Coupling methods for extreme quantile regression

B.Bobbia, joint work with C.Dombry and D.Varron

Université de Bourgogne Franche-Comté

May 13th 2019

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

Definition

Let P_1 and P_2 are two probability measure on $(\mathcal{X}, d)$, we define the Wasserstein distance of order $p \in [1, \infty)$ by

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

and if $p = \infty$, we define

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

A such (X_1, X_2) is called a coupling between P_1 and P_2.

Lemma

If L is a Lipschitz mapping on $(\mathcal{X}, d)$ then

$$W_p(P_1 \circ L^{-1}, P_2 \circ L^{-1}) \leq \|L\|_{Lip} W_p(P_1, P_2).$$

For the case $\mathcal{X} = \mathbb{R}$ we have an explicit expression of the Wasserstein distance:

Lemma

Let P_1 and P_2 be two probability measures on $\mathbb{R}$,

$$W_p(P_1, P_2) = \left(\int_0^1 |F_1^{\leftarrow}(x) - F_2^{\leftarrow}(x)|^p \mathrm{d}x \right)^{1/p}.$$

Definition

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

$$\mathcal{W}_p(\mathcal{X}) = \{P \in \mathcal{M}_1(\mathcal{X}) \text{ such that } W_p(P, \delta_{x_0}) < \infty\}$$

where $x_0 \in \mathcal{X}$.

The definition of $\mathcal{W}_p(\mathcal{X})$ does not depend on x_0 and W_p is a distance on $\mathcal{W}_p(\mathcal{X})$. Moreover, for $1 \leq p \leq q \leq \infty$ we have the inclusion

$$\mathcal{W}_q(\mathcal{X}) \subset \mathcal{W}_p(\mathcal{X}).$$

We can transfer properties from $\mathcal{X}$ to $\mathcal{W}_p(\mathcal{X})$

Theorem

Let $p \in [1, \infty)$

- *if $(\mathcal{X}, d)$ is complete separable, then so is $(\mathcal{W}_p(\mathcal{X}), W_p)$*
- *if $(\mathcal{X}, d)$ is compact, then so is $(\mathcal{W}_p(\mathcal{X}), W_p)$.*

Definition

A sequence $(P_n)_{n\in\mathbb{N}}$ converge weakly to P in $\mathcal{W}_p(\mathcal{X})$ if

- $P_n^* \underset{n\to\infty}{\overset{\text{weakly}}{\longrightarrow}} P$
- $\int_{\mathcal{X}} d(x, x_0)^p P_n(\mathrm{d}x) \underset{n\to\infty}{\longrightarrow} \int_{\mathcal{X}} d(x, x_0)^p P(\mathrm{d}x)$

We can metrize the weak convergence on $\mathcal{W}_p(\mathcal{X})$ thank to Wasserstein distance

Theorem

Let $p \in [1, \infty)$, $(P_n)_{n\geq 1}$ and $P \in \mathcal{W}_p(\mathcal{X})$. P_n converge weakly to P in $\mathcal{W}_p(\mathcal{X})$ is equivalent to

$$W_p(P_n, P) \underset{n\to\infty}{\longrightarrow} 0.$$

Lemma

Let $(P_n)_{n\in\mathbb{N}}$ and $(P_n^*)_{n\in\mathbb{N}}$ in $\mathcal{W}_p(\mathcal{X})$ and $P \in \mathcal{W}_p(\mathcal{X})$.

If $P_n^* \overset{weakly}{\underset{n\to\infty}{\longrightarrow}} P$ and $W_p(P_n, P_n^*) \underset{n\to\infty}{\longrightarrow} 0$ then $P_n \overset{weakly}{\underset{n\to\infty}{\longrightarrow}} P$.

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

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

Theorem

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

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

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

Let $\varepsilon > 0$ and $(\tilde{X}, \tilde{X}^*)$ a coupling between P_X and P_{X^*} such that

$$\|d(\tilde{X}, \tilde{X}^*)\|_{L^p} \leq (1+\varepsilon) W_p(P_X, P_{X^*})$$

Taking i.i.d copies $(\tilde{X}_i, \tilde{X}_i^*)_{1 \leq i \leq n}$ of $(\tilde{X}, \tilde{X}^*)$ we can built empirical measures $\tilde{\Pi}_n$ and $\tilde{\Pi}_n^*$ such that

$$\tilde{\Pi}_n \overset{d}{=} \Pi_n \quad \text{and} \quad \tilde{\Pi}_n^* \overset{d}{=} \Pi_n^*.$$

For fixed ω, set $(Z, Z^*) = (\tilde{X}_\iota(\omega, \tilde{X}_\iota^*(\omega))$ with $\iota \sim \mathcal{U}(\{1, ..., n\})$. So $Z \sim \tilde{\Pi}_n(\omega)$ and $Z^* \sim \tilde{\Pi}_n^*(\omega)$. That why

$$W_p(\tilde{\Pi}_n(\omega), \tilde{\Pi}_n^*(\omega)) \leq \|d(Z, Z^*)\|_{L^p}$$

We have

$$W_p(\tilde{\Pi}_n(\omega), \tilde{\Pi}_n^*(\omega)) \leq \|d(Z, Z^*)\|_{L^p} = \left(\frac{1}{n}\sum_{i=1}^{n} d(\tilde{X}_i \tilde{X}_i^*)^p\right)^{1/p}$$

and we deduce

$$\begin{aligned} W_p(P_{\Pi_n}, P_{\Pi_n^*}) &\leq \left[\mathbb{E}(W_p(\tilde{\Pi}_n, \tilde{\Pi}_n^*)^p)\right]^{1/p} \\ &\leq \left[\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} d(\tilde{X}_i \tilde{X}_i^*)^p\right)\right]^{1/p} \\ &\leq \left[\mathbb{E}(d(\tilde{X}_i \tilde{X}_i^*)^p)\right]^{1/p} \\ &\leq (1+\varepsilon) W_p(P, P^*). \end{aligned}$$

Letting $\varepsilon \to 0$, we have the result.

Theorem (Tippett-Fisher-Gnedenko)

The following statements are equivalent:

1. *There exist real constants* $a_n > 0$ *and* b_n *such that*

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G_\gamma(x) := \exp(-(1+\gamma x)^{-1/\gamma}),$$

for all x *with* $1 + \gamma x > 0$.

2. *There is a positive function* a *such that for* $x > 0$

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

where the right hand side interpreted as $\log x$ *when* $\gamma = 0$.

U is the quantile function defined as $U(t) = F^{\leftarrow}(1 - \frac{1}{t})$.

The previous theorem is equivalent to the existence of 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 $H_\gamma(x) := 1 - (1+\gamma x)^{-1/\gamma}$
Two questions:

- Can we quantify the approximation of $\mathcal{L}\left(\frac{X-u}{f(u)} | X > u\right)$ by H_γ in Wasserstein metric ?
- Can we prove the consistency an the asymptotic normality of an estimator of γ?

In this talk we focus on the heavy tail case, $\gamma > 0$.
In this case

$$\mathbb{P}(u^{-1}X > x | X > u) \underset{u\to\infty}{\longrightarrow} \mathcal{P}_\alpha(x).$$

Hence we are interested in the the approximation of $\mathcal{L}(u^{-1}X|X > u)$ by $\mathcal{P}_\alpha$ the Pareto distribution of index $\alpha := \gamma^{-1}$.

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

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

between the normalized exceedence distribution and the Pareto distribution equals

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

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

Let $X_1, ..., X_n$ be a i.i.d sample with distribution P_X and $X_1^*, ..., X_k^*$ be a i.i.d sample with $Pareto(\alpha)$-distribution. 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 want to measure the error done approximating the distribution of exceedences above a threshold by a Pareto distribution with the Wasserstein distance between the previous empirical measures.

Theorem

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

$$W_p(P_{\Pi_{n,k}}, P_{\Pi_k^*}) \leq \tilde{A}_p(n,k) := \mathbb{E}\left(A_p\left(\frac{1}{1-T}\right)\right)$$

With $T \sim \beta(n-k, k+1)$.

Under the first order condition and provided $k = k(n)$ satisfies $k/n \to 0$, the upper bound satisfies $\tilde{A}_p(n,k) \longrightarrow 0, \quad$ as $n \to \infty$. If in addition the second order condition holds, we have

$$\tilde{A}_p(n,k) \sim A_p\left(\frac{n}{k}\right), \quad \text{as } n \to \infty.$$

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

Corollary

Let $1 \leq p < \infty$. It holds, for all $1 \leq k \leq n$,

$$W_p(P_{\sqrt{k}(\hat{\gamma}_{n,k}^{Hill}-\gamma)}, \mathcal{N}(0,\gamma^2)) \leq \sqrt{k}\tilde{A}_p(n,k) + \left(4 + 3\sqrt{\frac{2}{\pi}}\right)\frac{\gamma^3}{\sqrt{k}}.$$

As a consequence, the asymptotically normality of the Hill estimator holds as soon as $k = k(n)$ satisfies

$$k \to \infty \quad \text{and} \quad \sqrt{k}\tilde{A}_p(n,k) \to 0.$$

Sketch of proof:

$$\hat{\gamma}_{n,k}^{HILL} = \varphi(\Pi_{n,k}) \quad \text{and} \quad \hat{\gamma}_k^* = \varphi(\Pi_k^*)$$

with $\varphi : \Pi \mapsto \int_1^\infty \log(x)\Pi(\mathrm{d}x)$ which is $1-$ Lipschitz. So

$$W_p(P_{\sqrt{k}(\hat{\gamma}_{n,k}^{Hill}-\gamma)}, P_{\sqrt{k}(\hat{\gamma}_k^*-\gamma)}) \leq \sqrt{k} W_p(P_{\Pi_{n,k}}, P_{\Pi_k^*}) \leq \sqrt{k}\tilde{A}_p(n,k).$$

And the bound

$$W_p(P_{\sqrt{k}(\hat{\gamma}_k^*-\gamma)}, \mathcal{N}(0,\gamma^2)) \leq \left(4 + 3\sqrt{\frac{2}{\pi}}\right)\frac{\gamma^3}{\sqrt{k}}$$

is given by Stein's method.

Regression Problem: Let $(X_i, Y_i)_{1\leq i\leq n}$ are n i.i.d copies of $(X, Y) \in \mathbb{R}^p \times \mathbb{R}$.

We want to estimate the conditional quantile of order $1-\alpha_n$ for Y given $X = x$, with $\alpha_n \to 0$. We want to estimate q_n such that

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

Formally,

$$q_n(x) = F_x^{\leftarrow}(1-\alpha_n)$$

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

Problem: There is few observation at the level $1-\alpha_n$, we need a tail extrapolation.

The Tippett-Fisher-Gnedenko Theorem gives

$$U(tx) \approx U(t) + a(t)\frac{x^\gamma - 1}{\gamma}$$

which give a new way to estimate q_n

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

with $k := k(n)$ a sequence such that $k \to \infty$ and $k/n \to 0$.

- In 2016, Einmahl et al. proposed their heteroscedastic extreme model where the covariate is one dimensional, deterministic and uniformly distributed on $[0, 1]$.
- **Framework:** $(Y_{i,n})_{1 \leq i \leq n}$ are independent random variables with distribution $F_{i,n}$ such that

$$\lim_{y \to \infty} \left| \frac{1 - F_{i,n}(y)}{1 - F(y)} - \sigma(i/n) \right| = 0 \text{ uniformly in } 1 \leq i \leq n$$

  with σ the skedasis function representing the frequency of extremes.
- 2 main goals:
  1. extend the model to a general quantile regression framework
  2. use coupling methods to provide consistency of estimators

Framework:$(X_i, Y_i)_{1\leq i\leq n}$ are n independents 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. The proportional tail

model consist in two main assumptions

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

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

- **Proportional tail assumption**

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

with σ the skedasis function satisfying $\int_{\mathcal{X}} \sigma(x) P_X(\mathrm{d}x) = 1$

The two conditions imply $F_x \in D(G_\gamma)$ with the same γ for all $x \in \mathcal{X}$.

We propose two estimators based on the k-upper order statistics.One for the integrated skedasis function $C(x) = \int_{\{u \leq x\}} \sigma(u) \mathrm{d}u$ and an other for the extreme value index γ.

$$\hat{C}_k(x) := \frac{1}{k} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}} \mathbb{1}_{\{Y_i \geq Y_{n-k:n}\}}$$

and

$$\hat{\gamma}_k := \frac{1}{k} \sum_{i=1}^{k} \log \left(\frac{Y_{n-k+i:n}}{Y_{n-k:n}} \right) \quad \text{(Hill's estimator)}$$

From technical point of view, we need a rate of convergence:

1. $$\sup_{x \in \mathcal{X}} \left| \frac{1 - F_x(y)}{\sigma(x)(1 - F(y))} - 1 \right| = \underset{y \to \infty}{O} \left(A \left(\frac{1}{1 - F(y)} \right) \right)$$

2. $$\sup_{z > 1/2} \left| \frac{1 - F_x(zy)}{z^{-\alpha}(1 - F(y))} - 1 \right| = \underset{y \to \infty}{O} \left(A \left(\frac{1}{1 - F(y)} \right) \right)$$

with A decreasing to 0.

Theorem

If $k_n \to +\infty$, $\frac{k_n}{n} \to 0$ *and* $\sqrt{k_n}A(\frac{n}{k_n}) \to 0$, *we have:*

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

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

Definition (Total variation norm)

Let P_1 and P_2 two probability measures

$$||P_1 - P_2||_{TV} = \sup_{A \text{ Borel}} |P_1(A) - P_2(A)|.$$

Lemma (Maximal coupling)

Given P_1, P_2 two probability measures on $\mathbb{R}^p$ space. One can construct random variables $X_1 \sim P_1$ and $X_2 \sim P_2$ such that

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

Lemma

Under the proportional tail model assumption, we have

$$||P_{X|Y>y} - \sigma(x)P_X(\mathrm{d}x)||_{TV} = \underset{y\to\infty}{O}\left(A\left(\frac{1}{1-F(y)}\right)\right).$$

Let $((X_1, Y_1), ..., (X_n, Y_n))$ random variable with distribution $P_{X,Y}$ and $((X_1^*, Y_1^*), ..., (X_n^*, Y_n^*))$ i.i.d with distribution $\sigma(x)P_X(\mathrm{d}x) \otimes \text{Pareto}(\alpha)$. Define the empirical measures

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

Endow $\mathbb{R}^p \times [1, +\infty)$ with the metric

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

Theorem

For $1 \leq p < \infty$ and $1 \leq k \leq n$ we have,

$$W_p(P_{\Pi_k}, P_{\Pi_k^*}) \leq MA\left(\frac{n}{k}\right)$$

Let φ the map from $(\mathcal{W}_p, W_p)$ to $L^\infty \times \mathbb{R}$ defined by

$$\varphi(\Pi) := \left(t \mapsto \int_{\mathbb{R}^p \times [1,+\infty)} \mathbb{1}_{x \leq t} \Pi(\mathrm{d}x, \mathrm{d}y), \int_{\mathbb{R}^p \times [1,+\infty)} \log(y) \Pi(\mathrm{d}x, \mathrm{d}y)\right).$$

φ is $1-$ Lipschitz and we have identities:

$$\varphi(\Pi_k) = (\hat{C}_n, \hat{\gamma}_k) \quad \text{and} \quad \varphi(\Pi_k^*) = (\hat{C}_n^*, \hat{\gamma}_k^*).$$

Theorem

For $1 \leq p < \infty$,

$$W_p(P_{\sqrt{k}(\hat{C}_n - C, \hat{\gamma} - \gamma)}, P_{\sqrt{k}(\hat{C}_n^* - C, \hat{\gamma}_k^* - \gamma)}) \leq \sqrt{k} MA\left(\frac{n}{k}\right)$$

Since (X_i^*, Y_i^*) are i.i.d, the weak convergence of $\sqrt{k}(\hat{C}_n^* - C, \hat{\gamma}_k^* - \gamma)$ is well known:

- $\sqrt{k}(\hat{C}_n^* - C) \xrightarrow{\mathcal{L}} B$ with B a C-Brownian bridge due to Donsker theorem.
- $\sqrt{k}(\hat{\gamma}_k^* - \gamma) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \gamma^2)$ is consequence of central limite theorem for exponnential random variables.
- $\hat{C}_n^*$ and $\hat{\gamma}_k^*$ are independents as X^* and Y^* are.

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

$$q_n(x) = F_x^{\leftarrow}(1-\alpha_n) \approx F^{\leftarrow}\left(1-\frac{\alpha_n}{\sigma(x)}\right) \approx U\left(\frac{n}{k}\right)+a\left(\frac{n}{k}\right)\frac{\left(\frac{\hat{\sigma}(x)k}{\alpha_n n}\right)^{\gamma}-1}{\gamma}.$$

We now have two ways to estimate q_n:

- Approximation by the inverse empirical distribution

$$\hat{q}_n(x) := \mathbb{F}_n^{\leftarrow}\left(1-\frac{\alpha_n}{\hat{\sigma}(x)}\right).$$

- GPD approximation

$$\hat{q}_n(x) := Y_{n-k:n} + \hat{a}\left(\frac{n}{k}\right)\frac{\left(\frac{\hat{\sigma}(x)}{\alpha_n}\frac{k}{n}\right)^{\hat{\gamma}_k}-1}{\hat{\gamma}_k}.$$

Future work:

- Prove the consistency and asymptotic normality of $\hat{\sigma}(x)$

$$\hat{\sigma}_k(x) := \frac{\sum_{i=1}^{n} \mathbb{1}_{\{|X_i - x| \leq \frac{h}{2}\}} \mathbb{1}_{\{Y_i > Y_{n-k:n}\}}}{\frac{k}{n} \sum_{i=1}^{n} \mathbb{1}_{\{|X_i - x| \leq \frac{h}{2}\}}}$$

  with h and k both dependents on n.

- Deduce the properties of $\hat{q}_n(x)$

$$\hat{q}_n(x) := Y_{n-k:n} + \hat{a}\left(\frac{n}{k}\right) \frac{\left(\frac{\hat{\sigma}(x)}{\alpha_n} \frac{k}{n}\right)^{\hat{\gamma}_k} - 1}{\hat{\gamma}_k}.$$

- Use the estimator of the integrated skedasis function C to build a test for model validation

Thank you for your attention!