Quantile Regression

Ruibin Xi

Interior Point Method: an example (1)

- Given a polygon inscribed in a circle, find the point on the polygon that maximizes the sum of its coordinates

$$\max\{e^\top u | X^\top d = u, \;\; e^\top d = 1, \;\; d \in \mathbb{R}^n_+\}$$

where e is the vector of ones, and X has rows representing the n vertices

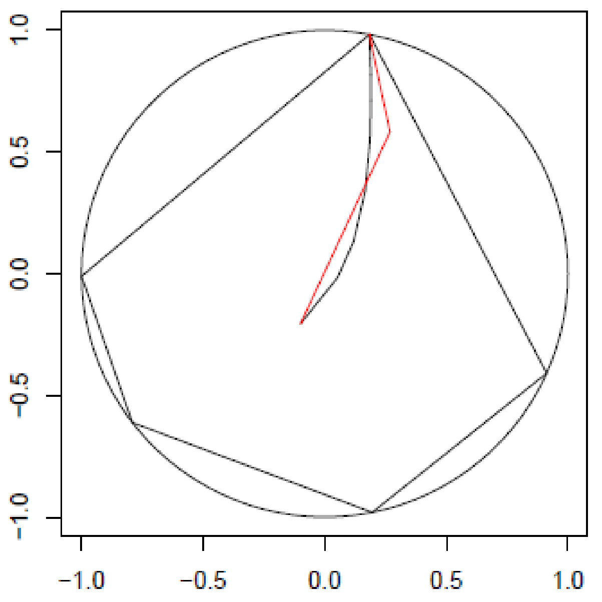
- Eliminating u, setting $s = Xe$, we can formulate the problem as

$$\max\{s^\top d | e^\top d = 1, \;\; d \in \mathbb{R}^n_+\}$$

Interior Point Method: an example (2)

- Simplex method goes around the outside of the polygon; interior point method search from the inside, solving a sequence of problems of the form

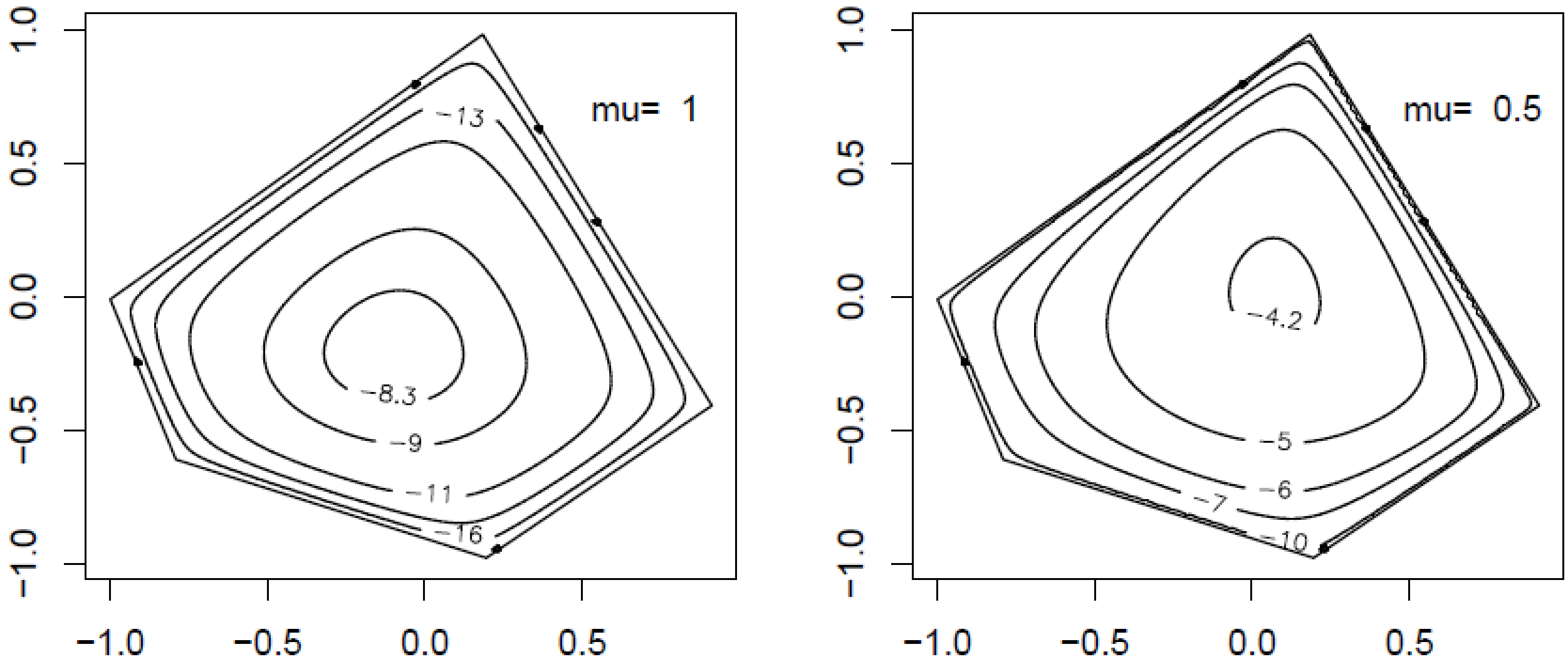
$$\max \left\{ s^\top d + \mu \sum_{i=1}^{n} \log d_i | e^\top d = 1 \right\}$$



Interior Point Method: an example (3)

By letting $\mu \to 0$ we get a sequence of smooth problems whose solutions approach the solution of the LP:

$$\max\left\{ s^\top d + \mu \sum_{i=1}^{n} \log d_i \,|\, e^\top d = 1 \right\}$$



Interior Point Method: an example (4)

- For the problem

$$\max \left\{ s^\top d + \mu \sum_{i=1}^{n} \log d_i \,|\, e^\top d = 1 \right\}$$

we can get the Newton direction

$$p = \mu^{-1} D^2 s + De - \hat{a} \mu^{-1} D^2 e$$

where $\hat{a} = (e^\top D^2 e)^{-1} (e^\top D^2 s + \mu e^\top De)$

Pursuing the iteration $d \leftarrow d + \lambda p$ yields the central path $d(\mu)$

Interior Point Method: an example (5)

- The dual of $\max\{s^\top d | e^\top d = 1,\ \ d \in \mathbb{R}^n_+\}$ is

$$\min\{a | ea - z = s, \quad z \geq 0\}$$

- This is simply equivalent to looking for the maximal elements in s.
- The primal-dual formulation is

$$e^\top d = 1$$

$$ea - z = s$$

$$Dz = \mu e$$

Interior Point Method: an example (6)

- For any feasible pair (z, d) , we have

$$s^\top d = a - z^\top d$$

- So $z^\top d$ is equal to the duality gap;
- At a solution, we have the complementary condition $z^\top d = 0$, thus implying a duality gap of zero
- We may take $\mu = z^\top d / n$ as a direct measure of progress toward a solution

Interior Point Method: an example (7)

- Newton's Method gives

$$\begin{pmatrix} Z & 0 & D \\ e^\top & 0 & 0 \\ 0 & e & -I \end{pmatrix} \begin{pmatrix} p_d \\ p_a \\ p_z \end{pmatrix} = \begin{pmatrix} \mu e - Dz \\ 0 \\ 0 \end{pmatrix}$$

- Solve for this equation we have

$$\hat{p}_a = (e^\top Z^{-1} De)^{-1} e^\top Z^{-1} (Dz - \mu e)$$

$$\hat{p}_d = Z^{-1}(\mu e - Dz - De\,\hat{p}_a)$$

$$\hat{p}_z = e\,\hat{p}_a$$

- Affine-scaling Newton direction corresponds to $\mu = 0$

Interior Point Method: an example (8)

- Newton's Method gives

$$\begin{pmatrix} Z & 0 & D \\ e^\top & 0 & 0 \\ 0 & e & -I \end{pmatrix} \begin{pmatrix} p_d \\ p_a \\ p_z \end{pmatrix} = \begin{pmatrix} \mu e - Dz \\ 0 \\ 0 \end{pmatrix}$$

- Solve for this equation we have

$$\hat{p}_a = (e^\top Z^{-1} D e)^{-1} e^\top Z^{-1} (Dz - \mu e)$$

$$\hat{p}_d = Z^{-1}(\mu e - Dz - De\,\hat{p}_a)$$

$$\hat{p}_z = e\,\hat{p}_a$$

Interior Point Method: an example (9)

- We may update d with $d + \lambda_d p_d$ and z with $z + \lambda_z p_z$ where

$$\lambda_d = \operatorname{argmax}\{\lambda \in [0, 1] | d + \lambda p_d \geq 0\}$$

$$\lambda_z = \operatorname{argmax}\{\lambda \in [0, 1] | z + \lambda p_z \geq 0\}$$

- If updating these two values with a full affine-scaling step, we have the new duality gap is

$$\hat{\mu} = (d + \lambda_d p_d)^\top (z + \lambda_z p_z)/n$$

- The original duality gap is

$$\mu = d^\top z/n$$

Interior Point Method: an example (10)

- If $\hat{\mu}$ is considerably smaller than μ, this means that the affine-scaling direction brought us considerably closer to the optimal solution
- Otherwise, the affine-scaling is not effective or not favorable
- Mehrotra proposed to update μ by

$$\mu \leftarrow \mu(\hat{\mu}/\mu)^3$$

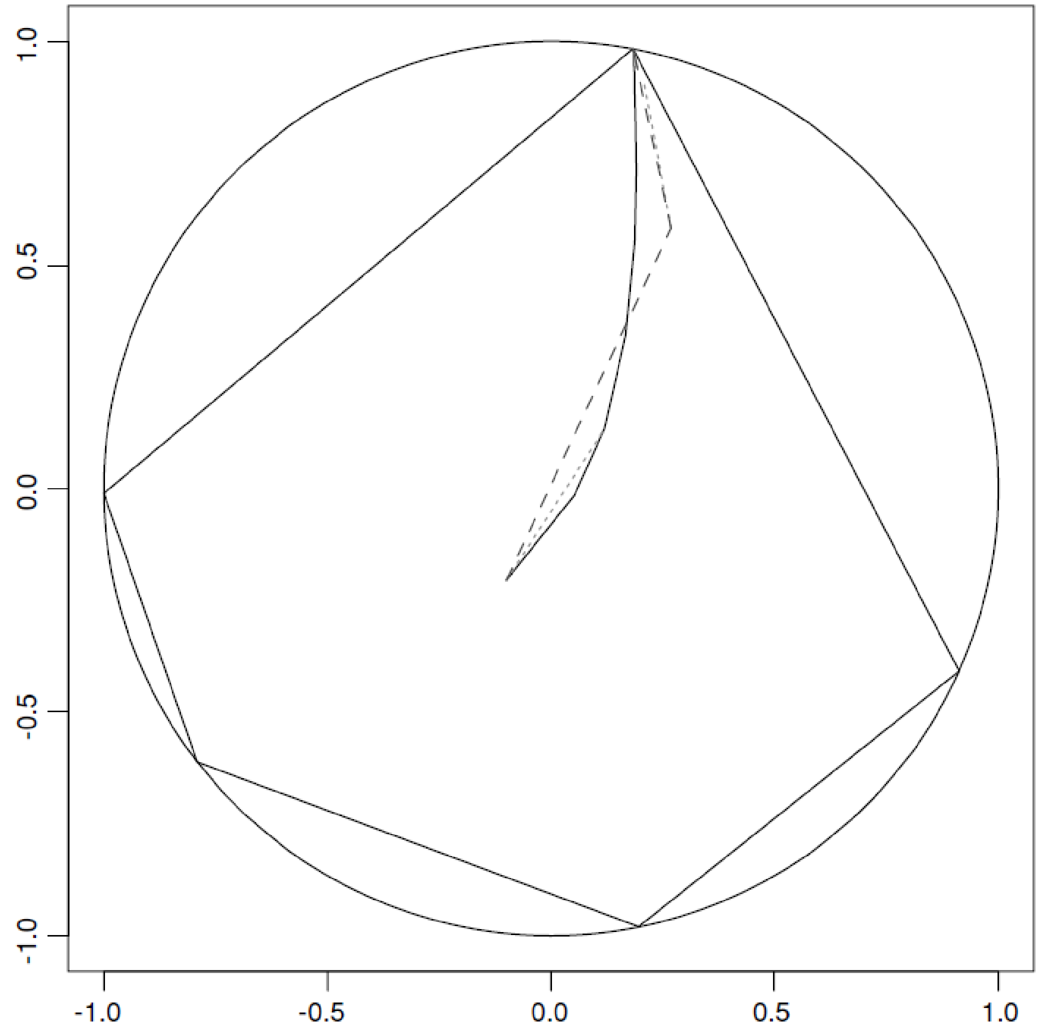
Interior Point Method: an example (10)

- To deal with the nonlinearity in the complementary condition, Mehroha proposed to modify the direction by solving

$$\begin{pmatrix} Z & 0 & D \\ e^\top & 0 & 0 \\ 0 & e & I \end{pmatrix} \begin{pmatrix} \delta_d \\ \delta_a \\ \delta_z \end{pmatrix} = \begin{pmatrix} \mu e - Dz - P_d p_z \\ 0 \\ 0 \end{pmatrix}$$

Interior Point Method: an example (11)

- Modified direction



QR: interior point method (1)

- QR

$$\min_{b \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau \left(y_i - x_i^\top b \right),$$

- Its equivalent LP

$$\min\{\tau e^\top u + (1-\tau) e^\top v \mid y = Xb + u - v, (u, v) \in \mathbb{R}_+^{2n}\}$$

- The dual

$$\max\{y^\top d \mid X^\top d = 0, \ \ d \in [\tau - 1, \tau]^n\}$$

- Setting $a = d + 1 - \tau$, we get

$$\max\{y^\top a \mid X^\top a = (1-\tau) X^\top e, \ \ a \in [0, 1]^n\}$$

QR: interior point method (2)

- Adding slack variables *s* and the constraint

$$a + s = e$$

- The barrier function is

$$B(a, s, \mu) = y^\top a + \mu \sum (\log a_i + \log s_i)$$

with constraints

$$X^\top a = (1 - \tau) X^\top e$$

$$a + s = e$$

QR: interior point method (3)

- The Lagrangian is

$$
\begin{aligned}
L(a, s, b, u, \mu) = B(a, s, \mu) - b^\top(X^\top a - (1-\tau)X^\top e) \\
- u^\top(a + s - e).
\end{aligned}
$$

- Set the derivative of the Lagrangian as zero and $v = \mu A^{-1}$ We have

$$
\begin{aligned}
X^\top a &= (1-\tau)X^\top e \\
a + s &= e \\
Xb + u - v &= y \\
USe &= \mu e \\
AVe &= \mu e.
\end{aligned}
$$

QR: interior point method (4)

- Applying Newton's method, we get

$$\begin{pmatrix} X^\top & 0 & 0 & 0 & 0 \\ I & I & 0 & 0 & 0 \\ 0 & 0 & I & -I & X \\ 0 & U & S & 0 & 0 \\ V & 0 & 0 & A & 0 \end{pmatrix} \begin{pmatrix} \delta_a \\ \delta_s \\ \delta_u \\ \delta_v \\ \delta_b \end{pmatrix} = \begin{pmatrix} (1-\tau)X^\top e - X^\top a \\ e - a - s \\ y - Xb - u + v \\ \mu e - USe \\ \mu e - AVe, \end{pmatrix}$$

- Solving for this,

$$\begin{aligned} \delta_b &= (X^\top W X)^{-1}((1-\tau)X^\top e - X^\top a - X^\top W \xi(\mu)) \\ \delta_a &= W(X\delta_b + \xi(\mu)) \\ \delta_s &= -\delta_a \\ \delta_u &= \mu S^{-1} e - Ue + S^{-1} U \delta_a \\ \delta_v &= \mu A^{-1} e - Ve + A^{-1} V \delta_s, \end{aligned}$$

$$\xi(\mu) = y - Xb + \mu(A^{-1} - S^{-1})e \qquad W = (S^{-1}U + A^{-1}V)^{-1}$$

QR: interior point method (5)

- Applying Newton's method, we get

$$\begin{pmatrix} X^\top & 0 & 0 & 0 & 0 \\ I & I & 0 & 0 & 0 \\ 0 & 0 & I & -I & X \\ 0 & U & S & 0 & 0 \\ V & 0 & 0 & A & 0 \end{pmatrix} \begin{pmatrix} \delta_a \\ \delta_s \\ \delta_u \\ \delta_v \\ \delta_b \end{pmatrix} = \begin{pmatrix} (1-\tau)X^\top e - X^\top a \\ e - a - s \\ y - Xb - u + v \\ \mu e - USe \\ \mu e - AVe, \end{pmatrix}$$

- Solving for this,

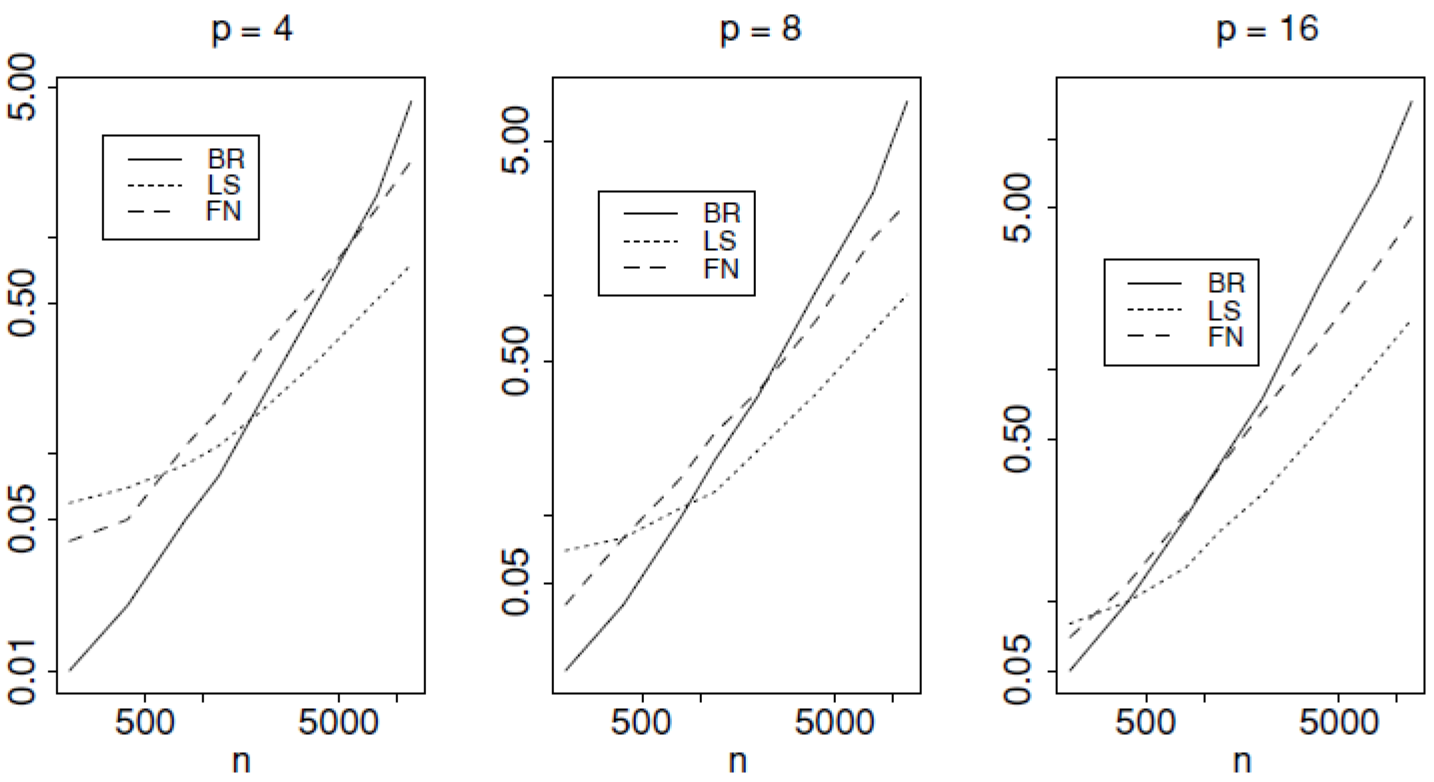
$$\begin{aligned} \delta_b &= (X^\top W X)^{-1}((1-\tau)X^\top e - X^\top a - X^\top W \xi(\mu)) \\ \delta_a &= W(X\delta_b + \xi(\mu)) \\ \delta_s &= -\delta_a \\ \delta_u &= \mu S^{-1} e - Ue + S^{-1} U \delta_a \\ \delta_v &= \mu A^{-1} e - Ve + A^{-1} V \delta_s, \end{aligned}$$

$$\xi(\mu) = y - Xb + \mu(A^{-1} - S^{-1})e \qquad W = (S^{-1}U + A^{-1}V)^{-1}$$

Mehrotra Primal-dual Predictor-corrector Algorithm

- Better numerical stability and efficiency due to better central path
- Easily generalized to exploit sparsity of the design matrix
- Used in the package quantreg

QR: Interior VS exterior



BR: Barrodale and Roberts algorithm
LS: Least Square
FN: Frisch-Newton

Globbing for median regression

- Consider the median regression

$$\min_b \sum_{i=1}^{n} |y_i - x_i^\top b|,$$

- Its directional derivative is

$$g(b, w) = \sum_{i=1}^{n} x_i^\top w \,\mathrm{sgn}^*(y_i - x_i^\top b, x_i^\top w),$$

$$\mathrm{sgn}^*(u, v) = \begin{cases} \mathrm{sgn}(u) & \text{if } u \neq 0 \\ \mathrm{sgn}(v) & \text{if } u = 0. \end{cases}$$

Globbing for median regression (1)

- Suppose that we "knew" that a certain subset of J_H fall above the optimal median plane and J_L fall below the median plane.
- Consider the revised problem

$$\min_{b \in \mathbb{R}^p} \sum_{i \in N \setminus (J_L \cup J_H)} |y_i - x_i^\top b| + |y_L - x_L^\top b| + |y_H - x_H^\top b|,$$

$$(x_L, y_L) = \Big(\sum_{i \in J_L} x_i, -\infty\Big), \quad (x_H, y_H) = \Big(\sum_{i \in J_H} x_i, +\infty\Big)$$

Globbing for median regression (2)

Preliminary estimation using random $m = n^{2/3}$ subset,

Construct confidence band $x_i^\top \hat{\beta} \pm \kappa \| \hat{V}^{1/2} x_i \|$.

Find $J_L = \{i | y_i \text{ below band } \}$, and $J_H = \{i | y_i \text{ above band } \}$,

Glob observations together to form pseudo observations:

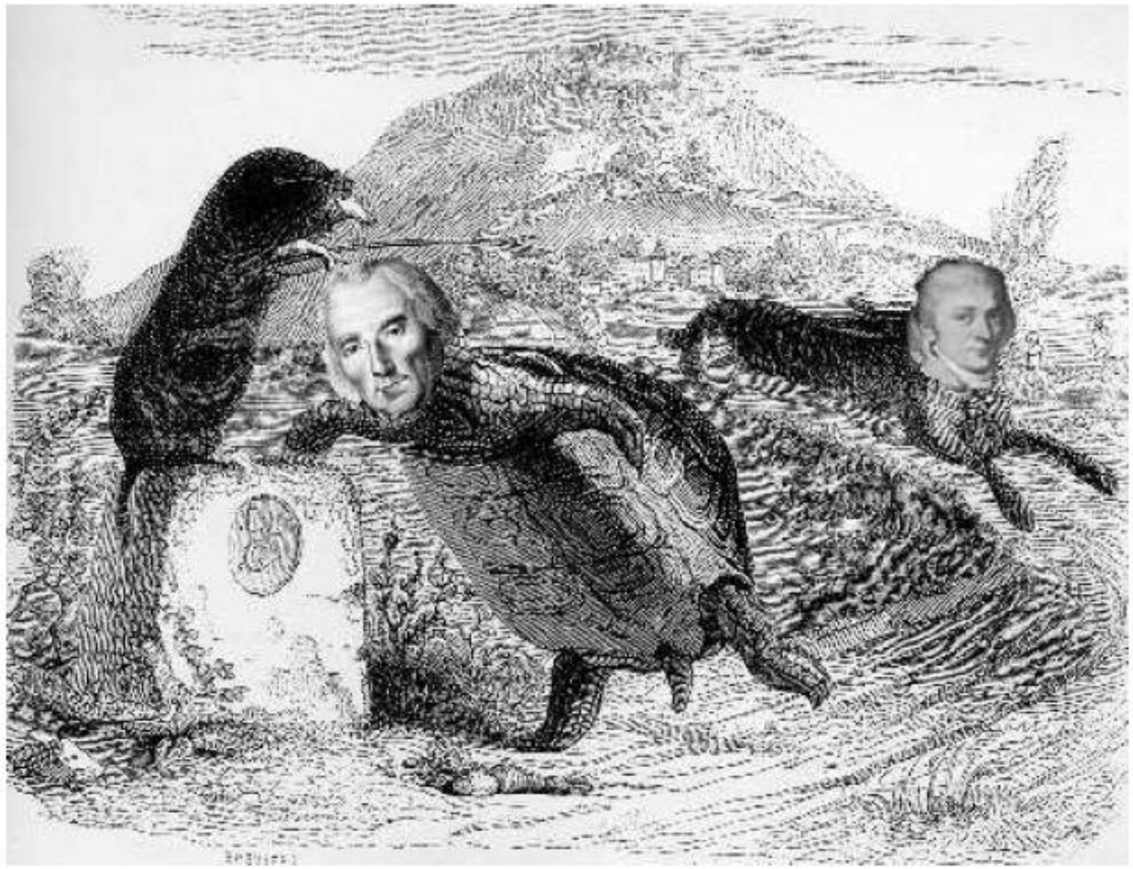
$$(x_L, y_L) = (\sum_{i \in J_L} x_i, -\infty), \quad (x_H, y_H) = (\sum_{i \in J_H} x_i, +\infty)$$

Solve the problem (with m+2 observations)

$$\min \sum |y_i - x_i b| + |y_L - x_L b| + |y_H - x_H b|$$

Verify that globbed observations have the correct predicted signs.

The Laplacian Tortoise and the Gausian Hare



Taken from Portnoy and Koenker (1997)

Locally polynomial quantile regression (1)

- Suppose we have bivariate observations

$$\{(x_i, y_i)\ i = 1, \ldots, n\}$$

- We would like to estimate the τth conditional quantile function of Y given X

$$g(x) = Q_Y(\tau | x).$$

Locally polynomial quantile regression (2)

- Let *K* be a positive, symmetric, unimodal kernal function
- We may consider

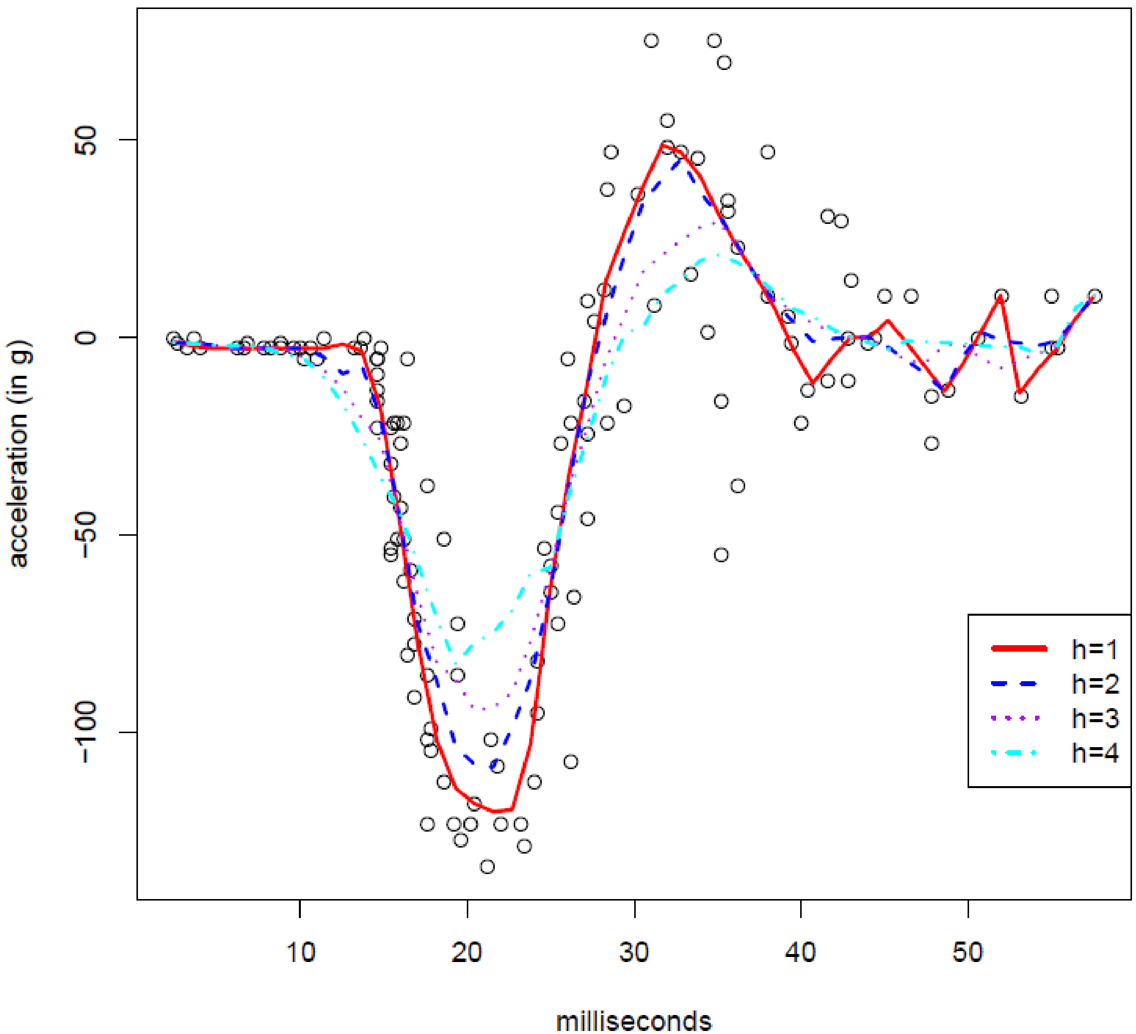
$$\min_{\beta \in \mathbb{R}^2} \sum_{i=1}^{n} w_i(x) \rho_\tau(y_i - \beta_0 - \beta_1(x_i - x))$$

$$w_i(x) = K((x_i - x)/h)/h$$

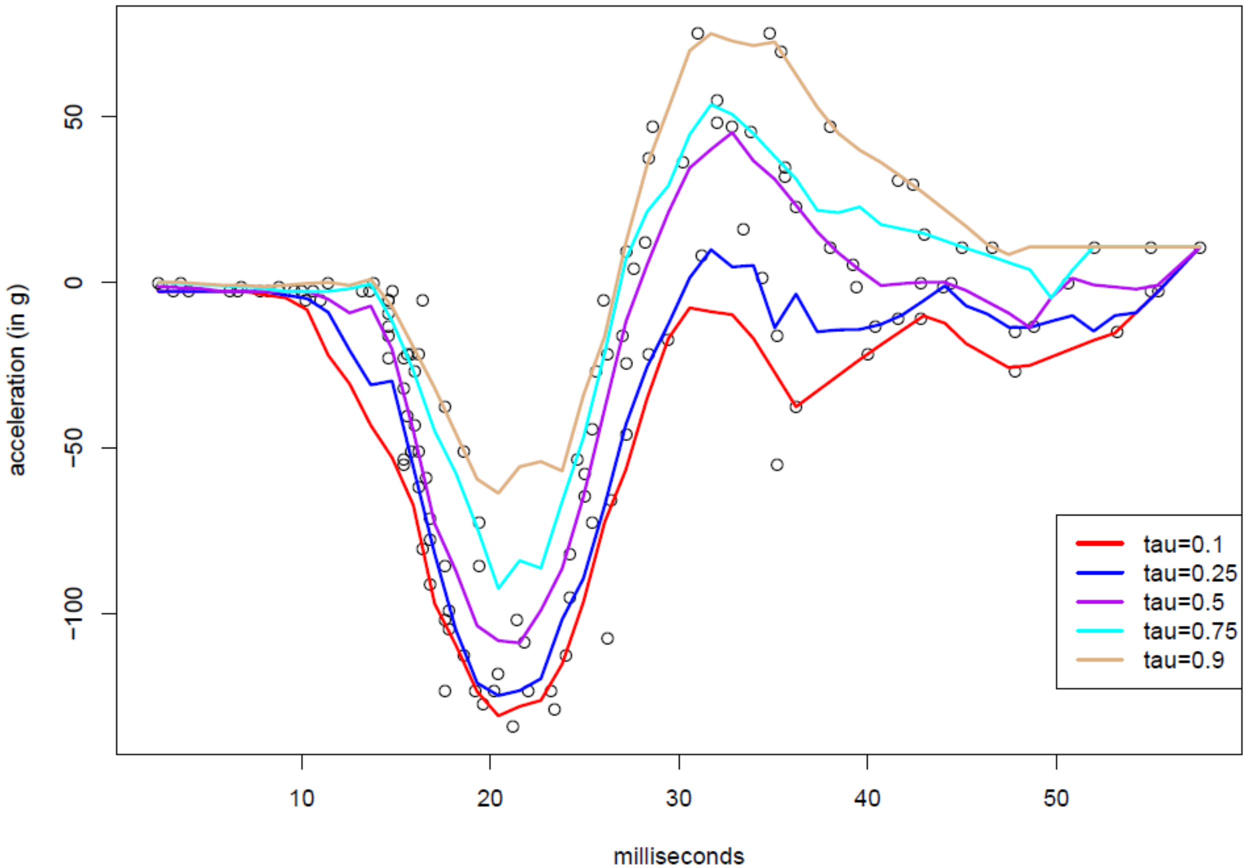
- More generally, we can consider

$$\min_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} w_i(x) \rho_\tau(y_i - \beta_0 - \beta_1(x_i - x) - \cdots - \beta_p(x_i - x)^p)$$

Locally polynomial quantile regression (3)



Locally polynomial quantile regression (3)



Locally polynomial quantile regression (4)

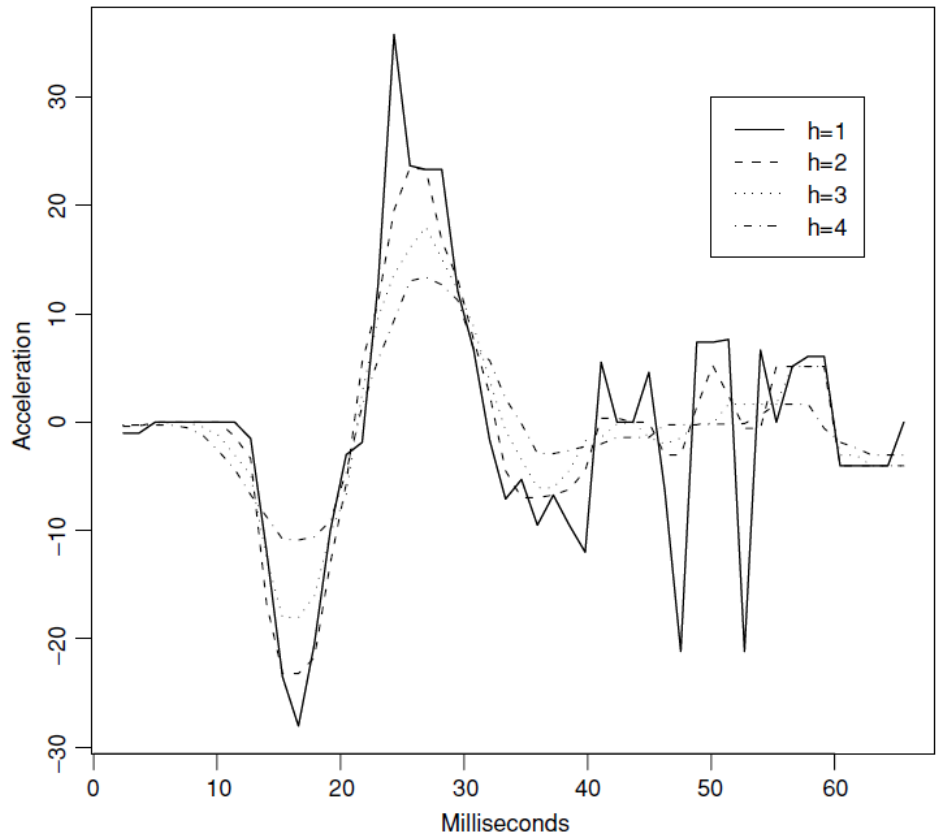


Figure 7.2. Locally linear median regression. Four estimates of the derivative of the acceleration curves for differing choices of the bandwidth parameter.