Lesson Plan 3/29

Thursday, March 29, 2018 3:06 PM

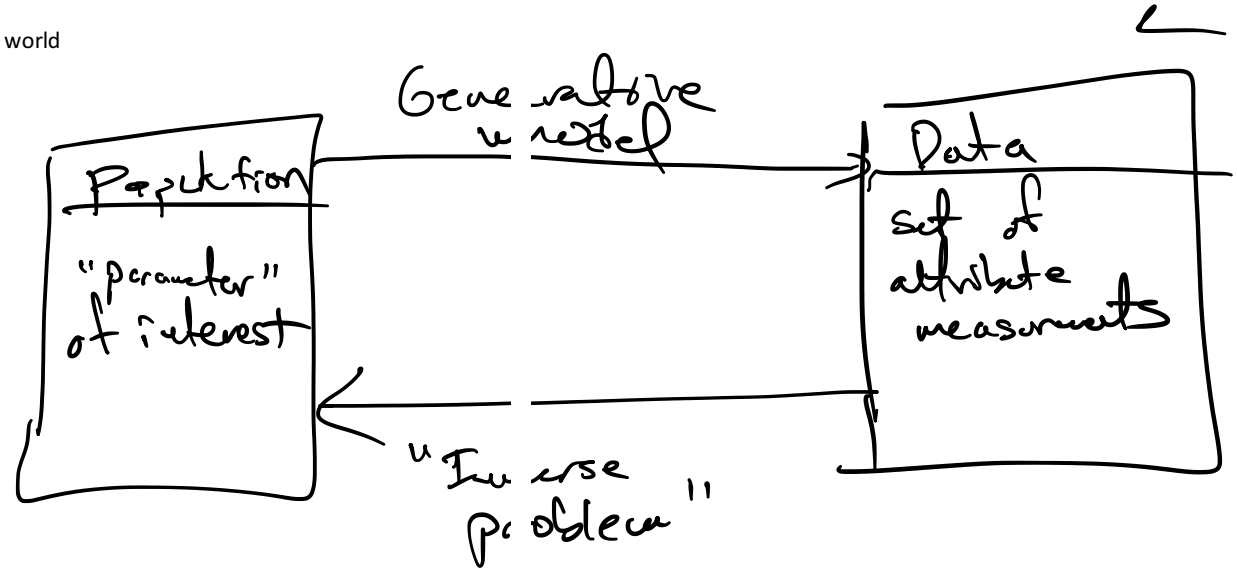
Admin
- Reminder: we switched midterm 2 to 4/24

HDSC
- About this Cambridge Analyitica Mess:
  - https://arstechnica.com/tech-policy/2018/03/facebooks-cambridge-analytica-sca ndal-explained/
  - https://motherboard.vice.com/en_us/article/mg9vvn/how-our-likes-helped-trun p-win
  - http://www.michalkosinski.com/
    - Psychological targeting as an effective approach to digital mass persuasion
    - Facebook as a Research Tool for the Social Sciences

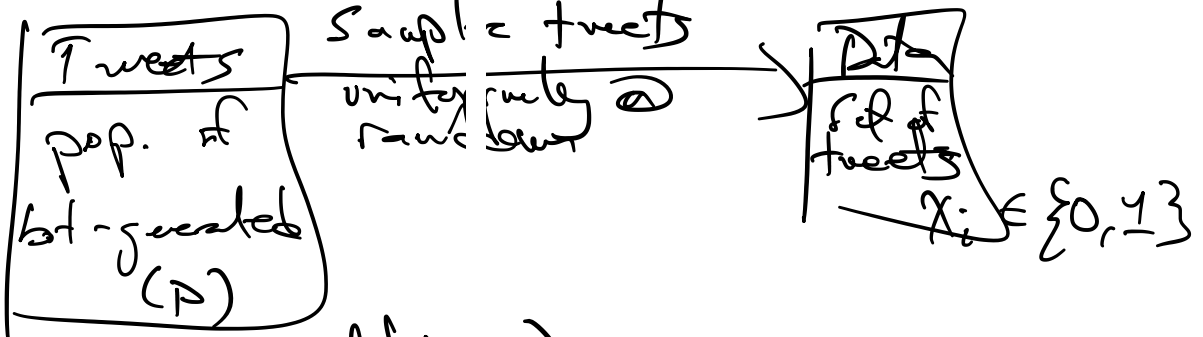
Project 2
- Reminder: Due 4/6
- Questions, clarifications?

Hypothesis Testing
- The "inverse problem" view of the world
- Null and alternative
- P-value

Multivariate Probability
- Joint Probability
- Conditional Probability
- Bayes' Rule



"Inverse Problem"



Inverse Problem (moment matching)

Decide EX under generative model as function of p

Then set it equal to $\frac{1}{n}\sum x_i$

solve for p

$$\Rightarrow \quad \hat{p} = \frac{1}{n} \sum_i x_i$$

What is the precision of your estimate?

Central Limit Theorem

$$\bar{X} \longrightarrow N\left(Ex, \frac{\sqrt{var(x)}}{\sqrt{n}}\right) \Longleftarrow \text{normal distribution}$$

In our case

$$\hat{p} \sim N\left(p, \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right)$$

by CLT

Hypothesis Testing

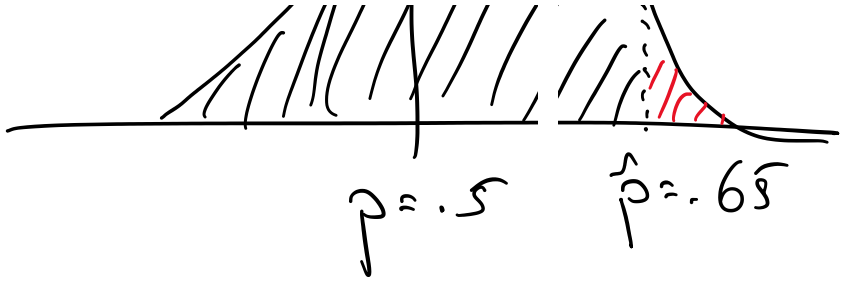
How to use "Inverse Problem" procedure to test hypothesis

$p = .5$

① Estimate $\hat{p}$ (for example $\hat{p} = .65$)

② What can we say about $\hat{p}$, under my hypothesis?

$\Longleftarrow N\left(p, \sqrt{\frac{p(1-p)}{}}\right)$



On /

(3) Reject Hypothesis or not.

based on probability of $\hat{p}$

$$1 - P\left(\hat{p};\ p \cdot \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) \leq 5\%$$

then reject this hypothesis

P- value :

Null Hypothesis

Alternative

p-value

-> A/B (a priori PA)

Had a priori hypothesis

$P_A = p$ ⟸ - Known

test

$p_B = p_A$

$\rightarrow$ A/B (p_A is estimated)

test

$p_B = p_A$

$\longrightarrow$ Bootstrap

- Use data to also estimate sampling distribution of p

Inference: Estimate some property by using "inverse problem" and make statements about their precision

Testing: Test a priori hypotheses based on inference

Probability Distributions:

→ Bernoulli:

D: $\{0, 1\}$

$X \sim Bernoulli(p)$

p: $P(X = 1)$

EX: p

varX: $p(1-p)$

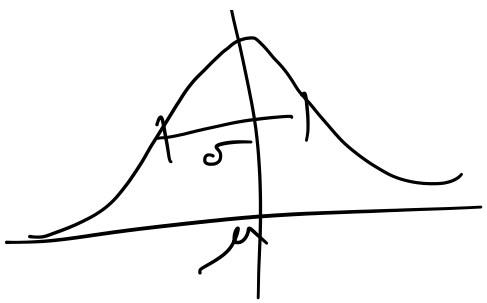
→ Binomial (sum of n trials of Bernoulli)

$X \sim Bin(n, p)$

D: $X = 1, 2, \ldots, n$

EX: np

varX: $np(1-p)$



→ Normal

$X \sim N(\mu, \sigma)$

$p \& X \in \mathbb{R}$

$EX = \mu$

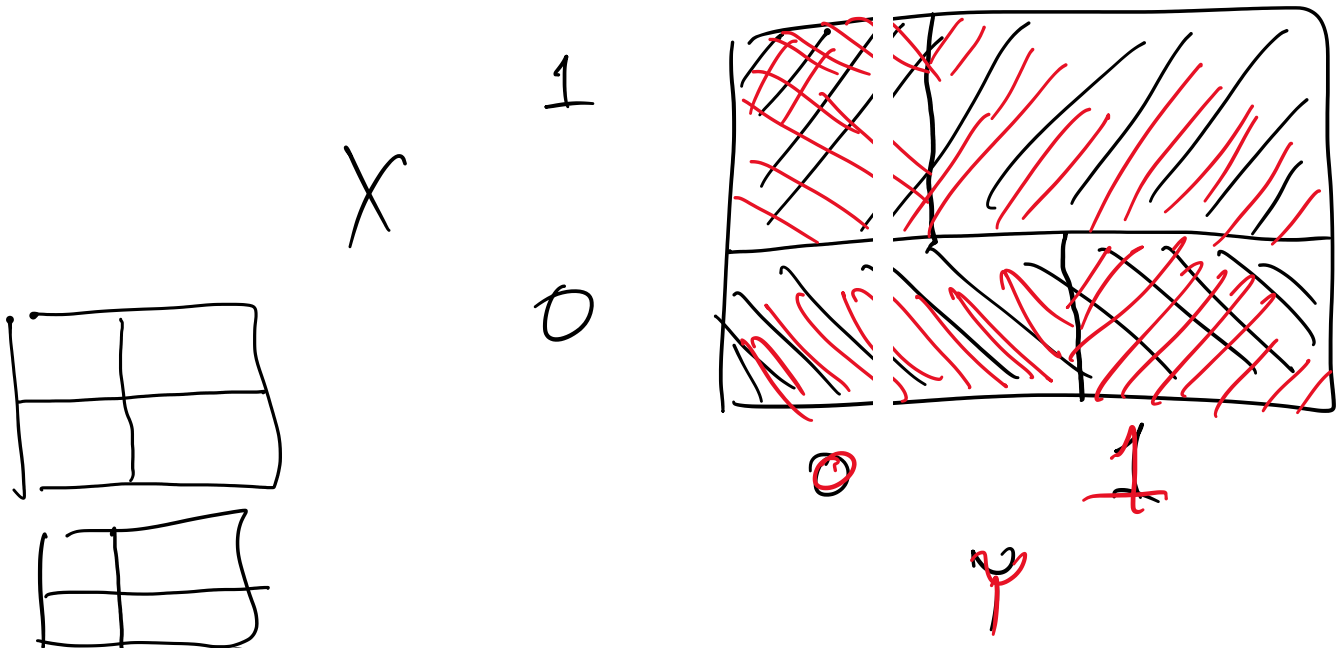
$var = \sigma^2$

→ Joint & Conditional probability

$X \in \{0, 1\}$: bof-generated

V. Soft. Lots of

$p \in [0,1]$ - retweets



1) $p(X=x, Y=y) \geq 0 \quad \forall \ (x,y) \in D_X \times D_Y$

2) $\sum_{(x,y) \in D_X \times D_Y} p(X=x, Y=y) = 1$

Indipendenza

Conditional Probability

$P(Y=y \mid X=1)$

Y ⊥ X

if $P(Y=y \mid X=x)$ is the same $\forall x$

Bayes Rule