Statistical Learning for Multivariate and Functional Extremes

PhD Defense, Nathan Huet
Télécom Paris, November 15th

Study of Extreme Values

Why? model, predict, understand, anticipate, or manage extreme phenomena such as heavy precipitation, devastating floods, stock market crashes...

Flood in Netherlands, 1953 (photo from *Watersnoodmuseum*).

Extreme Value Theory

Focus: observations outside the mass center of the distribution, *i.e.* in the tail of the distribution

Usual assumptions on X a random element

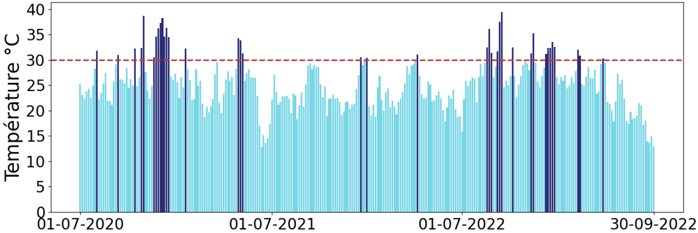
- convergence of maxima, *i.e.*

$$\lim_{n\to+\infty} \mathscr{L}\Big(\frac{\max_{i=1}^n X_i - b_n}{a_n}\Big) = \mathscr{L}(Z),$$

  with $X_i \overset{i.i.d.}{\sim} X$.
- convergence of excesses, *i.e.*

$$\lim_{t\to+\infty} \mathscr{L}(X/t \mid \|X\| \geq t) = \mathscr{L}(X_\infty).$$

Peaks-over-Threshold



Focus in my thesis: observations exceeding a high threshold

Regular Variation of $X \in \mathbb{X}$

PoT assumption

$X \in RV(\mathbb{X})$ if there exists a regularly varying function with index $\alpha > 0$ function b (*i.e.* $b(tx)/b(t) \underset{t \to +\infty}{\longrightarrow} x^\alpha$) and a nonzero Borel measure μ on $\mathbb{X} \backslash \{0\}$, finite on all Borelian sets bounded away from zero s.t.

$$\lim_{t \to +\infty} b(t)\mathbb{P}\big(X/t \in A\big) = \mu(A), \qquad (M_0\text{-convergence})$$

for all Borelian sets A bounded away from zero and s.t. $\mu(\partial A) = 0$.

Regular Variation of $X \in \mathbb{X}$

PoT assumption

$X \in RV(\mathbb{X})$ if there exists a regularly varying function with index $\alpha > 0$ function b (*i.e.* $b(tx)/b(t) \underset{t \to +\infty}{\longrightarrow} x^{\alpha}$) and a nonzero Borel measure μ on $\mathbb{X} \backslash \{0\}$, finite on all Borelian sets bounded away from zero s.t.

$$\lim_{t \to +\infty} b(t) \mathbb{P}(X/t \in A) = \mu(A), \qquad (M_0\text{-convergence})$$

for all Borelian sets A bounded away from zero and s.t. $\mu(\partial A) = 0$.

$\Leftrightarrow$ there exists a limit random variable X_∞ s.t.

$$\lim_{t \to +\infty} \mathscr{L}(X/t \mid \|X\| \geq t) = \mathscr{L}(X_\infty);$$

Regular Variation of $X \in \mathbb{X}$

PoT assumption

$X \in RV(\mathbb{X})$ if there exists a regularly varying function with index $\alpha > 0$ function b (*i.e.* $b(tx)/b(t) \underset{t\to+\infty}{\longrightarrow} x^\alpha$) and a nonzero Borel measure μ on $\mathbb{X}\backslash\{0\}$, finite on all Borelian sets bounded away from zero s.t.

$$\lim_{t\to+\infty} b(t)\mathbb{P}(X/t \in A) = \mu(A), \qquad (M_0\text{-convergence})$$

for all Borelian sets A bounded away from zero and s.t. $\mu(\partial A) = 0$.

$\Leftrightarrow$ there exists a limit random variable X_∞ s.t.

$$\lim_{t\to+\infty} \mathscr{L}(X/t \mid \|X\| \geq t) = \mathscr{L}(X_\infty);$$

$\Leftrightarrow$ there exist a limit radius R_∞ and limit angle Θ_∞ s.t.

$$\lim_{t\to+\infty} \mathscr{L}(X/\|X\|, \|X\|/t \mid \|X\| \geq t) = \mathscr{L}(\Theta_\infty, R_\infty).$$

Regular Variation of $X \in \mathbb{X}$

PoT assumption

$X \in RV(\mathbb{X})$ if there exists a regularly varying function with index $\alpha > 0$ function b (*i.e.* $b(tx)/b(t) \underset{t \to +\infty}{\longrightarrow} x^{\alpha}$) and a nonzero Borel measure μ on $\mathbb{X}\backslash\{0\}$, finite on all Borelian sets bounded away from zero s.t.

$$\lim_{t \to +\infty} b(t)\mathbb{P}(X/t \in A) = \mu(A), \qquad (M_0\text{-convergence})$$

for all Borelian sets A bounded away from zero and s.t. $\mu(\partial A) = 0$.

$\Leftrightarrow$ there exists a limit random variable X_∞ s.t.

$$\lim_{t \to +\infty} \mathscr{L}(X/t \mid \|X\| \geq t) = \mathscr{L}(X_\infty);$$

$\Leftrightarrow$ there exist a limit radius R_∞ and limit angle Θ_∞ s.t.

$$\lim_{t \to +\infty} \mathscr{L}(X/\|X\|, \|X\|/t \mid \|X\| \geq t) = \mathscr{L}(\Theta_\infty, R_\infty).$$

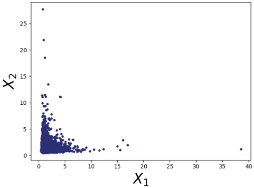
$$\boxed{X_\infty = R_\infty.\Theta_\infty \text{ and } R_\infty \perp\!\!\!\perp \Theta_\infty}$$

Regular Variation

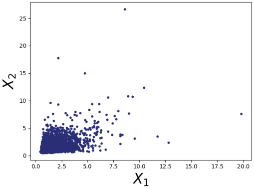
my angel angle...

$$\boxed{X_\infty = R_\infty . \Theta_\infty \text{ and } R_\infty \perp\!\!\!\perp \Theta_\infty}$$

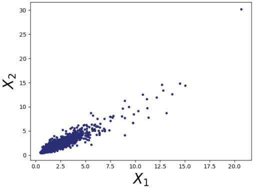
$\Rightarrow$ Θ_∞ **rules the extremal dependence structure of** X



Full independence



Partial dependence



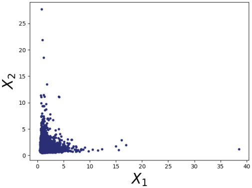
Full dependence

Regular Variation

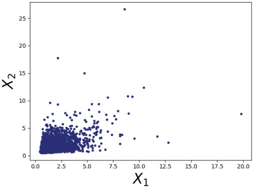
my angel angle...

$$\boxed{X_\infty = R_\infty.\Theta_\infty \text{ and } R_\infty \perp\!\!\!\perp \Theta_\infty}$$

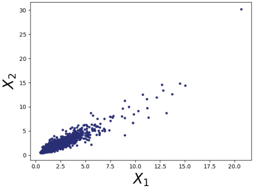
$\Rightarrow \Theta_\infty$ rules the extremal dependence structure of X



Full independence



Partial dependence

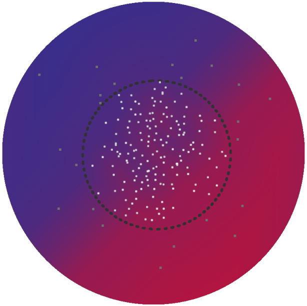


Full dependence

Focus in my thesis: How to obtain guarantees for **Extreme Values** through **Statistical Learning** methods?

Statistical learning for extremes?

◦ classic algorithms and concentration results focus on **the bulk of the distribution** (under boundedness or sub-Gaussianity assumptions)



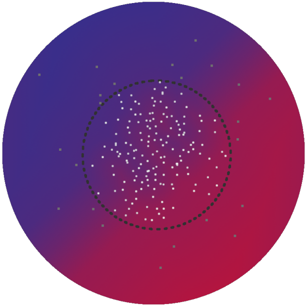
⚠ **'normal' behavior**

$\neq$

extreme behavior

Statistical learning for extremes?

∘ classic algorithms and concentration results focus on **the bulk of the distribution** (under boundedness or sub-Gaussianity assumptions)



⚠ **'normal' behavior**

$\neq$

extreme behavior

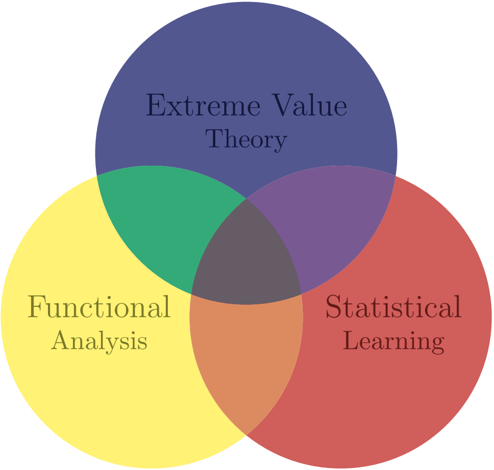
$\Rightarrow$ **classic statistical learning methods need to be adapted to well-perform in extreme regions**

Statistical learning for extremes in the literature

still fresh...

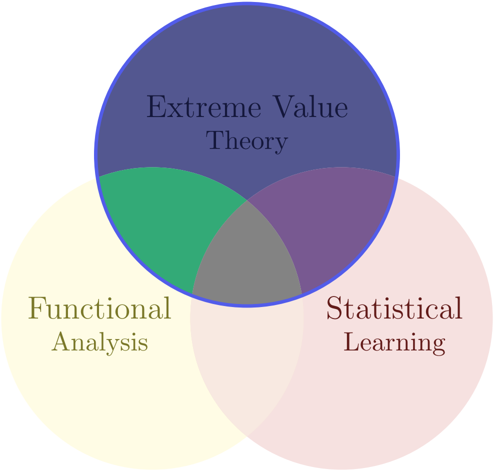
supervised learning in extreme input regions	**Classification** [Jalalzai et al.,2018] [Clémençon et al.,2023] ; **Regression**
functional data analysis	**with a RV assumption** [Kokoszka and Xiong,2018] [Kokoszka and Kulik,2023] [Kim and Kokoszka,2024] ; **of a RV element**
miscellanea	**Dimension reduction** [Goix et al.,2016] [Cooley and Thibaud,2019] [Drees and Sabourin,2021] ; **Anomaly detection** [Goix et al.,2017] [Chiapino et al.,2020] ; **Clustering** [Janßen and Wan,2020] [Vignotto et al.,2021] ; **Quantile regression** [Velthoen et al.,2023] [Gnecco et al.,2023] ; **Cross Validation** [Aghbalou et al.,2022] ; **Graphical models** [Engelke and Hitz,2020]
concentration	[Boucheron and Thomas,2012][Goix et al.,2015] [Lhaut and Segers,2021][Lhaut et al.,2022]

Outline



Extreme Value
Theory
Functional
Analysis
Statistical
Learning

Outline

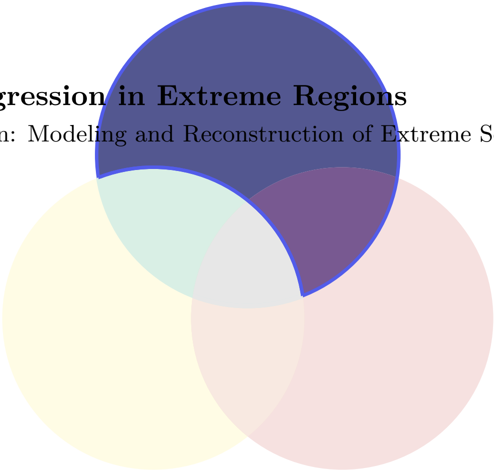


Extreme Value
Theory
Functional
Analysis
Statistical
Learning

Outline

I On Regression in Extreme Regions

Application: Modeling and Reconstruction of Extreme Sea Levels

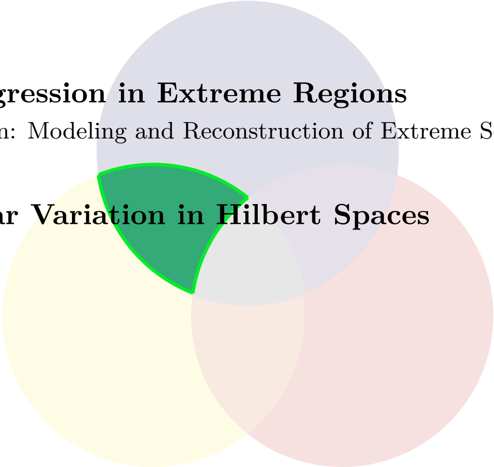


Outline

I **On Regression in Extreme Regions**

Application: Modeling and Reconstruction of Extreme Sea Levels

II **Regular Variation in Hilbert Spaces**



Outline

I **On Regression in Extreme Regions**

Application: Modeling and Reconstruction of Extreme Sea Levels

II **Regular Variation in Hilbert Spaces**

III **Principal Component Analysis of Extremes**

Regression in Extreme Regions

Goal and Motivation

Goal. for $(X, Y) \in \mathbb{R}^d \times [-M, M]$ input/output random pair, find f s.t. $f(X) \approx Y$ given that $\|X\|$ is large

Risk decomposition:

$$R(f) = \mathbb{P}(\|X\| \leq t)\mathbb{E}\Big[(Y - f(X))^2 \mid \|X\| \leq t\Big] + \mathbb{P}(\|X\| \geq t)\,\mathbb{E}\Big[(Y - f(X))^2 \mid \|X\| \geq t\Big]$$

Goal and Motivation

Goal. for $(X, Y) \in \mathbb{R}^d \times [-M, M]$ input/output random pair, find f s.t. $f(X) \approx Y$ given that $\|X\|$ is large

Risk decomposition:

$$R(f) = \mathbb{P}(\|X\| \leq t)\mathbb{E}\Big[(Y - f(X))^2 \mid \|X\| \leq t\Big] + \underbrace{\mathbb{P}(\|X\| \geq t)}_{\ll 1, \text{ if } t \gg 1}\mathbb{E}\Big[(Y - f(X))^2 \mid \|X\| \geq t\Big]$$

$\Rightarrow$ Extremes are negligible in standard Empirical Risk Minimization

Goal and Motivation

Goal. for $(X,Y) \in \mathbb{R}^d \times [-M,M]$ input/output random pair, find f s.t. $f(X) \approx Y$ given that $\|X\|$ is large

Risk decomposition:

$$R(f) = \mathbb{P}(\|X\| \leq t)\mathbb{E}\Big[(Y - f(X))^2 \mid \|X\| \leq t\Big] + \underbrace{\mathbb{P}(\|X\| \geq t)}_{\ll 1, \text{ if } t \gg 1}\mathbb{E}\Big[(Y - f(X))^2 \mid \|X\| \geq t\Big]$$

$\Rightarrow$ Extremes are negligible in standard Empirical Risk Minimization

$\Rightarrow$ focus on the minimization of the *Conditional Risk*

$$R_t(f) := \mathbb{E}\Big[(Y - f(X))^2 \mid \|X\| \geq t\Big].$$

Beyond observed data

⚠ minimizer of R_t depends on t

$\Rightarrow$ no performance guarantees in more distant regions (for $t' > t$).

Beyond observed data

⚠ minimizer of R_t depends on t

$\Rightarrow$ no performance guarantees in more distant regions (for $t' > t$).

$\Rightarrow$ focus on the minimization of the *Asymptotic Risk*

$$R_\infty(f) := \limsup_{t \to +\infty} R_t(f) = \limsup_{t \to +\infty} \mathbb{E}[(Y - f(X))^2 \mid \|X\| \geq t].$$

Beyond observed data

⚠ minimizer of R_t depends on t

$\Rightarrow$ no performance guarantees in more distant regions (for $t' > t$).

$\Rightarrow$ focus on the minimization of the *Asymptotic Risk*

$$R_\infty(f) := \limsup_{t \to +\infty} R_t(f) = \limsup_{t \to +\infty} \mathbb{E}[(Y - f(X))^2 \mid \|X\| \geq t].$$

Regular variation w.r.t. some component to save the day...

Regular Variation w.r.t. some component

Appropriate regularity/stability condition?

Reminder: $X \in RV(\mathbb{R}^d)$ if $\lim_{t\to+\infty} b(t)\mathbb{P}(X/t \in \cdot) = \mu$.

Regular Variation w.r.t. the covariates.

$$\lim_{t\to+\infty} b(t)\mathbb{P}\big(X/t \in A, Y \in C\big) = \mu(A \times C),$$

for all $C \in \mathscr{B}([-M, M])$ and $A \in \mathscr{B}(\mathbb{R}^d)$ bounded away from zero s.t. $\mu(\partial(A \times C)) = 0$.

◦ adaption of the classic assumption **to measure the extremality according to some component** (here the input variable).

Important example

Predicting a missing component in a regularly varying vector

Let $Z = (Z_1, ..., Z_{d+1}) \in RV(\mathbb{R}^{d+1})$. Under classic extreme-value assumptions on the density of Z, the pair (X, Y), defined as

$$X = (Z_1, ..., Z_d) \quad \text{and} \quad Y = Z_{d+1}/\|Z\|_p,$$

meets our assumptions.

Important example

Predicting a missing component in a regularly varying vector

Let $Z = (Z_1, ..., Z_{d+1}) \in RV(\mathbb{R}^{d+1})$. Under classic extreme-value assumptions on the density of Z, the pair (X, Y), defined as

$$X = (Z_1, ..., Z_d) \quad \text{and} \quad Y = Z_{d+1}/\|Z\|_p,$$

meets our assumptions.

$\Rightarrow$ **our framework is well-suited for predicting Z_{d+1} based on $Z_1, ..., Z_d$ given that $\|(Z_1, ..., Z_d)\|_p$ is large**

NB back to original scale through

$$Y = \frac{Z_{d+1}}{\|Z\|_p} \quad \Longleftrightarrow \quad Z_{d+1} = \frac{Y\|X\|_p}{(1 - |Y|^p)^{1/p}}.$$

Consequences

of regular variation w.r.t.X

- Existence of $(R_\infty, \Theta_\infty, Y_\infty)$ s.t.

$$\mathscr{L}(t^{-1}X, Y \mid \|X\| \geq t) \underset{t \to +\infty}{\longrightarrow} \mathscr{L}(R_\infty . \Theta_\infty, Y_\infty)$$

with $\boxed{R_\infty \perp\!\!\!\perp \Theta_\infty, Y_\infty}$

Consequences

of regular variation w.r.t.X

- Existence of $(R_\infty, \Theta_\infty, Y_\infty)$ s.t.

$$\mathscr{L}(t^{-1}X, Y \mid \|X\| \geq t) \underset{t\to+\infty}{\longrightarrow} \mathscr{L}(R_\infty.\Theta_\infty, Y_\infty)$$

with $\boxed{R_\infty \perp\!\!\!\perp \Theta_\infty, Y_\infty}$

$\rightsquigarrow$ Θ_∞ conveys all the information to predict Y_∞, *i.e.*

$$f^*_\infty(X_\infty) = \mathbb{E}[Y_\infty \mid X_\infty] = \mathbb{E}[Y_\infty \mid \Theta_\infty]$$

Propagation of this property to finite-distance extreme regions?

Propagation of the angular property

Notation: $\theta(x) = x/\|x\|$ and $\Theta = X/\|X\|$.

Proposition *(angular minimizer at finite-distance).*
With existence of densities and regularity conditions:

(*i*) **Convergence of minima:** $\inf_f R_t(f) \underset{t\to+\infty}{\longrightarrow} \inf_f R_\infty(f)$.

(*ii*) **Angular minimizer:** $\inf_f R_\infty(f) = R_\infty(f^*_\infty)$, with $f^*_\infty(x) = f^*_\infty(\theta(x))$.

Consequence: $\inf_h R_t(h \circ \theta) \underset{t\to+\infty}{\longrightarrow} \inf_f R_\infty(f)$.

$\Rightarrow$ suggests replacing the former minimization problem with

$$\min_h R_t(h \circ \theta).$$

Benefit: significant dimension reduction.

ROXANE algorithm

to handle regression in extreme regions

Input sample $\Big\{(X_1, Y_1), ..., (X_n, Y_n)\Big\}$ of input/output pairs; a class of angular regression functions $\mathscr{H}$; number $k \leq n$ of extreme observations.

Truncation keep the k 'largest' observations $\Big\{(X_{(1)}, Y_{(1)}), ..., (X_{(k)}, Y_{(k)}))\Big\}$.

Extreme ERM solve the minimization problem

$$\min_{h \in \mathscr{H}} \frac{1}{k} \sum_{i=1}^{k} \Big(Y_{(i)} - h(\theta(X_{(i)}))\Big)^2.$$

Output angular prediction function $\hat{f}$ for new examples such that $\|X\| \geq \|X_{(k)}\|$.

Ordered sample: $\Big\{(X_{(1)}, Y_{(1)}), ..., (X_{(n)}, Y_{(n)})\Big\}$ such that $\|X_{(1)}\| \geq \|X_{(2)}\| \geq ....$

$\rightsquigarrow$ *Empirical Asymptotic Risk* associated with the k largest obs.

$$\hat{R}_{n,k}(f \circ \theta) := \frac{1}{k} \sum_{i=1}^{k} \Big(Y_{(i)} - f(\theta(X_{(i)}))\Big)^2.$$

$\rightsquigarrow$ $\hat{h}_{\theta,k}$ solution of $\min_{h \in \mathscr{H}} \hat{R}_{n,k}(h \circ \theta)$ over a class $\mathscr{H}$

Risk decomposition

what can we expect?

$$R_\infty(\hat{h}_{\theta,k} \circ \theta) - \inf_f R_\infty(f) \leq \big(\inf_{h \in \mathscr{H}} R_{t_{n,k}}(h \circ \theta) - \inf_f R_{t_{n,k}}(f)\big)$$

$$+ \, 2 \sup_{h \in \mathscr{H}} |R_{t_{n,k}}(h \circ \theta) - R_\infty(h \circ \theta)| + \big(\inf_f R_{t_{n,k}}(f) - \inf_f R_\infty(f)\big)$$

$$+ \, 2 \sup_{h \in \mathscr{H}} |\hat{R}_{n,k}(h \circ \theta) - R_{t_{n,k}}(h \circ \theta)|$$

with $t_{n,k}$ the quantile s.t. $\mathbb{P}(\|X\| \geq t_{n,k}) = k/n$.

Risk decomposition

what can we expect?

$$R_\infty(\hat{h}_{\theta,k} \circ \theta) - \inf_f R_\infty(f) \le \underbrace{(\inf_{h \in \mathscr{H}} R_{t_{n,k}}(h \circ \theta) - \inf_f R_{t_{n,k}}(f))}_{\text{model bias}}$$

$$+\ 2 \sup_{h \in \mathscr{H}} |R_{t_{n,k}}(h \circ \theta) - R_\infty(h \circ \theta)| + (\inf_f R_{t_{n,k}}(f) - \inf_f R_\infty(f))$$

$$+\ 2 \sup_{h \in \mathscr{H}} |\hat{R}_{n,k}(h \circ \theta) - R_{t_{n,k}}(h \circ \theta)|$$

with $t_{n,k}$ the quantile s.t. $\mathbb{P}(\|X\| \ge t_{n,k}) = k/n$.

Risk decomposition

what can we expect?

$$R_\infty(\hat{h}_{\theta,k}\circ\theta) - \inf_f R_\infty(f) \leq \underbrace{\Big(\inf_{h\in\mathscr{H}} R_{t_{n,k}}(h\circ\theta) - \inf_f R_{t_{n,k}}(f)\Big)}_{\text{model bias}}$$

$$+\ \underbrace{2\sup_{h\in\mathscr{H}} |R_{t_{n,k}}(h\circ\theta) - R_\infty(h\circ\theta)|}_{\text{extreme bias 1}} + \underbrace{\Big(\inf_f R_{t_{n,k}}(f) - \inf_f R_\infty(f)\Big)}_{\text{extreme bias 2: } \underset{n,k\to+\infty}{\longrightarrow} 0}$$

$$+\ 2\sup_{h\in\mathscr{H}} |\hat{R}_{n,k}(h\circ\theta) - R_{t_{n,k}}(h\circ\theta)|$$

with $t_{n,k}$ the quantile s.t. $\mathbb{P}(\|X\| \geq t_{n,k}) = k/n$.

Risk decomposition

what can we expect?

$$R_\infty(\hat{h}_{\theta,k} \circ \theta) - \inf_f R_\infty(f) \leq \underbrace{(\inf_{h \in \mathscr{H}} R_{t_{n,k}}(h \circ \theta) - \inf_f R_{t_{n,k}}(f))}_{\text{model bias}}$$

$$+ \underbrace{2 \sup_{h \in \mathscr{H}} |R_{t_{n,k}}(h \circ \theta) - R_\infty(h \circ \theta)|}_{\text{extreme bias 1}} + \underbrace{(\inf_f R_{t_{n,k}}(f) - \inf_f R_\infty(f))}_{\text{extreme bias 2: } \underset{n,k \to +\infty}{\longrightarrow} 0}$$

$$+ \underbrace{2 \sup_{h \in \mathscr{H}} |\hat{R}_{n,k}(h \circ \theta) - R_{t_{n,k}}(h \circ \theta)|}_{\text{stochastic error}},$$

with $t_{n,k}$ the quantile s.t. $\mathbb{P}(\|X\| \geq t_{n,k}) = k/n$.

Uniform Statistical Guarantees

a concentration bound + a negligible bias

Assumption(*VC-class*): $\mathscr{H} \subset \mathscr{C}^0(\mathbb{S}, \mathbb{R})$ with VC-dimension $V_{\mathscr{H}} < +\infty$, uniformly bounded: $\exists M < +\infty$ s.t. $\forall h \in \mathscr{H}$, $\forall \omega \in \mathbb{S}$, $|h(\omega)| \leq M$.

Theorem(*Statistical Guarantees*)
(i) **Control of stochastic error:** With large probability:

$$\sup_{h \in \mathscr{H}} \left| \hat{R}_{n,k}(h \circ \theta) - R_{t_{n,k}}(h \circ \theta) \right| \leq C/\sqrt{k} + O(1/k).$$

(ii) **Control of extreme bias 1:** Under a mild additional assumption, we have:

$$\sup_{h \in \mathscr{H}} \left| R_{t_{n,k}}(h \circ \theta) - R_{\infty}(h \circ \theta) \right| \underset{n \to +\infty}{\longrightarrow} 0.$$

Sketch of proof of (i)

- **Intermediate risk functional**:

$$\tilde{R}_{t_{n,k}}(h \circ \theta) = \frac{1}{k} \sum_{i=1}^{n} \Big(h(\theta(X_i)) - Y_i \Big)^2 \mathbb{1}\{\|X_i\| \geq t_{n,k}\};$$

- **Sub-risk-decomposition**:

$$\sup_h \left| \hat{R}_{n,k} - R_{t_{n,k}} \right| \leq \sup_h \left| \hat{R}_{n,k} - \tilde{R}_{t_{n,k}} \right| + \sup_h \left| \tilde{R}_{t_{n,k}} - R_{t_{n,k}} \right|;$$

- **VC-bound**: $\mathbb{E}\left[\sup_h \left|\tilde{R}_{t_{n,k}} - R_{t_{n,k}}\right|\right] \leq O\Big(\sqrt{\frac{V_{\mathscr{H}}}{k}}\Big);$
- **Berstein's type inequality**:

$$\mathbb{P}\Big(\sup_h \left|\tilde{R}_{t_{n,k}} - R_{t_{n,k}}\right| - \mathbb{E}\Big[\sup_h \left|\tilde{R}_{t_{n,k}} - R_{t_{n,k}}\right|\Big] \geq \varepsilon\Big) \leq O\big(\exp(-Ck\varepsilon^2)\big);$$

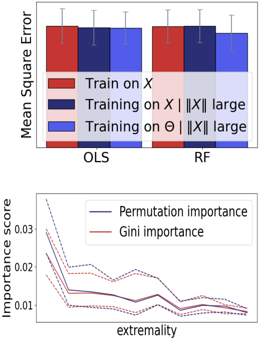
- **Concentration of the 1st term**:

$$\sup_h \left|\hat{R}_{n,k} - \tilde{R}_{t_{n,k}}\right| \leq \frac{C}{k} \sum \left(\mathbb{1}\{\|X_i\| \geq t_{n,k}\} - \mathbb{1}\{\|X_i\| \geq \|X_{(k)}\|\} \right)$$

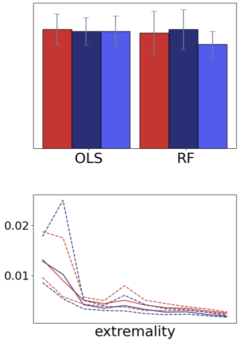
Experiments on a real dataset

the radius is irrelevant to predict extremes

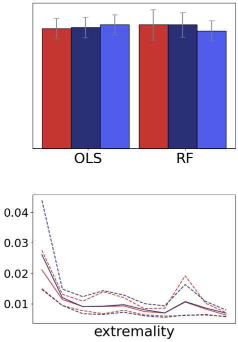
Comparison of three training models : on all observations, on extremes & on extreme angles.



Output : *Agric*



Output : *Food*



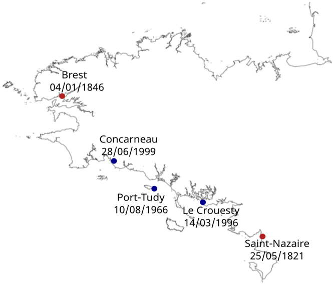
Output : *Soda*

Application:

Modeling and Reconstruction of Extreme Sea Levels

Prediction of extreme sea levels

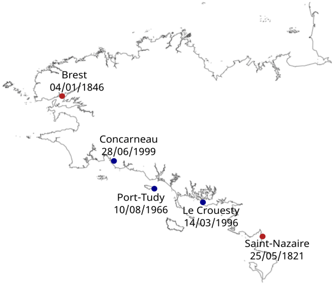
sea levels data (SHOM)



Goal: predict sea levels Y at some output tide gauges (●) given extreme sea levels $X = (X_B, X_N)$ measured at nearby input stations (●).

Prediction of extreme sea levels

sea levels data (SHOM)

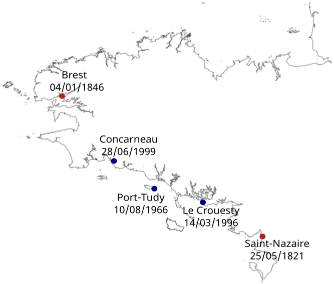


Goal: predict sea levels Y at some output tide gauges (•) given extreme sea levels $X = (X_B, X_N)$ measured at nearby input stations (•).

- **Output station**: Port-Tudy (10/08/1966 - 31/12/2023)
- **Extreme observations**: (X_B, X_N, Y) given that $\left\{X_B \geq t_B \text{ or } X_N \geq t_N\right\}$ with t_B, t_N large thresholds

Prediction of extreme sea levels

sea levels data (SHOM)



Goal: predict sea levels Y at some output tide gauges ($\bullet$) given extreme sea levels $X = (X_B, X_N)$ measured at nearby input stations ($\bullet$).

- **Output station**: Port-Tudy (10/08/1966 - 31/12/2023)
- **Extreme observations**: (X_B, X_N, Y) given that $\left\{X_B \geq t_B \text{ or } X_N \geq t_N\right\}$ with t_B, t_N large thresholds

comparison of ROXANE to a parametric method

Multivariate procedures

nonparametric vs *parametric*

ROXANE procedure:

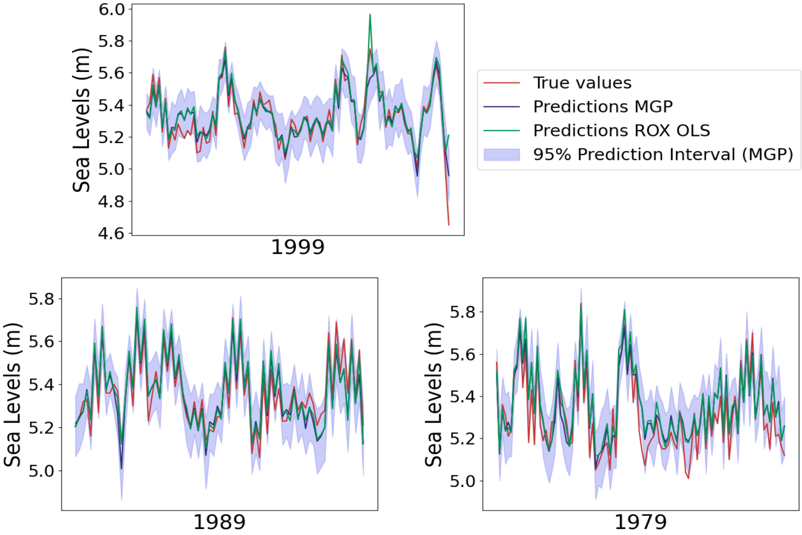
1. Pareto marginal transformation (to satisfy regular variation condition);
2. transformation as in the example "Predicting a missing component in a regularly varying vector" (to fit our framework);
3. predictions *via* predictive function estimated by RF or OLS.

MGP procedure:

1. procedure in [Kiriliouk et al., 2019] to deduce a well-fitted density;
2. conditional sampling given the values at the input stations;
3. predictions *via* Monte-Carlo average of the conditionally generated values.

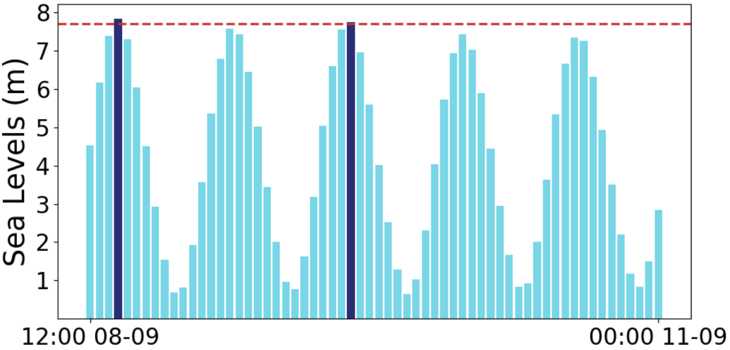
Time series prediction

of Extreme Sea Levels for 1999, 1989, and 1979



A different perspective

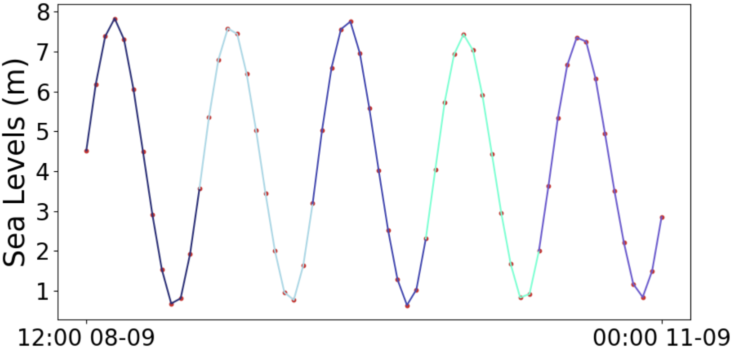
from finite to infinite dimension



in the latter study: extremes if a punctual value is large

A different perspective

from finite to infinite dimension



in the latter study: extremes if a punctual value is large

what could be done: represent data as a vector in $\mathbb{R}^{12}$ or as a **function**, with extremes defined by a large L^2-norm
$\rightsquigarrow$ better for precipitation or energy consumption analysis

Extreme FDA

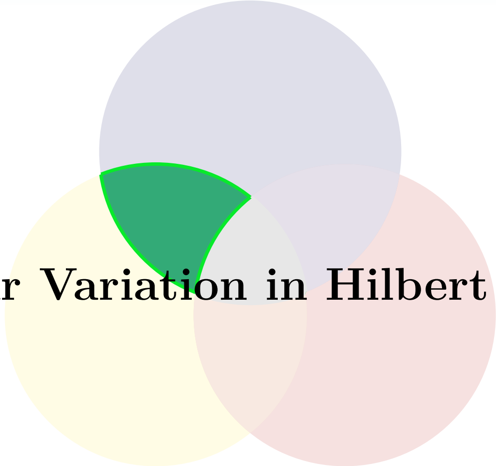
what exists?

- Vast majority of works in functional extremes concerns random objects in $\mathscr{C}[0,1], \mathbb{D}[0,1]$ or J_1;
  $\Rightarrow$ extremality measured w.r.t. the supremum norm $\|\cdot\|_\infty$

Target. Functional data with high energy, *i.e.* stochastic processes in $L^2[0,1]$ with large L^2-norm.

- Characterization of regular variation in general Polish spaces [Hult and Lindskog, 2006]
- **Existing works in Hilbert spaces**: [Kokozaska et al., 2018,2019,2022,2023,2024]

⚠ **Hilbert regular variation** as working assumption but not as study object.



Regular Variation in Hilbert Spaces

What does it mean to be regularly varying in a Hilbert space?

Focus(*more general*). Regularly varying stochastic processes with sample paths in a **separable** Hilbert space $\mathbb{H}$.

[Kim and Kokozska, 2022] prove that

$$X \in RV(\mathbb{H}) \Rightarrow \forall N \geq 1, \pi_N(X) \in RV(\mathbb{R}^N),$$

where $\pi_N(X) := (\langle X, e_1\rangle, ..., \langle X, e_N\rangle)$ is *fidi* projection on a Hilbert base $(e_i)_{i\geq 1}$

What does it mean to be regularly varying in a Hilbert space?

Focus(*more general*). Regularly varying stochastic processes with sample paths in a **separable** Hilbert space $\mathbb{H}$.

[Kim and Kokozska, 2022] prove that

$$X \in RV(\mathbb{H}) \Rightarrow \forall N \geq 1, \pi_N(X) \in RV(\mathbb{R}^N),$$

where $\pi_N(X) := (\langle X, e_1 \rangle, ..., \langle X, e_N \rangle)$ is *fidi* projection on a Hilbert base $(e_i)_{i \geq 1}$

What about the converse? Does finite-dimensional characterizations of regular variation in $\mathbb{H}$ exist?

fidi characterization of regular variation

⚠ regular variation of the *fidi* projections **not enough** for global regular variation, **relative compactness** is needed

fidi characterization of regular variation

⚠ regular variation of the *fidi* projections **not enough** for global regular variation, **relative compactness** is needed

***fidi* characterizations.** Equivalent assertions:

1. $X \in RV(\mathbb{H})$ such that $\mu_t(\cdot) := b(t)\mathbb{P}(X/t \in \cdot) \xrightarrow{M_0} \mu(\cdot)$.
2. the family (μ_t) is relatively compact w.r.t. the M_0-topology and for all $N \geq 1$, $\pi_N(X) \in RV(\mathbb{R}^N)$ such that $b(t)\mathbb{P}(\pi_N(X)/t \in \cdot) \overset{M_0}{\to} \mu_N(\cdot)$.

In particular, $\mu_N = \mu \circ \pi_N$.

⚠ relative compactness could be tricky to verify...

fidi polar characterization of regular variation

***fidi* polar characterizations.** Equivalent assertions:

1. $X \in RV(\mathbb{H})$ such that $\|X\| \in RV$ and $\mathscr{L}(\Theta_t) := \mathscr{L}(\Theta | \|X\| \geq t) \to \mathscr{L}(\Theta_\infty)$.
2. $\|X\| \in RV$ and for all $h \in \mathbb{H}$, $\langle \Theta_t, h \rangle \xrightarrow{w} \langle \Theta_\infty, h \rangle$.
3. $\|X\| \in RV$ and for all $N \geq 1$, $\pi_N(\Theta_t) \xrightarrow{w} \pi_N(\Theta_\infty)$.

fidi polar characterization of regular variation

***fidi* polar characterizations.** Equivalent assertions:

1. $X \in RV(\mathbb{H})$ such that $\|X\| \in RV$ and $\mathscr{L}(\Theta_t) := \mathscr{L}(\Theta | \|X\| \geq t) \to \mathscr{L}(\Theta_\infty)$.
2. $\|X\| \in RV$ and for all $h \in \mathbb{H}$, $\langle \Theta_t, h \rangle \xrightarrow{w} \langle \Theta_\infty, h \rangle$.
3. $\|X\| \in RV$ and for all $N \geq 1$, $\pi_N(\Theta_t) \xrightarrow{w} \pi_N(\Theta_\infty)$.

Hilbert regular variation translated in terms of finite-dimensional weak convergences.

Principal Component Analysis of Extremes

Adaptation of [Drees and Sabourin,2021]?

How to perform efficient **dimension reduction** tailored for **functional extremes**?

[Drees and Sabourin,2021]: results (with statistical guarantees) on the convergence of PCA elements of a regularly varying vector towards its limit vector **in finite-dimension** through an analysis of the **reconstruction error**

Adaptation of [Drees and Sabourin,2021]?

How to perform efficient **dimension reduction** tailored for **functional extremes**?

[Drees and Sabourin,2021]: results (with statistical guarantees) on the convergence of PCA elements of a regularly varying vector towards its limit vector **in finite-dimension** through an analysis of the **reconstruction error**

⚠ compactness of the unit sphere used;

⚠ bounds depend on the dimension;

Adaptation of [Drees and Sabourin,2021]?

How to perform efficient **dimension reduction** tailored for **functional extremes**?

[Drees and Sabourin,2021]: results (with statistical guarantees) on the convergence of PCA elements of a regularly varying vector towards its limit vector **in finite-dimension** through an analysis of the **reconstruction error**

⚠ compactness of the unit sphere used;

⚠ bounds depend on the dimension;

not suited for infinite dimension!

Our focus: Covariance Operators

$$\boxed{X \in RV(\mathbb{H}), \mathscr{L}(\Theta_t) := \mathscr{L}(\Theta | \|X\| \geq t) \to \mathscr{L}(\Theta_\infty)}$$

Rank one operator: $\forall h_1, h_2 \in \mathbb{H}, h_1 \otimes h_2(\cdot) = \langle h_1, \cdot \rangle h_2$.

Covariance operator of Θ_t: $C_t := \mathbb{E}[\Theta \otimes \Theta \mid \|X\| \geq t]$ with $\|C_t\|_{HS}^2 := \sum_{i=1}^{+\infty} \|C_t e_i\|^2 < \infty$.

PCA of Θ_t: with proba. one, $\Theta_t = \sum_{i=1}^{+\infty} \langle \Theta_t, \varphi_{i,t} \rangle \varphi_{i,t}$ where $\varphi_{i,t}$ eigenfunction of C_t and $\mathbb{E}[\langle \Theta_t, \varphi_{i,t} \rangle^2] = \lambda_{i,t}$ eigenvalue of C_t.

$$\boxed{\textbf{Convergence of } C_t \textbf{ and } (\varphi_{i,t}, \lambda_{i,t}) \textbf{ as } t \to +\infty\textbf{?}}$$

Candidates:
- $C_\infty = \mathbb{E}[\Theta_\infty \otimes \Theta_\infty]$;
- $(\varphi_{i,\infty}, \lambda_{i,\infty})$ ordered eigenelements of C_∞.

Deterministic convergences

is the limit of PCA the PCA of the limit?

Convergence of covariance operator.

$$\|C_t - C_\infty\|_{HS} \to 0, \quad \text{as } t \to +\infty.$$

positive eigengap: $\exists p \geq 1, \lambda_{p,\infty} > \lambda_{p+1,\infty}$
$\rightsquigarrow V_{p,\infty} = \text{span}(\varphi_{1,\infty}, ..., \varphi_{p,\infty})$ unique

Corollary(*convergence of eigenspaces*).

$$\rho(V_{p,t}, V_{p,\infty}) \to 0, \quad \text{as } t \to +\infty,$$

where $V_{p,t} = \text{span}(\varphi_{1,t}, ..., \varphi_{p,t})$ and $\rho(A, B) = \|\Pi_A - \Pi_B\|_{op}$ and Π_A orthogonal projection onto A.

Sketch of proof

Convergence of covariance operator:

- $\Theta_t \xrightarrow{w} \Theta_\infty \Rightarrow \Theta_t \otimes \Theta_t \xrightarrow{w} \Theta_\infty \otimes \Theta_\infty$ $(h \mapsto h \otimes h \in \mathscr{C}(\mathbb{H}, HS(\mathbb{H})))$;

$$\begin{aligned} \|C_t - C_\infty\|_{HS} &= \|\mathbb{E}[\Theta_t \otimes \Theta_t] - \mathbb{E}[\Theta_\infty \otimes \Theta_\infty]\|_{HS} && \text{Skorokhod} \\ &= \|\mathbb{E}[Y_t] - \mathbb{E}[Y_\infty]\|_{HS} && \text{Jensen} \\ &\leq \mathbb{E}\|Y_t - Y_\infty\|_{HS} && \text{DCT} \\ &\underset{t \to +\infty}{\longrightarrow} 0 \end{aligned}$$

- **Skorokhod's Th.**: $\exists Y_t, Y_\infty, Y_t \overset{d}{=} \Theta_t \otimes \Theta_t,\ Y_\infty \overset{d}{=} \Theta_\infty \otimes \Theta_\infty, Y_t \xrightarrow{as} Y_\infty$;
- **Jensen's inequality**: $\|C_t - C_\infty\|_{HS} \leq \mathbb{E}\|Y_t - Y_\infty\|_{HS}$;
- **DCT**: $Y_t \xrightarrow{as} Y_\infty + \|Y_t - Y_\infty\|_{HS} \leq 2 \Rightarrow \mathbb{E}\|Y_t - Y_\infty\|_{HS} \to 0$.

Sketch of proof

Convergence of covariance operator:

- $\Theta_t \overset{w}{\to} \Theta_\infty \Rightarrow \Theta_t \otimes \Theta_t \overset{w}{\to} \Theta_\infty \otimes \Theta_\infty$ $(h \mapsto h \otimes h \in \mathscr{C}(\mathbb{H}, HS(\mathbb{H})))$;

$$\begin{aligned}
\|C_t - C_\infty\|_{HS} &= \|\mathbb{E}[\Theta_t \otimes \Theta_t] - \mathbb{E}[\Theta_\infty \otimes \Theta_\infty]\|_{HS} & & \text{Skorokhod} \\
&= \|\mathbb{E}[Y_t] - \mathbb{E}[Y_\infty]\|_{HS} & & \text{Jensen} \\
&\leq \mathbb{E}\|Y_t - Y_\infty\|_{HS} & & \text{DCT} \\
&\underset{t \to +\infty}{\longrightarrow} 0
\end{aligned}$$

- **Skorokhod's Th.**: $\exists Y_t, Y_\infty, Y_t \overset{d}{=} \Theta_t \otimes \Theta_t,\ Y_\infty \overset{d}{=} \Theta_\infty \otimes \Theta_\infty, Y_t \overset{as}{\to} Y_\infty$;
- **Jensen's inequality**: $\|C_t - C_\infty\|_{HS} \leq \mathbb{E}\|Y_t - Y_\infty\|_{HS}$;
- **DCT**: $Y_t \overset{as}{\to} Y_\infty + \|Y_t - Y_\infty\|_{HS} \leq 2 \Rightarrow \mathbb{E}\|Y_t - Y_\infty\|_{HS} \to 0$.

Corollary: Th. 3 in [Zwald and Blanchard, 2005] gives

$$\rho(V_t^p, V_\infty^p) \leq \frac{\|C_t - C_\infty\|_{HS}}{\gamma_\infty^p}, \text{ with } \gamma_\infty^p = (\lambda_\infty^p - \lambda_\infty^{p+1})/2.$$

Statistical guarantees

a concentration bound

concentration of the **"empirical" covariance operator**

$$\hat{C}_k = \frac{1}{k}\sum_{i=1}^{k} \Theta_{(i)} \otimes \Theta_{(i)}$$

around $C_{t_{n,k}}$? and of the **"empirical" p-dim. eigenspace** $\hat{V}_k^p$ around $V_{t_{n,k}}^p$?

Statistical guarantees

a concentration bound

concentration of the **"empirical" covariance operator**

$$\hat{C}_k = \frac{1}{k} \sum_{i=1}^{k} \Theta_{(i)} \otimes \Theta_{(i)}$$

around $C_{t_{n,k}}$? and of the **"empirical" p-dim. eigenspace** $\hat{V}_k^p$ around $V_{t_{n,k}}^p$?

Concentration of the empirical covariance operator. with probability at least $1-\delta$,

$$\|\hat{C}_k - C_{t_{n,k}}\|_{HS} \leq C/\sqrt{k} + O(1/k),$$

and in particular, with $\gamma_{t_{n,k}}^p = (\lambda_{t_{n,k}}^p - \lambda_{t_{n,k}}^{p+1})/2$,

$$\rho(\hat{V}_k^p, V_{t_{n,k}}^p) \leq C/(\gamma_{t_{n,k}}^p \sqrt{k}) + O(1/k).$$

Conclusion and Perspectives

Conclusion

- development of a regression framework to handle regression in regions where the norm of the input is large;
- novel multivariate methods to tackle the problem of extreme sea level prediction;
- finite-dimensional characterization of the infinite-dimensional Hilbert regular variation;
- probabilistic and statistical guarantees for PCA of extremes in a general settings.

Contributions at the intersection of statistical learning and extreme value theory.

Perspectives

Regression in Extreme Regions

- weakening of the regular variation and density assumptions;
- statistical guarantees on the empirical marginal standardization in the ROXANE algorithm.

Modeling and Reconstruction of Extreme Sea Levels

- adjust the model for smallest extreme (by including meteorological variables?);
- analysis of our method to improve the inference of long return period.

Hilbertian extremes and PCA

- application to functional anomaly detection;
- generalization of the results to other convenient functional basis (wavelets?).

Contributions

- S. Clémençon, N.H., A. Sabourin (2024) *Regular variation in Hilbert spaces and principal component analysis for functional extremes*, Stochastic Processes and their Applications;
- N.H., S. Clémençon, A. Sabourin (2023) *On Regression in Extreme Regions*, submitted;
- N.H., P. Naveau, A. Sabourin (2024) *Multi-site modeling and reconstruction of past extreme sea levels along the Atlantic French coast*, in process.

Bibliography I

- A. Aghbalou, P. Bertail, F. Portier and A. Sabourin (2022), *Cross validation for rare events*, arXiv:2202.00488.
- S. Boucheron and M. Thomas (2012), *Concentration inequalities for order statistics.*
- M. Chiapino, S. Clémençon, V. Feuillard and A. Sabourin (2020). *A multivariate extreme value theory approach to anomaly clustering and visualization*, Computational Statistics.
- S. Clémençon, H. Jalalzai, S. Lhaut, A. Sabourin, and J. Segers (2023) *Concentration bounds for the empirical angular measure with statistical learning applications*, Bernoulli.
- D. Cooley and E. Thibaud (2019) *Decompositions of dependence for high-dimensional extremes*, Biometrika.
- C. Dombry and M. Ribatet (2015) *Functional regular variations, Pareto processes and peaks over threshold*, Statistics and its Interface.

Bibliography II

- H. Drees and A. Sabourin (2021) *Principal component analysis for multivariate extremes.*
- S. Engelke and A. S. Hitz (2020), *Graphical models for extremes*, Journal of the Royal Statistical Society Series B: Statistical Methodology.
- H. Hult and F. Lindskog (2006) *Regular Variation for Measures on Metric Spaces*, Publications de l'Institut Mathématique.
- N. Gnecco, J. Peters, S. Engelke and N. Pfister (2023), *Boosted control functions*, arXiv:2310.05805.
- N. Goix, A. Sabourin and S.Clémençon (2015) *Learning the dependence structure of rare events: a non-asymptotic study*, Conference on learning theory.
- N. Goix, A. Sabourin and S. Clémençon (2016) *Sparse representation of multivariate extremes with applications to anomaly ranking*, In Artificial intelligence and statistics.

Bibliography III

- N. Goix, A. Sabourin and S. Clémençon (2017) *Sparsity in multivariate extremes with applications to anomaly detection*, arXiv:1507.05899.
- H. Jalalzai, S. Clémençon and A. Sabourin, (2018) *On binary classification in extreme regions*, Advances in Neural Information Processing Systems.
- A. Janßen and P. Wan (2020), *K-means clustering of extremes*.
- M. Kim and P. Kokoszka (2022) *Extremal dependence measure for functional data*, Journal of Multivariate Analysis.
- A. Kiriliouk, H. Rootzén, J. Segers and J. L. Wadsworth (2019) *Peaks Over Thresholds Modeling With Multivariate Generalized Pareto Distributions*, Technometrics.
- P. Kokoszka and R. Kulik (2023) *Principal component analysis of infinite variance functional data*, Journal of Multivariate Analysis.

Bibliography IV

- P. Kokoszka, S. Stoev, Q. Xiong (2019) *Principal Component analysis of regularly varying functions*, Bernoulli.
- P. Kokoszka and Q. Xiong (2018) *Extremes of projections of functional time series on data-driven basis systems* , Extremes.
- S. Clémençon, H. Jalalzai, S. Lhaut, A. Sabourin and J. Segers (2023), *Concentration bounds for the empirical angular measure with statistical learning applications*, Bernoulli.
- S. Lhaut, A. Sabourin and J. Segers (2022), *Uniform concentration bounds for frequencies of rare events*, Statistics & Probability Letters.
- C. McDiarmid (1998), *Concentration*, Probabilistic methods for algorithmic discrete mathematics.

Bibliography V

- H. Rootzén, J. Segers and J. L. Wadsworth (2018) *Multivariate generalized Pareto distributions: Parametrizations, representations, and properties*, Journal of Multivariate Analysis.
- S. I. Resnick (1987) *Extreme Values, Regular Variation and Point Processes*, Springer-Verlag New York.

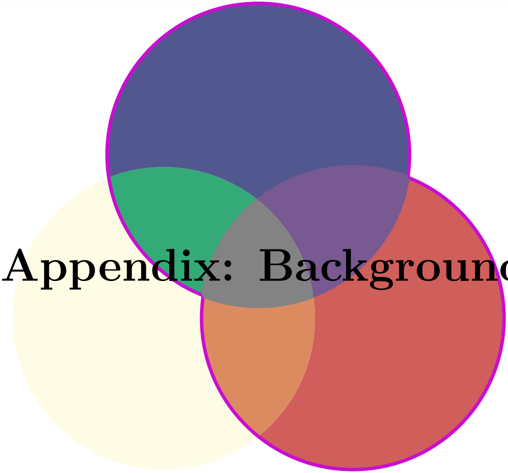
- H. Rootzén, J. Segers and J. L. Wadsworth (2018) *Multivariate generalized Pareto distributions: Parametrizations, representations, and properties*, Journal of Multivariate Analysis.

- V. N. Vapnik, and A. Y. Chervonenkis (2015), *On the uniform convergence of relative frequencies of events to their probabilities*, Measures of complexity: festschrift for Alexey Chervonenkis.
- J. Velthoen, C. Dombry, J. J. Cai and S. Engelke (2023), *Gradient boosting for extreme quantile regression*, Extremes.

Bibliography VI

- E. Vignotto, S. Engelke and J. Zscheischler (2021), *Clustering bivariate dependencies of compound precipitation and wind extremes over Great Britain and Ireland*, Weather and Climate Extremes.
- L. Zwald and G. Blanchard (2005) *On the convergence of eigenspaces in kernel principal component analysis*, Advances in neural information processing systems.

Appendix



Appendix: Background

VC-dimension

Goal: obtain nonasymptotic statistical guarantees on the deviation of an empirical measure $\hat{\nu}_n$ from the true measure ν over a class $\mathscr{A}$

$\rightsquigarrow$: control the deviation between risk functionals $|\hat{R} - R|$ over a class function

assumption on $\mathscr{A}$: **Vapnik-Chervonenkis** (VC) dimension $V_{\mathscr{A}}$ is finite ("$\mathscr{A}$ could be infinite but not too much")

$\Rightarrow$ **VC-inequality** [Vapnik and Chervonenkis,1971]: with probability at least $1 - \delta$

$$\sup_{A \in \mathscr{A}} |\nu_n(A) - \nu(A)| \leq 2\sqrt{\frac{2}{n}\Big(V_{\mathscr{A}} \log(2n+1) + \log(4/\delta)\Big)}.$$

Goal: obtain nonasymptotic statistical guarantees on the deviation of an empirical measure $\hat{\nu}_n$ from the true measure ν over a class $\mathscr{A}$
$\rightsquigarrow$: control the deviation between risk functionals $|\hat{R} - R|$ over a class function
assumption on $\mathscr{A}$: **Vapnik-Chervonenkis** (VC) dimension $V_{\mathscr{A}}$ is finite ("$\mathscr{A}$ could be infinite but not too much")
$\Rightarrow$ **VC-inequality** [Vapnik and Chervonenkis,1971]: with probability at least $1 - \delta$

$$\sup_{A \in \mathscr{A}} |\nu_n(A) - \nu(A)| \leq 2\sqrt{\frac{2}{n}\Big(V_{\mathscr{A}} \log(2n+1) + \log(4/\delta)\Big)}.$$

⚠ not adapted to extremes, *i.e.* if for all $A \in \mathscr{A}$ $\nu(A) \leq p \ll 1$

a better inequality for extremes [Anthony and Shawe-Taylor,1993]: with probability at least $1-\delta$

$$\sup_{A\in\mathscr{A}} \frac{\nu_n(A)-\nu(A)}{\sqrt{\nu(A)}} \le 2\sqrt{\frac{1}{n}\Big(V_{\mathscr{A}}\log(2n+1)+\log(4/\delta)\Big)}.$$

$\Rightarrow$ with $\nu(A)\le p$ for all $A\in\mathscr{A}$

$$\sup_{A\in\mathscr{A}} |\nu_n(A)-\nu(A)| \le 2\sqrt{\frac{p}{n}\Big(V_{\mathscr{A}}\log(2n+1)+\log(4/\delta)\Big)},$$

a better inequality for extremes [Anthony and Shawe-Taylor,1993]: with probability at least $1-\delta$

$$\sup_{A\in\mathscr{A}} \frac{\nu_n(A)-\nu(A)}{\sqrt{\nu(A)}} \leq 2\sqrt{\frac{1}{n}\Big(V_{\mathscr{A}}\log(2n+1)+\log(4/\delta)\Big)}.$$

$\Rightarrow$ with $\nu(A)\leq p$ for all $A\in\mathscr{A}$

$$\sup_{A\in\mathscr{A}} |\nu_n(A)-\nu(A)| \leq 2\sqrt{\frac{p}{n}\Big(V_{\mathscr{A}}\log(\mathbf{2n+1})+\log(4/\delta)\Big)},$$

[Lhaut et al.,2022]: $\nu(A)\leq p$ for all $A\in\mathscr{A}$, with probability at least $1-\delta$

$$\sup_{A\in\mathscr{A}} |\nu_n(A)-\nu(A)| \leq \sqrt{\frac{2p}{n}}\sqrt{\log(2)+V_{\mathscr{A}}\log(2np+1)} + \sqrt{\frac{2p}{n}}\left(\sqrt{2\log(1/\delta)}+\frac{\sqrt{2}}{2}\right)+\frac{2\log(1/\delta)}{3n}.$$

Concentration inequality for extremes

Theorem 3.8 in [McDiarmid,1998]

Lemma(*Bernstein's type inequality*) Let $\mathbf{X} = (X_1, \ldots, X_n)$ with X_i taking their values in a set $\mathscr{X}$ and let f be a real-valued function defined on $\mathscr{X}^n$. Let $Z = f(X_1, \ldots, X_n)$. Consider the positive deviation functions, defined for $1 \le i \le n$ and for $x_{1:i} := (x_1, \ldots, x_i) \in \mathscr{X}^i$

$$g_i(x_{1:i}) = \mathbb{E}[Z | X_{1:i} = x_{1:i}] - \mathbb{E}[Z | X_{1:i-1} = x_{1:i-1}].$$

Denote b the maximum deviation defined by

$$b := \max_{1 \le i \le n} \sup_{x_{1:i} \in \mathscr{X}^i} g_i(x_{1:i}).$$

Denote $\hat{v}$ the supremum of the sum of conditional variances defined by

$$\hat{v} := \sup_{(x_1, \ldots, x_n) \in \mathscr{X}^n} \sum_{i=1}^{n} \sigma_i^2(f(x_1, \ldots, x_n)),$$

where $\sigma_i^2(f(x_1, \ldots, x_n)) := \mathbf{Var}_{X_i' \sim X_i}[g_i(x_{1:i-1}, X_i')]$. If b and $\hat{v}$ are both finite, then

$$\mathbb{P}\{f(\mathbf{X}) - \mathbb{E}[f(\mathbf{X})] \ge t\} \le \exp\left(\frac{-t^2}{2(\hat{v} + bt/3)}\right),$$

for $u \ge 0$.

Appendix: Regression in Extreme Regions

Concrete examples

Noise model with heavy-tailed random design.

$$Y = g(X, \varepsilon),$$

where $X \in RV(\mathbb{R}^d)$ with $X \perp\!\!\!\perp \varepsilon$ and g bounded and continuous s.t. there exists g_θ satisfying for all z

$$\lim_{t \to +\infty} \sup_{\|x\| \geq t} |g(x, z) - g_\theta(x/\|x\|, z)| = 0.$$

Then (X, Y) are RV w.r.t. the X component.

Concrete examples

Noise model with heavy-tailed random design.

$$Y = g(X, \varepsilon),$$

where $X \in RV(\mathbb{R}^d)$ with $X \perp\!\!\!\perp \varepsilon$ and g bounded and continuous s.t. there exists g_θ satisfying for all z

$$\lim_{t \to +\infty} \sup_{\|x\| \geq t} |g(x, z) - g_\theta(x/\|x\|, z)| = 0.$$

Then (X, Y) are RV w.r.t. the X component.

with ε bounded:

- Additive model: $Y = f(X) + \varepsilon$.
- Multiplicative model: $Y = \varepsilon f(X)$.

Important example (details)

Predicting a missing component in a regularly varying vector

Let $Z = (Z_1, ..., Z_{d+1}) \in RV_\alpha(\mathbb{R}^{d+1})$ and a L^p-norm $\| \cdot \|_p$. Let $b(t) = \mathbb{P}(\|Z\|_p \geq t)^{-1}$.

Assume that Z have a continuous density p and that there exists a positive function q s.t. for all $z \neq 0$ s.t.

$$t^{d+1} b(t) p(tz) - q(z) \underset{t \to +\infty}{\longrightarrow} 0,$$

and

$$\sup_{\omega \in \mathbb{S}_{d+1}} |t^{d+1} b(t) p(t\omega) - q(\omega)| \underset{t \to +\infty}{\longrightarrow} 0.$$

Assume finally that $\min_{\omega \in \mathbb{S}_{d+1}} q(\omega) > 0$.

Then the pair $(X, Y) = \Big((Z_1, ..., Z_d), Z_{d+1}/\|Z\|\Big)$ satisfies all the necessary assumptions.

Appendix: Modeling and Reconstruction of Extreme Sea Levels

Sea levels data (SHOM)

- **Data**: maximum sea levels over each tide
- **Training set**: most recent observations (>01-01-2000);
  Test set: oldest observations (<01-01-2000);

scatterplots bonne couleur + legende

Margins modeling and threshold selection

common to both procedures

◦ margins are modeled by Extended Generalized Pareto distribution with cdf

$$F_{\sigma,\xi,\kappa}(x) = \left(1 - \left(1 + \frac{\xi x}{\sigma}\right)^{-1/\xi}\right)^{\kappa}$$

figure density fit sans le threshold, puis avec threshold

$\rightsquigarrow$ Generalized Pareto behavior in the right-tail

Margins modeling and threshold selection

common to both procedures

◦ margins are modeled by Extended Generalized Pareto distribution with cdf

$$F_{\sigma,\xi,\kappa}(x) = \left(1 - \left(1 + \frac{\xi x}{\sigma}\right)^{-1/\xi}\right)^{\kappa}$$

figure density fit sans le threshold, puis avec threshold

$\rightsquigarrow$ Generalized Pareto behavior in the right-tail

selected threshold lowest point above which the fitted density is convex, *i.e.* largest zero of $d^3F_{\sigma,\xi,\kappa}(x)/dx^3$.

Marginal modeling and threshold selection

Input Training dataset $\{X_1, \ldots, X_n\}$ with $X_i \in \mathbb{R}$ for all $1 \leq i \leq n$.

Marginal Fitting Fit an EGP distribution to the dataset to obtain a triplet of estimated parameters $(\sigma, \xi, \kappa) \in [0, +\infty[\times \mathbb{R} \times [0, +\infty[$.

Threshold Computation Compute the threshold t according to the EGP estimated terms

$$t = \frac{\sigma}{\xi}\left(\left(\frac{4\xi^2+3\kappa\xi+3\kappa+3\xi-1-\sqrt{(4\xi^2+3\kappa\xi+3\kappa+3\xi-1)^2-4(\kappa^2+2\xi^2+3\kappa\xi)(2\xi^2+3\xi+1)}}{2(\kappa^2+2\xi^2+3\kappa\xi)}\right)^{-\xi} - 1\right).$$

Output estimated marginal EGP parameters (σ, ξ, κ) and a threshold t.

ROXANE algorithm

Input Extreme training dataset $\mathscr{D}_{ext} = \{\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_k\}$ with $\boldsymbol{Z}_i = (\boldsymbol{X}_i, Y_i)^T \in \mathbb{R}^d \times \mathbb{R}$ input/target pair; fitted EGP parameters $(\boldsymbol{\sigma}, \boldsymbol{\xi}, \boldsymbol{\kappa}) = ((\boldsymbol{\sigma}_X, \boldsymbol{\xi}_X, \boldsymbol{\kappa}_X)^T, (\sigma_Y, \xi_Y, \kappa_Y)^T)^T \in \mathbb{R}^{3d+3}$; a multivariate threshold $\boldsymbol{t} = (\boldsymbol{t}_X, t_Y)^T$; a L^r-norm $\|\cdot\|_r$ for $r \in [1, +\infty[$; a class Γ of angular predictive function $\gamma : \mathbb{S}^{d-1} \to [0, 1]$.

Marginal Pareto Transformation Apply the Pareto transformation to each margin of the observations in $\mathscr{D}_{ext}$

$$\tilde{\boldsymbol{Z}}_i = (\tilde{\boldsymbol{X}}_i, \tilde{Y}_i) = p_{\boldsymbol{\sigma},\boldsymbol{\xi},\boldsymbol{\kappa}}(\boldsymbol{Z}_i) = \frac{1}{1 - F_{\boldsymbol{\sigma},\boldsymbol{\xi},\boldsymbol{\kappa}}(\boldsymbol{Z}_i)}, \quad \text{for all } i \in \{1, \ldots, k\}.$$

Angular rescaling Form the angular components of the Pareto scale observations,

$$\boldsymbol{\Theta}_{X,i} = \tilde{\boldsymbol{X}}_i / \|\tilde{\boldsymbol{X}}_i\|_r,$$

$$\Theta_{Y,i} = \tilde{Y}_i / \|\tilde{\boldsymbol{Z}}_i\|_r.$$

Empirical quadratic risk minimization based on the extreme transform dataset, solve the optimization problem

$$\min_{h \in \mathscr{H}} \sum_{i=1}^{k} \Big(\Theta_{Y,i} - h(\boldsymbol{\Theta}_{X,i})\Big)^2. \tag{1}$$

Output Solution $\hat{h}$ to problem (1) and a predictive function $\hat{g}$ given by

$$\hat{g} : \boldsymbol{x} \in \mathbb{R}^d \mapsto p^{-1}_{\sigma_Y, \xi_Y, \kappa_Y}\left(\left(\frac{\hat{h}(p_{\boldsymbol{\sigma}_X,\boldsymbol{\xi}_X,\boldsymbol{\kappa}_X}(\boldsymbol{x}) / \|p_{\boldsymbol{\sigma}_X,\boldsymbol{\xi}_X,\boldsymbol{\kappa}_X}(\boldsymbol{x})\|_r) \|p_{\boldsymbol{\sigma}_X,\boldsymbol{\xi}_X,\boldsymbol{\kappa}_X}(\boldsymbol{x})\|_r}{1 - \hat{h}(p_{\boldsymbol{\sigma}_X,\boldsymbol{\xi}_X,\boldsymbol{\kappa}_X}(\boldsymbol{x}) / \|p_{\boldsymbol{\sigma}_X,\boldsymbol{\xi}_X,\boldsymbol{\kappa}_X}(\boldsymbol{x})\|_r)^r}\right)^{1/r}\right),$$

to be used for predictions of Y_{n+1} based on new observation $\boldsymbol{X}_{n+1}$ such that $\boldsymbol{X}_{n+1} \nleq \boldsymbol{t}_X$.

MGP predictive algorithm

Input Extreme training dataset $\mathscr{D}_{ext} = \{\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_k\}$ with $\boldsymbol{Z}_i = (\boldsymbol{X}_i, Y_i)^T \in \mathbb{R}^d \times \mathbb{R}$ input/target pair; fitted GP parameters $(\boldsymbol{\sigma}, \boldsymbol{\xi}) = ((\boldsymbol{\sigma}_X, \boldsymbol{\xi}_X)^T, (\sigma_Y, \xi_Y)^T)^T \in \mathbb{R}^{2d+2}$; a multivariate threshold $\boldsymbol{t} = (\boldsymbol{t}_X, t_Y)^T$; $\mathscr{H} = \{\mathscr{H}_1, \ldots, \mathscr{H}_N\}$ set of N classes of density functions.
Marginal Exponential Transformation Apply the exponential transformation to each margin of the shifted observations in $\mathscr{D}_{ext}$

$$\tilde{\boldsymbol{Z}}_i = (\tilde{\boldsymbol{X}}_i, \tilde{Y}_i) = e_{\boldsymbol{\sigma},\boldsymbol{\xi}}(\boldsymbol{Z}_i - \boldsymbol{t}) = -\log\left(\left(1 + \frac{\boldsymbol{\xi}(\boldsymbol{Z}_i - \boldsymbol{t})}{\boldsymbol{\sigma}}\right)^{-1/\boldsymbol{\xi}}\right), \quad \text{for all } i \in \{1, \ldots, k\}.$$

Density selection Fit each density model $\mathscr{H}_j$, for all $1 \leq j \leq N$, to the transformed data $\{\tilde{\boldsymbol{Z}}_1, \ldots, \tilde{\boldsymbol{Z}}_k\}$ and select the density $\hat{h} \in \mathscr{H}$ with the smallest AIC.
Output Near-optimal density function $\hat{h}$ in $\mathscr{H}$ and a procedure to be used for predictions of Y_{k+1} based on new observations $\boldsymbol{X}_{k+1}$ such that $\boldsymbol{X}_{k+1} \nleq \boldsymbol{t}_X$ so that

- Generate a sample $(\hat{\tilde{Y}}^1_{k+1}, \ldots, \hat{\tilde{Y}}^L_{k+1})$ *via* rejection sampling from the conditional density

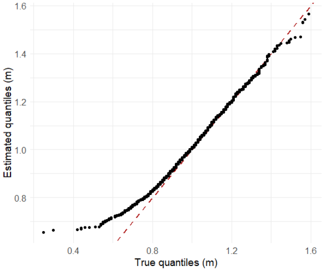
$$\hat{h}_{|\tilde{\boldsymbol{X}}_{k+1}}(\tilde{y}) := \frac{\hat{h}(\tilde{\boldsymbol{X}}_{k+1}, \tilde{y})}{\int_{\mathbb{R}} \hat{h}(\tilde{\boldsymbol{X}}_{k+1}, s)ds}.$$

- Backtransform the sample *via*

$$(\hat{Y}^1_{k+1}, \ldots, \hat{Y}^L_{k+1}) = (e^{-1}_{\sigma_Y,\xi_Y}(\hat{\tilde{Y}}^1_{k+1}) + t_Y, \ldots, e^{-1}_{\sigma_Y,\xi_Y}(\hat{\tilde{Y}}^L_{k+1}) + t_Y).$$

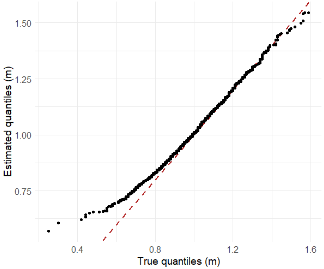
- Obtain a prediction of Y_{k+1} by the Monte Carlo average $\hat{Y}_{k+1} = (1/L)\sum_{l=1}^{L} \hat{Y}^l_{k+1}$.

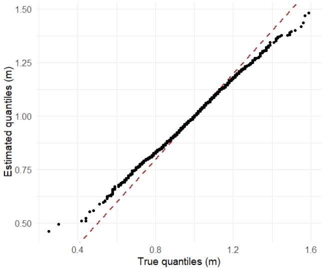
QQ-plots

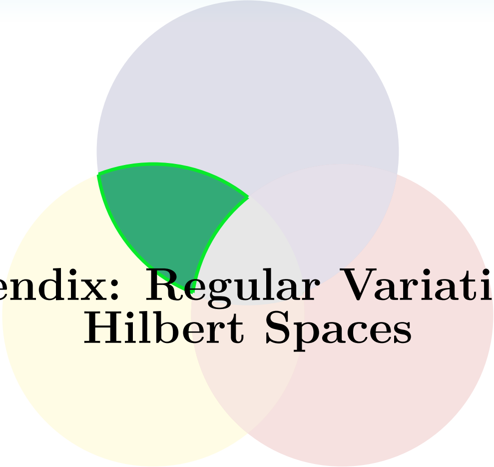


rajouter legendes

QQ-plots of the true values *vs* the estimated ones *via* the ROXANE procedure (with OLS and RF regression algorithms) and the MGP procedure.







Appendix: Regular Variation in Hilbert Spaces

Regular variation of *fidi* projections not sufficient 4

a counter-example

Let $R \sim Pareto(\alpha)$ and $\mathscr{L}(\Theta|R) = \frac{1}{\sum_{l=1}^{\lfloor R \rfloor} 1/l} \sum_{i=1}^{\lfloor R \rfloor} \frac{1}{i} \delta_{e_i}$.

Consider $X = R\Theta$.

- $\forall N \geq 1, \pi_N(X) \in RV(\mathbb{H})$;
- $\|X\| \in RV$;
- **but** $X \notin RV(\mathbb{H})$.

$RV(\mathscr{C}[0,1])$ *vs* $RV(L^2[0,1])$

equivalence?

Let X stochastic process in $\mathscr{C}[0,1]$.

$\Rightarrow$ since $\|\cdot\|_2$ continuous w.r.t. $\|\cdot\|_\infty$ in $\mathscr{C}[0,1]$ ([Dombry and Ribatet,2015]).

$\not\Leftarrow$ **a counter-example**: $Z \sim Pareto(\alpha_Z)$, $\rho \sim Pareto(\alpha_\rho)$ with $0 < \alpha_\rho < \alpha_Z$ and $Z \perp\!\!\!\perp \rho$. Let

$$Y(t) = \left(1 - \frac{t}{3Z^2 \exp(-2Z)}\right) \exp(Z) \mathbb{1}\{[0, 3Z^2 \exp(-2Z)[\}.$$

$\rightsquigarrow$ $Y \in \mathscr{C}[0,1] \cap RV(L^2[0,1])$ **but** $Y \notin RV(\mathscr{C}[0,1])$ (since $\|Y\|_\infty \notin RV$)

Appendix: Principal Component Analysis of Extremes

Skorokhod's representation Theorem. Let $(X_n)_{n\geq 1}$ and X_∞ r.v. defined on $(\Omega, \mathscr{F}, \mathbb{P})$ such that $X_n \xrightarrow{w} X_\infty$ and the support of X_∞ is separable. Then, there exist $(Y_n)_{n\geq 1}$ and Y_∞ defined on $(\Omega', \mathscr{F}', \mathbb{P}')$ such that $\mathscr{L}(X_n) = \mathscr{L}(Y_n)$ and $\mathscr{L}(X_\infty) = \mathscr{L}(Y_\infty)$ and $Y_n \xrightarrow{as} Y_\infty$.

Theorem 3 in [Zwald and Blanchard, 2005]**.** Let A be a symmetric positive $\mathbb{H}$-HS-operator with simple nonzero eigenvalues $\lambda_1 > \lambda_2 > \ldots$ Let $p > 1$ be an integer such that $\lambda_D > 0, \gamma^p = (\lambda_p - \lambda_{p+1})/2$. Let $B \in HS(\mathbb{H})$ be another symmetric operator such that $\|B\|_{HS(\mathbb{H})} < \gamma^p/2$ and $(A+B)$ is still a positive operator: Let $\Pi^p(A)$ (resp. $\Pi^p(A+B)$) denote the orthogonal projector onto the subspace spanned by the first p eigenvectors A (resp. $(A+B)$). Then these satisfy:

$$\|\Pi^p(A) - \Pi^p(A+B)\| \leq \frac{\|B\|}{\gamma^p}.$$

Concentration of the empirical covariance operator. with probability at least $1-\delta$,

$$\|\hat{C}_k - C_{t_{n,k}}\|_{HS(\mathbb{H})} \leq B(n,k,\delta),$$

with $B(n,k,\delta) := \frac{1+4\sqrt{\log(2/\delta)}+\sqrt{8\log(4/\delta)}}{\sqrt{k}} + \frac{8\log(2/\delta)+4\log(4/\delta)}{3k}$, and in particular,

$$\rho(\hat{V}_k^p, V_{t_{n,k}}^p) \leq \frac{B(n,k,\delta)}{\gamma_{t_{n,k}}^p}, \text{ with } \gamma_{t_{n,k}}^p = (\lambda_{t_{n,k}}^p - \lambda_{t_{n,k}}^{p+1})/2.$$