Thurs Jan 9

Last time, $X_1 \dots X_n \overset{iid}{\sim} f_X(x) = e^{-(x-\theta)} I(x > \theta)$

$$\hat{\Theta}_1 = \bar{X}_n - 1 \;,\quad \hat{\Theta}_3 = \text{Min}(X_i) - \frac{1}{n}$$

Both Unbiased

$$\text{Var}(\hat{\Theta}_1) = \frac{1}{n} \qquad \text{Var}(\hat{\Theta}_3) = \frac{1}{n^2}$$

$\hat{\Theta}_3$ is more <u>efficient</u>.

Sampling Distribution

The distribution of $T_n(X_1 \dots X_n)$ is sometimes called the sampling distribution of the statistic

Estimating Expected Value

Bernoulli $E(X) = \theta$

Normal $E(X) = \mu$

Poisson $E(X) = \lambda$

Natural estimate is $\bar{X}_n$

unbiased

consistent by LLN

$\bar{X}_n$ is linear combination of $X_1 \cdots X_n$

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}X_1 + \frac{1}{n}X_2 + \cdots + \frac{1}{n}X_n$$

General Linear Combination

$$L = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$$

that is unbiased for μ.

$X_1 \cdots X_n$ iid with $E(X_i) = \mu$, $Var(X_i) = \sigma^2$

L is unbiased for μ iff

$$\mu = E(L) = E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$$

$$= \sum_{i=1}^{n} a_i \mu \qquad \text{If } \mu \neq 0$$

$$\Longrightarrow \sum_{i=1}^{n} a_i = 1 \qquad \text{If } \mu = 0 \ \ a_i \text{ don't matter}$$

L will be unbiased for μ iff $\sum_{i=1}^{n} a_i = 1$

Choose $a_1, \ldots a_n$ so that $Var(L)$ is as small as possible (Efficiency)

Minimize $Var(L) = Var\left(\sum_{i=1}^{n} a_i X_i\right)$

$$\overset{ind}{=} \sum_{i=1}^{n} Var(a_i X_i) = \sum_{i=1}^{n} a_i^2 \, Var(X_i) = \sum_{i=1}^{n} a_i^2 \sigma^2$$

$$= \sigma^2 \sum_{i=1}^{n} a_i^2$$ So to minimize

$Var(L)$, minimize $\sum_{i=1}^{n} a_i^2$ subject to

$$\sum_{i=1}^{n} a_i = 1$$

$$\sum_{i=1}^{n} a_i^2 = \sum_{i=1}^{n} (\underbrace{a_i - \tfrac{1}{n}}_{a} + \underbrace{\tfrac{1}{n}}_{b})^2$$

$$= \sum_{i=1}^{n} \left[(a_i - \tfrac{1}{n})^2 + 2(a_i - \tfrac{1}{n})\tfrac{1}{n} + \tfrac{1}{n^2} \right]$$

$$= \sum_{i=1}^{n} (a_i - \tfrac{1}{n})^2 + \frac{2}{n} \sum_{i=1}^{n} (a_i - \tfrac{1}{n}) + \sum_{i=1}^{n} \frac{1}{n^2}$$

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

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

$$= \underbrace{\sum_{i=1}^{n} (a_i - \tfrac{1}{n})^2}_{= 0 \text{ iff } a_1 = a_2 = \cdots = a_n = \frac{1}{n}} + \frac{1}{n}$$

And $\sum_{i=1}^{n} a_i^2$ is MINIMIZED

$\overline{X}_n$ is Best Linear Unbiased Estimator

BLUE

Estimating $Var(X_i) = \sigma^2$

$$Var(X_i) = E(X-\mu)^2 \overset{\text{discrete}}{=} \sum_x (x-\mu)^2 p_X(x)$$

Random sample $X_1, \ldots, X_n$, $E(X) = \mu$, $Var(X) = \sigma^2$

Sample Variance

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

More natural would be use n

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

$n-1$ is to make it unbiased

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

$$=E\left(\sum_{i=1}^{n}(\underset{a}{X_i-\mu}+\underset{b}{\mu-\bar{X}_n})\right)^2$$

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

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

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

$$=\sum_{i=1}^{n}\sigma^2-2E(\bar{X}_n-\mu)(n\bar{X}_n-n\mu)+nE(\bar{X}_n-\mu)^2$$

$$=n\sigma^2-2nE(\bar{X}_n-\mu)^2+nE(\bar{X}_n-\mu)^2$$

$$=n\sigma^2-nE(\bar{X}_n-\mu)^2=n\sigma^2-n\,\mathrm{Var}(\bar{X}_n)$$

$$=n\sigma^2-\cancel{n}\frac{\sigma^2}{\cancel{n}}=(n-1)\sigma^2$$ So

$$E(S^2)=E\left(\frac{\sum_{i=1}^{n}(X_i-\bar{X}_n)^2}{n-1}\right)=\frac{\cancel{(n-1)}\sigma^2}{\cancel{n-1}}=\sigma^2$$

UNBIASED

To show S^2 consistent for σ^2, need...

Multi-variable convergence in probability

Let $\underset{\sim}{X} = \begin{pmatrix} X_{1n} \\ \vdots \\ X_{kn} \end{pmatrix}$ be jointly distributed RVs

$n = 1, 2 \ldots$

We will say $\underset{\sim}{X}_n \xrightarrow{P} \underset{\sim}{c}$ if for all $\varepsilon > 0$

$$\lim_{n \to \infty} P\{ \| \underset{\sim}{X}_n - \underset{\sim}{c} \| \geq \varepsilon \} = 0$$

Epsilon ball around $\underset{\sim}{c}$

Theorem

$\underset{\sim}{X}_n \xrightarrow{P} \underset{\sim}{c}$ iff $X_{jn} \xrightarrow{P} c_j$ for $j = 1, \ldots, k$