

# Quantiles multivariés par transport optimal et application à la régression quantile

Projet d'expertise en statistiques et probabilités 2020/2021

Théophile Baranger  
Sébastien GUIROY  
Exaucé Luweh Adjim Ngarti

18 mai 2021

Supervisé par Jérémie Bigot

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Rappels sur les quantiles . . . . .	2
1.2	Comment définir le quantile multivarié ? . . . . .	2
1.3	Objectifs et application . . . . .	3
<b>2</b>	<b>Transport optimal et quantiles multivariés</b>	<b>4</b>
2.1	Transport optimal . . . . .	4
2.1.1	Introduction et historique . . . . .	4
2.1.2	Transport de mesures de probabilité . . . . .	4
2.1.3	Le problème du transport optimal . . . . .	5
2.2	Lien entre transport et quantiles . . . . .	6
2.2.1	Étude théorique . . . . .	6
2.2.2	Généralisation à la dimension supérieure . . . . .	7
2.3	Quantiles de Monge-Kantorovich . . . . .	8
2.3.1	Définition et propriétés . . . . .	8
2.3.2	Cas discret . . . . .	9
2.3.3	Cas semidiscret . . . . .	9
<b>3</b>	<b>Application à la régression quantile</b>	<b>12</b>
3.1	Régression quantile . . . . .	12
3.1.1	Introduction . . . . .	12
3.1.2	Le quantile comme solution d'un problème d'optimisation . . . . .	12
3.1.3	Définition de la régression quantile . . . . .	14
3.1.4	Exemples et application aux données ANSUR II . . . . .	14
3.2	Régression quantile vectorielle par transport optimal . . . . .	18
3.2.1	Lien entre régression quantile et transport optimal . . . . .	18
3.2.2	Vers la dimension supérieure . . . . .	19
3.2.3	Application aux données ANSUR II . . . . .	20
<b>4</b>	<b>Conclusion</b>	<b>22</b>
	<b>Références</b>	<b>23</b>

# 1 Introduction

## 1.1 Rappels sur les quantiles

On se place dans un espace probabilisé  $(\Omega, \mathcal{F}, \mathbb{P})$ . Soit  $X$  une variable aléatoire réelle continue de densité de probabilité  $f$  strictement positive, et de fonction de répartition  $F$  continue et strictement croissante. Pour tout  $\alpha \in ]0, 1[$ , on appelle quantile d'ordre  $\alpha$  de  $X$ , l'unique racine  $\theta_\alpha$  de l'équation  $F(x) = \alpha$ . Ainsi, le quantile d'ordre  $\alpha$  correspond au seuil  $\theta_\alpha$ , tel que  $\mathbb{P}(X \leq \theta_\alpha) = \alpha$ . On note que dans ce cadre, la fonction de répartition  $F$  constitue une bijection de  $\mathbb{R}$  dans  $]0, 1[$ , et donc en notant  $F^{-1}$  son inverse, le quantile d'ordre  $\alpha$  de  $X$  est l'image par  $F^{-1}$  de  $\alpha$ . En effet, on a bien  $F^{-1}(\alpha) = \theta_\alpha$ . Cette remarque aura son importance par la suite.

Les quantiles les plus fréquemment utilisés en statistiques descriptives standard sont les quantiles d'ordre  $\frac{1}{4}$ ,  $\frac{1}{2}$ ,  $\frac{3}{4}$ , soit respectivement le premier quartile, la médiane, et le troisième quartile. Ils permettent de se donner une première idée de la distribution que l'on souhaite analyser. Il est fréquent de représenter cette information sous la forme de boîtes à moustaches, ou *boxplot*, tels que représentées en guise d'illustration dans la FIGURE 1.

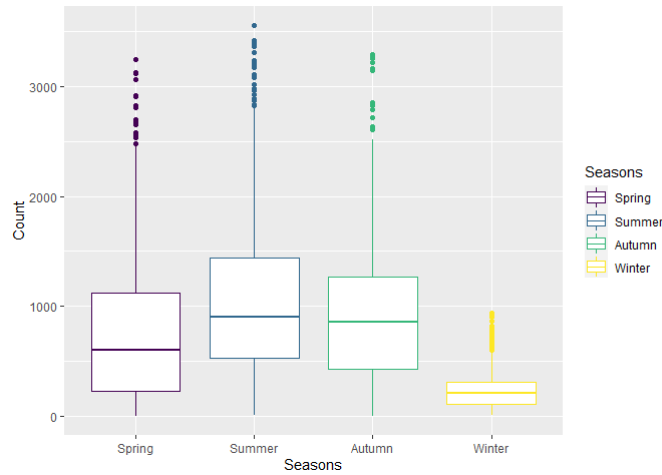


FIGURE 1 – Boîtes à moustaches représentant le nombre de vélos empruntés dans la ville de Séoul en 2019 selon les saisons. Données disponible à l'adresse suivante : <https://archive.ics.uci.edu/ml/datasets/Seoul+Bike+Sharing+Demand>.

On peut également définir le quantile d'ordre  $\alpha$  dans le cas d'une fonction de répartition non nécessairement continue et strictement croissante, à l'aide de l'inverse généralisée, définie de la manière suivante :

$$\forall y \in ]0, 1], \quad F^{-1}(y) = \inf\{x \in \mathbb{R} : F(x) \geq y\},$$

avec la convention  $\inf \emptyset = +\infty$ . Dans le cas où  $F$  est continue strictement croissante, l'inverse généralisé et son inverse coïncident. On définit alors le quantile d'ordre  $\alpha$  de  $X$ , pour  $\alpha \in ]0, 1]$ , par  $F^{-1}(\alpha)$ .

## 1.2 Comment définir le quantile multivarié ?

Nous avons rappelé dans la section précédente les définitions usuelles de quantiles univariés, ainsi que leur utilisation en statistiques descriptives. On souhaiterait

maintenant généraliser cette notion dans le cas multivarié. Cependant l'absence d'une relation d'ordre en dimension supérieure ne nous permet pas d'étendre naturellement ces définitions. Le problème est donc de généraliser les quantiles en dimension supérieure à 2. Une première idée pourrait être, face à un échantillon d'observations multivariées, de s'intéresser aux quantiles usuels selon chaque dimension de façon indépendante. Mais alors on dénature l'ensemble de données auquel on s'intéresse, et l'interprétation des résultats devient problématique.

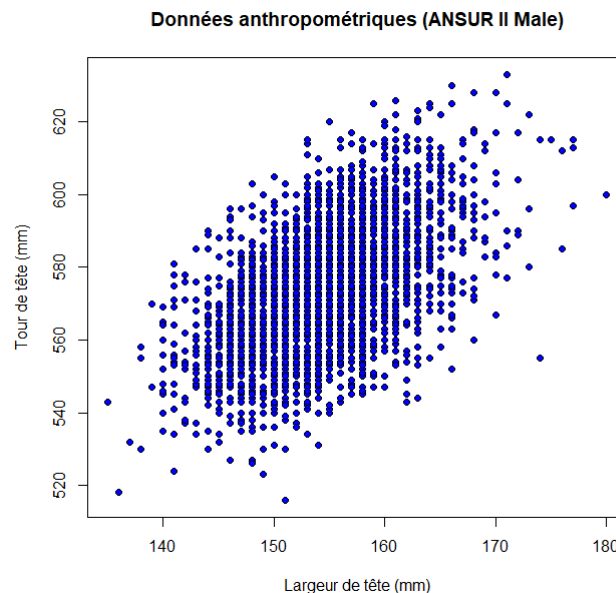


FIGURE 2 – Nuage de points en dimension 2. Il s'agit de données anthropométriques, en abscisses la largeur du crâne, en ordonnées sa circonférence, pour 4082 hommes. Comment définir un quantile (disons d'ordre 0.5, c'est-à-dire une médiane) dans ce cadre ?

### 1.3 Objectifs et application

Même en l'absence d'une relation d'ordre, il est possible de déterminer des ensembles qui, comme pour le cas des quantiles en dimension 1, « contiennent une proportion  $\alpha$  des observations ». Cela peut permettre, dans le cas des données de la FIGURE 2 par exemple, de déterminer des valeurs de référence et mettre en exergue une relation entre la circonférence du crâne et la largeur de la tête en considérant une partie des observations. Ainsi, dans la première partie de ce rapport, nous commencerons par une introduction rapide de la théorie du transport optimal, puis nous nous intéresserons à la définition de quantiles multivariées proposée par Chernozhukov, Galichon, Hallin et Henry [4], et nous illustrerons ces quantiles par des simulations.

Dans la seconde partie, nous présenterons une méthode de régression que nous n'avons pas eu l'occasion d'aborder en cours, la régression quantile. Nous nous appuierons sur l'ouvrage de référence sur ce sujet, *Quantile Regression*, par Roger Koenker [6]. Finalement, nous présenterons la généralisation à la régression quantile vectorielle par transport optimal proposée par Carlier, Chernozhukov et Galichon [3], et nous illustrerons les résultats obtenus sur des données anthropométriques.

## 2 Transport optimal et quantiles multivariés

### 2.1 Transport optimal

#### 2.1.1 Introduction et historique

Dans un mémoire publié en 1781, Gaspard Monge s'intéresse à un problème concret d'ingénierie militaire qui consiste à déterminer comment déplacer de manière optimale un volume de terre, un *déblai*, vers un endroit assigné à l'avance, le *remblai* [8]. De manière optimale signifie ici de telle sorte à minimiser l'effort nécessaire à cette opération, autrement dit la distance totale parcourue par chaque molécule de terre transportée. La théorie du transport a bien entendu beaucoup évolué depuis les travaux de Monge, et on ne retracera pas ici toute les étapes de cette évolution.

Dans sa formulation la plus basique, le problème de transport optimal consiste en un problème d'affectation, c'est-à-dire que l'on part d'un ensemble discret, par exemple des restaurants mettant à disposition des repas à livrer, et de destinations fixées à l'avance, par exemple de bureaux, et le problème consiste à répartir les repas de façon à minimiser le coût de transport global. Les restaurants et les bureaux peuvent ainsi être représentés comme des points du plan, auxquels sont assignés un nombre de repas, et on souhaite minimiser la somme des distances entre chaque restaurateur et l'adresse qui lui a été désignée. On représente cette situation en FIGURE 3.

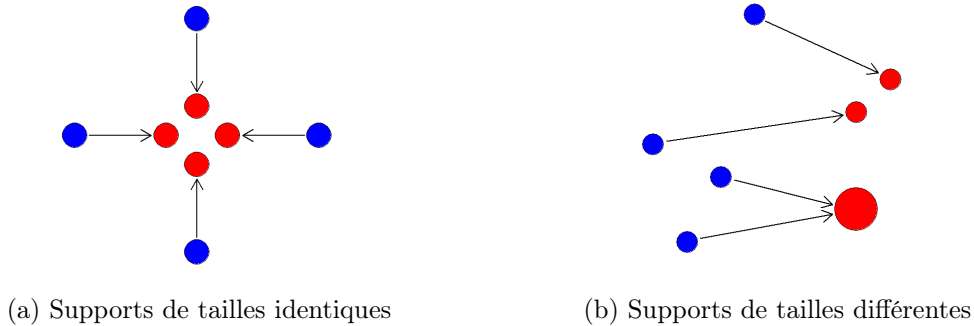


FIGURE 3 – Illustration d'une solution au problème de Monge dans le cas discret. Les points bleus représentent des restaurants, qui disposent ici tous de la même quantité de repas à livrer. Les points rouges représentent des bureaux, qui se font livrer ces repas par un transporteur, en minimisant le coût de transport.

Une évolution majeure dans la théorie du transport optimal fut de considérer non plus une fonction qui à chaque élément de la distribution de départ assigne un seul élément d'arrivée, mais éventuellement plusieurs. Cette seconde formulation, due à Kantorovich [5], est à la fois plus générale et plus simple, car elle permet de formuler le problème comme un problème d'optimisation convexe. Pour plus de détails sur la formulation du problème au sens de Monge ou au sens de Kantorovich, on renvoie notamment à la partie *Theoretical Foundations* du livre intitulé *Computational Optimal Transport* [9], par Gabriel Peyré et Marco Cuturi.

#### 2.1.2 Transport de mesures de probabilité

Nous allons nous concentrer à partir de maintenant sur le problème lié au transport de mesures de probabilités. La définition de transport que nous allons utiliser à partir

de maintenant est celle proposée par Villani [10]. Il définit le transport d'une mesure de probabilité de la façon suivante. On se donne deux mesures de probabilités,  $\mu$  et  $\nu$ , définies respectivement sur des espaces mesurables  $\mathcal{X}$  et  $\mathcal{Y}$ . Alors si l'on se donne une fonction mesurable  $T : \mathcal{X} \rightarrow \mathcal{Y}$ , on dit que  $\nu$  est la *push-forward*, ou la *mesure image* de  $\mu$  par  $T$ , ou encore que  $T$  pousse  $\mu$  sur  $\nu$ , si la condition suivante est vérifiée :

$$\text{Pour tout ensemble mesurable } B \subset \mathcal{Y}, \nu(B) = \mu(T^{-1}(B)).$$

On note alors  $\nu = T\#\mu$ . Une autre façon de caractériser le transport de mesure de probabilité est la suivante. Soit  $X$  une variable aléatoire sur  $\mathcal{X}$  de loi  $\mu$ . Alors dire que  $T\#\mu = \nu$  revient à dire que  $T(X) \sim \nu$ .

### 2.1.3 Le problème du transport optimal

Cependant, le transport d'une mesure de référence vers une mesure d'intérêt induit un coût  $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}_+$ . Par exemple, le coût du transport d'un point vers un autre dans le plan peut être la norme euclidienne entre les 2 points. Cette notion de coût se généralise au transport d'une mesure vers une autre. Selon la fonction de transport utilisée, le coût peut varier. Nous allons chercher celle qui, justement, minimise ce coût : il s'agit de la fonction de transport optimale.

Ainsi, on peut traduire le problème de transport au sens de Monge par la formulation [9] suivante :

$$\min_T \left\{ \int_{\mathcal{X}} c(x, T(x)) d\mu(x) : T\#\mu = \nu \right\}$$

Comme expliqué précédemment, Kantorovich a proposé une relaxation du problème de Monge en ne considérant plus une relation 1 à 1 entre les points de départ et d'arrivée, mais en englobant les cas où un point de départ peut conduire à plusieurs destinations différentes. On illustre le problème de Kantorovich dans le cas discret dans la FIGURE 4.

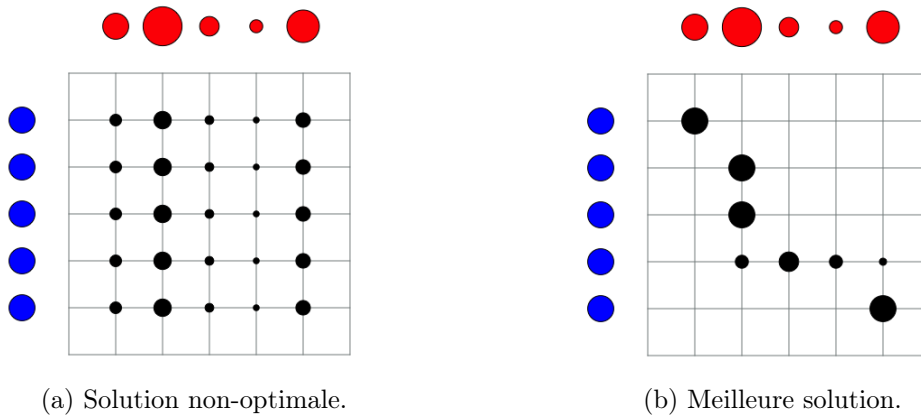


FIGURE 4 – Représentation schématique de solutions au problème de transport de Kantorovich. On cherche à transporter la masse des points rouges vers les points bleus. À gauche une solution naïve qui correspond à la loi jointe de deux variables discrètes indépendantes, et à droite une solution optimale au problème de Kantorovich.

Autrement dit, la résolution du problème de Kantorovich peut s'écrire [9] ainsi :

$$\min_{\pi \in \mathcal{U}(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\pi(x, y)$$

avec  $\mathcal{U}(\mu, \nu) = \{\pi \in \mathcal{M}_+^1(\mathcal{X} \times \mathcal{Y}) : P_{\mathcal{X}}\# \pi = \mu \text{ et } P_{\mathcal{Y}}\# \pi = \nu\}$ ,  $\mathcal{M}_+^1(\mathcal{X} \times \mathcal{Y})$  correspond à l'ensemble des densités jointes de l'espace produit, et  $P_{\mathcal{X}}$  et  $P_{\mathcal{Y}}$  sont les projections de  $\pi$  sur les espaces  $\mathcal{X}$  et  $\mathcal{Y}$  respectivement.

## 2.2 Lien entre transport et quantiles

### 2.2.1 Étude théorique

**Proposition 2.1.** *Soit  $X$  une variable aléatoire de fonction de répartition  $F_X$  continue et strictement croissante et  $U \sim \mathcal{U}([0, 1])$ . Alors la variable aléatoire  $F_X^{-1}(U)$  suit la même loi que  $X$ .*

*Démonstration.* Montrons que pour tout  $u \in [0, 1]$  et pour tout  $x \in \mathbb{R}$ ,

$$F_X^{-1}(u) \leq x \iff u \leq F_X(x).$$

On a :

1. Si  $F_X^{-1}(u) \leq x$ , comme  $F_X$  est croissante, on a aussi  $F_X(F_X^{-1}(u)) \leq F_X(x)$  et  $F_X$  est continue à droite donc  $F_X(F_X^{-1}(u)) = \lim_{t \rightarrow F_X^{-1}(u)} F_X(t)$  et comme  $t \geq F_X^{-1}(u)$ ,  $F_X(t) \geq u$  et  $F_X(F_X^{-1}(u)) \geq u$ . Ce qui entraîne  $u \leq F_X(F_X^{-1}(u)) \leq F_X(x)$ . Donc  $u \leq F_X(x)$ .
2. Si  $u \leq F_X(x)$ , alors  $x \in \{x \in \mathbb{R} : F_X(x) \geq u\}$  et donc  $x \geq \inf\{x \in \mathbb{R} : F_X(x) \geq u\}$  et donc  $F_X^{-1}(u) \leq x$ .

De plus, puisque

$$F_X(x) \in [0, 1] \text{ \& } U \sim \mathcal{U}([0, 1]),$$

alors on a

$$\mathbb{P}(F_X^{-1}(U) \leq x) = \mathbb{P}(U \leq F_X(x)) = F_X(x).$$

Finalement, on a bien  $F_X^{-1}(U) \sim X$ . □

Ainsi, on peut dire que la fonction  $F_X^{-1}$  transporte la loi uniforme sur  $[0, 1]$  sur la loi de  $X$ . De plus,  $\forall \alpha \in [0, 1]$ ,  $\mathbb{P}(X \leq F_X^{-1}(\alpha)) = \alpha$ , donc  $F_X^{-1}$  correspond bien dans ce cadre à la fonction quantile usuelle, et donc on peut dire que la fonction quantile est la fonction qui transporte la loi uniforme sur  $[0, 1]$  sur la loi de  $X$ . D'où le lien entre transport et quantile. De plus, par symétrie, la fonction  $F_X$  transporte la loi de  $X$  sur la loi uniforme sur  $[0, 1]$ .

*Démonstration.* Si  $F_X$  est continue partout, alors  $F_X^{-1}(F_X(u)) = u$ . Par suite,

1. Si  $u \in [0, 1]$ ,  $\mathbb{P}(F_X(X) \leq u) = \mathbb{P}(X \leq F_X^{-1}(u)) = F_X(F_X^{-1}(u)) = u$
2. Si  $u \leq 0$  alors  $\mathbb{P}(F_X(X) \leq u) = 0$  et si  $u \geq 1$  alors  $\mathbb{P}(F_X(X) \leq u) = 1$ ,

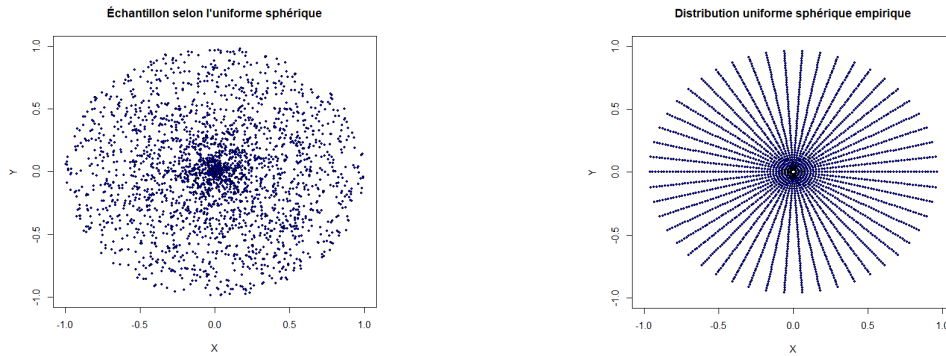
car  $F_X$  est une fonction de répartition.

Donc on a bien  $F_X(X) \sim \mathcal{U}([0, 1])$ . □

Nous avons montré que l'inverse de la fonction de répartition est la fonction quantile usuelle en dimension 1 et qu'elle transporte la loi uniforme sur  $[0, 1]$  vers la loi d'une variable aléatoire quelconque. On s'intéresse dans la suite à généraliser ce résultat en dimension supérieure à 1.

### 2.2.2 Généralisation à la dimension supérieure

Pour généraliser en dimension supérieure à 2, la distribution dite *de référence* qui jouera le rôle de la distribution uniforme sur  $]0, 1[$  dans le cas univarié est ce que les auteurs de [4] appellent la distribution *uniforme sphérique* en dimension  $d$ , Notée  $U_d$ , et définie comme la distribution d'un vecteur aléatoire  $r\varphi$ , où  $r$  suit la loi uniforme sur  $]0, 1[$ , et où  $\varphi$  suit la loi uniforme sur la sphère unité  $\mathcal{S}^{d-1} := \{x \in \mathbb{R}^d : \|x\| = 1\}$  avec  $r$  et  $\varphi$  indépendants. On illustre cette distribution dans la FIGURE 5.



(a) Échantillon de 2500 observations suivant la loi uniforme sphérique dans  $\mathbb{R}^2$ . (b) Distribution uniforme sphérique discrète dans  $\mathbb{R}^2$ , avec 2500 points.

FIGURE 5 – Illustration de la distribution uniforme sphérique. Il s'agit de la distribution de référence proposée dans [4] pour définir les quantiles multivariés.

On peut légitimement se poser la question du choix de la distribution uniforme sphérique comme généralisation à la dimension supérieure de la distribution uniforme sur  $]0, 1[$ . La raison donnée par les auteurs de l'article [4] est qu'en dimension supérieure, une famille de distribution pour laquelle on peut définir naturellement des quantiles multivariés est la famille  $\mathcal{P}_{\text{ell};g}^d$  des distributions elliptiques de plein rang dans  $\mathbb{R}^d$ , avec  $d > 1$ , et possédant une densité radiale  $g$  sur  $\mathbb{R}_+$ . Ces distributions sont caractérisées par le fait qu'à une transformation affine près, elles correspondent à des distributions sphériques. En notant  $X$  un vecteur aléatoire de  $\mathcal{P}_{\text{ell};g}^d$ , on dit que  $X$  a pour distribution  $P_{\mu,\Sigma,g}$ , et ainsi le *résidu*  $Y := \Sigma^{-1/2}(X - \mu)$  est de distribution sphérique  $P_{0,I,g}$ . Si l'on note  $G$  la fonction de répartition radiale de  $X$ , c'est-à-dire que pour tout  $x > 0$ ,  $G(x) = \int_0^x g(t) dt$ , alors les auteurs notent que  $R_P(Y) := (Y/\|Y\|)G(\|Y\|)$  suit la loi uniforme sphérique.

Par suite, en fixant  $\tau$  entre 0 et 1, on peut définir des contours quantiles de niveau  $\tau$  autour de la valeur centrale d'une distribution sphérique par des cercles concentriques, qui coïncident par définition de  $R_P$  avec la probabilité que  $Y$  se trouve à l'intérieur



de l'hypersphère de rayon  $\tau$  autour de l'origine. Les auteurs définissent ensuite les quantiles multivariés comme l'inverse de la fonction  $y \mapsto \mathbb{R}_P(y)$ , dans le cadre de ces distributions.

## 2.3 Quantiles de Monge-Kantorovich

Dans leur article intitulé *Monge-Kantorovich depth, quantiles, ranks and signs*, Chernozhukov, Galichon, Hallin et Henry [4] proposent une nouvelle définition de quantiles et de rangs multivariés, qui permettent de généraliser les définitions vues précédemment à la dimension supérieure. Ils appellent les quantiles ainsi définis les quantiles de Monge-Kantorovich, car ils se reposent sur la résolution d'un problème de transport optimal.

Nous avons vu que dans le cas univarié, la fonction quantile transportait la distribution uniforme sur  $]0, 1[$  vers la distribution d'une variable aléatoire réelle  $X$ . En dimension supérieure, l'absence de relation d'ordre nous pousse à définir des quantiles centrés. La généralisation proposée par [4] consiste à définir la fonction quantile comme le gradient d'une fonction convexe, dont l'existence et l'unicité sont assurées par l'extension de McCann du théorème de Brenier (voir Théorème 2.1) dans le cas où les distributions de référence et d'intérêt ont des moments d'ordre 2 finis.

### 2.3.1 Définition et propriétés

Avec les notations de l'article [4], on se place dans le cadre de deux distributions sur  $\mathbb{R}^d$ ,  $F$  et  $P$ . On s'intéresse alors au transport de la distribution connue  $F$ , vers une distribution d'intérêt,  $P$ . Commençons par présenter le théorème [4] dû à Brenier [1] et McCann [7], qui assure l'existence du gradient de fonctions convexes dont les propriétés permettront de définir les quantiles multivariés.

**Théorème 2.1** (Brenier-McCann). *Soient  $F$  et  $P$  deux distributions sur  $\mathbb{R}^d$ . Si  $F$  est absolument continue par rapport à la mesure de Lebesgue sur  $\mathbb{R}^d$ , alors pour tout ensemble convexe  $\mathcal{U} \subset \mathbb{R}^d$  contenant le support de  $F$ , il existe une fonction convexe  $\psi : \mathcal{U} \rightarrow \mathbb{R} \cup \{+\infty\}$  telle que  $\nabla \psi \# F = P$ . Le gradient  $\nabla \psi$  de cette fonction existe et est unique  $F$ -presque partout. Si, de plus,  $P$  est absolument continue sur  $\mathbb{R}^d$ , alors pour tout ensemble convexe  $\mathcal{Y} \subset \mathbb{R}^d$  contenant le support de  $P$ , il existe une fonction convexe  $\psi^* : \mathcal{Y} \rightarrow \mathbb{R} \cup \{+\infty\}$  telle que  $\nabla \psi^* \# P = F$ . Le gradient  $\nabla \psi^*$  de  $\psi^*$  existe, est unique, et  $\nabla \psi^* = \nabla \psi^{-1}$ ,  $P$ -presque partout.*

Les auteurs remarquent ensuite que si  $F$  et  $P$  ont des moments d'ordre 2 finis, alors  $\nabla \psi$  est un plan de transport optimal de  $F$  vers  $P$  pour le coût quadratique. C'est-à-dire que  $\nabla \psi$  est solution du problème

$$\inf_{\mathcal{Q}} \int \|u - \mathcal{Q}(u)\|^2 dF(u), \quad \mathcal{Q} \# F = P.$$

On remarque en développant le produit scalaire que ce problème est équivalent au problème de maximisation,

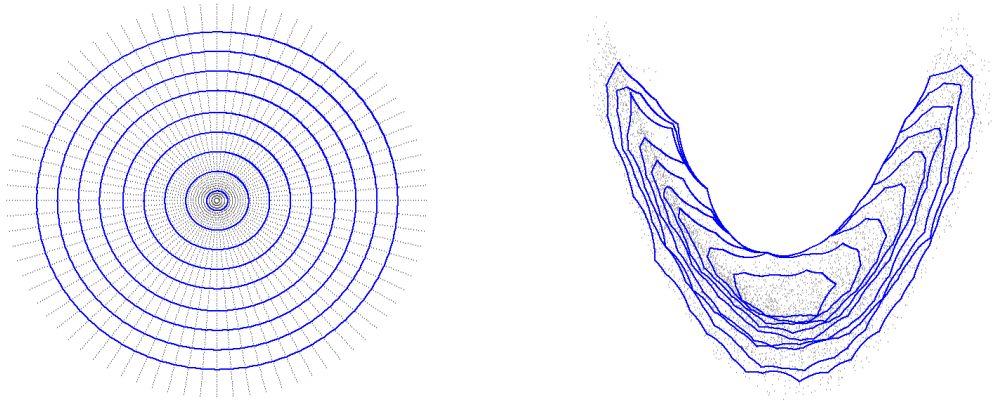
$$\sup_{\mathcal{Q}} \int \langle u, \mathcal{Q}(u) \rangle dF(u), \quad \mathcal{Q} \# F = P.$$

### 2.3.2 Cas discret

Un cas propice aux simulations est le cas où la distribution de référence et la distribution d'intérêt sont des mesures discrètes de même cardinalité, et uniformément distribuées sur leur supports. Autrement dit, en suivant les notations de [4], on note  $\hat{F}_n$  la distribution discrète de référence et  $\hat{P}_n$  la distribution empirique d'intérêt. Alors on a  $\hat{F}_n = \sum_{j=1}^n \delta_{u_j}/n$  pour un  $n$ -uplet  $\mathcal{U}_n = \{u_1, \dots, u_n\}$  d'éléments de  $\mathbb{R}^d$ , et  $\hat{P}_n = \sum_{j=1}^n \delta_{y_j}/n$  pour un autre  $n$ -uplet  $\mathcal{Y}_n = \{y_1, \dots, y_n\}$  d'éléments de  $\mathbb{R}^d$ . Alors dans ce cadre, la fonction quantile  $\hat{Q}_n$  est obtenue en résolvant le problème de programmation linéaire qui consiste à trouver la permutation optimale  $\sigma^*$  solution du problème d'affectation

$$\sigma^* = \min_{\sigma \in S_n} \sum_{j=1}^n \|u_j - y_{\sigma(j)}\|^2.$$

On illustre le transport des régions de probabilité concentriques de niveau  $\tau \in ]0, 1[$  de la distribution uniforme sphérique discrète vers la distribution empirique d'intérêt dans la FIGURE 6.



(a) Régions de probabilités de l'uniforme sphérique discrète. (b) Régions de probabilités de la distribution empirique d'intérêt.

FIGURE 6 – Illustration du transport de régions de probabilités par résolution du problème de Monge (10000 observations), pour les régions de probabilités de niveaux  $\tau \in \{0.1, 0.2, \dots, 0.9\}$ .

On remarque que l'emboîtement des régions de probabilités successives pour la distribution uniforme sphérique est maintenu lors du transport vers la distribution empirique d'intérêt.

### 2.3.3 Cas semidiscret

Dans le cas où la distribution d'intérêt  $\hat{P}_n$  est discrète mais que la distribution de référence est continue, ce qui est le cas de la distribution uniforme sphérique définie précédemment, alors on peut notamment utiliser un algorithme de descente de gradient stochastique pour déterminer la fonction de transport optimale. On s'intéressera dans la suite du projet à l'implémentation du transport optimal semidiscret dans le cadre des quantiles multivariés.

**Théorème 2.2** (Kantorovich-Brenier). *Soient  $F$  et  $P$  deux distributions sur  $\mathbb{R}^d$ . Si  $F$  est absolument continue par rapport à la mesure de Lebesgue sur  $\mathbb{R}^d$  et si  $F$  a un moment d'ordre 2 fini, alors on note  $(\psi, \psi^*)$  la paire de conjugués solution du problème d'optimisation*

$$\int \psi dF + \int \psi^* dP = \inf_{(\phi, \phi^*) \in \Theta} \left( \int \phi dF + \int \phi^* dP \right).$$

*L'ensemble  $\Theta$  sur lequel on minimise correspond à l'ensemble des paires de potentiels conjugués sur  $(\mathcal{U}, \mathcal{Y})$ . La solution  $(\psi, \psi^*)$  est appelé paire de potentiels optimal.*

Ce problème est le dual du problème de transport optimal et sous certaines conditions,  $\nabla \psi$  est l'unique fonction de transport optimal de  $F$  sur  $P$  pour le coût quadratique.

Soient  $y_1 \dots y_n \in \mathbb{R}^d$  des observations de notre distribution d'intérêt,  $\hat{P}_n = \sum_{j=1}^n \delta_{y_j}/n$  la mesure empirique des données et la distribution de référence  $F$ , une mesure uniforme sur la boule unité  $\mathcal{U} = \{u \in \mathbb{R} : \|u\| \leq 1\}$ . La fonction convexe  $\hat{\psi}_n$  telle que  $\nabla \hat{\psi}_n \# F = \hat{P}_n$  est donnée dans [4] à une constante additive près par

$$\hat{\psi}_n(u) = \max_{1 \leq k \leq n} \{u^\top y_k - v_k^*\} \quad \forall u \in \mathcal{U}.$$

Où  $(v_1^*, \dots, v_n^*)$  est le minimiseur de  $f : \{v_1, \dots, v_n\} \mapsto \int \hat{\psi}_n d\hat{F}_n + \sum_{k=1}^n \frac{v_k}{n}$ .

**Lemme 2.1.** *Pour tout  $u \in \mathcal{U}$ ,  $\nabla \hat{\psi}_n(u) = y_{k(u)}$ . Avec  $k(u) = \operatorname{argmax}_{1 \leq k \leq n} \{u^\top y_k - v_k^*\}$ .*

Pour pouvoir réaliser le transport de la boule unité vers notre distribution d'intérêt, il suffit de déterminer le vecteur  $v^* = (v_1^* \dots v_n^*)$ . Pour ce faire nous allons utiliser comme annoncé un algorithme de descente de gradient stochastique puisque  $v^*$  est le minimiseur de la fonction  $f$ .

$$\begin{aligned} v^* &= \operatorname{argmin}_{v \in \mathbb{R}^n} f(v_1 \dots v_n) \\ &= \operatorname{argmin}_{v \in \mathbb{R}^n} \int \hat{\psi}_n d\hat{F}_n + \sum_{k=1}^n \frac{v_k}{n} \\ &= \operatorname{argmin}_{v \in \mathbb{R}^n} \int \max_{1 \leq k \leq n} \{u^\top y_k - v_k\} d\hat{F}_n + \bar{v}_n \\ &= \operatorname{argmin}_{v \in \mathbb{R}^n} \int \left( \max_{1 \leq k \leq n} \{u^\top y_k - v_k\} + \bar{v}_n \right) d\hat{F}_n \\ &= \operatorname{argmin}_{v \in \mathbb{R}^n} \int g(u, v) d\hat{F}_n \end{aligned}$$

avec  $g(u, v) = \max_{1 \leq k \leq n} \{u^\top y_k - v_k\} + \bar{v}_n$ . Par suite, en posant  $U \sim \hat{F}_n$ ,

$$\begin{aligned} v^* &= \operatorname{argmin}_{v \in \mathbb{R}^n} \mathbb{E}[g(U, v)] \\ &= \operatorname{argmin}_{v \in \mathbb{R}^n} \frac{1}{m} \sum_{i=1}^m [g(u_i, v)] \\ &= \operatorname{argmin}_{v \in \mathbb{R}^n} g(u, v). \end{aligned}$$

Soient  $U_1 \dots U_m$ , les réalisations de notre distribution de référence, l'uniforme sphérique dans notre cas, alors la suite

$$\hat{V}_{n+1} = \hat{V}_n - \gamma_n \nabla_v g(U_{n+1}, \hat{V}_n)$$

converge presque sûrement vers  $v^*$ , la convergence est démontrée dans [9].

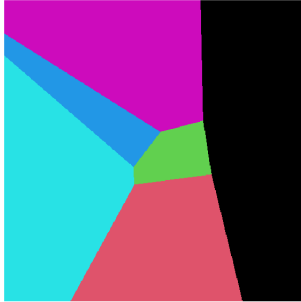
Déterminons le gradient de  $g$ ,

$$\begin{aligned} \nabla_v g(u, v) &= \nabla_v \left( \max_{1 \leq k \leq n} \{u^\top y_k - v_k\} + \bar{v}_n \right) \\ &= \nabla_v \max_{1 \leq k \leq n} \{u^\top y_k - v_k\} + \nabla_v \frac{1}{n} \sum_{k=1}^n v_k \\ &= \nabla_v f(u, v) + \nabla h(u, v) \end{aligned}$$

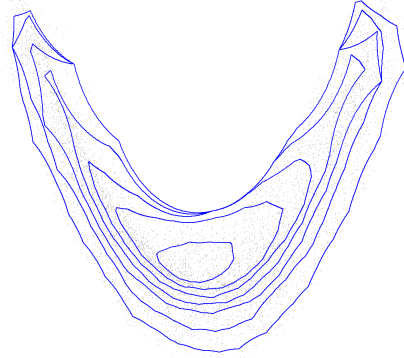
Grâce au théorème de l'enveloppe on calcule le max puis le gradient. Soit  $k^* = \operatorname{argmax}_{1 \leq k \leq n} \{u^\top y_k - v_k\}$  alors,

$$\frac{\partial f(u, v)}{\partial v_k} = -\mathbb{1}_{\{k=k^*\}} \quad \text{et} \quad \frac{\partial h(u, v)}{\partial v_k} = \frac{1}{n}.$$

Pour l'implémentation de l'algorithme, on peut considérer  $\nabla_v g = \nabla_v f$  car l'ajout du vecteur  $(\frac{1}{n}, \dots, \frac{1}{n})$  au gradient, ne change pas la direction de la descente et on peut bien se mettre dans un cas où  $v_n$  est tel que  $\bar{v}_n = 0$ .



(a) Transport semi-discret dans le cadre d'une distribution d'intérêt discrète avec très peu de points. On illustre ici l'espace de départ, et la distribution de référence est l'uniforme sur le carré  $[0, 1]^2$ .



(b) Régions de probabilités de la distribution empirique d'intérêt. Avec 20000 observations dans la distribution d'intérêt et une discrétisation du disque à 90000 points,  $10^6$  itérations de l'algorithme.

FIGURE 7 – Deux illustrations issues du transport optimal dans le cadre semi-discret. À gauche on transporte la distribution uniforme sur le carré  $[0, 1]^2$  vers une mesure discrète de cardinalité égale à 7, et on représente l'espace de départ, où les points d'une région de même couleur sont envoyés vers le même point d'arrivée. À droite, on illustre le transport semi-discret de l'uniforme sphérique vers la distribution banane décrite précédemment. L'approche semi-discrete permet d'obtenir des régions plus régulières.

## 3 Application à la régression quantile

### 3.1 Régression quantile

#### 3.1.1 Introduction

Nous commençons par introduire la régression quantile, qui est une approche nouvelle pour nous. Cette partie s'inspire grandement du premier chapitre du monographe de Roger Koenker, *Quantile Regression* [6]. L'ambition première de la régression quantile est d'offrir une analyse plus riche et plus nuancée que les méthodes classiques de régression telle que la régression linéaire. Ces méthodes se concentrent généralement sur l'analyse du comportement moyen d'une variable réponse, ou d'une fonction de celle-ci. On estime alors l'espérance d'une variable réponse conditionnellement à un certain nombre de variables explicatives. Malgré la richesse de la régression linéaire classique, beaucoup de subtilités peuvent être perdues dans cette approche.

Dans cette introduction, nous allons présenter les fondations de la régression quantile, puis nous illustrerons la richesse des interprétations qu'elle permet de fournir en proposant une petite étude de cas sur les données anthropométriques ANSUR + référence.

#### 3.1.2 Le quantile comme solution d'un problème d'optimisation

Nous avons rappelé dans l'introduction la définition classique du quantile d'ordre  $0 < \alpha < 1$  d'une variable aléatoire réelle, notée  $X$ , comme l'image de  $\alpha$  par l'inverse généralisée de la fonction de répartition de  $X$ ,

$$F^{-1}(\alpha) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}.$$

Le quantile peut également être obtenu comme la solution du problème de minimisation suivant. On considère une fonction de perte définie pour  $\alpha$  fixé, par

$$\forall x \in \mathbb{R}, \quad \rho_\alpha(x) = x(\alpha - \mathbb{1}_{\{x < 0\}}).$$

On illustre une telle fonction pour différentes valeurs de  $\alpha$  dans la FIGURE 8.

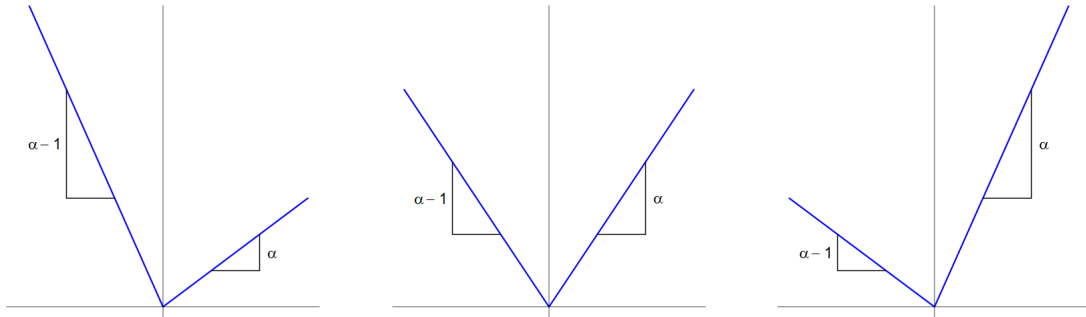


FIGURE 8 – Fonction de régression quantile  $\rho_\alpha$  pour  $\alpha = 0.25$ ,  $\alpha = 0.5$  et  $\alpha = 0.75$ .

On cherche alors  $x^*$  qui minimise l'espérance

$$\begin{aligned}\mathbb{E}[\rho_\alpha(X - x^*)] &= \int_{-\infty}^{+\infty} \rho_\alpha(x - x^*) dF(x) \\ &= \int_{-\infty}^{+\infty} (x - x^*)(\alpha - \mathbb{1}_{\{(x-x^*) < 0\}}) dF(x) \\ &= (\alpha - 1) \int_{-\infty}^{x^*} x - x^* dF(x) + \alpha \int_{x^*}^{+\infty} x - x^* dF(x).\end{aligned}$$

On cherche alors les points qui annulent la dérivée de cette espérance prise comme fonction de  $x^*$ , et par les théorèmes de dérivation sous le signe intégrale,

$$\begin{aligned}\frac{\partial}{\partial x^*}(\mathbb{E}[\rho_\alpha(X - x^*)]) &= (\alpha - 1) \int_{-\infty}^{x^*} \frac{\partial}{\partial x^*}(x - x^*) dF(x) + \alpha \int_{x^*}^{+\infty} \frac{\partial}{\partial x^*}(x - x^*) dF(x) \\ &= (1 - \alpha) \int_{-\infty}^{x^*} dF(x) - \alpha \int_{x^*}^{+\infty} dF(x) \\ &= \int_{-\infty}^{x^*} dF(x) - \alpha \int_{-\infty}^{+\infty} dF(x) \\ &= F(x^*) - \alpha.\end{aligned}$$

Et donc la dérivée s'annule dès lors que  $F(x^*) = \alpha$ . De plus  $F$  est croissante, donc si la solution est unique, on a  $x^* = F^{-1}(\alpha)$ , sinon c'est l'ensemble  $\{x \in \mathbb{R}, F(x) = \alpha\}$ . Dans ce cas, Koenker [6] indique qu'on prend le minimum de cette ensemble, pour que la fonction quantile soit continue à gauche. On retrouve bien la fonction quantile usuelle en résolvant ce problème de minimisation.

Dans le cas discret, c'est-à-dire où  $F$  est remplacée par la fonction de répartition empirique donnée par

$$\forall x \in \mathbb{R}, \quad F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}},$$

alors minimiser l'espérance de la perte selon  $x^*$  revient à minimiser la somme

$$\frac{1}{n} \sum_{i=1}^n \rho_\alpha(y_i - x^*),$$

et on obtient alors le quantile *empirique* de niveau  $\alpha$ . Le problème de trouver le quantile empirique de niveau  $\alpha$  s'écrit alors

$$\min_{x \in \mathbb{R}} \sum_{i=1}^n \rho_\alpha(y_i - x).$$

Le point clé est alors de reformuler ce problème de minimisation comme un programme linéaire, ce qui peut être fait par des méthodes d'optimisation classiques.

### 3.1.3 Définition de la régression quantile

Maintenant que nous avons vu que les quantiles pouvaient être exprimés comme solutions d'un problème d'optimisation, on présente l'idée à l'origine de la régression quantile, en suivant toujours Koenker [6]. Il explique que dans le cas de la régression linéaire classique, on s'appuie sur le fait que la moyenne empirique soit solution du problème des moindres carrés

$$\min_{\mu \in \mathbb{R}} \sum_{i=1}^n (y_i - \mu)^2.$$

Ainsi, si l'on pose le modèle linéaire standard

$$\mathbb{E}[Y|X] = \beta^\top X,$$

où  $X$  correspond à la matrice design à  $n$  lignes et  $p$  colonnes, alors on estime naturellement  $\beta$  en résolvant le problème de minimisation

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n (y_i - x_i^\top \beta)^2.$$

Pour définir la régression quantile, l'idée est alors de suivre le même modèle, en posant alors que le quantile conditionnel d'ordre  $\alpha$  satisfait une relation linéaire  $\mathcal{Q}_Y(\alpha|X) = X^\top \beta(\alpha)$ , et ainsi l'estimateur  $\hat{\beta}(\alpha)$  résout le problème

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho_\alpha(y_i - x_i^\top \beta).$$

Ce problème peut se réécrire comme un problème de programmation linéaire sous contrainte sur un polyèdre [6], en introduisant deux vecteurs de variables d'écart,  $u$  et  $v$ , respectivement la partie positive et la partie négative des résidus  $Y - X\beta$ ,

$$\min_{(\beta, u, v) \in \mathbb{R}^p \times \mathbb{R}_+^{2n}} \left\{ \alpha \mathbf{1}_n^\top u + (1 - \alpha) \mathbf{1}_n^\top v \mid X\beta + u - v = Y \right\}.$$

On a rappelé les fondations de la régression quantile, et dans la prochaine section on illustre cette approche sur quelques exemples simples.

La théorie de la régression quantile est très riche, et tout comme pour la régression linéaire classique, on dispose de nombreux tests statistiques, propriétés asymptotiques, intervalles de confiance, etc. qu'on ne mentionnera pas ici, mais dont on note qu'ils se reposent en général sur la définition précédente en tant que résolution d'un programme linéaire.

### 3.1.4 Exemples et application aux données ANSUR II

**Exemples théoriques.** Commençons par un exemple classique basé sur un échantillon issu d'un modèle de régression bivarié avec erreurs indépendantes et identiquement distribuées (iid)

$$y_i = \beta_0 + x_i \beta + \varepsilon_i.$$

Il est clair que les quantiles conditionnels de  $y$  sont données par

$$\mathcal{Q}_y(\alpha|x) = \beta_0 + x\beta_1 + F_\varepsilon^{-1}(\alpha),$$

où  $F_\varepsilon$  désigne la fonction de répartition des erreurs. On illustre dans la figure FIGURE 9 le résultat de la régression quantile pour certains niveaux  $\alpha$ . L'information importante à retenir dans ce cadre est qu'en cas d'erreurs iid, les droites de quantiles conditionnels théoriques sont simplement translatées verticalement en fonction de  $\alpha$ . On retrouve donc ce phénomène dans les droites de régression quantile, qui conservent une pente très similaire pour chaque valeur de  $\alpha$ , cette valeur oscillant autour de la valeur théorique.

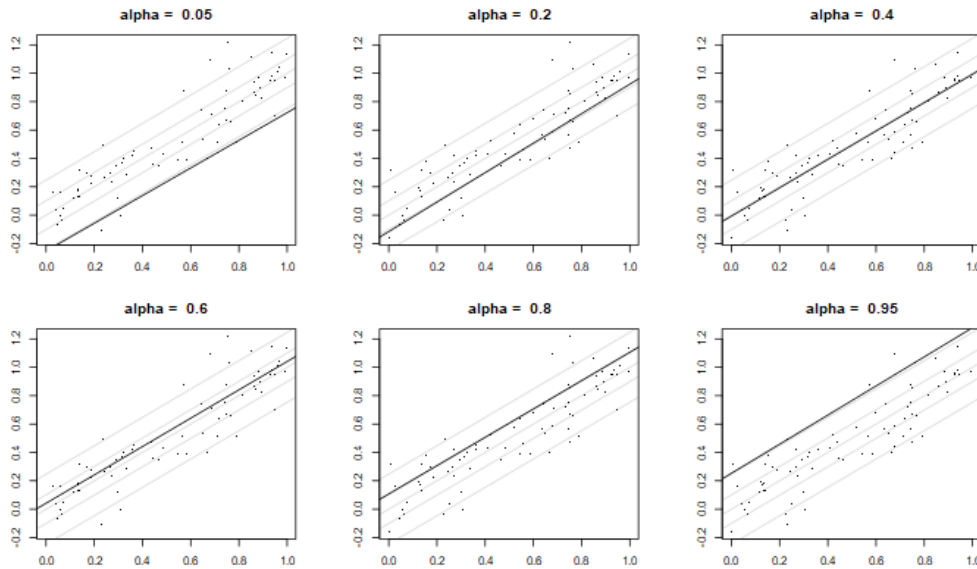


FIGURE 9 – Droites de régression quantile pour différentes valeurs de  $\alpha$  dans le cas d'un modèle linéaire avec erreurs iid. On voit que les droites de régression sont de pente presque identiques pour toutes les valeurs de  $\alpha$ , ce qui est un indicateur de l'homoscédasticité.

Lorsque l'hypothèse d'homoscédasticité n'est pas applicable, on obtient des droites ou courbes de régression qui ne sont plus simplement des translations verticales les unes des autres. On illustre un cas proche de celui donné dans [6] en FIGURE 10. On retrouve ce type de profils de courbes de quantiles dans les carnets de santé, pour surveiller par exemple la croissance des jeunes enfants.

**Application aux données réelles.** On illustre maintenant la régression quantile sur les données anthropométriques « ANSUR II » (*Anthropometric Survey of US Army Personnel*), disponibles en ligne<sup>1</sup>. Il s'agit d'un ensemble de 93 mesures effectuées sur plus de 6 000 militaires américains adultes, dont 4 082 hommes et 1 986 femmes. On utilisera dans les exemples suivants les données concernant les hommes. On utilise le package *quantreg* disponible en R, et développé par Roger Koenker pour effectuer la régression quantile. On cherche à expliquer dans cet exemple la variable *weightkg* en fonction de la variable *stature* et d'un intercept, autrement dit le poids en fonction de la taille. Le résultat de la régression est donné en FIGURE 11.

Dans cet exemple, les droites de régression sont de pente croissante, ce qui est un indicateur d'hétéroscédasticité, et plus précisément de variance plus grande pour les tailles élevées. On note que les profils de quantiles conditionnels sont de plus en plus étendus et pas simplement translatés verticalement lorsque la taille augmente,

1. <https://www.openlab.psu.edu/ansur2/>



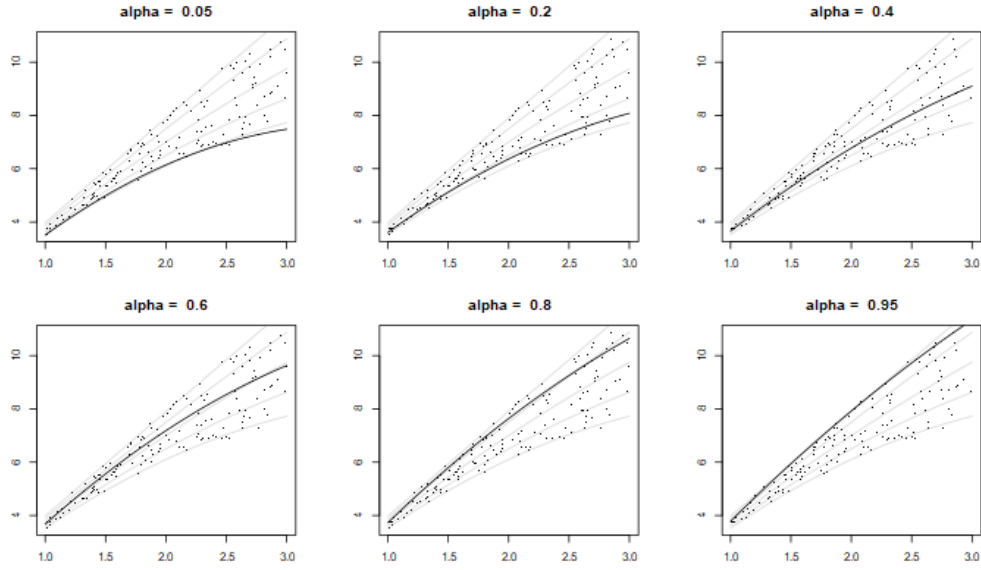


FIGURE 10 – Courbes de régression quantile pour différentes valeurs de  $\alpha$  dans le cas d'un modèle linéaire avec erreurs non iid. On note alors que les lignes de régression s'adaptent à l'allure du nuage de point selon les valeurs de  $\alpha$ .

ce que l'on retrouve dans les densités estimées. Ainsi, le poids est plus élevé pour les individus de taille plus grande, mais la densité est également plus aplatie, et on en déduit que la variabilité du poids est plus importante chez les individus de grande taille. Cet exemple reste très simple, et on renvoie à Koenker [6] pour d'autres exemples plus poussés, qui illustrent l'intérêt de la régression quantile.

Pour motiver la régression quantile multivariée de la prochaine partie, on peut également s'intéresser aux quantiles conditionnels d'une autre quantité, par exemple la circonférence des cuisses. On illustre le résultat de la régression quantile dans ce cadre en FIGURE 12.

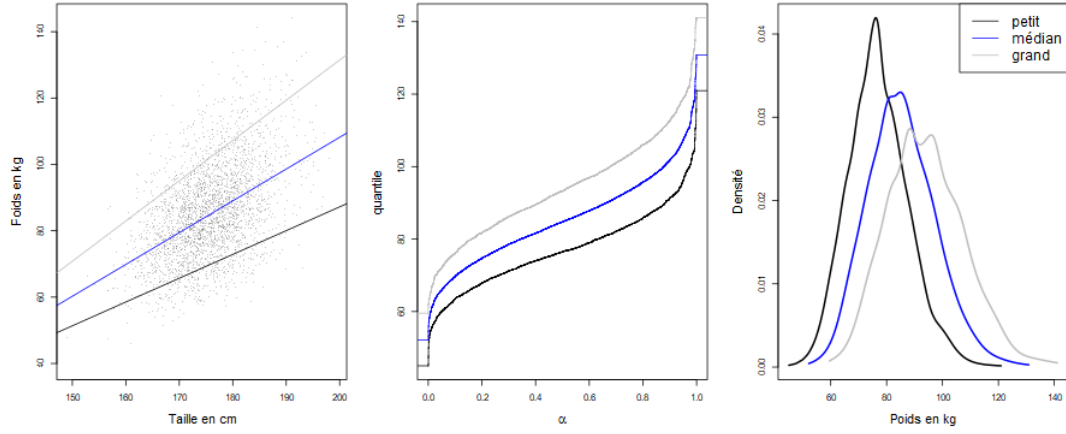


FIGURE 11 – Estimation de la fonction de quantiles conditionnels du poids (kg) pour trois déciles de taille (cm) (0.1, 0.5, 0.9), que l'on note respectivement « petit », « médian », « grand » (gauche), et densités estimées correspondantes (droite).

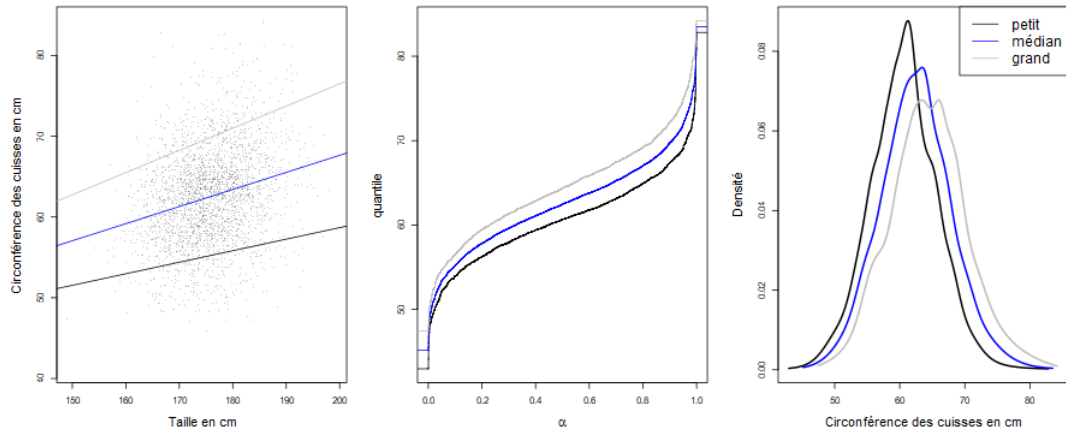


FIGURE 12 – Estimation de la fonction de quantiles conditionnels de la circonférence des cuisses (cm) pour trois déciles de taille (cm) (0.1, 0.5, 0.9), que l'on note respectivement « petit », « médian », « grand » (gauche), et densités estimées correspondantes (droite).

## 3.2 Régression quantile vectorielle par transport optimal

### 3.2.1 Lien entre régression quantile et transport optimal

Nous avons vu précédemment la notion de transport optimal et de régression quantile. Il est temps d'établir un lien entre ces deux notions. Dans cette partie nous allons nous appuyer fortement sur [2].

Soient  $X$  et  $Y$  deux variables alors nous avons montré précédemment que le quantile d'ordre  $\alpha$  de  $Y$  sachant  $X = x$  noté  $q_\alpha(x)$  pour  $\alpha \in [0, 1]$  est le minimiseur de la fonction  $\mathbb{E}[\rho_\alpha(Y - q_\alpha(x)|X = x)]$ . Si on considère une formulation linéaire du quantile c'est-à-dire  $q_\alpha(x) = \beta^\top X$  et la fonction coût  $\rho_\alpha$  telle que nous l'avons définie. Le problème de minimisation est équivalent à

$$\min_{\beta} \mathbb{E}[\alpha u + (1 - \alpha)v].$$

Avec  $u$  et  $v$  respectivement la partie positive et la partie négative du résiduel  $r = Y - q_\alpha(x)$ .

Or on peut écrire

$$\begin{aligned} \alpha u + (1 - \alpha)v &= \alpha u - (\alpha - 1)v \\ &= \alpha u - (\alpha - 1)u + (\alpha - 1)u - (\alpha - 1)v \\ &= \alpha u - \alpha u + u + (\alpha - 1)(u - v) \\ &= u + (\alpha - 1)r. \end{aligned}$$

Le problème est donc équivalent à

$$\begin{aligned} \min_{\beta} \mathbb{E}[u + (\alpha - 1)r] &\Leftrightarrow \min_{\beta} \mathbb{E}[(Y - \beta^\top X)^+ - (\alpha - 1)(Y - \beta^\top X)] \\ &\Leftrightarrow \min_{\beta} \mathbb{E}[(Y - \beta^\top X)^+ + (1 - \alpha)(\beta^\top X - Y)] \\ &\Leftrightarrow \min_{\beta} \mathbb{E}[(Y - \beta^\top X)^+ + (1 - \alpha)\beta^\top X] - (1 - \alpha)\mathbb{E}[Y] \\ &\Leftrightarrow \min_{\beta} \mathbb{E}[(Y - \beta^\top X)^+ + (1 - \alpha)\beta^\top X]. \end{aligned}$$

Pour résoudre ce problème, il est possible de le faire pour chaque  $\alpha$  fixé mais cette approche est moins robuste car ne garantit pas la croissance de la fonction quantile. Néanmoins il est proposé dans [2], pour estimer sur tout  $[0, 1]$  de poser le problème de minimisation intégré, en construisant la fonction  $\alpha \mapsto \beta_\alpha$  qui minimise

$$\min_{\beta_\alpha} \int_0^1 \mathbb{E}[(Y - \beta_\alpha^\top X)^+ + (1 - \alpha)\beta_\alpha^\top X] d\alpha.$$

Rappelons que  $U_\alpha = (Y - \beta_\alpha^\top X)^+$  est appelé déviation positive de  $Y$  par rapport à ses quantiles prédits  $\beta_\alpha^\top X$  et  $v_\alpha = (Y - \beta_\alpha^\top X)^- = U_\alpha - Y + \beta_\alpha^\top X$  est positif par définition, et donc le problème peut-être reformulé de la manière suivante

$$\begin{aligned} \min & \int_0^1 \mathbb{E}[U_\alpha + (1 - \alpha)\beta_\alpha^\top X] d\alpha \\ \text{s.c.} & \quad U_\alpha - Y + \beta_\alpha^\top X \geq 0, [V_\alpha]. \end{aligned}$$

Cette formulation est appelée dans [2] la formulation duale du problème initial, par conséquent la formulation suivante appelée formulation primale du problème est obtenue en prenant le dual du dual par une procédure de minmax.

$$\min_{U_\alpha \geq 0, \beta_\alpha} \max_{V_\alpha \geq 0} \int_0^1 \mathbb{E}[U_\alpha + (1 - \alpha)\beta_\alpha^\top X + V_\alpha Y - V_\alpha U_\alpha - V_\alpha \beta_\alpha^\top X] d\alpha,$$

ou encore

$$\max_{V_\alpha \geq 0} \left[ \int_0^1 \mathbb{E}[V_\alpha Y] d\alpha + \min_{U_\alpha \geq 0, \beta_\alpha} \int_0^1 \mathbb{E}[(1 - V_\alpha)u_\alpha + (1 - \alpha - V_\alpha)\beta_\alpha^\top X] d\alpha \right].$$

Il a été prouvé dans [2] et [3] que ce problème primal est équivalent au problème

$$\begin{aligned} \max_{(U, X, Y) \sim \pi} \mathbb{E}_\pi[UY] \\ \text{s.c. } U \sim \mathcal{U}([0, 1]) \\ (X, Y) \sim \nu \\ \mathbb{E}[X|Y] = \mathbb{E}[X]. \end{aligned}$$

Remarquons que cette dernière formulation sans la dernière contrainte est équivalente à la formulation du problème de transport optimal de Monge-Kantorovich.

### 3.2.2 Vers la dimension supérieure

Nous avons vu la notion de quantile multivarié, et avons parlé de la régression quantile dans le cas bivarié. Il serait intéressant, à présent, de voir comment on peut étendre cette dernière en dimension supérieure. C'est ce que l'on appelle la régression quantile vectorielle.

En dimension supérieure  $Y$  est une variable aléatoire prenant ses valeurs dans  $\mathbb{R}^d$  et on admet que  $Y$  admet un moment d'ordre 2 fini. L'idée proposée dans [3] est de construire la fonction  $(u, x) \mapsto Q_{Y|X}(u, x)$  de  $[0, 1]^d \times \mathbb{R}^q$  dans  $\mathbb{R}^d$ .

La définition de quantile utilisée dans le cadre de la régression quantile vectorielle est celle proposée par Chernozhukov, Galichon, Hallin et Henry [4], que nous avons présentée dans la première partie de ce rapport. La seule différence notable est que la distribution de référence sera ici non plus l'uniforme sphérique, mais l'uniforme sur l'hypercube  $[0, 1]^d$ . Le choix de la distribution de référence étant libre tant que les conditions de régularité sont respectées, ce choix apparaît naturel dans le cadre de la régression quantile.

Comme dans le cas univarié, pour des raisons de robustesse, on a les propriétés suivantes :

1. La monotonie de la fonction quantile, ce qui implique qu'elle est le gradient d'une certaine fonction convexe, et également
2. Les quantiles conditionnels vérifient

$$Y = Q_{Y|X}(U, X), \quad U|X \sim \mathcal{U}([0, 1]^d).$$

Déterminer une telle fonction  $Q_{Y|X}$  revient donc à trouver la fonction qui transporte optimalement pour chaque  $X$  fixé, la distribution  $U$  sur la distribution  $Y$ . En

considérant le coût quadratique, le théorème de Brenier nous garantit l'existence d'une telle fonction. De plus  $Q_{Y|X} = \nabla\psi$ , avec  $\psi$  une certaine fonction convexe.

En supposant toujours que le modèle est linéaire dans le cas multivarié, on a

$$Q_{Y|X}(U, X) = \beta(U)^\top X, \quad U|X \sim \mathcal{U}([0, 1]^d).$$

Il est proposé dans [3] de relaxer la dernière contrainte dans le cas multivarié, et les auteurs montrent alors que l'on a

$$Q_{Y|X}(U, X) = \beta(U)^\top X, \quad U \sim \mathcal{U}([0, 1]^d) \quad \text{et} \quad \mathbb{E}[X|U] = \mathbb{E}[X].$$

De plus

$$\forall (u, x) \in \mathbb{R}^d \times \mathbb{R}^q, \quad Q_{Y|X=x}(u, x) = \beta(u)^\top x \quad \text{et} \quad Q_{Y|X=x}(u, x) = \nabla_x \psi(u).$$

Donc finalement

$$\forall (u, x) \in \mathbb{R}^d \times \mathbb{R}^q, \beta(u)^\top x = \nabla_x \psi(u).$$

En s'appuyant sur ce résultat, Chernozhukov, Galichon, Hallin et Henry [4] proposent une méthode computationnelle pour réaliser la régression quantile vectorielle. Pour plus de détails concernant cette méthode, on se réfère à [4]. Dans la partie suivante nous allons présenter quelques résultats obtenus en utilisant cette méthode.

### 3.2.3 Application aux données ANSUR II

Dans la partie précédente, nous avons appliqué la régression quantile classique telle qu'introduite par Koenker aux données « ANSUR II ». On avait ainsi expliqué indépendamment deux variables en fonction du poids, tout d'abord le poids (kg) puis la circonférence de la cuisse (cm). On illustre maintenant la régression quantiles en dimension 2 sur ces mêmes données. Les quantiles conditionnels en dimension 2 obtenus par cette méthode sont représentés en FIGURE 13. Nous avons utilisé l'implémentation en Python développée par Pierre Delanoue et disponible sur Github <sup>2</sup> pour créer ces visualisations.

Dans ce cas, on observe la dépendance du couple (*Poids*, *Circonférence des cuisses*) en fonction de trois déciles de taille. En particulier, on remarque que les surfaces obtenues ont même allure pour les trois déciles de taille, et sont simplement translatées verticalement. On en déduit que le comportement des variables *Poids* et *Circonférence des cuisses* est indépendant de la taille. On note également que les deux variables sont corrélées positivement avec les deux axes  $U_1$  et  $U_2$ , ce qui signifie que non seulement les variables *Poids* et *Circonférence des cuisses* sont corrélées avec la taille comme nous l'avons observé en dimension 1, mais elles sont également corrélées positivement entre elles.

---

2. [https://github.com/DatenBiene/Vector\\_Quantile\\_Regression](https://github.com/DatenBiene/Vector_Quantile_Regression)

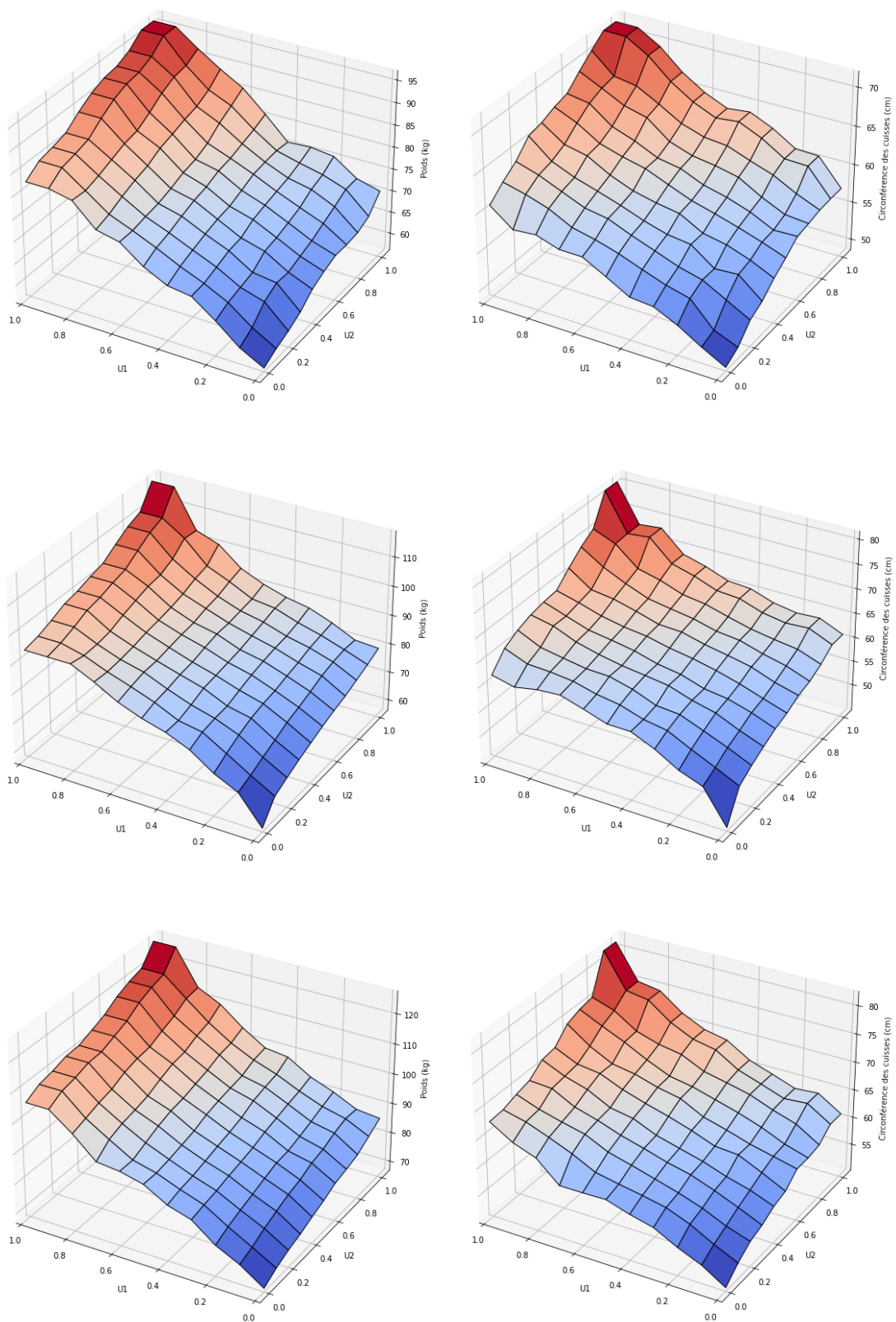


FIGURE 13 – Quantiles conditionnels en dimension 2 pour les trois déciles de tailles : 0.1 (en haut), 0.5 (au milieu), et 0.9 (en bas). À gauche on représente les quantiles du poids (kg), et à droite de la circonférence de la cuisse (cm).

## 4 Conclusion

La notion de quantile traditionnelle peut être étendue à la dimension supérieure. Cela implique une redéfinition de la notion de quantile comme un problème d'optimisation, qui nous permet ainsi de la relier au problème de transport optimal.

Il est alors naturel de vouloir étendre, de même, la régression quantile à la dimension supérieure.

Néanmoins, les concepts mathématiques

## Références

- [1] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4) :375–417, 1991.
- [2] Guillaume Carlier, Victor Chernozhukov, Gwendoline De Bie, and Alfred Galichon. Vector quantile regression and optimal transport, from theory to numerics. *Empirical Economics*, Aug 2020.
- [3] Guillaume Carlier, Victor Chernozhukov, and Alfred Galichon. Vector quantile regression : An optimal transport approach. *The Annals of Statistics*, 44(3) :1165 – 1192, 2016.
- [4] Victor Chernozhukov, Alfred Galichon, Marc Hallin, and Marc Henry. Monge–Kantorovich depth, quantiles, ranks and signs. *The Annals of Statistics*, 45(1) :223 – 256, 2017.
- [5] Leonid Kantorovich. On the transfert of masses (in russian). *Doklady Akademii Nauk*, 32(2) :227 – 229, 1942.
- [6] R. Koenker, A. Chesher, and M. Jackson. *Quantile Regression*. Econometric Society Monographs. Cambridge University Press, 2005.
- [7] Robert McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Mathematical Journal*, 80 :309–323, 1995.
- [8] G. Monge. *Mémoire sur la théorie des déblais et des remblais*. De l’Imprimerie Royale, 1781.
- [9] Gabriel Peyré and Marco Cuturi. Computational optimal transport, 2020.
- [10] Cédric Villani. *Topics in Optimal Transportation*. Graduate studies in mathematics. American Mathematical Society, 2003.