Statistical extreme-value analysis for climate data

Thomas Opitz, BioSP, INRAE Avignon

Master Climaths, École Centrale Marseille
2020–2021

Organisation details of the course

- Four sessions : 26/01/ 13.00–17.00, 02/01/ 13.30–17.30, 16/02/ 13.30–17.30, 16/03/ 13.30–17.30
- Depending on sanitary measures (Covid19), some or all of the sessions will be held online.
- Some supplementary material (lecture slides, data, R code) is provided online (Moodle platform)
- Graded evaluation based on two components (contrôle continu) :
  - written mini-test ($\approx$ 30mins) in session 4
  - mini-project (in groups of two student)
    - written report (such as a well-structured and commented RMarkdown output)
    - oral presentation $\sim$ 10 minutes + 5 minutes discussion (date to be fixed)
- Contact : `Thomas.Opitz@inrae.fr`

Outline of the course

1. Introduction : Motivation of statistical extreme-value analysis
2. The block maximum approach
3. The peaks-over-threshold approach
4. Dependent extremes
5. Discussion of approaches and extensions

Relevant literature :

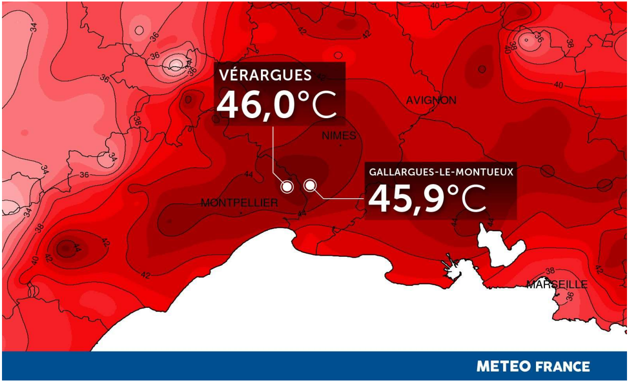
- Coles (2001). An Introduction to Statistical Modeling of Extreme Values
- Beirlant et al. (2006). Statistics of extremes : theory and applications
- Dey, Yan (2016). Extreme value modeling and risk analysis : methods and applications

A recent example : Alpes-Maritimes

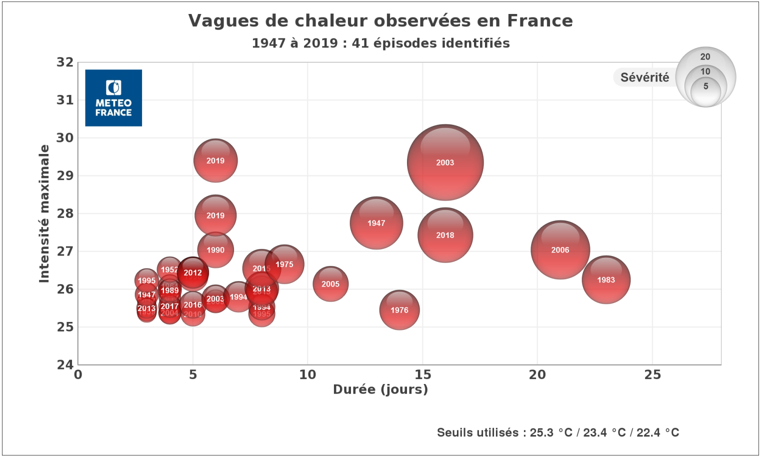
Storm Alex in October 2020 :
Extreme rainfalls and wind speeds led to an extreme flooding event in the valleys of the Alpes-Maritimes département

Another recent example : 2019 heatwave(s) in France

Extreme temperatures have been observed over a large area of France, with a new all-time record for France



Classification of heatwaves in France 1947–2019

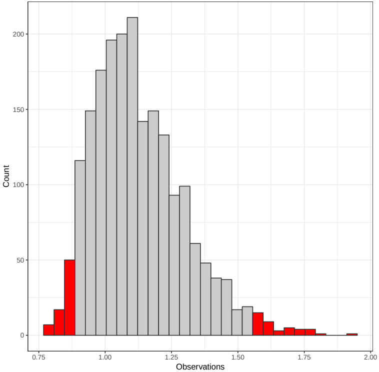


Environmental and meteorological extreme events

- heatwaves
- cold spells
- Mediterranean precipitation episodes
- floods (due to extreme precipitation)
- wave heights at the coast
- windstorms
- air pollution episodes
- large wildfires
- ...

We will analyze such events through a general probabilistic theory known as **Extreme-Value Theory**.

Distribution tails



- on the left : **lower tail**
- on the right : **upper tail**
- central region : **bulk**

There is no exact definition where the bulk "ends" and where the tail "starts" – it depends on the context !

⚠ Central limit theory is concerned with the grey area.

⚠ Extreme value theory is concerned with the red areas.

Two key concepts : return levels and return periods

Given a probability distribution F and independent observations

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

we can define the the following two important concepts :

- **return level** $\mathrm{RL}(T)$ (given a fixed return period T) = the quantile of the distribution that is exceeded once every T time units on average, i.e.,

$$1 - F(\mathrm{RL}(T)) = 1/T$$

- **return period** $\mathrm{RP}(x)$ (given a fixed return level x) = the number of time units that it takes on average to observe an exceedance above the level x, i.e.,

$$\mathrm{RP}(x) = 1/(1 - F(x))$$

Utility of extreme-value analysis

How can we understand and predict the statistical distribution of extreme events of environmental or climatic phenomena ?

- estimate return periods (given a fixed return level)

- estimate return levels (given a fixed return period)

- estimate **duration** and **spatial extent** of extreme event episodes

Aims :

- anticipate and prevent future damages and economic and human losses
- improve extreme risk management policies for better risk prevention

Extrapolation

In extreme-value analyses, we are interested in **rare events** characterized by return levels for relatively large T or by return periods for relatively large x.

Empirical estimation of such parameters based on the empirical distribution function is often not possible or not recommended.

Examples :

- return period for quantile x chosen larger than all values observed so far
  ⚠ empirical exceedance probability of x is 0 ("impossible event"), and the empirical return period is $+\infty$
- return level of $T = 100$ years
  ⚠ if we have less than 100 years of data, the empirical return level is $+\infty$

⇒ We need an estimation framework that allows for appropriate **extrapolation** beyond the range of observed values.

⇒ **Uncertainty quantification** of estimations and predictions is crucial in this data-scarce setting.

Notations used throughout this course

- $X \mid A$ refers to the distribution of the random variable X conditional to the event A ; for example, $X \mid X > u$ corresponds to X conditional to an exceedance of the threshold u

- $x_+ = \max(x, 0)$

- vector notations for componentwise operations ; for example, $(x_1, \ldots, x_d) > (y_1, \ldots, y_d)$ means that $x_j > y_j$ for all components $j = 1, \ldots, d$

Recall : The Central Limit Theorem

The CLT is the fundamental limit theorem for the central tendency of a distribution.

We suppose that $X_1, X_2, \ldots \sim F$ are independent, and $\mathbb{E}[X^2]$ is finite.

Central Limit Theorem :

$$\sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \xrightarrow{d} \mathcal{N}(0,1), \quad n \to \infty.$$

- characterizes the large-sample behavior of the sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$
- applies a linear location-scale normalisation :

  location $\mu = \mathbb{E}[X]$ and scale $\sigma = \sqrt{\mathbb{V}[X]}$
- convergence in distribution
- if $X \sim \mathcal{N}(\mu, \sigma^2)$, then the limit relation is exact
  (i.e., we can replace "$\to$" by "$=$" for any $n \in \mathbb{N}$)

⚠ The convergence of the sample mean is not informative about the behavior of observations in the extreme regions.
$\Rightarrow$ We need a *different* **limit theory adapted to extremes**.

The fundamental extreme-value limit theorem

For a sequence of independent and identically distributed (iid) random variables $X_i \sim F$, $i = 1, 2, \ldots$, consider the componentwise maximum

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

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

Fisher–Tippett–Gnedenko Theorem

If normalizing sequences a_n (location) and $b_n > 0$ (scale) exist such that

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

with a nondegenerate limit distribution G, then G is one of the three types of **extreme-value distributions** : Weibull, Gumbel or Fréchet.
If convergence $(\star)$ holds, we say that ***F* is in the maximum domain of attraction (MDA) of *G***.

The theorem is also known as the **Extremal Types theorem**, since it states the types of limit distributions that may arise.
Equivalently to $(\star)$, we can write

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

Characterizing the class of limit distributions

The class of limit distributions G coincides with the class of **max-stable distributions** : with appropriate choices of normalizing sequences α_n and $\beta_n > 0$,

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

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

The **Generalized Extreme-Value distributions (GEV)** uses threes parameter to jointly represent all three types of max-stable distributions :

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

- shape parameter $\xi \in \mathbb{R}$
- location parameter $\mu \in \mathbb{R}$
- scale parameter $\sigma > 0$

For $\xi = 0$, $(\star\star)$ is understood as the limit for $\xi \to 0$, i.e. $G(z) = \exp(-\exp(-(z - \mu)/\sigma))$.

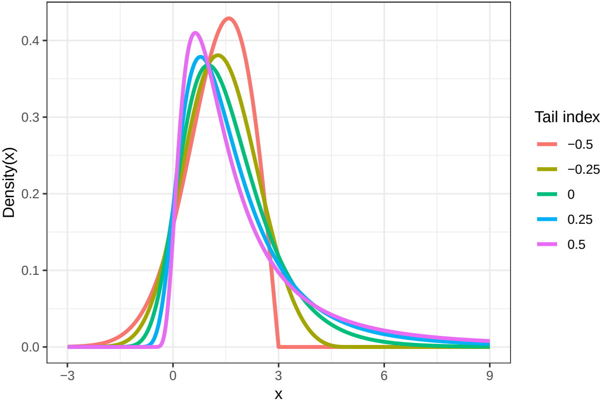
The $(\ldots)_+$-operator in $(\star\star)$ means that the distribution G has positive density dG/dz for values z satisfying $1 + \xi \frac{z-\mu}{\sigma} > 0$.

Details about the GEV distribution

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

Three types with very different structure arise depending on the tail index :

- Weibull MDA corresponding to $\xi < 0$: light tails with finite upper endpoint
- Gumbel MDA corresponding to $\xi = 0$: exponential tail behavior
- Fréchet MDA corresponding to $\xi > 0$: power-law tails, i.e., relatively heavy tails



Example : GEV limit of the exponential distribution

Standard exponential distribution

$$F(x) = \begin{cases} 0, & x \leq 0, \\ 1 - \exp(-x), & x > 0. \end{cases}$$

The distribution F^n of $M_n = \max_{i=1}^n X_i$, where $X_i \overset{iid}{\sim} F$, $i = 1, \ldots, n$, is

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

How can we find a_n and b_n such that $\lim_{n\to\infty} F^n(a_n + b_n x)$ exists and is nontrivial ?
A well-known results states that $(1 - x/n)^n \to \exp(-x)$ as $n \to \infty$, for any $x \in \mathbb{R}$.
Using this, we get that for all $x \in \mathbb{R}$,

$$\begin{aligned} \exp(-\exp(-x)) &= \lim_{n\to\infty} \left(1 - \frac{exp(-x)}{n}\right)^n = \lim_{n\to\infty} (1 - exp(-(x + \log(n)))^n \\ &= \lim_{n\to\infty} F^n(x + \log(n)). \end{aligned}$$

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 $\xi = 0, \mu = 0, \sigma = 1$.

Return levels and return periods for the GEV

The **return level** $\mathrm{RL}(T)$ for a given return period of $T > 0$ corresponds to a quantile whose value is exceeded on average once every T time units.

For the $\mathrm{GEV}(\xi, \mu, \sigma)$ distribution, the T-return level is given by

$$\mathrm{RL}(T; \xi, \mu, \sigma) = G^{-1}\left(1 - \frac{1}{T}\right) = \mu + \sigma \frac{(-\log(1 - 1/T))^{-\xi} - 1}{\xi}.$$

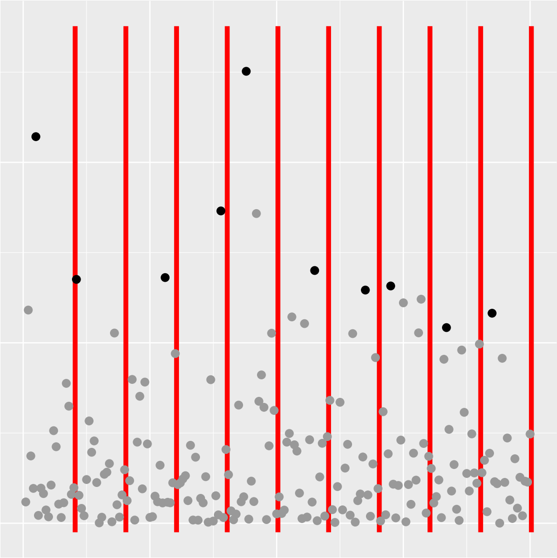
The **return period** for a given quantile z corresponds to a period over which the level z is exceeded once on average.

For the $\mathrm{GEV}(\xi, \mu, \sigma)$ distribution, the z-return period is given by

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

Estimation through the block maxima approach

To obtain a sample of maxima, we can divide the original dataset into m blocks of equal length, and then extract the maximum of each block (yearly maxima etc.).



Maximum likelihood estimation of the GEV distribution

Probability density function $\mathrm{d}G(z)/\mathrm{d}z$ of the GEV distribution :

$$f_{\mathrm{GEV}}(z;\xi,\mu,\sigma) = \frac{1}{\sigma}\, t(x)^{\xi+1} e^{-t(x)} \quad \text{with} \quad t(x) = \begin{cases} \left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi} & \text{if } \xi \neq 0 \\ e^{-(x-\mu)/\sigma} & \text{if } \xi = 0 \end{cases}$$

Log-likelihood function given a sample $z_1, \ldots, z_m$ of maxima for m blocks :

$$\ell(\xi,\mu,\sigma) = \ell(\xi,\mu,\sigma; z_1,\ldots,z_m) = \sum_{i=1}^{m} \log\left(f_{\mathrm{GEV}}(z_i;\xi,\mu,\sigma)\right)$$

The **maximum likelihood (ML) estimator of the GEV parameters** is obtained as follows :

$$(\hat{\xi},\hat{\mu},\hat{\sigma}) = \arg\max_{(\xi,\mu,\sigma)} \ell(\xi,\mu,\sigma). \quad (\star\star\star)$$

Practical implementation :

- The ML equation system $\mathrm{d}\ell/\mathrm{d}\xi = 0$, $\mathrm{d}\ell/\mathrm{d}\sigma = 0$, $\mathrm{d}\ell/\mathrm{d}\mu = 0$ does not have a closed-form solution.
- Instead, the maximization of $(\star\star\star)$ can be done numerically (for example, iteratively through *gradient descent*).

Estimation with covariates

The distribution of maxima may be nonstationarity such that $G_{i_1} \neq G_{i_2}$, where $Z_{i_1} \sim G_{i_1}$, $Z_{i_2} \sim G_{i_2}$ and $1 \leq i_1 \neq i_2 \leq m$.

We can model nonstationarity of G_i by allowing extreme-value parameters ξ, μ, σ to vary with time or with other auxiliary variables.

Example : a **time trend** model for the GEV distribution

- suppose that $t_i = t_0 + i$ represents the year of observation of Z_i, $i = 1, \ldots, m$
- a possible model structure is as follows :

$$\begin{aligned}
\mu(t) &= \mu_0 + \mu_1 \times t \\
\sigma(t) &= \exp(\sigma_0 + \sigma_1 \times t) \\
\xi(t) &= \xi
\end{aligned}$$

  with parameters $\mu_0, \mu_1, \sigma_0, \sigma_1, \xi \in \mathbb{R}$
- ⚠ models with nonstationary tail index ξ may lead to unstable numerical estimation, or may be difficult to interpret
- this model could be used to assess climate change effects

QQ-plots and PP-plots

Goal : check if an estimated distribution F shows a good fit to a sample $Z_1, \dots, Z_m$; that is, if

$$Z_i \overset{i.i.d.}{\sim} F, \quad i = 1, \dots, m.$$

We consider **empirical and theoretical quantiles** :

(i) ordered sample : $Z_{(1)} \leq Z_{(2)} \leq \cdots \leq Z_{(n)}$

(ii) theoretical quantiles of the model : $F^{-1}\left(\frac{1}{n+1}\right), F^{-1}\left(\frac{2}{n+1}\right), \cdots, F^{-1}\left(\frac{n}{n+1}\right)$

The **QQ-plot** (*quantile-quantile*) shows

(i) (usually on y-axis) plotted against (ii) (usually on x-axis).

If F is an adequate distribution for the sample, then points should be aligned along the diagonal, especially for intermediate order statistics $Z_{(k)}$ with k away from the extremes 1 and n.

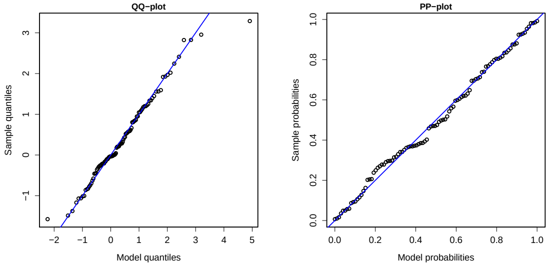
PP-plots (*probability-probability*) use a probability scale, that is, values in $[0, 1]$:

(i) ordered sample : $F\left(Z_{(1)}\right) \leq F\left(Z_{(2)}\right) \leq \cdots \leq F\left(Z_{(n)}\right)$

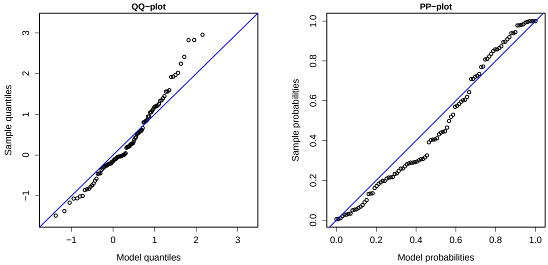
(ii) probabilities : $\frac{1}{n+1}, \frac{2}{n+1}, \cdots, \frac{n}{n+1}$

Examples of QQ- and PP-plots

Good fit :



Not satisfactory fit :



Other likelihood-based tools

The general **asymptotic theory of maximum likelihood estimation** applies specifically to the extreme-value context :

- asymptotic normality of estimators
- confidence intervals
- statistical tests
- information criteria : AIC, BIC...
- ...

Threshold exceedances

Suppose that $X_i \overset{i.i.d.}{\sim} F$.

Instead of considering maxima, we can also consider **threshold exceedances**

$$X_i - u \mid X_i \geq u$$

above a high threshold u.

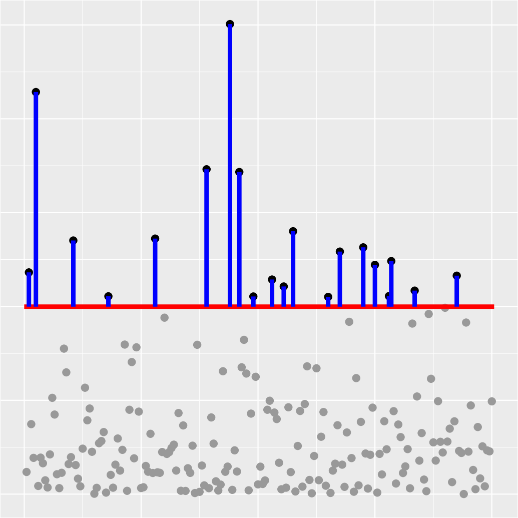
⚠ What "high" means depends on the context, but u should correspond to a high quantile of the distribution F of X_i, $i = 1, \ldots, n$.

Theoretical result : If F is in the maximum domain of attraction of a GEV distribution G, then there exists an equivalent limit for threshold exceedances.

The use of threshold exceedances is also known as the **Peaks-over-Threshold approach**.

Peaks-over-Threshold

To obtain a sample of threshold exceedances, we can set a threshold u and extract the excesses $X - u$ above the threshold.



The Peaks-over-Threshold limit theorem

Pickands–Balkema–de Haan Theorem

Suppose that $X \sim F$, where F is in the MDA of a (non-degenerate) $\mathrm{GEV}(\xi, \mu, \sigma)$-distribution ; that is,

$$F^n(a_n + b_n z) \to \mathrm{GEV}(z; \xi, \mu, \sigma), \quad n \to \infty, \quad z \in \mathbb{R}.$$

Then

$$\Pr(X \leq a_n + b_n(y+u) \mid X \geq a_n + b_n u) \to H(y), \quad n \to \infty, \quad y > 0,$$

where H is a **generalized Pareto distribution**, written $\mathrm{GPD}(\xi, \sigma_{\mathrm{GP}})$, given by

$$H(y) = 1 - \left(1 + \xi \frac{y}{\sigma_{\mathrm{GP}}}\right)_+^{-1/\xi}, \quad y > 0,$$

with $\sigma_{\mathrm{GP}} = \sigma + \xi(\mu - u) > 0$.

- The $\mathrm{GPD}(\xi, \sigma_{\mathrm{GP}})$ inherits the parameter constraints from the GEV distribution.
- As before, the case $\xi = 0$ is interpreted as the limit for $\xi \to 0$:
  $H(y) = 1 - \exp(-y/\sigma_{\mathrm{GP}})$ for $y > 0$ (exponential distribution with scale σ_{GP}).
- One can show that the MDA condition and the existence of the threshold exceedance limit are equivalent.

Sketch of the proof

1. $\Pr(X \geq a_n + b_n(y+u) \mid X \geq a_n + b_n u) = \frac{1-F(a_n+b_n(y+u))}{1-F(a_n+b_n u)}$

2. The MDA condition $F^n(a_n + b_n z) \to G(z)$ implies
$$n \log\left(F(a_n + b_n z)\right) \approx \log\left(G(z)\right).$$

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

4. By using 2) and 3) in 1), we obtain
$$\Pr(X \geq a_n + b_n(y+u) \mid X \geq a_n + b_n u) \to \frac{\log G(u+y)}{G(u)} = \mathrm{GPD}(\xi, \sigma_G), \quad n \to \infty.$$

Point 4) highlights the link between the formulas of the GEV distribution and the GPD.

Peaks-over-threshold stability

By analogy with **max-stability** of limit distributions for maxima, we have **threshold stability** for limit distributions of threshold exceedances :

Threshold stability of the GPD

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

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

- Application of the peaks-over-threshold approach to a GPD yields again a GPD.
- The link between the two distributions as as follows :

$$\mathrm{GPD}((\tilde{\sigma}_{\mathrm{GP}}/\sigma_{\mathrm{GP}})y; \xi, \tilde{\sigma}_{\mathrm{GP}}) = \mathrm{GPD}(y; \xi, \sigma_{\mathrm{GP}}).$$

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

Return levels of the GPD

The return level of the excesses $Y = (X - u) \mid (X > u)$ modeled by the GPD for a given a return period T is

$$\mathrm{RL}_{\mathrm{GP}}(T; \xi, \sigma_{\mathrm{GP}}) = \sigma_{\mathrm{GP}} \frac{T^{\xi} - 1}{\xi}.$$

In most cases, we want to **predict quantiles or return levels at the original scale** (for $X \sim F$) and not for excesses Y. Therefore, we combine the GPD parameters with the exceedance probability p_u of the threshold u, where $p_u = 1 - F(u)$.

Assume that we want to predict the return level $x_T = \mathrm{RL}(T; \xi, \sigma_{\mathrm{GP}}, p_u)$ on the original scale for a return period $T > 1/p_u$. Since

$$1/T = \mathrm{Pr}(X > x_T) = \mathrm{Pr}(X > x_T \mid X > u)\mathrm{Pr}(X > u) = (1 - H(x_T - u))p_u,$$

with $H = \mathrm{GPD}(\xi, \sigma_{\mathrm{GP}})$, we have to replace T in $(\star)$ by Tp_u and add the threshold u :

$$x_T = \mathrm{RL}(T; \xi, \sigma_{\mathrm{GP}}, p_u) = u + \sigma_{\mathrm{GP}} \frac{(Tp_u)^{\xi} - 1}{\xi}.$$

Similar aspects must be considered for the calculation of return periods.

Usage of the generalized Pareto distribution

Suppose that X_i, $i = 1, \ldots, n$ is an i.i.d. sample of F.

A potential advantage of the GPD is that it refers more directly to **original events**, and not to maxima.

How to estimate a GPD model

1. Fix a high threshold u
   (for example, corresponding to the empirical 95%-quantile).
2. Use the following asymptotically motivated modeling assumption :

$$X - u \mid X > u \sim \mathrm{GPD}(\xi, \sigma_{\mathrm{GP}})$$

3. Extract the sample of exceedances Y_j, $j = 1, \ldots, m$ where $Y_j = X_i - u$ for $X_i > u$.
4. Compute estimators $\hat{\xi}, \hat{\sigma}_{\mathrm{GP}}$ of the GPD parameters based on the sample $\{Y_j\}$
   (for instance, using maximum likelihood).

By choice of u, the sample size m represents only the "extreme" sample fraction and is much smaller than n.

Maximum likelihood estimation of the GPD

- We use a **sample of exceedances** Y_j, $j = 1, \ldots, m$.
- We can proceed as for the GEV by using the corresponding **log-likelihood function** $\ell(\xi, \sigma_{\mathrm{GP}})$ of the GPD.
- As for the GEV distribution, no closed-form expressions are available for the ML estimators, therefore we have to resort to **numerical optimization**.
- Fitting the GPD to exceedances yields the ML estimators, denoted $\hat{\xi}$, $\hat{\sigma}_{\mathrm{GP}}$. The usual tools associated to ML estimators are available.
- Note : The numerical values of the GEV and GPD estimators of $\hat{\xi}$ are of course different since we do not use the same data. However, they should be relatively close if models are consistent with asymptotic theory.

Threshold selection and validation

How to select an *appropriate* threshold ?

- There is a **bias-variance compromise** in estimation :
  - too low a threshold may lead to bias of the asymptotic models
  - too high a threshold leaves too few exceedance observations and leads to higher estimation variance

  (a similar statement can be made with respect to the block size for block maxima)

- In some cases, threshold values may be suggested by physical or policy considerations.

  Example : Air Quality Standards of the European Union, such as *not more than* 35 *days with more than* 50 μg *of* PM_{10} $\Rightarrow$ application-relevant threshold $u = 50$.

- Another possibility is to estimate parameters for different thresholds, and to check if the estimated parameters satisfy (approximately) the threshold stability properties ($\Rightarrow$ **parameter stability plot**).

- There are visual tools based on theoretical considerations, such as the **Mean Excess plot**.

Mean excess plots are also known as **mean residual life plot**.

Theoretical property of the GPD **:** if $Y \sim \mathrm{GPD}(\xi, \sigma_{\mathrm{GP}})$ with $\xi < 1$, then

$$\mathrm{ME}(\tilde{u}) = \mathbb{E}\left[Y - \tilde{u} \mid Y > \tilde{u}\right] = \frac{\sigma_{\mathrm{GP}}}{1-\xi} + \frac{\xi}{1-\xi}\tilde{u}.$$

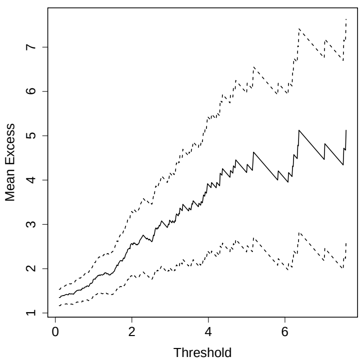
(this is a consequence of threshold stability)
$\Rightarrow$ the **mean excess** $\mathrm{ME}(\tilde{u})$ **is a linear function of the threshold** $\tilde{u}$.

Remark : The expectation of the GPD and the GEV is well-defined for $\xi < 1$, and the variance is well-defined for $\xi < 1/2$.

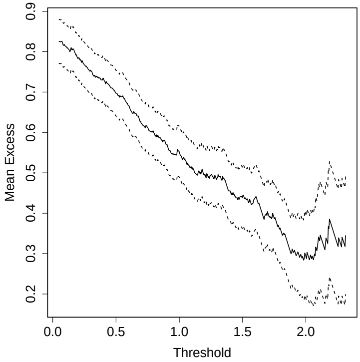
Mean-excess-based threshold selection in practice

1. Define a sequence of thresholds u_k, $k = 1, \ldots, K$, for the data $X \sim F$, where u_k increases towards the upper endpoint of X.
2. Compute the empirical mean excess $\mathrm{ME}(u) = \mathbb{E}\left[X - u \mid X > u\right]$ for each threshold.
3. Plot $\mathrm{ME}(u)$ against u (using the computed values).
4. Select the lowest u_{k_0} for which you see a linear behavior (up to statistical uncertainty) for $k \geq k_0$

Examples of mean excess plots



(data from student's t distribution)



(data from Gaussian distribution)

In the above MEPs, the lowest threshold u_1 was set to 0.

Remark : $\xi > 0$ for the student's t distribution ; $\xi = 0$ for the Gaussian distribution

Nonstationary modeling of threshold exceedances

- The distribution of threshold exceedances, and the exceedance probability above a threshold u, may be **nonstationary**. We can also fix a nonstationary threshold u_i.

- As in GEV-based modeling of maxima, we can include covariates into the GPD model.

- In contrast to the modeling of block maxima, nonstationary behavior within blocks must be taken into account when using threshold exceedances.

  Example : **seasonal behavior** is not relevant for yearly maxima, but it is for threshold exceedances of daily data.

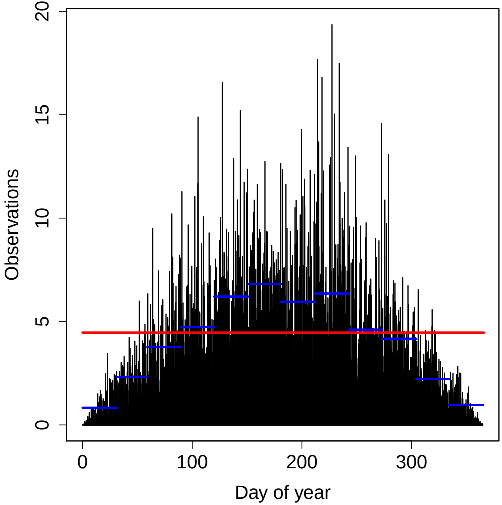
- We may want to use a nonstationary threshold u_i such that the exceedance probability $p_{u_i} = \mathrm{Pr}(X_i > u_i)$ is (approximately) constant (for instance, use a month-specific threshold).

- ⚠ Nonstationary peaks-over-threshold modeling is relatively complex.

- To apply stationary POT models, we can restrict modeling to a subset of data where data can be assumed to be (approximately) stationary.

  Example : consider only summer temperature extremes, or only autumn precipitation extremes.

Example : Nonstationary setting



Constant threshold at 85% level
Monthly threshold at 85% level

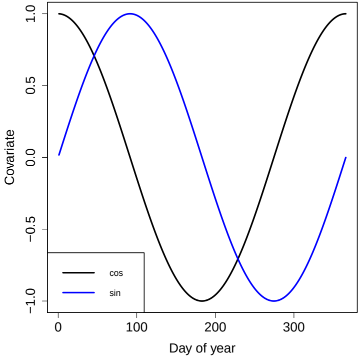
Example : a POT model for seasonal nonstationarity

- If $t = 1, \ldots, 366$ refers to the day of the year, we can use covariates defined as

$$c_t = \cos(2\pi \times t/366), \quad s_t = \sin(2\pi \times t/366).$$

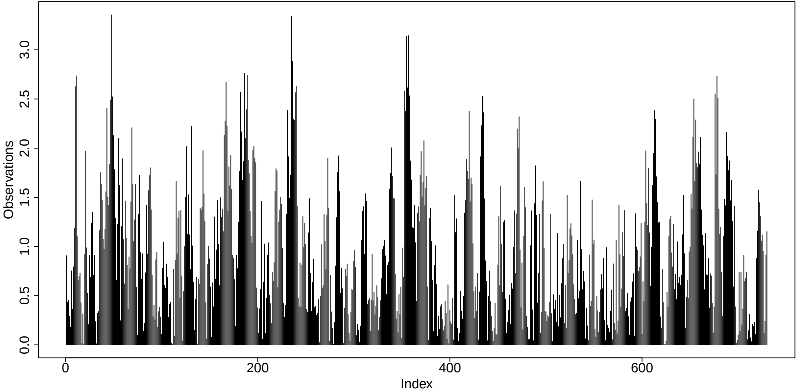
- A possible model with seasonal variation for excesses $Y = X_t - u > 0$ above a threshold u_t is as follows :

$$\sigma_{\mathrm{GP}}(t) = \exp(\sigma_0 + \sigma_1 c_t + \sigma_2 s_t)$$
$$\xi(t) = \xi$$



Here, approximately, c_t could capture differences between winter and summer, and s_t differences between spring and autumn.

Illustration : Stationary dependent sequences



Time series setting

We assume that the variables $X_1, X_2, X_3 \ldots$ define a **stationary time series**, such that blocks of the same length possess the same joint distribution :

$$(X_{k_1+1}, \ldots, X_{k_1+m})^T \stackrel{d}{=} (X_{k_2+1}, \ldots, X_{k_2+m})^T \quad \text{for any } k_1, k_2, m \geq 1.$$

⚠ The assumption of a natural ordering of the indices $1, 2, \ldots$ was not necessary for the previous approaches based on i.i.d. variables.

⚠ We do not not assume that variables X_i, $i = 1, 2 \ldots$ are mutually independent.

⚠ For simplicity, we first do not consider nonstationarities (which we often encounter in practice).

Possible consequences of dependent observations :

- the asymptotic distribution of maxima $M_n = \max_{i=1}^n X_i$ over n consecutive variables may not exist, or may be different from the GEV distribution arising in the i.i.d. case ;
- consecutive excesses over a high threshold may be dependent, for instance by arising in **clusters**.

A mixing condition for extremes

Mixing properties refer to events tending towards some form of independence when they become farther separated in time.

Definition : The $D(u_n)$-condition for extremes

A stationary series $X_1, X_2, \ldots$ satisfies the $D(u_n)$-condition if, for all $i_1 < i_2 < \ldots i_p < j_1 < j_2 < \ldots j_q$ with $j_1 - i_p > \ell$ we have that

$$|\mathrm{Pr}(X_{i_1} \leq u_n, X_{i_2} \leq u_n, \ldots, X_{i_p} \leq u_n, X_{j_1} \leq u_n, X_{j_2} \leq u_n, \ldots, X_{j_q} \leq u_n)$$
$$- \mathrm{Pr}(X_{i_1} \leq u_n, X_{i_2} \leq u_n, \ldots, X_{i_p} \leq u_n)\mathrm{Pr}(X_{j_1} \leq u_n, X_{j_2} \leq u_n, \ldots, X_{j_q} \leq u_n)| \leq \alpha(n, \ell),$$

where $\alpha(n, \ell_n) \to 0$ for some sequence ℓ_n with $\ell_n/n \to 0$ as $n \to \infty$.

Remarks :

- The structure of the sequence of thresholds u_n is not (yet) specified here.
- The probability terms in the first and second row are equal if the two blocks $(X_{i_1}, X_{i_2}, \ldots, X_{i_p})$ and $(X_{j_1}, X_{j_2}, \ldots, X_{j_p})$ are exactly independent.
- The gap ℓ_n between the two blocks $(X_{i_1}, X_{i_2}, \ldots, X_{i_p})$ and $(X_{j_1}, X_{j_2}, \ldots, X_{j_p})$ must increase less fast than n.
- **Interpretation :** the vectors $(X_{i_1}, X_{i_2}, \ldots, X_{i_p})$ and $(X_{j_1}, X_{j_2}, \ldots, X_{j_p})$ are close to being independent at levels u_n if they are separated by a large enough gap ℓ_n.
- ⚠ The $D(u_n)$-condition is important for theoretical results, but it is usually hard to verify in practice.

Extreme-value limits under serial dependence

Theorem

Suppose that $X_1, X_2, \ldots$ is a stationary time series, and define $M_n = \max_{i=1}^n X_i$. If there are sequences a_n and $b_n > 0$ such that

$$\Pr\left(\frac{M_n - a_n}{b_n} \leq z\right) \to G(z), \quad n \to \infty, \quad z \in \mathbb{R},$$

with a nondegenerate limit distribution G, and if G satisfies the $D(u_n)$-condition with $u_n = a_n + b_n z$ for all z, then G is a generalized extreme value distribution.

Counterexample : if $X_1 = X_2 = \ldots = X_n$ almost surely (perfect dependence) and $X_1 \sim F$, then $M_n \sim F$, and F can be any distribution.

Remark : The GEV parameters ξ, μ, σ of the GEV distribution G may be different from the case of an i.i.d. series $X_1, X_2 \ldots$

The extremal index

We keep the same assumptions as in the previous theorem.

In addition, assume that $\tilde{X}_1, \tilde{X}_2 \ldots$ is an i.i.d. sequence with $\tilde{X}_1 \overset{d}{=} X_1$, and define $\tilde{M}_n = \max_{i=1}^n \tilde{X}_i$.

Then we have

$$\Pr\left(\frac{M_n - a_n}{b_n} \le z\right) \to G(z), \quad n \to \infty, \quad z \in \mathbb{R},$$

if and only if

$$\Pr\left(\frac{\tilde{M}_n - a_n}{b_n} \le z\right) \to \tilde{G}(z), \quad n \to \infty, \quad z \in \mathbb{R}.$$

In that case, the two GEV limit distributions G and $G^\star$ are linked as follows :

$$G(z) = \tilde{G}^\theta(z),$$

where $\theta \in]0,1]$ is the **extremal index**.

$\Rightarrow$ if G is the $\mathrm{GEV}(\xi, \mu, \sigma)$, then $\tilde{G}$ is the $\mathrm{GEV}(\xi, \tilde{\mu}, \tilde{\sigma})$ with

$$\mu = \tilde{\mu} - \frac{\tilde{\sigma}}{\xi}\left(1 - \theta^\xi\right), \quad \sigma = \tilde{\sigma}\theta^\xi.$$

Interpretation of the extremal index for maxima

- With $\theta \in]0,1]$, we have $G(z) \geq \tilde{G}(z)$ for all z
  $\Rightarrow$ **stochastic dominance** of the distribution $\tilde{G}$ over G.

- If $\theta < 1$, then the maxima M_n of the dependent series are on average smaller than the ones of $\tilde{M}_n$.

- **Example :** standard Gumbel distribution ($= \mathrm{GEV}(\xi = 0, \mu = 1, \sigma = 1)$) for $\tilde{G}$
  $\Rightarrow \tilde{G}^{\theta}(z) = \exp(-\exp(-z))^{\theta} = \exp(-\theta \exp(-z)) = \exp(-\exp(-(z - \log\theta)))$
  $\Rightarrow$ the distribution G is shifted by $\log(\theta)$.

Extremal index and threshold exceedances

An alternative definition of the extremal index is as the **reciprocal of the limit of the expected cluster size** :

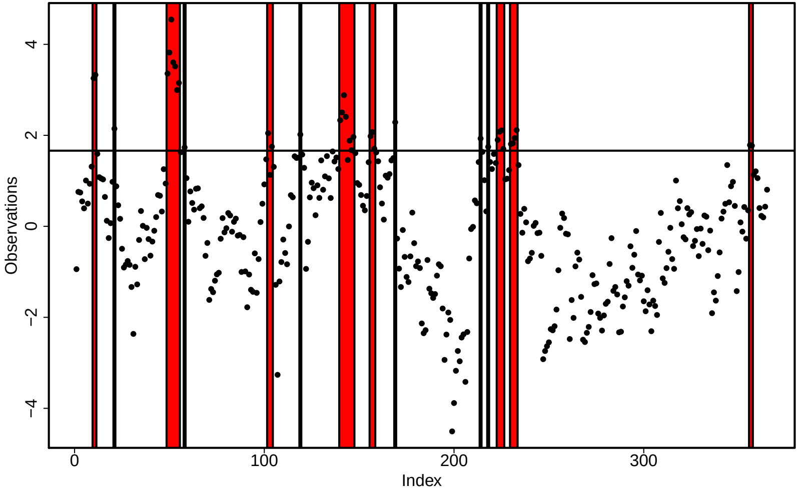
$$\frac{1}{\theta} = \lim_{n\to\infty} \mathbb{E}\left(\sum_{i=1}^{\rho_n} \mathbb{I}(X_i > u_n) \mid M_{\rho_n} > u_n\right)$$

where

- $\mathbb{I}(A) = 1$ if the event A occurs, and 0 otherwise, is the indicator function,
- as before, $u_n = a_n + b_n z$ with a_n, b_n the normalizing sequences for the maximum domain of attraction,
- $n(1 - F(u_n)) \to \lambda$ with (arbitrary) $0 < \lambda < \infty$, and
- $\rho_n = o(n)$, that is, $\rho_n/n \to 0$ as $n \to \infty$.

Interpretation : If at least one X_i exceeds a very high threshold in a block of consecutive observations, then on average there are $1/\theta$ observations that exceed this threshold.

Illustration : Clusters of exceedances



The average length of the intervals in red tends to $1/\theta$ when we increase the threshold (black vertical line).

Tail autocorrelation function

Suppose that $X_i \sim F$, $i = 1, 2 \dots$ Given that we exceed a high threshold, what is the probability that we exceed the same threshold again exactly h time steps later ?

Given a temporal lag $h \in \{0, 1, 2 \dots\}$, consider the conditional exceedance probability

$$\chi(h; u) = \Pr(F(X_{i+h}) > u \mid F(X_i) > u) = \frac{\Pr(F(X_{i+h}) > u, F(X_i) > u)}{\Pr(F(X_i) > u)}, \quad u \in]0, 1[\,.$$

We define the **tail autocorrelation function** as the limit (if it exists)

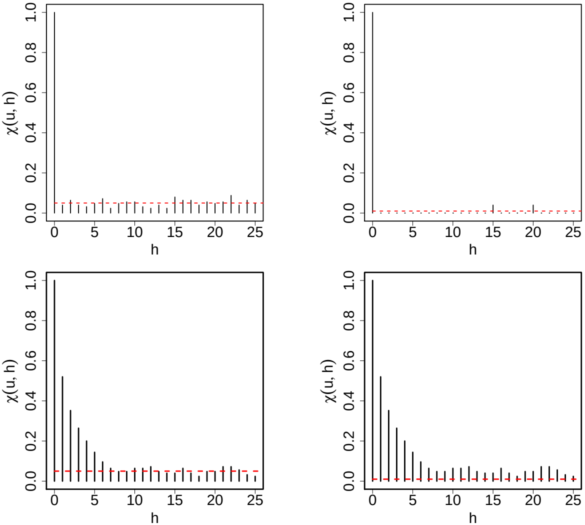
$$\chi(h) = \lim_{u \to 1} \chi(h; u) \in [0, 1], \quad h = 0, 1, 2 \dots$$

- $\chi(h)$ characterizes co-occurrence probabilities of high values at temporal lag h.
- By definition, $\chi(0) = 1$.
- With independent observations, $\chi(h) = 0$ for $h > 0$ (but $\chi(h; u) > 0$ for $u < 1$).
- The series $\{X_i\}$ is called **asymptotically independent at lag h if $\chi(h) = 0$**.
- For data, an **empirical version** can be estimated using $\chi(h; u)$ with empirical probabilities and with F replaced by the empirical distribution funtion F_n.

The function $\chi(h)$ is sometimes also called **auto-tail dependence function** or **extremogram**.

Illustration : Empirical tail autocorrelation function

Top row : independence ; bottom row : asymptotic dependence
Left column : $u = 0.95$; right column : $u = 0.99$
Dashed red line corresponds to theoretical $\chi(h; u)$ for independence.



Block maximum approach

Assuming that the $D(u_n)$-condition holds, we still get a GEV **distribution for blockwise maxima**.

As before, we can extract the sample of blockwise maxima.

There could be dependence between the maxima of consecutive blocks, but in practice we often ignore and assume that we have an i.i.d. sample of maxima
$\Rightarrow$ **maximum likehood estimation** works as in the i.i.d. setting.

Remark : in the case of **nonstationarities**, we can choose the blocks such as to avoid that block bounds fall into periods where we often observe block maxima.

Examples :

- For temperature maxima in mainland France, take years from January to December as blocks.
- For maximum snow width in mainland France, take years from July to June as blocks.

Cluster-based estimation of the extremal index

Estimation of the extremal index through threshold exceedances :

$$\hat{\theta} = \frac{1}{\text{average cluster length}}$$

Question : How can we **define and identify clusters in practice** ?

The runs method

Principle : two exceedances are part of the same cluster if there are less than k consecutive non-exceeding observations in-between, with some $k \geq 1$.

1. Fix a high threshold u, such as the empirical 95%-quantile of data.
2. Fix k. Often $k = 1$ or $k = 2$.
3. Look for the index i_0 with first threshold exceedance $X_{i_0} > u$ in X_i, $i = 1, 2 \ldots$ $\Rightarrow$ the first cluster begins at i_0.
4. If at least one of $X_{i_0+1}, \ldots, X_{i_0+k}$ exceeds u, then $i_0 + 1$ is still part of the first cluster.
5. Iterate until **k consecutive non exceedances** above u are found.
6. The first cluster goes from i_0 until the last found exceedance.
7. Continue and iterate through Steps 3 to 6 to detect the second cluster, third cluster, and so on.

Declustering in peaks-over-threshold modeling

Exceedances in threshold-exceedance models may be dependent. To obtain an (approximately) independent sample, we can consider only the most extreme observation among each "cluster" of extremes.

Peaks-over-threshold with declustering :

1. Use an **empirical rule to define clusters of exceedances**, for instance the runs method.
2. Identify the **maximum excess of each cluster**.
3. Assume cluster maxima to be independent, with their excesses $X_j - u > 0$ following the GPD
4. Estimate the **GPD** for the sample of cluster maxima.

⚠ The results of the model have to be interpreted with respect to cluster maxima.

Discussion of approaches for dependent extremes

- **Extremal index** : a summary parameter that can be estimated and interpreted.
- **Tail autocorrelation function :** an alternative way to explore the strength of temporal dependence among extremes at fixed lags $h = 1, 2 \dots$
- With the block maximum approach, we can proceed as in the i.i.d. case, if the dependence between blocks is negligible.
- With the POT-approach using the GPD, we can also proceed as in the i.i.d. case if we "decluster" the threshold exceedances.
- It is possible to design and estimate parametric time series models for extremes, for instance Markov models (**not discussed here**).
- **Nonstationarities :**
  - Here we have established theoretical for *stationary* sequences, but often data are nonstationary.
  - Estimation and interpretation of clusters and of the extremal index still makes sense in a nonstationary setting.
  - In case of strong nonstationarities, it makes sense to use a nonstationary threshold, for instance in the runs method and in the GPD estimation.

Bivariate and multivariate extremes

Often, several series of environmental or climatic variables are interdependent.

Examples :

- Different variables observed at the same location, such as minimum temperature, maximum temperature, precipitation, wind speed.
- The same variable observed at different locations, such as precipitation at different locations of a river catchment.

⚠ **Studying co-occurrences of extreme events in several variables is important :**

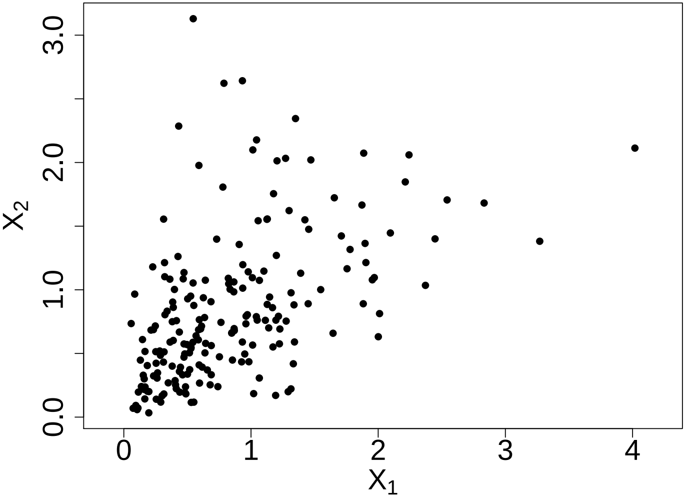
- **aggregation** of extreme observations in several components (example : cumulated precipitation over several sites or a catchment $\Rightarrow$ flood risk)
- understand and predict the **spatial extent** and **temporal duration** of extreme events
- **reliability :** probability that several critical components fail simultaneously ?

Topics of this section :

- exploratory statistical tools
- multivariate limit theory
- estimation of simple parametric models for extreme-value dependence

Illustration : a bivariate sample with dependence

Scatterplot of $X_{1,i}$ and $X_{2,i}$, $i = 1, 2, \ldots, n$.



Tail correlation

Given a bivariate random vector (X_1, X_2) with $X_1 \sim F_1$ and $X_2 \sim F_2$, we consider the conditional probability

$$\chi(u) = \Pr(F_2(X_2) > u \mid F_1(X_1) > u) = \frac{\Pr(F_2(X_2) > u, F_1(X_1) > u)}{\Pr(F_1(X_1) > u)}, \quad u \in]0,1[.$$

(**Remark :** $F_i(X_i) \sim \mathrm{Unif}(0,1)$).

We define the following limit (if it exists) : $\chi = \lim_{u \to 1} \chi(u)$.

The **coefficient** $\chi \in [0,1]$ is symmetric with respect to X_1 and X_2 and is known as **χ-measure** or **tail correlation**. We say that

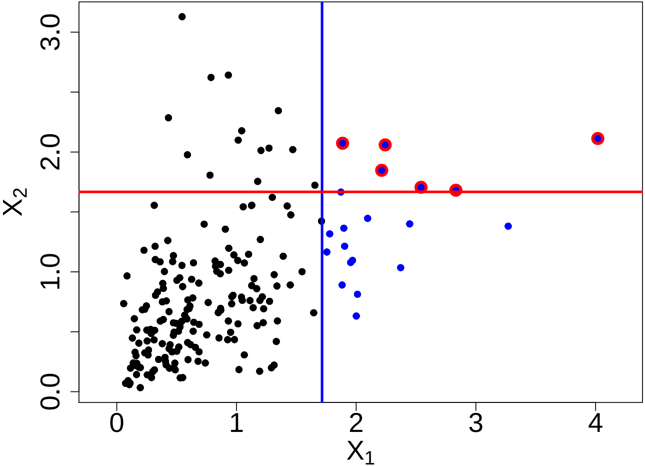
- X_1 and X_2 are **aysymptotically dependent if** $\chi > 0$;
- X_1 and X_2 are **asymptotically independent if** $\chi = 0$.

Examples :

- Bivariate Gaussian distribution with linear correlation coefficient $\rho < 1 \Rightarrow \chi = 0$
- Consider three i.i.d. variables X_0, X_{11}, X_{12} in MDA with tail index $\xi > 0$, and $X_1 = cX_0 + X_{11}$, $X_2 = cX_0 + X_{12}$ with $c > 0 \Rightarrow \chi = c/(1+c) > 0$.

Illustration : empirical tail correlation

Setting : $n = 200$, $u = 0.9$. Blue points : exceedances of $\hat{F}_1(X_1)$ above u.
Red points : exceedances of $\hat{F}_2(X_2)$ above u given that $\hat{F}_1(X_1)$ above u.
Empirical tail correlation : $\hat{\chi}(u) = \frac{6}{20} = 0.3$.



Componentwise maxima of random vectors

Consider a sequence of independent and identically distributed random vectors

$$\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,d})^T \stackrel{d}{=} \boldsymbol{X} \sim F_{\boldsymbol{X}},$$

where $F_{\boldsymbol{X}}$ is the joint distribution of the components of $\boldsymbol{X}$.

The **componentwise maximum**

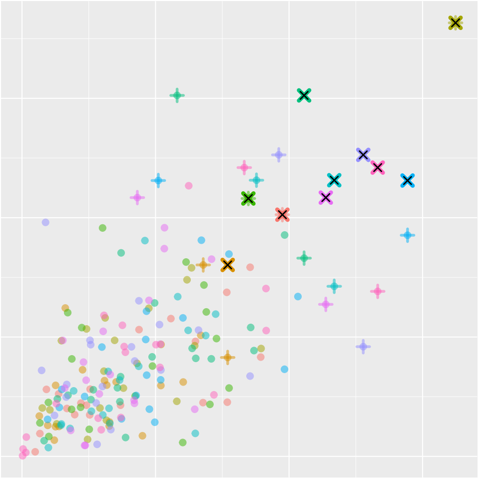
$$\boldsymbol{M}_n = (M_{n,1}, \ldots, M_{n,d})^T = \left(\max_{i=1}^{n} X_{i,1}, \ldots, \max_{i=1}^{n} X_{i,d}\right)^T$$

has distribution $F_{\boldsymbol{X}}^n$.

⚠ The componentwise maximum $\boldsymbol{M}_n$ can be composed of values $X_{i,j}$ from different indices i.

Illustration : bivariate componentwise block maxima

The scatterplot shows values $\boldsymbol{X}_i = (X_{i,1}, X_{i,2})$. Different colors correspond to different blocks of size n. Bivariate componentwise block maxima $\boldsymbol{M}_n$ are shown by crosses $\times$. Plus-symbols $+$ show vectors $\boldsymbol{X}_i$ that contribute to the maxima.



Multivariate max-stable distributions

Definition : max-stable distribution

A ***d*-dimensional multivariate distribution** G is called **max-stable** if, for each $n \in \mathbb{N}$, there exist deterministic vectors $\boldsymbol{\alpha}_n = (\alpha_{n,1}, \ldots, \alpha_{n,d})$ and $\boldsymbol{\beta}_n = (\beta_{n,1}, \ldots, \beta_{n,d}) > \mathbf{0}$ such that

$$G^n(\boldsymbol{\alpha}_n + \boldsymbol{\beta}_n \boldsymbol{z}) = G(\boldsymbol{z}), \quad \boldsymbol{z} \in \mathbb{R}^d.$$

Equivalently, if $\boldsymbol{X}_1 \sim G$, then

$$\frac{\boldsymbol{M}_n - \boldsymbol{\alpha}_n}{\boldsymbol{\beta}_n} \stackrel{d}{=} \boldsymbol{X}_1, \quad n \in \mathbb{N}.$$

⚠ Multivariate max-stability is stronger than max-stability of the univariate marginal distributions (of GEV type)

$$G_i(z_i) = \Pr(X_i \leq z_i) = G(\infty, \ldots, \infty, z_i, \infty, \ldots, \infty),$$

since it additionally implies a stability property for the dependence structure.

Multivariate maximum domain of attraction

Theorem : Multivariate maximum domain of attraction

If there are normalizing sequences of deterministic vectors $\boldsymbol{a}_n = (a_{n,1}, \ldots, a_{n,d})$ and $\boldsymbol{b}_n = (b_{n,1}, \ldots, b_{n,d}) > 0$, $n \in \mathbb{N}$, such that we observe the convergence of all finite-dimensional distributions in

$$\frac{\boldsymbol{M}_n - \boldsymbol{a}_n}{\boldsymbol{b}_n} \to \boldsymbol{Z} = (Z_1, \ldots, Z_d) \sim G, \quad n \to \infty,$$

where $\boldsymbol{Z}$ has non-degenerate marginal distributions, then G is **multivariate extreme-value distribution**, that is, a multivariate max-stable distribution.

Standardized marginal distributions

To focus on the extremal dependence structure, it is useful to **standardize the marginal distributions** F_j of X_j and G_j of Z_j.

Often, the **unit Fréchet marginal distribution** is used :

$$G_j^\star(z) = \mathrm{GEV}(z;\, \xi = 1, \mu = 1, \sigma = 1) = \exp\left(-\frac{1}{z}\right), \quad z > 0.$$

We can transform any continuous random variable $X \sim F$ towards unit Fréchet marginal distribution as follows : $X^\star = -\frac{1}{\log F(X)} \sim G^\star$.

If $X \sim \mathrm{GEV}(\xi, \mu, \sigma)$, then $X^\star = \left(1 + \xi \frac{X-\mu}{\sigma}\right)^{1/\xi}$.

Domain of attraction

If the distribution of the $\boldsymbol{X} \sim F$ is in the maximum domain of attraction of G, then $\boldsymbol{X}^\star = (X_1^\star, \ldots, X_d^\star) \sim F_{\boldsymbol{X}^\star}$ is in the maximum domain of attraction of $\boldsymbol{G}^\star$ with unit Fréchet marginal distributions.
With normalized marginal distributions, we can choose normalizing sequences $\boldsymbol{a}_n^\star = (0, \ldots, 0)$ and $\boldsymbol{b}_n^\star = (1/n, \ldots, 1/n)$.

Pickands dependence function

A multivariate max-stable distribution $G^\star$ with unit Fréchet marginal distributions $G_j^\star$ is called **simple**, and it has representation

$$G^\star(\boldsymbol{z}) = \exp(-V^\star(\boldsymbol{z})), \quad \boldsymbol{z} > \boldsymbol{0}.$$

The **exponent function** $V^\star > 0$ is **(-1)-homogeneous** : $tV^\star(t\boldsymbol{z}) = V^\star(\boldsymbol{z})$, $t > 0$.

For $d = 2$, the **Pickands dependence function** A can be defined from $V^\star$ as follows :

$$V^\star(1/z_1, 1/z_2) = (z_1 + z_2)\, A(z_2/(z_1 + z_2))$$

where

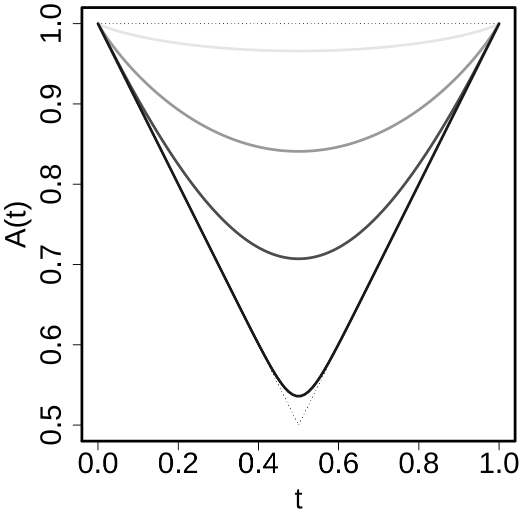
- $A(w)$ is defined for $w \in [0, 1]$,
- A is convex,
- $\max(w, 1 - w) \leq A(w) \leq 1$
- $A(w) = 1$ corresponds to independence with

$$G^\star(z_1, z_2) = G_1^\star(z_1) G_2^\star(z_2) = \exp(-1/z_1 - 1/z_2), \quad z_1, z_2 > 0,$$

- $A(w) = \max(w, 1 - w)$ corresponds to perfect dependence with $G^\star(z_1, z_2) = \exp\left(-\frac{1}{\min(z_1, z_2)}\right)$.

Illustration : Pickands dependence function

Lower curves correspond to stronger dependence.
Upper dashed curve corresponds to independence.
Lower dashed curve corresponds to full dependence.



Example : parametric bivariate max-stable models

⚠ In general, there is no parametric form of the dependence structure.

Bivariate logistic model

$$G^\star(z_1, z_2) = \exp\left(-\left(z_1^{-1/\kappa} + z_2^{-1/\kappa}\right)^\kappa\right), \quad z_1, z_2 > 0,$$

with parameter $0 < \kappa \leq 1$ and

- perfect dependence for $\kappa \to 0$;
- independence for $\kappa = 1$

Huesler–Reiss model

$$G^\star(z_1, z_2) = \exp\left(-\frac{1}{z_1}\Phi\left(\frac{1}{\kappa} + \frac{r}{2}\log\frac{z_2}{z_1}\right) - \frac{1}{z_2}\Phi\left(\frac{1}{\kappa} + \frac{r}{2}\log\frac{z_1}{z_2}\right)\right), \quad z_1, z_2 > 0,$$

with parameter $\kappa > 0$, the univariate standard Gaussian distribution function Φ, and :

- perfect dependence for $\kappa \to \infty$;
- independence for $\kappa \to 0$.

The **extremal coefficient** $\theta_d \in [1, d]$ is a summary measure for multivariate extremes defined as $\theta_d = V^\star(1, \ldots, 1)$.

- For $d = 2$, we have $\theta_2 = 2A(1/2)$ and $\theta_2 = 2 - \chi$.
- For componentwise maxima, we obtain

$$\Pr(M^\star_{n,1}/n \leq z, \ldots, M^\star_{n,d}/n \leq z) = \Pr\left(\max_{j=1}^{d} M^\star_{n,j}/n \leq z\right) \to \exp\left(-\frac{\theta_d}{z}\right), \quad n \to \infty.$$

  for $z > 0$.

- For threshold exceedances, we obtain

$$x\,P\left(\max_{j=1,\ldots,d} X^\star_j > x\right) \to \theta_d, \quad x \to \infty.$$

Interpretation : θ_d is the **average number of independent clusters** in the components $1, \ldots, d$ when an extreme event occurs, and d/θ_d is the average cluster size.

Direct estimation of θ_d is possible based on empirical versions of the above probabilities.

Estimation for multivariate extremes

- Studying some **summary statistics** (tail correlation, extremal coefficients...) is usually recommended before estimating more complex models.

- For a full model, we have to **estimate marginal distributions (GEV) and the dependence structure**.

- With bivariate block maxima, joint estimation of all parameters is feasible via **maximum likelihood estimation**, and it can even be combined with nonstationary modeling of marginal distributions.

- The Pickands dependence function can also be estimated nonparametrically based on maxima or threshold exceedances.

- Parametric models for multivariate threshold exceedances exist, but we do not discuss them here.

Maximum likelihood estimation for bivariate maxima

Consider a general **bivariate extreme-value distribution** parametrized by

- marginal parameters ξ_j, μ_j, σ_j, $j = 1, 2$ and
- dependence parameter(s), such as κ in the logistic model.

Define the marginal parametric transformation $T(z; \xi, \mu, \sigma) = \left(1 + \xi \frac{z-\mu}{\sigma}\right)^{1/\xi}$.

The distribution function of a bivariate extreme-value model with marginal $\mathrm{GEV}(\xi_j, \mu_j, \sigma_j)$ distributions and with standardized bivariate distribution $G^\star$ is

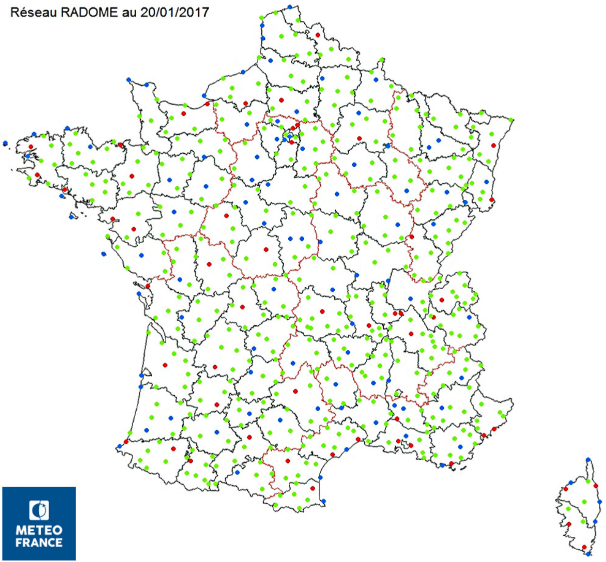
$$G(z_1, z_2; \xi_1, \mu_1, \sigma_1, \xi_2, \mu_2, \sigma_2, \kappa) = G^\star(T^{-1}(z_1; \xi_1, \mu_1, \sigma_1), T^{-1}(z_2; \xi_1, \mu_2, \sigma_2)).$$

Given a bivariate sample of block maxima $(z_{i,1}, z_{i,2})$, $i = 1, \ldots, m$, the **maximum likelihood estimator** is obtained by maximizing the likelihood function

$$(\xi_1, \mu_1, \sigma_1, \xi_2, \mu_2, \sigma_2, \kappa) \mapsto \prod_{i=1}^{m} \frac{d^2}{\mathrm{d}z_{1,i}\mathrm{d}z_{2,i}} G(z_{1,i}, z_{2,i}; \xi_1, \mu_1, \sigma_1, \xi_2, \mu_2, \sigma_2, \kappa).$$

Models can be compared and validated based on **visual inspection** (QQ-plots, Pickands dependence function...) and **information criteria** such as AIC.

Illustration : The RADOME network of Météo France



As space, we here consider the two-dimensional geographical space, and we denote it by $\mathbb{R}^2$.

Spatial extreme-value modeling can be considered as an **extension of multivariate modeling towards stochastic processes**.

⚠ A key difference is that the number of spatial locations in $\mathbb{R}^2$ is infinite.

- In multivariate modeling with a random vector $X = (X_1, \ldots, X_d) \sim F$, the model must be valid only for the d components.
- In spatial modeling, this is slightly different : we observe data only for a finite number of spatial locations $s_1, \ldots, s_d$, but we want the model to be valid for every $s \in \mathbb{R}^2$ (**spatial interpolation**).

Componentwise maxima

Given independent and identically distributed copies of stochastic processes, we can consider their **componentwise maxima**, and we can then use **max-stable limit processes** as models.

Consider a sequence of independent and identically distributed (iid) stochastic processes

$$\{X_i(s),\ s \in \mathcal{S}\}, \quad i = 1, 2, \ldots$$

defined over a non-empty spatial domain $\mathcal{S} \subset \mathbb{R}^D$ with some $D \geq 1$. Then we can define the componentwise maximum process

$$\{M_n(s),\ s \in \mathcal{S}\} \quad \text{with} \quad M_n(s) = max_{i=1}^n X_i(s).$$

Remark : random vectors can be considered as special cases of stochastic processes with $\mathcal{S} = \{1, 2, \ldots, d\}$.

⚠ We set $D = 2$ in the following for geographical coordinates $s \in \mathcal{S}$.

Theorem : Maximum domain of attraction

If there are sequences of **normalizing functions** $a_n(s)$ and $b_n(s) > 0$, $s \in \mathcal{S}$, $n \in \mathbb{N}$, such that we observe the convergence of all finite-dimensional distributions in

$$\left\{ \frac{M_n(s) - a_n(s)}{b_n(s)},\ s \in \mathcal{S} \right\} \to \{Z(s),\ s \in \mathcal{S}\}, \quad n \to \infty, \qquad (\star),$$

where $Z(s)$ has non-degenerate marginal distributions for all $s \in \mathcal{S}$, then $\{Z(s)\}$ is a **max-stable process**.

A max-stable process is characterized by its univariate and multivariate **max-stable marginal distributions**.

As before, **max-stability** of a process $\boldsymbol{Z}$ means that $(\star)$ holds exactly when using i.i.d. copies $\boldsymbol{Z}_i$ of $\boldsymbol{Z}$ in the construction of $\boldsymbol{M}_n$.

The spectral construction of max-stable processes

Any max-stable process $\boldsymbol{Z}^\star$ with unit Fréchet marginal distributions $\mathrm{GEV}(1,1,1)$, that is,

$$\Pr(Z^\star(s) \leq z) = \exp(-1/z), \quad z > 0,$$

can be represented through a **spectral construction**

$$Z^\star(s) = \max_{i=1,2,\ldots} \varepsilon_i(s)/U_i, \quad 0 < U_1 < U_2 < \ldots \sim \mathrm{PPP}(\mathrm{d}u), \quad s \in \mathcal{S},$$

where

- the **profile processes** $\varepsilon_i(s)$, $i = 1, 2, \ldots$ are i.i.d. with $\mathbb{E}\,\varepsilon_i(s)_+ = 1$,
- $\mathrm{PPP}(\mathrm{d}u)$ is a unit Poisson process on $[0, \infty[$, that is,

$$\#\{U_i \in [a, b]\} \sim \mathrm{Poisson}(b - a), \quad b > a > 0.$$

This spectral construction can be used to **construct spatial max-stable models**.

Extremal-t processes are max-stable processes where the **profile processes ε_i in the spectral construction corresponds to a (powered) Gaussian process** :

$$Z^\star(s) = \max_{i=1,2,\dots} \left(W_i(s)_+\right)^\alpha / U_i, \quad s \in \mathcal{S},$$

with **degrees of freedom parameter $\alpha > 0$**, and $\boldsymbol{W}_i(s)$ an i.i.d. sequence of **centered Gaussian processes** with variance σ^2 chosen to ensure $\mathbb{E}\left[\left(W_i(s)_+\right)^\alpha\right] = 1$.

The Gaussian processes $\boldsymbol{W}_i$ are parametrized through a covariance function $C(s_1, s_2)$ such that

$$(W(s_1), \dots, W(s_d)) \sim \mathcal{N}(\mathbf{0}, \Sigma)$$

with

$$\Sigma = \begin{pmatrix} C(s_1, s_1) & C(s_1, s_2) & \dots & C(s_1, s_d) \\ \vdots & \vdots & \vdots & \vdots \\ C(s_d, s_1) & C(s_d, s_2) & \dots & C(s_d, s_d) \end{pmatrix}.$$

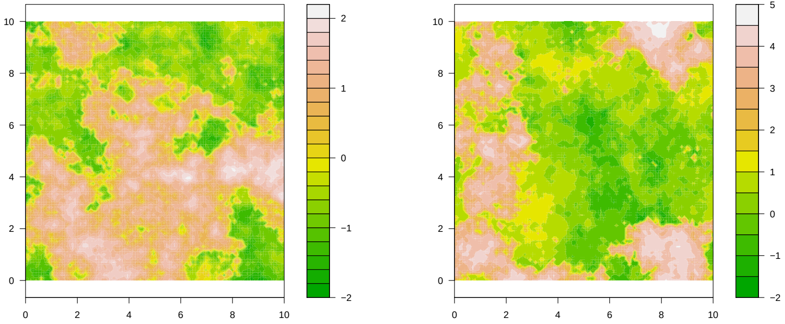
and $C(s_j, s_j) = \sigma^2$, $j = 1, \dots, d$.

If the covariance function is stationary and isotropic, then it depends only on the distance $h = \|s_2 - s_1\|$ between sites : $C(s_1, s_2) = C(h)$.

To obtain parametric models of max-stable processes, we can use a **parametric covariance function**.

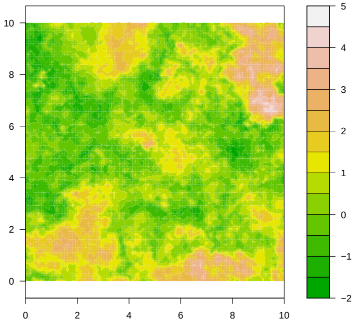
Example : Extremal-t process with $\alpha = 1$

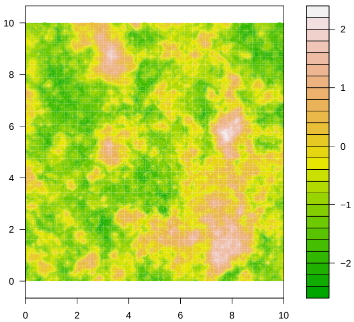
- exponential covariance function $C(h) = \sigma^2 \exp(-h/3)$ in the profile process $W(s)$
- plot of $\log(Z^\star(s))$



Example : Extremal-t with $\alpha = 4$

- exponential covariance function $C(h) = \sigma^2 \exp(-h/3)$ in the profile process $W(s)$
- plot of $\log(Z^\star(s))$





Functional summary statistics for spatial extremes

An important aspect of spatial extreme-value analysis is how fast the **tail dependence between two locations** decreases as their distance increases.

Given $X(s_j) \sim F_j$, $j = 1, 2$, we can consider the **tail correlation** χ and the **bivariate extremal coefficient** θ_2 as a function of distance $h = \|s_2 - s_1\|$ **between two locations** $s_1, s_2 \in \mathbb{R}^2$:

$$\chi(h; u) = \Pr(F_2(X(s_2)) > u \mid F_1(X(s_1)) > u) = \frac{\Pr(F_1(X(s_1)) > u, F_2(X(s_2)) > u)}{\Pr(F_1(X(s_1))) > u)},$$

where $u \in]0, 1[$ and the distance $h \geq 0$.

The following limit, if it exists, is called the **tail correlation function** :

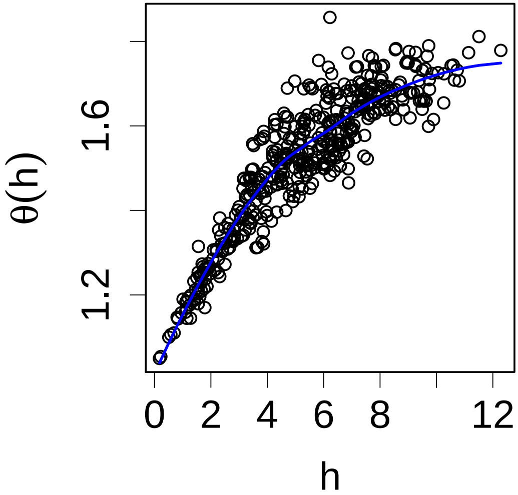
$$\chi(h) = \lim_{u \to 1} \chi(h; u), \quad h > 0.$$

By analogy with the bivariate case, we can consider the **extremal coefficient function** $\theta(h) = 2 - \chi(h)$, where

$$\Pr(Z^\star(s_1) \leq z, Z^\star(s_2) \leq z) = \exp\left(-\theta(h)/z\right), \quad z > 0.$$

Example : Empirical extremal coefficient function

Based on data for an extremal-t process at 30 (uniformly sampled) locations with 100 replicates of the max-stable process.



Parametric inference for max-stable processes

We consider the max-stable distribution $G_d(z_1, \ldots, z_d)$ of a max-stable random vector $(Z(s_1), \ldots, Z(s_d))$ corresponding to the max-stable process at the locations $s_1, \ldots, s_d$.

As the number of observed locations d increases, the number of terms in the probability density function $\frac{\partial^d}{\partial z_1 \cdots \partial z_d} G_d(z_1, \ldots, z_d)$ increases exponentially fast
$\Rightarrow$ ⚠ standard full likelihood estimation of parametric models is not possible for d much larger than 2.

A useful alternative is the **pairwise likelihood** :

$$\text{params} \mapsto \prod_{j=1}^{d-1} \prod_{k=j+1}^{d} \frac{\partial^2}{\partial z_j \partial z_k} G_2(z_j, z_k; \text{params}).$$

- Implicit assumption is that pairs of variables are independent among each other.
- An estimator of the parameters is obtained by maximizing the pairwise likelihood.
- The estimator is known to be well-behaved (**consistency**, **asymptotic normality**...).
- To reduce the number of pairs, we can choose to keep only some of the pairs, for instance only pairs closer than a maximum distance.

Multivariate threshold exceedances

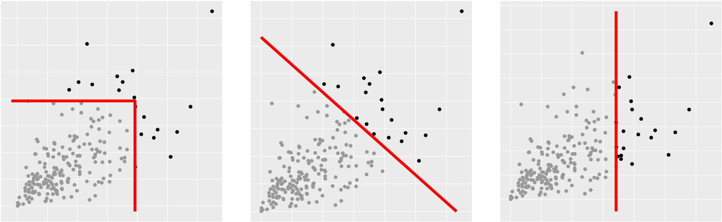
For bivariate extremes, and more generally multivariate extremes, we have so far only considered the approach of blockwise maxima.

There exist also a large number of approaches for modeling threshold exceedances in a multivariate or spatial setting, but the methods are often more involved, in particular with respect to the **choice of a multivariate threshold**.

Multivariate and functional Peaks-Over-Threshold

⚠ There exists no unique ordering of multivariate vectors or functions !

- exceedances in at least one component $\Rightarrow$ multivariate GPDs
- exceedances of the sum of components $\Rightarrow$ spectral measures
- exceedances in a fixed component $\Rightarrow$ conditional extremes



Multivariate/spatial Peaks-Over-Threshold limits

What is a useful limit theory for POT-limits of dependent extremes ?

- homogeneous **risk functional** r with $r(t\boldsymbol{x}) = t\,r(\boldsymbol{x})$ for $t > 0$
- data normalized to standard Pareto margins : $\Pr(X^\star(s) > x) = \frac{1}{x}$, $x > 1$

Existence of max-stable limit $\Leftrightarrow$ existence of POT limits :

$$\boldsymbol{X}^\star/u \mid r(\boldsymbol{X}^\star) \geq u \quad \to \quad \boldsymbol{Y}, \quad u \to \infty$$

$\Rightarrow$ $\boldsymbol{Y}$ defines a **Pareto process**.

Peaks-over-theshold-stability :

$$\boldsymbol{Y}/u \mid r(\boldsymbol{Y}) \geq u \quad \stackrel{d}{=} \quad \boldsymbol{Y}/\tilde{u} \mid r(\boldsymbol{Y}) \geq \tilde{u}, \qquad u, \tilde{u} > 0$$

Tail index estimation based on order statistics

Construction principle : estimate extreme-value parameters based on the k **highest order statistics**

$$\cdots < X_{(n-k+1)} < X_{(n-k+2)} < \cdots < X_{(n-1)} < X_{(n)}$$

for $k \in \mathbb{N}$ chosen such that

- the sample of size k is large enough to construct useful estimators,
- the sample fraction k/n is close to 0.

Benefits and inconvenients :

- Widely used estimators, often for exploratory statistical analysis.
- The classical estimators are of this type, and new estimators are still developed.
- Such estimators are often very fast and easy to compute.
- Such estimators may be more robust than likelihood-based estimators, for example with respect to extreme outliers, or with respect to assumptions on the behavior of normalizing sequences.

The Hill estimator of $\xi > 0$

The **Hill estimator**, proposed in 1975, is one of the oldest and most used estimators of the tail index :

- It is based on the highest k order statistics, typically with $k \ll n$.
- **Formula of the estimator :**

$$\hat{\xi}_{\mathrm{Hill}} = \left(\frac{1}{k} \sum_{j=1}^{k} (\log(X_{(n-j+1)}) - \log(X_{(n-k)})) \right)^{-1}.$$

- It corresponds to the maximum likelihood estimator of ξ if the tail of X is exactly Pareto-distributed, that is,

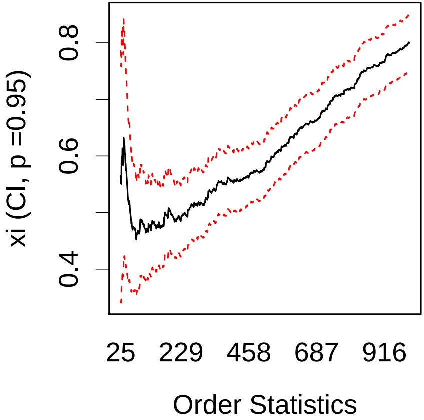
$$P(X \geq x) = ax^{-1/\xi}, \quad x > a^{\xi},$$

  with Pareto shape parameter $1/\xi > 0$.

Many extensions have been developed for this estimator, such as for $\xi \leq 0$.
Rank-based estimators exist also for the normalizing sequences a_n and b_n.

Example : Hill plots

Hill plot : Hill estimator of $\xi > 0$ for different numbers k of highest order statistics (note : the sample size is 4000 here)



Interpretation : Values become stable at $\hat{\xi}_{\text{Hill}} \approx 0.5$ for $k \leq 200$ (approximately, taking into account estimation uncertainty).
$\Rightarrow$ We could choose to use $\hat{\xi}_{\text{Hill}}$ for $k = 200$, corresponding to using approximately the $k/4000 = 5\%$ of largest values for estimation.

Residual dependence in asymptotic independence

For assessing the strength of tail dependence, we have considered the tail correlation coefficient χ.

For many bivariate probability distributions, the tail coefficient is 0, such that we obtain **asymptotic independence**. This means that the probability of one component exceeding a very high threshold, given that the other component exceeds this threshold, tends to 0.

For example, the bivariate Gaussian distribution with correlation $\rho \in [-1, 1[$ is asymptotically independent.

For many environmental and climatic data of interest, the estimates of χ also point towards asymptotic independence.

Question : Can we get a more detailed characterization of the **joint tail behavior in the case of asymptotic independence** ?

Coefficient of tail dependence

Coefficient of tail dependence $\eta \in]0,1]$: define

$$\eta(u) = \frac{\log \Pr(F_1(X_1) > u)}{\log \Pr(F_1(X_1) > u, F_2(X_2) > u)}$$

then

$$\eta = \lim_{u \to 1} \eta(u),$$

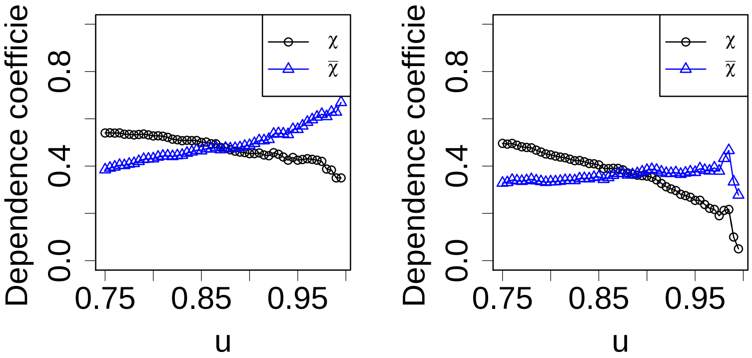
if the limit exists.

- If $\chi > 0$ (asymptotic dependence) then $\eta = 1$; if $\eta < 1$ then $\xi = 0$ (asymptotic dependence).
- An alternative, often-used coefficient is defined as $\overline{\chi} = 2\eta - 1 \in]-1,1]$.
- For a standard **bivariate Gaussian random vector** $(X_1, X_2) \sim \mathcal{N}((0,0), \Sigma_2)$ with $\Sigma_2 = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, we obtain $\overline{\chi} = \rho$.
- **Empirical estimation :** fix a threshold u close to 1, and consider empirical distribution functions in $\eta(u)$ and $\overline{\chi}(u) = 2\eta(u) - 1$.

⚠ If data are asymptotically dependent, we get interesting information from χ ; if data are asymptotically independent, then we get interesting information from η (or $\overline{\chi}$).

Example : Empirical extremal coefficient function

One plot corresponds to asymptotic dependence in data, the other to asymptotic independence.



⚠ Values $\chi(u)$ and $\overline{\chi}(u)$ can remain far from the limit, even for u very close to 1.

von Mises conditions for extreme-value limits

We have considered the maximum domain of attraction condition, in particular for univariate distributions F.

Given a distribution F, are there relatively **simple theoretical tools**

- to check if a distribution F (for instance, normal distribution, gamma distribution, Weibull distribution) is in the maximum domain of attraction of an extreme-value limit ?
- to determine appropriate normalizing sequences a_n and b_n ?

The **von-Mises conditions** are useful **sufficient conditions** for the maximum domain of attraction : if F has probability density function f, we can consider the Mill's ratio

$$r(x) = \frac{1 - F(x)}{f(x)}.$$

If $\xi = \lim_{x \to x^\star} r'(x)$ exists (where $x^\star \in], \infty]$ is the essential supremum F), and given

$$a_n = F^{-1}(1 - 1/n), \quad b_n = r(a_n),$$

then the normalized maximum $(M_n - a_n)/b_n$ tends to the GEV distribution with shape parameter ξ.

Example : standard normal distribution $r(x) \to 0 = \xi$ as $x \to \infty$ (by l'Hôpital's rule).

Type of available observations

The variable of interest may have been observed, or its observations may have been transformed, as follows :

1. Observation series of **original events** (*e.g.*, hourly or daily data)
2. **Block maxima** over blocks of the same size (*e.g.*, yearly maxima)
3. **Exceedances** above a high but fixed threshold (*Peaks-over-Threshold*)
4. other data types : Records, observation series with missing data...

Remarks :

- We put focus on estimation using block maxima and threshold exceedances.
- The transformation of original series to a sample of threshold exceedances requires some care :
  - choice of the threshold ;
  - if the series is dependent, how do we treat clusters of threshold exceedances ?
- In case of nonstationarities (*e.g.*, seasonality) and serial dependence within blocks, the block maximum approach has the benefit of avoiding to deal with these complex structures.

Some general remarks on extreme-value statistics

- Observed samples are finite, but theoretical results are asymptotic, for instance the natural limit distributions (GEV, GPD) :
    - For using the GEV distribution with maxima data, we must take maxima over blocks of relatively large number of observations (e.g., yearly maxima of variables).
    - For using the GPD with threshold exceedances, we must fix a relatively large threshold.
- There is often a **bias-variance trade-off in extreme-value statistics** :
    - By using a smaller block size (e.g., monthly instead of yearly blocks), we obtain a larger sample of maxima data, but the asymptotic GEV distribution may be less accurate.
    - By using a lower threshold we obtain more threshold exceedances, but the approximation through the asymptotic GPD may be less accurate.
    - In practice, it makes sense to check the sensitivity of estimation results with respect to the choice of the threshold or the block size.

Final remarks

- Focus of this course was on likelihood-based estimation. A large variety of other estimators have been proposed and may be preferrable depending on the context.
- Handling **nonstationarities** and **dependence** is quite challenging, and often different approaches can lead to sensible analyses. We have seen some theory and practical techniques to take them into account.
- We have studied the models based on block maxima and peaks-over-threshold limits. A third popular approach is known as the **point process approach**, which also provides a natural link between the two approaches.
- Currently, a lot of ongoing research aims to "hybridize" extreme-value theory with statistical learning concepts (for instance, graphical models, classification, high-dimensional variable selection...).
- Systematic use of extreme-value methods is recommended for studying extreme events related to climate and the environment.