



HAL
open science

Projection de mesures de probabilité sous contraintes de quantile par distance de Wasserstein et approximation monotone polynomiale

Marouane Il Idrissi, Nicolas Bousquet, Fabrice Gamboa, Bertrand Iooss,
Jean-Michel Loubes

► To cite this version:

Marouane Il Idrissi, Nicolas Bousquet, Fabrice Gamboa, Bertrand Iooss, Jean-Michel Loubes. Projection de mesures de probabilité sous contraintes de quantile par distance de Wasserstein et approximation monotone polynomiale. 53èmes Journées de Statistique de la Société Française de Statistique (SFdS), Jun 2022, Lyon, France. hal-03597059

HAL Id: hal-03597059

<https://hal.science/hal-03597059>

Submitted on 4 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PROJECTION DE MESURES DE PROBABILITÉ SOUS CONTRAINTES DE QUANTILE PAR DISTANCE DE WASSERSTEIN ET APPROXIMATION MONOTONE POLYNOMIALE

Marouane Il Idrissi ^{1,2,3,*}, Nicolas Bousquet ^{1,3,4}, Fabrice Gamboa ², Bertrand Iooss ^{1,2,3},
Jean-Michel Loubes ²

¹ EDF R&D, 6 Quai Watier, 78400 Chatou, France

² Institut de Mathématiques de Toulouse, 31062, Toulouse, France

³ SINCLAIR AI Laboratory, Saclay, France

⁴ Sorbonne Université, LPSM, 4 place Jussieu, Paris, France

* Auteur correspondant - marouane.il-idrissi@edf.fr

Résumé. Motivés par des applications en analyse de robustesse pour la quantification d'incertitude et l'apprentissage statistique, nous nous intéressons au problème de projection d'une mesure de probabilité univariée donnée. L'ensemble de projection est composé de mesures de probabilité respectant certaines contraintes sur leur quantiles. Ce travail explore le cas de la distance de Wasserstein pour des mesures de probabilité portées par \mathbb{R} . Ce problème particulier de projection dans des espaces de mesures sous contraintes quantile se réécrit comme un problème de projection dans $L^2([0, 1])$ avec contraintes de monotonie et d'interpolation. Ce dernier problème se résout analytiquement. Nous présentons également une régularisation de la solution en nous restreignant aux fonctions polynomiales par morceaux. Finalement, nous illustrons ces résultats sur des données réelles.

Mots-clés. Projection de mesures de probabilité, quantiles, transport optimal, approximation monotone.

Abstract. Motivated by applications in robustness analysis for uncertainty quantification and statistical learning, this paper focuses on the problem of univariate probability measure projection, under quantile constraints. The choice of the Wasserstein distance is explored. The initial problem can be equivalently written as a projection in $L^2([0, 1])$ with interpolation and monotonicity constraints, and can be solved analytically. We leverage this equivalence to ensure some smoothness in the solution, by further restricting the solution to be a continuous piece-wise non-decreasing polynomial. Finally, some illustrations on a dataset are presented.

Keywords. Probability measure projection, quantiles, optimal transport, isotonic approximation.

1 Projection W_2 sous contraintes quantile

Soit P une mesure de probabilité supportée par $\mathcal{X} \subseteq \mathbb{R}$. Sans distinction dans la suite, P peut être une mesure de probabilité *empirique*, i.e., composée d'atomes issus d'un n -

échantillon réel noté $X = (X_1, \dots, X_n)$, ou une mesure de probabilité *continue*, i.e., qui ne compte aucun atome, et supposée appartenir à l'ensemble des mesures de probabilités réelles de moment d'ordre 2 fini, noté $\mathcal{P}_2(\mathbb{R})$. La fonction de répartition de P est notée $F_P(x) = P((-\infty, x])$ et on note $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$. Nous pouvons définir deux inverses généralisées de F_P : son inverse à gauche, définie pour tout $y \in [0, 1]$ par

$$F_P^{\leftarrow}(y) := \sup \{x \in \overline{\mathbb{R}} \mid F_P(x) < y\} = \inf \{x \in \overline{\mathbb{R}} \mid F_P(x) \geq y\}$$

qui est croissante et continue à gauche, et qui correspond à la fonction quantile classique associée à P . Son inverse à droite, définie pour tout $y \in [0, 1]$ par

$$F_P^{\rightarrow}(y) := \sup \{x \in \overline{\mathbb{R}} \mid F_P(x) \leq y\} = \inf \{x \in \overline{\mathbb{R}} \mid F_P(x) > y\}$$

qui est croissante, continue à droite, et est égale presque partout à F_P^{\leftarrow} . Notons également l'espace fonctionnel

$$\mathcal{F}^{\rightarrow} = \{G : [0, 1] \rightarrow \overline{\mathbb{R}} \mid G \text{ est continue à droite, croissante, } \lim_{y \rightarrow 0} G(y) = -\infty, \lim_{y \rightarrow 1} G(y) = +\infty\}.$$

$\mathcal{F}^{\rightarrow}$ est l'espace des inverses généralisées *valides*, i.e., chaque fonction dans $\mathcal{F}^{\rightarrow}$ est l'unique inverse généralisée à droite d'une fonction de répartition, associée à une unique mesure de probabilité. La Figure 1 illustre les différences entre ces deux inverses généralisées : elles ne diffèrent qu'aux points dont l'image est un atome de P (de La Fortelle 2015).

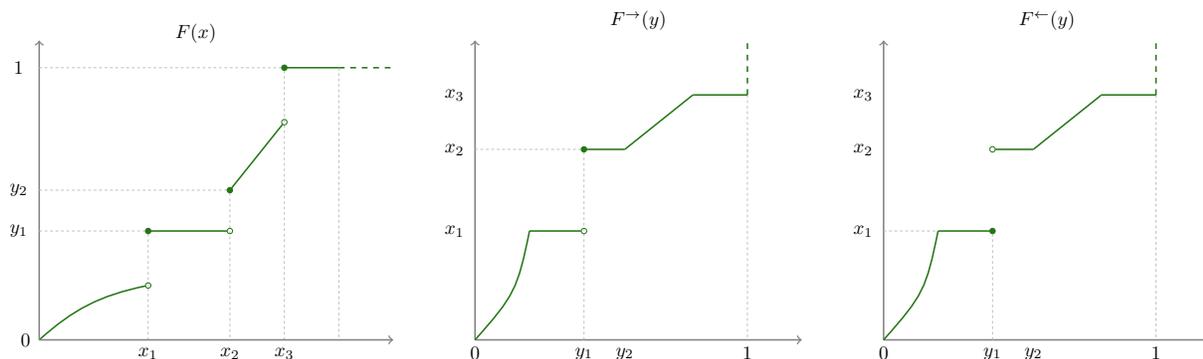


Figure 1: Une fonction de répartition et ses inverses généralisées.

Notre intérêt pour une perturbation des quantiles de F_P est issu d'une réflexion plus large sur la nécessité, dans des applications variées en quantification d'incertitude ou en apprentissage statistique, de tester la robustesse d'une prévision dépendante de X vis-à-vis de la loi de X , que celle-ci soit connue explicitement (Da Veiga et al. 2021), ou au travers de réalisations (Bachoc et al. 2020). Ce type de perturbation peut s'interpréter plus aisément que des contraintes usuelles de type moment généralisé. Plus précisément, nous imposons des contraintes sur les seuils de quantile, par essence bornés sur $[0, 1]$.

Afin de définir de telles contraintes, prenons plusieurs niveaux de probabilités $\alpha = (\alpha_1, \dots, \alpha_K)^\top$ et $\beta = (\beta_1, \dots, \beta_K)^\top$ avec $\alpha, \beta \in (0, 1)^K$, où $K \in \mathbb{N}^*$ représente le nombre de contraintes quantiles souhaitées. Afin d'être valides, les niveaux doivent respecter la condition

$$\beta_1 < \beta_2 < \dots < \beta_K, \quad \text{et} \quad \alpha_1 < \alpha_2 < \dots < \alpha_K.$$

Pour $i = 1, \dots, K$, dans le cas où $\alpha_i = \beta_i$, nous parlerons de *conservation de quantiles*, et si $\alpha_i \neq \beta_i$, nous parlerons de *perturbation de quantile*. Une mesure de probabilité Q respecte une contrainte quantile lorsque

$$\begin{cases} F_Q^\rightarrow(\beta) = F_P^\leftarrow(\alpha) & \text{si } \beta < \alpha, \\ F_Q^\leftarrow(\beta) = F_P^\leftarrow(\alpha) & \text{si } \beta \geq \alpha. \end{cases}$$

Ainsi, pour une mesure P , et des niveaux valides fixés, notons \mathcal{C} la restriction de $\mathcal{P}_2(\mathbb{R})$ des mesures de probabilités respectant ces contraintes quantiles. Nous nous intéressons au problème de projection suivant

$$Q = \operatorname{argmin}_{G \in \mathcal{P}_2(\mathbb{R})} W_2^2(P, G) \quad \text{s.c.} \quad G \in \mathcal{C}, \quad (1)$$

où W_2 dénote la distance de Wasserstein d'ordre 2. Essentiellement, nous recherchons une mesure de probabilité Q respectant les contraintes quantiles, et étant la plus proche de la mesure de probabilité initiale P , au sens de la distance de Wasserstein d'ordre 2. A l'aide de la reformulation de la distance de Wasserstein dans le cas particulier des mesures réelles (Villani 2003), et par unicité entre les fonctions de répartition et leurs inverses généralisées, ce problème de projection se réécrit de la manière suivante :

$$\begin{aligned} F_Q^\rightarrow = \operatorname{argmin}_{L \in \mathcal{F}^\rightarrow} & \left\{ \int_0^1 (F_P^\rightarrow(x) - L(x))^2 dx \right\} \\ \text{s.c.} & \begin{cases} L(\beta_i) = F_P^\leftarrow(\alpha_i) & \text{si } \beta_i < \alpha_i, \\ L(\beta_i^-) = F_P^\leftarrow(\alpha_i) & \text{si } \beta_i \geq \alpha_i, \end{cases} \quad i = 1, \dots, K, \end{aligned} \quad (2)$$

où F_Q^\rightarrow est définie sur la mesure de probabilité Q , minimisant le problème de projection introduit en (1). Cette formulation équivalente peut être résolue de manière analytique : soient $a, b \in (0, 1)^K$ définis par:

$$\begin{aligned} a_1 &= \min(\alpha_1, \beta_1), & a_i &= \min\left[\max(\beta_{i-1}, \alpha_i), \beta_i\right], \quad i = 2, \dots, K, \\ b_K &= \max(\alpha_K, \beta_K), & b_j &= \max\left[\min(\alpha_j, \beta_{j+1}), \beta_j\right], \quad j = 1, \dots, K-1, \end{aligned}$$

et notons $A_i = [a_i, b_i)$, et $\bar{A} = [0, 1] \setminus \left(\bigcup_{i=1}^K A_i\right)$. Alors, la solution du problème (2) s'écrit

$$F_Q^\rightarrow(x) = \begin{cases} F_P^\rightarrow(x) & \text{si } x \in \bar{A}, \\ F_P^\leftarrow(\alpha_i) & \text{si } x \in A_i, \quad i = 1, \dots, K. \end{cases}$$

En d'autres termes, Q diffère de P dans le sens où elle présente des atomes aux points $F_P^{\leftarrow}(\alpha_i)$, de masse $Q(\{F_P^{\leftarrow}(\alpha_i)\}) = b_i - a_i$, pour $i = 1, \dots, K$.

2 Polynômes monotones continus par morceaux

La reformulation (2) permet également de mettre en avant le lien entre projection de mesures de probabilité par W_2 sous contraintes quantiles et les projections de fonctions dans $L^2([0, 1])$ sous contraintes conjointes d'interpolation et de monotonie. La solution peut être lissée en restreignant l'espace de recherche $\mathcal{F}^{\rightarrow}$ à des fonctions plus régulières.

Considérons une restriction de $\mathcal{F}^{\rightarrow}$ aux fonctions continues polynomiales par morceaux. Soit $\alpha, \beta \in (0, 1)^K$, notons $d \in \mathbb{N}^*$ un degré de polynôme et intéressons-nous aux fonctions de la forme :

$$G(x) = \begin{cases} G_0(x) & \text{si } 0 \leq x < \beta_1, \\ \vdots & \\ G_i(x) & \text{si } \beta_i \leq x < \beta_{i+1}, \\ \vdots & \\ G_K(x) & \text{si } \beta_K \leq x \leq 1, \end{cases} \quad (3)$$

où chaque G_i est un polynôme de degré inférieur ou égal à d . Notons \mathcal{G} la famille paramétrique des fonctions s'écrivant comme (3). Supposons maintenant que le support de P soit un compact $[a, b] \subset \mathbb{R}$, avec $a < b$ deux réels donnés. Le problème de projection de P par distance W_2 sur $\mathcal{G} \cap \mathcal{F}^{\rightarrow}$ peut s'écrire de manière équivalente comme

$$S = \operatorname{argmin}_{G \in \mathcal{G}} \left\{ \int_0^1 (F_P^{\rightarrow}(x) - G(x))^2 dx \right\} \quad \text{s.c.} \quad \begin{cases} G(0) = a, & G(1) = b, \\ G_i(\beta_i) = G_{i+1}(\beta_i) = F_P^{\leftarrow}(\alpha_i) \\ \quad \text{pour } i = 0, \dots, K-1, \\ G'(x) \geq 0, \forall x \in [0, 1]. \end{cases} \quad (4)$$

Ce problème initial est séparable sur chaque intervalle et chaque sous-problème peut être reformulé comme un problème de recherche des coefficients de chaque G_i sur l'intervalle correspondant. De manière générique, cela revient à résoudre le problème quadratique convexe, avec les contraintes convexes suivantes, pour un intervalle $[\beta_i, \beta_{i+1}]$, $i = 0, \dots, K$ (où $\beta_0 = 0$, et $\beta_{K+1} = 1$) :

$$s = \operatorname{argmin}_{g \in \mathbb{R}^{d+1}} g^\top M g - 2g^\top r \quad \text{s.c.} \quad \begin{cases} g^\top \tilde{\beta}_i = F_P^{\leftarrow}(\alpha_i), & g^\top \tilde{\beta}_{i+1} = F_P^{\leftarrow}(\alpha_{i+1}), \\ g \in \mathcal{K}, \end{cases}$$

où $\tilde{\beta} = (1, \beta, \beta^2, \dots, \beta^d)^\top$, M représente la matrice $((d+1) \times (d+1))$ des moments de la mesure de Lebesgue sur l'intervalle considéré, i.e.,

$$M_{ij} = \int_{\beta_i}^{\beta_{i+1}} x^{i+j-2} dx = \frac{(\beta_{i+1})^{i+j-1} - (\beta_i)^{i+j-1}}{i+j-1},$$

et r est le vecteur des moments de F_P^\rightarrow , i.e., pour $i = 0, \dots, d$, $r_i = \int_{t_0}^{t_1} x^i A(x) dx$. L'ensemble \mathcal{K} représente une sous-partie convexe de \mathbb{R}^{d+1} , étant l'image d'une composition d'applications linéaires, notée $T : \mathbb{S}_k \times \mathbb{S}_l \rightarrow \mathbb{R}^{d+1}$, pouvant s'écrire :

$$\mathcal{K} = \{g \in \mathbb{R}^{d+1} \mid T(\Gamma) = g, \quad \Gamma \in \Sigma_k \times \Sigma_l\}$$

où \mathbb{S}_k représente le sous-espace vectoriel des matrices $(k \times k)$ symétriques et Σ_k représente le cône convexe fermé des matrices semi-définies positives dans \mathbb{S}_k . Les tailles k, l des matrices dépendent de la parité de d : si d est pair, $k = d$ et $l = d - 2$; si d est impair, alors $k = d - 1$ et $l = d - 1$. Ce résultat est issu de la représentation successive de la dérivée du polynôme $G_i(x)$ par le biais de polynômes *sum-of-squares* (SOS) (Lasserre 2015), garantissant sa positivité sur l'intervalle considéré, puis de la représentation des polynômes SOS par le biais de matrices symétriques semi-définies positives (Parrilo 2010, 2012). Ce problème étant convexe, il peut être résolu par le biais d'outils de résolution numérique, tel CVXR, une interface en langage R de l'outil CVX (Grant and Boyd 2014).

3 Approximation continue de mesure empirique : exemple des données d'Ozone

Afin d'illustrer nos résultats, nous allons nous intéresser aux données `airquality`, et en particulier aux mesures journalières d'Ozone à New York entre Mai et Septembre 1973. Nous avons accès à 116 observations complètes, définissant notre mesure empirique initiale. Nous appliquons la projection de cette mesure, avec régularisation polynomiale par morceaux, en conservant les valeurs des quartiles empirique (Figure 2 haut) et en perturbant ces mêmes quartiles (Figure 2 bas).

Par ailleurs, cette approximation continue de la fonction quantile empirique permet, par le biais d'observations distribuées uniformément, de simuler facilement de nouveaux échantillons selon la loi Q résultante de la projection sous contraintes. Le contrôle des valeurs des quantiles permet également d'injecter des informations extérieures aux données, tout en préservant l'allure de l'échantillon observé. Des applications de cette méthode pour l'étude de la robustesse de prévisions issues de modèles *boîte noire* sont en cours.

Bibliographie

- Bachoc, F., F. Gamboa, M. Halford, J-M. Loubes, and L. Risser. 2020. "Explaining Machine Learning Models using Entropic Variable Projection." *arXiv:1810.07924 [cs, stat]*.
- Da Veiga, S., F. Gamboa, B. Iooss, and C. Prieur. 2021. *Basics and Trends in Sensitivity Analysis. Theory and Practice in R*. SIAM.

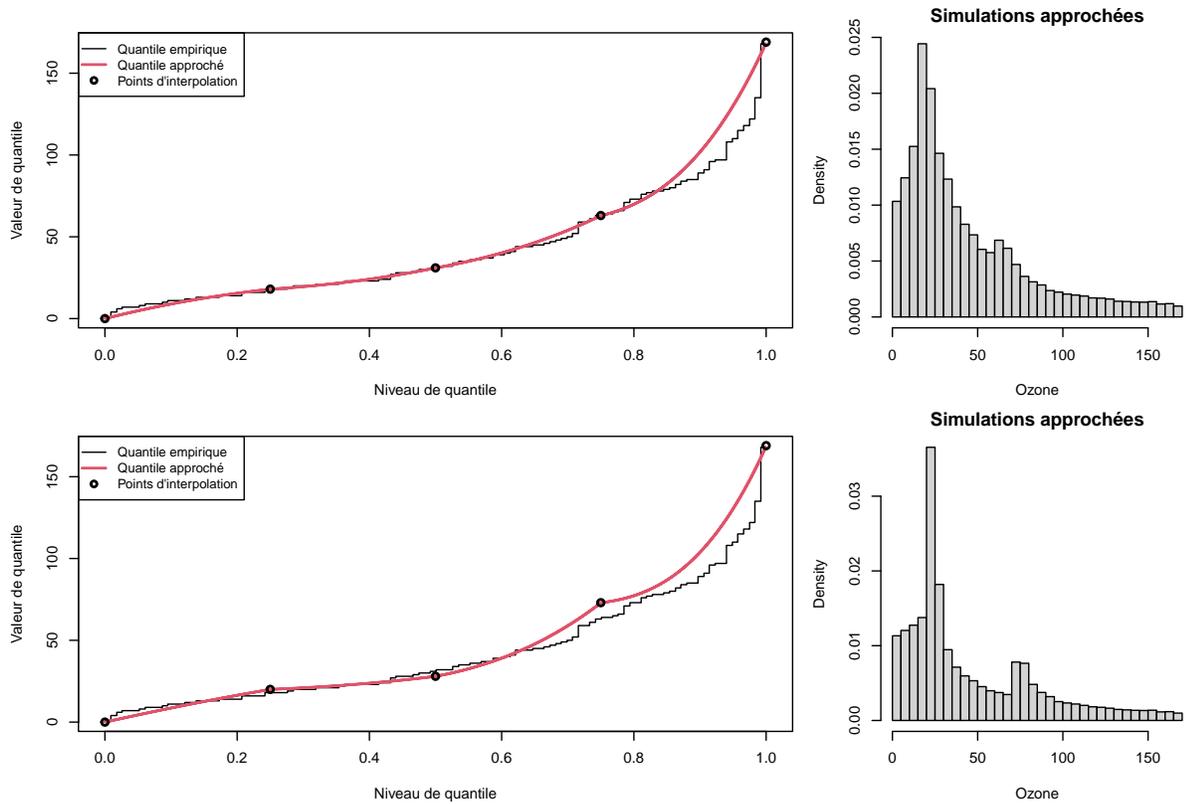


Figure 2: Approximations de la fonction de quantile empirique des données Ozone et histogrammes de simulations, par le résultat de la projection par conservation (haut) et perturbation (bas) des quartiles empiriques.

de La Fortelle, A. 2015. “A study on generalized inverses and increasing functions Part I: generalized inverses,” 14.

Grant, M., and S. Boyd. 2014. *CVX: Matlab Software for Disciplined Convex Programming, version 2.1*. <http://cvxr.com/cvx>.

Lasserre, J-B. 2015. *An Introduction to Polynomial and Semi-Algebraic Optimization*. Cambridge University Press.

Parrilo, P. A. 2010. *Algebraic Optimization and Semidefinite Optimization : EIDMA Mini-course*.

———. 2012. “Chapter 3: Polynomial Optimization, Sums of Squares, and Applications.” Chap. 3 in *Semidefinite Optimization and Convex Algebraic Geometry*, 47–157.

Villani, C. 2003. *Topics in Optimal Transportation*. Vol. 58. American Mathematical Society.