Density Estimation

ECDF

Aim: estimate $f(x)$ ⟶ difficult than $F(x)$

$$\hat{p}_n(x) = \sum_{i=1}^{M} \mathbb{1}\{x \in B_i\} \cdot M \cdot \hat{p}_i$$

$$\hat{p}_i = \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}\{X_k \in B_i\}$$

Outline:

Histogram

① $MSE(x) = b^2(x) + v(x)$

② $MISE(\hat{p}_n) = \int MSE(x)\,dx$

$= \mathbb{E}\left[\int (\hat{f} - f)^2\,dx\right]$

③ Confidence Set → not accurate

④ CV → estimate of $MISE(\hat{p}_n)$ → infeasible

CV ⇒ Shortcuts

CV → choose h_{cv}

Kernel

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

① $\hat{p}_n(x) \xrightarrow{p} p(x)$

② $MSE(x)$

③ $MISE(\hat{p}_n)$

④ CV → optimal bandwidth h_{cv}

$\hat{h}_{cv} \xrightarrow{p} h_{MSE}$