STAT 593
Quantile regression

Joseph Salmon
http://josephsalmon.eu

Télécom Paristech, Institut Mines-Télécom
&
University of Washington, Department of Statistics
(Visiting Assistant Professor)

Outline

Definition and reminder

Properties

Non-parametric extension / crossing

Various properties of quantiles

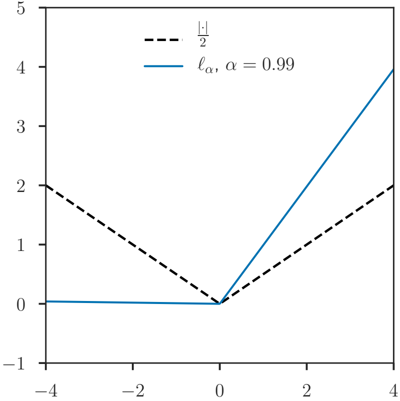
Limits of quantile and more meaningful risk measure

Table of Contents

"Pinball loss" / quantile regression

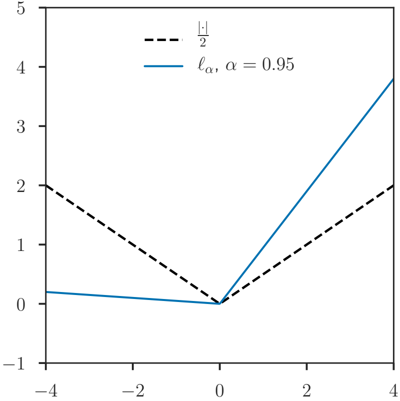
$$\ell_\alpha(x) = \begin{cases} -(1-\alpha)x & \text{if } x \leq 0 \\ \alpha x & \text{if } x \geq 0 \end{cases}$$

$$= \alpha|x|\mathbb{1}_{\{x\geq 0\}} + (1-\alpha)|x|\mathbb{1}_{\{x\leq 0\}}$$



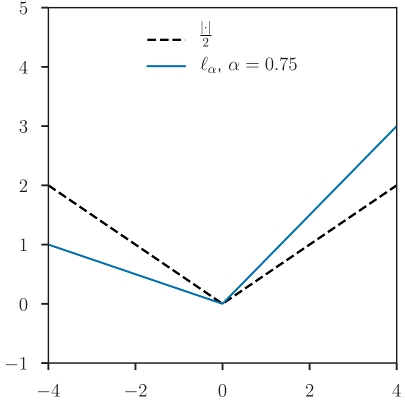
"Pinball loss" / quantile regression

$$\ell_\alpha(x) = \begin{cases} -(1-\alpha)x & \text{if } x \le 0 \\ \alpha x & \text{if } x \ge 0 \end{cases}$$
$$= \alpha |x| \mathbb{1}_{\{x \ge 0\}} + (1-\alpha)|x| \mathbb{1}_{\{x \le 0\}}$$



"Pinball loss" / quantile regression

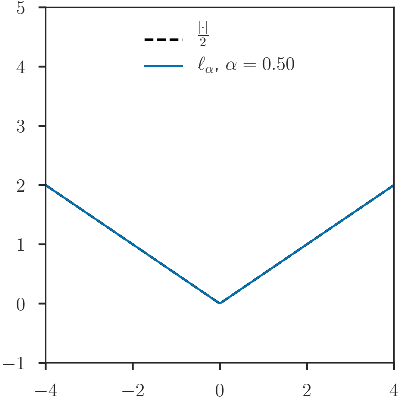
$$\ell_\alpha(x) = \begin{cases} -(1-\alpha)x & \text{if } x \leq 0 \\ \alpha x & \text{if } x \geq 0 \end{cases}$$
$$= \alpha|x|\mathbb{1}_{\{x\geq 0\}} + (1-\alpha)|x|\mathbb{1}_{\{x\leq 0\}}$$



"Pinball loss" / quantile regression

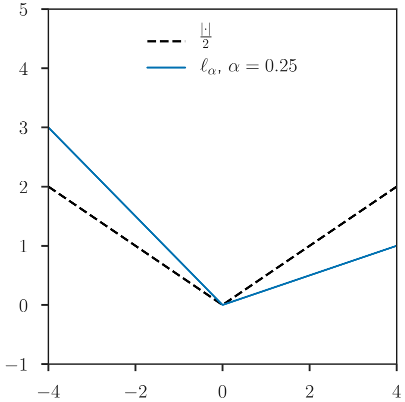
$$\ell_\alpha(x) = \begin{cases} -(1-\alpha)x & \text{if } x \le 0 \\ \alpha x & \text{if } x \ge 0 \end{cases}$$

$$= \alpha|x|\mathbb{1}_{\{x\ge 0\}} + (1-\alpha)|x|\mathbb{1}_{\{x\le 0\}}$$



"Pinball loss" / quantile regression

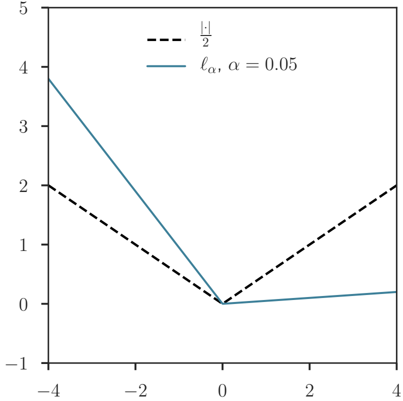
$$\ell_\alpha(x) = \begin{cases} -(1-\alpha)x & \text{if } x \leq 0 \\ \alpha x & \text{if } x \geq 0 \end{cases}$$
$$= \alpha|x|\mathbb{1}_{\{x\geq 0\}} + (1-\alpha)|x|\mathbb{1}_{\{x\leq 0\}}$$



"Pinball loss" / quantile regression

$$\ell_\alpha(x) = \begin{cases} -(1-\alpha)x & \text{if } x \le 0 \\ \alpha x & \text{if } x \ge 0 \end{cases}$$

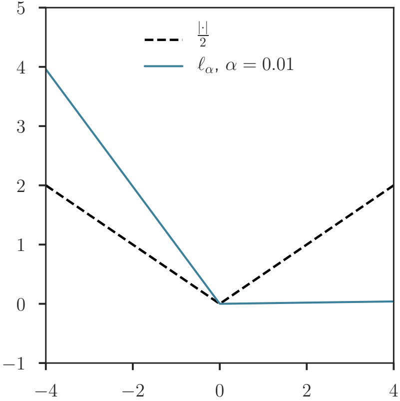
$$= \alpha|x|\mathbb{1}_{\{x\ge 0\}} + (1-\alpha)|x|\mathbb{1}_{\{x\le 0\}}$$



"Pinball loss" / quantile regression

$$\ell_\alpha(x) = \begin{cases} -(1-\alpha)x & \text{if } x \leq 0 \\ \alpha x & \text{if } x \geq 0 \end{cases}$$

$$= \alpha|x|\mathbb{1}_{\{x\geq 0\}} + (1-\alpha)|x|\mathbb{1}_{\{x\leq 0\}}$$



Link between pinball loss and quantile regression

Theorem

Define

$$\breve{\mu} \in \operatorname*{arg\,min}_{\mu \in \mathbb{R}} \mathbb{E}_F\big(\ell_\alpha(X-\mu)\big)$$

when $\ell_\alpha(x) := \alpha|x|\mathbb{1}_{\{x \geq 0\}} + (1-\alpha)|x|\mathbb{1}_{\{x \leq 0\}}$ then

$$\breve{\mu}(F,\rho) = F^{-1}(\alpha) := q_X(\alpha) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}$$

is[1] the α-quantile of the distribution F.

<u>Rem</u>: when $\alpha = 1/2$ then one recovers the ℓ_1 loss / median

(1) R. Koenker and G. Bassett. "Regression quantiles". In: *Econometrica* 46.1 (1978), pp. 33–50.

Quantile regression

Recall the regression setting: $X \in \mathbb{R}^{n \times p}$ and $\mathbf{y} \in \mathbb{R}^n$

Definition

For $\ell_\alpha(x) := \alpha |x| \mathbb{1}_{\{x \geq 0\}} + (1-\alpha)|x| \mathbb{1}_{\{x \leq 0\}}$ the α-th quantile regression estimator is defined as follows:

$$\hat{\beta}^\alpha \in \underset{\beta \in \mathbb{R}^p}{\arg\min} \sum_{i=1}^n \ell_\alpha(y_i - \langle x_i , \beta \rangle)$$

<u>Rem</u>: one recovers the LAD when $\alpha = 1/2$

Quantile regression

Recall the regression setting: $X \in \mathbb{R}^{n \times p}$ and $\mathbf{y} \in \mathbb{R}^n$

Definition

For $\ell_\alpha(x) := \alpha|x|\mathbb{1}_{\{x \geq 0\}} + (1-\alpha)|x|\mathbb{1}_{\{x \leq 0\}}$ the α-th quantile regression estimator is defined as follows:

$$\hat{\beta}^\alpha \in \operatorname*{arg\,min}_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \ell_\alpha(y_i - \langle\, x_i \,,\, \beta \,\rangle)$$

Rem: one recovers the LAD when $\alpha = 1/2$

Rem: it is the MLE for the asymmetric Laplace distribution

$$f_\theta(x) = \theta(1-\theta)\exp(-\ell_\alpha(x))$$

Equivariance properties

- the α-th quantile regression is scale equivariant, *i.e.*,

$$\forall c > 0, \quad \hat{\beta}^{\alpha}(X, c \cdot \mathbf{y}) = c \cdot \hat{\beta}^{\alpha}(X, \mathbf{y})$$

- the α-th quantile switches for negative values, *i.e.*,

$$\forall c < 0, \quad \hat{\beta}^{\alpha}(X, c \cdot \mathbf{y}) = c \cdot \hat{\beta}^{1-\alpha}(X, \mathbf{y})$$

(1)

Equivariance properties

- the α-th quantile regression is scale equivariant, *i.e.,*

$$\forall c > 0, \quad \hat{\beta}^{\alpha}(X, c \cdot \mathbf{y}) = c \cdot \hat{\beta}^{\alpha}(X, \mathbf{y})$$

- the α-th quantile switches for negative values, *i.e.,*

$$\forall c < 0, \quad \hat{\beta}^{\alpha}(X, c \cdot \mathbf{y}) = c \cdot \hat{\beta}^{1-\alpha}(X, \mathbf{y})$$

proof: take $c < 0$

$$\begin{aligned}\ell_\alpha(cy_i - \langle\, x_i \,,\, \beta\,\rangle) = \alpha|cy_i - \langle\, x_i \,,\, \beta\,\rangle\,|\mathbb{1}_{\{cy_i \geq \langle\, x_i \,,\, \beta\,\rangle\}}\\ + (1-\alpha)|cy_i - \langle\, x_i \,,\, \beta\,\rangle\,|\mathbb{1}_{\{cy_i \leq \langle\, x_i \,,\, \beta\,\rangle\}}\end{aligned}$$

[1] $c < 0$ reverses the inequality

Equivariance properties

- the α-th quantile regression is scale equivariant, *i.e.*,

$$\forall c > 0, \quad \hat{\beta}^{\alpha}(X, c \cdot \mathbf{y}) = c \cdot \hat{\beta}^{\alpha}(X, \mathbf{y})$$

- the α-th quantile switches for negative values, *i.e.*,

$$\forall c < 0, \quad \hat{\beta}^{\alpha}(X, c \cdot \mathbf{y}) = c \cdot \hat{\beta}^{1-\alpha}(X, \mathbf{y})$$

proof: take $c < 0$

$$\begin{aligned}
\ell_\alpha(cy_i - \langle x_i , \beta \rangle) &= \alpha |cy_i - \langle x_i , \beta \rangle| \mathbb{1}_{\{cy_i \geq \langle x_i , \beta \rangle\}} \\
&\quad + (1-\alpha)|cy_i - \langle x_i , \beta \rangle| \mathbb{1}_{\{cy_i \leq \langle x_i , \beta \rangle\}} \\
&\overset{(1)}{=} c\Big(\alpha \Big| y_i - \Big\langle x_i , \frac{\beta}{c} \Big\rangle \Big| \mathbb{1}_{\{y_i \leq \langle x_i , \frac{\beta}{c} \rangle\}} \\
&\quad + (1-\alpha) \Big| y_i - \Big\langle x_i , \frac{\beta}{c} \Big\rangle \Big| \mathbb{1}_{\{y_i \geq \langle x_i , \frac{\beta}{c} \rangle\}}\Big)
\end{aligned}$$

---

(1) $c < 0$ reverses the inequality

Table of Contents

Motivation

- Quantile regression minimizes a sum that gives asymmetric penalties:
  - weight is $\alpha \, |y_i - \langle \, x_i \, , \, \beta \, \rangle|$ for over-prediction
  - weight is $(1 - \alpha) \, |y_i - \langle \, x_i \, , \, \beta \, \rangle|$ for under-prediction

Motivation

- Quantile regression minimizes a sum that gives asymmetric penalties:
  - weight is $\alpha\,|y_i - \langle\, x_i\,,\,\beta\,\rangle|$ for over-prediction
  - weight is $(1-\alpha)\,|y_i - \langle\, x_i\,,\,\beta\,\rangle|$ for under-prediction
- More robust than OLS to non-normal errors and outliers (in only to y-outliers though!)

Motivation

- Quantile regression minimizes a sum that gives asymmetric penalties:
  - weight is $\alpha\,|y_i - \langle\, x_i\,,\, \beta\,\rangle|$ for over-prediction
  - weight is $(1-\alpha)\,|y_i - \langle\, x_i\,,\, \beta\,\rangle|$ for under-prediction

- More robust than OLS to non-normal errors and outliers (in only to y-outliers though!)

- "Richer" characterization of the data when considering a path (several values) of α. Allows considering impact of a covariate entire conditional distribution of $\mathbf{y}$ given X (OLS/LAD would give only a "central" snapshot)

Equivariance properties continued

- the α-th quantile regression is regression equivariant, *i.e.*,

$$\forall v \in \mathbb{R}^p, \quad \hat{\beta}^\alpha(X, \mathbf{y} + Xv) = \hat{\beta}^\alpha(X, \mathbf{y}) + v$$

- the α-th quantile regression is affine equivariant, *i.e.*, for any non-singular matrix $A \in \mathbb{R}^{p \times p}$

$$\hat{\beta}^\alpha(XA, \mathbf{y}) = A^{-1}\hat{\beta}^\alpha(X, \mathbf{y})$$

Monotonic equivariance

Recall that for a nondecreasing function h over $\mathbb{R}$

$$\begin{aligned} q_{h(Y)}(\alpha) := F_{h(Y)}^{-1}(\alpha) &= \inf\left\{y' \in \mathbb{R} \ : \ F_{h(Y)}(y') \geq \alpha\right\} \\ F_{h(Y)}^{-1}(\alpha) &= \inf\left\{y' \in \mathbb{R} \ : \ \mathbb{P}(h(Y) \geq y') \geq \alpha\right\} \\ F_{h(Y)}^{-1}(\alpha) &= \inf\left\{y' \in \mathbb{R} \ : \ \mathbb{P}(Y \geq h^{-1}(y')) \geq \alpha\right\} \\ F_{h(Y)}^{-1}(\alpha) &= h(F_Y^{-1}(\alpha)) =: h(q_Y(\alpha)) \end{aligned}$$

Conclusion: quantiles are equivariant w.r.t. nondecreasing transformations; so are conditional quantiles

Rem: this might be of interest when no model has clear physical/linear interpretation (could be only after a log/exp/power transform)

Examples where monotonicity helps: censored regression[2]

Context: assume than one do not observe y but only $\max(y, a)$ for a constant a
OLS would fail, need for instance to use an MLE for censored data

BUT: the quantile regression approach would work: consider $h(\cdot) = \max(\cdot, a)$ as a monotonic transform,

$$q_{h(Y)}(\alpha) =: h(q_Y(\alpha))$$

and similarly for conditional quantile

$$q_{h(Y)|X}(\alpha) =: h(q_{Y|X}(\alpha))$$

Hence, one can consider the following consistent approach:

$$\operatorname*{arg\,min}_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \ell_\alpha(y_i - \max(0, \langle\, x_i\,,\, \beta\,\rangle))$$

[2] J. L. Powell. "Censored regression quantiles". In: *J. Econometrics* 32.1 (1986), pp. 143–155.

Examples where monotonicity helps: heteroscedasticity

Take the case where the ε_i's are *i.i.d.*:

$$y_i = \langle \beta , x_i \rangle + \varepsilon_i$$

Then, in terms of quantiles:

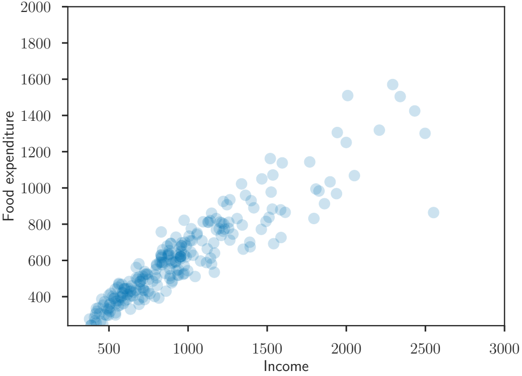
$$q_{y_i|x_i}(\alpha) = \langle \beta , x_i \rangle + q_{\varepsilon_i}(\alpha)$$

Rem: the regression quantile estimate for α and α' are just translated: $q_{y_i|x_i}(\alpha) - q_{y_i|x_i}(\alpha') = q_{\varepsilon_i}(\alpha) - q_{\varepsilon_i}(\alpha')$ (does not depend on x_i)

But now if $\varepsilon_i = \sigma(x_i)\varepsilon_i'$ in the previous model, then

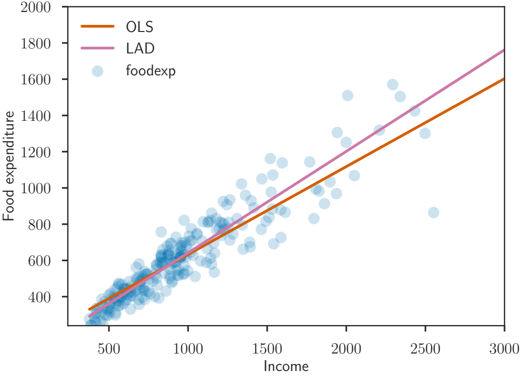
$$q_{y_i|x_i}(\alpha) = \langle \beta , x_i \rangle + \sigma(x_i) q_{\varepsilon_i'}(\alpha)$$

Example[3]



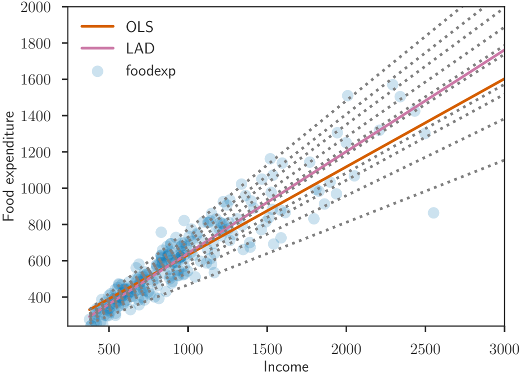
[3] R. Koenker and K. F. Hallock. "Quantile regression". In: *J. Econ. Perspect.* 15.4 (2001), pp. 143–156.

Example[3]



[3] R. Koenker and K. F. Hallock. "Quantile regression". In: *J. Econ. Perspect.* 15.4 (2001), pp. 143–156.

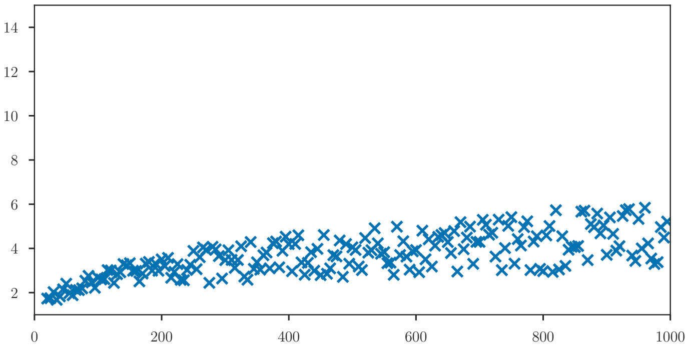
Example[3]



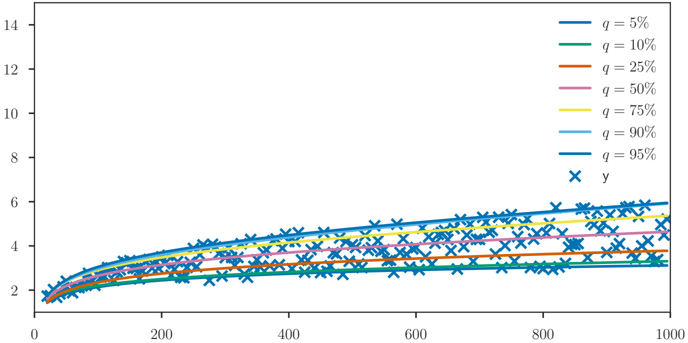
- Food expenditure increases with income
- Food expenditure dispersion increases with income
- the OLS fits over estimate it for low income

(3) R. Koenker and K. F. Hallock. "Quantile regression". In: *J. Econ. Perspect.* 15.4 (2001), pp. 143–156.

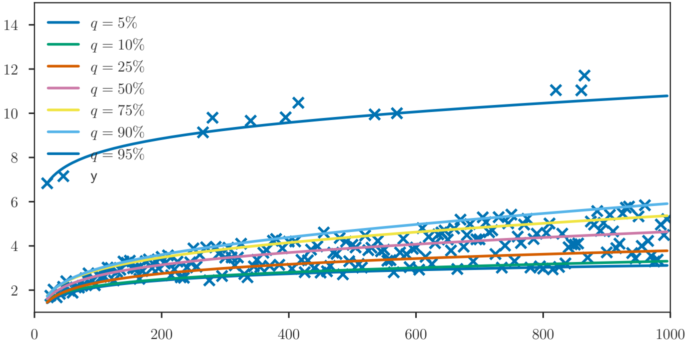
Examples with y-outliers



Examples with y-outliers



Examples with y-outliers



Computation

Several approaches are possibles:

Computation

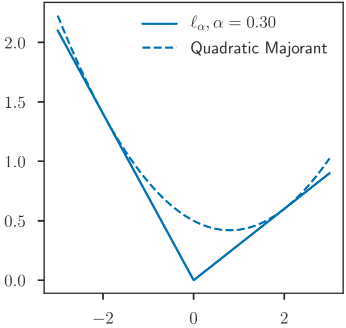
Several approaches are possibles:

- reformulate as a Linear Programming

Computation

Several approaches are possibles:

- reformulate as a Linear Programming
- Majorization - Minimization (`statsmodel` approach)

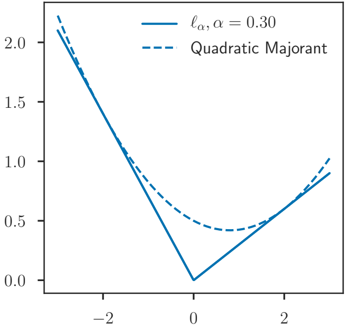


the parabola can be obtained as the best Majorization quadratic function being sharp, with gradient matching and providing the best decrease (see next slide)

Computation

Several approaches are possibles:

- reformulate as a Linear Programming
- Majorization - Minimization (`statsmodel` approach)



the parabola can be obtained as the best Majorization quadratic function being sharp, with gradient matching and providing the best decrease (see next slide)

- More later with regularization $\rightarrow$ quadratic program

Solution of the quadratic majorization

Let's look for a majorizing parabola of f at $x_0 > 0$, under the form

$$F(x) = a(x - x_0)^2 + b(x - x_0) + c$$

with $a > 0$.
We need $F(x_0) = qx_0$, so we must have $c = qx_0$.
We need $F'(x_0) = q$ so we must have $b = q$.
Hence F must be of the form

$$F(x) = a(x - x_0)^2 + qx$$

It is clear that $F(x) \geq qx$ for $x \geq 0$.
Let us find the smallest a s.t. $F(x) \geq (q - 1)x$ for $x \leq 0$.
Let $\phi(x) = F(x) - (q - 1)x = a(x - x_0)^2 + qx$.
$\phi'(x) = 0 \Leftrightarrow x = x_0 - \frac{1}{2a}$
We want $\phi(x_0 - \frac{1}{2a}) = 0$, this gives $a = \frac{1}{4x_0}$

$$F(x) = \frac{1}{4x_0}(x - x_0)^2 + qx$$

Table of Contents

Non-parametric quantile regression[4]

Let $\mathcal{H}$ be a Reproducing Kernel Hilbert Space (RKHS) with kernel K, then the non-parametric quantile regression is

$$\arg\min_{\phi \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} \ell_\alpha(y_i - \phi(x_i)) + \frac{\lambda}{2} \|\phi\|_{\mathcal{H}}^2$$

where $\|\cdot\|_{\mathcal{H}}$ is the norm over the RKHS $\mathcal{H}$

Rem: this is a similar optimization problem to the "Support Vector Machine" SVM one.

Rem: centering $\mathbf{y}$ / handling intercept is needed in practice

[4] I. Takeuchi et al. "Nonparametric quantile estimation". In: *Journal of Machine Learning Research* 7.Jul (2006), pp. 1231–1264.

Reformulation

With $f(x) = \langle \phi(x) , \beta \rangle$, the previous problem is equivalent to

$$\begin{cases} \underset{\beta,\xi,\xi^*}{\arg\min} \left(\frac{\lambda}{2} \|\beta\|^2 + \sum_{i=1}^{n} \alpha\xi_i + (1-\alpha)\xi_i^* \right) & \\ \text{s.c} \quad \xi_i \geq 0, \xi_i^* \geq 0 & \forall i \in [n], \\ \qquad y_i - \langle \beta , \phi(x_i) \rangle \leq \xi_i, & \forall i \in [n]. \\ \qquad \langle \beta , \phi(x_i) \rangle - y_i \leq \xi_i^*, & \forall i \in [n]. \end{cases}$$

Rem: this is a quadratic programming problem

Dual formulation

K is the **kernel matrix** obtained via
$K_{i,j} = k(x_i, x_j) = \langle\, \phi(x_i)\,,\, \phi(x_j)\,\rangle$ for $i, j \in [n] \times [n]$.

$$
\begin{cases}
\underset{\gamma \in \in \mathbb{R}^n}{\arg\min} \left(\dfrac{1}{2} \displaystyle\sum_{1 \leq i,\, j \leq n} \gamma_i \gamma_j y_i y_j K_{i,j} \right) - \displaystyle\sum_{i=1}^{n} \gamma_i y_i & \\
\text{s.c} \quad \frac{\lambda}{n}(\alpha - 1) \leq \gamma_i \leq \frac{\lambda}{n}\alpha, & \forall i \in \{1, \cdots, n\}, \\
\qquad \sum_{i=1}^{n} \gamma_i y_i = 0 &
\end{cases}
$$

Rem: one recovers $f(x) = \sum_{i=1}^{n} \gamma_i k(x_i, x)$ and $\beta = \sum_{i=1}^{n} \gamma_i \phi(x_i)$

Computation

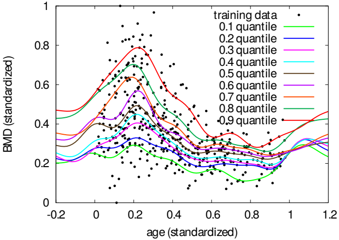
- Fast solvers like LibSVM[5] can be adapted
- Speed-up for standard Kernel method can also be considered:
  - Random Kernel Features[6]
  - Nyström methods[7]

[5] R.-E. Fan et al. "LIBLINEAR: A library for large linear classification". In: *J. Mach. Learn. Res.* 9 (2008), pp. 1871–1874.

[6] P. Drineas and M. W. Mahoney. "On the Nyström method for approximating a Gram matrix for improved kernel-based learning". In: *J. Mach. Learn. Res.* 6 (2005), pp. 2153–2175.

[7] A. Rahimi and B. Recht. "Random features for large-scale kernel machines". In: *NIPS*. ed. by J.C. Platt et al. Curran Associates, Inc., 2008, pp. 1177–1184.

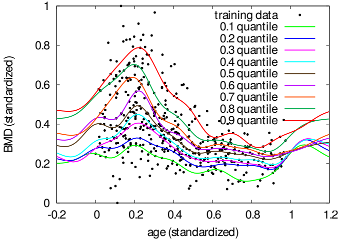
Examples (extracted from[8])



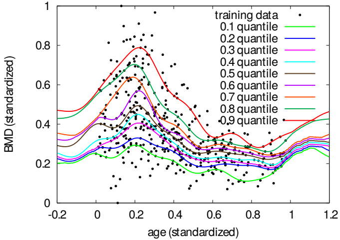
(a) Without *non-crossing* constraints

[8] I. Takeuchi et al. "Nonparametric quantile estimation". In: *Journal of Machine Learning Research* 7.Jul (2006), pp. 1231–1264.

Examples (extracted from[8])



(a) Without *non-crossing* constraints



(b) With *non-crossing* constraints

[8] I. Takeuchi et al. “Nonparametric quantile estimation”. In: *Journal of Machine Learning Research* 7.Jul (2006), pp. 1231–1264.

Crossing issue

Crossing: quantile level "lines" can cross in non-parametric quantile regression when the optimization problem over several α's are solved separately.

Natural solution: couple the optimization problems for the α's of interest (say $T = 10$ or $T = 20$ values of α's)

Crossing resolution

Enforce "ordered" constraints for instance for all the points in the sample, and all the T quantiles $\alpha_1 \leq \cdots \leq \alpha_T$ targeted, *i.e.*,

$$f_t(x) = \langle \phi(x_i) \, , \, \beta_t \rangle \leq f_{t+1}(x) = \langle \phi(x_i) \, , \, \beta_{t+1} \rangle \, , \forall i, t \in [n] \times [T-1]$$

Crossing resolution

Enforce "ordered" constraints for instance for all the points in the sample, and all the T quantiles $\alpha_1 \leq \cdots \leq \alpha_T$ targeted, *i.e.*,

$$f_t(x) = \langle\, \phi(x_i)\,,\, \beta_t \,\rangle \leq f_{t+1}(x) = \langle\, \phi(x_i)\,,\, \beta_{t+1} \,\rangle\,, \forall i, t \in [n] \times [T-1]$$

$$\begin{cases} \underset{\beta_t, \xi_t, \xi_t^*}{\arg\min} \displaystyle\sum_{t=1}^{T} \left(\frac{\lambda}{2} \|\beta_t\|^2 + \sum_{i=1}^{n} \alpha_t \xi_{i,t} + (1-\alpha_t)\xi_{i,t}^* \right) & \\ \text{s.c} \quad \xi_{i,t} \geq 0, \xi_{i,t}^* \geq 0 & \forall i, t \in [n] \times [T-1] \\ \qquad y_i - \langle\, \beta\,,\, \phi(x_i) \,\rangle \leq \xi_{i,t}, & \forall i, t \in [n] \times [T-1] \\ \qquad \langle\, \beta\,,\, \phi(x_i) \,\rangle - y_i \leq \xi_{i,t}^*, & \forall i, t \in [n] \times [T-1] \end{cases}$$

Crossing resolution

Enforce "ordered" constraints for instance for all the points in the sample, and all the T quantiles $\alpha_1 \leq \cdots \leq \alpha_T$ targeted, *i.e.*,

$$f_t(x) = \langle \phi(x_i) \, , \, \beta_t \rangle \leq f_{t+1}(x) = \langle \phi(x_i) \, , \, \beta_{t+1} \rangle \, , \forall i, t \in [n] \times [T-1]$$

$$\begin{cases} \underset{\beta_t, \xi_t, \xi_t^*}{\arg\min} \sum_{t=1}^{T} \left(\frac{\lambda}{2} \|\beta_t\|^2 + \sum_{i=1}^{n} \alpha_t \xi_{i,t} + (1-\alpha_t) \xi_{i,t}^* \right) & \\ \text{s.c} \quad \xi_{i,t} \geq 0, \xi_{i,t}^* \geq 0 & \forall i, t \in [n] \times [T-1] \\ \quad y_i - \langle \beta \, , \, \phi(x_i) \rangle \leq \xi_{i,t}, & \forall i, t \in [n] \times [T-1] \\ \quad \langle \beta \, , \, \phi(x_i) \rangle - y_i \leq \xi_{i,t}^*, & \forall i, t \in [n] \times [T-1] \\ \quad \color{red}{\langle \phi(x_i) \, , \, \beta_t \rangle \leq \langle \phi(x_i) \, , \, \beta_{t+1} \rangle \, ,} & \color{red}{\forall i, t \in [n] \times [T-1]} \end{cases}$$

Crossing resolution

Enforce "ordered" constraints for instance for all the points in the sample, and all the T quantiles $\alpha_1 \leq \cdots \leq \alpha_T$ targeted, *i.e.*,

$$f_t(x) = \langle \phi(x_i) \, , \, \beta_t \rangle \leq f_{t+1}(x) = \langle \phi(x_i) \, , \, \beta_{t+1} \rangle \, , \forall i, t \in [n] \times [T-1]$$

$$\begin{cases} \underset{\beta_t, \xi_t, \xi_t^*}{\arg\min} \displaystyle\sum_{t=1}^{T} \left(\frac{\lambda}{2} \|\beta_t\|^2 + \sum_{i=1}^{n} \alpha_t \xi_{i,t} + (1-\alpha_t) \xi_{i,t}^* \right) & \\ \text{s.c} \quad \xi_{i,t} \geq 0, \xi_{i,t}^* \geq 0 & \forall i, t \in [n] \times [T-1] \\ \qquad y_i - \langle \beta \, , \, \phi(x_i) \rangle \leq \xi_{i,t}, & \forall i, t \in [n] \times [T-1] \\ \qquad \langle \beta \, , \, \phi(x_i) \rangle - y_i \leq \xi_{i,t}^*, & \forall i, t \in [n] \times [T-1] \\ \qquad {\color{red} \langle \phi(x_i) \, , \, \beta_t \rangle \leq \langle \phi(x_i) \, , \, \beta_{t+1} \rangle \, ,} & {\color{red} \forall i, t \in [n] \times [T-1]} \end{cases}$$

Rem: larger problem, but can be solve similarly

Rem: variants of constraints can be envisioned too

https://operalib.github.io/operalib/documentation/auto_examples/

Table of Contents

Density quantile function[10]

Definition

For a pdf f associated to a quantile function f, the density quantile function (p-pdf) is define as $f_p = f \circ q$

- in particular using derivative of inversion: $f_p = \frac{1}{q}$

Theorem[9]

A positive linear combination of a finite number of quantile functions is a quantile function.

- for q_1 and q_2 two quantile distribution, the sum $q = q_1 + q_2$ associated to p-pdf $f_{p,1}$ and $f_{p,2}$ then $f_p = \left(\frac{1}{f_{p,1}} + \frac{1}{f_{p,2}}\right)^{-1}$

(9) B. W. Powley. "Quantile function methods for decision analysis". PhD thesis. Stanford University, 2013.

(10) W. Gilchrist. *Statistical modelling with quantile functions*. CRC Press, 2000.

Logistic detour[11]

Cumulative density function for the logistic distribution:

$$F(t) = \frac{\exp(t)}{\exp(t) + 1}$$

Probability density function for the logistic distribution:

$$f(t) = \frac{\exp(t)}{(\exp(t) + 1)^2}$$

Quantile function (it is the **logit** function!):

$$q(\alpha) = \log\left(\frac{\alpha}{1 - \alpha}\right)$$

[11] W. Gilchrist. "Regression Revisited". In: *International Statistical Review / Revue Internationale de Statistique* 76.3 (2008), pp. 401–418.

Asymmetric logit

Reminder:

- for the exponential distribution $f(x) = \mathbb{1}_{\mathbb{R}^+} \exp(-x)$, then $q(\alpha) = -\log(1-\alpha)$
- for the reverse exponential distribution $f(x) = \mathbb{1}_{\mathbb{R}^-} \exp(x)$, then $q(\alpha) = \log(\alpha)$

the quantile function of the logistic is the sum a right-tail and left tail exponential.
Possible alternative: put asymmetric weight ω and get as quantile function:

$$\begin{aligned} q(\alpha) &= (1-\omega)\log(\alpha) - (1+\omega)\log(1-\alpha) \\ &= \log\left(\frac{\alpha}{1-\alpha}\right) - \omega\log[\alpha(1-\alpha)] \end{aligned}$$

Table of Contents

Limits of quantiles (for risk measure)

Let us assume that : $X \sim F$ (cdf)

Definition

the α-quantile of X is given by

$$q_X(\alpha) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}$$

- $q_X(\cdot)$ non-sub-additive and non convex
- $q_X(\cdot)$ is discontinuous for discrete distributions
- for quantile regression, the solution only depends on the sign of the residuals (Fermat's rule), *e.g.*, for median, one targets as many errors on each side of the estimator: it is overly "robust", a large proportion of the dataset can change without changing at all the solution (see example with 90% quantile before)

Risk

Risk (Wikipedia): **Risk** is the potential of losing something of value . . . Uncertainty is a potential, unpredictable, and uncontrollable outcome; risk is a consequence of action taken in spite of uncertainty.

Risk aversion (Wikipedia): . . . risk aversion is the behavior of humans, when exposed to uncertainty, to attempt to reduce that uncertainty. . . .

Where risk aversion matters?

- Financial portfolios
- Health-care decisions
- Agriculture
- Public infrastructure
- Robotics (Mars rover :)
- Self-driving cars

Historical examples: Markowitz[12] portfolio

X_i are risky asset.

$$\min_{c \geq 0} \quad \mathrm{Var}\left(\sum_i c_i X_i\right)$$
$$\text{s.t. } \mathbb{E}\left(\sum_i c_i X_i\right) = \mu, \sum_i c_i = 1$$

Limited modeling capability and also penalizes upside

[12] H. Markowitz. "Portfolio selection". In: *The Journal of Finance* 7.1 (1952), pp. 77–91.

Risk measures

Definition

A **risk measure** is a function ρ mapping a random variable X to a real number $\rho(X)$

In practice: one aim at minimizing the risk ($\hookrightarrow$ hence convexity might be nice), while having a large output X

- expectation $\mathbb{E}(X)$ (risk neutral), minimum $\min(X)$ (very risk averse)
- Value at Risk in finance $(\mathrm{V@R})$ / quantile
- Conditional Value at Risk in finance $(\mathrm{CV@R})$ / superquantile
- Coherent measure of risk

Rem: interpretation connects to financial scenarios and the risk needs to be minimize

Translation equivariance

Definition

We say that a risk measure ρ is risk **translation equivariance** if for any real random variable X and any constant $c \in \mathbb{R}$ one has:

$$\rho(X+c) = \rho(X) - c$$

Interpretation: adding a sure amount of capital reduces the risk by the same amount

Subadditivity

Definition

We say that a risk measure ρ is **subadditive** if for any two real random variables X and Y one has

$$\rho(X+Y) \leq \rho(X) + \rho(Y)$$

Interpretation: adding risky position can only be riskier

Positive homogeneity

Definition

We say that a risk measure ρ is **positively homogeneous** if for any real random variable X and $c > 0$ one has

$$\rho(c \cdot X) = c \cdot \rho(X)$$

Interpretation: in financial risk management, the risk of a position is proportional to its size.

Rem: positive homogeneity + subadditivity $\implies$ convexity, and the interpretation is that combining risk together decreases risks

Monotonicity

Definition

We say that a risk measure ρ is **monotonic** if for any real random variables X and Y,

$$X \geq Y \implies \rho(X) \leq \rho(Y)$$

Interpretation: a systematically better portofolio must be less risky

Coherence of a risk measure[13]

In what follows X and Y are real random variables

Definition

A risk measure, *i.e.*, a function $\rho : \mathcal{X} \to \mathbb{R}$ satisfying:

1. translation equivariance
2. subadditivity
3. positive homogeneity
4. monotonicity

is called **coherent**

Rem: hence it is convex

[13] P. Artzner et al. "Coherent Measures of Risk". In: *Mathematical Finance* 9.3 (1999), pp. 203–228.

Super-quantile

Definition

the α-superquantile of X is given by

$$\bar{q}_X(\alpha) := \mathbb{E}[X | X \geq q_X(\alpha)]$$

Rem: superquantile is a coherent risk measure[14] not true for quantile

[14] A. Ben-Tal and M. Teboulle. "An old-new concept of convex risk measures: the optimized certainty equivalent". In: *Math. Finance* 17.3 (2007), pp. 449–476.

Superquantile alternative formulation

Proposition:

$$\bar{q}_X(\alpha) = \frac{1}{1-\alpha}\mathbb{E}[X\mathbb{1}_{\{X \geq q_X(\alpha)\}}] = \frac{1}{1-\alpha}\int_\alpha^1 q_X(\gamma)d\gamma$$

Motivation[15],[16]: as Conditional Value-at-Risk ($\mathrm{CV@R}$), average the risk over the tail of the distribution

Proof:

$$\begin{aligned}\mathbb{E}[X\mathbb{1}_{\{X \geq q_X(\alpha)\}}] &= \mathbb{E}[X|X \geq q_X(\alpha)] \cdot \mathbb{P}[X \geq q_X(\alpha)] \\ &= \mathbb{E}[X|X \geq q_X(\alpha)] \cdot (1-\alpha) \\ &= \bar{q}_X(\alpha) \cdot (1-\alpha)\end{aligned}$$

□

[15] R. T. Rockafellar and S. Uryasev. "Optimization of conditional value-at-risk". In: *Journal of Risk* 2 (2000), pp. 21–42.

[16] R. T. Rockafellar and S. Uryasev. "The fundamental risk quadrangle in risk management, optimization and statistical estimation". In: *Surveys in Operations Research and Management Science* 18.1-2 (2013), pp. 33–53.

Properties

Theorem:

Let $\nu_\alpha(X) = \frac{1}{1-\alpha}\mathbb{E}[(X)_+]$, where $(X)_+ = \max(X, 0)$, then

$$q_X(\alpha) \in \arg\min_{q\in\mathbb{R}} (q + \nu_\alpha(X - q))$$

$$\bar{q}_X(\alpha) = \min_{q\in\mathbb{R}} (q + \nu_\alpha(X - q))$$

proof:

$$h_\alpha(q) = q + \frac{1}{1-\alpha}\mathbb{E}[(X - q)_+]$$

$$\begin{aligned} h_\alpha(q_X(\alpha)) &= q_X(\alpha) + \tfrac{1}{1-\alpha}\mathbb{E}[(X - q_X(\alpha))_+] \\ &= q_X(\alpha) + \tfrac{1}{1-\alpha}\mathbb{E}[(X - q_X(\alpha))|X \geq q_X(\alpha)]\mathbb{P}(X \geq q_X(\alpha)) \\ &= q_X(\alpha) + \mathbb{E}[(X - q_X(\alpha))|X \geq q_X(\alpha)] \\ &= \mathbb{E}[X|X \geq q_X(\alpha)] \end{aligned}$$

Usage for regression

See Rockafellar *et al.* (2014) for details

Other recent extensions

- Robust extension of logistic regression: Shafieezadeh *et al.* (2015)
- Book on quantile functions Gilchrist (2000)
- More general approach and primal dual algorithms for non-parametric quantile regression Sangnier *et al.* (2016)

References I

- Artzner, P. et al. "Coherent Measures of Risk". In: *Mathematical Finance* 9.3 (1999), pp. 203–228.
- Ben-Tal, A. and M. Teboulle. "An old-new concept of convex risk measures: the optimized certainty equivalent". In: *Math. Finance* 17.3 (2007), pp. 449–476.
- Drineas, P. and M. W. Mahoney. "On the Nyström method for approximating a Gram matrix for improved kernel-based learning". In: *J. Mach. Learn. Res.* 6 (2005), pp. 2153–2175.
- Fan, R.-E. et al. "LIBLINEAR: A library for large linear classification". In: *J. Mach. Learn. Res.* 9 (2008), pp. 1871–1874.
- Gilchrist, W. "Regression Revisited". In: *International Statistical Review / Revue Internationale de Statistique* 76.3 (2008), pp. 401–418.
- – .*Statistical modelling with quantile functions*. CRC Press, 2000.
- Koenker, R. and G. Bassett. "Regression quantiles". In: *Econometrica* 46.1 (1978), pp. 33–50.

References II

- Koenker, R. and K. F. Hallock. "Quantile regression". In: *J. Econ. Perspect.* 15.4 (2001), pp. 143–156.
- Markowitz, H. "Portfolio selection". In: *The Journal of Finance* 7.1 (1952), pp. 77–91.
- Powell, J. L. "Censored regression quantiles". In: *J. Econometrics* 32.1 (1986), pp. 143–155.
- Powley, B. W. "Quantile function methods for decision analysis". PhD thesis. Stanford University, 2013.
- Rahimi, A. and B. Recht. "Random features for large-scale kernel machines". In: *NIPS*. Ed. by J.C. Platt et al. Curran Associates, Inc., 2008, pp. 1177–1184.
- Rockafellar, R. T. and S. Uryasev. "Optimization of conditional value-at-risk". In: *Journal of Risk* 2 (2000), pp. 21–42.

References III

- Rockafellar, R. T. and S. Uryasev. "The fundamental risk quadrangle in risk management, optimization and statistical estimation". In: *Surveys in Operations Research and Management Science* 18.1-2 (2013), pp. 33–53.
- Rockafellar, R.T., J.O. Royset, and S.I. Miranda. "Superquantile regression with applications to buffered reliability, uncertainty quantification, and conditional value-at-risk". In: *European J. Oper. Res.* 234.1 (2014), pp. 140–154.
- Sangnier, M., O. Fercoq, and F. d'Alché-Buc. "Joint quantile regression in vector-valued RKHSs". In: *NIPS*. 2016, pp. 3693–3701.
- Shafieezadeh-Abadeh, S., P. M. Esfahani, and D. Kuhn. "Distributionally Robust Logistic Regression". In: *NIPS*. Ed. by C. Cortes et al. Curran Associates, Inc., 2015, pp. 1576–1584.
- Takeuchi, I. et al. "Nonparametric quantile estimation". In: *Journal of Machine Learning Research* 7.Jul (2006), pp. 1231–1264.