Monte Carlo Simulation

Stat 432

Spring 2020

Dalpiaz

Setup

Statistics for random samples of size n

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

Population distribution

Random Variables

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

2.1, 1.3, ..., 3.4 Realized values

$E[X]$

Parameter

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

Estimator

Statistic

Random Variable*

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

How to calculate estimate with data

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

Estimate for a particular dataset

Function of distribution

Function of sample data

Parameters	Estimators	Sampling Dist
$\mu = E[X]$	$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$	?
$m : P[X > m] = 0.5$	$\text{median}(x_1, x_2, \ldots x_n)$	?
$P[X > 8]$	$\hat{p}[x > 8] = \frac{1}{n}\sum_{i=1}^{n} I(x_i > 8)$	?
$\sigma^2 = \text{Var}[X]$	$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2$	?

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 $p(x|\theta)$

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

→ R magic!

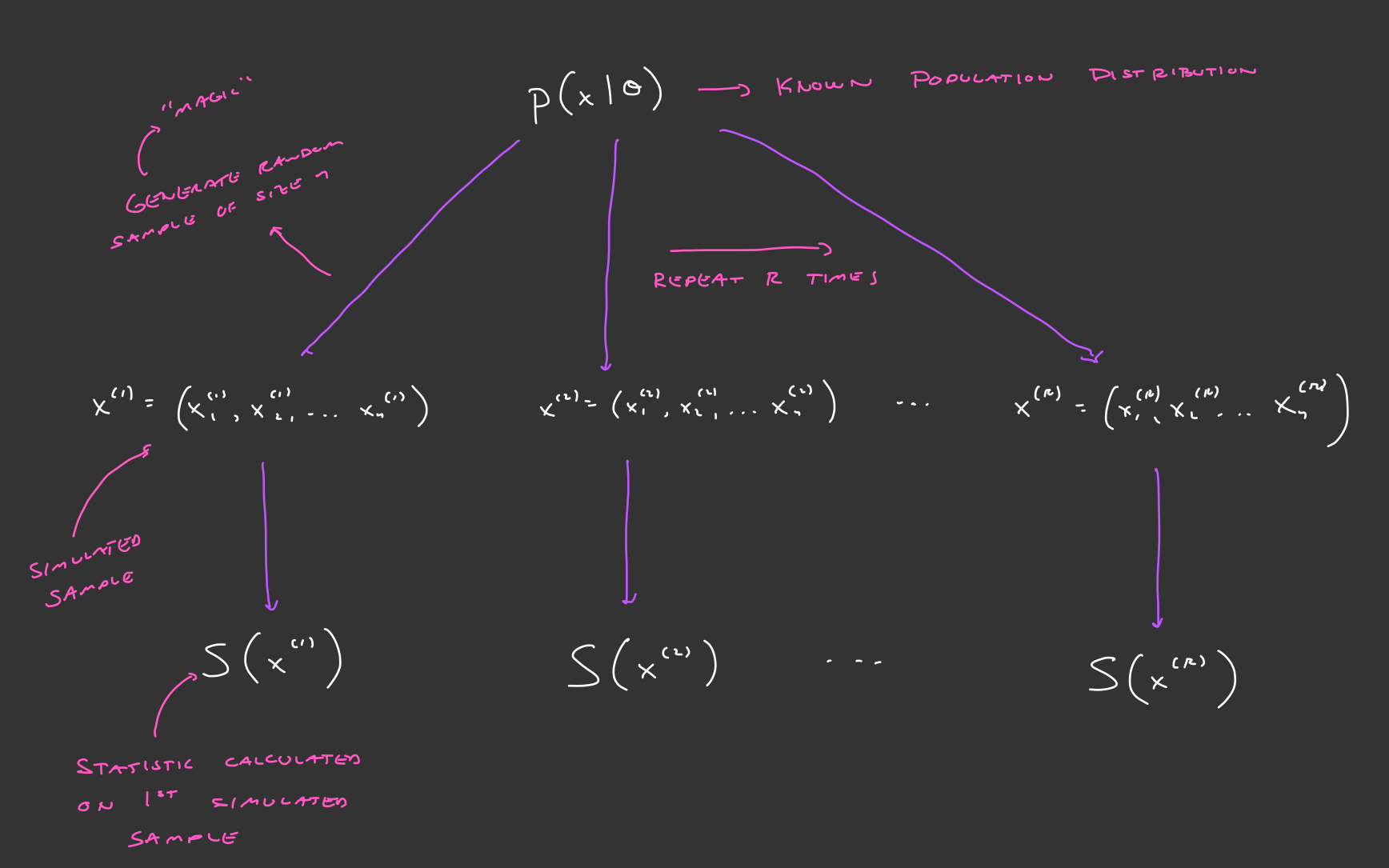
- Calculate statistic of interest, $S(x^{(i)})$

↳ estimator

R^{th} simulated statistic

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

Use empirical distribution to estimate true distribution


"MAGIC"
GENERATE RANDOM SAMPLE OF SIZE n
p(x|θ) → KNOWN POPULATION DISTRIBUTION
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