SIMULATION

STAT 510

Setup

$$X_1, X_2, \dots X_n \sim F$$

- How to estimate F? Empirical distribution
- How to estimate properties of F, $T(F)$? plug-in

$$T_n = g(X_1, X_2, \dots X_n)$$

↑

statistic

Sampling Distribution

$$X_1, X_2, \ldots\ldots X_n \sim F$$

$$T_n = g(X_1, X_2, \ldots X_n)$$

↑

STATISTIC

- WHAT IS THE DISTRIBUTION OF T_n?
  - $\mathbb{E}[T_n] = ?$
  - $\mathbb{V}[T_n] = ?$
  - ETC

Easy Example

$$X_1, X_2, \ldots\ldots, X_n \overset{\text{IID}}{\sim} N(\mu, \sigma^2)$$

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

$$\overline{X}_n \sim N(\mu, \sigma^2/n)$$ ✓

"HARD" EXAMPLE

$$X_1, X_2, \ldots, X_n \sim \text{EXP}(\lambda)$$

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

$$S_n^2 \sim \, ?$$

ALSO MEDIAN, QUANTILE, ETC

Simulation

REPEAT B TIMES:

1. * SIMULATE DATA $X_1^{(i)}, X_2^{(i)}, \ldots, X_n^{(i)} \sim F$
2. CALCULATE STATISTIC $T_n^{(i)} = g\left(X_1^{(i)}, X_2^{(i)}, \ldots, X_n^{(i)}\right)$

$$\longrightarrow \underbrace{T_n^{(1)}, T_n^{(2)}, T_n^{(3)}, T_n^{(4)}, \ldots T_n^{(B)}}$$

USE EMPIRICAL DISTRIBUTION AS ESTIMATE OF SAMPLING DISTRIBUTION

$$\hat{F}_{T_n} \xrightarrow{\text{ESTIMATES}} F_{T_n}$$

* MAGIC

$$F(x)$$ ← Population distribution

"magic"

Random sample of size n from F

Repeat B times

$$X^{(1)} = \left(X_1^{(1)}, X_2^{(1)}, \ldots, X_n^{(1)}\right) \qquad X^{(2)} = \left(X_1^{(2)}, X_2^{(2)}, \ldots X_n^{(2)}\right) \quad \cdots \quad X^{(B)} = \left(X_1^{(B)}, X_2^{(B)}, \ldots, X_n^{(B)}\right)$$

Simulated sample

$$T^{(1)} = g\left(X^{(1)}\right) \qquad T^{(2)} = g\left(X^{(1)}\right) \quad \cdots \quad T^{(B)} = g\left(X^{(B)}\right)$$

Value of statistic for first simulated sample

WHY DOES THIS WORK?

RECALL GLIVENKO-CANTELLI THM

$X_1, X_2, \ldots X_n \sim F$ THEN

$T_n^{(1)}, \ldots\ldots T_n^{(B)} \sim F_{T_n}$

$$\sup_x \left| \underset{\underset{\hat{F}_{T_n}}{\uparrow}}{\hat{F}(x)} - \underset{\underset{F_{T_n}}{\uparrow}}{F(x)} \right| \xrightarrow{P} 0$$

IN R

- WRITE FUNCTION THAT
  - GENERATES DATA
  - CALCULATES STATISTIC
- REPLICATE!

Examples

- $X_1, X_2, \ldots X_n \sim N(\mu = 3, \sigma^2 = 4)$, $n = 20$, $T_n = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

- $X_1, X_2, \ldots X_n \sim Pois(\lambda = 4.2)$, $n = 13$, $T_n = S_n^2 = \frac{1}{n-1}\sum(X_i - \overline{X})^2$

- $X_1, X_2, \ldots X_n \sim Exp(\beta = 3.1)$, $n = 17$, $T_n =$ "SAMPLE MEDIAN"