

# Ciencias de Datos con R: Fundamentos Estadísticos

Ana M. Bianco, Jemina García y Mariela Sued.

Estimación No Paramétrica de la Densidad

# Enfoque Paramétrico

- $X$  v.a. continua con densidad  $f(x)$ : queremos estimar  $f(x)$
- Muestra Aleatoria:  $X_1, \dots, X_n$ , i.i.d.,  $X_i \sim X$  donde  $X_i \sim X$ .
- Familia paramétrica: asumimos que  $f$  pertenece a una familia determinada y que sólo desconocemos sus parámetros

$$f \in \mathcal{M} = \{f(\cdot, \theta), \theta \in \Theta\}.$$

# Enfoque Paramétrico

- $X$  v.a. continua con densidad  $f(x)$ : queremos estimar  $f(x)$
- Muestra Aleatoria:  $X_1, \dots, X_n$ , i.i.d.,  $X_i \sim X$  donde  $X_i \sim X$ .
- Familia paramétrica: asumimos que  $f$  pertenece a una familia determinada y que sólo desconocemos sus parámetros

$$f \in \mathcal{M} = \{f(\cdot, \theta), \theta \in \Theta\}.$$

- Plug-in:
  - 1)  $\hat{\theta}_n$  estimador de  $\theta$
  - 2) En particular,  $\hat{\theta}_n$  EMV de  $\theta$

# Enfoque Paramétrico

- $X$  v.a. continua con densidad  $f(x)$ : queremos estimar  $f(x)$
- Muestra Aleatoria:  $X_1, \dots, X_n$ , i.i.d.,  $X_i \sim X$  donde  $X_i \sim X$ .
- Familia paramétrica: asumimos que  $f$  pertenece a una familia determinada y que sólo desconocemos sus parámetros

$$f \in \mathcal{M} = \{f(\cdot, \theta), \theta \in \Theta\}.$$

- Plug-in:

1)  $\hat{\theta}_n$  estimador de  $\theta$

$$\Rightarrow \hat{f} = f_{\hat{\theta}}(x)$$

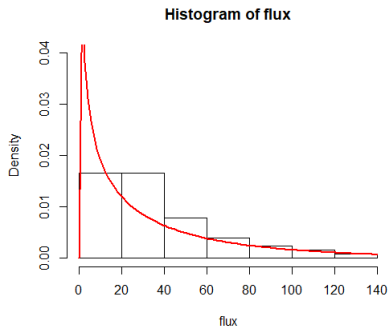
2) En particular,  $\hat{\theta}_n$  EMV de  $\theta$

- Así , por ejemplo:

- $X \sim \mathcal{E}(\lambda)$ ,  $\hat{f}(x) = f_{\hat{\lambda}}(x)$ .
- $X \sim N(\mu, \sigma^2)$ ,  $\hat{f}(x) = f_{\hat{\mu}, \hat{\sigma}^2}(x)$ .

## Ejemplo: Datos de Flux

```
hist(flux, freq=FALSE, ylim=c(0,0.04))  
curve(dgamma(x, shape=alpha.MV, rate = lambda.MV), add=TRUE,  
col="red", lwd=2, main="Histograma de Flux")
```



# Enfoque

$X$  v.a. continua con densidad  $f(x)$

**Paramétrico:**  $X \sim F_\theta$

$$\hat{F}_\theta = \hat{F}_{\hat{\theta}}$$

$$\hat{f}(x) = f_{\hat{\theta}}(x)$$

# Enfoque

$X$  v.a. continua con densidad  $f(x)$

**Paramétrico:**  $X \sim F_\theta$

$$\hat{F}_\theta = \hat{F}_{\hat{\theta}}$$

$$\hat{f}(x) = f_{\hat{\theta}}(x)$$

**No Paramétrico:**  $X \sim F$

$$\hat{F}_n = \text{"la empírica"}$$

$$\hat{f}(x) = ?$$

# Enfoque No Paramétrico

- $X$  con densidad  $f(x)$ : queremos estimar  $f(x)$
- $X_1, \dots, X_n$ , i.i.d.,  $X_i \sim X$  donde  $X_i \sim X$ .
- Queremos estimar  $f$  sin asumir una determinada forma: sólo asumimos que es  $f$  es suave.



# Enfoque No Paramétrico

- $X$  con densidad  $f(x)$ : queremos estimar  $f(x)$
- $X_1, \dots, X_n$ , i.i.d.,  $X_i \sim X$  donde  $X_i \sim X$ .
- Queremos estimar  $f$  sin asumir una determinada forma: sólo asumimos que es  $f$  es suave.
- La forma más sencilla: **Histograma**

# Histograma

$X_1, \dots, X_n$ , i.i.d.,  $X_i \sim X$  donde  $X_i \sim X$

- Sea  $\mathcal{A}_j$  una partición de intervalos o clases acotados (bins) disjuntos tales que:

$$\mathbb{R} = \cup_j \mathcal{A}_j$$

- Para cada  $x \in \mathcal{A}_j$

$$\hat{f}(x) = \frac{\#\{X_i : X_i \in \mathcal{A}_j\}}{n|\mathcal{A}_j|}$$

con  $|\mathcal{A}_j|$  ancho del bin  $\mathcal{A}_j$

# Histograma

$X_1, \dots, X_n$ , i.i.d.,  $X_i \sim X$  donde  $X_i \sim X$

- Sea  $\mathcal{A}_j$  una partición de intervalos o clases acotados (bins) disjuntos tales que:

$$\mathbb{R} = \cup_j \mathcal{A}_j$$

- Para cada  $x \in \mathcal{A}_j$

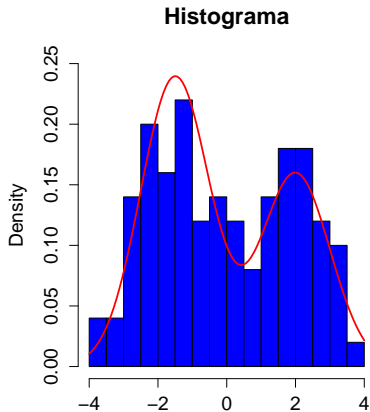
$$\hat{f}(x) = \frac{\#\{X_i : X_i \in \mathcal{A}_j\}}{n|\mathcal{A}_j|}$$

con  $|\mathcal{A}_j|$  ancho del bin  $\mathcal{A}_j$

- El histograma requiere dos parámetros:
  - i) ancho del bin
  - ii) punto inicial del primer bin

Vamos a las tareas de Clase: items 1 a 3.

## Ejemplo: datos simulados



# Desventajas del histograma

- el estimador de la densidad depende del punto inicial de los bins: para un número de bins fijo, la forma puede cambiar moviendo la ubicación de los bins
- la densidad estimada no es suave, es *escalonada* y esto no es propio de la densidad sino de la herramienta de estimación
- por estas razones, el histograma es usado sólo para visualización

# Busquemos otra idea...

$X_1, \dots, X_n$ , i.i.d.,  $X_i \sim X$  donde  $X_i \sim X$

- $X$  con densidad  $f(x)$ : queremos estimar  $f(x)$
- Queremos estimar  $f$  sin asumir una determinada forma: sólo asumimos que  $f$  es suave.

# Busquemos otra idea...

$X_1, \dots, X_n$ , i.i.d.,  $X_i \sim X$  donde  $X_i \sim X$

- $X$  con densidad  $f(x)$ : queremos estimar  $f(x)$
- Queremos estimar  $f$  sin asumir una determinada forma: sólo asumimos que es  $f$  es suave.
- **Idea frecuentista**: por la LGN

$$\mathbb{P}(X \in (x - h, x + h)) \approx \frac{\#\{X_i \in (x - h, x + h)\}}{n}$$



# Busquemos otra idea...

$X_1, \dots, X_n$ , i.i.d.,  $X_i \sim X$  donde  $X_i \sim X$

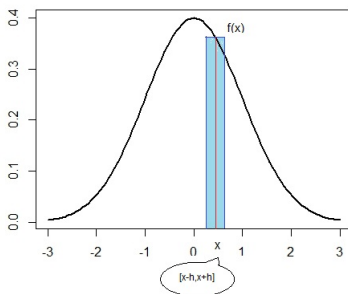
- $X$  con densidad  $f(x)$ : queremos estimar  $f(x)$
- Queremos estimar  $f$  sin asumir una determinada forma: sólo asumimos que es  $f$  es suave.
- **Idea frecuentista**: por la LGN

$$\mathbb{P}(X \in (x - h, x + h)) \approx \frac{\#\{X_i \in (x - h, x + h)\}}{n}$$

$$\mathbb{P}(X \in (x - h, x + h)) = \int_{x-h}^{x+h} f(t) dt$$

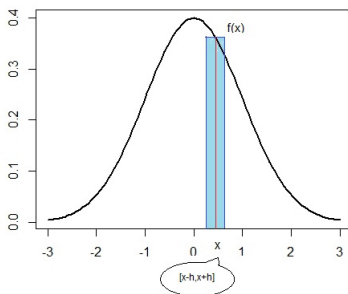
## Aproximando analíticamente...

- $\mathbb{P}(X \in (x-h, x+h)) = \int_{x-h}^{x+h} f(t) dt$
- Si  $h$  es pequeño y  $f$  continua en  $x$ ,



## Aproximando analíticamente...

- $\mathbb{P}(X \in (x-h, x+h)) = \int_{x-h}^{x+h} f(t) dt$
- Si  $h$  es pequeño y  $f$  continua en  $x$ ,



$$\int_{x-h}^{x+h} f(t) dt \approx 2hf(x)$$

## Juntando todo...

$X_1, \dots, X_n$ , i.i.d.,  $X_i \sim X$  donde  $X_i \sim X$

- $\mathbb{P}(X \in (x - h, x + h)) \approx \frac{\#\{X_i \in (x - h, x + h)\}}{n}$  por la LGN

- $\mathbb{P}(X \in (x - h, x + h)) = \int_{x-h}^{x+h} f(t) dt$

# Juntando todo...

$X_1, \dots, X_n$ , i.i.d.,  $X_i \sim X$  donde  $X_i \sim X$

- $\mathbb{P}(X \in (x - h, x + h)) \approx \frac{\#\{X_i \in (x - h, x + h)\}}{n}$  por la LGN
- $\mathbb{P}(X \in (x - h, x + h)) = \int_{x-h}^{x+h} f(t) dt$
- Si  $h$  es pequeño y  $f$  continua en  $x$ ,

$$\mathbb{P}(X \in (x - h, x + h)) \approx 2h f(x)$$

## Juntando todo...

$X_1, \dots, X_n$ , i.i.d.,  $X_i \sim X$  donde  $X_i \sim X$

- $\mathbb{P}(X \in (x - h, x + h)) \approx \frac{\#\{X_i \in (x - h, x + h)\}}{n}$  por la LGN

- $\mathbb{P}(X \in (x - h, x + h)) = \int_{x-h}^{x+h} f(t) dt$

- Si  $h$  es pequeño y  $f$  continua en  $x$ ,

$$\mathbb{P}(X \in (x - h, x + h)) \approx 2h f(x)$$

- Entonces, podemos aproximar analíticamente

$$2h f(x) \approx \mathbb{P}(X \in (x - h, x + h)) \approx \frac{\#\{X_i \in (x - h, x + h)\}}{n}$$

# Juntando todo...

$X_1, \dots, X_n$ , i.i.d.,  $X_i \sim X$  donde  $X \sim f$

- $\mathbb{P}(X \in (x - h, x + h)) \approx \frac{\#\{X_i \in (x - h, x + h)\}}{n}$  por la LGN

- $\mathbb{P}(X \in (x - h, x + h)) = \int_{x-h}^{x+h} f(t) dt$

- Si  $h$  es pequeño y  $f$  continua en  $x$ ,

$$\mathbb{P}(X \in (x - h, x + h)) \approx 2h f(x)$$

- Entonces, podemos aproximar analíticamente

$$2h f(x) \approx \mathbb{P}(X \in (x - h, x + h)) \approx \frac{\#\{X_i \in (x - h, x + h)\}}{n}$$

$$f(x) \approx \frac{\#\{X_i \in (x - h, x + h)\}}{2h n}$$

# Propuesta

$X_1, \dots, X_n$ , i.i.d.,  $X_i \sim X$  donde  $X_i \sim X$

$$\hat{f}(x) = \frac{\#\{X_i \in (x - h, x + h)\}}{2h n}$$



# Propuesta

$X_1, \dots, X_n$ , i.i.d.,  $X_i \sim X$  donde  $X_i \sim X$

$$\hat{f}(x) = \frac{\#\{X_i \in (x-h, x+h)\}}{2h n}$$

Notemos que

- $\hat{f}(x) \geq 0$

# Propuesta

$X_1, \dots, X_n$ , i.i.d.,  $X_i \sim X$  donde  $X_i \sim X$

$$\hat{f}(x) = \frac{\#\{X_i \in (x-h, x+h)\}}{2h n}$$

Notemos que

- $\hat{f}(x) \geq 0$
- $\int \hat{f}(x) dx = 1$

# Propuesta

$X_1, \dots, X_n$ , i.i.d.,  $X_i \sim X$  donde  $X_i \sim X$

$$\hat{f}(x) = \frac{\#\{X_i \in (x-h, x+h)\}}{2h n}$$

Notemos que

- $\hat{f}(x) \geq 0$
- $\int \hat{f}(x) dx = 1$

# Propuesta

$X_1, \dots, X_n$ , i.i.d.,  $X_i \sim X$  donde  $X_i \sim X$

$$\hat{f}(x) = \frac{\#\{X_i \in (x-h, x+h)\}}{2h n}$$

Notemos que

- $\hat{f}(x) \geq 0$
- $\int \hat{f}(x) dx = 1$

$$\hat{f}(x) = \frac{1}{2h n} \sum_{i=1}^n \mathcal{I}_{(x-h, x+h)}(X_i)$$

- Estimador de Parzen

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n \frac{1}{2} \mathcal{I}_{[-1,1]} \left( \frac{x - X_i}{h} \right)$$

# Propuesta

$X_1, \dots, X_n$ , i.i.d.,  $X_i \sim X$  donde  $X_i \sim X$

$$\hat{f}(x) = \frac{\#\{X_i \in (x-h, x+h)\}}{2h n}$$

Notemos que

- $\hat{f}(x) \geq 0$
- $\int \hat{f}(x) dx = 1$

$$\hat{f}(x) = \frac{1}{2h n} \sum_{i=1}^n \mathcal{I}_{(x-h, x+h)}(X_i)$$

- Estimador de Parzen

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n \frac{1}{2} \mathcal{I}_{[-1,1]} \left( \frac{x - X_i}{h} \right)$$

- si  $K(t) = \frac{1}{2} \mathcal{I}_{[-1,1]}(t) \Rightarrow$

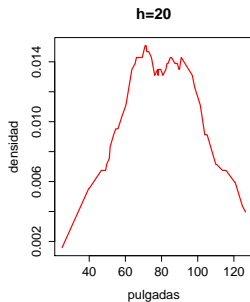
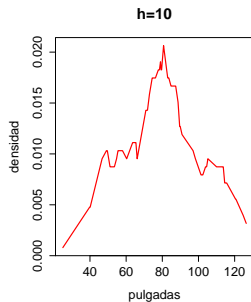
$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right)$$

## Juntando todo...

- $K(t) = \frac{1}{2}\mathcal{I}_{[-1,1]}(t) \Rightarrow \hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)$

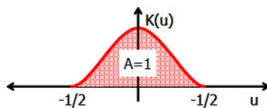
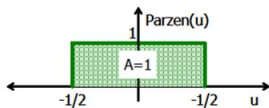
- $K$  : núcleo

- $h$  : ventana



Vayamos a terminar las tareas de Clase: ítems 4 a 7.

# Núcleos

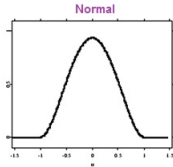
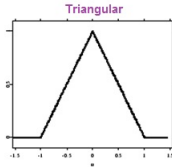
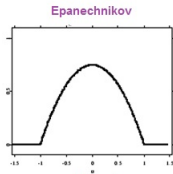
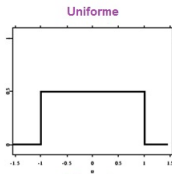




# Tipos de núcleos

- Núcleo Rectangular:  $K(t) = \frac{1}{2}\mathcal{I}_{[-1,1]}(t)$
- Núcleo Triangular:  $K(t) = (1 - |t|)\mathcal{I}_{[-1,1]}(t)$
- Núcleo Gaussiano:  $K(t) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}$
- Núcleo Epanechnikov:  $K(t) = \frac{3}{4}(1 - t^2)\mathcal{I}_{[-1,1]}(t)$

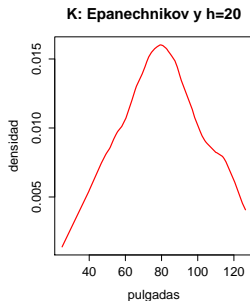
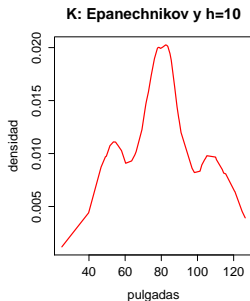
# Núcleos



# Estimadores de núcleos (Rosenblatt-Parzen)

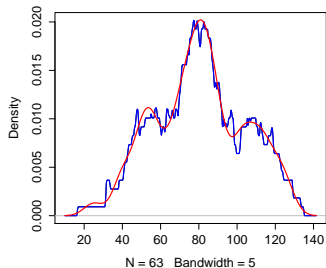
$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)$$

- $K$  núcleo: \*  $K \geq 0$  y \*  $\int K(x)dx = 1$ .
- $h$ : ventana o parámetro de suavizado
- Notemos que  $\hat{f}(x)$  depende de  $n$ , del núcleo  $K$  y de  $h$



# Comandos de R

```
nieve=scan()  
126.4 82.4 78.1 51.1 90.9 76.2 104.5 ...  
  
density(nieve, from=40, to=40, n=1, kernel="rectangular", bw=5)$y  
[1] 0.003665716  
  
pp.rec=density(nieve, kernel="rectangular", bw=5)  
pp.nor=density(nieve, kernel="gaussian", bw=5)  
  
plot(pp.rec)  
lines(pp.rec$x, pp.rec$y, type="l", col="black", lwd=2)  
lines(pp.nor$x, pp.nor$y, type="l", col="red", lwd=2)
```



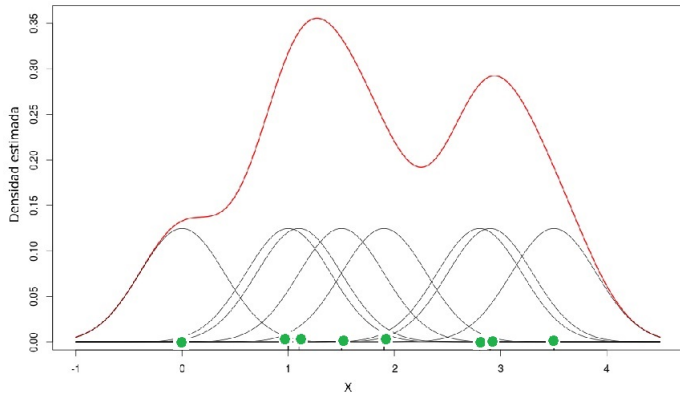
# Estimadores de núcleos: Selección de ventana

**En el gráfico anterior se muestra que la elección de la ventana es crucial.**

- Una ventana  $h$  pequeña dará un estimador muy rugoso, con muchos picos y difícil de interpretar
- una ventana  $h$  grande sobresuaviza al estimador de la densidad y enmascara estructuras de los datos.

# Interpretación del estimador de núcleos

Fuente: Tesis de Lic. en Cs. Matem. de Sofía Ruiz, 2016.



## Sesgo y Varianza de $\hat{f}(x)$

Veamos en el pizarrón en un caso fácil que para que todo funcione bien deben pasar dos cosas:

- $h \longrightarrow 0$
- $nh \longrightarrow \infty$

## Sesgo y Varianza de $\hat{f}(x)$

Veamos en el pizarrón en un caso fácil que para que todo funcione bien deben pasar dos cosas:

- $h \longrightarrow 0$
- $nh \longrightarrow \infty$
- Se puede probar que bajo condiciones generales del núcleo  $K$

$$\begin{aligned}\text{Sesgo}[\hat{f}(x)] &\approx \frac{h^2}{2} C_1(K) f''(x) \\ \mathbb{V}[\hat{f}(x)] &\approx \frac{1}{nh} C_2(K) f(x)\end{aligned}$$



## Sesgo y Varianza de $\hat{f}(x)$

Veamos en el pizarrón en un caso fácil que para que todo funcione bien deben pasar dos cosas:

- $h \longrightarrow 0$
- $nh \longrightarrow \infty$
- Se puede probar que bajo condiciones generales del núcleo  $K$

$$\begin{aligned}\text{Sesgo}[\hat{f}(x)] &\approx \frac{h^2}{2} C_1(K) f''(x) \\ \mathbb{V}[\hat{f}(x)] &\approx \frac{1}{nh} C_2(K) f(x)\end{aligned}$$

- El sesgo es proporcional a  $h^2 \Rightarrow$  elijamos  $h$  pequeña

## Sesgo y Varianza de $\hat{f}(x)$

Veamos en el pizarrón en un caso fácil que para que todo funcione bien deben pasar dos cosas:

- $h \longrightarrow 0$
- $nh \longrightarrow \infty$
- Se puede probar que bajo condiciones generales del núcleo  $K$

$$\begin{aligned}\text{Sesgo}[\hat{f}(x)] &\approx \frac{h^2}{2} C_1(K) f''(x) \\ \mathbb{V}[\hat{f}(x)] &\approx \frac{1}{nh} C_2(K) f(x)\end{aligned}$$

- El sesgo es proporcional a  $h^2 \Rightarrow$  elijamos  $h$  pequeña
- El sesgo depende de  $f''(x)$  que mide la curvatura de  $f$  en  $x$

## Sesgo y Varianza de $\hat{f}(x)$

Veamos en el pizarrón en un caso fácil que para que todo funcione bien deben pasar dos cosas:

- $h \longrightarrow 0$
- $nh \longrightarrow \infty$
- Se puede probar que bajo condiciones generales del núcleo  $K$

$$\begin{aligned}\text{Sesgo}[\hat{f}(x)] &\approx \frac{h^2}{2} C_1(K) f''(x) \\ \mathbb{V}[\hat{f}(x)] &\approx \frac{1}{nh} C_2(K) f(x)\end{aligned}$$

- La varianza disminuye a medida que  $nh$  crece

## Sesgo y Varianza de $\hat{f}(x)$

Veamos en el pizarrón en un caso fácil que para que todo funcione bien deben pasar dos cosas:

- $h \longrightarrow 0$
- $nh \longrightarrow \infty$
- Se puede probar que bajo condiciones generales del núcleo  $K$

$$\begin{aligned}\text{Sesgo}[\hat{f}(x)] &\approx \frac{h^2}{2} C_1(K) f''(x) \\ \mathbb{V}[\hat{f}(x)] &\approx \frac{1}{nh} C_2(K) f(x)\end{aligned}$$

- La varianza disminuye a medida que  $nh$  crece
- Para disminuir la varianza necesitamos  $h$  o  $n$  grandes.

# Error Cuadrático Medio de $\hat{f}(x)$

## Compromiso Sesgo-Varianza

Tenemos que

$$\text{ECM}[\hat{f}(x)] = \text{Sesgo}^2[\hat{f}(x)] + \mathbb{V}[\hat{f}(x)]$$

Por otro lado:

$$\text{Sesgo}[\hat{f}(x)] \approx \frac{h^2}{2} C_1(K) f''(x)$$

$$\mathbb{V}[\hat{f}(x)] \approx \frac{1}{nh} C_2(K) f(x)$$

# Error Cuadrático Medio de $\hat{f}(x)$

## Compromiso Sesgo-Varianza

Tenemos que

$$\text{ECM}[\hat{f}(x)] = \text{Sesgo}^2[\hat{f}(x)] + \mathbb{V}[\hat{f}(x)]$$

Por otro lado:

$$\text{Sesgo}[\hat{f}(x)] \approx \frac{h^2}{2} C_1(K) f''(x)$$

$$\mathbb{V}[\hat{f}(x)] \approx \frac{1}{nh} C_2(K) f(x)$$

$$\Rightarrow \text{ECM}[\hat{f}(x)] \approx \frac{h^4}{4} C_1^2(K) (f''(x))^2 + \frac{1}{nh} C_2(K) f(x)$$

Si  $h \rightarrow 0$  y  $nh \rightarrow \infty$   $\hat{f}(x)$  es un estimador consistente de  $f(x)$ .

$$\hat{f}_h(x)$$

Por lo que vimos, el efecto de la ventana sobre la estimación de la densidad puede ser crucial, por lo tanto haremos explícita esta dependencia, denotando

$$\hat{f}_h(x)$$

al estimador basado en una ventana  $h$ .

# Error Cuadrático Medio de $\hat{f}_h(x)$

## Compromiso Sesgo-Varianza

Tenemos que

$$\text{ECM}[\hat{f}_h(x)] = \text{Sesgo}^2[\hat{f}_h(x)] + \mathbb{V}[\hat{f}_h(x)]$$

Por otro lado:

$$\text{Sesgo}[\hat{f}_h(x)] \approx \frac{h^2}{2} C_1(K) f''(x)$$

$$\mathbb{V}[\hat{f}_h(x)] \approx \frac{1}{nh} C_2(K) f(x)$$

$$\Rightarrow \text{ECM}[\hat{f}_h(x)] \approx \frac{h^4}{4} C_1^2(K) (f''(x))^2 + \frac{1}{nh} C_2(K) f(x)$$

Si  $h \rightarrow 0$  y  $nh \rightarrow \infty$   $\hat{f}_h(x)$  es un estimador consistente de  $f(x)$ .



# Error Cuadrático Medio Integrado de $\hat{f}_h$

Como medida global:

$$\text{ECMI}[\hat{f}_h] = \int \text{ECM}[\hat{f}_h(x)] dx$$

Integrando resulta:

$$\text{ECMI}[\hat{f}_h(x)] \approx \frac{h^4}{4} C_1^2(K) \int (f''(x))^2 dx + \frac{1}{nh} C_2(K)$$

# Error Cuadrático Medio Integrado de $\hat{f}_h$

Como medida global:

$$\text{ECMI}[\hat{f}_h] = \int \text{ECM}[\hat{f}_h(x)] dx$$

Integrando resulta:

$$\text{ECMI}[\hat{f}_h(x)] \approx \frac{h^4}{4} C_1^2(K) \int (f''(x))^2 dx + \frac{1}{nh} C_2(K)$$

Si llamamos  $\|g\|_2^2 = \int g(t)^2 dt$

$$\text{ECMI}[\hat{f}_h(x)] \approx \frac{h^4}{4} C_1^2(K) \|f''\|_2^2 + \frac{1}{nh} \|K\|_2^2$$

## Error Cuadrático Medio Integrado de $\hat{f}_h$

$$\text{ECMI}[\hat{f}_h(x)] \approx \frac{h^4}{4} C_1^2(K) \|f''\|_2^2 + \frac{1}{nh} \|K\|_2^2$$

- Notemos que ECMI es una función de la ventana  $h$ .
- Derivando a ECMI respecto de  $h$  e igualando a 0, obtenemos que el valor de  $h$  que minimiza el ECMI es

$$h = \left\{ \frac{\|K\|_2^2}{C_1^2(K) \|f''\|_2^2 n} \right\}^{\frac{1}{5}}$$

## Error Cuadrático Medio Integrado de $\hat{f}_h$

$$\text{ECMI}[\hat{f}_h(x)] \approx \frac{h^4}{4} C_1^2(K) \|f''\|_2^2 + \frac{1}{nh} \|K\|_2^2$$

- Notemos que ECMI es una función de la ventana  $h$ .
- Derivando a ECMI respecto de  $h$  e igualando a 0, obtenemos que el valor de  $h$  que minimiza el ECMI es

$$h = \left\{ \frac{\|K\|_2^2}{C_1^2(K) \|f''\|_2^2 n} \right\}^{\frac{1}{5}} \sim n^{-\frac{1}{5}}$$

## Selección de $h$ : Regla de Silverman

El  $\text{ECMI}[\hat{f}_h(x)]$  se minimiza en

$$h = \left\{ \frac{\|K\|_2^2}{C_1^2(K) \|f''\|_2^2 n} \right\}^{\frac{1}{5}}$$

- Silverman propone reemplazar a  $\|f''\|_2^2$  por su valor cuando  $f$  es normal:

$$\|f''\|_2^2 = \sigma^{-5} \frac{3}{8\sqrt{\pi}} \approx 0.212 \sigma^{-5}$$

y a  $\sigma$  por un estimador  $\hat{\sigma}$

# Selección de $h$ : Regla de Silverman

El  $\text{ECMI}[\hat{f}_h(x)]$  se minimiza en

$$h = \left\{ \frac{\|K\|_2^2}{C_1^2(K) \|f''\|_2^2 n} \right\}^{\frac{1}{5}}$$

- Silverman propone reemplazar a  $\|f''\|_2^2$  por su valor cuando  $f$  es normal:

$$\|f''\|_2^2 = \sigma^{-5} \frac{3}{8\sqrt{\pi}} \approx 0.212 \sigma^{-5}$$

y a  $\sigma$  por un estimador  $\hat{\sigma}$

$$h_{Sil} = \left( \frac{4\hat{\sigma}^5}{3n} \right)^{\frac{1}{5}} \approx 1.06 \hat{\sigma} n^{-\frac{1}{5}}$$

- Si  $f$  es normal, la ventana  $h_{Sil}$  es óptima.
- Si  $f$  no es normal,  $h_{Sil}$  dará una ventana no muy alejada de la óptima cuando la distribución no es muy diferente a la normal.

## Selección de $h$ : Regla de Silverman

$$h_{Sil} = \left( \frac{4\hat{\sigma}^5}{3n} \right)^{\frac{1}{5}} \approx 1.06 \hat{\sigma} n^{-\frac{1}{5}}$$

- $\sigma$  puede estimarse por  $S$  (sd(datos) en R).  
o
- $\sigma$  puede estimarse por la distancia intercuartil  $IQR$  (IQR(datos) en R). Para que coincida con  $\sigma$  bajo la distribución normal debe dividirse por 1.349
- La ventana óptima de acuerdo a la regla de Silverman resulta:

$$h_{Sil} = 1.06 \min\left(S, \frac{IQR}{1.349}\right) n^{-\frac{1}{5}}$$

# Convalidación Cruzada por Máxima Verosimilitud (CV)

- El EMV de  $h$  es degenerado: da  $h_{MV} = 0$ , resultando una densidad que da masa 1 a cada uno de los datos.
- Alternativa: maximizar una *pseudo-verosimilitud* computada sacando de los datos una observación a la vez (leave-one-out cross-validation).

Si observamos los datos  $x_1, \dots, x_n$

$$h_{MV}^* = \operatorname{argmax}_h \left\{ \frac{1}{n} \sum_{i=1}^n \log \hat{f}_h^{(-i)}(x_i) \right\}$$

siendo  $\hat{f}_h^{(-i)}(x_i)$  la densidad estimada en el punto  $x_i$  sin utilizar al punto  $x_i$ :

$$\hat{f}_h^{(-i)}(x_i) = \frac{1}{(n-1)h} \sum_{j \neq i} K\left(\frac{x_i - x_j}{h}\right)$$

- En general, buscamos el máximo sobre una grilla  $h_1, \dots, h_q$  y luego, eventualmente, se refina.