



Random Variables

Chris Piech

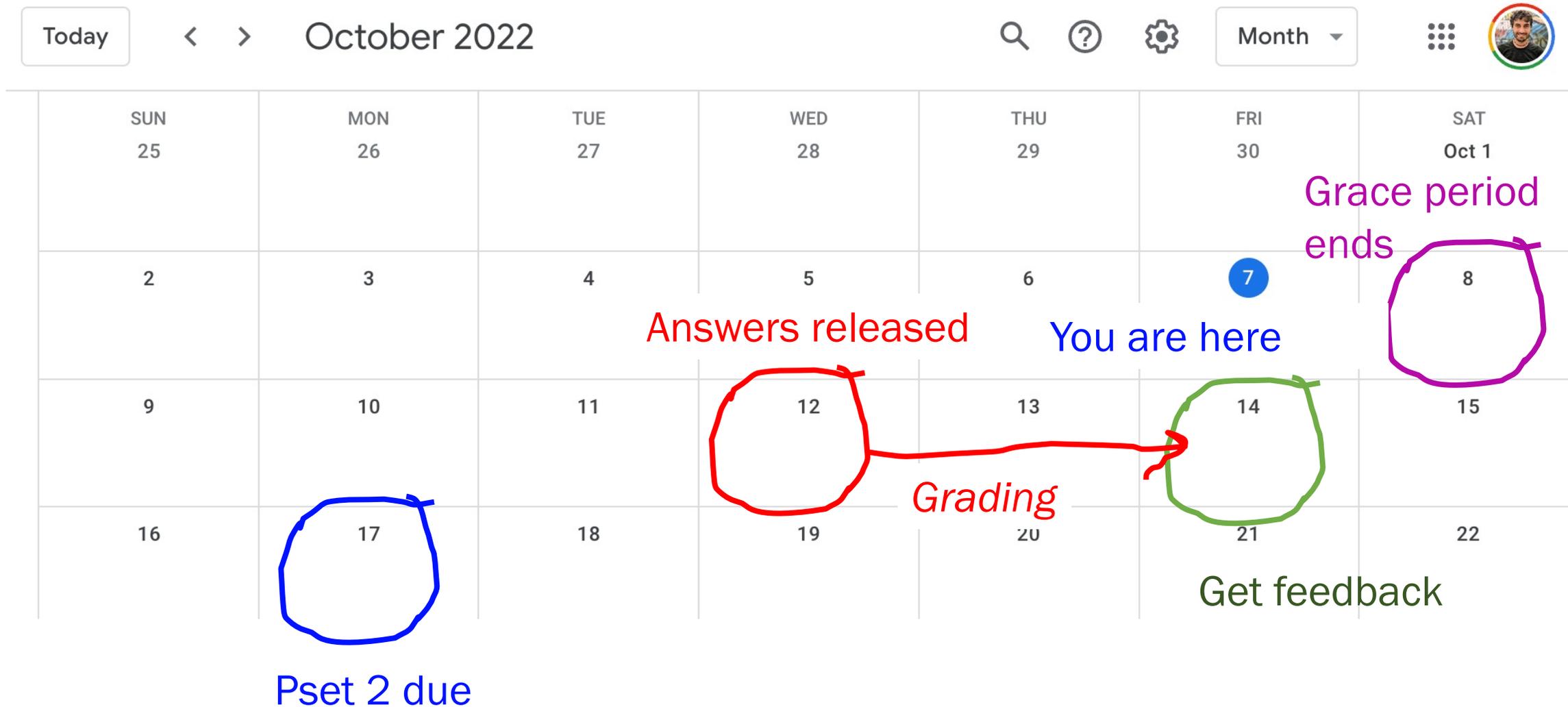
CS109, Stanford University

Announcements

- Pset #1 in today
- Pset #2 out today
- Happy Friday!



Problem Set #1 timeline



Problem Set #2

Browser tabs: Pset 2 - Core Probability

Address bar: cs109psets.netlify.app/win22/pset2/medical_diagnosis

Navigation: Home, Back, Forward, Refresh, Search, Print, Share, Settings, Profile

Bookmarks: gmail, drive, Pixlr, Send Email, Interviews, neat bookmarks, Other Bookmarks, Reading List

PS2 Medical Test

Write a function:

```
def predict_positive_given_test_result(
    prior_disease,
    p_true_given_disease,
    p_true_given_no_disease,
    test_result):
```

That can be used for any noisy (binary) medical test, such as a Covid-19 test, or an Ebola test. Your function takes in a prior belief that a patient has a disease, statistics on a noisy test, and the test result from the noisy test. Based off this information, you should compute the probability that the patient is "positive" for the disease (in other words, they have the disease). Your return value must be a number between 0 and 1, not a boolean prediction. This problem requires you to code up a general implementation of Bayes' Theorem for a binary prediction!

Hint: you might find it helpful to read the medical example from the [Bayes Theorem](#) chapter.



Navigation: Previous Question, Next Question

Answer Editor Solution

Agent Code:

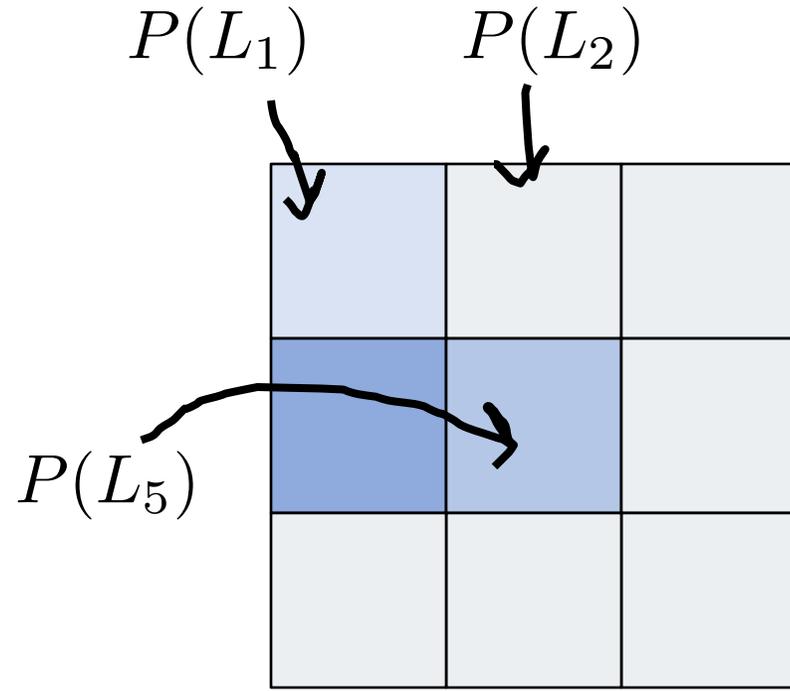
```
1 def predict_positive_given_test_result(
2     prior_disease, # prior prob a patient has the condition
3     p_true_given_disease, # the "true positive" probability
4     p_true_given_no_disease, # the "false positive" probability
5     test_result): # True/False test result
6
7 # improve this function
8     return 0.5
```

Run One Game Test Agent

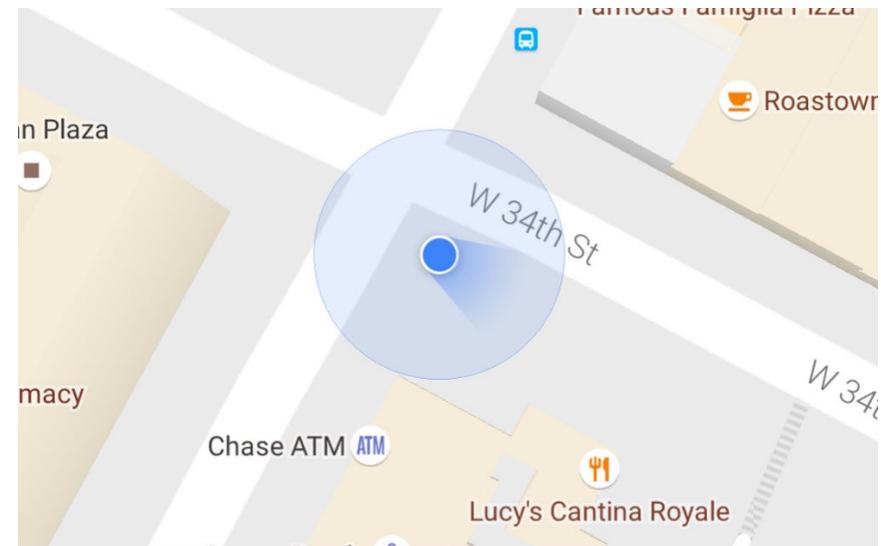
Console



Bayes' Theorem and Location

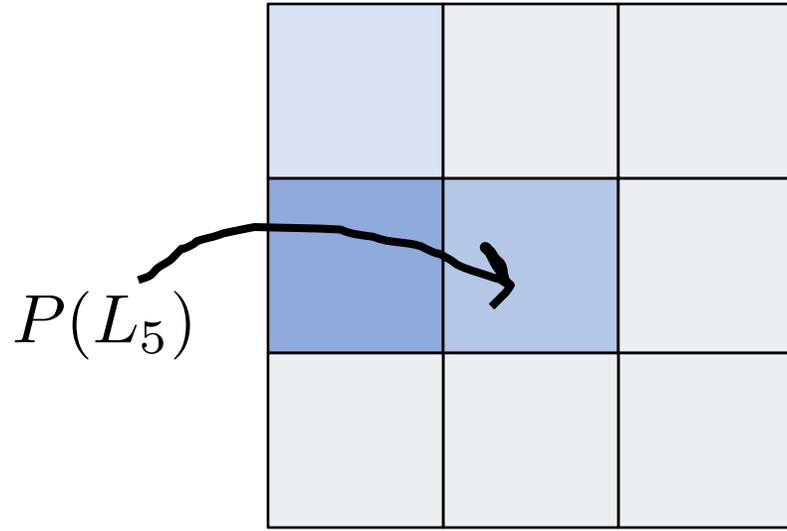
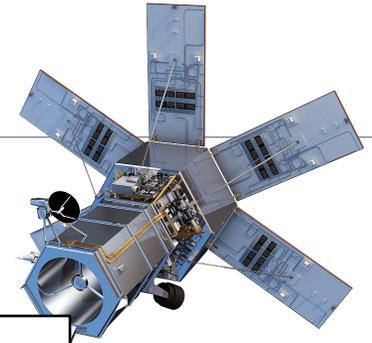


Before Observation

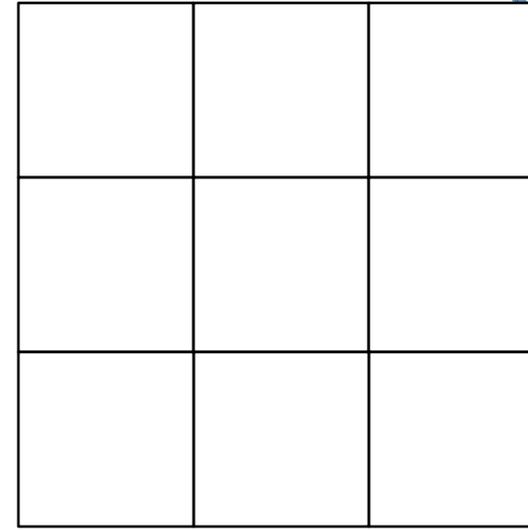


Bayes' Theorem and Location

Know: $P(O|L_i)$



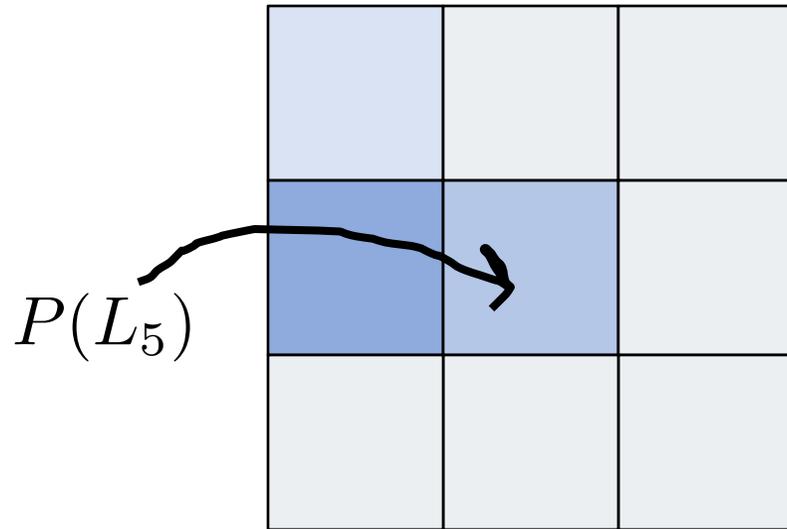
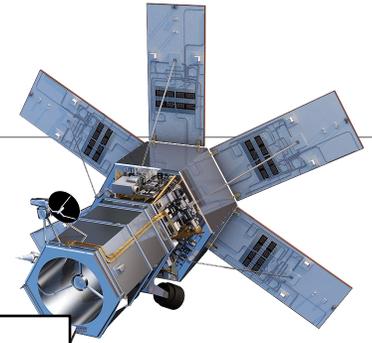
Before Observation



After Observation

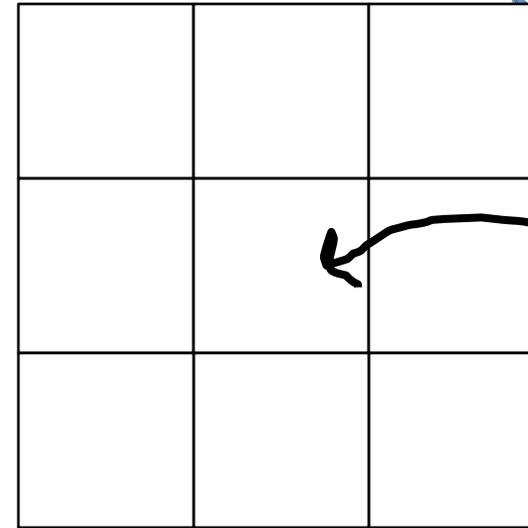
Bayes' Theorem and Location

Know: $P(O|L_i)$



$P(L_5)$

Before Observation

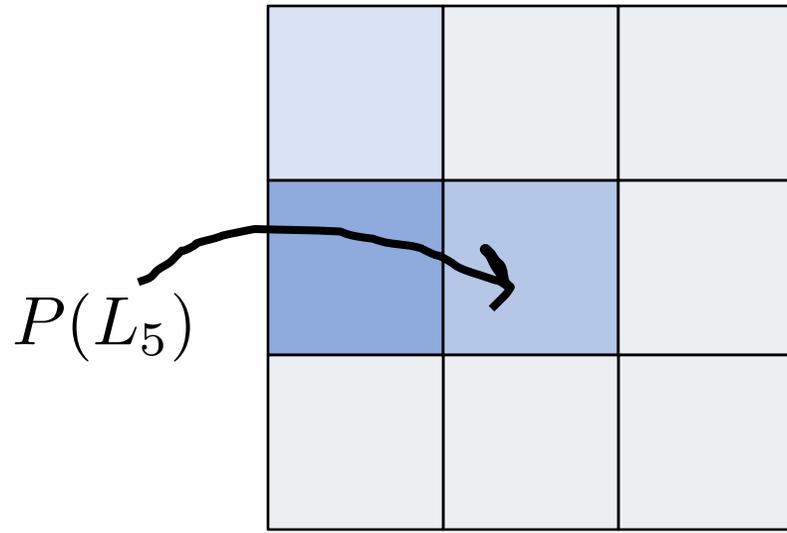
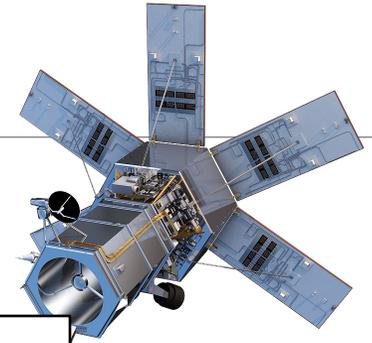


$P(L_5|O)$

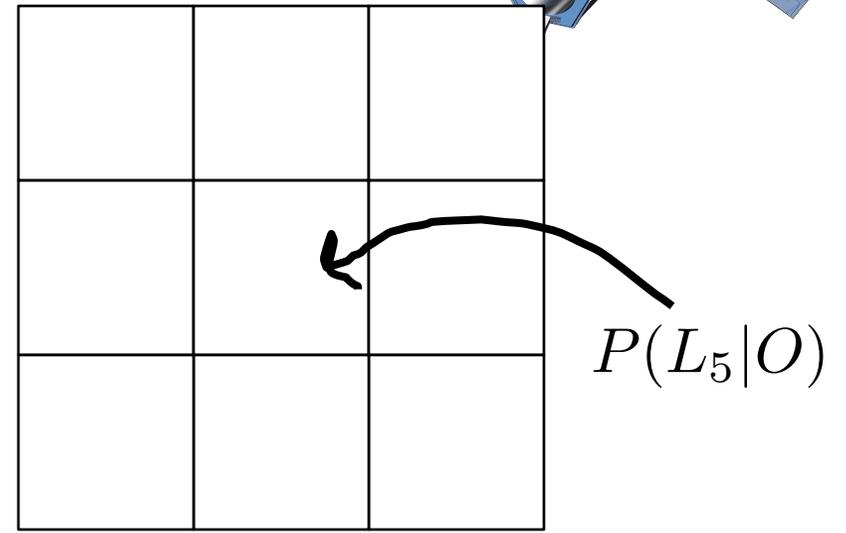
After Observation

$$P(L_5|O) = \frac{P(O|L_5)P(L_5)}{P(O)}$$

Bayes' Theorem and Location



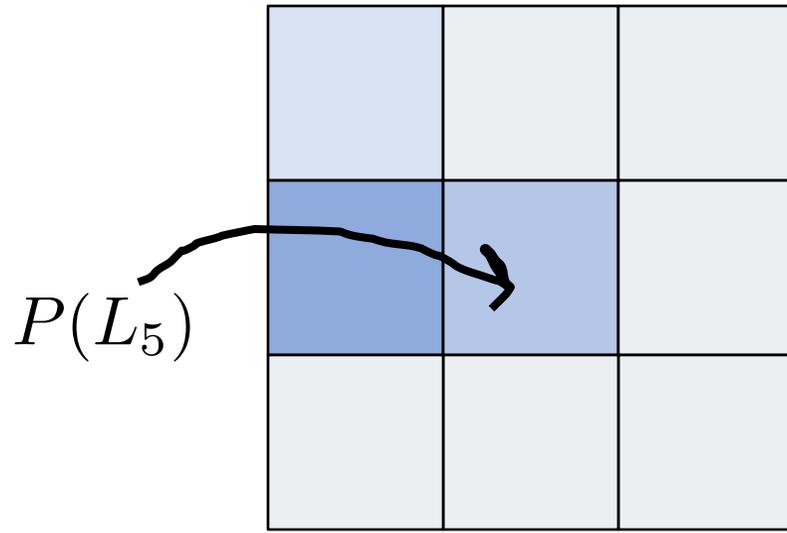
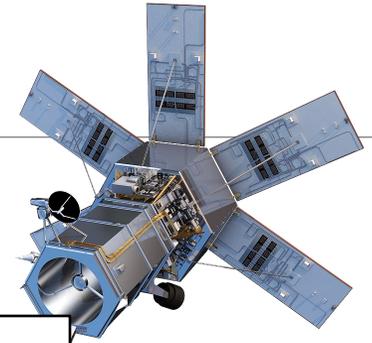
Before Observation



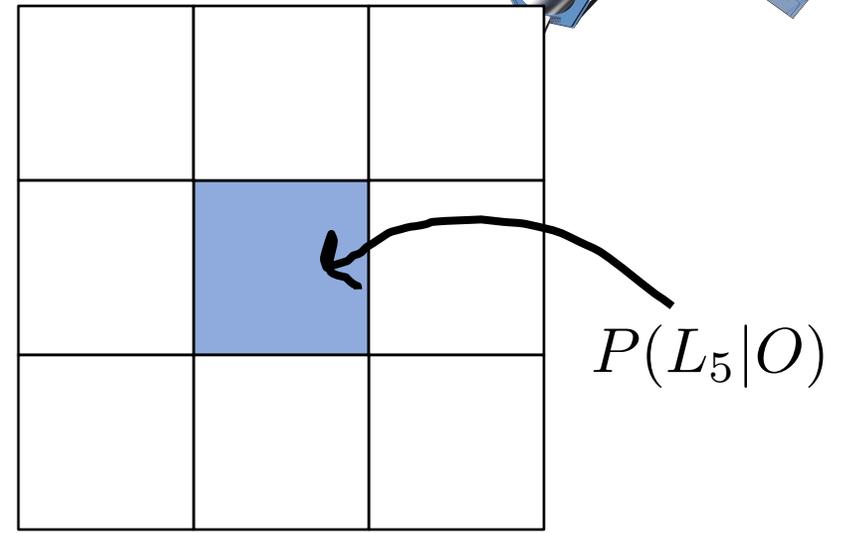
After Observation

$$P(L_5|O) = \frac{P(O|L_5)P(L_5)}{\sum_i P(O|L_i)P(L_i)}$$

Bayes' Theorem and Location



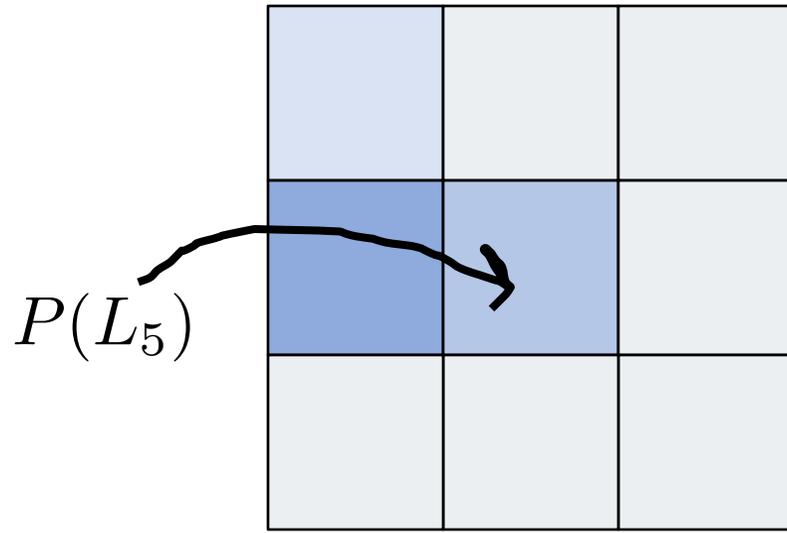
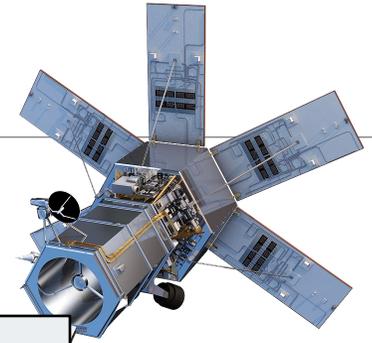
Before Observation



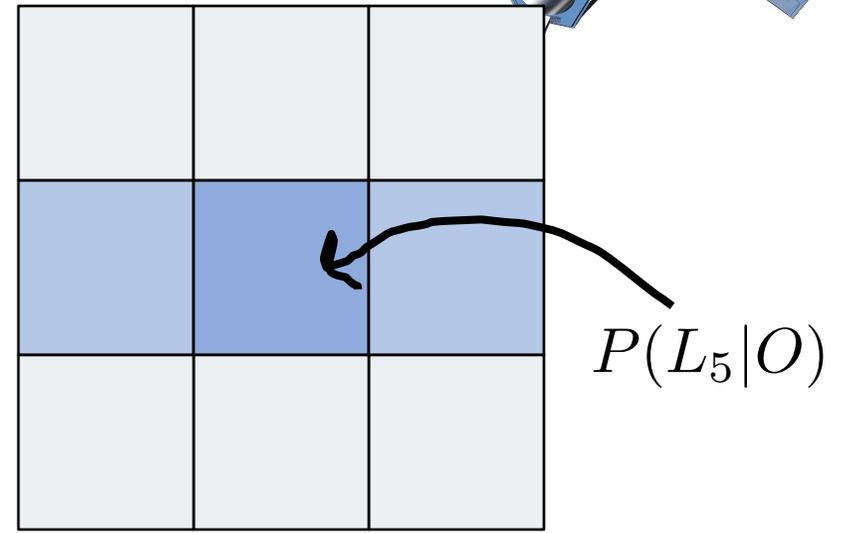
After Observation

$$P(L_5|O) = \frac{P(O|L_5)P(L_5)}{\sum_i P(O|L_i)P(L_i)}$$

Bayes' Theorem and Location

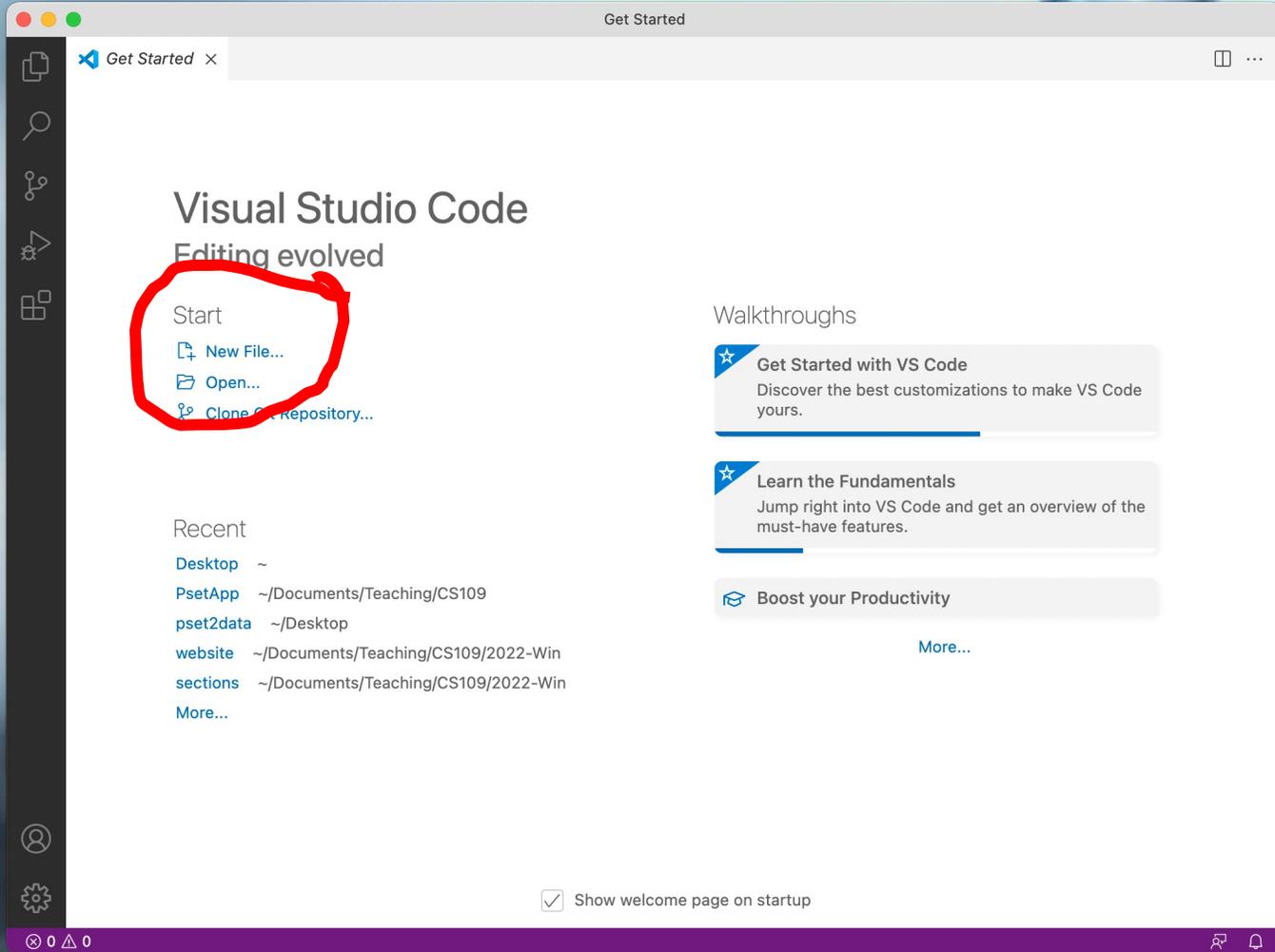
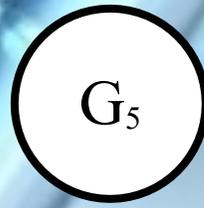
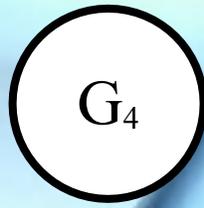
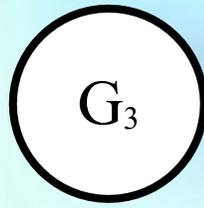
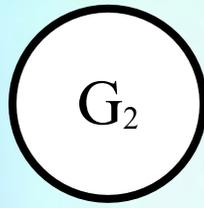


Before Observation



After Observation

$$P(L_5|O) = \frac{P(O|L_5)P(L_5)}{\sum_i P(O|L_i)P(L_i)}$$



Story: Ultimate Probability



Ultimate Probability

3,290 views • 1 Dec 2018

👍 14 💬 0 ➦ SHARE ⌵ SAVE ...



Maika Isogawa
21 subscribers

SUBSCRIBE

<https://www.maikaisogawa.com/ultimate-frisbee-probability/>

<https://www.youtube.com/watch?v=H2lfTwGisOg>

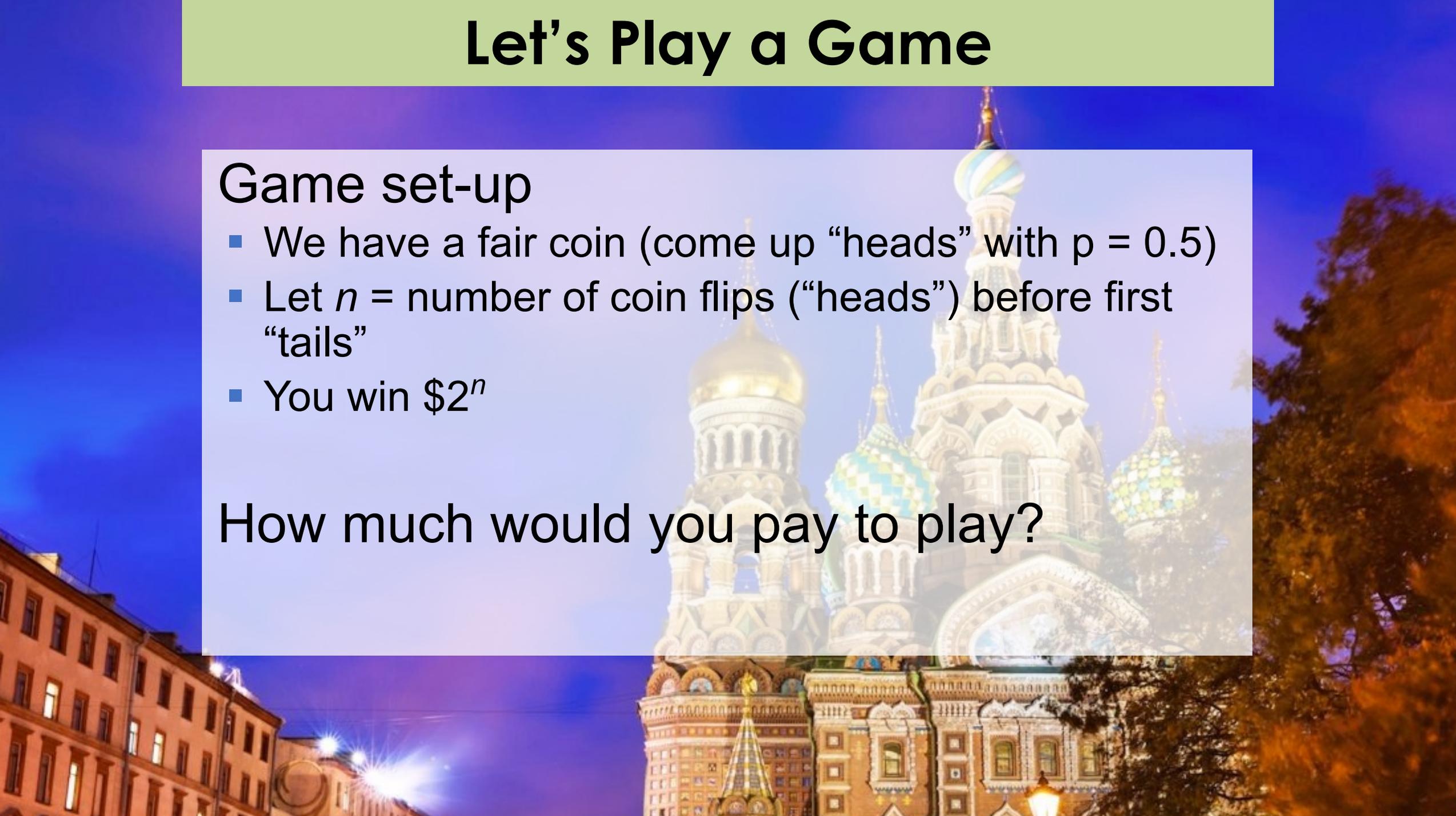


Let's Play a Game

Game set-up

- We have a fair coin (come up “heads” with $p = 0.5$)
- Let n = number of coin flips (“heads”) before first “tails”
- You win $\$2^n$

How much would you pay to play?



Review!

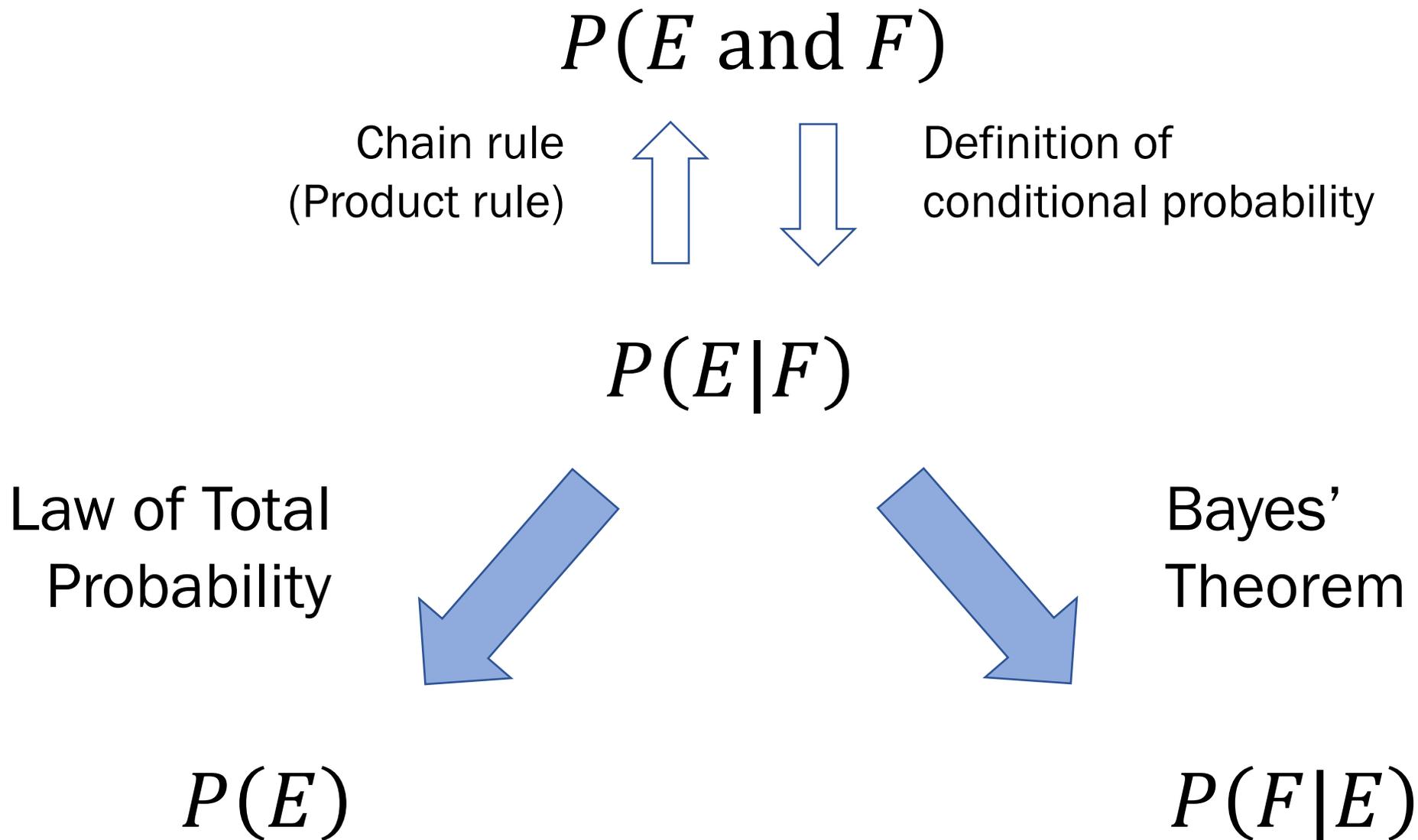
1. Review: Axiom Probability Tools



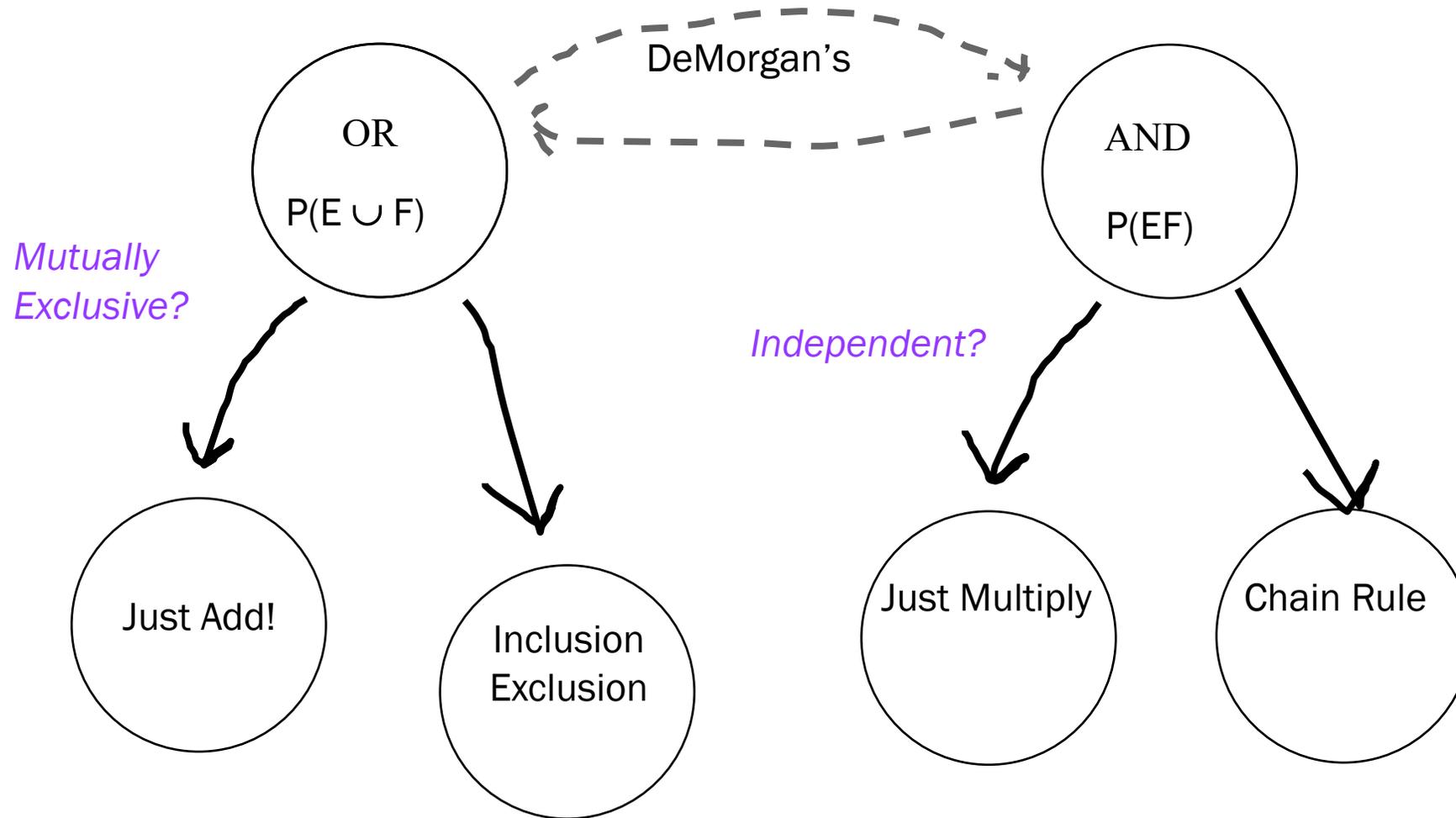
- Axiom 1: $0 \leq P(E) \leq 1$
- Axiom 2: $P(S) = 1$
- Identity 3: $P(E^c) = 1 - P(E)$



2. Review: Conditional Probability Tools



3. Review: Probability of Or and And



Practice: Lets do the Frisbee Problem.

You flip **two frisbees**. For each frisbee, the probability that it lands “**heads**” is **0.6**. The two frisbees are considered “**even**” if both frisbees are heads **or** both frisbees are tails.

What is the probability that the frisbees are even?



Lets go solve it!

Let p be the probability of a 1 from `unknown_random`.
What is the probability of a 1 from `fair_random` if $p = .45$?

```
def fair_random():  
    """  
    There are four outcomes for assignments to r1 and r2:  
    [0, 0], [0, 1], [1, 0], [1, 1]. Return 1 if the  
    outcomes are [0, 0] or [1, 1]  
    """  
    r1 = unknown_random()  
    r2 = unknown_random()  
    return r1 == r2
```

https://cs109psets.netlify.app/fall22/pset2/fairRandom_a





Conditional Paradigm: In the conditional paradigm, the *formulas* of probability are preserved.



BAE's Theorem

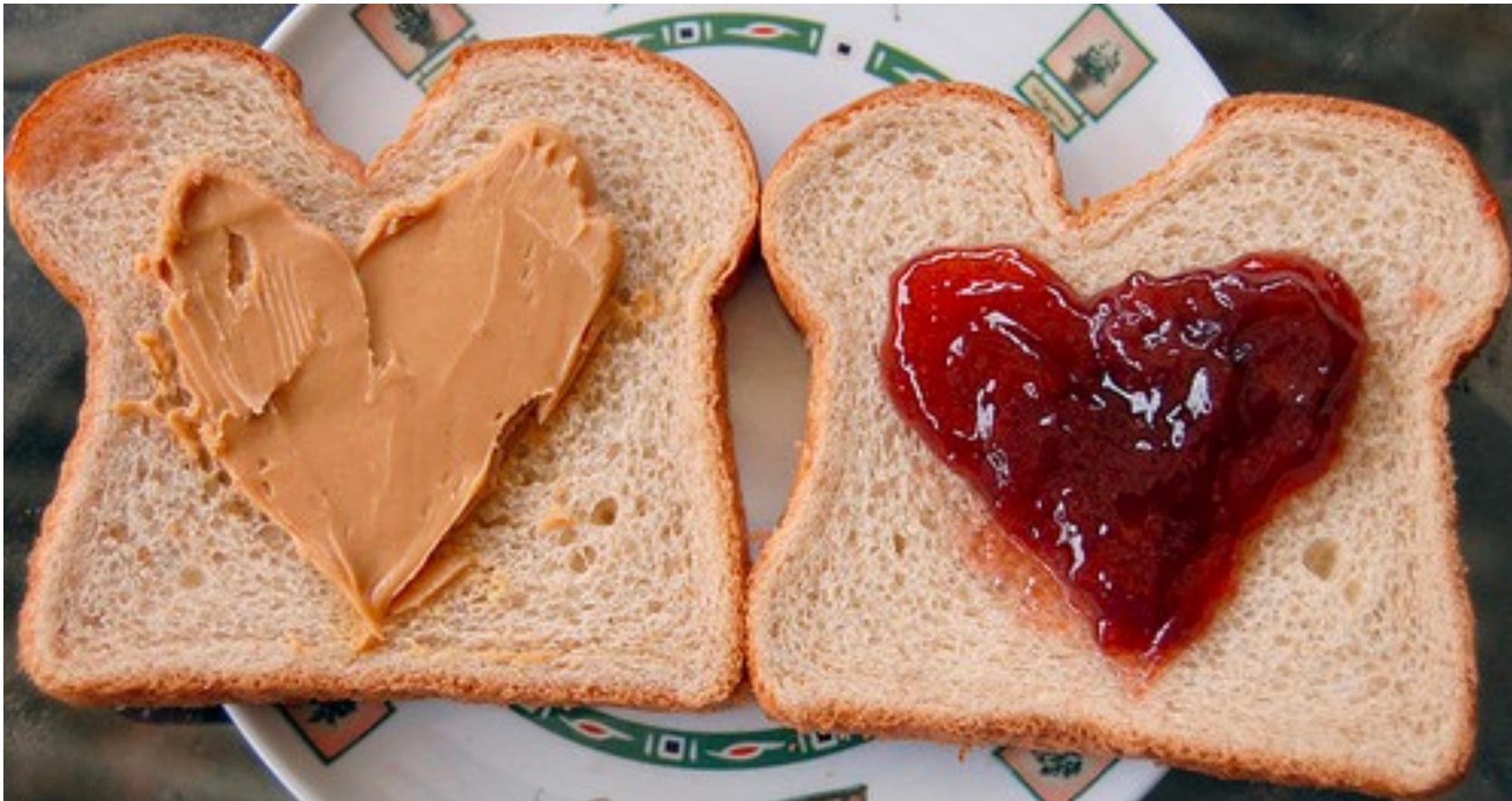
$$P(A | B \cap E) = \frac{P(B | A \cap E) P(A | E)}{P(B | E)}$$



Two Great Tastes

Conditional Probability

Independence



G_1

G_2

G_3

G_4

G_5

T



G₁

G₂

G₃

G₄

G₅

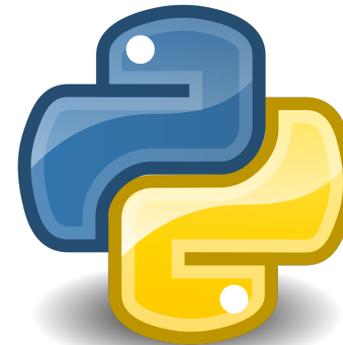
T

```
dna.txt — dna
dna.txt
1 False, True, False, False, True, False
2 True, True, False, True, True, False
3 True, True, False, True, True, True
4 False, True, False, True, True, False
5 False, True, False, False, True, False
6 True, True, False, True, True, True
7 False, False, True, False, False, False
8 False, False, True, False, True, False
9 True, False, False, True, False, False
10 False, True, False, True, True, False
11 True, False, False, True, False, False
12 True, False, True, True, False, False
13 False, True, False, False, True, False
14 False, False, True, True, False, False
15 True, True, False, False, True, True
16 True, False, True, True, False, False
17 True, True, True, True, True, True |
18 True, False, True, False, False, True
19 False, True, False, True, True, True
20 False, False, True, False, False, False
21 False, False, False, True, True, False
22 False, True, False, False, True, False
23 True, True, False, True, True, True
24 False, True, False, True, True, False
25 True, False, False, False, False, True
26 False, False, True, True, False, True
27 False, False, False, True, False, False
28 False, True, True, False, False, True
29 False, True, False, False, True, True
30 False, False, False, False, False, True
31 False, True, False, True, True, False
32 True, False, False, True, False, False
33 True, True, False, True, True, True
34 True, True, False, False, True, True
35 True, True, False, True, True, True
36 False, False, False, True, False, False
--
```



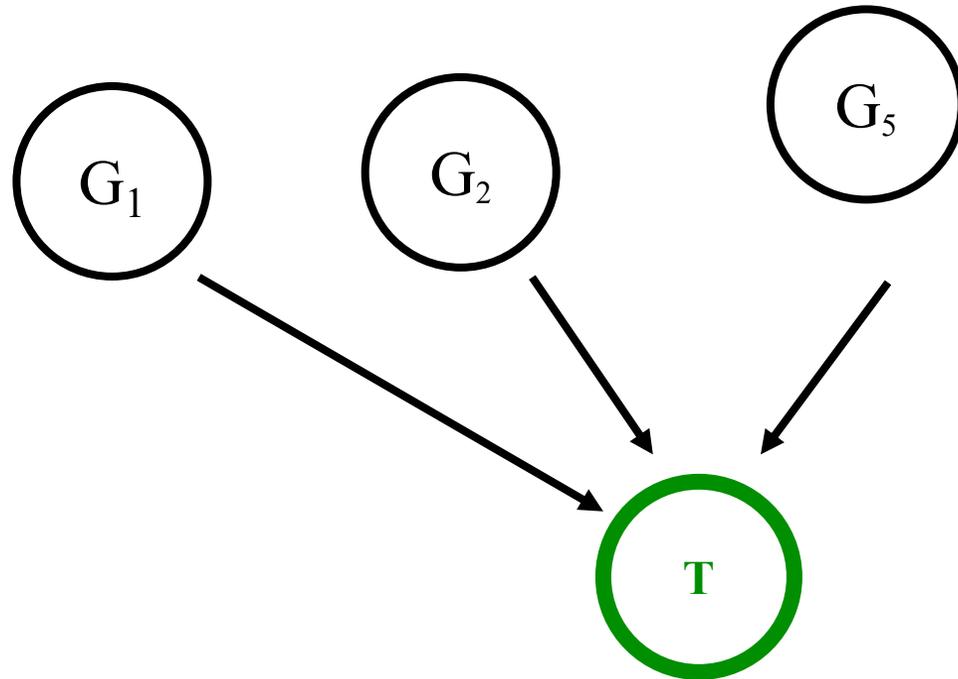
100,000
samples

6 observations per sample

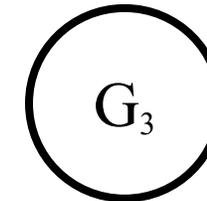
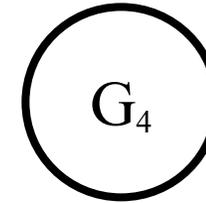


Independence Rules out Causality Structures

These genes do
impact T (are not
independent)



These genes don't impact T
(are independent)



Teaser for Future Classes: Conditional Independence

```
[Piech-2:dna piech$ python findStructure.py
size data = 100000
p(G1) = 0.500
p(G2) = 0.545
p(G3) = 0.299
p(G4) = 0.701
p(G5) = 0.600
p(T) = 0.290
```

```
p(T and G1) = 0.291 , P(T)p(G1) = 0.195
p(T and G2) = 0.300 , P(T)p(G2) = 0.213
p(T and G3) = 0.116 , P(T)p(G3) = 0.117
p(T and G4) = 0.273 , P(T)p(G4) = 0.273
p(T and G5) = 0.309 , P(T)p(G5) = 0.234
```

...

```
p(T and G5 | G2) = 0.450
p(T | G2)p(G5 | G2) = 0.450
```



Advanced Concept:
Conditional Independence

Teaser for Future Classes: Conditional Independence

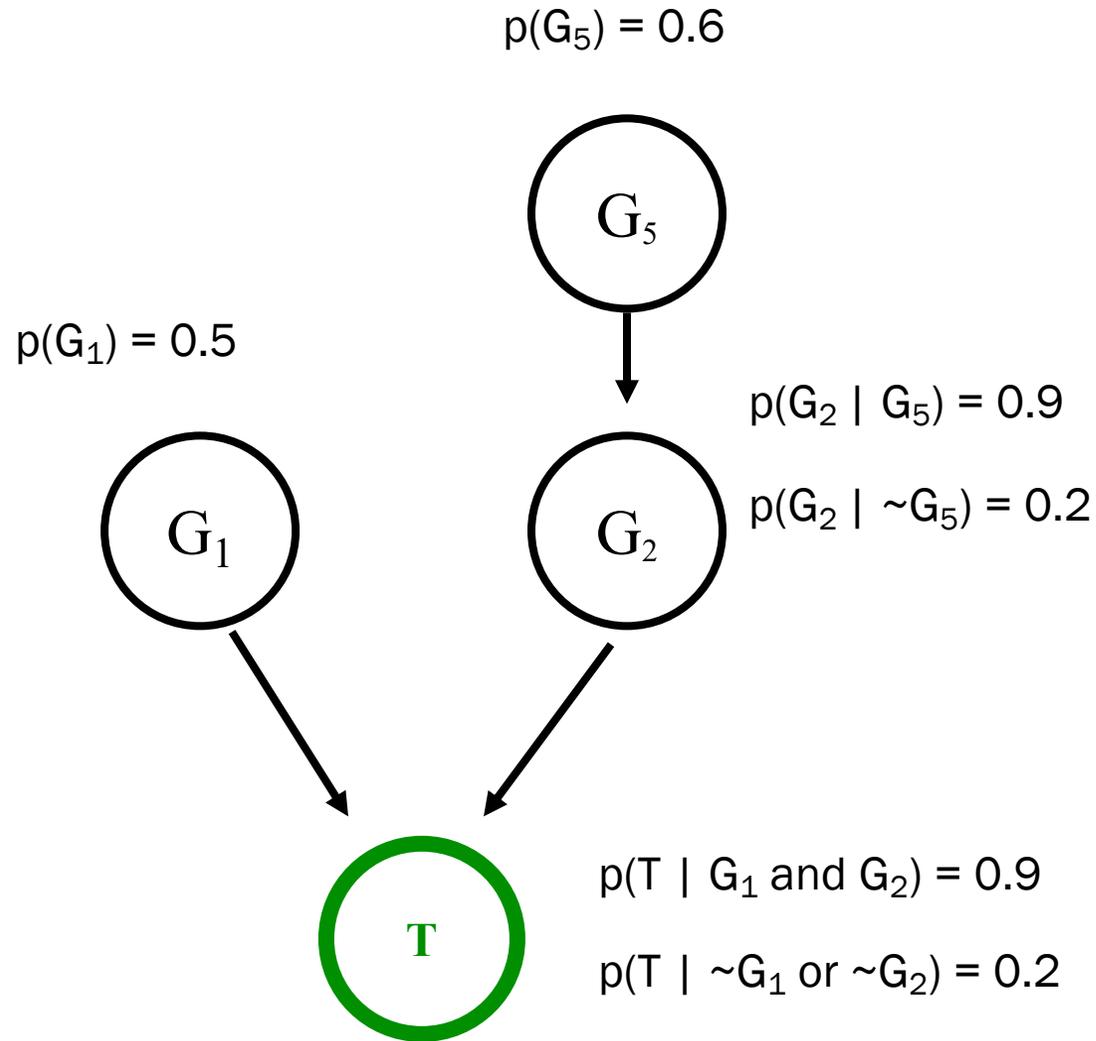
```
[Piech-2:dna piech$ python findStructure.py
size data = 100000
p(G1) = 0.500
p(G2) = 0.545
p(G3) = 0.299
p(G4) = 0.701
p(G5) = 0.600
p(T) = 0.390
p(T and G1) = 0.291 , P(T)p(G1) = 0.195
p(T and G2) = 0.300 , P(T)p(G2) = 0.213
p(T and G3) = 0.116 , P(T)p(G3) = 0.117
p(T and G4) = 0.273 , P(T)p(G4) = 0.273
p(T and G5) = 0.309 , P(T)p(G5) = 0.234
```

...

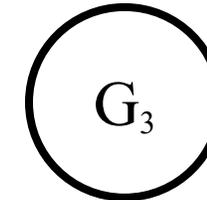
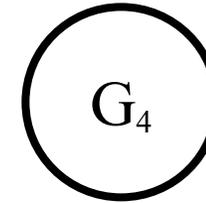
$$p(T \text{ and } G5 \mid G2) = 0.450$$
$$p(T \mid G2)p(G5 \mid G2) = 0.450$$



Only Causal Structure That Fits



These genes don't impact T
(are independent)



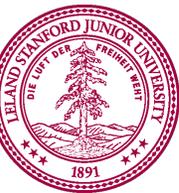
Advanced: Conditional Independence

Two events E and F are called **conditionally independent given G** , if

$$P(EF|G) = P(E|G)P(F|G)$$

Or, equivalently if:

$$P(E|FG) = P(E|G)$$

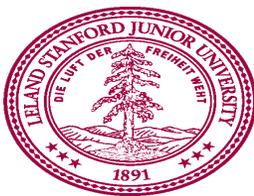


Conditional Paradigm: In the conditional paradigm, the *formulas* of probability are preserved.



Independence
relationships can change
with conditioning.

If E and F are independent, that does not mean they will still be independent given another event G .



NETFLIX

And Learn

Netflix Learning

What is the probability
that a user will watch
Life is Beautiful?

$$P(E)$$



$$P(E) = \lim_{n \rightarrow \infty} \frac{n(E)}{n} \approx \frac{\text{\#people who watched movie}}{\text{\#people on Netflix}}$$

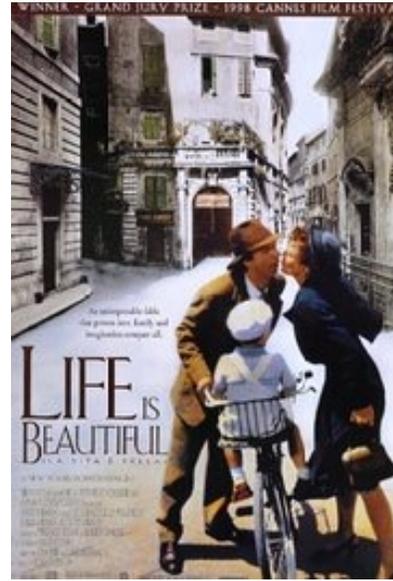
$$P(E) = 10,234,231 / 50,923,123 = 0.20$$



Netflix Learning

What is the probability that a user will watch Life is Beautiful, given they watched Coda?

$$P(E|F)$$



$$P(E|F) = \frac{P(EF)}{P(F)} = \frac{\text{\#people who watched both}}{\text{\#people who watched } F}$$

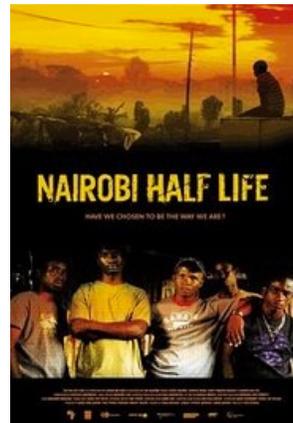
$$P(E|F) = 0.42$$



Conditioned on liking a set of movies?

Netflix Learning

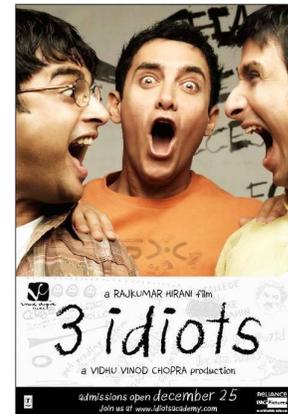
Each event corresponds to liking a particular movie



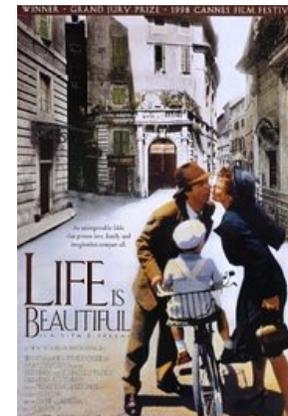
E_1



E_2



E_3



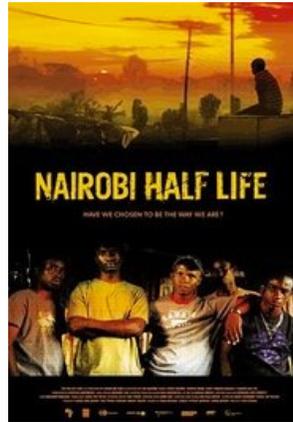
E_4

$$P(E_4 | E_1, E_2, E_3)?$$

Is E_4 independent of E_1, E_2, E_3 ?

Netflix Learning

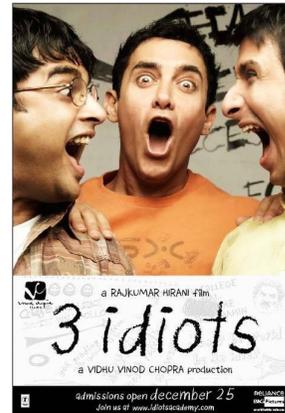
Is E_4 independent of E_1, E_2, E_3 ?



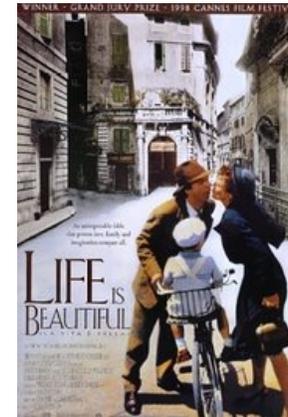
E_1



E_2



E_3

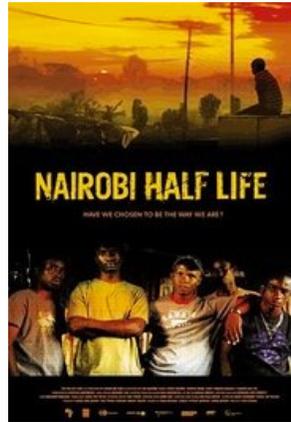


E_4

$$P(E_4|E_1, E_2, E_3) \stackrel{?}{=} P(E_4)$$

Netflix Learning

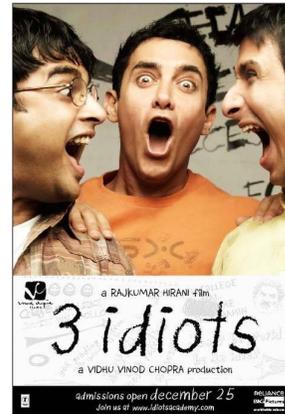
Is E_4 independent of E_1, E_2, E_3 ?



E_1



E_2



E_3



E_4

$$P(E_4|E_1, E_2, E_3) = \frac{P(E_1 E_2 E_3 E_4)}{P(E_1 E_2 E_3)}$$



Netflix Learning

What is the probability that a user watched four particular movies?

- There are 13,000 titles on Netflix
- The user watches 30 random titles.
- E = movies watched include the given four.

Solution:

$$P(E) = \frac{\binom{4}{4} \binom{12996}{24}}{\binom{13000}{30}} = 10^{-11}$$

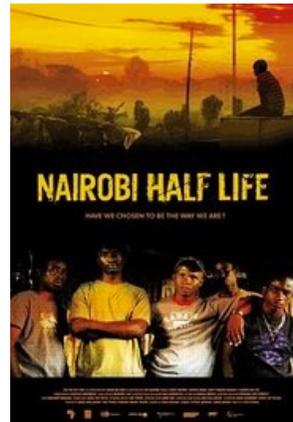
Watch those four

Choose 24 movies not in the set

Choose 30 movies from netflix



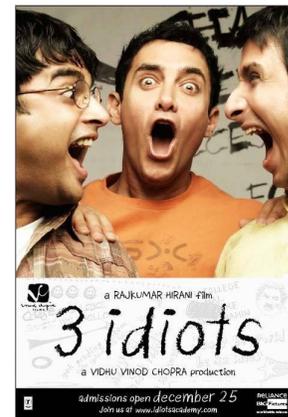
Netflix Learning



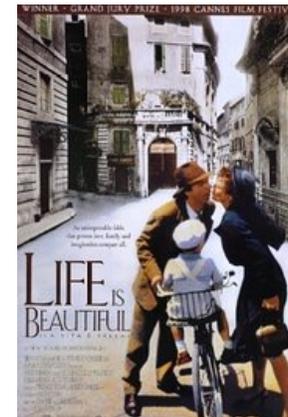
E_1



E_2



E_3

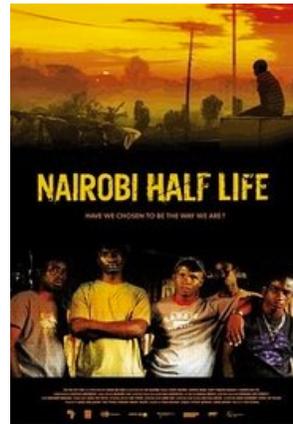
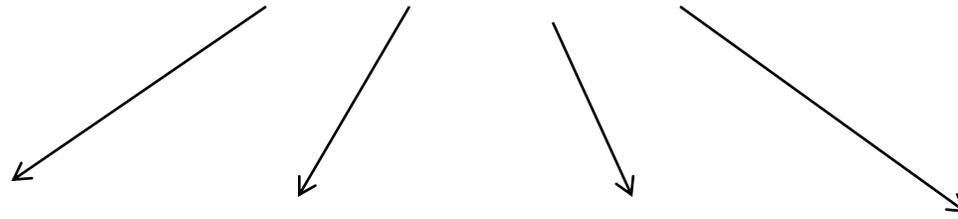


E_4

Netflix Learning: Advanced, Conditional Independence

K_1

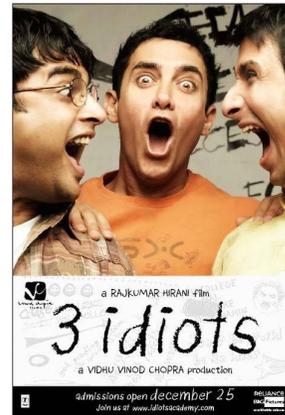
Like foreign emotional comedies



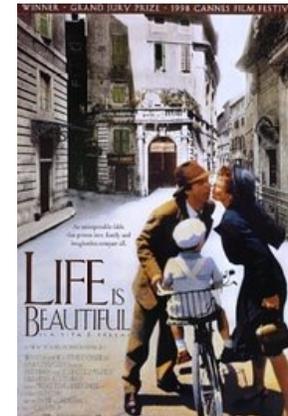
E_1



E_2



E_3



E_4

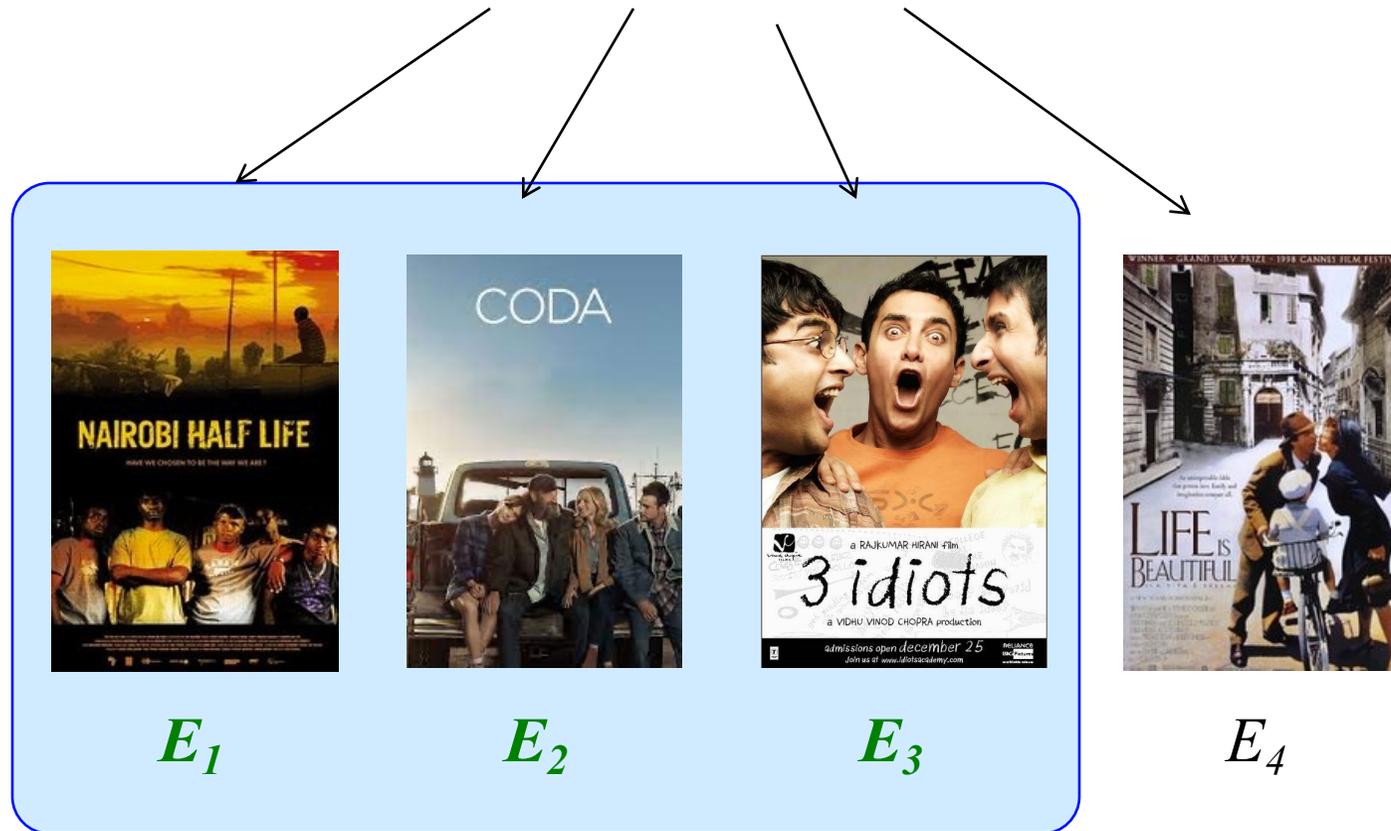
Assume E_1 , E_2 , E_3 and E_4 are conditionally independent given K_1



Netflix Learning: Advanced, Conditional Independence

K_1

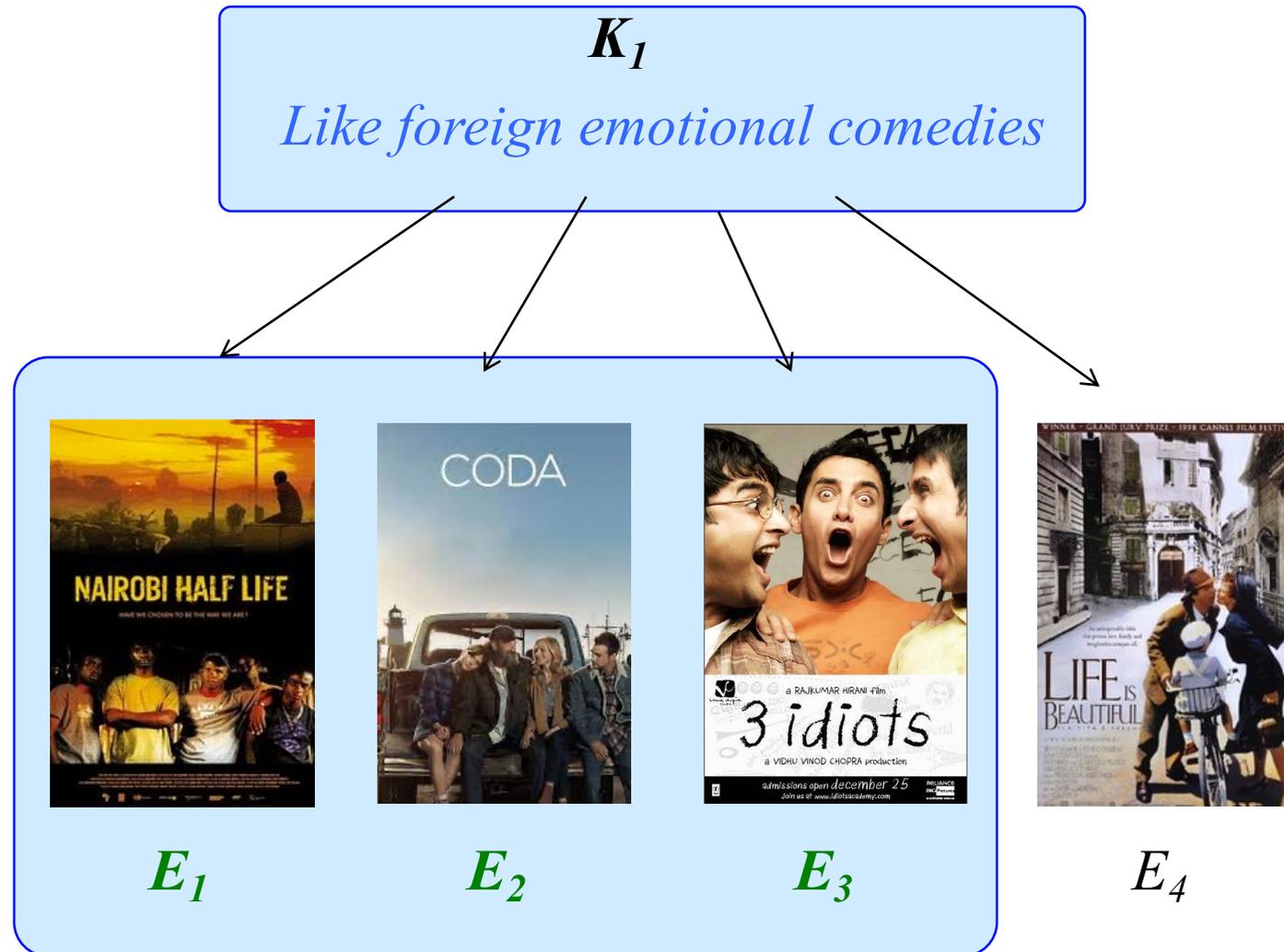
Like foreign emotional comedies



Assume E_1, E_2, E_3 and E_4 are conditionally independent given K_1

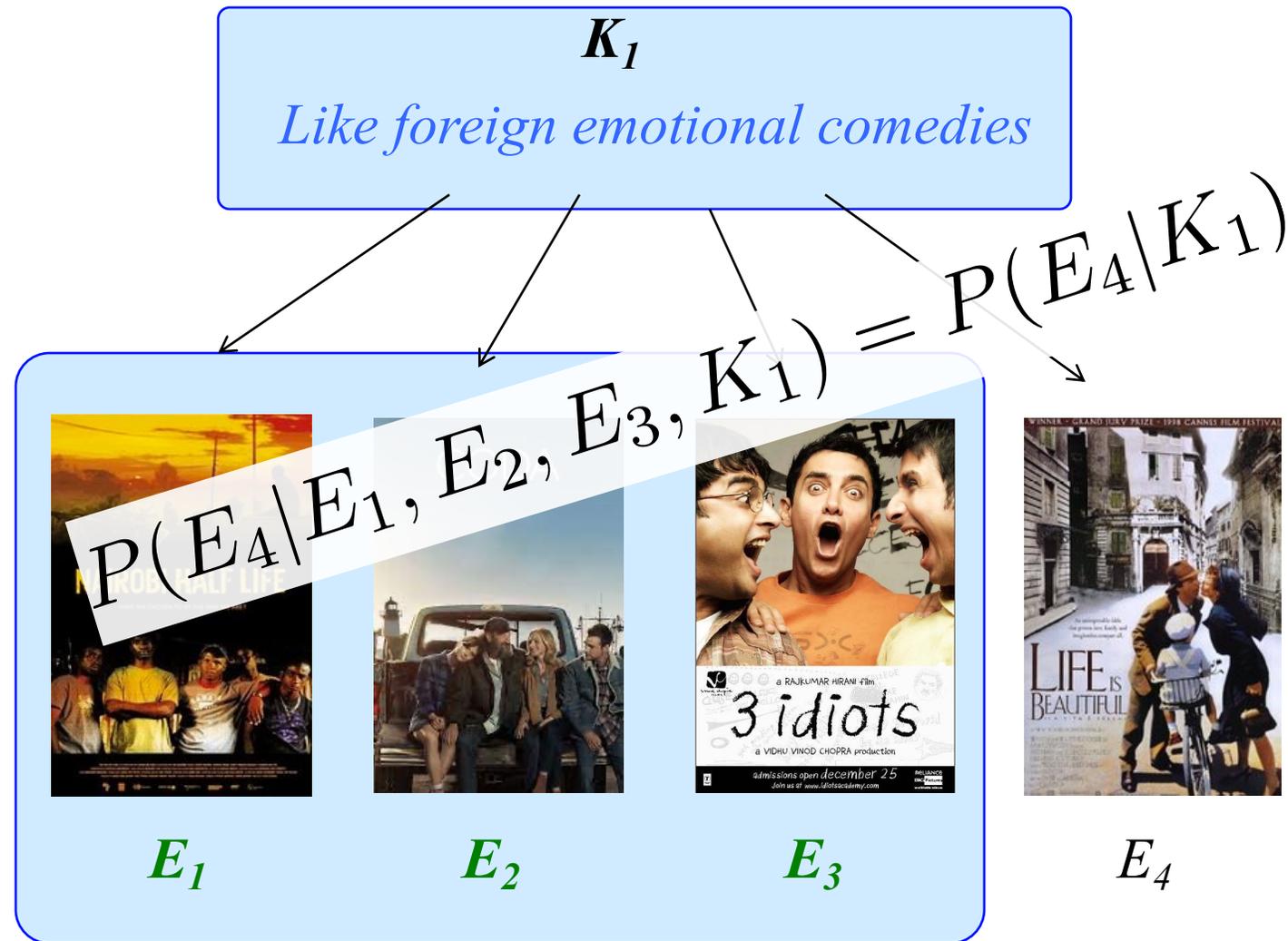


Netflix Learning: Advanced, Conditional Independence



Assume E_1 , E_2 , E_3 and E_4 are conditionally independent given K_1

Netflix Learning: Advanced, Conditional Independence



Assume E_1, E_2, E_3 and E_4 are conditionally independent given K_1

Conditional independence is a practical, real world way of decomposing hard probability questions.

Conditional Independence



If E and F are
dependent,

that does not mean E and
 F will be dependent
when another event is
observed.



Conditional Dependence



If E and F are independent,

that does not mean E and F will be independent when another event is observed.

Big Deal!

“Exploiting *conditional independence* to generate fast probabilistic computations is one of the main contributions CS has made to probability theory”

-Judea Pearl wins 2011 Turing Award, “*For fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning*”



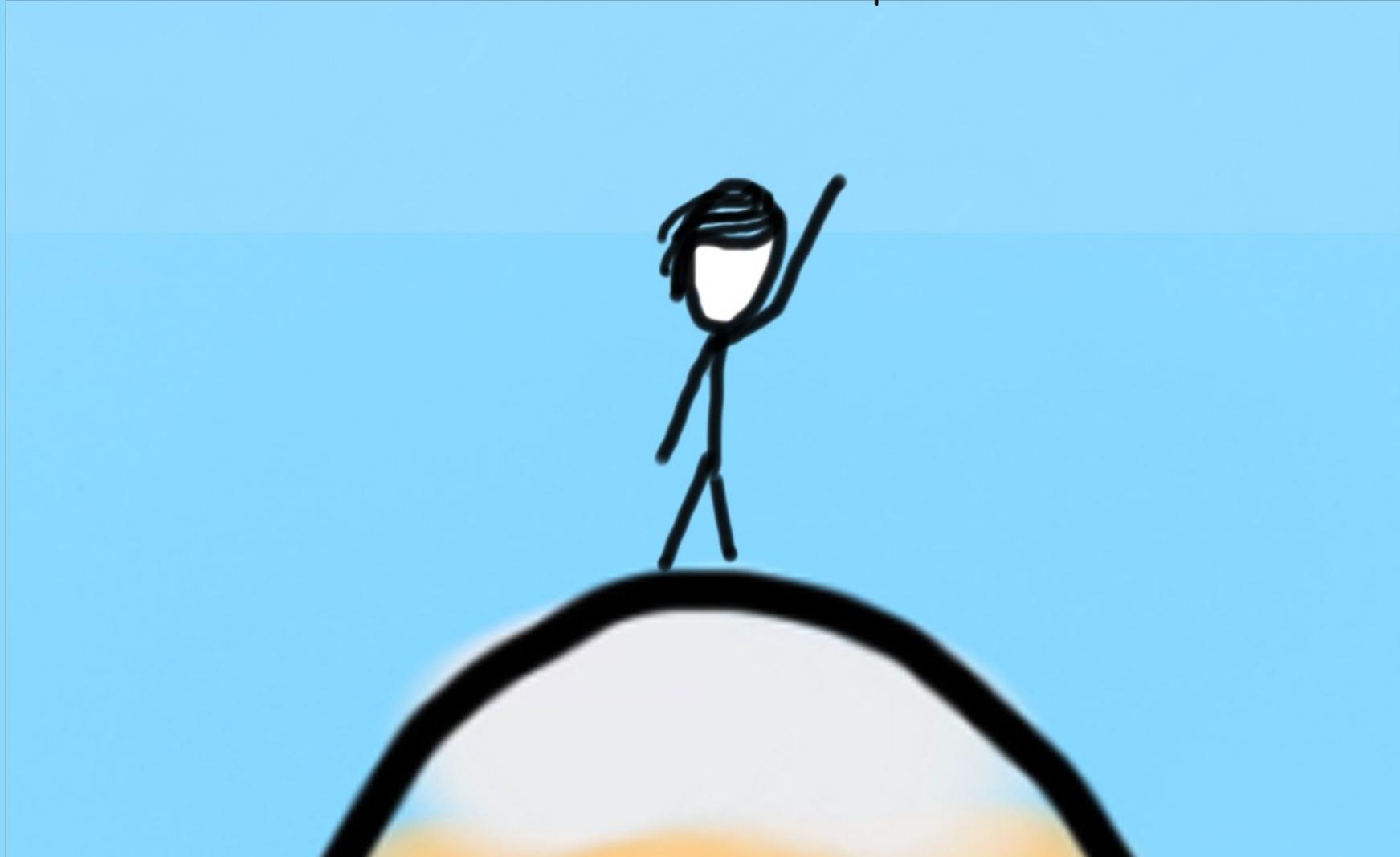
Ready for the next (cs109) episode

Pedagogical Pause!

Random Variables

Learning Goals

1. Be able to define a random variable (R.V.)
2. Be able to use and produce a PMF of a R.V.
3. Be able to calculate the expectation of the R.V.



Remember Learning to Code

type

name

value

```
int a = 5;  
double b = 4.2;  
bit c = 1;  
choice d = medium;
```

$z \in \{\text{high, medium, low}\}$

Random variables are like programming variables, with uncertainty



Pirates of the Variables

```
int a = 5;
```

A is the number of pirate ships in our *future* armada.

$$A \in \{1, 2, \dots, 10\}$$



```
double b = 4.2;
```

B is the amount of money we get *after* we defeat Blackbeard.

$$B \in \mathbb{R}^+$$



```
bit c = 1;
```

C is 1 *if* we successfully raid Isla de Muerta. 0 otherwise.

$$C \in \{0, 1\}$$



Random Variable

A **Random Variable** is a variable that will have a value. But there is uncertainty as to what value.

Example:

- 3 fair coins are flipped.
- Y = number of “heads” on 3 coins
- **Y is a random variable**
- $P(Y = 0) = 1/8$ (T, T, T)
- $P(Y = 1) = 3/8$ (H, T, T), (T, H, T), (T, T, H)
- $P(Y = 2) = 3/8$ (H, H, T), (H, T, H), (T, H, H)
- $P(Y = 3) = 1/8$ (H, H, H)
- $P(Y \geq 4) = 0$



It is confusing that both random variables
and events use the same notation

Example of Random Variables

Consider 5 coin flips, each which independently come up heads with probability p

- Recall:

$$P(2 \text{ heads}) = \binom{5}{2} p^2 (1 - p)^3$$

$$P(3 \text{ heads}) = \binom{5}{3} p^3 (1 - p)^2$$

- $Y =$ number of “heads” on 5 flips

$$Y \in \{1, 2, \dots, 5\}$$

$$P(Y = k) = \binom{5}{k} p^k (1 - p)^{5-k}$$

* Pro tip: Not really about coins. Many real world binary events work like this.



Properties of Random Variables

Probability Mass Function:

$$P(X = a)$$

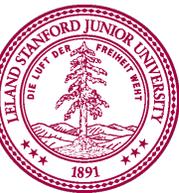
Expectation:

$$E[X]$$

Variance:

$$\text{Var}(X)$$

Learning
goals for
today



1. Probability Mass Function

The relationship between values a random variable can take on, and the corresponding probability, is a *function*!

Let Y be a random variable



Y

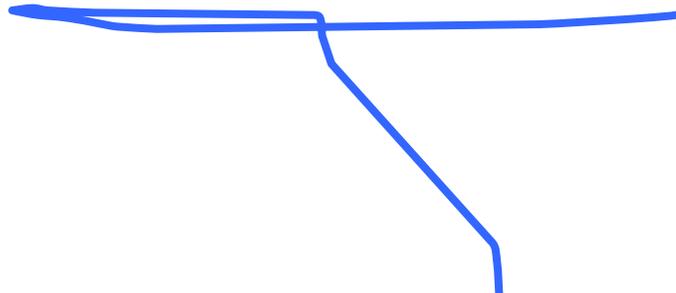
For example Y is the number of heads in 5 coin flips

$$Y = 2$$

It is an *event* when
Y takes on a value

For example Y is the number of heads in 5 coin flips

If this is a number


$$P(Y = 2)$$


Then this is a number
(between 0 and 1)

For example Y is the number of heads in 5 coin flips

If this is a variable

$$P(Y = k)$$

Then this is a function

For example Y is the number of heads in 5 coin flips

Random Variables -> Functions

$$P(Y = k)$$

The diagram illustrates the evaluation of a probability function. A blue arrow points from the expression $k = 5$ to the variable k in the function $P(Y = k)$. A second blue arrow points from the function $P(Y = k)$ to the numerical value 0.03125 .

$$k = 5$$
$$0.03125$$

For example Y is the number of heads in 5 coin flips

Random Variables -> Functions

$$P(Y = k)$$

```
def event_probability(k):  
    # probability mass function of Y in python  
    N = 5    # number of coin flips  
    P = 0.5  # probability of heads  
  
    ways = math.comb(N, k);  
    prob_heads = math.pow(P, k)  
    prob_tails = math.pow(P, N-k)  
    return ways * a * b
```

For example Y is the number of heads in 5 coin flips



If a random variable is discrete we call this function the **Probability Mass Function**



Probability Mass Function (PMF)

Let X be a random variable that represents the result of a **single dice roll**. X can take on the values $\{1, 2, 3, 4, 5, 6\}$

$$P(X = x)$$

$$p(x)$$

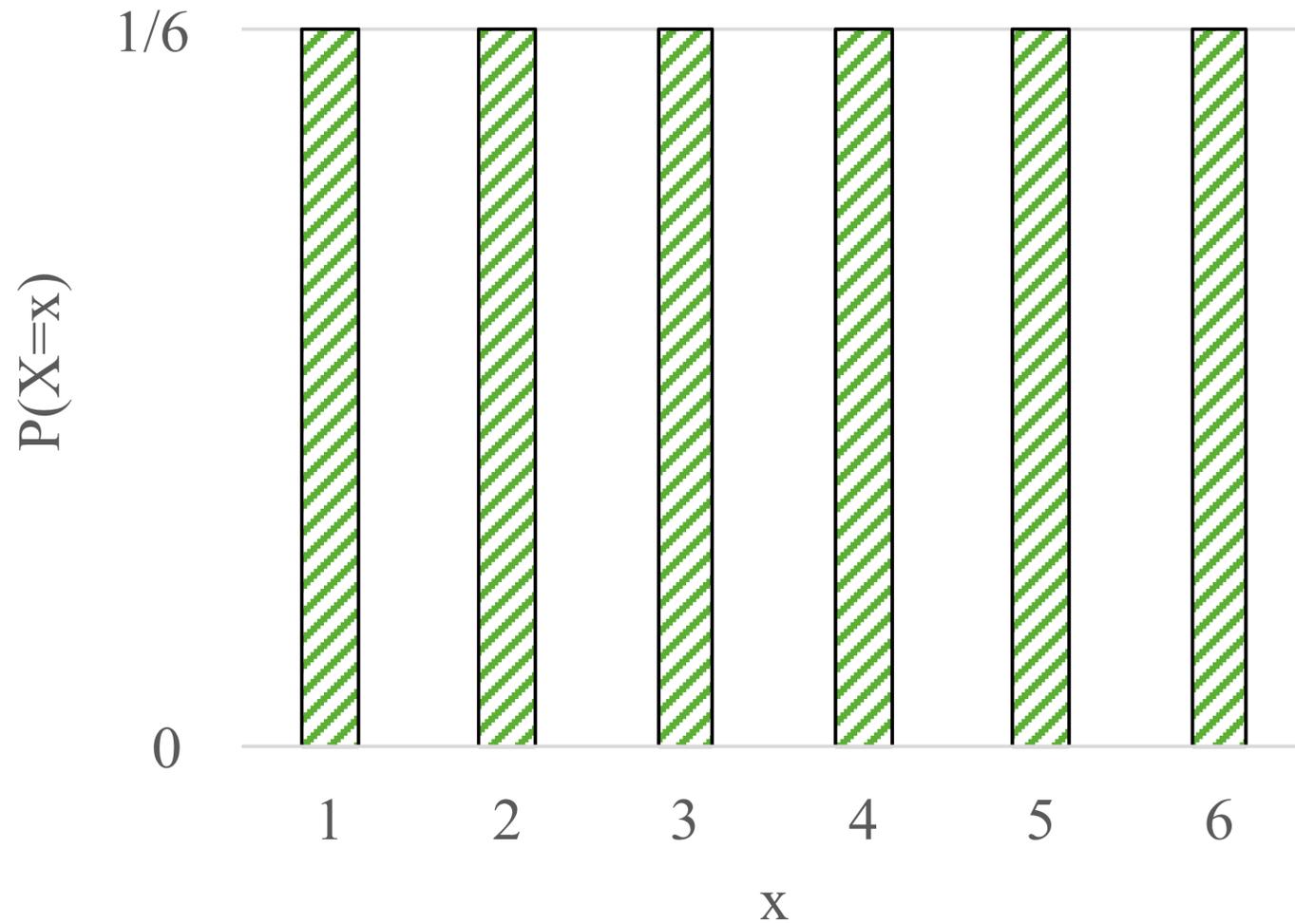
This is shorthand notation for the PMF

$$p_X(x)$$

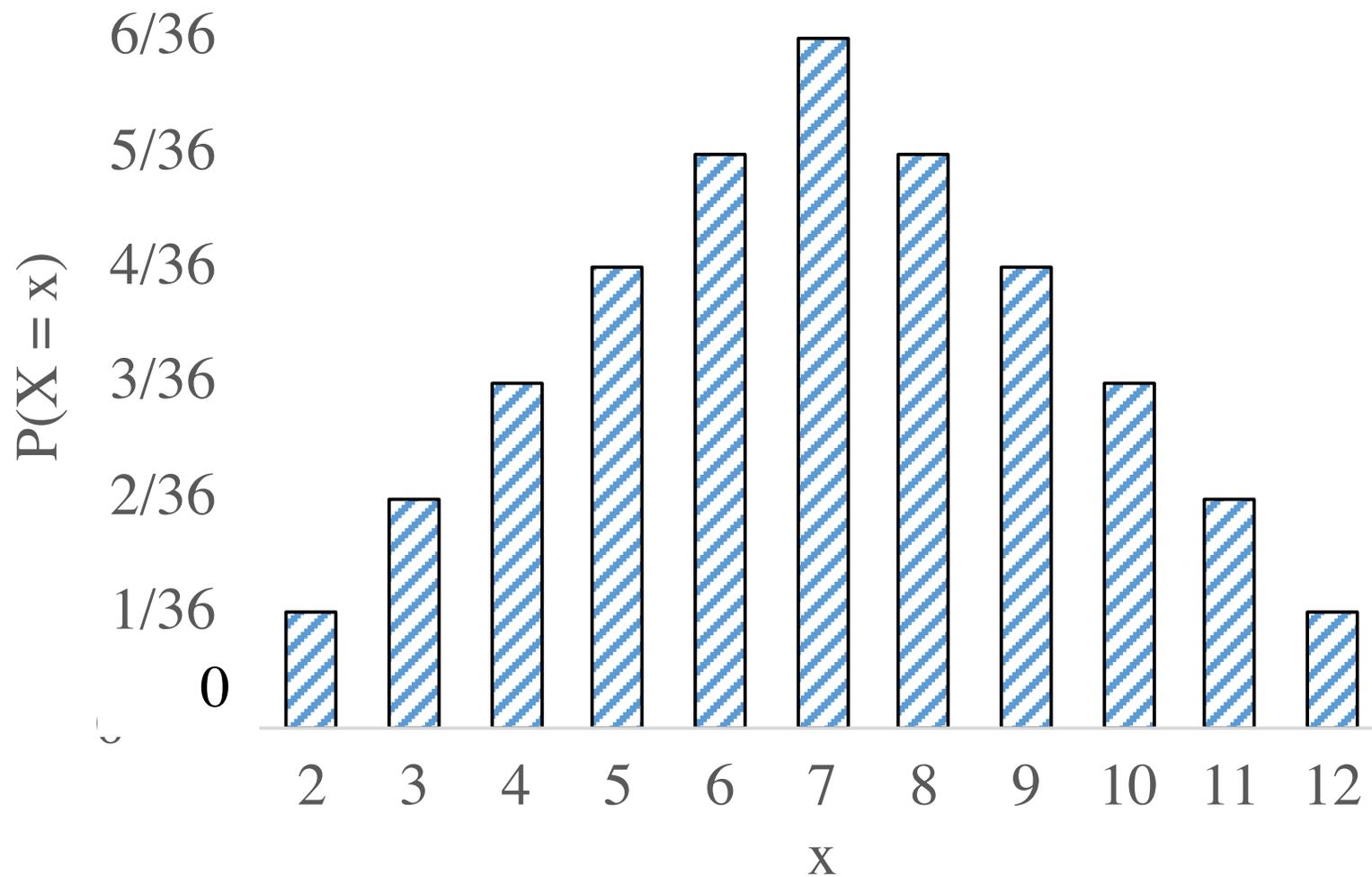
This is also shorthand notation for the PMF



PMF for X the outcome of a die roll



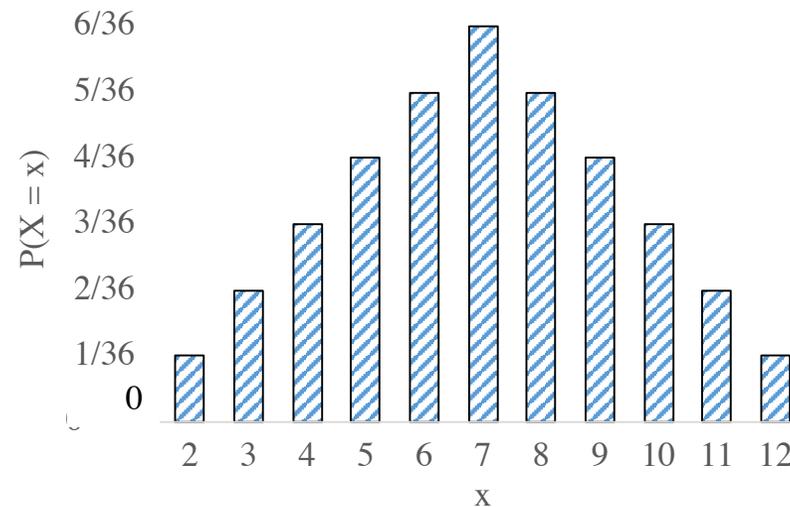
PMF for X the sum of two dice rolls



PMF as an equation

$$p(X = x) = \begin{cases} \frac{x-1}{36} & \text{if } x \in \mathbb{Z}, 1 \leq x \leq 6 \\ \frac{13-x}{36} & \text{if } x \in \mathbb{Z}, 7 \leq x \leq 12 \\ 0 & \text{else} \end{cases}$$

Again, this is the probability for the sum of two dice



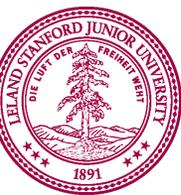
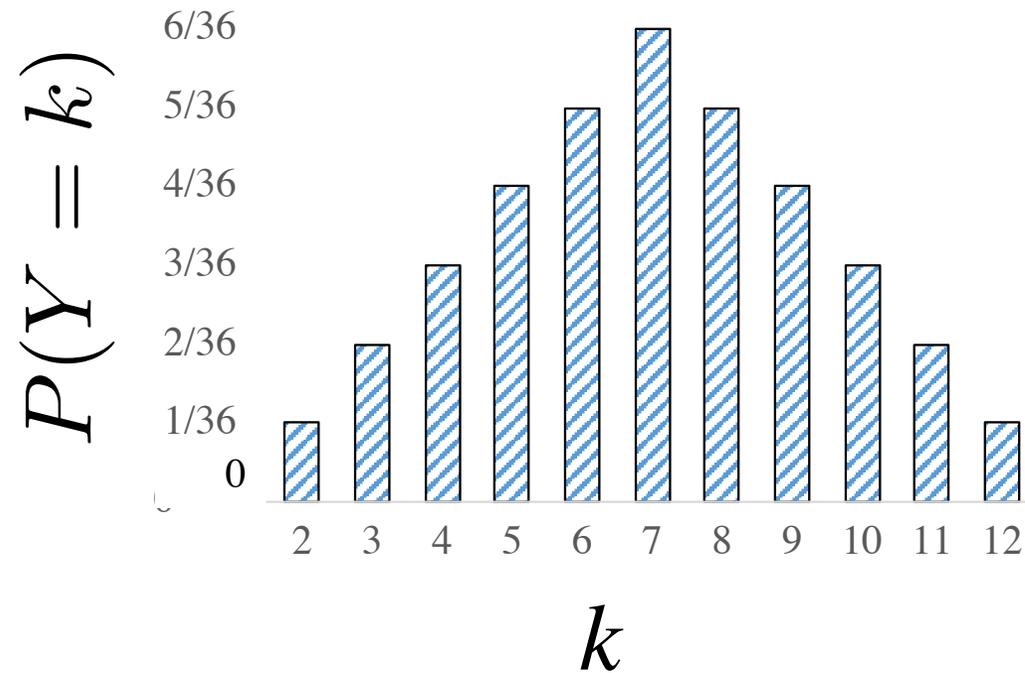
This is Fine Check

$$\sum_{\text{all } k} P(Y = k) \stackrel{?}{=} 1$$



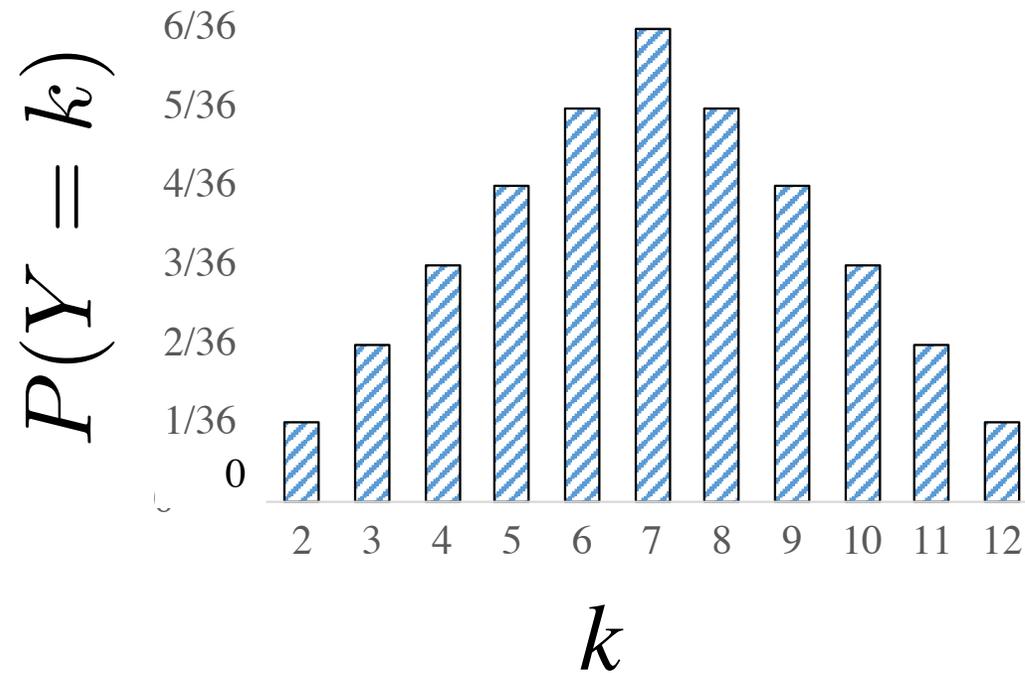
This is Fine Check

$$\sum_{\text{all } k} P(Y = k) \stackrel{?}{=} 1$$



This is Fine Check

$$\sum_k P(Y = k) = 1$$



2. Expectation

Properties of Random Variables

Probability Mass Function:

$$P(X = a)$$

Expectation:

$$E[X]$$

Variance:

$$\text{Var}(X)$$

Learning
goals for
today



Expected Value

The Expected Values for a discrete random variable X is defined as:

$$E[X] = \sum_{x:p(x)>0} x \cdot p(x)$$

Note: sum over all values of x that have $p(x) > 0$.

Expected value also called: **Mean**, *Expectation*, **Weighted Average**, **Center of Mass**, *1st Moment*



Expected Value

Roll a 6-Sided Die. X is outcome of roll

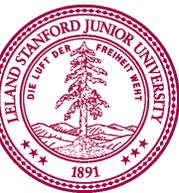
- $p(X=1) = p(X=2) = p(X=3) = p(X=4) = p(X=5) = p(X=6) = 1/6$

$$E[X] = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) = \frac{7}{2}$$

Y is random variable

- $P(Y = 1) = 1/3, \quad P(Y = 2) = 1/6, \quad P(Y = 3) = 1/2$

$$E[Y] = 1 (1/3) + 2 (1/6) + 3 (1/2) = 13/6$$



Lying With Statistics

“There are three kinds of lies:
lies, damned lies, and statistics”

– *Mark Twain*

School has 3 classes with 5, 10 and 150 students

Randomly choose a class with equal probability

X = size of chosen class

What is $E[X]$?

- $E[X] = 5 (1/3) + 10 (1/3) + 150 (1/3)$
 $= 165/3 = 55$



Lying With Statistics Part 2

“There are three kinds of lies:
lies, damned lies, and statistics”

– *Mark Twain*

School has 3 classes with 5, 10 and 150 students

Randomly choose a student with equal probability

Y = size of class that student is in

What is $E[Y]$?

- $E[Y] = 5 (5/165) + 10 (10/165) + 150 (150/165)$
 $= 22635/165 \approx 137$

Note: $E[Y]$ is students' perception of class size

- But $E[X]$ is what is usually reported by schools!



Properties of Expectation (more on this later)

Linearity:

$$E[aX + b] = aE[X] + b$$

- Consider $X = 6$ -sided die roll, $Y = 2X - 1$.
- $E[X] = 3.5$ $E[Y] = 6$

Expectation of a sum is the sum of expectations

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

Unconscious statistician:

$$E[g(x)] = \sum_x g(x)p(x)$$



Properties of Random Variables

Probability Mass Function:

$$P(X = a)$$

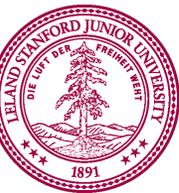
Expectation:

$$E[X]$$

Variance:

$$\text{Var}(X)$$

Learning
goals for
today



Wonderful

St Petersburg

Game set-up

- We have a fair coin (come up “heads” with $p = 0.5$)
- Let n = number of coin flips (“heads”) before first “tails”
- You win $\$2^n$

How much would you pay to play?

St Petersburg

Game set-up

- We have a fair coin (come up “heads” with $p = 0.5$)
- Let n = number of coin flips (“heads”) before first “tails”
- You win $\$2^n$

How much would you pay to play?

Solution

- Let X = your winnings
- $$E[X] = \left(\frac{1}{2}\right)^1 2^0 + \left(\frac{1}{2}\right)^2 2^1 + \left(\frac{1}{2}\right)^3 2^2 + \left(\frac{1}{2}\right)^4 2^3 + \dots = \sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^{i+1} 2^i$$
$$= \sum_{i=0}^{\infty} \frac{1}{2} = \infty$$
- I'll let you play for \$1 thousand... but just once!
Takers?



St Petersburg + Reality

What if Chris has only \$65,536?

- Same game
- If you win over \$65,536 I leave the country.

Solution

- Let X = your winnings

- $$\begin{aligned} E[X] &= \left(\frac{1}{2}\right)^1 2^0 + \left(\frac{1}{2}\right)^2 2^1 + \left(\frac{1}{2}\right)^3 2^2 + \left(\frac{1}{2}\right)^4 2^3 + \dots \\ &= \sum_{i=0}^k \left(\frac{1}{2}\right)^{i+1} 2^i \text{ s.t. } k = \log_2(65,536) \\ &= \sum_{i=0}^{16} \frac{1}{2} = 8.5 \end{aligned}$$



Learning Goals

1. Know what is meant by Conditional Independence
2. Be able to define a random variable (R.V.)
3. Be able to use + produce a PMF of a R.V.
4. Be able to calculate the expectation of the R.V.

