

Tema 3. Estimación de la función de densidad

Parte 2

Rosa M. Crujeiras
Alberto Rodríguez



Dpto. de Estadística e Investigación Operativa
Máster en Técnicas Estadísticas
Curso 2009-2010

Ejercicio. Para el estimador Naive que has programado, vamos a programar una aproximación del MISE de la siguiente forma:

$$MISE(h) = \mathbb{E} \int (\hat{f}_{n,N}(x) - f(x))^2 dx$$

- Consideramos una secuencia de ventanas H entre 0.2 y 1 , con paso 0.05 .
- Generamos una muestra de una $N(0, 1)$ de tamaño $n = 100$ para construir el Naive.
- Evaluamos el Naive, para cada $H[i]$ (y la densidad Normal) en una secuencia de puntos x entre -3 y 3 , con espaciado 0.01 .
- Aproximamos $\int (\hat{f}_{n,N}(x) - f(x))^2 dx$ como suma de las diferencias al cuadrado.
- Repetimos el procedimiento anterior para $M = 100$ muestras.
- Para cada $H[i]$ promediamos el MISE en las M muestras.
- Representamos $MISE(h) \sim h$. ¿Dónde se encuentra el

El estimador Naive viene dado por la expresión

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

que se puede reescribir como

$$\hat{f}_{n,N}(x) = \frac{1}{nh} \sum_{i=1}^n \frac{1}{2} \mathbb{I}(x - h < X_i < x + h) = \frac{1}{nh} \sum_{i=1}^n \omega \left(\frac{x - X_i}{h} \right)$$

donde, ω es la densidad de la distribución uniforme en $(-1, 1)$.

Por tanto la aportación de cada observación X_i al Estimador Naive en el punto x viene determinado por el valor de

$$\omega\left(\frac{x - X_i}{h}\right),$$

que vale $1/2$ para todos los puntos en $(x - h, x + h)$ independientemente de su proximidad a x .

Si se reemplaza ω por una densidad K (denominada núcleo) con una única moda en cero y simétrica se obtiene el estimador tipo núcleo

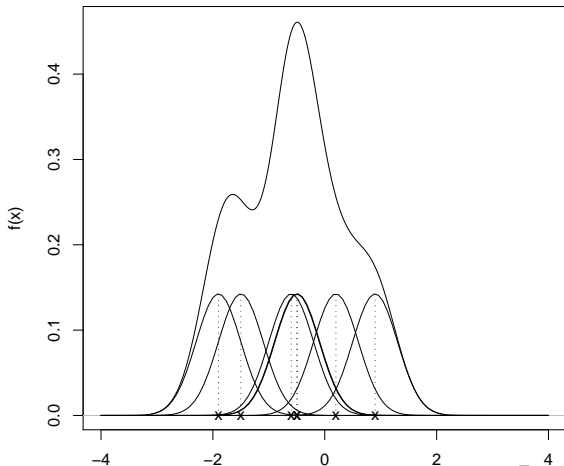
$$\hat{f}_{n,K}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i),$$

donde

$$K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right)$$

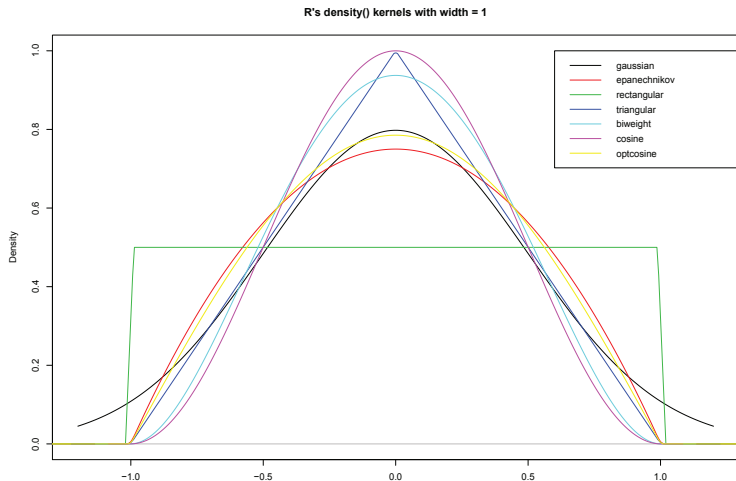
es la densidad de hX_K siendo X_K una variable con función de densidad K .

¿Cómo se interpreta el estimador tipo núcleo?

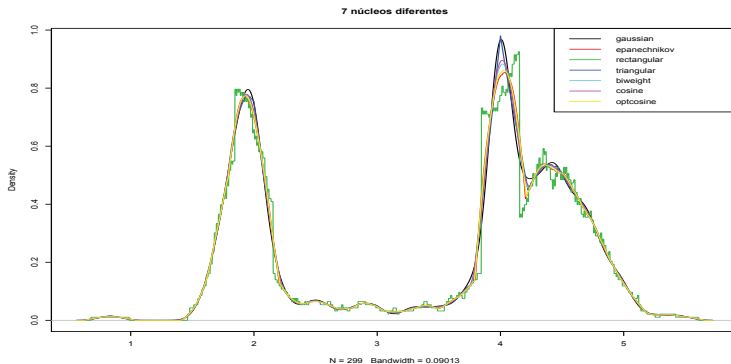


A continuación se muestran algunos núcleos utilizados en la práctica:

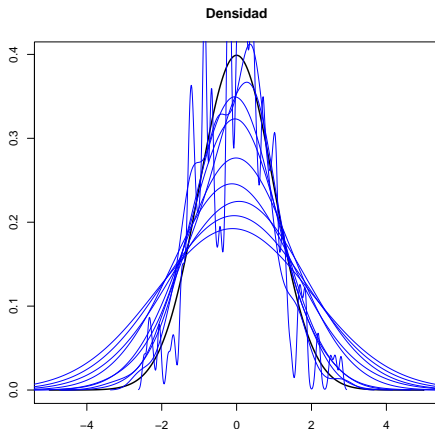
- ▶ Uniforme: $1/2\mathbb{I}(|x| < 1)$.
- ▶ Triangular: $(1 - |x|)\mathbb{I}(|x| < 1)$.
- ▶ Epanechnikov: $3/4(1 - x^2)\mathbb{I}(|x| < 1)$.
- ▶ Biweight: $15/16(1 - x^2)^2\mathbb{I}(|x| < 1)$.
- ▶ Triweight: $35/32(1 - x^2)^3\mathbb{I}(|x| < 1)$.
- ▶ Gaussiano: $(2\pi)^{-1/2} \exp(-x^2/2)$.



El estimador tipo núcleo hereda las propiedades de suavidad del núcleo. Sin embargo la forma del núcleo no tiene un papel tan determinante como la ventana h .



Efecto de la ventana en la estimación de una densidad $N(0,1)$, con el estimador tipo núcleo:



Para evaluar $\hat{f}_{n,K}(x)$ como estimador de $f(x)$, utilizamos el Error Cuadrático Medio:

$$MSE(x) = \mathbb{E}(\hat{f}_{n,K}(x) - f(x))^2$$

que sabemos se puede descomponer en sesgo al cuadrado y varianza:

$$MSE(x) = \left(\mathbb{E}(\hat{f}_{n,K}(x)) - f(x) \right)^2 + \text{Var}(\hat{f}_{n,K}(x)).$$

Al igual que derivamos las expresiones del sesgo y la varianza del Naive, podemos hacer lo mismo con las del estimador tipo núcleo, obteniendo:

► Para el sesgo:

$$\mathbb{E}(\hat{f}_{n,K}(x)) = \int K(u)f(x - hu)du$$

Bajo condiciones de regularidad sobre f (Wand y Jones, 1995) podemos obtener la siguiente aproximación:

$$f(x - hu) = f(x) - huf'(x) + \frac{1}{2}(hu)^2 f''(x) + o(h^2)$$

Si suponemos que $\mu_2(K) = \int u^2 K(u) du < \infty$, entonces:

$$\mathbb{E}(\hat{f}_{n,K}(x)) = f(x) + \frac{1}{2}h^2\mu_2(K) + o(h^2)$$

donde hemos utilizado que K es una densidad simétrica, y por tanto:

$$\int K(u) du = 1, \quad \int uK(u) du = 0.$$

¿Qué relación tiene con el sesgo del Naive que obtuvimos antes?

Para la varianza tenemos:

$$\begin{aligned}
 \text{Var}(\hat{f}_{n,K}(x)) &= \frac{1}{n} \text{Var}(K_h(x - X_1)) \\
 &= \frac{1}{n} [\mathbb{E}(K_h^2(x - X_1)) - \mathbb{E}^2(K_h(x - X_1))] \\
 &= \frac{1}{n} \left[\int K_h^2(x - y) f(y) dy - \left(\int K_h(x - y) f(y) dy \right)^2 \right] \\
 &= \frac{1}{nh} \int K^2(u) f(x - hu) du - \frac{1}{n} \left(\int K(u) f(x - hu) du \right)^2 \\
 &= \frac{1}{nh} \int K^2(u) (f(x) + o(1)) du - \frac{1}{n} (f(x) + o(1))^2 \\
 &= \frac{1}{nh} R(K) f(x) + o((nh)^{-1})
 \end{aligned}$$

Utilizando las expresiones anteriores del sesgo y la varianza tenemos:

$$MSE(x) = \frac{1}{nh} R(K) f(x) + \frac{1}{4} h^4 \mu_2(K)^2 \left(f''(x) \right)^2 + o((nh)^{-1} + h^4).$$

El AMSE en un punto x viene dado por:

$$AMSE(x) = \frac{1}{nh} R(K) f(x) + \frac{1}{4} h^4 \mu_2(K)^2 \left(f''(x) \right)^2$$

Integrando el MSE en x tenemos:

$$MISE(h) = \frac{1}{nh}R(K) + \frac{1}{4}h^2\mu_2(K)^2R(f'') + o((nh)^{-1} + h^4)$$

y la versión asintótica:

$$AMISE(h) = \frac{1}{nh}R(K) + \frac{1}{4}h^2\mu_2(K)^2R(f'')$$

Ejercicio (Estimador Kernel)

Prueba que el valor que minimiza el $AMISE$ es:

$$h_{AMISE} = \left(\frac{R(K)}{n\mu_2(K)^2 R(f'')} \right)^{1/5}$$

y por tanto:

$$\inf_{h>0} AMISE(h) = \frac{5}{4} \left(\mu_2(K)^2 R^4(K) R(f'') \right)^{1/5} n^{-4/5}$$

El AMISE del Histograma (Wand y Jones, 1995, p.23) viene dado por:

$$AMISE_H(h) = \frac{1}{nh} + \frac{1}{12}R(f')$$

y el menor error que podemos cometer en muestras grandes, utilizando el Histograma es:

$$\inf_{h>0} AMISE_H(h) = \frac{1}{4}(36R(f'))^{1/3}n^{-2/3}.$$

¿Es mejor o peor que el estimador núcleo?

Ejercicio (Estimador Kernel) Prueba que si f es $N(\mu, \sigma^2)$, entonces:

$$h_{AMISE} = \left(\frac{8\sqrt{\pi}R(K)}{3\mu_2(K)^2n} \right)^{1/5} \sigma$$

Ejercicio (Estimador Kernel) En el caso anterior, prueba que si K es la densidad de la $N(0, 1)$ entonces:

$$h_{AMISE} = \left(\frac{4}{3n} \right)^{1/5} \sigma$$

Ejercicio (Estimador Kernel) Calcula h_{AMISE} con f densidad de una $N(\mu, \sigma^2)$ y K el núcleo de Epanechnikov.

Ventana Normal

Si f es una densidad Normal con desviación típica σ , entonces:

$$h_{AMISE} = \left(\frac{8\sqrt{\pi}R(K)}{3\mu_2(K)^2n} \right)^{1/5} \sigma$$

La ventana Normal se obtiene reemplazando σ por una estimación $\hat{\sigma}$:

$$\hat{h}_{NS} = \left(\frac{8\sqrt{\pi}R(K)}{3\mu_2(K)^2n} \right)^{1/5} \hat{\sigma}$$

Por ejemplo $\hat{\sigma} = S$, cuasi-desviación típica.

Se puede utilizar un estimador más robusto de la desviación, que viene dado por el rango intercuartílico estandarizado:

$$\hat{\sigma}_{IQR} = \frac{IQR}{\Phi^{-1}(0.75) - \Phi^{-1}(0.25)},$$

donde Φ^{-1} es la función cuantil de la $N(0, 1)$. Para reducir el riesgo de sobresuavizar, Silverman (1986, p.47) recomienda utilizar el mínimo entre $\hat{\sigma}_{IQR}$ y S .

Validación cruzada

El Error Cuadrático Integrado (ISE) se define como:

$$ISE(h) = \int \left(\hat{f}_{n,K}(x) - f(x) \right)^2 dx$$

que se puede escribir como:

$$ISE(h) = \int \hat{f}_{n,K}^2(x) dx - 2 \int \hat{f}_{n,K}(x) f(x) dx + \int f^2(x) dx$$

Podemos buscar la ventana que haga el ISE lo más pequeño posible.

A partir de la expresión anterior, bastará con minimizar:

$$CV(h) = \int \hat{f}_{n,K}^2(x) dx - 2 \int \hat{f}_{n,K}(x) f(x) dx$$

El primer término de CV es fácil de calcular a partir de la muestra:

$$\int \hat{f}_{n,K}^2(x) dx = \frac{1}{n^2 h} \sum_{i,j} K * K \left(\frac{X_i - X_j}{h} \right),$$

donde

$$K * K(x) = \int K(y) K(x - y) dy$$

es la convolución de K con K .

El segundo término no se puede calcular directamente a partir de la muestra, pero podemos estimarlo:

$$2 \int \hat{f}_{n,K}(x) f(x) dx = 2\mathbb{E} \left(\hat{f}_{n,K}(X) \right)$$

donde X es una v.a. con densidad f . Entonces, si tuviésemos una m.a.s. Y_1, \dots, Y_n de X (independiente de la muestra de los X_1, \dots, X_n) podríamos construir el estimador:

$$\frac{1}{n} \sum_{i=1}^n \hat{f}_{n,K}(Y_i)$$

El procedimiento anterior lo podemos *imitar* construyendo para cada X_i el estimador del segundo término utilizando toda la muestra excepto esa observación:

$$\hat{f}_{n,K}^{-i}(X_i) = \frac{1}{n-1} \sum_{j \neq i} K_h(X_i - X_j)$$

Entonces:

$$\frac{1}{n} \sum_{i=1}^n \hat{f}_{n,K}^{-i}(X_i) = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j>i} K_h(X_i - X_j)$$

será un estimador de $\int \hat{f}_{n,K}(x) f(x) dx$.

Por tanto, la función CV se puede estimar mediante:

$$CV(h) = \frac{1}{n^2 h} \sum_{i,j} K * K \left(\frac{X_i - X_j}{h} \right) - \frac{4}{n(n-1)h} \sum_{i=1}^n \sum_{j>i} K_h(X_i - X_j)$$

El método de validación cruzada selecciona como parámetro de suavizado el valor h_{CV} que minimiza CV .

Recordemos la expresión del $MISE$:

$$MISE(h) = \mathbb{E} \int \hat{f}_{n,K}^2(x) dx - 2\mathbb{E} \int \hat{f}_{n,K}(x) f(x) dx + \int f^2(x) dx$$

Siguiendo la misma idea que para el ISE, consideremos la función a minimizar:

$$LSCV(h) = \int \hat{f}_{n,K}^2(x) dx - \frac{2}{n} \sum_{i=1}^n \hat{f}_{n,K}^{-i}(X_i)$$

donde:

$$\hat{f}_{n,K}^{-i}(X_i) = \frac{1}{n-1} \sum_{j \neq i} K_h(X_i - X_j)$$

La ventana h_{LSCV} es la que minimiza $LSCV(h)$.

En lugar de utilizar la fórmula del MISE se considera su expresión asintótica:

$$AMISE(h) = \frac{1}{nh}R(K) + \frac{1}{4}h^4\mu_2(K)^2R(f'')$$

donde la función objetivo a minimizar (*BCV*) se obtiene substituyendo $R(f'')$ por el estimador:

$$\hat{R}(f'') = \frac{1}{n^2} \sum_{i=1}^n \sum_{j \neq i} \left(K_h'' * K_h'' \right) (X_i - X_j)$$

Recordemos que estamos trabajando con la suposición de que $R(f'') < \infty$. Bajo condiciones de regularidad sobre f , y aplicando integración por partes, se puede demostrar que:

$$\int (f''(x))^2 dx = \int f^{(4)}(x) f(x) dx = \psi_4$$

donde $\psi_4 = \mathbb{E}(f^{(4)}(X))$.

Jones y Seather (1991) propusieron estimar ψ_r como:

$$\hat{\psi}_r(g) = \frac{1}{n} \sum_{i=1}^n f_{n,L,g}^{(r)}(X_i) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n L_g^{(r)}(X_i - X_j)$$

donde L es una función núcleo y g es una ventana, no necesariamente iguales a K y h (aunque se suele tomar $L = K$). Pero tenemos que seleccionar g ...

Para seleccionar g , buscamos el que minimiza el AMSE (Wand y Jones, 1995):

$$g_{AMSE} = \left(-\frac{2K^r(0)}{\mu_2(K)n\psi_{r+2}} \right)^{1/(r+3)}$$

Volviendo al problema inicial, recordemos que:

$$h_{AMISE} = \left(\frac{R(K)}{\mu_2(K)n\psi_4} \right)^{1/5}.$$

Si reemplazamos ψ_4 por su estimador, utilizando la ventana óptima g_{AMSE} tendríamos:

$$\hat{h}_{PI} = \left(\frac{R(K)}{\mu_2(K)n\hat{\psi}_4(g_{AMSE})} \right)^{1/5},$$

donde:

$$g_{AMSE} = \left(-\frac{2K^4(0)}{\mu_2(K)n\psi_6} \right)^{1/7}.$$

Pero esto vuelve a depender de ψ_6 , y si hacemos lo mismo, volvemos a tener ψ_8 ...

Normalmente, se para después de varias etapas (dos etapas, usualmente). En la última etapa, se reemplaza el valor de ψ_r por lo que valdría para la densidad Normal. En el caso de una Normal con desviación σ , para r par se tiene:

$$\psi_r = \frac{(-1)^{r/2} r!}{(2\sigma)^{r+1} (r/2)! \sqrt{\pi}}.$$

Ejercicio. Detalla el método de Seather-Jones en dos etapas, para la selección de ventana.