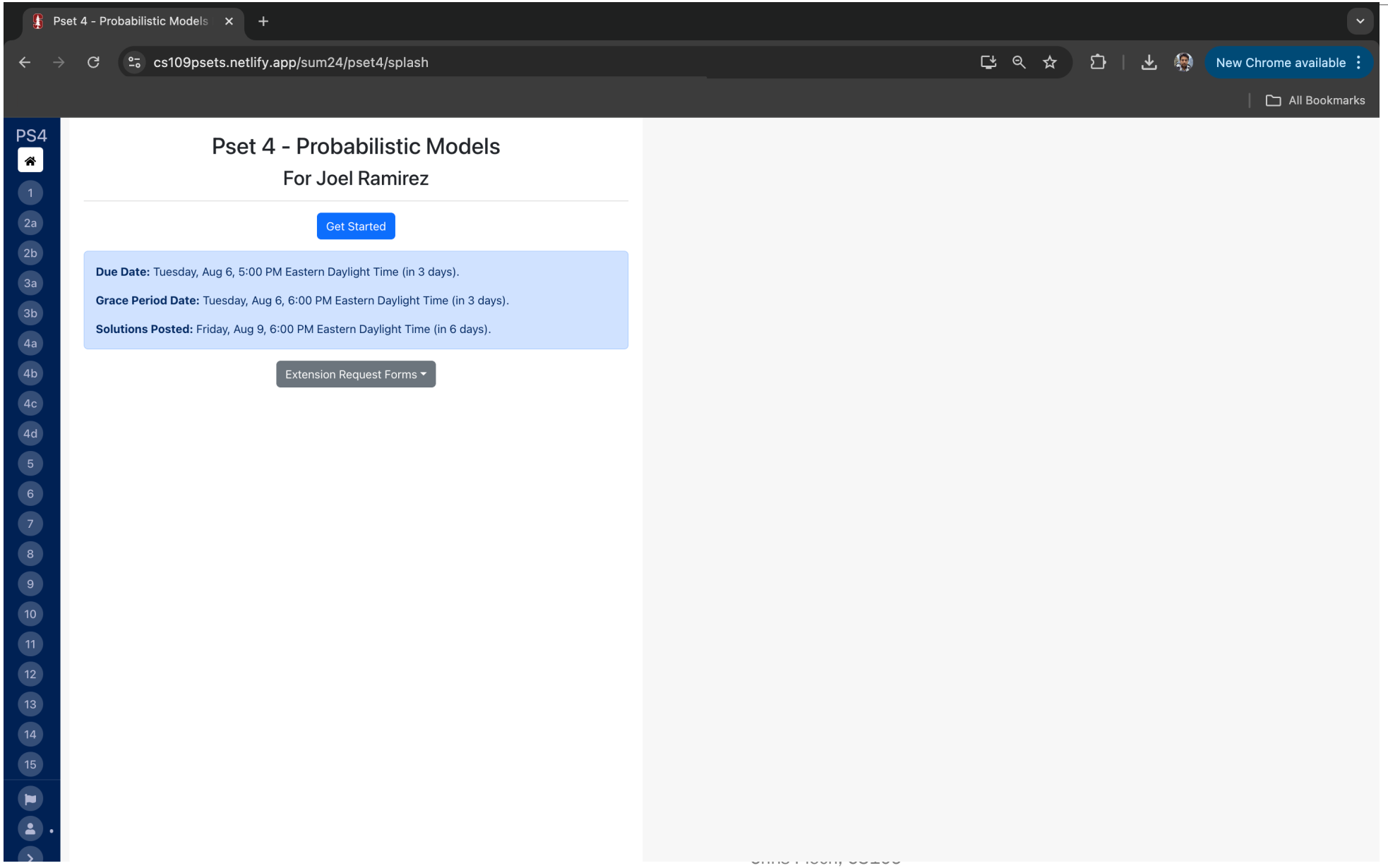
Algorithmic Analysis

CS109, Stanford University

Pset #4 is Due Tomorrow!



<review>

Expectation (just cool)

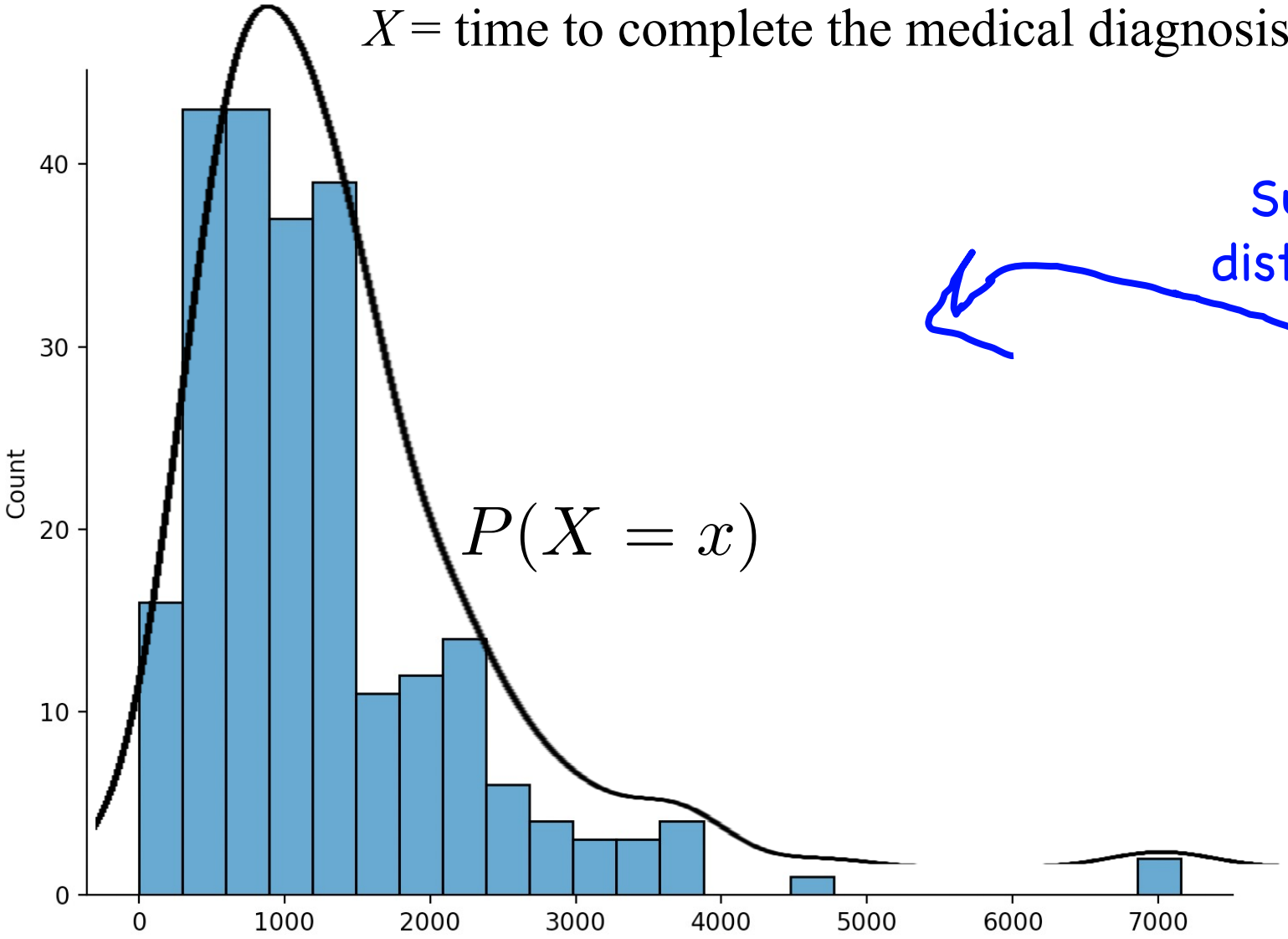
$$\mathrm{E}[X] = \sum_{x} x \cdot P(X = x)$$

The probability that X takes on that value

All the values that X can take on

Limitation of Expectation (not cool?)

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 (much cool)

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

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

We shouldn't be surprised about this though…

Expectation of a Sum

What does this even mean?

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

Whenever you have two variables inside of the expectation you should think about how they interact *jointly.*

We enumerate every possible value of x + y.

Just like you've already been doing with individual variables).

$$\mathbb{E}[X + Y] = \sum_{x,y}(x + y)P(X = x, Y = y)$$

Let's exercise our math muscles to show that they're equivalent!
(Or maybe you can already tell!)

Expectation of a Function

Law of unconscious statistician

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

So for example...

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

Expectation of a Function

Law of the unconscious statistician

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

This one I won't prove...it's sorta lengthy...

Consider what we would do if a function outputs the same values for some x (e.g. x^2)...

This is why we call it the LOTUS

Bool was Cool

Bool was Cool

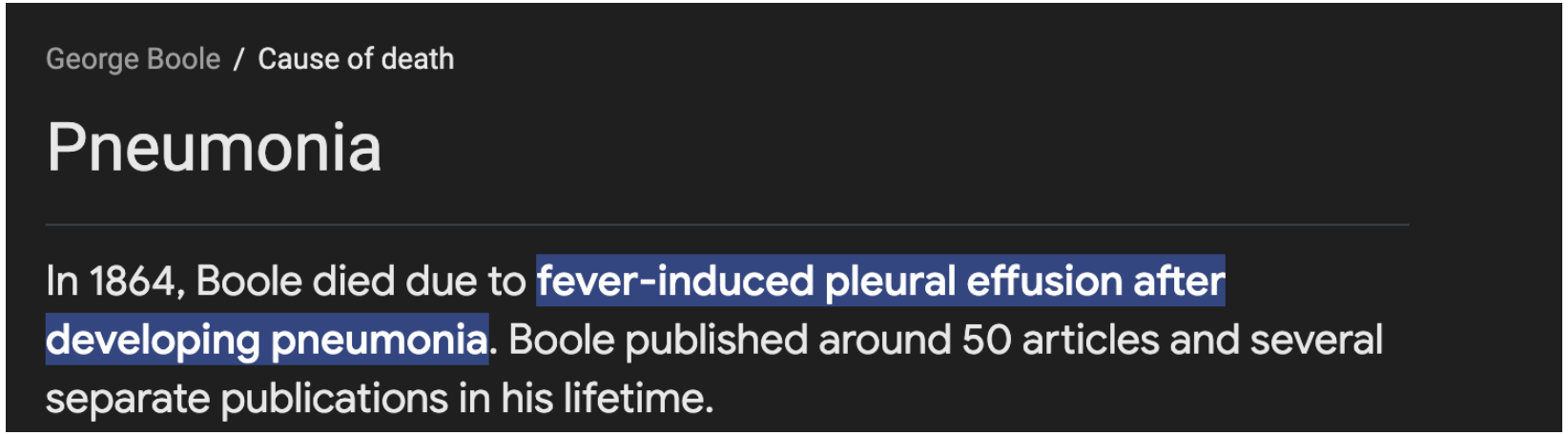
Let E_1, E_2, ... E_n be events with indicator RVs X_i

- Indicators are Bernoulli RVs that represent whether an event took place
- If event E_i occurs, then $X_i = 1$, else $X_i = 0$
- Hmmm: $E[X_i] = P(E_i)$

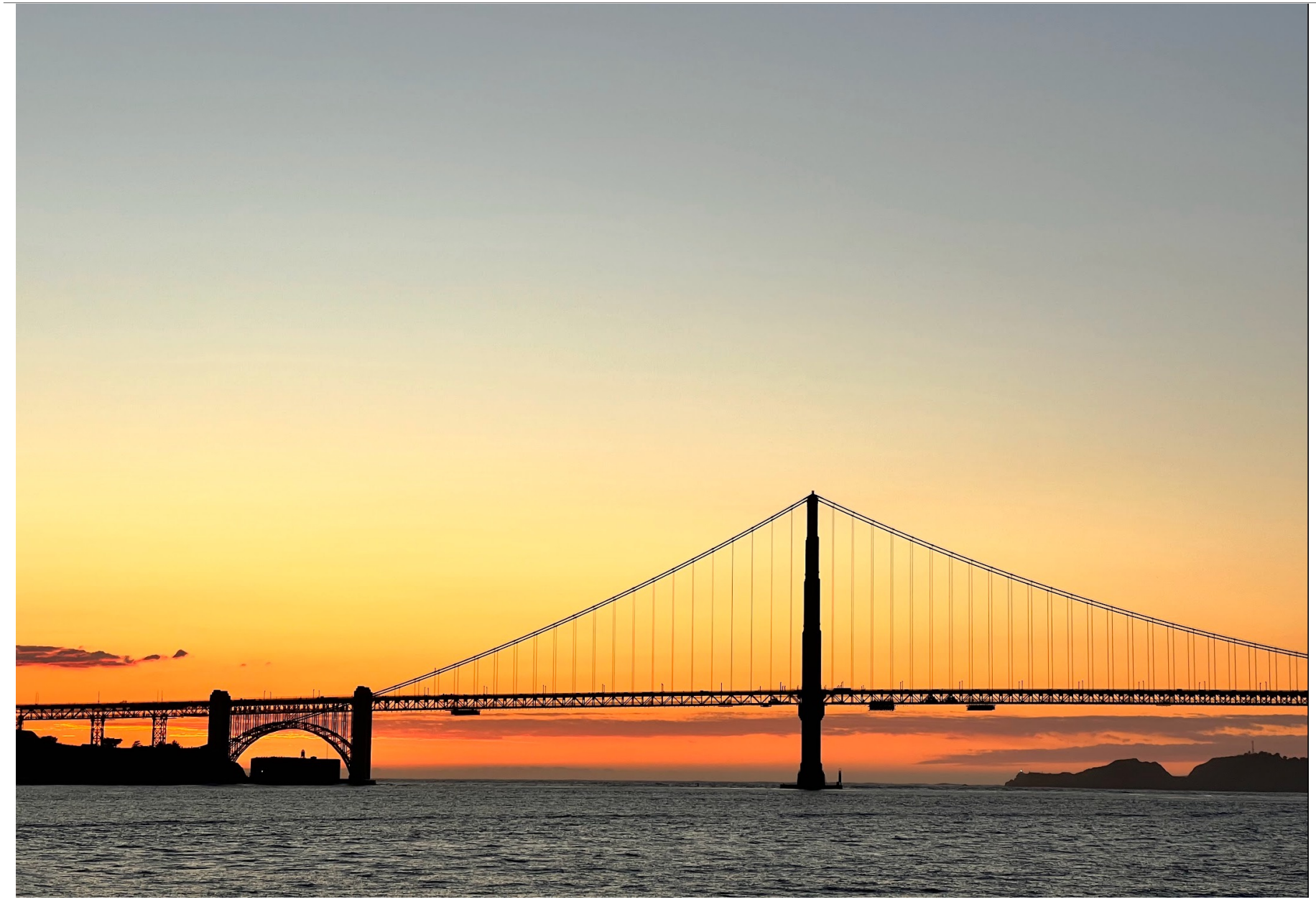
$$E[X_i] = 0 \cdot (1 - P(E_i)) + 1 \cdot P(E_i)$$

Bernoulli aka Indicator Random Variables were studied extensively by George Boole
Boole died of being too cool

Bool was Cool until he died



An aside…things to care about



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

Y_i

What is returned

```
# 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 ... X_{100} where $X_i \sim \text{Bern}(p)$

Y_i

What is returned

```
# 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 $\text{E}[Y_i]$?

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

Differential Privacy

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

Y_i

What is returned

```
# 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

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

P(Y=1)=P(Y=1| obfuscate=True)·P(obfuscate=True)+P(Y =1| obfuscate=False)·P(obfuscate=False)

Let's derive this to convince you!

Differential Privacy

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

Y_i

What is returned

```
# 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 E[Z]?

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

Differential Privacy

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

Y_i

What is returned

```
# 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 $Z = \sum_{i=1}^{100} Y_i$ $E[Z] = 50p + 25$ How do you estimate p?

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

End Review

And aside....

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
- X = # of events $A_1, A_2, \ldots A_k$ that occur
- E[X] after first *n* requests?

Since X is an expectation, can you express it as a sum?

- Let Bernoulli B_i be an **indicator** for A_i
- Since requests independent:

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

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

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

Amazon Monetized This

* ~~52%~~ 74% of Amazons Profits
**More profitable than Amazon's North America commerce operations

https://www.fool.com/investing/2022/07/07/aws-chief-says-amazons-most-profitable-segment-is/

When stuck, brainstorm about random variables

Toy Collection

You are trying to collect *n* distinct toys (and there are only n distinct toys)

- Each purchase, each toy is equally likely (buying with replacement)
- Let X = # purchases until you have ≥ 1 of each toy. What is E[X]?

Let X_i = # of **trials to get success *after i-th success*** where "success" is getting an unseen toy.

$$X = X_0 + X_1 + \ldots + X_{n-1} \Rightarrow \mathrm{E}[X] = \mathrm{E}[X_0] + \mathrm{E}[X_1] + \ldots + \mathrm{E}[X_{n-1}]$$

$$X_i \sim \mathrm{Geo}\left(p = \frac{n-i}{n}\right)$$

After i successes, the probability of the next success is $p = (n-i)/n$

$$\mathbb{E}[X_i] = \frac{1}{p} = \frac{n}{n-i}$$

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

Break

Conditional Expectation

Conditional Expectation

X and Y are discrete random variables:

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

Analogously, continuous random variables:

$$E[X|Y=y] = \int_x xP(X=x|Y=y)$$

Conditional Expectation

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

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 | Y = 6]?

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

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

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

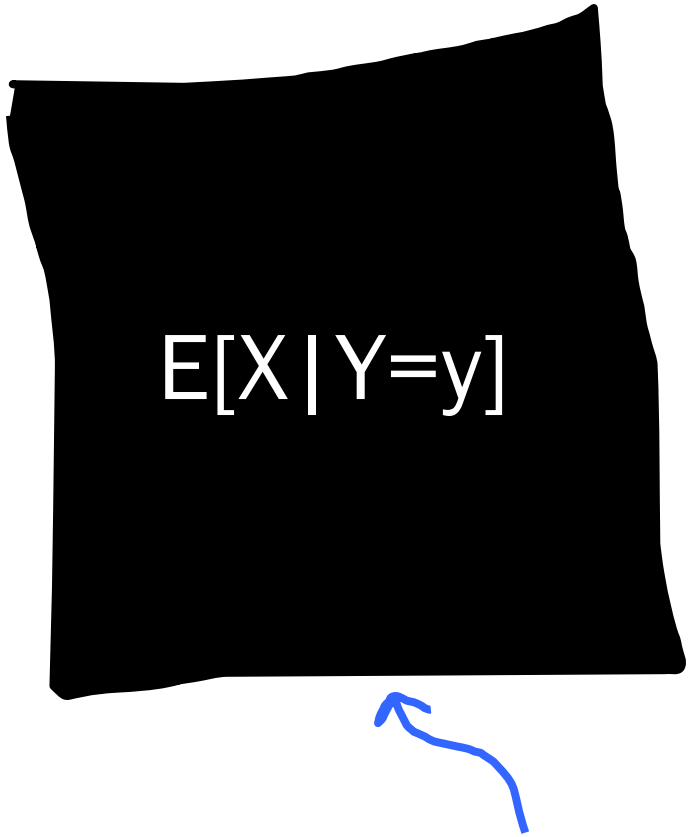
Law of Total Expectation

Conditional Expectation

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

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

This is a function of Y



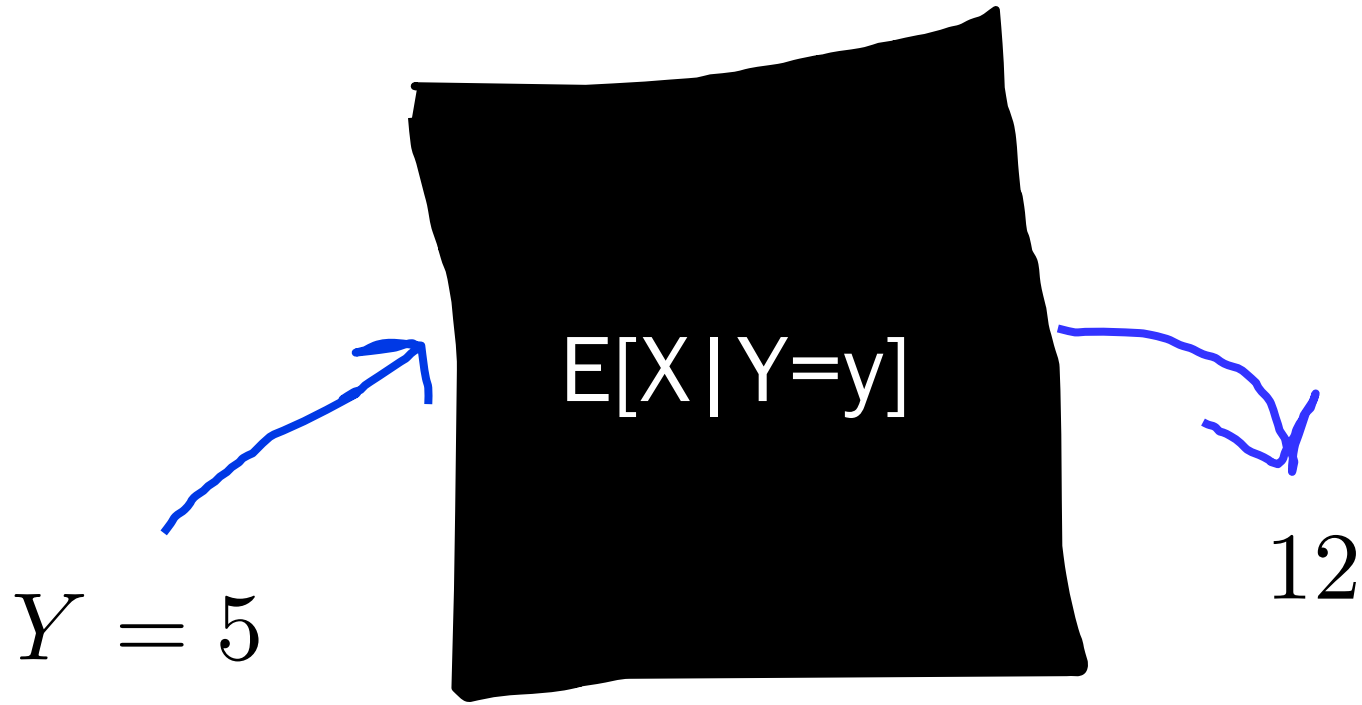
This is a function with Y as input

Conditional Expectation

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

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

This is a function of Y

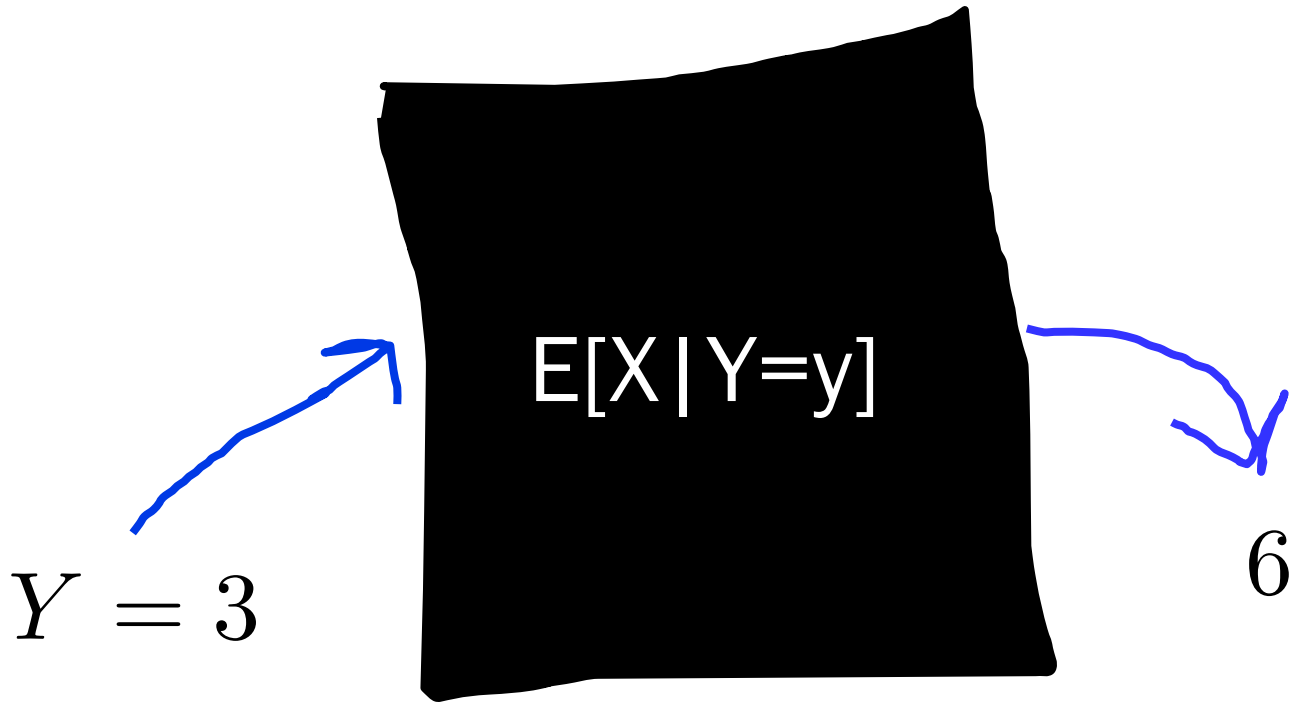


Conditional Expectation

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

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

This is a 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:

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

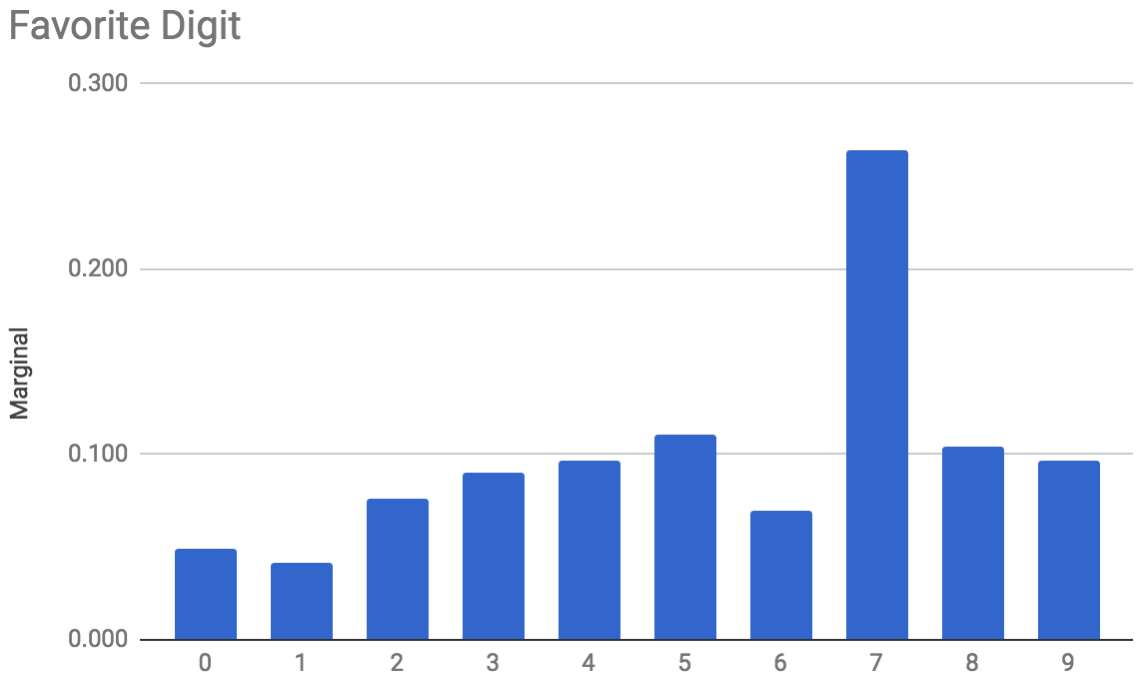
$E[X=5]$

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

Expectation

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

X = favorite number
Y = year in school



E[X] = 0 * 0.05 + ... + 9 * 0.10 = 5.38

Conditional Expectation

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

X = favorite number
Y = year in school

E[X | Y] ?

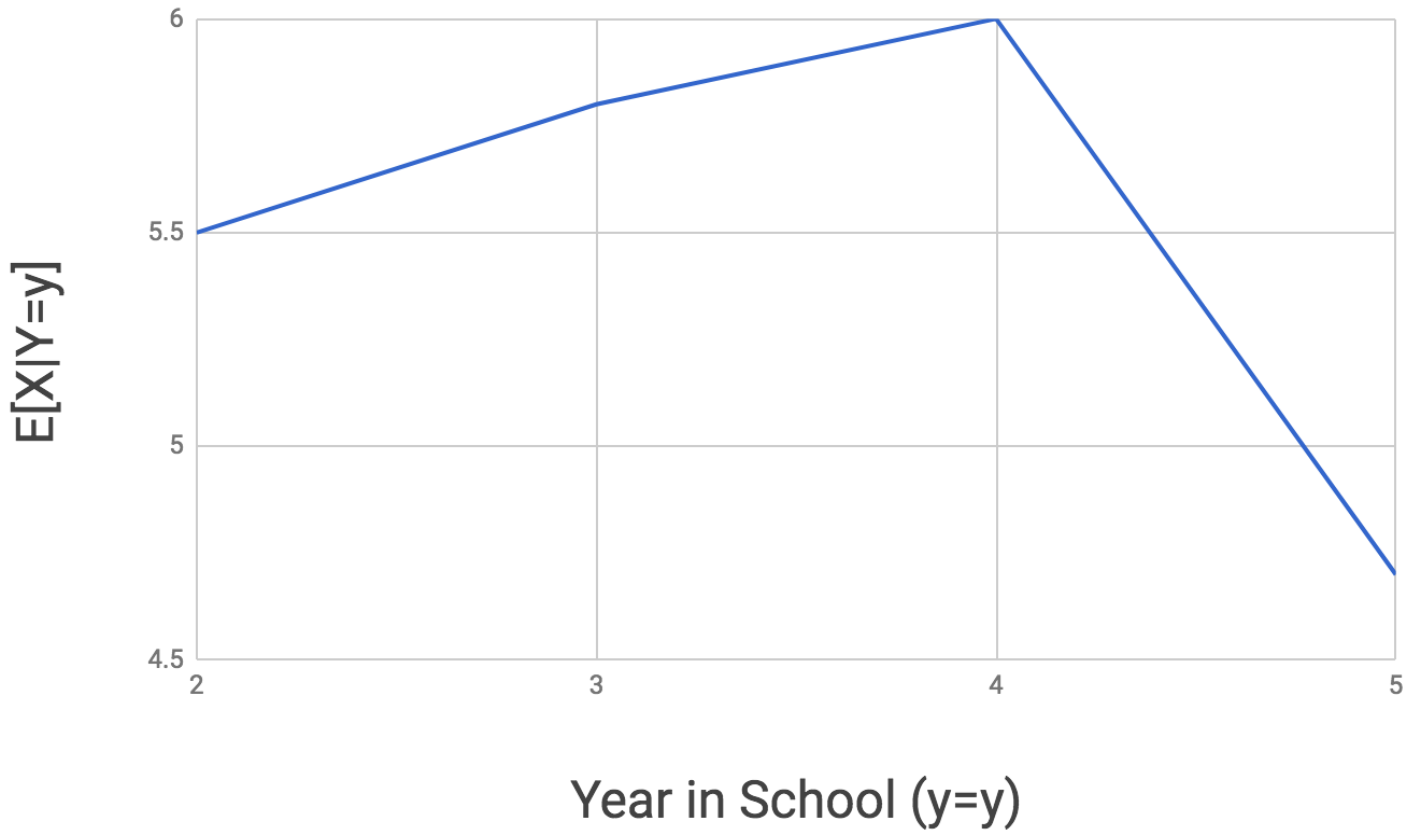
Year in school, Y = y	E[X | Y = y]
2	5.5
3	5.8
4	6.0
5	4.7

Conditional Expectation

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

X = favorite number
Y = year in school

E[X | Y] ?

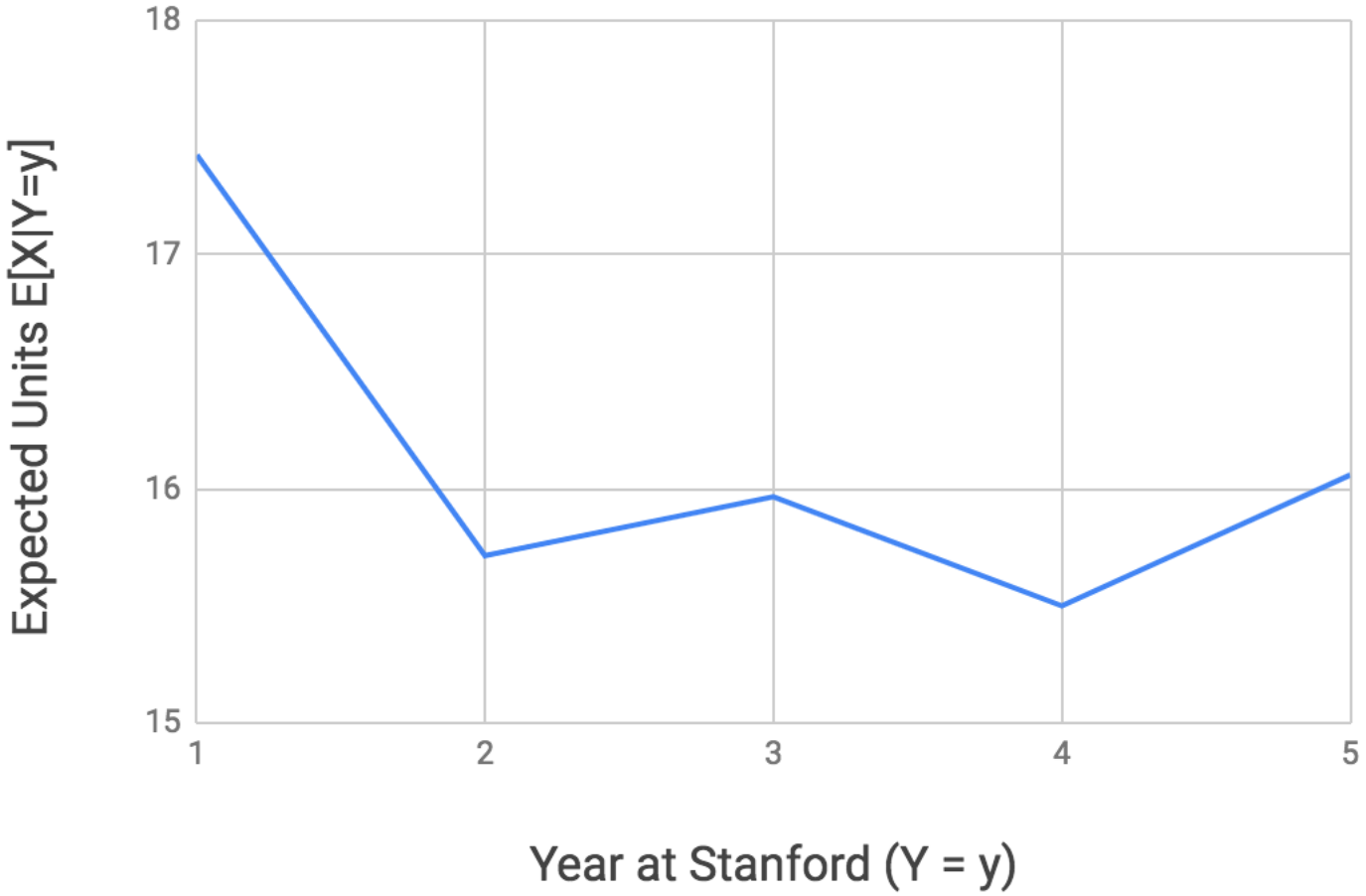


Conditional Expectation

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

X = units in fall quarter
Y = year in school

E[X | Y] ?



Want to see something cool?

What the heck does this give you?

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

Law of Total Expectation

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

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

Law of Total Expectation

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

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

Recall the Law of Total Probability

$$P(X=x) = \sum_y P(X=x|Y=y)P(Y=y)$$

(The Streaming Part)

How long does this code take to run?

Netflix streams millions of hours of videos per day. They REALLY care about the speed of the following code:

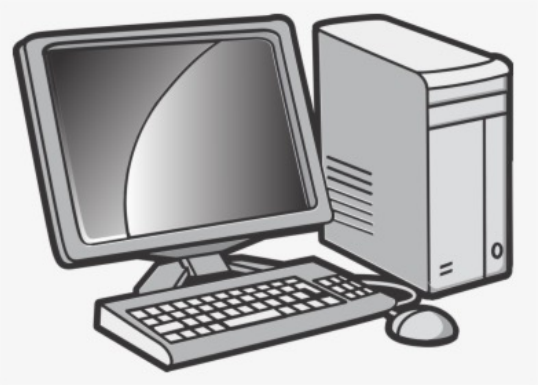
```
database.get_movie(movie_name)
```

How long does this line of code take? Say 512 MB movie.

1. 0.3s
2. 1.6 mins
3. 5 mins
4. 2 hours

All are correct! It is a RV!

How long does this code take to run?



```
database.get_movie(movie_name)
```

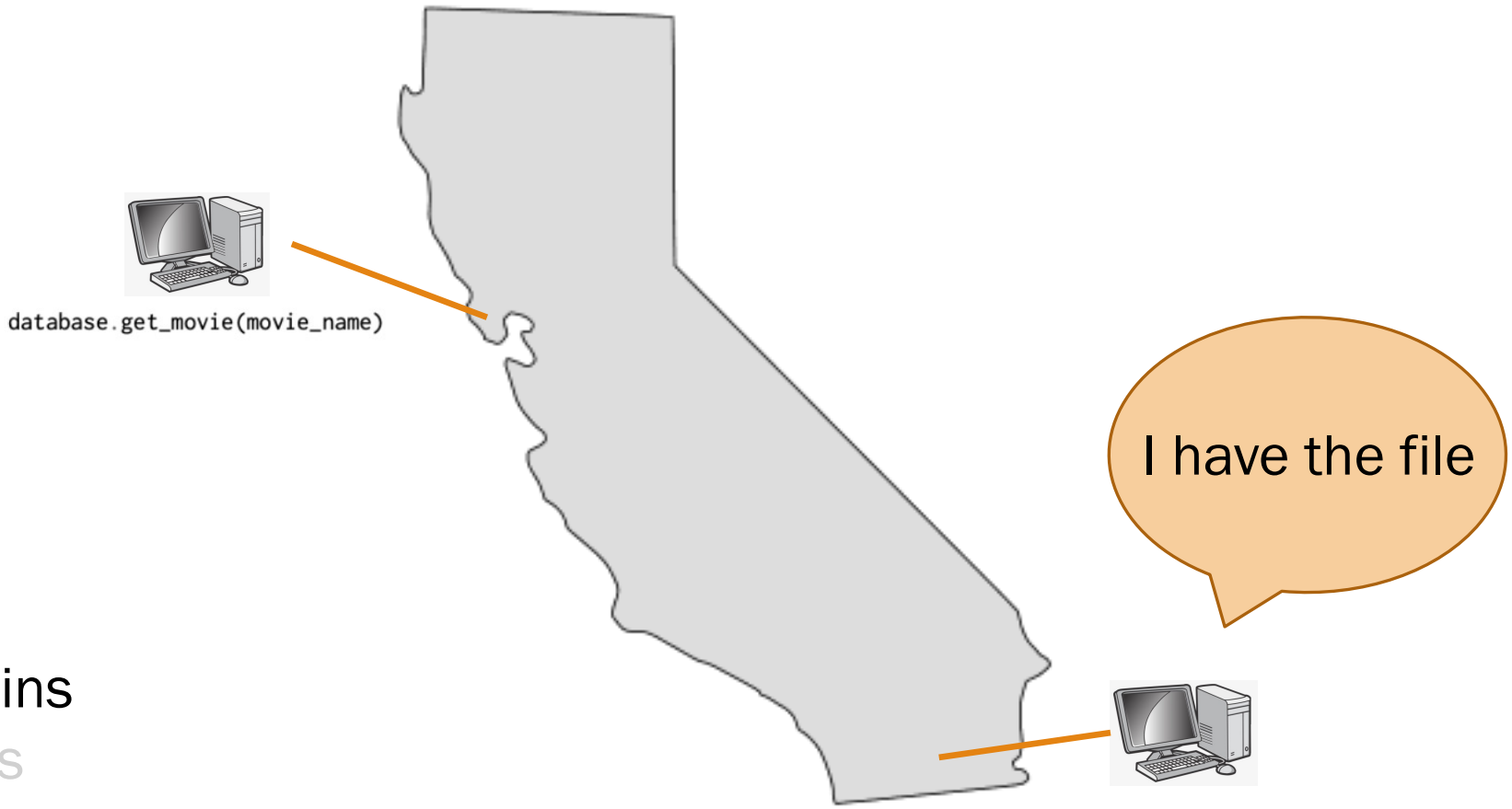
1. 0.3s
2. 1.6 mins
3. 5 mins
4. 2 hours

Millisecond Latency


```
database.get_movie(movie_name)
```

1. 0.3s
2. 1.6 mins
3. 5 mins
4. 2 hours

Minute Latency



1. 0.3s
2. 1.6 mins
3. 5 mins
4. 2 hours

Many Minutes Latency

1. 0.3s
2. 1.6 mins
3. 5 mins
4. 2 hours

Are we done?

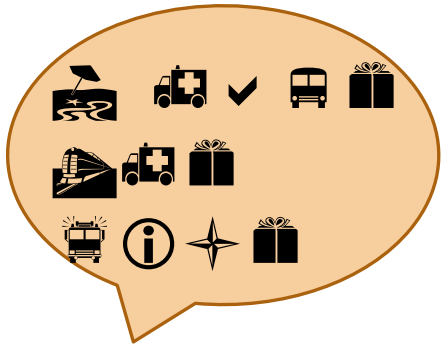
```
database.get_movie(movie_name)
```

5mins across the world!!!!!!


database.get_movie(movie_name)

Anyways

```
database.get_movie(movie_name)
```



1. 0.3s
2. 1.6 mins
3. 5 mins
4. 2 hours

Expected Run Time

Expected runtime (single file from database)

Assume the file location is distributed with the PMF:

- P(file on computer) = 0.10
- P(file in SoCal) = 0.50
- P(file in Japan) = 0.37
- P(file in Space) = 0.03

What is the expected runtime of `database.get_movie(movie_name)`

$$
\begin{aligned}
\mathbb{E}[\texttt{get\_movie\_database\_time}] &= \mathbb{E}[\texttt{get\_movie\_database\_time}|\text{Home}] \cdot \mathbb{P}(\text{Home}) \\
&+ \mathbb{E}[\texttt{get\_movie\_database\_time}|\text{SoCal}] \cdot \mathbb{P}(\text{SoCal}) \\
&+ \mathbb{E}[\texttt{get\_movie\_database\_time}|\text{Japan}] \cdot \mathbb{P}(\text{Japan}) \\
&+ \mathbb{E}[\texttt{get\_movie\_database\_time}|\text{Space}] \cdot \mathbb{P}(\text{Space}) \\
&= 0.1 \cdot 0.3s + 0.5 \cdot 1.6min + 0.37 \cdot 5min + 0.03 \cdot 2hours \\
&\approx 6.25mins
\end{aligned}
$$

Times From Before:

1. Home: 0.3s
2. Socal: 1.6 mins
3. Japan: 5 mins
4. Space: 2 hours

Analyze Recursive Code

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]?

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

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

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

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

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

$$E[Y] = 15$$

Uncertainty Theory

Beta Distributions

Thompson Sampling

Adding Random Vars

Central Limit Theorem

Sampling

Bootstrapping

Algorithmic Analysis

Where are we in CS109?

On Wednesday...