1. ¿Cuál de las siguientes funciones sirven como kernel para estimación de densidades? Probar su afirmación.

   a. $K(x) = I(|x| < 1)/2$
   b. $K(x) = I(0 < x < 1)$
   c. $K(x) = 1/x$
   d. $K(x) = \frac{3}{2}(2x+1)(1-2x)I(-\frac{1}{2} < x < \frac{1}{2}$
   e. $K(x) = \frac{3}{4}(1-x^2)I(|x| < 1)$.

1.a) Redangular 1. $\int_{\mathbb{R}} K_1(x)\,dx = \int_{-1}^{1} \frac{dx}{2} = 1.$

2. K_1 es par 3. $K_1(x) \geq 0 \quad \forall x$

4. $\int_{\mathbb{R}} x K_1(x)\,dx = \int_{-1}^{1} \underbrace{x}_{\text{impar}} = 0$

5. $\int_{\mathbb{R}} \underbrace{x^2}_{\text{PAR}} \underbrace{K_1(x)}_{\text{PAR}}\,dx = \int_{-1}^{1} x^2\,dx = 2x^3\Big]_0^1 = 2 < \infty$

Sí sirve

1.b) 1. $\int_{\mathbb{R}} K_2(x)\,dx = \int_0^1 dx = 1.$

2. $K_2(x) \neq K_2(-x) = 0$

NO sirve

1.c) 1. $\int_{\mathbb{R}} K_3(x)\,dx = \int_{\mathbb{R}} 1/x\,dx = \ln(x)\Big|_{-\infty}^{\infty} \overset{x\in\mathbb{R}^+}{=} \ln(x)\Big]_0^{\infty} = \lim_{\substack{a\to\infty \\ b\to 0}} \ln\left(\frac{a}{b}\right) = \infty$ ∴ NO sirve

1.d) 1. $\int_{\mathbb{R}} K_4(x)\,dx = \int_{-1/2}^{1/2} \frac{3}{2}(2x+1)(1-2x)\,dx \underset{\substack{u(-1/2)=-1 \\ u(1/2)=1}}{\overset{\substack{u=2x \\ \frac{1}{2}du=dx}}{=}} \frac{3}{4}\int_{-1}^{1}(u+1)(1-u)\,du = -\frac{3}{4}\int_{-1}^{1}\underbrace{(u^2-1)}_{\text{PAR}}\,du = -\frac{3}{2}\int_0^1 (u^2-1)\,du = -\frac{3}{2}\left(\frac{u^3}{3}-u\right)\Big]_0^1$

$= -\frac{3}{2}\left(\frac{1}{3}-1\right) = -\frac{3}{2}\left(\frac{1-3}{3}\right) = 1.$

2. $K_4(x) \leq 0 \quad \forall x$

NO sirve

1.e) Epanechnikov 1. $\int_{\mathbb{R}} K_5(x)\,dx = \frac{3}{4}\int_{-1}^{1} \underbrace{(1-x^2)}_{\text{PAR}}\,dx = \frac{3}{2}\int_0^1 (1-x^2)\,dx = \frac{3}{2}\left(x-\frac{x^3}{3}\right)\Big]_0^1 = \frac{3}{2}\left(1-\frac{1}{3}\right) = \frac{3}{2}\cdot\frac{2}{3} = 1.$

2. $K_5(x) = \underbrace{K_5(-x)}_{\text{PAR}}$ 3. $K_5(x) \geq 0 \quad \forall x$

4. $\int_{\mathbb{R}} \underbrace{x}_{\text{IMPAR}} \underbrace{K_5(x)}_{\text{PAR}} = \int_{-1}^{1} x K_5(x)\,dx = 0$

5. $\int_{\mathbb{R}} \underbrace{x^2}_{\text{PAR}} \underbrace{K_5(x)}_{\text{PAR}}\,dx = 2\int_0^1 x^2\left(\frac{3}{2}(1-x^2)\right)dx < \infty$

Sí sirve

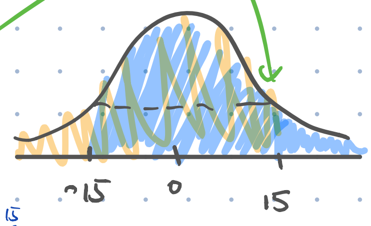
2. Con el siguiente conjunto de datos: $12, 13, 12, 15, 15, 16, 19, 20, 25$ estimar $p^* = P(X < 15)$ usando un estimador de densidad basado en un kernel gaussiano con $h_n = \sqrt{3/n}$ "a mano" (pueden hacer los cálculos con R, pero sin usar ninguna de las funciones predefinidas).

• $n = 9$

• Queremos $\mathbb{P}[X<15] \approx \int_{-\infty}^{15} \hat{f}_n^K(x)\,dx = \int_{-\infty}^{15} \frac{1}{n h_n} \sum_{i=1}^{n} K_0\left(\frac{x_i - x}{h_n}\right) dx$

¿No es una integral "negativa"?

$$= \int_{-\infty}^{15} \frac{1}{3\cdot\sqrt{9}} \sum_{i}^{9} K_0\left(\frac{x_i - x}{\sqrt{3/9}}\right) dx = \frac{1}{3\sqrt{3}} \int_{-\infty}^{15} \sum_{i}^{9} \varphi\left(\frac{x_i - x}{\sqrt{1/3}}\right) dx$$



$$= \frac{1}{3\sqrt{3}} \sum_{i}^{9} \int_{-\infty}^{15} \varphi\left(\frac{x_i - x}{\sqrt{1/3}}\right) dx \overset{\varphi \text{ par}}{=} \frac{1}{3\sqrt{3}} \sum_{i}^{9} \int_{-\infty}^{15} \varphi\left(\frac{x - x_i}{\sqrt{1/3}}\right) dx$$

- $u = (x - x_i)/\sqrt{1/3}$
- $\sqrt{1/3}\,du = dx$
- $u(-\infty) = -\infty$
- $u(15) = (15 - x_i)/\sqrt{1/3}$

$$\left.\right\} = \frac{1}{9} \sum_{i}^{9} \Phi\left(\frac{15 - x_i}{\sqrt{1/3}}\right)$$

Tarea 4. Fecha de entrega: Jueves 30 de abril de 2020.

Problemas

1. ¿Cuál de las siguientes funciones sirven como kernel para estimación de densidades? Probar su afirmación.

   a. $K(x) = I(|x| < 1)/2$
   b. $K(x) = I(0 < x < 1)$
   c. $K(x) = 1/x$
   d. $K(x) = \frac{3}{2}(2x+1)(1-2x)I(-\frac{1}{2} < x < \frac{1}{2}$
   e. $K(x) = \frac{3}{4}(1-x^2)I(|x| < 1)$.

2. Con el siguiente conjunto de datos: $12, 13, 12, 15, 15, 16, 19, 20, 25$ estimar $p^* = P(X < 15)$ usando un estimador de densidad basado en un kernel gaussiano con $h_n = \sqrt{3/n}$ "a mano" (pueden hacer los cálculos con `R`, pero sin usar ninguna de las funciones predefinidas).

3. Generar 20 observaciones de una distribución mezcla donde la mitad de las observaciones son normales estándar y la otra mitad son datos normales $\mathcal{N}(1, 0.64)$. Usar la función `density` de `R` para crear un estimador de la densidad.

4. Usar datos de Michelson del paquete `HistData` (contiene conjuntos de datos históricos). Estos datos son 100 mediciones de la velocidad de la luz en el aire que se obtuvieron por Albert Michelson en 1879.

```
library(HistData)
data(Michelson)
head(Michelson)


  velocity
1      850
2      740
3      900
4     1070
5      930
6      850
```

   Construir un estimador de densidad de kernel que sientan que da una mejor imagen visual de la información que proveen los datos. ¿Qué paramétros escogieron?

5. Consideren la función

$$s(x) = \begin{cases} 1 - x + x^2 - x^3 & 0 < x \leq 1 \\ -2(x-1) - 2(x-1)^2 & 1 < x \leq 2 \\ -4 - 6(x-2) - 2(x-2)^2 & 2 < x \leq 3 \end{cases}$$

   ¿ $s(x)$ define un spline suave cúbico en $[0, 3]$ con nodos 1 y 2? Si lo es, graficar los 3 polinomios en $[0, 3]$.

6. Simular un conjunto de datos como sigue:

```
u <- runif(100)
x <- sort(u)
y <- x^2 + 0.1*rnorm(100)
```

Ajustar un spline interpolado a los datos simulados.

7. Con los datos del archivo `motor.dat`. Estos datos corresponden a datos simulados de accidentes de motocicleta. La variable `time` es el tiempo en milisegundos desde el impacto y `accel` es la aceleración de la cabeza del muñeco (g).

   - Construir un estimador `LOESS` a los datos.
   - ajustar un spline a los datos.

8. Usar el teorema de Taylor para expandir las siguientes ssucesiones a términos de orden h_n^2, donde $h = h_n \to 0$ cuando $n \to \infty$:

   a. $\sin^2(\pi/4 + h)$
   b. $e^{h^2+h} - e^{h^2}$
   c. $\exp(-e^{x-h})$
   d. $\frac{x-h}{x+h}$

9. Basado en una muestra $X_1, \ldots, X_n$, un estimador M-bipeso de ubicación se define como la solución $\hat{\theta}_c$, a

$$\sum_{i=1}^{n} \xi_c \left(\frac{X_i - \theta}{\hat{\sigma}} \right) = 0$$

   donde $\xi_c(x) = x(c^2 - x^2)^2 I(|x| < c)$ y $\hat{\sigma}$ es un estimador de escala.

   Mostrar que $\hat{\theta}_c$ es igual a la moda del estimado del kernel de densidad basado en $X_1, \ldots, X_n$ con ancho de banda $h = c\hat{\sigma}$ y el kernel bipeso $K(x) = \frac{15}{16}(1 - x^2)^2 I(|x| < 1)$

10. Usar `R` para construir una gráfica para los datos de radiocarbón que se encuentran en el conjunto de datos `radioc` del paquete `sm` que muestra un estimador de la varianza del error σ^2 basado en la suma de los residuales al cuadrado y grados de libertad aproximados como una función del parámetro h. La matriz de pesos S se puede construir a través de la función `sm.weight(x,y,h)`. Agregar a la gráfica una línea horizontal correspondiente a la diferencia basado en el estimador de Gasser *et al.* (1986), que es disponible en la función `sm.sigma(x, y)`.

    ¿A qué valor el estimador basado en la suma de los residuales al cuadrado converge cuando el parámetro de suavizamiento se hace muy grande?