Estimación no-paramétrica

Máximo Camacho Alonso

Universidad de Murcia

www.um.es/econometria/tecpre

mcamacho@um.es

Contenido del tema

- Introducción: ventajas e inconvenientes
- Estimación función de densidad
  - Algunos tipos de estimación
  - Propiedades asintóticas
  - El papel del kernel
  - El papel de la ventana
- Estimación de funciones
  - Nadaraya-Watson
  - Predicción

1. Introducción

- Estimación paramétrica
- Estimación no-paramétrica
- Ventajas
- Inconvenientes

1. Introducción

- Supongamos que
  - Disponemos de n observaciones de y, x.
  - Pensamos

$$y_t = m(x_t) + e_t$$

- Deseamos estimar
  - Función de densidad de y
  - La función m

1. Introducción

1.1. Estimación paramétrica

- Supuesto 1: m es lineal $y_t = a + bx_t + e_t$
- Supuesto 2: $e \approx N(0, \sigma^2)$
  - Implica $y_t / x_t \approx N(a + bx_t, \sigma^2)$
- Estimamos a, b, σ

$$\hat{y}_t / x_t \approx N(\hat{a} + \hat{b}x_t, \hat{\sigma}^2)$$

1. Introducción

1.2. Estimación no-paramétrica

- No exigimos a priori que
  - la densidad de y
  - la función m

  pertenezcan a una familia paramétrica

- Trataremos de estimarlas dejando que los datos “hablen por sí mismos”

1. Introducción

1.3. Ventajas

- Principio de parquedad
- Método flexible
  - ¿Qué método paramétrico?
- Fácil de implementar
- Evitamos el error al elegir
  - funciones de densidad erróneas
  - m errónea

1. Introducción

1.4. Inconvenientes

- Muestras grandes
- Existen varios no-paramétricos
  - ¿Qué método usar?
- Resultados menos intuitivos
- Inferencia

2. Funciones de densidad

- Introducción
- Tipos de estimación no-paramétrica
  - Histograma
  - Kernel
  - Kernel adaptativa
  - "Vecino más cercano"
- Estimación kernel
  - Propiedades
  - Datos multivariantes

2. Estimación funciones densidad

2.1. Introducción

- Densidad univariante
- Ejemplo
  - Copas y Fryer (1980)
  - y_t duración 86 pacientes tras suicidio
  - Nótese:
    - y > 0
    - y suele ser pequeña

2. Estimación funciones densidad

2.2. Tipos de estimación no-paramétrica

- Histograma
- Kernel: Rosenblatt (1956)
- Vecino más cercano: Loftsgaarcen y Quesenberry (1965)
- Kernel variable: Breiman, Meisel y Purcell (1977)

2. Estimación funciones densidad

2.2. Tipos de estimación no-paramétrica

1. Histograma

- Sean:
  - Un punto de partida y_0
  - Un radio h
- Para cada natural m se define
  - La bola $[\, y_0+mh,\, y_0+(m+1)h\,)$, $m = 0,1,2,\ldots$

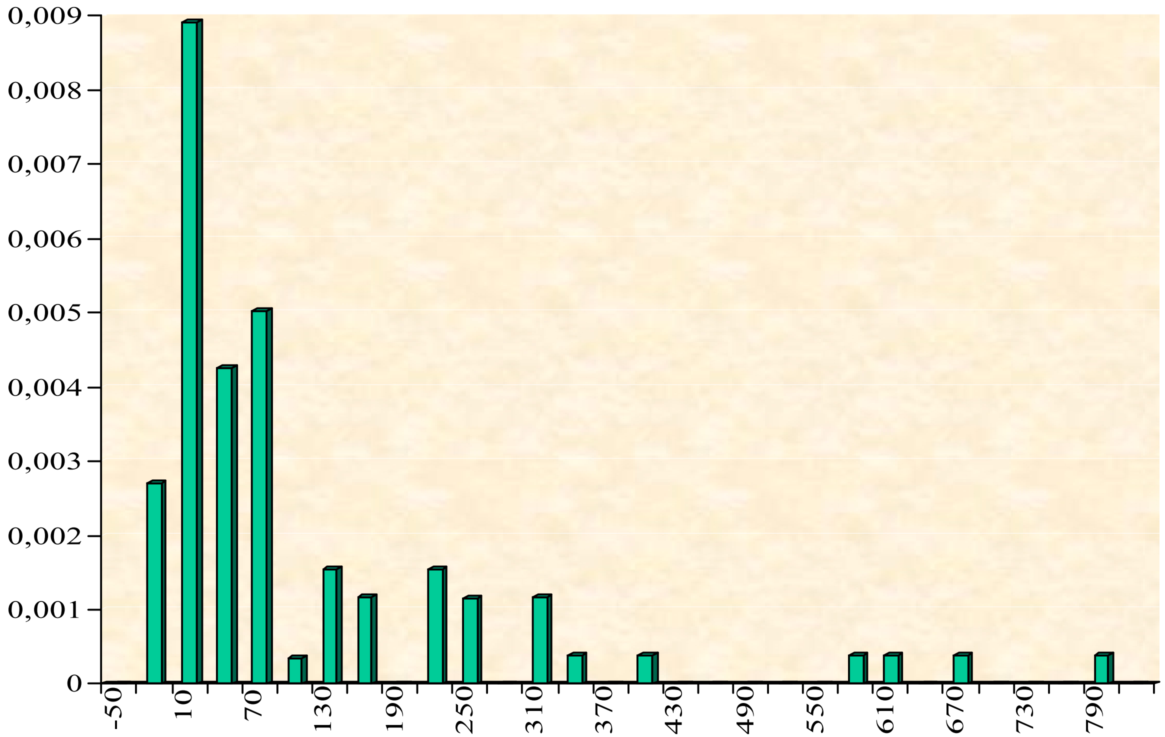
$$\hat{f}(y_i) = \frac{\# B_i}{nh}$$

2. Estimación funciones densidad

2.2. Tipos de estimación no-paramétrica

1. Histograma

- $y_0 = -50$
- $h = 30$



2. Estimación funciones densidad

2.2. Tipos de estimación no-paramétrica

1. Histograma

- **Ventajas**
  – Fácil aplicación

- **Inconvenientes**
  – Depende de y_0
  – Discontinuidades: no derivadas
  – Múltiples explicativas

2. Estimación funciones densidad

2.2. Tipos de estimación no-paramétrica

2. Estimación kernel: definiciones

- Una función de densidad f(y) satisface:

1. $\int_{-\infty}^{\infty} f(y)dy = 1$

2. $f(y) \geq 0$

- Una función kernel K(y) sólo satisface 1

2. Estimación funciones densidad

2.2. Tipos de estimación no-paramétrica

2. Estimación kernel: estimador

- Definimos estimador kernel:

$$\hat{f}(y) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{y - y_i}{h} \right)$$

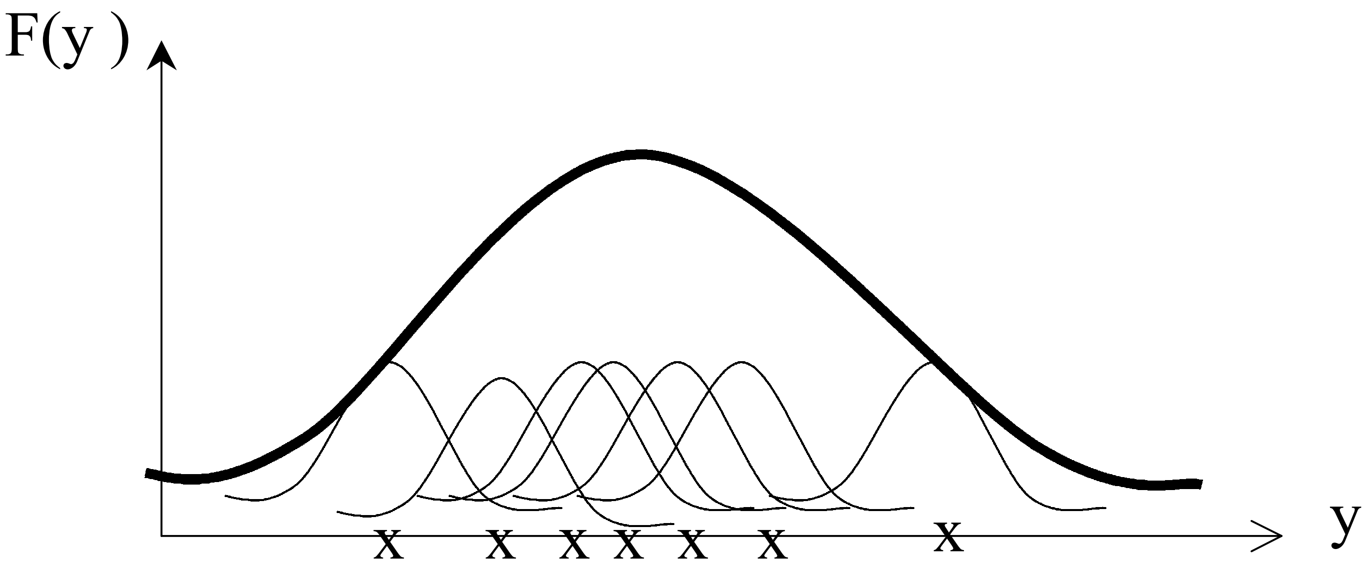
- Simplificamos: K es la normal estandarizada

2. Estimación funciones densidad

2.2. Tipos de estimación no-paramétrica

2. Estimación kernel: intuición

$$\hat{f}(y) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{y - y_i}{h}\right)$$

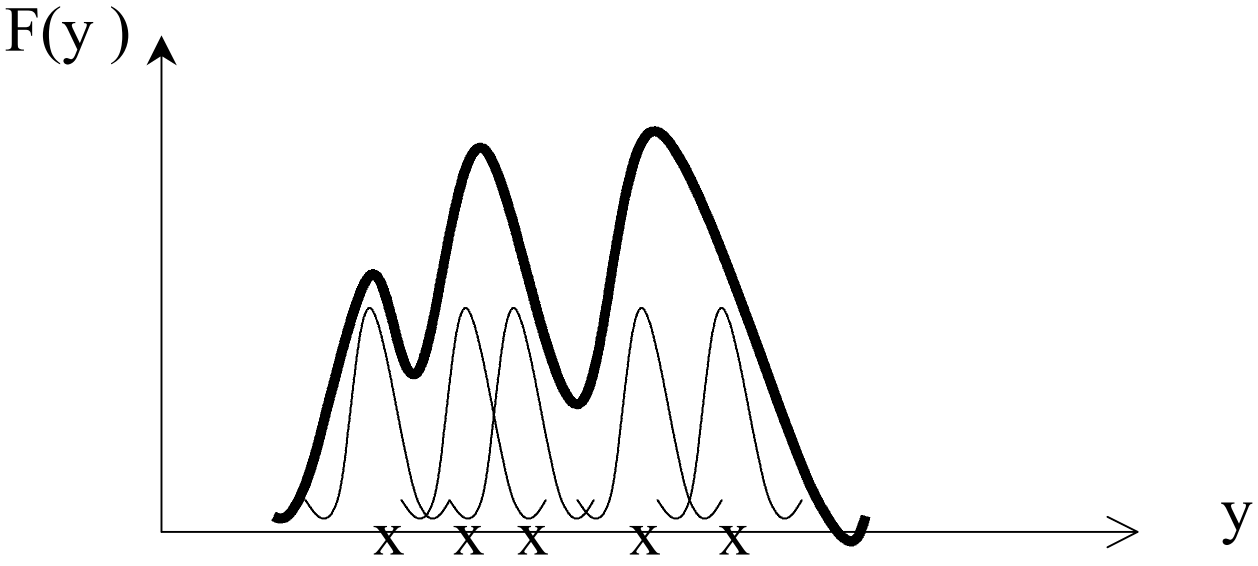


2. Estimación funciones densidad

2.2. Tipos de estimación no-paramétrica

2. Estimación kernel: el papel de h

$$\hat{f}(y) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{y - y_i}{h}\right)$$



2. Estimación funciones densidad

2.2. Tipos de estimación no-paramétrica

2. Estimación kernel: kernel a examen

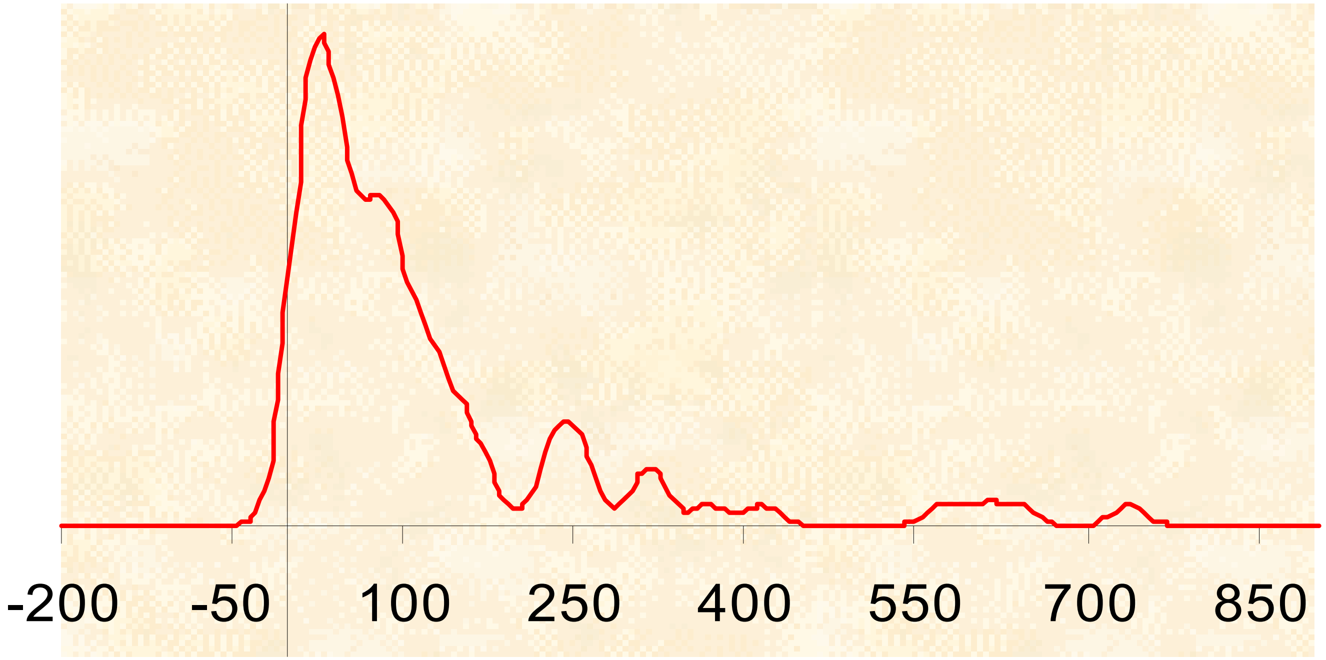
- **Ventajas: ciertos K**
  - $\hat{f}(y)$ función densidad
  - $\hat{f}(y)$ continua
  - $\hat{f}(y)$ diferenciable
- **Inconvenientes**
  - Elección de K, h
  - h fija: problemas en distribiciones asimétricas

2. Estimación funciones densidad

2.2. Tipos de estimación no-paramétrica

2. Estimación kernel: kernel a examen

h=15

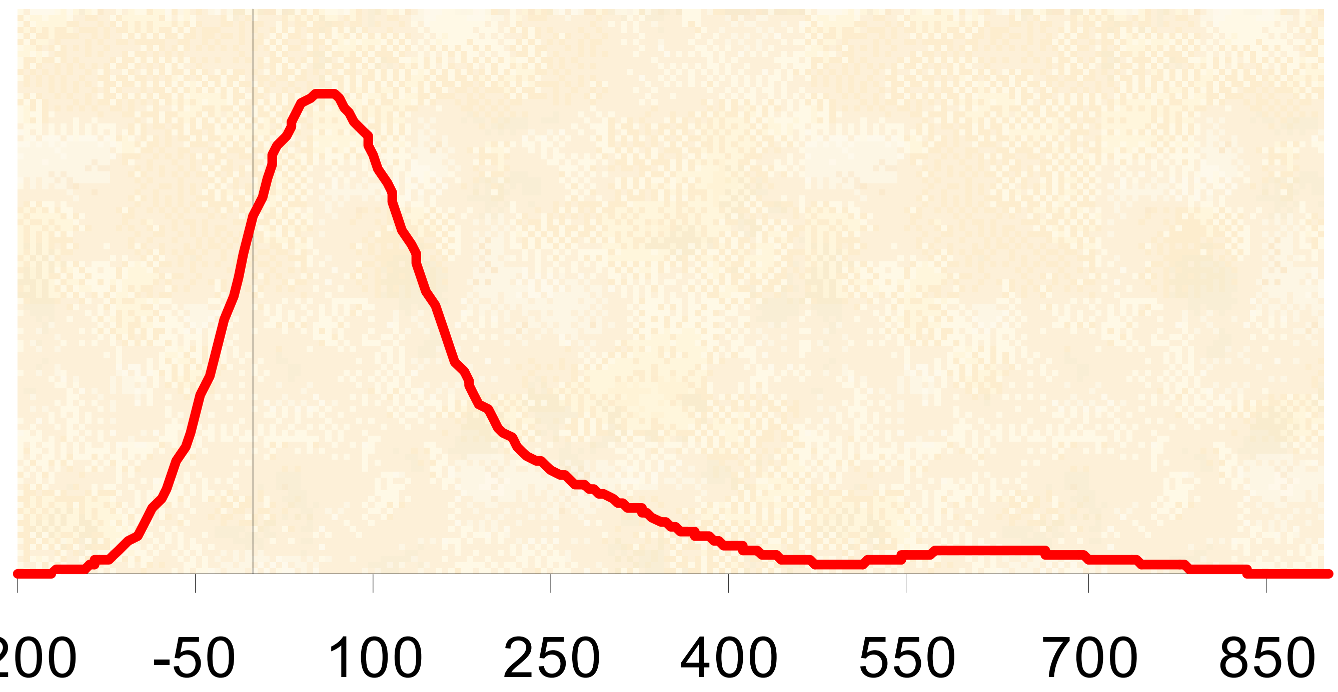


2. Estimación funciones densidad

2.2. Tipos de estimación no-paramétrica

2. Estimación kernel: kernel a examen

h=60



2. Estimación funciones densidad

2.2. Tipos de estimación no-paramétrica

3. Estimación kernel adaptativa

- Definimos

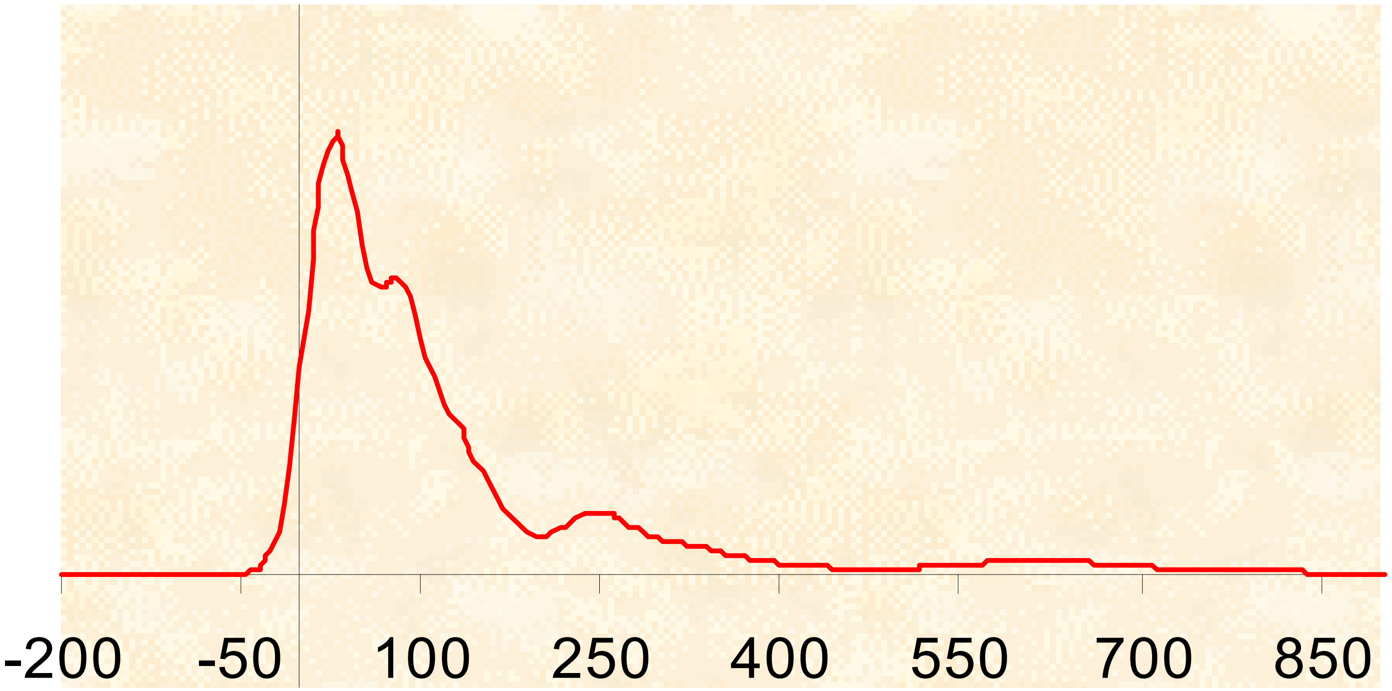
$$\hat{f}(y) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h\lambda_i}K\left(\frac{y-y_i}{h\lambda_i}\right)$$

- λ_i adapta la ventana a cada observación: resta peso a observaciones “aisladas”

2. Estimación funciones densidad

2.2. Tipos de estimación no-paramétrica

3. Estimación kernel adaptativa



2. Estimación funciones densidad

2.2. Tipos de estimación no-paramétrica

4. Estimación "vecino más cercano"

- Consideremos: $d(x,y) = | x - y |$
- Ordenamos: $d_1(y) \leq d_2(y) \leq \ldots \leq d_n(y)$

  $d_n(y)$ = distancia entre (observación más lejana a y, y)

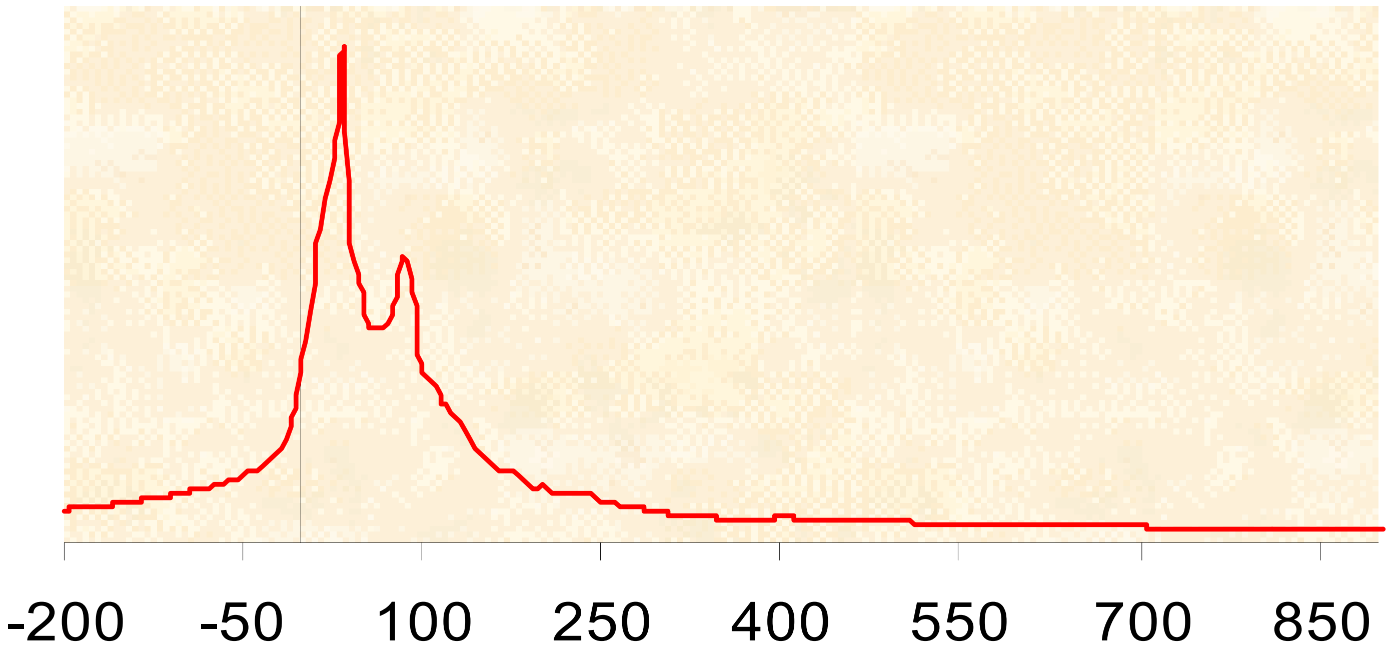
- Definimos $\boxed{\hat{f}(y) = \frac{1}{n d_k(y)} \sum_{i=1}^{n} K\left(\frac{y - y_i}{d_k(y)} \right)}$ $\quad k \sim n$ ^ (1/2)

- Ventana grande (curva suave) en obs. atípicas

2. Estimación funciones densidad

2.2. Tipos de estimación no-paramétrica

4. Estimación "vecino más cercano"



2. Estimación funciones densidad

2.3. Estimación Kernel

1. Supuestos adicionales

- K simétrica

$$\int K(t)\,dt = 1$$

$$\int tK(t)\,dt = 0$$

$$\int t^2 K(t)dt = k_2 \neq 0$$

- f derivadas continuas de cualquier orden

2. Estimación funciones densidad

2.3. Estimación Kernel

2. Error Cuadrático Medio: definición

$$\text{ECM} = \int E\left(f(y) - \hat{f}(y)\right)^2 dy$$

$$= \int \left\{f(y) - E\hat{f}(y)\right\}^2 dy \qquad \int (\text{sesgo})^2\, dy$$

$$+ \int E\left\{\hat{f}(y) - E\hat{f}(y)\right\}^2 dy \qquad \int (\text{varianza})\, dy$$

2. Estimación funciones densidad

2.3. Estimación Kernel

2. Error Cuadrático Medio: sesgo

$$\int (\text{sesgo})^2\, dy$$

Aplicar propiedades de K
Expansión de Taylor

$$\cong \frac{1}{4} h^4 (k_2)^2 \int f''(y)^2\, dy$$

(Aumenta con h)

2. Estimación funciones densidad

2.3. Estimación Kernel

2. Error Cuadrático Medio: varianza

$$\int \text{varianza}\, dy$$

Aplicar propiedades de K
Expansión de Taylor

$$\cong \frac{1}{nh} \int K(y)^2 \, dy$$

(Disminuye con h)

2. Estimación funciones densidad

2.3. Estimación Kernel

3. El papel de la ventana

- h: aumenta sesgo y diminuye varianza
- $h_{opt} = \left(k_2\right)^{-2/5} \left(\int K(y)^2 dt\right)^{1/5} \left(\int f''(y)^2 dy\right)^{-1/5} n^{-1/5}$
  - minimiza el ECM
  - pequeña para
    - n grande
    - densidades fluctuación rápida
- $h_{Nopt} = 1.06\ \sigma\ n^{-1/5}$

2. Estimación funciones densidad

2.3. Estimación Kernel

4. El papel del Kernel

$$\mathrm{ECM}_{\mathrm{opt}} = \quad g\left(n, f'', \int K(y)^2\,dy\right)$$

- Es mínimo para Epanechnikov (Ke)
- Compararemos ECM de otras Kernel con Ke para analizar eficiencia relativa

2. Estimación funciones densidad

2.3. Estimación Kernel

4. El papel del Kernel

Kernel	K(t)	Eficiencia
Epanechnikov	$\begin{cases} \frac{3}{4}\left(1-\frac{1}{5}y^2\right)/\sqrt{5} & \mid y \mid < \sqrt{5} \\ 0 & \text{resto} \end{cases}$	1
Gausiana	$1/\sqrt{2\pi}\ \exp((1/2)y^2)$	0.95
Rectangular	1/2 para $\|y\| < 1$, y 0 resto	0.93

2. Estimación funciones densidad

2.3. Estimación Kernel

4. Elección de h

- Validación cruzada máximo verosímil
- h tal que maximiza

$$VCMV(h) = \frac{1}{n}\sum_{i=1}^{n} \log \hat{f}_{-i}(y_i)$$

- Buscando entre (1/4 h_{Nopt} , 3/2 h_{Nopt})

2. Estimación funciones densidad

2.3. Estimación Kernel

5. Propiedades asintóticas

- Consistencia puntual. Bajo f continua y:
  - $\int |K(y)|\, dy < \infty$, $\int K(y)dy = 1$
  - $|\, y\, K(y)\, | \rightarrow 0$ si $|\, y\, | \rightarrow \infty$
  - $h_n \rightarrow 0$ y $nh_n \rightarrow \infty$ si $n \rightarrow \infty$

entonces

$$\hat{f}(y) \xrightarrow{p} f(y)$$

2. Estimación funciones densidad

2.3. Estimación Kernel

5. Propiedades asintóticas

- Consistencia funcional.
  - Bajo los supuestos anteriores
  - K no negativa

entonces, con probabilidad 1:

$$\int |\hat{f}(y) - f(y)|\, dy \to 0$$

2. Estimación funciones densidad

2.3. Estimación Kernel

6. Datos multivariantes

- Suponemos Y_i vector d-dimensional
- Definición

$$\hat{f}(Y) = \frac{1}{nh^d} \sum_{i=1}^{n} \underline{K}\left(\frac{1}{h}(Y - Y_i) \right)$$

- Ejemplo: si el Kernel es gausiano

$$\underline{K}(Y) = (2\pi)^{-d/2} \exp(-1/2Y'Y)$$

2. Estimación funciones densidad

2.3. Estimación Kernel

6. Datos multivariantes: consideraciones

- Gran importancia de las colas
- n grande. Ej.: para asegurar ECM (y = 0) < 0.1
  - d = 3............n > 67
  - d = 5............n > 770
  - d = 10...........n > 850.000

3. Estimación de funciones

- Introducción
- Nadaraya-Watson
  - Estimación
  - Propiedades asintóticas
- "Vecino más cercano"
- Muchas variables explicativas
- Predicción

3. Estimación de funciones

3.1. Introducción

- Disponemos $\{y_t , x_t \}$ i=1,...,n
- Suponemos $y_t = m(x_t) + e_t$
- Tratamos de estimar m (x) sin hacer ningún supuesto previo sobre la naturaleza de m

3. Estimación de funciones

3.2. Nadaraya-Watson : estimador

- Si $E\ (e\ /\ x) = 0$
- $m\ (x) = E\ (y\ /\ x) = \int y\ f\ (y\ /\ x)\ dy$

$$= \int y\ f\ (y\ ,\ x)\ /\ f\ (x)\ dy$$

- El estimador natural de
  - $f\ (x)$ es $K\ (x)$
  - $f\ (y\ ,\ x) = \underline{K}\ (y\ ,\ x)$

3. Estimación de funciones

3. 2. Nadaraya-Watson: estimador

- Bajo los supuestos $\underline{K}$
  - Media 0: $\int y \, \underline{K}\,(y\,,\,x)\,dy = 0$
  - Marginalizable: $\int \underline{K}\,(y\,,\,x)\,dy = K\,(x)$

- Nadaraya y Watson (1964)

$$\hat{m}(x) = \frac{\sum K\left(\frac{x - x_i}{h}\right) y_i}{\sum K\left(\frac{x - x_i}{h}\right)} = \sum w(x_i) y_i$$

3. Estimación de funciones

3. 2. Nadaraya-Watson: propiedades asintóticas

- Si K satiface
  - $\int |K(u)|\, du < \infty$
  - $\lim u\, K(u) = 0$, cuando $|u| \to \infty$
- m(x), f(x), σ^2 (x) continuas en x
- $f(x) > 0$

entonces

$$\hat{m}(x) \xrightarrow{p} m(x) \quad \text{si} \quad \begin{array}{l} h \longrightarrow 0 \\ nh \xrightarrow{n \longrightarrow \infty} \infty \end{array}$$

3. Estimación de funciones

3. 2. Nadaraya-Watson: ECM

$$\text{ECM} = \int E\left(m(x) - \hat{m}(x)\right)^2 dy$$

$$= \int (\text{sesgo})^2\, dy + \int \text{varianza}\, dy$$

$$= h^4\, B^2(x) + n^{-1}\, h^{-1}\, V(x)$$

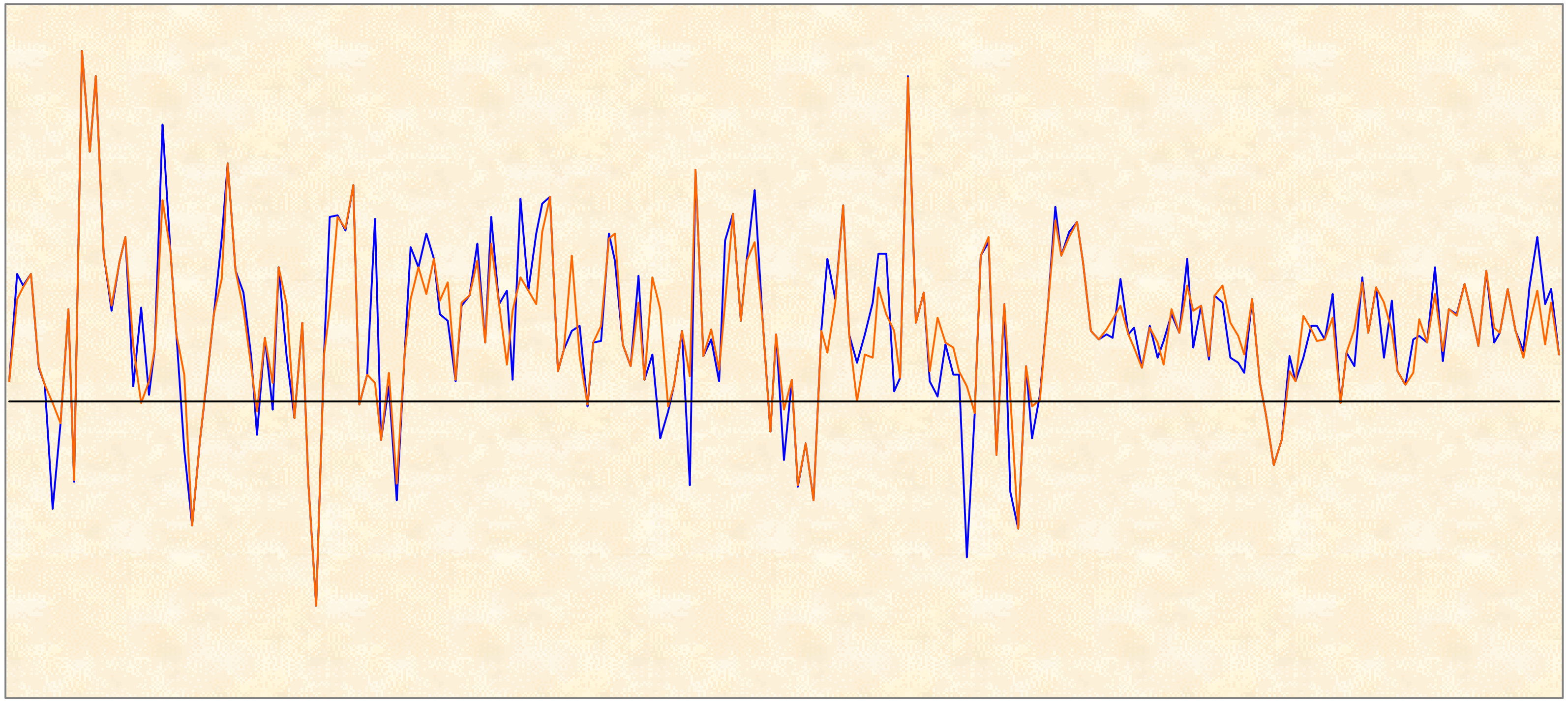
donde $B^2(x)$ y $V(x)$ dependen de:

m, f y var(x)

3. Estimación de funciones

3. 2. Nadaraya-Watson: papel de h

Serie: tasa crecimiento US GDP 1947.2-2000.3



h = 0.01

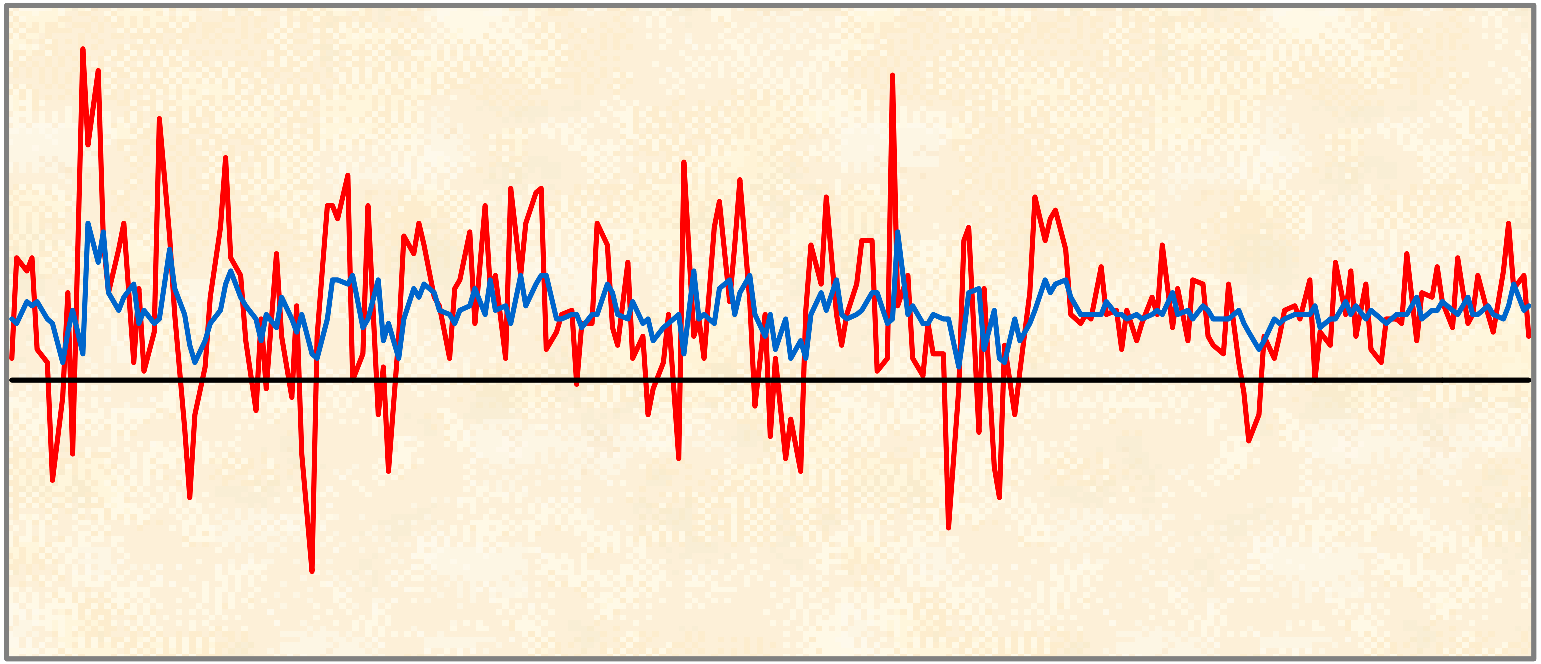
Serie real

Serie estimada

3. Estimación de funciones

3. 2. Nadaraya-Watson: papel de h

Serie: tasa crecimiento US GDP 1947.2-2000.3



h = 2 — Serie real — Serie estimada

3. Estimación de funciones

3. 2. Nadaraya-Watson: elección de h

- Härdle y Vieu (1990)
- Para series temporales: validación cruzada
- h que minimiza

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{m}_i(x_i)\right)^2$$

con $$\hat{m}_i(x_i) = \frac{1}{n-1}\sum_{i \neq j} y_j w(x_j)$$

3. Estimación de funciones

3. 3. Estimación vecino más cercano

$$\hat{m}(x) = n^{-1} \sum w_{ki}(x) y_i$$

donde $w_{ki}(x) = n / k$ si x_i es una de las k observaciones más cercanas a x (0 en otro caso)

- Efectos de k
  - $k >= n \Rightarrow$ serie estimada = media de y
  - $k = 1 \Rightarrow$ serie estimada = serie real

3. Estimación de funciones

3. 3. Estimación vecino más cercano

- Ejemplo. Queremos m(4) con k = 3 en:

x	1	2	3	4	5
y	2	4	5	5	6
w	0	0	1/3	1/3	1/3

- m(4) = (5+5+6)/3

3. Estimación funciones

3.4. Datos multivariantes

- $\uparrow$ variables explicativas $\Rightarrow$ $\uparrow\uparrow$ n
- Reducción de la dimensión d $\rightarrow$ d' $\leq$ 5
  - Proyection Pursuit
    (eliminar las que más disminuyan la varianza)
  - Análisis de componentes informativos
    (eliminar las que menos aumentan el sesgo)

3. Estimación funciones

3.5. Predicción

- Ver comentarios de predicción en STAR
- Sólo a título orientativo, supongamos
  - $y_1, y_2, \ldots, y_n$
  - $y_t = m(y_{t-1}) + e_t$
- Para predecir:
  - Estimamos m con información hasta n
  - Predecimos : $\hat{y}_{n+1} = \hat{m}(y_n)$

$$\hat{y}_{n+2} = \hat{m}(\hat{y}_{n+1})$$

$$\hat{y}_{n+3} = \hat{m}(\hat{y}_{n+2})\ldots$$