Non-Parametric Simulation of Multivariate Extreme Events via Spectral Bootstrap

joint with Juliette Legrand (Université de Bretagne Occidentale) and Nisrine Madhar (Université Paris Cité & Natixis)

Maud Thomas

ISFA, Université Lyon 1

Elements of extreme value theory

Goals of Extreme Value Theory

Goals of Extreme Value Theory

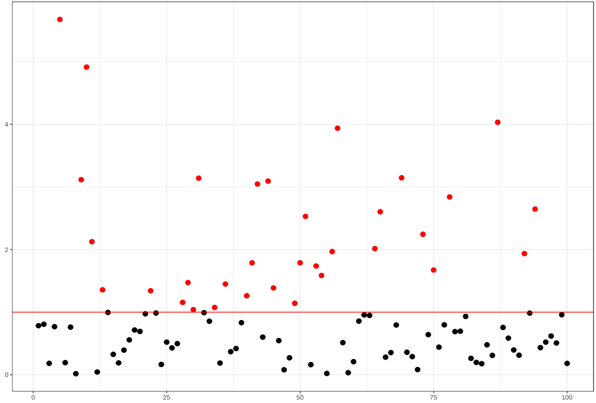
1. Estimate the probability of occurrence of an event more severe/extreme than previously observed
2. Estimate an extreme quantile

⇒ Inference outside the sample support

Univariate Peaks-over-Threshold method

- $Y_1, Y_2, \ldots$ a series of i.i.d. random variables
- Fix a (high) threshold u
- Extreme event = Y_i exceeds u

  $\rightarrow$ Given that $Y_i > u$, an excess is defined by $Z_i = Y_i - u$



Univariate Peaks-over-Threshold method

- $Y_1, Y_2, \ldots$ a series of i.i.d. random variables
- Fix a (high) threshold u
- Extreme event = Y_i exceeds u

  $\rightarrow$ Given that $Y_i > u$, an excess is defined by $Z_i = Y_i - u$
- Excess distribution

$$\overline{F}_u(z) = P[Y_1 - u > z \mid Y_1 > u] = \frac{\overline{F}(u+z)}{\overline{F}(u)},\ z > 0\,.$$

Balkema et de Haan (1974), Pickands (1975)

Under certain conditions, the distribution of excesses F_u converges, as $u \to \infty$, to a generalized Pareto distribution (GPD) whose distribution function is

$$H_{\sigma,\gamma}(z) = \begin{cases} 1 - \left(1 + \frac{\gamma}{\sigma} z\right)^{-1/\gamma} & \text{if } \gamma \neq 0 \\ 1 - \exp\left(-\frac{z}{\sigma}\right) & \text{if } \gamma = 0 \end{cases}$$

- Families of possible distributions for excesses = parametric family

  $\hookrightarrow$ Generalized Pareto distributions (GPD)

Generalized Pareto distributions

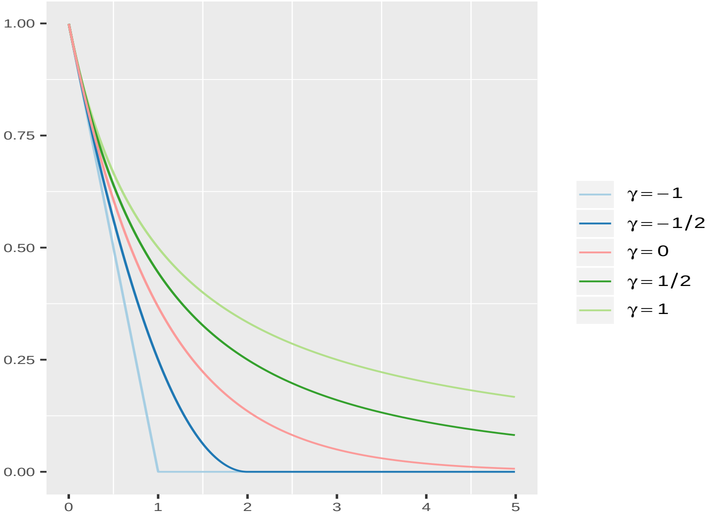


Figure: GPD survival functions

3 domains of attraction

1. Fréchet domain ($\gamma > 0$): **heavy-tailed distributions**

$$1 - H_\gamma(z) \underset{+\infty}{\sim} \gamma^{-1/\gamma} z^{-1/\gamma}$$

   Examples: Cauchy, Log-gamma, Student

2. Gumbel domain ($\gamma = 0$): **thin tail distributions**

$$1 - H_0(z) \underset{+\infty}{\sim} \exp(-z)$$

   Examples: Gaussian, Gamma, Exponential

3. Weibull domain ($\gamma < 0$): **finite tail distributions**

$$1 - H_\gamma(z) = 0 \quad \text{for } x \geq -1/\gamma$$

   Examples: Uniform, Beta

Non parametric simulation of multivariate extremes via spectral bootstrap

Motivation and goal

- Extreme Value Theory (EVT) focuses on behaviour in the tail of multivariate distributions.
- Few extreme observations → difficult inference
  → Natural idea = bootstrap to increase the number of data
- BUT, standard bootstrap fails for EVT

Goal

Non-parametric simulation approach to expand the number of observations above the extreme level
→ Extension of the non-parametric simulation algorithm for multivariate extremes, developed by Legrand et al. [2023] in the dimension 2 case to larger dimensions

Main contributions

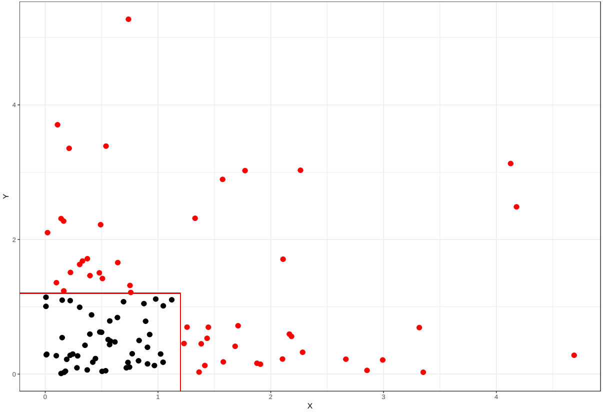
- Expands the number of observation above extreme level
- Ensures more reliable estimations
- Enables extrapolation beyond the range of observed data

→ Makes use of a **spectral representation** of multivariate extreme vectors

Lyon 1 Université Claude Bernard

Multivariate Generalized Pareto Distributions

- $\mathbf{X} = (X_1, \ldots, X_d)$ observations
- Choose (high) thresholds $\mathbf{u} = (u_1, \ldots, u_d)$
- Extreme event = AT LEAST one of the X_j exceeds its threshold u_j



Multivariate Generalized Pareto Distributions

- $\mathbf{X} = (X_1, \ldots, X_d)$ observations
- Choose (high) thresholds $\mathbf{u} = (u_1, \ldots, u_d)$
- Extreme event = AT LEAST one of the X_j exceeds its threshold u_j
- Theory: asymptotically (when $\boldsymbol{u} \to \infty$), exceedances occur according to a Poisson process and $\mathbf{Z} = \mathbf{X} - \mathbf{u} \mid \boldsymbol{X} \not\leq \boldsymbol{u}$ follows a multivariate Generalized Pareto Distribution (MGPD) with a scale parameter $\boldsymbol{\sigma}$ and a shape parameter $\boldsymbol{\gamma}$.

- Standard MGPD $\rightarrow \boldsymbol{\sigma} = \mathbf{1}$ and $\boldsymbol{\gamma} = \mathbf{0}$
  $\rightarrow$ Exponential marginals

⚠ NO parametric family of limits distributions

- Rootzén et al. [2018], Kiriliouk et al. [2019] have proposed explicit density formula for specific models
- Non parametric models are difficult to fit (lack of data)

Multivariate generalized Pareto Vectors

- Let $\boldsymbol{X}$ be a d-dimensional random vector, the vector of excesses is defined as

$$\boldsymbol{Z} = \boldsymbol{X} - \boldsymbol{u} \mid \boldsymbol{X} \not\leq \boldsymbol{u} \tag{1}$$

  where $\boldsymbol{u} \in \mathbb{R}^d$ is a vector of suitably chosen thresholds and $\not\leq$ means that at least one of the components of $\boldsymbol{X} - \boldsymbol{u}$ is positive.

- Rootzén et al. [2018] have shown that a standard MGP vector $\boldsymbol{Z}$ can be decomposed as follows

$$\boldsymbol{Z} = E + \boldsymbol{T} - \max(\boldsymbol{T}),$$

  where
  - E is a unit exponential variable ;
  - $\boldsymbol{T}$ a d-dimensional random vector
  - $\boldsymbol{T}$ and E are independent.

Non-parametric joint MGP simulation in dimension 2

Legrand et al. [2023]

- From Rootzén et al. [2018b],

$$\begin{cases} Z_1 & = E + T_1 - \max(T_1, T_2) \\ Z_2 & = E + T_2 - \max(T_1, T_2) \end{cases}$$

- Noting $\Delta = Z_1 - Z_2 = T_1 - T_2$,

$$\begin{cases} Z_1 & = E + \Delta \mathbf{1}_{\Delta < 0} \\ Z_2 & = E - \Delta \mathbf{1}_{\Delta \geq 0} \end{cases}$$

- Simulate values of Δ and E independently
- Simulate E is trivial
- Difficulty : simulate Δ

  $\Rightarrow$ **Bootstrapping on Δ**

Spectral Representation of MGP en dimenson d

Key theorem (Rootzén et al., 2018)

A standard MGP vector $\boldsymbol{Z}$ admits representation:

$$\boldsymbol{Z} = E + \boldsymbol{T} - \max_k T_k,$$

where $E \sim \text{Exp}(1)$ and $\boldsymbol{T}$ is a spectral random vector.

- Define differences:
$$\Delta_j = Z_j - \max_k Z_k.$$
- Then $\boldsymbol{Z} = E + \boldsymbol{\Delta}$.
- Sepctral boostrap uses bootstrap on observed $\boldsymbol{\Delta}$ values.

Algorithm 1: Multivariate Extreme Spectral Bootstrap

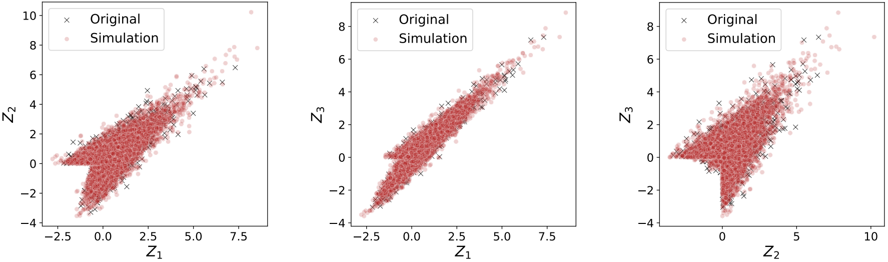
Algorithm

require n i.i.d. observations $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ of a standard MGP distribution in $\mathbb{R}^d$
output m standard MGP simulated samples $\boldsymbol{Z}_1^*, \ldots, \boldsymbol{Z}_m^*$ in $\mathbb{R}^d$
procedure
    Compute $\Delta_{i,j} \leftarrow Z_{i,j} - \max_{1 \le k \le d} Z_{i,k}$, for $1 \le i \le n$, $1 \le j \le d$
    Generate $E_1, \ldots, E_m \overset{\text{i.i.d.}}{\sim} \text{Exp}(1)$, independently of $\boldsymbol{\Delta}_1, \ldots, \boldsymbol{\Delta}_n$
    where $\boldsymbol{\Delta}_i = (\Delta_{i,1}, \ldots, \Delta_{i,d})$ for $1 \le i \le n$
    Generate m bootstrap samples $\boldsymbol{\Delta}_1^*, \ldots, \boldsymbol{\Delta}_m^*$ from $(\boldsymbol{\Delta}_1, \ldots, \boldsymbol{\Delta}_n)$
end procedure
return $\boldsymbol{Z}_\ell^* \leftarrow E_\ell + \boldsymbol{\Delta}_\ell^*$ for $1 \le \ell \le m$

- Produces synthetic MGP observations preserving tail dependence.
- Theoretical guarantees that the samples obtained with Algorithm **??** are distributed according to the same distribution as the original observations.
- The SB simulation procedure is applied to standard MGP vectors so that the procedure itself ensures the preservation of the dependence structure of the original MGP vectors.

Non-parametric joint MGP simulation

Bivariate representations of the simulated sample of size 2,000 of a MGP vector $\boldsymbol{Z} \in \mathbb{R}^3$ with $\boldsymbol{T}$



Non-parametric joint MGP simulation

Bivariate representations of the simulated sample of size 2,000 of a MGP vector $\boldsymbol{Z} \in \mathbb{R}^3$ with $\boldsymbol{T}$

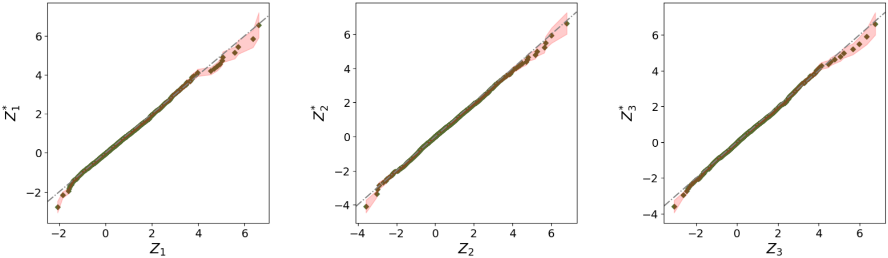


Illustration on Risk management

Illustration on Risk management

- **Risk management** = Crucial in various sectors, specific to each sector
- **Risk factor** = any variable that could result in a loss or damage
  → can be represented by random variables that quantify the magnitude of potential losses
- **Examples**
  - Climatology
    → Meteorological and marine hazards can cause significant damages
    → e.g., drought, floods, landslides
    → Risk factor = any physical quantity (wave heights, wind gusts, precipitation)
  - Finance
    → Market movements can result to substantial losses
    → Risk factor = typically, market parameters, interest rates or exchange rates
- **Tail risk**: Events with very severe magnitudes and that occurred with very low probability

Aim

For a given target risk factor X_j, accurately quantify its risk through the estimation of tail risk metrics (TRM).

Tail risk metrics

- $\mathbf{X} = (X_1, \ldots, X_d) \in \mathbb{R}^d$ vector of risk factors, with X_j of density f_j
- We consider 3 TRMs defined as
  - **Expected Shortfall** [Artzner et al., 1999] at level $\alpha \in (0,1)$

$$\mathrm{ES}_\alpha(X_j) = \mathbb{E}\left[X_j \mid X_j > \mathrm{VaR}_\alpha(X_j)\right] = \frac{1}{1-\alpha}\int_{\mathrm{VaR}_\alpha(X_j)}^{\infty} x_j f(x_j)\mathrm{d}x_j$$

    where $\mathrm{VaR}_\alpha(X_j) := \inf\{x_j \in \mathbb{R} : \mathbb{P}\left(X_j \leq x_j\right) \geq 1-\alpha\}$.
  - **Multivariate marginal ES** of X_j at level α

$$\mathrm{MMES}_\alpha(X_j; \boldsymbol{X}) = \mathbb{E}\left[X_j \mid \boldsymbol{X}_{-j} \geq \boldsymbol{v}^\alpha_{-j}\right] = \int_{\mathbb{R}} x_j f_{X_j \mid \boldsymbol{X}_{-j} \geq \boldsymbol{v}^\alpha_{-j}}(x_j)\mathrm{d}x_j,$$

    with $\boldsymbol{v}^\alpha = \mathrm{VaR}_\alpha(\boldsymbol{X}) \in \mathbb{R}^d$
  - **Dependent conditional tail expectation** of X_j at level α

$$\mathrm{DCTE}_\alpha(X_j; \boldsymbol{X}) = \mathbb{E}\left[X_j \mid \boldsymbol{X} \geq \boldsymbol{v}^\alpha\right] = \int_{v^\alpha_j}^{\infty} x_j f_{X_j \mid \boldsymbol{X} \geq \boldsymbol{v}^\alpha}(x_j)\mathrm{d}x_j,$$

Simulation framework

- Let $\boldsymbol{X} = (X_1, X_2, X_3)$ be a random vector with marginals distributed as a Student t-distribution with degrees of freedom $\nu_1 = 2$, $\nu_2 = 3$, $\nu_3 = 2.5$.
- An underlying assumption of our simulation framework is that the components of $\boldsymbol{X}$ are asymptotically dependent
- To ensure that this hypothesis is satisfied, we consider the Gumbel copula [Nelsen, 2006] to obtain dependent extremes in the upper tail

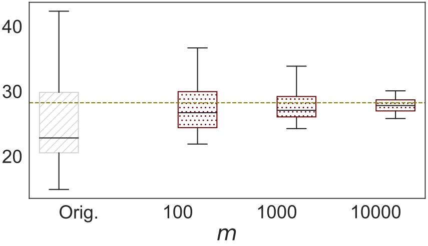
$$\mathscr{C}(\boldsymbol{y}) := \exp\left(-\left(\sum_{i=1}^{3}\left[-\log(y_i)\right]^{\theta}\right)^{1/\theta}\right),$$

where $\theta \geq 1$ is the copula parameter. The larger θ, the stronger the asymptotic dependence structure between the components of $\boldsymbol{X}$.

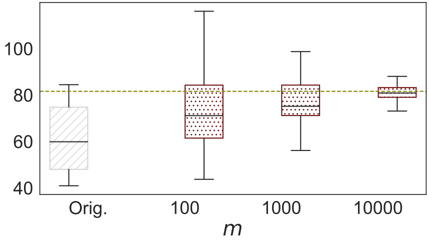
- The numerical experiments are performed on simulated data sets $\mathscr{D} \in \mathbb{R}^{1500\times 3}$,

TRMs estimation through SB simulation procedure

Estimations of the TRM on 100 original samples *(grey hatched box)* **and on 100 extended samples** *(red dotted boxes)*. **On each graph, the dashed green horizontal line corresponds to the associated theoretical value.**



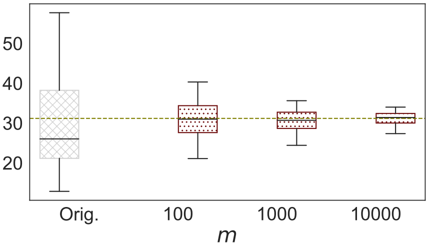
a) $ES_{0.0025}$



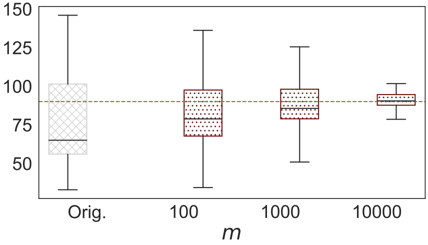
b) $ES_{0.0003}$

TRMs estimation through SB simulation procedure

Estimations of the TRM on 100 original samples *(grey hatched box)* **and on 100 extended samples** *(red dotted boxes)*. **On each graph, the dashed green horizontal line corresponds to the associated theoretical value.**



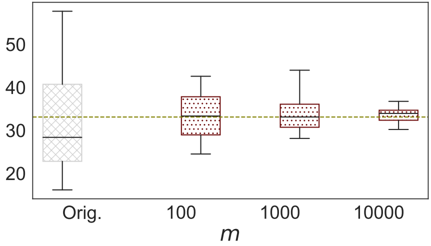
c) $MES_{0.0025}$



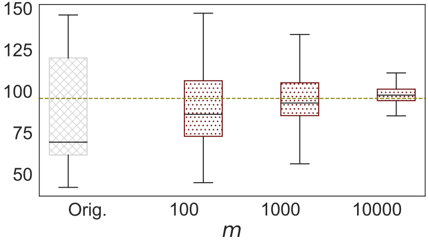
d) $MES_{0.0003}$

TRMs estimation through SB simulation procedure

Estimations of the TRM on 100 original samples *(grey hatched box)* **and on 100 extended samples** *(red dotted boxes)*. **On each graph, the dashed green horizontal line corresponds to the associated theoretical value.**



e) $DCTE_{0.0025}$



f) $DCTE_{0.0003}$

Thank you for your attention!

Reference I

P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical finance*, 9(3):203–228, 1999.

A. Kiriliouk, H. Rootzén, J. Segers, and J. L. Wadsworth. Peaks over thresholds modeling with multivariate generalized pareto distributions. *Technometrics*, 61 (1):123–135, 2019.

J. Legrand, P. Ailliot, P. Naveau, and N. Raillard. Joint stochastic simulation of extreme coastal and offshore significant wave heights. 2023.

R. B. Nelsen. *An introduction to copulas.* Springer, 2006.

H. Rootzén, J. Segers, and J. L. Wadsworth. Multivariate generalized pareto distributions: Parametrizations, representations, and properties. *Journal of Multivariate Analysis*, 165:117–131, 2018.