# Estimación con histogramas

Sabemos que dada una función de densidad de probabilidad $f$ y $X \sim \mathcal{L}(f)$

$$\boxed{f(x)} = \lim_{h \to 0} \frac{1}{2h} \mathbb{P}\left[x - h < X < x + h\right] \quad = \int_{x-h}^{x+h} f(x)\, dx$$

Por otro lado dada una muestra $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{L}(f)$, por Monte-Carlo se tiene la paroximación

$$\mathbb{P}\left[x - h < X < x + h\right] \approx \frac{\#\left\{1 \leq i \leq n \,:\, X_i \in (x - h, x + h)\right\}}{n}$$

Lo cual motiva el uso para $h$ pequeño de

$$\hat{f}\,(x) = \frac{\#\left\{1 \leq i \leq n \,:\, X_i \in (x - h, x + h)\right\}}{2hn} \approx f(x)$$

Si definimos

$$w(x) = \begin{cases} \frac{1}{2} & \text{si } |x| < 1 \\ 0 & \text{en otro caso.} \end{cases}$$

con lo que

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

$$\boxed{\frac{\#\{1 \le i \le n \,:\, X_i \in (x-h, x+h)\}}{2hn}}$$

$$\frac{1}{n}\sum_{i=1}^{n} \frac{1}{h} w\left(\frac{x - X_i}{h}\right)$$

$$= \frac{1}{n}\frac{1}{h}\sum_{i=1}^{n} \frac{1}{2} \mathbb{1}\left\{\left|\frac{x - X_i}{h}\right| \le 1\right\}$$

$$= \frac{1}{n}\frac{1}{h} \frac{\#\{1 \le i \le n : X_i \in (x-h, x+h)\}}{2}$$

$$\frac{\#\{1 \le i \le n : X_i \in (x-h, x+h)\}}{2hn}$$

$$X_i \in (x-h, x+h)$$

$$x - h \le X_i \le x + h$$

$$-h < X_i - x \le h$$

$$|X_i - x| \le h$$

$$\left|\frac{X_i - x}{h}\right| \le 1$$

En el caso de kernel Gaussiano

$$\boxed{h_{\text{opt}} = 1.06\sigma n^{-\frac{1}{5}}}$$

La regla de Scott para elegir $h$ es

$$h_{\text{opt}} = 1.06\hat{\sigma} n^{-\frac{1}{5}}$$

alternativamente se puede utilizar $R$ el rango intercuantil del kernel Gaussiano asociado obteniendi estimadores más robustos para cuando se tiene comportamiento alejada al Gaussiano.

$$h_{\text{opt}} = 0.79 R n^{-\frac{1}{5}}.$$

Finalmente la regla de Silverman está dada por

$$h_{\text{opt}} = 0.9 A n^{-\frac{1}{5}}$$

con $A = \min\{\hat{\sigma}, R\}$.

$$\frac{1}{h}K\left(\frac{x-x_i}{h}\right) = \frac{1}{h}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\left(\frac{x-x_i}{h}\right)^2}{2\sigma^2}\right)$$

$$= \frac{1}{\sqrt{2\pi}\,(h\sigma)}\exp\left(\frac{(x-x_i)^2}{2h^2\sigma^2}\right)$$

$$= K'(x-x_i)$$

$$K' = \mathcal{L}(\text{Normal}(0,\sigma^2 h^2))$$