



Conditional Probability and Bayes

Review

What is a Probability?

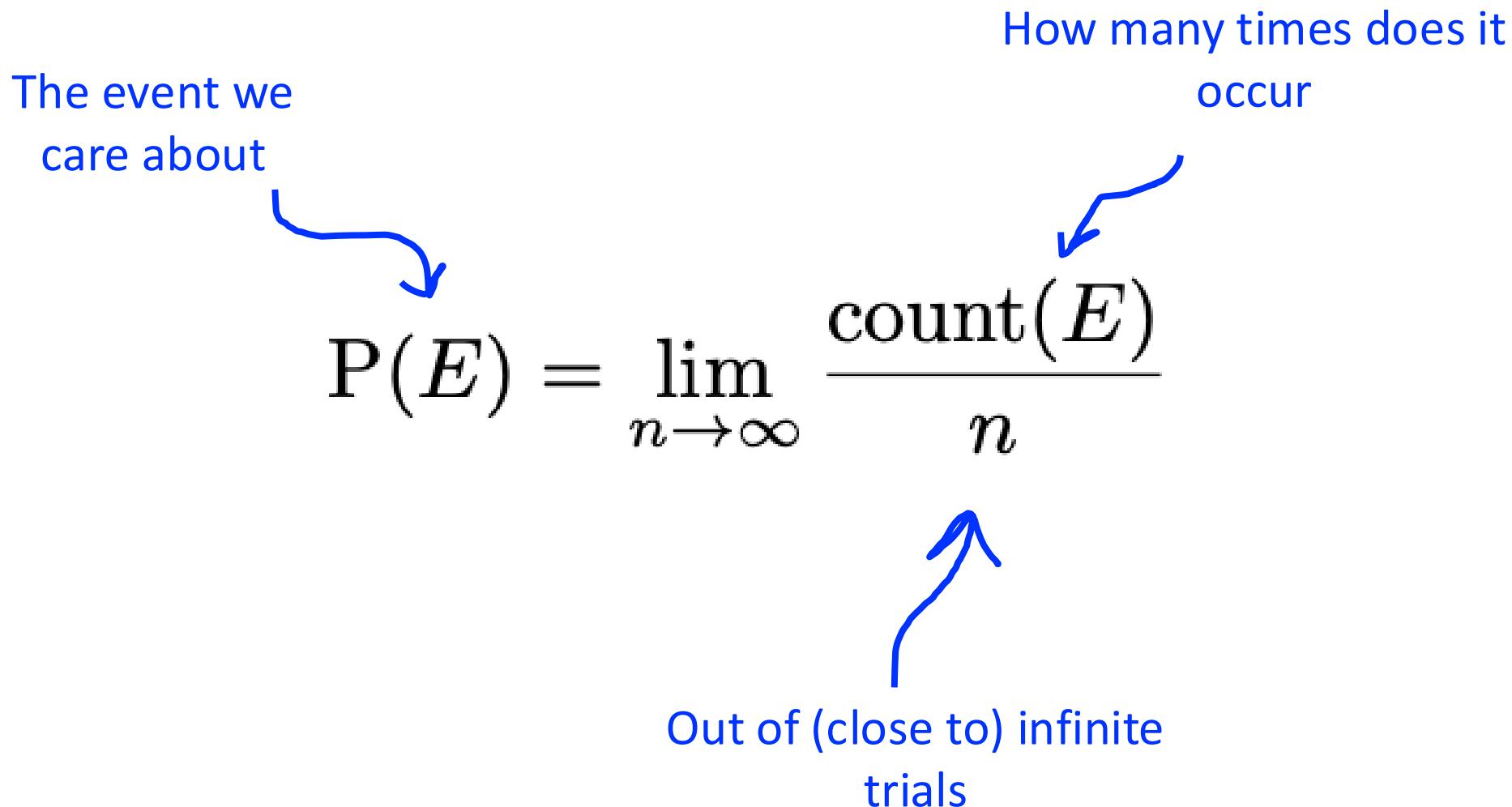
A number $[0, 1]$ to which we ascribe meaning

$$P(E) = \lim_{n \rightarrow \infty} \frac{\text{count}(E)}{n}$$

The event we care about

How many times does it occur

Out of (close to) infinite trials

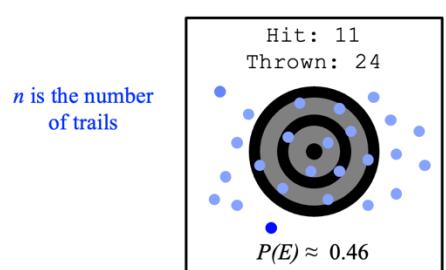




Sources of Probability

Infinite Trials

$$P(E) = \lim_{n \rightarrow \infty} \frac{\text{count}(E)}{n}$$



The “event” E is that you hit the target

Dataset of weather

Trial	Value
1	Rainy
2	Sunny
3	Rainy
4	Cloudy
5	Rainy
6	Sunny
...	
10000	Cloudy

Let E be the event that it is **Sunny**

$$P(E) = \lim_{n \rightarrow \infty} \frac{n(E)}{n}$$

$$\approx \frac{\text{Count}(E)}{10000}$$

$$\approx \frac{3332}{10000} \approx 0.3332$$

Equally Likely Outcomes

Sum of Two Die = 7?

Roll two 6-sided dice. What is probability the sum = 7?
Let E be the event that the sum is 7

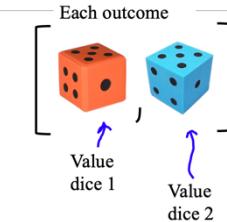
S = {	[1,1]	[1,2]	[1,3]	[1,4]	[1,5]	[1,6]
	[2,1]	[2,2]	[2,3]	[2,4]	[2,5]	[2,6]
	[3,1]	[3,2]	[3,3]	[3,4]	[3,5]	[3,6]
	[4,1]	[4,2]	[4,3]	[4,4]	[4,5]	[4,6]
	[5,1]	[5,2]	[5,3]	[5,4]	[5,5]	[5,6]
	[6,1]	[6,2]	[6,3]	[6,4]	[6,5]	[6,6] }

E is in blue

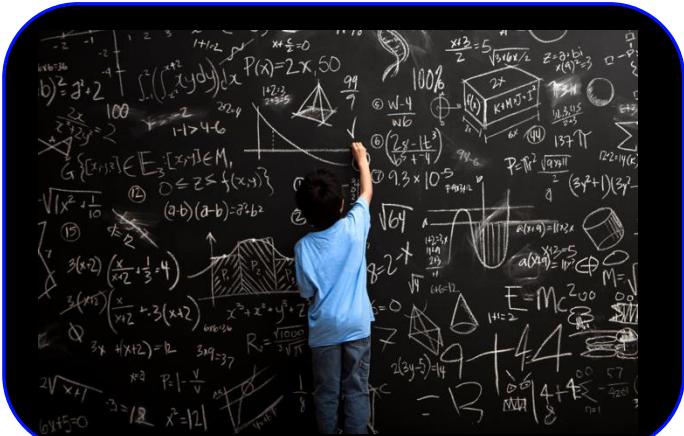
$$P(E) = \frac{|E|}{|S|} = \frac{6}{36} = 0.16\bar{6}$$

Piech, CS109, 2021

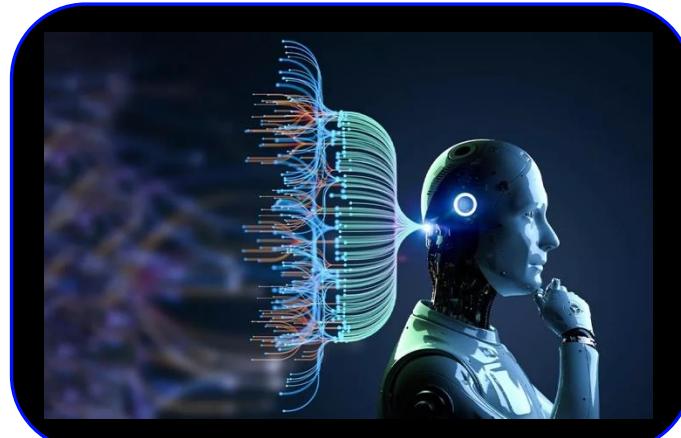
Stanford University 184



Analytics



AI Models



How Does This Work?

A screenshot of a web browser showing a poker probability calculator. The URL is probabilitycoders.stanford.edu/fall25/poker. The page is titled "Poker Game". It displays a hand of cards: J♠, K♣, 4♦, 10♣. Below the cards, it says "Straight, K High". The player's hand is shown as "You" with 100 chips. Other players shown are Fabian, Emir, Nisha, Jade, and Isabel, each with 100 chips. To the right is a bar chart titled "Events" showing the probability of winning at different stages: deal (0.15), flop (0.15), turn (0.15), and river (0.82). The chart has "Probability" on the y-axis (0 to 1.0) and "Events" on the x-axis (deal, flop, turn, river). The page also contains a sidebar with "Part 1: Core Probability" and "Part 2: Random Variables" sections, and a "Poker" section is currently selected.

A screenshot of a WhatsApp conversation. The title is "Code In place Runner". The message history shows:

- M3ei51 8:11 AM: Running Poker Probability Quiz, a program created by Chris Piech...
- M3ei51 8:11 AM: Type `quit` to stop the program early.
- M3ei51 8:11 AM: > This program generates random Texas Holdem situations and has you guess your probability of winning.
- M3ei51 8:11 AM: > Guess the probability!
- M3ei51 8:11 AM: > Number of opponents: 5
Your cards are:
Queen of Hearts ❤️
King of Diamonds ♦️
Cards on the table:
8 of Spades ♠️
5 of Hearts ❤️
3 of Spades ♠️
- M3ei51 8:11 AM: > What is the probability that you win?
- 0.6

WhatsApp m3ei51 to +1-415-728-3856





Axioms of Probability

Recall: S = all possible outcomes. E = the event.

- Axiom 1: $0 \leq P(E) \leq 1$
- Axiom 2: $P(S) = 1$
- Axiom 3: If events E and F are mutually exclusive:

$$P(E \cup F) = P(E) + P(F)$$



$$P(E^c) = 1 - P(E)?$$

$$P(E \cup E^c) = P(E) + P(E^c)$$

Axiom 3. Since E and E^c are mutually exclusive

$$P(S) = P(E) + P(E^c)$$

Since everything must either be in E or E^c

$$1 = P(E) + P(E^c)$$

Axiom 2

$$P(E^c) = 1 - P(E)$$

Rearrange

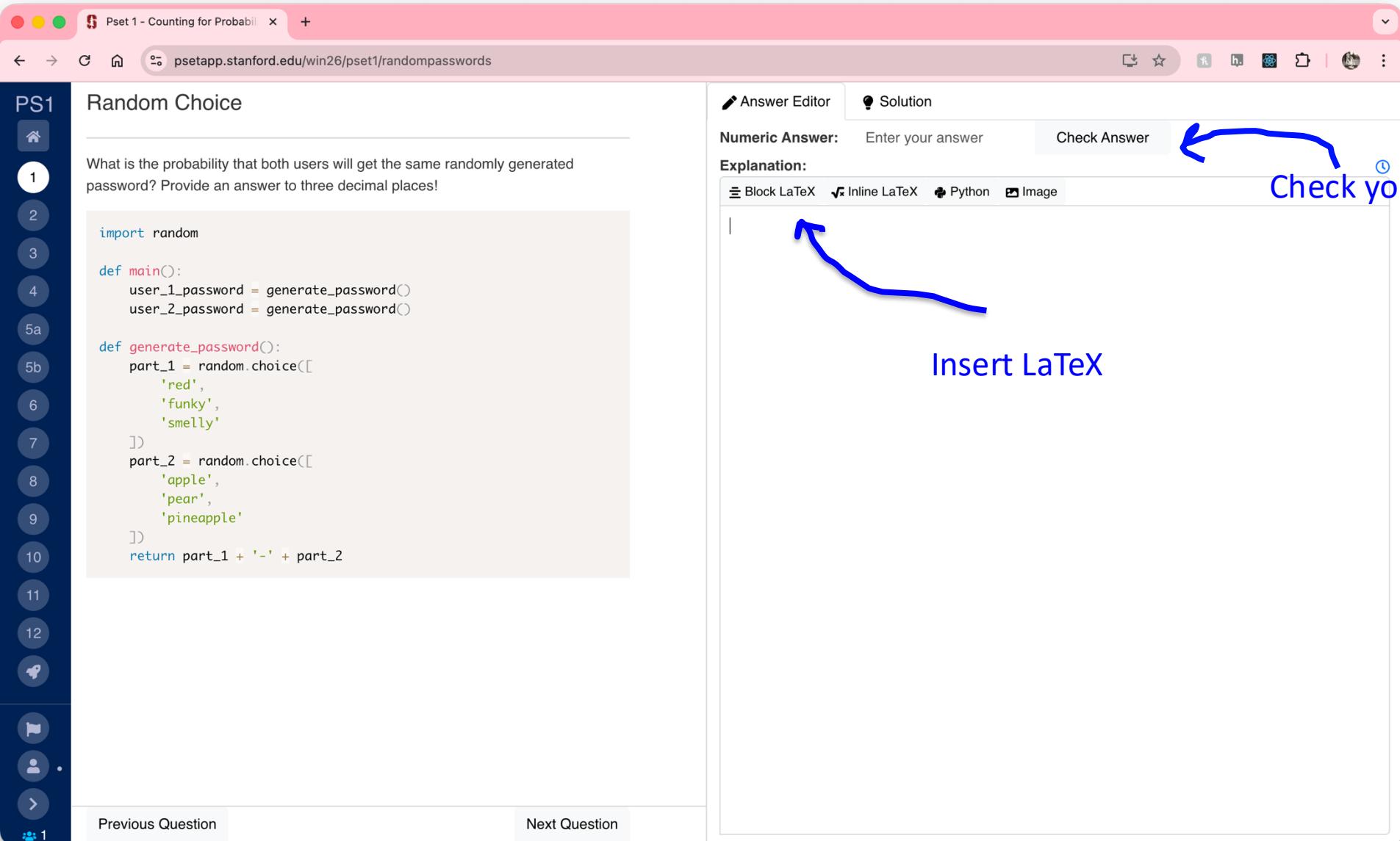


End Review

Announcements

Problem Set #1 is out

newish



PS1

1 2 3 4 5a 5b 6 7 8 9 10 11 12

Auto Submission

Random Choice

What is the probability that both users will get the same randomly generated password? Provide an answer to three decimal places!

```
import random

def main():
    user_1_password = generate_password()
    user_2_password = generate_password()

def generate_password():
    part_1 = random.choice([
        'red',
        'funky',
        'smelly'
    ])
    part_2 = random.choice([
        'apple',
        'pear',
        'pineapple'
    ])
    return part_1 + '-' + part_2
```

Answer Editor Solution

Numeric Answer: Enter your answer Check Answer

Explanation:

Block LaTeX Inline LaTeX Python Image

Check your answer

Insert LaTeX



Write Agents

newish

Pset 1 - Counting for Probabil + 1

psetapp.stanford.edu/win26/pset1/countingcards

PS1

1 2 3 4 5a 5b 6 7 8 9 10 11 12

Counting Cards

Counting cards refers to when a player keeps track of what cards have already been played during a card-game, in order to have a better estimate of how likely they are to win. Counting cards was successfully used by probability students from MIT to beat casinos worldwide: [MIT Blackjack Team](#) a heist which was popularized by the movie [21](#). The key to counting cards in blackjack is to keep track of the probability of high cards.



In this problem we are going to consider a simpler game called High Card played on a standard 52 card deck. The game works as follows: You decide if you want to play. If you do, the casino deals you a single card. If the card is a high card, (10, Jack, Queen, King or Ace), you win \$20. If it is not, you lose \$20. Another player is playing as well and each game they will play (thus revealing a card). You can play even if you have negative dollars (we assume you will borrow money to pay it back).

If you were given a truly random card out of the deck of 52, your chance of winning would be $20/52 \approx 0.38$ since 20 of the 52 cards are high. Not very good! But you

[Previous Question](#) [Next Question](#)

Answer Editor Solution

Agent:

```
1 1000
2 counting_agent.py
3 This file defines an agent "counting_agent" which plays the game of
4 High Card. The function gets called each time it is the agents turn.
5 The cards_played list has all cards which have been played so far. You must return
6 "1000"
7
8 def counting_agent(cards_played):
9     # default strategy: always play
10    return 'play'
```

Run One Game Test Agent



Above and Beyond

new

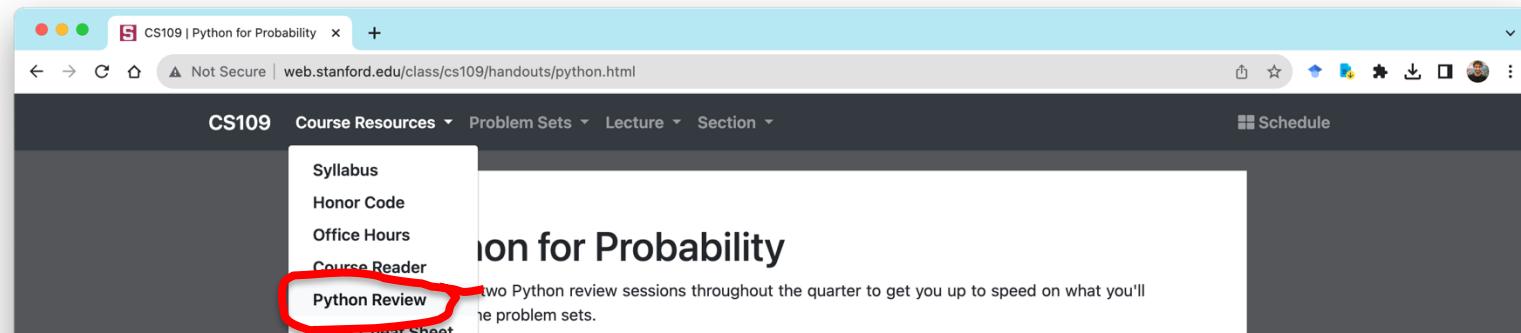
The screenshot shows a web browser window with a pink header bar. The title bar says "Pset 1 - Counting for Probabil" and the address bar says "psetapp.stanford.edu/win26/pset1/above-and-beyond". The main content area is titled "Above and Beyond". On the left, a sidebar titled "PS1" lists numbered items from 1 to 12, with a "Create" icon at the bottom. A central box contains a welcome message: "Welcome to Above and Beyond! Here you can submit a project (videos/links/files) that goes beyond the scope of the Pset. You can create as many projects as you like, but you can only submit one project per Pset. You can submit a project and continue editing it until Monday, Jan 19, 10:00 PM Pacific Standard Time." Below this is a "Create Project" button. A "Projects" section shows a table with columns: Name, Submitted, Updated, and Edit. The table content is "No projects yet. Click \"Create Project\" to get started.".



Python Review Session

Friday at 4:30-5:30pm PT, recorded

Find links, recordings, and setup here



Learn LaTeX

```
1 \begin{aligned}
```

```
2 P(E) |
```

```
3 &= \sum_{i=0}^n e^i \\
```

```
4 &= 0.25
```

```
5 \end{aligned}
```

$$\begin{aligned} P(E) &= \sum_{i=0}^n e^i \\ &= 0.25 \end{aligned}$$

Done



CS109 | Python for Probability

Not Secure | web.stanford.edu/class/cs109/handouts/latex/

CS109 Course Resources Problem Sets Lecture Section Schedule

Syllabus Honor Code Office Hours Course Reader Python Review Latex Cheat Sheet Fall 2022 Videos

Quick Guide to LaTeX

by Roshini Ravi and Chris Piech

LaTeX is a typesetting system that creates beautiful scientific documents. It is the digital language of science, and it was invented right here at Stanford. You can still submit handwritten homeworks, but we encourage you to use LaTeX.

Some examples that should help you get started! As a helpful tip, you can access the LaTeX code for any equation in the course reader by right clicking the equation and clicking "Show TeX Commands".



Inline LaTeX

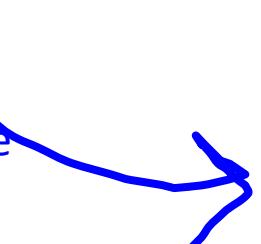
new

Explanation:

≡ Block LaTeX ✓ Inline LaTeX ⚡ Python Image

This is an example of inline latex. Let $\$ \$ Y \$$

When you type
the closing \$



Explanation:

≡ Block LaTeX ✓ Inline LaTeX ⚡ Python Image

This is an example of inline latex. Let Y

CS109 Course Resources Problem Sets Lecture Section Schedule

Block Guide to LaTeX

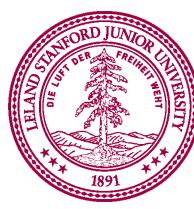
by Roshini Ravi and Chris Piech

LaTeX is a typesetting system that creates beautiful scientific documents. It is the digital language of science, and it was invented right here at Stanford. You can still submit handwritten homeworks, but we encourage you to use LaTeX.

[Latex Cheat Sheet](#)

Fall 2022 Videos

Midterm Final



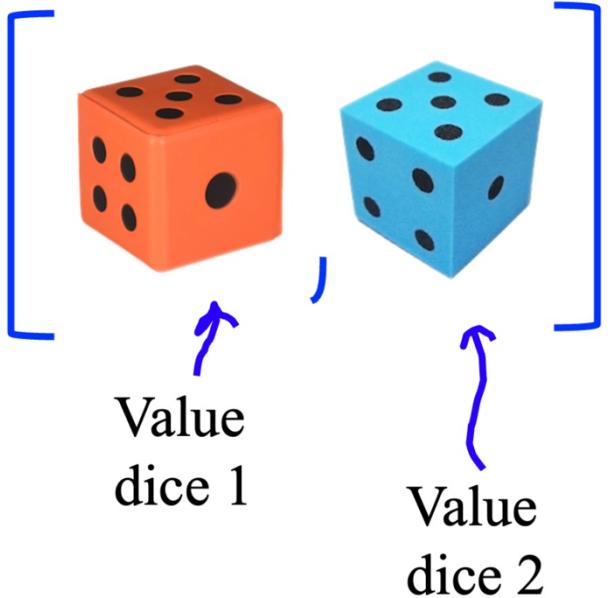
What Makes for a Good Answer

https://psetapp.stanford.edu/win26/lecture2/dice_probability

L1 Dice Probability

This problem is a warmup to get used to the CS109 pset app.

If you roll two fair six-sided die, what is the probability that you get a sum that is **not** 6. Report your answer to three decimal places.



Value
dice 1

Value
dice 2

Answer Editor Solution

Numeric Answer: 0.5 Check Answer

Explanation:

Block LaTeX Inline LaTeX Python Image

Let E be the event that you roll a 6 on two dice. We want $P(E^C)$.

There are 36 equally likely outcomes for throwing two dice (if you think of each outcome as a tuple with value on dice 1, value on dice 2). Out of those 36 equally likely outcomes 5 are ones that have a sum of 6:
(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)

As such $P(E) = 5/36$. Using the first identity proved in class:

$$\begin{aligned} P(E^C) &= 1 - P(E) \\ &= 1 - \frac{5}{36} \end{aligned}$$

1 p_E = 5/36
2 print(1 - p_E)

Run Show

The answer is: 0.8611111...

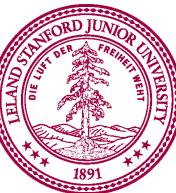
Solution Previous Question Next Question



If you notice a bug?

It should be robust, but things can happen.

Let me know: send an email to jwoodrow@stanford.edu. I need your email and the approximate time you encountered the bug.



Honor Code

Always remember: You need to be able to recreate your ability on an exam. And in the real world. This is a foundation course.

Cheating in CS109 is cheating yourself and your friends.

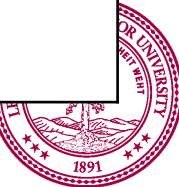
Talk to your friends about the **concepts**, not the solution. Words must be your own.

Practice the **art of teaching**. Three most important things to know:

1. Do not give away the answer
2. Always be respectful
3. Know what you don't know

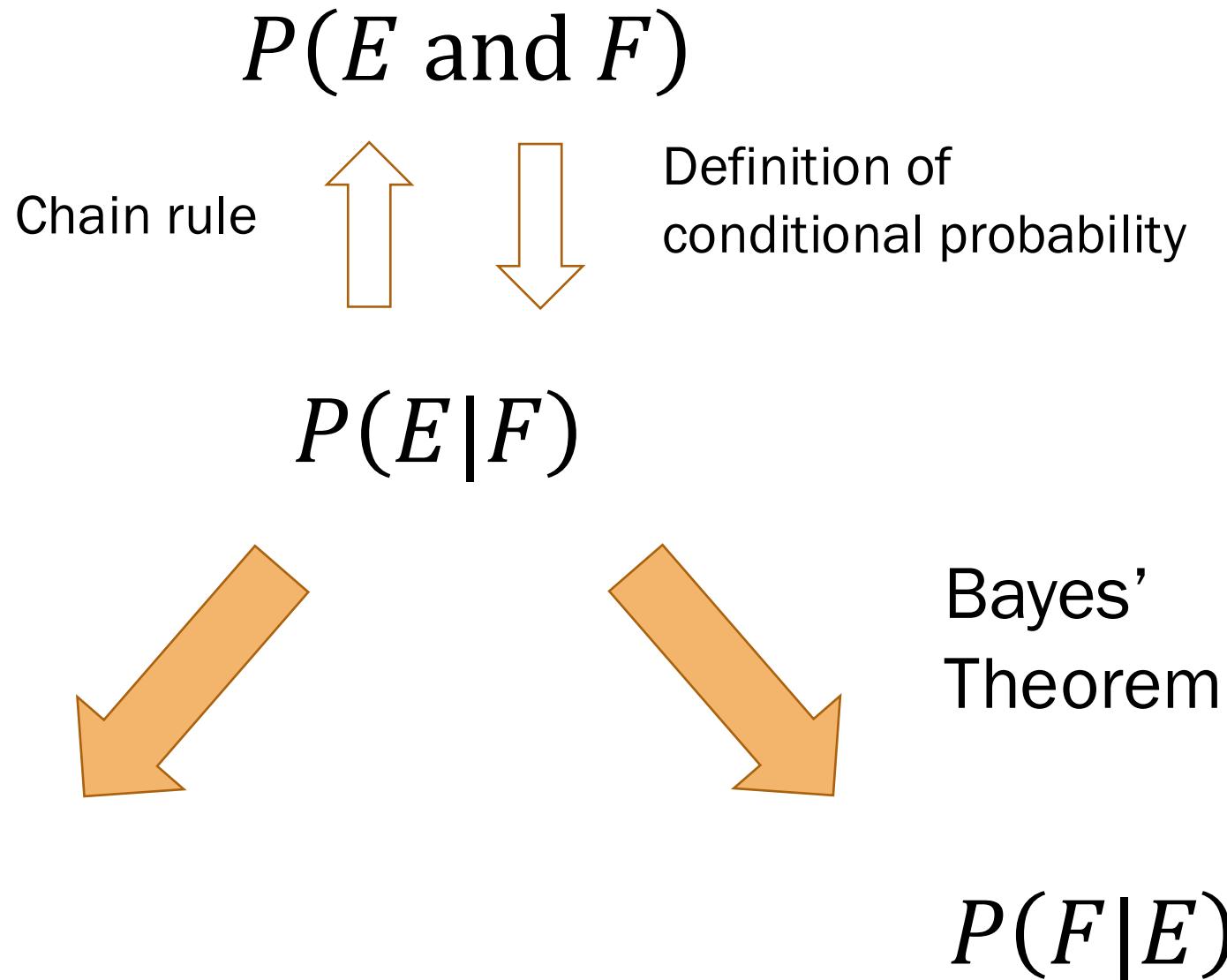


Section Signups Open Tomorrow

Sunday	Monday	Tuesday	Wed	Thursday	Friday	Saturday
			Jan 7 You are here	Jan 8 Section signup opens	Jan 9	Jan 10
Jan 11 Section signups close at 5pm	Jan 12	Jan 13 Sections announced	Jan 14	Jan 15, Jan 16 First section!	Jan 17	

End Announcements

Learning Goal for Today: Conditional Probability

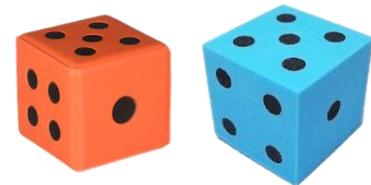


Conditional Probability

Roll two dice

$$P(E) = \frac{|E|}{|S|}$$
 Equally likely outcomes

Roll two 6-sided fair dice. What is $P(\text{sum} = 7)$?



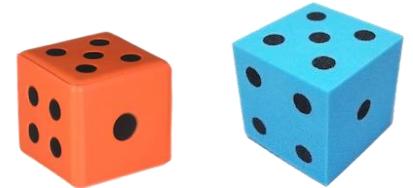
$$S = \{(1,1) (1,2) (1,3) (1,4) (1,5) \mathbf{(1,6)} \\ (2,1) (2,2) (2,3) (2,4) \mathbf{(2,5)} (2,6) \\ (3,1) (3,2) (3,3) \mathbf{(3,4)} (3,5) (3,6) \\ (4,1) (4,2) \mathbf{(4,3)} (4,4) (4,5) (4,6) \\ (5,1) \mathbf{(5,2)} (5,3) (5,4) (5,5) (5,6) \\ \mathbf{(6,1)} (6,2) (6,3) (6,4) (6,5) (6,6) \}$$

$$E = \text{In blue}$$



Dice, our misunderstood friends

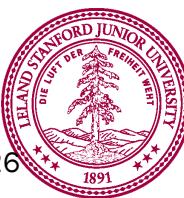
Roll two 6-sided dice, yielding values D_1 and D_2 .
You want them to sum to 4.



What is the best outcome for $P(D_1)$?

Your Choices:

- A. 1 and 3 tie for best
- B. 1, 2 and 3 tie for best
- C. 2 is the best
- D. Other/none/more than one



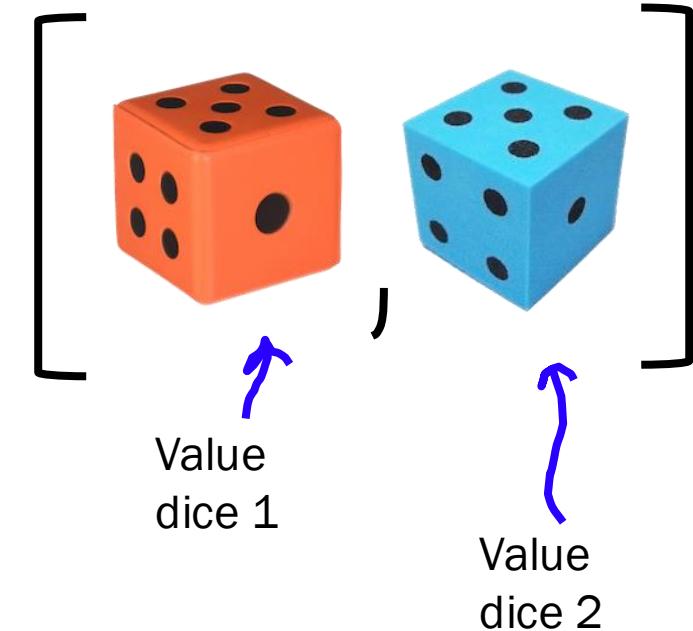
Sum of Two Die = 4?

Roll two 6-sided dice. What is probability the sum = 4?

Let E be the event that the sum is 4

$$S = \{ [1,1] [1,2] [1,3] [1,4] [1,5] [1,6] \\ [2,1] [2,2] [2,3] [2,4] [2,5] [2,6] \\ [3,1] [3,2] [3,3] [3,4] [3,5] [3,6] \\ [4,1] [4,2] [4,3] [4,4] [4,5] [4,6] \\ [5,1] [5,2] [5,3] [5,4] [5,5] [5,6] \\ [6,1] [6,2] [6,3] [6,4] [6,5] [6,6] \}$$

Each outcome



E = *In red*

$$P(E) = \frac{|E|}{|S|} = \frac{3}{36} = 0.08\bar{3}$$

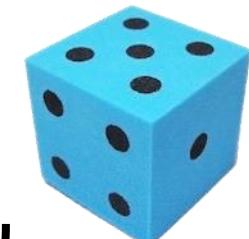
Sum of Two Die = 4? Condition on F: $D_1 = 2$

Roll two 6-sided dice. What is probability the sum = 4?

Let E be the event that the sum is 4

$$S = \{ [1,1] [1,2] [1,3] [1,4] [1,5] [1,6] \\ [2,1] [2,2] [2,3] [2,4] [2,5] [2,6] \\ [3,1] [3,2] [3,3] [3,4] [3,5] [3,6] \\ [4,1] [4,2] [4,3] [4,4] [4,5] [4,6] \\ [5,1] [5,2] [5,3] [5,4] [5,5] [5,6] \\ [6,1] [6,2] [6,3] [6,4] [6,5] [6,6] \}$$

Each outcome



Value
dice 1

Value
dice 2

$E = \text{In red}$

$$P(E) = \frac{|E|}{|S|} = \frac{3}{36} = 0.08\bar{3}$$



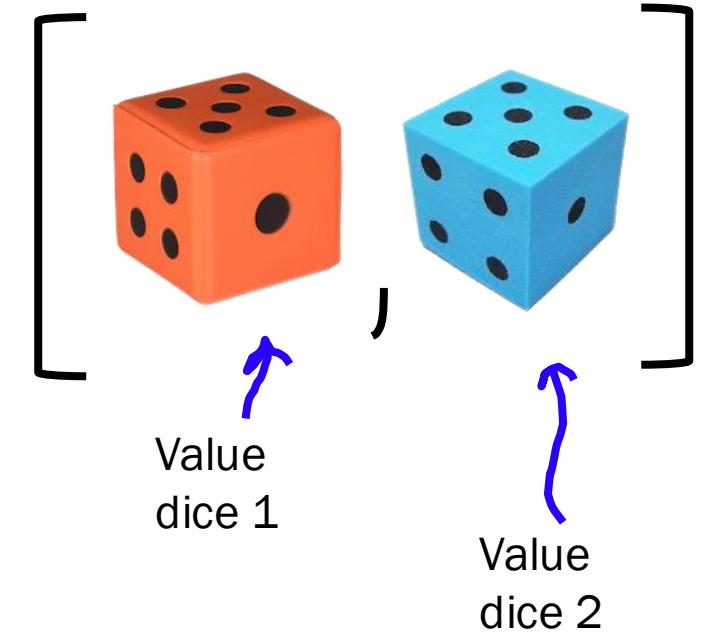
Sum of Two Die = 4? Condition on F: $D_1 = 2$

Roll two 6-sided dice. What is probability the sum = 4?

Let E be the event that the sum is 4

$$S = \{ [1,1] [1,2] [1,3] [1,4] [1,5] [1,6] \\ [2,1] [2,2] [2,3] [2,4] [2,5] [2,6] \\ [3,1] [3,2] [3,3] [3,4] [3,5] [3,6] \\ [4,1] [4,2] [4,3] [4,4] [4,5] [4,6] \\ [5,1] [5,2] [5,3] [5,4] [5,5] [5,6] \\ [6,1] [6,2] [6,3] [6,4] [6,5] [6,6] \}$$

Each outcome

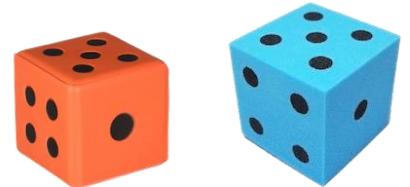


$E = \text{In red}$

$$P(E) = \frac{|E|}{|S|} = \frac{3}{36} = 0.08\bar{3}$$

Dice, our misunderstood friends

Roll two 6-sided dice, yielding values D_1 and D_2 .



Let E be event: $D_1 + D_2 = 4$.

Let F be event: $D_1 = 2$.

What is $P(E)$?

What is $P(E, \text{ given } F \text{ already observed})$?

$$|S| = 36$$

$$E = \{(1,3), (2,2), (3,1)\}$$

$$P(E) = 3/36 = 1/12$$

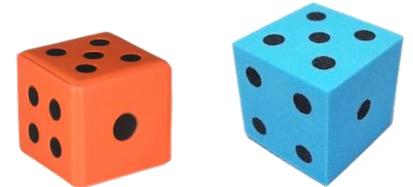
$$S = \{(2,1), (2,2), (2,3), (2,4), (2,5), (2,6)\}$$

$$E = \{(2,2)\}$$

$$P(E) = 1/6$$

Dice, our misunderstood friends

Roll two 6-sided dice, yielding values D_1 and D_2 .



Let E be event: $D_1 + D_2 = 4$.

Let F be event: $D_1 = 3$.

What is $P(E)$?

What is $P(E, \text{ given } F \text{ already observed})$?

$$|S| = 36$$

$$E = \{(1,3), (2,2), (3,1)\}$$

$$P(E) = 3/36 = 1/12$$

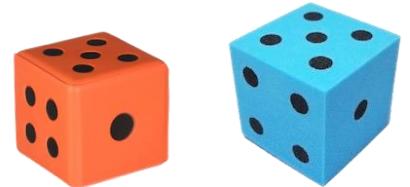
$$S = \{(3,1), (3,2), (3,3), (3,4), (3,5), (3,6)\}$$

$$E = \{(3,1)\}$$

$$P(E) = 1/6$$

Dice, our misunderstood friends

Roll two 6-sided dice, yielding values D_1 and D_2 .



Let E be event: $D_1 + D_2 = 4$.

Let F be event: $D_1 = 5$.

What is $P(E)$?

What is $P(E, \text{ given } F \text{ already observed})$?

$$|S| = 36$$

$$E = \{(1,3), (2,2), (3,1)\}$$

$$P(E) = 3/36 = 1/12$$

$$S = \{(5,1), (5,2), (5,3), (5,4), (5,5), (5,6)\}$$

$$E = \{ \quad \}$$

$$P(E) = 0/6$$

Conditional Probability

The **conditional probability** of E given F is the probability that E occurs given that F has already occurred. This is known as conditioning on F .

Written as:

$$P(E|F)$$

Means:

“ $P(E$, given F already observed)”

Sample space \rightarrow

all possible outcomes consistent with F (i.e. S and F)

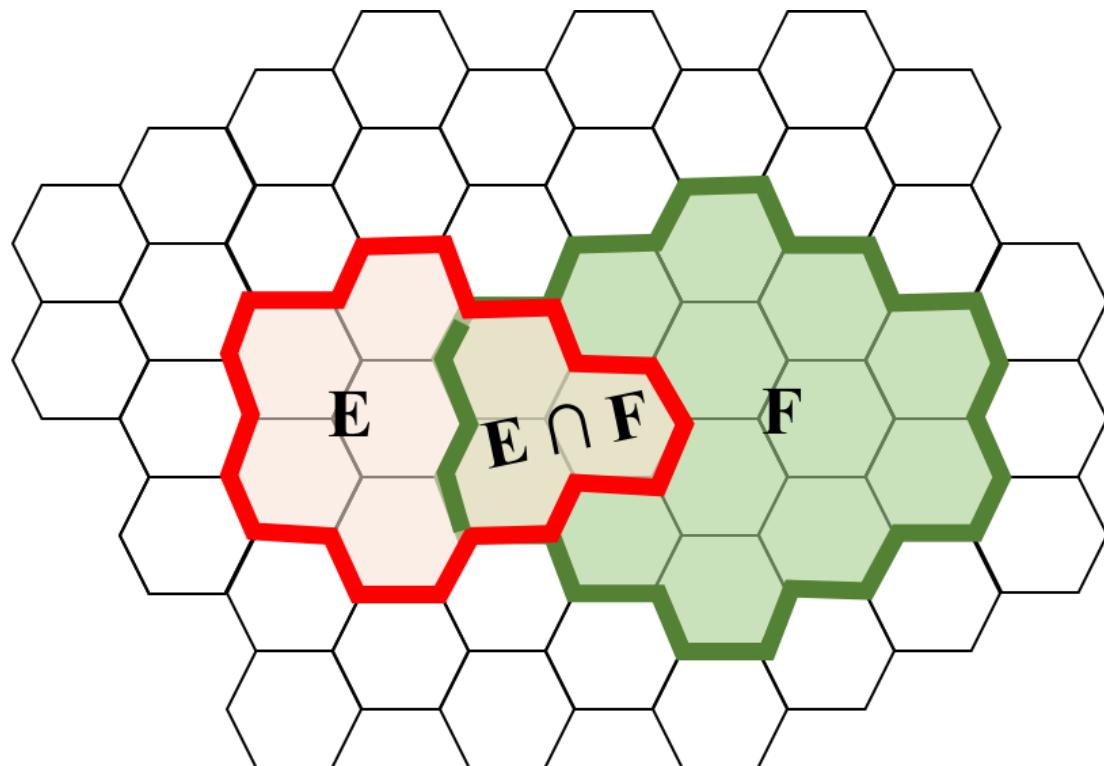
Event \rightarrow

all outcomes in E consistent with F (i.e. E and F)



Conditional Probability, visual intuition

The **conditional probability** of E given F is the probability that E occurs given that F has already occurred. This is known as conditioning on F .

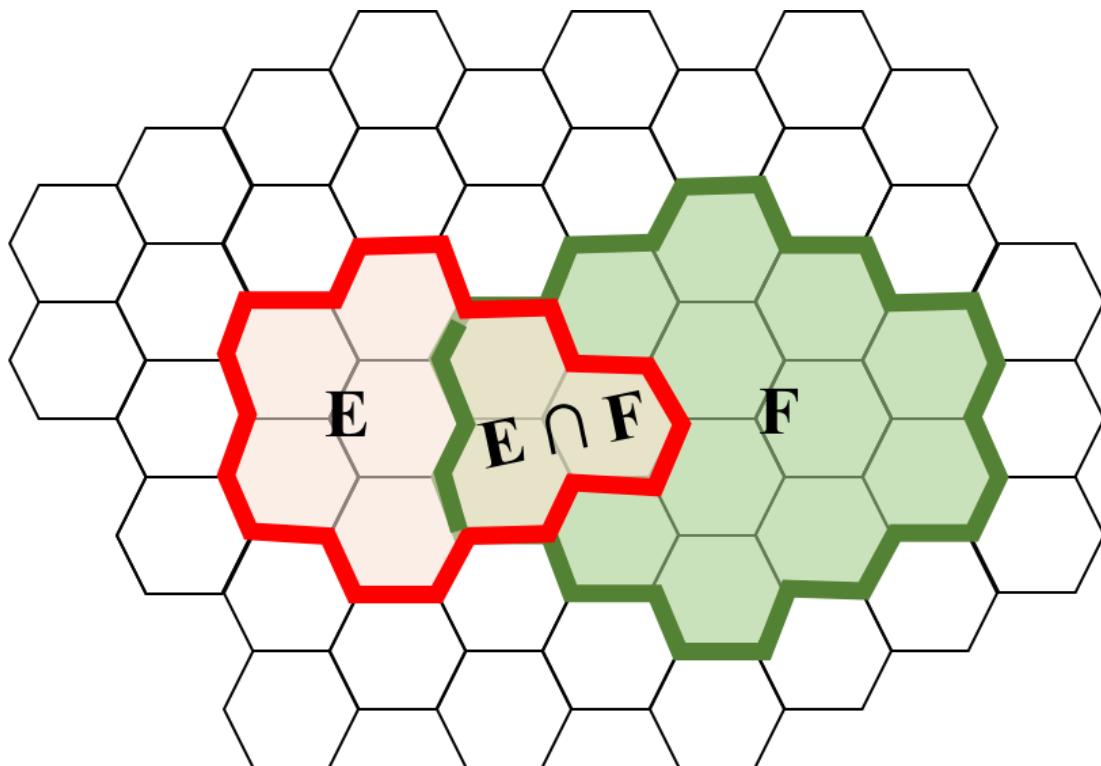


$$P(E) = \frac{8}{50} \approx 0.16$$

$$P(E|F) =$$

Conditional Probability, visual intuition

The **conditional probability** of E given F is the probability that E occurs given that F has already occurred. This is known as conditioning on F .



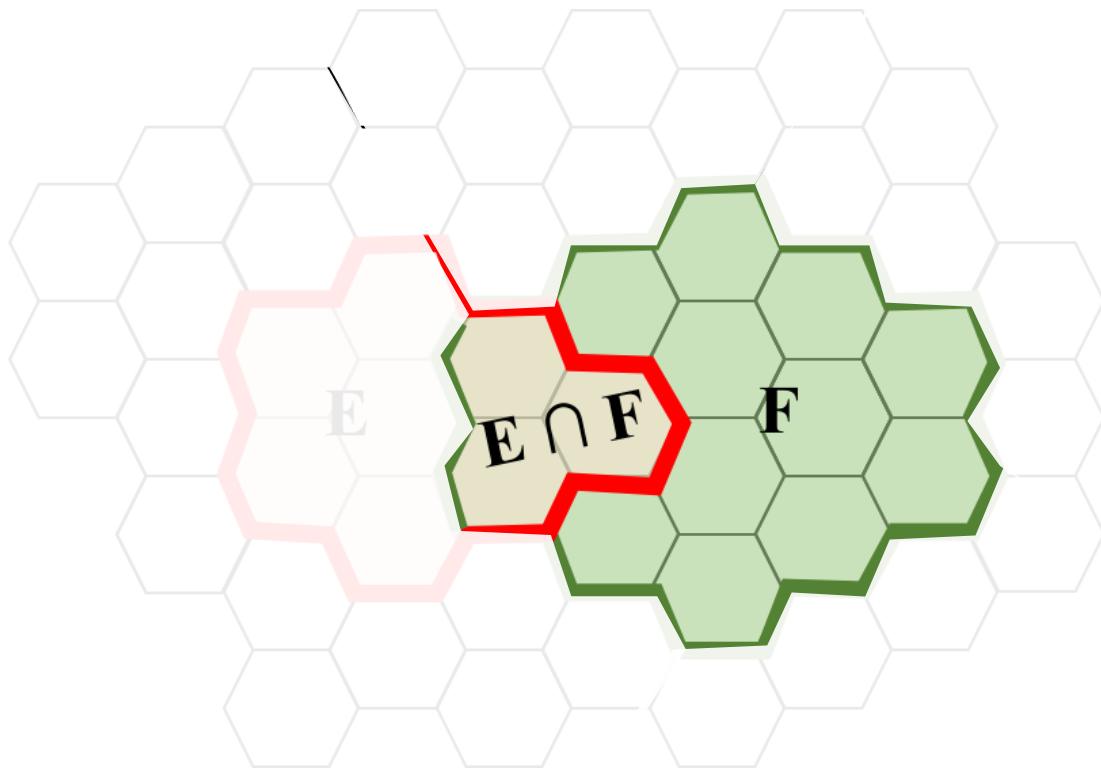
$$P(E) = \frac{8}{50} \approx 0.16$$

$$P(E|F) = \frac{3}{14} \approx 0.21$$



Conditional Probability, visual intuition

The **conditional probability** of E given F is the probability that E occurs given that F has already occurred. This is known as conditioning on F .



$$P(E) = \frac{8}{50} \approx 0.16$$

$$P(E|F) = \frac{3}{14} \approx 0.21$$



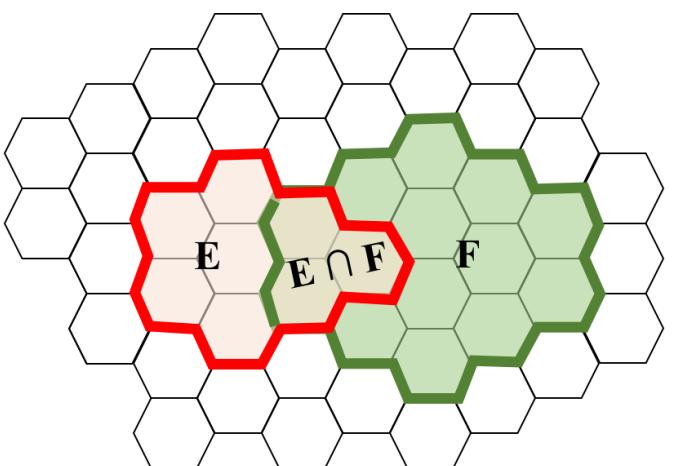
Conditional Probability, visual intuition

The **conditional probability** of E given F is the probability that E occurs given that F has already occurred. This is known as conditioning on F .

Shorthand notation for set intersection (aka set “and”)

$$\Pr(E|F) = \frac{\text{\# of outcomes in } E \text{ consistent with } F}{\text{\# of outcomes in } S \text{ consistent with } F}$$

$$= \frac{|EF|}{|SF|} = \frac{|EF|}{|F|}$$



$$P(E) = \frac{8}{50} \approx 0.16$$

$$P(E|F) = \frac{3}{14} \approx 0.21$$



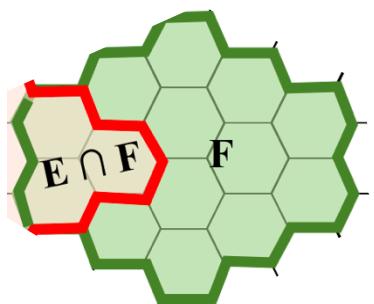
Conditional Probability, visual intuition

The **conditional probability** of E given F is the probability that E occurs given that F has already occurred. This is known as conditioning on F .

Shorthand notation for set intersection (aka set “and”)

$$\Pr(E|F) = \frac{\text{\# of outcomes in } E \text{ consistent with } F}{\text{\# of outcomes in } S \text{ consistent with } F}$$

$$= \frac{|EF|}{|SF|} = \frac{|EF|}{|F|}$$



$$P(E) = \frac{8}{50} \approx 0.16$$

$$P(E|F) = \frac{3}{14} \approx 0.21$$



Conditional probability in general

These properties hold even when outcomes are not equally likely.

General **definition** of conditional probability:

$$P(E|F) = \frac{P(EF)}{P(F)}$$

Shorthand for E and F

The **Chain Rule** (aka Product rule):

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



What if $P(F) = 0$?

- $P(E|F)$ undefined
- *Congratulations! Observed impossible*





Bye, land of equally likely outcomes

NETFLIX

and Learn

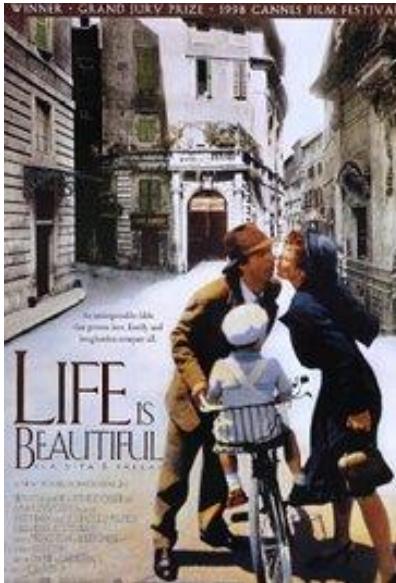
Netflix and Learn

$$P(E|F) = \frac{P(EF)}{P(F)}$$

Definition of
Cond. Probability

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

$P(E)$



S = {Watch, Not Watch}

E = {Watch}

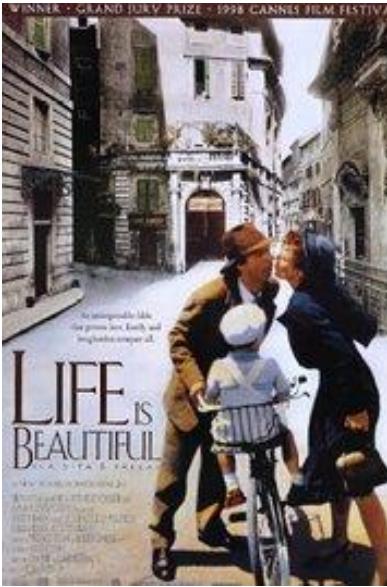
$P(E) = 1/2 ?$



Netflix and Learn

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

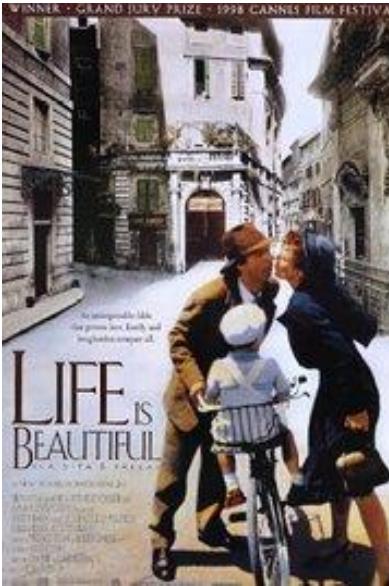
$$P(E)$$



Netflix and Learn

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

$$P(E)$$



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

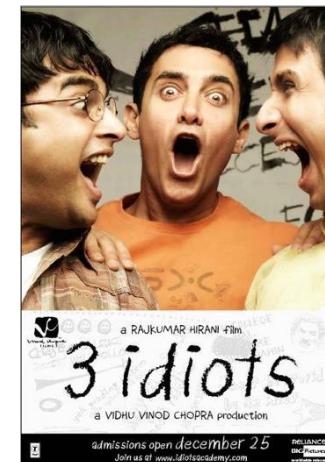
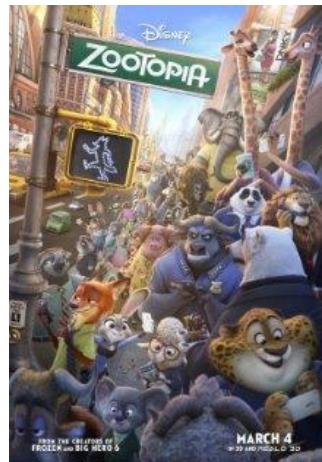
$$P(E) = 10,234,231 / 50,923,123 = 0.20$$



Netflix and Learn

$$P(E|F) = \frac{P(EF)}{P(F)}$$
 Definition of Cond. Probability

Let E be the event that a user watches the given movie.



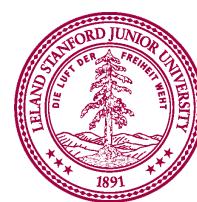
$$P(E) = 0.19$$

$$P(E) = 0.32$$

$$P(E) = 0.20$$

$$P(E) = 0.09$$

$$P(E) = 0.20$$



Netflix and Learn

$$P(E|F) = \frac{P(EF)}{P(F)}$$
 Definition of Cond. Probability

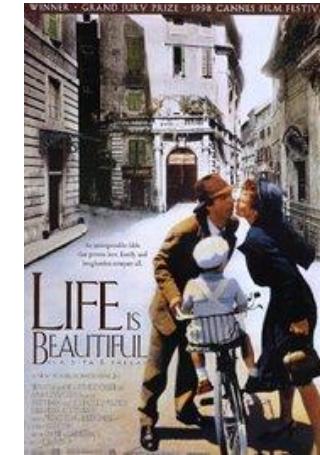
Let E = a user watches Life is Beautiful.

Let F = a user watches CODA.

What is the probability that a user watches Life is Beautiful, given they watched CODA?

$$P(E|F)$$

$$P(E|F) = \frac{P(EF)}{P(F)} \approx \frac{\frac{\# \text{ people who have watched both}}{\# \text{ people on Netflix}}}{\frac{\# \text{ people who have watched CODA}}{\# \text{ people on Netflix}}}$$



Netflix and Learn

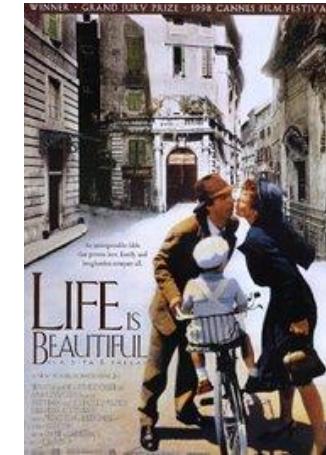
$$P(E|F) = \frac{P(EF)}{P(F)}$$
 Definition of Cond. Probability

Let E = a user watches Life is Beautiful.

Let F = a user watches CODA.

What is the probability that a user watches Life is Beautiful, given they watched CODA?

$$P(E|F)$$



$$P(E|F) = \frac{P(EF)}{P(F)} \approx \frac{\frac{\# \text{ people who have watched both}}{\# \text{ people on Netflix}}}{\frac{\# \text{ people who have watched CODA}}{\# \text{ people on Netflix}}}$$

$$\approx \frac{\# \text{ people who have watched both}}{\# \text{ people who have watched CODA}}$$

$$\approx 0.42$$



Netflix and Learn

$$P(E|F) = \frac{P(EF)}{P(F)}$$
 Definition of Cond. Probability

Let E = a user watches Life is Beautiful.

Let F = a user watches CODA.

What is the probability that a user watches Life is Beautiful, given they watched CODA?

$$P(E|F)$$



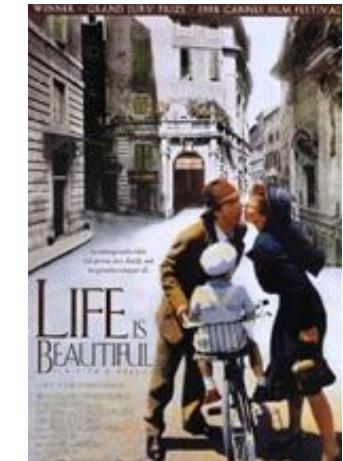
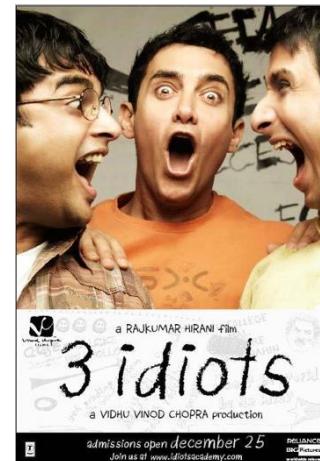
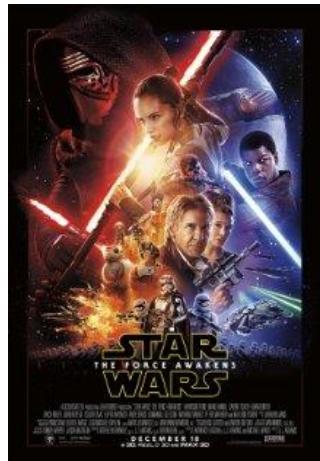
$$\begin{aligned} P(E|F) &= \frac{P(EF)}{P(F)} \approx \frac{\frac{\# \text{ people who have watched both}}{\# \text{ people on Netflix}}}{\frac{\# \text{ people who have watched CODA}}{\# \text{ people on Netflix}}} \\ &\approx \frac{\# \text{ people who have watched both}}{\# \text{ people who have watched CODA}} \\ &\approx 0.42 \end{aligned}$$



Netflix and Learn

$$P(E|F) = \frac{P(EF)}{P(F)}$$
 Definition of Cond. Probability

Let E be the event that a user watches the given movie.
Let F be the event that the same user watches CODA (2021).



$$P(E) = 0.19$$

$$P(E) = 0.32$$

$$P(E) = 0.20$$

$$P(E) = 0.09$$

$$P(E) = 0.20$$

$$P(E|F) = 0.14$$

$$P(E|F) = 0.35$$

$$P(E|F) = 0.20$$

$$P(E|F) = 0.72$$

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

Machine Learning

Machine Learning is:
Probability + Data + Computers



Notation

And

$P(E \text{ and } F)$

$P(E, F)$

$P(EF)$

$P(E \cap F)$

Or

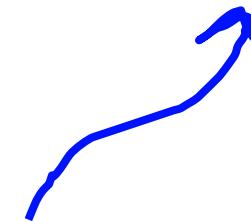
$P(E \text{ or } F)$

$P(E \cup F)$

Given

$P(E|F)$

$P(E|F, G)$



Probability of E given
F and G



Chain Rule via Baby Poop

<https://psetapp.stanford.edu/win26/lecture2/poop>

$$P(E|F) = \frac{P(EF)}{P(F)}$$

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

In the morning when she wakes up, a baby has a 50% chance of having pooped. The chance that a baby cries **given** that she has pooped is 50%. What is the probability that a baby **has pooped, and cries**.

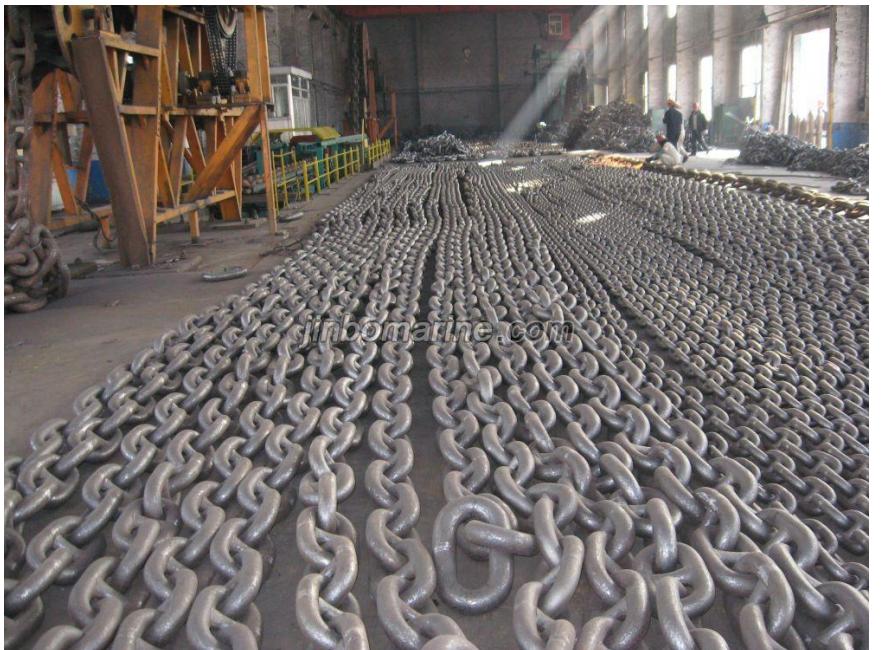


A screenshot of a web browser window. The title bar says "CS109 | Conditioning And Bayes". The address bar says "Not Secure | web.stanford.edu/class/cs109/lectures/4-ConditioningAndBayes/". The page content shows the "Lecture 4: Conditioning And Bayes" page. The top navigation bar includes "CS109", "Course Resources", "Problem Sets", "Lecture", "Section", and "Schedule". A dropdown menu under "Lecture" lists "1. Welcome", "2. Combinatorics", "3. Probability", and "4. Conditioning and Bayes". The main content area displays the lecture title "Lecture 4: Conditioning And Bayes" and the date "OCT 4TH, 2023 HEWLETT 200, 3:30P". Below this is a section titled "Lecture Materials" with icons for a document and a rocket.

Generalized Chain Rule

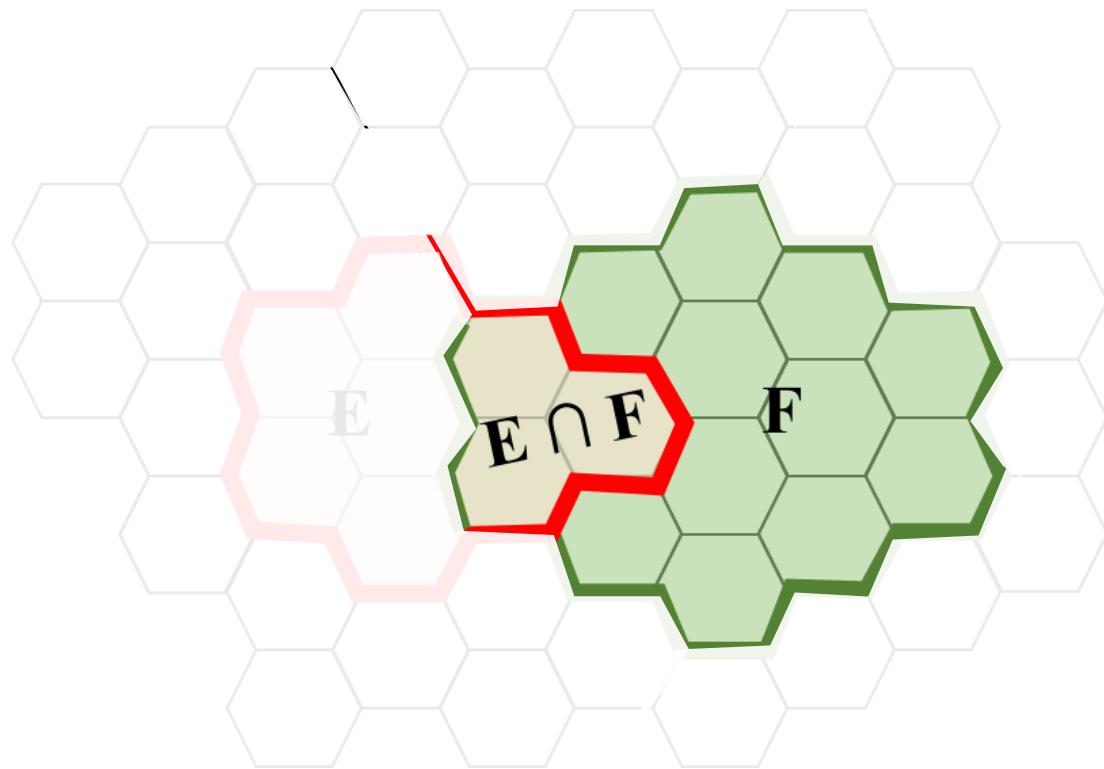
$$\Pr(E_1 \text{ and } E_2 \text{ and } E_3 \text{ and } \dots E_n)$$

$$= \Pr(E_1) \cdot \Pr(E_2|E_1) \cdot \Pr(E_3|E_1, E_2) \cdots \Pr(E_n|E_1, E_2 \dots E_{n-1})$$



Conditional Paradigm

When you condition on an event (or multiple events), you enter a world where all the rules of probability still hold.



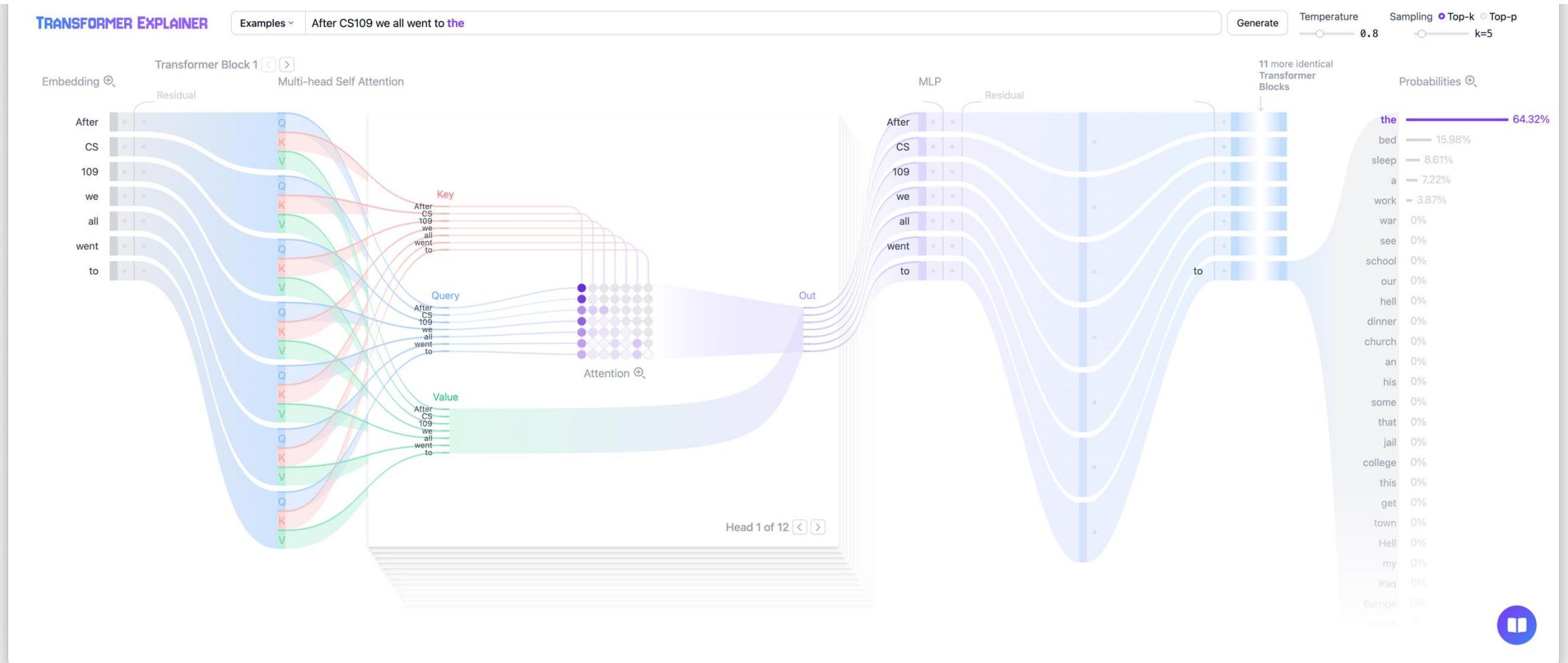
For example:

$$P(E^C|F) = 1 - P(E|F)$$



and Learn

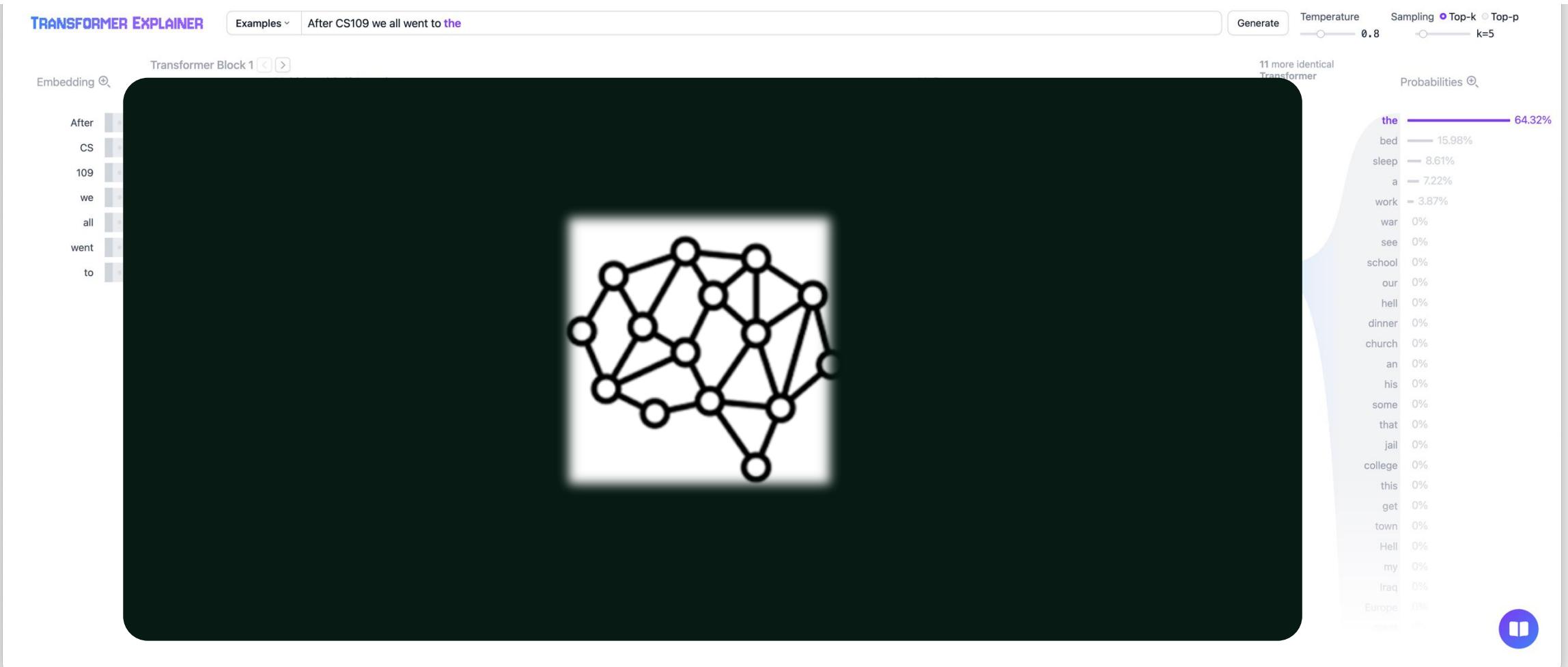
Under the hood of a Large Language Model



<https://poloclub.github.io/transformer-explainer/>



Under the hood of a Large Language Model

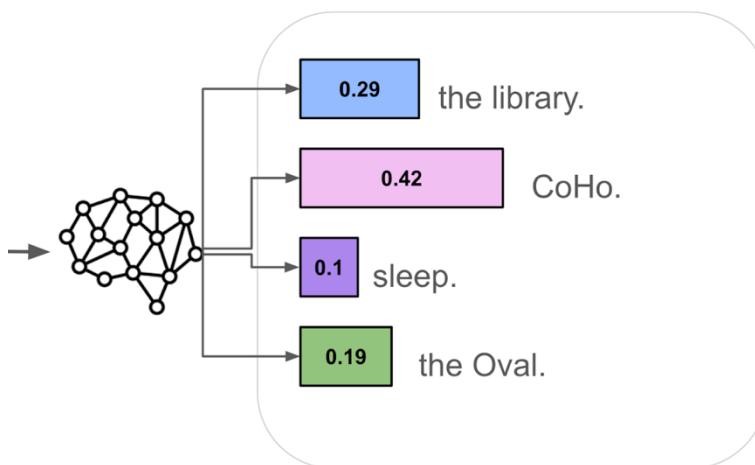


<https://poloclub.github.io/transformer-explainer/>



LLM is a Conditional Probability Machine

After CS109 we all went to



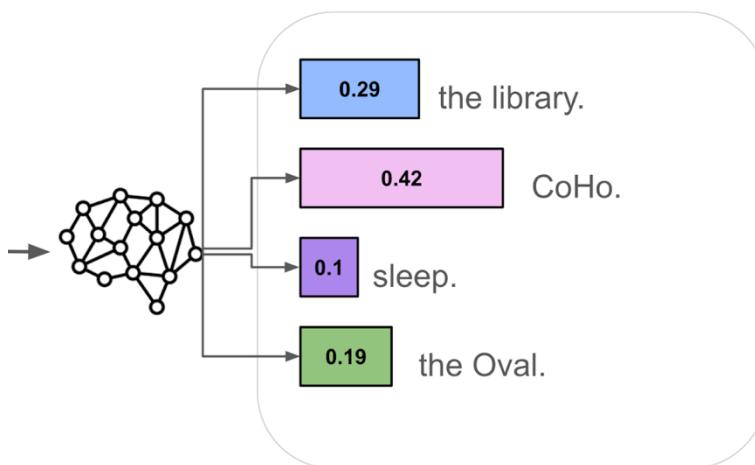
LLM is a Conditional Probability Machine

Let T_i be the i th token in a prompt.

An LLM is built to compute:

$$P(T_i|T_1, \dots, T_{i-1})$$

After CS109 we all went to



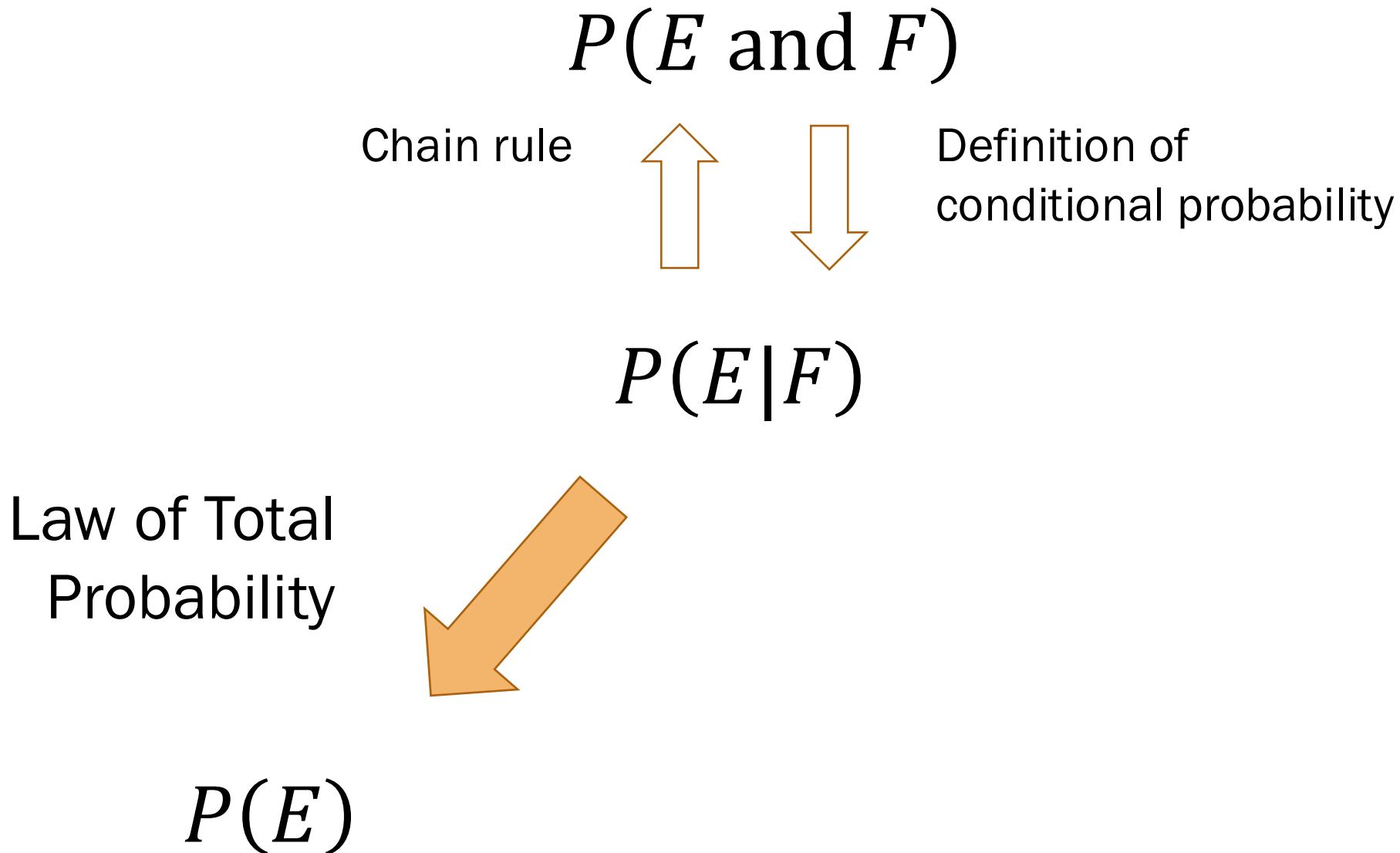
Q1: What is the probability of the string “**After CS109 we**”

Q2: the string “**went dancing**” coming after the string “**After CS109 we**”



Law of Total Probability

Relationship Between Probabilities



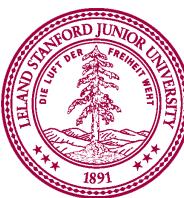
Baby Poop Revisited

In the morning when she wakes up, a baby has a 50% chance of having pooped. The chance that a baby cries given that she has pooped is 50%.



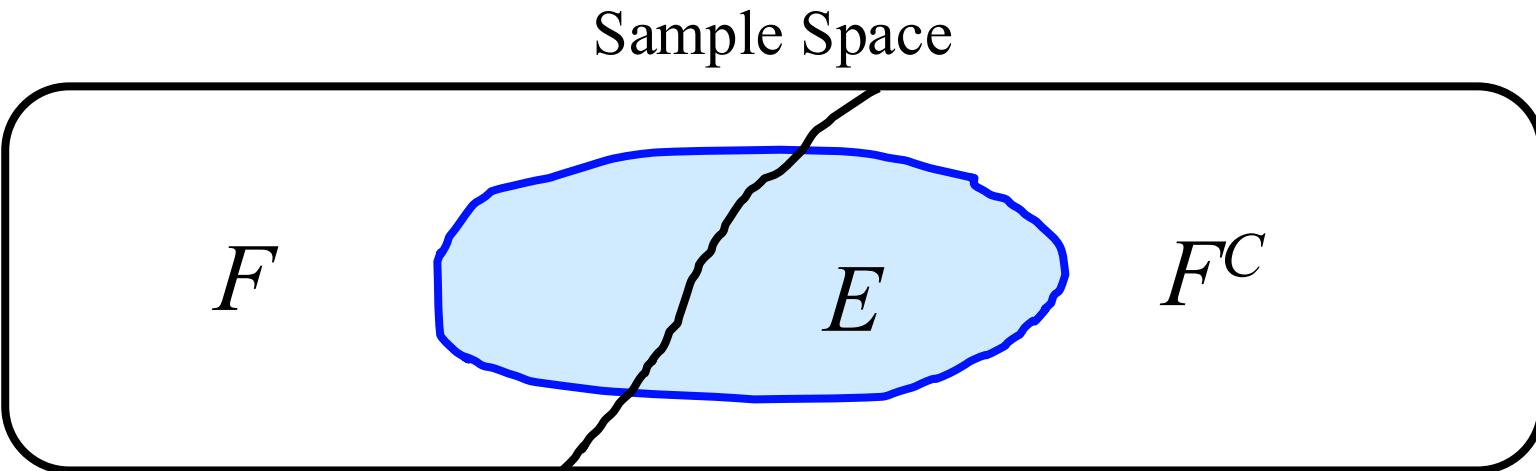
What is the probability of crying, unconditioned?

What information do you need?



Law of Total Probability

Say E and F are events in S

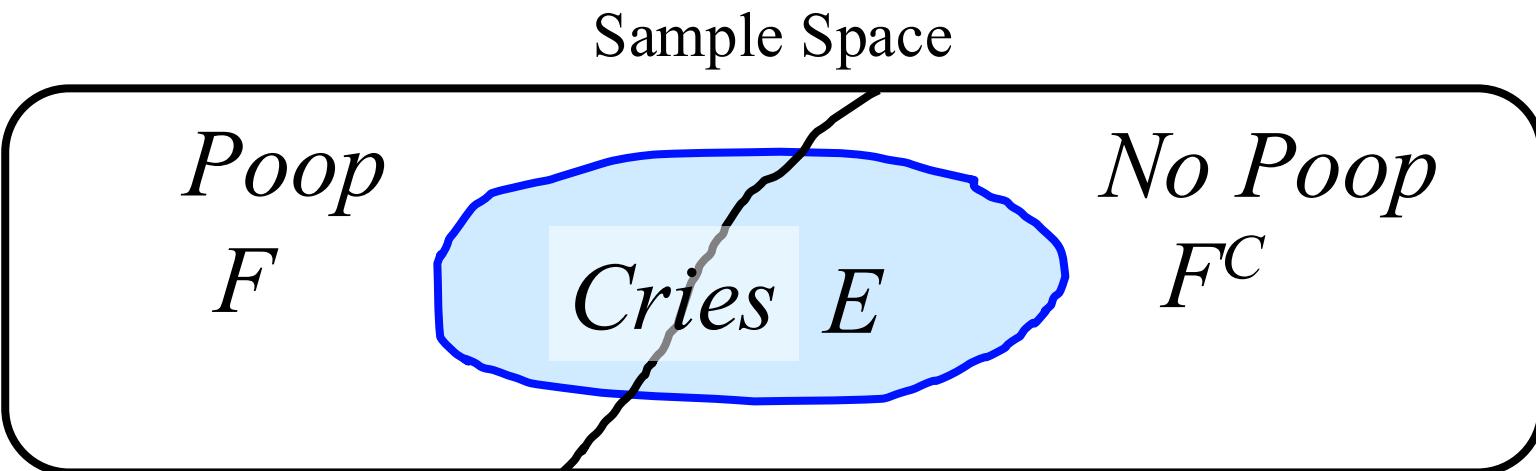


$$\begin{aligned} P(E) &= P(EF) + P(EF^C) \\ &= P(E|F)P(F) + P(E|F^C)P(F^C) \end{aligned}$$



Law of Total Probability

Say E and F are events in S

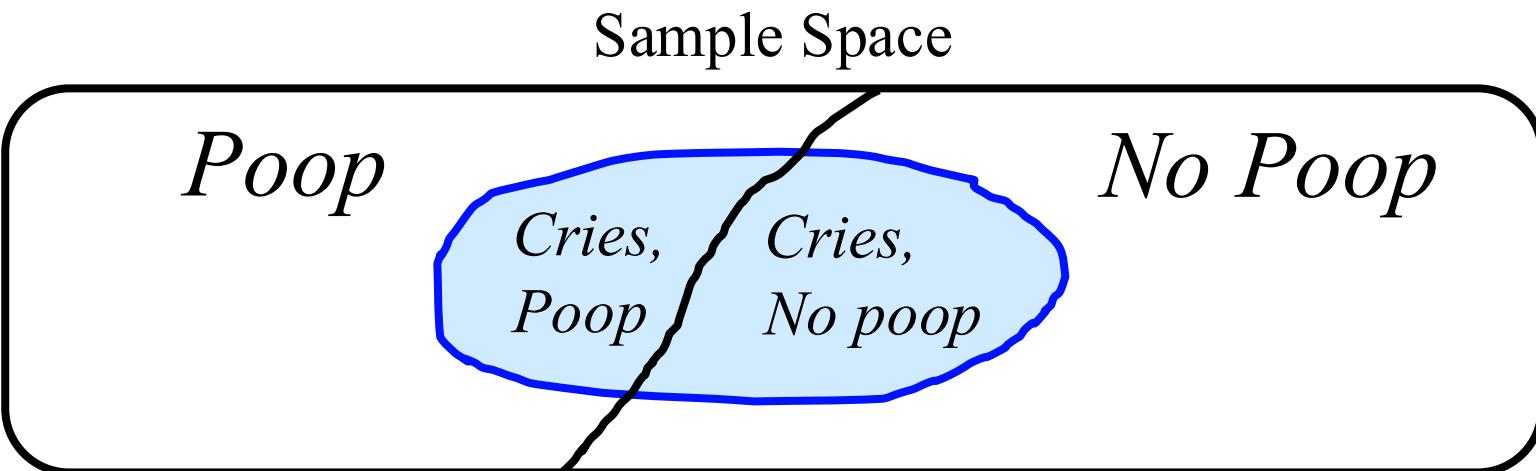


$$\begin{aligned} P(E) &= P(EF) + P(EF^C) \\ &= P(E|F)P(F) + P(E|F^C)P(F^C) \end{aligned}$$



Law of Total Probability

Say E and F are events in S

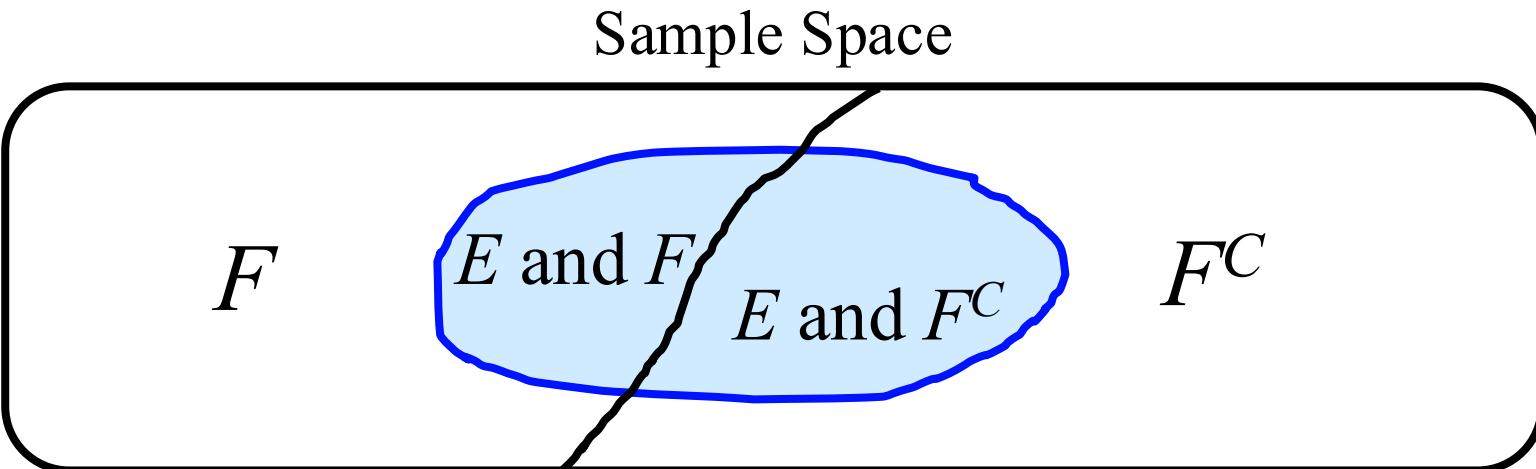


$$\begin{aligned} P(E) &= P(EF) + P(EF^C) \\ &= P(E|F)P(F) + P(E|F^C)P(F^C) \end{aligned}$$



Law of Total Probability

Say E and F are events in S



$$\begin{aligned} P(E) &= P(EF) + P(EF^C) \\ &= P(E|F)P(F) + P(E|F^C)P(F^C) \end{aligned}$$



Law of Total Probability

Thm Let F be an event where $P(F) > 0$. For any event E ,

$$P(E) = P(E|F)P(F) + P(E|F^C)P(F^C)$$

Proof

1. $E = (EF) \text{ or } (EF^C)$

2. $P(E) = P(EF) + P(EF^C)$

3. $P(E) = P(E|F)P(F) + P(E|F^C)P(F^C)$

Since F and F^C are disjoint
Probability of **or** for disjoint
Chain rule (product rule)



Baby Poop

In the morning when she wakes up, a baby has a 50% chance of having pooped. The chance that a baby cries given that she has pooped is 50%.



Probability of crying (E)?

What information do you need?

Probability of crying given no poop.

Recall that E is crying and F is poop

$$P(E) = P(E|F)P(F) + P(E|F^C)P(F^C)$$



Evolution of Bacteria

$$P(E) = P(E|F)P(F) + P(E|F^C)P(F^C)$$

Law of Total Probability



- You have bacteria in your gut which is causing a disease.
- 10% have a mutation which makes them resistant to anti-biotics
- You take half a course of anti-biotics...

Probability a bacteria survives given it has the mutation: 20%

Probability a bacteria survives given it doesn't have the mutation: 1%

What is the probability that a randomly chosen bacteria survives?

Let E be the event that a bacterium survives. Let M be the event that a bacteria has the mutation. By the Law of Total Probability (LOTP):

$$\begin{aligned} \Pr(E) &= \Pr(E \text{ and } M) + \Pr(E \text{ and } M^C) && \text{LOTP} \\ &= \Pr(E|M)\Pr(M) + \Pr(E|M^C)\Pr(M^C) && \text{Chain Rule} \\ &= 0.20 \cdot 0.10 + 0.01 \cdot 0.90 && \text{Substituting} \\ &= 0.029 \end{aligned}$$



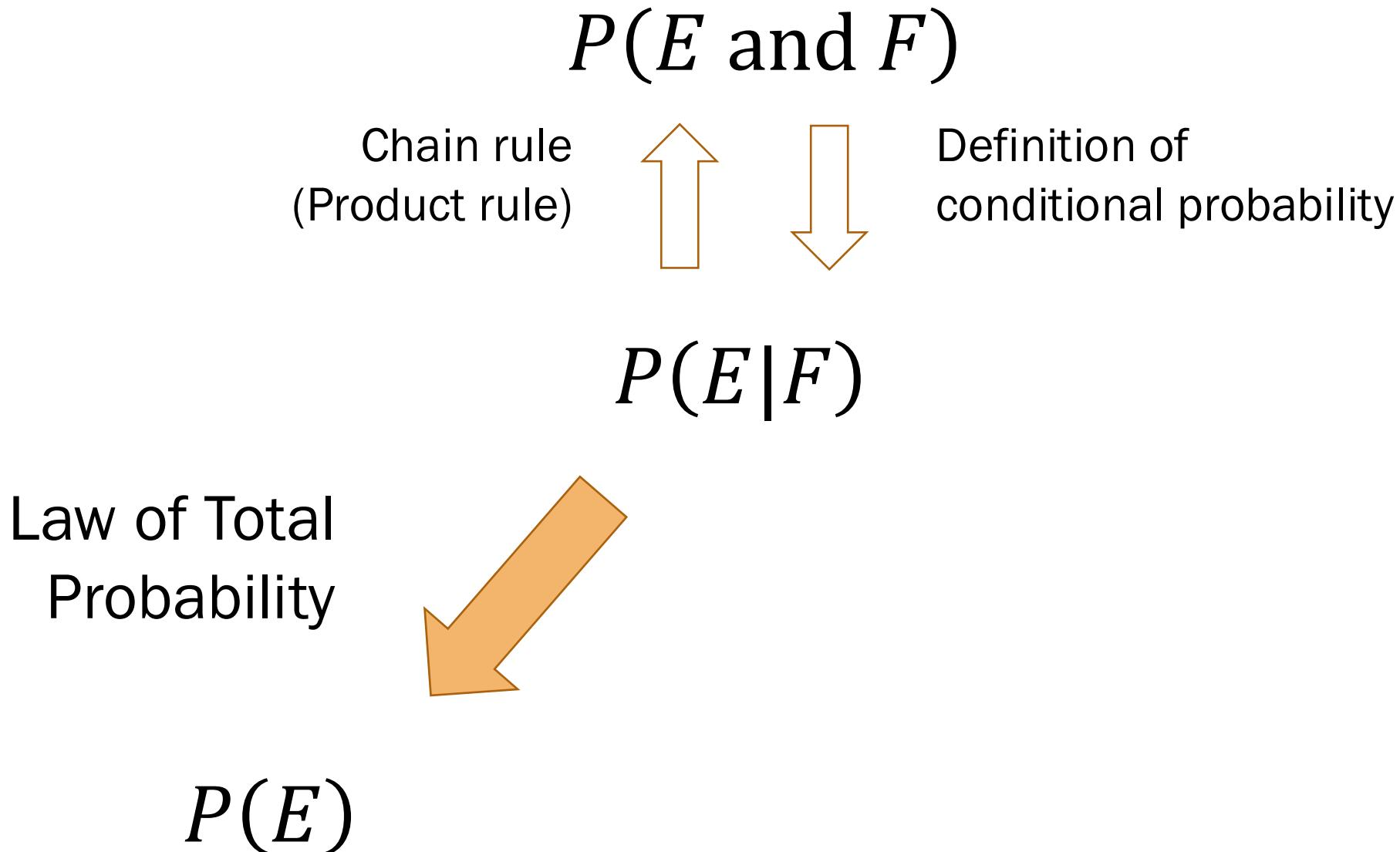
Know:

$P(\text{Survive} \mid \text{Mutation})$, $P(\text{Survive})$, $P(\text{Mutation})$

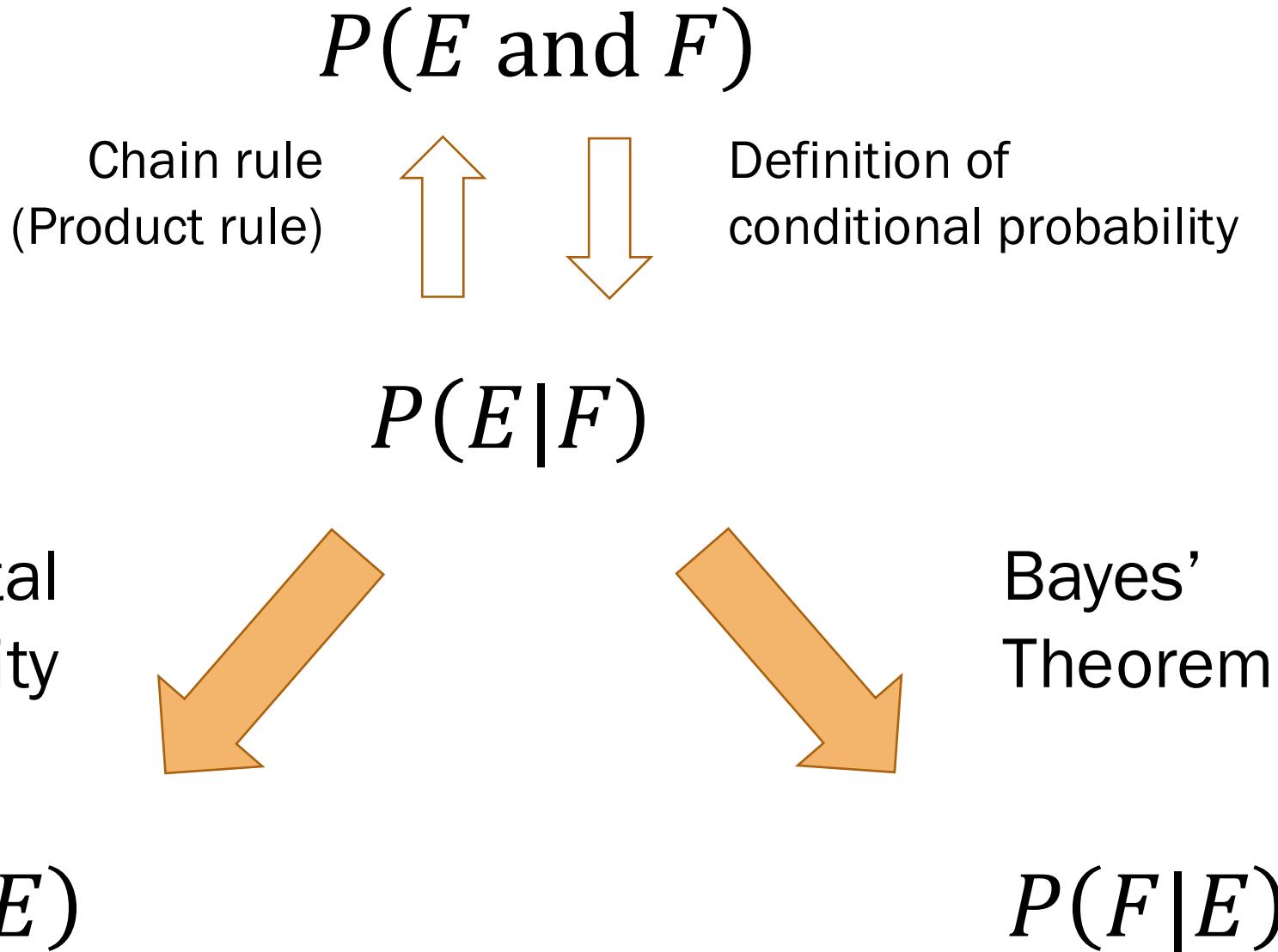
Real question. What is the probability of mutation given the bacteria survived?

$P(\text{Mutation} \mid \text{Survive})$

Relationship Between Probabilities



Relationship Between Probabilities



Bayes' Theorem

Thomas Bayes

Rev. Thomas Bayes (~1701-1761):
British mathematician and Presbyterian minister



He looked remarkably similar to Sean Astin
(but that's not important right now)



Thomas Bayes



$$P(F | E)$$

I want to calculate

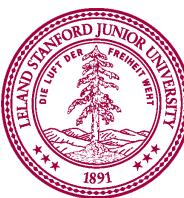
$$P(\text{State of the world } F | \text{Observation } E)$$

It seems so tricky!...

The other way around is easy

$$P(\text{Observation } E | \text{State of the world } F)$$

$$P(E | F)$$





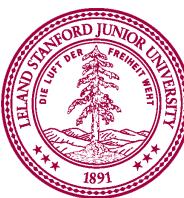
$$P(F|E) = \frac{P(E|F)P(F)}{P(E)} \quad \text{Def. of Conditional Prob.}$$

A little while later...

$$= \frac{P(E|F)P(F)}{P(E)} \quad \text{Chain Rule}$$

A little while later...

$$= \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^C)P(F^C)} \quad \text{LOTP}$$



(silent drumroll)



Bayes' Theorem

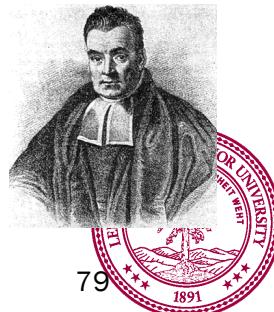
$$P(E|F) \xrightarrow{\text{orange arrow}} P(F|E)$$

Thm For any events E and F where $P(E) > 0$ and $P(F) > 0$,

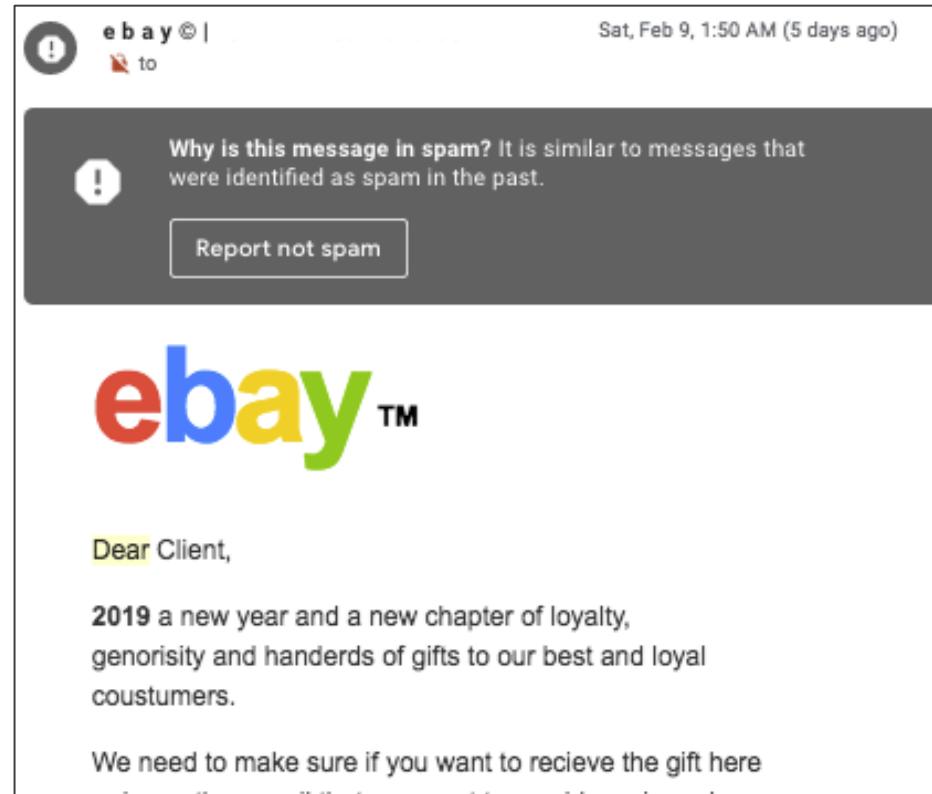
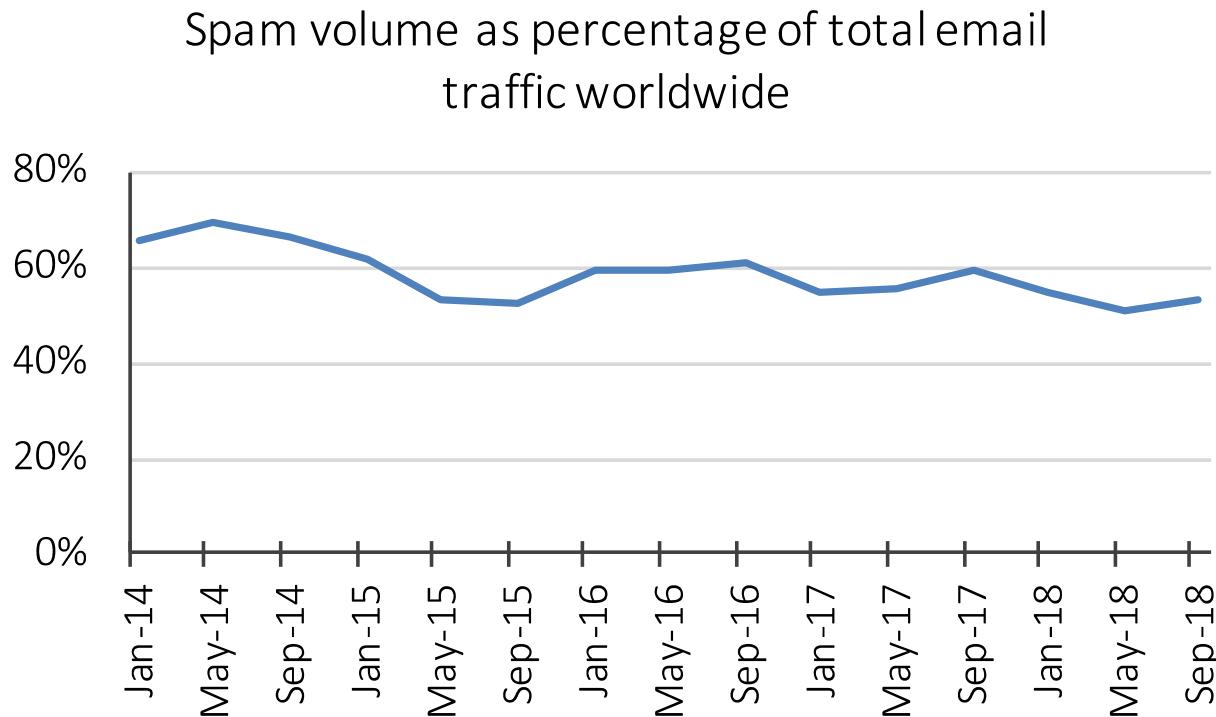
$$P(F|E) = \frac{P(E|F)P(F)}{P(E)}$$

Expanded form:

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^C)P(F^C)}$$



Detecting spam email



We can easily calculate how many spam emails contain “Dear”:

$$P(E|F) = P(\text{“Dear”} \mid \text{Spam email})$$

But what is the probability that an email containing “Dear” is spam?

$$P(F|E) = P(\text{Spam email} \mid \text{“Dear”})$$



Detecting spam email

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^C)P(F^C)}$$
 Bayes' Theorem

- 60% of all email in 2016 is spam.
- 20% of spam has the word “Dear”
- 1% of non-spam (aka ham) has the word “Dear”

You get an email with the word “Dear” in it.

What is the probability that the email is spam?

1. Define events
& state goal

2. Identify known
probabilities

3. Solve

Let: E : “Dear”, F : spam

Want: $P(\text{spam}|\text{“Dear”})$
 $= P(F|E)$



Bayes' Theorem terminology

Let: E : “Dear”, F : spam
Want: $P(F | E)$

- 60% of all email in 2016 is spam.
- 20% of spam has the word “Dear”
- 1% of non-spam (aka ham) has the word “Dear”

$P(F)$

$P(E|F)$

$P(E|F^C)$

You get an email with the word “Dear” in it.

What is the probability that the email is spam?

Want: $P(F|E)$

$$P(F|E) = \frac{\text{likelihood} \quad \text{prior}}{\text{normalization constant}}$$
$$P(F|E) = \frac{P(E|F)P(F)}{P(E)}$$



SARS Virus Testing

A test is 98% effective at detecting SARS

- However, test has a “false positive” rate of 1%
- 0.5% of US population has SARS
- Let E = you test positive for SARS with this test
- Let F = you actually have SARS
- What is $P(F | E)$?

Solution:

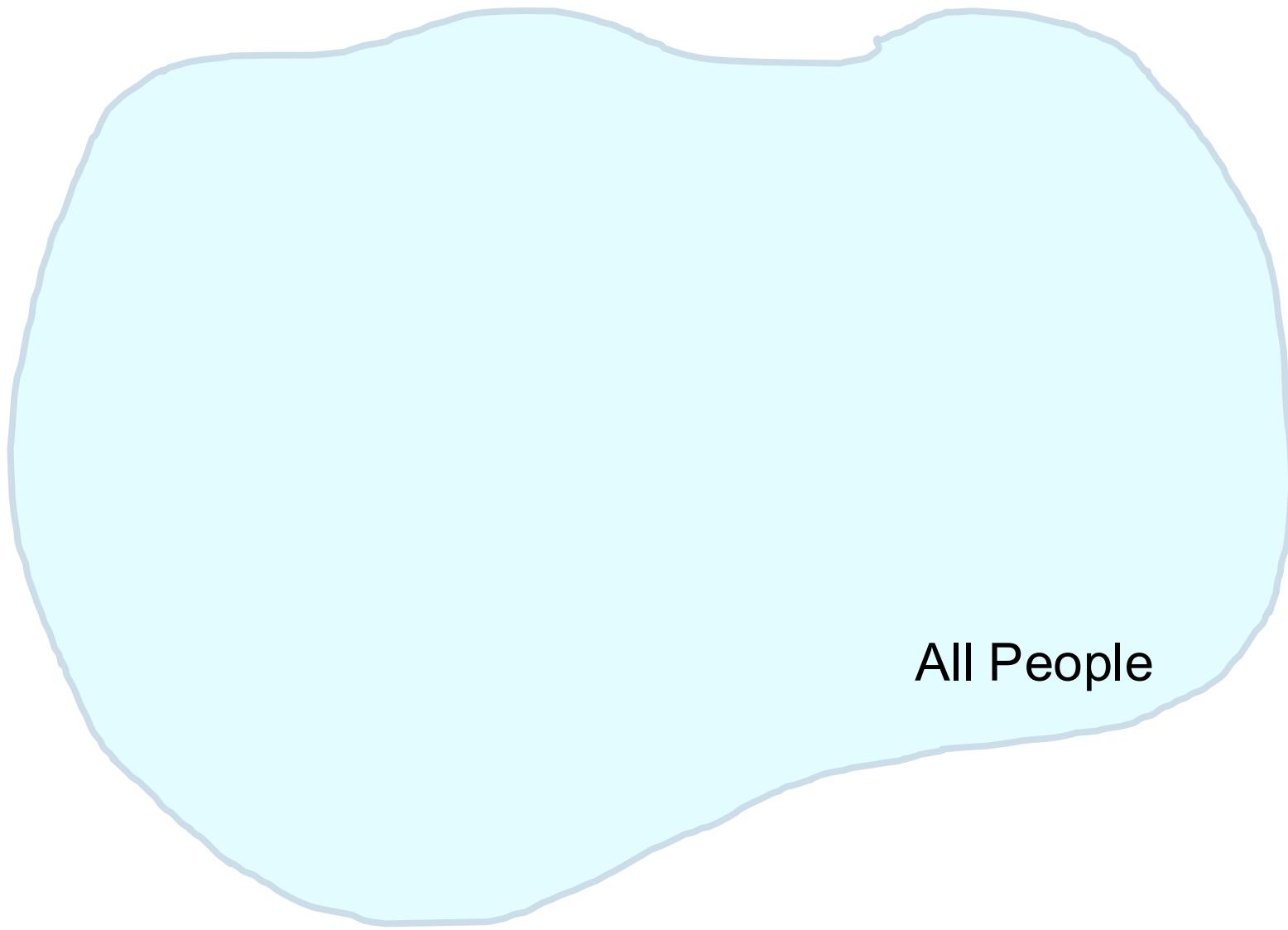
$$P(F | E) = \frac{P(E | F) P(F)}{P(E | F) P(F) + P(E | F^c) P(F^c)}$$

$$P(F | E) = \frac{(0.98)(0.005)}{(0.98)(0.005) + (0.01)(1 - 0.005)} \approx 0.330$$

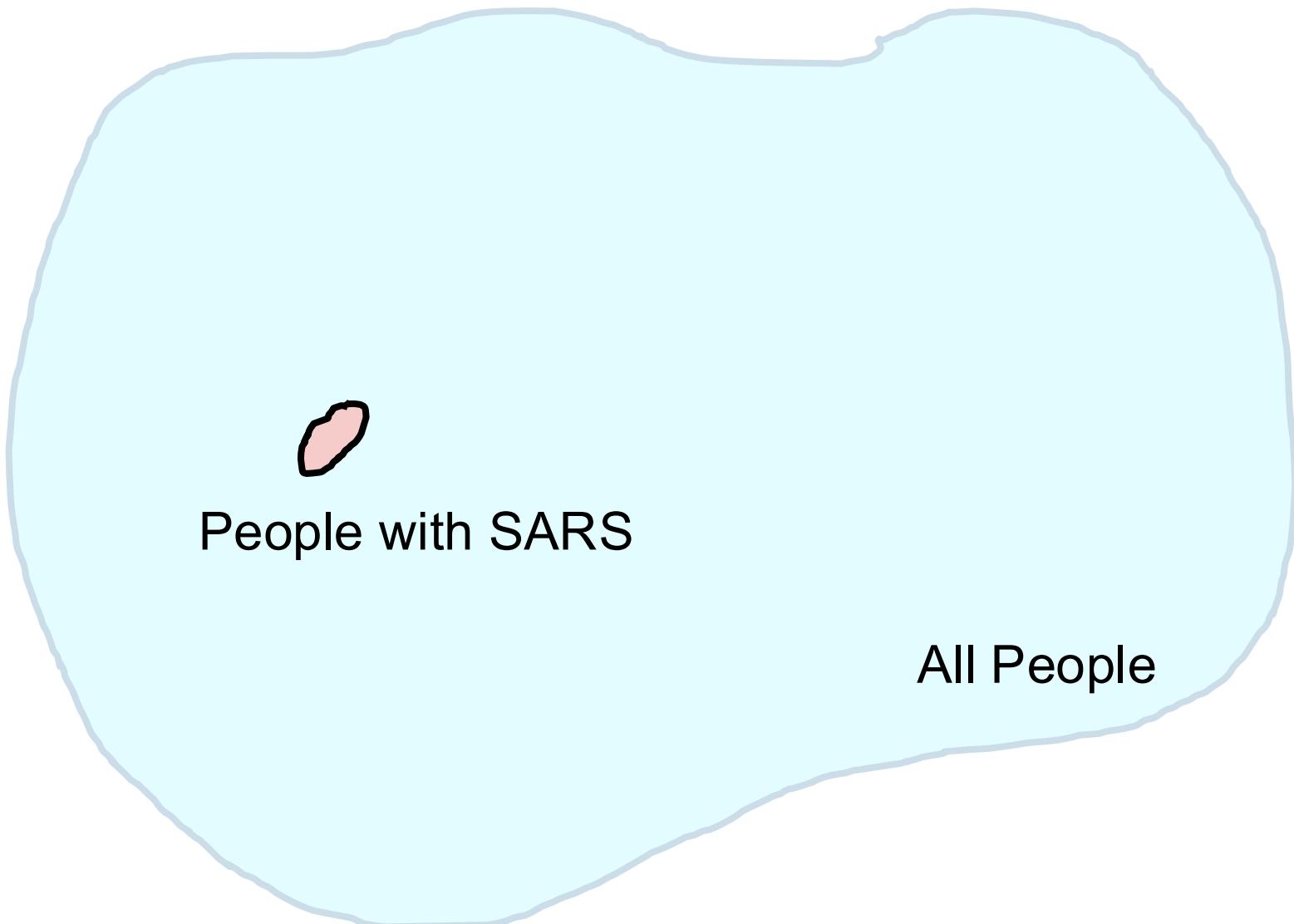


Intuition Time

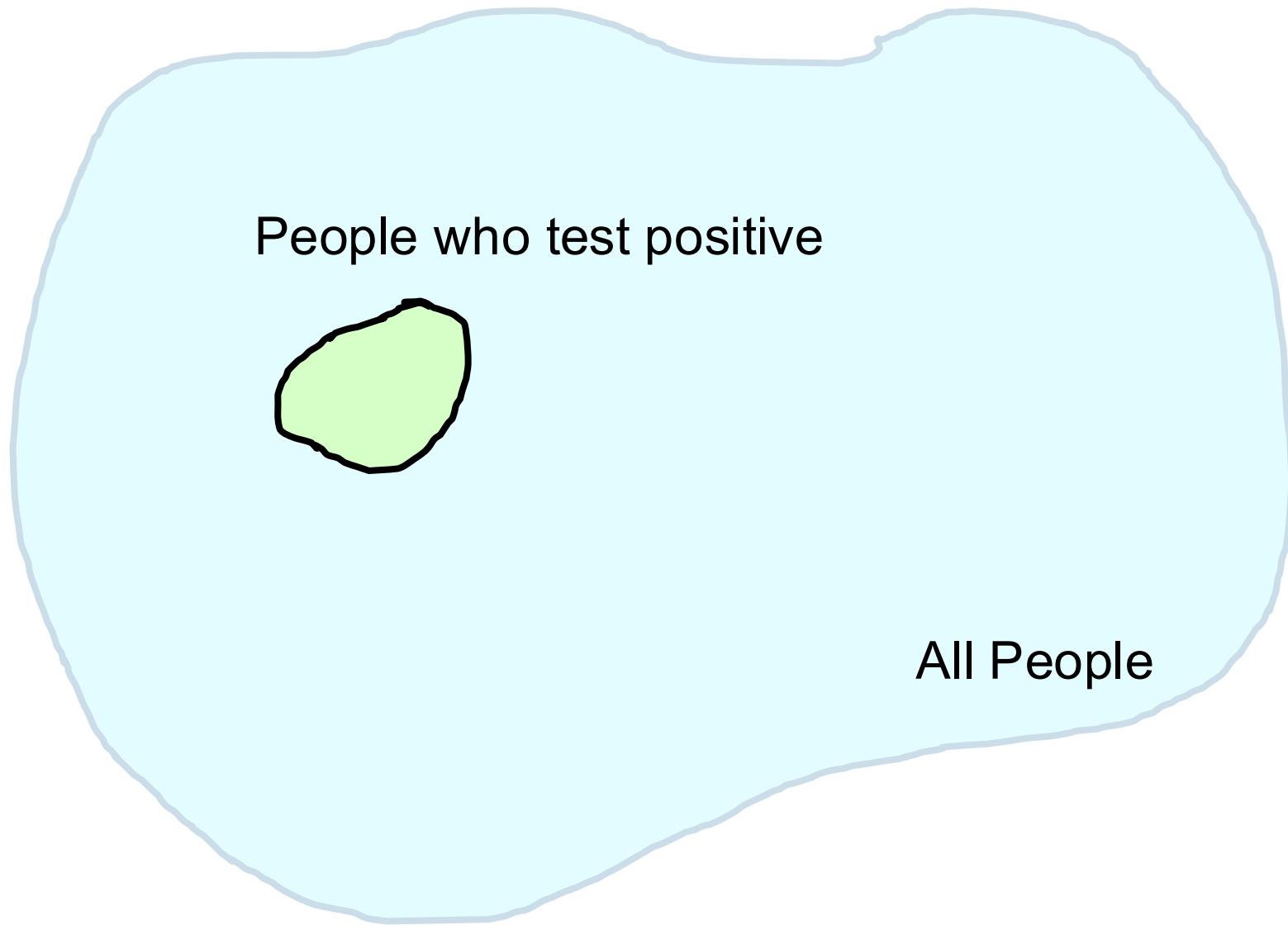
Bayes Theorem Intuition



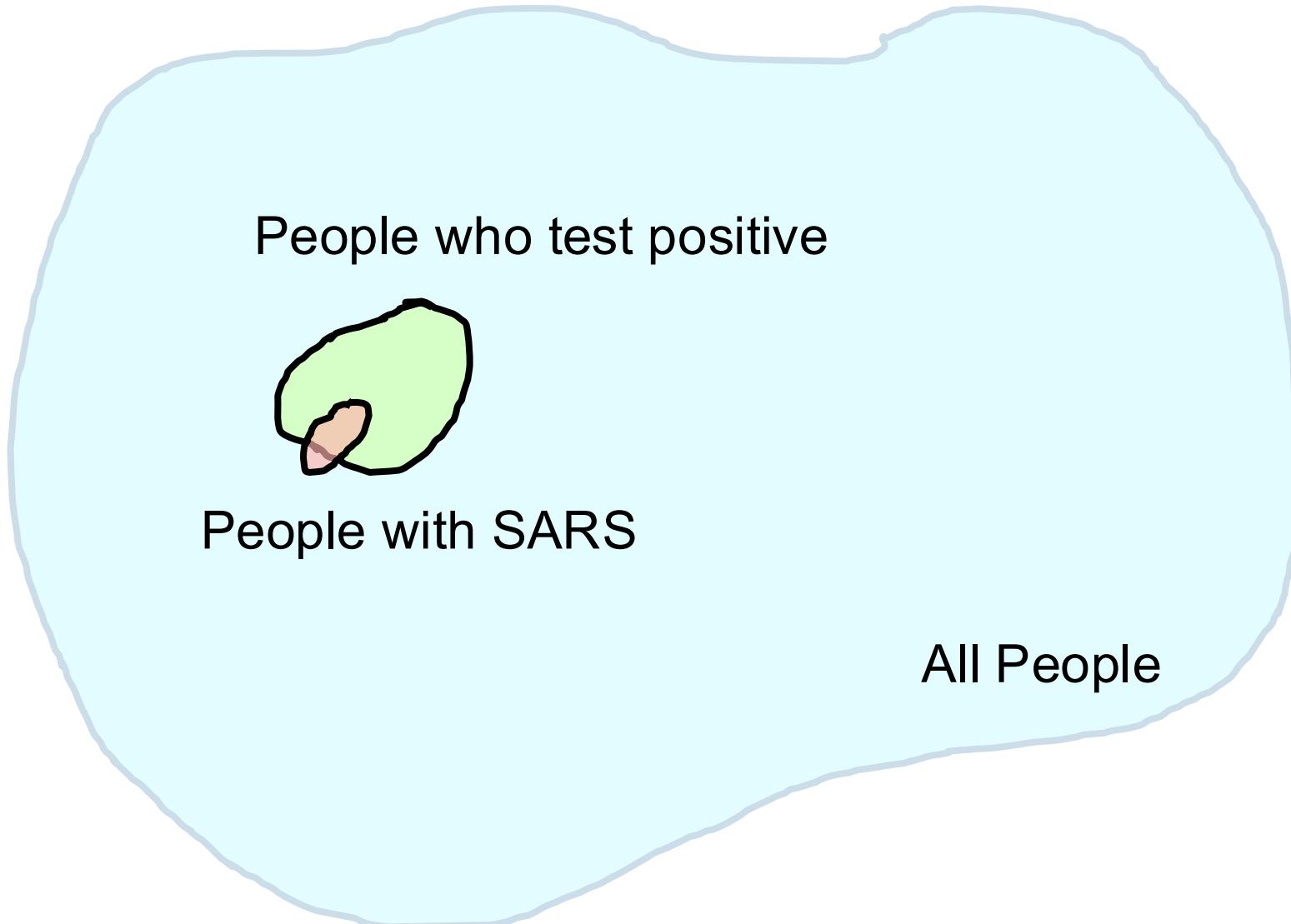
Bayes Theorem Intuition



Bayes Theorem Intuition

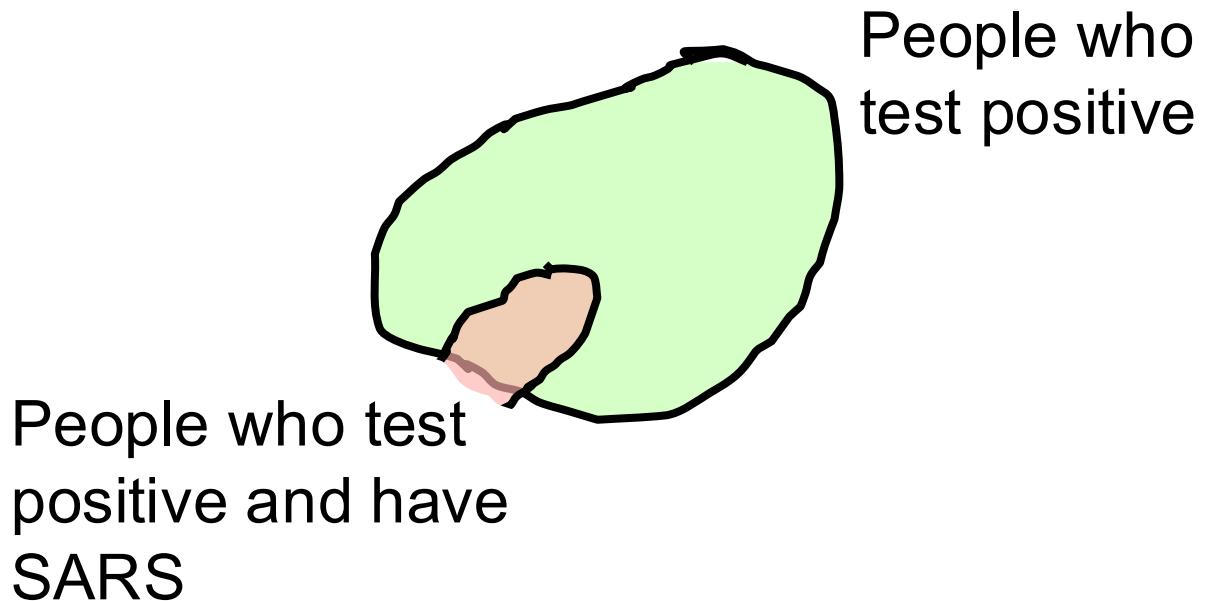


Bayes Theorem Intuition

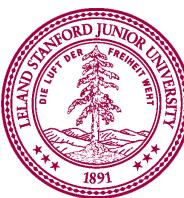


Bayes Theorem Intuition

Conditioning on a positive result changes the sample space to this:

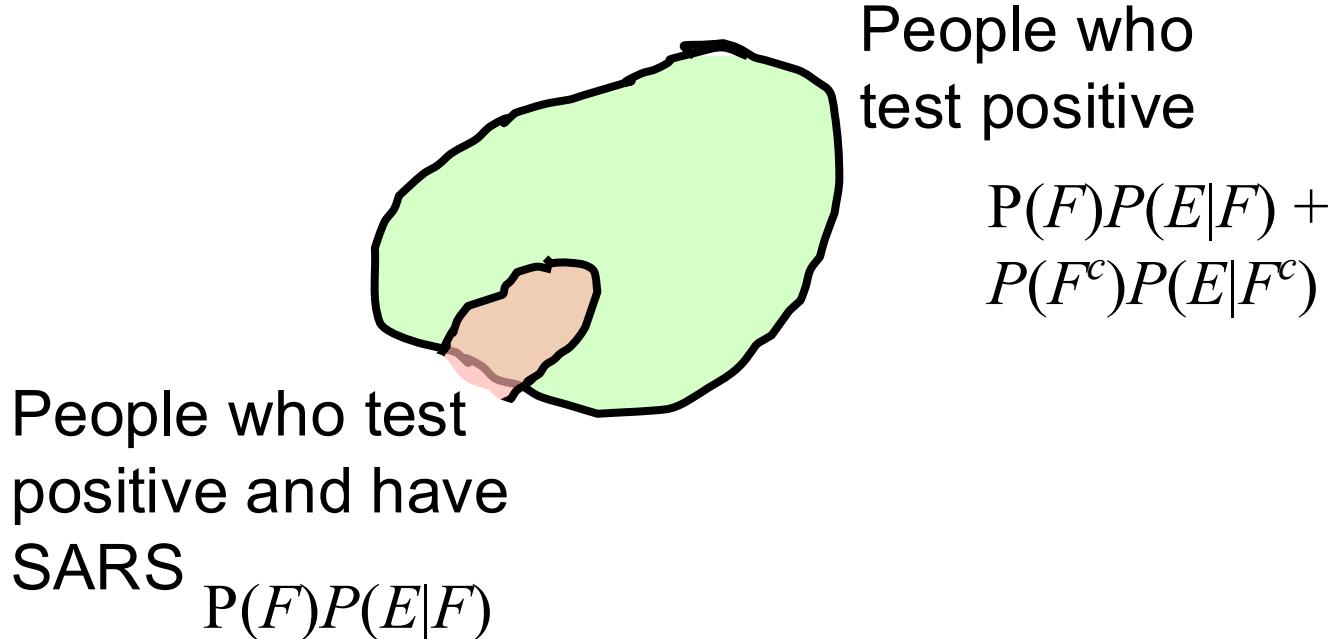


≈ 0.330



Bayes Theorem Intuition

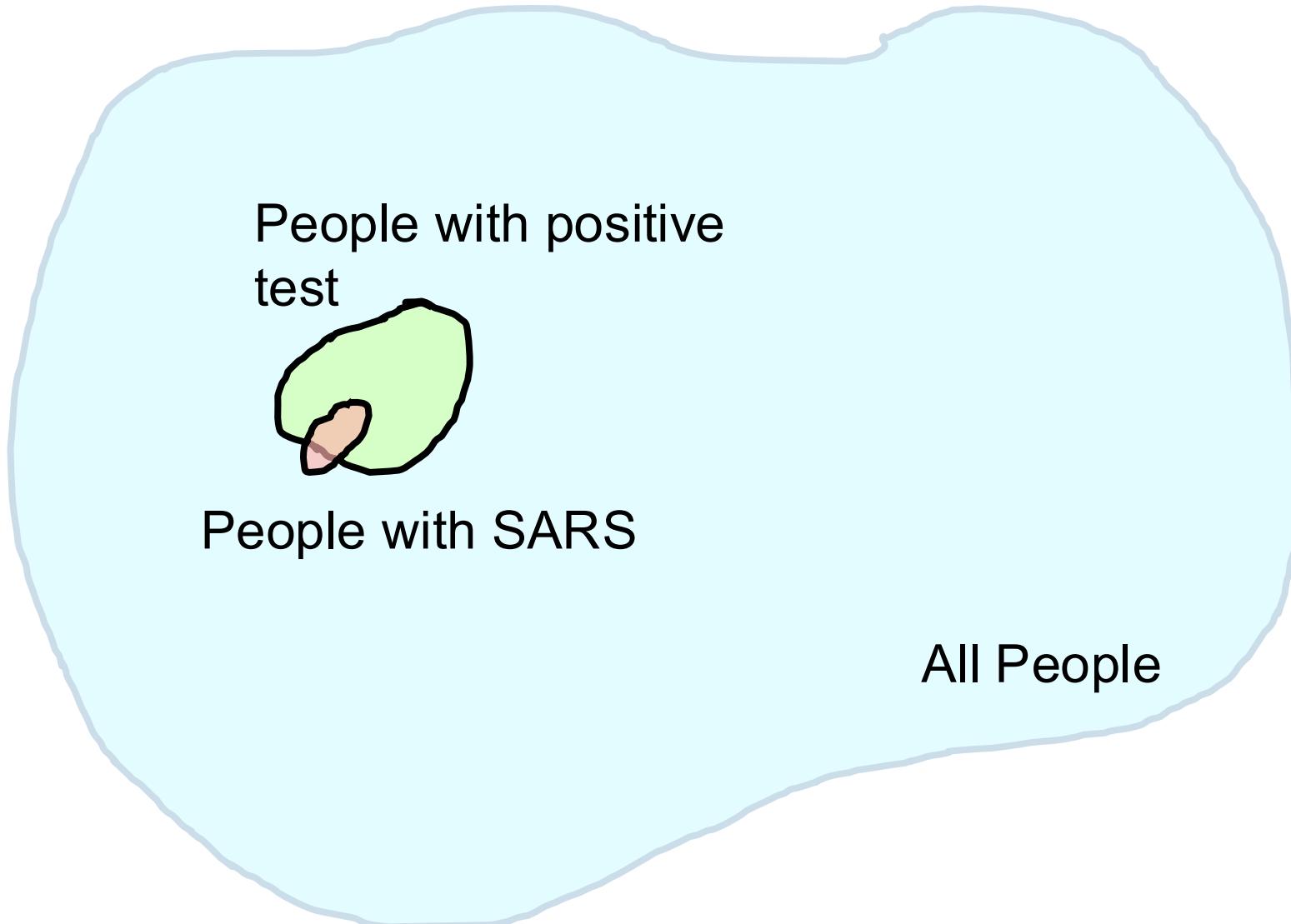
Conditioning on a positive result changes the sample space to this:



≈ 0.330

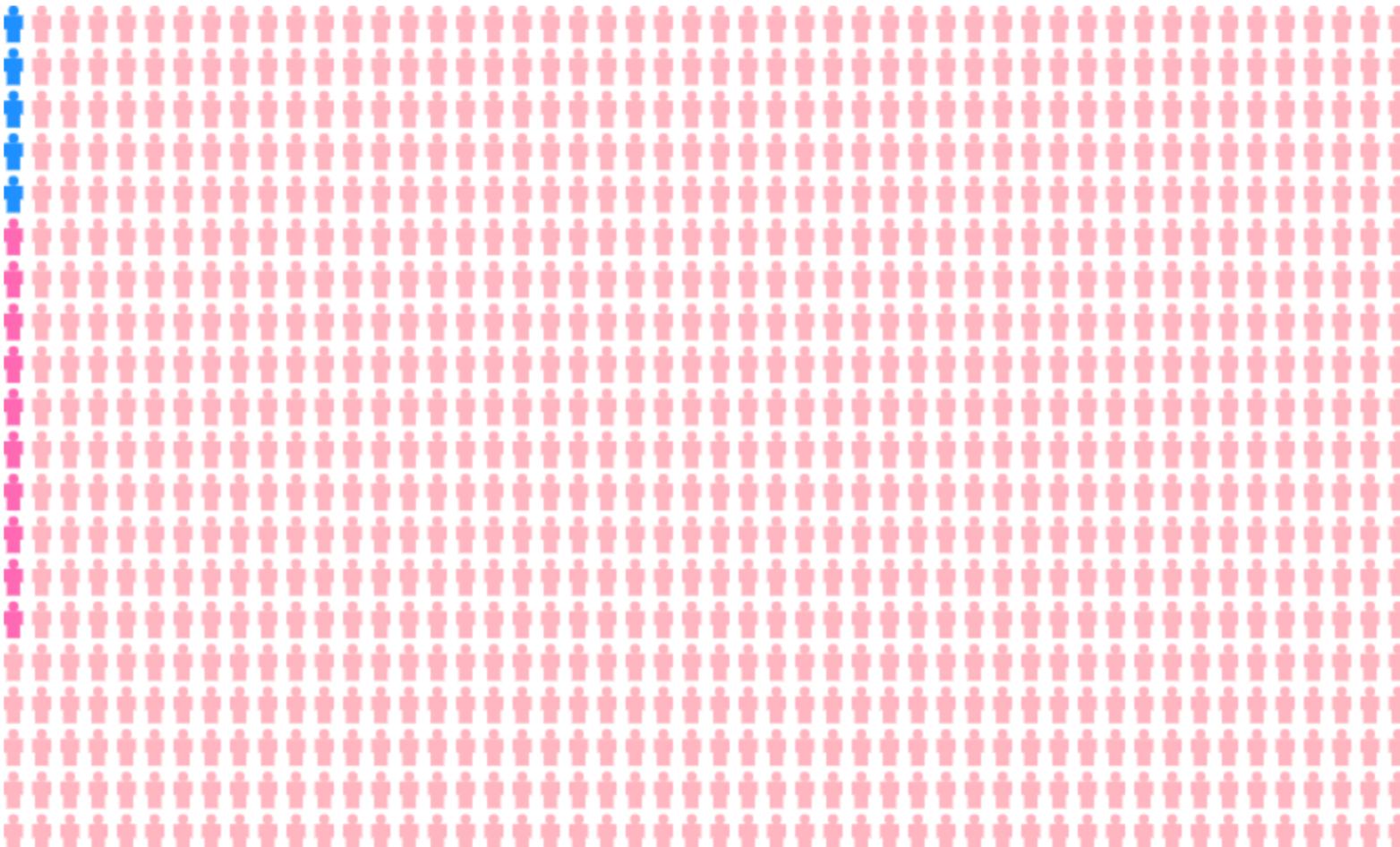


Bayes Theorem Intuition



Bayes Theorem Intuition

Say we have 1000 people:

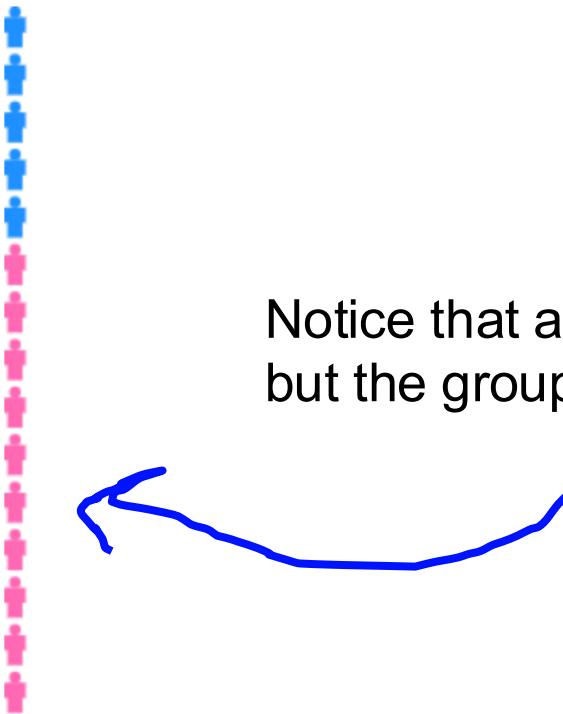


5 have SARS and test positive, 985 **do not** have SARS and test negative.
10 **do not** have SARS and test positive. ≈ 0.333



Bayes Theorem Intuition

Conditioned on just those that test positive:



Notice that all the people with SARS are here, but the group is still mainly folks without SARS

5 have SARS and test positive, 985 **do not** have SARS and test negative.
10 **do not** have SARS and test positive. ≈ 0.333



Why it is still good to get tested

	SARS +	SARS -
Test +	$0.98 = P(E F)$	$0.01 = P(E F^c)$
Test -	$0.02 = P(E^c F)$	$0.99 = P(E^c F^c)$

- Let E^c = you test negative for SARS with this test
- Let F = you actually have SARS
- 0.5% of population has SARS
- What is $P(F | E^c)$?

$$P(F | E^c) = \frac{P(E^c | F) P(F)}{P(E^c | F) P(F) + P(E^c | F^c) P(F^c)}$$

$$P(F | E^c) = \frac{(0.02)(0.005)}{(0.02)(0.005) + (0.99)(1 - 0.005)} \approx 0.0001$$



The End for Wednesday

Spam Revisited



How would you detect Spam using an LLM?

Real spam email:

“Pay for Viagra with a credit-card. Viagra is great.
Risk free Viagra. Click for free.”



1 Let E be email text. Let F be event the email is Spam.



$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^C)P(F^C)}$$

60%

60%

40%



How would you detect Spam using an LLM?

1

Let E be email text. Let F be event the email is Spam.

2

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^C)P(F^C)}$$

60% 40%



How would you detect Spam using an LLM?

- 1 Let E be email text. Let F be event the email is Spam.

- 2
$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^C)P(F^C)}$$

60%	
60%	40%

- 3 Make assumption that LLM understands probability

```
prompt = "This email is spam: "
```

```
p_email_given_spam = string_pr(prompt + email) / string_pr(prompt)
```

```
prompt = ("This email is NOT spam: ")
```

```
p_email_given_ham = string_pr(prompt + email) / string_pr(prompt)
```



Mysteries

Whats an LLMs (real) belief?

Llama-3.3-70B

```
from llm import string_pr
pr = string_pr("Hello, world")
print(pr)
```

If you used a prompt like this:

```
f"Dice simulator output. Sum of two random 6 sided dice:
{outcome}"
```

Could you test an LLMs understanding of probability?

Outcome	string_pr
2	3.0×10^{-30}
:	:
7	9.9×10^{-29}
:	:
12	2.4×10^{-30}

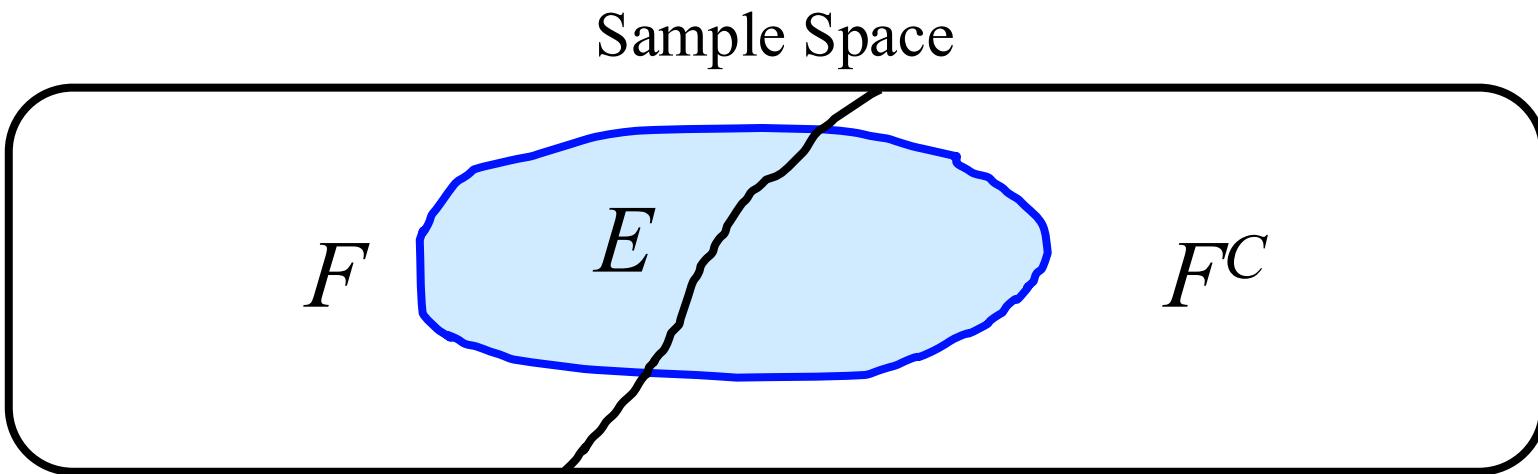


Come on Friday!

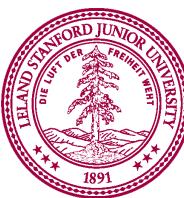
Stories + Make History

Sneak Peek...

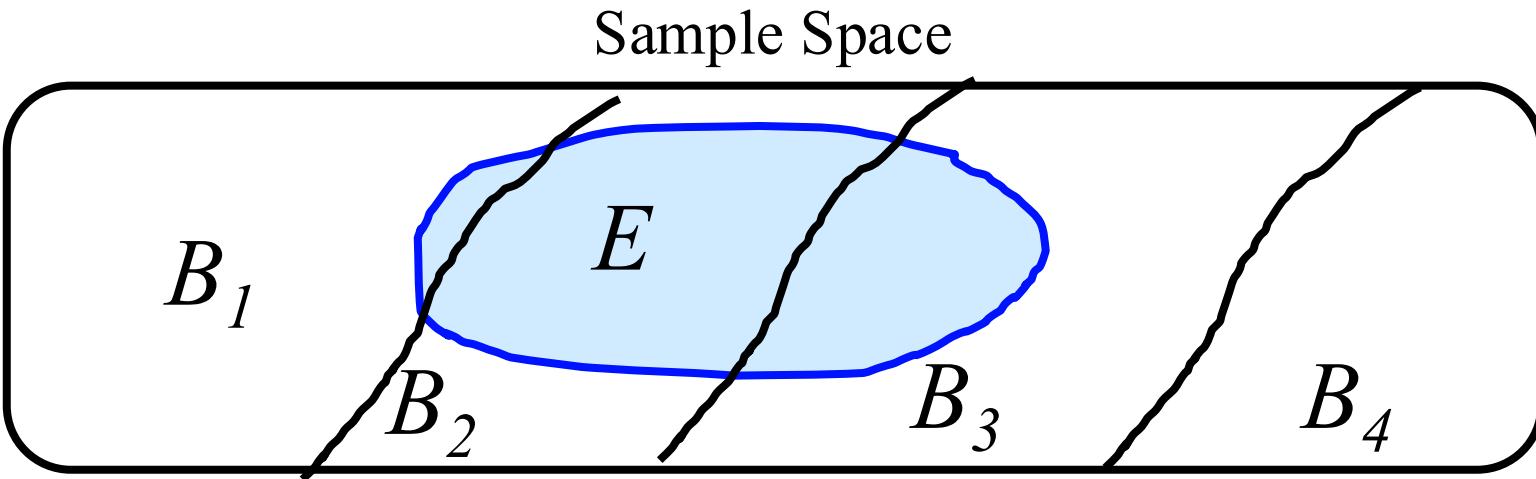
Law of Total Probability



$$\begin{aligned} P(E) &= P(EF) + P(EF^C) \\ &= P(E|F)P(F) + P(E|F^C)P(F^C) \end{aligned}$$



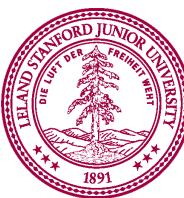
Law of Total Probability



Thm

For **mutually exclusive events** B_1, B_2, \dots, B_n
s.t. $B_1 \cup B_2 \cup \dots \cup B_n = S$,

$$\begin{aligned} P(E) &= \sum_i \underline{P(B_i \cap E)} \\ &= \sum_i P(E|B_i)P(B_i) \end{aligned}$$



Results for **San Francisco, CA**



49

°F | °C

Precipitation: 90%

Humidity: 71%

Wind: 8 mph

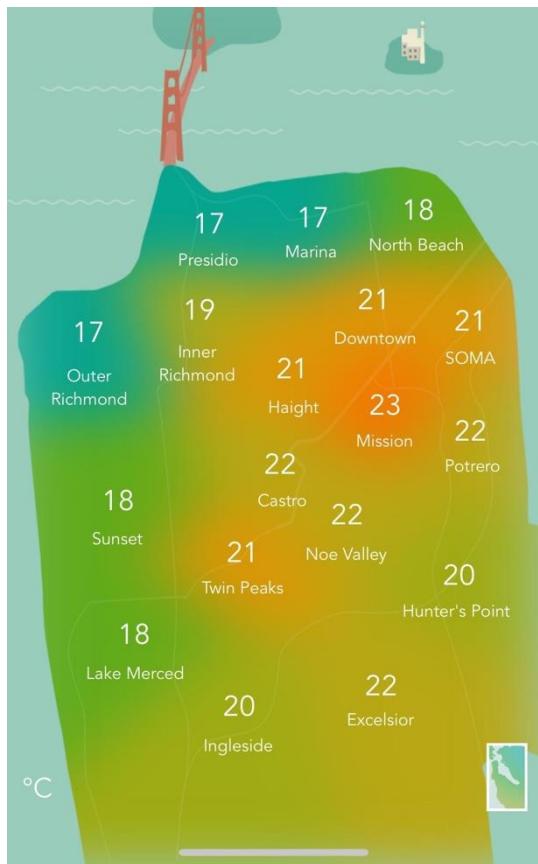
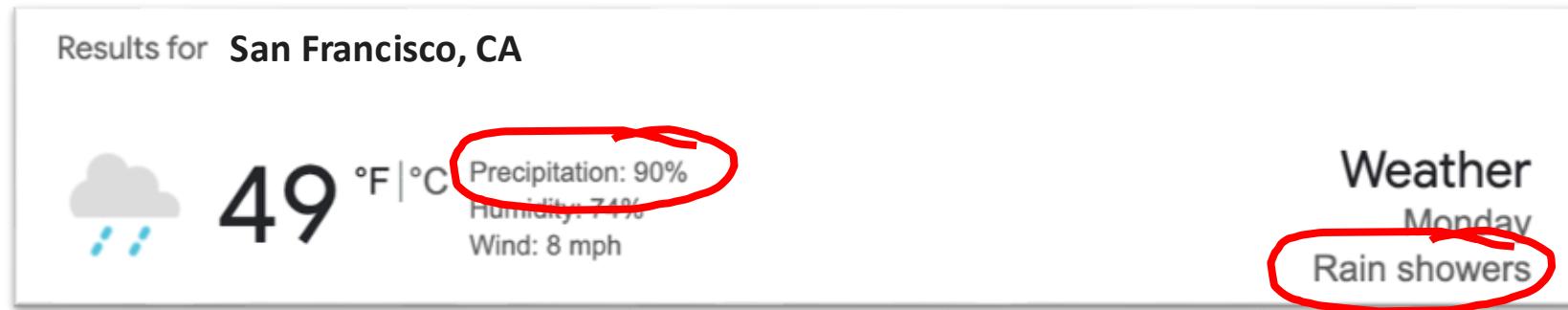
Weather

Monday

Rain showers

Background event.
Where is the person in
San Francisco?

Results for **San Francisco, CA**



From Google's Perspective:
There are 18 different “districts” in San Francisco.

Know:

It rains tomorrow

$$P(R|D_i)$$

Person is in district i

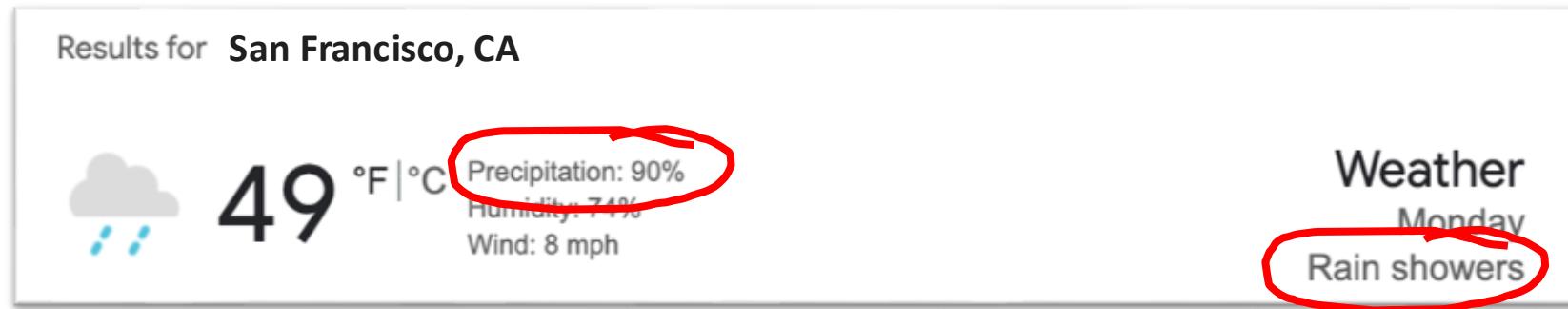
$$P(D_i)$$

Want:

$$P(R)$$

Background event.
Where is the person in
San Francisco?

Results for **San Francisco, CA**



Background event.
Where is the person in
San Francisco?

From Google's Perspective:
There are 18 different “districts” in San Francisco.

Know:

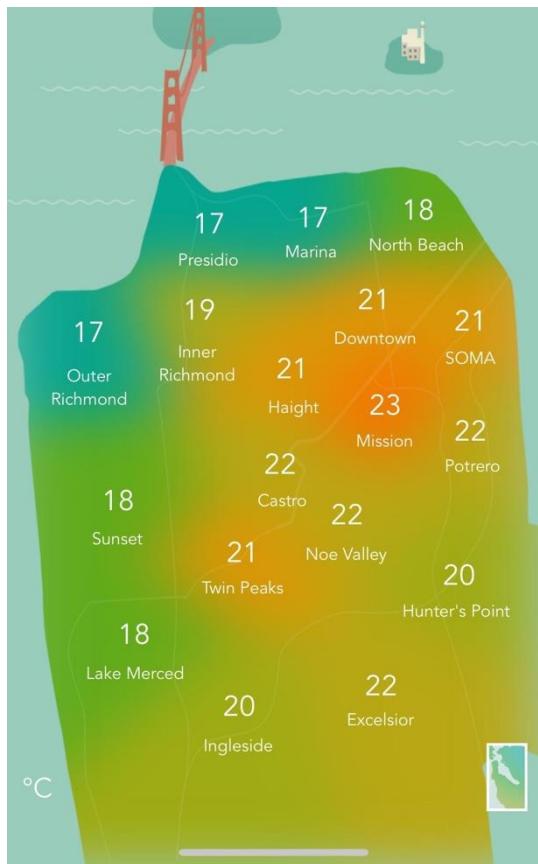
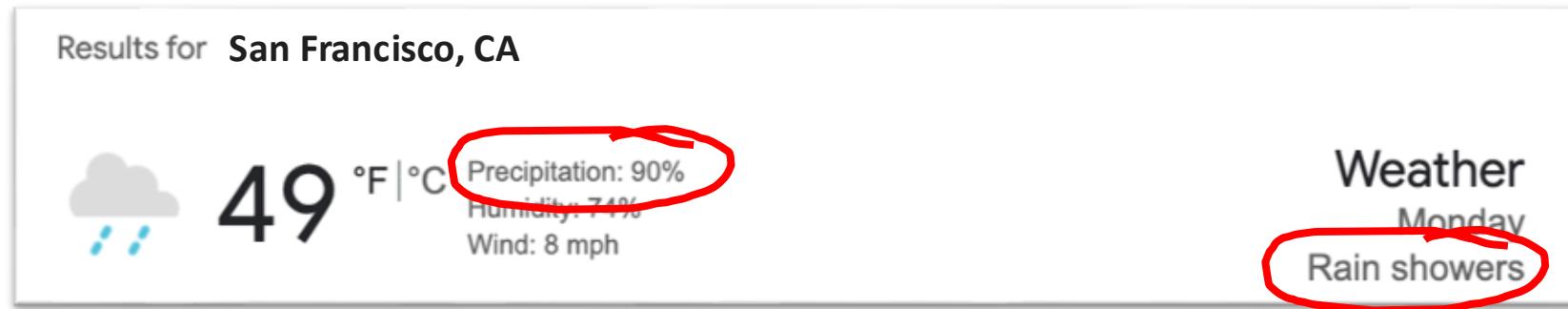
	Mission District	Presidio	SOMA
$P(R D_i)$	0.23	0.84	0.52
$P(D_i)$	0.15	0.02	0.24

Want:

$$P(R)$$



Results for **San Francisco, CA**



Background event.
Where is the person in
San Francisco?

From Google's Perspective:
There are 18 different “districts” in San Francisco.

Know:

	Mission District	Presidio	SOMA
$P(R D_i)$	0.23	0.84	0.52
$P(D_i)$	0.15	0.02	0.24

Want:

$$P(R) = \sum_{\text{district } i} P(R \text{ and } D_i) = \sum_{\text{district } i} P(R|D_i) \cdot P(D_i)$$

SARS Virus Testing

A test is 98% effective at detecting SARS

- However, test has a “false positive” rate of 1%
- 0.5% of US population has SARS
- Let E = you test positive for SARS with this test
- Let F = you actually have SARS
- What is $P(F | E)$?

Solution:

$$P(F | E) = \frac{P(E | F) P(F)}{P(E | F) P(F) + P(E | F^c) P(F^c)}$$

$$P(F | E) = \frac{(0.98)(0.005)}{(0.98)(0.005) + (0.01)(1 - 0.005)}$$

≈ 0.330



Pedagogic Pause

Multiple Choice Theory

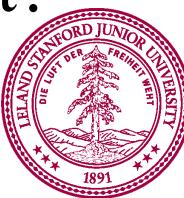
Let's consider the relationship between **knowing** the concepts used in a multiple choice midterm question, and getting the question correct, taking into account guessing and making silly mistakes.

Let $3/4$ be the prior probability that a learner knows the concepts to a midterm question.

Let $1/4$ be the probability that a learner gets the answer **correct** if they **don't** know the concepts.

Let $9/10$ be the probability that a learner gets the question **correct** given they **do** know the concepts.

What is the probability they know the concept, given they answered correct?



Monty Hall Problem

Monty Hall Problem



Monty Hall Problem aka Let's Make a Deal

Behind one door is a prize (equally likely to be any door).

Behind the other two doors is nothing

1. We choose a door
2. Host opens 1 of other 2 doors, revealing nothing
3. We are given an option to change to the other door.

Should we switch?

Note: If we don't switch,

$$P(\text{Win}) = 1/3$$



Doors A,B,C



In the world where we switch

Without loss of generality, say we pick A (out of Doors A,B,C).

1/3
↓

A: Prize in Door A

- Host opens B or C
- We switch
- We always lose

$$P(\text{Win} | A) = 0$$

1/3
↓

B: Prize in Door B

- Host must open C
- We switch to B
- We always win

$$P(\text{Win} | B) = 1$$

1/3
↓

C: Prize in Door C

- Host must open B
- We switch to C
- We always win

$$P(\text{Win} | C) = 1$$

$$\begin{aligned}P(\text{Win}) &= P(\text{Win}|A)P(A) + P(\text{Win}|B)P(B) + P(\text{Win}|C)P(C) \\&= \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 1 = \frac{2}{3}\end{aligned}$$

You should switch!

Marilyn Vos Savant



Ask Marilyn™

BY MARILYN VOS SAVANT



Monty Hall, 1000 envelope version

Start with 1000 envelopes (of which 1 is the prize).

1. You choose 1 envelope.

$$\left\{ \begin{array}{l} \frac{1}{1000} = P(\text{envelope is prize}) \\ \frac{999}{1000} = P(\text{other 999 envelopes have prize}) \end{array} \right.$$

2. I open 998 of remaining 999 (showing they are empty).

$$\begin{aligned} \frac{999}{1000} &= P(998 \text{ empty envelopes had prize}) \\ &\quad + P(\text{last other envelope has prize}) \\ &= P(\text{last other envelope has prize}) \end{aligned}$$

3. Should you switch?

No: $P(\text{win without switching}) =$

$$\frac{1}{\text{original \# envelopes}}$$

Yes: $P(\text{win with new knowledge}) =$

$$\frac{\text{original \# envelopes} - 1}{\text{original \# envelopes}}$$