# Modeling multivariate extremes

Thomas Opitz, BioSP, INRAE Avignon

ATHENS course on Extreme Value Analysis
MinesParisTech

16 March 2022

## Plan for this session

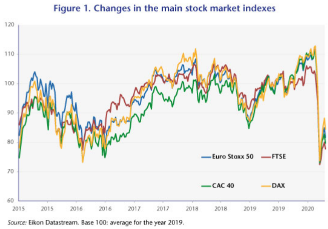
1. Multivariate extreme-value theory
2. Summary statistics for multivariate extremes
   - Practicals Part 1
   - Practicals Part 2
3. Some parametric models for multivariate extremes
4. Estimation of multivariate models
5. Discussion and extensions
   - Practicals Part 3

## Applications of multivariate extreme-value theory

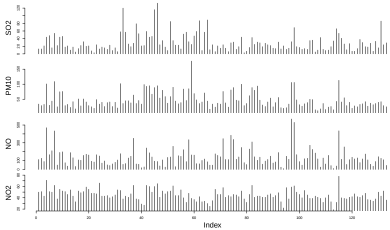
- **aggregation** of extreme observations in several components
  (*e.g.*, cumulated precipitation over several sites or a catchment $\Rightarrow$ flood risk)
- **spatial extent** and **temporal duration** of extreme events
  (heat waves, cold spells, windstorms, air pollution episodes,...)
- financial portfolio theory and **systemic risk** in finance and actuarial sciences
- **reliability** : probability that several critical components fail simultaneously

and many other applications...

**Stock market indices**




**Air pollution variables**



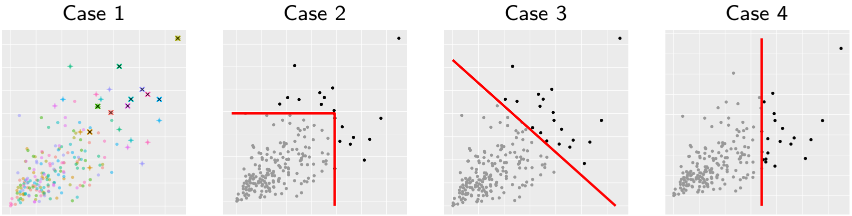
## How to characterize multivariate extreme events, and their asymptotic distribution ?

⚠ There is no unique ordering of vectors in $\mathbb{R}^d$ for $d > 1$ ⚠

Various cases can be of interest :

1. component-wise maxima over blocks of data ⇒ **mv. extreme-value distributions**
2. exceedances in at least one component ⇒ **mv. generalized Pareto distributions**
3. exceedances of the sum of components ⇒ **angular measures**
4. exceedances in a fixed component ⇒ **conditional extremes**

Methods 2–4 are **peaks-over-threshold approaches**.



## Vector notations used in the following

- bold symbols for vector-valued objects, for instance $\boldsymbol{x} = (x_1, \ldots, x_d)^T$

- $\boldsymbol{X} = (X_1, \ldots, X_d)^T$ ; $X_j \sim F_j$, $j = 1, \ldots, d$ ; $\boldsymbol{X} \sim F_{\boldsymbol{X}}$
  random vector with $d \geq 1$ components with joint (cumulative) distribution function $F_{\boldsymbol{X}}$, such that $\Pr(X_1 \leq x_1, \ldots, X_d \leq x_d) = F_{\boldsymbol{X}}(x_1, \ldots, x_d) = F_{\boldsymbol{X}}(\boldsymbol{x})$

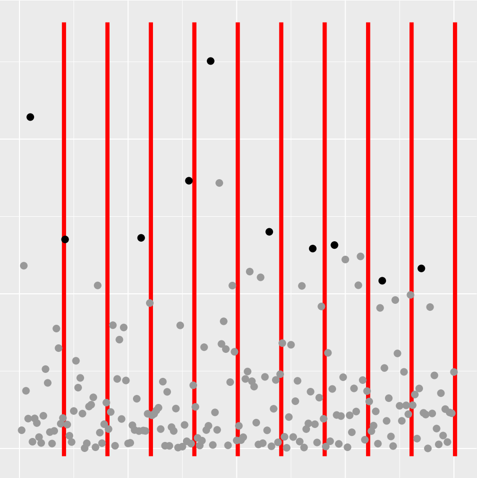
- $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{id})^T$, $i = 1, 2, \ldots$
  a series of random vectors, which could represent observations for times $i$

- $\boldsymbol{Z} = (Z_1, \ldots, Z_d)^T$ ; $Z_j \sim G_j$, $j = 1, \ldots, d$ ; $\boldsymbol{Z} \sim G$ ;
  $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{id})^T$, $i = 1, 2, \ldots$
  max-stable random vectors and distributions

- vector notations for componentwise operations :
  $\boldsymbol{x} > \boldsymbol{y}$ means that $x_j > y_j$ for all components $j = 1, \ldots, d$ ;
  $\boldsymbol{x} \not\leq \boldsymbol{y}$ means that $x_j > y_j$ in at least one component $j \in \{1, \ldots, d\}$
  $\boldsymbol{x} > \boldsymbol{0}$ means that $x_j > 0$ for all components $j = 1 = 1, \ldots, d$

## Recall : univariate block maximum approach

- Extract the maximum from each block to obtain a sample of block maxima
- Fit the Generalized Extreme-Value (GEV) limit distribution to this sample (which arises asymptotically as the block size tends to infinity)



**How to generalize theory and statistical methods to more than one data series ?**

## Multivariate maximum domain of attraction

For a sequence of independent and identically distributed (iid) random vectors $\boldsymbol{X}_i \sim F_{\boldsymbol{X}}$, $i = 1, 2, \ldots$, 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 $F_{\boldsymbol{X}}^n$ with $F_{\boldsymbol{X}}^n(\boldsymbol{x}) = (F_{\boldsymbol{X}}(\boldsymbol{x}))^n$ for $\in \mathbb{R}^d$.

### Multivariate maximum-domain-of-attraction (MMDA) condition

If vector sequences $\boldsymbol{a}_n = (a_{n,1}, \ldots, a_{n,d})^T$ and $\boldsymbol{b}_n = (b_{n1}, \ldots, b_{n,d})^T > \boldsymbol{0}$ exist such that

$$F_{\boldsymbol{X}}^n(\boldsymbol{a}_n + \boldsymbol{b}_n \boldsymbol{z}) \to G(\boldsymbol{z}) \quad (n \to \infty) \quad (\star)$$

with a nondegenerate multivariate limit distribution $G$, then $G$ is called a **multivariate extreme-value distribution**.

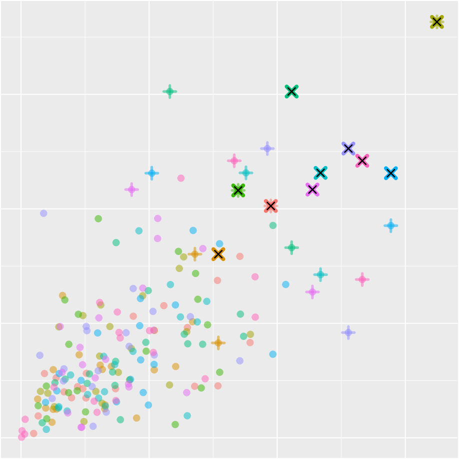
Equivalently with $(\star)$, $(\boldsymbol{M}_n - \boldsymbol{a}_n)/\boldsymbol{b}_n \to \boldsymbol{Z} \sim G$ for $n \to \infty$.

The class of limit distributions $G$ coincides with the class of **max-stable distributions**, which satisfy $G^n(\boldsymbol{\alpha}_n + \boldsymbol{\beta}_n \boldsymbol{z}) = G(\boldsymbol{z})$ for any $n \in \mathbb{N}$ and $\boldsymbol{z}$ with appropriate choices of $\boldsymbol{\alpha}_n, \boldsymbol{\beta}_n$.

With convergence $(\star)$, we say that **$F$ is in the multivariate domain of attraction of $G$**.

# Illustration of componentwise bivariate 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.



## Standardized marginal distributions

To focus on the extremal dependence structure, we often pretransform marginal distributions $F_j$ of $X_j$ by transforming them to the same **standardized marginal distribution** $F^\star$.

We transform a continuous random variable $X \sim F$ to a **unit Fréchet marginal distribution** as follows :

$$X^\star = -\frac{1}{\log F(X)}, \text{with} \quad P(X^\star \leq x) = \exp(-1/x), \; x > 0.$$

Alternatively, we could also transform to **standard Pareto marginal distributions**

$$X^\star = \frac{1}{1 - F(X)}, \quad \text{with} \quad P(X^\star \leq x) = 1 - 1/x, \; x > 0.$$

In both cases, $X^\star \geq 0$ and $x\, P(X^\star > x) \to 1$ as $x \to \infty$.

The subsequent limit results hold in both cases.

**Notation :** Given a random vector $\boldsymbol{X} = (X_1, \ldots, X_d)^T$, we write $F^\star_{\boldsymbol{X}}$ for the distribution of the random vector $\boldsymbol{X}^\star = (X_1^\star, \ldots, X_d^\star)^T$ with standardized margins :

$$X^\star \sim F^\star_{\boldsymbol{X}}, \quad F^\star_{\boldsymbol{X}}(\boldsymbol{x}) = F^\star_{\boldsymbol{X}}(x_1, \ldots, x_d) = F_{\boldsymbol{X}}(F_1^{-1}(F^\star(x_1)), \ldots, F_d^{-1}(F^\star(x_d))).$$

# Convergence of standardized maxima

The MMDA condition is equivalent to the following two conditions :

1. **univariate domain of attraction condition** for $X_j$, $j = 1, \dots, d$, such that

$$F_j^n(a_n + b_n z) \to G_j(z) \quad (n \to \infty)$$

with a generalized extreme value distribution $G_j$ ;

2. **multivariate domain of attraction condition for standardized data $\boldsymbol{X}^\star$**, such that

$$(F^\star_{\boldsymbol{X}})^n(n\boldsymbol{z}) \to G^\star(\boldsymbol{z}) \quad (n \to \infty);$$

equivalently, we observe the convergence in distribution

$$\frac{\boldsymbol{M}_n^\star}{n} = \left(\max_{i=1}^{n} \frac{X_{i1}^\star}{n}, \dots, \max_{i=1}^{n} \frac{X_{id}^\star}{n}\right)^T \to \boldsymbol{Z}^\star \sim G^\star \quad (n \to \infty).$$

Then, **$G^\star$ has unit Fréchet marginal distributions** $G_j^\star(z) = \exp(-1/z)$ for $z > 0$, and is max-stable satisfying $(G^\star)^n(n\boldsymbol{z}) = G^\star(\boldsymbol{z})$ for any $n \in \mathbb{N}$.
We say that **$G^\star$ is simple max-stable**.

A simple max-stable distribution $G^\star$ has representation

$$G^\star(\boldsymbol{z}) = \begin{cases} \exp(-V^\star(\boldsymbol{z})), & \boldsymbol{z} > \boldsymbol{0}, \\ 0, & \text{otherwise.} \end{cases}$$

where the **exponent function** $V^\star$ is **homogeneous of order** $-1$ :

$$tV^\star(t\boldsymbol{z}) = V^\star(\boldsymbol{z}).$$

The homogeneity property reflects max-stability :

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

So far, we have a convergence result for maxima of multivariate distributions.

What about **threshold exceedances** ?
$\Rightarrow$ We use the concept of **multivariate regular variation**.

### Multivariate regular variation condition (MRV)

A multivariate distribution $F^{\star}_{\boldsymbol{X}}$ with standardized marginal distributions is in the MMDA of a max-stable distribution $G^{\star}$ with exponent function $V^{\star}$ if and only if

$$t\,P(\boldsymbol{X}^{\star} \not\leq t\boldsymbol{x}) = t\,(1 - F^{\star}_{\boldsymbol{X}}(t\boldsymbol{x})) \to V^{\star}(\boldsymbol{x}), \quad \boldsymbol{x} > \boldsymbol{0} \quad (t \to \infty).$$

By exploiting the $(-1)$-homogeneity of $V^{\star}$,
**MRV suggests using the following practical approximation** :

$$P(\boldsymbol{X}^{\star} \not\leq \boldsymbol{x}) = \overline{F}^{\star}_{\boldsymbol{X}}(\boldsymbol{x}) = 1 - F^{\star}_{\boldsymbol{X}}(\boldsymbol{x}) \approx V^{\star}(\boldsymbol{x}) \quad \text{if} \quad \min_{j=1}^{d} x_j \gg 0.$$

$\Rightarrow$ If we know $V^{\star}$, we directly obtain an approximation for the probability that

*at least one of the components $X^{\star}_j$ exceeds a high threshold $x_j$.*

# Multivariate threshold exceedances using structure variables

Using the **maximum norm** $\|y\|_\infty = \max_{j=1}^d |y_j|$,
the condition $\boldsymbol{X}^\star \not\leq \boldsymbol{x}$ for $\boldsymbol{x} > \boldsymbol{0}$ can be reformulated as

$$\|\boldsymbol{X}^\star/\boldsymbol{x}\|_\infty \geq 1,$$

and we can directly use the MRV approximation.

We will generalize the MRV condition by considering more general norms $\|\cdot\|$ on $\mathbb{R}^d$, and we define extreme events as

$$\|\boldsymbol{X}^\star\| \geq r$$

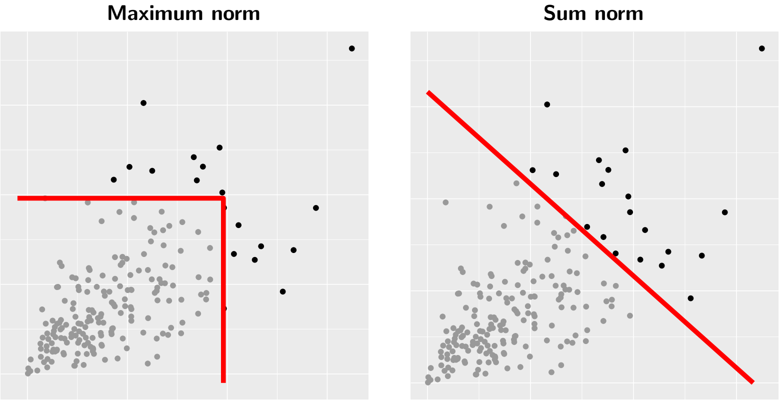
above some "high enough" threshold $r > 0$ for the **aggregated value** $\|\boldsymbol{X}^\star\|$.

**Examples :**

- sum-norm (or $L^1$-norm) : $\|\boldsymbol{y}\|_1 = |y_1| + \ldots + |y_d|$,
- weighted sum norm : $\|\boldsymbol{y}\|_{1,\boldsymbol{u}} = |y_1/u_1| + \ldots + |y_d/u_d|$ for $\boldsymbol{u} = (u_1, \ldots, u_d) > \boldsymbol{0}$
- weighted maximum norm : $\|\boldsymbol{y}\|_{\infty,\boldsymbol{u}} = \max_{j=1}^d |y_j/u_j|$ for $\boldsymbol{u} > \boldsymbol{0}$
  $\Rightarrow \boldsymbol{y} \not\leq \boldsymbol{u}$ is equivalent to $\|\boldsymbol{y}\|_{\infty,\boldsymbol{u}} > 1$,
- Euclidean norm (or $L^2$-norm) : $\|\boldsymbol{y}\|_1 = (y_1^2 + \ldots + y_d^2)^{1/2}$,
- and a wide variety of other relevant norms.

The **unit sphere** of a norm $\|\cdot\|$ consists of all points $\boldsymbol{y}$ satisfying $\|\boldsymbol{y}\| = 1$.

Black dots are threshold exceedances of observations $\|\boldsymbol{x}^\star\|$ above the threshold.



# Pseudo-polar decomposition and angular measures

Given a norm $\|\cdot\|$, theory shows that the MRV condition is equivalent to :

$$t\,P\left(\frac{\boldsymbol{X}^\star}{\|\boldsymbol{X}^\star\|}\in A,\ \|\boldsymbol{X}^\star\|>tr\right)\to H(A)\times\frac{\kappa}{r}\quad(t\to\infty)$$

with

- the **angular measure** $H(\cdot)$, a probability distribution defined on the unit sphere $\Xi$ of $\|\cdot\|$ ;
- the **extremal coefficient** $\kappa=\kappa_{\|\cdot\|}>0$, a constant that depends on $V^\star$ and $\|\cdot\|$.

The angular measure indicates along which **directions** (or **angles**) $\boldsymbol{W}=\boldsymbol{X}^\star/\|\boldsymbol{X}^\star\|$ the extreme values in $\boldsymbol{X}^\star$ concentrate as the **magnitude** $R=\|\boldsymbol{X}^\star\|$ increases towards infinity.

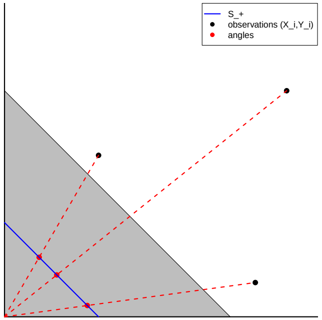
The **classical extremal coefficient** $\kappa=\theta_d=V^\star(1,\ldots,1)$ arises for the **maximum norm** $\|\boldsymbol{y}\|_\infty=\max_{j=1}^d|y_j|$.

⚠ Sometimes, one defines the angular measure as $\kappa\,H(\cdot)$ with overall mass $\kappa$.
⚠ We could also choose different norms for the direction and the magnitude.

# Illustration of pseudo-polar decomposition

Left : sum-norm for magnitudes and the unit sphere
Right : maximum norm for magnitudes, sum-norm for the unit sphere

Observed directions $\boldsymbol{W}_i$ correspond to red dots on the unit sphere.



## Interpretation of magnitudes and directions

Recall that we have the general multivariate regular variation condition given as

$$t\,P\left(\frac{\boldsymbol{X}^\star}{\|\boldsymbol{X}^\star\|} \in A,\ \|\boldsymbol{X}^\star\| > tr\right) \to H(A) \times \frac{\kappa}{r} \quad (t \to \infty)$$

- The magnitude $R = \|\boldsymbol{X}^\star\|$ has tail behavior according to a **Pareto distribution with shape $1$ and scale $\kappa$** :

$$r\,P(R/\kappa > r) \to 1 \quad (r \to \infty).$$

- The direction $\boldsymbol{W} = \boldsymbol{X}^\star/\|\boldsymbol{X}^\star\|$ tends to the angular measure as $R$ increases :

$$P\left(\boldsymbol{W} \in A \mid R > r\right) \to H(A) \quad (r \to \infty).$$

- **$R$ and $W$ become stochastically independent for exceedances $R > r$** when $r \to \infty$.

# Angular measures on the unit simplex

Often, it is practical to use the **sum-norm** $\|\boldsymbol{x}\|_1 = x_1 + x_2 + \ldots + x_d$, $\boldsymbol{x} \geq \mathbf{0}$.

In this case, the **angular measure** $H$ **is defined on the unit simplex**

$$\Xi = \left\{ \boldsymbol{x} \in \mathbb{R}^d \mid \boldsymbol{x} \geq \mathbf{0}, \ \sum_{j=1}^{d} x_j = 1 \right\},$$

and $\kappa_{\|\cdot\|_1} = d$ **does not depend on the extremal dependence structure**.

For coherence, $H$ must satisfy the **moment constraints**

$$\mathbb{E} W_j = 1/d, \quad \text{if} \quad \boldsymbol{W} \sim H.$$

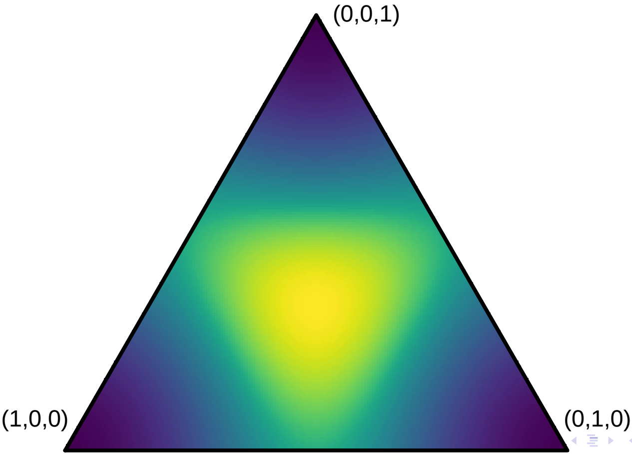
**Why ?**
Since $G_j^\star(z) = \exp(-1/z)$ for $z > 0$,
we have $V^\star(\infty, \ldots, \infty, z, \infty, \ldots, \infty) = 1/z$ for any $z > 0$ and for all $j = 1, \ldots, d$.

## Example : Dirichlet angular measure for $d = 3$

The Dirichlet distribution is a generalization of the Beta distribution to $d > 2$.

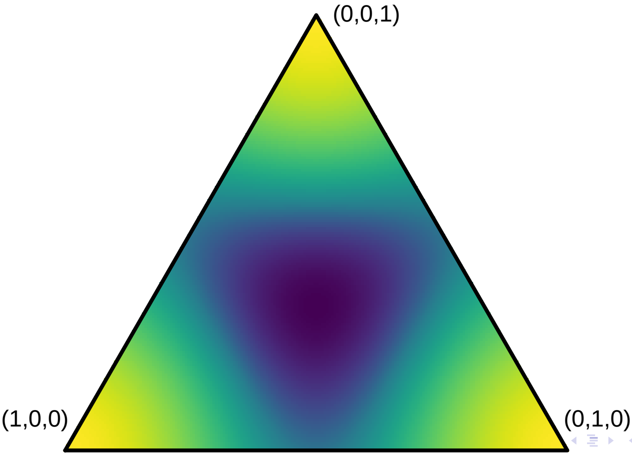
**Example :** $d = 3$, **sum-norm** $\|\boldsymbol{x}\|_1 = x_1 + \ldots + x_d$, Dirichlet shape parameter $\alpha = 2$.

Mass concentrates in the middle $\Rightarrow$ relatively strong extremal dependence.



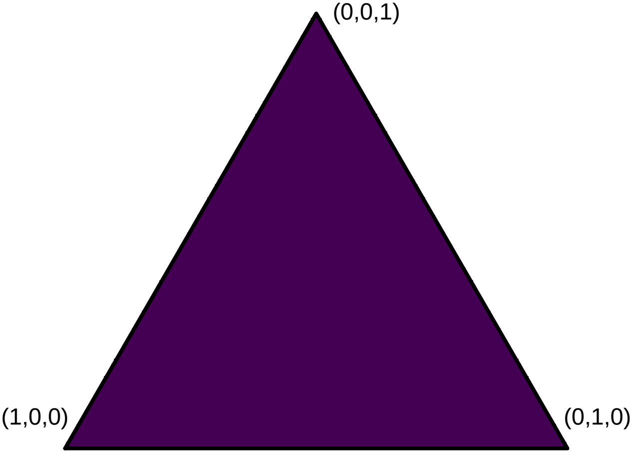
# Example : Dirichlet angular measure

$d = 3$, sum-norm, Dirichlet shape parameter $\alpha = 0.5$.
Mass concentrates towards the coordinate axes around $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$
$\Rightarrow$ relatively weak extremal dependence.



## Example : Dirichlet angular measure

$d = 3$, sum-norm, Dirichlet shape parameter $\alpha = 1 \Rightarrow$ uniform distribution (same density of the angular measure throughout the simplex)

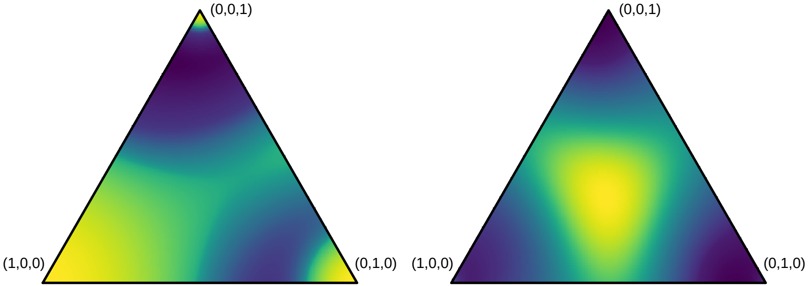


## Example : Pairwise beta model ($d = 3$, sum-norm)

**Pairwise beta model** for $H$ :

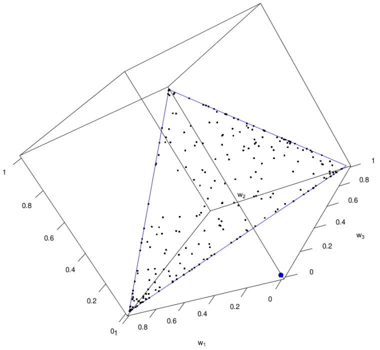
- not "fully symmetric" as the Dirichlet
- appropriately combines univariate Beta distributions to a valid angular measure $H$
- 1 **global concentration parameter** + **1 parameter for each component pair** $j_1, j_2$

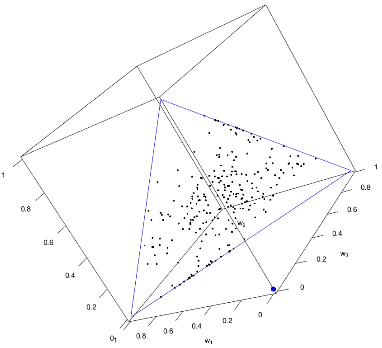
Illustrations for two different parameter configurations :
left : weak concentration ; right : strong concentration



# Example : Simulation of the pairwise beta model

Perspective-plots of 250 realisations $\boldsymbol{W}_i$ of $H$ for the parameter configurations of the preceding slide



# How to define multivariate extreme events in practice

**Approach :**

1. Standardize marginal distributions in data.
2. Choose a norm $\|\cdot\|$ and fix a high threshold $r$.
3. Consider events $\boldsymbol{X}$ with
$$\|\boldsymbol{X}^\star\| \geq r$$
as being extreme.

**Example : (generalized) maximum norm**

- We consider an event $\boldsymbol{X}$ as extreme if we observe a **threshold exceedance** $X_j > u_j$ **in at least one of the components** $j = 1, \ldots, d$.
- Standardized marginal thresholds are $u_j^\star = 1/(1 - F_j(u_j))$

$\Rightarrow$ we use the generalized maximum norm defined by

$$\|x^\star\|_{\infty,\boldsymbol{u}^\star} = \max_{j=1,\ldots,d} x_j^\star / u_j^\star$$

and fix the threshold to $r = 1$.

The extremal coefficient associated to this norm is $\kappa = V^\star(u_1^\star, \ldots, u_d^\star)$.

# Even more general choices for defining extreme events

In the **pseudo-polar decomposition**, we can even generalize the norm $\|\cdot\|$ to **any homogeneous function $\ell$ continuous in $\mathbf{0}$**. Then,

$$t\,P\left(\frac{\boldsymbol{X}^\star}{\ell(\boldsymbol{X}^\star)} \in A,\ \ell(\boldsymbol{X}^\star) > tr\right) \to H_\ell(A) \times \frac{\kappa_\ell}{r}, \quad t \to \infty$$

Note : homogeneity of $\ell$ means that $\ell(t\boldsymbol{x}) = t\,\ell(\boldsymbol{x})$ for any $t > 0$, $\boldsymbol{x} \geq \mathbf{0}$.

**Examples for $\ell$ that are not norms :**

- $\ell(\boldsymbol{x}) = \min_{j=1,\ldots,d} x_j$,
- $\ell(\boldsymbol{x}) = x_{(j_0)}$ for $j_0 < d$ ($j_0$th order statistics), or
- $\ell(\boldsymbol{x}) = x_{j_0}$ for a fixed component $j_0$.

In some cases, $\kappa_\ell H_\ell \equiv 0$ is possible
(for instance, with $\ell = \min$ and asymptotic independence).

The choice of $\ell(\boldsymbol{x}) = x_{j_0}$ corresponds to the **conditional extremes approach**.

# Calculating extreme event probabilities

We assume that we know (or have estimated) the angular measure $H_a$ and the extremal coefficient $\kappa_a$ for a specific norm $\|\cdot\|_a$.

**Question :** What is the **probability of an extreme event** $E = \{\|X^\star\|_b \geq r\}$, defined with respect to a different norm $\|\cdot\|_b$ ?

**Response :** we have to **reweight the probabilities of** $H_a$ according to the new norm $\|\cdot\|_b$ :

$$P(E) = \frac{\kappa_b}{r} = \frac{\kappa_a}{r} \int_{\Xi_a} \|\boldsymbol{w}\|_b \, H_a(\mathrm{d}\boldsymbol{w}), \quad \Xi_a = \{\boldsymbol{w} \in [0,\infty)^d \mid \|\boldsymbol{w}\|_a = 1\}$$

**Examples :**

- $E = \{X_j > u_j \text{ for at least one } j\}$
  $\Rightarrow$ set $u_j^\star = 1/(1 - F^\star(u_j))$, $\|\boldsymbol{x}\|_b = \max_{j=1,\ldots,d} x_j/u_j^\star$ and $r = 1$

  Remark : it follows that $V^\star(\boldsymbol{z}) = \kappa_a \int_{\Xi_a} \max_{j=1,\ldots,d} \frac{w_j}{x_j} H_a(\mathrm{d}\boldsymbol{w})$

- $E = \{X_j > u_j \text{ for all } j\}$
  $\Rightarrow$ set $u_j^\star = 1/(1 - F^\star(u_j))$, $\ell_b(\boldsymbol{x}) = \min_{j=1,\ldots,d} x_j/u_j^\star$ and $r = 1$

## Wrap-up : theory for multivariate extremes

- Multivariate limit theory is based on the **maximum domain of attraction condition** and on **multivariate regular variation** :
  - **max-stable limit distributions** for **componentwise maxima** ;
  - **angular measures** when considering **threshold exceedances**.
- ⚠ Componentwise block maxima may be easy to extract from data, but they may contain components from different extreme events
  $\Rightarrow$ interpretation can be difficult
- ⚠ There is **no unique definition of multivariate extreme events**.

  $\Rightarrow$ How to characterize the magnitude of extreme events in practice ?

  $\Rightarrow$ Appropriate **norms** $\|\cdot\|$ **or aggregation functions** $\ell$ provide high flexibility : an **extreme event arises if** $\|\boldsymbol{X}^\star\|$ **or** $\ell(\boldsymbol{X}^\star)$ **exceeds a high fixed threshold** $r$.

The standard **multivariate extremal coefficient $\theta_d$** is defined as $\theta_d = V^\star(1,\dots,1)$.

Let's assume that $\boldsymbol{X}$ is in a multivariate domain of attraction.

For component maxima $\boldsymbol{M}_n^\star = \left(\max_{i=1}^n X_{i1}^\star, \dots, \max_{i=1}^n X_{id}^\star,\right)^T$, we obtain

$$P(M_{n,1}^\star/n \le x, \dots, M_{n,d}^\star/n \le x) \to \exp\left(-\frac{\theta_d}{x}\right), \quad x > 0 \quad (n \to \infty)$$

(whereas $P(M_{n,i}^\star/n \le x) \to \exp\left(-\frac{1}{x}\right),\ x > 0\ (n \to \infty)$).

For threshold exceedances, we obtain

$$x\,P\left(\max_{j=1,\dots,d} X_j^\star > x\right) \to \theta_d \quad (x \to \infty).$$

**Interpretation :** $\theta_d \in [1, d]$ is the **average number of independent clusters** among the components $1, \dots, d$ when an extreme event occurs, and $d/\theta_d$ is the average cluster size.

# Asymptotic dependence and asymptotic independence

**Asymptotic independence** between two variables $X_1$ and $X_2$ arises if

$$P(X_2^\star > x \mid X_1^\star > x) \to 0 \quad (x \to \infty).$$

It is a property of bivariate distributions ($d = 2$).

If $d > 2$, we can check on asymptotic (in)dependence by considering all pairs $(X_{j_1}, X_{j_2})$, $1 \leq j_1 < j_2 \leq d$.

If all pairs of $\boldsymbol{X}$ are asymptotically independent, then we can say that we have **full asymptotic independence**, and in this case we have $\theta_d = d$.

For $d \geq 2$, we say that $X_1, \ldots, X_d$ is **jointly asymptotically independent** if

$$x\, P\left(\min_{j=1,\ldots,d} X_j^\star > x\right) \to 0, \quad x \to \infty.$$

⚠ Some pairs of $\boldsymbol{X}$ may still be asymptotically dependent if $d > 2$ !
Example : $(X_1, X_2)^T$ asymptotically dependent, and $X_3$ independent of $(X_1, X_2)^T$

# Multivariate coefficient of tail dependence

In the case of **joint asymptotic independence**,
the following assumption on joint tail behavior is flexible and useful :

$$P(\min_{j=1,\dots,d} X_j^\star > x) = \mathcal{L}(x) x^{-1/\eta_d} \quad (x \to \infty),$$

where $\mathcal{L}(\cdot)$ is slowly varying and $\eta_d > 0$ **is called the coefficient of tail dependence**.

- slow variation : $\mathcal{L}(tx)/\mathcal{L}(t) \to 1$ for all $x > 0$ as $t \to \infty$
- $\eta_d$ is the extreme-value index of $\min_{j=1,\dots,d} X_j^\star$
- for $d = 2$, we have $\eta_2 = (1 + \overline{\chi})/2$ with $\overline{\chi} \in (-1, 1]$
- joint asymptotic dependence of $\boldsymbol{X} \Rightarrow \eta_d = 1$
- classical independence of components of $\boldsymbol{X} \Rightarrow \eta_d = 1/d$
- joint asymptotic independence with positive association $\Rightarrow \eta_d \in [1/d, 1]$

**Simulation experiments with the pairwise beta model**

1. Simulate and visualize the pairwise beta angular measure.
   1. Load the `BMAmevt` package and use the function `rpairbeta` to simulate $n = 200$ points $\boldsymbol{w}_i$ for $d = 3$ according to the angular measure $H$ associated to the sum-norm.
   2. Use the function `persp3d` from the `rgl` package (or any other nice 3D plotting function) to interactively visualize the simulated points.
   3. Explore the role of the parameter vector pars in `rpairbeta(n,dimData=3,par=pars)`, for instance for `pars=c(a,b,b,b)` with $a \in \{0.1, 25\}$ and $b \in \{0.1, 10\}$.

2. Based on the $n$ points $\boldsymbol{w}_i$ simulated in Exercice 1, calculate an empirical estimate of $P(E) = P(X_j > F_j^{-1}(0.9), j = 1, 2, 3)$, that is, of the event $E$ of joint exceedance above the marginal 90% quantile.
   - Each $\boldsymbol{w}_i$ has weight $1/n$ in the empirical angular measure $\hat{H}$.
   - The thresholds on the standard scale are $u^\star = 1/(1 - 0.9) = 10$.

   Therefore,
   1. Implement a function `ell.min=function(w,u.star=10){...}` which returns $\min(w_1/u.star, w_2/u.star, w_3/u.star)$.
   2. Implement the estimation of the extreme event probability $P(E)$ as $\frac{d}{n}\sum_{i=1}^{n} \ell_{\min}(\boldsymbol{w}_i)$.
   3. Apply this estimation for `pars=c(5,1,1,1)` and `pars=c(0.1,1,1,1)`, and interpret the results.

**Dependence summaries of air pollution data**

Load the $NO_2$ air pollution data for 5 measurement stations :
`load("nitrogenDioxideMsmts.RData")`
(see the Practicals of Days 1 and 2)

1. Implement the rank-based empirical cdf to transform to standard Pareto margins :
   1. $r_{ij}$ = rank of component $j$ of observation $i$ among all observations $i = 1, \dots, n$ ;
   2. $x_{ij}^\star = 1/(1 - r_{ij}/(n+1))$.
2. Implement the estimation of the extremal coefficient $\theta_5$ through the likelihood estimator
$$\hat{\theta}_5 = u^\star \times \text{(number of observations with } \max_{j=1,\dots,5} x_{ij}^\star > u^\star)/n,$$
   and plot the estimates $\hat{\theta}_5(p)$ for $u^\star = 1/(1-p)$ with $p = 0.8, 0.81, \dots, 0.99$. For each threshold value, you can provide 95% pointwise confidence intervals $[\hat{\theta}_5 - 1.96 \times \mathrm{se}, \hat{\theta}_5 + 1.96 \times \mathrm{se}]$ with
$$\mathrm{se} = \sqrt{n^{-1}(\hat{\theta}_5/u^\star) \times (1 - \hat{\theta}_5/u^\star)}.$$
   Is the estimate stable with respect to the quantile level ?
3. Estimate the coefficient of tail dependence $\eta$ by using the Hill estimator for the sample of minima $\min_{j=1,\dots,5} x_{ij}^\star$. Interpret the result.
   Remark : the `hillPlot` function in the `fExtremes` package is useful.

# Parametric models for multivariate extremes

**Good extreme-value models** should have a **sound asymptotic motivation** and be **flexible** for accurately modeling of extreme event probabilities.

We **focus on asymptotic models** which arise as **limit distributions**.
Such models focus on the situation of **asymptotical dependence**.

Moreover, we focus on estimation using **threshold exceedances, not block maxima**.

# Specification of multivariate dependence models

A multivariate dependence model can be specified in different ways, for instance :

- through the formula of the exponent function $V^\star(\cdot)$
- through the density of the angular measure $H$ and the extremal coefficient $\kappa$ associated to some norm $||\cdot||$
- constructively, through the so-called spectral construction of max-stable random vectors

# Parametric models for the angular measure

Models for the angular measure must be defined over the unit sphere $\Xi$ of a norm $\|\cdot\|$.

⚠ Since margins of $V$ are standardized,

$$V^\star(\infty,\ldots,\infty,z,\infty,\ldots,\infty) = 1/z,$$

angular measures $H$ must fulfill certain **moment constraints**.

**Example :** sum-norm $\|\boldsymbol{x}\|_1 = x_1 + \ldots + x_d \Rightarrow \mathbb{E}W_j = 1/d$ for $\boldsymbol{W} \sim H$.

Some models are naturally specified through the density $h$ of $H$ on $\|\cdot\|_1$ :

- symmetric Dirichlet distribution (1 global parameter $\alpha > 0$) with probability density

$$h(\boldsymbol{w}) = \frac{\Gamma(\alpha)^d}{\Gamma(d\alpha)} \prod_{j=1}^{d} w_j^{\alpha-1}$$

- pairwise beta distribution
  (1 global parameter + 1 parameter for each pair $1 \leq j_1 < j_2 \leq d$ )

and many others...

## Example : multivariate logistic model

The **multivariate logistic model** was introduced by Emil J. Gumbel in 1960 and can be easily defined through its exponent function

$$V^\star(\mathbf{z}) = \left(z_1^{-1/\alpha} + \ldots + z_d^{-1/\alpha}\right)^\alpha ,$$

such that

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

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

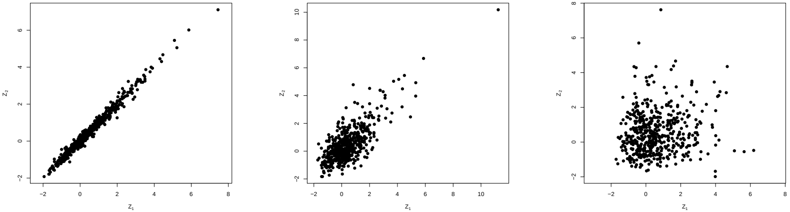
- perfect dependence for $\alpha \to 0$ ;
- independence for $\alpha = 1$.

# Example : Simulations of bivariate logistic distribution

$n = 500$ realizations

Bivariate scatterplots show $\log \boldsymbol{Z}^\star$ (standard Gumbel margins)

$\alpha = 0.1$ $\alpha = 0.5$ $\alpha = 0.9$



**Huesler–Reiss distributions** are related to multivariate Gaussian distributions. Consider

$$\tilde{\boldsymbol{Y}}_i \sim \mathcal{N}\left(-(\sigma_{11}, \ldots, \sigma_{dd})^T/2, \Sigma\right)$$

with the variance-covariance matrix $\Sigma = \left(\sigma_{j_1 j_2}\right)_{1 \leq j_1, j_2 \leq d}$.

**Bivariate distribution function** with parameter $r = 2/\sqrt{\mathrm{Var}(\tilde{Y}_2 - \tilde{Y}_1)} > 0$ :

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

$\Rightarrow$ independence for $r \to 0$, perfect dependence for $r \to \infty$

**Exponent function for general** $d$ **:** $V^\star(\boldsymbol{z}) = \sum_{j=1}^d z_j^{-1} F_{|j}(\boldsymbol{z}/z_j \mid \Sigma)$,
where the multivariate distribution functions $F_{|j}$ are log-Gaussian as follows :

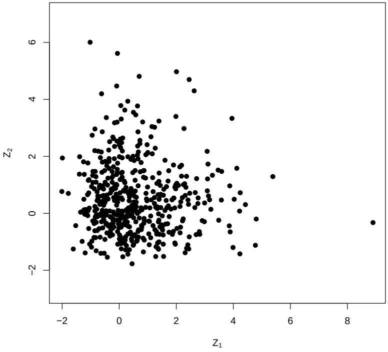
$$\exp(\tilde{\boldsymbol{Y}} - \tilde{Y}_j - 0.5\,\mathrm{Var}(\tilde{Y} - \tilde{Y}_j)) \sim F_{|j}(\cdot \mid \Sigma)$$

**Model parameters :** $d(d-1)/2$ values $\mathrm{Var}(\tilde{Y}_{j_2} - \tilde{Y}_{j_1})$ for $1 \leq j_1 < j_2 \leq d$

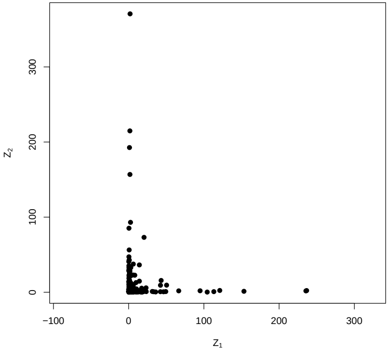
# Example : Simulations of bivariate Huesler–Reiss distribution

$n = 500$ realizations
$r = 0.25$

$\log \boldsymbol{Z}^\star$ (Gumbel margins)



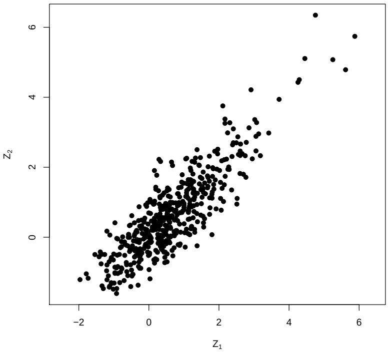

$\boldsymbol{Z}^\star$ (Fréchet margins)



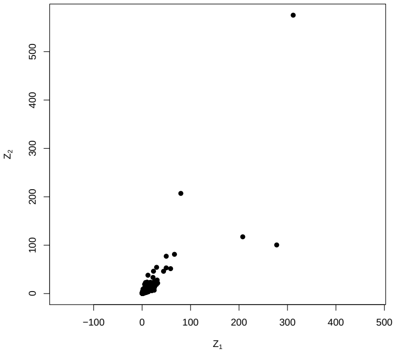
# Example, cont'd

$n = 500$ realizations
$r = 3$

$\log \boldsymbol{Z}^{\star}$ (Gumbel margins)




$\boldsymbol{Z}^{\star}$ (Fréchet margins)



# Estimation approaches

**We broadly distinguish two types of estimation approaches :**

- **nonparametric estimation :** no parametric assumptions at all

  $\Rightarrow$ useful in relatively small dimension $d$ (**curse of dimensionality for larger** $d$)

  $\Rightarrow$ in the following, we consider empirical angular measures

- **parametric estimation :** a relatively small number of parameters determines the dependence structure

# Marginal standardization in practice

⚠ Estimating marginal parameters and the dependence structure jointly in a single step is often not feasible, especially when $d > 2$.

Instead, we often adopt a **two-step procedure** :

1. Estimate a univariate extremes model for the margins $j = 1, \dots, d$.
2. Estimate the extremal dependence model using marginally standardized data.
   - **Option 1 :** Use the estimated marginal model to standardize data $\boldsymbol{X}$ to $\boldsymbol{X}^\star$.
   - **Option 2 :** Use the **rank transform**

     $\Rightarrow$ this means that the empirical distribution function is the marginal model :

     - $R_{ij}$ = rank of $X_{ij}$ among $X_{kj}$, $k = 1, \dots, n$
       $\Rightarrow R_{ij} \in \{1, 2, \dots, n\}$
     - $X^\star_{ij} = 1/(1 - R_{ij}/(n+1)))$
       (empirical transformation to standard Pareto scale)

## Empirical angular measures

**Recall :** Multivariate regular variation and pseudo-polar decomposition give

$$t\,P\left(\frac{\boldsymbol{X}^\star}{\|\boldsymbol{X}^\star\|}\in A,\ \|\boldsymbol{X}^\star\|>tr\right)\to H(A)\times\frac{\kappa}{r}\quad(t\to\infty)$$

- angular measure $H$ defined on the unit sphere with respect to $\|\cdot\|$
- extremal coefficient $\kappa>0$ associated to $\|\cdot\|$

**Estimation :** **empirical angular measure $\hat{H}$**

- fix a threshold value $r_0$ for $r=\|\boldsymbol{x}_i^\star\|$
- count the number of exceedances $n_0=\sum_{i=1}^n 1(R_i>r_0)$
- calculate angles $\boldsymbol{w}_i=\boldsymbol{x}_i^\star/\|\boldsymbol{x}_i^\star\|$ and define the estimator

$$\hat{H}(A)=\frac{1}{n_0}\sum_{i=1}^n 1(R_i>r_0)1(\boldsymbol{w}_i\in A)$$

  of the angular measure of sets $A$ on the unit sphere

The **extremal coefficient $\kappa$ of $\|\cdot\|$** (if it is not invariant) can be estimated as follows :

$$\hat{\kappa}=r_0\frac{n_0}{n}.$$

# Parametric maximum likelihood estimation

For models with a parametric **probability density** $f_\gamma$ and only **moderately many parameters in** $\gamma$ to estimate, we can apply **maximum likelihood estimation**.

**Example :** maximum likelihood estimation of **angular measures**

- We need the density $f_\gamma = h$ of the angular measure $H$.
- Observations : angles $\{\boldsymbol{w}_i \mid r_i > r_0\}$ associated to threshold exceedances of the magnitude value $r_i = \|\boldsymbol{x}_i^\star\|$.

**Principle :** We maximize the log-likelihood function (usually done numerically)

$$\gamma \mapsto \sum_{\mathrm{obs}} \log\left(f_\gamma(\mathrm{obs})\right)$$

to obtain the **maximum likelihood estimator** $\hat{\gamma}$.

# Multivariate likelihood estimation with censoring

What parametric models make sense for the distribution function $F^\star_{\boldsymbol{X}}$ of standardized data $\boldsymbol{X}^\star$ ?

Based on the **MRV condition**

$$t\,(1 - F^\star_{\boldsymbol{X}}(t\boldsymbol{x})) \to V^\star(\boldsymbol{x}) \quad (t \to \infty),$$

useful **modeling assumptions** are

1. $F^\star_{\boldsymbol{\gamma}}(\boldsymbol{x}) = 1 - V^\star(\boldsymbol{x})$ with parameter vector $\boldsymbol{\gamma}$ for data $\boldsymbol{x}$ with $\min_j x_j \gg 0$,

   or, very similarly,

2. $F^\star_{\boldsymbol{\gamma}}(\boldsymbol{x}) = \exp(-V^\star_{\boldsymbol{\gamma}}(\boldsymbol{x})) = G^\star(\boldsymbol{x})$ for data $\boldsymbol{x}$ with $\min_j x_j \gg 0$ (this is based on the approximation $1 - \varepsilon \approx \exp(-\varepsilon)$ for small $\varepsilon$)

$\Rightarrow$ **For estimation, we have to censor non-extreme observations with $\|\boldsymbol{x}^\star_i\| \not\gtrsim \boldsymbol{u}^\star$ that fall below some high threshold $\boldsymbol{u}^\star$.**

## Censored univariate log-likelihood

To explain the **censoring approach**, let us consider the univariate case $d = 1$.

We need :

- distribution function $F_\gamma$ with probability density $f_\gamma$
- data $x_i \sim F_\gamma$, $i = 1, \dots, n$
- a **threshold** $u$

Classical (uncensored) likelihood function :

$$\gamma \mapsto \sum_{i=1}^{n} \log\left(f_\gamma(x_i)\right)$$

By contrast, in the **censored log-likelihood** $\Rightarrow$, the log-likelihood contribution of $x_i$ is either

- $\log\left(f_\gamma(x_i)\right)$ if $x_i > u$ (uncensored data points), or
- $\log(F_\gamma(u))$ if $x_i \leq u$ (censored data point).

$\Rightarrow$ Censoring prevents low values from influencing the estimates $\hat{\gamma}$ of tail parameters.

# Partially censored multivariate likelihood estimation

- data vectors $x_{ij}$, $i = 1, \ldots, n$, $j = 1, \ldots, d$
- threshold vector $\boldsymbol{u} = (u_1, \ldots, u_d)$
- **three censoring scenarios** are possible :
  - $\boldsymbol{x}_i \leq \boldsymbol{u} \Rightarrow$ **fully censored data vector** with log-likelihood contribution
  $$\log \left(F_\gamma(\boldsymbol{u})\right)$$
  - $\boldsymbol{x}_i > \boldsymbol{u} \Rightarrow$ **uncensored data vector** with log-likelihood contribution
  $$\log \left(f_\gamma(\boldsymbol{x}_i)\right)$$
  - $\boldsymbol{x}_i \not\leq \boldsymbol{u}$ and $\boldsymbol{x}_i \not> \boldsymbol{u} \Rightarrow$ **partially censored data vector** :

    assume (without loss of generality) that $x_{ij} > u_j$ for $j = 1, \ldots, d_1$, and $x_{ij} < u_j$ for $j = d_1 + 1, \ldots, d \Rightarrow$ log-likelihood contribution is
  $$\log \left( \frac{\partial^{d_1}}{\partial y_1 \times \ldots \times \partial y_{d_1}} F_\gamma(y_1, \ldots, y_{d_1}, u_{d_1+1}, \ldots, u_d) \,|_{y_j = x_{ij}, j=1,\ldots,d_1} \right)$$

**Implementation example :** function `fbvpot` in the `evd` package for a number of bivariate extreme value models : `fbvpot(...,likelihood="censored",...)`

# Block maxima vs. threshold exceedances

- **Block maxima** data (such as yearly or monthly blocks)
  - very relevant when we have **long observation periods**
  - goal is often to make **long-term predictions**
  - we do not care about the behavior within blocks
  - theoretical results often based on unrealistic **iid assumption for original event data**
  - potentially strong **loss of information** since there is only one observation for each block
- **Threshold exceedance** data
  - stronger focus on the **original event data**
  - modeling intra-block behavior such as **seasonal** and **temporal dependence**

## Choice of thresholds

**Good estimator properties** (consistency, asymptotic normality) have been shown **for many estimators** of extreme value parameters when the sample size grows to infinity.

**Bias-variance trade-off** : larger bias for lower thresholds, but higher estimation variance for higher thresholds.

Focus of this part of the course was mostly on **asymptotically stable models with asymptotic dependence**.

**Modeling extremal dependence in air pollution data**

We continue working on the $NO_2$ data in `nitrogenDioxideMsmts.RData`. Previous results indicate that the 5 variables are asymptotically independent. Asymptotically dependent models will therefore tend to provide "conservative" estimates of joint risks.

1. Estimate the empirical angular measure $\hat{H}$ for the sum-norm, with magnitude threshold $r_0$ chosen as the empirical 95% quantile of observations $r_i = \|\boldsymbol{x}_i^\star\|$. Represent the empirical angular measure through a matrix where each row shows the coordinates $\boldsymbol{w}_i$ of an atom of $\hat{H}$.

2. Estimate the probability of joint exceedance of all components $X_j$ above the overall 95% quantile of all observed values (on the original marginal scale) by reweighting the atoms of $\hat{H}$, and compare the result to the empirically observed probability of this event.

3. Estimate the shape parameter $\alpha$ of the Dirichlet angular measure $H$ for the sum norm. Note : the density $h$ of $H$ is available through the function `ddirichlet` in the `MCMCpack` package.

   1. Define a function `llfun(alpha)` that returns the log-likelihood for the angle atoms of $\hat{H}$ above. If $\alpha \leq 0$, the function should return `-Inf`.
   2. Use the `maxLik` function from the `maxLik` package to calculate the maximum likelihood estimate of $\alpha$. Use $\alpha_0 = 1$ as initial parameter.

4. Consider the original observations $\boldsymbol{x}$ with marginal thresholds $u_j$, $j = 1, \ldots, d$, fixed to the empirical 0.95-quantile. For each pair of indices $1 \leq j_1 < j_2 \leq d$, use the bivariate censored likelihood maximization implemented in the `fbvpot` function of the `evd` package to estimate the dependence parameters $r_{j_1,j_2}$ of the Huesler–Reiss distribution for all pairs of variables. Interpret the resulting set of dependence parameter estimates with respect to the extremal dependence strength.