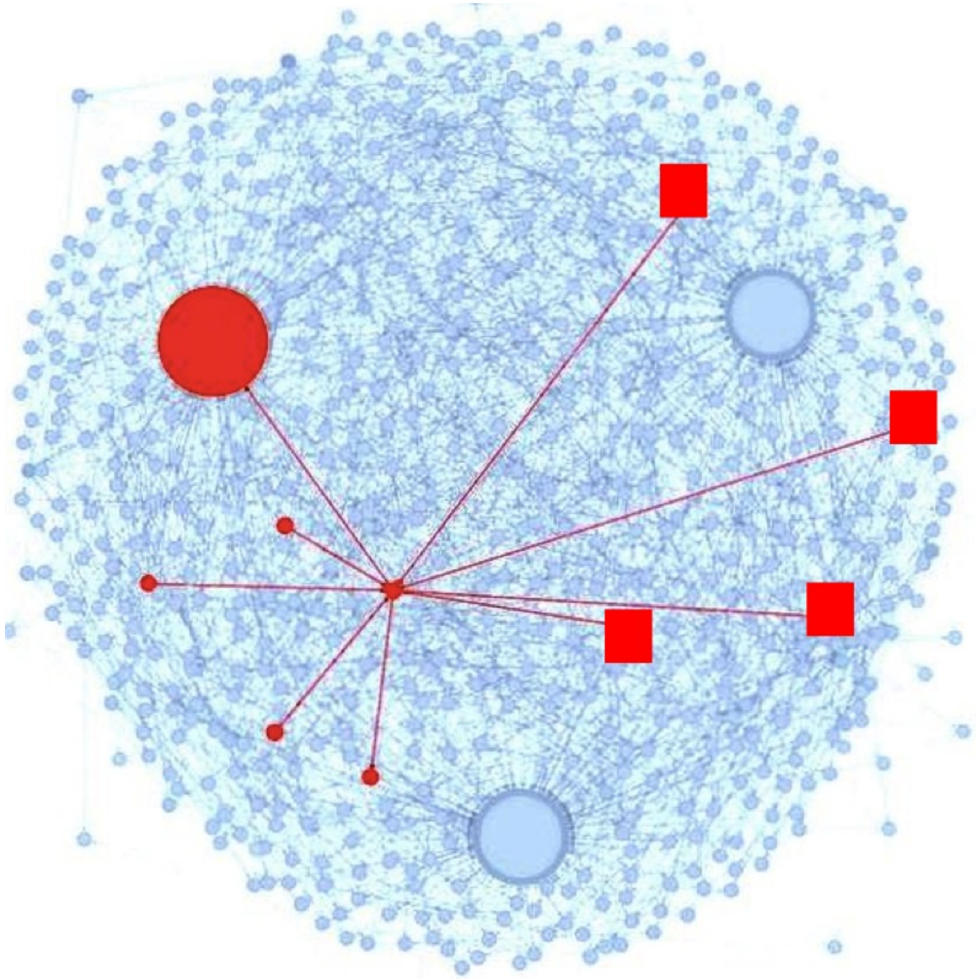
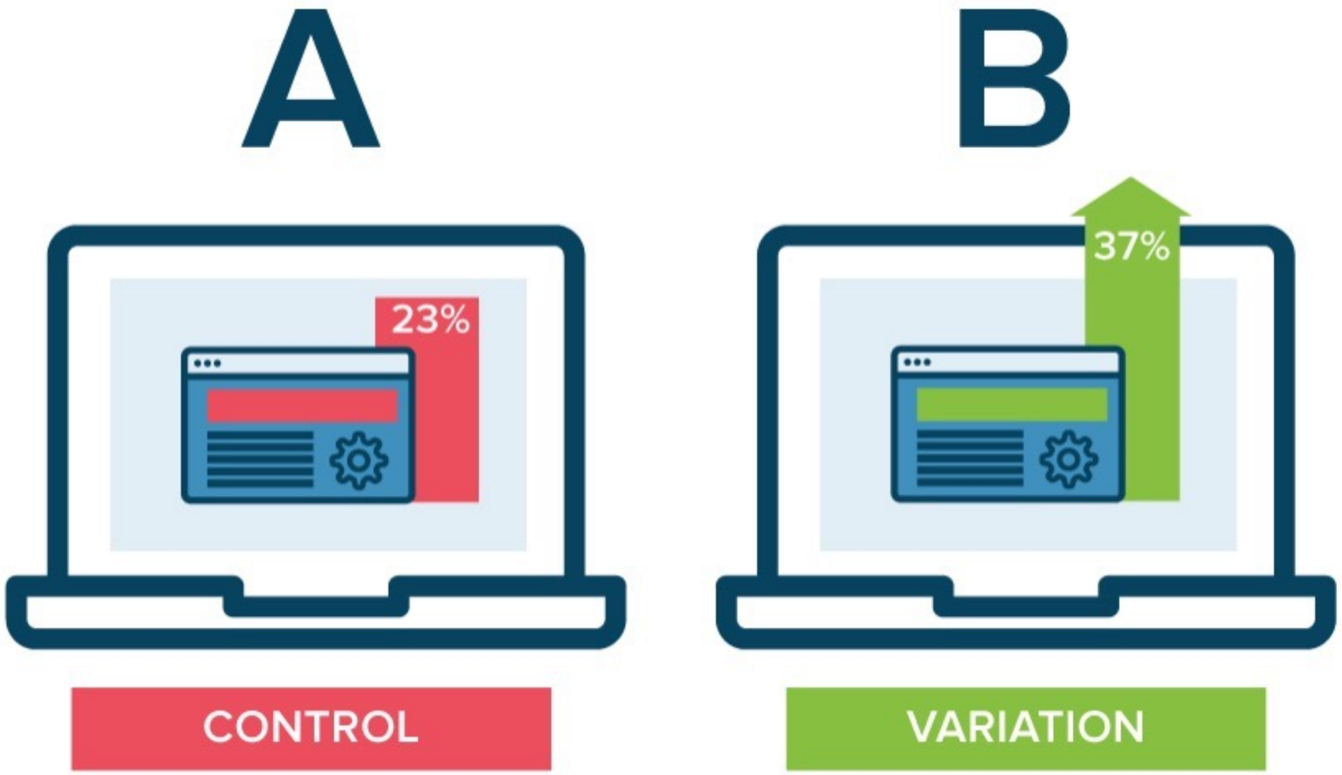
Algorithmic Analysis

Chris Piech
CS109, Stanford University

Pset #5 is out





Pset 5 - Uncertainty Theory

cs109psets.netlify.app/win22/pset5/code_analysis_flu

PS5

1
2
3
4
5
6
7
8
9

Llama Flu

Heads up: This problem uses material that we are going to cover on Feb 23rd

Our ability to fight contagious diseases depends on our ability to model them. One person is exposed to llama-flu (a made up disease). The method below returns the number of individuals who will get infected.

```
def num_infected():
    """
    Returns the number of people infected by one individual
    """

    # most people are immune to llama-flu
    immune = bernoulli(p = 0.99)
    if immune: return 0

    # people who are not immune spread the disease far
    spread = 0

    # they make contact with k people (up to 100)
    k = binomial(n = 100, p = 0.25)
    for i in range(k)
      spread += num_infected():

    # total infections should include this individual
    return spread + 1
```

What is the expected return value of \code{num_infected}?

Previous Question

Next Question

Answer Editor

Solution

Numeric Answer: Enter your answer

Check Answer

Explanation:

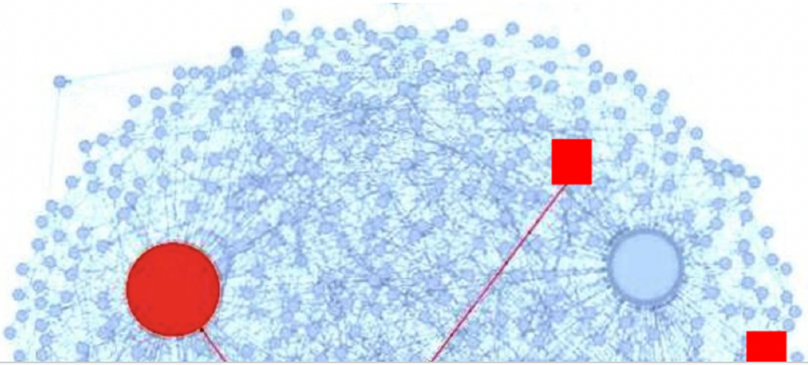
Block LaTeX Image

Better Peer Grading

Stanford's HCI class runs a massive online class that was taken by ten thousand students. The class used peer assessment to evaluate student's work. We are going to use their data to learn more about peer graders. In the class, each student has their work evaluated by 5 peers and every student is asked to evaluate 6 assignments: five peers and the control assignment (the graders were un-aware of which assignment was the control). All 10,000 students evaluated the same control assignment and the scores they gave are in the file peerGrades.csv in the pset5 data zip:

pset5.zip

Would you use the **mean** or the **median** of 5 peer grades to assign scores in the online version of Stanford's HCI class? Explain why. You may use simulations to solve any part of this question. Hint: it might help to visualize the scores in peerGrades.csv. In order to make your decision compute the statistics in part a) and b).



Answer Editor
Solution

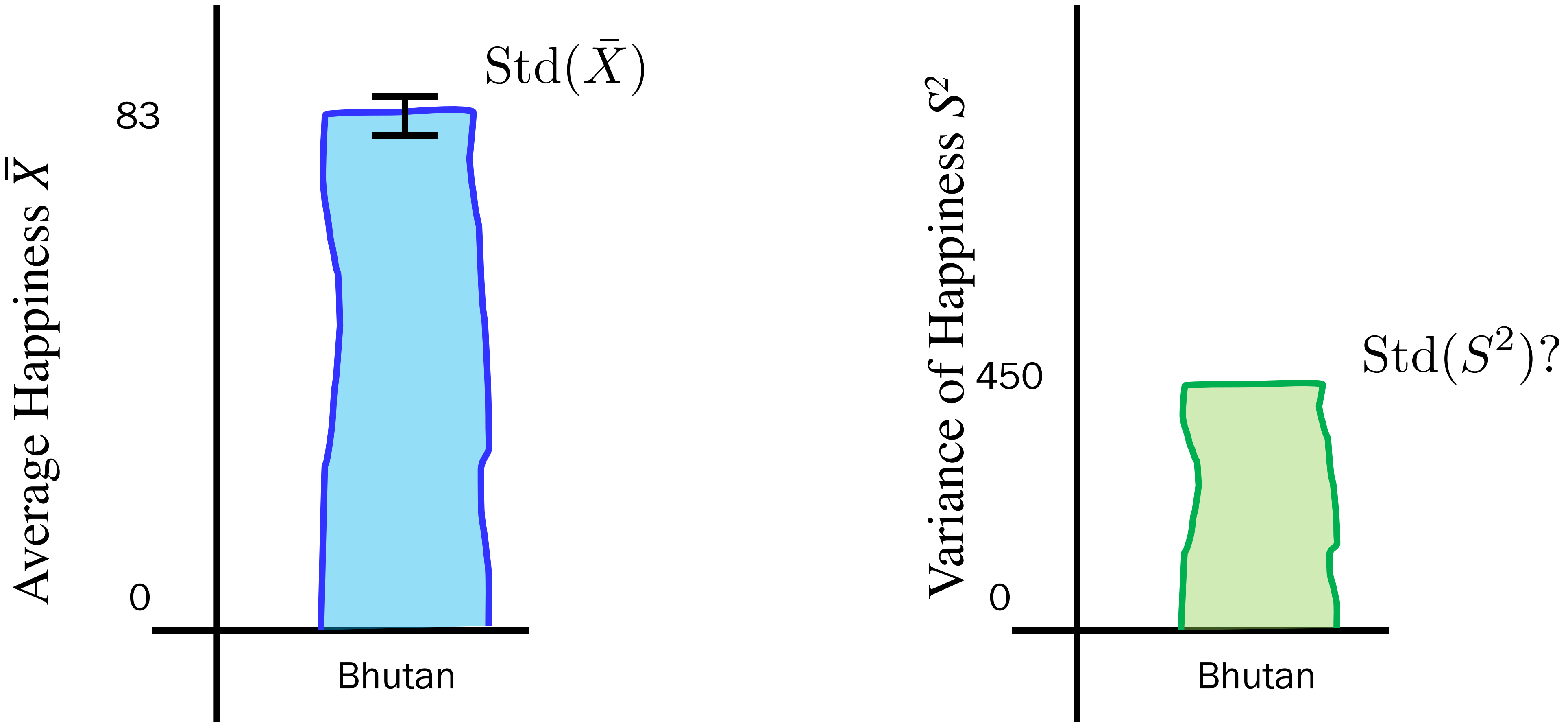
Numeric Answer: Enter your answer
Check Answer

Explanation:

Block LaTeX
Image
B
I
U

<review>

Our Report to Bhutan Government

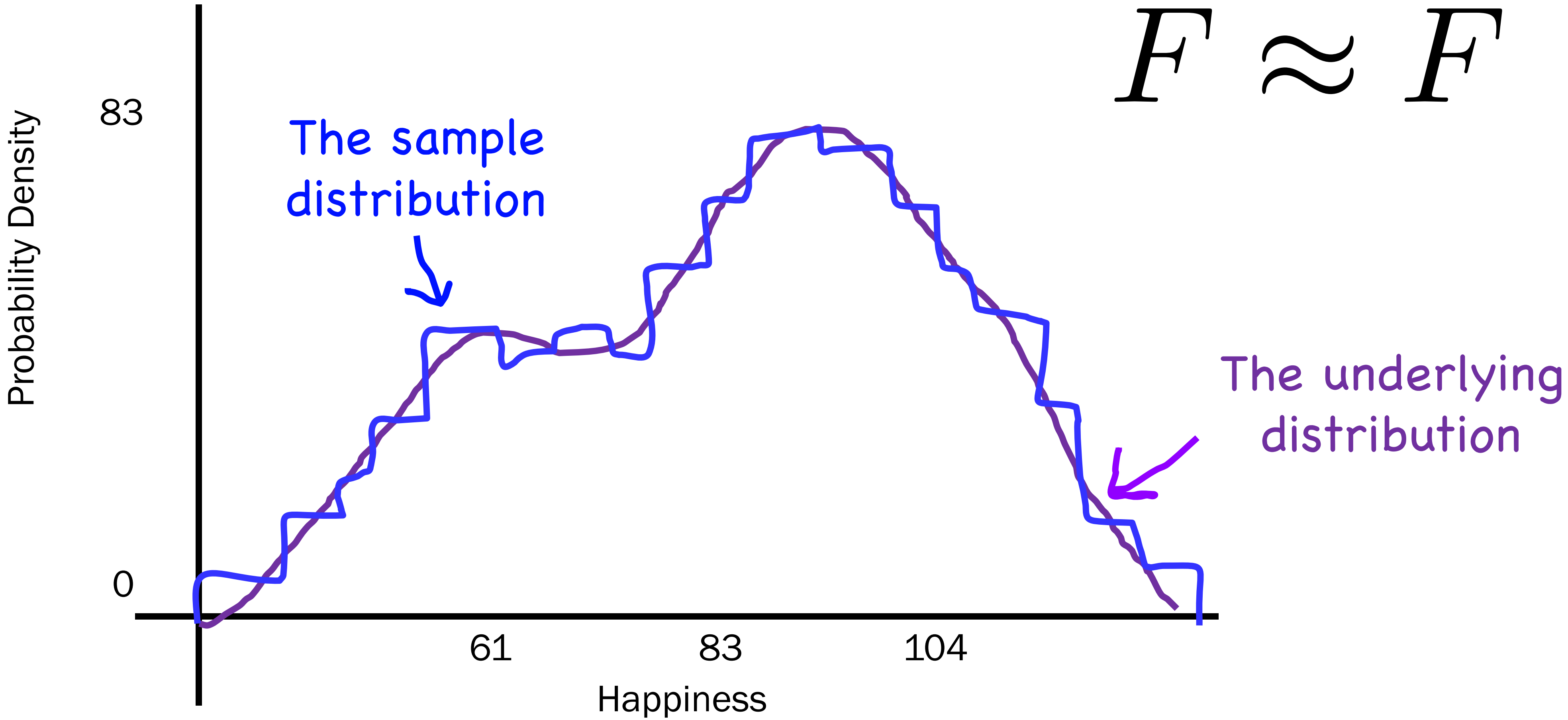


Claim: The average happiness of Bhutan is 83 ± 2

But Wait – What If You Actually Have a Good Estimate?

You can estimate the PMF of the underlying distribution, using your sample.*

$$F \approx \hat{F}$$

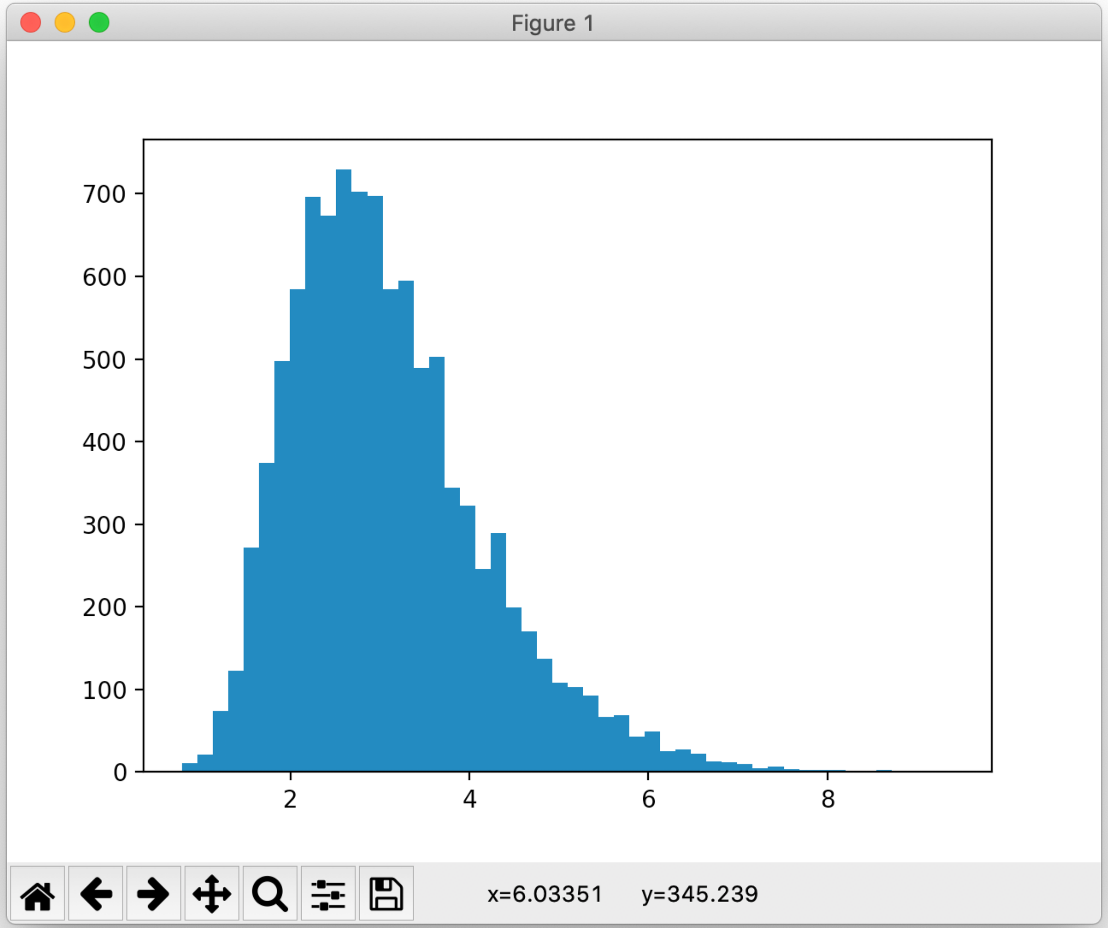


* This is just a histogram of your data!!

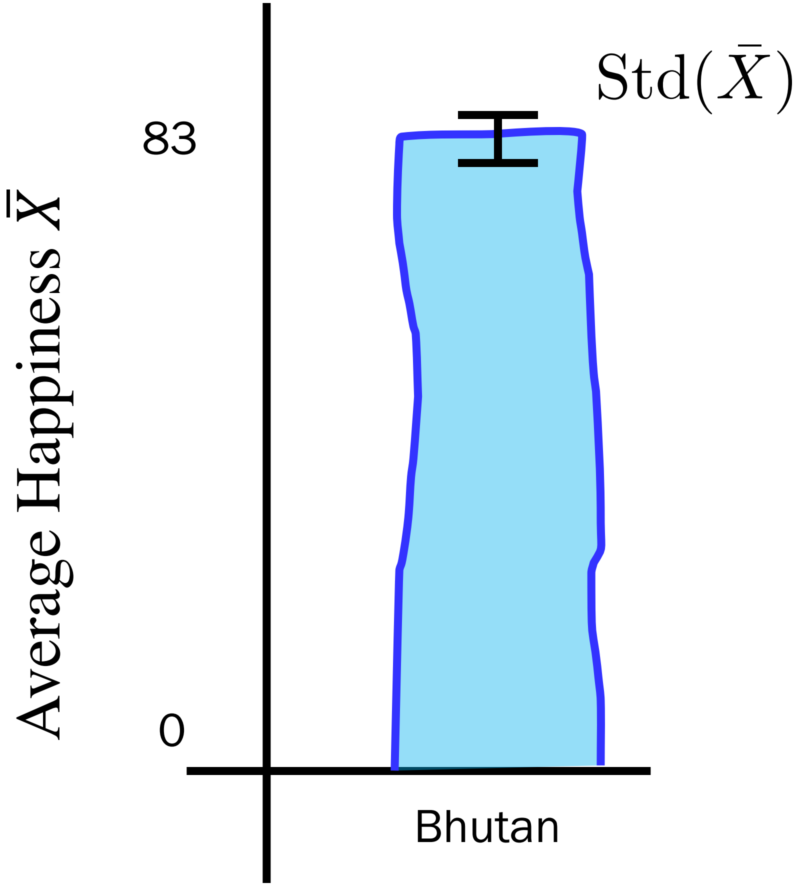
Bootstrapping in Practice

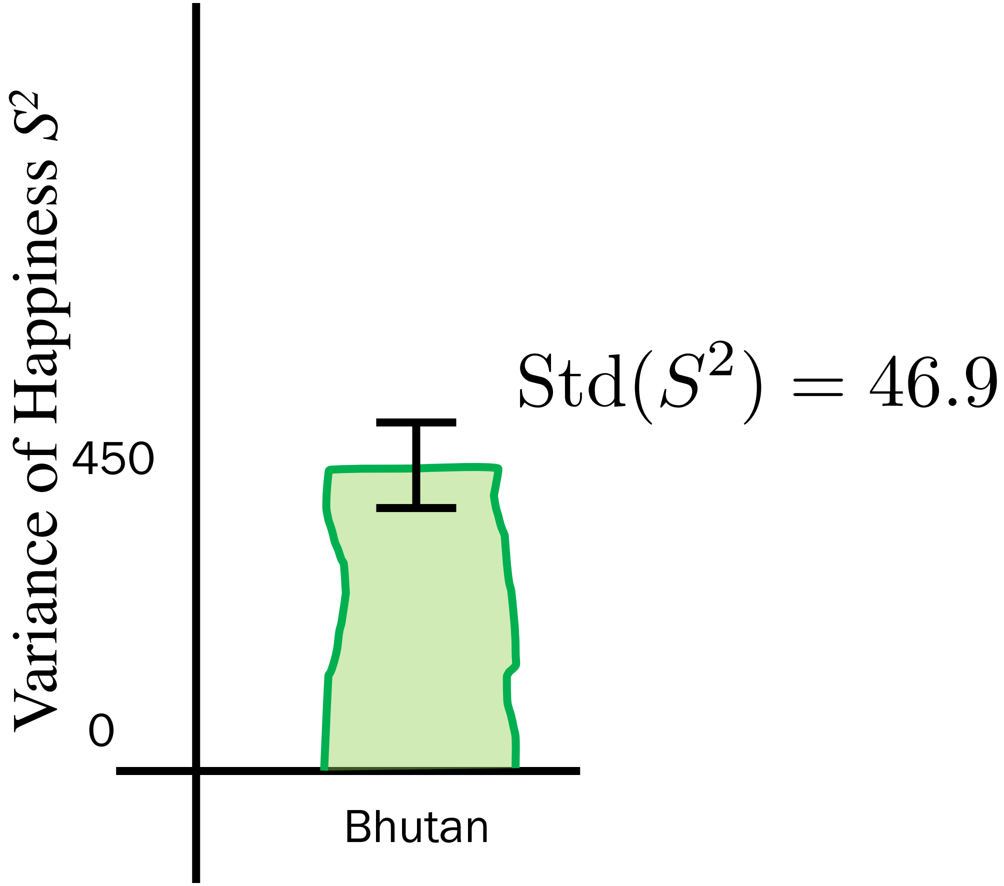
```
Bootstrap Algorithm (sample):
  1.   Repeat 10,000 times:
    a.  Choose len(sample) elems from sample, with replacement
    b.  Recalculate the stat on the resample
  2.   You now have a distribution of your stat
```

The Distribution of the Sampling Variance



Our Report to Bhutan Government





Claim: The average happiness of Bhutan is 83 ± 2

Bootstraping allows you to:

- Know the **distribution of *statistics***
- Calculate **p values**

A real difference?

18 students

Learning in Context A
4.44
3.36
5.87
2.31
…
3.70

$\mu_1 = 3.1$

Learning in Context B
2.15
3.01
2.02
1.43
…
1.83

23 students

$\mu_2 = 2.4$

Claim: Group 1 and Group 2 are samples from **different distributions** with a 0.7 difference of means.

How confident are you in this claim?

End Review

The Classic Science Test

Group 1
4.44
3.36
5.87
2.31
...
3.70

Group 2
2.15
3.01
2.02
1.43
...
1.83

$\mu_1 = 3.1$

$\mu_2 = 2.4$

Claim: Group 1 and Group 2 are samples from **different distributions** with a 0.7 difference of means.

How confident are you in this claim?

A real difference?

18 students

Learning in Context A
4.44
3.36
5.87
2.31
…
3.70

$\mu_1 = 3.1$

23 students

Learning in Context B
2.15
3.01
2.02
1.43
…
1.83

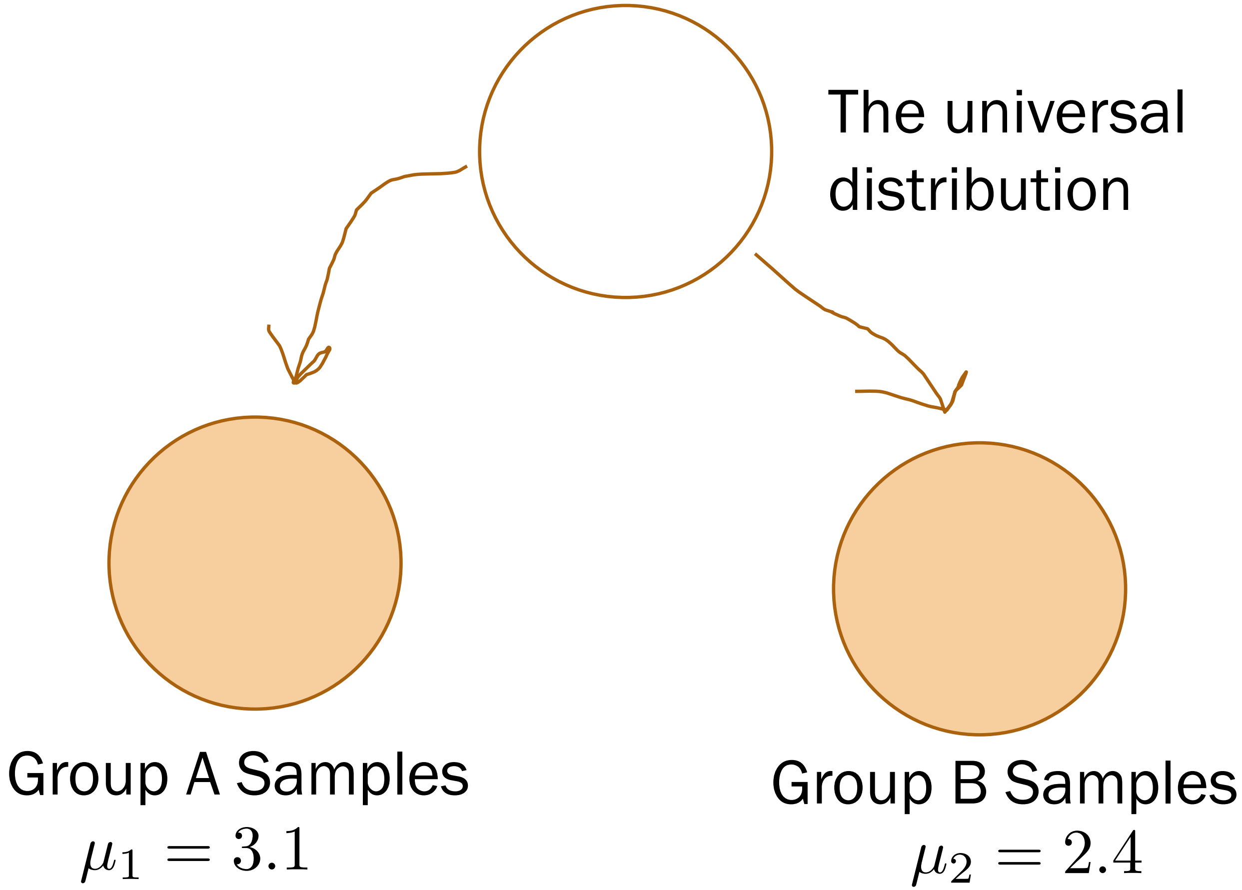
$\mu_2 = 2.4$

Claim: Group 1 and Group 2 are samples from **different distributions** with a 0.7 difference of means.

How confident are you in this claim?

The Null Hypothesis. How can we use bootstrapping here?

There is no difference between the two groups, so everyone is drawn from the same distribution. Any difference you observe is due to sampling error.

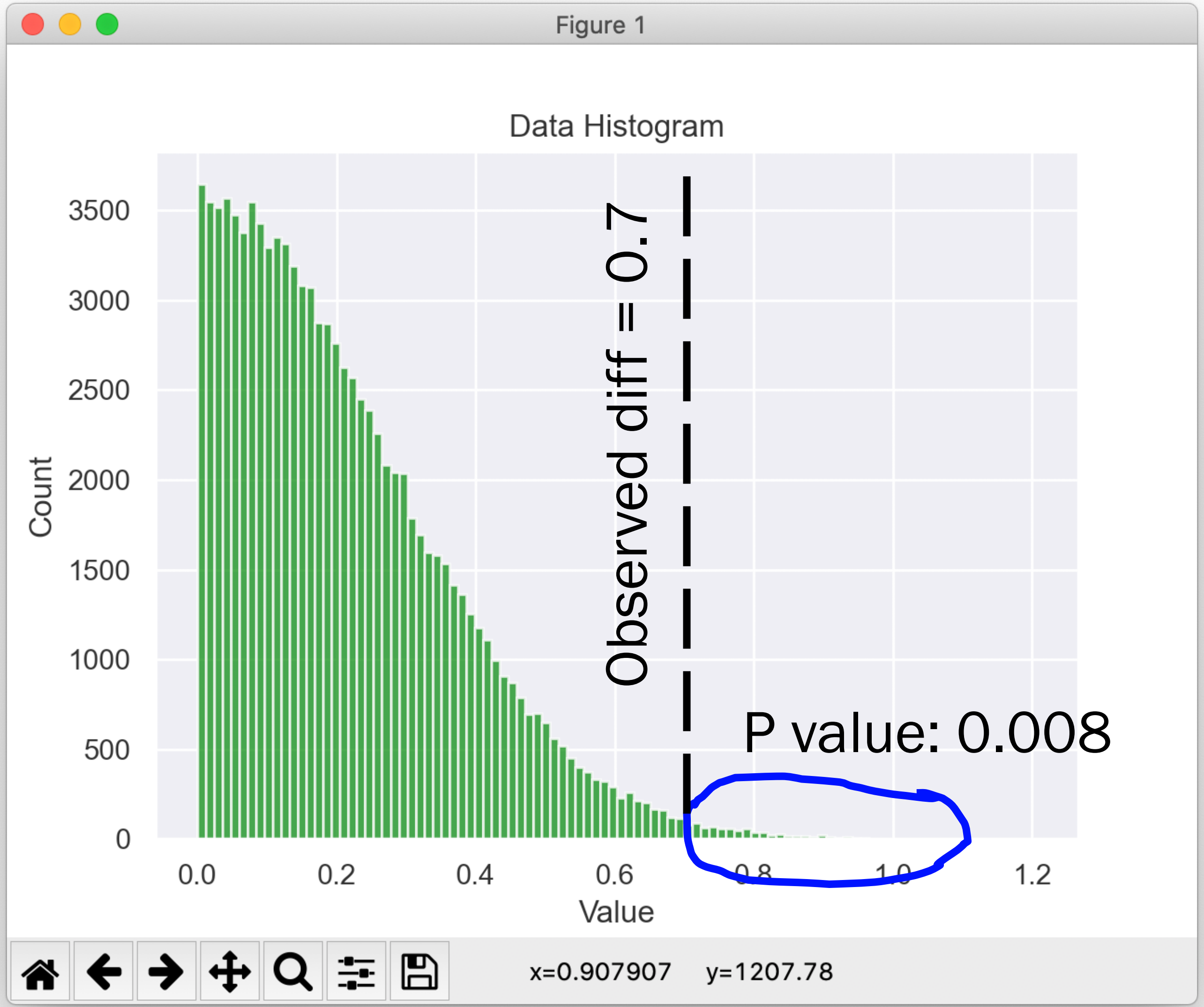


To the code!

Distribution of Mean Diffs under Null Hypothesis

Q: But what distribution is this??

A: Folded Normal. Abs of a CLT process



Food for thought: Two Opinions on Distributions

Results of flipping a coin 20 times. Give your belief distribution of p:

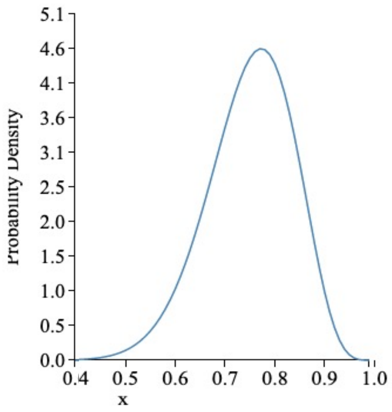
H, H, H, T, H, T, H, H, H, H, H, T, H, H, H, H, H, H, T, H

4 tails, 16 heads

Bayesian:

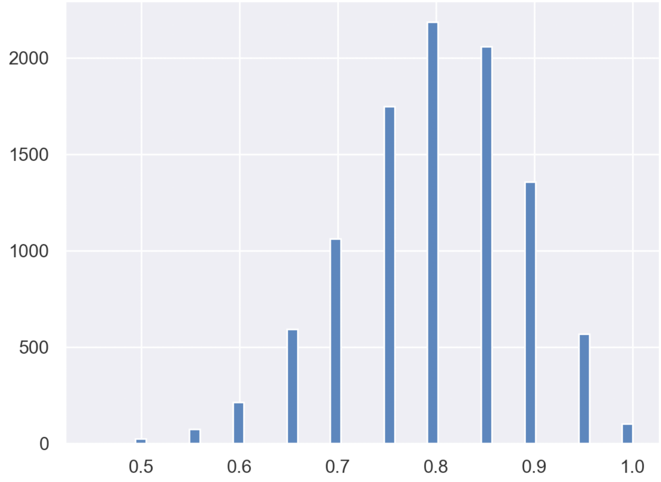
Let's use Laplace prior

X ~ Beta(a = 18, b = 6)



Frequentist:

Let's bootstrap



Algorithmic Analysis

Expectation

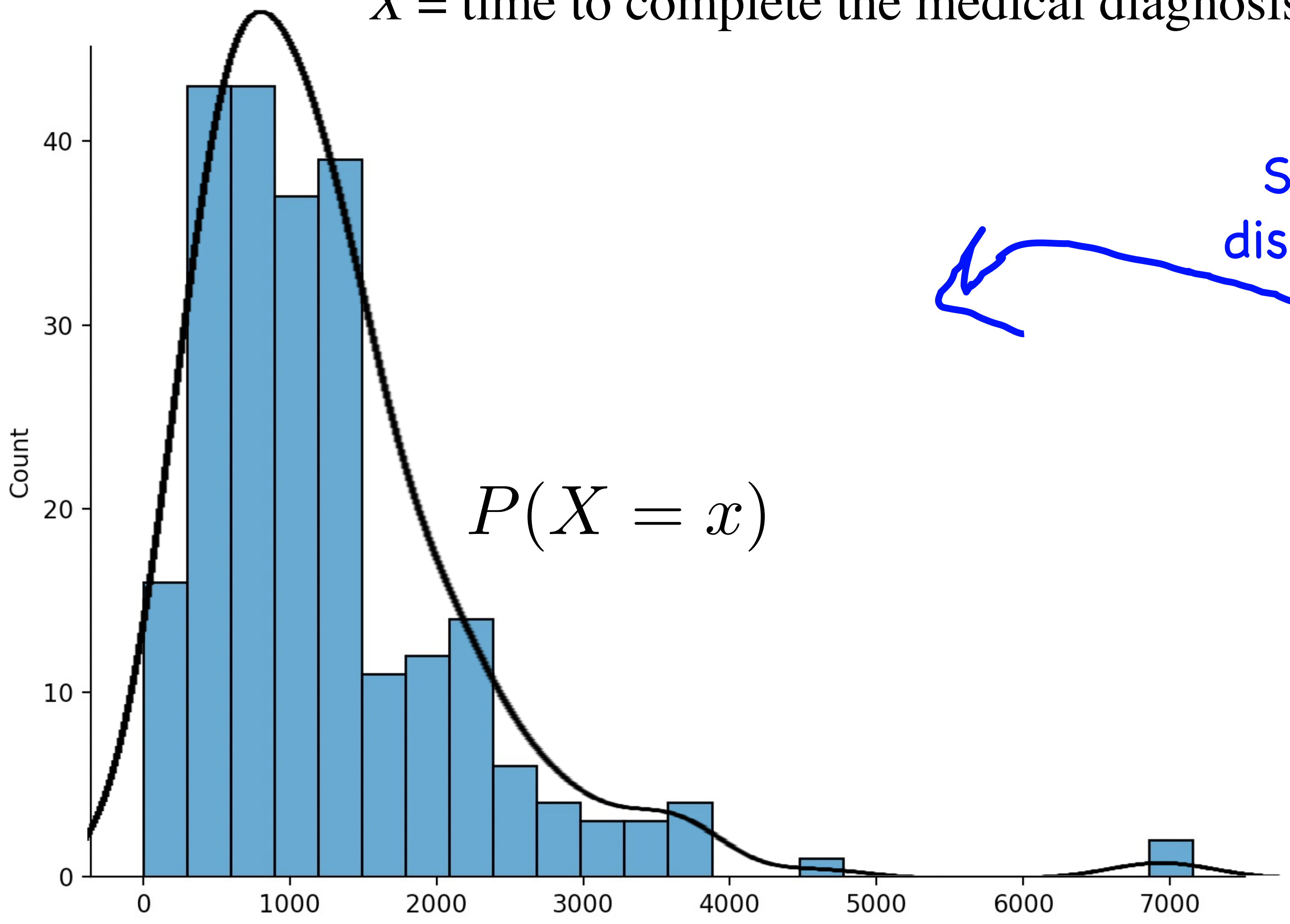
The probability that X takes on that value

$$E[X] = \sum_{x} x \cdot P(X = x)$$

All the values that X can take on

Limitation of Expectation

X = time to complete the medical diagnosis problem (in seconds)



$P(X = x)$

Summarize the distribution in one number

$\mathrm{E}[X] = 12600$

Expectation of a Sum

$$E[X + Y] = E[X] + E[Y]$$

Generalized: $$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

Holds regardless of dependency between X_i's

Expectation of a Function

Law of unconscious statistician

$$E[g(X)] = \sum_x g(x) \cdot P(X = x)$$

So for example…

$$E[X^2] = \sum_x x^2 \cdot P(X = x)$$

Bool was Cool

Let E_1, E_2, ... E_n be events with indicator RVs X_i

- If event E_i occurs, then X_i = 1, else X_i = 0
- Recall $E[X_i] = P(E_i)$

- Why?

$$E[X_i] = 0 \cdot (1 - P(E_i)) + 1 \cdot P(E_i)$$

Bernoulli aka Indicator Random Variables were studied extensively by George Boole

Boole died of being too cool

Expectation of the Binomial

Let $Y \sim \text{Bin}(n, p)$

- n independent trials
- Let X_i = 1 if i-th trial is "success", 0 otherwise
- $X_i \sim \text{Ber}(p)$ $\quad E[X_i] = p$

$$Y = X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{n} X_i$$

$$E[Y] = E[\sum_{i=1}^{n} X_i]$$

$$= \sum_{i=1}^{n} E[X_i]$$

$$= E[X_1] + E[X_2] + \ldots E[X_n]$$

$$= np$$

Expectation of the Negative Binomial

Let Y ~ NegBin(r, p)

- Recall Y is number of trials until r "successes"
- Let X_i = # of trials to get success after (i – 1)st success
- X_i ~ Geo(p) (i.e., Geometric RV)

$$Y = X_1 + X_2 + \cdots + X_r = \sum_{i=1}^{r} X_i \qquad E[X_i] = \frac{1}{p}$$

$$\begin{aligned} E[Y] &= E[\sum_{i=1}^{r} X_i] \\ &= \sum_{i=1}^{r} E[X_i] \\ &= E[X_1] + E[X_2] + \ldots E[X_r] \\ &= \frac{r}{p} \end{aligned}$$

Differential Privacy

Aims to provide means to **maximize the accuracy** of probabilistic queries while minimizing the **probability** of identifying its records.

Cynthia Dwork's celebrity lookalike is Cynthia Dwork.

Differential Privacy

100 independent values $X_1 \ldots X_{100}$ where $X_i \sim \text{Bern}(p)$

```
# Maximize accuracy, while preserving privacy.
def calculateYi(Xi):
    obfuscate = random()
    if obfuscate:
        return indicator(random())
    else:
        return Xi
```

`random()` returns `True` or `False` with equal likelihood

Differential Privacy

100 independent values $X_1 \ldots X_{100}$ where $X_i \sim \text{Bern}(p)$

```
# Maximize accuracy, while preserving privacy.
def calculateYi(Xi):
   obfuscate = random()
   if obfuscate:
      return indicator(random())
   else:
      return Xi
```

`random()` returns `True` or `False` with equal likelihood

What is $E[Y_i]$?

$$E[Y_i] = P(Y_i = 1) = \frac{p}{2} + \frac{1}{4}$$

Differential Privacy

100 independent values X_1 ... X_{100} where $X_i \sim \text{Bern}(p)$

```
# Maximize accuracy, while preserving privacy.
def calculateYi(Xi):
    obfuscate = random()
    if obfuscate:
        return indicator(random())
    else:
        return Xi
```

`random()` returns `True` or `False` with equal likelihood

Let $Z = \sum_{i=1}^{100} Y_i$

What is the E[Z]?

$$E[Z] = E\left[\sum_{i=1}^{100} Y_i\right] = \sum_{i=1}^{100} E[Y_i] = \sum_{i=1}^{100} \left(\frac{p}{2} + \frac{1}{4}\right) = 50p + 25$$

Differential Privacy

100 independent values $X_1 \ldots X_{100}$ where $X_i \sim \text{Bern}(p)$

```
# Maximize accuracy, while preserving privacy.
def calculateYi(Xi):
    obfuscate = random()
    if obfuscate:
        return indicator(random())
    else:
        return Xi
```

`random()` returns `True` or `False` with equal likelihood

Let $Z = \sum_{i=1}^{100} Y_i \quad E[Z] = 50p + 25$

How do you estimate *p*?

$$p \approx \frac{Z - 25}{50}$$

Challenge: What is the probability that our estimate is good?

Differential Privacy

Story which continues to unfold...

Generalization in Adaptive Data Analysis and Holdout Reuse*

Cynthia Dwork
Microsoft Research

Vitaly Feldman
IBM Almaden Research Center†

Moritz Hardt
Google Research

Toniann Pitassi
University of Toronto

Omer Reingold
Samsung Research America

Aaron Roth
University of Pennsylvania

Now at Stanford

More Practice!

Computer Cluster Utilization

Computer cluster with *k* servers

- Requests independently go to server *i* with probability p_i
- Let event A_i = server *i* receives no requests
- Let Bernoulli B_i be an indicator for A_i
- X = # of events $A_1, A_2, \ldots A_k$ that occur
- Y = # servers that receive ≥ 1 request = *k* – X
- E[Y] after first *n* requests?
- Since requests independent:

$$X = \sum_{i=1}^{k} B_i$$

$$P(A_i) = (1 - p_i)^n$$

$$E[X] = E\left[\sum_{I=1}^{k} B_i\right] = \sum_{i=1}^{k} E[B_i] = \sum_{i=1}^{k} P(A_i) = \sum_{i=1}^{k} (1 - p_i)^n$$

$$E[Y] = k - E[X] = k - \sum_{i=1}^{k} (1 - p_i)^n$$

Amazon Monetized This

amazon

amazon web services™

* 52% of Amazons Profits

**More profitable than Amazon's North America commerce operations

http://www.zdnet.com/article/amazons-finds-its-profit-horse-in-aws-why-its-so-disruptive-to-its-old-guard/

When stuck, brainstorm about random variables

Hash Tables (aka Toy Collection)

Consider a hash table with *n* buckets

- Each string equally likely to get hashed into any bucket
- Let X = # strings to hash until each bucket ≥ 1 string
- What is E[X]?
- Let X_i = # of trials to get success after *i*-th success
  - where "success" is hashing string to previously empty bucket
  - After *i* buckets have ≥ 1 string, probability of hashing a string to an empty bucket is *p* = (*n* – *i*) / *n*
  - $P(X_i = k) = \frac{n-i}{n}\left(\frac{i}{n}\right)^{k-1}$ equivalently: $X_i \sim \text{Geo}((n - i) / n)$
  - $E[X_i] = 1 / p = n / (n - i)$
- $X = X_0 + X_1 + \ldots + X_{n-1} \Rightarrow E[X] = E[X_0] + E[X_1] + \ldots + E[X_{n-1}]$

$$E[X] = \frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + \ldots + \frac{n}{1} = n\left[\frac{1}{n} + \frac{1}{n-1} + \ldots + 1\right] = O(n \log n)$$

This is your final answer

Break

Conditional Expectation

Conditional Expectation

X and Y are jointly discrete random variables

- Recall conditional PMF of X given Y = y:

$$p_{X|Y}(x \mid y) = P(X = x \mid Y = y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

Define conditional expectation of X given Y = y:

$$E[X \mid Y = y] = \sum_x x P(X = x \mid Y = y) = \sum_x x\, p_{X|Y}(x \mid y)$$

Analogously, jointly continuous random variables:

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} \qquad E[X \mid Y = y] = \int_{-\infty}^{\infty} x\, f_{X|Y}(x \mid y)\, dx$$

Conditional Expectation

$$E[X|Y=y] = \sum_x x \cdot P(X=x|Y=y)$$

Roll two 6-sided dice D_1 and D_2

- X = value of $D_1 + D_2$ Y = value of D_2
- What is $E[X \mid Y = 6]$?

$$E[X \mid Y=6] = \sum_x xP(X=x \mid Y=6)$$

$$= \left(\frac{1}{6}\right)(7+8+9+10+11+12) = \frac{57}{6} = 9.5$$

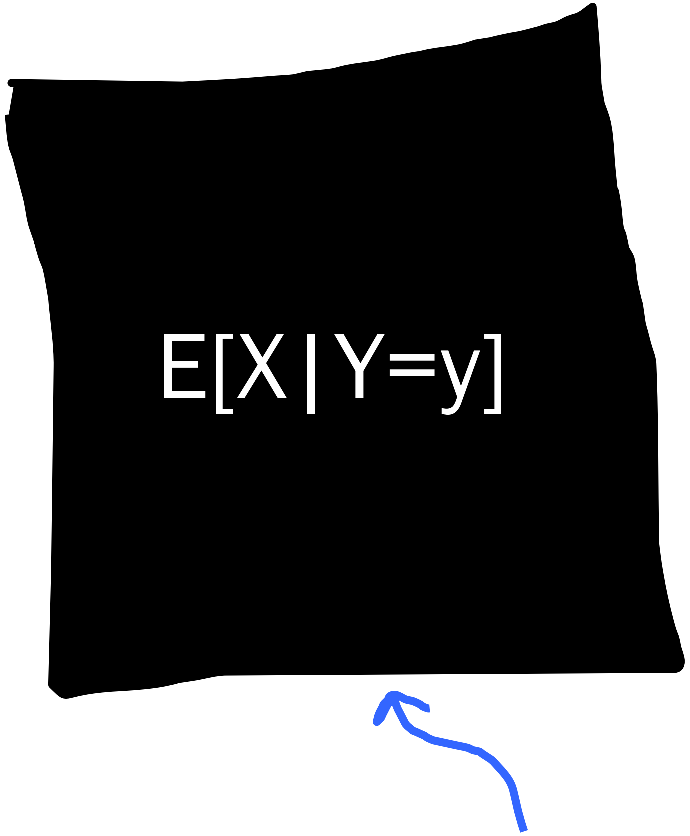
- Intuitively makes sense: 6 + E[value of D_1] = 6 + 3.5

Conditional Expectation

$$E[X|Y=y] = \sum_x x \cdot P(X=x|Y=y)$$

Define $g(Y) = E[X \mid Y]$

This is just function of Y

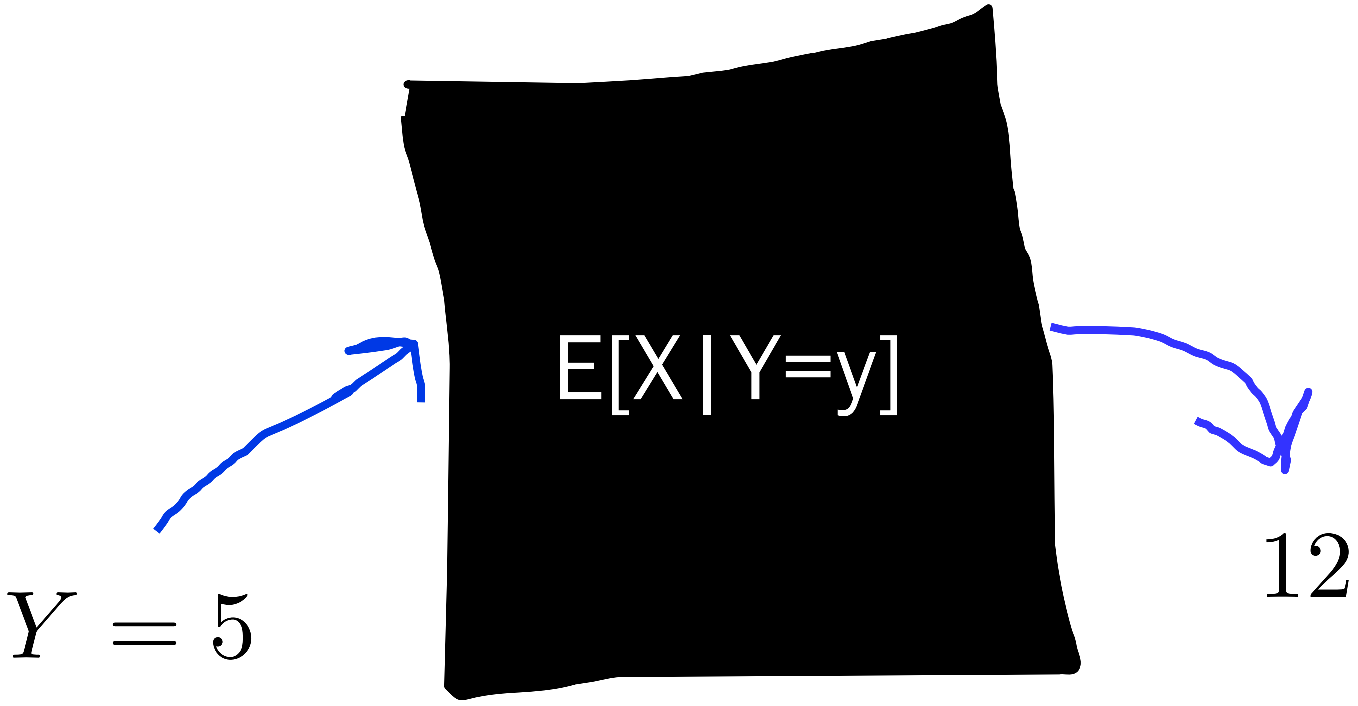


This is a function with Y as input

Conditional Expectation

$$E[X|Y=y] = \sum_x x \cdot P(X=x|Y=y)$$

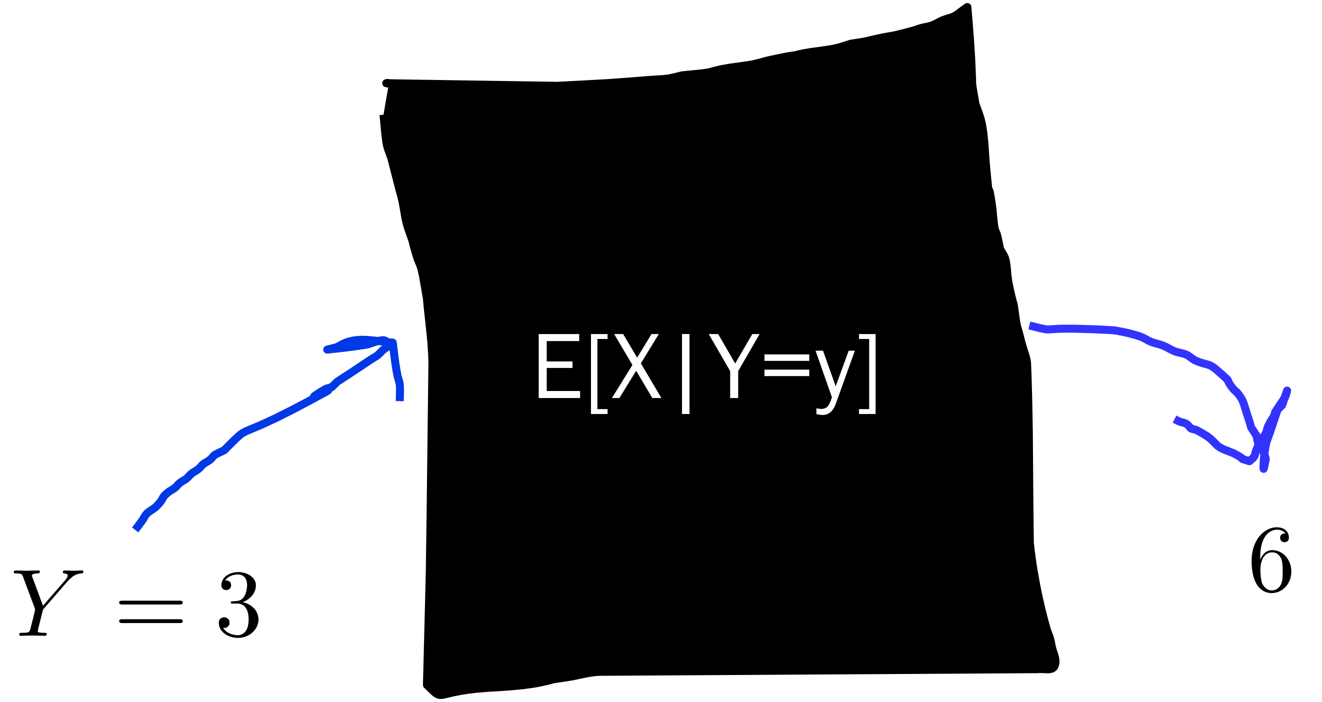
- Define $g(Y) = E[X \mid Y]$
- This is just function of Y



Conditional Expectation

$$E[X|Y = y] = \sum_x x \cdot P(X = x|Y = y)$$

- Define $g(Y) = E[X \mid Y]$
- This is just function of Y



Conditional Expectation

$$\mathrm{E}[X|Y=y] = \sum_x x \cdot P(X=x|Y=y)$$

This is a number:

$$E[X]$$

This is a function of y:

$$E[X|Y=y]$$

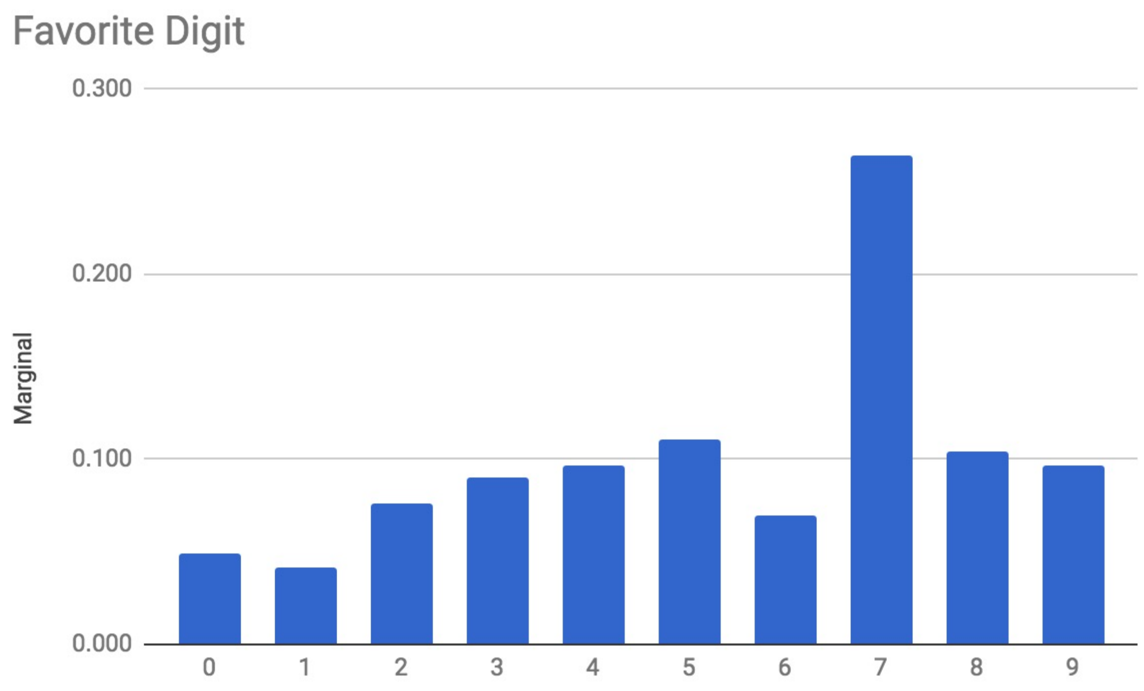
$$E[X=5]$$

Doesn't make sense. Take expectation of random variables, not events

Expectation

$$E[X|Y=y] = \sum_x x \cdot P(X=x|Y=y)$$

X = favorite number
Y = year in school



E[X] = 0 * 0.05 + … + 9 * 0.10 = 5.38

Conditional Expectation

$$E[X|Y = y] = \sum_x x \cdot P(X = x|Y = y)$$

X = favorite number
Y = year in school

E[X | Y] ?

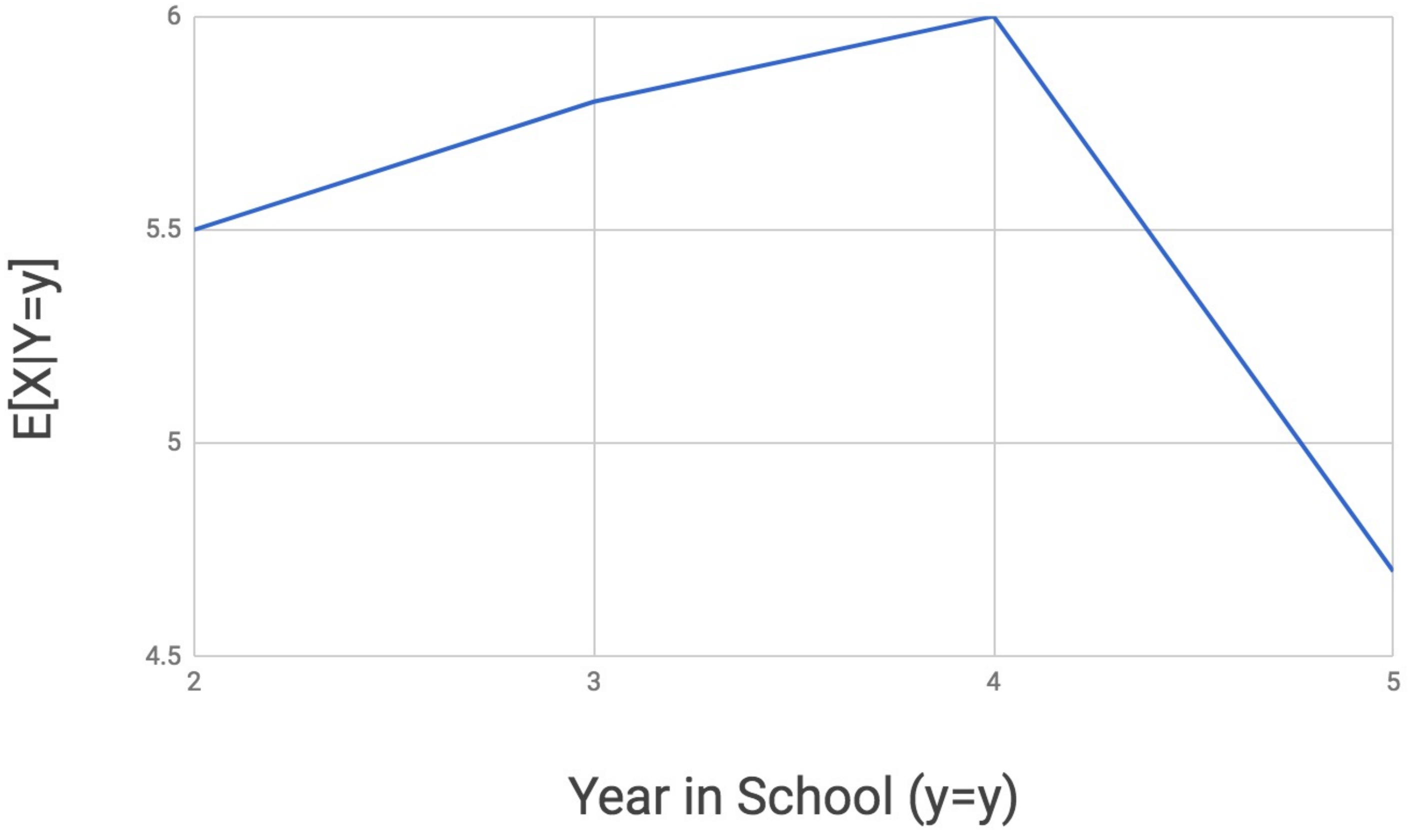
Year in school, Y = y	E[X | Y = y]
2	5.5
3	5.8
4	6.0
5	4.7

Conditional Expectation

$$E[X|Y=y] = \sum_x x \cdot P(X=x|Y=y)$$

X = favorite number
Y = year in school

E[X | Y] ?



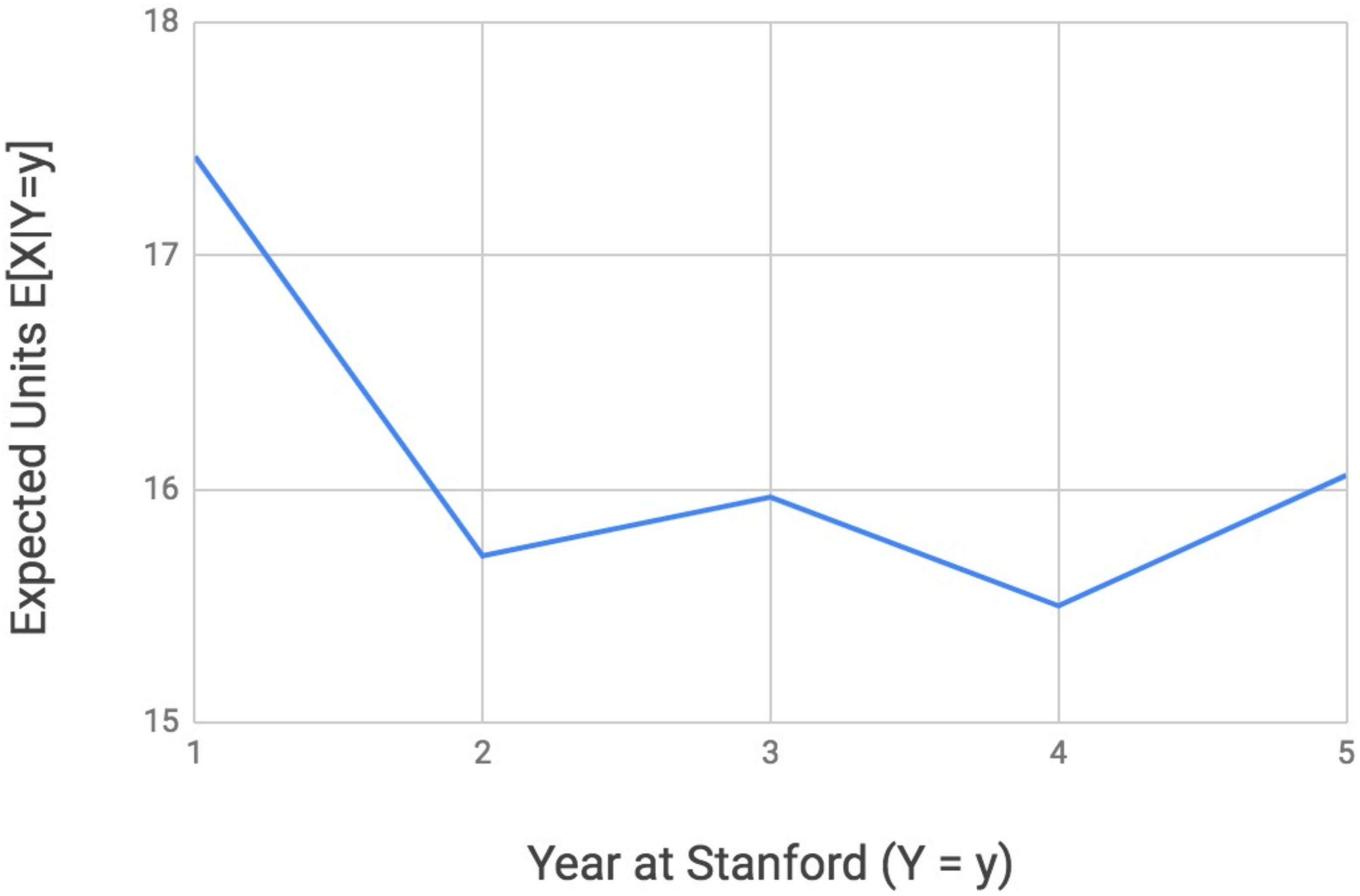
Conditional Expectation

$$E[X|Y = y] = \sum_x x \cdot P(X = x|Y = y)$$

X = units in fall quarter

Y = year in school

E[X | Y] ?



What is this???

$$E[E[X|Y]]$$

Function of Y

Law of Total Expectation

$$E[E[X|Y]] = E[X]$$

$$
\begin{aligned}
E[E[X|Y]] &= \sum_y E[X|Y=y]P(Y=y) && \text{g(Y) = E[X|Y]} \\
&= \sum_y \sum_x xP(X=x|Y=y)P(Y=y) && \text{Def of E[X|Y]} \\
&= \sum_y \sum_x xP(X=x, Y=y) && \text{Chain rule!} \\
&= \sum_x \sum_y xP(X=x, Y=y) && \text{I switch the order of the sums} \\
&= \sum_x x \sum_y P(X=x, Y=y) && \text{Move that x outside the y sum} \\
&= \sum_x xP(X=x) && \text{Marginalization} \\
&= E[X] && \text{Def of E[X]}
\end{aligned}
$$

Law of Total Expectation

For any random variable *X* and any discrete random variable *Y*

$$E[X] = \sum_y E[X|Y = y]P(Y = y)$$

Analyzing Recursive Code

```
int Recurse() {
    int x = randomInt(1, 3);  // Equally likely values

    if (x == 1) return 3;
    else if (x == 2) return (5 + Recurse());
    else return (7 + Recurse());
}
```

Let Y = value returned by `Recurse()`. What is E[Y]?

$$E[Y] = E[Y \mid X=1]P(X=1) + E[Y \mid X=2]P(X=2) + E[Y \mid X=3]P(X=3)$$

$$E[Y \mid X=1] = 3$$

$$E[Y \mid X=2] = E[5+Y] = 5 + E[Y]$$

$$E[Y \mid X=3] = E[7+Y] = 7 + E[Y]$$

$$E[Y] = 3(1/3) + (5+E[Y])(1/3) + (7+E[Y])(1/3) = (1/3)(15 + 2E[Y])$$

$$E[Y] = 15$$

Protip: do this in CS161

Uncertainty Theory

- Beta Distributions
- Thompson Sampling
- Adding Random Vars
- Central Limit Theorem
- Sampling
- Bootstrapping
- Algorithmic Analysis

Where are we in CS109?

On Friday...

Counting Theory	Core Probability	Random Variables	Probabilistic Models	Uncertainty Theory	Machine Learning