

Conditional Probability and Bayes

Review

What is a Probability?

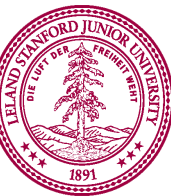
A number $[0, 1]$ to which we ascribe meaning

The event we
care about

How many times does it
occur

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

Out of (close to) infinite
trials

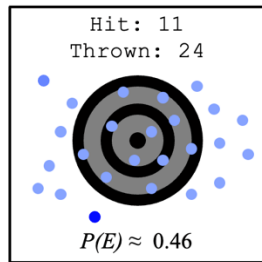


Sources of Probability

Infinite Trials

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

n is the number of trials



The "event" E is that you hit the target

Datasets

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

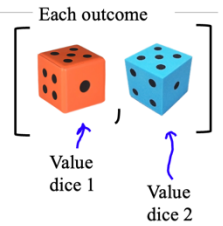
Sum of Two Die = 7?

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]

$\}$



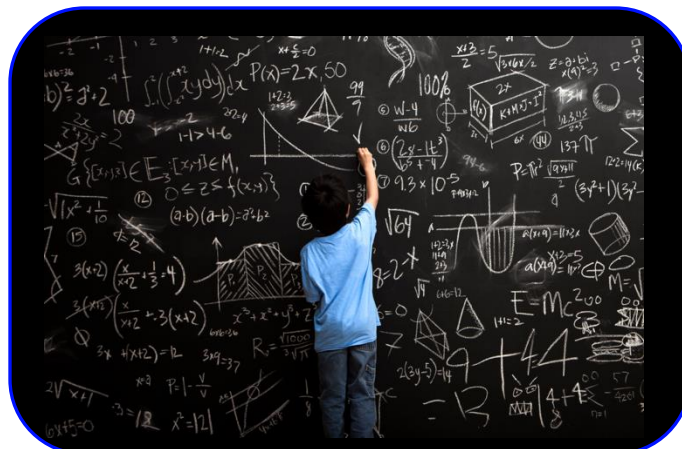
$E =$ in blue

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

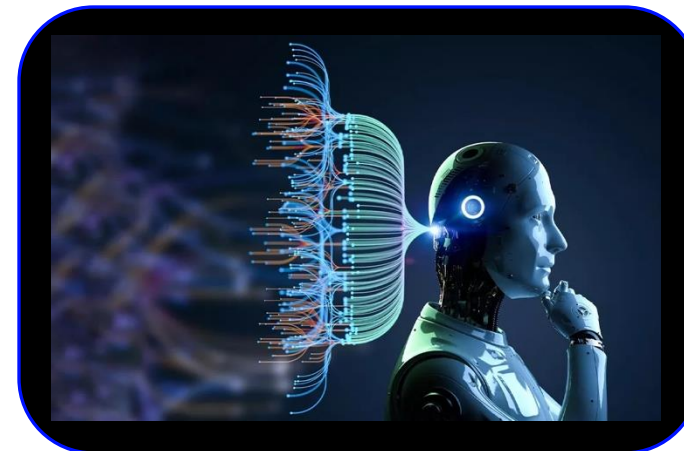
Prob, CS109, 2021

Stanford University 184

Analytics



AI Models



How Does This Work?

Part 1: Core Probability

- Probability
- Equally Likely Outcomes
- Axioms of Probability
- Probability of or
- Conditional Probability
- Independence
- Probability of and
- De Morgan's Law
- Law of Total Probability
- Bayes' Theorem
- Log Probabilities
- Many Coin Flips
- Stories
- Bacteria Evolution
- Google Rain Prediction
- Random Walks
- Binomial with Different Probs
- Netflix Genres
- Poker**
- Core Probability Practice

Part 2: Random Variables

- Random Variables
- Probability Mass Functions
- Expectation
- Variance
- Bernoulli Distribution
- Binomial Distribution
- Bayesian Distribution

Poker Game

P(Win): 82.8%
P(Win | One Opponent): 96.0%
E[EarningsPerDollar]: 3.97

Probability

Events: deal, flop, turn, river

Continue

Isabel 100
Fabian 100
Jade 100
Emir 100
Nisha 100
You 100

Community Cards: J♠, K♣, 4♦, 10♣

Hand: Straight, K High (Q♠, 9♥)

Code In place Runner

Today

M3ei51 8:11 AM ✓

Running **Poker Probability Quiz**, a program created by Chris Piech...

Type **quit** to stop the program early. 8:11 AM

> This program generates random Texas Holdem situations and has you guess your probability of winning. 8:11 AM

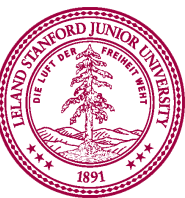
> Guess the probability! 8:11 AM

> 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 ♠ 8:11 AM

> What is the probability that you win? 8:11 AM

0.6

WhatsApp m3ei51 to +1-415-728-3856

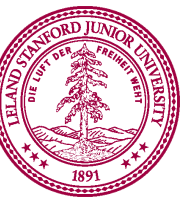


Axioms of Probability

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

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

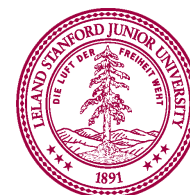
Since everything must either be in E or E^c

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

Axiom 2

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

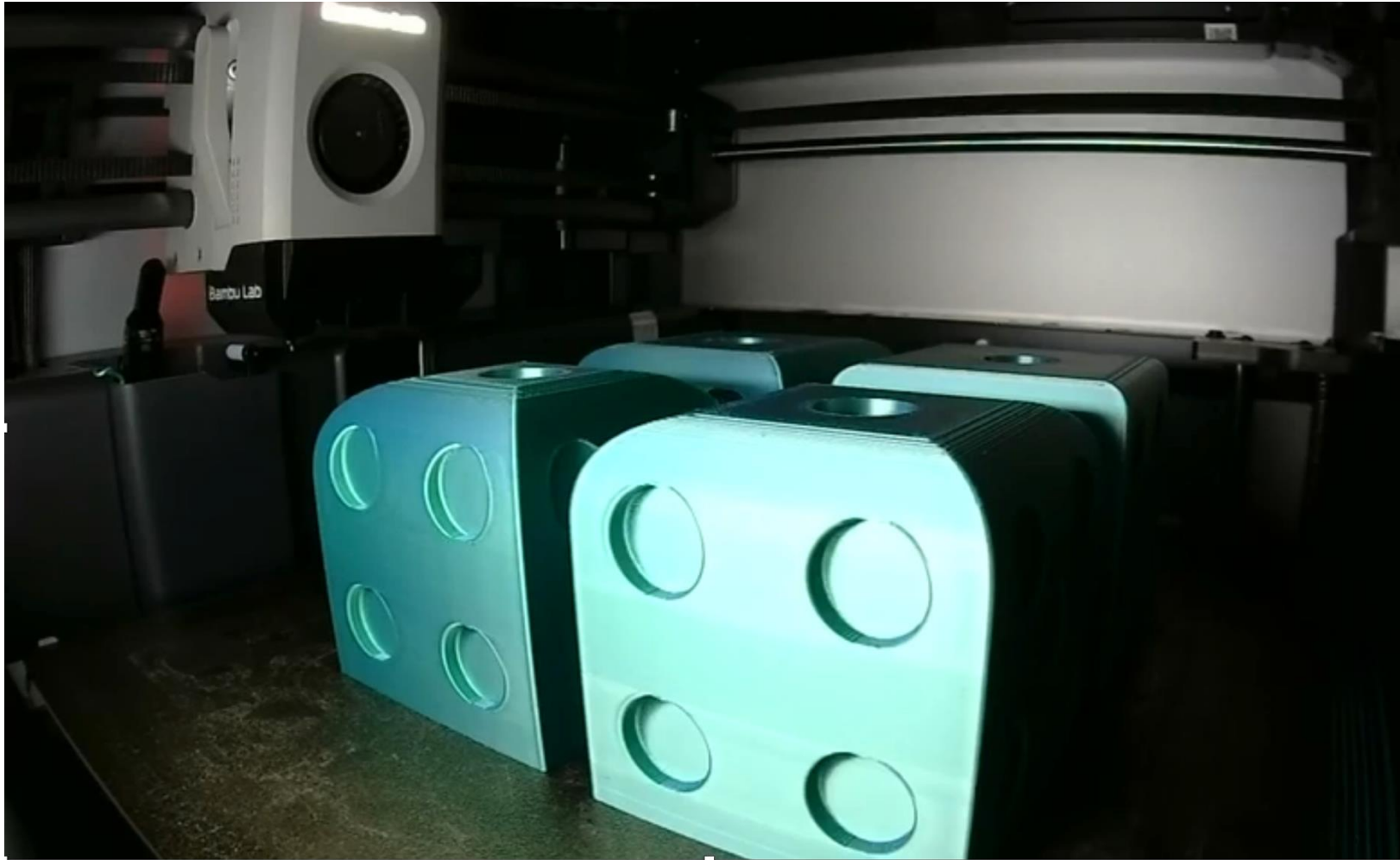
Rearrange



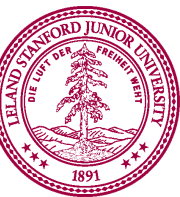


Mysteries

Sum of Two of These Die = 7?



I printed this last week



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}



End Review

Announcements

Problem Set #1 is out

newish

PS1 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: 0.5 Check Answer

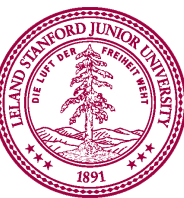
Explanation:

Block LaTeX Inline LaTeX Python Image

Auto Submission

Check your answer

Insert LaTeX




Write an Agent **newish**

The screenshot shows a web browser window with the URL `cs109psets.netlify.app/fall25/pset1/countingcards`. The page title is "Counting Cards" under the heading "PS1".

The main content area contains the following text:

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).

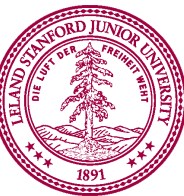
If you were given a truly random card out of the deck of 52, your chance of

Navigation: Previous Question, Next Question

The right side of the browser shows an "Answer Editor" with a "Solution" tab. The code is as follows:

```
Agent:
1 """
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.
6 """
7
8 def counting_agent(cards_played):
9     print(cards_played)
10    # default strategy: always play
11    return 'play'
```

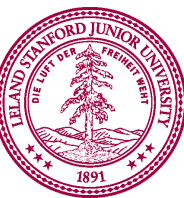
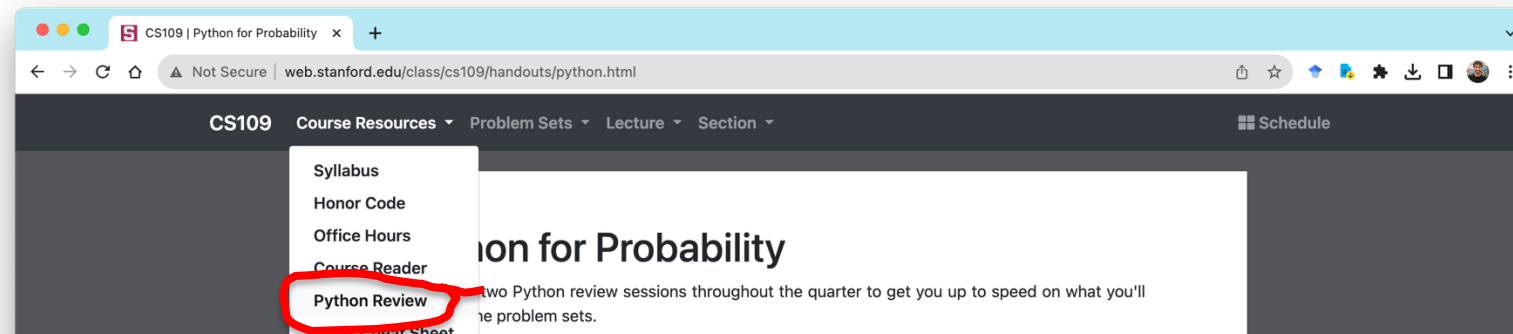
Buttons: Run One Game, Test Agent



Python Review Session

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

Find links, recordings, and setup here

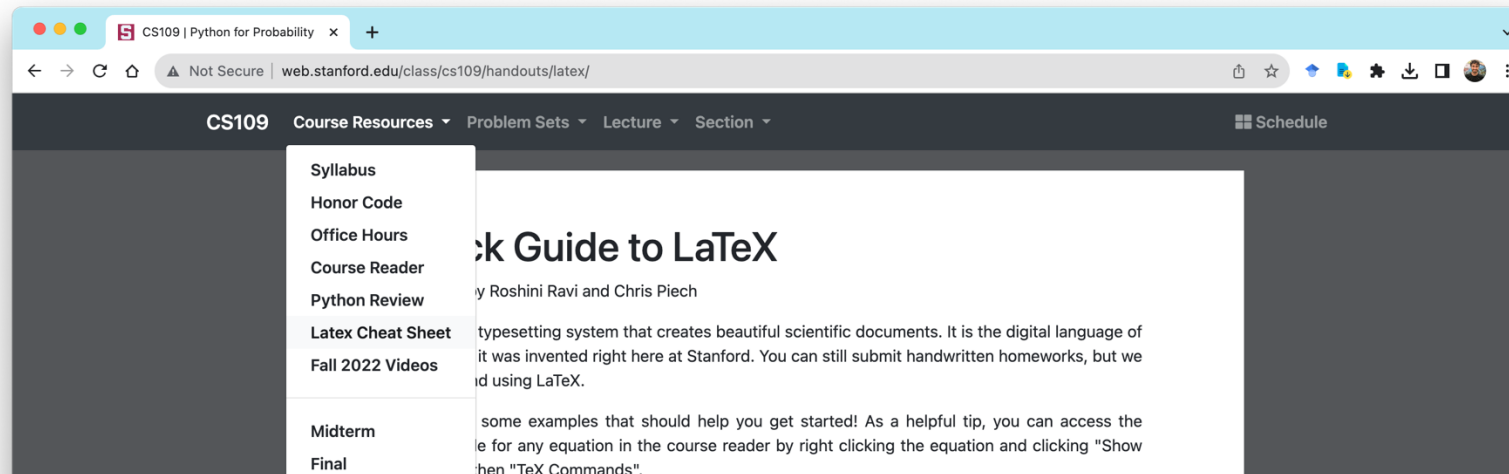


Learn LaTeX

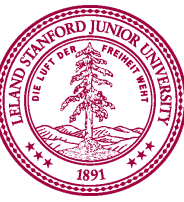
```
1 \begin{aligned}
2 P(E) |
3 &= \sum_{i=0}^n e^i \\
4 &= 0.25
5 \end{aligned}
```

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

Done



The screenshot shows a web browser window with the URL `web.stanford.edu/class/cs109/handouts/latex/`. The page title is "CS109 | Python for Probability". A navigation menu is visible with items: "Course Resources", "Problem Sets", "Lecture", "Section", and "Schedule". A dropdown menu is open, listing: "Syllabus", "Honor Code", "Office Hours", "Course Reader", "Python Review", "Latex Cheat Sheet", and "Fall 2022 Videos". The main content area displays the heading "Quick Guide to LaTeX" and the text: "LaTeX is a typesetting system that creates beautiful scientific documents. It is the digital language of mathematics. It was invented right here at Stanford. You can still submit handwritten homeworks, but we encourage you to use LaTeX." Below this, there is a section for "Midterm" and "Final" with some introductory text.



Inline LaTeX

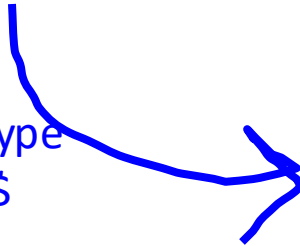
new

Explanation:

☰ Block LaTeX Inline LaTeX  Python  Image

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

When you type
the closing \$

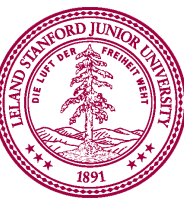


Explanation:

☰ Block LaTeX Inline LaTeX  Python  Image

This is an example of inline latex. Let Y

The screenshot shows a web browser window with the URL `web.stanford.edu/class/cs109/handouts/latex/`. The page title is "CS109" and the main heading is "Block Guide to LaTeX". A dropdown menu is open, listing various course resources: Syllabus, Honor Code, Office Hours, Course Reader, Python Review, Latex Cheat Sheet, Fall 2022 Videos, Midterm, and Final. The text "Block Guide to LaTeX" is partially visible behind the menu.



What Makes for a Good Answer

The screenshot shows a web browser window with the URL `cs109psets.netlify.app/fall25/lecture1/dice_probability`. The page title is "Dice Probability".

Problem Statement:
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.

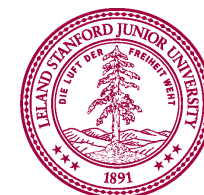
Diagram: Two dice are shown: an orange die on the left and a blue die on the right. Blue arrows point from the text "Value dice 1" and "Value dice 2" to the respective dice. A large blue bracket encompasses both dice.

Solution Editor:
The "Solution" tab is active. The "Numeric Answer" field contains "0.5". A "Check Answer" button is visible.
The "Explanation" section contains the following text:
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:
$$P(E^C) = 1 - P(E)$$
$$= 1 - \frac{5}{36}$$

A code editor shows the following Python code:

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

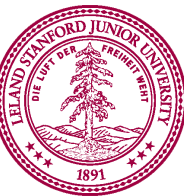
Buttons for "Run" and ">_ Show" are present.
The final output of the code is "The answer is: 0.861111111...".



If you notice a bug?

It should be robust, but things can happen.

Let me know: send an email to cpiech@stanford.edu or message me on slack. I need your email and the approximate time you encountered the bug.



Honor Code

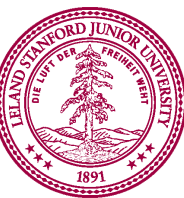
Always remember: You need to be able to recreate your ability on an exam. And in the real world. This is a foundation course.

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:

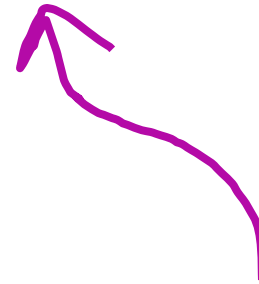
1. Do not give away the answer
2. Always be respectful
3. Know what you don't know



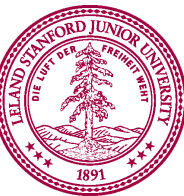
Midterm Moved by One Day!

Before: ~~Oct 28, 7pm~~

Now: **Oct 27th, 7pm**

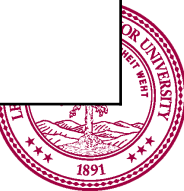


If you have an academic conflict,
we will offer you an alternate time



Section Signups Open Tomorrow

Sunday	Monday	Tuesday	Wed	Thursday	Friday	Saturday
			Sept 24 You are here	Sept 25 Section signup opens	Sept 26	Sept 27
Sept 28 Section signups close at 5pm	Sept 29	Sept 30 Sections announced	Oct 1	Oct 2, Oct 3 First section!		Oct 4



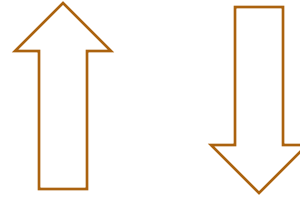
End Announcements

Learning Goal for Today: Conditional Probability



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

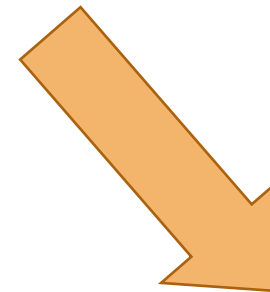
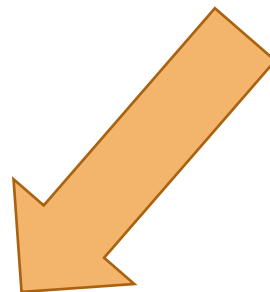
Chain rule



Definition of conditional probability

$$P(E|F)$$

Law of Total Probability



Bayes' Theorem

$$P(E)$$

$$P(F|E)$$

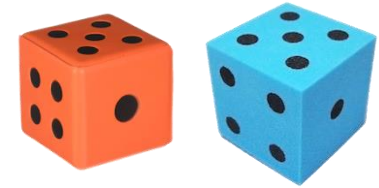


Conditional Probability

Roll two dice

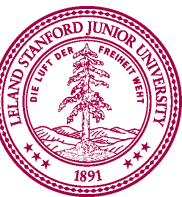
$$P(E) = \frac{|E|}{|S|} \text{ Equally likely outcomes}$$

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



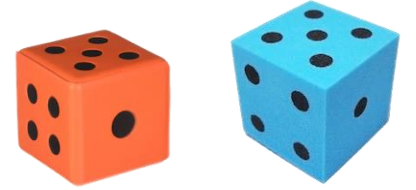
$S = \{(1,1) (1,2) (1,3) (1,4) (1,5) \mathbf{(1,6)}$
 $(2,1) (2,2) (2,3) (2,4) \mathbf{(2,5)} (2,6)$
 $(3,1) (3,2) (3,3) \mathbf{(3,4)} (3,5) (3,6)$
 $(4,1) (4,2) \mathbf{(4,3)} (4,4) (4,5) (4,6)$
 $(5,1) \mathbf{(5,2)} (5,3) (5,4) (5,5) (5,6)$
 $\mathbf{(6,1)} (6,2) (6,3) (6,4) (6,5) (6,6) \}$

$E =$ *In blue*



Dice, our misunderstood friends

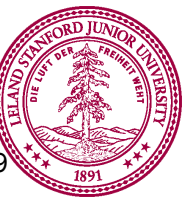
Roll two 6-sided dice, yielding values D_1 and D_2 .
You want them to sum to 4.



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

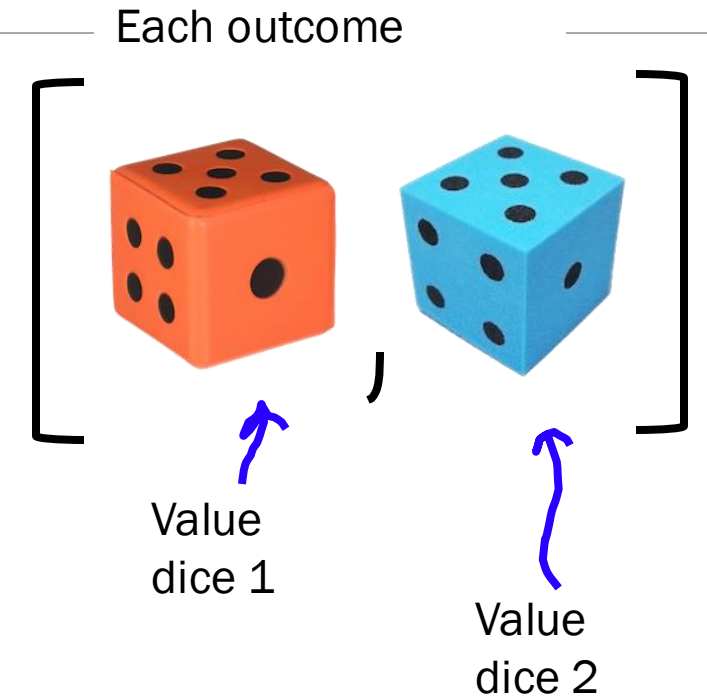


Sum of Two Die = 4?

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

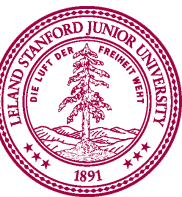
Let E be the event that the sum is 4

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] }



E = *In red*

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

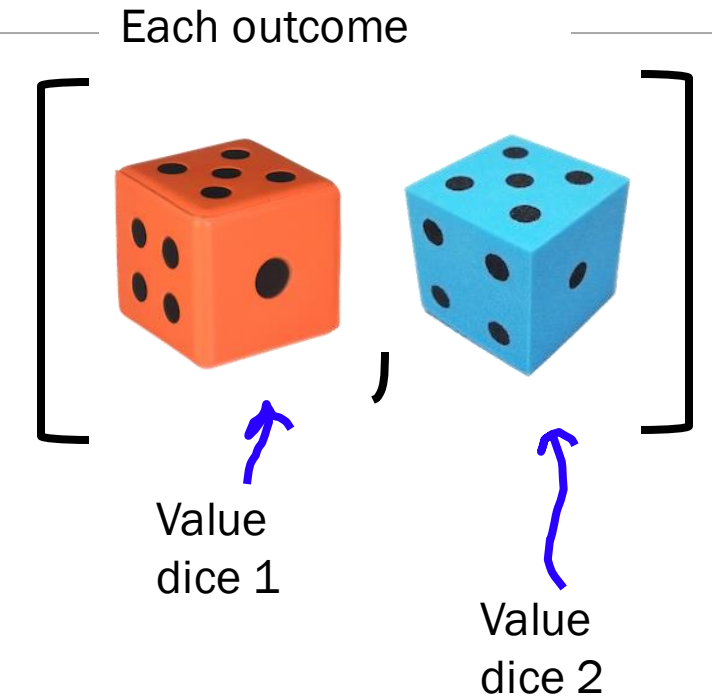


Sum of Two Die = 4? Condition on F: $D_1 = 2$

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

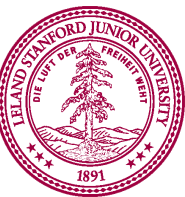
Let E be the event that the sum is 4

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] }



E = *In red*

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

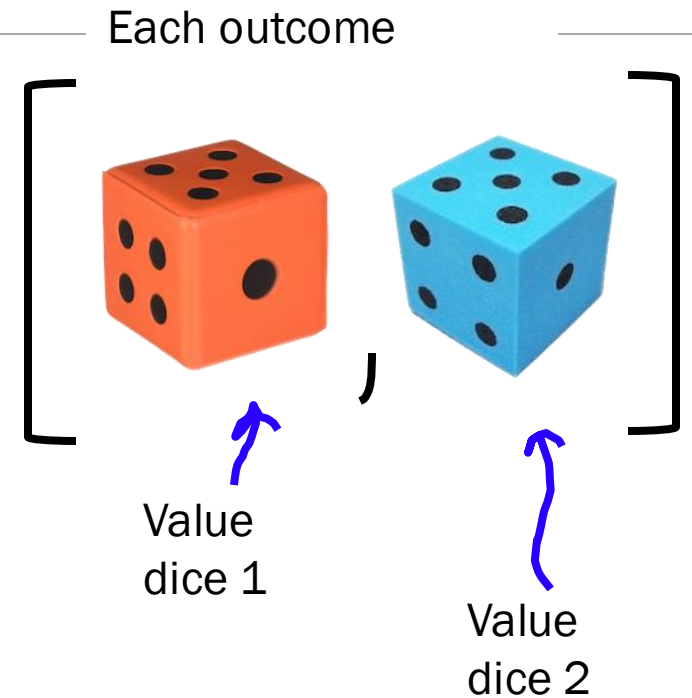


Sum of Two Die = 4? Condition on F: $D_1 = 2$

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]
	}					



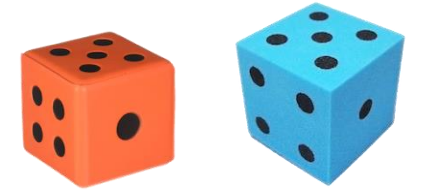
E = *In red*

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



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

$$|S| = 36$$

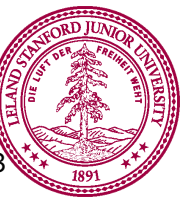
$$E = \{(1,3), (2,2), (3,1)\}$$

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

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

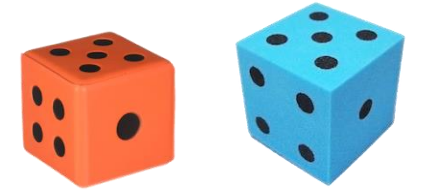
$$E = \{(2,2)\}$$

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



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

$$|S| = 36$$

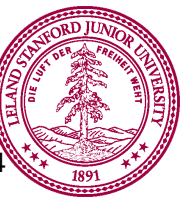
$$E = \{(1,3), (2,2), (3,1)\}$$

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

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

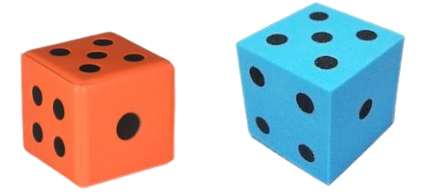
$$E = \{(3,1)\}$$

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



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

$$|S| = 36$$

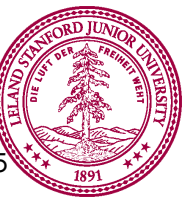
$$E = \{(1,3), (2,2), (3,1)\}$$

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

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

$$E = \{ \}$$

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



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:

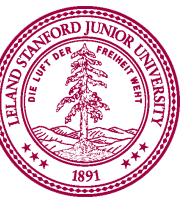
“ $P(E, \text{ given } F \text{ already observed})$ ”

Sample space \rightarrow

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

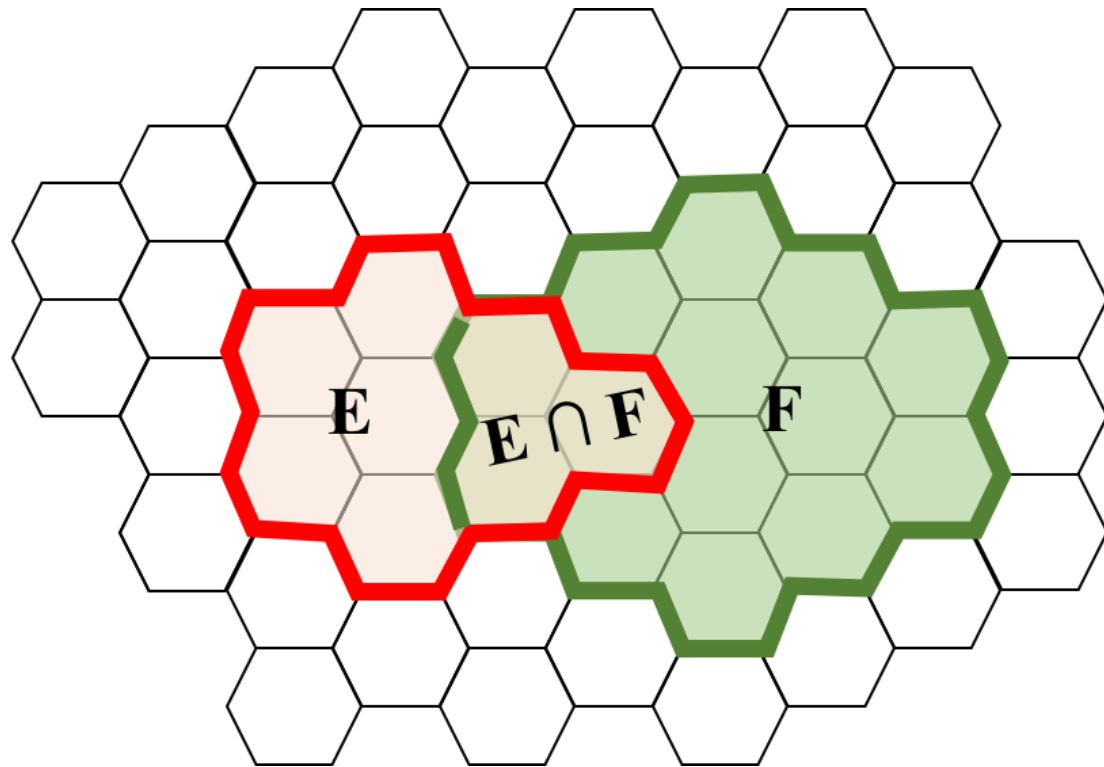
Event \rightarrow

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 .

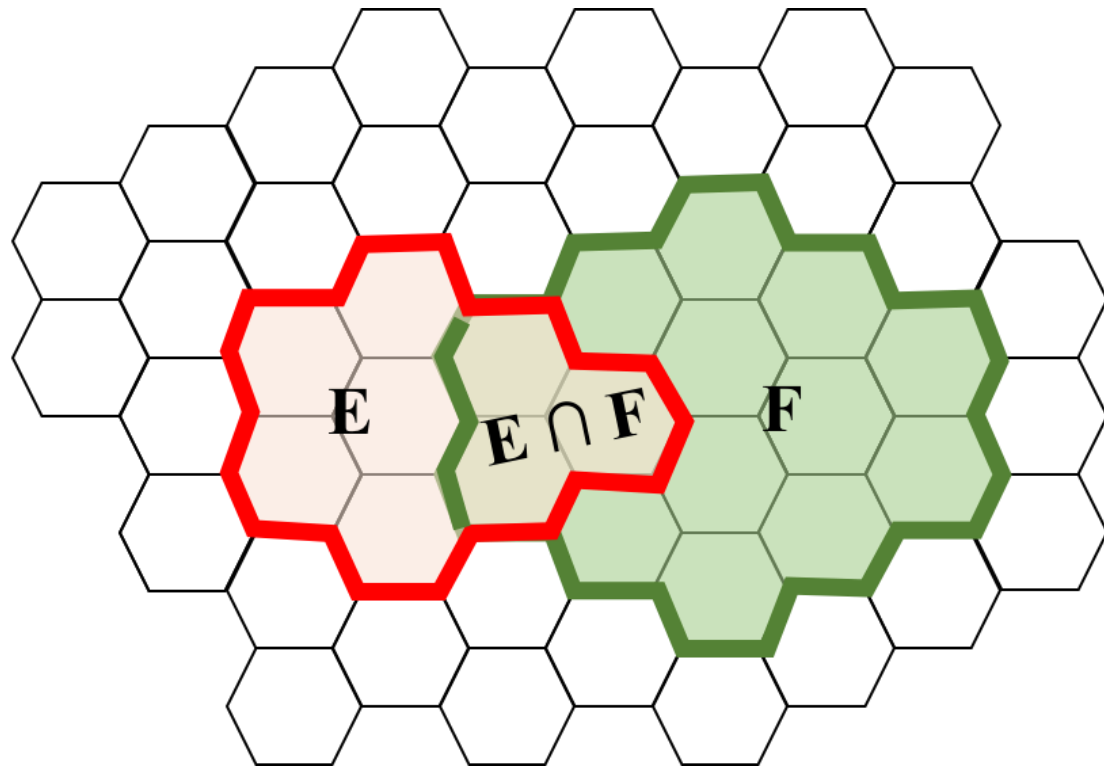


$$P(E) = \frac{8}{50} \approx 0.16$$

$$P(E|F) =$$

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 .

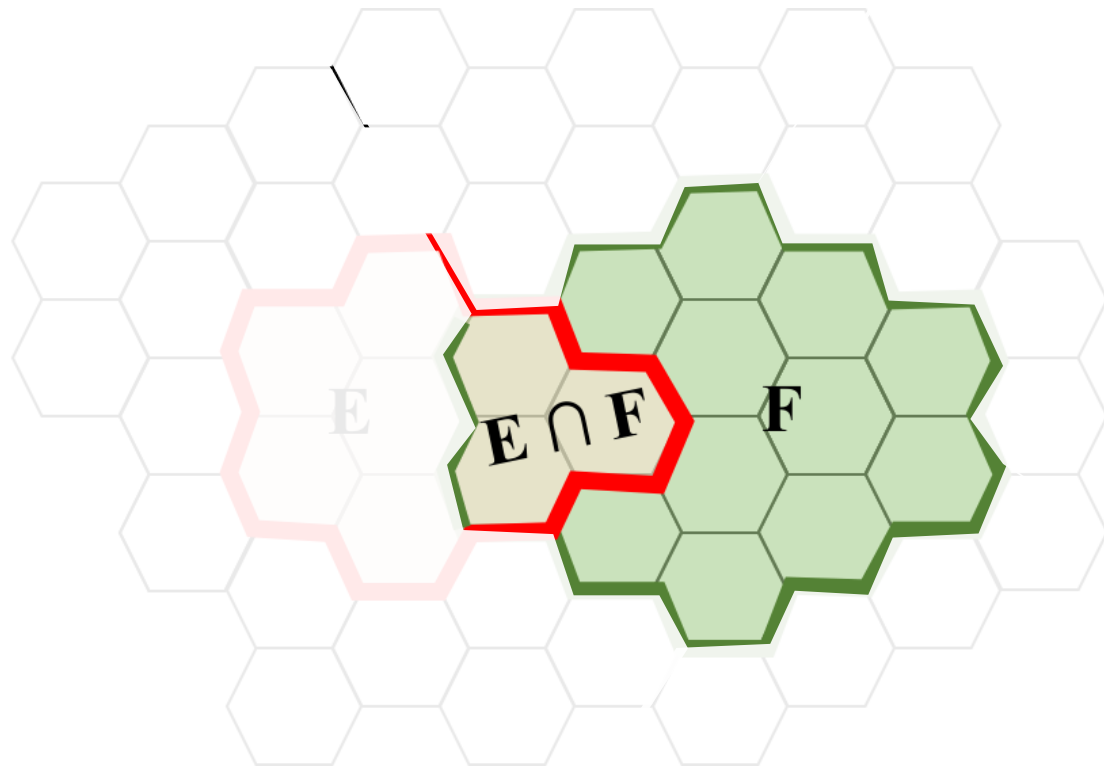


$$P(E) = \frac{8}{50} \approx 0.16$$

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

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 .



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

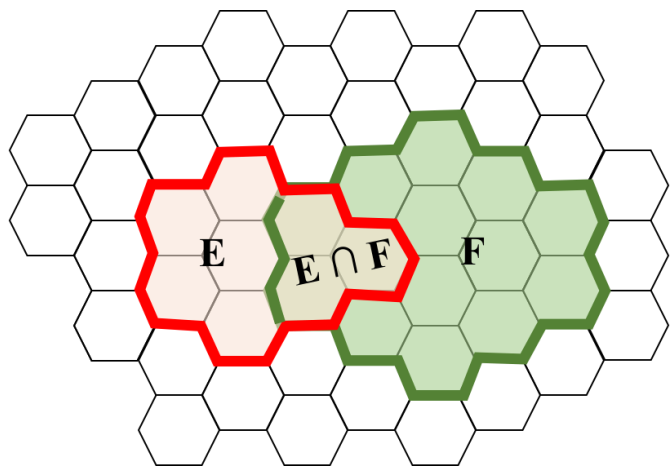
$$P(E|F) = \frac{3}{14} \approx 0.21$$

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”)

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



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

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

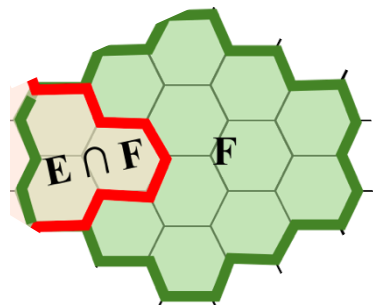


Conditional Probability, visual intuition

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

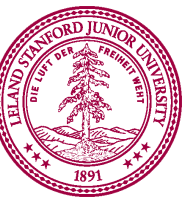
What is the probability
that a user will watch
Life is Beautiful?

$$P(E)$$

$$S = \{\text{Watch, Not Watch}\}$$

$$E = \{\text{Watch}\}$$

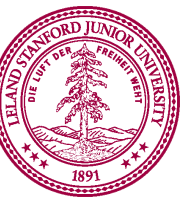
$$P(E) = 1/2 ?$$



Netflix and Learn

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

$$P(E)$$



Netflix and Learn

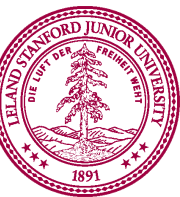
What is the probability
that a user will watch
Life is Beautiful?

$$P(E)$$



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

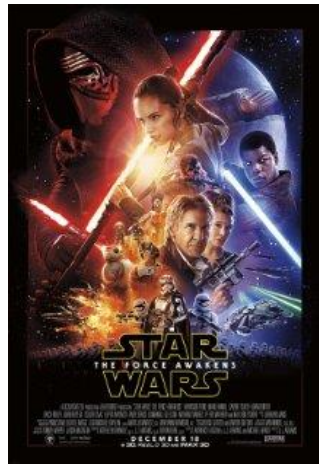
$$P(E) = 10,234,231 / 50,923,123 = 0.20$$



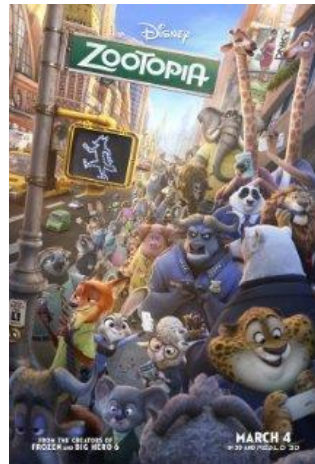
Netflix and Learn

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

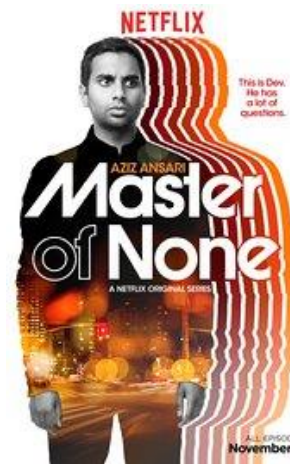
Let E be the event that a user watches the given movie.



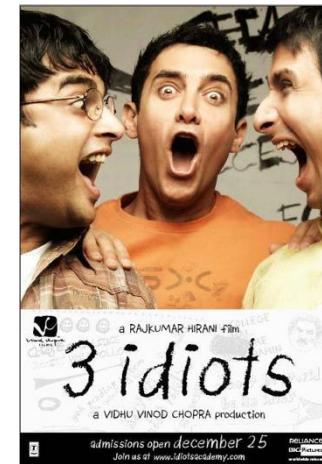
$$P(E) = 0.19$$



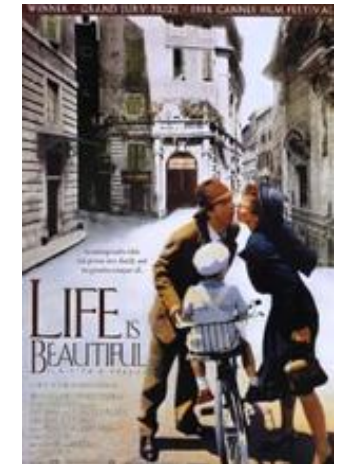
$$P(E) = 0.32$$



$$P(E) = 0.20$$



$$P(E) = 0.09$$



$$P(E) = 0.20$$

Netflix and Learn

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

Let E = a user watches Life is Beautiful.

Let F = a user watches CODA.

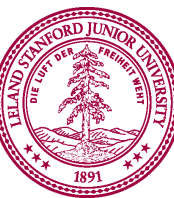
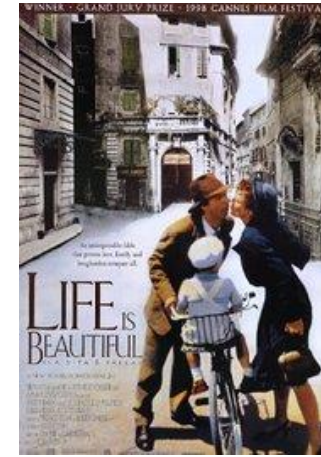
What is the probability that a user watches Life is Beautiful, given they watched CODA?

$$P(E|F)$$

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

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

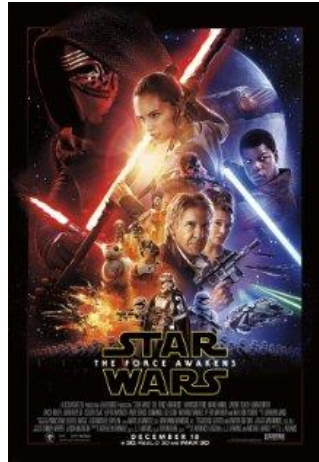
$$\approx 0.42$$



Netflix and Learn

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

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



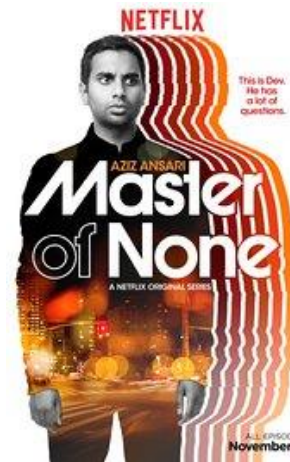
$$P(E) = 0.19$$

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



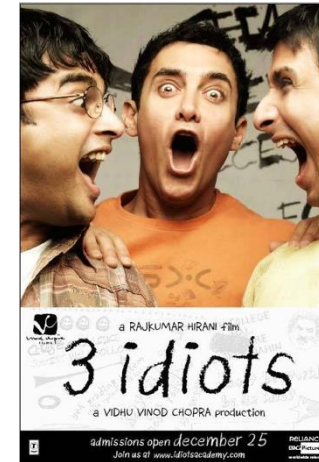
$$P(E) = 0.32$$

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



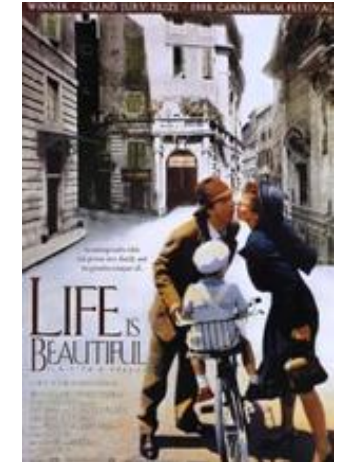
$$P(E) = 0.20$$

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



$$P(E) = 0.09$$

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

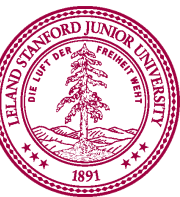


$$P(E) = 0.20$$

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

Machine Learning

Machine Learning is:
Probability + Data + Computers



Notation

And

Or

Given

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

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

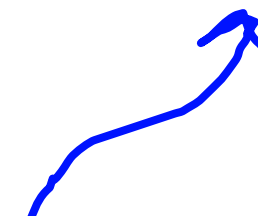
$$P(E|F)$$

$$P(E, F)$$

$$P(E \cup F)$$

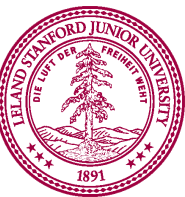
$$P(E|F, G)$$

$$P(EF)$$



$$P(E \cap F)$$

Probability of E given
F and G



Chain Rule via Baby Poop

<https://cs109psets.netlify.app/fall23/lecture4/poop>

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

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

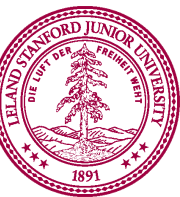
In the morning when she wakes up, a baby has a 50% chance of having pooped. The chance that a baby cries **given** that she has pooped is 50%. What is the probability that a baby **has pooped, and cries**.

A screenshot of a web browser displaying the CS109 course website. The browser's address bar shows the URL: web.stanford.edu/class/cs109/lectures/4-ConditioningAndBayes/. The website header includes "CS109" and navigation links for "Course Resources", "Problem Sets", "Lecture", and "Section". A dropdown menu is open under "Lecture", listing: 1. Welcome, 2. Combinatorics, 3. Probability, and 4. Conditioning and Bayes. The main content area displays "Lecture 4: Conditioning and Bayes" with the date "OCT 4TH, 2023" and location "HEWLETT 200, 3:30P". Below this, the "Lecture Materials" section is visible, featuring a document icon and a blue rocket icon circled in purple.

Generalized Chain Rule

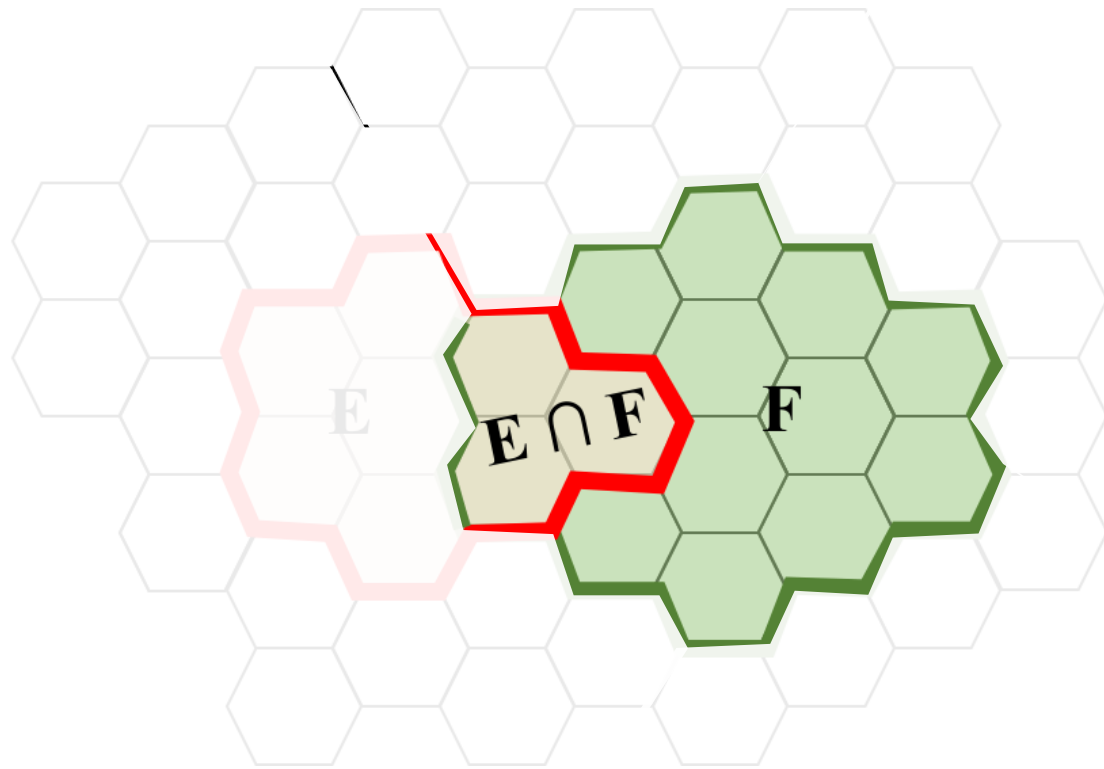
$$\Pr(E_1 \text{ and } E_2 \text{ and } E_3 \text{ and } \dots E_n)$$

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



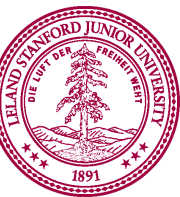
Conditional Paradigm

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



For example:

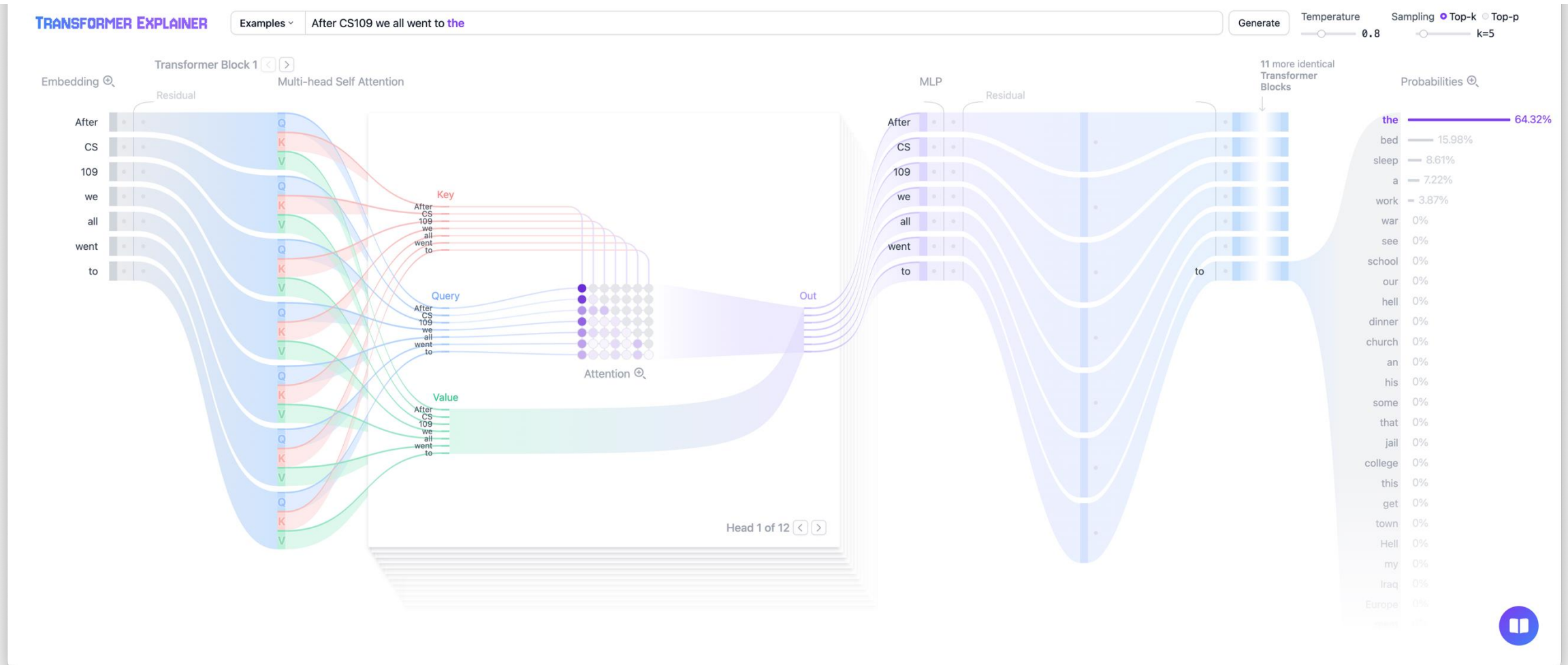
$$P(E^C | F) = 1 - P(E | F)$$



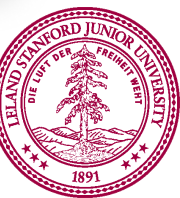


and Learn

Under the hood of a Large Language Model



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



Under the hood of a Large Language Model

TRANSFORMER EXPLAINER

Examples ▾

After CS109 we all went to the

Generate

Temperature 0.8

Sampling Top-k Top-p
k=5

Embedding 🔍

Transformer Block 1 ◀ ▶

11 more identical
Transformer

Probabilities 🔍

After
CS
109
we
all
went
to



the	64.32%
bed	15.98%
sleep	8.61%
a	7.22%
work	3.87%
war	0%
see	0%
school	0%
our	0%
hell	0%
dinner	0%
church	0%
an	0%
his	0%
some	0%
that	0%
jail	0%
college	0%
this	0%
get	0%
town	0%
Hell	0%
my	0%
Iraq	0%
Europe	0%
...	0%

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

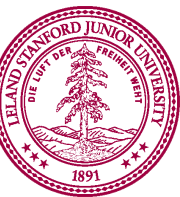


LLM is a Conditional Probability Machine

After CS109 we all went to



- 0.29 the library.
- 0.42 CoHo.
- 0.1 sleep.
- 0.19 the Oval.



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



0.29

the library.

0.42

CoHo.

0.1

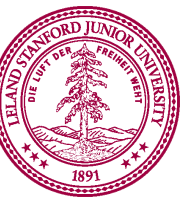
sleep.

0.19

the Oval.

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

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



LLM is a Conditional Probability Machine. Some uses

Prompt optimization:

What prompt makes expected responses look as likely as possible?

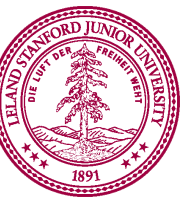
Understanding summaries:

What is the mutual information between a summary and future productions?

Classification tasks:

Combine with Bayes' to infer unobserved.

Many more creative applications...

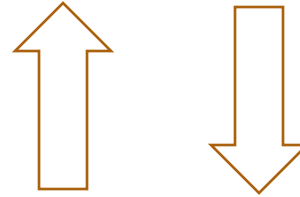


Law of Total Probability

Relationship Between Probabilities

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

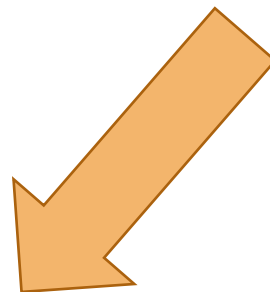
Chain rule



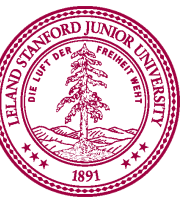
Definition of
conditional probability

$$P(E|F)$$

Law of Total
Probability



$$P(E)$$



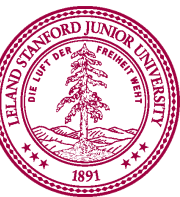
Baby Poop Redux

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%.



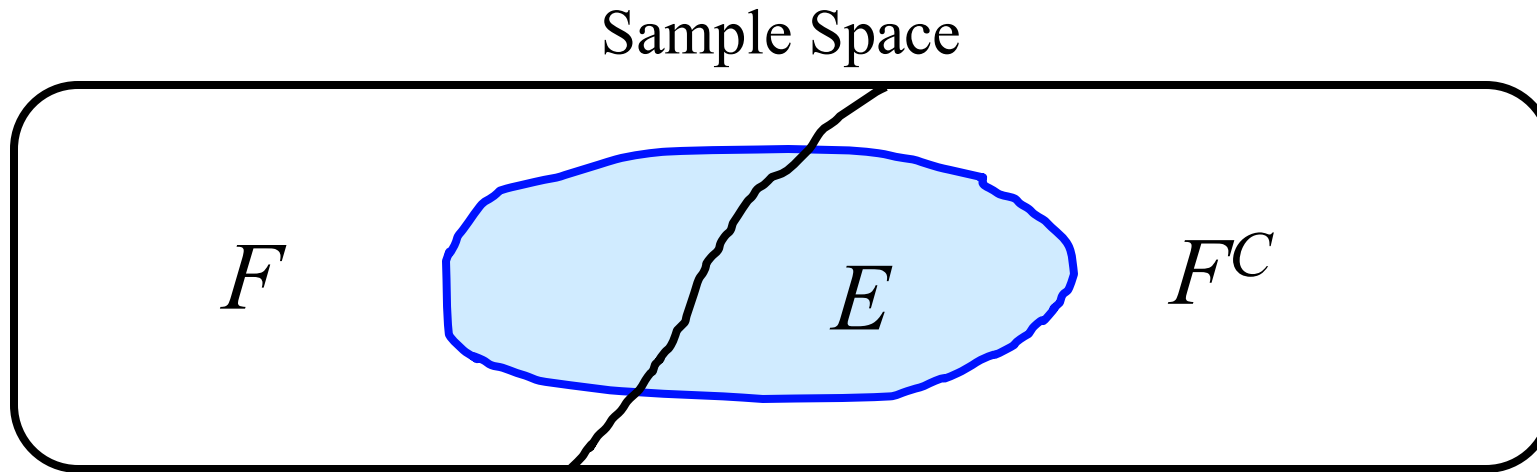
What is the probability of crying, unconditioned?

What information do you need?

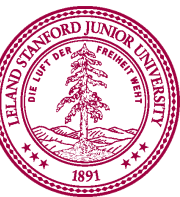


Law of Total Probability

Say E and F are events in S

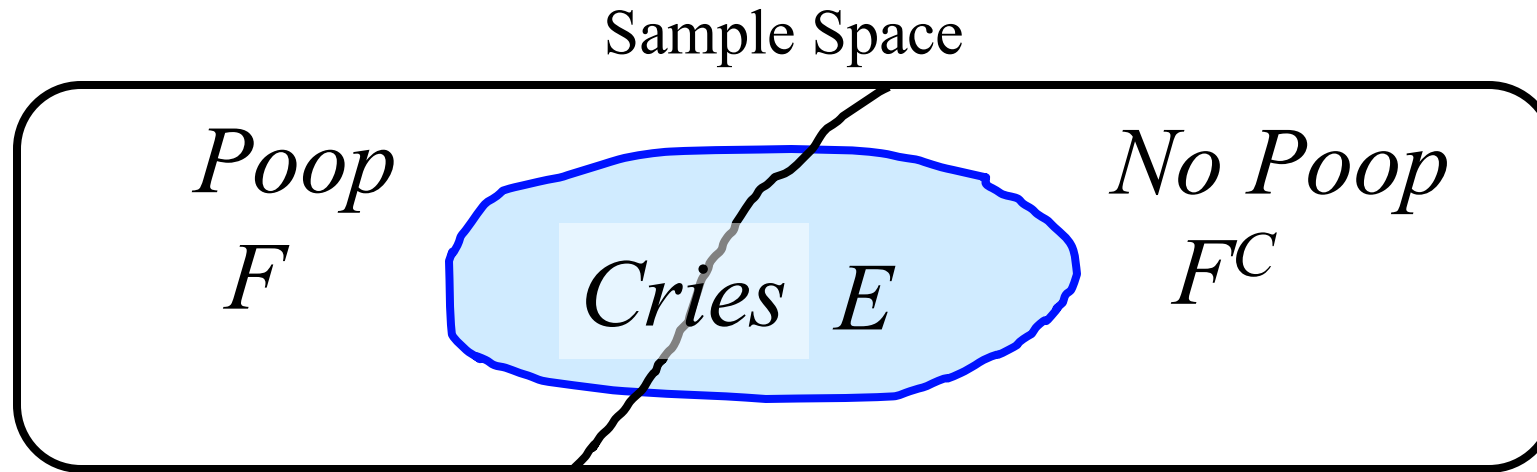


$$\begin{aligned} P(E) &= P(EF) + P(EF^C) \\ &= P(E|F)P(F) + P(E|F^C)P(F^C) \end{aligned}$$

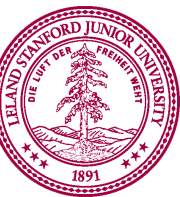


Law of Total Probability

Say E and F are events in S

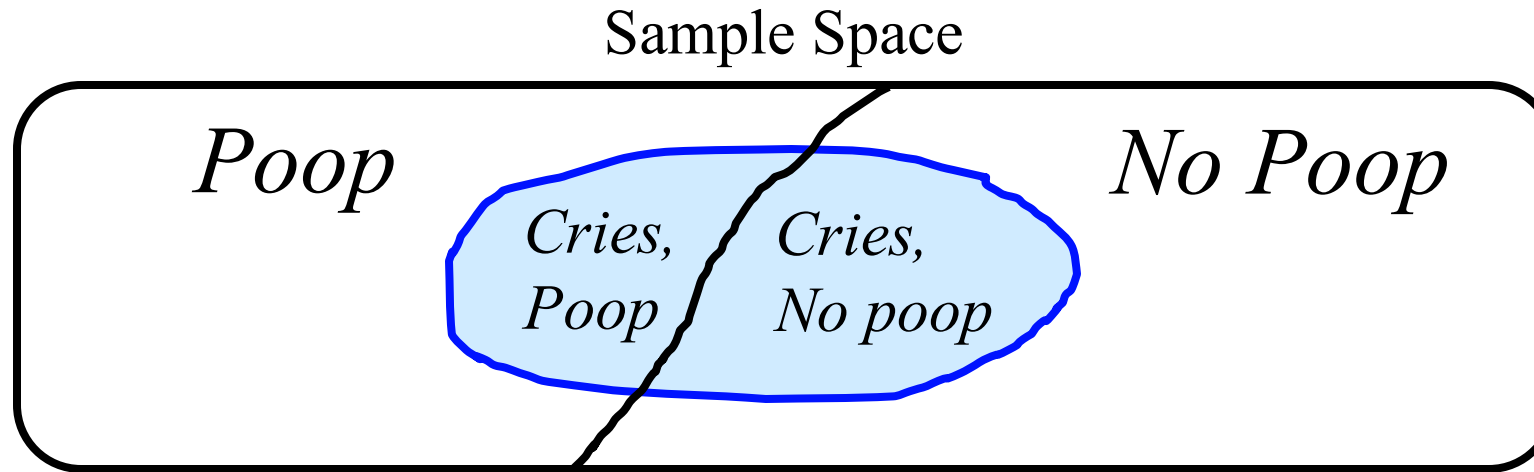


$$\begin{aligned} P(E) &= P(EF) + P(EF^C) \\ &= P(E|F)P(F) + P(E|F^C)P(F^C) \end{aligned}$$

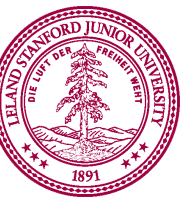


Law of Total Probability

Say E and F are events in S

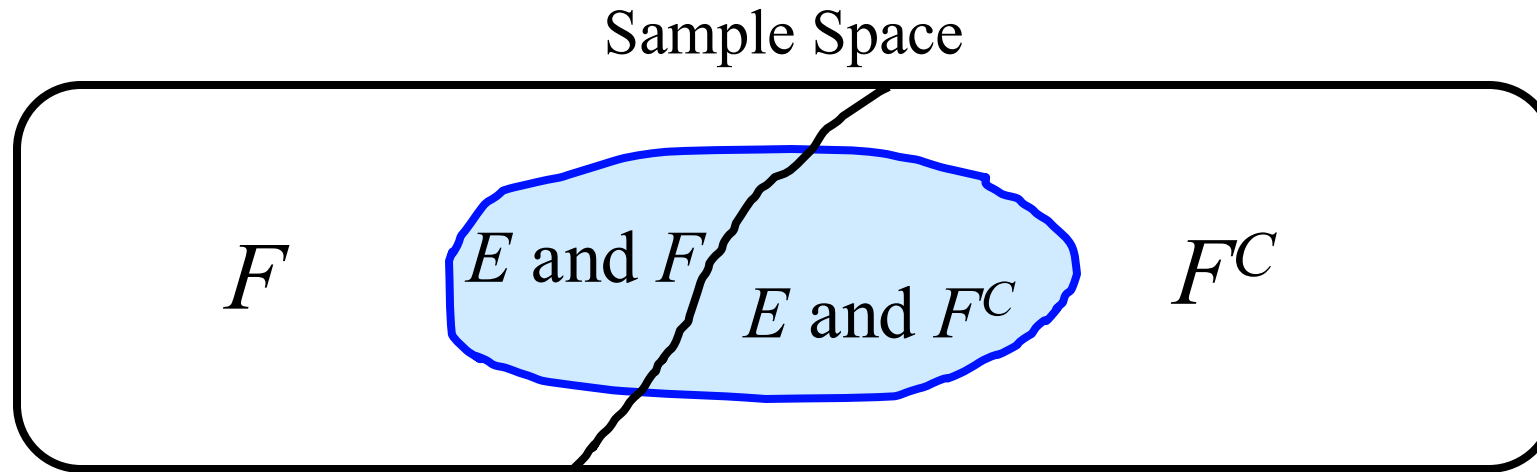


$$\begin{aligned} P(E) &= P(EF) + P(EF^C) \\ &= P(E|F)P(F) + P(E|F^C)P(F^C) \end{aligned}$$

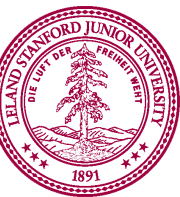


Law of Total Probability

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 ,

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

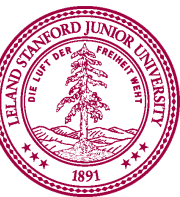
Proof

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%.



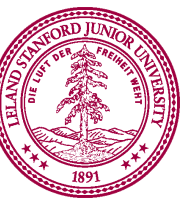
Probability of crying (T)?

What information do you need?

Probability of crying given no poop.

Recall that T is crying and E is poop

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



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 ~~By~~ Law of Total Probability (LOTP):

$$\Pr(E) = \Pr(E \text{ and } M) + \Pr(E \text{ and } M^C)$$

$$= \Pr(E|M)\Pr(M) + \Pr(E|M^C)\Pr(M^C)$$

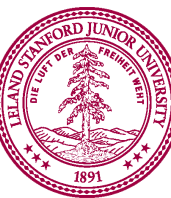
$$= 0.20 \cdot 0.10 + 0.01 \cdot 0.90$$

$$= 0.029$$

LOTP

Chain Rule

Substituting



Know:

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

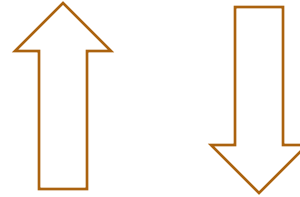
Real question. What is the probability of poop
given the baby cries...

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

Relationship Between Probabilities

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

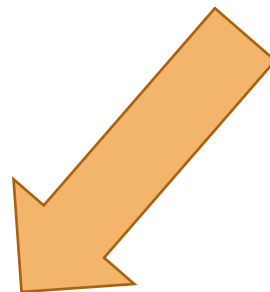
Chain rule
(Product rule)



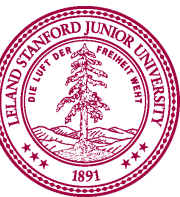
Definition of
conditional probability

$$P(E|F)$$

Law of Total
Probability



$$P(E)$$

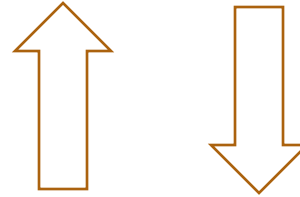


Relationship Between Probabilities



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

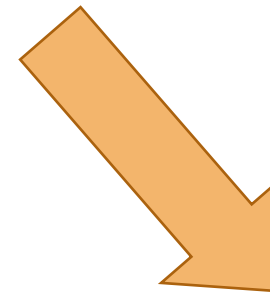
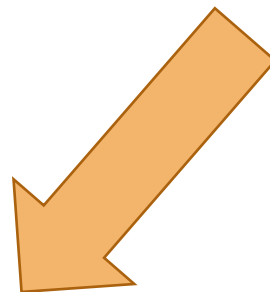
Chain rule
(Product rule)



Definition of
conditional probability

$$P(E|F)$$

Law of Total
Probability



Bayes'
Theorem

$$P(E)$$

$$P(F|E)$$



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

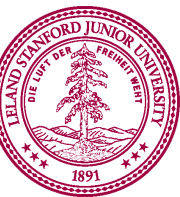
$$P(F | E)$$



I want to calculate
 $P(\text{State of the world } F | \text{Observation } E)$
It seems so tricky!...

The other way around is easy
 $P(\text{Observation } E | \text{State of the world } F)$
What options to I have, chief?

$$P(E | F)$$



Thomas Bayes

Want $P(F|E)$. Know $P(E|F)$

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



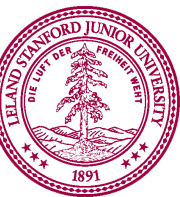
A little while later...

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

A little while later...

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

LOTP



(silent drumroll)



Bayes' Theorem

$$P(E|F) \Rightarrow P(F|E)$$

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

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

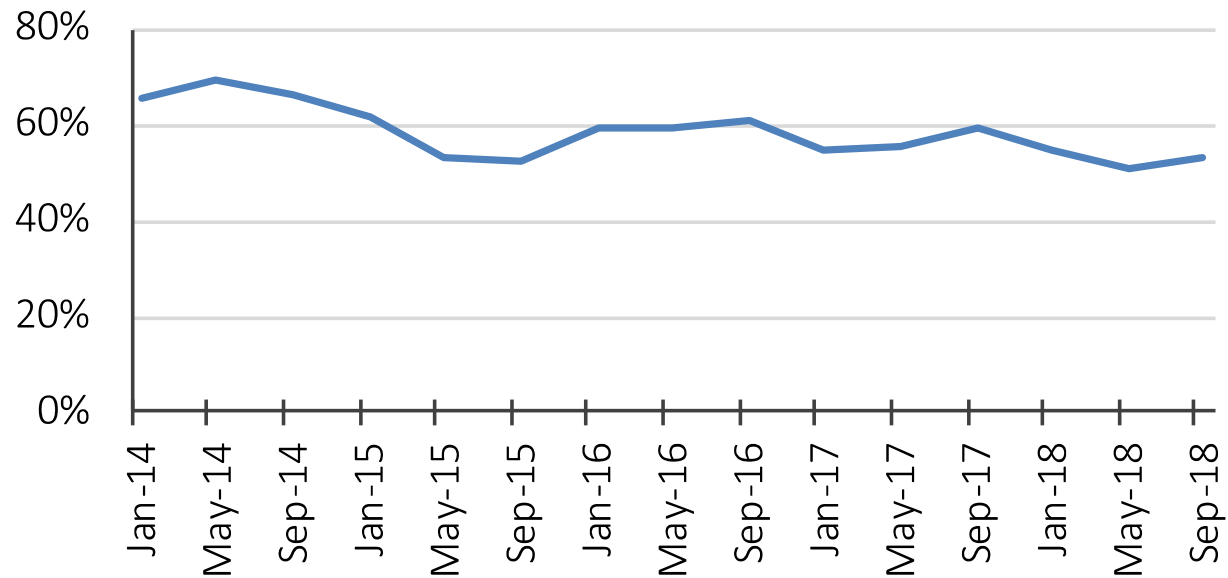
Expanded form:

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



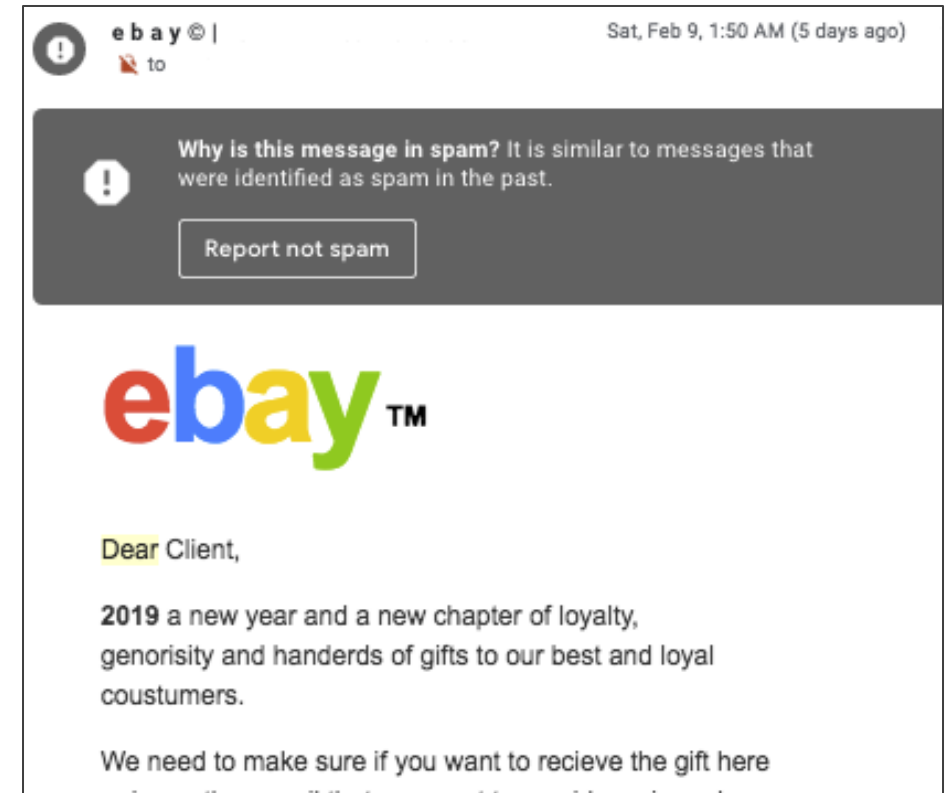
Detecting spam email

Spam volume as percentage of total email traffic worldwide



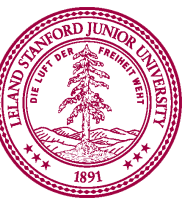
We can easily calculate how many spam emails contain “Dear”:

$$P(E|F) = P\left(\text{“Dear”} \mid \begin{array}{l} \text{Spam} \\ \text{email} \end{array}\right)$$



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

$$P(F|E) = P\left(\begin{array}{l} \text{Spam} \\ \text{email} \end{array} \mid \text{“Dear”}\right)$$



Detecting spam email

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

- 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?

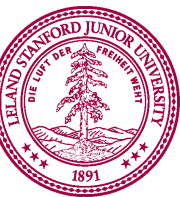
1. Define events
& state goal

2. Identify known
probabilities

3. Solve

Let: E : “Dear”, F : spam

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



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”

$P(F)$ ✓

$P(E|F)$ ✓

$P(E|F^C)$ ✓

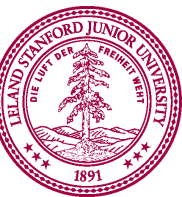
You get an email with the word “Dear” in it.

What is the probability that the email is spam?

Want: $P(F|E)$

$$\text{posterior } P(F|E) = \frac{\text{likelihood } P(E|F) \text{ prior } P(F)}{P(E)}$$

normalization constant



Spam Revisited



How would you detect Spam using an LLM?

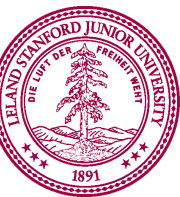
Real spam email:

“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.

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

60% 60% 40%

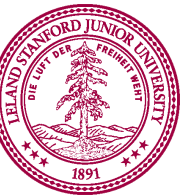


How would you detect Spam using an LLM?

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

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

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.

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

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

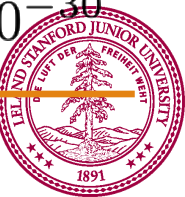
```
from llm import string_pr
pr = string_pr("Hello, world")
print(pr)
```

If you used a prompt like this:

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

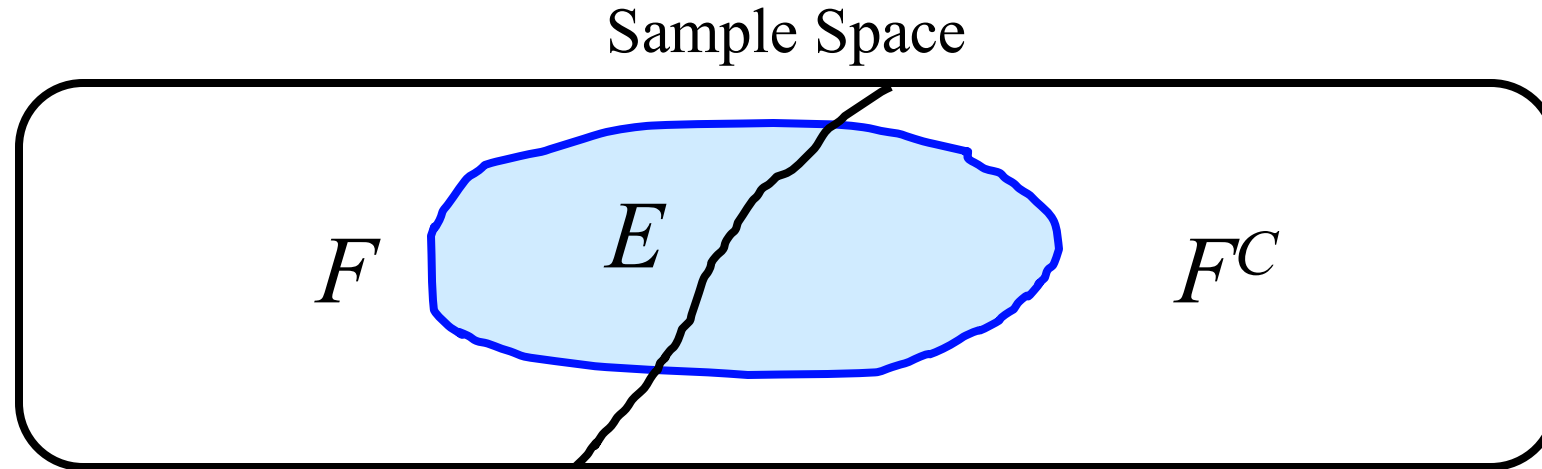
Could you test an LLMs understanding of probability?

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

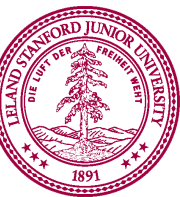


Come on Friday!
Stories + Make History

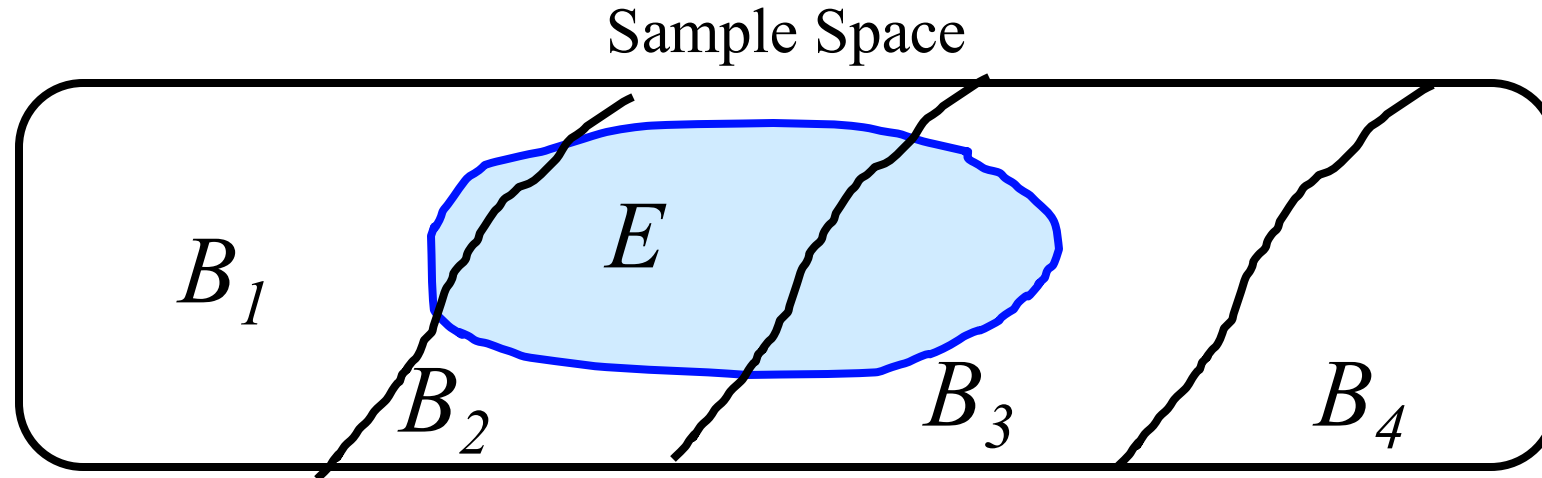
Law of Total Probability



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

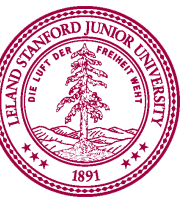


Law of Total Probability



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: 74%
Wind: 8 mph

Weather

Monday

Rain showers

Background event.
Where is the person in
San Francisco?

Results for San Francisco, CA



49 °F | °C

Precipitation: 90%

Humidity: 74%

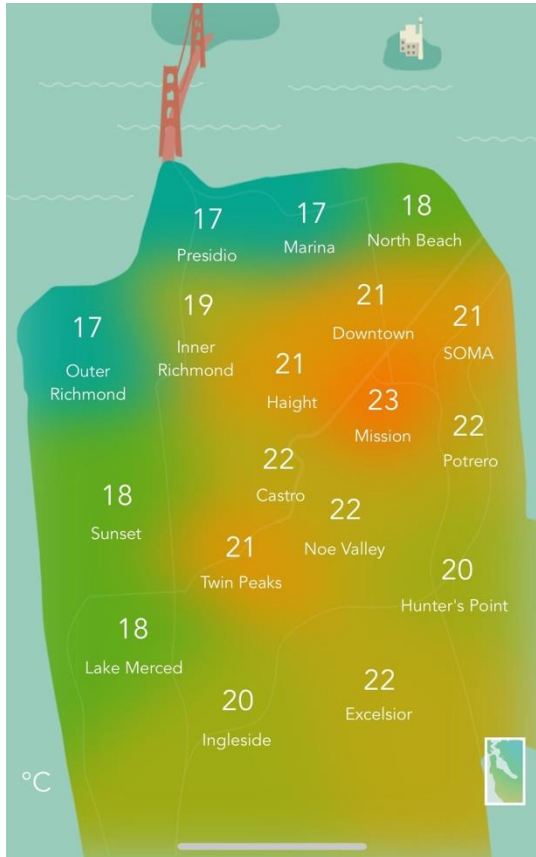
Wind: 8 mph

Weather

Monday

Rain showers

Background event.
Where is the person in
San Francisco?



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:

$$P(R)$$

Results for San Francisco, CA



49 °F | °C

Precipitation: 90%

Humidity: 74%

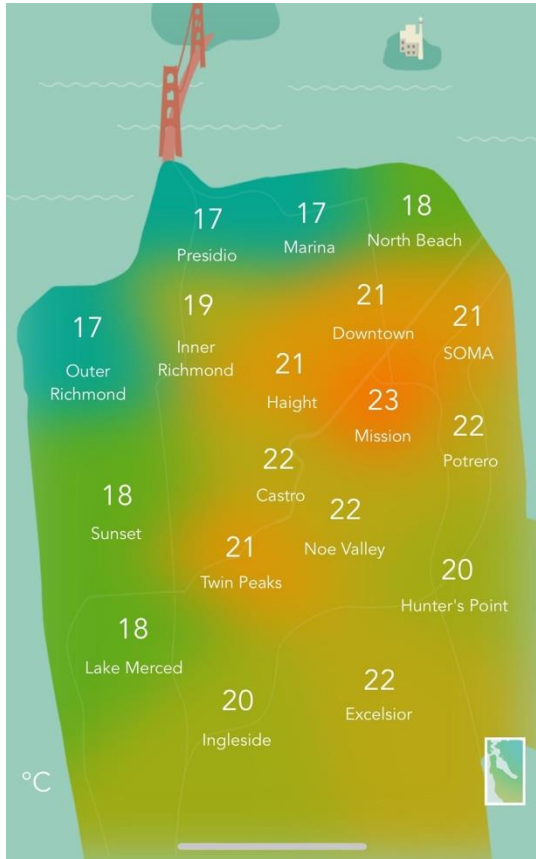
Wind: 8 mph

Weather

Monday

Rain showers

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



49 °F | °C

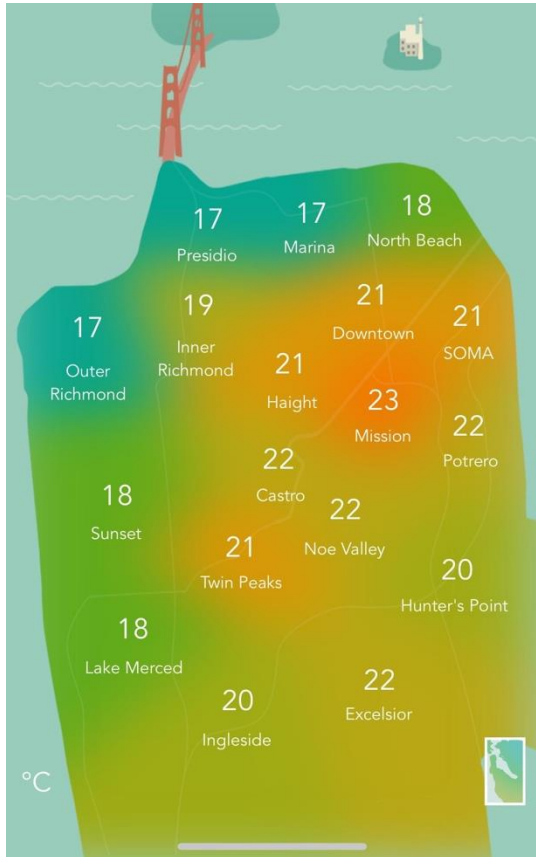
Precipitation: 90%
Humidity: 74%
Wind: 8 mph

Weather

Monday

Rain showers

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:

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

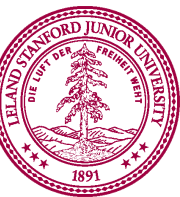
SARS Virus Testing

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 $P(F | E)$?

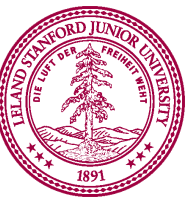
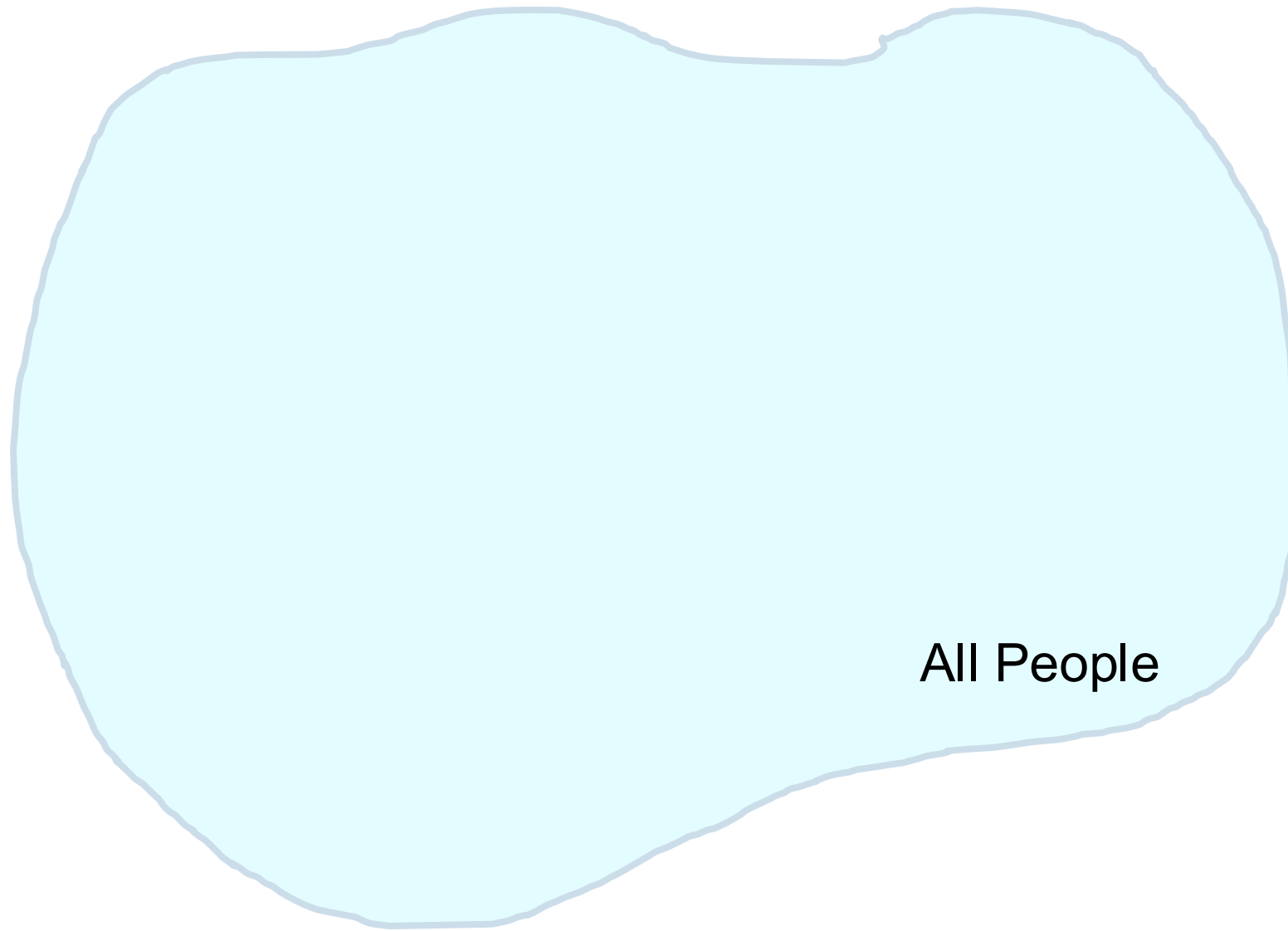
Solution:

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

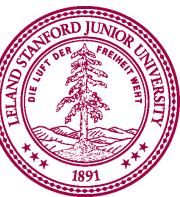
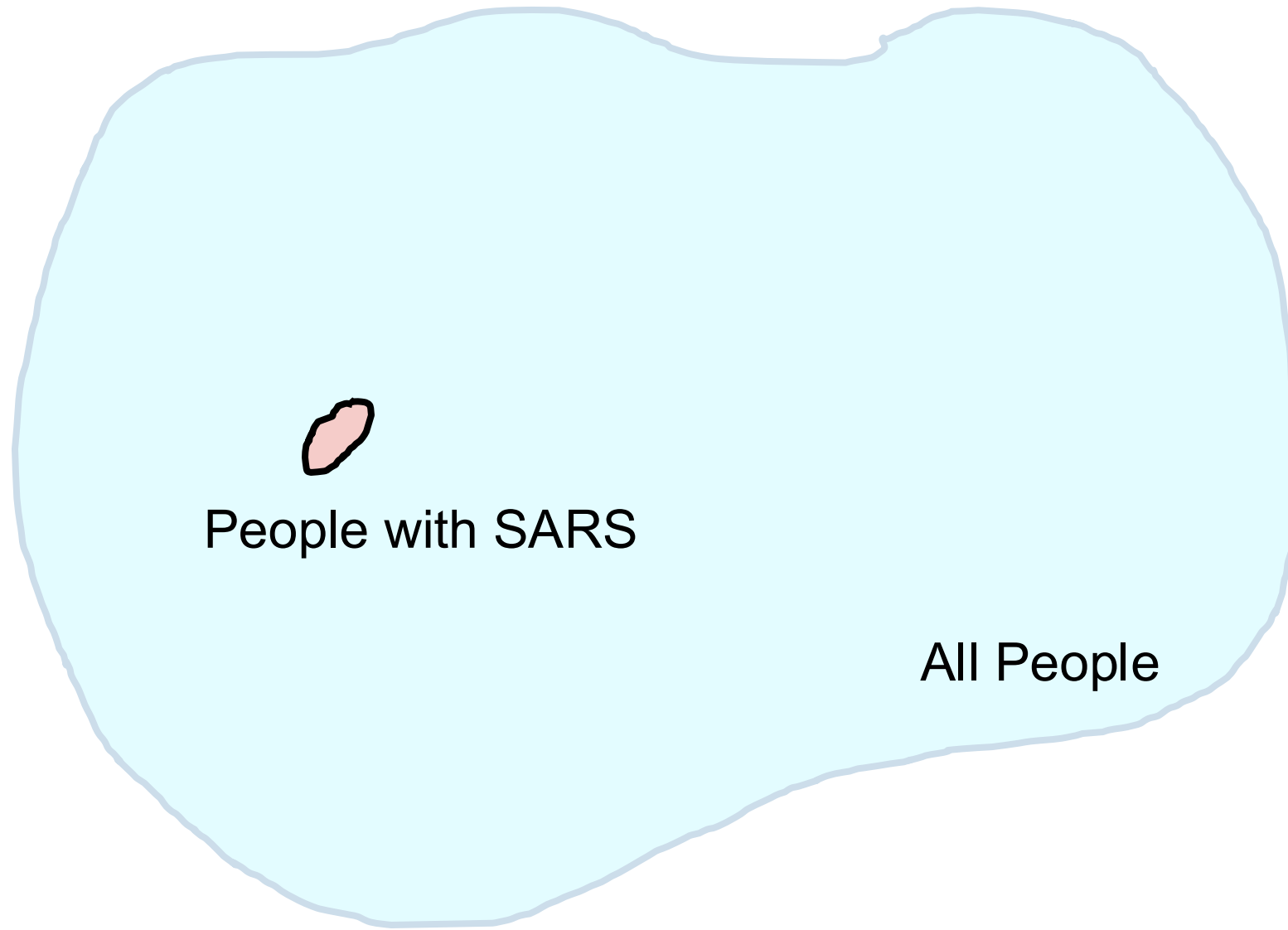


Intuition Time

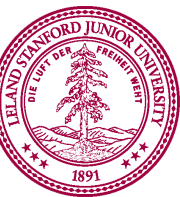
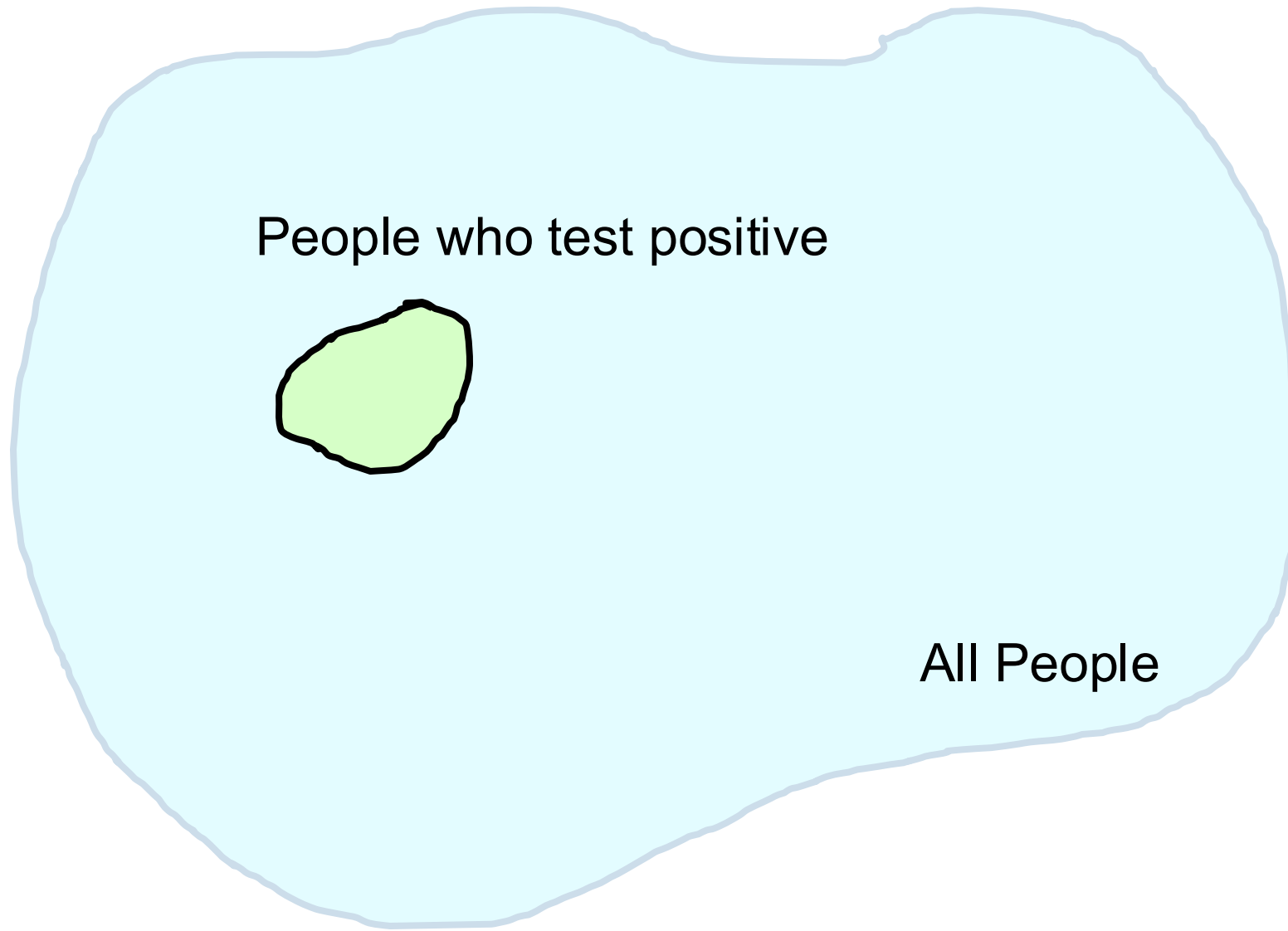
Bayes Theorem Intuition



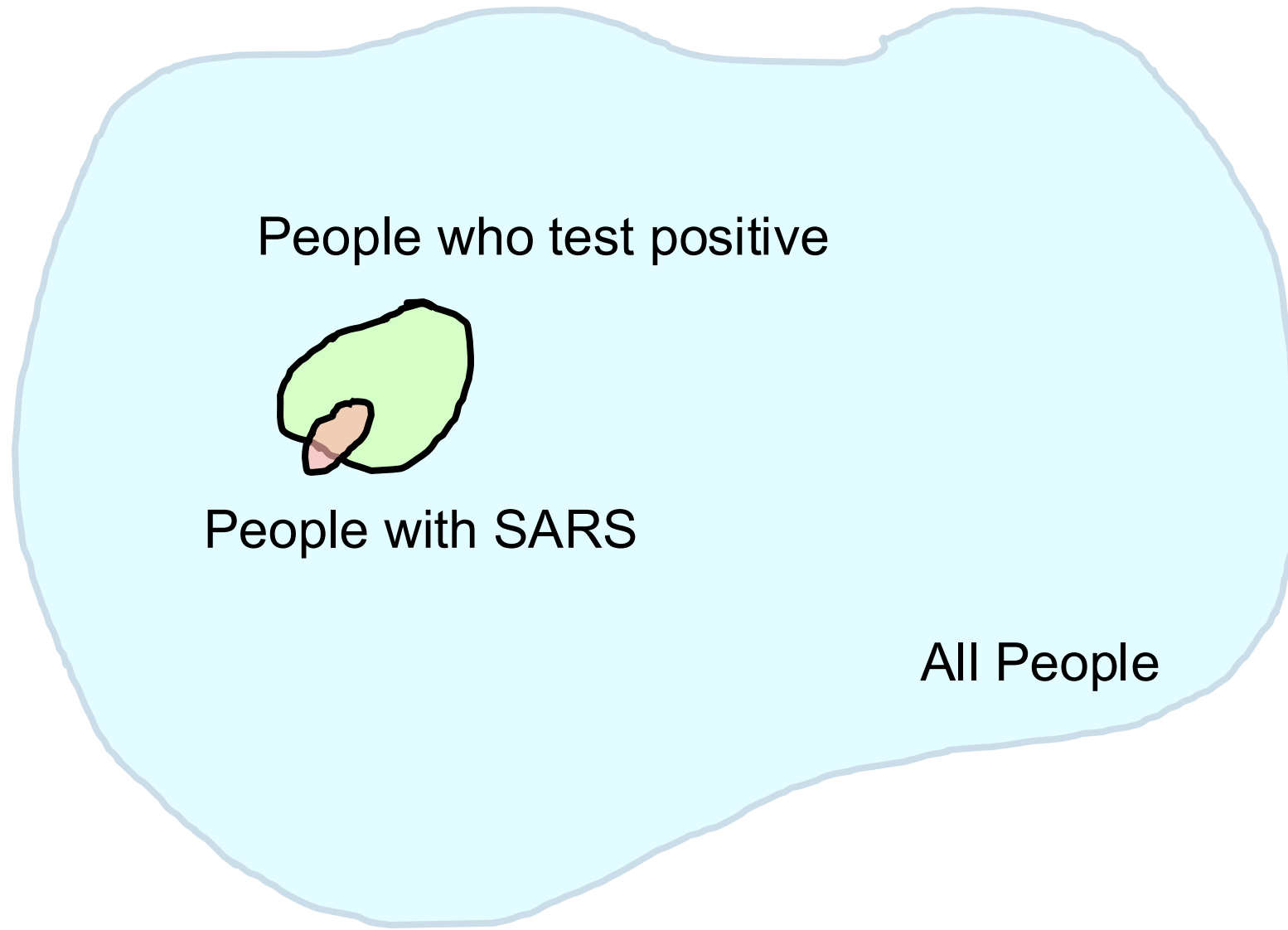
Bayes Theorem Intuition



Bayes Theorem Intuition

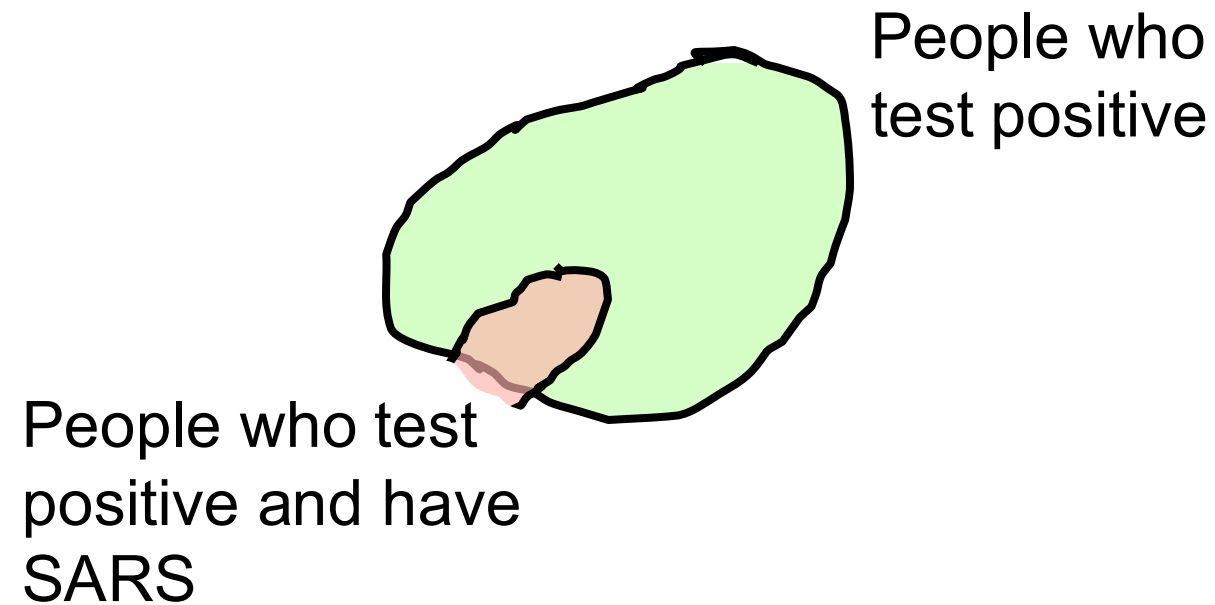


Bayes Theorem Intuition

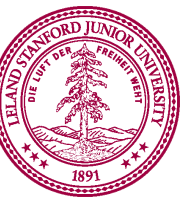


Bayes Theorem Intuition

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

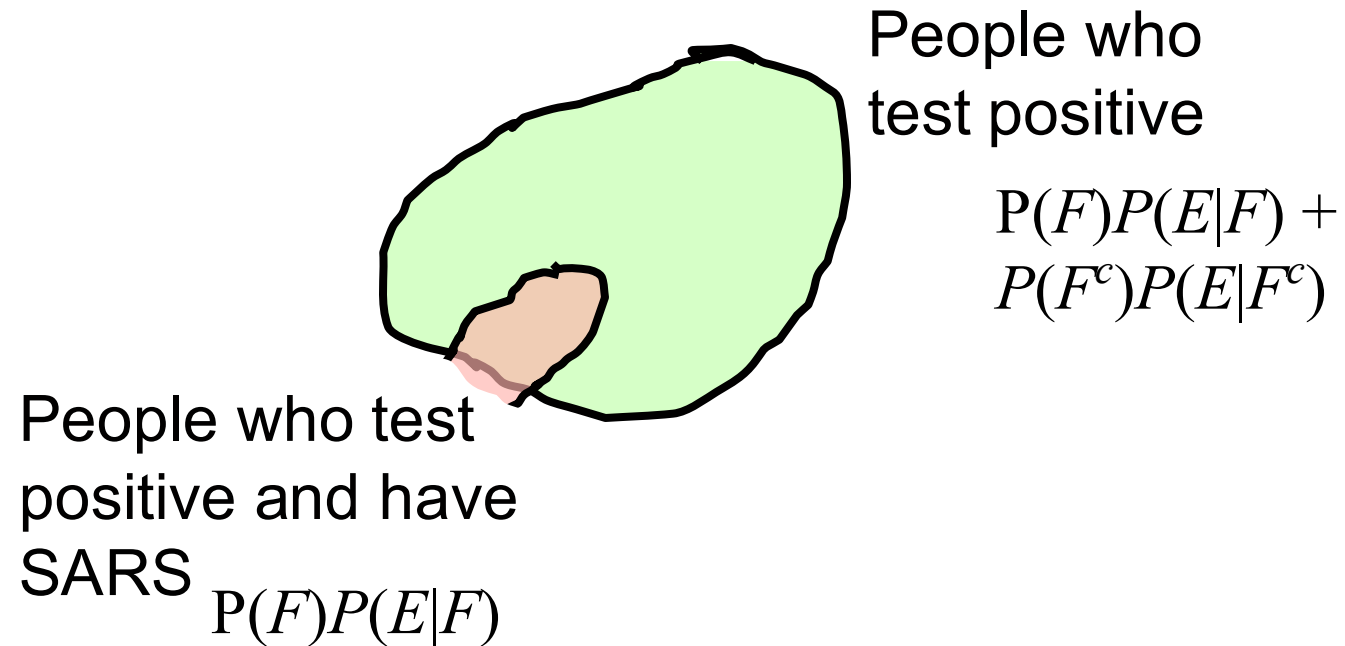


≈ 0.330

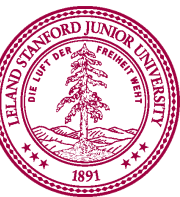


Bayes Theorem Intuition

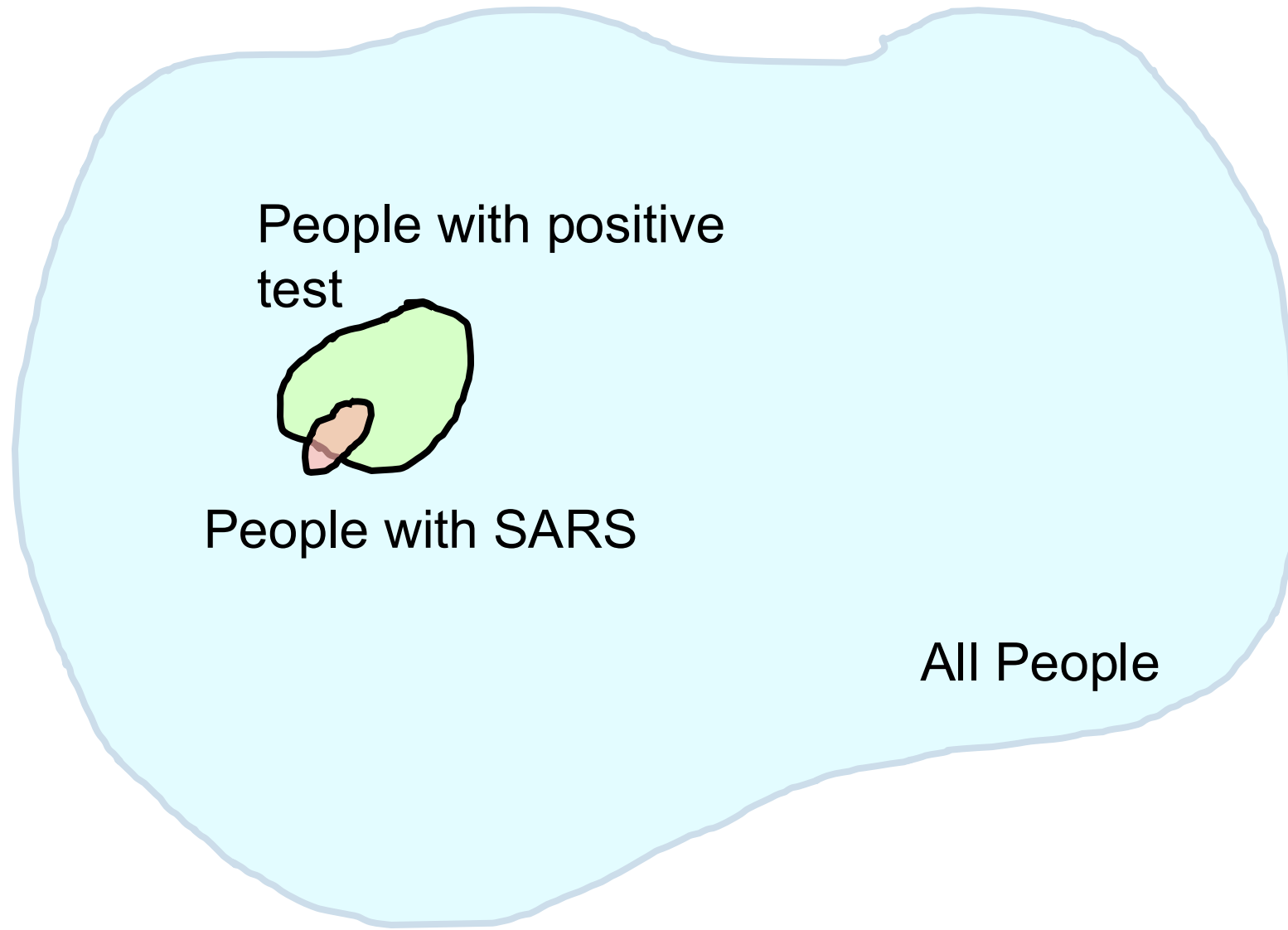
Conditioning on a positive result changes the sample space to this:



≈ 0.330

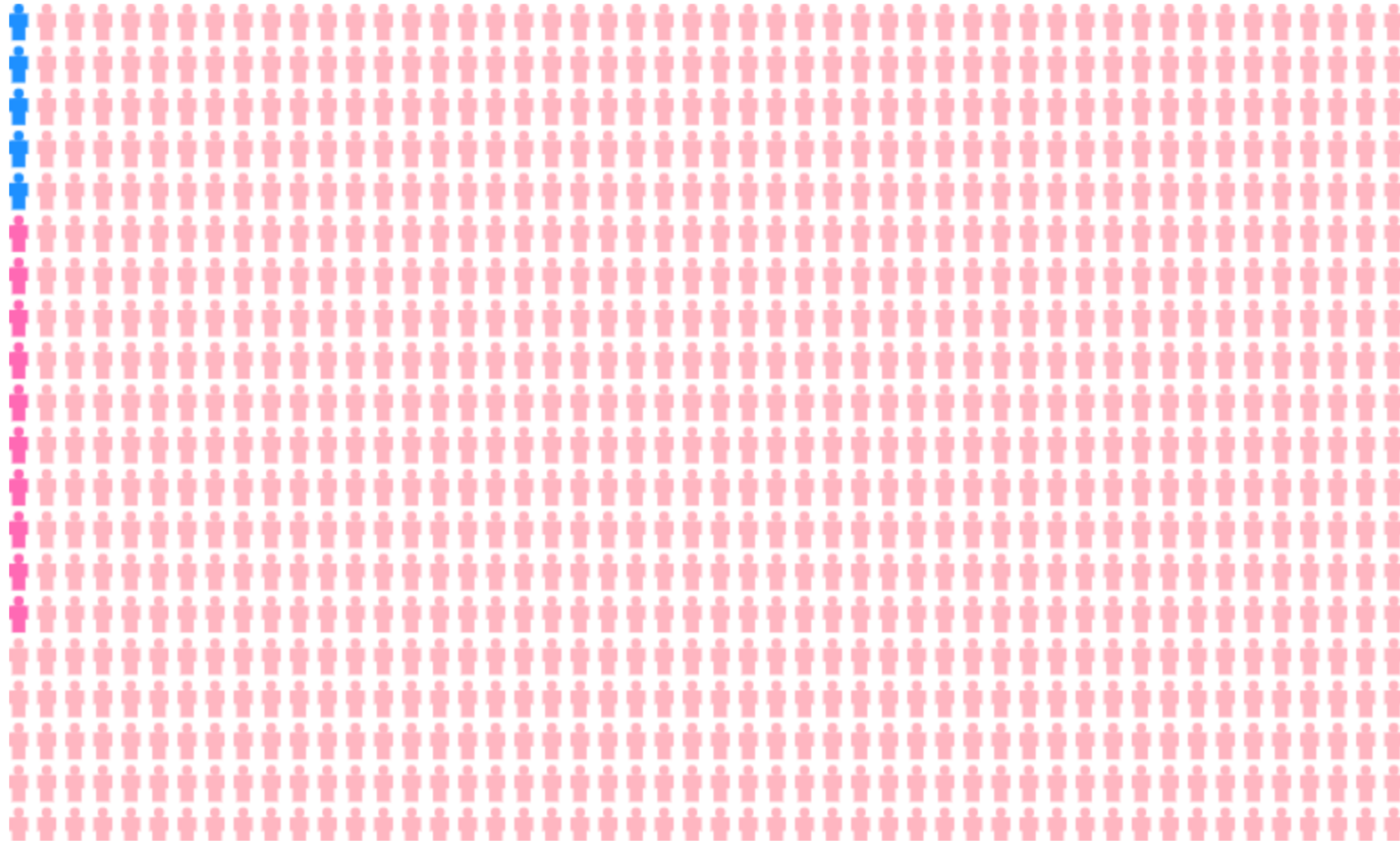


Bayes Theorem Intuition

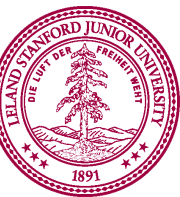


Bayes Theorem Intuition

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



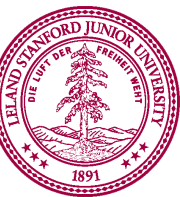
Bayes Theorem Intuition

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



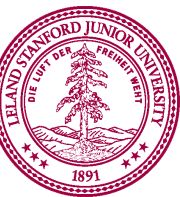
Why it is still good to get tested

	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
- What is $P(F | E^c)$?

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

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



Pedagogic Pause

Multiple Choice Theory

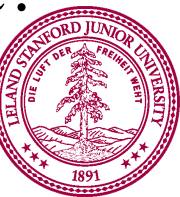
Let's consider the relationship between **knowing** the concepts used in a multiple choice midterm question, and getting the question correct, taking into account guessing and making silly mistakes.

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.

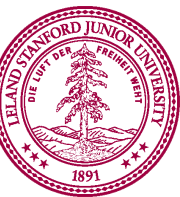
Should we switch?

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



1/3



1/3



A: Prize in Door A

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

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

B: Prize in Door B

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

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

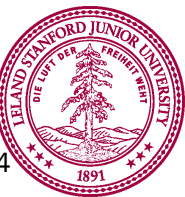
C: Prize in Door C

- 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!

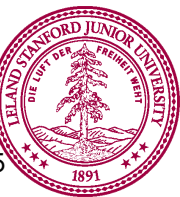


Marilyn Vos Savant



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?

$$\text{No: } P(\text{win without switching}) = \frac{1}{\text{original \# envelopes}}$$

$$\text{Yes: } P(\text{win with new knowledge}) = \frac{\text{original \# envelopes} - 1}{\text{original \# envelopes}}$$