Monte Carlo Simulation

Stat 432

Spring 2020

Dalpiaz

Setup

STATISTICS FOR RANDOM SAMPLES OF SIZE n

POPULATION DISTRIBUTION

$$X_1, X_2, \ldots, X_n \sim p(x \mid \theta)$$

RANDOM VARIABLES

$x_1, x_2, \ldots, x_n$ (POTENTIAL) REALIZED VALUES

$2.1, 1.3, \ldots, 3.4$ REALIZED VALUES

$E[X]$

PARAMETER

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

ESTIMATOR

STATISTIC

RANDOM VARIABLE*

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

HOW TO CALCULATE ESTIMATE WITH DATA

$$\bar{x} = \frac{1}{10}(2.1 + 1.3 + \ldots + 3.4) = 2.6$$

ESTIMATE FOR A PARTICULAR DATASET

FUNCTION OF DISTRIBUTION

PARAMETERS

$\mu = E[X]$

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

$P[X > 8]$

$\sigma^2 = Var[X]$

FUNCTION OF SAMPLE DATA

ESTIMATORS

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

$median(x_1, x_2, \ldots x_n)$

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

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

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:

- Generate sample of size n from $p(x \mid \theta)$ → R magic!

$$x^{(i)} = (x_1^{(i)}, x_2^{(i)}, \ldots\ldots, x_n^{(i)})$$

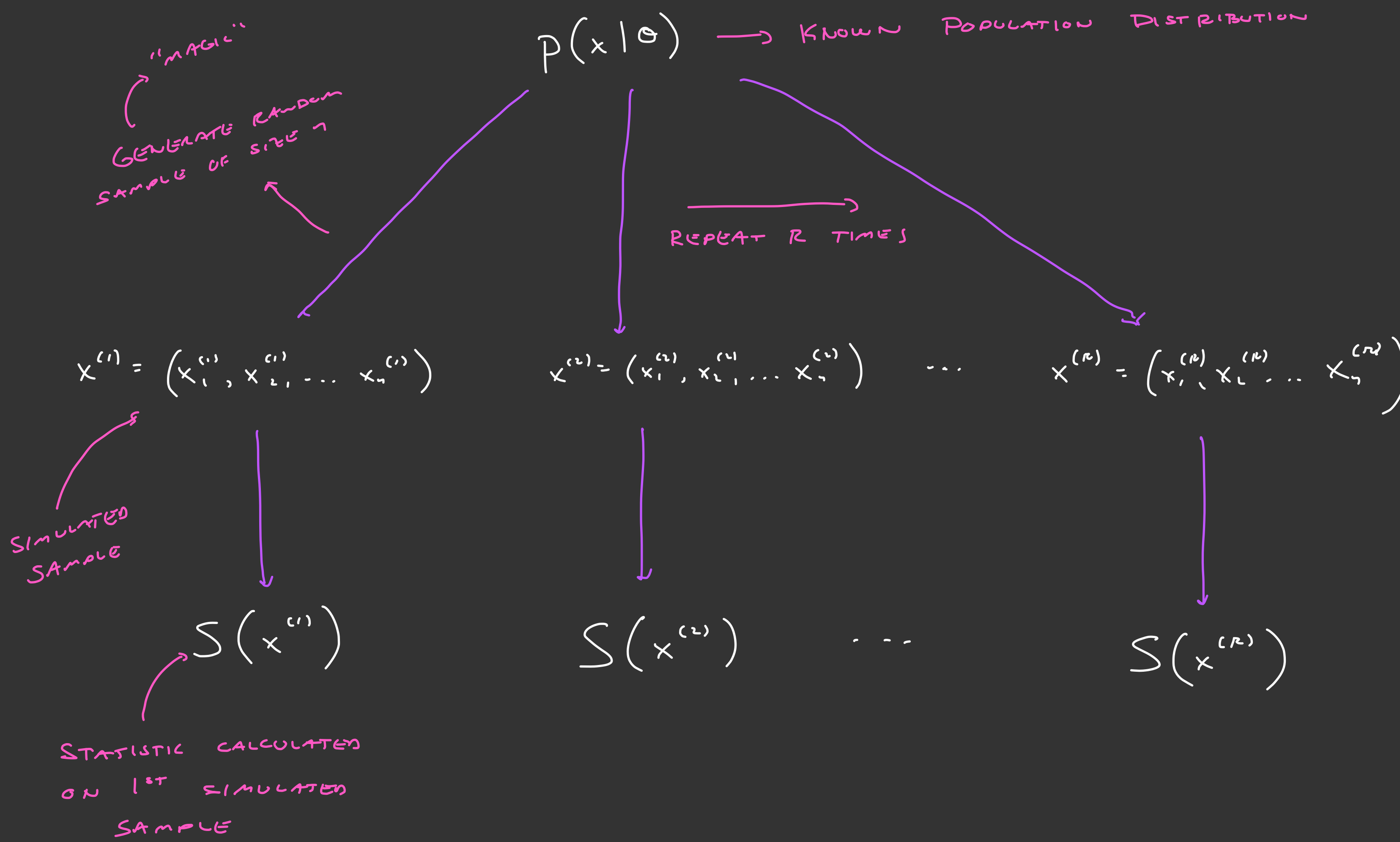
- Calculate statistic of interest, $S(x^{(i)})$

↳ estimator

R^{th} simulated statistic

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

Use empirical distribution to estimate true distribution



$p(x|\theta)$
KNOWN POPULATION DISTRIBUTION
"MAGIC"
GENERATE RANDOM SAMPLE OF SIZE n
REPEAT R TIMES
$x^{(1)} = (x_1^{(1)}, x_2^{(1)}, \ldots x_n^{(1)})$
$x^{(2)} = (x_1^{(2)}, x_2^{(2)}, \ldots x_n^{(2)})$
$x^{(R)} = (x_1^{(R)}, x_2^{(R)} \ldots x_n^{(R)})$
SIMULATED SAMPLE
$S(x^{(1)})$
$S(x^{(2)})$
$S(x^{(R)})$
STATISTIC CALCULATED ON 1ST SIMULATED SAMPLE

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