Extreme quantile regression: a coupling approach and Wasserstein distance

Benjamin Bobbia
Joint work with C.Dombry and D.Varron

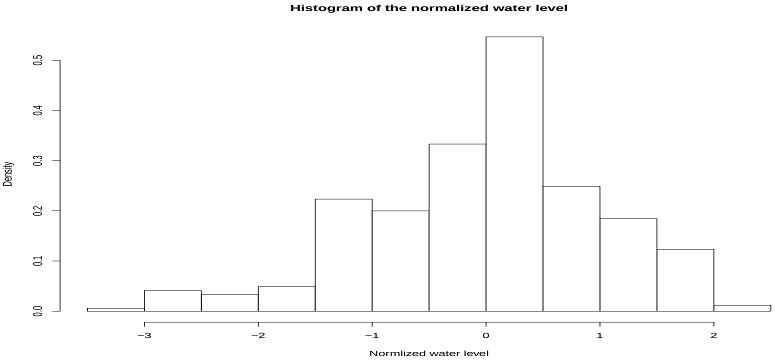
CYU and Université de Bourgogne Franche-Comté

EcoDep, January 27th 2021

Structure of the talk

Introduction

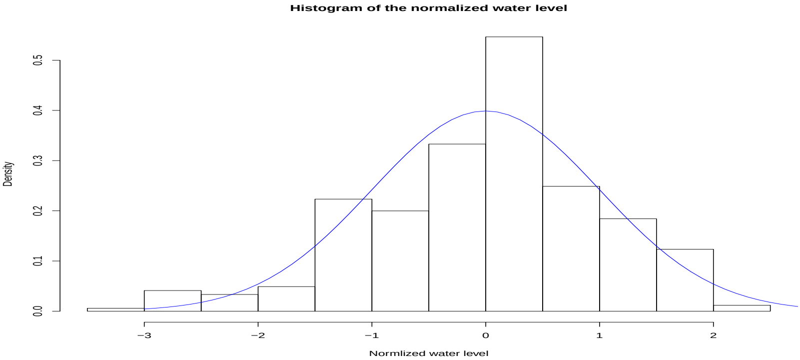
We consider the height of the Saône River in the city of Cendrecourt[1] (France), and we want to estimate the return level for a given period T.



[1] Data from "https://www.vigicrues.gouv.fr"

Introduction

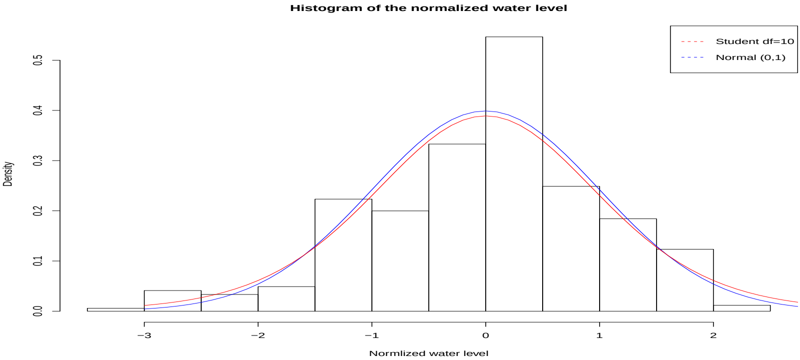
We consider the height of the Saône River in the city of Cendrecourt[2](France), and we want to estimate the return level for a given period T.



[1]Data from "https://www.vigicrues.gouv.fr"

Introduction

We consider the height of the Saône River in the city of Cendrecourt[3](France), and we want to estimate the return level for a given period T.

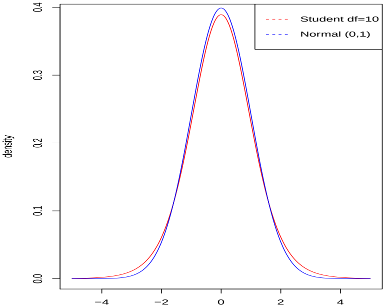


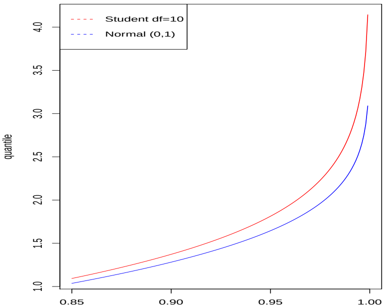
[1]Data from "https://www.vigicrues.gouv.fr"

Introduction

We compare the estimated return level (in meters) for a period T with fitting a normal distribution $\mathcal{N}(0,1)$ and a student distribution with 10 degrees of freedom.

	$\mathcal{N}(0,1)$	$t(df = 10)$
T=1 month	1.29	1.42
T=1 year	1.39	1.66
T=10 years	1.46	1.91





Quantile estimation

Let $(Y_i)_{1\leq i\leq n}$ be i.i.d copies of $Y \in \mathbb{R}$.

We want to estimate the quantile $q(\alpha_n)$ of order $1-\alpha_n$ of Y, with $\alpha_n \to 0$:

$$F(q(\alpha_n)) = \mathbb{P}(Y \leq q(\alpha_n)) \approx 1 - \alpha_n.$$

Formally,

$$q(\alpha_n) = F^{\leftarrow}(1-\alpha_n)$$

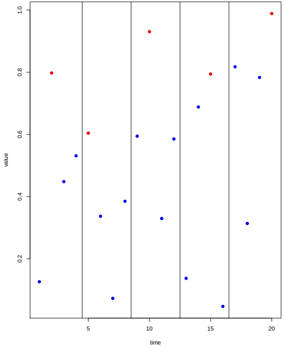
can be estimated by $\mathbb{F}_n^{\leftarrow}(1-\alpha_n)$ with $\mathbb{F}_n$ an estimator of the cdf.

Problem: If there are few observations at the level $1-\alpha_n$, we need a tail extrapolation.

Two main approaches

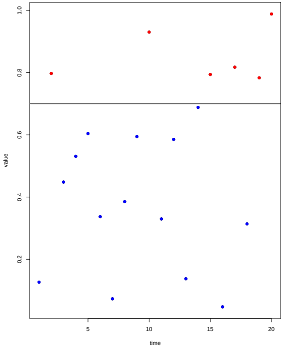
Block Maxima:

- divide the sample into blocks
- subsample : maximum of each block



Peak over Threshold:

- set a high threshold
- subsample : observations above the threshold



Theorem (Fisher-Tippett-Gnedencko/Balkema and de Haan)

For $(X_n)_{i\in\mathbb{N}}$ *an i.i.d sequence of random variables with cdf* F*, the following statements are equivalent:*

- *There is a positive function* a *such that for* $y > 0$

$$\lim_{t\to\infty} \frac{U(ty) - U(t)}{a(t)} = \frac{y^\gamma - 1}{\gamma},$$

  with $U(t) := F^{\leftarrow}(1 - \frac{1}{t})$.

- *There exist real constants* $a_n > 0$ *and* b_n *such that* $\lim_{n\to\infty} F^n(a_n y + b_n) = G_\gamma(y)$*, with* $G_\gamma(y) := \exp(-(1+\gamma y)^{-1/\gamma})$.

- *There exists a function* f *such that*

$$\mathbb{P}\left(\frac{X-u}{f(u)} > x | X > u\right) \underset{u\to x^*}{\longrightarrow} 1 - H_\gamma(x),$$

  with $H_\gamma(x) := 1 - (1+\gamma x)^{-1/\gamma}$.

We say that F belongs to the domain of attraction of G_γ, noted $F \in D(G_\gamma)$ and γ is called the extreme value index.

Extreme quantile estimation

From now we suppose that $\gamma > 0$. In this case the extreme value first order condition simplifies into

$$U(ty)/U(t) \underset{t\to\infty}{\longrightarrow} y^{\gamma}, \quad y > 1.$$

This gives a new way to estimate q_n

$$q(\alpha_n) = U(1/\alpha_n) \approx U\left(\frac{n}{k}\right)\left(\frac{k}{\alpha_n n}\right)^{\gamma},$$

with $k := k(n)$ a sequence such that $k \to \infty$ and $k/n \to 0$.
We can propose the estimator, known as Weissman estimator,

$$\hat{q}(\alpha_n) = Y_{n-k:n}\left(\frac{k}{n\alpha_n}\right)^{\hat{\gamma}_n}$$

with random threshold $Y_{n-k:n}$.

Structure of the talk

The Wasserstein metric

The Wasserstein distance is a distance on the space of probability measures $\mathcal{M}_1(\mathcal{X})$ on a metric space $(\mathcal{X}, d)$.

Definition

Let P_1 and P_2 be two probability measures on $(\mathcal{X}, d)$. The Wasserstein distance of order $p \in [1, \infty)$ is defined by

$$W_p(P_1, P_2) = \inf\{\mathbb{E}(d(X_1, X_2)^p)^{1/p}, X_1 \sim P_1, X_2 \sim P_2\}.$$

For $p = \infty$,

$$W_\infty(P_1, P_2) = \inf\{\text{ess sup } d(X_1, X_2), X_1 \sim P_1, X_2 \sim P_2\}.$$

We call (X_1, X_2) a coupling between P_1 and P_2.

Wasserstein spaces

Definition

We define the Wasserstein space of order $p \in [1, \infty)$ by

$$\mathcal{W}_p(\mathcal{X}) = \{P \in \mathcal{M}_1(\mathcal{X}) \text{ such that } \int_{\mathcal{X}} d(x, x_0)^p P(dx) < \infty\},$$

for some $x_0 \in \mathcal{X}$.

For a sequence $(P_n)_{n \in \mathbb{N}}$ and P in $\mathcal{W}_p(\mathcal{X})$,

$$W_p(P_n, P) \underset{n \to \infty}{\longrightarrow} 0 \text{ implies that } P_n \xrightarrow[n \to \infty]{d} P.$$

A particular case

If we consider Π_1 and Π_2 two random variables taking their values in $\mathcal{W}_p(\mathcal{X})$.

It is natural to considere their distribution in $\mathcal{W}_p(\mathcal{W}_p(\mathcal{X})) = \mathcal{W}_p^{(2)}(\mathcal{X})$ and endow this space with

$$W_p^{(2)}(P_{\Pi_1}, P_{\Pi_2}) = \inf\{\mathbb{E}(W_p(\Pi_1, \Pi_2)^p)^{1/p}, \Pi_1 \sim P_{\Pi_1}, \Pi_2 \sim P_{\Pi_2}\}.$$

We respectively call $\mathcal{W}_p^{(2)}(\mathcal{X})$ and $W_p^{(2)}$ second order Wassertein space and distance.

Wasserstein distance between emprirical measures

Let $X_1, ..., X_n$ and $X_1^*, ..., X_n^*$ be i.i.d samples with distribution P_X and P_{X^*}. Define the empirical measures

$$\Pi_n = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_{X_i} \quad \text{and} \quad \Pi_n^* = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_{X_i^*}.$$

Both Π_n and Π_n^* take their values in $\mathcal{W}_p(\mathcal{X})$ and define measurable maps $P_{\Pi_n}, P_{\Pi_n^*}$ on $\mathcal{W}_p^{(2)}(\mathcal{X})$.

Theorem (B. et al. (2020))

Let $p \in [1, \infty]$ and assume $W_p(P_X, P_{X^*}) < \infty$. Then,

$$W_p^{(2)}(P_{\Pi_n}, P_{\Pi_n^*}) \leq W_p(P_X, P_{X^*}).$$

Furthermore we have equality for p finite if $\mathcal{X}$ is separable and complete.

We can provide the same result with weighted empirical measures.

Wasserstein distance and PoT inference

- We consider the Peak over Threshold method and focus on the heavy tail case $\gamma > 0$.
- Under the first order condition

$$\lim_{t\to\infty} \frac{U(tz)}{U(t)} = z^{\gamma}, \quad z > 0,$$

  the rescaled exceedences above a high threshold satisfy

$$\mathcal{L}(u^{-1}X > x | X > u) \xrightarrow{d} Pareto(\gamma^{-1}) \quad \text{as } u \to \infty.$$

- Two questions:
  - Can we quantify the approximation of $\mathcal{L}(u^{-1}X|X > u)$ by $Pareto(\gamma^{-1})$ in Wasserstein metric ?
  - Can we use this to prove the asymptotic normality of an estimator of γ?

On $\mathcal{X} = [1, \infty)$ with distance $d(x, x') = |\log(x) - \log(x')|$, the Wasserstein distance

$$A_p(t) := W_p(P_{U(t)^{-1}X|X>U(t)}, Pareto(\gamma^{-1}))$$

can be explicitly determined.

Proposition (B. et al. (2020))

$$A_p(t) = \begin{cases} \left(\int_1^\infty \left|\log \frac{U(tx)}{x^\gamma U(t)}\right|^p \frac{\mathrm{d}x}{x^2}\right)^{1/p} & \text{for } 1 \leq p < \infty \\ \sup_{x>1} \left|\log \frac{U(tx)}{x^\gamma U(t)}\right| & \text{for } p = \infty \end{cases}.$$

Under the first order condition with $\gamma > 0$, $A_p(t) \to 0$ *as* $t \to \infty$ *for all* $p \in [1, \infty)$.

We can even provide a rate of convergence under second order regular variation condition.

Let $X_1, \ldots, X_n$ be a i.i.d sample with distribution P_X and $X_1^*, \ldots, X_k^*$ be an i.i.d sample with distribution $Pareto(\gamma^{-1})$.
With

$$X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n},$$

the upper order statistics, we define

$$\Pi_{n,k} = \frac{1}{k}\sum_{i=1}^{k} \varepsilon_{\frac{X_{n+1-i:n}}{X_{n-k:n}}} \quad \text{and} \quad \Pi_k^* = \frac{1}{k}\sum_{i=1}^{k} \varepsilon_{X_i^*}.$$

We measure the error between the two empirical measures in Wasserstein distance in the space $\mathcal{W}_p^{(2)}(\mathcal{X})$.

Theorem

Assume $\mathcal{X} = [1, \infty)$ is endowed with the logarithmic distance and F continuous. Then, for $p \in [1, \infty)$, $t > 1$ and $1 \leq k \leq n$, we have

$$W_p^{(2)}(P_{\Pi_{n,k}|X_{n-k:n}}, P_{\Pi_k^*}) = A_p\left(U^{-1}(X_{n-k:n})\right).$$

Under the first order condition and provided $k/n \to 0$, $A_p(U^{-1}(X_{n-k:n})) \longrightarrow 0$ as $n \to \infty$.
Under the second order condition, $A_p(U^{-1}(X_{n-k:n})) \sim A_p\left(\frac{n}{k}\right)$ as $n \to \infty$.

The Hill estimator

The Hill estimator is defined by

$$\hat{\gamma}_{n,k}^{HILL} = \frac{1}{k}\sum_{i=1}^{k} \log\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right)$$

Corollary

Let $1 \leq p < \infty$ and $t > 1$. For all $1 \leq k \leq n$,

$$W_p(P_{\sqrt{k}(\hat{\gamma}_{n,k}^{Hill}-\gamma)|X_{n-k:n}=U(t)}, \mathcal{N}(0,\gamma^2)) \leq \sqrt{k}A_p(t) + \left(4 + 3\sqrt{\frac{2}{\pi}}\right)\frac{\gamma^3}{\sqrt{k}}.$$

As a consequence, the Hill estimator is asymptotically normal as soon as

$$k \to \infty \quad \text{and} \quad \sqrt{k}A_p\left(\frac{n}{k}\right) \to 0.$$

The Weissman estimator

The Weissman estimator is defined by

$$\hat{q}(\alpha_n) = X_{n-k:n} \left(\frac{k}{n\alpha_n} \right)^{\hat{\gamma}^{Hill}_{n,k}} .$$

Corollary

Let $1 \leq p < \infty$, for $k = k(n)$ an intermediate sequence and $\alpha_n = o(k/n)$

$$W_p \left(P_{v_n^{-1} \log\left(\frac{\hat{q}(\alpha_n)}{q(\alpha_n)}\right) | X_{n-k:n}}, \mathcal{N}(0, \gamma^2) \right) \underset{n\to\infty}{=} O_P \left(\sqrt{k} A_p(n/k) + 1/\sqrt{k} \right),$$

with $v_n = \log(k/n\alpha_n)/\sqrt{k}$.

Under second order conditions, if $v_n \to 0$ we recover

$$v_n^{-1} \left(\frac{\hat{q}(\alpha_n)}{q(\alpha_n)} - 1 \right) \xrightarrow{d} \mathcal{N}(0, \gamma^2).$$

Extensions :

- Asymtpotic normality of $\hat{\gamma}_{n,k}^{Hill}$ in presence of bias $\left(\sqrt{k}A\left(\frac{n}{k}\right) \to \lambda \in \mathbb{R}\right)$
- Asymptotic normality of Probability Weighted Moments estimators.
- Caracterisation of the Block Maxima approach in term of Wasserstein distance.

Structure of the talk

Extreme Quantile Regression Problem

Use the information available in a covariate $X \in \mathbb{R}^p$.
In a regression framework, we observe $(X_i, Y_i)_{1\leq i\leq n}$ be n i.i.d copies of $(X, Y) \in \mathbb{R}^p \times \mathbb{R}$, with F the cdf of Y, F_x the conditional cdf of Y given $X = x$ and P_X the distribution of X.

We want to estimate the conditional quantile $q(\alpha_n|x)$ of Y given $X = x$ of order $1 - \alpha_n$, with $\alpha_n \to 0$.
It satisfies

$$F_x(q(\alpha_n|x)) = \mathbb{P}(Y \leq q(\alpha_n|x)|X = x) \approx 1 - \alpha_n.$$

and is defined by

$$q(\alpha_n|x) = F_x^{\leftarrow}(1 - \alpha_n).$$

Framework:

The proportional tail model (closely related to the heteroscedastic extremes by Einmahl et al. '16) assumes two main conditions:

- **Extreme value condition:** $F \in D(G_\gamma)$ with $\gamma > 0$ i.e

$$\lim_{t\to\infty} \frac{U(tz)}{U(t)} = z^\gamma, \quad z > 0$$

- **Asymptotically proportional tails:**

$$\lim_{y\to\infty} \frac{1-F_x(y)}{1-F(y)} = \sigma(x) \text{ uniformly in } x \in \mathbb{R}^p,$$

  with σ the skedasis function satisfying $\int_{\mathbb{R}^p} \sigma(x) P_X(\mathrm{d}x) = 1$

The two conditions imply $F_x \in D(G_\gamma)$ with the same γ for all $x \in \mathbb{R}^p$.

Second order asumptions

- **Extreme value condition:** $F \in D(G_\gamma)$ with $\gamma > 0$ and

$$\lim_{t\to\infty} \frac{\frac{U(tz)}{U(t)} - z^\gamma}{A_{ext}(t)} = z^\gamma \frac{z^\rho - 1}{\rho}, \quad z > 0$$

  with $\rho < 0$.

- **Asymptotically proportional tails:**

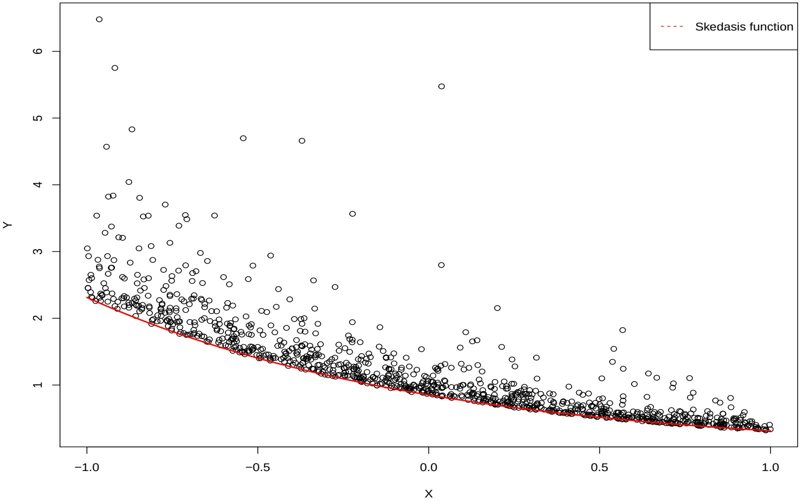
$$\sup_{x\in\mathbb{R}^d} \sup_{y'>y} \left| \frac{1 - F_x(y')}{\sigma(x)(1 - F(y'))} \right| = A_{prop}(y) \underset{y\to\infty}{\longrightarrow} 0.$$

Moreover, we assume that there exists $\varepsilon_\sigma > 0$ such that

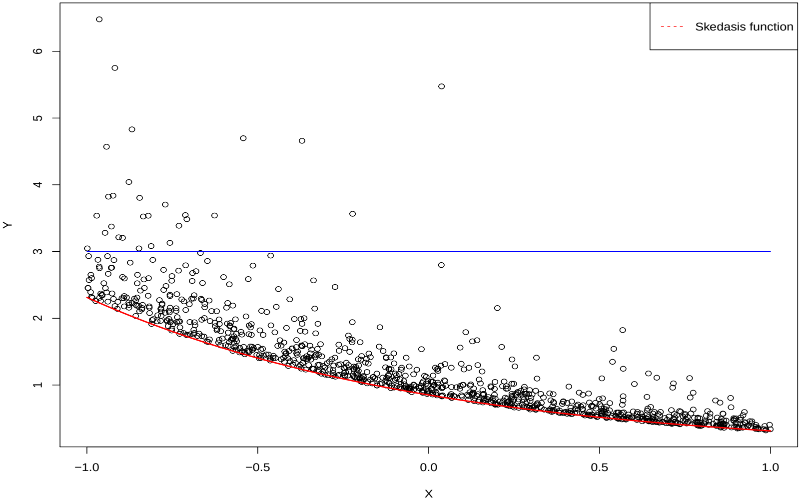
$$\varepsilon_\sigma \le \sigma(x) \le \varepsilon_\sigma^{-1}, \quad x \in \mathbb{R}^d.$$

We define $A := \max(A_{prop}, A_{ext})$.

Data simulated with $X \sim \mathcal{U}([-1,1])$, $\gamma = 1/5$ and $\sigma(x) = e^{-x}/\sinh(1)$.
The sample size is $n = 1000$.



Data simulated with $X \sim \mathcal{U}([-1,1])$, $\gamma = 1/5$ and $\sigma(x) = e^{-x}/\sinh(1)$. The sample size is $n = 1000$.



A Weissman-type estimator

The proportional tail assumption gives a new approximation for the conditional quantile q_n of order $1-\alpha_n$:

$$q(\alpha_n|x) = F_x^{\leftarrow}(1-\alpha_n) \approx F^{\leftarrow}\left(1-\frac{\alpha_n}{\sigma(x)}\right) \approx U\left(\frac{\sigma(x)}{\alpha_n}\right).$$

Thanks to the extreme value first order condition , we can now estimate

$$\hat{q}(\alpha_n|x) := y_n \left(\frac{\hat{p}_n \hat{\sigma}_n(x)}{\alpha_n}\right)^{\hat{\gamma}_k}$$

with a threshold y_n and $p_n = \mathbb{P}(Y > y_n)$ estimated by $\hat{p}_n$.

Our first estimators

We consider the threshold y_n, random or deterministic. Now define :

- The estimator of the integrated skedasis $C(x) = \int_{\{u \le x\}} \sigma(u) P_X(\mathrm{d}u)$:

$$\hat{C}_n(x) := \frac{1}{k_n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \le x\}} \mathbb{1}_{\{Y_i > y_n\}}.$$

- The estimator of γ (of Hill type):

$$\hat{\gamma}_n := \frac{1}{k_n} \sum_{i=1}^{n} \log\left(\frac{Y_i}{y_n}\right) \mathbb{1}_{\{Y_i > y_n\}}$$

with k_n the number of exceedances.

Their asymptotic normality

This theorem extends the results of Einmahl et al.(2016):

Theorem

If $k_n \xrightarrow{\mathbb{P}} +\infty$, $\frac{k_n}{n} \xrightarrow{\mathbb{P}} 0$ *and* $\sqrt{k_n}^{1+\varepsilon} A(\frac{n}{k_n}) \xrightarrow{\mathbb{P}} 0$, *we have:*

$$\begin{pmatrix} \sqrt{k}(\widehat{C}_n(\cdot) - C(\cdot)) \\ \sqrt{k}(\hat{\gamma}_n - \gamma) \end{pmatrix} \xrightarrow[n \to +\infty]{\mathcal{L}} \begin{pmatrix} B(\cdot) \\ N \end{pmatrix}$$

where $N \sim \mathcal{N}(0, \gamma^2)$ *and* B *is a* C*-Brownian bridge. Moreover* N *and* $B(\cdot)$ *are independent.*

Idea of coupling approach

$$(Z_i)_{1 \leq i \leq n} \sim Pareto(1) \text{ and } (E_{i,n})_{1 \leq i \leq n} \sim \mathcal{B}(1 - F(y_n))$$

Idea of coupling approach

$$(Z_i)_{1\leq i\leq n} \sim Pareto(1) \text{ and } (E_{i,n})_{1\leq i\leq n} \sim \mathcal{B}(1-F(y_n))$$

$$(X_i, Y_i)_{1\leq i\leq n} \qquad (X_i^*, Y_i^*)_{1\leq i\leq n}$$

- If $E_{i,n} = 1$, we set
  - $X_i \sim P_{X|Y>y_n}$, $X_i^* \sim \sigma P_X$

    such that $\mathbb{P}(X_i \neq X_i^*) = \|P_{X|Y>y_n} - \sigma P_X\|_{TV}$.
  - $Y_i = U_{X_i}(\frac{Z_i}{(1-F(y_n))})$, $Y_i^* = y_n Z_i^\gamma$
- If $E_{i,n} = 0$, we set
  - $X_i = X_i^* \sim P_{X|Y\leq y_n}$
  - $Y_i = Y_i^* \sim \mathcal{L}(Y|Y<y_n,\ X = X_i)$.

Idea of coupling approach

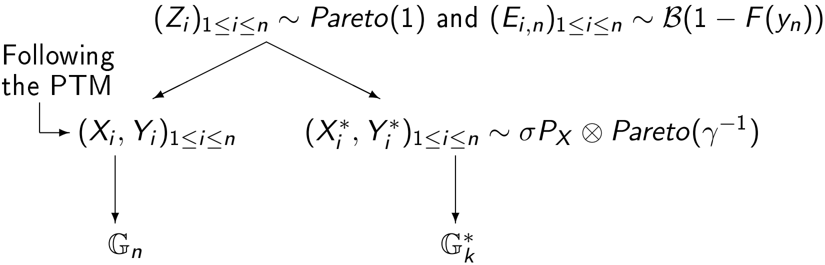
$$(Z_i)_{1\leq i\leq n} \sim Pareto(1) \text{and } (E_{i,n})_{1\leq i\leq n} \sim \mathcal{B}(1-F(y_n))$$

Following the PTM

$$(X_i, Y_i)_{1\leq i\leq n} \qquad (X_i^*, Y_i^*)_{1\leq i\leq n} \sim \sigma P_X \otimes Pareto(\gamma^{-1})$$

- $\max_{i:\ E_{i,n=1}} \left| \frac{Y_i^*}{Y_i} - 1 \right| = O\left(A\left(\frac{1}{1-F(y_n)}\right)\right)$
- $\mathbb{P}(X_1 \neq X_1^*) = O\left(A\left(\frac{1}{1-F(y_n)}\right)\right)$

Idea of coupling approach



- $\mathbb{P}(X_1 \neq X_1^*) = O(A)$
- $\max_{i:\ E_{i,n}=1} |Y_i^*/Y_i - 1| = O(A)$

Idea of coupling approach

$(Z_i)_{1\le i\le n} \sim Pareto(1)$ and $(E_{i,n})_{1\le i\le n} \sim \mathcal{B}(1 - F(y_n))$

Following the PTM $\to (X_i, Y_i)_{1\le i\le n}$

$(X_i^*, Y_i^*)_{1\le i\le n} \sim \sigma P_X \otimes Pareto(\gamma^{-1})$

$(X_i, Y_i)_{1\le i\le n} \to \mathbb{G}_n$

$(X_i^*, Y_i^*)_{1\le i\le n} \to \mathbb{G}_k^*$

$\mathbb{G}_k^* \downarrow \mathcal{L}$ by Donsker's theorem

$\mathbb{W}$ tight Borel

- $\mathbb{P}(X_1 \neq X_1^*) = O(A)$
- $\max_{i:\ E_{i,n=1}} |Y_i^*/Y_i - 1| = O(A)$

Idea of coupling approach

$(Z_i)_{1\le i\le n} \sim Pareto(1)$ and $(E_{i,n})_{1\le i\le n} \sim \mathcal{B}(1-F(y_n))$

Following the PTM → $(X_i, Y_i)_{1\le i\le n}$

$(X_i^*, Y_i^*)_{1\le i\le n} \sim \sigma P_X \otimes Pareto(\gamma^{-1})$

$\varphi(\mathbb{G}_n) = (\hat\gamma_n, \widehat{C}_n(.))$

$\varphi(\mathbb{G}_k^*) = (\hat\gamma_k^*, \widehat{C}_k^*(.))$

$\downarrow \mathcal{L}$ by Donsker's theorem

$\varphi(\mathbb{W}) = (N, B)$ Brownian bridge

φ Lipschitz

- $\mathbb{P}(X_1 \neq X_1^*) = O(A)$
- $\max_{i:\ E_{i,n}=1} |Y_i^*/Y_i - 1| = O(A)$

Idea of coupling approach

$(Z_i)_{1\le i\le n} \sim Pareto(1)$ and $(E_{i,n})_{1\le i\le n} \sim \mathcal{B}(1-F(y_n))$

Following the PTM

$(X_i, Y_i)_{1\le i\le n}$ $\quad$ $(X_i^*, Y_i^*)_{1\le i\le n} \sim \sigma P_X \otimes Pareto(\gamma^{-1})$

$\varphi(\mathbb{G}_n) = (\hat{\gamma}_n, \widehat{C}_n(.))$ $\quad$ $\varphi(\mathbb{G}_k^*) = (\hat{\gamma}_k^*, \widehat{C}_k^*(.))$

$\xrightarrow{\mathcal{L}}$ $\quad$ $\downarrow \mathcal{L}$ by Donsker's theorem

Coupling arguments :

- $\mathbb{P}(X_1 \neq X_1^*) = O(A)$
- $\max_{i:\ E_{i,n}=1} |Y_i^*/Y_i - 1| = O(A)$

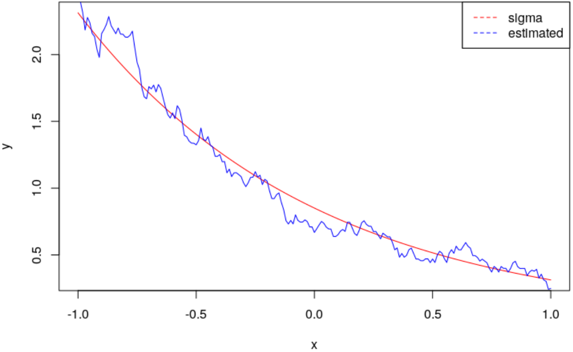
$\varphi(\mathbb{W}) = (N, B)$ Brownian bridge

φ Lipschitz

Estimation of the skedasis function

From this moment, we consider that the sequence of thresholds $(y_n)_{n\in\mathbb{N}}$ is deterministic. We can estimate the skedasis function by the kernel estimator

$$\hat{\sigma}_n(x) := \frac{\frac{1}{k}\sum_{i=1}^{n} \mathbb{1}_{\{|x-X_i|\leq h_n\}}\mathbb{1}_{\{Y_i\geq y_n\}}}{\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{|x-X_i|\leq h_n\}}}$$



Theorem

Let $(h_n)_{n\geq 1}$ be the bandwidth and $p_n := \bar{F}(y_n)$. Then, under technical assumptions and the proportional tail model assumption, for each $x \in \mathcal{O}$, we have

$$\sqrt{np_n h_n^d}\Big(\hat{\sigma}_n(x) - \sigma(x)\Big) \underset{n\to+\infty}{\longrightarrow} N$$

where $N \sim \mathcal{N}\left(0, \frac{\sigma(x)}{f(x)}\right)$.

Technical assumptions:
Let $(h_n)_{n\geq 1}$ be a deterministic sequence such that $h_n \underset{n\to\infty}{\longrightarrow} 0$. Assume that σ is continuous and that the density f of X is continuous and bounded away from zero on a open set $\mathcal{O} \subset \mathbb{R}^d$. Suppose in addition that

$$np_n h_n^d \to +\infty,\;\; p_n \to 0, \sqrt{np_n h_n^d} A(p_n^{-1}) \to 0.$$

Asymptotic normality of the extreme quantile estimator

Theorem

Under the assumptions of the previous theorem, if, in addition

$$\alpha_n = o(p_n), \log(n\alpha_n) = o(\sqrt{np_n}).$$

then we have

$$\sqrt{np_n h_n^d} \log\left(\frac{\hat{q}(\alpha_n|x)}{q(\alpha_n|x)}\right) \underset{n\to+\infty}{\longrightarrow} N,$$

where $N \sim \mathcal{N}\left(0, \gamma^2\right)$.

Structure of the talk

The Wasserstein distance in PTM

Let $t > 1$, consider the couple $((\tilde{X}, \tilde{Y}), (X^*, Y^*))$ defined by

- $\tilde{X} \sim P_{X|Y>U(t)}$ and $X^* \sim \sigma P_X$ satisfying maximal coupling.
- Given $\tilde{X} = \tilde{x}$ we set $\tilde{Y} = \frac{U_x(\bar{F}_x(U(t)))}{U(t)}$ and $Y^* \sim Pareto(\gamma^{-1})$.

Lemma

$((\tilde{X}, \tilde{Y}), (X^*, Y^*))$ is a coupling between $P_{(X, U(t)^{-1}Y)|Y \geq U(t)}$ and $\boldsymbol{P}^* := \sigma P_X \otimes Pareto(\gamma^{-1})$.

On $\mathcal{W}_p(\mathbb{R}^d \times [1, +\infty))$ with the underlying distance

$$d\left((x_1, y_1), (x_2, y_2)\right) = \mathbb{1}_{\{x_1 \neq x_2\}} + \left|\log(y_1) - \log(y_2)\right|.$$

Under the second order assumption of proportional tail model, we have

$$W_p(P_{(X, U(t)^{-1}Y)|Y>U(t)}, \mathbf{P}^*) = O(A(t)), \quad \text{as } t \to \infty.$$

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ i.i.d. with distribution $P_{(X,Y)}$ and $(X_1^*, Y_1^*), \ldots, (X_n^*, Y_n^*)$ i.i.d with distribution $\sigma P_X \otimes \text{Pareto}(\gamma^{-1})$.

Define the empirical measures

$$\Pi_{n,k} := \frac{1}{k}\sum_{i=1}^{k} \varepsilon_{\left(X_{(n-k+i)}, \frac{Y_{n-k+i:n}}{Y_{n-k:n}}\right)} \quad \text{and} \quad \Pi_k^* := \frac{1}{k}\sum_{i=1}^{k} \varepsilon_{(X_i^*, Y_i^*)}$$

Theorem

For $1 \leq p < \infty$ and (k_n) and intermediary sequence, under assumptions of the proportional tail, we have,

$$W_p^{(2)}(P_{\Pi_{n,k}|Y_{n-k:n}=U(t)}, P_{\Pi_k^*}) = O\left(A\left(t\right)\right)$$

In order to handle the bias we provide a similar result, comparing $\Pi_{n,k}$ with $\Pi_{k,t}^* = \frac{1}{k}\sum_{i=1}^{k} \varepsilon_{\left(X_i^*, Y_i^*\left(1+A_{ext}\left(\frac{1}{\bar{F}_X(U(t))}\right)\right)\frac{(Y_i^*)^{\rho/\gamma}-1}{\rho}\right)}, \quad k \geq 1, t \geq 1.$

The estimators $\widehat{C}_n$ and $\hat{\gamma}_n$ are built with threshold $Y_{n-k:n}$.

Theorem

If $k \to \infty$, $k/n \to 0$ *and* $\sqrt{k} A_{ext}(n/k) \to \lambda \in \mathbb{R}$ *and* $\sqrt{k} A_{prop}(n/k) \to 0$, *we have:*

$$\begin{pmatrix} \sqrt{k}(\widehat{C}_n(\cdot) - C(\cdot)) \\ \sqrt{k}(\hat{\gamma}_n - \gamma) \end{pmatrix} \xrightarrow[n \to +\infty]{\mathcal{L}} \begin{pmatrix} B(\cdot) \\ N \end{pmatrix},$$

where $N \sim \mathcal{N}(\lambda b(\rho), \gamma^2)$ *with*

$$b(\rho) = \int_0^1 \frac{z^\rho - 1}{\rho} \frac{\mathrm{d}z}{z^2} \times \int_{\mathbb{R}^p} \sigma(x)^{-\rho} \mathrm{P}_X(\mathrm{d}x),$$

and B *is a* C*-Brownian bridge. Moreover* N *and* $B(\cdot)$ *are independent.*

Validation procedure

The assumptions of proportional tail model implies that γ does not depend on the value of the covariate X.

$\hookrightarrow$ We test whether γ is constant over $X \in \mathbb{R}^d$

We define, for every $u \in \mathbb{R}^d$,

$$\Delta_n(u) := \frac{1}{k}\sum_{i=1}^{k}\left(\log\left(\frac{Y_{n+1-i:n}}{Y_{n-k:n}}\right) - \hat{\gamma}_n\right)\mathbb{1}_{\{X_{(n+1-i)}\leq u\}}.$$

Theorem

Under proportional tail model, for a sequence $k = k(n)$ satisfying $k \to +\infty, \frac{k}{n} \to 0$ and $\sqrt{k}A\left(\frac{n}{k}\right) \to 0$, we have

$$\sqrt{k}\frac{\Delta_n(.)}{\hat{\gamma}_n} \xrightarrow[k\to\infty]{d} B(.), \quad in\ L^\infty\left(\mathbb{R}^d\right),$$

with B a $C-$Brownian bridge.

Problem: The limit B depends on the unknown function C.

Bootstrap theorem

With $(\xi_{n,1}, \ldots, \xi_{n,k})$ non negative exchangeable random variables, independent of $(X_i, Y_i)_{1 \leq i \leq n}$, we define the bootstrap estimators as

$$\widehat{C}_{\xi,n}(x) := \frac{1}{k} \sum_{i=1}^{k} \xi_{n,i} \mathbb{1}_{\{X_{(n+1-i)} \leq x\}}, \quad x \in \mathbb{R}^d,$$

and

$$\hat{\gamma}_{\xi,n} := \frac{1}{k} \sum_{i=1}^{k} \xi_{n,i} \log\left(\frac{Y_{n+1-i:n}}{Y_{n-k:n}}\right).$$

For example, multinomial or exponential weights are of great interest in our framework. They respectively correspond to the Efron's and Bayesian bootstrap.

Theorem

Under technical assumptions for the weights and the proportional tail model, for a sequence $k = k(n)$ satisfying $k \to +\infty$, $\frac{k}{n} \to 0$ and $\sqrt{k} A\left(\frac{n}{k}\right) \to 0$, we have

$$\sqrt{k} \begin{pmatrix} \widehat{C}_{\xi,n}(\cdot) - \widehat{C}_n(\cdot) \\ \hat{\gamma}_{\xi,n} - \hat{\gamma}_n \end{pmatrix} \underset{\xi}{\rightsquigarrow} \begin{pmatrix} B(\cdot) \\ N \end{pmatrix}, \tag{1}$$

where $N \sim \mathcal{N}(0, \gamma^2)$, B is a C-Brownian bridge and N and B are independent.

The notation $\underset{\xi}{\rightsquigarrow}$ means conditional weak convergence.

Simulation study

Testing procedure:

We compare $\sup_{u \in \mathbb{R}^d} \sqrt{k} \left| \frac{\Delta_n(u)}{\hat{\gamma}_n} \right|$ and the empirical quantile of $\sup_{u \in \mathbb{R}^d} \sqrt{k} \left| \widehat{C}_{\xi,n}(u) - \widehat{C}_n(u) \right|$.

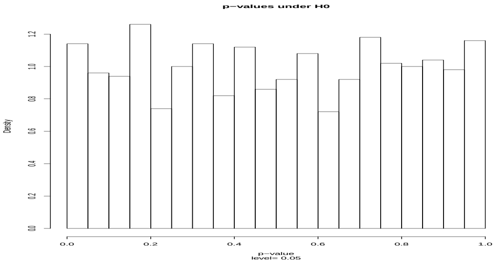
Experimental design

- $X \sim \mathcal{U}([0,1]^2)$
- $\sigma : (x_1, x_2) \mapsto e^{(x_1+x_2)/2}$
- $\gamma_0 = 1/4$ and $\gamma_\varepsilon : (x_1, x_2) \mapsto \gamma_0 + \varepsilon(x_1 + x_2)/2$
- $Y = \sigma(X) Y_0$ with Y_0 following a distribution belonging the domain of attraction of an extreme value distribution G_γ.

Level: Bayesian bootstrap the under proportional tail model.

α	5%			
n	5000		10000	
k	500	1000	1000	2000
Pareto	0.949	0.96	0.949	0.95
Frechet	0.939	0.965	0.962	0.949
Burr	0.953	0.967	0.952	0.952

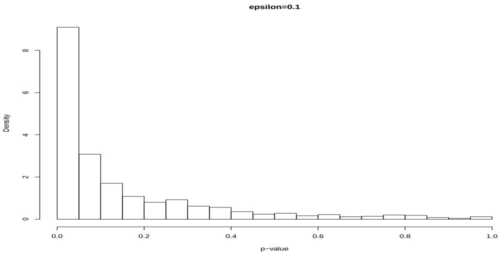
Table: Distribution of p-values under H_0



Power: Bayesian bootstrap for Fréchet distribution with γ depending on X.

α	5%			
n	5000		10000	
k	500	1000	1000	2000
Frechet ($\varepsilon = 0.1$)	0.803	0.546	0.588	0.239
Frechet ($\varepsilon = 0.5$)	0.54	0.306	0.267	0.056
Frechet ($\varepsilon = 1$)	0.064	0	0	0

Table: Distribution of p-values under H_1



Future work :

- Variable selection: build a test testing whether if a given covariate has an effect on extremes.
- Build an estimator of σ when the threshold is random, mainly when it is $Y_{n-k:n}$ and provide a quantile estimator.
- Extend the proportional tail model to the case $\gamma = 0$.

Thank you for your attention !

Handle the bias:

Comparing $\Pi_{n,k}$ with

$$\frac{1}{k}\sum_{i=1}^{k}\delta_{X_i^*\left(1+A(t)\frac{X_i^{*\,\rho/\gamma}-1}{\rho}\right)}$$

we have the asymptotic normality of Hill estimator when $\sqrt{k}A(n/k) \to \lambda \in \mathbb{R}$.

Asymptotic normality of PWM estimator:

Theorem

Assume $F \in D(G_\gamma)$ with $\gamma < 1$ and $p \in [1, 1/\gamma_+)$ we have

$$W_p\left(P_{(X-U(t))/a(t)|X>U(t)}, H_\gamma\right) = A'_p(t), \quad t > 1$$

with H_γ the Generalized Pareto distribution and

$$A'_p(t) := \left(\int_1^\infty \left|\frac{U(zt)-U(t)}{a(t)} - \frac{z^\gamma - 1}{\gamma}\right| \frac{\mathrm{d}z}{z^2}\right)^{\frac{1}{p}}, \quad t > 1.$$

Considering the Lipschitz on $\mathcal{W}_p(\mathbb{R})$

$$\pi \mapsto \left(\int_{\mathbb{R}} F_\pi^{\leftarrow}(u)^q u^r (1-u)^s \mathrm{d}u\right)^{1/q}$$

we also have the asymptotic normality of Probability Weighted Moments estimators.

Coupling in Block Maxima:

We divide the sample into k blocks of size m, and define M_i the maximum of the block $1 \leq i \leq k$. Considering $V(t) := F^{\leftarrow}(e^{-1/t})$ instead of U we have :

Theorem

Assume $F \in D(G_\gamma)$ with $\gamma < 1$ and $p \in [1, 1/\gamma_+)$ we have the following theorem:

$$W_p\left(P_{(M_i - b_m)/a_m|, G_\gamma}\right) = A''_p(m), \quad m > 1$$

with

$$A''_p(t) := \left(\int_1^\infty \left|\frac{V(zt) - V(t)}{a(t)} - \frac{z^\gamma - 1}{\gamma}\right| e^{-\frac{1}{z}} \frac{\mathrm{d}z}{z^2}\right)^{\frac{1}{p}}, \quad t > 1.$$

Conditions about weights:

$$\sup_{n \geq 1} \|\xi_{1,n} - 1\|_{2,1} < \infty,$$

$$\frac{1}{\sqrt{k}} \mathbb{E}\left(\max_{1 \leq i \leq k} |\xi_{i,n} - 1| \right) \to 0,$$

$$\frac{1}{k} \sum_{i=1}^{k} (\xi_{i,n} - 1)^2 \xrightarrow{\mathbb{P}} c^2 > 0.$$

With the norm defined by

$$\|\xi\|_{2,1} = \int_0^\infty \sqrt{\mathbb{P}(|\xi| > x)} \mathrm{d}x.$$

Note that the Efron and Bayesian bootstrap satisfy those assumptions with $c = 1$.

Conditional weak convergence:

We said that a process $\mathbb{G}_{\xi,n}$ converge weakly conditionally to ξ to a tight limit $\mathbb{G}$ if

$$\sup_{h \in BL_1} \left| \mathbb{E}_\xi(h(\mathbb{G}_{\xi,n})) - \mathbb{E}(h(\mathbb{G})) \right| \xrightarrow{\mathbb{P}^*} 0 \quad \text{and}$$

$$\mathbb{E}_\xi(h(\mathbb{G}_{\xi,n}))^* - \mathbb{E}_\xi(h(\mathbb{G}_{\xi,n}))_* \xrightarrow{\mathbb{P}} 0, \quad \text{for all } h \in BL_1,$$

where BL_1 is the set of 1-Lipschitz functions that are uniformly bounded by 1, and where $\mathbb{E}_\xi(X)^*$ and $\mathbb{E}_\xi(X)_*$ are inner and outer measurable cover functions.

Maximal coupling and total variations:

For P_1 and P_2 two probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, we define the total variations distance by

$$\|P_1 - P_2\|_{TV} = \sup\{|P_1(A) - P_2(A)|,\ A \in \mathcal{B}(\mathcal{X})\}.$$

Theorem (Maximal coupling)

Let P_1 and P_2 two probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, there exists a coupling (X_1, X_2) between P_1 and P_2 such that

$$\|P_1 - P_2\|_{TV} = \mathbb{P}(X_1 \neq X_2).$$

What about bootstrap measure?

We now consider $(\xi_1, ..., \xi_k)$ random variables taking values in $(0, +\infty)$ with expectation 1 and independent of $(X_1, ..., X_k)$ and $(X_1^*, ..., X_k^*)$ satisfying $\sum \xi_i = k$ almost surely. Define bootstrap versions of Π and Π^*:

$$\Pi_{\xi,n} := \frac{1}{k}\sum_{i=1}^{k} \xi_i \delta_{X_i} \quad \text{and} \quad \Pi^*_{\xi,n} := \frac{1}{k}\sum_{i=1}^{k} \xi_i \delta_{X_i^*}.$$

Since bootstrap theorems involve conditional weak convergence, we consider the data $(X_1, \dots, X_k)$ and $(X_1^*, \dots, X_k^*)$ as fixed.

Theorem (B. et al. (2020))

Conditionnaly on $(X_1, ..., X_k)$ and $(X_1^*, ..., X_k^*)$, for $p \in [1, \infty)$, we have

$$W_p^{(2)}(P_{\Pi_{\xi,n}}, P_{\Pi^*_{\xi,n}}) = W_p(\Pi_n, \Pi_n^*).$$