Monte Carlo Simulation

Stat 432

Spring 2020

Dalpiaz

Setup

Statistics for random samples of size n

Population distribution

$X_1, X_2, \ldots, X_n \sim \underbrace{p(x \mid \theta)}$ Random Variables

$x_1, x_2, \ldots, x_n$ (Potential) Realized values

$2.1, 1.3, \ldots, 3.4$ Realized values

$E[X]$ $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ $\bar{x} = \frac{1}{10}(2.1 + 1.3 + \ldots + 3.4) = 2.6$

Parameter

Estimator

Statistic

Random Variable

How to calculate estimate with data

Estimate for a particular dataset

FUNCTION OF DISTRIBUTION

PARAMETERS

$\mu = E[X]$

$m : P[X > m] = 0.5$

$P[X > 8]$

$\sigma^2 = \text{Var}[X]$

FUNCTION OF SAMPLE DATA

ESTIMATORS

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

$\text{median}(x_1, x_2, \ldots, x_n)$

$\hat{P}[X > 8] = \frac{1}{n} \sum_{i=1}^{n} I(x_i > 8)$

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

SAMPLING DIST

?

?

?

?

LOTS OF MATH

Example

$X_1, X_2, \ldots, X_{25} \sim N(\mu = 5, \sigma^2 = 9)$

Parameters	Estimators		Dist?
$\mu = E[X]$	$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$	$\longrightarrow$	$N(\mu = 5, \sigma^2 = 9/25)$
$m : P[X > m] = 0.5$	$\text{median}(X_1, X_2, \ldots, X_n)$	$\longrightarrow$	Something with order statistics?
$P[X > 8]$	$\hat{P}[X > 8] = \frac{1}{n}\sum_{i=1}^{n} I(X_i > 8)$	$\longrightarrow$	Sum of Bernoullis?
$\sigma^2 = \text{Var}[X]$	$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^2$	$\longrightarrow$	Something w/ χ^2?

New Idea → Monte Carlo Simulation

Repeat many times

$$\left\{ \begin{array}{l} \bullet \text{ Generate sample of size } n \text{ from } p(x \mid \theta) \\ \qquad x^{(i)} = (x_1^{(i)}, x_2^{(i)}, \ldots, x_n^{(i)}) \\ \bullet \text{ Calculate statistic of interest, } S(x^{(i)}) \end{array} \right.$$

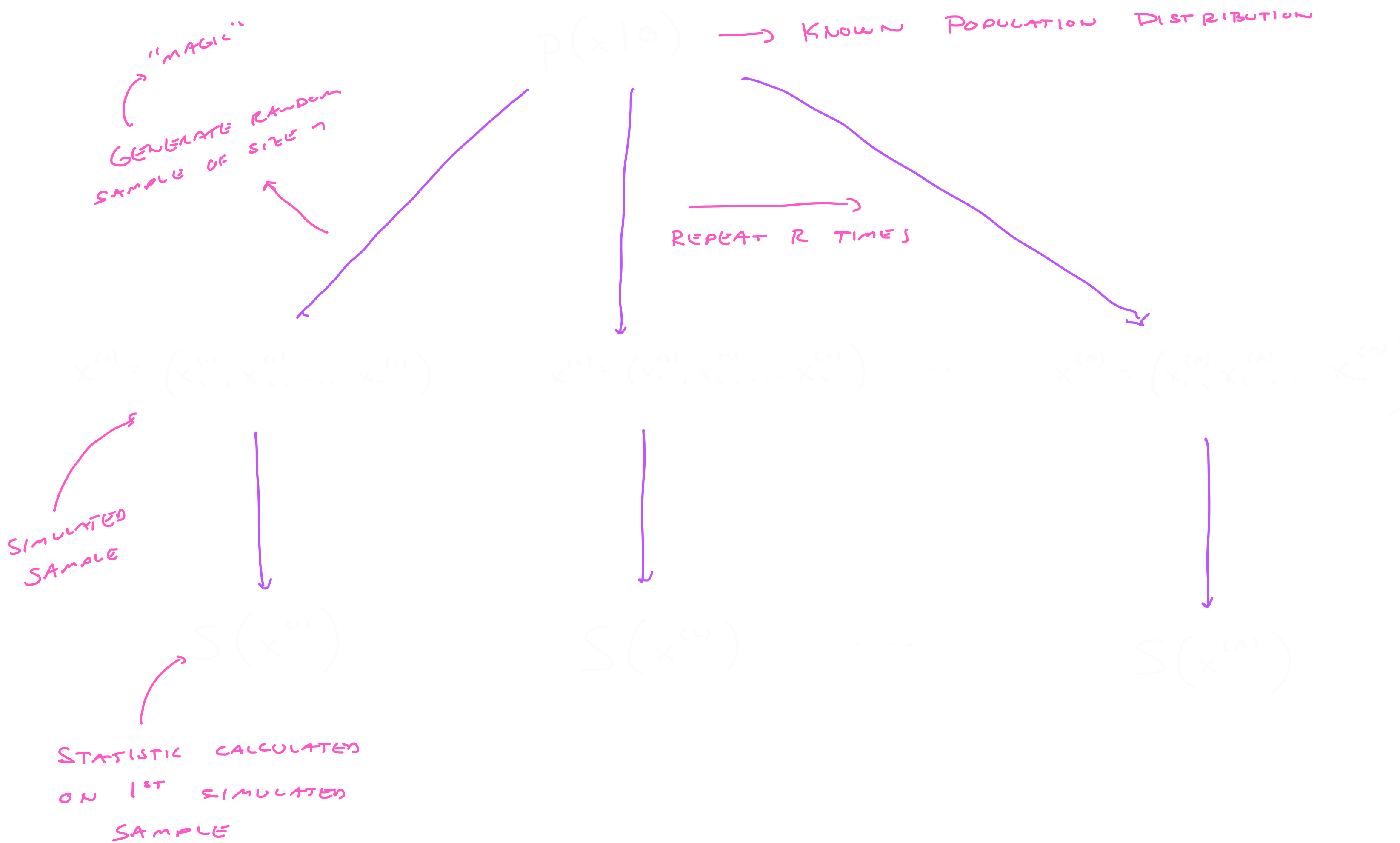
→ R magic!

→ Estimator

R^{th} simulated statistic

$$\underbrace{S(x^{(1)}),\ S(x^{(2)}),\ \ldots,\ S(x^{(R)})}$$

Use empirical distribution to estimate true distribution



KNOWN POPULATION DISTRIBUTION
"MAGIC"
GENERATE RANDOM SAMPLE OF SIZE n
REPEAT R TIMES
SIMULATED SAMPLE
STATISTIC CALCULATED ON 1ST SIMULATED SAMPLE

- See examples in R → Normal example
  → Exponential example
- Why?
  - Math is hard
  - Helps explain bootstrap