Regresión cuantil lineal local
Un novo paquete de R

Mercedes Conde Amboage

Departamento de Análisis Matemático, Estatística e Optimización
Universidade de Santiago de Compostela

Definición

Dada unha variable aleatoria X, pódese definir o **cuantil de orde** τ como o valor que verifica :

$$\mathbb{P}(X \leq c_\tau) \geq \tau$$
$$\mathbb{P}(X \geq c_\tau) \geq 1 - \tau$$

Cálculo con R

```
> x=c(1.68,1.85,1.63,1.65,1.65)
> quantile(x,prob=c(0.25,0.5,0.75))
25%  50%  75%
1.65 1.65 1.68
```

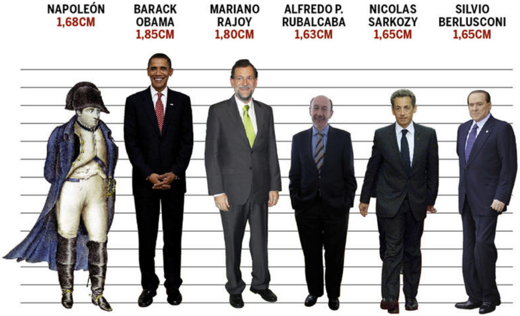
Definición

Dada unha variable aleatoria X, pódese definir o **cuantil de orde** τ como o valor que verifica :

$$\mathbb{P}(X \leq c_\tau) \geq \tau$$
$$\mathbb{P}(X \geq c_\tau) \geq 1 - \tau$$

Cálculo con R

```
> x=c(1.68,1.85,1.63,1.65,1.65)
> quantile(x,prob=c(0.25,0.5,0.75))
25%  50%  75%
1.65 1.65 1.68
```



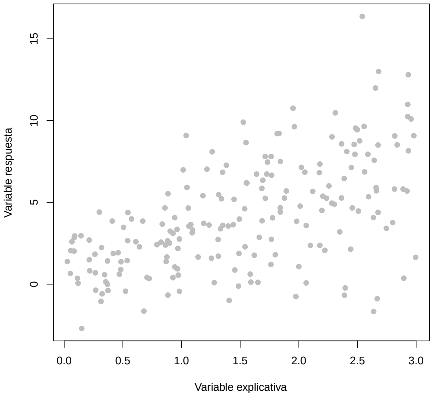
Definición

Dada unha variable aleatoria X, pódese definir o **cuantil de orde** τ como o valor que verifica :

$$\mathbb{P}(X \leq c_\tau) \geq \tau$$
$$\mathbb{P}(X \geq c_\tau) \geq 1 - \tau$$

Cálculo con R

```
> x=c(1.68,1.85,1.63,1.65,1.65)
> quantile(x,prob=c(0.25,0.5,0.75))
25%  50%  75%
1.65 1.65 1.68
```



Regresión cuantil

Sexa un modelo de regresión

$$Y_i = X_i'\beta + \varepsilon_i$$

onde $\mathbb{P}(\varepsilon_i \leq 0|X) = \tau$.

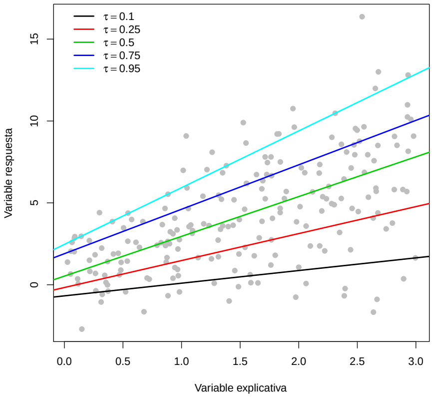
ESTIMACIÓN

$$\widehat{\beta} = \min_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} \rho_\tau(Y_i - X_i'\beta)$$

Función de perda cuantílica

Cálculo con R

```
> library(quantreg)
> mod=rq(y ~ x,tau=0.5)
```



Regresión cuantil

Sexa un modelo de regresión

$$Y_i = X_i'\beta + \varepsilon_i$$

onde $\mathbb{P}(\varepsilon_i \leq 0|X) = \tau$.

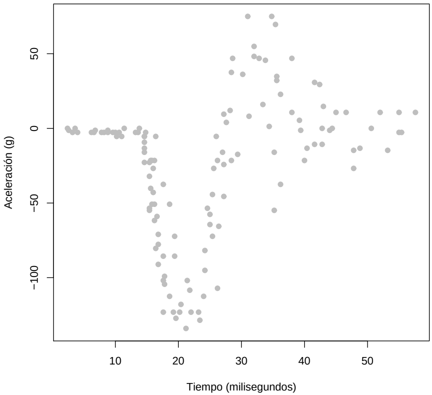
ESTIMACIÓN

$$\widehat{\beta} = \min_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} \rho_\tau(Y_i - X_i'\beta)$$

Función de perda cuantílica

Cálculo con R

```
> library(quantreg)
> mod=rq(y ~ x,tau=0.5)
```



Regresión cuantil non paramétrica

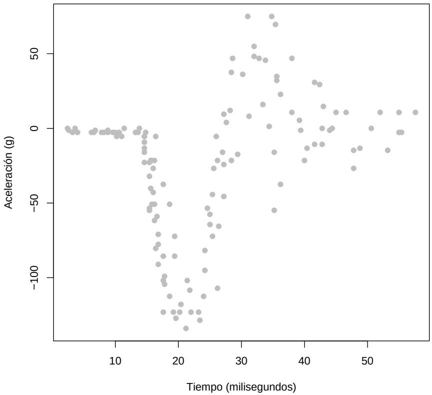
Sexa un modelo de regresión

$$Y_i = q_\tau(X_i) + \varepsilon_i$$

onde $\mathbb{P}(\varepsilon_i \leq 0|X) = \tau$.

Regresión cuantil lineal local

$$\widehat{q}_\tau(x) = \arg\min_a \sum_{i=1}^{n} \rho_\tau(Y_i - a - b(X_i - x))K\left(\frac{x - X_i}{h_\tau}\right)$$



Regresión cuantil non paramétrica

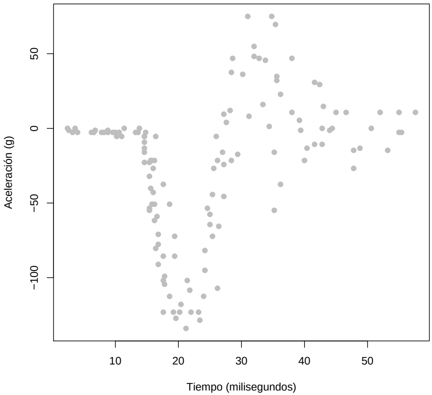
Sexa un modelo de regresión

$$Y_i = q_\tau(X_i) + \varepsilon_i$$

onde $\mathbb{P}(\varepsilon_i \leq 0|X) = \tau$.

Regresión cuantil lineal local

$$\widehat{q}_\tau(x) = \arg\min_a \sum_{i=1}^{n} \rho_\tau(Y_i - a - b(X_i - x))K\left(\frac{x - X_i}{h_\tau}\right)$$



Regresión cuantil non paramétrica

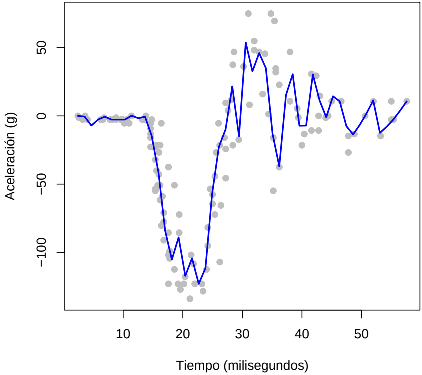
Sexa un modelo de regresión

$$Y_i = q_\tau(X_i) + \varepsilon_i$$

onde $\mathbb{P}(\varepsilon_i \leq 0|X) = \tau$.

Regresión cuantil lineal local

$$\widehat{q}_\tau(x) = \arg\min_a \sum_{i=1}^{n} \rho_\tau(Y_i - a - b(X_i - x))K\left(\frac{x - X_i}{h_\tau}\right)$$



Regresión cuantil non paramétrica

Sexa un modelo de regresión

$$Y_i = q_\tau(X_i) + \varepsilon_i$$

onde $\mathbb{P}(\varepsilon_i \leq 0|X) = \tau$.

Regresión cuantil lineal local

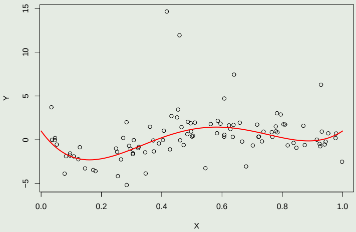
$$\widehat{q}_\tau(x) = \arg\min_a \sum_{i=1}^{n} \rho_\tau(Y_i - a - b(X_i - x))K\left(\frac{x - X_i}{h_\tau}\right)$$

Erro cadrático medio (ECM)

$$ECM(\widehat{q}_\tau(x)) \cong \frac{1}{4}h^4\mu_2(K)^2 q''_\tau(x)^2 + \frac{R(K)\tau(1-\tau)}{nh_\tau g(x) f(q_\tau(x)|x)^2}$$

sendo:

- g a densidade marxinal de X
- $f(q_\tau(x)|x)$ a densidade condicional da resposta no cuantil
- $\mu_2(K) = \int u^2 K(u)du$
- $R(K) = \int K^2(u)du$



O parámetro de suavizado óptimo

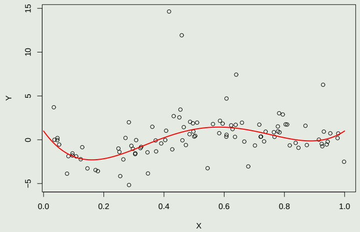
$$h_\tau^5 = \frac{R(K)\tau(1-\tau)}{n\mu_2(K)^2 \int q''_\tau(x)^2 g(x)dx} \int \frac{1}{f(q_\tau(x)|x)^2}dx$$

Erro cadrático medio (ECM)

$$ECM(\widehat{q}_\tau(x)) \cong \frac{1}{4} h^4 \mu_2(K)^2 q''_\tau(x)^2 + \frac{R(K)\tau(1-\tau)}{nh_\tau g(x) f(q_\tau(x)|x)^2}$$

sendo:

- g a densidade marxinal de X
- $f(q_\tau(x)|x)$ a densidade condicional da resposta no cuantil
- $\mu_2(K) = \int u^2 K(u)du$
- $R(K) = \int K^2(u)du$



O parámetro de suavizado óptimo

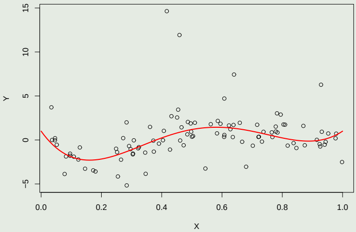
$$h_\tau^5 = \frac{R(K)\tau(1-\tau)}{n\mu_2(K)^2 \int q''_\tau(x)^2 g(x)dx} \int \frac{1}{f(q_\tau(x)|x)^2} dx$$

Erro cadrático medio (ECM)

$$ECM(\widehat{q}_\tau(x)) \cong \frac{1}{4}h^4\mu_2(K)^2 q_\tau''(x)^2 + \frac{R(K)\tau(1-\tau)}{nh_\tau g(x) f(q_\tau(x)|x)^2}$$

sendo:

- g a densidade marxinal de X
- $f(q_\tau(x)|x)$ a densidade condicional da resposta no cuantil
- $\mu_2(K) = \int u^2 K(u)du$
- $R(K) = \int K^2(u)du$



O parámetro de suavizado óptimo

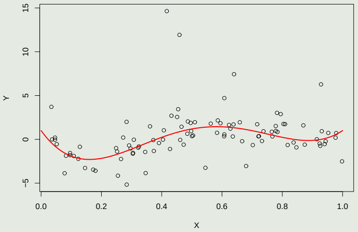
$$h_\tau^5 = \frac{R(K)\tau(1-\tau)}{n\mu_2(K)^2 \boxed{\int q_\tau''(x)^2 g(x)dx}} \boxed{\int \frac{1}{f(q_\tau(x)|x)^2}dx}$$

Erro cadrático medio (ECM)

$$ECM(\widehat{q}_\tau(x)) \cong \frac{1}{4} h^4 \mu_2(K)^2 q_\tau''(x)^2 + \frac{R(K)\tau(1-\tau)}{nh_\tau g(x) f(q_\tau(x)|x)^2}$$

sendo:

- g a densidade marxinal de X
- $f(q_\tau(x)|x)$ a densidade condicional da resposta no cuantil
- $\mu_2(K) = \int u^2 K(u) du$
- $R(K) = \int K^2(u) du$



O parámetro de suavizado óptimo

$$h_\tau^5 = \frac{R(K)\tau(1-\tau)}{n\mu_2(K)^2 \boxed{\int q_\tau''(x)^2 g(x) dx}} \boxed{\int \frac{1}{f(q_\tau(x)|x)^2} dx}$$

Package 'BwQuant'

May 10, 2016

Type Package

Title Bandwidth selectors for local linear quantile regression

Version 1.0

Date 2016-05-10

Author Mercedes Conde-Amboage and César Sánchez-Sellero

Maintainer Mercedes Conde-Amboage <mercedes.amboage@usc.es>

Depends quantreg, KernSmooth, numDeriv

Description Different bandwidth selectors for local liner quantile regression

License What license is it under? (GLP-2)

Archs i386, x64

R topics documented:

BwQuant-package *Bandwidth selectors for local linear quantile regression*

Description

Selector plug-in propuesto por Yu e Jones (1998)

Primeira suposición

$$q''_{\tau_1}(x) = q''_{\tau_2}(x)$$

Segunda suposición

Normalidade dos erros

$$h_\tau = \sqrt[5]{\frac{\tau(1-\tau)}{\phi(\Phi^{-1}(\tau))^2}}\, h_{media}$$

Cálculo con R

```
> data(mcycle)
> attach(mcycle)
> bwYJ(times,accel,0.75)
```

Selector plug-in propuesto por Yu e Jones (1998)

Primeira suposición

$$q''_{\tau_1}(x) = q''_{\tau_2}(x)$$

Segunda suposición

Normalidade dos erros

$$h_\tau = \sqrt[5]{\frac{\tau(1-\tau)}{\phi(\Phi^{-1}(\tau))^2}}\, h_{media}$$

Cálculo con R

```
> data(mcycle)
> attach(mcycle)
> bwYJ(times,accel,0.75)
```

Selector plug-in propuesto por Yu e Jones (1998)

Primeira suposición

$$q''_{\tau_1}(x) = q''_{\tau_2}(x)$$

Segunda suposición

Normalidade dos erros

$$h_\tau = \sqrt[5]{\frac{\tau(1-\tau)}{\phi(\Phi^{-1}(\tau))^2}}\, h_{media}$$

Cálculo con R

```
> data(mcycle)
> attach(mcycle)
> bwYJ(times,accel,0.75)
```

Selector basado en validación cruzada (Abberger (1998))

$$\widehat{h}_{CV} = \arg\min_{h} \sum_{i=1}^{n} \rho_\tau \left(Y_i - \widehat{q}_{\tau,h}^{-i}(X_i) \right)$$

clásico estimador **leave-one-out**

Cálculo con R

```
> data(mcycle)
> attach(mcycle)
> hseq=seq(0.5,3,length=50)
> bwCV(times,accel,hseq,0.75)
```

Selector basado en validación cruzada (Abberger (1998))

$$\widehat{h}_{CV} = \arg\min_{h} \sum_{i=1}^{n} \rho_\tau \left(Y_i - \widehat{q}^{\,-i}_{\tau,h}(X_i) \right)$$

clásico estimador **leave-one-out**

Cálculo con R

```
> data(mcycle)
> attach(mcycle)
> hseq=seq(0.5,3,length=50)
> bwCV(times,accel,hseq,0.75)
```

Selector basado na regra do pulgar

Procedemento

- Dividimos a mostra en bloques.
- En cada bloque axustamos un modelo cuantil de orde 4.
- Construímos **estimacións paramétricas** da curvatura e da sparsity.

$$\widehat{h}_{RP} = \left(\frac{R(K)\tau(1-\tau)}{\mu_2(K)^2} \frac{\widehat{s}^b(b-a)}{\widehat{\theta}_{22}^b n} \right)^{1/5}$$

Cálculo con R

```
> data(mcycle)
> attach(mcycle)
> bwRT(times,accel,0.75)
```

Selector basado na regra do pulgar

Procedemento

- Dividimos a mostra en bloques.
- En cada bloque axustamos un modelo cuantil de orde 4.
- Construímos **estimacións paramétricas** da curvatura e da sparsity.

$$\widehat{h}_{RP} = \left(\frac{R(K)\tau(1-\tau)}{\mu_2(K)^2} \frac{\widehat{s}^b(b-a)}{\widehat{\theta}_{22}^b n} \right)^{1/5}$$

Cálculo con R

```
> data(mcycle)
> attach(mcycle)
> bwRT(times,accel,0.75)
```

Selector basado na regra do pulgar

Procedemento

- Dividimos a mostra en bloques.
- En cada bloque axustamos un modelo cuantil de orde 4.
- Construímos **estimacións paramétricas** da curvatura e da sparsity.

$$\widehat{h}_{RP} = \left(\frac{R(K)\tau(1-\tau)}{\mu_2(K)^2} \, \frac{\widehat{s}^b(b-a)}{\widehat{\theta}_{22}^b n} \right)^{1/5}$$

Cálculo con R

```
> data(mcycle)
> attach(mcycle)
> bwRT(times,accel,0.75)
```

Novo selector plug-in

Procedemento

- Construímos **estimacións non paramétricas** da curvatura e da sparsity sen asumir ningunha suposición.

$$\widehat{h}_{PI} = \left(\frac{R(K)\tau(1-\tau)}{\mu_2(K)^2} \frac{\widehat{s}_{\widehat{h}_s,\widehat{d}_s}}{\widehat{\theta}_{22,\widehat{h}_c}\, n} \right)^{1/5}$$

Cálculo con R

```
> data(mcycle)
> attach(mcycle)
> bwPI(times,accel,0.75)
```

Novo selector plug-in

Procedemento

– Construímos **estimacións non paramétricas** da curvatura e da sparsity sen asumir ningunha suposición.

$$\widehat{h}_{PI} = \left(\frac{R(K)\tau(1-\tau)}{\mu_2(K)^2} \, \frac{\widehat{s}_{\widehat{h}_s,\widehat{d}_s}}{\widehat{\theta}_{22,\widehat{h}_c} \, n} \right)^{1/5}$$

Cálculo con R

```
> data(mcycle)
> attach(mcycle)
> bwPI(times,accel,0.75)
```

Novo selector plug-in

Procedemento

– Construímos **estimacións non paramétricas** da curvatura e da sparsity sen asumir ningunha suposición.

$$\widehat{h}_{PI} = \left(\frac{R(K)\tau(1-\tau)}{\mu_2(K)^2} \, \frac{\widehat{s}_{\widehat{h}_s,\widehat{d}_s}}{\widehat{\theta}_{22,\widehat{h}_c} \; n} \right)^{1/5}$$

Cálculo con R

```
> data(mcycle)
> attach(mcycle)
> bwPI(times,accel,0.75)
```

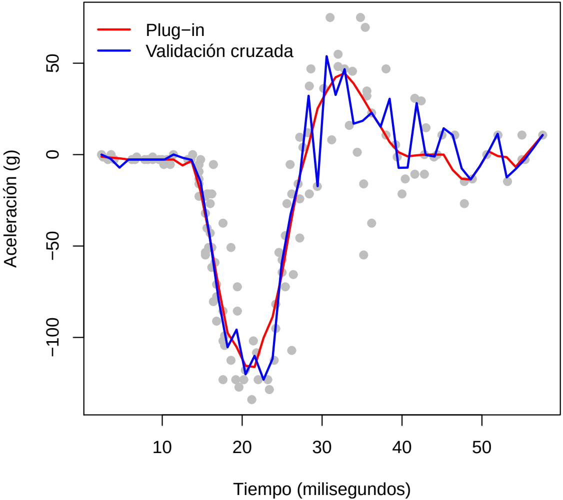


Figura: Representación do axuste para diferentes selectores do parámetro de suavización asociados a $\tau = 0.5$.

Regresión cuantil lineal local
Un novo paquete de R

Mercedes Conde Amboage

Departamento de Análisis Matemático, Estatística e Optimización
Universidade de Santiago de Compostela

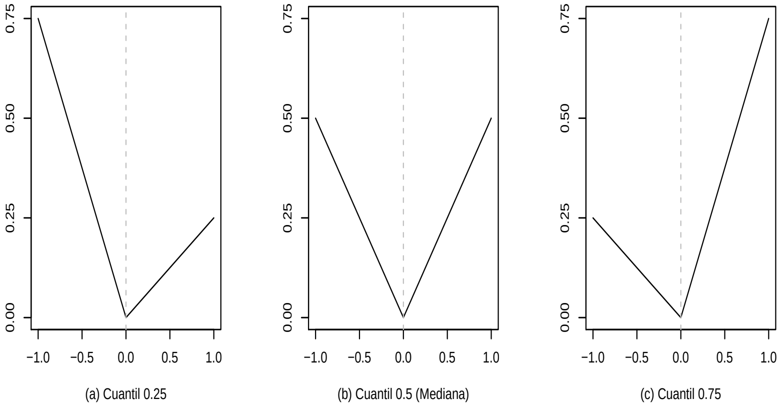


Figura: Representación da función de perda cuantílica para diferentes valores do cuantil de interese.

Volver