gaussian, dependence properties = fixed by a "correlation" matrix).
  - *Product of univariate distributions* (margins = arbitrary, dependence properties = independence).
- **Copulas:** Tool to build multivariate distributions with arbitrary margins and dependence properties. $\rightsquigarrow$ notebook4#1

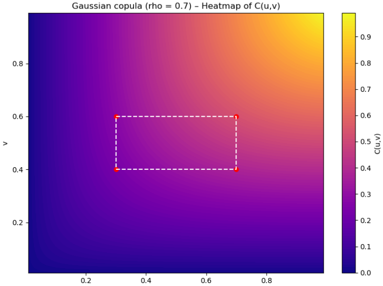
Copula

Definition

A d-variate copula is a cumulative distribution function (cdf) on $[0,1]^d$ with uniform margins.

Characterization. Assume $d = 2$ for the sake of simplicity.

- $C(u,0) = C(0,v) = 0$, $\forall (u,v) \in [0,1]^2$,
- $C(u,1) = u$ and $C(1,v) = v$, $\forall (u,v) \in [0,1]^2$,
- 2-increasing property:
  $C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0$,
  $\forall (u_1,u_2,v_1,v_2) \in [0,1]^4$, such that $u_1 \leq u_2$ and $v_1 \leq v_2$.



Main result

Theorem (Sklar, 1959)

Any (bivariate) cdf H with margins F and G can be written as

$$H(x, y) = C(F(x), G(y))$$

where C is a copula. Moreover, C is unique if the margins are continuous.

Consequence. Copulas are a convenient tool to build multivariate distributions with arbitrary margins (via the choice of F and G) and arbitrary dependence structure (via the choice of C).

Definition

The **survival copula** $\tilde{C}$ associated with C is given by, for all $(u, v) \in [0, 1]^2$.

$$\tilde{C}(u, v) = u + v - 1 + C(1 - u, 1 - v).$$

Application to the computation of the probability of joint exceedances:

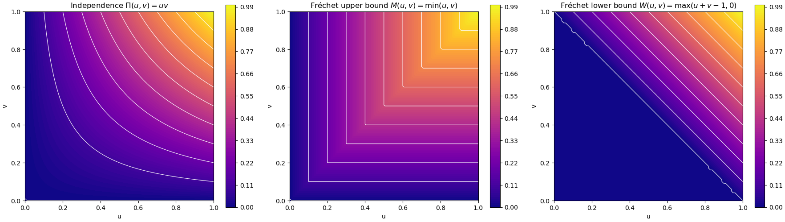
$$\mathbb{P}(X > x,\ Y > y) = \tilde{C}(\bar{F}(x), \bar{G}(y)). \tag{1}$$

Three special copulas

- *Independent* copula $\Pi(u, v) = uv$.
- Fréchet copulas for *comotonic* $M(u, v) = \min(u, v)$ and *anti-comotonic* dependence $W(u, v) = \max(u + v - 1, 0)$.

Property: For all copula C and $(u, v) \in [0, 1]^2$:

$$W(u, v) \leq C(u, v) \leq M(u, v).$$

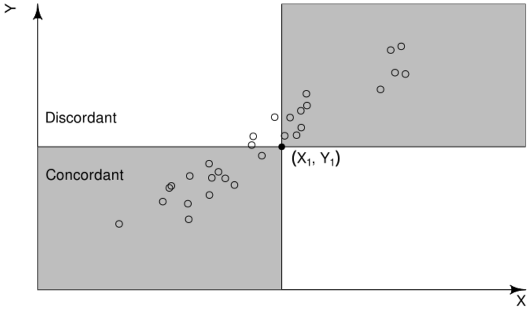


Left: Π, center: M, right: W.

Definition

The **Spearman rho** is three times the probability of concordance minus the probability of discordance between two random pairs (X_1, Y_1) from $H = C(F, G)$ and (X_2, Y_2) from $F \times G$:

$$\rho = 3[\mathbb{P}((X_1 - X_2)(Y_1 - Y_2) > 0) - \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) < 0)].$$



(source: Wikipedia)

Spearman rho

Definition

The **Spearman rho** is three times the probability of concordance minus the probability of discordance between two random pairs (X_1, Y_1) from $H = C(F, G)$ and (X_2, Y_2) from $F \times G$:

$$\rho = 3[\mathbb{P}((X_1 - X_2)(Y_1 - Y_2) > 0) - \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) < 0)].$$

Proposition

- $-1 \leq \rho \leq 1$.
- ρ *only depends on* C *[3, Theorem 5.1.6]:*

$$\rho = 12\mathbb{E}(UV) - 3 = 12 \int_0^1 \int_0^1 C(u, v)\mathrm{d}u\mathrm{d}v - 3,$$

  where (U, V) *is a random pair from* C*.*
- *For special copulas:* $\rho_\Pi = 0$*,* $\rho_M = 1$ *and* $\rho_W = -1$*.*

Spearman rho: inference

Let $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$ be a random sample of (X, Y) from $H = C(F, G)$. Our goal is to use the previous result:

$$\rho = 12\mathbb{E}(UV) - 3 \text{ where } (U = F(X), V = G(Y)) \sim C.$$

The idea is to generate a pseudo-sample $\{(\tilde{U}_1, \tilde{V}_1), \dots, (\tilde{U}_n, \tilde{V}_n)\}$ from C by letting

$$(\tilde{U}_i, \tilde{V}_i) := \left(\frac{\text{Rank}(X_i)}{n+1}, \frac{\text{Rank}(Y_i)}{n+1}\right),$$

for all $i \in \{1, \dots, n\}$ and then

$$\hat{\rho} = \frac{12}{n} \sum_{i=1}^{n} \tilde{U}_i \tilde{V}_i - 3.$$

Kendall tau

Definition

The **Kendall tau** is the probability of concordance minus the probability of discordance between two random pairs (X_1, Y_1) and (X_2, Y_2) from the same cdf:

$$\tau = \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) > 0) - \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) < 0).$$

Proposition

- *Clearly,* $-1 \leq \tau \leq 1$*.*
- τ *only depends on* C *[3, Theorem 5.1.3]*

$$\tau = 4\mathbb{E}\{C(U, V)\} - 1 = 4 \int_0^1 \int_0^1 C(u, v)\mathrm{d}C(u, v) - 1,$$

  where (U, V) *is a random pair from* C*.*
- *For special copulas:* $\tau_\Pi = 0$, $\tau_M = 1$ *and* $\tau_W = -1$*.*

Kendall tau: inference

Similarly, the idea is to use the previous result:

$$\tau = 4\mathbb{E}\{C(U,V)\} - 1 \text{ where } (U = F(X), V = G(Y)) \sim C.$$

as well as the pseudo-sample $\{(\tilde{U}_1, \tilde{V}_1), \ldots, (\tilde{U}_n, \tilde{V}_n)\}$ from C.
The copula is estimated by its empirical counterpart:

$$\hat{C}(u,v) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\{\tilde{U}_i \leq u,\ \tilde{V}_i \leq v\}$$

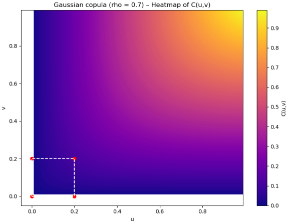
leading to

$$\hat{\tau} = \frac{4}{n}\sum_{j=1}^{n} \hat{C}(\tilde{U}_j, \tilde{V}_j) - 1 = \frac{4}{n^2}\sum_{i,j=1}^{n} \mathbb{I}\{\tilde{U}_i \leq \tilde{U}_j,\ \tilde{V}_i \leq \tilde{V}_j\} - 1. \quad (2)$$

Tail dependence coefficients

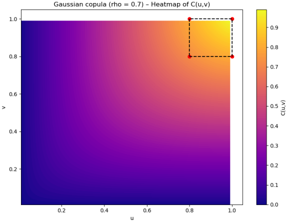
Lower Tail Dependence Coefficient:

$$\begin{aligned}\mathrm{LTDC} &= \lim_{t\to 0} \mathbb{P}(X \le F^{-1}(t) \mid Y \le G^{-1}(t)) \\ &= \lim_{t\to 0} \frac{\mathbb{P}(X \le F^{-1}(t) \cap Y \le G^{-1}(t))}{t} \\ &= \lim_{t\to 0} \frac{C(t,t)}{t}.\end{aligned}$$



Upper Tail Dependence Coefficient:

$$\begin{aligned}\mathrm{UTDC} &= \lim_{t\to 1} \mathbb{P}(X > F^{-1}(t) \mid Y > G^{-1}(t)) \\ &= \lim_{t\to 1} \frac{\mathbb{P}(X > F^{-1}(t) \cap Y > G^{-1}(t))}{1-t} \\ &= 1 + \lim_{t\to 1} \frac{C(t,t)-t}{1-t}.\end{aligned}$$



Tail dependence coefficients

Properties

- $0 \leq \text{LTDC} \leq 1$,
- $0 \leq \text{UTDC} \leq 1$.

Special copulas

- Independence: $\text{LTDC}(\Pi)=\text{UTDC}(\Pi) = 0$.
- Fréchet upper bound: $\text{LTDC}(M)=\text{UTDC}(M) = 1$.
- Fréchet Lower bound: $\text{LTDC}(W)=\text{UTDC}(W) = 0$.
- Gaussian: $\text{LTDC}(\rho)=\text{UTDC}(\rho) = 0$ if the correlation $\rho \neq 1$.
  "*The formula that killed Wall Street*" [2].

Tail dependence coefficients: inference

Focusing on the LTDC defined as

$$\mathrm{LTDC} = \lim_{t \to 0} \frac{C(t,t)}{t},$$

one can replace the copula C by its empirical counterpart $\hat{C}$ as previously. The problem is the computation of the limit. The usual practice is to consider instead the function

$$t \in [0,1] \mapsto \mathrm{LTDC}(t) = \frac{C(t,t)}{t},$$

and to draw, for small values of t:

$$\widehat{\mathrm{LTDC}}(t) = \frac{1}{nt} \sum_{i=1}^{n} \mathbb{I}\{\tilde{U}_i \le t,\ \tilde{V}_i \le t\}.$$

Similarly,

$$\widehat{\mathrm{UTDC}}(t) = \frac{1}{n(1-t)} \sum_{i=1}^{n} \mathbb{I}\{\tilde{U}_i > t,\ \tilde{V}_i > t\}.$$

for t close to 1. $\rightsquigarrow$ notebook4#2

Two examples

Definition

X and Y are **Positively Quadrant Dependent** (PQD) if

$$\forall (x,y) \in \mathbb{R}^2, \;\; \mathbb{P}(X \leq x, \; Y \leq y) \geq \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y).$$

This property can be rewritten in term of copulas as $C \geq \Pi$ *i.e.*

$$\forall (u,v) \in [0,1]^2, \;\; C(u,v) \geq uv.$$

Definition

Y is **Left Tail Decreasing** (LTD) with respect to X if

$$\forall y \in \mathbb{R}, \;\; x \mapsto \mathbb{P}(Y \leq y \mid X \leq x) \text{ is non-increasing.}$$

This property can be rewritten in term of copulas as

$$\forall v \in [0,1], \;\; u \mapsto C(u,v)/u \text{ is non-increasing.}$$

Proposition

LTD $\Rightarrow$ *PQD* $\Rightarrow \tau \geq 0$ *and* $\rho \geq 0$.

Definition

Archimedean copulas are defined by a univariate function $\varphi : [0,1] \to \mathbb{R}_+$ called **generator**:

$$C(u,v) = \varphi^{-1}(\varphi(u) + \varphi(v)). \tag{3}$$

Theorem (Theorem 4.1.4 in [3])

Let $\varphi : [0,1] \to \mathbb{R}_+$ be a continuous, **strictly decreasing** *function such that $\varphi(1) = 0$. Then, the function C given by (3) is a copula if and only if φ is* **convex**.

Examples

- Π and W are Archimedean copulas with generators $\varphi_\Pi(t) = -\log(t)$ and $\varphi_W(t) = 1 - t$.
- M is not an Archimedean copula.
- 22 examples of one-parameter families of Archimedean copulas in Table 4.1 of [3].

Quantitative measures of dependence

Spearman rho:

$$\rho = 12\int_0^1 \varphi^{-1}(2\varphi(t))\mathrm{d}t - 3.$$

Kendall tau:

$$\tau = 1 + 4\int_0^1 \frac{\varphi(t)}{\varphi'(t)}\mathrm{d}t.$$

Lower Tail Dependence Coefficient:

$$\mathrm{LTDC} = \lim_{x\to\infty}\frac{\varphi^{-1}(2x)}{\varphi^{-1}(x)}.$$

Upper Tail Dependence Coefficient:

$$\mathrm{UTDC} = 2 - \lim_{x\to 0}\frac{\varphi^{-1}(2x)}{\varphi^{-1}(x)}.$$

Examples

Gumbel copula

$$C_\theta(u,v) = \exp\left\{-\left[(-\log u)^\theta + (-\log v)^\theta\right]^{1/\theta}\right\}, \quad \theta \in (1,\infty).$$

- Generator: $\varphi_\theta(t) = (-\log t)^\theta$.
- Special/limit cases: $C_1 = \Pi$ and $C_\infty = M$.
- PQD and LTD.
- Kendall tau: $\tau_\theta = (\theta - 1)/\theta$.
- Tail dependence coefficients: $\mathrm{UTDC}_\theta = 2 - 2^{1/\theta}$ and $\mathrm{LTDC}_\theta = 0$.
- For graphical diagnostic (see later): $\lambda_\theta(t) = t\log(t)/\theta$.

Clayton copula

$$C_\theta(u,v) = \left[u^{-\theta} + v^{-\theta} - 1\right]^{-1/\theta}, \quad \theta \in (0,\infty).$$

- Generator: $\varphi_\theta(t) = (t^{-\theta} - 1)/\theta$.
- Special/limit cases: $C_0 = \Pi$ and $C_\infty = M$.
- PQD and LTD.
- Kendall tau: $\tau_\theta = \theta/(\theta + 2)$.
- Tail dependence coefficients: $\mathrm{UTDC}_\theta = 0$ and $\mathrm{LTDC}_\theta = 2^{-1/\theta}$.
- For graphical diagnostic (see later): $\lambda_\theta(t) = t(t^\theta - 1)/\theta$.

Inference

Methodology

- Choose C_θ an Archimedean copula with generator φ_θ where $\theta \in \mathbb{R}$. The choice of this parametric family should be done following the desired dependence properties (PQD, non zero UTDC, etc ...)
- The parameter θ is then estimated from the pseudo-sample $\{(\tilde{U}_1, \tilde{V}_1), \ldots, (\tilde{U}_n, \tilde{V}_n)\}$.

Two main methods are available:

- (pseudo) **Maximum likelihood** estimator,
- **Method of moments**:
    - Compute $\hat{\tau}$ from the pseudo-sample using (2).
    - Assume that an explicit formula $\tau = \Phi(\theta)$ of the Kendall tau is available. Otherwise, one may consider ρ, LTDC, or UTDC.
    - Set $\hat{\theta} = \Phi^{-1}(\hat{\tau})$.

Graphical diagnostic

Definition

The **Kendall distribution function** of the copula C is the cdf

$$t \in [0,1] \mapsto K(t) := \mathbb{P}(C(U,V) \leq t),$$

where (U,V) is a random pair from C.

– If C is an **Archimedean copula** with generator φ_θ then

$$K_\theta(t) = t - \frac{\varphi_\theta(t)}{\varphi'_\theta(t)}. \tag{4}$$

– **Special copulas**

- Independence: $K_\Pi(t) = t - t \log t$.
- Fréchet upper bound: $K_M(t) = t$.
- Fréchet lower bound: $K_W(t) = 1$.

Graphical diagnostic

The graphical diagnostic [1] consists in comparing two estimations of $\lambda(t) := t - K(t)$:

- A parametric estimator $\lambda_{\hat{\theta}}(t) = t - K_{\hat{\theta}}(t)$ based on (4) and an estimator $\hat{\theta}$;
- A non-parametric estimator (without Archimedean assumption) based on the pseudo-observations from the cdf K:

$$W_i = \frac{1}{n-1} \sum_{j \neq i}^{n} \mathbb{I}\left\{ \tilde{U}_j < \tilde{U}_i,\ \tilde{V}_j < \tilde{V}_i \right\},$$

for all $i \in \{1, \ldots, n\}$, leading to

$$\hat{K}(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\{W_i \leq t\}.$$

$\rightsquigarrow$ notebook4#3

Application to the estimation of small tail probabilities

- From (1), one can estimate the probability of **joint extreme exceedances**

$$p_n = \mathbb{P}(X > x_n,\ Y > y_n)$$

with $x_n > \max(X_1, \ldots, X_n)$ and $y_n > \max(X_1, \ldots, Y_n)$ by

$$\hat{p}_n = \tilde{C}_{\hat{\theta}}(\hat{\bar{F}}(x_n), \hat{\bar{G}}(y_n)),$$

where $\hat{\bar{F}}(x_n)$ and $\hat{\bar{G}}(y_n)$ are computed either by the Peaks Over Threshold or the Block Maxima approach, See Chapter 3.

- Similarly, the probability of **conditional extreme exceedances**

$$p'_n = \mathbb{P}(X > x_n \mid Y > y_n)$$

can be estimated by

$$\hat{p}'_n = \frac{\tilde{C}_{\hat{\theta}}(\hat{\bar{F}}(x_n), \hat{\bar{G}}(y_n))}{\hat{\bar{G}}(y_n)}.$$

$\rightsquigarrow$ notebook4#4

References

- Genest, C., & Rivest, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association*, 88(423), 1034–1043.
- MacKenzie, D., & Spears, T. (2014). "The formula that killed Wall Street": The Gaussian copula and modelling practices in investment banking. *Social Studies of Science*, 44(3), 393–417.
- Nelsen, R. (2006). *An Introduction to Copulas*, Springer.