

# MÉTHODE DE COUPLAGE EN DISTANCE DE WASSERSTEIN POUR LA THÉORIE DES VALEURS EXTRÊMES

Benjamin Bobbia <sup>1</sup> & Clément Dombry <sup>2</sup> & Davit Varron <sup>3</sup>

<sup>1</sup> *Laboratoire de Mathématiques, Université de Franche-Comté, 16 route de Gray, 25030  
Besançon, benjamin.bobbia@univ-fcomre.fr*

<sup>2</sup> *Laboratoire de Mathématiques, Université de Franche-Comté, 16 route de Gray, 25030  
Besançon, clement.dombry@univ-fcomte.fr*

<sup>3</sup> *Laboratoire de Mathématiques, Université de Franche-Comté, 16 route de Gray, 25030  
Besançon, davit.varron@gmail.com*

**Résumé.** Nous proposons une relecture de résultats classiques de la théorie des valeurs extrêmes, que nous étudions grâce aux outils que nous fournit la théorie du transport optimal. Dans ce cadre, nous pouvons voir la normalité des estimateurs comme une convergence de mesures dans un espace métrique muni de la distance de Wasserstein. Il s'agit d'une approche par couplage. Soient  $(X_1, \dots, X_n)$  et  $(X_1^*, \dots, X_n^*)$  deux échantillons i.i.d, nous nous intéressons aux relations qui lient la distance de Wasserstein entre les mesures empiriques sur les échantillons et la distance de Wasserstein entre les lois qui ont généré ces échantillons. Ce résultat nous permet de redémontrer la normalité du célèbre estimateur de Hill et de donner une vitesse de convergence.

**Mots-clés.** Valeurs extrêmes, distance de Wasserstein, mesure aléatoire, processus empiriques.

**Abstract.** We propose a relecture of classical results of extreme value theory with the tools given by the optimal transport theory. The asymptotic normality of estimators is seen as convergence of measures in a metric space endowed by the Wasserstein metric. This approach relies on coupling method. Let two i.i.d samples  $(X_1, \dots, X_n)$  and  $(X_1^*, \dots, X_n^*)$ , we are interested in the investigation about relations between, the Wasserstein distance between the empiricals measures on the two samples, and the Wasserstein distance between the distribution which generate the two samples. We apply this result to show the normality of the well known Hill's estimator and give a rate of convergence.

**Keywords.** Extreme value, wasserstein's distance, random measure, empiricals processes.

## 1 Généralités sur les espaces de Wasserstein

### 1.1 La métrique de Wasserstein

Soit  $(\mathcal{X}, d)$  un espace métrique muni de sa tribu Borélienne. On désignera par  $\mathcal{M}(\mathcal{X})$  l'ensemble des mesures de probabilités sur  $\mathcal{X}$  et on notera  $\delta_x$  la mesure de Dirac en  $x$ .

**Définition 1.** On définit un couplage entre deux mesures de probabilités  $P_1$  et  $P_2$  dans  $\mathcal{M}(\mathcal{X})$  comme une paire de variables aléatoires à valeur dans  $\mathcal{X}$ , définies sur le même espace probabilisé, telle que  $X_1 \sim P_1$  et  $X_2 \sim P_2$ .

Nous pouvons alors définir la distance de Wasserstein d'ordre  $p \in [1, +\infty]$  entre deux mesures  $P_1, P_2 \in \mathcal{M}(\mathcal{X})$  par

$$W_p(P_1, P_2) = \inf \{ \|d(X_1, X_2)\|_{L^p} : (X_1, X_2) \text{ couplage entre } P_1 \text{ et } P_2 \},$$

avec  $\|Z\|_{L^p}$  la norme  $L^p$  de la variable aléatoire réelle  $Z$ ,

$$\|Z\|_{L^p} = \begin{cases} (\mathbb{E}[|Z|^p])^{1/p} & \text{pour } 1 \leq p < \infty \\ \text{ess sup } |Z| & \text{pour } p = \infty \end{cases}.$$

Dans le cas simple des mesures de Dirac, nous avons pour  $x_1$  et  $x_2$  dans  $\mathcal{X}$ .

$$W_p(\delta_{x_1}, \delta_{x_2}) = d(x_1, x_2).$$

Dans le cas où  $\mathcal{X}$  est la droite réelle, Ambrosio et al.(2008) donnent une formulation explicite de la distance de Wasserstein.

$$d_{W_p(\mathbb{R})}(P_1, P_2) = \left( \int_0^1 |F_1^{\leftarrow}(x) - F_2^{\leftarrow}(x)|^p dx \right)^{\frac{1}{p}}$$

où  $F_1$  et  $F_2$  désignent respectivement les fonctions de répartitions de  $P_1$  et  $P_2$ .

## 1.2 Les espaces de Wasserstein

Nous pouvons définir un sous espace de  $\mathcal{M}(\mathcal{X})$  par

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

pour un certain  $x_0 \in \mathcal{X}$  fixé. Il s'agit de l'espace de Wasserstein d'ordre  $p$  sur  $(\mathcal{X}, d)$ . Le choix de l'origine  $x_0$  n'a aucun impact sur la définition de l'espace de Wasserstein. En effet, pour tout point  $x$  de  $\mathcal{X}$  l'inégalité triangulaire nous fournit  $W_p(P, \delta_{x_0}) < \infty$  si et seulement si  $W_p(P, \delta_x) < \infty$ .

Dans le cas particulier où  $\mathcal{X} = \mathbb{R}^d$  muni de sa norme euclidienne, l'espace de Wasserstein d'ordre  $p$  est simplement l'espace des mesures de probabilités qui admettent un moment d'ordre  $p$  fini. C'est-à-dire

$$\mathcal{W}_p(\mathbb{R}^d) = \left\{ P \in \mathcal{M}_1(\mathbb{R}^d) : \int \|x\|^p P(dx) < \infty \right\}.$$

Villani (2009) détaille les intérêts de ces espaces. D'une part, il est possible de comprendre la convergence en loi de variables aléatoires à valeurs dans  $\mathcal{X}$  comme la convergence dans

$\mathcal{W}_p$  au sens de la distance  $W_p$ . Plus précisément, si une suite de mesures  $P_k$  converge faiblement dans  $\mathcal{W}_p$  vers une mesure  $p$  alors  $W_p(P_k, P)$  converge vers 0. La réciproque est également vraie.

D'autre part certaines propriétés de  $\mathcal{X}$  sont transférées à son espace de Wasserstein. Si  $\mathcal{X}$  est séparable et complet,  $\mathcal{W}_p(\mathcal{X})$  l'est aussi. Il en va de même pour la compacité.

## 2 Méthode de couplage en théorie des valeurs extrêmes

### 2.1 Distribution empirique et distance de Wasserstein

Considérons deux échantillons indépendants et identiquement distribués  $X_1, \dots, X_n$  et  $X_1^*, \dots, X_n^*$ ,  $n \geq 1$ , prenant leurs valeurs dans  $(\mathcal{X}, d)$  et admettant pour distributions  $P_X$  et  $P_{X^*} \in \mathcal{W}_p(\mathcal{X})$ ,  $p \in [1, \infty]$  respectivement. Nous pouvons définir les mesures empiriques correspondantes par

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

Ces mesures peuvent être vues comme des éléments aléatoires de  $\mathcal{W}_p(\mathcal{X})$ ,  $p \in [1, \infty]$ . On note  $P_{\Pi_n}$  et  $P_{\Pi_n^*}$  leurs distributions. On note  $W_p(P_{\Pi_n}, P_{\Pi_n^*})$  leur distance de Wasserstein dans l'espace  $\mathcal{W}_p(\mathcal{W}_p(\mathcal{X}))$ . Dans un souci de clarté des énoncés, nous utiliserons la même notation pour la distance de Wasserstein que ce soit sur  $\mathcal{W}_p(\mathcal{X})$  ou sur  $\mathcal{W}_p(\mathcal{W}_p(\mathcal{X}))$ .

**Théorème 1.** *Soit  $p \in [1, \infty]$ . Si  $W_p(P_X, P_{X^*}) < \infty$ , alors*

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

*Si, de plus,  $(\mathcal{X}, d)$  est séparable et complet et  $p < \infty$ , alors*

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

La preuve de ce théorème repose essentiellement sur la formule de dualité de Kantorovich-Rubinstein fournie par Villani (2009) qui affirme que pour deux mesures  $P_1$  et  $P_2$  dans  $\mathcal{W}_1(\mathcal{X})$  avec  $\mathcal{X}$  séparable et complet on a

$$W_p(P_1, P_2) = \sup \left( \int_{\mathcal{X}} \varphi dP_1 - \int_{\mathcal{X}} \varphi dP_2 \right)$$

où le supremum est pris sur toutes les fonction  $\varphi$  qui sont 1-Lipschitz. On peut remarquer que même si l'énoncé original demandait que  $\mathcal{X}$  soit polonais, cette dualité reste vraie tant que le second membre est fini.

## 2.2 Approximation en Wasserstein dans la méthode "peaks-over-threshold"

Ici, nous nous intéressons aux excès d'une loi  $P_X$  au dessus d'un seuil élevé que nous appellerons  $u$ . La théorie classique des valeurs extrêmes que l'on peut trouver dans le livre de de Haan et Ferreira (2006) nous affirme que, sous les hypothèses du premier ordre, la loi des excès converge vers une distribution de Pareto généralisée.

Nous nous concentrons ici sur le cas où l'indice des valeurs extrême  $\gamma$  de la loi de  $P_X$  est strictement positif. Ce qui permet de simplifier les conditions du premier ordre en

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

avec  $U(t) = F^{\leftarrow}(1 - 1/t)$ . Et la loi des excès converge alors vers une loi de Pareto.

$$P_{u^{-1}X|X>u} \xrightarrow{d} P_\alpha, \quad \text{pour } u \rightarrow \infty,$$

avec  $P_\alpha(x) = x^{-\alpha}$ ,  $x > 1$ , la distribution de Pareto d'indice  $\alpha = 1/\gamma > 0$ . Comme la distribution de Pareto a pour support  $[1, +\infty)$  et des moments d'ordre  $p < \alpha$  fini, il est naturel d'introduire la distance logarithmique

$$d(x, x') = |\log(x) - \log(x')|, \quad x, x' \in [1, \infty[.$$

De plus, dans le but d'étudier les vitesses de convergence, il est habituel d'introduire les conditions du second ordre qui s'expriment comme suit

$$\lim_{t \rightarrow +\infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho}, \quad x > 0$$

avec le paramètre  $\rho < 0$  et la fonction  $A$  positive, à variations régulières d'indice  $\rho$  à l'infini et tendant vers 0.

**Proposition 1.** *Sur  $\mathcal{X} = [1, \infty)$  muni de la distance logarithmique  $d$ , la distance de Wasserstein*

$$A_p(t) := d_{\mathcal{W}_p([1, \infty))}(P_{U(t)^{-1}X|X>U(t)}, P_\alpha)$$

*entre la distribution des excès normalisés et la distribution de Pareto est donnée par*

$$A_p(t) = \begin{cases} \left( \int_1^\infty \left| \log \frac{U(t'x)}{x^\gamma U(t')} \right|^p \frac{dx}{x^2} \right)^{1/p} & \text{for } 1 \leq p < \infty \\ \sup_{x>1} \left| \log \frac{U(t'x)}{x^\gamma U(t')} \right| & \text{for } p = \infty \end{cases}$$

avec  $t' = 1/(1 - F(U(t)))$ . De plus sous les conditions du second ordre, on a

$$\lim_{t \rightarrow +\infty} \frac{A_p(t)}{A(t)} = \begin{cases} \left( \int_1^\infty \left| \frac{x^\rho - 1}{\rho} \right|^p \frac{dx}{x^2} \right)^{1/p} & \text{si } p \in [1, \infty[, \\ \frac{1}{|\rho|} & \text{si } p = \infty \text{ et } \rho < 0, \\ +\infty & \text{si } p = \infty \text{ et } \rho = 0. \end{cases}$$

Si la fonction de répartition de  $X$  est continue, on peut alors choisir  $t' = t$ . Dans tous les cas,  $t$  et  $t'$  sont équivalents en l'infini et  $U(t) = U(t')$ .

Sous les conditions du premier ordre,  $A_p$  est continue par morceaux sur  $]1, +\infty[$  et admet un nombre fini de discontinuités qui sont toutes de première espèce. Ainsi  $A_p$  est bornée sur tout intervalle de la forme  $[1 + \varepsilon, +\infty[$  avec  $\varepsilon > 0$ . De plus,  $A_p$  tend vers 0 en l'infini.

### 2.3 Application à l'estimateur de Hill

Soit  $(X_1, \dots, X_n)$  un échantillon de loi  $P_X$  d'indice des valeurs extrêmes  $\gamma$ . Nous souhaitons appliquer la méthode de couplage pour estimer  $\gamma$ . Nous utilisons l'estimateur de Hill basé sur les  $k$  plus grandes statistiques d'ordre. Dans le cas  $\gamma > 0$  l'estimateur de Hill s'exprime

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

Pour prouver la convergence de cet estimateur, nous le regarderons sur le modèle Pareto limite et nous contrôlerons l'erreur grâce à la distance de Wasserstein entre la mesure empirique des excès

$$\Pi_{n,k} = \frac{1}{k} \sum_{i=1}^n \delta_{\frac{X_{n+1-i;n}}{X_{n-k;n}}}$$

et la mesure empirique d'un échantillon  $(X_1^*, \dots, X_n^*)$  i.i.d de loi Pareto de paramètre  $\alpha = 1/\gamma$

$$\Pi_k^* = \frac{1}{k} \sum_{i=1}^n \delta_{X_i^*}.$$

Montrer la normalité asymptotique de  $\gamma_{n,k}^{HILL}$  revient à montrer la convergence vers 0 de

$$W_p(P_{\sqrt{k}(\hat{\gamma}_{n,k}^{HILL} - \gamma)}, \mathcal{N}(0, \gamma^2)).$$

Avec  $\hat{\gamma}_k^*$  l'estimateur de Hill défini sur l'échantillon Pareto  $(X_1^*, \dots, X_n^*)$ .

Ce qui revient, par inégalité triangulaire, à contrôler séparément les deux termes

$$W_p(P_{\sqrt{k}(\hat{\gamma}_k^* - \gamma)}, \mathcal{N}(0, \gamma^2)) \quad \text{et} \quad W_p(P_{\sqrt{k}(\hat{\gamma}_{n,k}^{HILL})}, P_{\sqrt{k}(\hat{\gamma}_k^* - \gamma)}).$$

Le premier terme peut facilement être majoré par la méthode de Stein détaillée par Ross (2005). Alors que le second peut être contrôlé grâce au couplage.

**Théorème 2.** *Soit  $\mathcal{X} = [1, +\infty[$  muni de la distance logarithmique. Alors pour  $p \in [1, +\infty[$  et  $1 \leq k \leq n$  on a*

$$W_p(P_{\Pi_{n,k}}, P_{\Pi_k^*}) \leq \tilde{A}_p(n, k) := \left( \int_0^\infty \int_0^\infty \left| \log \frac{U(tz)}{z^\gamma U(t)} \right| \beta_{n-k, k+1} \left( 1 - \frac{1}{t} \right) \frac{dt dz}{t^2 z^2} \right)^{1/p}$$

Ces résultats nous fournissent la borne suivante

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

On a donc la normalité de l'estimateur de Hill dès que  $k = k(n)$  vérifie les conditions classiques en extrême

$$k \rightarrow \infty, \quad \frac{k}{n} \rightarrow 0 \quad \text{et} \quad \sqrt{k}\tilde{A}_p n, k \rightarrow 0.$$

Ce théorème énoncé comme tel ne permet, en réalité, pas de conclure. En effet la convergence vers 0 de  $\tilde{A}_p n, k$  n'est pas assurée de façon théorique. En effet, pour avoir convergence de la borne il faut s'assurer que  $X$  soit "loin" de 1. On peut donc soit établir le théorème sur  $\mathcal{X}[1 + \varepsilon, +\infty[$  ou alors s'intéresser à  $W_p(P_{\Pi_{n,k}|X \geq 1+\varepsilon}, P_{\Pi_k^*})$ . Dans la pratique, cela ne pose aucun problème, car nous nous intéressons à des valeurs extrêmes pour  $x$  qui seront, de fait, loin de 1.

L'estimation de l'indice des valeurs est un premier pas dans l'utilisation de telles méthodes de couplage en théorie des valeurs extrêmes. Elle nous permet, dans un travail en cours, de généraliser le modèle des extrêmes hétéroscedastiques de Einmahl, de Haan et Zhou (2016).

## Bibliographie

- De Haan, L. et Ferreira, A. (2006), *Extreme value theory*, Springer, New York.
- Villani, C. (2009), *Optimal transport*, Springer-Verlag, Berlin.
- Ambrosio, L. et Gigli, N. et Savaré, G. (2008), *Gradient flow in Metric spaces and in the Space of Probability Measures*, Springer, Berlin.
- Einmahl, J.H.J. et de Haan, L. et Zhou, C. (2016). Statistics of heteroscedastic extremes, *Journal of the Royal Society. Serie B. Statistical Methodology*, 78, pp. 31-51.
- Ross, N. (2011). Fundamentals of stein's method, *ArXiv*.