Overview of parameter estimation

- We have a density function f(X; θ) whose parameters θ are unknown.
- We have a dataset D of n independent observations from f
  - D is random variable denoting $X_1, X_2, \ldots, X_n$ where each $X_i \sim f(X; \theta)$
- We use any method to get an estimate $\hat{\theta} = A(D)$ as some function of D. Thus $\hat{\theta}$ is also a R.V.
  - Alternative notations $\hat{\theta}_n, \hat{\theta}_D$ to stress that estimate depends on D.

$f(X;\theta)$ hidden

$D \quad X_1 \quad X_2 \cdots X_n$

$X_1 = v_1 \quad X_2 = v_2 \cdots X_n = v_n$

$\hat{\theta} \equiv A(D)$

$\hat{\theta}_D \quad \hat{\theta}_n$

Example: $f(x;\theta) \equiv N(x;\ \theta \equiv (\mu, \sigma^2)$ $x \equiv$ height of people.

$D \equiv \{X_1, X_2 \cdots X_{10}\}$ $\hat{\theta}_{10} \equiv A(D) = \frac{v_1 + v_2 + \cdots v_{10}}{10}$ Sample-mean

Risk of an estimate — Theoretical exercise.

- Let $\theta, \hat{\theta}_n$ be actual and estimated quantities.
- Risk = Expected square error

$$E_D\left[(\hat{\theta}_n - \theta)^2\right]$$

$$\int (\hat{\theta}_n - \theta)^2 p(D)\, dD \qquad D \equiv \{x_1 \cdots x_n\}$$

$$\int_{x_1}\int_{x_2}\cdots\int_{x_n} \left(\underbrace{A(x_1, x_2 \cdots x_n)}_{\hat{\theta}_n} - \theta\right)^2 f(x_1) f(x_2) \cdots f(x_n)\, dx_1 \cdots dx_n .$$

$\hat{\theta}_n^1 = 5.7$ ft

$\hat{\theta}_n^2 = 5.9$ ft.

$\hat{\theta}^\infty \equiv 6.0$ ft

Bias and Variance

$\hat{\theta}_n$ is a R.V

- Expected value of $\mathrm{E}_D[\hat{\theta}_n] =$
  $\int_D A(D)f(D) = \int_{X_1} \ldots \int_{X_n} A(X_1, \ldots X_n) \prod_i f(X_i) dX_1 dX_2 \ldots dX_n$

Bias $E_D[\hat{\theta}_n] - \theta$

Variance $E_D[(\hat{\theta}_n - E_D[\hat{\theta}_n])^2] = E_D[\hat{\theta}_n^2] - (E_D[\hat{\theta}_n])^2$

Risk = Bias^2 + Variance

To show

$$E_D[(\hat{\theta}_n - \theta)^2] = (E_D[\hat{\theta}_n] - \theta)^2 + Var(\hat{\theta}_n)$$

Proof:

$$E_D[(\hat{\theta}_n - \theta)^2] = E_D[(\hat{\theta}_n - E_D[\hat{\theta}_n] + E_D[\hat{\theta}_n] - \theta)^2]$$

$$= E_D[(\hat{\theta}_n - E_D[\hat{\theta}_n])^2] + E_D[(E_D[\hat{\theta}_n] - \theta)^2] - 2E_D[(\hat{\theta}_n - E_D[\hat{\theta}_n])(E_D[\hat{\theta}_n] - \theta)]$$

$$= Var(\hat{\theta}_n) + (E_D[\hat{\theta}_n] - \theta)^2 - 2(E_D(\hat{\theta}_n) - E_D(\hat{\theta}_n))(\cdots)$$

$= (E_D[\hat{\theta}_n] - \theta)^2$ is the Bias^2; $2(E_D(\hat{\theta}_n) - E_D(\hat{\theta}_n))(\cdots) = 0$

Estimating CDF of any scalar random variable

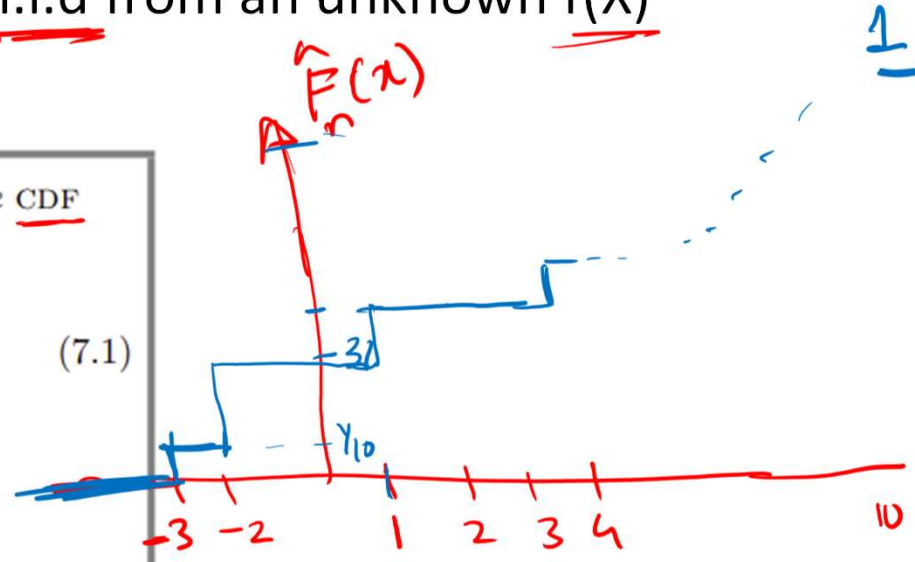
- Given sample: $D = \{X_1, X_2, \ldots, X_n\}$ sampled i.i.d from an unknown f(X)
- Goal: estimate CDF function using D

7.1 Definition. *The* **empirical distribution function** $\widehat{F}_n$ *is the* CDF *that puts mass* $1/n$ *at each data point* X_i. *Formally,*

$$\widehat{F}_n(x) = \frac{\sum_{i=1}^{n} I(X_i \leq x)}{n} \tag{7.1}$$

where

$$I(X_i \leq x) = \begin{cases} 1 & \text{if } X_i \leq x \\ 0 & \text{if } X_i > x. \end{cases}$$



Example: n = 10, D={-3,-2,-2,1,2,2,4,6,8,10}

Analyzing Bias, Variance, and Risk of empirical CDF

Bias: at a x

$$E_D[\hat{F}_n(x)] = E_D\left[\frac{\sum_{i=1}^{n} I(X_i \le x)}{n}\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} E_D[I(X_i \le x)] = \frac{1}{n}\sum_{i=1}^{n} \int_{x_1} \cdots \int_{x_n} I(X_i \le x) f(x_1) \cdots f(x_n)\, dx_n \cdots dx_1$$

$\hat{F}_n(x)$ is unbiased.

$$= \frac{1}{n}\sum_{i=1}^{n} \int_{x_i} I(X_i \le x) f(x_i)\, dx_i \int_{x_1} f(x_1)\, dx_1 \int_{x_2} f(x_2)\, dx_2 \cdots \int_{x_n} f(x_n)\, dx_n$$

$$\int_{x_1} f(x_1)\, dx_1 = 1, \quad \int_{x_2} f(x_2)\, dx_2 = 1, \quad \int_{x_n} f(x_n)\, dx_n = 1 \quad \text{except } "i"$$

$$= \frac{1}{n}\sum_{i=1}^{n} \int_{x_i} I(X_i \le x) f(x_i)\, dx_i$$

$$= \frac{1}{n}\sum_{i=1}^{n} \int_{x_i=-\infty}^{x} f(x_i)\, dx_i = \frac{n}{n} F(x) = F(x)$$

Variance $\hat{F}_n(x)$

$$\mathrm{Var}_D\left(\left[\frac{1}{n}\sum_{i=1}^{n} I(X_i \le x)\right]\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}_{X_i}\left(\left[I(X_i \le x)\right]\right)$$

$$= \frac{1}{n^2} \cdot n \, \mathrm{Var}_{X_1}\left(I(X_1 \le x)\right)$$

$$= \frac{1}{n}\left(\left[E_{X_1}\left(I(X_1 \le x)^2\right)\right] - F(x)^2\right)$$

$$\int_{X_1=-\infty}^{x} 1^2 \cdot f(x_1)\, dx_1$$

$$= \frac{1}{n}\left(1 \cdot F(x) - F(x)^2\right)$$

$$= \frac{1}{n}\left(F(x) - F(x)^2\right)$$

$$= \frac{1}{n} F(x)\left(1 - F(x)\right)$$

Non-parametric density estimation

Reading material: https://faculty.washington.edu/yenchic/18W_425/Lec6_hist_KDE.pdf

Motivation

- $D = \{X_1, X_2, \ldots, X_n\}$ sampled i.i.d from an unknown f(X)
- Estimate $\hat{f}(x)$ without committing on a specific parametric form of f(X)
- Why not use empirical CDF?

  - Too inefficient to maintain. Need to store entire data
  - Too jerky. Non-zero density at observed points, zero elsewhere.

- Density estimation: assume some form of smoothness of f(X)