An Introduction to Quantile Regression

Matthew Durey

University of Bath

June 1, 2015

Definitions

- For $\tau \in (0,1)$, the τ^{th} quantile is a value $x(\tau)$ such that

$$\int_{-\infty}^{x} f(u)\,\mathrm{d}u = \tau,$$

  where $f(\cdot)$ is a probability density function.

- Equivalently,

$$\begin{aligned} & P(X \leq x) = \tau && \text{for R.V. } X \\ \text{or} \quad & F(x) = \tau && \text{for c.d.f. } F(\cdot) \end{aligned}$$

Examples

- $\tau = 1/2$ gives the median $x(1/2)$.
- $\tau = 1/4$ and $\tau = 3/4$ give the lower and upper quartiles, respectively.
- The inter-quartile range is $x(3/4) - x(1/4)$.

Motivational example

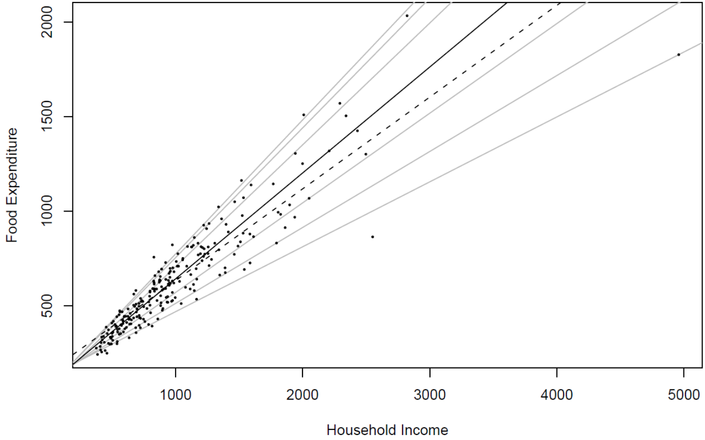

Figure: Engel's 1857 data with median (black), quartiles for $\tau = 0.05, 0.1, 0.25, 0.75, 0.9, 0.95$ (grey) and linear least squares (dashed).

Finding the median through minimisation

- Univariate observations y_i for $i = 1, \ldots, n$.
- Sample mean
$$\hat{\mu} = \text{argmin}_{\mu \in \mathbb{R}} \sum_{i=1}^{n} (y_i - \mu)^2.$$
- The sample median $\hat{\xi}$ is the 'middle value' of the observations.
$$\hat{\xi} = \text{argmin}_{\xi \in \mathbb{R}} \sum_{i=1}^{n} |y_i - \xi|.$$

Finding other quantiles

- Adjust median minimisation problem.
- Consider $\tau = 0.25$, so $\hat{\xi}$ is the lower quartile.
- To 'reduce' $\hat{\xi}$ from median to lower quartile, weight the data.
- For $y_i < \hat{\xi}$, weight by more.
- For $y_i > \hat{\xi}$, weight by less.
- So have problem of the form

$$\hat{\xi} = \text{argmin}_{\xi \in \mathbb{R}} \sum_{i=1}^{n} w_i |y_i - \xi|$$

for given weights $w_i > 0$.

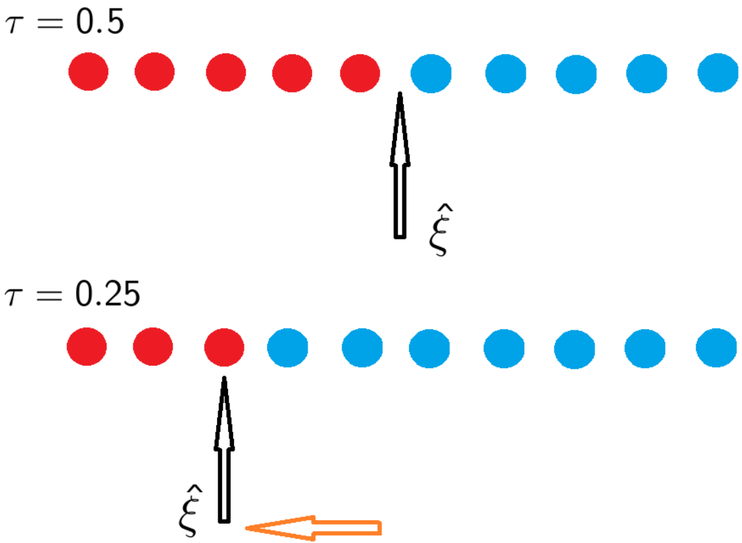

Figure: The red and blue dots are heavier and lighter dots, respectively.

Formulation

- Find that

$$\hat{\xi}(\tau) = \text{argmin}_{\xi \in \mathbb{R}} \left((1-\tau) \sum_{i: y_i \leq \xi} |y_i - \xi| + \tau \sum_{i: y_i > \xi} |y_i - \xi| \right).$$

- Write compactly as

$$\hat{\xi}(\tau) = \text{argmin}_{\xi \in \mathbb{R}} \sum_{i=1}^{n} \rho_\tau(y_i - \xi),$$

$$\text{where} \quad \rho_\tau(u) \equiv u\left(\tau - \mathbb{1}_{\{u<0\}}\right).$$

- For $\tau = 1/2$, have $\rho_{1/2}(u) = |u|/2$.
- Minimisation problem is unaffected by proportionality factor.

Weighted function

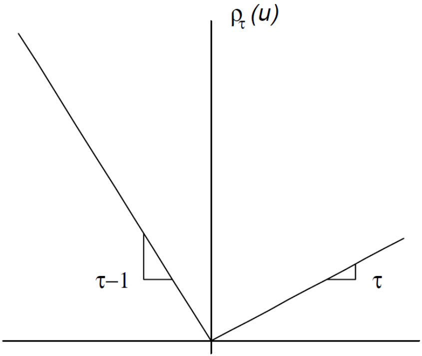

Figure: Assymetric weighting function $\rho_\tau(\cdot)$.

Sketch proof (1)

- Univariate case
- For given τ, define the objective function

$$R_\tau(\xi) = \sum_{i=1}^{n} \rho_\tau(y_i - \xi).$$

- $R_\tau(\cdot)$ is a sum of convex functions and so is convex.
- At minimiser, want non-negative directional derivatives. Define

$$R'(\xi^+) \equiv \lim_{h \to 0^+} \left(\frac{R(\xi + h) - R(\xi)}{h} \right)$$
$$R'(\xi^-) \equiv \lim_{h \to 0^+} \left(\frac{R(\xi - h) - R(\xi)}{h} \right)$$

Sketch proof (2)

- After some tedious algebra...

$$R'(\xi^+) = \sum_{i=1}^{n} \left(\mathbb{1}_{\{y_i < \xi^+\}} - \tau\right) = N^+(\xi) - \tau n,$$
$$R'(\xi^-) = \sum_{i=1}^{n} \left(\tau - \mathbb{1}_{\{y_i < \xi^-\}}\right) = \tau n - N^-(\xi),$$

where

$$N^+(\xi) = |\{y_i : y_i \leq \xi\}|,$$
$$N^-(\xi) = |\{y_i : y_i < \xi\}|.$$

Sketch proof (3)

- At a minimiser, have

$$\frac{N^-(\hat{\xi})}{n} \leq \tau \leq \frac{N^+(\hat{\xi})}{n}.$$

- For $\tau n \notin \mathbb{Z}$, $N^-(\hat{\xi})$ and $N^+(\hat{\xi})$ are unique.
- For $\tau n \in \mathbb{Z}$, $\hat{\xi}$ lies between two data points.
- Example: Find median of the data $\{1, 2\}$.

Multivariate extension

- Response observations y_i and p-dimensional covariates $\mathbf{x}_i$.
- Express the conditional mean as a linear combination of the covariates $\mu(\mathbf{x}) = \mathbf{x}^T \boldsymbol{\beta}$, where $\boldsymbol{\beta} \in \mathbb{R}^p$.
- For conditional sample mean, find

$$\hat{\boldsymbol{\beta}} = \text{argmin}_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i=1}^{n} (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2.$$

- Specifying the τ^{th} conditional quantile function $Q_y(\tau|\mathbf{x}) = \mathbf{x}^T \boldsymbol{\beta}(\tau)$, find that

$$\hat{\boldsymbol{\beta}}(\tau) = \text{argmin}_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x}_i^T \boldsymbol{\beta}).$$

- Can also parameterise quantile function by splines, for example.

Example 2

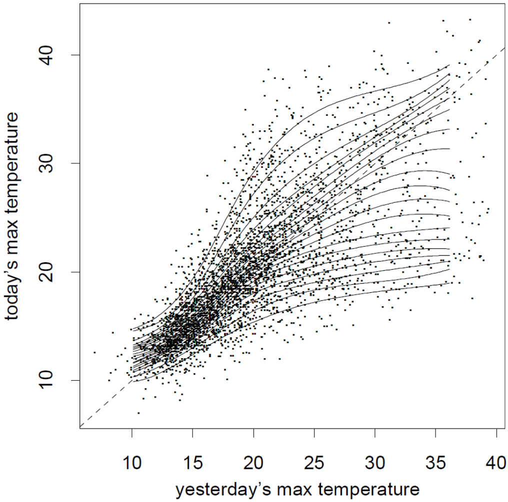

Figure: Melbourne temperature data. Quantiles using spline basis function (solid curves) and linear least squares fit (dashed line).

Example 2 (cont.)

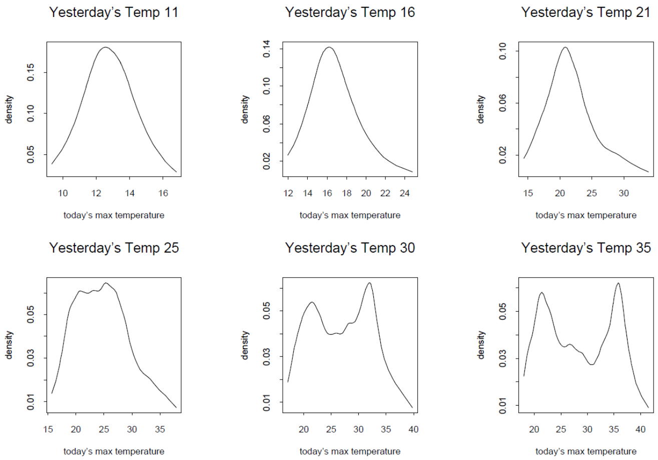

Figure: Associated probability density functions given Yesterday's Temperature.

Equivariance property

- For $h(\cdot)$ a non-decreasing function on $\mathbb{R}$,

$$Q_{h(Y)}(\tau) = h(Q_Y(\tau)).$$

- This property follows from

$$P(Y \leq y) = P(h(Y) \leq h(y)).$$

- Compare to $\mathbb{E}[h(Y)] \neq h(\mathbb{E}[Y])$ for general $h(\cdot)$.

Linear programming

- General problem: For given $\mathbf{c} \in \mathbb{R}^k$, $A \in \mathbb{R}^{m \times k}$ and $\mathbf{b} \in \mathbb{R}^m$:

$$\begin{array}{lll} \text{minimise} & \mathbf{c}^T \mathbf{z} & \text{over } \mathbf{z} \in \mathbb{R}^k \\ \text{subject to} & A\mathbf{z} \geq \mathbf{b}. & \end{array}$$

- For canonical form, add in slack variables $\mathbf{s} \in \mathbb{R}_+^m$:

$$\begin{array}{lll} \text{minimise} & \mathbf{c}^T \mathbf{z} & \text{over } \mathbf{z} \in \mathbb{R}^k \text{ and } \mathbf{s} \in \mathbb{R}_+^m \\ \text{subject to} & A\mathbf{z} - \mathbf{s} = \mathbf{b}. & \end{array}$$

- Can solve using Simplex Method (for example).

Quantile Regression with Linear programming

- Recall that we wish to find

$$\hat{\boldsymbol{\beta}}(\tau) = \text{argmin}_{\boldsymbol{\beta} \in \mathbb{R}^p} \left(\sum_{i : y_i \geq \mathbf{x}_i^T \boldsymbol{\beta}} \tau(y_i - \mathbf{x}_i^T \boldsymbol{\beta}) - \sum_{i : y_i \leq \mathbf{x}_i^T \boldsymbol{\beta}} (1 - \tau)(y_i - \mathbf{x}_i^T \boldsymbol{\beta}) \right),$$

subject to

$$y_i \geq \mathbf{x}_i^T \boldsymbol{\beta} \quad \forall i \in S \subset \{1, \dots, n\},$$
$$y_i \leq \mathbf{x}_i^T \boldsymbol{\beta} \quad \forall i \in \{1, \dots, n\} \setminus S.$$

Quantile Regression with Linear programming (2)

- Define the model matrix $X \in \mathbb{R}^{n \times p}$ to be $X = [\mathbf{x}_1 | \dots | \mathbf{x}_n]^T$.
- Introduce $\mathbf{e} \in \mathbb{R}^n$ to be $\mathbf{e} = (1, \dots, 1)$.
- Introduce slack variables $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n_+$.
- Have linear programming problem

$$\begin{aligned} &\text{minimise} && [\tau \mathbf{e}^T \mathbf{u} + (1 - \tau)\mathbf{e}^T \mathbf{v}] \qquad \text{over } \boldsymbol{\beta} \in \mathbb{R}^n \text{ and } \mathbf{u}, \mathbf{v} \in \mathbb{R}^n_+ \\ &\text{subject to} && X\boldsymbol{\beta} + \mathbf{u} - \mathbf{v} = \mathbf{y}, \\ & && u_i v_i = 0 \;\; \forall i = 1, \dots, n. \end{aligned}$$

- The final condition means that slack can only be added for one of $y_i \geq \mathbf{x}_i^T \boldsymbol{\beta}$ and $y_i \leq \mathbf{x}_i^T \boldsymbol{\beta}$ for each i.