

## 1 Introduction

La régression ordinaire (OLS) ne s'intéresse qu'à la moyenne de la distribution conditionnelle, or il est possible de se trouver dans le cas de données où la moyenne ne suffit pas à tout modéliser. C'est dans ce type de contexte que la régression quantile peut être utile puisqu'elle prend en compte l'ensemble de la distribution conditionnelle de la variable d'intérêt.

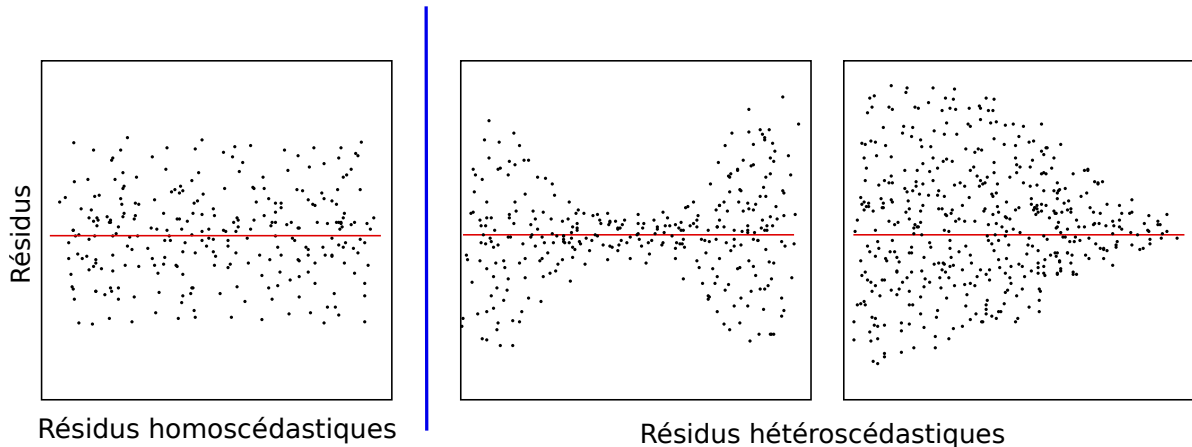


FIGURE 1 – Exemple de jeux de données où la régression classique est limitée.

(XXX : reconnecter avec le code)

## 2 Rappels : Médiane / quantiles

### 2.1 Médiane

**Définition 2.1** (Médiane). Soit  $y_1, \dots, y_n \in \mathbb{R}$ , on a :

$$\text{Med}_n(y_1, \dots, y_n) \in \arg \min_{\mu \in \mathbb{R}} \frac{1}{n} \sum_{i=1}^n |y_i - \mu| . \quad (1)$$

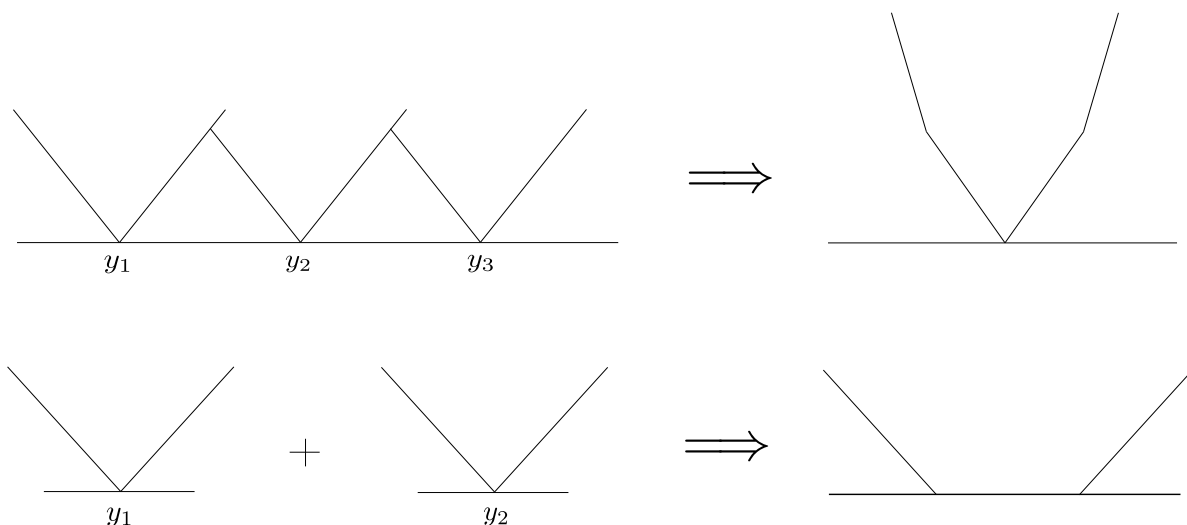


FIGURE 2 – Fonction d’optimisation  $f : \mu \rightarrow \frac{1}{n} \sum_{i=1}^n |y_i - \mu|$  pour  $n = 2$  et  $n = 3$

Considérons deux cas canoniques :

- Pour 2 observations  $y_1, y_2$  tous les points entre  $y_1$  et  $y_2$  sont des minimums globaux.
- Pour 3 observations  $y_1, y_2, y_3$  la fonction d’optimisation admet un unique minimum global.

Le problème d’optimisation défini en (2.1) dépend de la parité des points :

En pratique, pour  $y_{(1)} \leq \dots \leq y_{(n)}$  on prend :

$$\text{Med}_n(y_1, \dots, y_n) = \begin{cases} y_{\lfloor \frac{n}{2} \rfloor + 1}, & \text{si } n \text{ est impair} \\ \frac{y_{\lfloor \frac{n}{2} \rfloor} + y_{\lfloor \frac{n}{2} \rfloor + 1}}{2}, & \text{si } n \text{ est pair} \end{cases} \quad (2)$$

**Remarque 2.1.** Considérons la fonction

$$f : \begin{cases} \mathbb{R} & \rightarrow & \mathbb{R} \\ \mu & \mapsto & \frac{1}{n} \sum_{i=1}^n |y_i - \mu| \end{cases} \quad (3)$$

Cette fonction est une fonction continue, convexe, et donc un minimum local est un minimum global. De plus la fonction est coercive ce qui assure qu’un minimum global existe. La fonction n’est pas “lisse” (en certains points la fonction est non-différentiable).

**Remarque 2.2.** Nous pouvons également reformuler le problème donné par (2.1) en :

$$\begin{aligned} \min_{\mu, t_1, \dots, t_n} & \frac{1}{n} \sum_{i=1}^n t_i \\ \text{s.c.} & \quad \forall i \in \llbracket 1, n \rrbracket, -t_i \leq y_i - \mu \leq t_i \\ & \quad \forall i \in \llbracket 1, n \rrbracket, 0 \leq t_i \end{aligned} \quad (4)$$

Les deux problèmes ainsi posés sont équivalents. (XXX à montrer)

Dans les années 1940, George Dantzig s'intéresse à la programmation linéaire (PL). L'intérêt principal de ces méthodes est leur robustesse (cf. par exemple <http://josephsalmon.eu/STAT593.html>).

Le package CVXPY de Python contient diverses méthodes numériques de résolution pour ces problèmes.

**Contrepartie "théorique" :** Soit  $Y$  une variable aléatoire réelle, alors

$$\text{Med}(Y) \in \arg \min_{\mu \in \mathbb{R}} \mathbb{E}_Y (|Y - \mu|) . \tag{5}$$

Rappel : cas important de la moyenne/espérance

$$\mathbb{E}(Y) = \arg \min_{\mu \in \mathbb{R}} \mathbb{E}_Y (|Y - \mu|^2) \tag{6}$$

## 2.2 Quantile

**Définition 2.2** (Quantile). (Usuelle) Le quantile de niveau  $\alpha$  est défini,  $\forall \alpha \in ]0, 1[$ ,

$$q_\alpha(y_1, \dots, y_n) = \inf \{t \in \mathbb{R} : F_n(t) \geq \alpha\} . \tag{7}$$

avec

$$F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{y_i \leq t\}} . \tag{8}$$

Interprétation pour  $\alpha = 0.75$  :

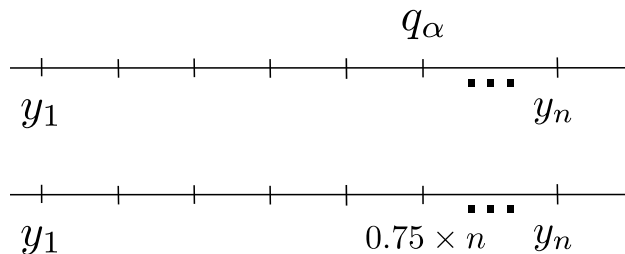


FIGURE 3 – Schéma d'explication du quantile  $\alpha = 0.75$

(XXX schéma peu clair ? message ?)

Pour une variable aléatoire réelle  $Y$ , on note le quantile de niveau  $\alpha$  de cette variable aléatoire de la manière suivante :

$$q_\alpha(Y) = \inf \{t \in \mathbb{R} : F_Y(t) \geq \alpha\} \tag{9}$$

où  $F_Y$  est la fonction de répartition de  $Y$ .

**Proposition 2.1.**

$$q_\alpha(Y) \in \arg \min_{\mu \in \mathbb{R}} \underbrace{\alpha \int_{y > \mu} |y - \mu| dF_Y(y) + (1 - \alpha) \int_{y \leq \mu} |y - \mu| dF_Y(y)}_{\psi(\mu)}$$

*Démonstration.*

$$\begin{aligned}\psi(\mu) &= \alpha \int_{\mu}^{+\infty} (y - \mu) dF_Y(y) + (1 - \alpha) \int_{-\infty}^{\mu} (\mu - y) dF_Y(y) \\ \psi'(\mu) &= -\alpha \int_{\mu}^{+\infty} dF_Y(y) + (1 - \alpha) \int_{-\infty}^{\mu} dF_Y(y) \\ \psi'(\mu) &= -\alpha \mathbb{P}(Y \geq \mu) + (1 - \alpha) \mathbb{P}(Y \leq \mu) \\ \psi'(\mu) &= -\alpha(1 - F_Y(\mu)) + (1 - \alpha)F_Y(\mu) .\end{aligned}$$

(XXX à peaufiner)

On veut trouver un point critique  $\psi'(\mu) = 0$  et résout pour cela :

$$\begin{aligned}\psi'(\mu) = 0 &\iff 0 = F_Y(\mu)(\alpha + 1 - \alpha) - \alpha \\ &\iff \alpha = F_Y(\mu) \\ \text{(inversion)} &\iff \mu = F_Y^{-1}(\alpha)\end{aligned}$$

□

**Remarque 2.3.** On définit la fonction perte  $l_\alpha$  :

$$l_\alpha : \begin{cases} \mathbb{R} & \longrightarrow & \mathbb{R} \\ x & \longmapsto & \begin{cases} -(1 - \alpha)x, & \text{si } x \leq 0 \\ \alpha x, & \text{si } x \geq 0 \end{cases} \end{cases}$$


On peut reformuler l'expression précédente et obtenir  $l_\alpha(x) = \alpha|x|\mathbb{1}_{x>0} + (1 - \alpha)|x|\mathbb{1}_{x\leq 0}$ . On a ainsi :

$$\psi(\mu) = \mathbb{E}_Y [l_\alpha(Y - \mu)] . \quad (10)$$

*Démonstration.*

$$\begin{aligned}\psi(\mu) &= \alpha \int_{y>\mu} |y - \mu| dF_Y(y) + (1 - \alpha) \int_{y\leq\mu} |y - \mu| dF_Y(y) \\ &= \int \alpha|y - \mu|\mathbb{1}_{y>\mu} + (1 - \alpha)|y - \mu|\mathbb{1}_{y\leq\mu} dF_Y(y) \\ &= \int (l_\alpha(y - \mu)) dF_Y(y) \\ &= \mathbb{E}_Y(l_\alpha(Y - \mu))\end{aligned}$$

□

La fonction coût que l'on a défini  $l_\alpha$ , s'appelle la “perte Flipper” ( : *pinball loss*), sa forme est une trajectoire de balle de flipper, de pente qui dépend de  $\alpha$ .

Interprétation :

- Cas  $\alpha$  petit  $\simeq 0$  : Les erreurs positives ne sont pas ou peu pénalisées  $\Rightarrow$  on peut prendre un point vers la gauche de l'échantillon
- Cas  $\alpha \simeq \frac{1}{2}$  : On récupère la médiane
- Cas  $\alpha$  grand  $\simeq 1$  : Les erreurs négatives ne sont pas ou peu pénalisées  $\Rightarrow$  on peut prendre un point à droite de l'échantillon

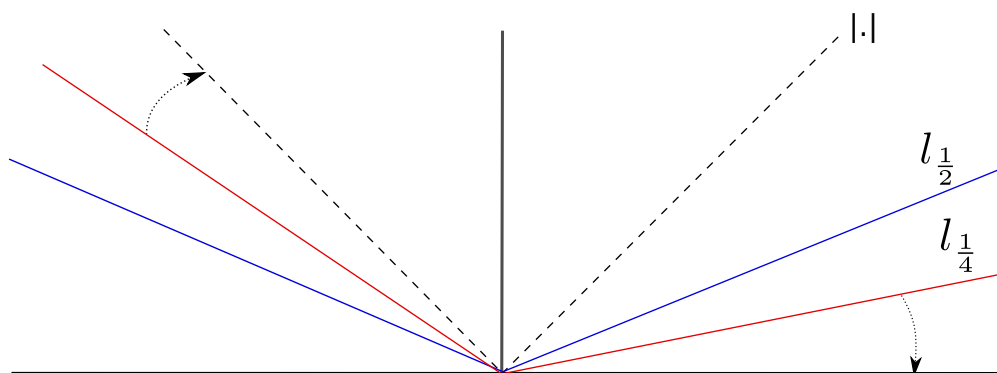


FIGURE 4 – La perte flipper

**Explications** On cherche à minimiser la fonction  $f_\alpha$  définie comme suit :

$$f_\alpha : \mu \rightarrow \frac{1}{n} \sum_{i=1}^n l_\alpha(y_i - \mu) . \tag{11}$$

On note  $\hat{\mu}_\alpha$  un tel minimiseur.

- Si  $\alpha \simeq 0$ , il y a une pénalisation avec la distance du point à  $\hat{\mu}_\alpha$ . Comme on cherche à minimiser la fonction, on va avoir tendance à faire bouger le  $\hat{\mu}_\alpha$  vers la gauche.
- Si  $\alpha \simeq \frac{1}{2}$  alors on va avoir tendance à équilibrer les erreurs autour de la cible  $\hat{\mu}_\alpha$ , on se rapproche de la médiane.
- Si  $\alpha \simeq 1$ , il va falloir faire bouger le  $\hat{\mu}_\alpha$  vers la droite pour minimiser la fonction.

(XXX améliorer le texte précédent)

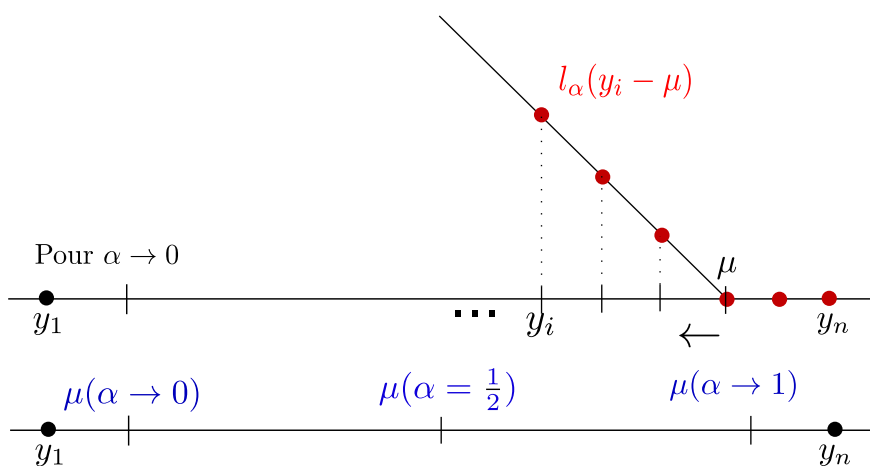


FIGURE 5 – Explication pour le choix de  $\hat{\mu}_\alpha$ , cas où  $\alpha = 0$  pour le schéma du haut

La perte "flipper",  $l_\alpha$ , permet de donner plus de poids aux observations les plus lointaines de  $\hat{\mu}_\alpha$ .

La régression quantile étant une régression "déséquilibrée", le paramètre  $\alpha$  règle le niveau de la pente de pondération ( $\alpha$  d'un côté,  $(1 - \alpha)$  de l'autre) de la fonction de perte associée à la régression et reflète ainsi le déséquilibre désiré dans la prévision entre sur/sous-estimation.

**Théorème 2.1.**

$$\hat{\mu}_\alpha \in \arg \min_{\mu} \mathbb{E}_Y(l_\alpha(Y - \mu)) . \quad (12)$$

est le quantile  $q_\alpha(Y)$ .

**Remarque 2.4.**  $\mu \rightarrow \mathbb{E}_Y(l_\alpha(Y - \mu))$  est convexe pour tout  $\alpha \in ]0, 1[$ . (Preuve : une moyenne de fonction convexe est convexe).

Contre partie empirique :

$$q_{\alpha,n}(y_1, \dots, y_n) \in \arg \min_{\mu \in \mathbb{R}} \frac{1}{n} \sum_{i=1}^n l_\alpha(y_i - \mu) . \quad (13)$$

**Remarque 2.5.** Ce problème est aussi formulable comme un problème de programmation linéaire (🇬🇧 : linear programming (LP)). (XXX mettre les espaces dans lesquels vivent les indices dans le min dessous)

$$\begin{aligned} \min_{u,v,\mu} \quad & \alpha \sum_{i=1}^n u_i + (1 - \alpha) \sum_{i=1}^n v_i \\ \text{s.c.} \quad & u_i - v_i = y_i - \mu \\ & u_i \geq 0 \\ & v_i \geq 0 \end{aligned} \quad (14)$$

Preuve : on sépare les parties positives et négatives dans les valeurs absolues, en se servant des relations suivantes :

$$\begin{cases} x & = x_+ - x_- \\ |x| & = x_+ + x_- \end{cases}$$

### 3 Régression quantile

$y_1, \dots, y_n \in \mathbb{R}$  observations

$x_1, \dots, x_n \in \mathbb{R}^p$  variables explicatives

**Définition 3.1.** Soit  $\alpha \in ]0, 1[$ . On appelle régression quantile les coefficients :

$$\hat{\beta}^\alpha \in \operatorname{argmin}_{\beta \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^n l_\alpha(y_i - x_i^\top \beta) \quad (15)$$

$$\text{avec } X = \begin{bmatrix} x_1^\top \\ \vdots \\ x_n^\top \end{bmatrix} \in \mathbb{R}^{n \times p} . \quad (16)$$

- $\hat{\beta}^\alpha = \hat{\beta}^\alpha(X, y)$
- $\hat{\beta}^\alpha(X, c.y) = c\hat{\beta}^\alpha(X, y)$ , pour toute constante  $c > 0$  : équivariance d'échelle (🇬🇧 : scale equivariant)
- $\hat{\beta}^\alpha(X, y + Xv) = \hat{\beta}^\alpha(X, y) + v$ , pour  $v \in \mathbb{R}^p$  : équivariance en régression.

XXX : texte manquant...  $h : \mathbb{R} \rightarrow \mathbb{R}$  fonction croissante.

$$q_{h(Y)}(\alpha) = h(q_Y(\alpha)) . \quad (17)$$

*Démonstration.* Utiliser la définition des quantiles.

□

**Exemple :**

$$h(y) = \min(\max(y, \tau); \bar{\tau}) . \quad (18)$$

C'est le cas où l'on n'observe pas  $y_1, \dots, y_n$  mais

$$\begin{cases} \tau & \text{si } y_i \leq \tau \\ y_i & \text{si } \tau \leq y_i \leq \bar{\tau} \\ \bar{\tau} & \text{si } y_i \geq \bar{\tau} \end{cases}$$

(Kroenker 2005) XXX : faire bibliographie.

Autre cas d'intérêt de la régression quantile : le cas hétéroscédastique c'est à dire  $(y_i - x_i^\top \beta^*)$  ont des variances différentes.

**Remarque 3.1.** *L'estimateur  $\beta$  dépend du niveau de quantile  $\alpha$  que l'on spécifie.*

Comportement avec des variables catégorielles :

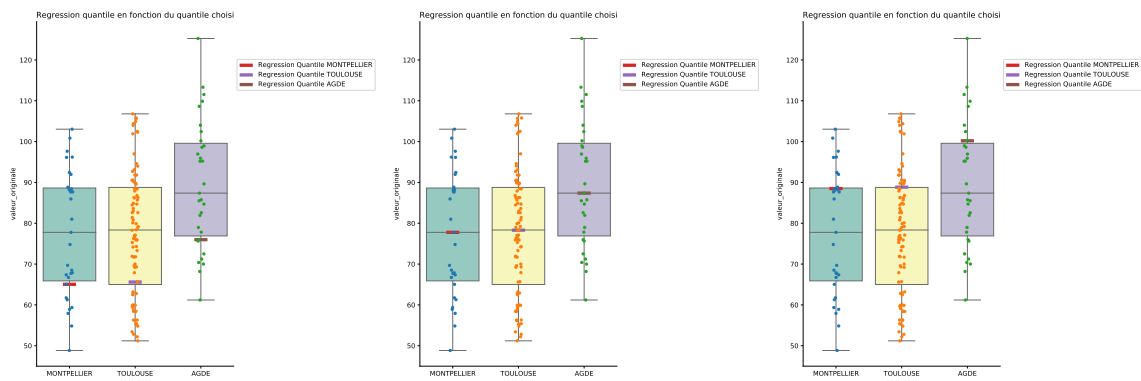


FIGURE 6 – Régression quantile pour des variables catégorielles ( $q_{0.25}$ ,  $q_{0.5}$  et  $q_{0.75}$ )

En prenant comme exemple le jeu de données de la pollution en fonction des villes de France, on observe que notre régression quantile retourne bien les mêmes valeurs qu'une étude basique des données. (Médiane et quantile)

**Exemple:** Prenons l'exemple du jeu de données de Boscovich sur l'aplanissement de la Terre. XXX source à citer. On observe les régressions quantile (le quantile est de niveau  $\tau$ ).

(XXX taille figure trop petite sur la figure régression quantile.)

A l'aide du package 'CVXPY', nous avons estimé le paramètre "Arc length of 1 degree of Latitude" en fonction du  $\sin^2$  de la Latitude des villes. Nous avons ainsi obtenu 4 droites de régressions, correspondant à toutes les valeurs possibles de  $\tau$ . Il n'y a, en effet, que 4 régressions différentes pour toutes les valeurs possibles des quantiles. Pour plus de précisions, se référer au notebook associé.

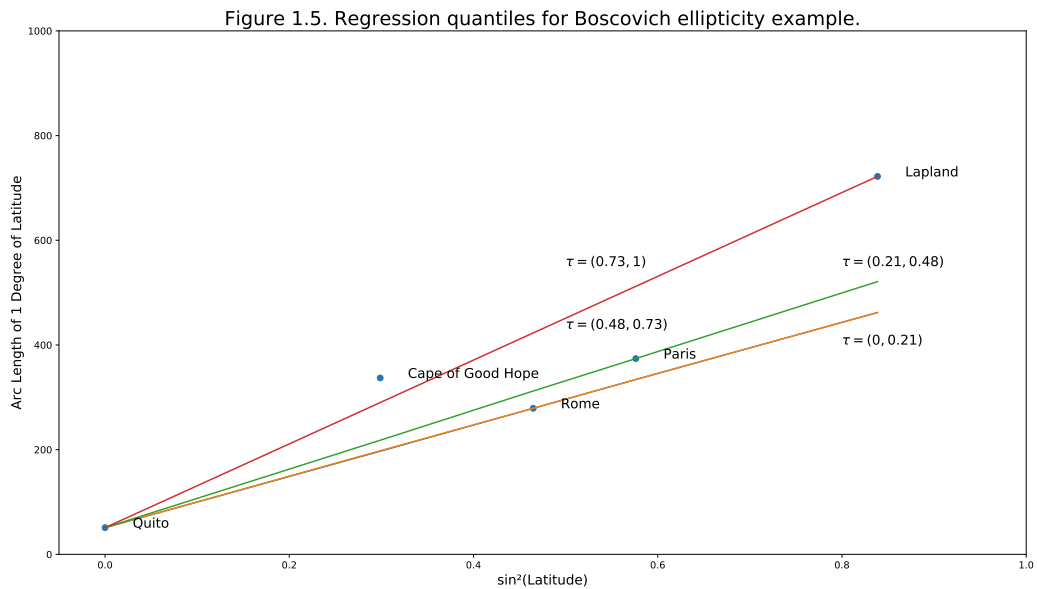


FIGURE 7 – Régressions quantiles

## 4 Conclusion

Comparée à la régression linéaire classique, la régression quantile nous offre deux avantages principaux : la robustesse face aux valeurs aberrantes et la flexibilité par rapport à la distribution des données.

## 5 Source

La figure 1 est inspirée de : [[researchgate.net/figure/Example-of-homoscedasticity-Ideally-residuals-are-randomly-scattered-around-0-the\\_fig2\\_319091775](https://researchgate.net/figure/Example-of-homoscedasticity-Ideally-residuals-are-randomly-scattered-around-0-the_fig2_319091775)]

La figure 7 est inspirée de : [<https://www.cambridge.org/core/books/quantile-regression/C18AE7BCF3EC43C16937390D44A328B1#>]