Conditional Probability and Bayes

Announcements

- Pset #1 due next Monday.
- Section assignments will be sent out today. Can't make your time or need a swap? Reach out to us!

Review

Combinatorial

How many unique shuffles of a card deck are there?

52!

80658175170
64037669752
00000000000

A fun story we didn't get to explore last Friday. #28

Yunsung Kim STAFF
Now in General

UNPIN STAR WATCHING 1 VIEW

Last Friday, I announced in the beginning of class that we'll be "making history" at the end of lecture, but we didn't get to the part where we actually did. In this post, I just briefly wanted to explain the fun story that I wanted to tell you.

In 1992, Stanford Statistics professor Persi Diaconis proved in a paper titled "Trailing the Dovetail Shuffle to Its Lair" that doing a riffle-shuffle 7 times on a deck of playing cards yields a shuffling scheme that is very close to a truly random shuffle.

Let's assume that we just shuffled a deck of playing cards in a truly random fashion and ask the following bold question:

> **"What is the probability that nobody in the history of mankind has ever seen the exact same deck?"**

Review, Axioms of Probability

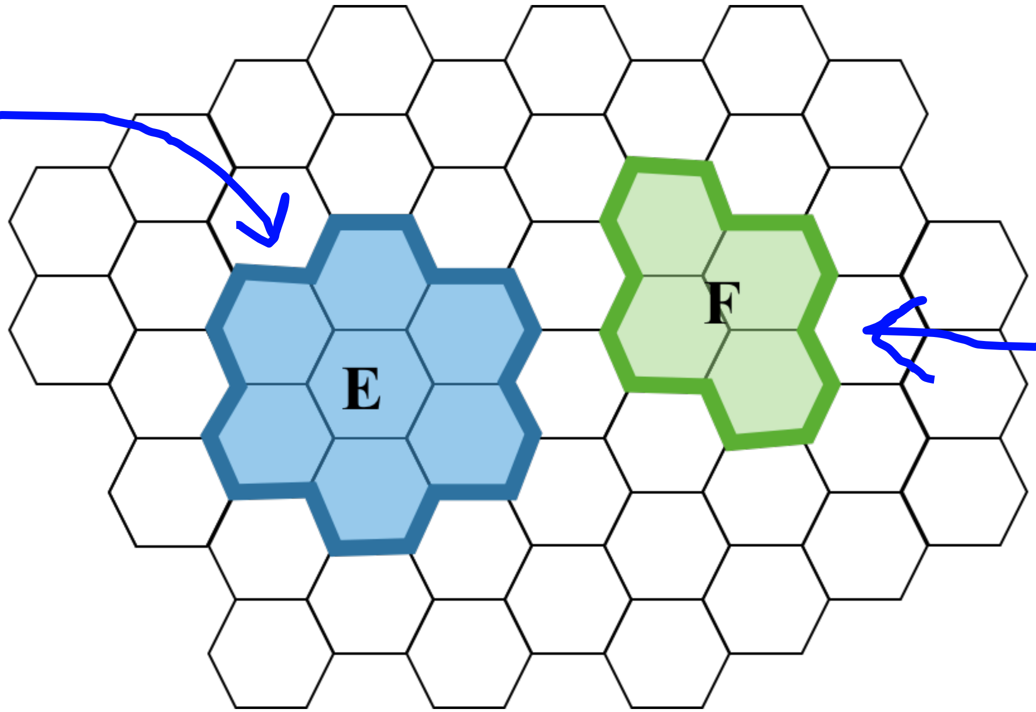
Recall: S = all possible outcomes. E = the event.

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

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

Review, Mutually Exclusive Events

7/50



4/50

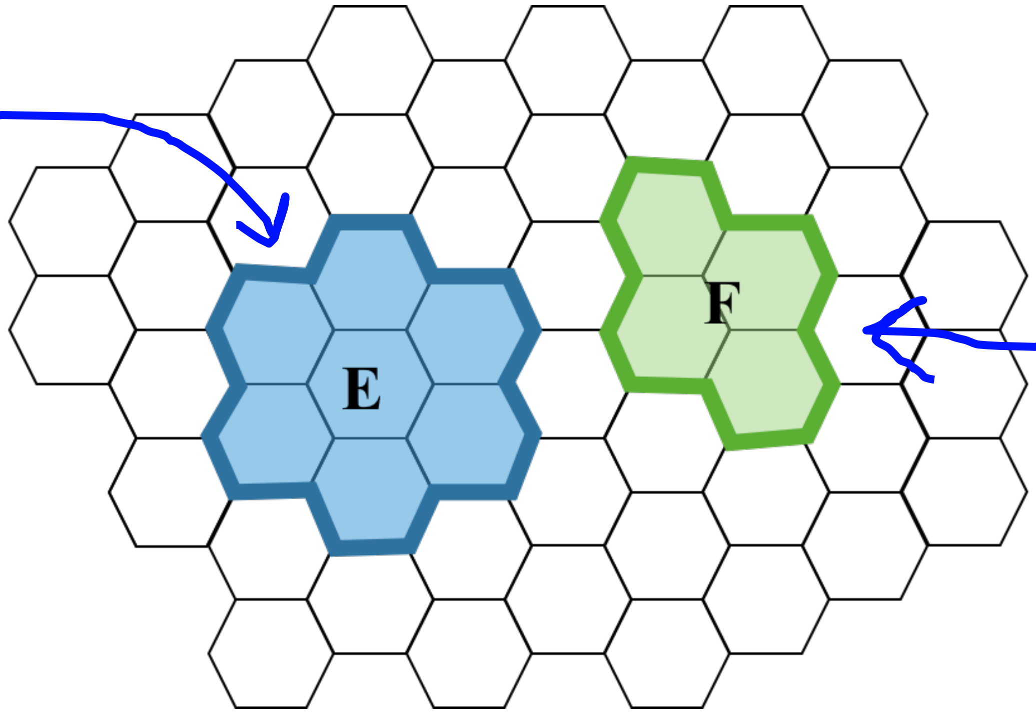
If events are mutually exclusive, probability of OR is simple:

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

Review, Mutually Exclusive Events

7/50

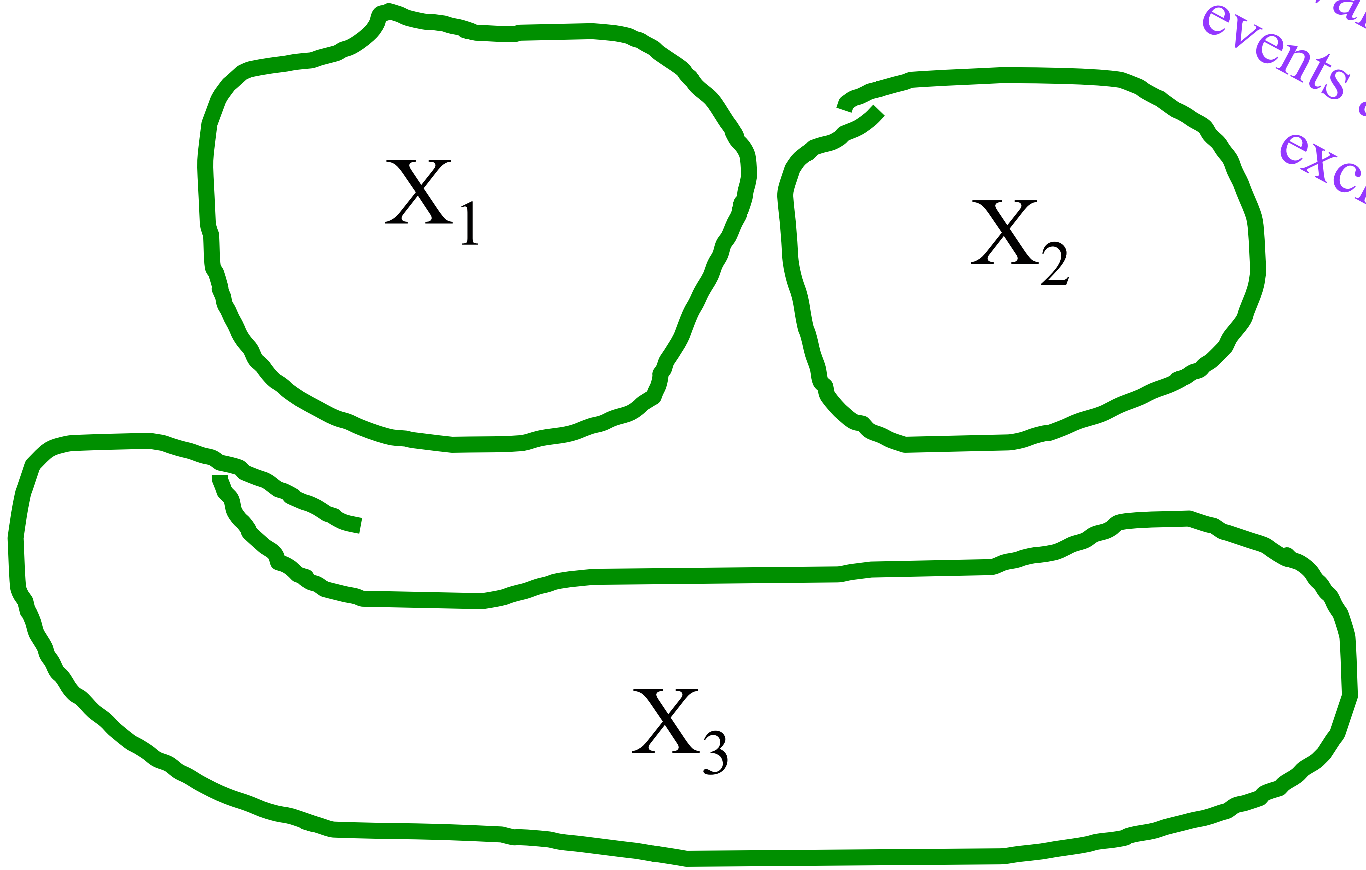
4/50



If events are mutually exclusive, probability of OR is simple:

$$P(E \cup F) = \frac{7}{50} + \frac{4}{5} = \frac{11}{50}$$

Review, Mutually Exclusive Events



Wahoo! All my events are mutually exclusive

$$P(X_1 \cup X_2 \cup \cdots \cup X_n) = \sum_{i=1}^{n} P(X_i)$$

Review, Mutually Exclusive Events

If events are *mutually exclusive* probability of OR is easy!

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

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

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

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

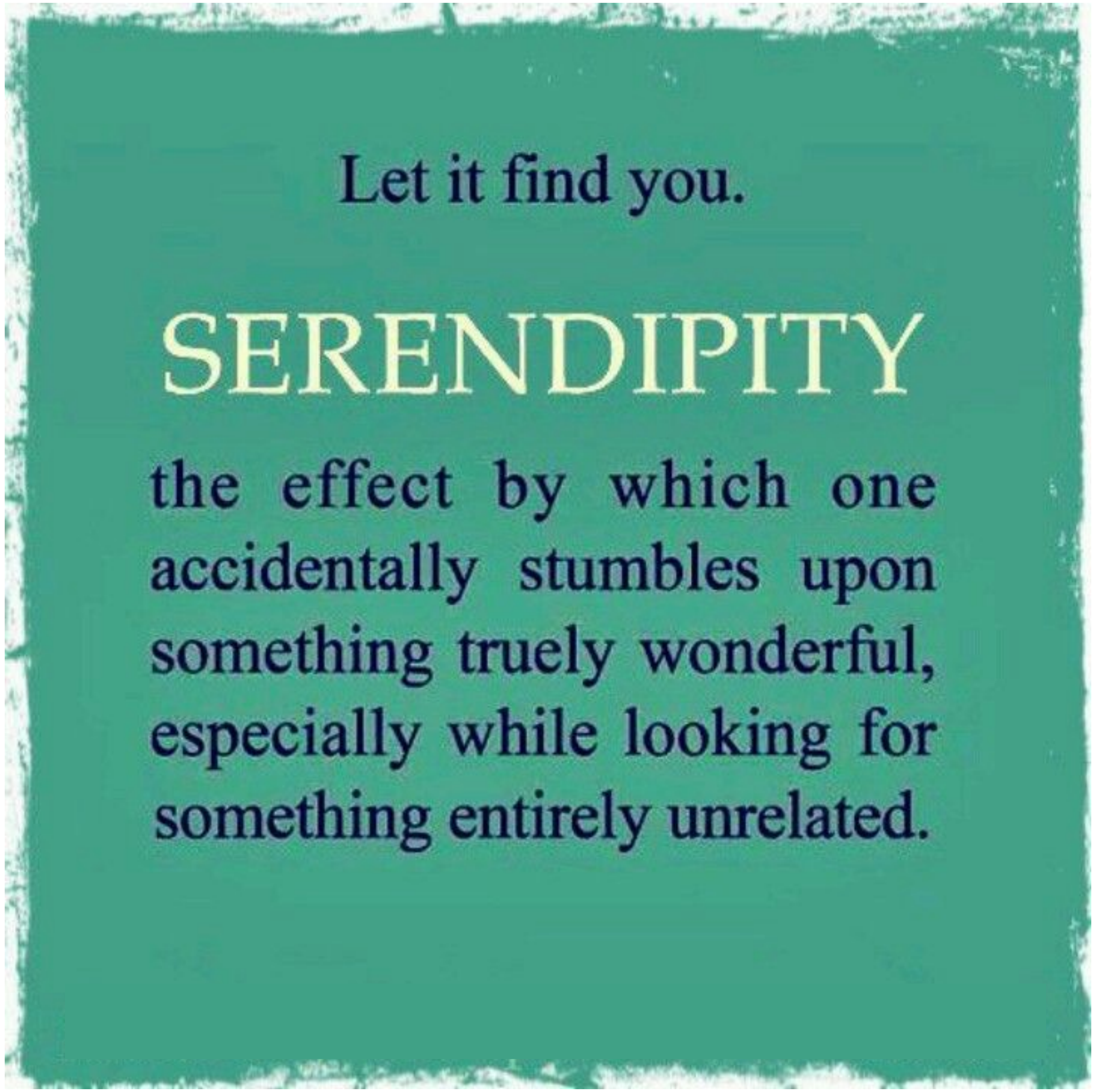
Since everything must either be in E or E^c

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

Axiom 2

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

Rearrange

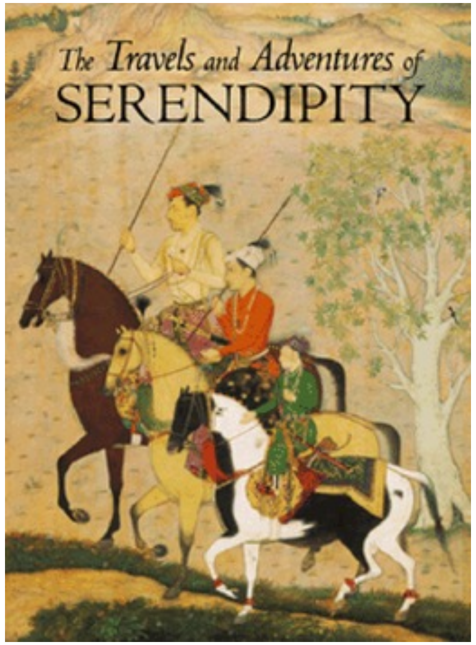

Let it find you.
SERENDIPITY
the effect by which one accidentally stumbles upon something truely wonderful, especially while looking for something entirely unrelated.

WHEN YOU MEET YOUR BEST FRIEND

Somewhere you didn't expect to.

Serendipity

- Say the population of Stanford is 17,000 people
  - You are friends with 80 people?
  - Walk into a room, see 64 random people.
  - What is the probability that you see someone you know?
  - Assume you are equally likely to see each person at Stanford



Many times it is easier to calculate $P(E^C)$.

End Review

Learning Goal for Today: Conditional Probability

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

Chain rule
(Product rule)

Definition of
conditional probability

$$P(E|F)$$

Law of Total
Probability

Bayes'
Theorem

$$P(E)$$

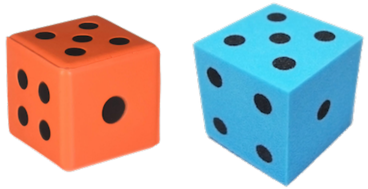
$$P(F|E)$$

Conditional Probability

Roll two dice

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

Roll two 6-sided fair dice. What is P(sum = 7)?

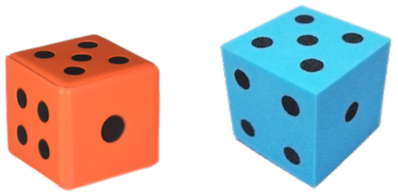


S = {(1,1) (1,2) (1,3) (1,4) (1,5) **(1,6)**
(2,1) (2,2) (2,3) (2,4) **(2,5)** (2,6)
(3,1) (3,2) (3,3) **(3,4)** (3,5) (3,6)
(4,1) (4,2) **(4,3)** (4,4) (4,5) (4,6)
(5,1) **(5,2)** (5,3) (5,4) (5,5) (5,6)
(6,1) (6,2) (6,3) (6,4) (6,5) (6,6) }

$E =$ *In blue*

Dice, our misunderstood friends

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



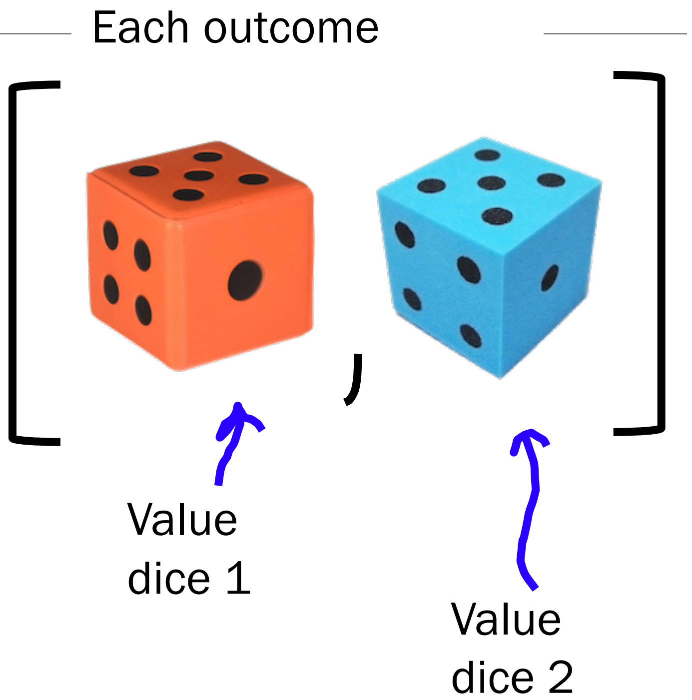
Which of the following situations would make you most/least hopeful, if you could know the value of D_1?

Your Choices:

- A. $D_1 = 1$
- B. $D_1 = 2$
- C. $D_1 = 3$
- D. $D_1 = 4$

Sum of Two Die = 4?

Roll two 6-sidex dice. What is probability the sum = 4?
Let E be the event that the sum is 4

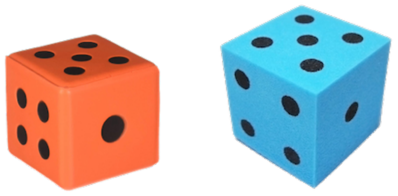


S = { [1,1] [1,2] [1,3] [1,4] [1,5] [1,6]
[2,1] [2,2] [2,3] [2,4] [2,5] [2,6]
[3,1] [3,2] [3,3] [3,4] [3,5] [3,6]
[4,1] [4,2] [4,3] [4,4] [4,5] [4,6]
[5,1] [5,2] [5,3] [5,4] [5,5] [5,6]
[6,1] [6,2] [6,3] [6,4] [6,5] [6,6] }

E =

Dice, our misunderstood friends

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



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

What is $P(E)$?

$|S| = 36$

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

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

Let F be event: $D_1 = 2$.

What is $P(E$, given F already observed$)$?

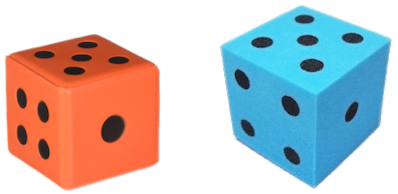
$S = \{(2,1), (2,2), (2,3), (2,4), (2,5), (2,6)\}$

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

$P(E) = 1/6$

Dice, our misunderstood friends

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



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

What is $P(E)$?

$|S| = 36$

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

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

Let F be event: $D_1 = 4$.

What is $P(E$, if F happens$)$?

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

$E = \{ \quad \}$

$P(E) = \frac{0}{6} = 0$

Conditional Probability

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

Written as: $P(E|F)$

Means: "$P(E$, given F already observed)"

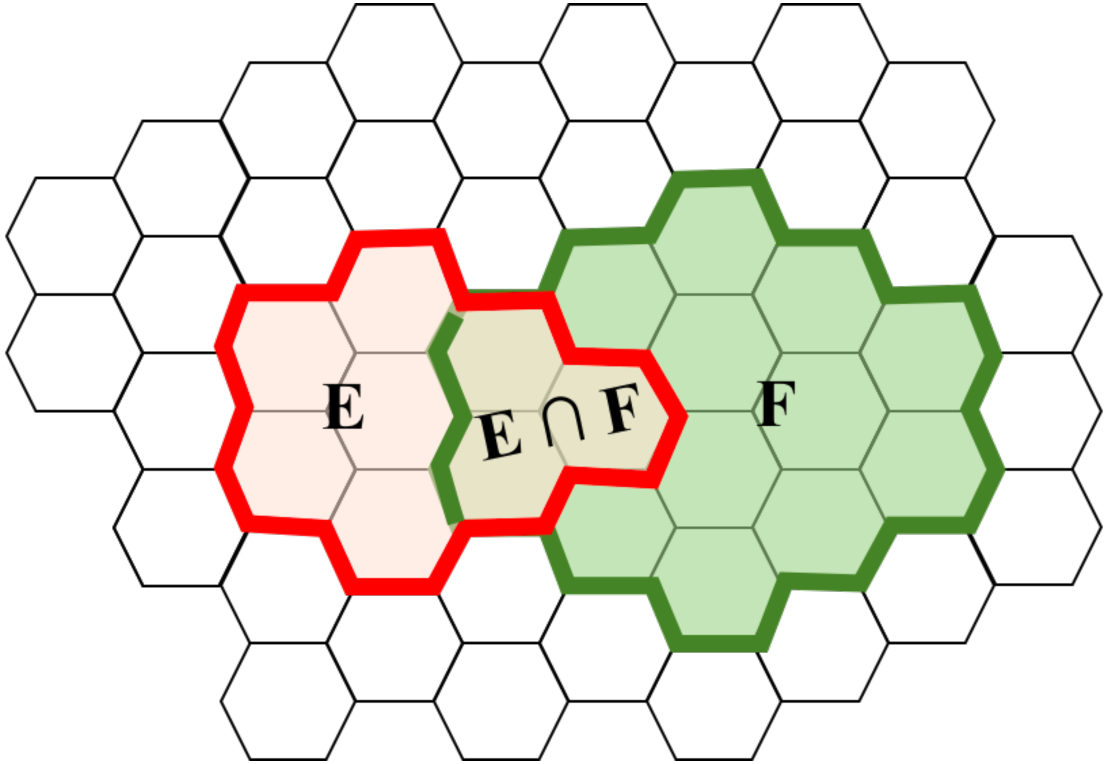
Sample space → all possible outcomes consistent with F (i.e. $S \cap F$)

Event → all outcomes in E consistent with F (i.e. $E \cap F$)

Some words that imply conditional probabilities are:
"given," "assuming," "if," "when"...

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.



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

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

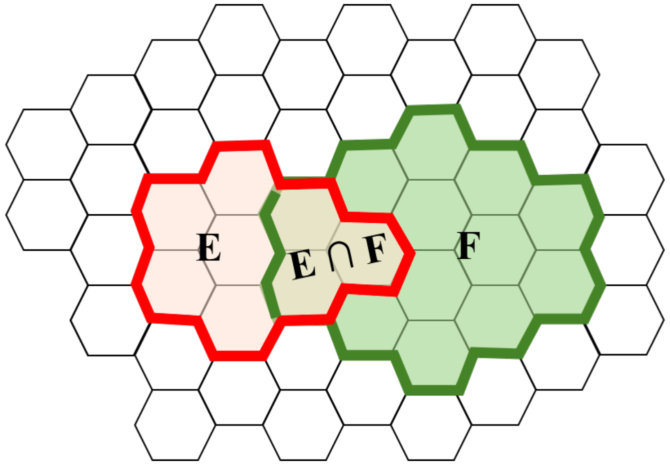
Conditional Probability, equally likely outcomes

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")

With equally likely outcomes:

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



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

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

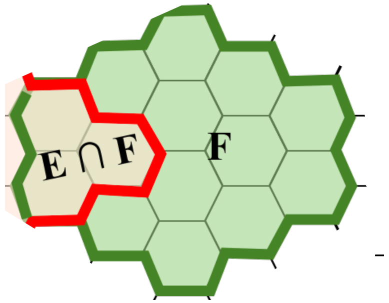
Conditional Probability, equally likely outcomes

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")

With equally likely outcomes:

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



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

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

Conditional probability in general

These properties hold even when outcomes are not equally likely.

General definition of conditional probability:

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

The Chain Rule (aka Product rule):

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

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

What if P(F) = 0?

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

Direction of conditioning matters

Direction of conditioning matters:

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

(Often, conditional probability in one direction (observation|cause) is more "natural" and easier than conditioning in the other direction (cause|observation), which is more "backward" and "inferential")

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) = ½ ?

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?

$$P(E)$$

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

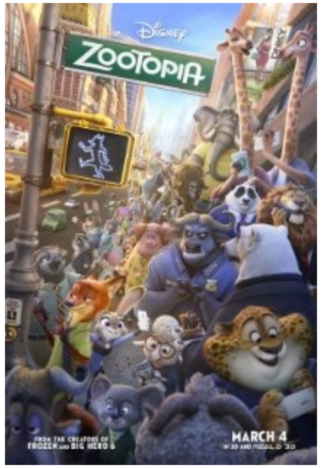
$P(E) = 10{,}234{,}231 \,/\, 50{,}923{,}123 = 0.20$

Netflix and Learn

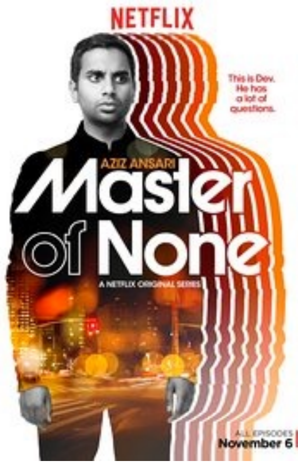
$P(E|F) = \frac{P(EF)}{P(F)}$ Definition of Cond. Probability

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

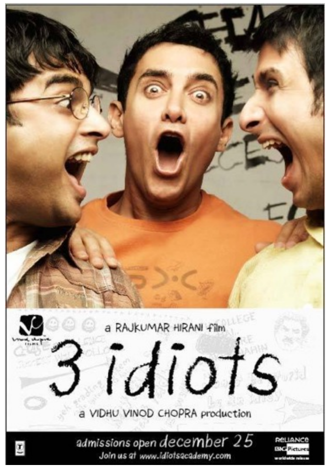
$P(E) = 0.19$



$P(E) = 0.32$



$P(E) = 0.20$



$P(E) = 0.09$

$P(E) = 0.20$

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?

$$P(E|F)$$

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

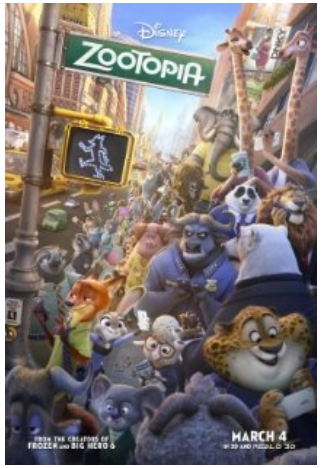
$$= \frac{\text{\# people who have watched both}}{\text{\# people who have watched CODA}}$$

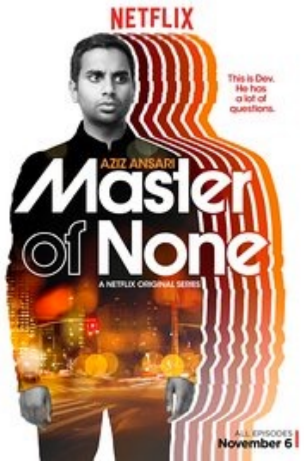
$$\approx 0.42$$

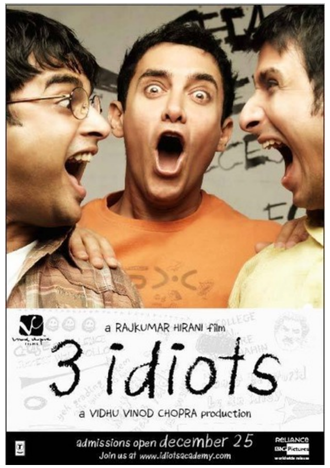
Netflix and Learn

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







$P(E) = 0.19$	$P(E) = 0.32$	$P(E) = 0.20$	$P(E) = 0.09$	$P(E) = 0.20$
$P(E\|F) = 0.14$	$P(E\|F) = 0.35$	$P(E\|F) = 0.20$	$P(E\|F) = 0.72$	$P(E\|F) = 0.42$

Machine Learning

Machine Learning is:
Probability + Data + Computers

Notation

And	**Or**	**Given**
$P(E \text{ and } F)$	$P(E \text{ or } F)$	$P(E\|F)$
$P(E,F)$	$P(E \cup F)$	$P(E\|F,G)$
$P(EF)$		Probability of E given F and G
$P(E \cap F)$		

Chain Rule via Playful Doggie

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

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

Just before bedtime, a dog has a 50% chance of bringing a toy to bed. The chance that the dog wants to be petted **given** that she has brought a toy is 50%. What is the probability that the dog **brought a toy, and wants to be petted?**

Generalized Chain Rule

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

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

Law of Total Probability

Playful Doggie Redux

Just before bedtime, a dog has a 50% chance of bringing a toy to bed. The chance that the dog comes asking to be petted **given** that she has brought a toy is 50%.

What is the probability of wanting to be petted, unconditioned?

What information do you need?

Relationship Between Probabilities

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

Chain rule
(Product rule)

Definition of
conditional probability

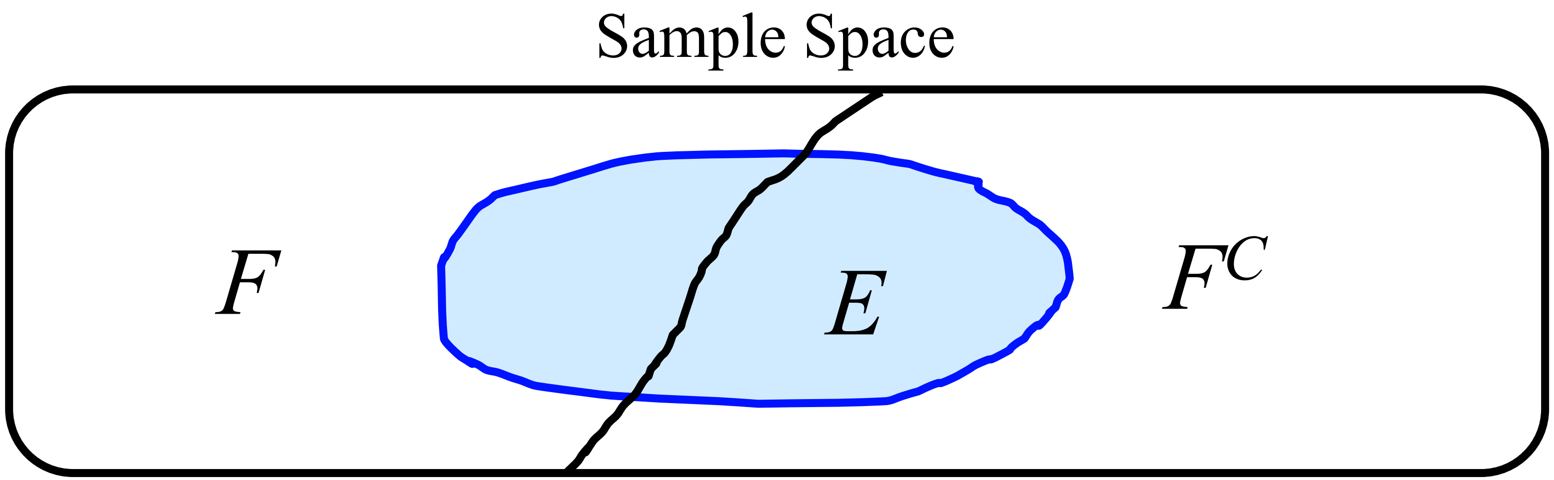
$$P(E|F)$$

Law of Total
Probability

$$P(E)$$

Law of Total Probability

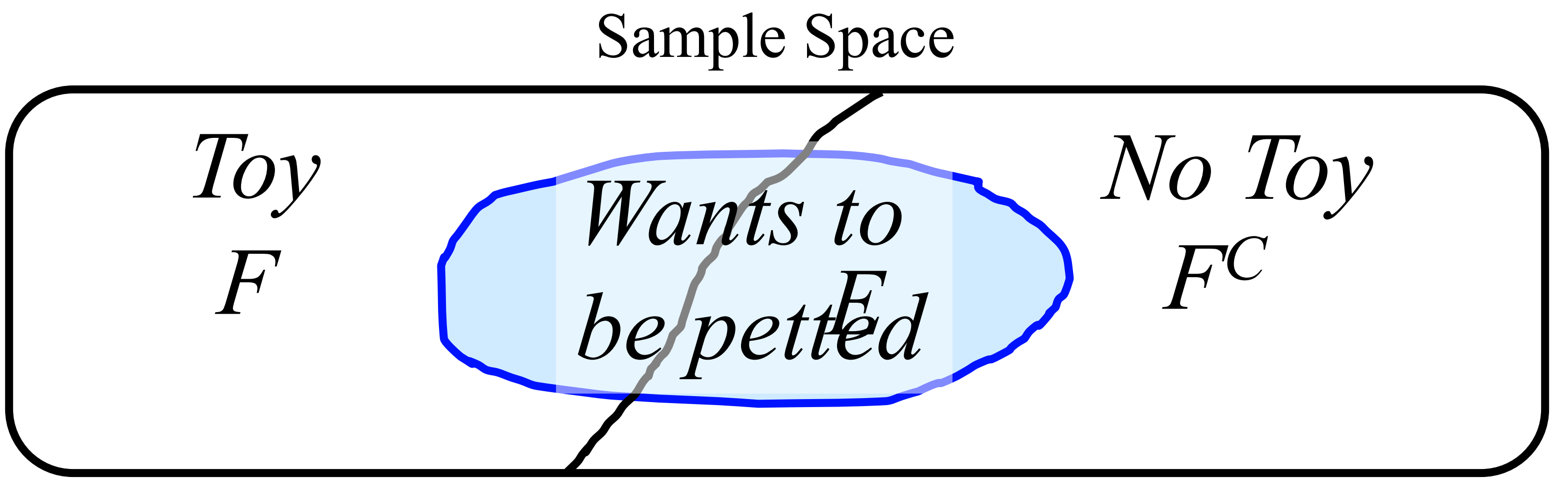
Say E and F are events in S



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

Law of Total Probability

Say E and F are events in S

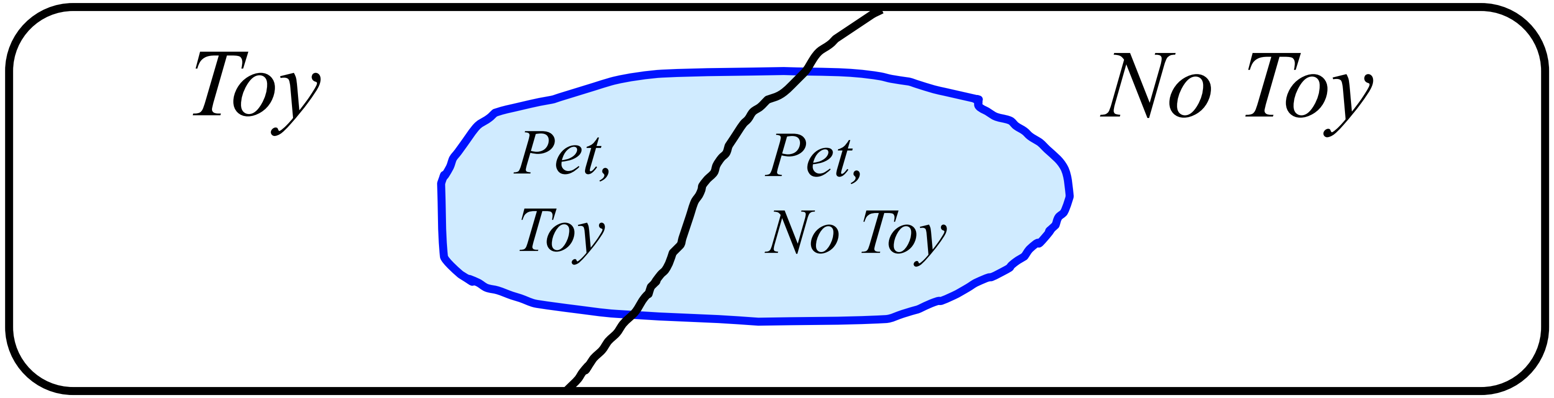


$$P(E) = P(EF) + P(EF^C)$$
$$= P(E|F)P(F) + P(E|F^C)P(F^C)$$

Law of Total Probability

Say E and F are events in S

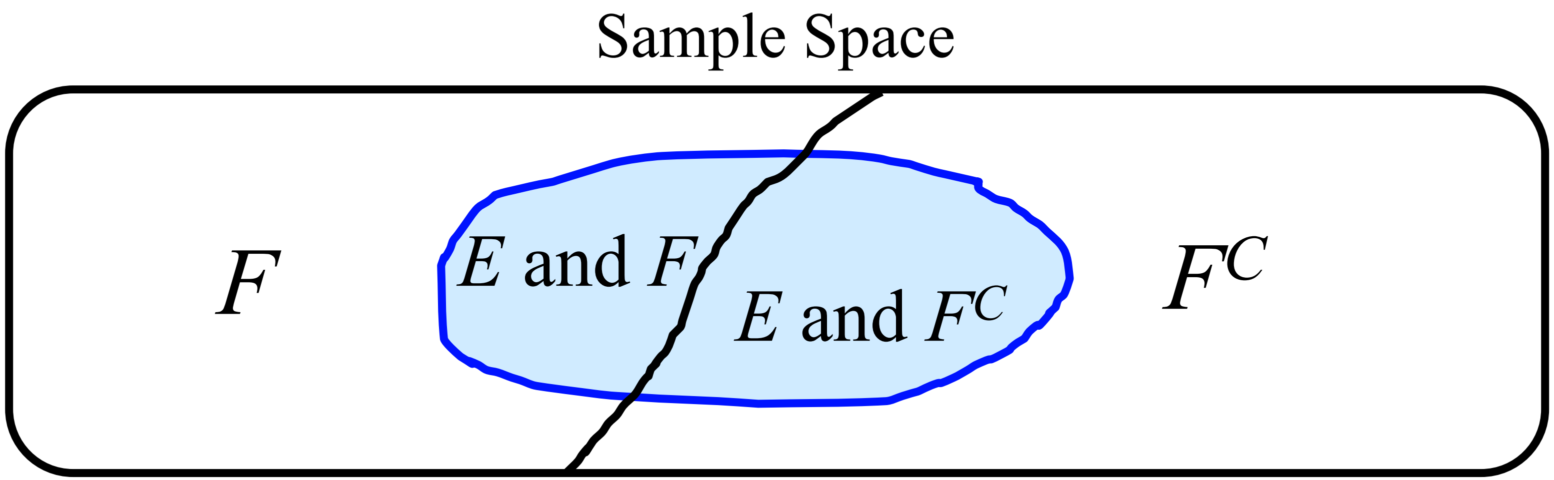
Sample Space



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

Law of Total Probability

Say E and F are events in S



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

Law of Total Probability

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

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

<u>Proof</u>

1. $E = (EF)$ or (EF^C)	Since F and F^C are disjoint
2. $P(E) = P(EF) + P(EF^C)$	Probability of **or** for disjoint
3. $P(E) = P(E|F)P(F) + P(E|F^C)P(F^C)$	Chain rule (product rule)

When Conditional probabilities are easier to obtain than unconditional,
Law of Total Probability allows you to get unconditional from conditional.

Playful Doggie

Just before bedtime, a dog has a 50% chance of bringing a toy to bed. The chance that the dog comes asking to be petted **given** that she has brought a toy is 50%.

Probability of wanting to be petted (W)?

What information do you need?

Probability of wanting petting given no toy.

Recall that W is petting and T is toy

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

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

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?

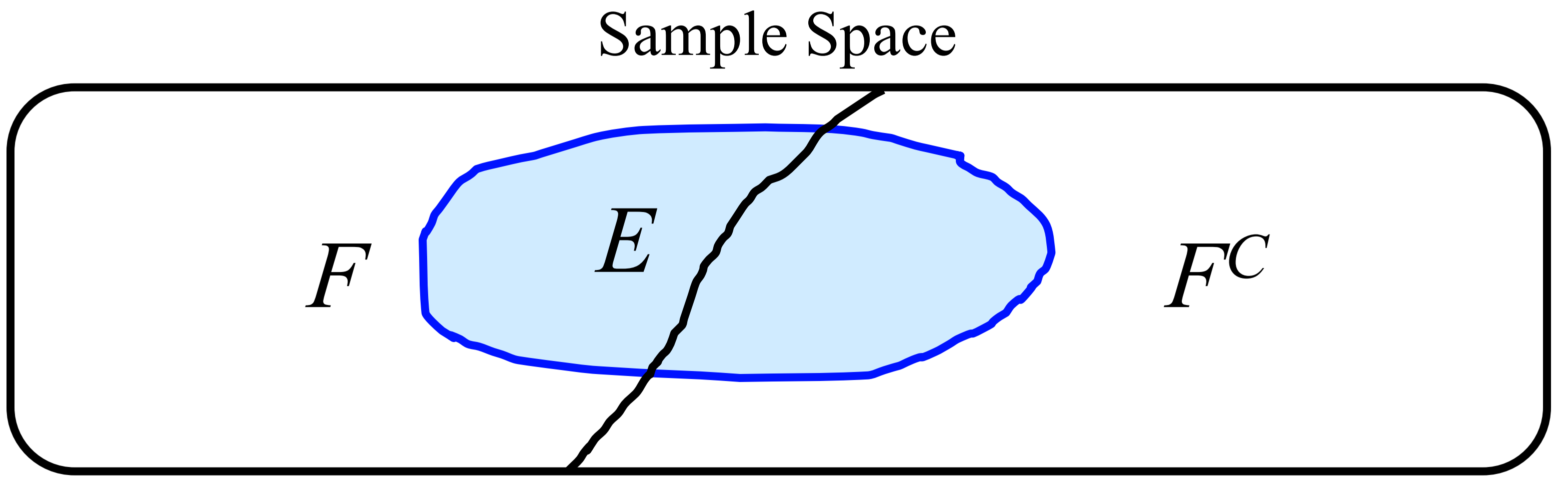
Let E be the event that a bacterium survives. Let M be the event that a bacteria has the mutation. By the By Law of Total Probability (LOTP):

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

Steps for solving word probability problems

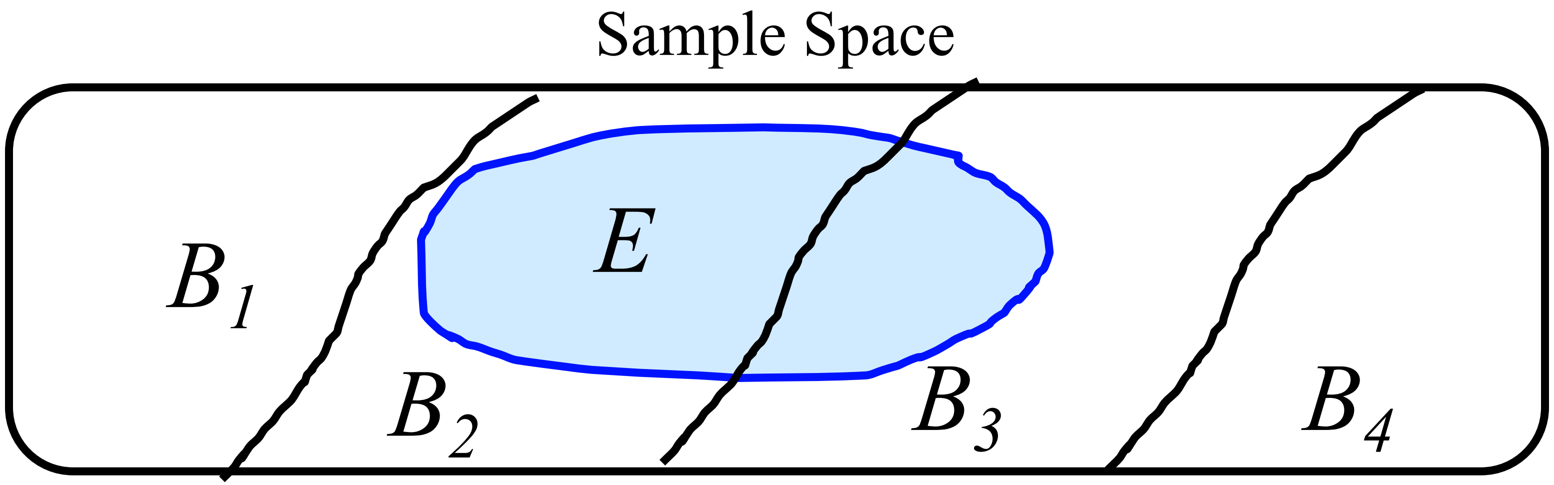
1. Define your events
2. Find what you need to solve for
3. List all probabilities you are given (Including the complements!)
4. Think about what law can be used to get to the answer

Law of Total Probability



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

Law of Total Probability



Thm For mutually exclusive events $B_1, B_2, \ldots, B_n$ s.t. $B_1 \cup B_2 \cup \cdots \cup B_n = S$,

$$P(E) = \sum_i P(B_i \cap E)$$

$$= \sum_i P(E|B_i)P(B_i)$$

Real question. What is the probability that a surviving bacteria has the mutation?

$$\Pr(\text{Mutation} \mid \text{Survives})$$

$$\Pr(M \mid S)$$

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

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

Bayes’ Theorem

Thomas Bayes

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

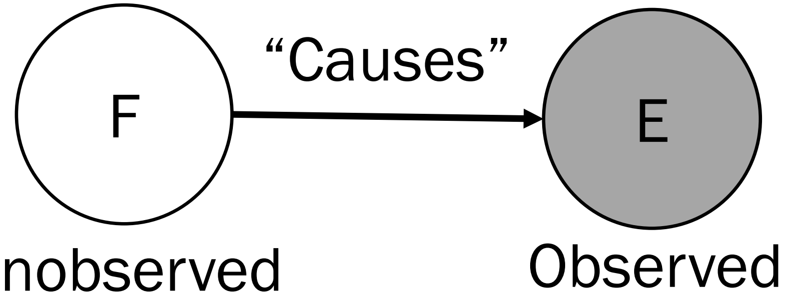
Thomas Bayes

$P(F \mid E)$

I want to calculate
P(State of the world F | Observation E)
It seems so tricky!…

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

$P(E \mid F)$



Thomas Bayes

Want $P(F \mid E)$. Know $P(E \mid F)$

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

A little while later…

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

A little while later…

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

Bayes' Theorem

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

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

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

Proof

2 steps! See board

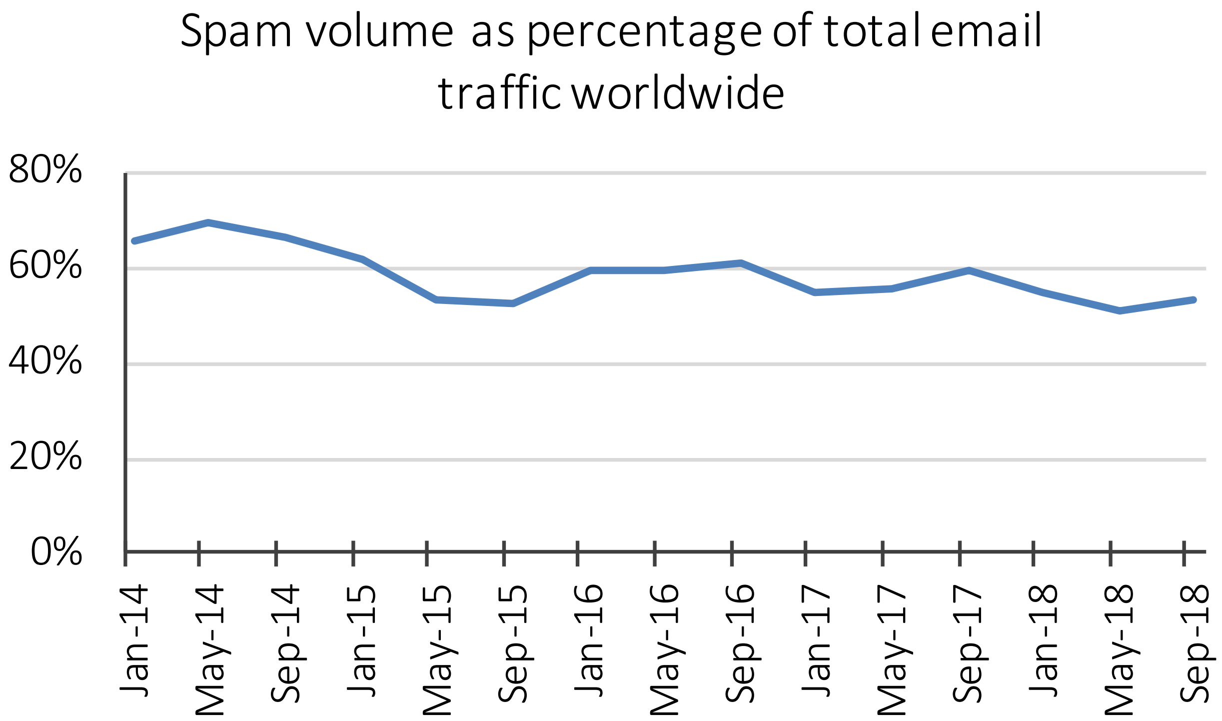
Expanded form:

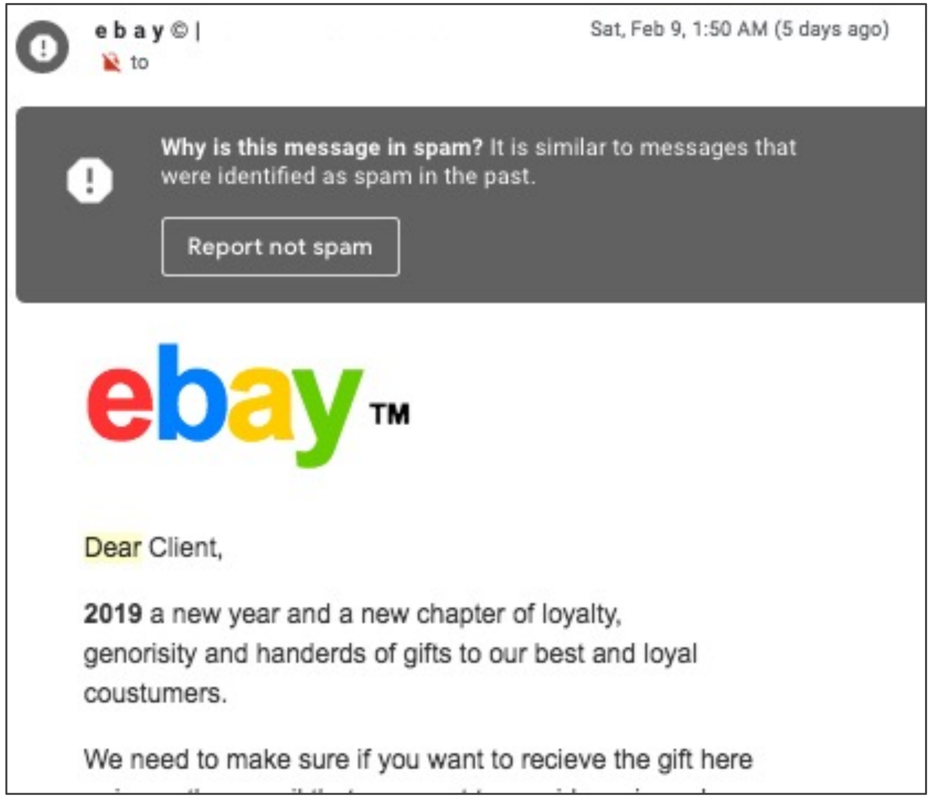
$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^C)P(F^C)}$$

Proof

1 more step! See board

Detecting spam email





We can easily calculate how many spam emails contain "Dear":

$$P(E|F) = P\left(\text{“Dear”}\,\middle|\,\begin{matrix}\text{Spam}\\ \text{email}\end{matrix}\right)$$

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

$$P(F|E) = P\left(\begin{matrix}\text{Spam}\\ \text{email}\end{matrix}\,\middle|\,\text{“Dear”}\right)$$

(silent drumroll)

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(spam | "Dear")

$= P(F|E)$

Bayes' Theorem terminology

- 60% of all email in 2016 is spam. $P(F)$
- 20% of spam has the word "Dear" $P(E|F)$
- 1% of non-spam (aka ham) has the word "Dear" $P(E|F^C)$

You get an email with the word "Dear" in it.

What is the probability that the email is spam? Want: $P(F|E)$

likelihood prior

posterior

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

normalization constant

Observe E

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

"Updating" your belief

Prior → Posterior

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)?

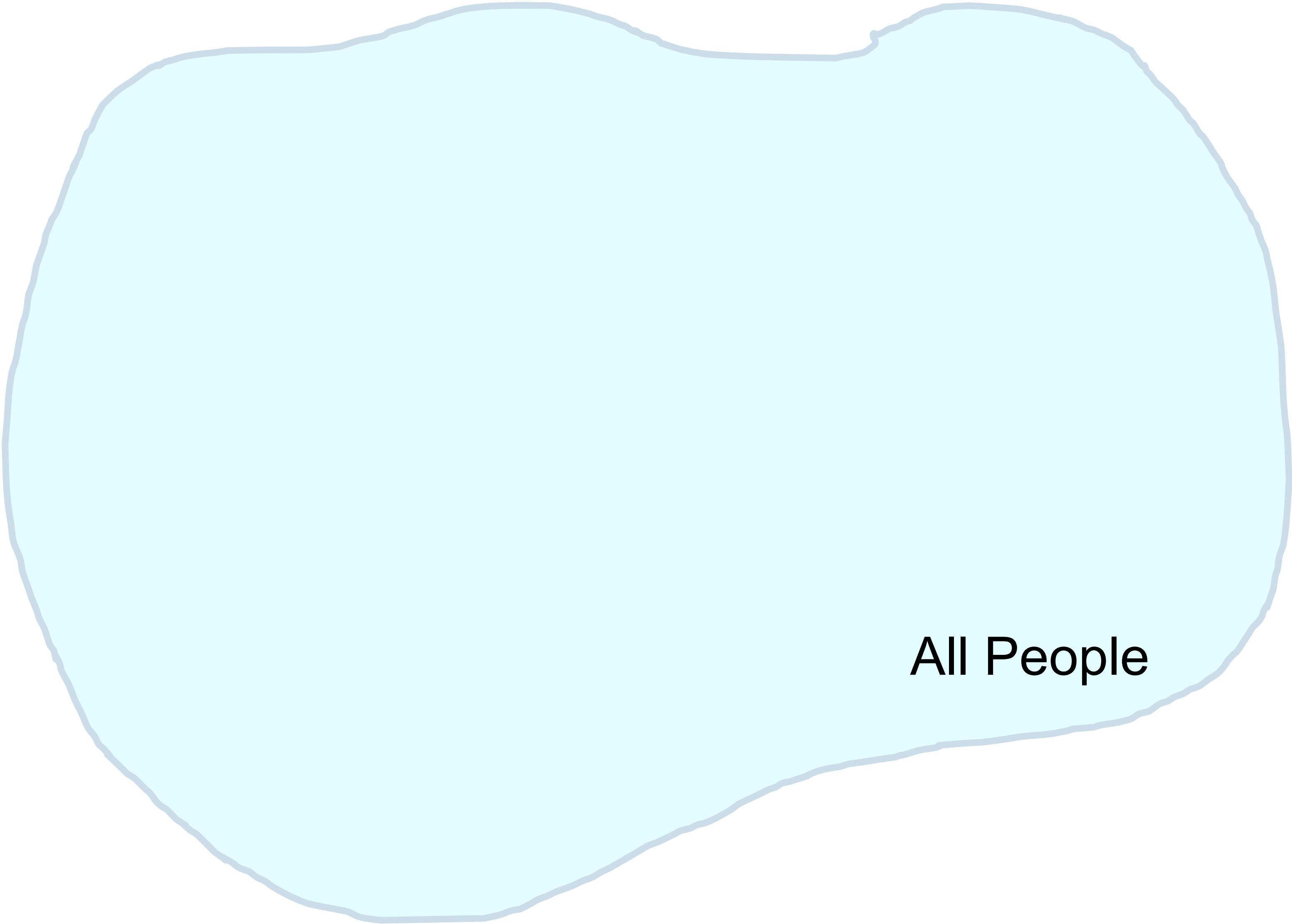
Solution:

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

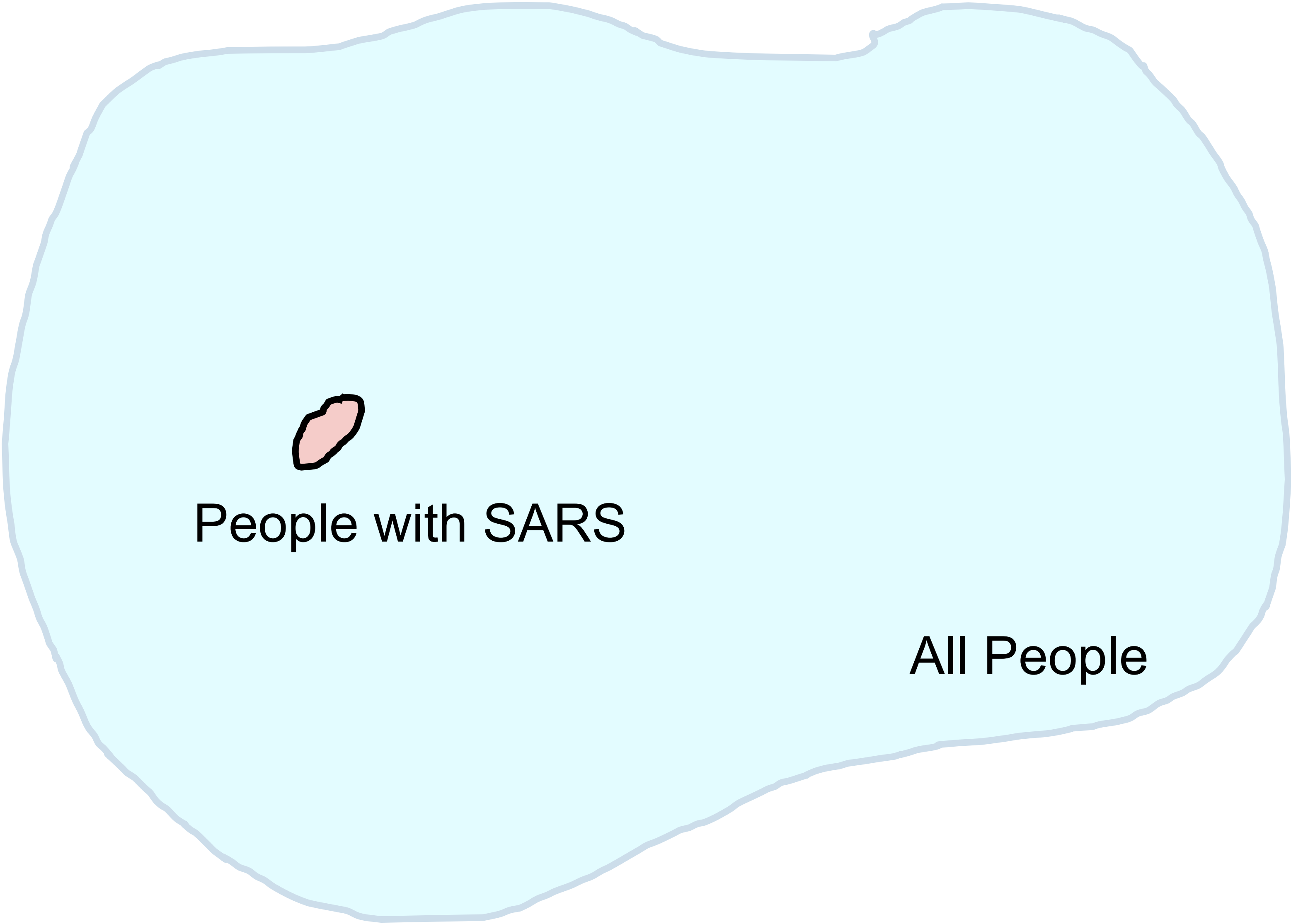
$$P(F \mid E) = \frac{(0.98)(0.005)}{(0.98)(0.005) + (0.01)(1 - 0.005)} \approx 0.330$$

Intuition Time

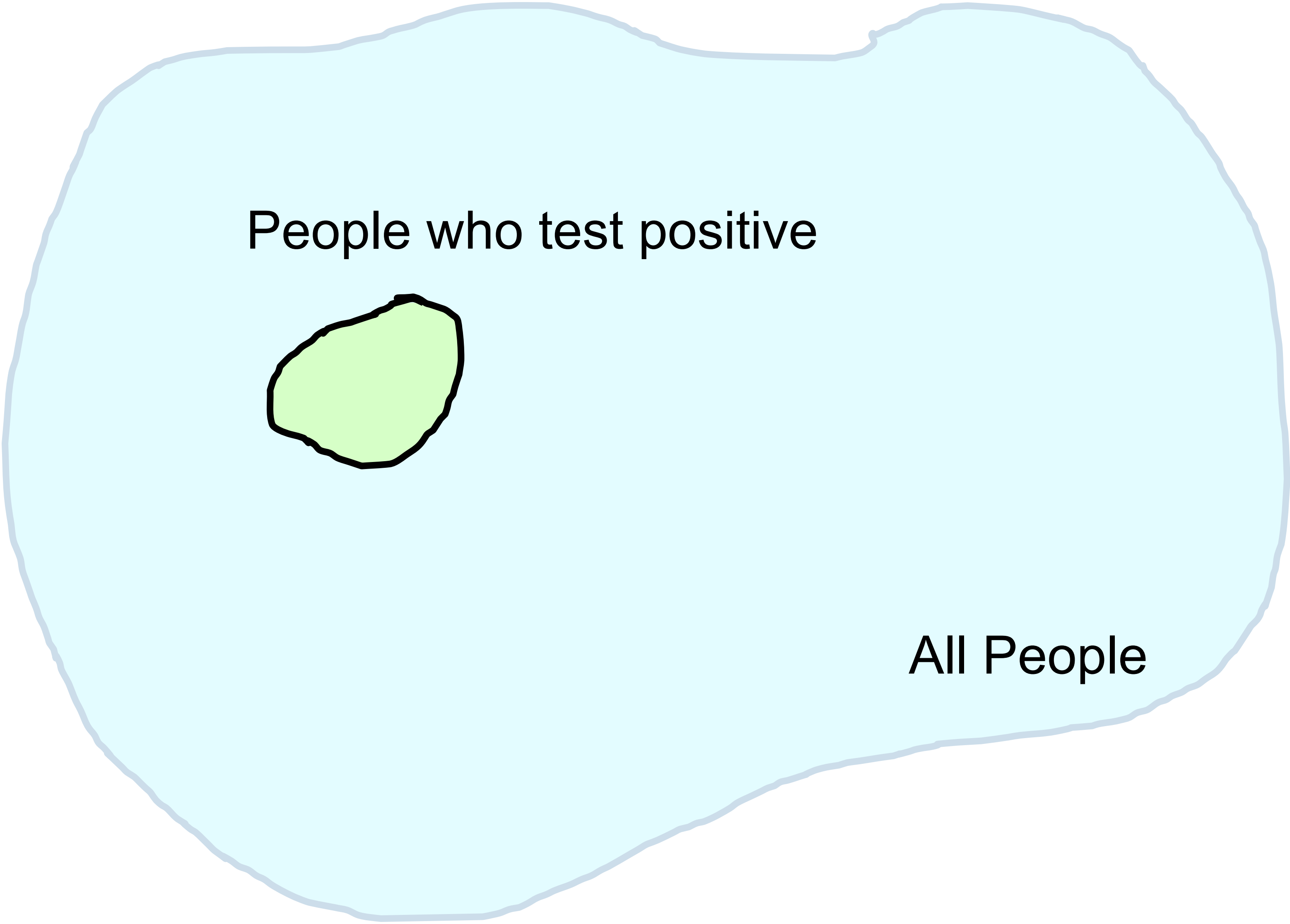
Bayes Thorem Intuition



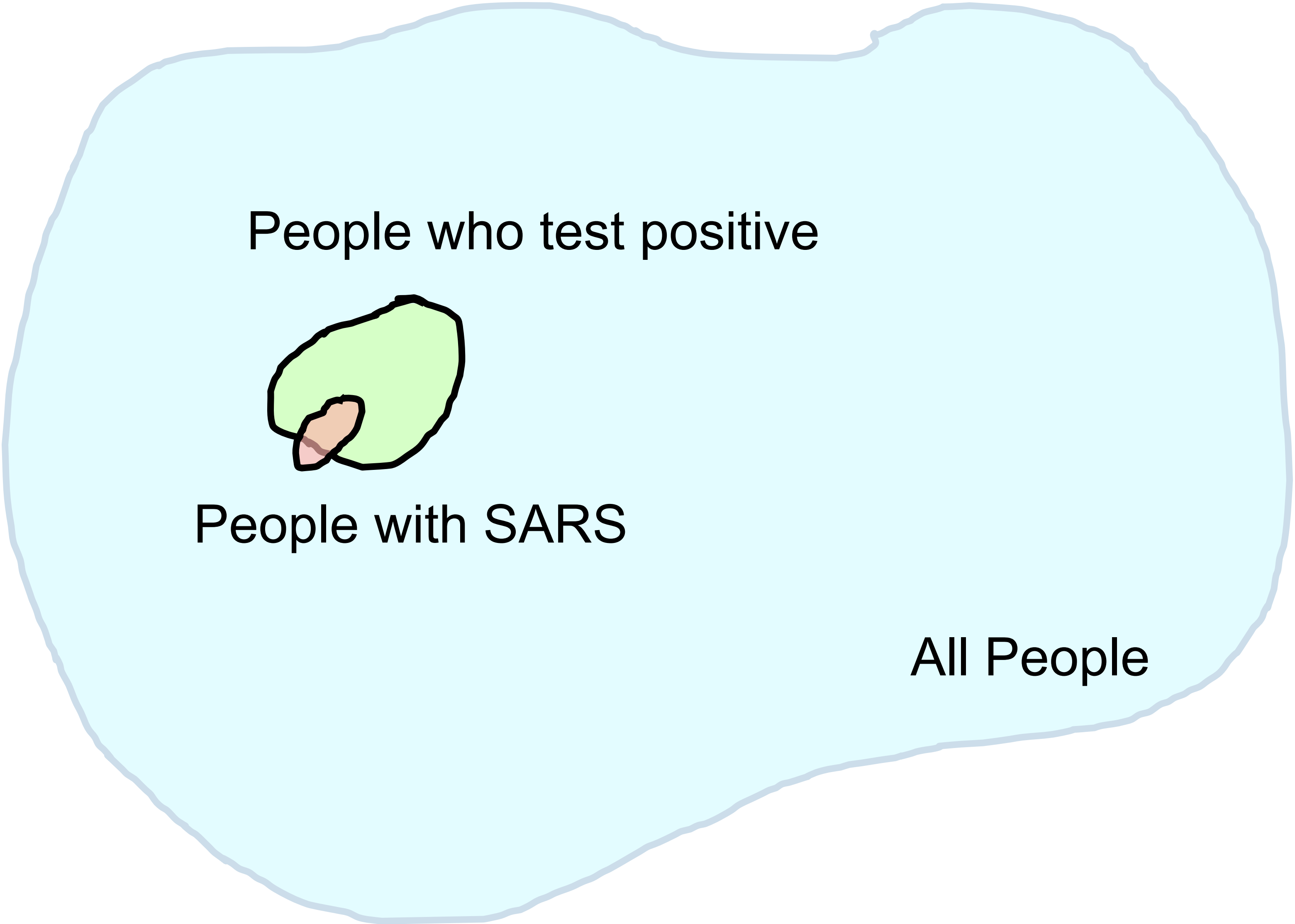
Bayes Thorem Intuition



Bayes Thorem Intuition

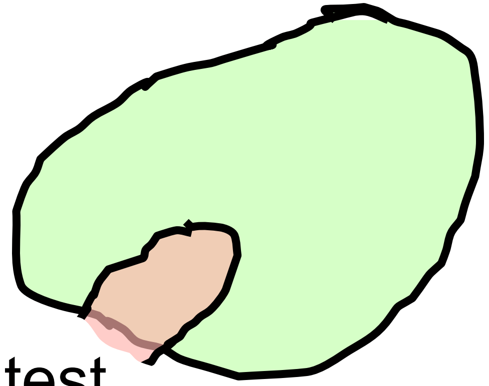


Bayes Thorem Intuition



Bayes Thorem Intuition

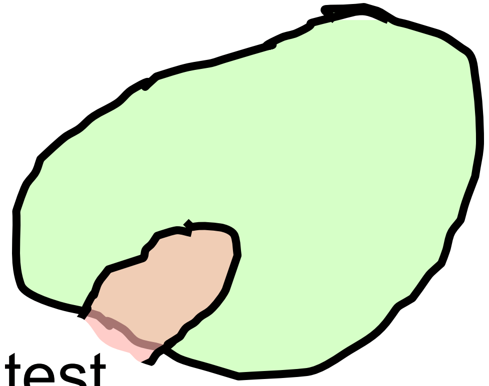
Conditioning on a positive result changes the sample space to this:



≈ 0.330

Bayes Thorem Intuition

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



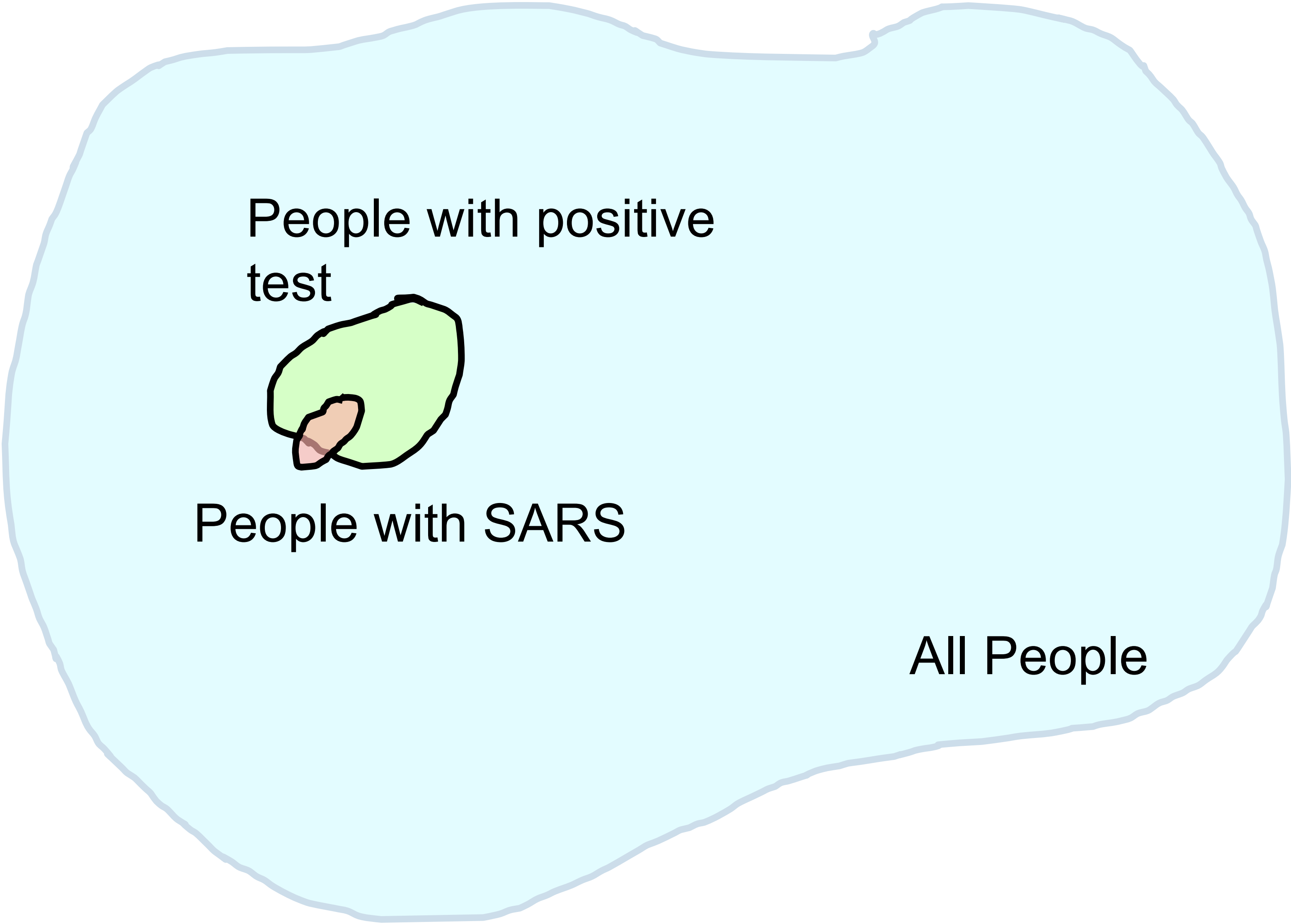
People who test positive

$$P(F)P(E|F) + P(F^c)P(E|F^c)$$

People who test positive and have SARS $P(F)P(E|F)$

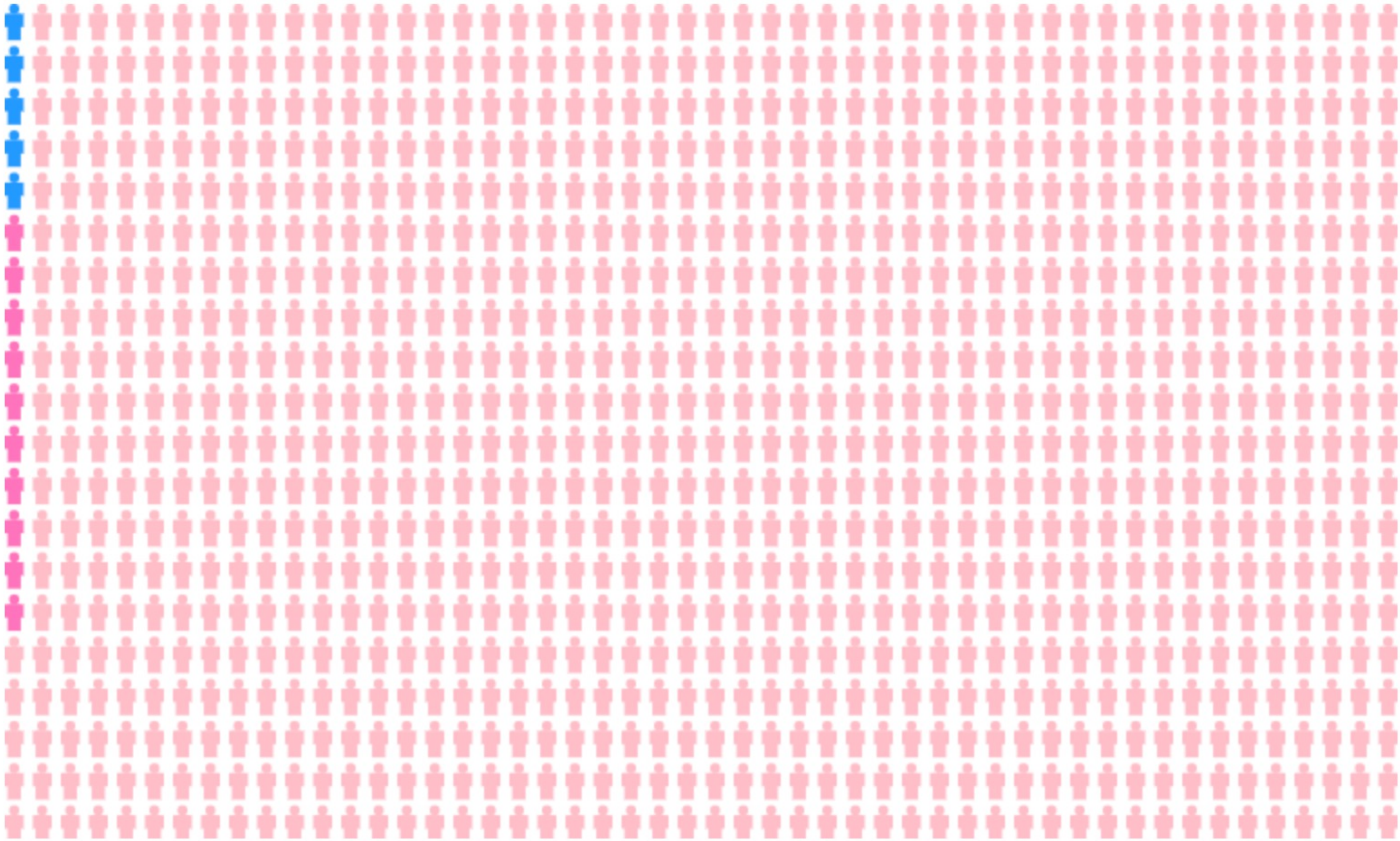
≈ 0.330

Bayes Thorem Intuition



Bayes Thorem 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 Thorem 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 +	$0.98 = P(E \mid F)$	$0.01 = P(E \mid F^c)$
Test –	$0.02 = P(E^c \mid F)$	$0.99 = P(E^c \mid F^c)$

- Let E^c = you test <u>negative</u> for SARS with this test
- Let F = you actually have SARS
- What is $P(F \mid E^c)$?

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

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

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 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 1/10 be the probability that a learner gets the question incorrect given they **do** know the concepts.

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

Come on Wednesday!