Table of Contents

04: Conditional Probability and Bayes

Jerry Cain

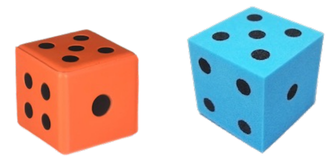
January 17th, 2024

Lecture Discussion on Ed

Conditional Probability

Dice, our misunderstood friends

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



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

What is $P(E)$?

$|D_1| = 6$
$|D_2| = 6$
$|S| = |D_1| \cdot |D_2| = 36$

$|S| = 36$

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

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

Let F be event: $D_1 = 2$.

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

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

$E' = \{(2,2)\}$ when F is the "new" sample space

$P(E,$ knowing F already occurred$) = \boxed{1/6}$

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:	"$P(E,$ knowing 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.

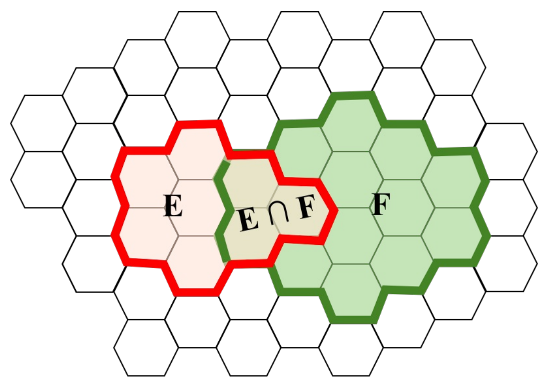
$|E| = 8$ $|S| = 50$

$|F| = 14$ $|E \cap F| = 3$

With **equally likely outcomes**:

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

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



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

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

Slicing up the spam

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

assume emails are distinct, but order doesn't matter.

Let E = user 1 receives 3 spam emails.

What is $P(E)$?

$|E| = \binom{10}{3}\binom{14}{3}$

$|S| = \binom{24}{6}$

Let F = user 2 receives 6 spam emails.

What is $P(E|F)$?

knowing F happened, only 4 spam emails available to user 1, but all 14 non-spam still available

Let G = user 3 receives 5 spam emails.

What is $P(G|F)$?

user 3 has difficult time selecting 5 spam emails when only 4 are left! :)

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 $P(G|F)$?

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

$$= 0$$

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

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?

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

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

$P(E) = 1/2$?

data-driven estimate of what true probability is.

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

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

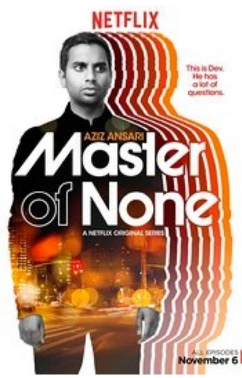
Netflix and Learn

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

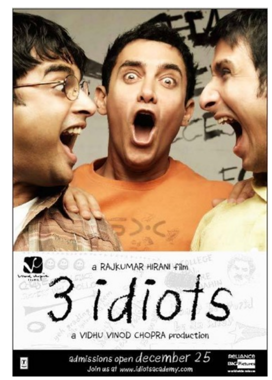
Let E be the event that a user watches the given movie.

$P(E) = 0.19$

$P(E) = 0.32$



$P(E) = 0.20$



$P(E) = 0.09$

$P(E) = 0.20$

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?

$$P(E|F)$$

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

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

$$\approx 0.42$$

again, compute by analyzing data sets available from Netflix

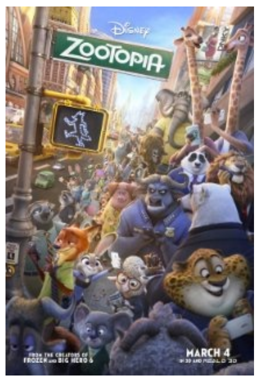
Netflix and Learn

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

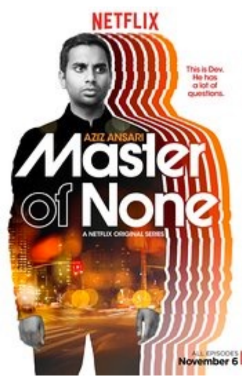
$P(E) = 0.19$

$P(E|F) = 0.14$



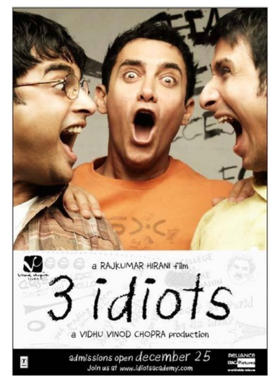
$P(E) = 0.32$

$P(E|F) = 0.35$



$P(E) = 0.20$

$P(E|F) = 0.20$



$P(E) = 0.09$

$P(E|F) = 0.72$

$P(E) = 0.20$

$P(E|F) = 0.42$

Law of Total Probability

Today's tasks

$$P(EF)$$

Chain rule
(Product rule)

Definition of
conditional probability

$$P(E|F)$$

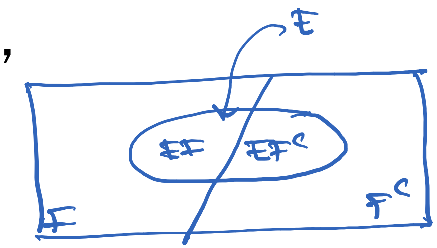
Law of Total
Probability

$$P(E)$$

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, F^C are disjoint such that $F \cup F^C = S$	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 are, by definition, mutually exclusive events

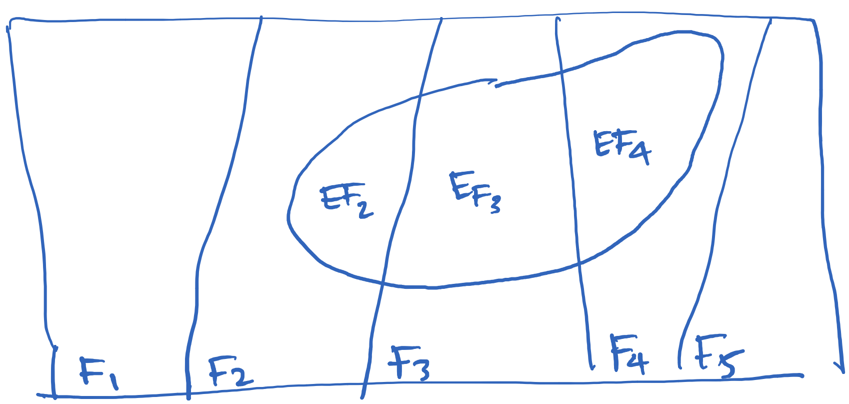
General Law of Total Probability

<u>Thm</u> For **mutually exclusive events** $F_1, F_2, \ldots, F_n$ such that $F_1 \cup F_2 \cup \cdots \cup F_n = S$,

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

need all $P(E|F_i)$ such that $F_1 \cup F_2 \cup F_3 \cup F_4 \cdots = S$

and you also need all $P(F_i)$ for all i



$EF_1 = EF_5 = \emptyset$ in this example

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)?

{1,2,3}

1. Define events & state goal

Let: E: win, F: flip heads
Want: $P(\text{win})$
$= P(E)$

2. Identify known 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

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

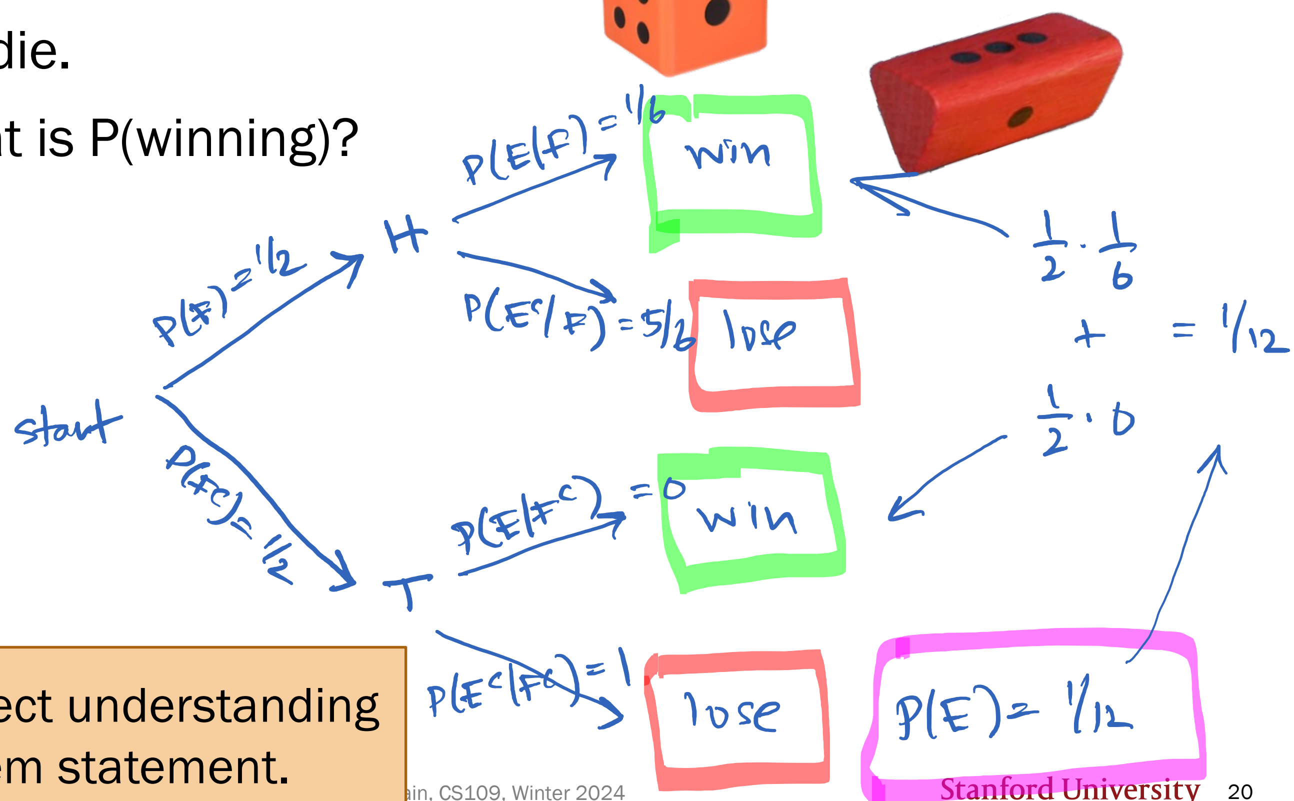
Finding $P(E)$ from $P(E|F)$, an understanding

- 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 understanding of the experiment with the problem statement.

Bayes' Theorem I

Today's tasks

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

$$P(EF)$$

Chain rule
(Product rule)

Definition of
conditional probability

$$P(E|F)$$

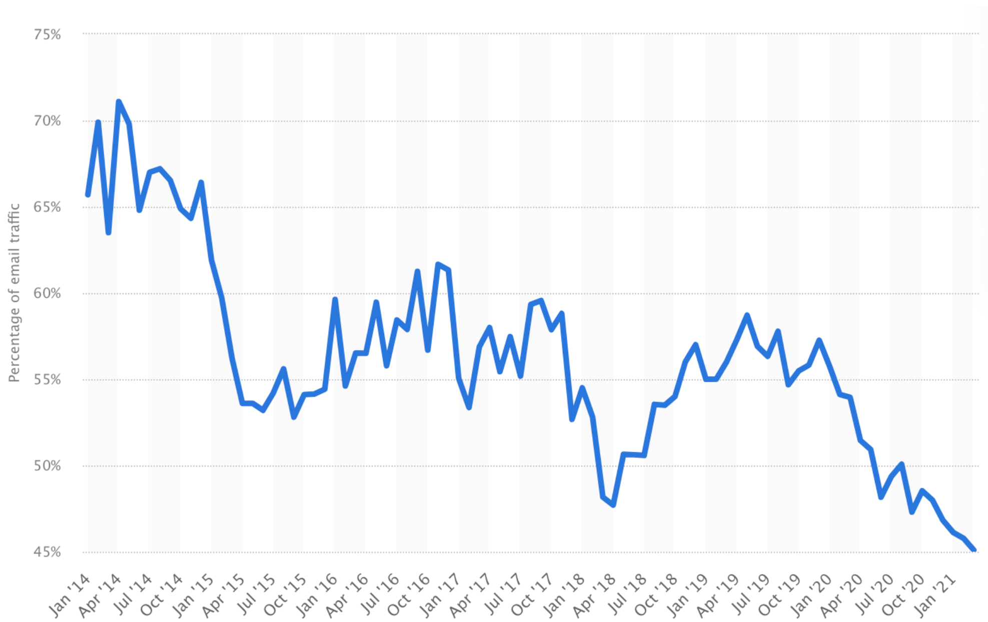
Law of Total
Probability

Bayes'
Theorem

$$P(E)$$

$$P(F|E)$$

Detecting spam email



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?

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

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!

$$P(F|E) \overset{①}{=} \frac{P(EF)}{P(E)}$$

$$\overset{②}{=} \frac{P(E|F)P(F)}{P(E)}$$

Expanded form:

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

Proof

1 more step!

expand $P(E)$ using LOTP

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"

$P(F) = 0.6$
$P(E|F) = 0.2$
$P(E|F^c) = 0.01$

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(spam | "Dear")

$= P(F|E)$

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^c)P(F^c)} = \frac{(0.2)(0.6)}{(0.2)(0.6) + (0.01)(0.4)}$$

$$\boxed{\approx 0.967}$$

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”

Note: You should know how to use Bayes/ Law of Total Prob., but drawing a tree can help.

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

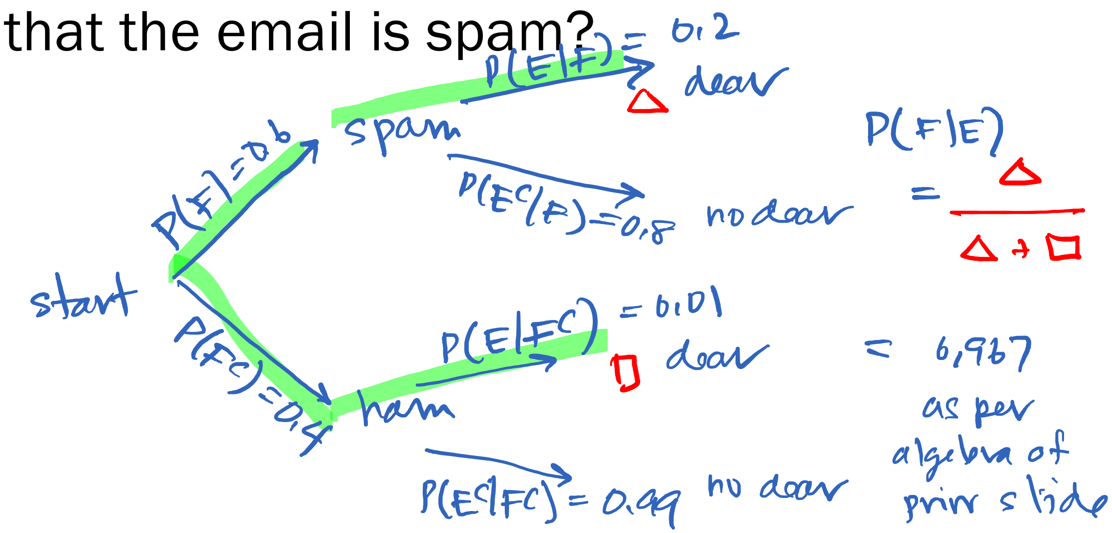
What is the probability that the email is spam?

1. Define events & state goal

Let: E: "Dear", F: spam

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

$= P(F|E)$



Bayes' Theorem terminology

- 60% of all email in 2016 is spam. $P(F)$ prior
- 20% of spam has the word "Dear" $P(E|F)$ likelihood
- 1% of non-spam (aka ham) has the word "Dear" $P(E|F^C)$ no special term

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

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

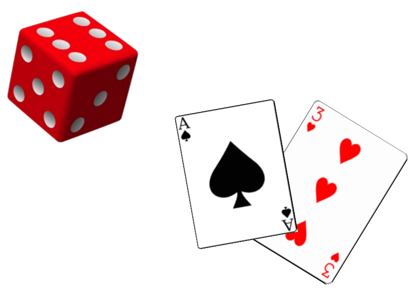
F: fact
E: evidence

$$\underbrace{P(F|E)}_{\text{posterior}} = \frac{\overbrace{P(E|F)}^{\text{likelihood}}\,\overbrace{P(F)}^{\text{prior}}}{\underbrace{P(E)}_{\text{normalization constant}}}$$

This class going forward

<u>Last week</u>

Equally likely events



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

(counting, combinatorics)

<u>Today and for most of this course</u>

Events not always equally likely

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

(collected from data)

Bayes'

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

(categorize a new datapoint)

Bayes' Theorem II

Bayes' Theorem

posterior likelihood prior

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

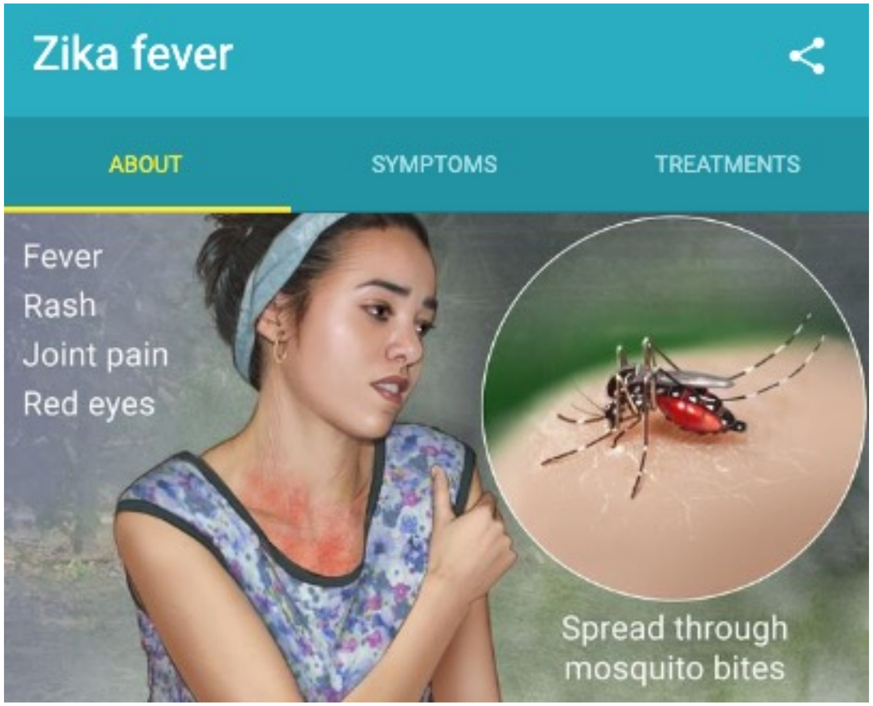
Mathematically:

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

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.
Generally, no symptoms, but can cause paralysis in very, very rare cases. During pregnancy: may cause birth defects. Very serious news story in 2015/2016.

Ziika Forest, Uganda

Rhesus monkeys

https://www.nytimes.com/2016/04/06/world/africa/uganda-zika-forest-mosquitoes.html

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

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 $P(E\|F)$	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

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?

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

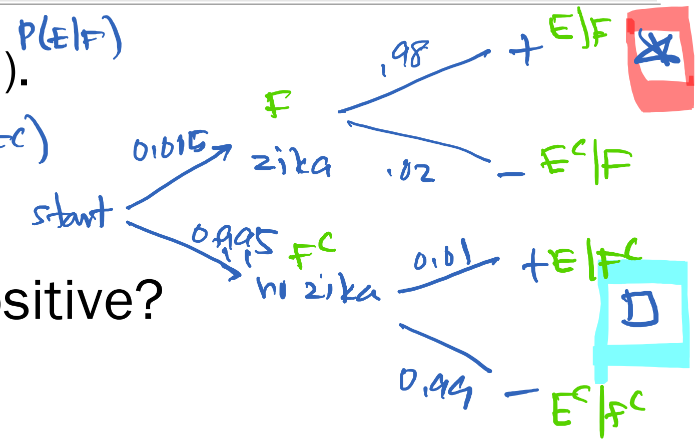
Let: E = you test positive
F = you actually have the disease

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"). $P(E|F)$
- However, the test has a "false positive" rate of 1%. $P(E|F^C)$
- 0.5% of the US population has Zika. $P(F)$



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

Why would you expect this number?

1. Define events & state goal
2. Identify known probabilities
3. Solve

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

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

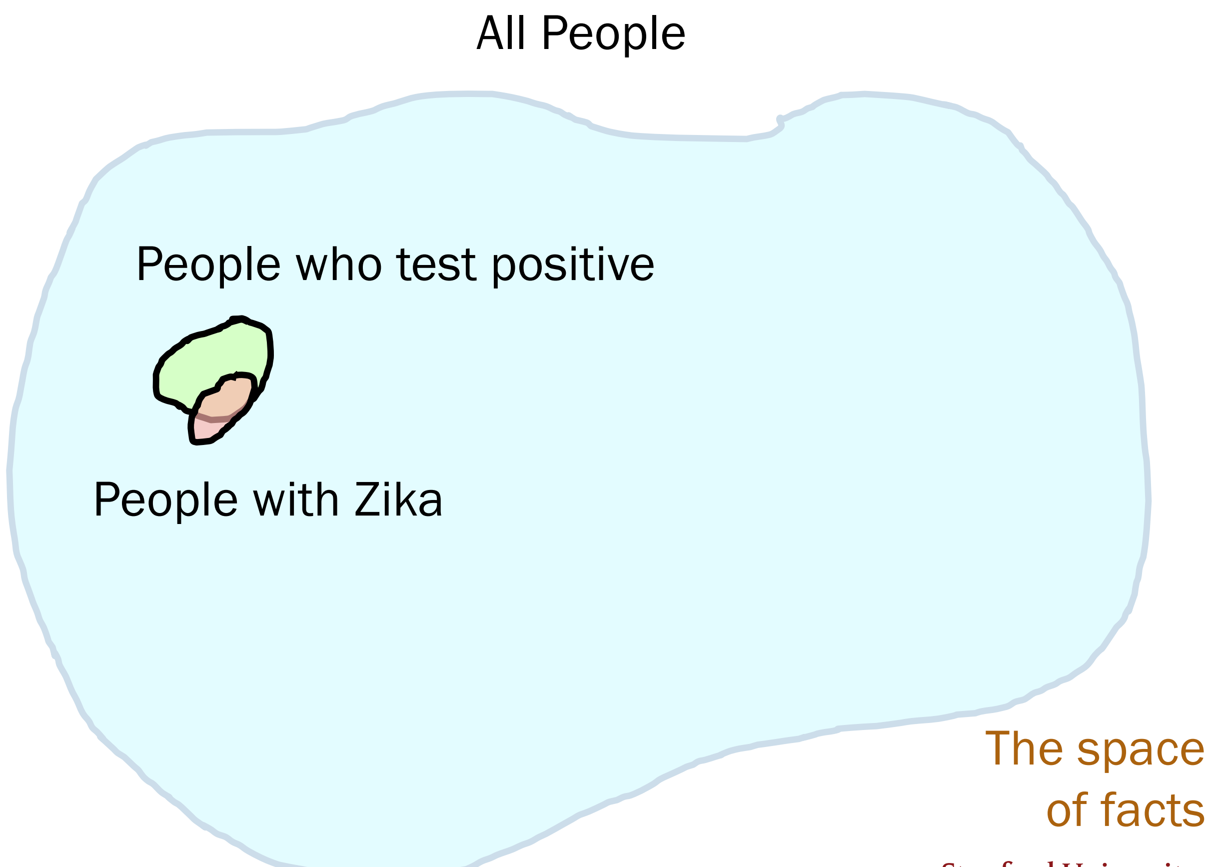
$$P(F|E) = \frac{(0.005)(0.98)}{(0.005)(0.98) + (0.995)(0.01)}$$

$$\simeq 0.330$$

Bayes' Theorem intuition

<u>Original question</u>:

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



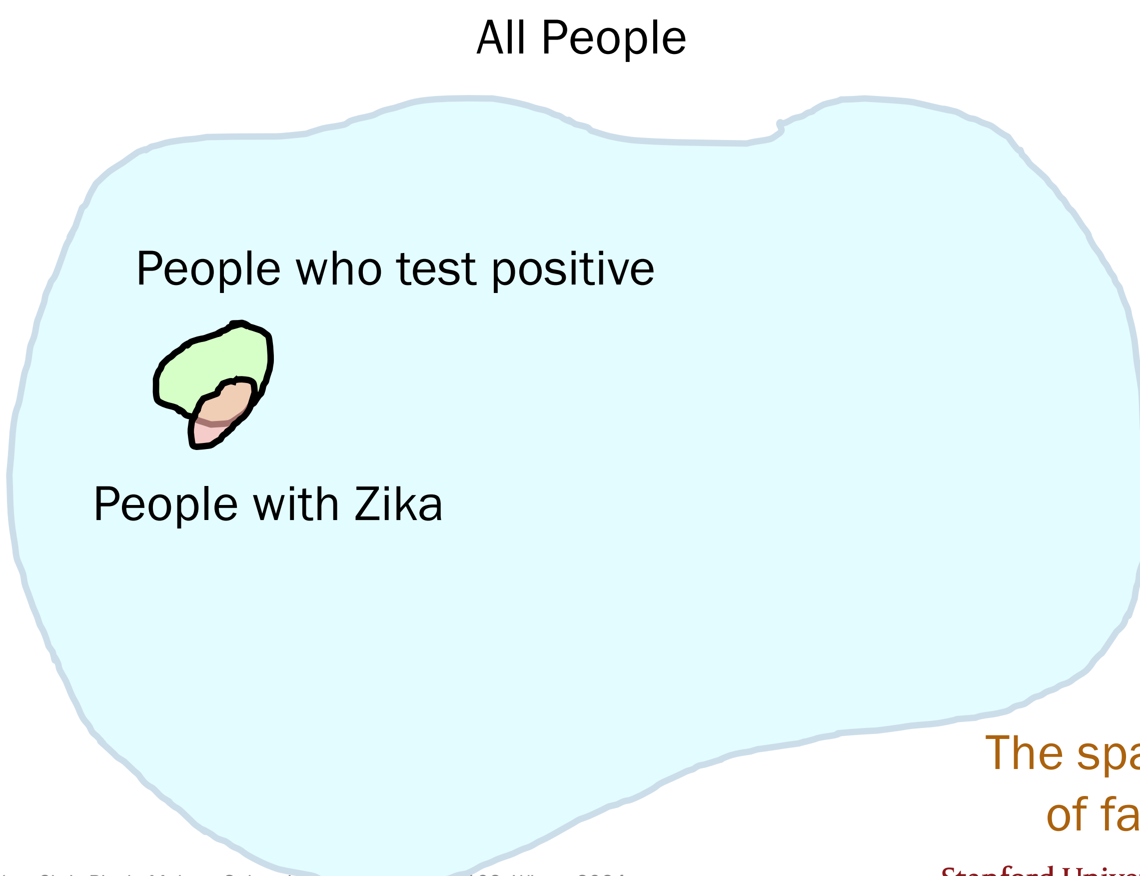
Bayes' Theorem intuition

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?



The space of facts

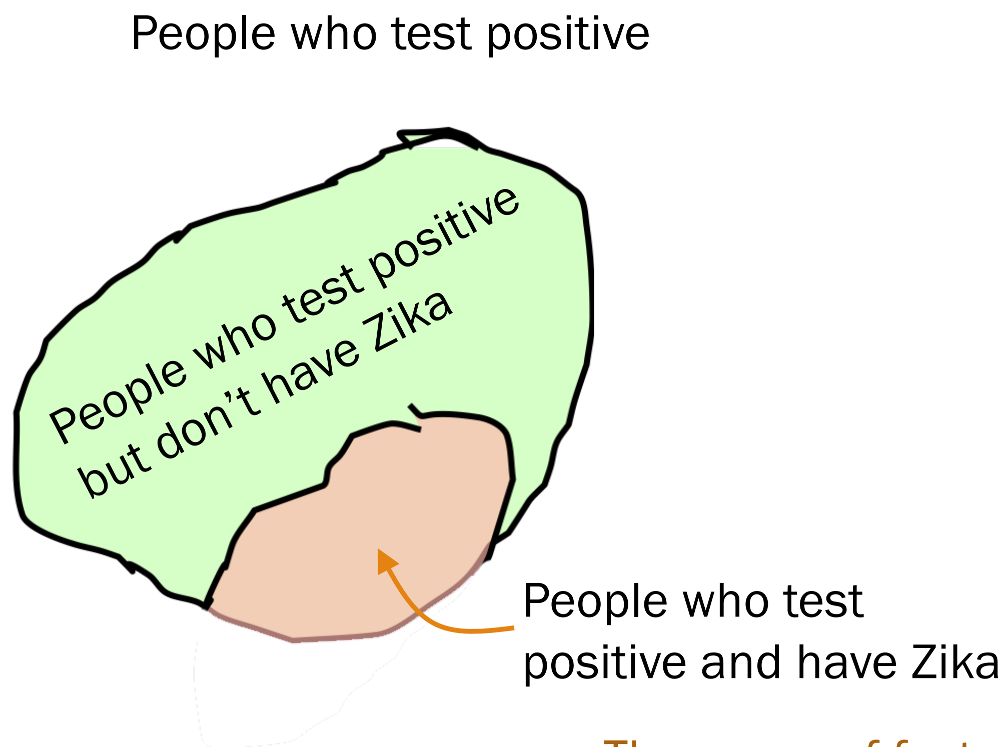
Bayes' Theorem intuition

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.

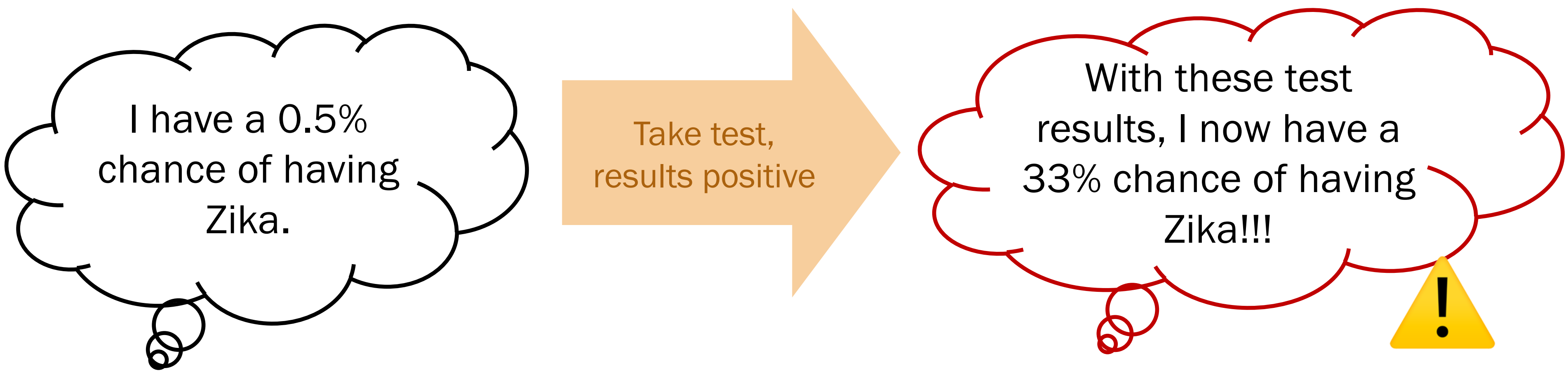
10 **do not** have Zika
and test positive.

≈ 0.333

[Demo (class website)](#)

Update your beliefs with Bayes' Theorem

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



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

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 $P(E\|F^C) = 0.01$

(ruminating)

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 $\boxed{P(E^C\|F)} = 0.02$	True negative $\boxed{P(E^C\|F^C)} = 0.99$

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

≈ 0.0001

via similar math

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!!!

$P(F|E)$

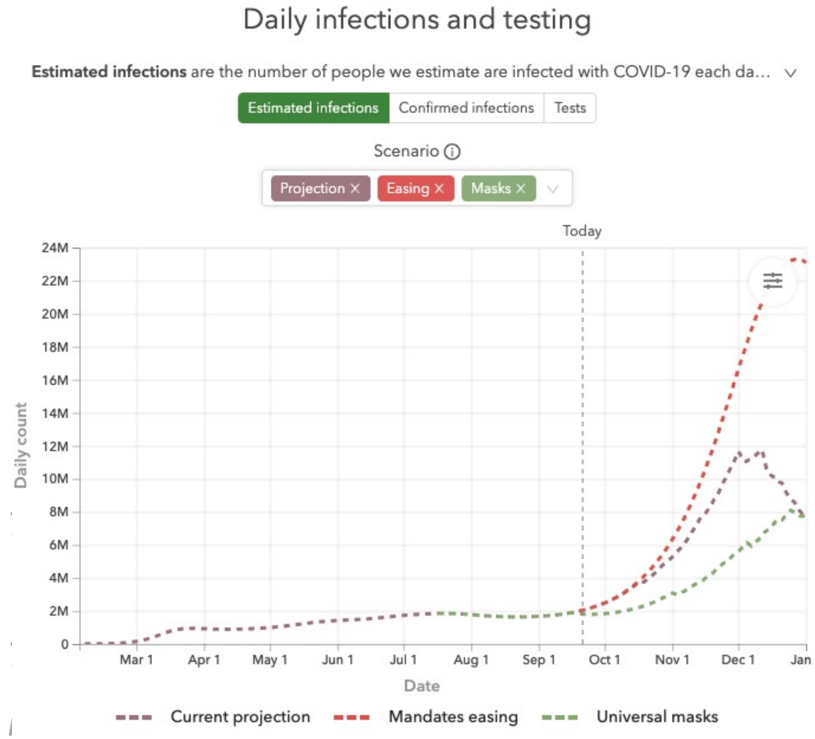
Take test, results negative

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.



Why test if there are errors?

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: 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, and 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.

$$\begin{cases} \frac{1}{1000} = P(\text{envelope is prize}) \\ \frac{999}{1000} = P(\text{other 999 envelopes have prize}) \end{cases}$$

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

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

3. Should you switch?

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

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