# More Practice
## Independence, Conditioning, Bayes

CSE 312 Winter 25
Lecture 8

# Today

Practice (independence and conditioning)

One more independence definition

Bayes' Rule in the real world!

# More Bayes Practice

# A contrived example

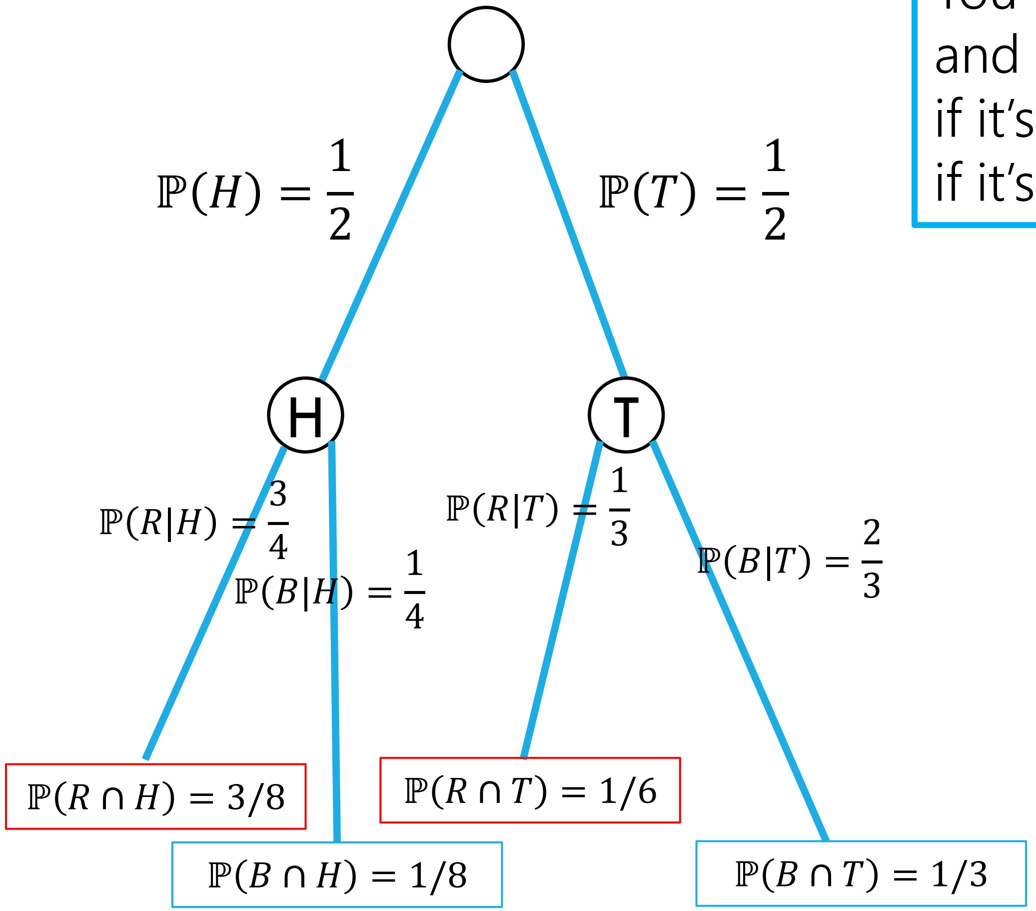
You have three red marbles and one blue marble in your left pocket, and one red marble and two blue marbles in your right pocket.

You will flip a fair coin; if it's heads, you'll draw a marble (uniformly) from your left pocket, if it's tails, you'll draw a marble (uniformly) from your right pocket.

Let $B$ be you draw a blue marble. Let $T$ be the coin is tails.

What is $\mathbb{P}(B|T)$ what is $\mathbb{P}(T|B)$ ?

# Updated Sequential Processes

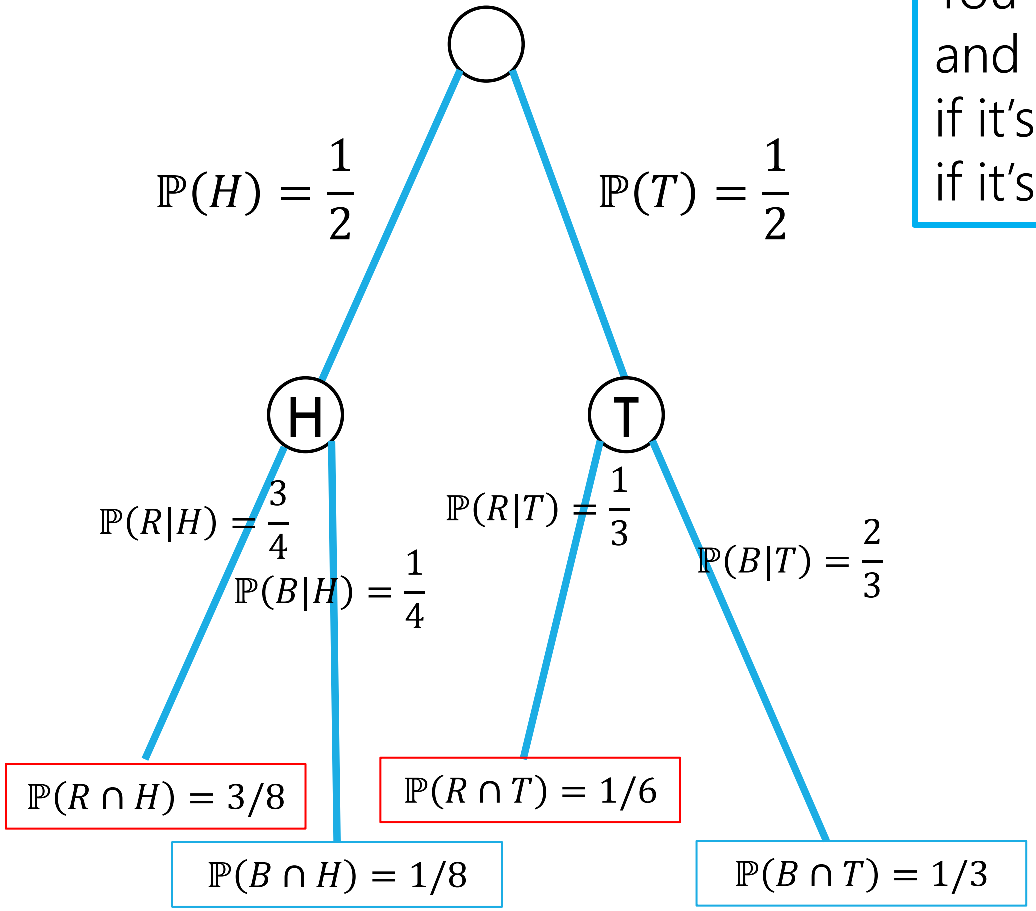



You have three red marbles and one blue marble in your left pocket, and one red marble and two blue marbles in your right pocket.
if it's heads, you'll draw a marble (uniformly) from your left pocket,
if it's tails, you'll draw a marble (uniformly) from your right pocket.

For sequential processes with probability, at each step multiply by
$\mathbb{P}(\text{next step } | \text{all} \cap \text{prior} \cap \text{steps})$

# Updated Sequential Processes

You have three red marbles and one blue marble in your left pocket, and one red marble and two blue marbles in your right pocket.
if it's heads, you'll draw a marble (uniformly) from your left pocket,
if it's tails, you'll draw a marble (uniformly) from your right pocket.




For sequential processes with probability, at each step multiply by
$\mathbb{P}(\text{next step }|\text{all} \cap \text{prior} \cap \text{steps})$

$$\mathbb{P}(B|T) = 2/3;\ \mathbb{P}(B) = \frac{1}{8} + \frac{1}{3} = \frac{11}{24}$$

# Flipping the conditioning

What about $\mathbb{P}(T|B)$?

You have three red marbles and one blue marble in your left pocket, and one red marble and two blue marbles in your right pocket.
if it's heads, you'll draw a marble (uniformly) from your left pocket,
if it's tails, you'll draw a marble (uniformly) from your right pocket.

Pause, what's your intuition?

Is this probability

A. less than ½

B. equal to ½

C. greater than ½

# Flipping the conditioning

What about $\mathbb{P}(T|B)$?

You have three red marbles and one blue marble in your left pocket, and one red marble and two blue marbles in your right pocket.
if it's heads, you'll draw a marble (uniformly) from your left pocket,
if it's tails, you'll draw a marble (uniformly) from your right pocket.

Pause, what's your intuition?

Is this probability

A. less than ½

B. equal to ½

C. greater than ½

The right (tails) pocket is far more likely to produce a blue marble if picked than the left (heads) pocket is. Seems like $\mathbb{P}(T|B)$ should be greater than ½.

# Flipping the conditioning

What about $\mathbb{P}(T|B)$?

You have three red marbles and one blue marble in your left pocket, and one red marble and two blue marbles in your right pocket.
if it's heads, you'll draw a marble (uniformly) from your left pocket,
if it's tails, you'll draw a marble (uniformly) from your right pocket.

Bayes' Rule says:

$$\mathbb{P}(T|B) = \frac{\mathbb{P}(B|T)\mathbb{P}(T)}{\mathbb{P}(B)}$$

$$= \frac{\frac{2}{3} \cdot \frac{1}{2}}{11/24} = 8/11$$

# The Technical Stuff

# Proof of Bayes' Rule

$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ by definition of conditional probability

Now, imagining we get $A \cap B$ by conditioning on $A$, we should get a numerator of $\mathbb{P}(B|A) \cdot \mathbb{P}(A)$

$= \frac{\mathbb{P}(B|A) \cdot \mathbb{P}(A)}{\mathbb{P}(B)}$

As required.

# A Technical Note

After you condition on an event, what remains is a probability space.

With $B$ playing the role of the sample space,

$\mathbb{P}(\omega|B)$ playing the role of the probability measure.

All the axioms are satisfied (it's a good exercise to check)

That means any theorem we write down has a version where you condition everything on $B$.

# An Example

Bayes Theorem still works in a probability space where we've already conditioned on $S$.

$$\mathbb{P}(A|[B \cap S]) = \frac{\mathbb{P}(B|[A \cap S]) \cdot \mathbb{P}(A|S)}{\mathbb{P}(B|S)}$$

Complementary law still works in a probability space where we've already conditioned on $S$

$$\mathbb{P}(A|C) = 1 - \mathbb{P}(\bar{A}|C)$$

# A Quick Technical Remark

I often see students write things like

$\mathbb{P}([A|B]|C)$

This is not a thing.

You probably want $\mathbb{P}(A|[B \cap C])$

$A|B$ isn't an event – it's describing an event **and** telling you to restrict the sample space. So you can't ask for the probability of that conditioned on something else.

# More Independence

# Independence of events

Recall the definition of independence of **events**:

## Independence

**Two events $A, B$ are independent if**

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

# Independence for 3 or more events

For three or more events, we need two kinds of independence

## Pairwise Independence

Events $A_1, A_2, \dots, A_n$ are pairwise independent if
$\mathbb{P}(\boldsymbol{A_i} \cap \boldsymbol{A_j}) = \mathbb{P}(\boldsymbol{A_i}) \cdot \mathbb{P}(\boldsymbol{A_j})$ for all $\boldsymbol{i, j}$

## Mutual Independence

Events $A_1, A_2, \dots, A_n$ are mutually independent if
$\mathbb{P}(\boldsymbol{A_{i_1}} \cap \boldsymbol{A_{i_2}} \cap \cdots \cap \boldsymbol{A_{i_k}}) = \mathbb{P}(\boldsymbol{A_{i_1}}) \cdot \mathbb{P}(\boldsymbol{A_{i_2}}) \cdots \mathbb{P}(\boldsymbol{A_{i_k}})$
for every subset $\{\boldsymbol{i_1, i_2, \dots, i_k}\}$ of $\{\boldsymbol{1, 2, \dots, n}\}$.

# Pairwise Independence vs. Mutual Independence

Roll two fair dice (one red one blue) independently

$R =$"red die is 3"

$B =$"blue die is 5"

$S =$"sum is 7"

How should we describe these events?

# Pairwise Independence

$R, B, S$ are pairwise independent

$\mathbb{P}(R \cap B) \,?\!= \mathbb{P}(R)\mathbb{P}(B)$

$\frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6}$ Yes! (These are also independent by the problem statement)

$\mathbb{P}(R \cap S) \,?\!= \mathbb{P}(R)\mathbb{P}(S)$

$\frac{1}{36} \,?\!= \frac{1}{6} \cdot \frac{1}{6}$ Yes!

$\mathbb{P}(B \cap S) \,?\!= \mathbb{P}(B)\mathbb{P}(S)$

$\frac{1}{36} \,?\!= \frac{1}{6} \cdot \frac{1}{6}$ Yes!

Since all three pairs are independent, we say the random variables are pairwise independent.

# Mutual Independence

$R, B, S$ are not mutually independent.

$\mathbb{P}(R \cap B \cap S) = 0$; if the red die is 3, and blue die is 5 then the sum is 8 (so it can't be 7)

$$\mathbb{P}(R)\mathbb{P}(B)\mathbb{P}(S) = \left(\frac{1}{6}\right)^3 = \frac{1}{216} \neq 0$$

# Checking Mutual Independence

It's not enough to check just $\mathbb{P}(A \cap B \cap C)$ either.

Roll a fair 8-sided die.

Let $A$ be $\{1,2,3,4\}$

$B$ be $\{2,4,6,8\}$

$C$ be $\{2,3,5,7\}$

$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\{2\}) = \frac{1}{8}$

$\mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$

# Checking Mutual Independence

It's not enough to check just $\mathbb{P}(A \cap B \cap C)$ either.

Roll a fair 8-sided die.

Let $A$ be $\{1,2,3,4\}$

$B$ be $\{2,4,6,8\}$

$C$ be $\{2,3,5,7\}$

$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\{2\}) = \frac{1}{8}$

$\mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$

But $B$ and $C$ aren't independent. Because there's a subset that's not independent, $A, B, C$ are not mutually independent.

# Checking Mutual Independence

To check mutual independence of events:

Check **every** subset.

To check pairwise independence of events:

Check **every** subset of size two.

# Why Two Versions?

Pairwise independence is often all you need and is easier to design an experiment/code to achieve it.

"Pairwise independent hash functions" are a theoretical example.

Mutual Independence would let us vastly simplify the chain rule computation.

$P(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_2 \cap A_1) \cdots \mathbb{P}(A_n|A_1 \cap \cdots \cap A_{n-1})$

Simplifies to $\mathbb{P}(A_1) \cdot \mathbb{P}(A_2) \cdot \mathbb{P}(A_3) \cdots \mathbb{P}(A_n)$

# Bayes in the real world

# Application 1: Medical Tests

Helping Doctors and Patients Make Sense of Health Statistics

A researcher posed the following scenario to a group of 160 doctors:

Assume you conduct a disease screening using a standard test in a certain region. You know the following information about the people in this region:

The probability that a person has the disease is 1% (prevalence)

If a person has the disease, the probability that she tests positive is 90% (sensitivity)

If a person does not have the disease, the probability that she nevertheless tests positive is 9% (false-positive rate)

A person tests positive. She wants to know from you whether that means that she has the disease for sure, or what the chances are. What is the best answer?

A. The probability that she has the disease is about 81%.
B. Out of 10 people with a positive test, about 9 have the disease.
C. Out of 10 people with a positive test, about 1 have the disease.
D. The probability that she has the disease is about 1%

# Let’s do the calculation!

Let $D$ be “the patient has the disease”, $T$ be the test was positive.

$$\mathbb{P}(D|T) = \mathbb{P}(T|D) \cdot \mathbb{P}(D)/\mathbb{P}(T)$$

$$= \frac{.9 \cdot .01}{.99 \cdot .09 + .01 \cdot .9} \approx 0.092$$

Calculation tip: for Bayes’ Rule, you should see one of the terms on the bottom exactly match your numerator (if you’re using the LTP to calculate the probability on the bottom)

# Pause for vocabulary

Physicians have words for just about everything

Let $D$ be has the disease; $T$ be test is positive

$\mathbb{P}(D)$ is “prevalence”

$\mathbb{P}(T|D)$ is “sensitivity”

A ‘sensitive’ test is one which picks up on the disease when it’s there
(high sensitivity -> few false negatives)

$\mathbb{P}(\overline{T}|\overline{D})$ is “specificity”

A ‘specific’ test is one that is positive specifically because of the disease, and for no other reason (high specificity -> few false positives)

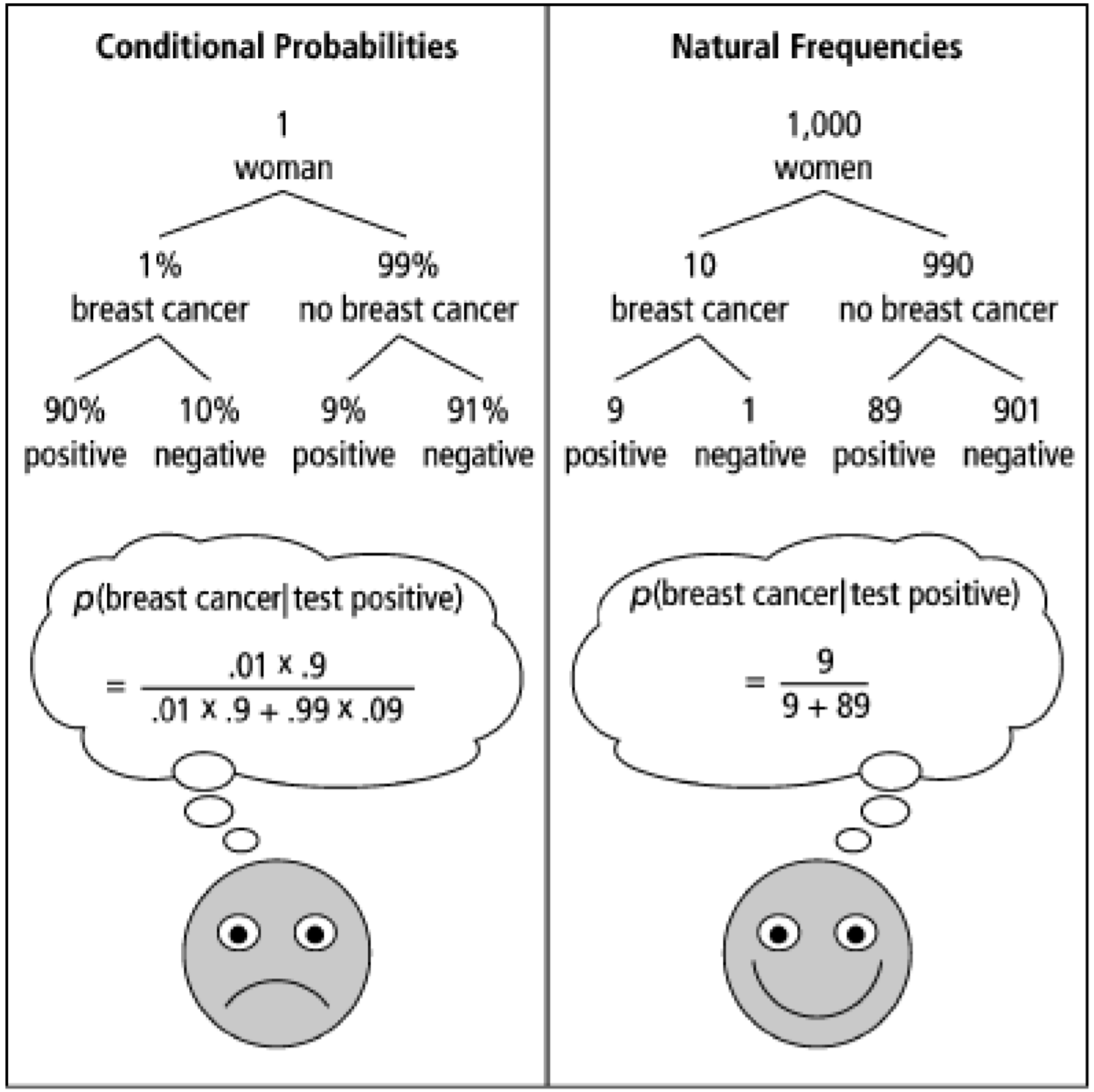
# How did the doctors do

C (about 1 in 10) was the correct answer.

Of the doctors surveyed, less than ¼ got it right (so worse than random guessing).

After the researcher taught them his calculation trick, more than 80% got it right.

# One Weird Trick!




Calculation Trick: imagine you have a large population (not one person) and ask how many there are of false/true positives/negatives.

# What about the real world?

When you’re older and have to do more routine medical tests, don’t get concerned (yet) when they ask to run another test.*

It’s usually fine.*

*This is not medical advice, Robbie is not a physician.

# Infinite Sample Spaces

# Implicitly defining $\Omega$

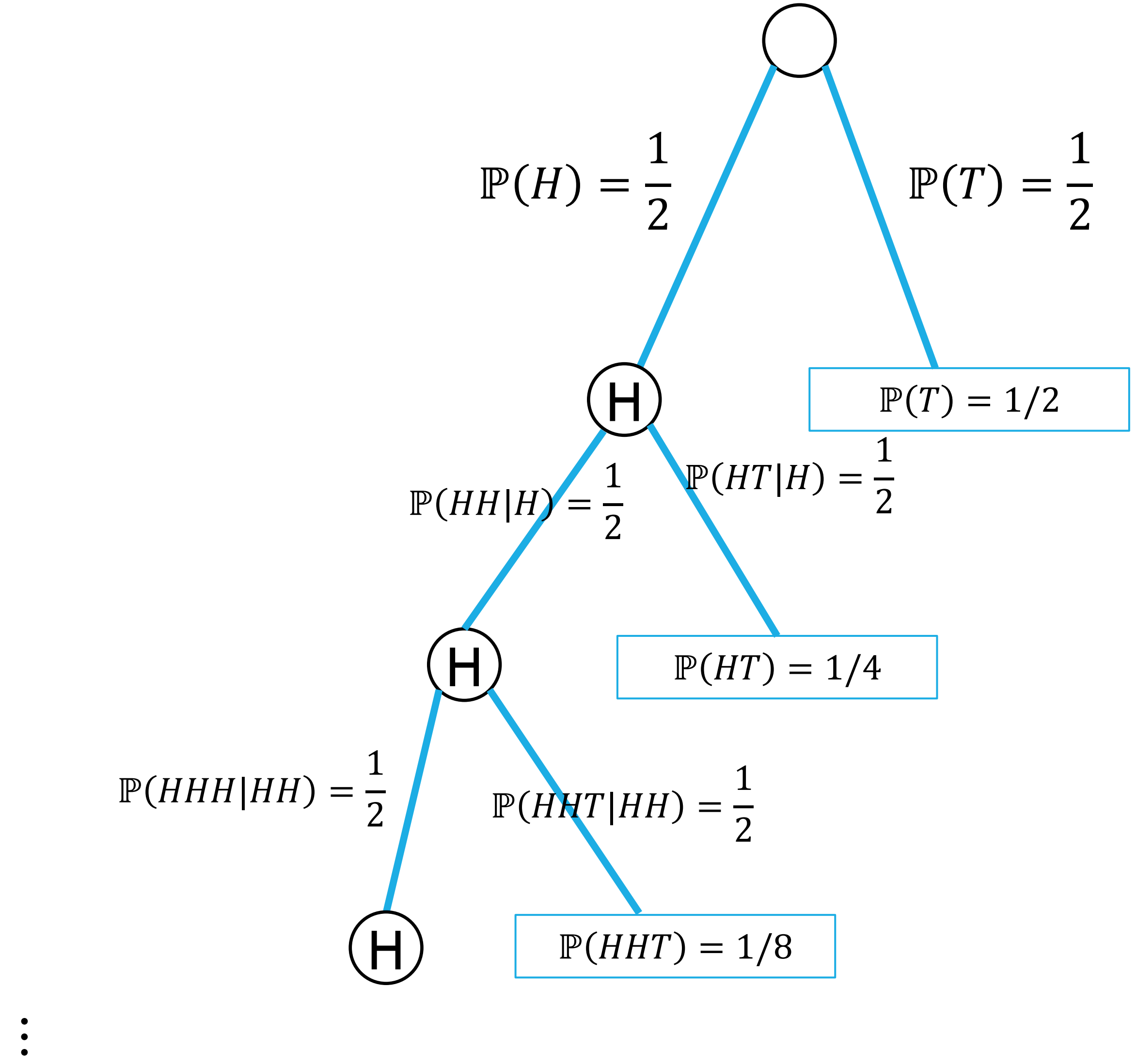
We've often skipped an explicit definition of $\Omega$.

Often $|\Omega|$ is infinite, so we really couldn't write it out (even in principle).

How would that happen?

Flip a fair coin (independently each time) until you see your first tails.
what is the probability that you see at least 3 heads?

# An infinite process.




$\Omega$ is infinite.

A sequential process is also going to be infinite…

But the tree is "self-similar"

To know what the next step looks like, you only need to look back a finite number of steps.

From every node, the children look identical (H with probability ½, continue pattern; T to a leaf with probability ½)

# Finding $\mathbb{P}$(at least 3 heads)

Method 1: infinite sum.

$\Omega$ includes $H^iT$ for every $i$. Every such outcome has probability $1/2^{i+1}$

What outcomes are in our event?

$\sum_{i=3}^{\infty} 1/2^{i+1} = \frac{\frac{1}{2^4}}{1-1/2} = \frac{1}{8}$

Infinite geometric series, where common ratio is between $-1$ and $1$ has closed form $\frac{\text{first term}}{1-\text{ratio}}$

# Finding $\mathbb{P}$(at least 3 heads)

Method 2:

Calculate the complement

$\mathbb{P}$(at most 2 heads) $= \frac{1}{2} + \frac{1}{4} + \frac{1}{8}$

$\mathbb{P}$(at least 3 heads)$= 1 - \left(\frac{1}{2} + \frac{1}{4} + \frac{1}{8}\right) = \frac{1}{8}$

# Optional: Careful Surveys

Another Real-World Bayes example

# Application 2: An Imbalanced Survey

In 2014, a paper was published

"Do non-citizens vote in U.S. elections?"

This is a real paper (peer-reviewed). It claims that

1. In a survey, about 4% (of a few hundred) of non-U.S.-citizens surveyed said they voted in the 2008 federal election (which isn't allowed).

2. Those non-citizen voters voted heavily (estimate 80+%) for democrats.

3. "It is likely though by no means certain that John McCain would have won North Carolina were it not for the votes for Obama cast by non-citizens"

# Application 2: What is this survey?

The “Cooperative Congressional Election Study” was run in 2008 and 2010.

It interviews about 20,000 people about how/whether they voted in federal elections.

Two strange observations:

1. The noncitizens are a very small portion of those surveyed. Feels a little strange.

2. Those people…maybe accidentally admitted to a crime?

# Application 2: Another Red Flag

A response paper (by different authors)

[“The perils of cherry picking low frequency events in large sample surveys”]

**Table 1**
Response to citizenship question across two-waves of CCES panel.

| Response in 2010 | Response in 2012 | Number of respondents | Percentage |
|---|---|---|---|
| Citizen | Citizen | 18,737 | 99.25 |
| Citizen | Non-Citizen | 20 | 0.11 |
| Non-Citizen | Citizen | 36 | 0.19 |
| Non-Citizen | Non-Citizen | 85 | 0.45 |

# An Explanation

Suppose $0.1\%$ of people check the wrong check-box on any individual question (independently)

Suppose you really interviewed $20{,}000$ people, of whom $300$ are really non-citizens (none of whom voted), and the rest are citizens, of whom $70\%$ voted. What is the probability someone appears to have voted

$$\mathbb{P}(say\ V|say\ NC) = \frac{\mathbb{P}(say\ NC|say\ V)\cdot\mathbb{P}(say\ V)}{\mathbb{P}(say\ NC)} = \frac{.001\cdot.7}{.999\cdot(\frac{300}{20000})+.001\cdot(\frac{19700}{20000})} \approx$$

$4.38\%$

# Conclusion

The authors of the original paper did know about response error…

…and they have an appendix that argues the population of “non-citizen” voters isn’t distributed exactly like you’d expect.

But with it being such a small number of people, this isn’t surprising.

And even they admit response bias played more of a role than they initially thought.

Though they still think they found some evidence of non-citizens voting (but not enough to flip North Carolina anymore).

# Takeaways

When talking about rare events (rare diseases, rare prize-winning-golden-tickets), think carefully about whether a test is really as informative as you think.

Do the explicit calculation

Intuition is easier if thinking about a large population of repeated tests, not just one.

Be careful of small subparts of large datasets

People from a large majority group (accidentally) clicking the wrong demographic information can “drown out” signal of a very small group.

# Optional: Bayes Factor

A way to **estimate Bayes** calculations quickly

# Bayes Factor

Another Intuition Trick: from 3Blue1Brown

When you test positive, you (**approximately**) multiply the prior by the "Bayes Factor" (aka likelihood ratio)

$$\frac{\text{sensitivity}}{\text{false positive rate}} = \frac{1 - FNR}{FPR}$$

# Bayes Factor

Does it work?

Let's try it…

Find

$$\mathrm{prior} \cdot \frac{\mathrm{Sensitivity}}{FPR}$$

# Wonka Bars

Willy Wonka has placed golden tickets on 0.1% of his Wonka Bars.

You want to get a golden ticket. You could buy a 1000-or-so of the bars until you find one, but that's expensive…you've got a better idea!

You have a test – a very precise scale you've bought.

- If the bar you weigh **does** have a golden ticket, the scale will alert you 99.9% of the time.
- If the bar you weigh does not have a golden ticket, the scale will (falsely) alert you only 1% of the time.

If you pick up a bar and it alerts, what is the probability you have a golden ticket?

# Wonka Bars

Bayes Factor

$$\frac{99.9}{1}$$

Prior: $.1\%$

Product: $9.99$, so about $10\%$

About what Bayes Rule gets!

# Application 1: Medical Tests

Helping Doctors and Patients Make Sense of Health Statistics

A researcher posed the following scenario to a group of 160 doctors:

Assume you conduct a disease screening using a standard test in a certain region. You know the following information about the people in this region:

The probability that a person has the disease is 1% (prevalence)

If a person has the disease, the probability that she tests positive is 90% (sensitivity)

If a person does not have the disease, the probability that she nevertheless tests positive is 9% (false-positive rate)

A person tests positive. She wants to know from you whether that means that she has the disease for sure, or what the chances are. What is the best answer?

A. The probability that she has the disease is about 81%.
B. Out of 10 people with a positive test, about 9 have the disease.
C. Out of 10 people with a positive test, about 1 have the disease.
D. The probability that she has the disease is about 1%

# Bayes Factor

What about with the doctors?

$$1\% \cdot \frac{90\%}{9\%} = 10\%$$

Again about right!

# Caution

Multiplying by the Bayes Factor is an **approximation**

It gives you the exact numerator for Bayes, but the denominator is "the number of false positives if the prevalence (/prior) were 0"

When the prior is close to 0, this is a fine approximation!

But plug in a prior of 15% on the last slide, and we get 150% chance.

# What about negative tests?

For negative tests, the Bayes Factor is

$$\frac{FNR}{\text{specificity}}$$

Specificity is $(1 - \text{false negative rate})$