

Table of Contents

2 Conditional Probability
14 Law of Total Probability
21 Bayes' Theorem, Take I
31 Bayes' Theorem, Take II
38 The Monty Hall Problem

04: Conditional Probability and Bayes

Jerry Cain
April 10, 2023

Ed Discussion: <https://edstem.org/us/courses/38517/discussion/2914366>



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

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., $S \cap F$)

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

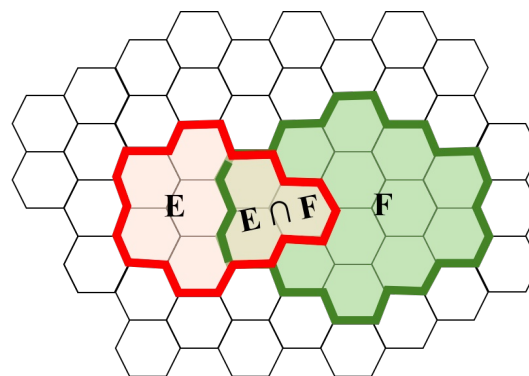
Conditional Probability, equally likely outcomes

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

With **equally likely outcomes**:

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

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



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

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

Slicing up the spam

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

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

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

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!

Stanford University 7

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

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

Let E = a user watches Life is Beautiful.

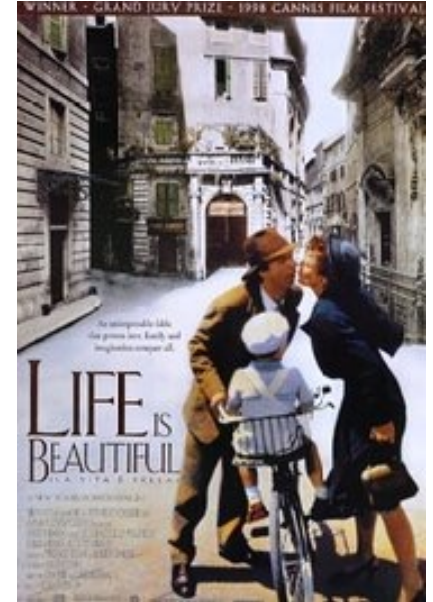
What is $P(E)$?

✗ Equally likely outcomes?

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

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

$P(E) = 1/2$?



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

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

Netflix and Learn

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

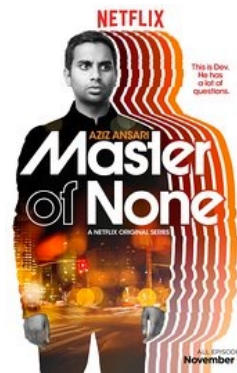
Let E be the event that a user watches the given movie.



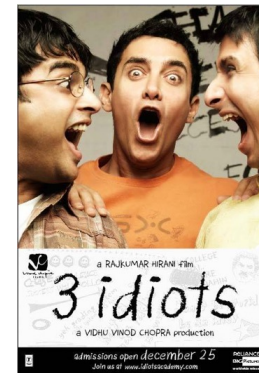
$$P(E) = 0.19$$



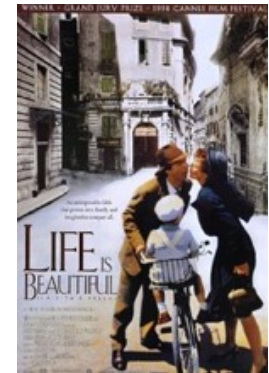
$$P(E) = 0.32$$



$$P(E) = 0.20$$



$$P(E) = 0.09$$



$$P(E) = 0.20$$

Netflix and Learn

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

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?

$$P(E|F)$$

$$\begin{aligned} P(E|F) &= \frac{P(EF)}{P(F)} = \frac{\frac{\# \text{ people who have watched both}}{\# \text{ people on Netflix}}}{\frac{\# \text{ people who have watched Amelie}}{\# \text{ people on Netflix}}} \\ &= \frac{\# \text{ people who have watched both}}{\# \text{ people who have watched Amelie}} \\ &\approx 0.42 \end{aligned}$$

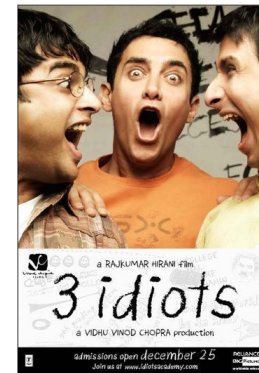
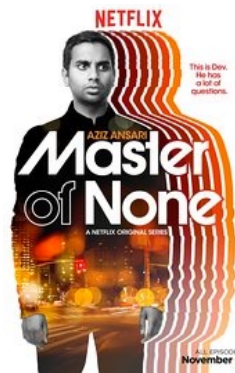


Netflix and Learn

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

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.



$$P(E) = 0.19$$

$$P(E) = 0.32$$

$$P(E) = 0.20$$

$$P(E) = 0.09$$

$$P(E) = 0.20$$

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

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

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

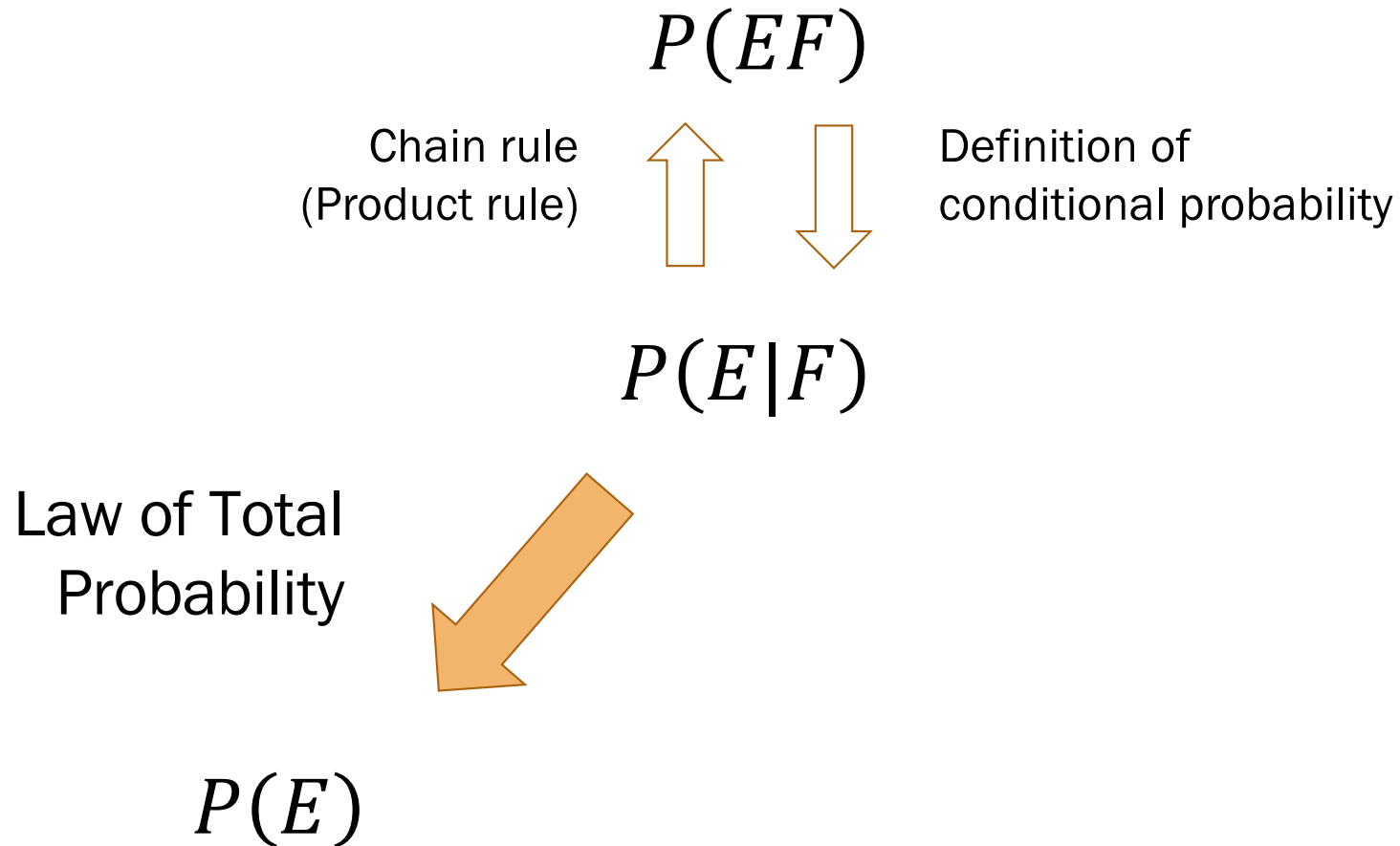
$$P(E|F) = 0.72$$

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



Law of Total Probability

Today's tasks



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. 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

Thm For **mutually exclusive events** F_1, F_2, \dots, F_n such that $F_1 \cup F_2 \cup \dots \cup F_n = S$,

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

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 $P(\text{winning})$?



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

$$P(E) = P(E|F)P(F) + P(E|F^C)P(F^C) \quad \text{Law of Total Probability}$$

- 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 $P(\text{winning})$?

1. Define events & state goal

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

2. Identify known probabilities

$$\begin{aligned} P(\text{win}|H) &= P(E|F) = 1/6 \\ P(H) &= P(F) = 1/2 \\ P(\text{win}|T) &= P(E|F^C) = 0 \\ P(T) &= P(F^C) = 1 - 1/2 \end{aligned}$$

3. Solve

$$\begin{aligned} P(E) &= (1/6)(1/2) \\ &\quad + (0)(1/2) \\ &= \frac{1}{12} \approx 0.083 \end{aligned}$$

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 $P(\text{winning})$?

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



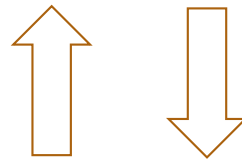
Rev. Thomas Bayes (~1701-1761):
British mathematician and Presbyterian minister

Law of Total
Probability

$$P(E)$$

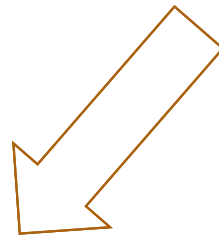
Chain rule
(Product rule)

$$P(EF)$$

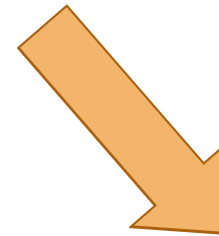


Definition of
conditional probability

$$P(E|F)$$

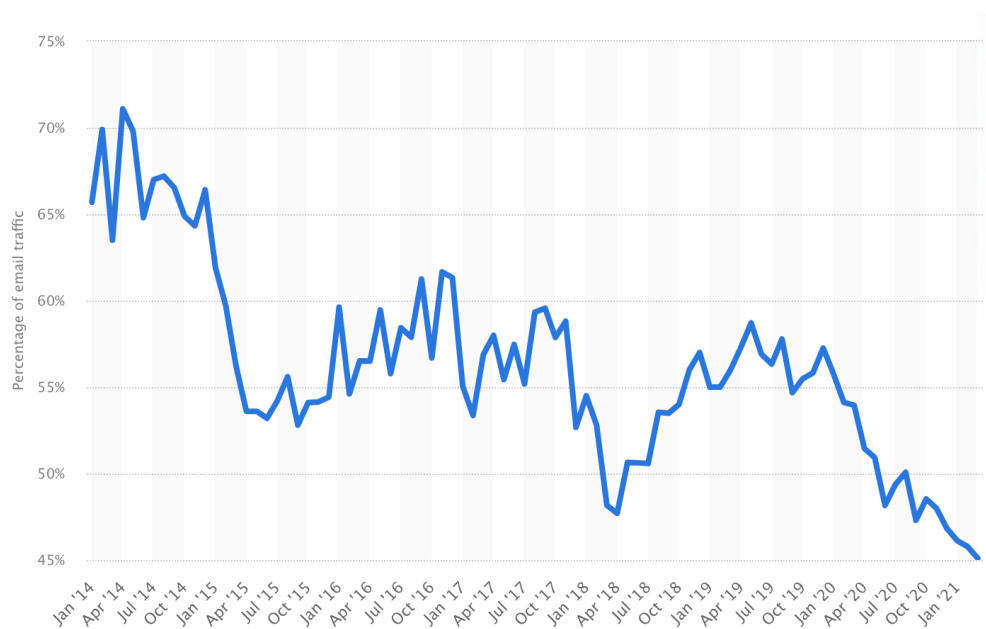


Bayes'
Theorem



$$P(F|E)$$

Detecting spam email



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

$$P(E|F) = P(\text{"Dear"} \mid \text{Spam email})$$

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

$$P(F|E) = P(\text{Spam email} \mid \text{"Dear"})$$

INVOICE

Geek SQUAD

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

Dear Geek Squad Customer,

Thank you for using Geek Squad Antivirus for the last one year. Your **Geek SQUAD Antivirus 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 : 24th 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.

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!

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!



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: $P(\text{spam} | \text{"Dear"})$
 $= P(F|E)$

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?

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

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. $P(F)$
- 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.

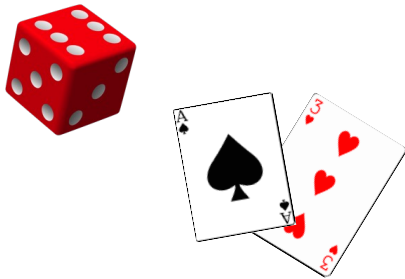
What is the probability that the email is spam? **Want: $P(F|E)$**

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

normalization constant

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
Events not always equally likely

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

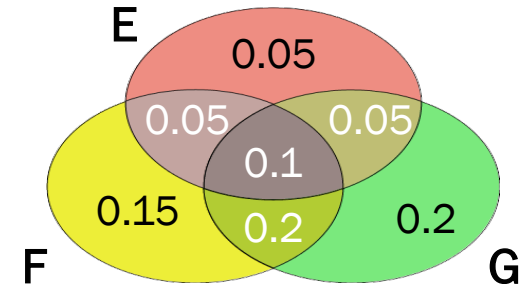
Bayes'

$P(F = \text{Fact} \mid E = \text{Evidence})$
(categorize
a new datapoint)

Brainstorm with neighbors

You have a flowering plant.

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



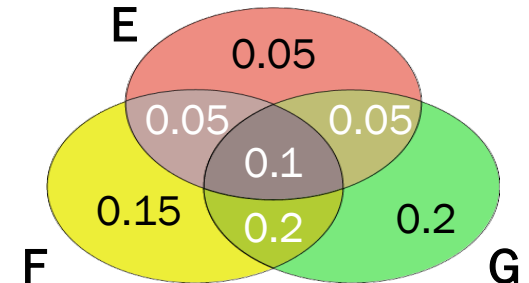
- How would you write
 - the probability that the plant got sun, given that it was watered, and flowers bloomed?
 - the probability that the plant got sun and flowers bloomed given that it was watered?
- Using the Venn diagram, compute the above probabilities.
- Chain Rule: Fill in the blanks.
 - $P(GE)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



Brainstorm with neighbors

You have a flowering plant.

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



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



Bayes' Theorem II

Bayes' Theorem

Review

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

Mathematically:

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

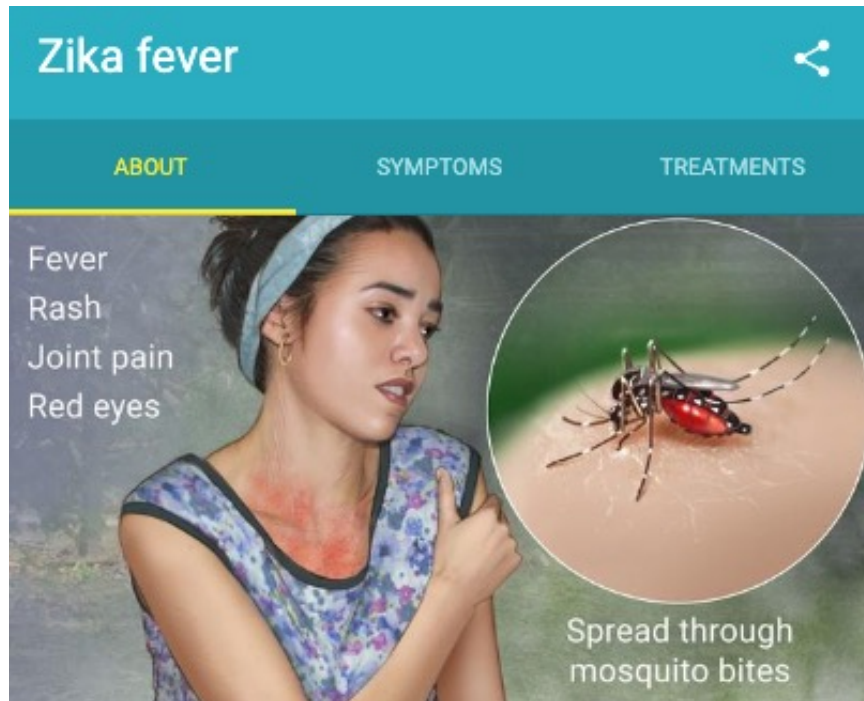
Real-life application:

Given new evidence E , update belief of fact F

Prior belief \rightarrow 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-a-zika-forest-mosquitoes.html>

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



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



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?

Zika Testing

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

- 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:
 $P(\text{disease} \mid \text{test}^+)$
 $= P(F|E)$



Zika Testing

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

- 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 known
probabilities

3. Solve

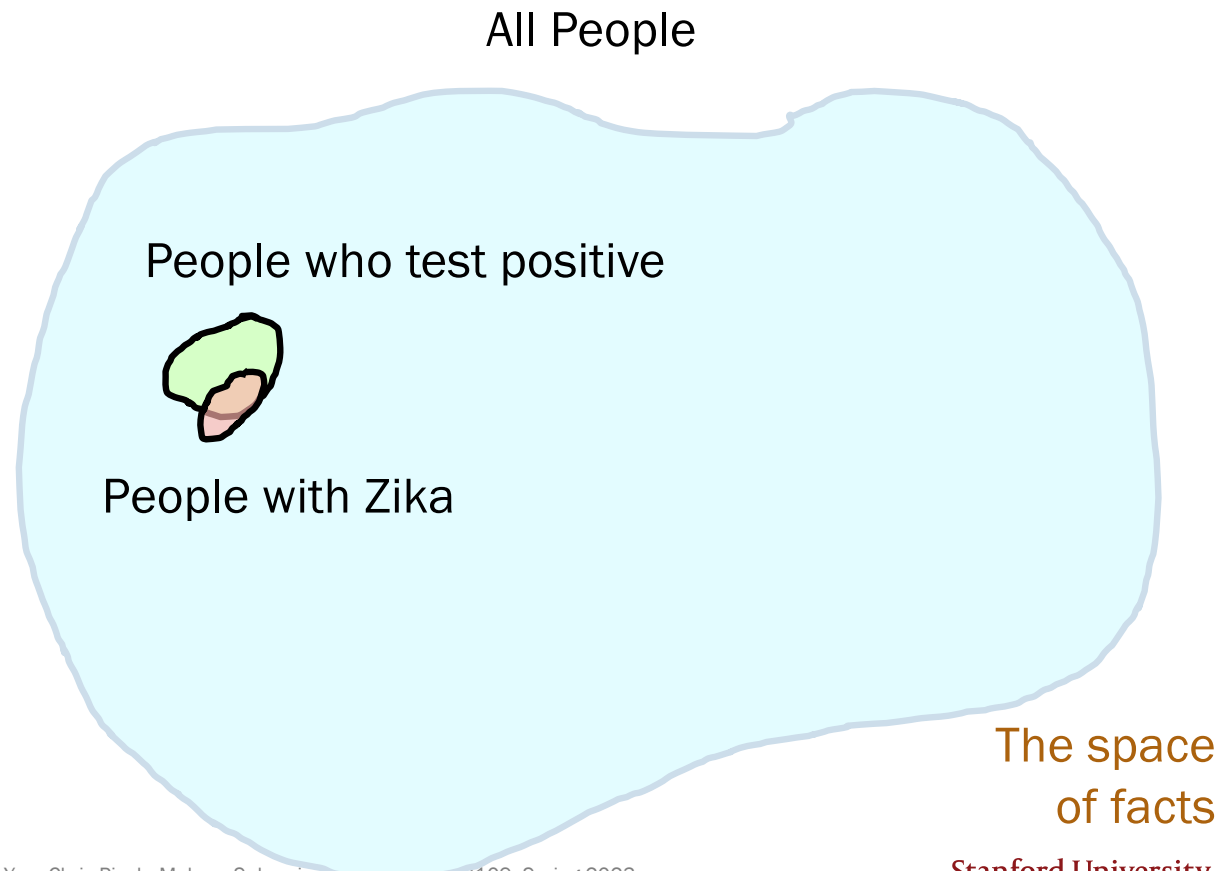
Let: E = you test positive
 F = you actually have
the disease

Want:
 $P(\text{disease} \mid \text{test}^+)$
 $= P(F|E)$

Bayes' Theorem intuition

Original question:

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



Bayes' Theorem intuition

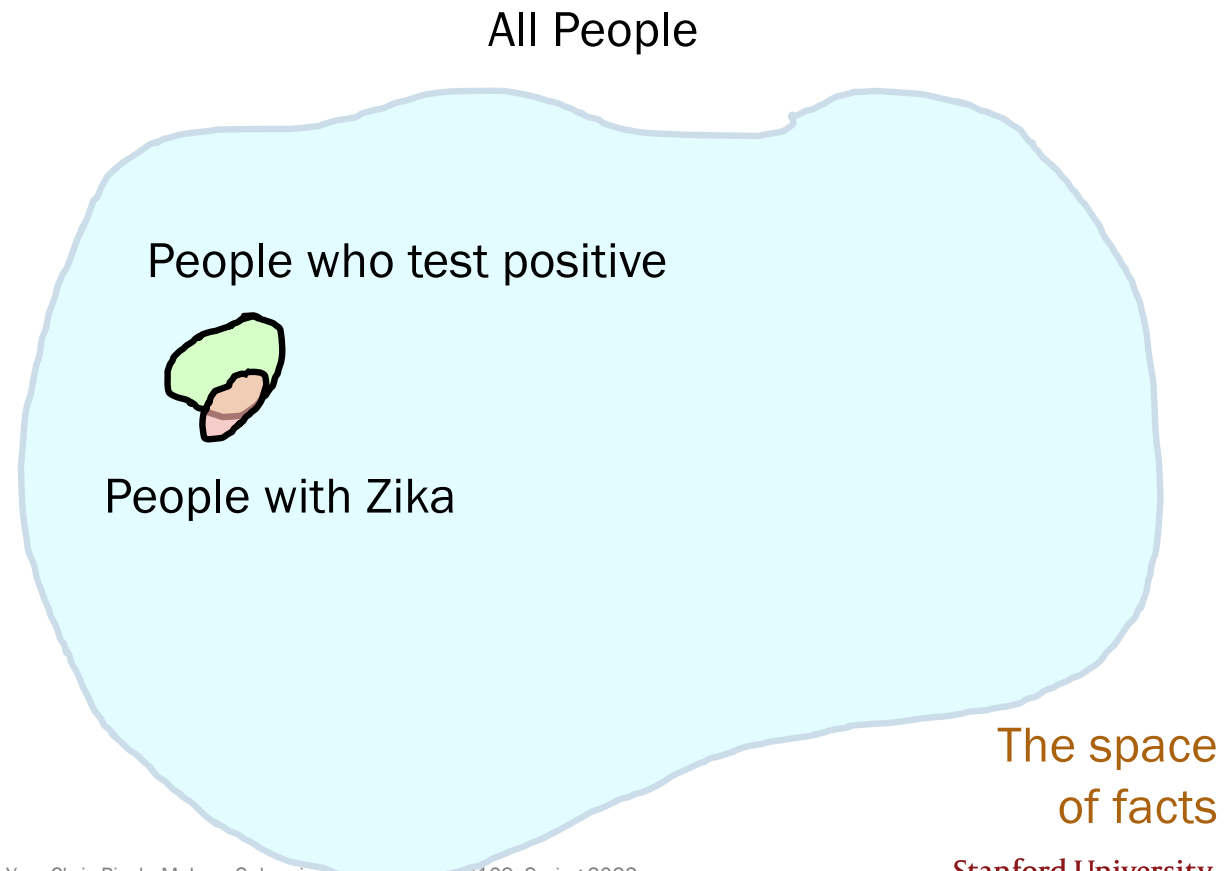
Original question:

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:

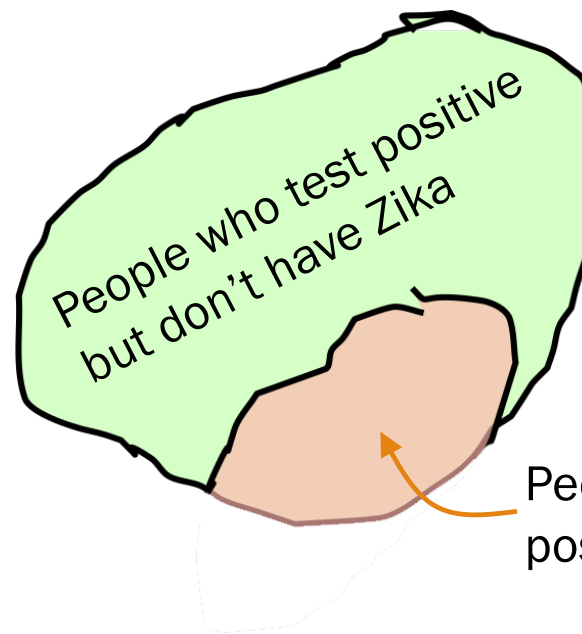
What is the likelihood you have Zika if you test positive for the disease?

Interpret

Interpretation:

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

People who test positive



People who test positive and 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:

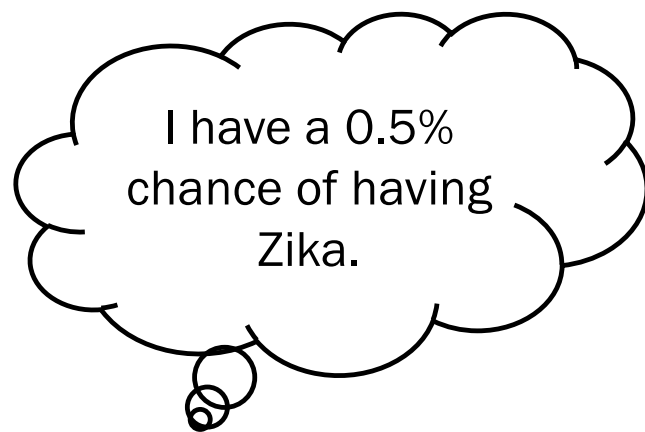


5 have Zika
and test positive
985 do not have Zika
and test negative.
10 do not have Zika
and test positive.
 ≈ 0.333

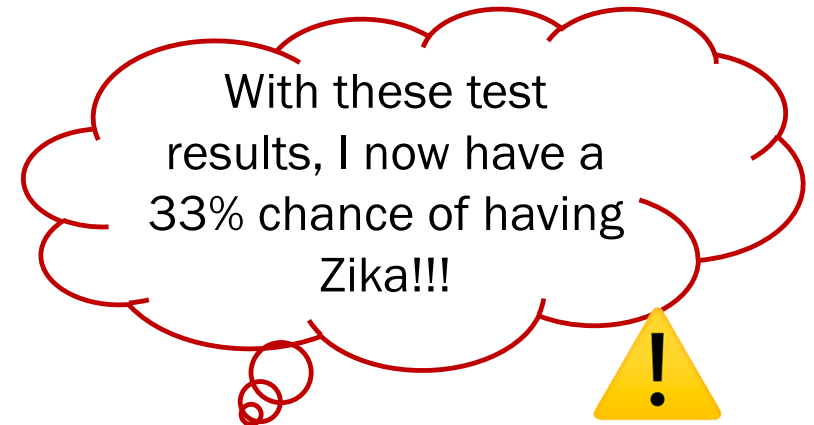
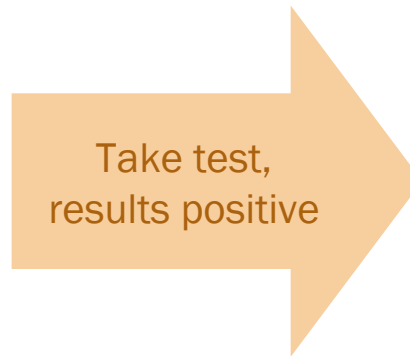
Update your beliefs with Bayes' Theorem

E = you test positive for Zika

F = you have the disease



$P(F)$



$P(F|E)$

Why it's still good to get tested

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

- 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 $P(F|E^C)$?



Why it's still good to get tested

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

- 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 $P(F|E^C)$?

Why it's still good to get tested

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

- 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 $P(E^C F^C) = 0.99$

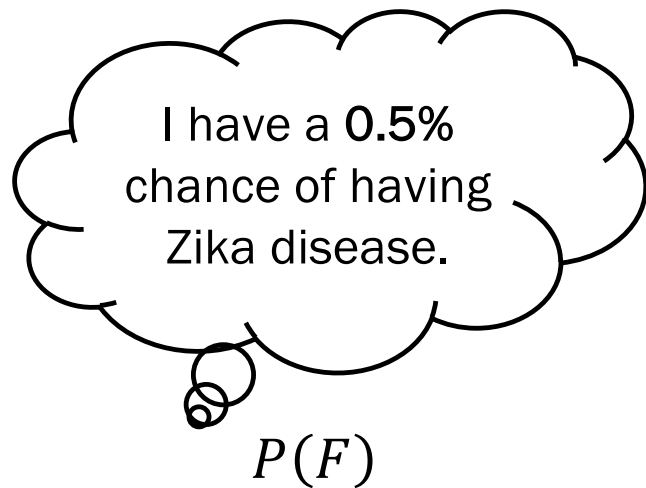
$$P(F|E^C) = \frac{P(E^C|F)P(F)}{P(E^C|F)P(F) + P(E^C|F^C)P(F^C)}$$

Why it's still good to get tested

E = you test positive for Zika

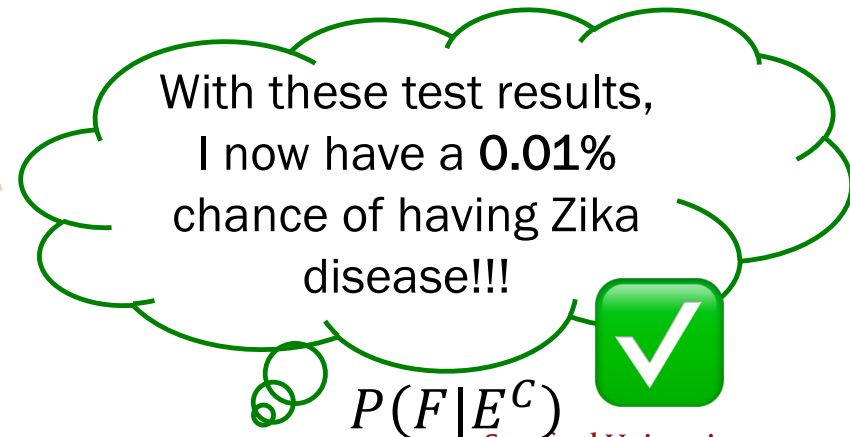
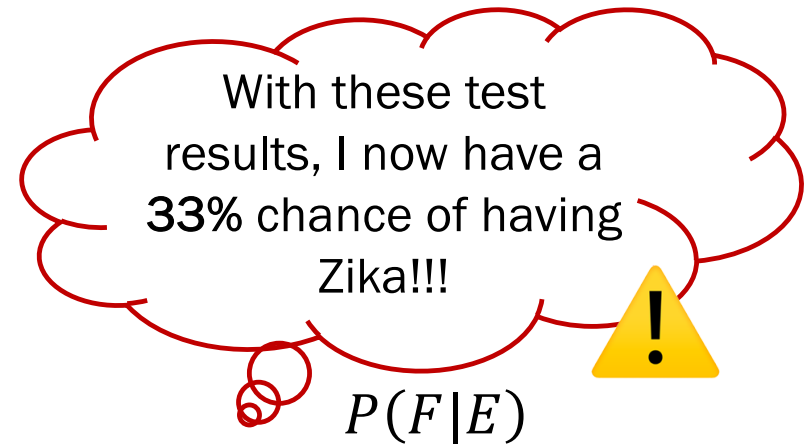
F = you actually have the disease

E^c = you test **negative** for Zika



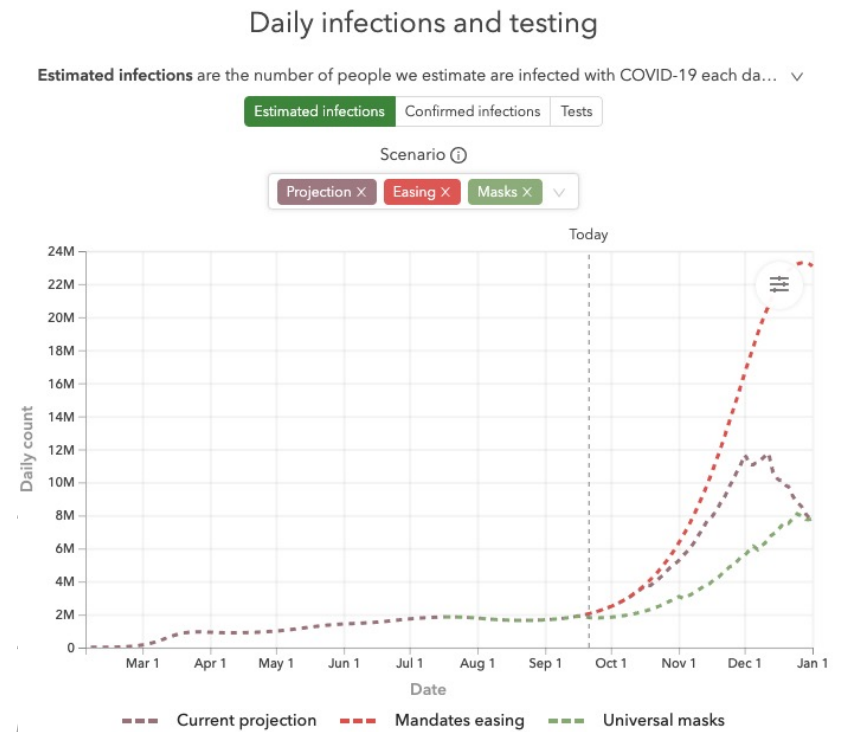
Take test,
results positive

Take test,
results negative



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.

Should we switch?



Note: If we don't switch, $P(\text{win}) = 1/3$ (random)

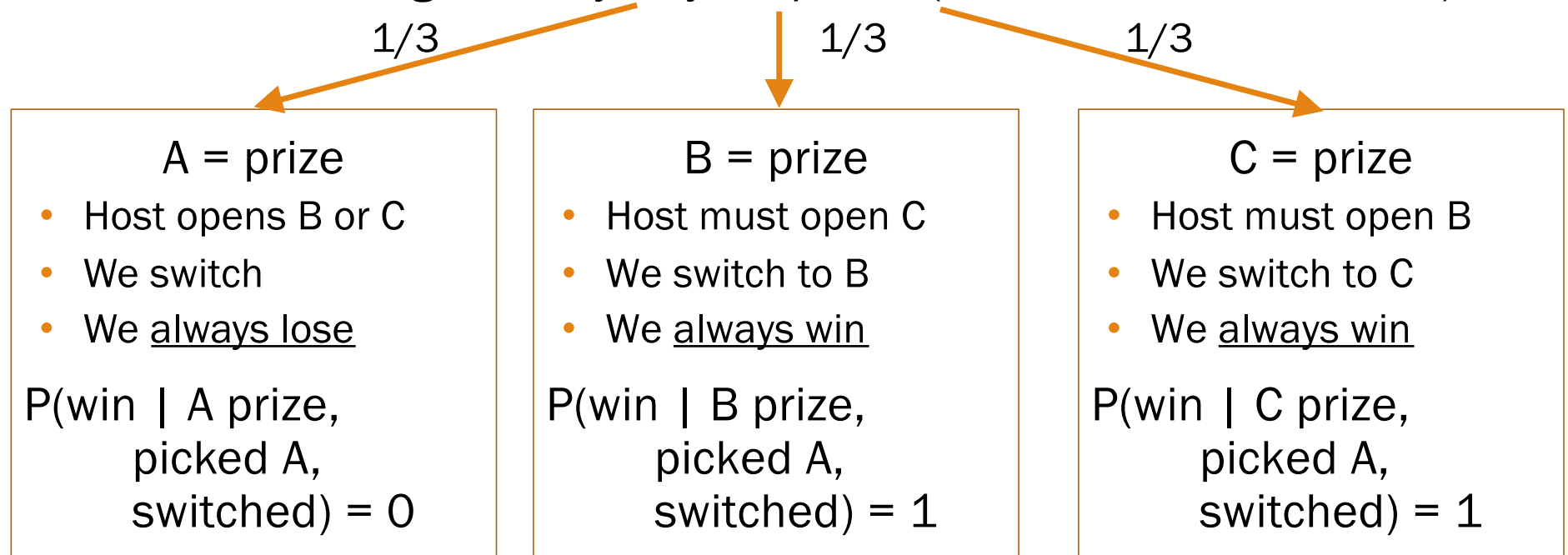
We are comparing $P(\text{win})$ and $P(\text{win} | \text{switch})$.



Doors A,B,C

If we switch

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



$$P(\text{win} \mid \text{picked A, switched}) = \frac{1}{3} * 0 + \frac{1}{3} * 1 + \frac{1}{3} * 1 = \frac{2}{3}$$

You should switch.

Monty Hall, 1000 envelope version

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

1. You choose 1 envelope.
$$\left\{ \begin{array}{l} \frac{1}{1000} = P(\text{envelope is prize}) \\ \frac{999}{1000} = P(\text{other 999 envelopes have prize}) \end{array} \right.$$
2. I open 998 of remaining 999 (showing they are empty).
$$\begin{aligned} \frac{999}{1000} &= P(998 \text{ empty envelopes had prize}) \\ &\quad + P(999^{\text{th}} \text{ envelope has prize}) \\ &= P(999^{\text{th}} \text{ envelope has prize}) \end{aligned}$$
3. Should you switch?
No: $P(\text{win without switching}) = \frac{1}{\text{original \# envelopes}}$
Yes: $P(\text{win with new knowledge}) = \frac{\text{original \# envelopes} - 1}{\text{original \# envelopes}}$