05: Conditional Probability

Amrita Kaur

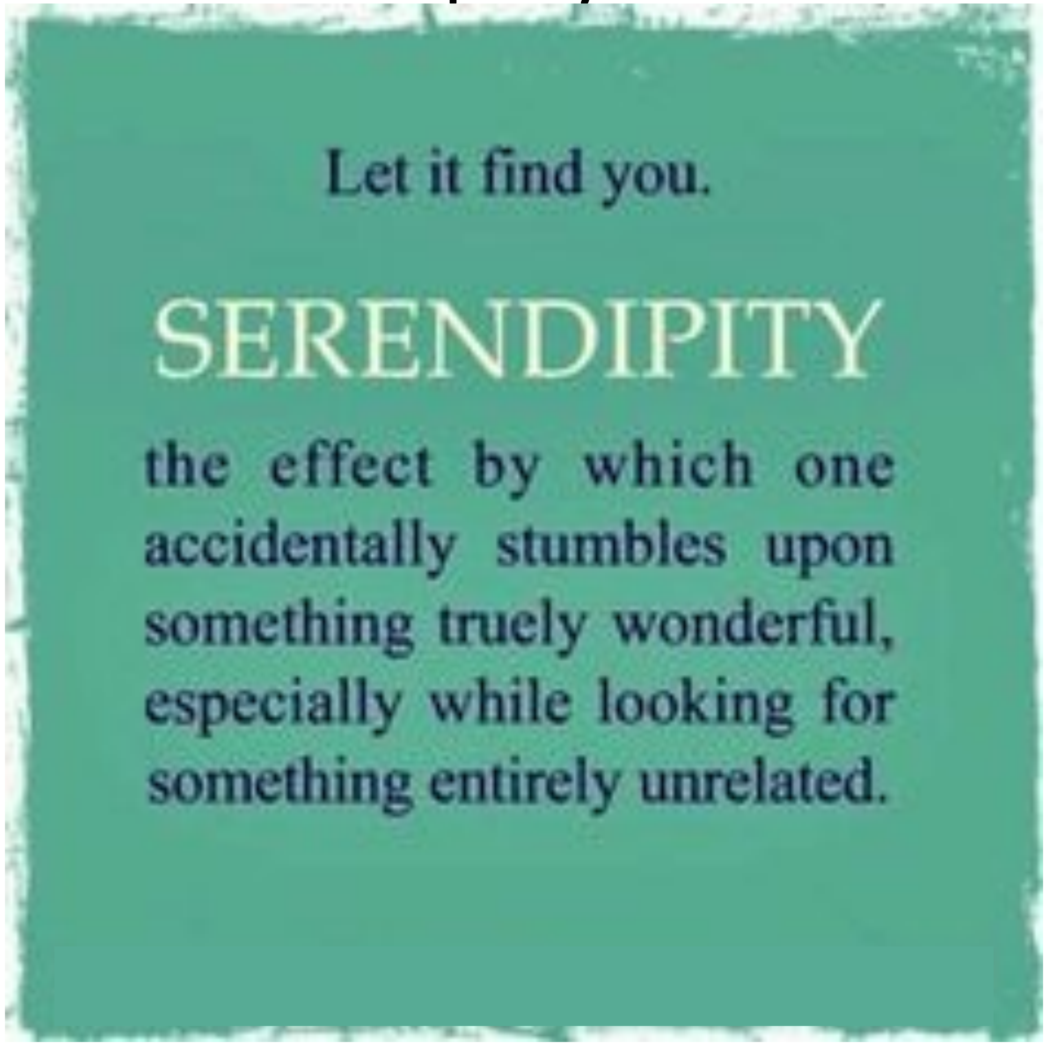
June 29, 2022

Announcements

- Switched over to using QueueMeIn for OH starting today – more info on “Office Hours” tab of website
  - My OH are now in my office: Durand 319
- PSet 2 timeline pushed back by 24 hours!
  - Out: 6/30
  - Due: 7/9

Review!

Serendipity



Serendipity

- The population of Stanford is $n = 17{,}000$ people.
- You are friends with $r = 54$ people.
- Walk into a room, see $k = 250$ random people.
- Assume each group of k Stanford people is equally likely to be in the room.

What is the probability that you see someone you know in the room?

http://web.stanford.edu/class/cs109/demos/serendipity.html

Serendipity

- The population of Stanford is $n = 17{,}000$ people.
- You are friends with $r = 54$ people.
- Walk into a room, see $k = 250$ random people.
- Assume each group of k Stanford people is equally likely to be in the room.

What is the probability that you see at least one friend in the room?

Define

- S (unordered)
- E: ≥ 1 friend in the room

What strategy would you use?

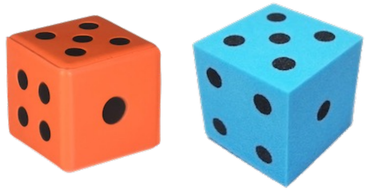
A. $P(\text{exactly } 1) + P(\text{exactly } 2)$ $P(\text{exactly } 3) + \cdots$

B. $1 - P(\text{see no friends})$

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

Conditional Probability

Roll two dice



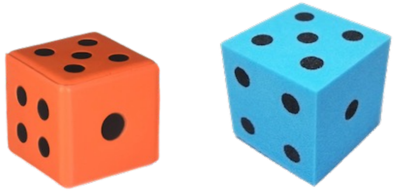
Roll two 6-sided fair dice. What is P(sum = 7)?

$S = \{$ (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)}

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

E = *in red*

Dice, our misunderstood friends



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

You want them to sum to 4.

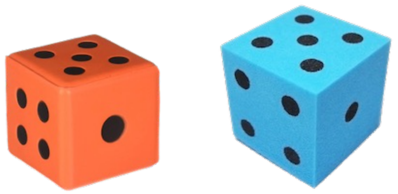
What is the best outcome for D_1?

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/trick question

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$)$?

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

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

$P(E) = 1/6$

Slicing up the spam

24 emails are sent, 6 each to 4 users.

- 10 of the 24 emails are spam.
- All possible outcomes are equally likely.

Let E = user 1 receives 3 spam emails.

What is $P(E)$?

Let F = user 2 receives 6 spam emails.

What is $P(E|F)$?

Let G = user 3 receives 5 spam emails.

What is $P(G|F)$?

Arts and Crafts

On any given day, there is a 50% chance that our artist-in-residence will decide to paint.

The chance that our artist crochets given that she has also painted on that same day is 50%.

What is the probability that, in one day, she decides to paint and crochet?

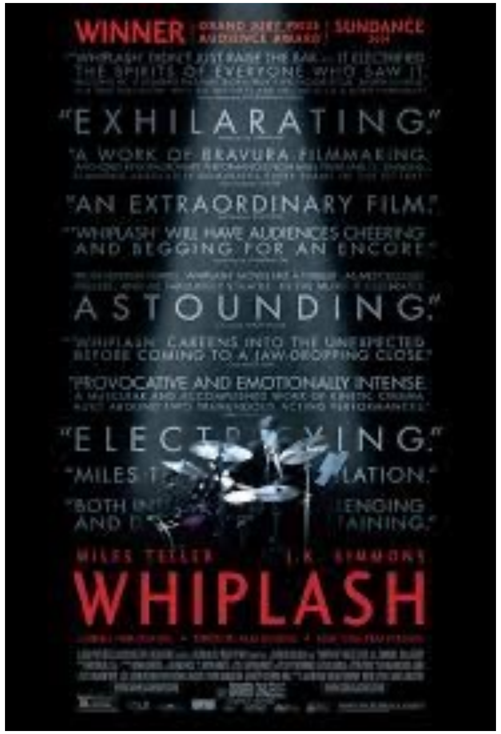
NETFLIX

and Learn

Netflix and Learn

Let E = a user watches Whiplash

What is $P(E)$?



❌ Equally likely outcomes?

$$S = \{\text{watch, not watch}\}$$

$$E = \{\text{watch}\}$$

$$P(E) = 1/2 \text{ ?}$$

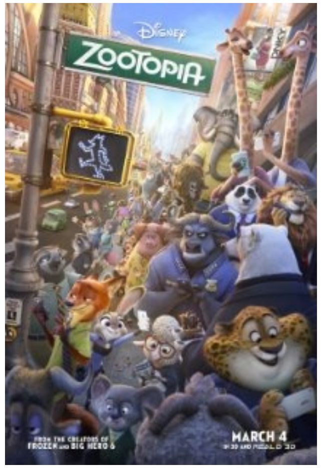
✅ $$P(E) = \lim_{n \to \infty} \frac{n(E)}{n} \approx \frac{\#\text{ people who have watched movie}}{\#\text{ people on Netflix}}$$

$$= 13{,}384{,}625 / \ 66{,}923{,}123 \approx 0.20$$

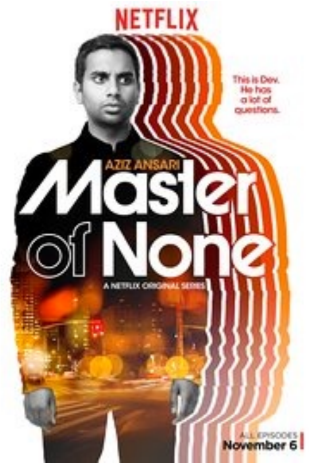
Netflix and Learn

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

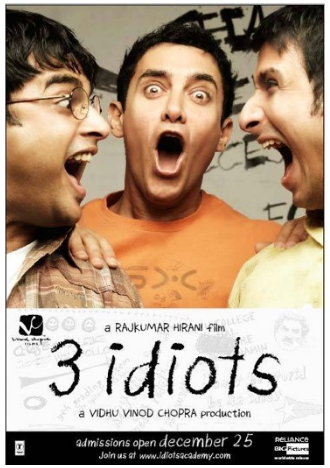
$P(E) = 0.19$



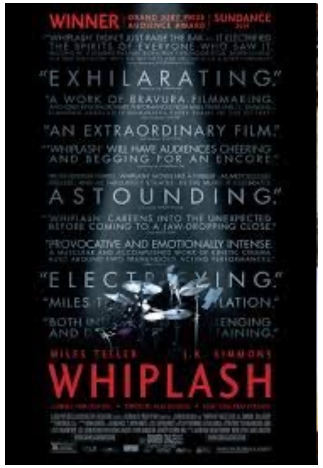
$P(E) = 0.32$



$P(E) = 0.20$



$P(E) = 0.09$

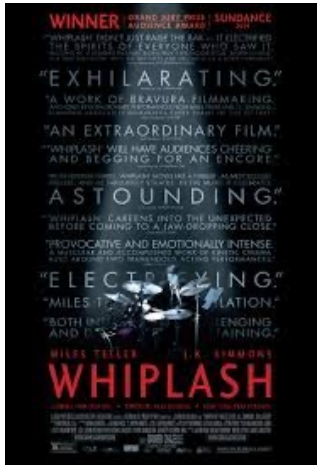


$P(E) = 0.20$

Netflix and Learn

Let E = a user watches Whiplash.
Let F = a user watches CODA.

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



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

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

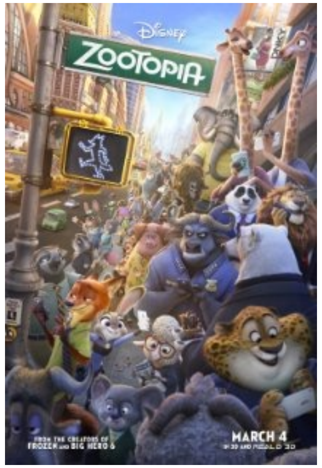
$$\approx 0.42$$

Netflix and Learn

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

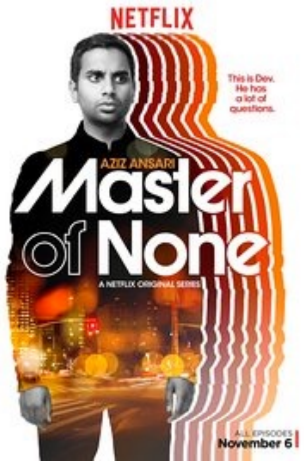
$P(E) = 0.19$

$P(E|F) = 0.14$



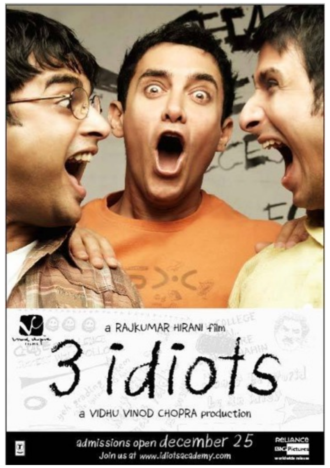
$P(E) = 0.32$

$P(E|F) = 0.35$



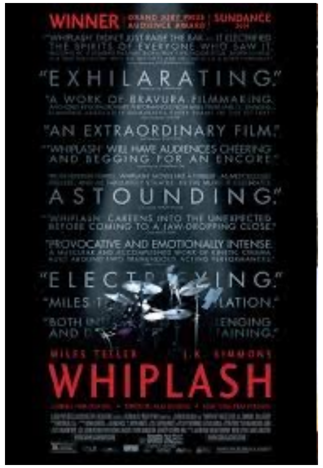
$P(E) = 0.20$

$P(E|F) = 0.20$



$P(E) = 0.09$

$P(E|F) = 0.72$



$P(E) = 0.20$

$P(E|F) = 0.42$

$$P(EF)$$

Chain rule
(Product rule)

Definition of
conditional probability

$$P(E|F)$$

Law of Total
Probability

$$P(E)$$

Practice Problems!

Evolution of Bacteria

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…(oh no)

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?

$$P(EF)$$

Chain rule
(Product rule)

Definition of
conditional probability

$$P(E|F)$$

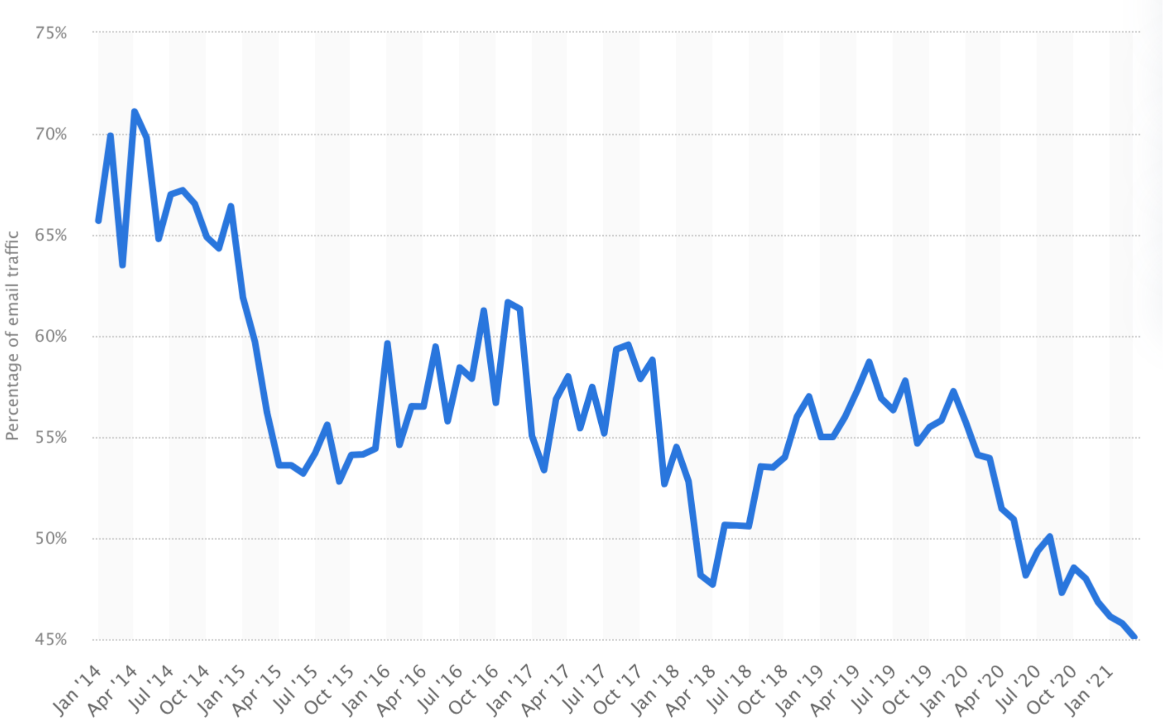
Law of Total
Probability

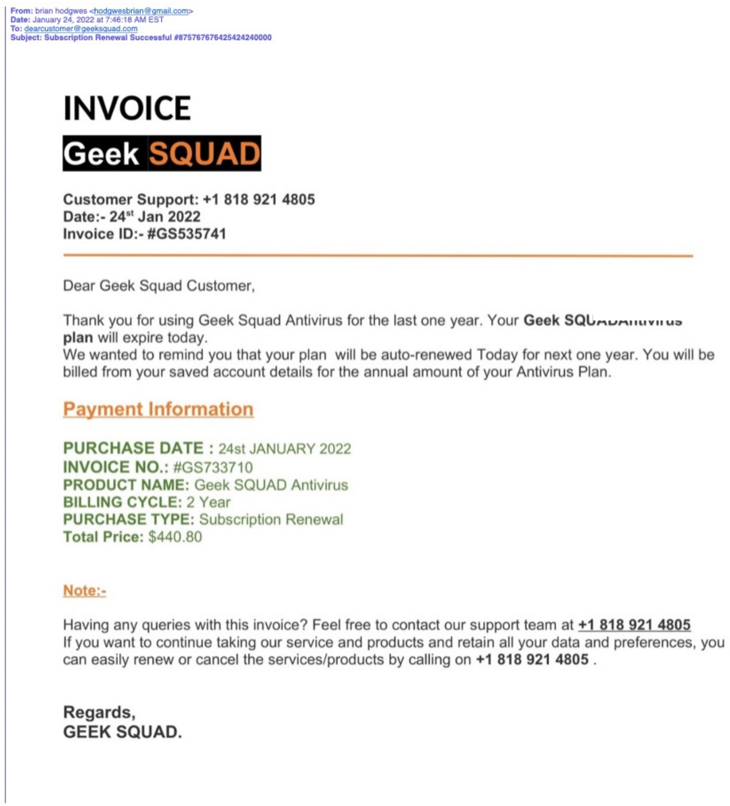
Bayes'
Theorem

$$P(E)$$

$$P(F|E)$$

Detecting spam email





From: brian hodgwes <hodgwesbrian@gmail.com>
Date: January 24, 2022 at 7:46:18 AM EST
To: dearcustomer@geeksquad.com
Subject: Subscription Renewal Successful #875767676425424240000

INVOICE

Geek SQUAD

Customer Support: +1 818 921 4805
Date:- 24st Jan 2022
Invoice ID:- #GS535741

Dear Geek Squad Customer,

Thank you for using Geek Squad Antivirus for the last one year. Your **Geek SQUADAntivirus plan** will expire today.
We wanted to remind you that your plan will be auto-renewed Today for next one year. You will be billed from your saved account details for the annual amount of your Antivirus Plan.

Payment Information

PURCHASE DATE : 24st JANUARY 2022
INVOICE NO.: #GS733710
PRODUCT NAME: Geek SQUAD Antivirus
BILLING CYCLE: 2 Year
PURCHASE TYPE: Subscription Renewal
Total Price: $440.80

Note:-

Having any queries with this invoice? Feel free to contact our support team at **+1 818 921 4805**
If you want to continue taking our service and products and retain all your data and preferences, you can easily renew or cancel the services/products by calling on **+1 818 921 4805** .

Regards,
GEEK SQUAD.

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

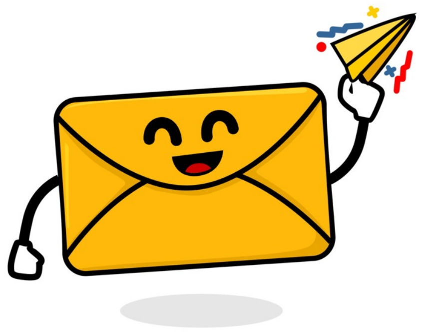
$$P(E|F) = P\left(\text{"Dear"} \,\middle|\, \begin{matrix}\text{Spam}\\ \text{email}\end{matrix}\right)$$

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

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

Detecting spam email

- 45% of all email in 2022 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.

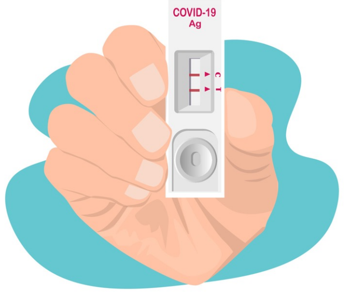
What is the probability that the email is spam?

SARS Testing

- A test is 98% effective at detecting SARS (“true positive”).
    - True positive – probability of positive result given that you have the disease.
- However, the test has a “false positive” rate of 1%.
    - False positive – probability of a positive result given that you don’t have the disease
- 0.5% of the US population has SARS.

Taking tests: Confusion matrix

Take test



Fact, F Has disease
or F^C No disease

Evidence, E Test positive
or E^C Test negative

		Fact	
		F, disease +	F^C, disease –
Evidence	E, Test +	True positive $P(E \mid F)$	False positive $P(E \mid F^C)$
	E^C, Test –	False negative $P(E^C \mid F)$	True negative $P(E^C \mid F^C)$

If a test returns positive, what is the likelihood you have the disease?

SARS Testing

- A test is 98% effective at detecting SARS (“true positive”).
  - True positive – probability of positive result given that you have the disease.
- However, the test has a “false positive” rate of 1%.
  - False positive – probability of a positive result given that you don’t have the disease
- 0.5% of the US population has SARS.

Definitions

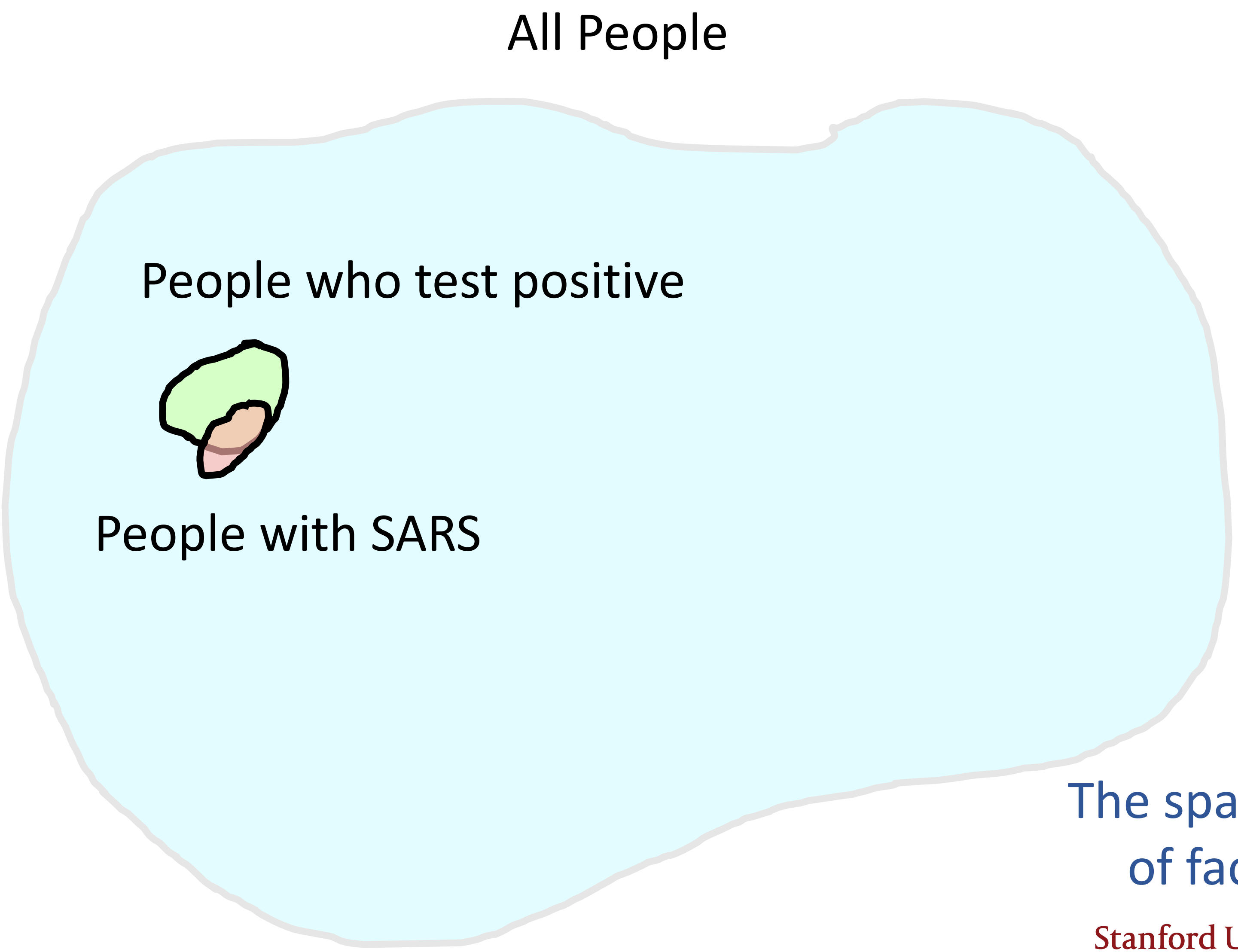
- Let E (evidence) be the event that the test comes back positive
- Let F (fact) be the event that you actually have SARS

What is the chance you have SARS if you test positive?

Bayes' Theorem intuition

Original question:

What is the likelihood you have SARS if you test positive for the disease?



The space of facts

Bayes' Theorem intuition

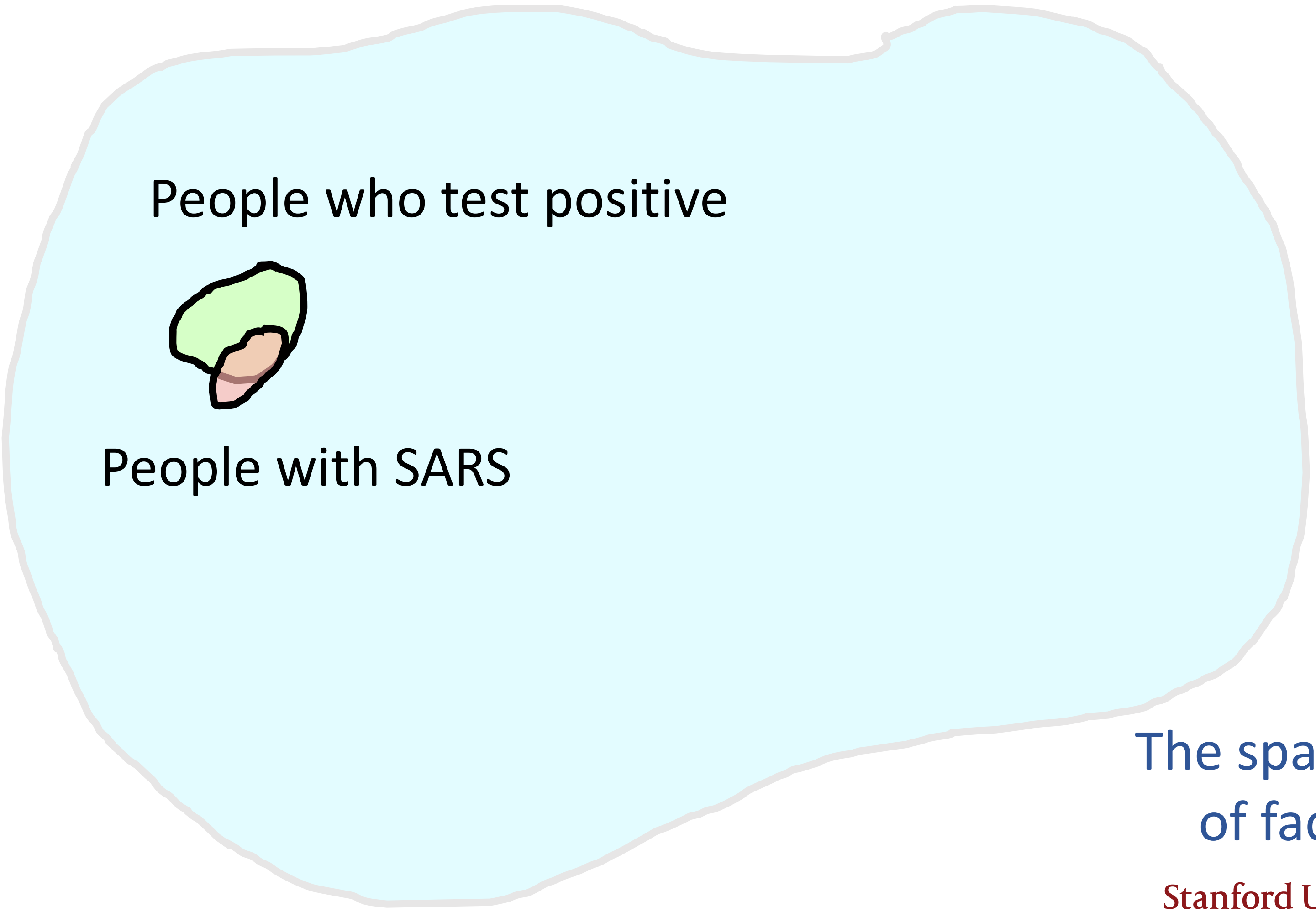
Original question:

What is the likelihood you have SARS if you test positive for the disease?

Interpret

Interpretation:
Of the people who test positive, how many actually have SARS?

All People



The space of facts

Bayes' Theorem intuition

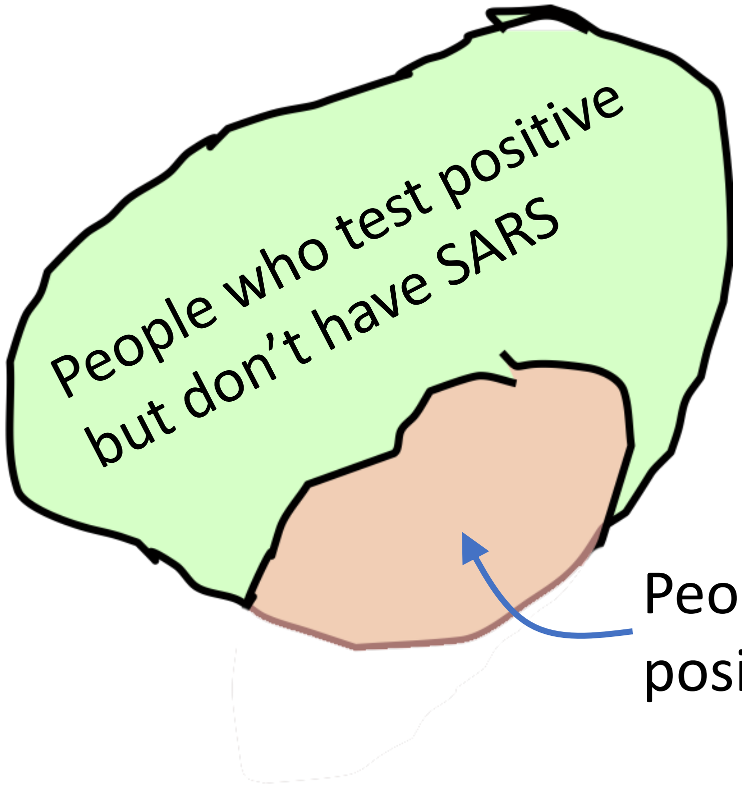
Original question:

What is the likelihood you have SARS if you test positive for the disease?

Interpret

Interpretation:
Of the people who test positive, how many actually have SARS?

People who test positive

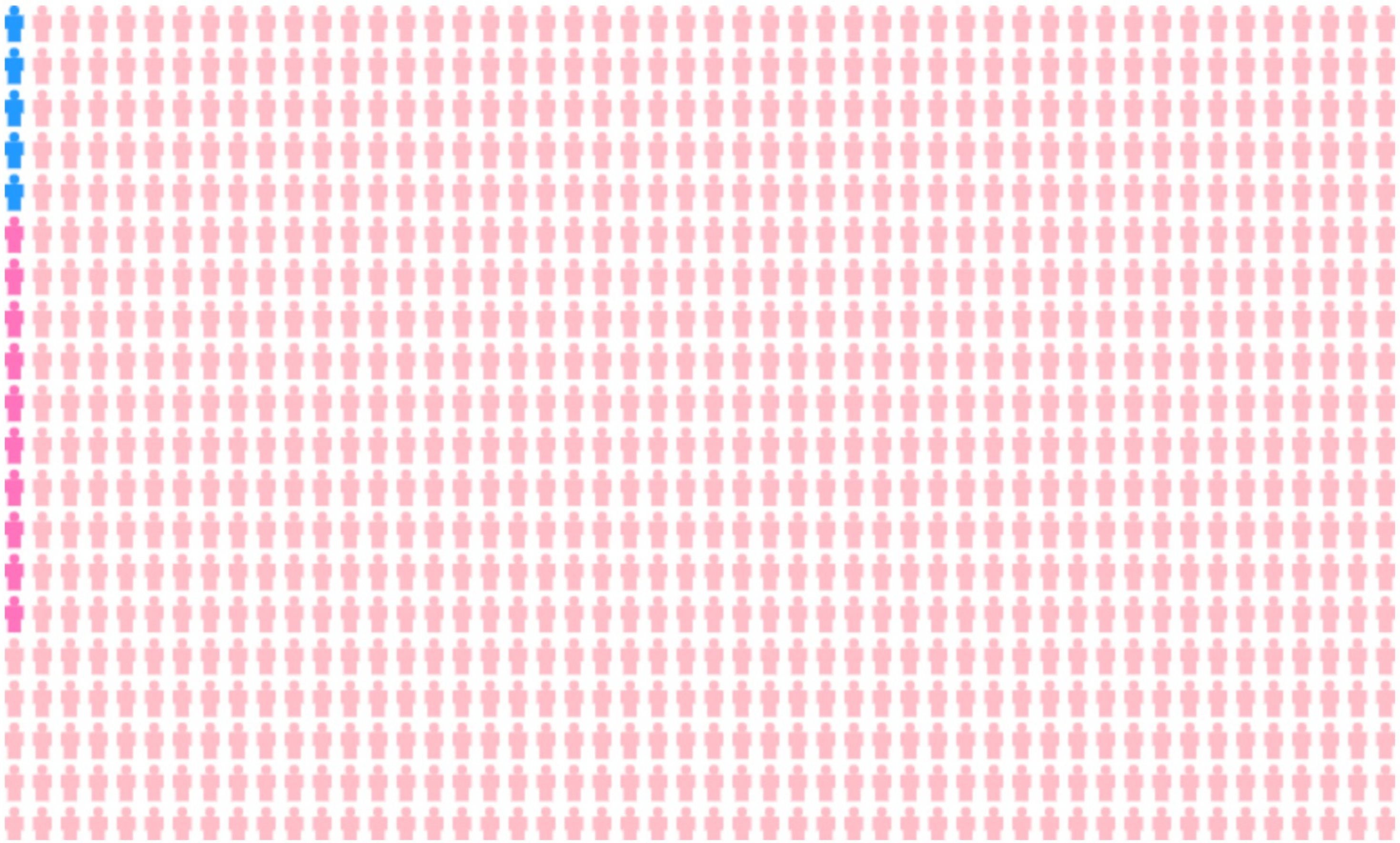


People who test positive and have SARS

The space of facts, **conditioned** on a positive test result

SARS Testing

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.

$$\approx 0.333$$

Demo (class website)

SARS Testing

Conditioning on having a positive result:

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's still good to get tested

Let: E = you test positive
F = you actually have the disease

Let: E^C = you test negative for SARS with this test.

	F, disease +	F^C, disease –
E, Test +	True positive $P(E\|F) = 0.98$	False positive $P(E\|F^C) = 0.01$

What is $P(F|E^C)$?

Why it's still good to get tested

Let: E = you test positive
F = you actually have the disease

Let: E^C = you test negative for SARS with this test.

	F, disease +	F^C, disease –
E, Test +	True positive $P(E\|F) = 0.98$	False positive $P(E\|F^C) = 0.01$
E^C, Test –	False negative $P(E^C\|F) = 0.02$	True negative $P(E^C\|F^C) = 0.99$

What is $P(F|E^C)$?

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

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.

Doors A,B,C

Should we switch?

Note: If we don't switch, P(win) = 1/3 (random)

We are comparing P(win) and P(win|switch).

If we switch

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

1/3 | 1/3 | 1/3

A = prize

- Host opens B or C
- We switch
- We <u>always lose</u>

- P(win | A prize, picked A, switched) = 0

B = prize

- Host must open C
- We switch to B
- We <u>always win</u>

- P(win | B prize, picked A, switched) = 1

C = prize

- Host must open B
- We switch to C
- We <u>always win</u>

- P(win | C prize, picked A, switched) = 1

P(win | picked A, switched) = 1/3 * 0 + 1/3 * 1 + 1/3 * 1 = 2/3

You should switch.

Monty Hall, 1000 envelope version

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

1. You choose 1 envelope.

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

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

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

$$= \text{P(last other envelope has prize)}$$

3. Should you switch?

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

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