Estimación empírica de distribuciones:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \hline \mathbf{x}_2 \\ \hline \mathbf{x}_3 \\ \hline \vdots \\ \hline \mathbf{x}_n \end{bmatrix}_{n \times d}$$

$\mathbf{x}_i \in \mathbb{R}^d$

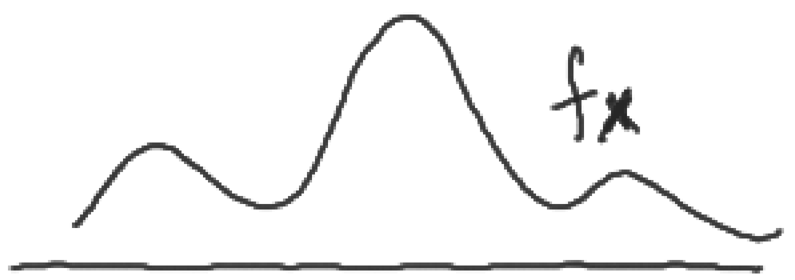
$\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{id})$

conjunto de datos
(muestra de observaciones)

$n = \#$ observaciones
$d = \#$ variables

X variable aleatoria $\quad (\Omega, \mathcal{E}, \mathbb{P}) \xrightarrow{X} \mathbb{R}$

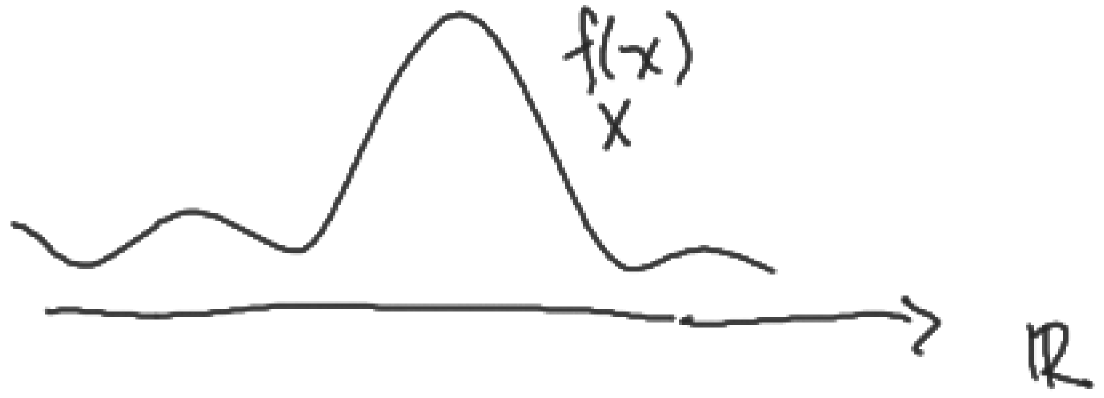
X tiene asociada una distribución de prob.



f_X = función de densidad de X.

$f_X : \mathbb{R} \longrightarrow \mathbb{R}$

$X =$ variable aleatoria real



$f_X : \mathbb{R}^d \longrightarrow \mathbb{R}$

$X =$ vector aleatorio

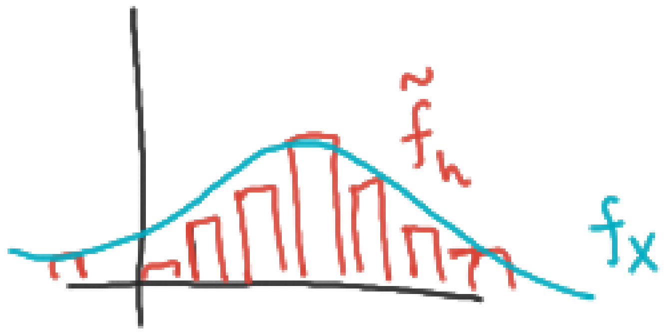
$$\mathbb{X} = \begin{bmatrix} x_1 \\ \hline x_2 \\ \hline \vdots \\ \\ x_n \end{bmatrix}_{n \times d}$$

x_i son observaciones de cierta variable aleatoria X. $\leftarrow$ f_X ¿cuál?

Histograma

frec/total

$$\left.\begin{matrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{matrix}\right\} \text{datos}$$



$$f_X \approx \tilde{f}_h$$

Dado un conjunto de datos, queremos estimar f_X.

Estimador de densidad por kernel

Estimador de Parzen-Rosenblatt / ventanas de Parzen.

aprox. de f_X

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i)$$

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

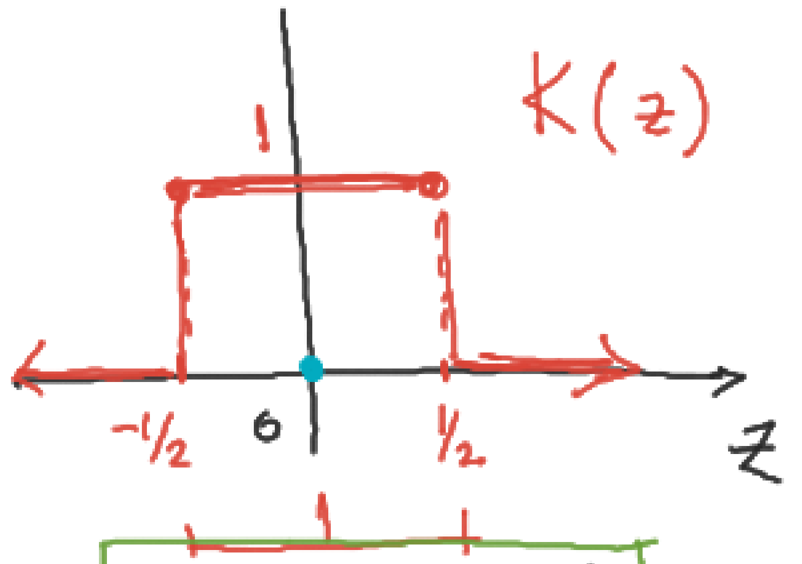
$$K_h(x - x_i) = \frac{1}{h} K\left(\frac{x - x_i}{h}\right).$$

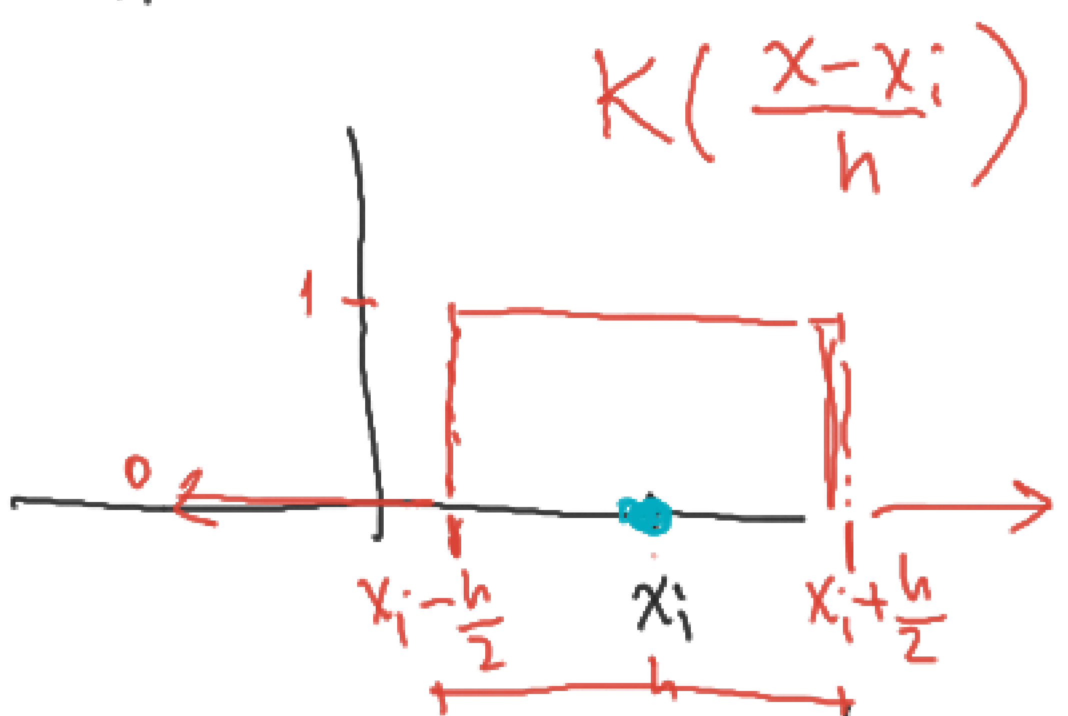
K es una función de kernel

- $K \geq 0$ (K es no-negativa)
- $h > 0$ (parámetro de suavizamiento)

Función kernel

$$z = \frac{x - x_i}{h}$$



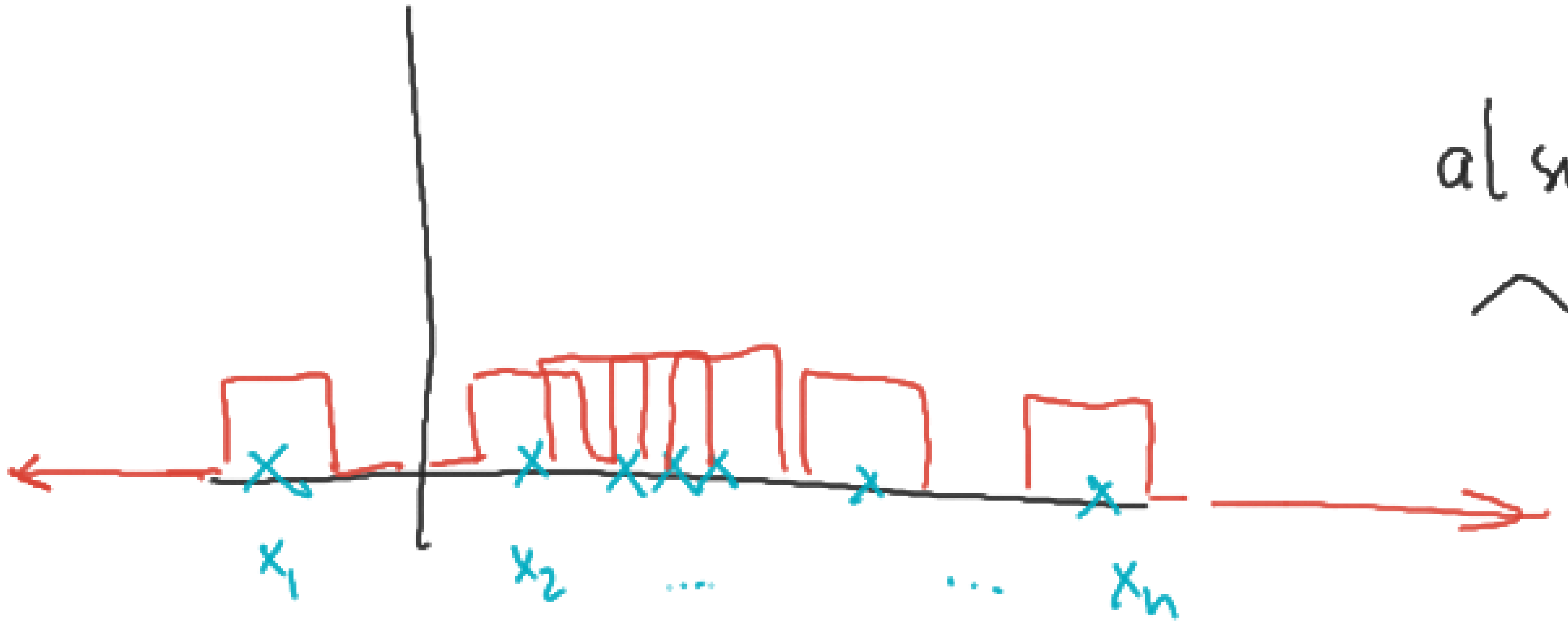


$$\boxed{z = \frac{x - x_i}{h}} \iff zh = x - x_i \iff x = x_i + zh$$

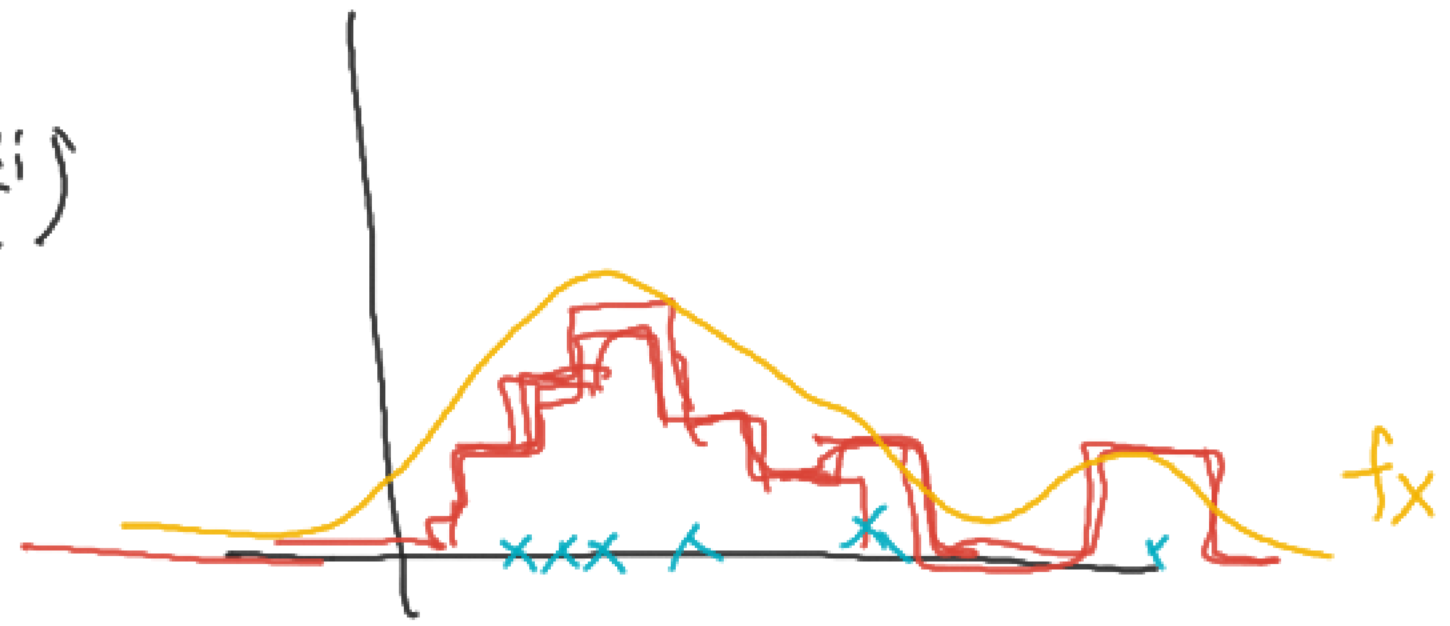
$$z = -\tfrac{1}{2} \Rightarrow x = x_i - h/2$$

$$z = \tfrac{1}{2} \Rightarrow x = x_i + h/2$$

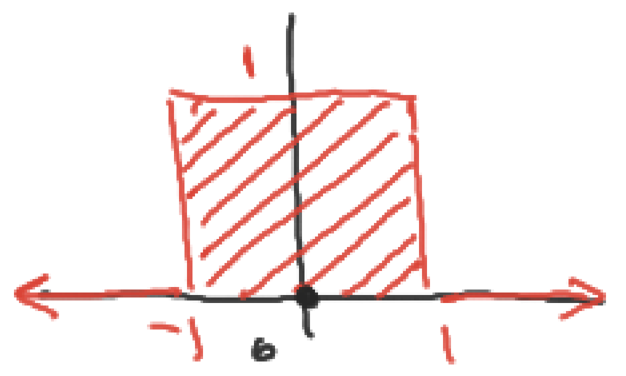
h



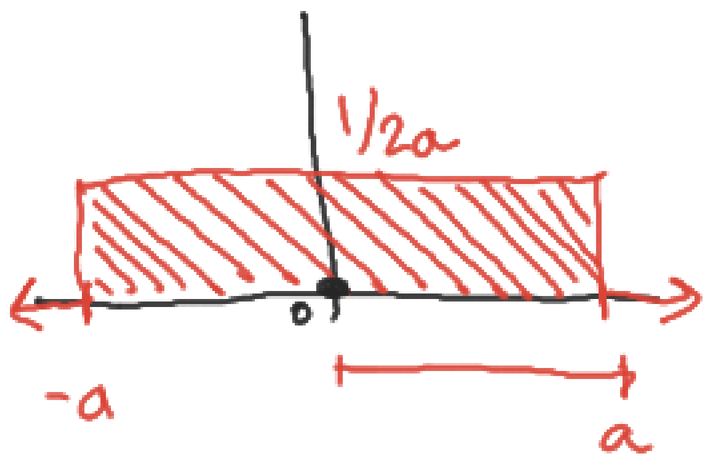
al sumar $\frac{1}{nh}\sum K\left(\frac{x - x_i}{h}\right)$ $\rightsquigarrow$



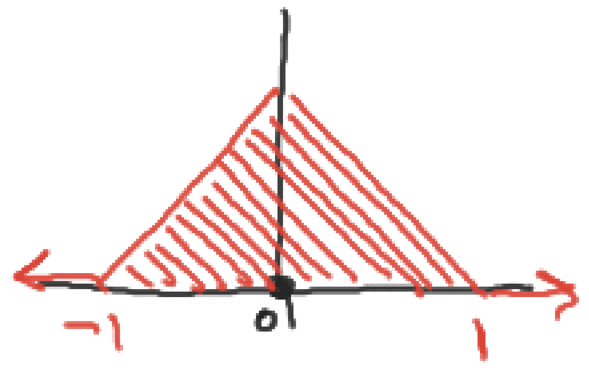
Tipos de Kernel



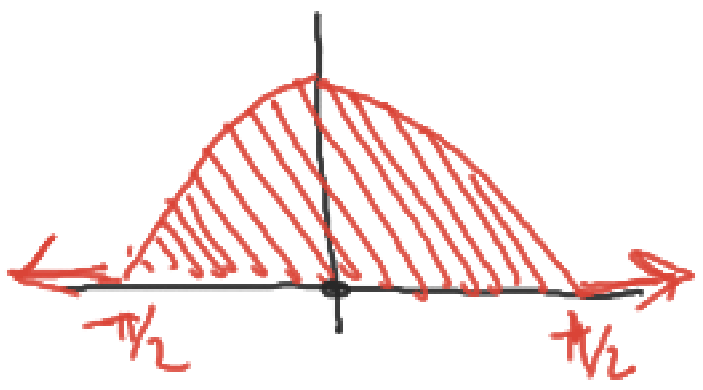
square
top-hat



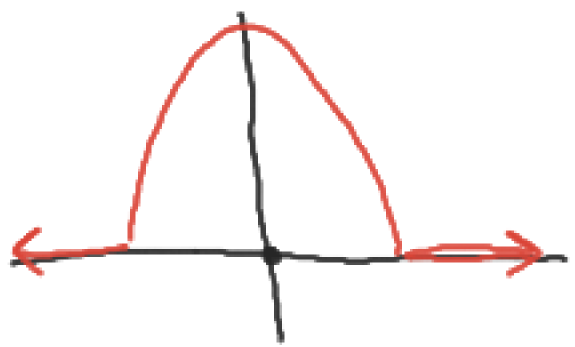
rectangular



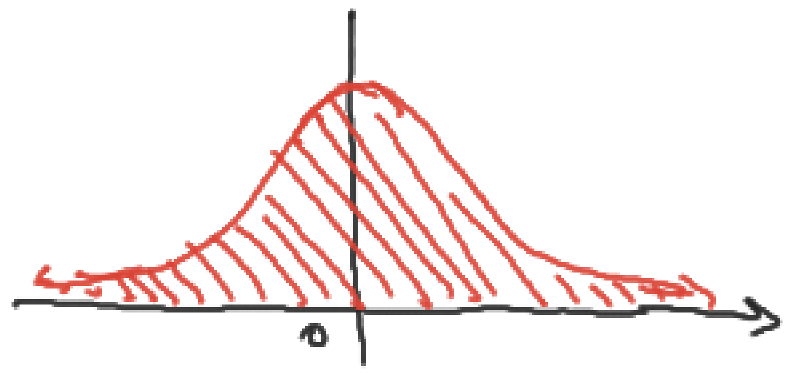
triangular



cosine

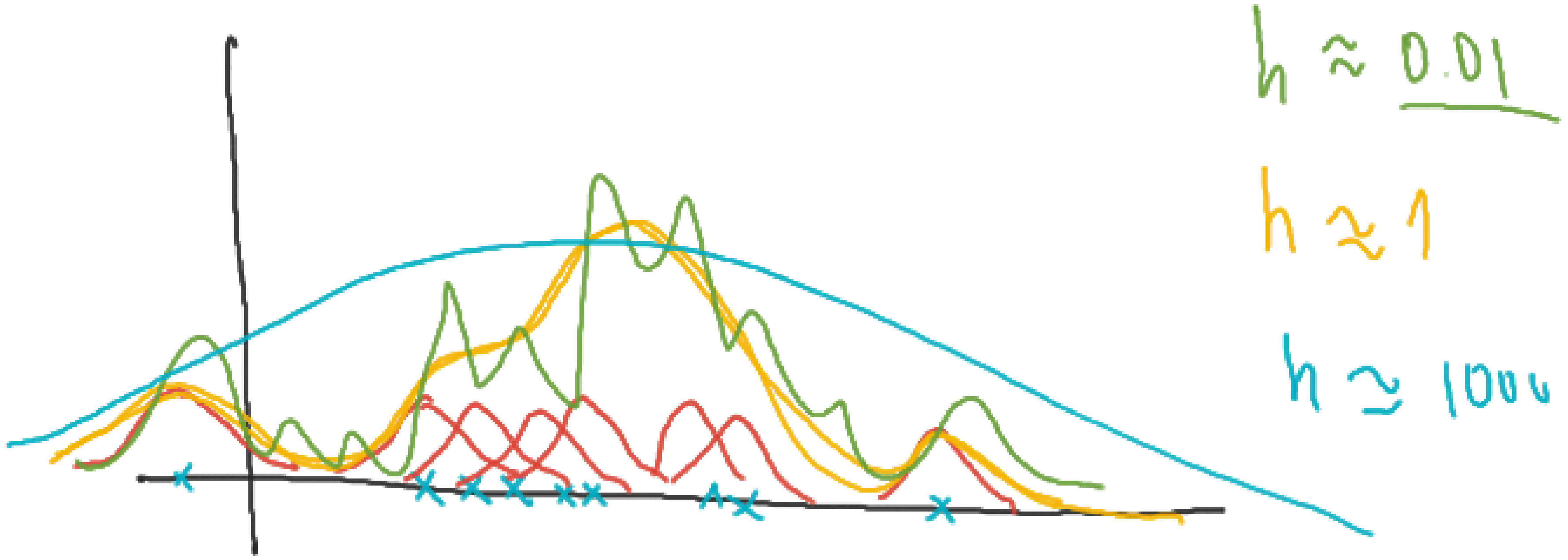


Epanechnikov



gaussian

laplacian



h ≈ 0.01
h ≈ 1
h ≃ 1000