Prob Stats L13

Estimation, Bias + Variance

Mar 23, 2023

Central Limit Theorem

R.Vs $X_1, X_2, \ldots, X_n \overset{\sim}{\underline{iid}} f$

- $E[X_i] = \mu$
- $Var[X_i] = \sigma^2 < \infty$

Define $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$

① $E[\bar{X}_n] = \mu$

② $Var[\bar{X}_n] = \frac{\sigma^2}{n}$

③ as $n \to \infty$ $\bar{X}_n$ converges to $N(\mu, \frac{\sigma^2}{n})$

③ alt

$$Z_n = \frac{\bar{X}_n - \mu}{\underbrace{\sigma / \sqrt{n}}_{\text{Stdev}(\bar{X}_n)}}$$

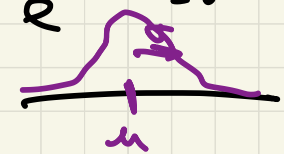
in limit $n \to \infty$

$Z_n \sim N(0, 1)$

Parameters of Distribution

Normal

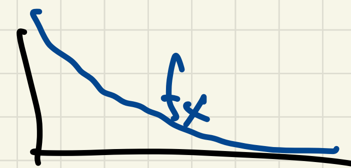
$$X \sim N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{\|x-\mu\|^2}{2\sigma^2}}$$



Exponential

$$X \sim Exp(\lambda)$$

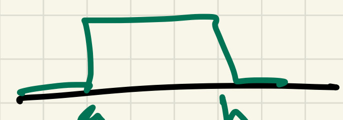
$$f_X(x) = \lambda e^{-\lambda x}$$



Unif

$$X \sim Unif(a, b)$$

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in (a, b) \\ 0 & \text{o.w.} \end{cases}$$



Generically

$$X \sim f(\theta)$$

generic parameters

Estimation : of a Distribution's Parameter

(1)

(0) Get / Gather / Have Data

observations / realization $x_1, x_2, \ldots x_n$ lower case

(1) Choose Distribution $f(\theta)$

- go to R, graph it.
- throw out the world

Modeling

(2) Estimate Parameters θ

RVs $X_1, X_2, \ldots X_n$ iid $f(\theta)$ constant

hat means estimates

estimator $\hat{\theta} = T(X_1, X_2, \ldots, X_n)$

R.V. $f_{\hat{\theta}}$

what I want:

$E[\hat{\theta}] = \theta \Leftarrow$ **unbiased** estimator

$$\text{bias}(\hat{\theta}) = E[\hat{\theta}] - \theta$$

Ex: $X_1 \ldots X_n \overset{iid}{\sim} N(\mu, 1)$

$$\hat{\mu} = T(X_1, \ldots X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n$$

unbiased $E[\hat{\mu}] = \mu$

Efficiency Estimator

$X_1, \ldots X_n \sim f(\theta)$

$$\hat{\theta}_1 = T_1(X_1, \ldots X_n)$$

$$\hat{\theta}_2 = T_2(X_1, \ldots X_n)$$

$$E[\hat{\theta}_1] = E[\hat{\theta}_2] = \theta$$

both be unbiased.

Ex: $\theta = \mu$ mean

$\bar{X}_n = \hat{\theta}_1$ $\quad \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

$X_1 = \hat{\theta}_2$ also unbiased.

more efficient

$$\underbrace{Var[\bar{X}_n]}_{\sigma^2/n} < \underbrace{Var[X_1]}_{\sigma^2}$$

Two goals of parameter estimation

1. Small bias (bias = 0 $\Rightarrow$ unbiased)

2. Small variance

$\{X_i\} \overset{iid}{\sim} N(\mu, \sigma^2)$

$E[\hat{\sigma}^2] = \sigma^2$

sample variance $S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 = \hat{\sigma}^2$

- $E[X_i^2] = Var(X_i) + E[X_i]^2 = \sigma^2 + \mu^2$

- $E[(\bar{X}_n)^2] = Var(\bar{X}_n) + E[\bar{X}_n]^2 = \frac{\sigma^2}{n} + \mu^2$

- $E[X_i \bar{X}_n] = E\left[X_i \cdot \frac{1}{n} \sum_{j=1}^{n} X_j\right] = \boxed{\frac{1}{n} \sum_{j=1}^{n} E[X_i X_j]}$

$= \mu^2 + \frac{\sigma^2}{n}$

n times

$E[X_j X_i] = Cov(X_i, X_j) + E[X_i] E[X_j]$

$i \neq j$ $= 0$ $\qquad$ $= \mu^2$

$i = j$ $= \sigma^2$

$$E[S_n^2] = E\left[\frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \bar{X}_n)^2\right]$$

$$= \frac{1}{n-1} \cdot \sum_{i=1}^{n} E\left[(X_i - \bar{X}_n)^2\right]$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} E\left[X_i^2 - 2X_i\bar{X}_n + \bar{X}_n^2\right]$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} \left(E[X_i^2] - 2E[X_i\bar{X}_n] + E[\bar{X}_n^2]\right)$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} \left((\cancel{\mu^2} + \sigma^2) - \cancel{2}\left(\cancel{\mu^2} + \frac{\sigma^2}{n}\right) + \cancel{\left(\mu^2 + \frac{\sigma^2}{n}\right)}\right)$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} \left(\sigma^2 - \frac{\sigma^2}{n}\right) = \sigma^2\left(\frac{n-1}{n}\right)$$

$$= \sigma^2$$ $\Longleftarrow$ S_n^2 unbiased estimator of σ^2

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)$$

$$E[\hat{\sigma}_n^2] = \frac{n-1}{n}\sigma^2$$

$$\text{bias}(\hat{\sigma}_n^2) = \frac{\sigma^2}{n}$$

not unbiased.

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Review

1. iid = identical, independently distributed

2. Central Limit Theorem CLT

$X_1, \dots X_n \underset{iid}{\sim} f$ $\quad E[X_i] = \mu \quad \text{Var}[X_i] = \sigma^2$

$\bar{X}_n = \frac{1}{n}\sum X_i$ (a) $E[\bar{X}_n] = \mu$ (b) $\text{Var}[\bar{X}_n] = \frac{\sigma^2}{n}$

(c) $\lim_{n \to \infty} Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

3. Estimation $\quad \hat{\theta} = T(X_1 \dots X_n)$ statistic

$\text{bias}(\hat{\theta}) = E[\hat{\theta}] - \theta$ $\quad$ if $\text{bias}(\hat{\theta}) = 0$ unbiased.