[SQUEAKING]

[RUSTLING]

[CLICKING]

BRYNMOR CHAPMAN: Hello. Can people hear me in the back? Yeah, thumbs up. OK. So today, we're going to be continuing our discussion of probability. In the last lecture, we saw how to compute probabilities--

[BANG]

--using the-- oh, dear-- using the tree method. And in this lecture, we are going to see how some of the tools that we learned in the last unit for counting, for reasoning about the sizes of sets and things like that, can also carry over naturally into the realm of probability.

And we'll also learn how to express mathematically statements like, if you recall, I was saying things last time like if the prize is behind door A, what is the probability that Marilyn selects door B or something like that? So we're going to see more formally what these statements are trying to express and how to reason about them.

So recall how we defined the probability of an event A. So an event is just a set of outcomes. Our probability function, Pr, is a function from the set of outcomes, the sample space, to the interval 0 to 1. And we can extend this naturally to events by saying that the probability of an event is just the sum of the probabilities of all of the outcomes in that event.

So an immediate consequence of this definition, or several immediate consequences I suppose, are the following. So first, we have the sum rule. So if we have two events, A and B, and they're disjoint-- so A and B are disjoint events-- this tells us that the probability of the union of A and B is, what? What do we think that is? Any ideas? Yeah?

AUDIENCE: Sum of the probabilities.

BRYNMOR CHAPMAN: Yeah, the sum of the probabilities. So probability of A plus the probability of B. So very similar to what we saw with sets. If you have two disjoint sets, the size of their union is the same as the sum of the individual sizes. The same thing happens with probability.

We also have the complement rule. So if we have some event A, what do we think the probability of the complement of A is, so the probability that A doesn't occur? So I should probably-- yeah. What's the probability of not A? Anybody? Yeah?

AUDIENCE: 1 minus the probability of A.

BRYNMOR CHAPMAN: Yeah, 1 minus the probability of A because A and not A are disjoint. And one of them must happen. So with certainty, one of A or not A happens. So the probability of their union is 1. So this kind of just follows from the sum rule.

What else do we have? The difference rule. So in this case, suppose we have, again, two events, A and B. And this time, we'd like to look at the difference A set minus B. What do we think that might be in terms of the probabilities of A and B? Yeah?

AUDIENCE: The probability of A minus the intersection of A and B.

BRYNMOR CHAPMAN: Exactly. So the probability of A minus the probability of the intersection. Why is that? Well, A is just the disjoint union of A minus B, and A intersects B.

We also have an analog of principle of inclusion/exclusion. So the probability of A union B for more general A and B. So they're not necessarily disjoint anymore.

Just like we had with sets, this is going to be the probability of A plus the probability of B minus the probability of the intersection. And we have a couple of new rules that don't really have analogs in the case of counting sets.

So first, we have the union bound. Oops. So probability of A union B is less than or equal to probability of A plus probability of B. Or I suppose I shouldn't say it doesn't really have an analog, but we didn't really look at the analog. And why might that be the case? Yeah?

AUDIENCE: I have a question.

BRYNMOR CHAPMAN: Oh, yeah.

AUDIENCE: What's the difference between the sum rule and pi?

BRYNMOR CHAPMAN: So the question is, what's the difference between the sum rule and pi? So in the sum rule, we assume that A and B are disjoint. So if A and B are disjoint, then their intersection is empty. So the intersection has zero probability.

So pi is just a way to generalize the sum rule to the case where they're not necessarily disjoint anymore. They may have a nontrivial intersection. Good question. Any other questions? No? OK. So can anybody give a reason why the union bound might be true? Yeah?

AUDIENCE: Because the probability of the intersection is non-negative.

BRYNMOR CHAPMAN: Yeah, so the answer was the probability of the intersection is non-negative. So if we just take pi, so this holds with equality. And if we just get rid of this negative term here, we can only increase the probability because all probabilities have to be between 0 and 1.

We also have monotonicity. Again, I don't think we saw an analog of this with sets. But if A is a subset of B, what do you think that tells us about the probabilities? Any ideas?

I'm seeing some hand motions. So as you might expect from the name, if is A subset of B, then its probability should be smaller than the probability of B. So do these make sense to everybody?

So just as we saw with sets, both the sum rule, the union bound, and the principle of inclusion/exclusion can all generalize to arbitrary numbers of events or even countably many. I suppose principle of inclusion/exclusion, you should probably only think about finite cases. But the other two, they fairly, naturally, generalize to countable sets of events.

So do these rules make sense to everybody? You'll probably be using them quite a lot. So make sure you're familiar with them. Any questions? So in that case, let us move on to the main topic of today's lecture-- conditional probability.

So if you recall on Tuesday when we were talking about the Monty Hall problem, when we were drawing our tree diagram, we were making statements such as, if the car is behind door 1, what's the probability that we pick door 2? Or I guess it was purple, gold, and green-- or purple, gold, silver? Silver, yeah.

So how do we express this mathematically? So that's where conditional probability comes in. We're going to make use of the following definition. So if we have two events, A and B, we define the conditional probability of A given B, which we denote as the following.

So we say A conditional B. We denote that with a vertical bar. We define this to be the probability of A intersect B divided by the probability of B.

So intuitively, you can think of it a little bit as we're trying to scale our probability space so that B becomes our entire sample space. So we kind of want to divide out this probability of B so that we still get a probability space, all of the probabilities sum to 1. Does that make sense to everybody?

And then the analog of A, we're only looking at everything that's already within B. So we're looking at the intersection of A and B. Now, if we rewrite this, we can get a product rule.

So if we want to know the probability of A intersect B, we can just move this probability of B over to the other side. And we end up with probability of A intersect B equals probability of A given B times the probability of B.

And as you might expect, you can extend this to multiple events. Or I guess this is already multiple events. You can extend it to more than two events. So you can also have, e.g., the probability of A intersect B intersect C equals the probability of A given B and C times the probability of B-- or maybe I should write this out more explicitly-- B intersect C.

And then if we apply our binary product rule again, we can expand out this term as the probability of A given B and C, probability of B given C, probability of C. So that's basically what we were doing on Tuesday written out more formally.

So probability of C corresponds to what is the first edge we chose, the first edge going away from the root. And then conditioned on reaching a particular vertex, like conditioned on Monty hiding the prize in the gold box, what's the probability that Marilyn chooses the silver one?

So that's the conditional probability of choosing the silver box, conditioned on the prize being in the gold box. And then, as you continue along the path to a leaf, all of the edge probabilities that you're looking at are secret-- they were secretly just basic conditional probabilities. Does that make sense to everybody? Any questions?

So another extension of this that you may find useful is the following-- A intersect B given C is going to be probability of A given B intersect C probability of B given C. So can anybody see why that might be the case?

Well, first maybe we should look at, what exactly is it saying? So if you squint a little bit and get rid of the C's, it kind of looks a little bit like the product rule. A intersect B is probability of A given B times probability of B. It looks kind of like that. But we've added a bunch of C's everywhere.

So basically what we're saying here is this is the product rule in the probability space where we're assuming C. It's the product rule conditioned on C. Does that make sense to everybody? And we can obtain it from the previous line just by dividing out by the probability of C.

Do people have questions? Are we happy to move on to some examples? Yeah?

AUDIENCE: Sorry.

BRYNMOR CHAPMAN: Sorry.

AUDIENCE: Is it sets on the left board? Or is it still events [INAUDIBLE] on the board?

BRYNMOR CHAPMAN: Oh, OK. So the question was, are we talking about sets or events? Yes. So remember that events are defined to be sets of outcomes. So we're talking about events. But events are sets, and we can treat them as such.

So that's why we're using lots of notation here that you would typically associate with sets, like union intersection, et cetera. Any other questions? Oh, yeah?

AUDIENCE: So if you have A and B and C [INAUDIBLE] or is it more like a single node of [INAUDIBLE]?

BRYNMOR CHAPMAN: So these are not nodes in the tree. Oh, sorry. Question was, if we have an event that's like A union B union C or something, are these general sets? Or are they nodes in the tree? That is a good observation.

Some of the nodes in the tree naturally correspond to events. Like if we draw out our tree diagram, this node here fairly naturally corresponds to this set of events-- or sorry, this set of outcomes, those two leaves in its subtree. But not every event necessarily corresponds to an internal node of the tree like that.

For instance, you could have these two outcomes. If you consider the set of those two outcomes, that's still a valid event. But it doesn't really naturally correspond to a tree node in the same way. Does that make sense?

AUDIENCE: I think you misunderstood my question.

BRYNMOR CHAPMAN: Oh, that is quite possible.

AUDIENCE: I was meaning the probability of A given B or A given C, [INAUDIBLE] is that still a similar set? Or how that would be different? But I think it would just be helpful to see what happened on the tree.

BRYNMOR CHAPMAN: OK. So asking about the conditional probabilities as they appear in the tree.

TA: Brynmor?

BRYNMOR CHAPMAN: Yes?

TA: What is A given B? Is that an event? Is it a set? How can we describe it, not as a probability but--

BRYNMOR CHAPMAN: Oh, I see. OK. So I guess the way that I like to think about A given B, it is an event in a different probability space. So you can think of-- if you limit the scope-- rather than talking about your entire sample space S, if you limit the scope to just C, so you scale all of the probabilities down by the probability of C, you end up with a different probability space. So now the probability of-- or sorry, the event B given C is an event in that probability space. Does that make sense?

AUDIENCE: Yeah.

BRYNMOR CHAPMAN: OK. Yeah, sorry. [CHUCKLES] Thank you for the clarification. So why don't we move on to some examples then if this makes reasonable amount of sense. Hopefully it'll make more sense after examples.

So suppose Ash and Gary are having a bunch of Pokemon battles against each other. And they're doing a tournament. The first to win two battles wins the-- or I guess the series. The first to win two battles wins the series.

And the probabilities of victory are not entirely uniform. So for the first match, maybe it's a toss-up. Ash wins with probability one half. Garry wins with probability one half.

But after the first battle, there's an advantage to whoever won the last battle. Their morale is high or something. They do better. So they now have probability of 2/3 of winning the next battle. And there aren't any draws. They're bitter rivals. They will fight to the death.

So let's define a couple of events. We'll say that A is the event Ash wins the series. And B is the event Ash wins the first battle.

So now we can ask for the probability of A given B. So does the problem setup make sense? Yeah? OK. So how could we compute this? Any ideas? Where do we want to start? Should we just throw some numbers out there? Oh, yeah?

AUDIENCE: Draw a tree diagram.

BRYNMOR CHAPMAN: Yeah, let's draw a tree diagram. Nice one. OK, so drawing a tree diagram, what's the first random thing that happens? What's the first source of randomness? Yeah?

AUDIENCE: [INAUDIBLE].

BRYNMOR CHAPMAN: Yeah. So Ash could win or lose the first battle. So we can actually label these as B and not B. What's the next random thing that could happen? Yeah?

AUDIENCE: [INAUDIBLE].

BRYNMOR CHAPMAN: Yeah, that's right. So actually, maybe it's better to call these just W and L so we can use the same thing everywhere. So win, lose, win, lose.

Regardless of whether he wins or loses the first battle, he can still either win or lose the second one. Now, what's the third layer going to be?

AUDIENCE: If he's won two, it's over. If he's won one and he's lost one, you go again and split it into two separate outcomes.

BRYNMOR CHAPMAN: Exactly.

AUDIENCE: Lost two, that means that he's lost two.

BRYNMOR CHAPMAN: Yeah, exactly. So if he's won two or lost two, those are just leaves now. We don't do anything more. The series is over. But if he's won one and lost one in either order, we can now split again. So that's a little bit long, but, oh, well. W, L, W, L.

So now what are our outcomes? Again, let's look at the paths from root to leaves and just label with a string of edge labels. So this is going to be W comma W; W, L, W; W, L, L; L, W, W; L, W, L; and L, L.

Now, what are the probabilities on each of the edges? So in the first layer, what's the probability that Ash wins the first game? First match. Yeah?

AUDIENCE: One half.

BRYNMOR CHAPMAN: One half. So each of these is probability one half. Now, the second layer, we said, are now conditional probabilities. So conditioned on winning the first match, what's the probability of winning the second match? What's the probability on this edge here? Do we remember?

AUDIENCE: 2/3.

BRYNMOR CHAPMAN: 2/3, yeah. And so the other branch is going to be the complement of that, so one third. And then for the bottom branch, it's going to be exactly the opposite. It's going to be one third chance of winning, 2/3 chance of losing. Now, conditioned on winning the first match and losing the second match, what's the probability of winning the third? Yeah?

AUDIENCE: One third.

BRYNMOR CHAPMAN: One third, yeah. So one third win, 2/3 loss. And this will be exactly the opposite, so 2/3 win, one third loss. So now we can assign probabilities to all of the outcomes. What's the probability of winning twice? How do we compute these probabilities? Anyone? Yeah?

AUDIENCE: Why is the probability of winning [INAUDIBLE]?

BRYNMOR CHAPMAN: Oh, so this was just what we assumed in the problem. If it's not the first game, you win with probability 2/3 if you won the last game or with probability one third if you lost the last game.

So the first branch is 50-50. But the subsequent ones are going to be 2/3, one third, in one direction or the other, depending on the outcome of the previous game-- maybe I shouldn't use "outcome"-- depending on the result of the previous game. OK, so how do we compute probabilities of outcomes? Does anybody remember? Yeah?

AUDIENCE: Do you just multiply to that each outcome-- or, yeah, you multiply the probabilities along the tree to that outcome.

BRYNMOR CHAPMAN: Yeah, so you multiply probabilities along paths to outcomes. So for this top outcome, we've got one half times 2/3. So this is probability one third. Win, loss, win has probability one third by one half by one third. So that's 1/18.

Win, loss, loss says probability one half times one third times 2/3. So that's going to be one ninth. And then these all reverse-- 1/18 and one third. Have I done that correctly? Yes, that looks fine. If I make an error, please shout at me.

OK, so we've got our tree diagram. We've got our sample space. We've got our probability function. What's the next thing we need? Do we remember?

So we've got, step 1, tree; step 2, the probabilities; 3 was the events of interest. So what are the events we care about here? Yeah?

AUDIENCE: Winning the tournament [INAUDIBLE].

BRYNMOR CHAPMAN: Pardon?

AUDIENCE: Winning the tournament and winning the first game.

BRYNMOR CHAPMAN: OK, winning the tournament and winning the first game. So A was-- yep, A was winning the tournament. B was winning the first game. Is that all? Are there any other events that we care about? Yeah?

AUDIENCE: A intersect B.

BRYNMOR CHAPMAN: Yes, exactly. So we also want A intersect B. We also care about the event where you win the first battle and the series. Why is that? Would you like to explain?

AUDIENCE: [INAUDIBLE].

BRYNMOR CHAPMAN: Yeah, exactly. So ultimately, we want this conditional probability, the probability of A given B. And if we go back to our definition, the probability of A given B is the probability of A intersect B divided by the probability of B.

So as it turns out, we don't really need A. The two probabilities we really need are these two, are the two events I suppose. So finally-- 4, answer.

What is our answer? Well, probability of A given B, as we just said, is the probability of A intersect B divided by probability of B. What are each of these probabilities?

Well, which events-- or sorry, which outcomes are in the event A intersect B? It's W, W in A intersect B. Yeah? Win, loss, win-- also yes. Win, loss, loss-- nope. And all of these, he lost the first game. So these are also not in it.

So we care about these two outcomes. So this is going to be, by the sum rule, one third plus 1/18. So that doesn't look much like an 8. Oh, well. Now what's the probability of B? Yeah?

AUDIENCE: One half.

BRYNMOR CHAPMAN: Yeah, it's just one half, either by symmetry, or you can actually look at what those outcomes are. It's these top three. And if we do this arithmetic, we've got 7/18 divided by one half, which is 7/9.

So the probability that Ash wins the entire series given that he won the first battle is going to be 7/9. So far this is pretty straightforward, fairly routine mathematical-- well, arithmetic. But what if, instead, we asked for the probability of B given A?

What is that probability? Well, numerically, we can compute it in exactly the same way. It's going to be the probability of the intersection divided by, this time, the probability of A rather than the probability of B.

But still, by symmetry, that probability is going to be one half. So this is also going to be 7/9. But what exactly does it mean? Yeah?

AUDIENCE: The probability that-- he already won the series. What was the probability that he won the first battle?

BRYNMOR CHAPMAN: Yeah, so the answer was, given that he already won the series, what's the probability that he won the first battle? Now, this might sound a little bit weird to some of you. Like, if we know that he's won the series, everything's already happened. He either won the battle or he didn't.

But when we're modeling probability questions, we don't really care about any temporal relations or anything like that, or even causal ones. It's all just numerical. You can think of it a bit like an inference. You've got some uncertainty.

Maybe you're told that he won the battle. Or, sorry, you're told that he won the series. You want to update your belief that maybe he won the first battle, maybe he didn't. And maybe you're not entirely confident in it. But you'd like to estimate how confident you should be. That might be one way to think about it.

So in a sense, it's like looking at a backwards probability, that something in the past has happened given that we observed something in the future. So in general, they won't always be the same like this. You can relate the probability of B given A to the probability of A given B. They won't always be equal.

Let's see how to relate them. It just so happens that in this case, these two things-- these two things had the same probability. So that's really the core reason why they were equal.

But more generally, we have the following. This is called Bayes' rule. You can think of it as it's just the product rule in disguise. So the probability of B given A equals the probability-- oops, why don't I split this into two lines-- probability of A given B, probability of B divided by probability of A.

So in the example we just saw, these two are the same. So they just cancel. But more generally, you want to multiply by the quotient. And you can think of this as the product rule in disguise. Like if you just move this probability of A over to this side, now you're just computing the probability of the intersection of A and B on both sides. Does that make sense?

But it can often be useful-- even though it's not really saying anything new, it can often be useful to think about it in this way. If you want to reverse the direction of an inference, you want to multiply by the quotient of the two probabilities of the events that you care about.

And one particularly useful consequence of this is the following. So suppose we have two events, B and C. And we're trying to condition them both on the same event. So what is the probability of C given A-- or B given A divided by the probability of C given A?

And in particular, often we would take C to be the complement of B or something like that. And that will then help us to compute this probability of B given A. So as it turns out, this is going to be the probability of A given B times the probability of B divided by probability of A given C-- oops-- probability of C.

So does that make sense to everybody? And the reason that this in particular might be useful-- in order to compute the probability of B, if we use B complement to C, this right-hand side here does not depend on the probability of A. We don't have to compute the probability of A.

So maybe if A is some really nasty event and we don't actually want to compute its probability but we can still condition on it, we might still be able to use this formula more easily than we could use this one. Does that make sense to everybody? So let's take a look at a couple of applications of Bayes' rule.

Or actually, first-- mm, first, maybe I should give you a brief lesson in statistics. So terminology-- the probability of A given B, this would often be called the likelihood of A given B. So if any of you have ever seen maximum likelihood estimations or anything like that, that's where this comes from. Or this is where that comes from, sorry.

Probability of B we call the prior probability of B. And the left-hand side, the probability of B given A, is the posterior of B. Actually, we don't usually specify it with that. Oops.

So this is how you would normally talk about it in statistics or something like that-- the likelihood of A given B, the prior probability, and the posterior probability-- because basically what you're doing is you're starting with some prior. If you don't know anything, this is what you think the probability of B should be. And now you make some observation A. And so now you update, and you get a posterior probability of B conditioned on that observation. Does that make sense?

OK, so why don't we take a look at a couple of examples then? Hopefully everybody is OK with those. So first example, suppose I've got two coins. I've got one fair coin. I toss this coin, and it'll come up heads with 50% probability, tails with 50% probability.

And I've also got a trick coin. It's double heads. I flip it, I always get heads. Suppose I pick one of these coins uniformly at random, I flip that coin, and it comes up heads. What is the probability that I picked the fair coin to begin with, as opposed to the biased one?

Any ideas? Does anybody have any guesses as to what the probability is? Who wants to just draw the tree diagram? A few people. People aren't as enthusiastic about it as they were on Tuesday. That's slightly unfortunate. Let's do it anyway.

Now, let's make precise the events that we care about. So let's let H be the event that we flipped heads. And F is the event picked the fair coin. So what is the probability that we're trying to compute in terms of these two quantities? Yeah?

AUDIENCE: Probability of F given H.

BRYNMOR CHAPMAN: Probability of F given H, yes. OK, so how might we compute this? People didn't seem too enthusiastic about the tree diagram. So what else could we do instead? Does anybody see a way to use Bayes' rule? Yeah? Or just a stretch? Sorry. [CHUCKLES]

Yeah?

AUDIENCE: The probability of H [INAUDIBLE].

BRYNMOR CHAPMAN: Yeah, the probability of H given F is easy. That's 50%. So that's an easy known quantity. So if we use Bayes' rule, that might make it easier to compute.

So actually, why don't we use the slightly more complicated version of Bayes' rule? So let's compute the ratio of the probability that we picked the fair coin given that we observed a heads versus the probability of picking the biased coin given that we observed a heads.

So the probability of F given H over probability of not F given H. So what was that? So Bayes' rule is about flipping the order. So instead of F conditioned on H, we want H conditioned on F, multiplied by what? Yeah?

AUDIENCE: Probability of F.

BRYNMOR CHAPMAN: Yeah, probability of f. And on the bottom, we can do the same thing. Probability of H given not F times the probability of not F. So this expression is much easier to work with. What's the probability of H given F? Yep?

AUDIENCE: One half.

BRYNMOR CHAPMAN: One half, yeah. What about the probability of H given not F? Yeah?

AUDIENCE: One.

BRYNMOR CHAPMAN: One. And probability of F and not F? Those are each one half by assumption. So those just cancel. And we are left with one half.

So now, what is this probability? We computed the ratio of this probability to its complement. So what does that tell us the probability is? Yeah?

AUDIENCE: One third.

BRYNMOR CHAPMAN: One third, yeah. So observe here that if our prior changes, if the chance that we are picking the fair coin to begin with changes, this posterior probability also changes. So if I originally pick the fair coin with very high probability and then I observe a heads, do you think the probability that the posterior probability should be the same, higher, or lower?

AUDIENCE: Just a question.

BRYNMOR CHAPMAN: Oh, yeah?

AUDIENCE: How did you get a third?

BRYNMOR CHAPMAN: Oh, question. How did we get a third? So the ratio of this to its complement is one half. They sum to 1. So they're one third and 2/3. Yeah?

AUDIENCE: Can you write the numerical values of H and F?

BRYNMOR CHAPMAN: The numerical values of H and F. So H and F are events. They're not numbers exactly. So H is the event that whichever coin I flipped, it got a heads on it. And F is the event that the coin I picked was the fair one. So they're not really numbers. They're just sets of outcomes that could have happened. Yeah?

AUDIENCE: [INAUDIBLE] the assumptions. Can you go over the assumptions?

BRYNMOR CHAPMAN: Oh, the assumptions. So pick fair or biased uniformly, and then flip it. So "fair" means it comes up with 50/50 heads or tails. Biased, it's a two-headed coin. Yeah?

AUDIENCE: Could you re-explain where the one third came from again?

BRYNMOR CHAPMAN: Explain where the one third came from again. So the probability of F given H over not F given H, these are complementary probabilities. They should sum to 1.

So we've figured that the ratio right is one half. So if we want them to sum to 1, then they should be one third and 2/3. Does that make sense? Yeah. OK. Yeah?

AUDIENCE: [INAUDIBLE] do we care which direction it's biased, or--

BRYNMOR CHAPMAN: Sorry?

AUDIENCE: If it's biased [INAUDIBLE] do we care [INAUDIBLE]?

BRYNMOR CHAPMAN: Oh, question. If it's biased, do we care in which direction it's biased? Yes, absolutely. [CHUCKLES] So if it's a double-tailed coin instead of a double-heads coin, if I observe a heads, it must be the fair coin because the double-tails coin cannot get a heads. So then my probability is just 1.

OK, are people reasonably happy with this example? Shall we move on to a second one? OK, second example-- COVID testing. Woo-hoo, everybody loves COVID testing, right?

Oh, I suppose you-- most of you probably weren't here when there were those daily or near-daily nasal swabs, right? Well, those were fun times. So in the next couple of examples, we're going to see some rather counterintuitive behavior that arises from Bayes' rule.

So just as we were discussing last time, you want to throw away intuitive reasoning and just go with the numbers. So suppose 10% of MIT has COVID. And everybody has to take a COVID test.

They have a false-positive rate of 30% and false-negative rate of 10%. So false-positive rate means if I am healthy and I take a COVID test, there's a 30% chance that it's going to be wrong and it's going to say that I have COVID. So it's falsely reporting a positive result.

And false negative is the opposite. So if I'm sick and I take a COVID test, there is a 10% chance that it does not detect it, and it reports that I'm healthy. Does this setup make sense? Yeah?

AUDIENCE: Are the false positives included in the 10%? Or is the 10% the actual--

BRYNMOR CHAPMAN: Oh, question was, are the false positives included in the 10%? Or is the 10% when you actually have it? So this 10% is conditioned on being sick. The conditional probability that it gets it wrong is 10%.

And the false-positive rate is conditioned on being-- oh, sorry, false negative is conditioned on being healthy. False positive is conditioned on being sick. No. No, positive is sick. Yeah, it was right the first time. Sorry. I'm confused.

Yeah, so these are both talking about conditional probabilities. Conditioned on being healthy or sick, what's the probability that the test is accurate or wrong? So let's define some events.

Let's say that H is healthy. And S can be its complement-- sick. Plus is going to be positive result. And minus is going to be a negative result.

So suppose I take this COVID test and it comes out positive. What's the probability that I have COVID? So first, what is the probability that we're trying to compute? Yeah?

AUDIENCE: Probability of S.

BRYNMOR CHAPMAN: Probability of S. Unconditionally?

AUDIENCE: If [INAUDIBLE] positive.

BRYNMOR CHAPMAN: Yeah, so S given positive. The unconditional probability of S is just 10% by assumption. What we're actually looking for is the probability that I'm sick given that I tested positive because that's the observation that we made.

So how would we compute this? Well, let's do the same thing we just did over here. Let's look at the ratio of this probability to its complement and apply Bayes' rule.

So we're going to flip this conditional here-- positive given sick times probability of sick divided by probability of positive given healthy times probability of healthy. Now, those four probabilities are probabilities that we know. What are they? What's the probability of plus given that I'm sick? Yeah?

AUDIENCE: 90.

BRYNMOR CHAPMAN: Yeah, 90%. So let's write that as 0.9. What's the prior probability that I'm sick? Yeah?

AUDIENCE: 10.

BRYNMOR CHAPMAN: 10. 10%. And now what's the probability of a positive result conditioned on being healthy? Yeah?

AUDIENCE: 0.3.

BRYNMOR CHAPMAN: Yeah, that's the false positive rate-- 0.3 times-- and what's the probability that I'm healthy, the prior probability that I'm healthy? Well, that's just going to be the complement of the event that I'm sick. So that's going to be 0.9.

So these 0.9's cancel. This is going to be one third. So I am one third as likely to be sick as healthy. So the probability that I'm sick conditioned on getting the positive result is actually only 1 in 4. So probability S given plus equals one fourth.

Does that make sense to everybody? So it might seem a little bit weird, right? This test is a pretty good one. It's usually giving the right answer.

But I took the test. I got a positive result. But I'm still probably healthy. Why is that? Can anybody explain what's going on here? Yeah?

AUDIENCE: There's a slim chance that you have COVID.

BRYNMOR CHAPMAN: Yeah, exactly. So there's a very slim chance that I had COVID to begin with. So it's actually much higher now that I've got a positive result. But the prior was overwhelmingly in favor of being healthy. And so it's still very likely that I'm healthy, just slightly less so.

So often, people underestimate the impact of the prior in this way. Of course, in the real world, you generally don't do this. So if you get a positive result, you can be more confident in it because normally you only test when you have symptoms. So the probability that you're sick, given that you tested positive and you have symptoms, that's probably going to be much higher than one fourth.

Are people happy with that? Yeah? OK, why don't we look at another interesting application called Simpson's paradox? So there's actually kind of a funny story behind this one.

So about 50 years ago, Berkeley, UC Berkeley, got sued because it was claimed that they were being very sexist in their admissions policies. And the claim was that the admissions rate was much higher for men than it was for women. So that's clearly bad, right?

But then Berkeley counterclaimed by saying that if you look at each individual department, it's actually the opposite. Women are being admitted with much higher rates than men. So what's going on? Was somebody lying? Can these both be true? Any ideas? Yeah?

AUDIENCE: Maybe the sizes of the departments.

BRYNMOR CHAPMAN: Yeah, exactly. So the observation was that the sizes of the departments could be very different. And what else? Yeah?

AUDIENCE: [INAUDIBLE].

BRYNMOR CHAPMAN: Yeah, so the admissions rates-- the overall admissions rates for the departments can also be very different. And so maybe there's some departments-- let's take a look at a very simple example.

So suppose we have two departments, EE and CS. And maybe both of them mildly favor women. But one of them is much more popular with men. One of them is much more popular with women. And they have very different admissions rates generally.

So maybe we have a hundred men and a hundred women applying to this two-department UC Berkeley. And maybe we have 99 women and one man apply to CS. And in double E, it's the opposite-- one woman and 99 men.

And maybe in CS, we're a bunch of snobs. We hate everybody. We take one applicant. So maybe we take-- one woman is accepted. Everybody else is rejected.

And double E, on the other hand, has no standards at all. They'll just take anybody with a pulse. But unfortunately, one of these men is dead. So they take one woman and 98 men.

So now, if you look at the individual department's admissions rates, CS has accepted a woman. So their admissions rate for women is non-zero. They haven't accepted any men. So they've got a zero admissions rate for men. So it's higher for women.

And in double E, they've accepted all of the women but not all of the men. So again, it's higher for women. But if you look at what's happened overall, we've got two women being accepted and 98 men.

So you can argue that, rather than being a problem with the University or the departments, it's more of a systemic problem. Maybe they're encouraging men and women to apply to different departments at different rates. And it's all a bit messy.

And also, this is 50 years ago. So they only really considered men and women. Didn't really consider anybody who identified as anything else. But we're not saying anything about that. This is just to illustrate the maths. So hopefully it's not terribly offensive.

But yeah, even though, for each department, the women are doing better than men, the men are doing much better overall. So that's what we call Simpson's paradox. Does anybody have any questions about that?

OK, so last example. Has anybody seen *The Dragon Prince?* A couple of hands. OK, well, so last example. Suppose a cake has gone missing from the bakery. And Barius the baker has a prime suspect.

He thinks that Prince Ezran stole the cake. And he wants to present evidence to the king. He would like to tell the king that this deviant is always stealing jelly tarts from my bakery. That makes it much more likely that he stole the cake, right? Jelly tart thieves are 10 times as likely as randos to steal cakes.

That seems fairly reasonable. Does anybody think that this is a good reason to admit the jelly tart thievery into evidence? Who thinks that it's not a compelling reason? No? Who has no idea? More hands. OK.

So maybe Prince Ezran's counterclaim is, well, if you look at all jelly tart thieves, only one in, say, 2,500 of them actually go on to steal cakes. So even though the probability of being a cake thief has increased dramatically from the baseline in terms of a multiplicative factor, it can't actually have increased very much because it went from 1 in 25,000 to 1 in 2,500.

So the probative value here is essentially 0. You're trying to prove guilt beyond a reasonable doubt. And so a 1-in-2,500-chance increase is kind of negligible.

But on the other hand, if you admit this into evidence, the king is going to be biased against Ezran. He's going to be like, oh, this kid's a thief. This is terrible. We got to punish him, regardless of whether he stole the cake.

So who's right? Is the baker right? Or is the suspect right? Who thinks baker is right? Who thinks suspect is right? Who thinks neither? [CHUCKLES]

Yeah, I've got a few hands. OK. Well, both parties here are trying to reason about a conditional probability. What is that conditional probability? Anybody? Well, maybe we should start by writing out some events.

So let's let J be the event suspect steals jelly tarts. C can be the event suspect steals cake. And X can be the event cake was stolen.

So both the baker and the suspect were making claims about conditional probabilities. What specific conditional probability in terms of these events were they making claims about? Yeah?

AUDIENCE: Can you repeat the assumptions?

BRYNMOR CHAPMAN: Sorry?

AUDIENCE: Can you repeat the assumptions?

BRYNMOR CHAPMAN: Repeat the assertions?

AUDIENCE: Assumptions.

BRYNMOR CHAPMAN: Oh, the assumptions. So the assumptions were that a cake has gone missing from the bakery. The baker has a prime suspect. And the baker wants to admit into evidence that this suspect is always stealing jelly tarts.

And the reasoning is that jelly tart thieves are much more likely to steal cakes than random other people. And the suspect counters by saying, well, sure, but only 1 in 2,500 jelly tart thieves is actually a cake thief. So the actual increase is less than 1 in 2,500. So what are the probabilities that they're reasoning about? What are the conditional probabilities? Yeah?

AUDIENCE: [INAUDIBLE].

BRYNMOR CHAPMAN: Yeah. Well, you're both right. Sorry, it's-- [CHUCKLES] geometry. They're both trying to reason about probability of C given J. So we're conditioning on the suspect having stolen jelly tarts. And they're both trying to make claims about, what is the probability of C conditioned on that?

So, well, the baker is arguing that probability of C given J divided by probability of C given not J is high. So because this is about 10, J is very relevant to the case. So it should be admitted.

The suspect is saying that probability of C given J is low. So J cannot actually dramatically increase the probability. The increase is going to be less than this probability. Yeah?

AUDIENCE: [INAUDIBLE] single person, a single piece. If this were an actual event, wouldn't these probabilities only really work in aggregate [INAUDIBLE]?

BRYNMOR CHAPMAN: Yeah, so that's a good point. So the point was that we're talking about one individual suspect, one individual cake, one individual case. So as we were saying before when we first started looking at Bayes' rule, either it happened or it didn't. You could say that it doesn't make a whole lot of sense to talk about probabilities. But you're trying to convince the king or a jury or somebody.

So this is kind of similar to when you present DNA evidence in court or something like that. The fact that there's a match means that there's overwhelming probability that the suspect was at the crime scene or something like that. So you do still talk about probabilities. But you could think of it as like, how confident are you in the inference that the suspect did something? Does that make sense?

AUDIENCE: Yeah.

BRYNMOR CHAPMAN: OK. So if you were the judge in this case determining what should be presented to the king, what would your verdict be? Do we allow jelly tart thievery into evidence or no?

AUDIENCE: Draw a tree diagram.

BRYNMOR CHAPMAN: Pardon?

AUDIENCE: Draw a tree diagram.

BRYNMOR CHAPMAN: [CHUCKLES] Draw a tree diagram. Well, that's not actually necessary anymore. We've already figured out what the probabilities are. The question is, is the probative value good enough to justify putting it into evidence? Or is the probative value small enough that the bias is going to be more of a factor? Yeah?

AUDIENCE: [INAUDIBLE] the evidence about the jelly part because, also, like, an MIT student, for instance, given an amazing student, there's a very low chance of them being a cake thief [INAUDIBLE]. I guess this is what I'm wondering about an individual instance where it doesn't really count for them.

BRYNMOR CHAPMAN: Yeah, so it will not generally capture the full story. And this is why lawyers in a courtroom are trying to find all kinds of evidence to bolster their cases. So, yeah, something like that might be something that the defense would want to present. Or the prosecution might be like, oh, but even though it's an MIT student, it's an MIT EECS student who do steal cakes all the time, so-- [CHUCKLES] I'm getting some nods. That's mildly concerning. [LAUGHING]

So, yeah, it's going to be very difficult to capture the entire picture. For right now, we're just trying to limit the scope to, these are the things that are known. We know that a cake is stolen. We know that this suspect steals jelly tarts. But that's it. We want to try and figure out what the probative value of this evidence is just based on those. Does that make sense?

So who thinks the jelly tarts should be admitted? Who thinks no? OK, I'm getting more hands this time, less abstentions. Does anybody want to say why? Yeah?

AUDIENCE: Probability of C given [INAUDIBLE] low [INAUDIBLE] low probability of C given [INAUDIBLE]. So that could account for an inflated probability as well.

BRYNMOR CHAPMAN: Yeah, so the observation was the probability of C given not J is also very low. And in fact, it is much lower than C given J. So that's where the ratio there, probability of C given J over C given not J, is coming from.

Probability of C given J is 1 in 2,500. Probability of C given not J is 1 in 25,000. So that's where you're getting this factor of 10. So you agree with the defense, that it should not be admitted into evidence. Does anybody have any counterclaims? Yeah?

AUDIENCE: So wouldn't you consider the part where the thief also steals where the cake is stolen before?

BRYNMOR CHAPMAN: Yeah. [CHUCKLES]

AUDIENCE: [INAUDIBLE].

BRYNMOR CHAPMAN: Yeah, so that's a good point. We've kind of forgotten about something, right? [CHUCKLES]

The cake was stolen, right? We don't care that the probability of cakes being stolen is low. It happened. [LAUGHS] Right? So we don't really care about these probabilities at all. What is the actual probability we care about? What we really want is probability of C given, what? Yeah?

AUDIENCE: Given that they stole the jelly tarts and they stole a cake?

BRYNMOR CHAPMAN: Well, not that they stole a cake, but that a cake was stolen. Yeah. So C given J and X. So that's what we really care about. The fact that the cake is stolen is really important.

If you don't know whether the cake is stolen, why are you even having a trial? And as it turns out, this probability is something like 80%. Jelly tart thieves found in bakeries with missing cakes, usually they're the guilty party.

So, yeah, I've kind of made this a little bit more palatable. But this is based on an actual courtroom case. This is the kind of thing that people-- people try and use and end up misusing probabilities and conditional probabilities all the time. This was actually from a very high profile case, the OJ Simpson case, where both the defense and the prosecution attorneys, they're supposed to be experts. But they were both making bogus claims about conditional probabilities.

It's very easy to make these kinds of mistakes. So fall back on fundamentals. When you're in doubt, fall back on the fundamentals. Even if you're not, it's probably a good idea anyway. OK, feel free to come up if you have any questions. And otherwise, we will see you next week.