Independence

CSE 312 Winter 26
Lecture 7

Quiz 1 Logistics: tomorrow in section

You must go to your officially-registered section for the quiz.

We'll give you a reference sheet

No other notes allowed for quizzes (midterm/final will allow your own notes).

Bring an ID to the quiz.

The directions are here, in case you want to know what that will look like

Explanations not required (unless specified), but they can help us give partial credit.

combinations, permutations, etc. ok in final answer; summations with … or Σ isn't fully simplified.

Conditioning Tool: Bayes’ Rule

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?

Willy Wonka

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

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.

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

Which of these is closest to the right answer?

A. 0.1%
B. 10%
C. 50%
D. 90%
E. 99%
F. 99.9%

Conditioning (Wonka Bars)

Let S be the event that the **S**cale alerts you

Let G be the event your bar has a **G**olden ticket.

What conditional probabilities are each of these?

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

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.

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

Conditioning (labels)

Let S be the event that the **S**cale alerts you

Let G be the event your bar has a **G**olden ticket.

What conditional probabilities are each of these?

Willy Wonka has placed golden tickets on 0.1% of his Wonka Bars. $\mathbb{P}(G)$

If the bar you weigh **does** have a golden ticket, the scale will alert you 99.9% of the time. $\mathbb{P}(S|G)$

If the bar you weigh does not have a golden ticket, the scale will (falsely) alert you only 1% of the time. $\mathbb{P}(S|\bar{G})$

You pick up a bar and it alerts, what is the probability you have a golden ticket? $\mathbb{P}(G|S)$

Reversing the Conditioning

All of our information conditions on whether G happens or not – does your bar have a golden ticket or not?

But we're interested in the "reverse" conditioning. We know the scale alerted us – we know the test is positive – but do we have a golden ticket?

Bayes' Rule

Bayes' Rule

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

Bayes' Rule (2)

Bayes' Rule

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

What do we know about Wonka Bars?

$$\mathbb{P}(G|S) = \frac{\mathbb{P}(S|G) \cdot \mathbb{P}(G)}{\mathbb{P}(S)}$$

Bayes' Rule (3)

Bayes' Rule

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

What do we know about Wonka Bars?

$$\mathbb{P}(G|S) = \frac{.999 \cdot .001}{\mathbb{P}(S)}$$

Filling In

What's $\mathbb{P}(S)$?

We'll use a trick called "the law of total probability":

$$\mathbb{P}(S) = \mathbb{P}(S|G) \cdot \mathbb{P}(G) + \mathbb{P}(S|\bar{G}) \cdot \mathbb{P}(\bar{G})$$

$$= 0.999 \cdot .001 + .01 \cdot .999$$

$$= .010989$$

Partition

Let $A_1, A_2, \dots, A_k$ be a **partition** of Ω.

A partition of a set S is a family of subsets $S_1, S_2, \dots, S_k$ such that:

$S_i \cap S_j = \emptyset$ for all i, j and

$S_1 \cup S_2 \cup \cdots \cup S_k = S$.

i.e. every element of Ω is in exactly one of the A_i.

Law of Total Probability

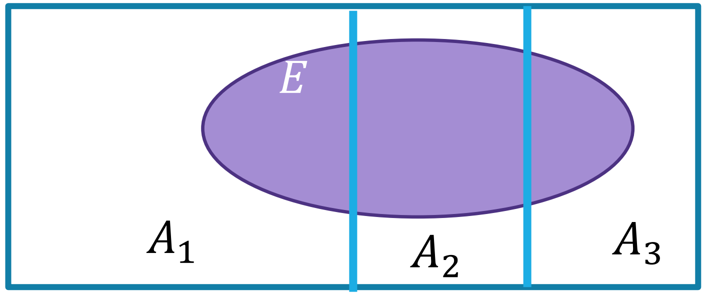
Law of Total Probability

Let $A_1, A_2, \ldots, A_k$ be a **partition** of Ω.

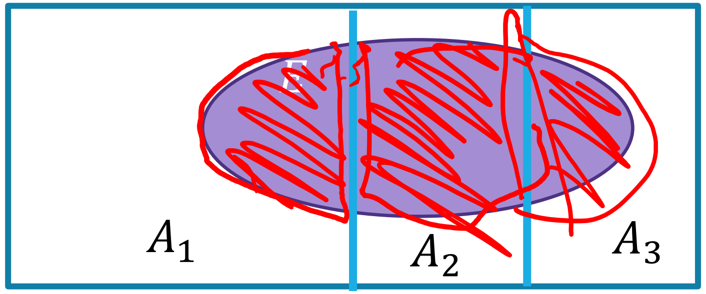
For any event E,

$$\mathbb{P}(E) = \sum_{\text{all } i} \mathbb{P}(E|A_i)\mathbb{P}(A_i)$$

Ω, split into partition A_1, A_2, A_3 with event E inside.



Why?



The Proof is actually pretty informative on what's going on.

$\sum_{\text{all } i} \mathbb{P}(E|A_i)\mathbb{P}(A_i)$

$= \sum_{\text{all } i} \frac{\mathbb{P}(E \cap A_i)}{\mathbb{P}(A_i)} \cdot \mathbb{P}(A_i)$ (definition of conditional probability)

$= \sum_{\text{all } i} \mathbb{P}(E \cap A_i)$

$= \mathbb{P}(E)$

The A_i partition Ω, so $E \cap A_i$ partition E. Then we just add up those probabilities.

Ability to add follows from the "countable additivity" axiom.

Wonka Bars Answer

What do we know about Wonka Bars?

$$\mathbb{P}(G|S) = \frac{.999 \cdot .001}{.010989}$$

Solving $\mathbb{P}(G|S) = \frac{1}{11}$, i.e. about 0.0909.

Only about a 10% chance that the bar has the golden ticket!

Wait a minute…

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

That doesn't fit with many of our guesses. What's going on?

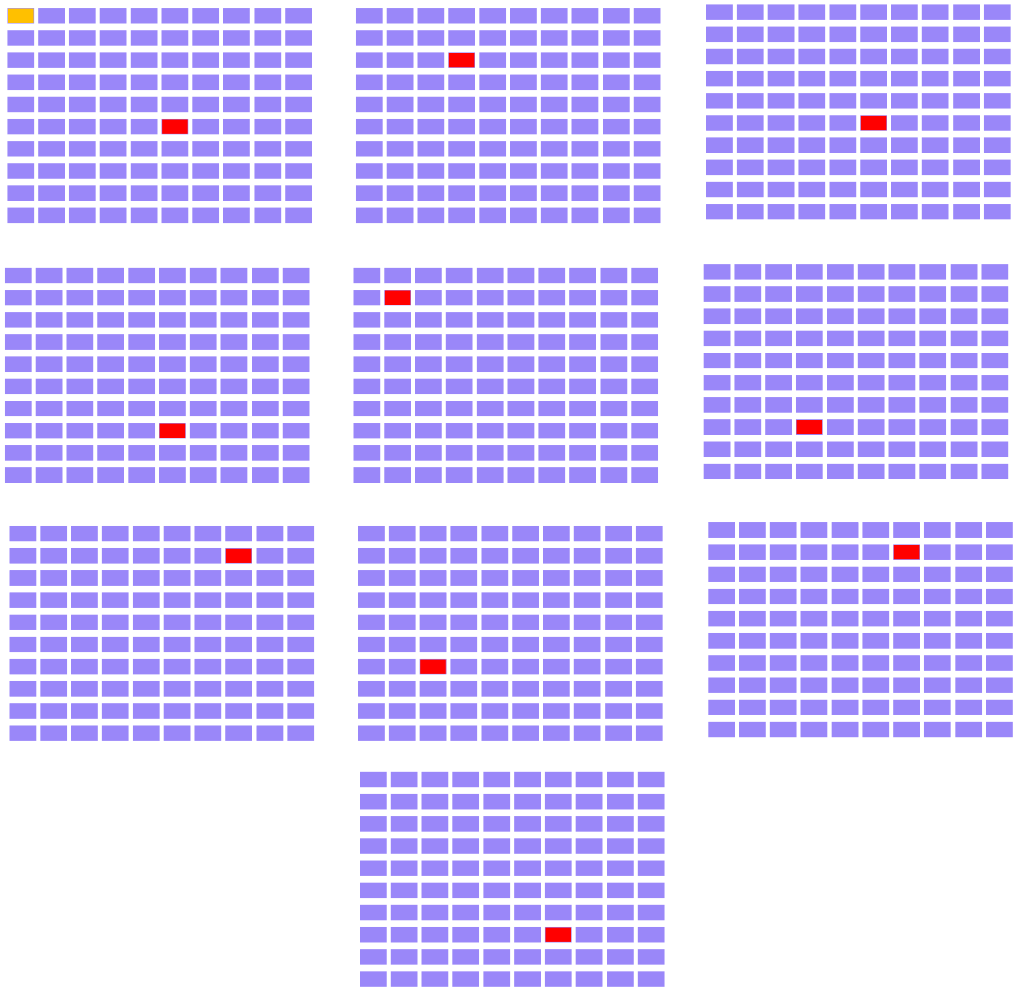
Instead of saying "we tested one and got a positive" imagine we tested 1000. **ABOUT** how many bars of each type are there?

(about) 1 with a golden ticket 999 without. Let's say those are exactly right.

Let's just say that one golden is truly found.

(about) 1% of the 999 without would be a positive. Let's say it's exactly 10.

Visually



Gold bar is the one (true) golden ticket bar.
Purple bars don't have a ticket and tested negative.
Red bars don't have a ticket, but tested positive.

The test is, in a sense, doing really well.
It's almost always right.

The problem is it's also the case that the correct answer is almost always "no."

Updating Your Intuition (Take 1)

🔥 Take 1: The test is **actually good** and has VASTLY increased our belief that there IS a golden ticket when you get a positive result.

If we told you "your job is to find a Wonka Bar with a golden ticket" without the test, you have 1/1000 chance, with the test, you have (about) a 1/11 chance. That's (almost) 100 times better!

This is actually a huge improvement!

Updating Your Intuition (Take 2)

🔥 Take 2: Humans are really bad at intuitively understanding very large or very small numbers.

When I hear “99% chance”, “99.9% chance”, “99.99% chance” they all go into my brain as “well that’s basically guaranteed” And then I forget how many 9’s there actually were.

But the number of 9s matters because they end up “cancelling” with the “number of 9’s” in the population that’s truly negative.

Updating Your Intuition (Take 3)

🔥 Take 3: View tests as updating your beliefs, not as revealing the truth.

Bayes' Rule says that $\mathbb{P}(B|A)$ has a factor of $\mathbb{P}(B)$ in it. You have to translate "The test says there's a golden ticket" to "the test says you should increase your estimate of the chances that you have a golden ticket."

A test takes you from your "prior" beliefs of the probability to your "posterior" beliefs.

Independence of Events

Definition of Independence

We've calculated conditional probabilities.

Sometimes conditioning – getting some partial information about the outcome – doesn't actually change the probability.

We already saw an example like this…

Conditioning Practice $P(A|B) = P(A)$

A: Red die 6 conditioned on sum 7 1/6

B

Red die 6 conditioned on sum 9 1/4

Sum 7 conditioned on red die 6 1/6

Red die 6 has probability 1/6 before or after conditioning on sum 7.

	D2=1	D2=2	D2=3	D2=4	D2=5	D2=6
D1=1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1.6)
D1=2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
D1=3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
D1=4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
D1=5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
D1=6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Independence (definition)

Independence

Two events A, B are independent if

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

You'll sometimes see this called "statistical independence" to emphasize that we're talking about probabilities (not, say, physical interactions).

If A, B both have non-zero probability then

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

$P(A|B) = P(A)$

$\frac{P(A \cap B)}{P(B)}$

Examples

We flip a fair coin three times. Each flip is independent. (both in the statistical independence sense and in the "doesn't affect the next one" sense).

Is $E = \{HHH\}$ independent of $F =$"at most two heads"?

Are $A =$"the first flip is heads" and $B =$"the second flip is tails" independent?

Examples (answers)

Is $E = \{HHH\}$ independent of $F =$"at most two heads"?

$\mathbb{P}(E \cap F) = 0$ (can't have all three heads and at most two heads).

$\mathbb{P}(E) = 1/8$, $\mathbb{P}(F) = 7/8$, $\mathbb{P}(E \cap F) \neq \mathbb{P}(E)\mathbb{P}(F)$.

Are $A =$"the first flip is heads" and $B =$"the second flip is tails" independent?

$\mathbb{P}(A \cap B) = 2/8$ (uniform measure, and two of eight outcomes meet both A and B.

$\mathbb{P}(A) = 1/2$, $\mathbb{P}(B) = 1/2$; $\frac{2}{8} = \frac{1}{2} \cdot \frac{1}{2}$. These are independent!

Hey Wait

I said “the flips are independent” why aren’t E, F independent?

“the flips are independent” means events like “the first flip is <blah>” are independent of events like “the second flip is <blah>”

But if you have an event that involves both flip one and two that might not be independent of an event involving flip one or two.

Mutual Exclusion and independence

Two of these statements are true, one is false. Explain to each other which ones are true, and find a counter-example to the false one.

1. If A, B both have nonzero probability and they are mutually exclusive, then they cannot be independent.

2. If A has zero probability, then A, B are independent (for any B).

3. If two events are independent, then at least one has nonzero probability.

Conditional Independence

Conditional Independence (definition)

We say A and B are conditionally independent on C if

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

i.e. if you condition on C, they are independent.

Conditional Independence Example

You have two coins. Coin A is fair, coin B comes up heads with probability 0.85.

You will roll a (fair) die, if the result is odd flip coin A twice (independently); if the result is even flip coin B twice (independently)

Let C_1 be the event "the first flip is heads", C_2 be the event "the second flip is heads", O be the event "the die was odd"

Are C_1 and C_2 independent? Are they independent conditioned on O?

(Unconditioned) Independence

$\mathbb{P}(C_1) = \mathbb{P}(O)\mathbb{P}(C_1|O) + \mathbb{P}(\bar{O})\mathbb{P}(C_1|\bar{O})$

$= \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot 0.85 = .675$

$\mathbb{P}(C_2) = .675$ (the same formula works)

$\mathbb{P}(C_1)\mathbb{P}(C_2) = .675^2 = .455625$

$\mathbb{P}(C_1 \cap C_2) = \mathbb{P}(O)\mathbb{P}(C_1 \cap C_2|O) + \mathbb{P}(\bar{O})\mathbb{P}(C_1 \cap C_2|\bar{O})$

$= \frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot .85^2 = .48625$

Those aren't the same! They're not independent!

Intuition: seeing a head gives you information – information that it's more likely you got the biased coin and so the next head is more likely.

Conditional Independence (computation)

$\mathbb{P}(C_1|O) = 1/2$

$\mathbb{P}(C_2|O) = 1/2$

$\mathbb{P}(C_1 \cap C_2|O) = \frac{1}{2} \cdot \frac{1}{2} = 1/4$

$\mathbb{P}(C_1|O)\mathbb{P}(C_2|O) = \mathbb{P}(C_1 \cap C_2|O)$

Yes! C_1 and C_2 are conditionally independent, conditioned on O.

Takeaway

Read a problem carefully – when we say “these steps are independent of each other” about some part of a sequential process, it’s usually “conditioned on all prior steps, these steps are conditionally independent of each other.”

Those conditional steps are usually dependent (without conditioning) because they might give you information about which branch you took.

Chain Rule

Chain Rule (definition)

We defined conditional probability as: $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$

Which means $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$

Chain Rule

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n)$$
$$= \mathbb{P}(A_n | A_1 \cap \cdots \cap A_{n-1}) \cdot \mathbb{P}(A_{n-1} | A_1 \cap \cdots \cap A_{n-2}) \cdots \mathbb{P}(A_2 | A_1) \cdot \mathbb{P}(A_1)$$

$\mathbb{P}(A_3 | A_1 \cap A_2)$

Chain Rule Example

Shuffle a standard deck of 52 cards (so every ordering is equally likely).

Let A be the event "The top card is a K♦"

Let B be the event "the second card is a J ♠

Let C be the event "the third card is a 5 ♠

What is $\mathbb{P}(A \cap B \cap C)$?

Use the chain rule!

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

$$\frac{1}{52} \cdot \frac{1}{51} \cdot \frac{1}{50}$$

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.

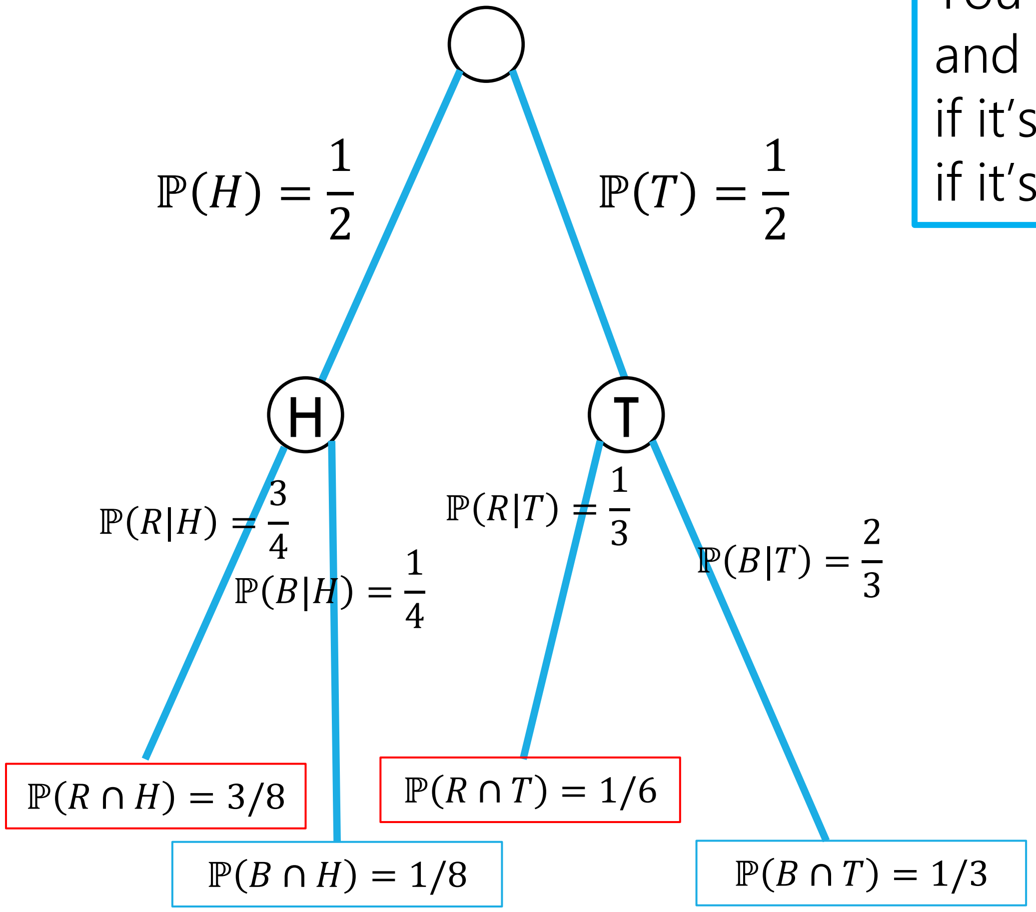
You will flip a fair coin; if it's heads, you'll draw a marble (uniformly) from your left pocket, if it's tails, you'll draw a marble (uniformly) from your right pocket.

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

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

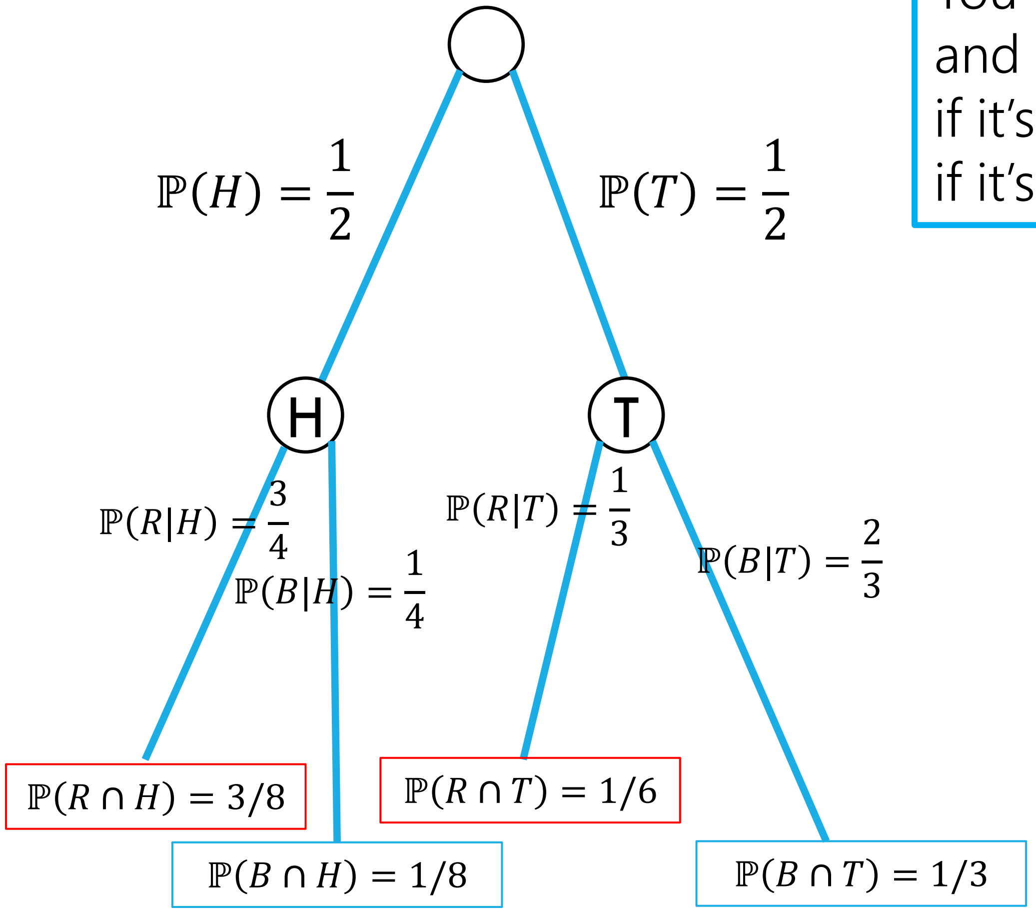
Updated Sequential Processes

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
$\mathbb{P}(\text{next step} \mid \text{all} \cap \text{prior} \cap \text{steps})$

Updated Sequential Processes (answer)



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 (marbles)

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 (marbles, answer)

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.

Setting the stage: Random Variables

Implicitly defining Ω

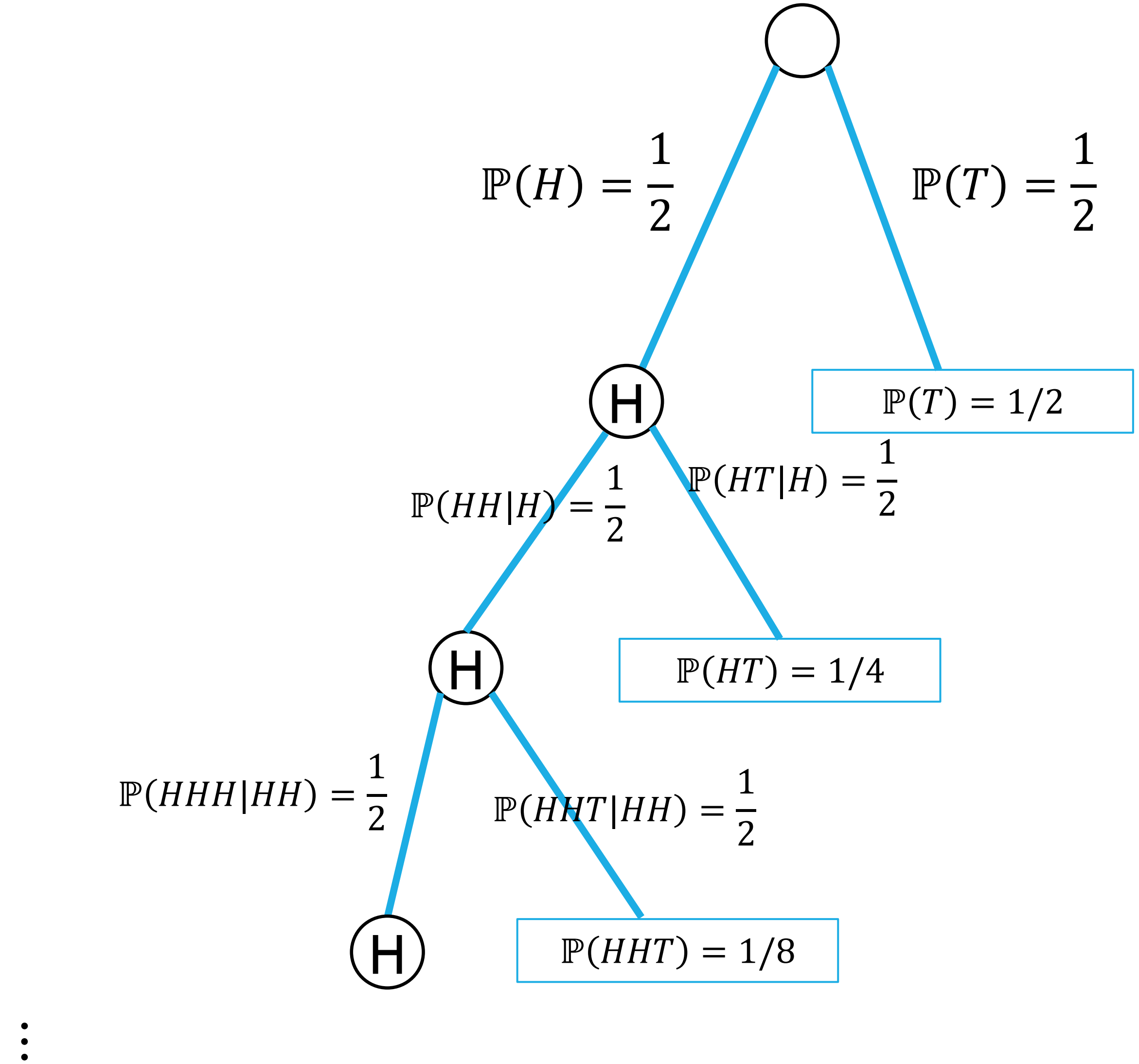
We've often skipped an explicit definition of Ω.

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.



Ω is infinite.

A sequential process is also going to be infinite…

But the tree is "self-similar"

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

Method 1: infinite sum.

Ω 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

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)$= 1 - \left(\frac{1}{2} + \frac{1}{4} + \frac{1}{8}\right) = \frac{1}{8}$