Estimación Empírica de Distribuciones

muestra de datos

$(x_1, x_2, \dots, x_n)$ ⇝ observaciones de una v.a. X. → i.i.d.

(independientes + identicamente distrib)

X y los x_i tienen asociada una distribución $f_X(t)$ ó $F_X(t)$

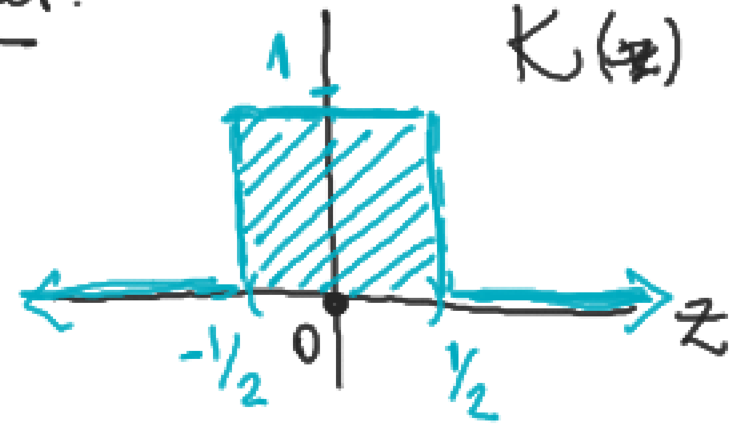
Queremos estimar $f_X(t)$.

Def: El estimador de densidad por kernel de $f_X(t)$ es

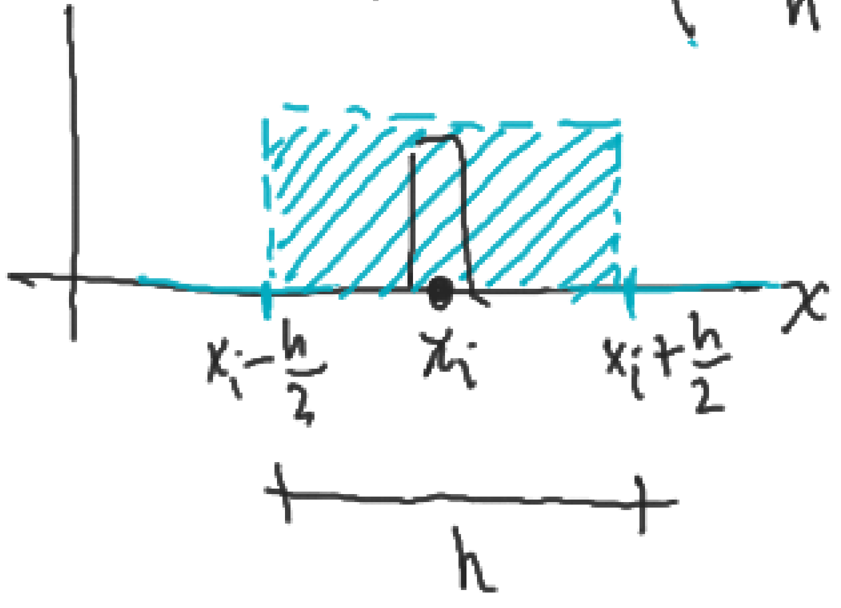
$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x-x_i) = \frac{1}{hn}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)$$

- K es una función kernel (no-negativa, $K \geq 0$)
- $h > 0$, parámetro de suavizamiento.

Kernel:



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



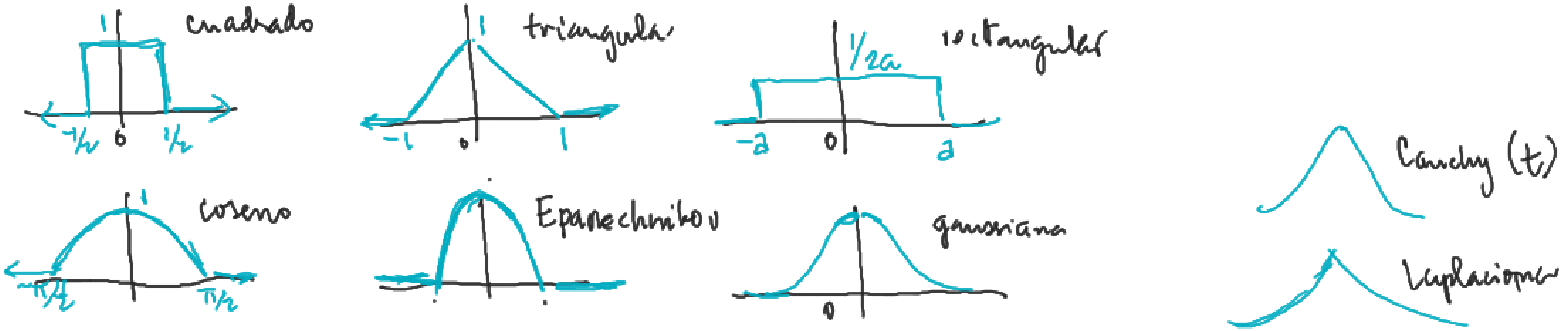
$$z = \frac{x-x_i}{h}$$

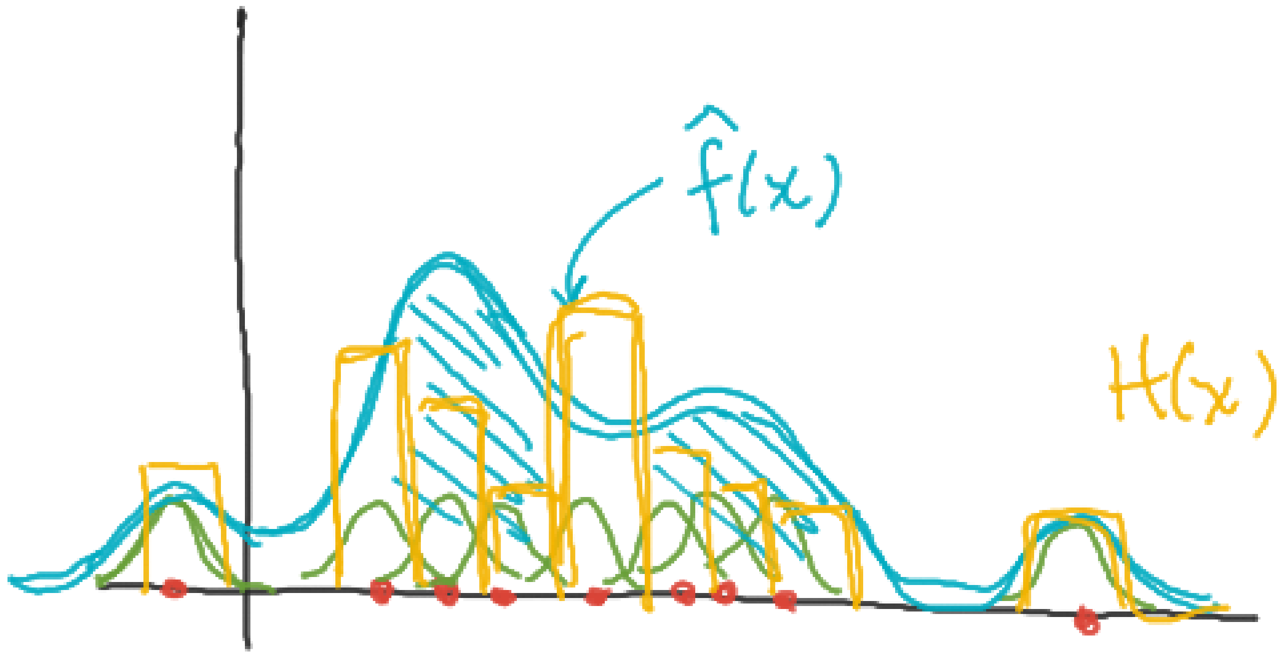
$$-1/2 = \frac{x-x_i}{h} \Rightarrow x = x_i - \frac{h}{2}$$

$$+1/2 = \frac{x-x_i}{h} \Rightarrow x = x_i + \frac{h}{2}$$

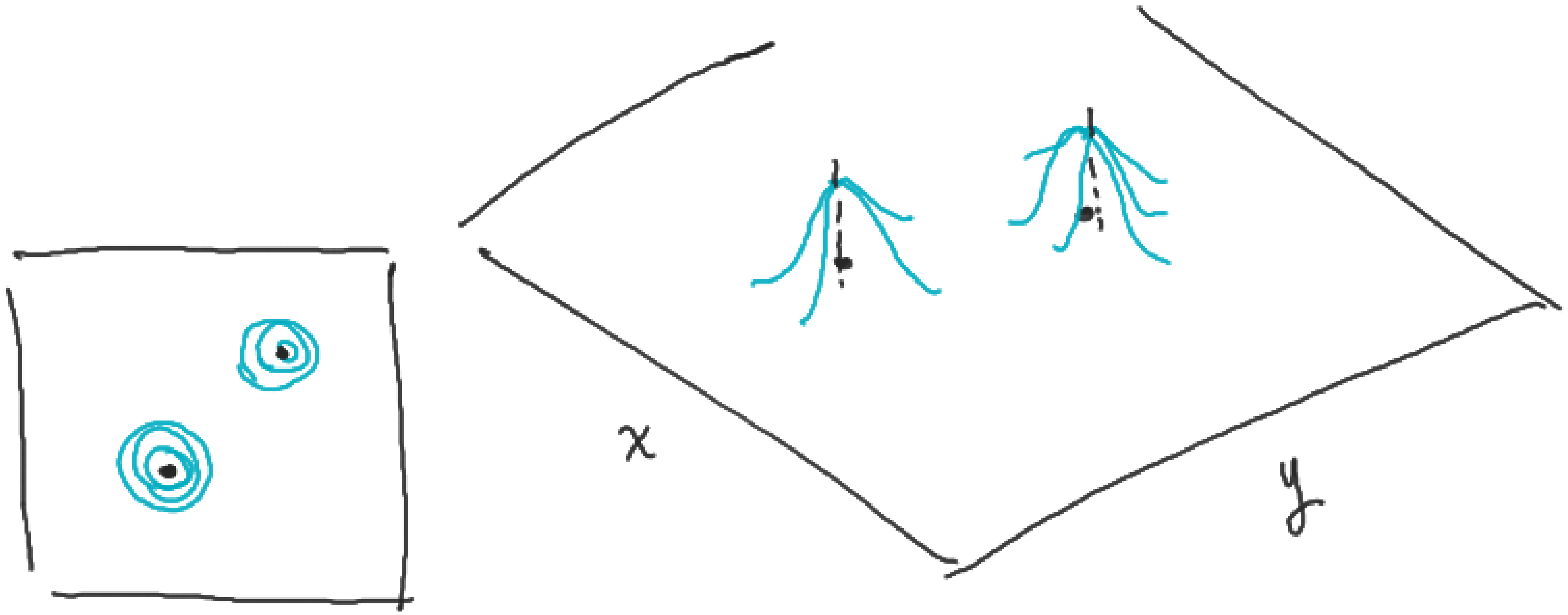
Tipos de Kernel:

$$K(x) = e^{-x^2/h^2}$$





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


x
y