

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

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

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



STATISTIC

• WHAT IS THE DISTRIBUTION OF T_n ?

• $E[T_n] = ?$

• $V[T_n] = ?$

• ETC

EASY EXAMPLE

$$X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$$

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

$$\bar{X}_n \sim \mathcal{N}(\mu, \sigma^2/n) \quad \checkmark$$

"HARD" EXAMPLE

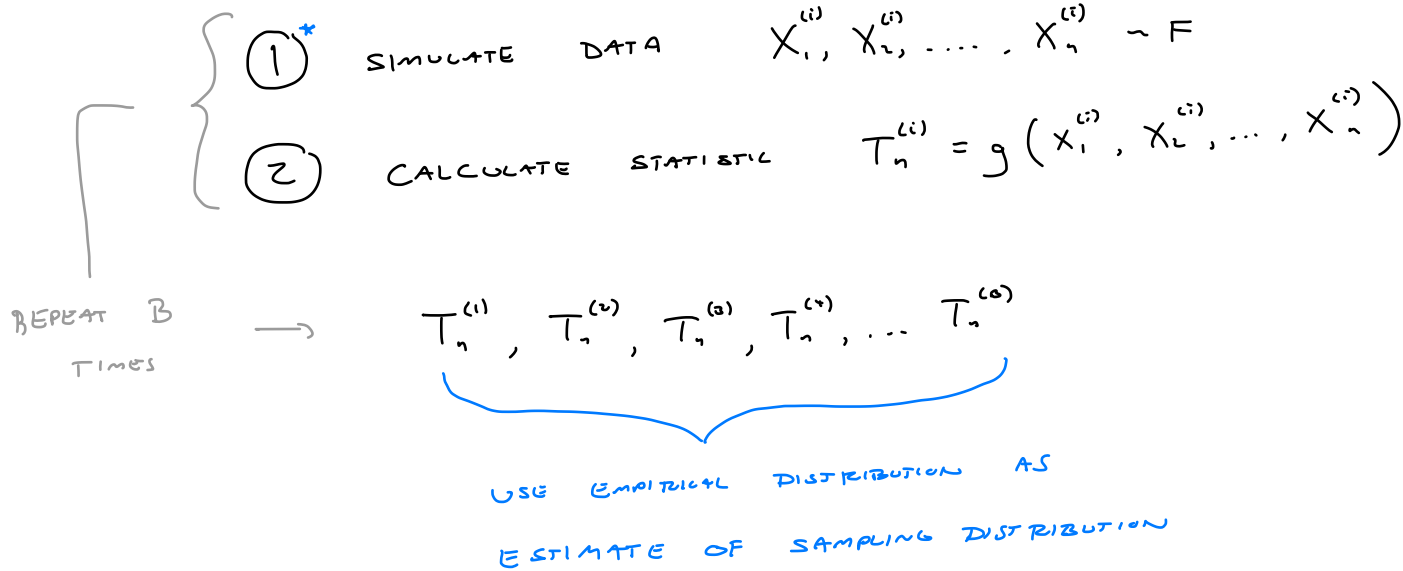
$$X_1, X_2, \dots, X_n \sim \text{EXP}(\lambda)$$

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

$$S_n^2 \sim ?$$

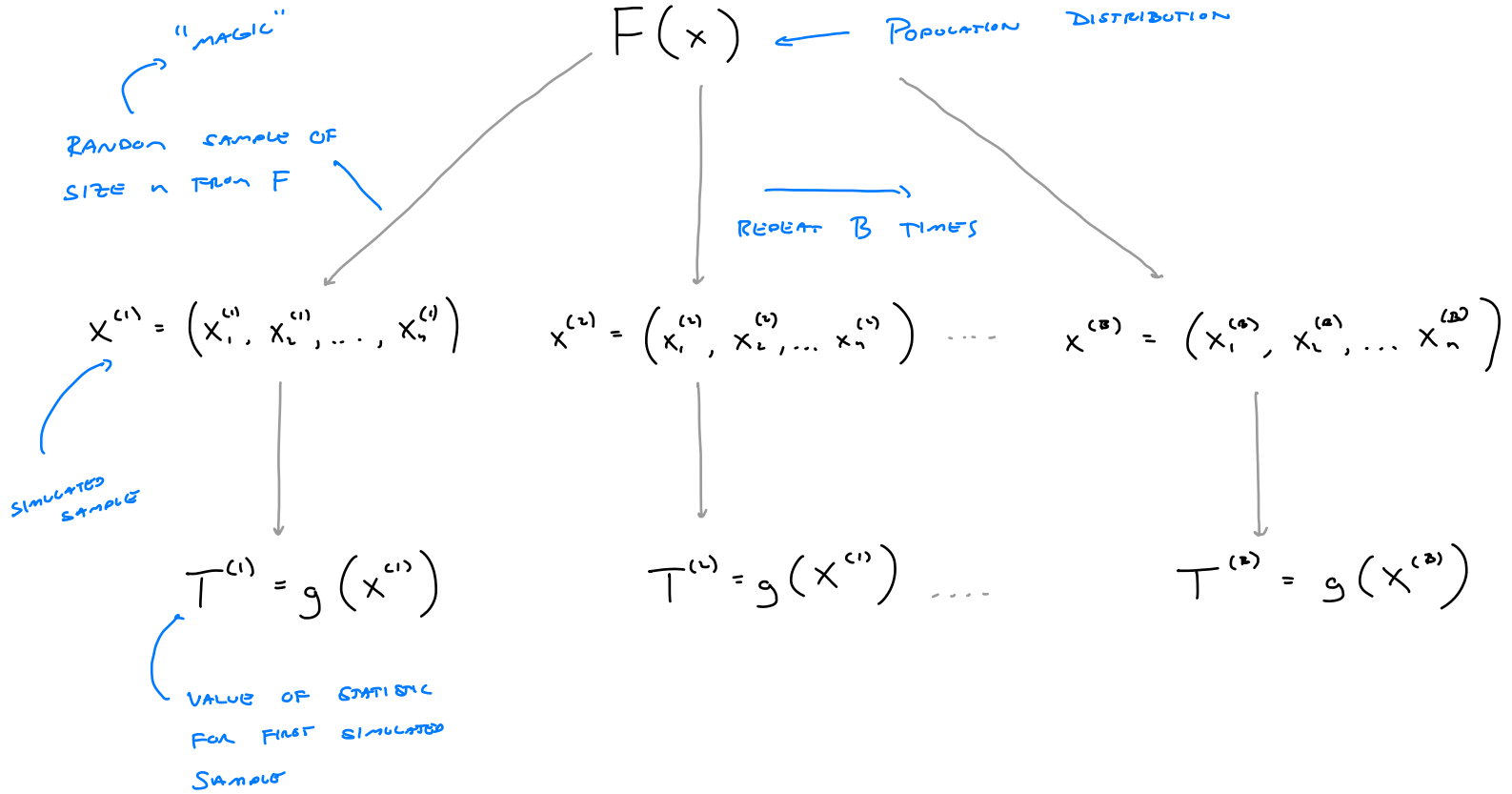
ALSO MEDIAN, QUANTILE, ETC

SIMULATION



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

* MAGIC



WHY DOES THIS WORK?

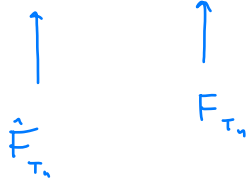
RECALL GLIVENKO-CANTELLI THM

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

THEN

$$T_n^{(1)}, \dots, T_n^{(n)} \sim F_{T_n}$$

$$\sup_x \left| \hat{F}(x) - F(x) \right| \xrightarrow{P} 0$$



IN R

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

EXAMPLES

- $X_1, X_2, \dots, X_n \sim N(\mu=3, \sigma^2=4)$, $n=20$, $T_n = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$
- $X_1, X_2, \dots, X_n \sim \text{Pois}(\lambda=4.2)$, $n=13$, $T_n = S_n^2 = \frac{1}{n-1} \sum (X_i - \bar{x})^2$
- $X_1, X_2, \dots, X_n \sim \text{EXP}(\beta=3.1)$, $n=17$, $T_n = \text{"same as before"}$