04: Conditional Probability and Bayes

Lisa Yan

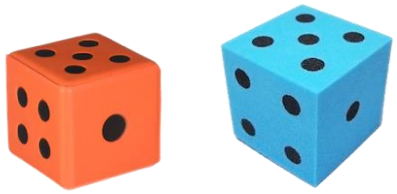
April 13, 2020

Quick slide reference

Conditional Probability

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$)$?

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. $E \cap F$)

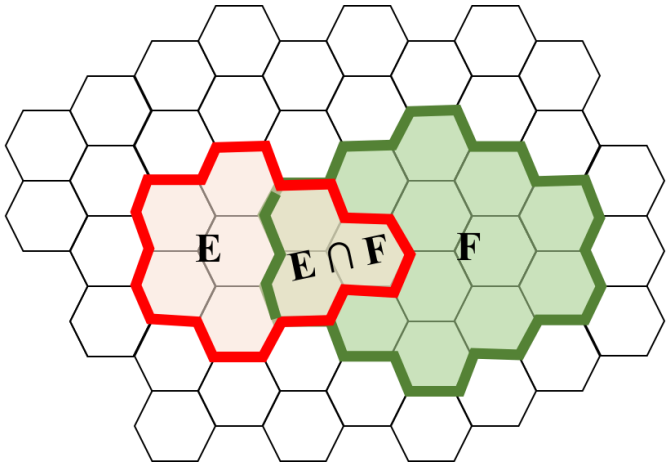
Conditional Probability, equally likely outcomes

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.

With equally likely outcomes:

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

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



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

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

Slicing up the spam

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

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 $P(G|F)$?

Slicing up the spam

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

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 $P(E)$?

$$P(E) = \frac{\binom{10}{3}\binom{14}{3}}{\binom{24}{6}}$$

$$\approx 0.3245$$

Let F = user 2 receives 6 spam emails.

What is $P(E|F)$?

$$P(E|F) = \frac{\binom{4}{3}\binom{14}{3}}{\binom{18}{6}}$$

$$\approx 0.0784$$

Let G = user 3 receives 5 spam emails.

What is $P(G|F)$?

$$P(G|F) \overset{⚠}{=} \frac{\binom{4}{5}\binom{14}{1}}{\binom{18}{6}}$$

$$= 0$$

No way to choose 5 spam from 4 remaining spam emails!

Conditional probability in general

General definition of conditional probability:

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

The Chain Rule (aka Product rule):

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

These properties hold even when outcomes are not equally likely.

NETFLIX

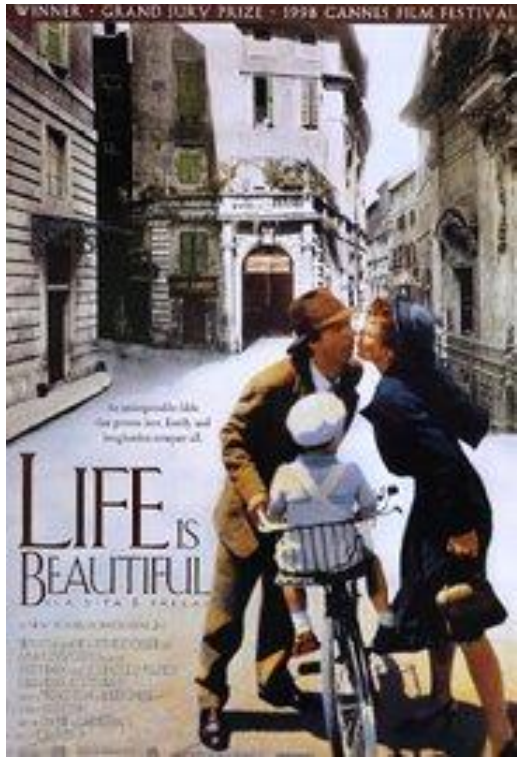
and Learn

Netflix and Learn

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

Let E = a user watches Life is Beautiful.

What is $P(E)$?



✗ Equally likely outcomes?

S = {watch, not watch}

E = {watch}

$P(E) = 1/2$?

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

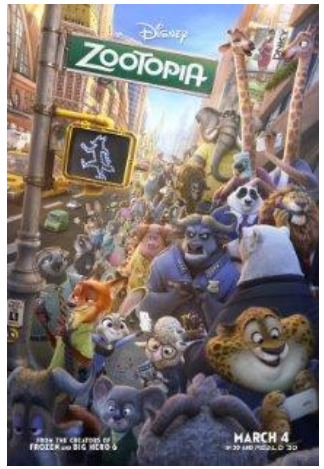
$= 10{,}234{,}231 \,/\, 50{,}923{,}123 \approx 0.20$

Netflix and Learn

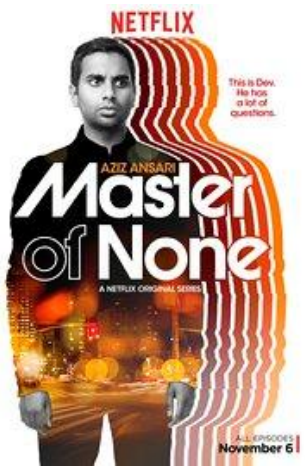
$$P(E|F) = \frac{P(EF)}{P(F)}$$ Definition of Cond. Probability

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

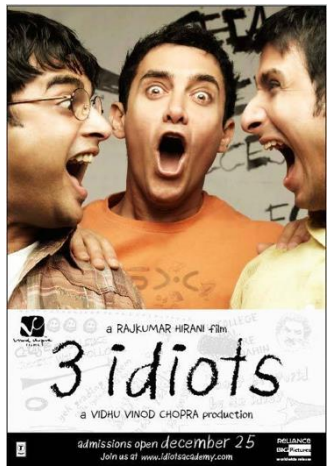
$P(E) = 0.19$



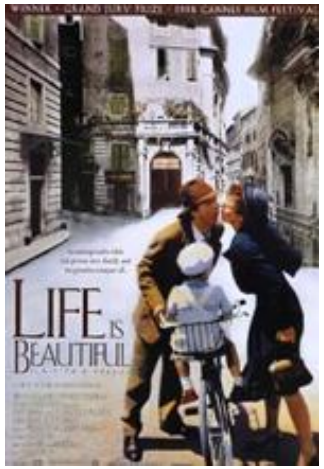
$P(E) = 0.32$



$P(E) = 0.20$



$P(E) = 0.09$



$P(E) = 0.20$

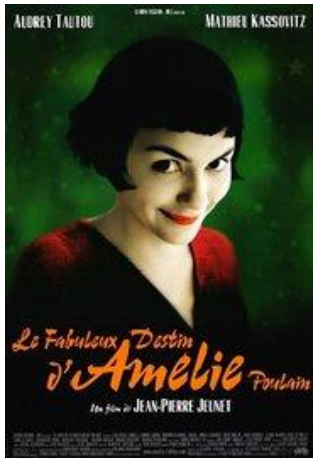
Netflix and Learn

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

Let E = a user watches Life is Beautiful.

Let F = a user watches Amelie.

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



$$P(E|F)$$

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

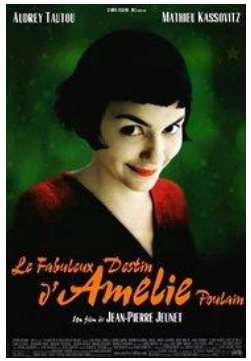
$$= \frac{\text{\# people who have watched both}}{\text{\# people who have watched Amelie}}$$

$$\approx 0.42$$

Netflix and Learn

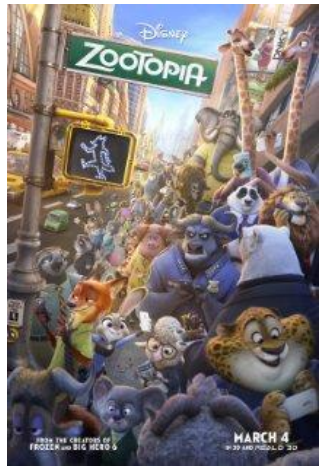
$$P(E|F) = \frac{P(EF)}{P(F)}$$ Definition of Cond. Probability

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



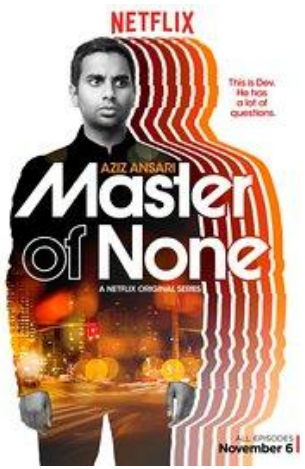
$P(E) = 0.19$

$P(E|F) = 0.14$



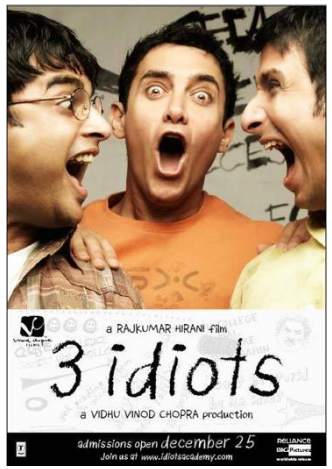
$P(E) = 0.32$

$P(E|F) = 0.35$



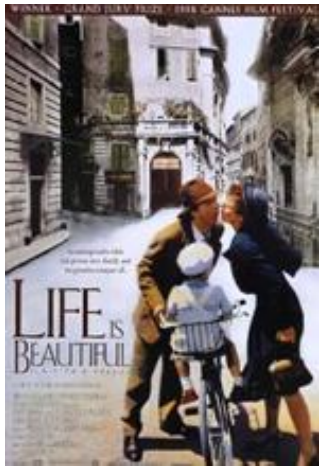
$P(E) = 0.20$

$P(E|F) = 0.20$



$P(E) = 0.09$

$P(E|F) = 0.72$



$P(E) = 0.20$

$P(E|F) = 0.42$

Law of Total Probability

Today's tasks

$$P(EF)$$

Chain rule
(Product rule)

Definition of
conditional probability

$$P(E|F)$$

Law of Total
Probability

$$P(E)$$

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. F and F^C are disjoint s.t. $F \cup F^C = S$	Def. of complement
2. $E = (EF) \cup (EF^C)$	(see diagram)
3. $P(E) = P(EF) + P(EF^C)$	Additivity axiom
4. $P(E) = P(E\|F)P(F) + P(E\|F^C)P(F^C)$	Chain rule (product rule)

Note: disjoint sets by definition are mutually exclusive events

General Law of Total Probability

<u>Thm</u> For mutually exclusive events $F_1, F_2, \ldots, F_n$ s.t. $F_1 \cup F_2 \cup \cdots \cup F_n = S$,

$$P(E) = \sum_{i=1}^{n} P(E|F_i)P(F_i)$$

Finding $P(E)$ from $P(E|F)$

$P(E) = P(E|F)P(F) + P(E|F^C)P(F^C)$ Law of Total Probability

- Flip a fair coin.
- If heads: roll a fair 6-sided die.
- Else: roll a fair 3-sided die.

You win if you roll a 6. What is P(winning)?

Finding $P(E)$ from $P(E|F)$

$P(E) = P(E|F)P(F) + P(E|F^C)P(F^C)$ Law of Total Probability

- Flip a fair coin.
- If heads: roll a fair 6-sided die.
- Else: roll a fair 3-sided die.

You win if you roll a 6. What is P(winning)?

1. Define events & state goal

Let: E: win, F: flip heads

Want: $P(\text{win})$

$= P(E)$

2. Identify <u>known</u> probabilities

$P(\text{win}|\text{H}) = P(E|F) = 1/6$

$P(\text{H}) = P(F) = 1/2$

$P(\text{win}|\text{T}) = P(E|F^C) = 0$

$P(\text{T}) = P(F^C) = 1 - 1/2$

3. Solve

$$P(E) = (1/6)(1/2) + (0)(1/2) = \frac{1}{12} \approx 0.083$$

Finding $P(E)$ from $P(E|F)$, an understanding

- Flip a fair coin.
- If heads: roll a fair 6-sided die.
- Else: roll a fair 3-sided die.

You win if you roll a 6. What is P(winning)?

1. Define events & state goal

Let: E: win, F: flip heads

Want: $P(\text{win})$

$= P(E)$

"Probability trees" can help connect your understanding of the experiment with the problem statement.

Bayes' Theorem I

Today's tasks

$$P(EF)$$

Chain rule
(Product rule)

Definition of
conditional probability

$$P(E|F)$$

Law of Total
Probability

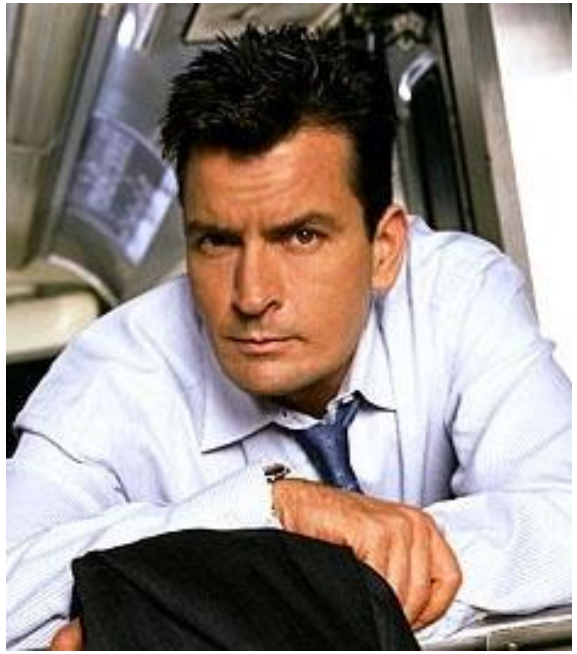
Bayes'
Theorem

$$P(E)$$

$$P(F|E)$$

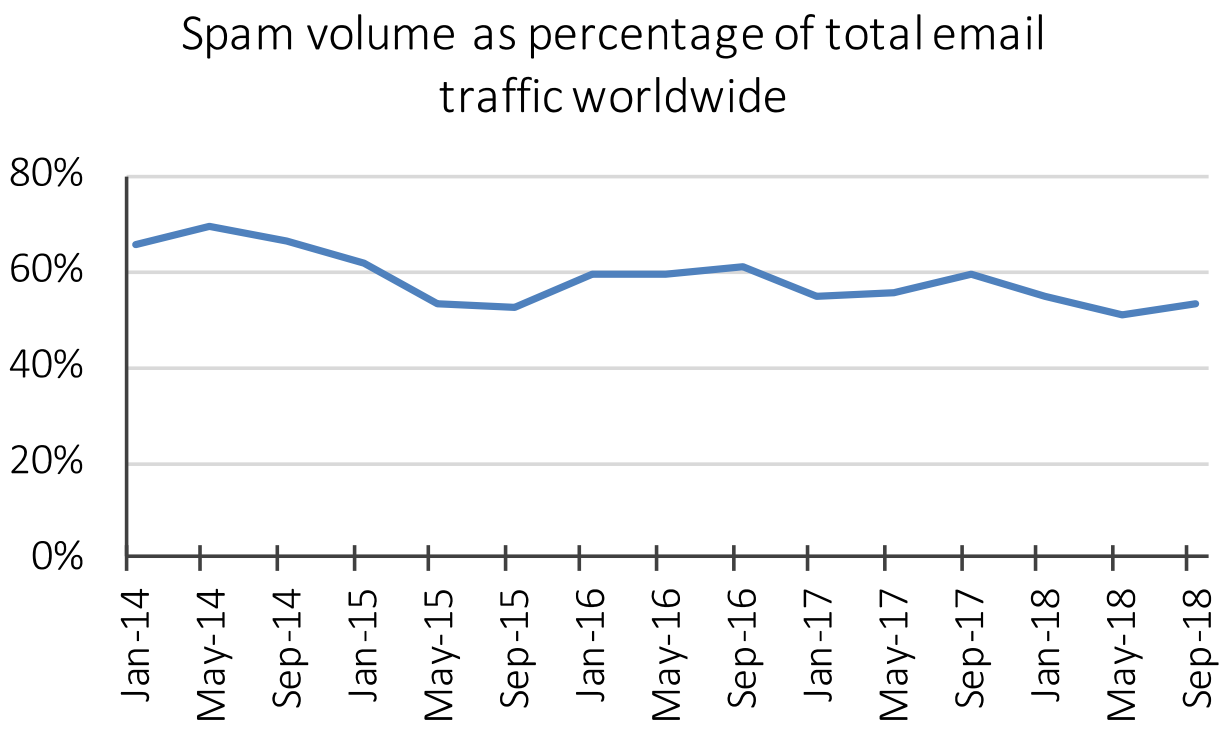
Thomas Bayes

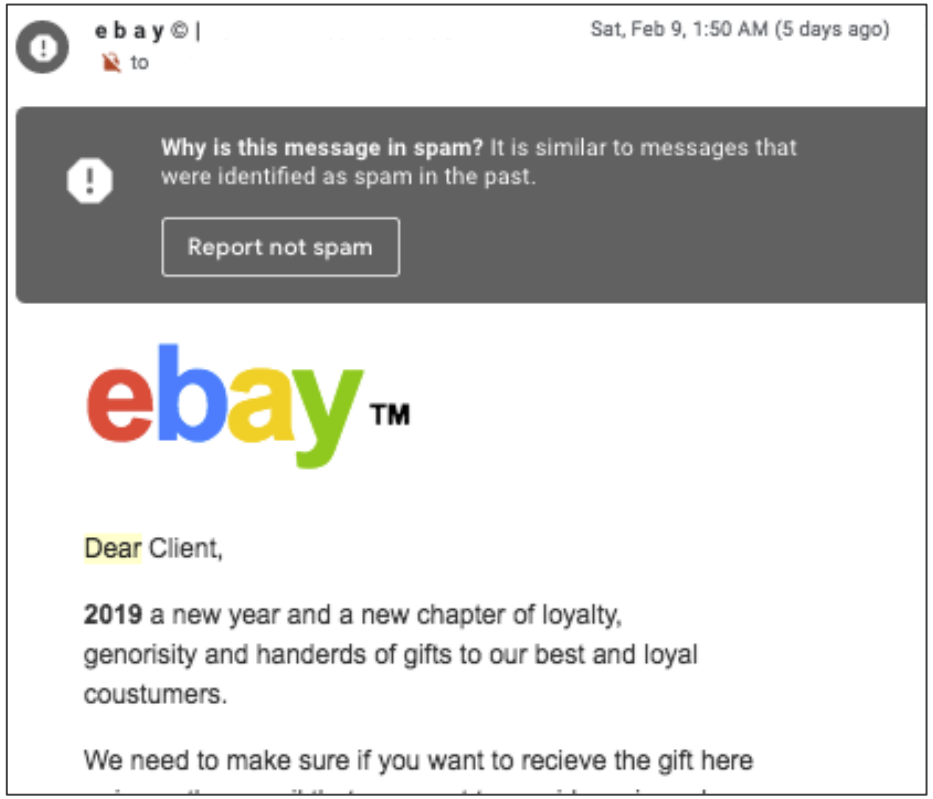
Rev. Thomas Bayes (~1701-1761):
British mathematician and Presbyterian minister



He looked remarkably similar to Charlie Sheen
(but that's not important right now)

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)

Bayes' Theorem

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

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

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

<u>Proof</u>

2 steps! See board

<u>Expanded form</u>:

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

<u>Proof</u>

1 more step! See board

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 <u>known</u> probabilities
3. Solve

Let: E: "Dear", F: spam

Want: $P(\text{spam} \mid \text{"Dear"})$

$= P(F|E)$

Detecting spam email, an understanding

- 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

Let: E: "Dear", F: spam

Want: $P(\text{spam} \mid \text{"Dear"})$

$= P(F|E)$

Note: You should still know how to use Bayes/ Law of Total Probab., but drawing a probability tree can help you identify which probabilities you have. The branches are determined using the problem setup.

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

(live)

04: Conditional Probability and Bayes

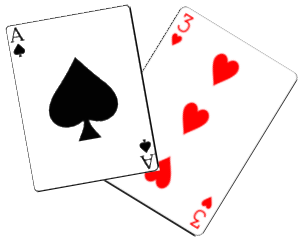
Slides by Lisa Yan

June 29, 2020

This class going forward

Last week

Equally likely events



$P(E \cap F)$ $P(E \cup F)$

(counting, combinatorics)

Today and for most of this course

Not equally likely events

$P(E = \text{Evidence} \mid F = \text{Fact})$

(collected from data)

Bayes'

$P(F = \text{Fact} \mid E = \text{Evidence})$

(categorize a new datapoint)

Conditional probability in general

Review

General definition of conditional probability:

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

The Chain Rule (aka Product rule):

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

These properties hold even when outcomes are not equally likely.

Think, then Breakout Rooms

Then check out the question on the next slide (Slide 35). Post any clarifications here!

https://us.edstem.org/courses/667/discussion/83250

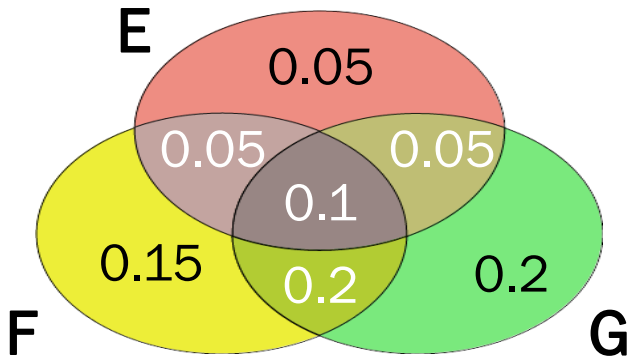
Think by yourself: 1 min

Breakout rooms: 4 min. Introduce yourself!

Think, then groups

You have a flowering plant.

Let E = Flowers bloom
F = Plant was watered
G = Plant got sun



1. How would you write
   i. the probability that the plant got sun, given that it was watered and flowers bloomed?
   ii. the probability that the plant got sun and flowers bloomed given that it was watered?
2. Using the Venn diagram, compute the above probabilities.
3. Chain Rule: Fill in the blanks.
   i. $P(GE) = \_\_\_\_\_ \cdot P(E)$
   ii. $P(GE|F) = P(G|EF) \cdot \_\_\_\_\_$

LIVE

Bayes' Theorem II

Why is Bayes' so important?

It links belief to evidence in probability!

Bayes' Theorem

Review

posterior likelihood prior

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

F: Fact

E: Evidence

Mathematically:

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

Real-life application:

Given new evidence E, update belief of fact F
Prior belief → Posterior belief
$P(F) \rightarrow P(F|E)$

Zika, an autoimmune disease

A disease spread through mosquito bites.
Usually no symptoms; worst case paralysis. During pregnancy: may cause birth defects

Ziika Forest, Uganda

Rhesus monkeys

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

Taking tests: Confusion matrix

Fact, F Has disease
or F^C No disease

Take test

Evidence, E Test positive
or E^C Test negative

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

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

Taking tests: Confusion matrix

Fact, F Has disease
or F^C No disease

Take test

Evidence, E Test positive
or E^C Test negative

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

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

Breakout Rooms

Check out the question on the next slide (Slide 43). Post any clarifications here!

https://us.edstem.org/courses/667/discussion/83250

Breakout rooms: 5 minutes

Zika Testing

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

- A test is 98% effective at detecting Zika ("true positive").
- However, the test has a "false positive" rate of 1%.
- 0.5% of the US population has Zika.

What is the likelihood you have Zika if you test positive?

Why would you expect this number?

1. Define events & state goal

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

Want:
P(disease | test+)
$= P(F|E)$

Zika Testing

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

- A test is 98% effective at detecting Zika ("true positive").
- However, the test has a "false positive" rate of 1%.
- 0.5% of the US population has Zika.

What is the likelihood you have Zika if you test positive?

Why would you expect this number?

1. Define events & state goal
2. Identify <u>known</u> probabilities
3. Solve

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

Want:
P(disease | test+)
$= P(F|E)$

Bayes' Theorem intuition

<u>Original question</u>:

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

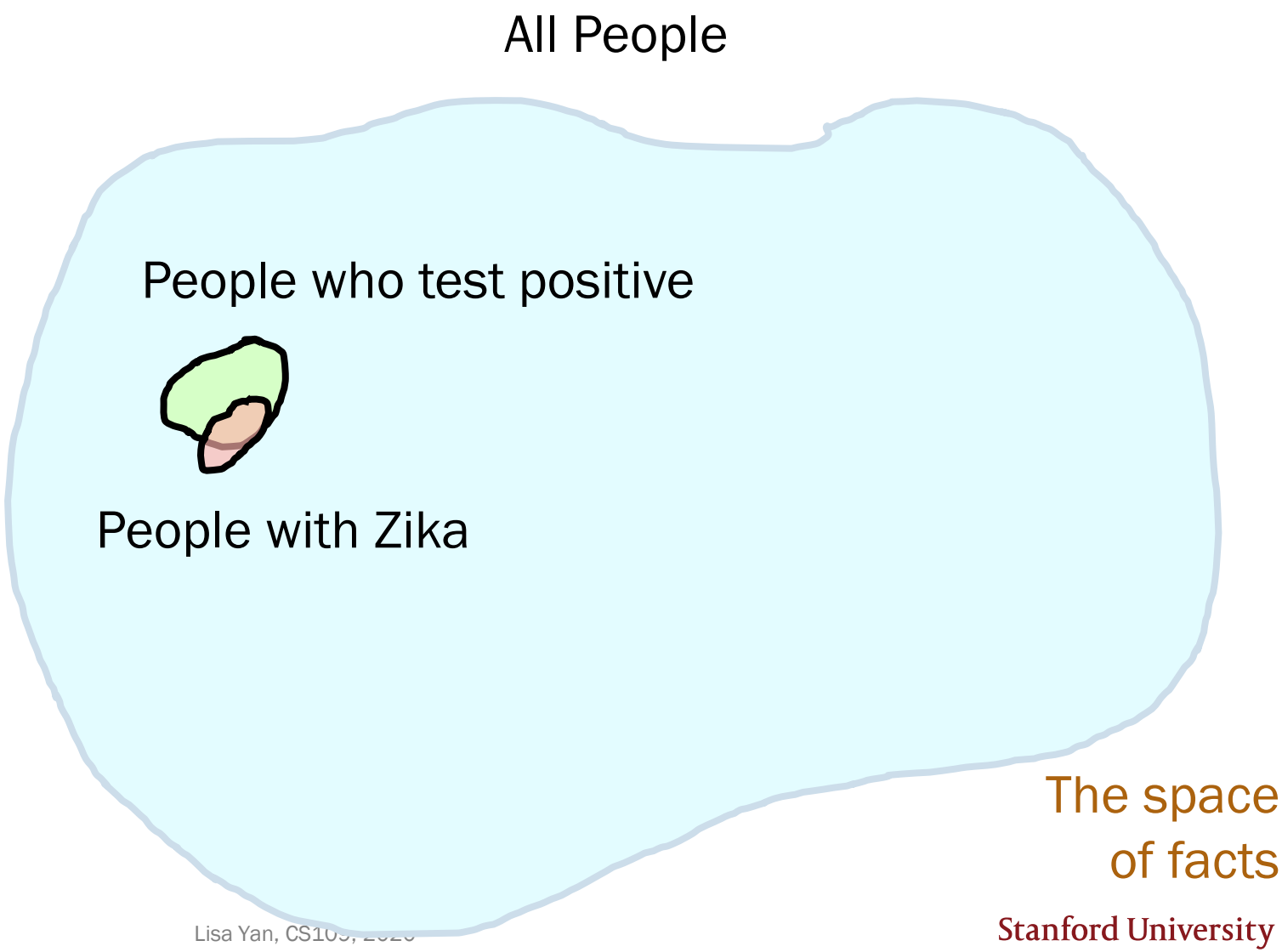
Bayes' Theorem intuition

Original question:

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

Interpret

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



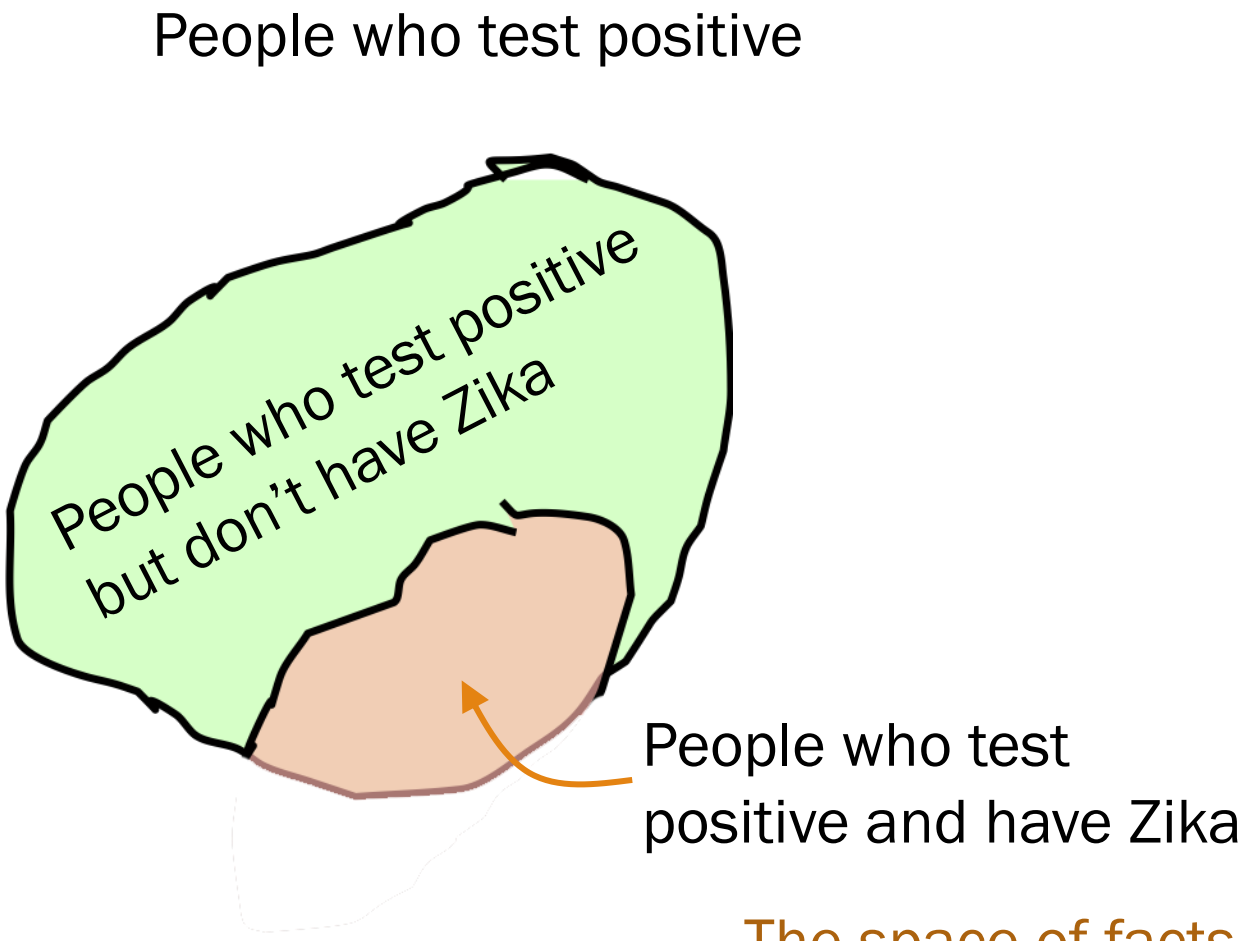
Bayes' Theorem intuition

Original question:

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

Interpret

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



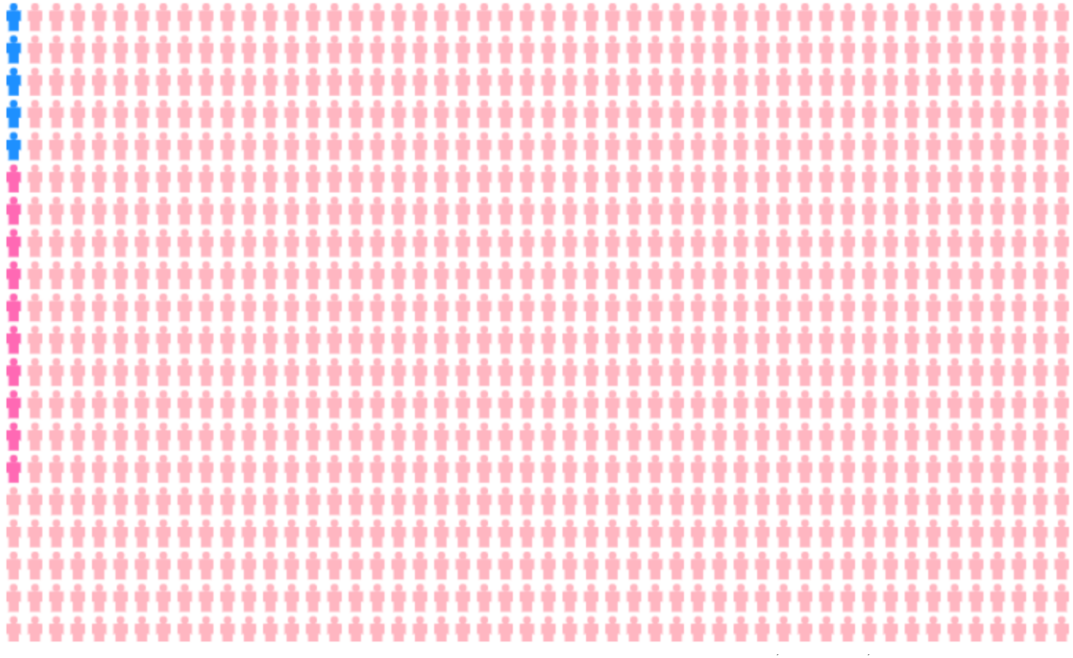
The space of facts, **conditioned** on a positive test result

Zika Testing

- A test is 98% effective at detecting Zika ("true positive").
- However, the test has a "false positive" rate of 1%.
- 0.5% of the US population has Zika.

What is the likelihood you have Zika if you test positive?

Say we have 1000 people:



5 have Zika
and test positive
985 **do not** have Zika
and test negative.
10 **do not** have Zika
and test positive.

≈ 0.333

Demo (class website)

Update your beliefs with Bayes' Theorem

E = you test positive for Zika
F = you actually have the disease

I have a 0.5% chance of having Zika.

Take test, results positive

With these test results, I now have a 33% chance of having Zika!!!

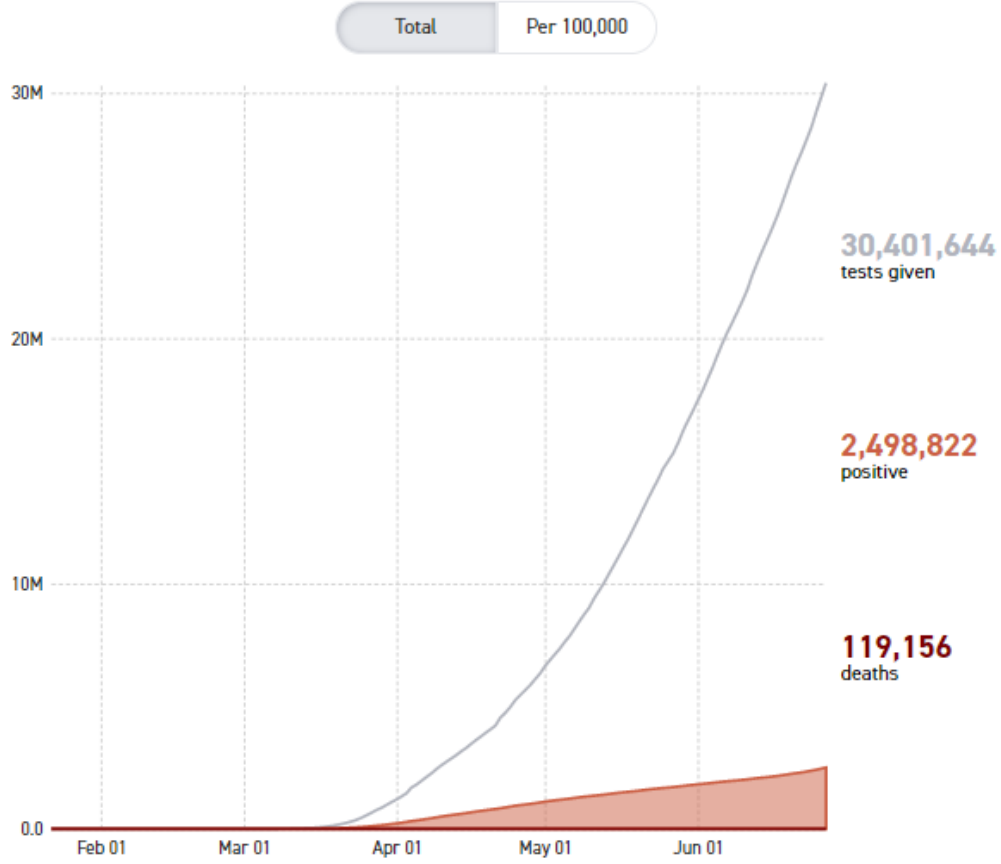
$P(F)$

$P(F|E)$

Interlude for fun/announcements

Topical probability news: Bayes for COVID-19 testing

By **BEATRICE JIN** | 03/16/2020 8:10 PM EDT | Updated 06/28/2020 12:25 PM EDT



	Deaths	Positive tests	Total tests
New York	24,830	391,923	3,754,579
Kentucky	553	14,859	350,754

How representative are today's testing rates?

How do we know if a positive test is a true positive or a false positive?

Why test if there are errors?

https://www.politico.com/interactives/2020/coronavirus-testing-by-state-chart-of-new-cases/

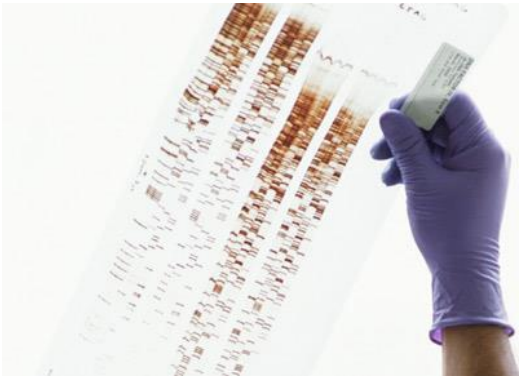
Ethics in Probability: Bayes and the Prosecutor's Fallacy

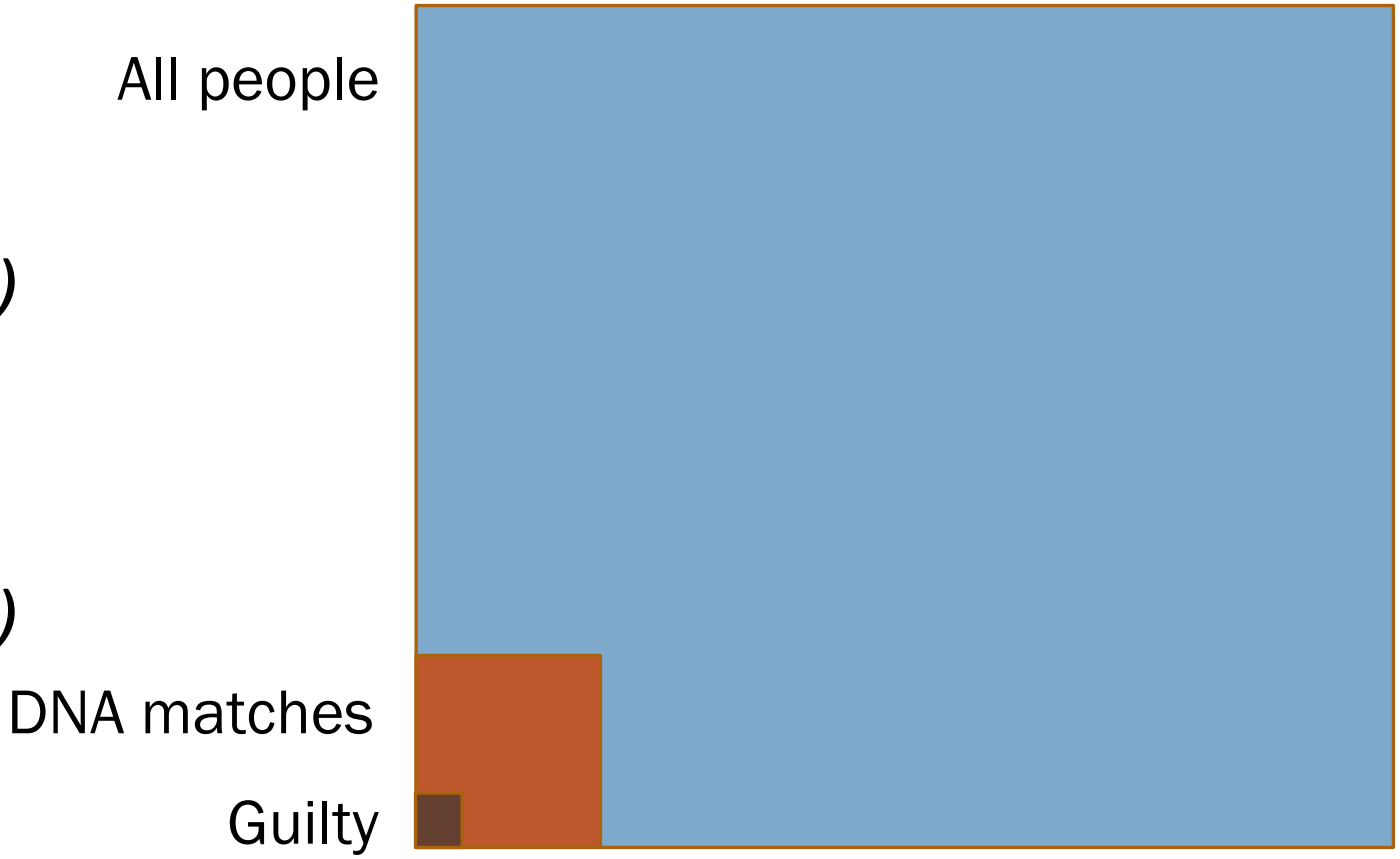
Confusing two probabilities:

P(Innocent | Evidence)

versus

P(Evidence | Innocent)





https://www.cebm.net/2018/07/the-prosecutors-fallacy/

Think

Slide 57 is a question to think over by yourself.

We'll go over it together afterwards.

Post any clarifications here!

https://us.edstem.org/courses/667/discussion/83250

Think by yourself: 2 minutes

(by yourself)

Why it's still good to get tested

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

- A test is 98% effective at detecting Zika ("true positive").
- However, the test has a "false positive" rate of 1%.
- 0.5% of the US population has Zika.

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

Let: E^C = you test negative for Zika 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 $P(F|E^C)$?

(by yourself)

Why it's still good to get tested

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

- A test is 98% effective at detecting Zika ("true positive").
- However, the test has a "false positive" rate of 1%.
- 0.5% of the US population has Zika.

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

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

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

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

Why it's still good to get tested

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

- A test is 98% effective at detecting Zika ("true positive").
- However, the test has a "false positive" rate of 1%.
- 0.5% of the US population has Zika.

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

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

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

	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 $P(E^C\|F^C) = 0.99$

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

Why it's still good to get tested

E = you test positive for Zika
F = you actually have the disease
E^C = you test negative for Zika

I have a **0.5%** chance of having Zika disease.

$P(F)$

Take test, results positive

With these test results, I now have a **33%** chance of having Zika!!!

$P(F|E)$

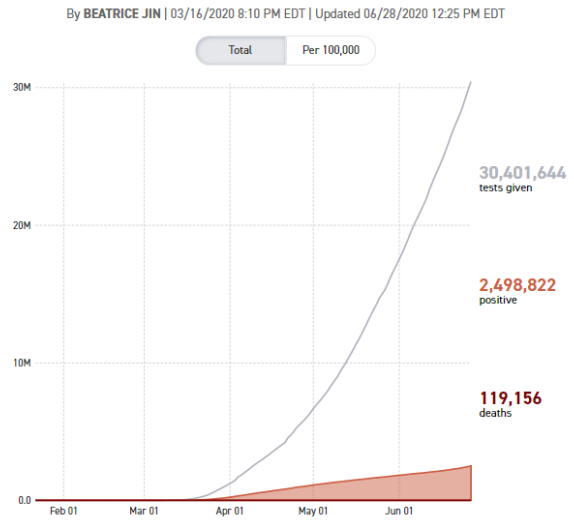
Take test, results negative

With these test results, I now have a **0.01%** chance of having Zika disease!!!

$P(F|E^C)$

Topical probability news: Bayes for COVID-19 testing

- Antibody tests (blood samples) have higher false negative, false positive rates than RT-PCR tests (nasal swab). However, they help explain/identify our body's reaction to the virus.
- The real world has many more "**givens**" (current symptoms, existing medical conditions) that improve our belief **prior** to testing.
- Most importantly, testing gives us a noisy signal of the spread of a disease.



How representative are today's testing rates?

How do we know if a positive test is a true positive or a false positive?

Why test if there are errors?

LIVE

Monty Hall Problem

Monty Hall Problem

and Wayne Brady

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: P(win|no switch) = 1/3 (random)

We are comparing P(win|no switch) and P(win|switch).

(Vote in the chat! ☺)

(by yourself)

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 door version

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

1. You choose 1 door.

$$\frac{1}{1000} = \text{P(door is prize)}$$

$$\frac{999}{1000} = \text{P(other 999 doors have prize)}$$

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

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

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

3. Should you switch?

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

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

LIVE

Next Time: Independence!