

The following content is provided under a Creative Commons license. Your support will help MIT OpenCourseWare continue to offer high quality educational resources for free. To make a donation or view additional materials from hundreds of MIT courses, visit MIT OpenCourseWare at [ocw.mit.edu](http://ocw.mit.edu).

**PROFESSOR:** Just a reminder, drop day is tomorrow. So if you were thinking about dropping the course or in danger of a bad grade or something, tomorrow's the last chance to bail out. Last time we began our discussion on probability with the Monty Hall game-- the Monty Hall problem. And as part of the analysis, we made assumptions of the form that given that Carol placed the prize in box 1, the probability that the contestant chooses box 1 is  $1/3$ . Now, this is an example of something that's called a conditional probability. And that's what we're going to study today.

Now, in general, you have something like the conditional probability that an event, A, happens given that some other event, B, has already taken place. And you write that down as a probability of A given B. And both A and B are events. Now, the example from Monty Hall-- and actually, we had several-- but you might have B being the event that Carol places the prize in box 1. And A might be the event that the contestant chooses box 1.

And we assumed for the Monty Hall game that the probability of A given B in this case was  $1/3$  third because the contestant didn't know where the prize was. Now in general, there's a very simple formula to compute the probability of A given B. In fact, we'll treat it as a definition. Assuming the probability of B is non-zero then the probability of A given B is just the probability of A and B happening, both happening, divided by the probability of B happening.

And you can see why this makes sense when the picture-- say this is our sample space. And let this be the event, A, and this be the event, B. Now we're conditioning on the fact that B happened. Now once we've conditioned on that, all this stuff outside of B is no longer possible.

All those outcomes are no longer in the space of consideration. The only outcomes left are in B. So in some sense we've shrunk the sample space to be B. And all we care about is the probability that A happens inside this new sample space. And that is, we're asking the probability 1 of these outcomes happens given that this is the sample space.

Well, this is just  $A \cap B$  because you still have to have A happen, but now you're inside

of B. And then we divide by probability of B. So we normalize this to be probability one.

OK. Because we're saying B happened-- we're conditioning on that. Therefore, the probability of these outcomes must be 1. So we divide by the probability of B. So we normalize.

This now becomes-- the probability of A given B is this share of B weighted by the outcomes.

OK. All right. For example then, what's the probability of B given B? what's that equal? 1.

OK. Because we said it happened-- so it happens with probability 1. Or, using the formula, that's just probability of B and B divided by probability of B. Well, that equals the probability of B divided by the probability of B, which is 1.

All right. Any questions about the definition of the conditional probability? Very simple. And it's easy to work with using the formulas.

Now, there's a nice rule called the product rule, which follows from the definition very simply. The product rule says that the probability of A and B for two events is equal to the probability of B times the probability of A given B. And that's just follow straightforwardly from this definition.

Just multiply by probability of B on both sides. All right. So now you have a rule of computing a probability of two events simultaneously happening. So for example, in the Monty Hall problem, what's the probability that Carol places the prize in box one and that's the box the contestant chooses? All right?

So if we took A and B as defined up there, that's the probability that Carol places it in box one and the contestant chose it. Well, that's the probability that the contestant chooses it is  $1/3$  times the probability that Carol put it there, given the contestant chose it, or actually, vice versa, is  $1/9$ . OK?

And this extends to more events. It is called the general product rule. So if you want to compute the probability of  $A_1$  and  $A_2$  and all the way up to  $A_n$ , that's simply the probability of  $A_1$  happening all by itself times the probability of  $A_2$  given  $A_1$  times-- well, I'll do the next one-- times the probability of  $A_3$  given  $A_1$  and  $A_2$ , dot, dot dot, times, finally, the probability of  $A_n$  given all the others.

So that starts to look a little more complicated. But it gives you a handy way of computing the probability that an intersection of events takes place. I do This is proved by induction on  $n$ , just

taking that rule and using induction on  $n$ . It's not hard. But we won't go through it.

All right. Let's do some examples. We'll start with an easy one. Say you're playing a playoff series and you're going to play best 2 out of 3. All right. So you have a best 2 out of 3 series. So whoever wins the first two games, best two out of three wins. And say you're told that the probability of winning the first game is  $1/2$ .

So the teams are matched 50-50 for the first game. But then you're told that the probability of winning a game after a victory is higher. It's  $2/3$ . So the probability of winning immediately after a game following a win is two thirds. And similarly, the probability of winning after a loss is  $1/3$ .

All right. And the idea here is that you win a game, you're sort of psyched, you've got momentum, and going into the next day you're more likely to win. Similarly, if you lost you're sort of down and the other guy has a better chance of beating you. Now, what we're going to try to figure out is the probability of winning the series given you won the first game. All right?

Now, conditional probability comes up in two places in this problem. Anybody tell me places where it's come up? So I got the problem statement and the that's the goal is to figure out the probability you win the series given you won the first game. So what's one place conditional probability is entering into this problem? Yeah?

**AUDIENCE:** The probability changes depending on the result of the previous game.

**PROFESSOR:** That's true. The probability of winning any particular game is influenced by the previous game. So you're using conditional probability there. All right. And where else? Yeah.

**AUDIENCE:** [INAUDIBLE] you have to take into account [INAUDIBLE].

**PROFESSOR:** That's interesting. That will be another question we're going to look at. What's the probability of playing three games? Yep. That's one.

OK. Well, the question we're after, what's the probability of winning the series given that you won the first game. We're going to compute a conditional probability there. So it's coming up in a couple of places here.

All right. Let's figure this out. It's easy to do given the tree method. So let's make the tree for this. So we have possibly three games there's game one, game two, and game three.

Game one, you can win or lose. There's two branches. Game two you can win or lose. And now, game three-- well, it doesn't even take place here.

But it does here. You can win or lose here. And you could win or lose here. And here the series is over. So there is no game three in that case.

The probabilities are next we put a probability of every branch here. Game one is 50-50. What's the probability you take this branch?  $2/3$ , because you're on the path where you won the first game.

You win the second game with  $2/3$ . You lose with  $1/3$ . Now here you're on the path where you lost the first game. So this has  $1/3$  and this has  $2/3$ .

All right? And then lastly, what's the probability I have the win on the third game here?  $1/3$ , because I just lost the last game. That's all I'm conditioning on.

So that becomes  $1/3$ . And this is  $2/3$  now. And then here I just won a game. So I've got  $2/3$  and  $1/3$ .

All right. So I got all the probabilities. And now I need to figure out for the sample points what's their probability. So this sample point we'll call win-win.

This sample point is win-lose-win. This one's win-lose-lose. Then we have lose-win-win, lose-win-lose, and then lose-lose.

So I got six sample points. And let's figure out the probability for each one. Now remember the rule we had for the tree method. I just multiply these things.

Well, in fact, the reason we have that rule is because that is the same as the product rule. Because what I'm asking here to compute the probability of this guy is-- so the product rule gives the probability of a win-win scenario-- win the first game, win the second game. By the product rule is the probability that I win the first game times the probability that I win the second game given that I won the first game. That's what the product rule says.

Probability I win the first game is  $1/2$  times the probability I win the second given that I won the first is  $2/3$ . So that equals  $1/3$ . So what we're doing here now is giving you the formal justification for that rule that we had last time and that you'll always use-- is the probability of a sample point is the product of the probabilities on the edges leading to it. It's just the product

rule.

Now the next example is this one. And here we're going to use the general product rule to get it. The probability of win-lose-win by the general product rule is the probability that you win the first game times the probability you lose the second game given that you win the first times the probability you win the third given what?

What am I given on the product rule? Won the first, lost the second. All right. Well, now we can fill in the numbers.

The probability I win the first is a  $1/2$ . The probability that I lose the second given that I won the first, that's  $1/3$ . And then this one here, the probability that I win the third given that I won the first and lost the second, that simplifies the probability I win the third given that I lost the second. Doesn't matter what happened on the first. And that's  $1/3$ .

So this is  $1/2$  times  $1/3$  times  $1/3$  is  $1/18$ . And that's  $1/18$ . And it's just the product because the product rule saying product of the first probability times this one, which is the conditional probability of being here times this one, which is a conditional probability if these events happened before.

Any questions about that? Very simple to do, which is good. Yeah. Is there a question? OK. All right. So let's fill in the other probabilities here.

I got  $1/2$ ,  $1/3$ , and  $2/3$ . That's  $1/9$ . Same thing here is  $1/9$ . This is  $1/18$  and  $1/3$ . OK. So those are the probabilities in the sample points.

Now, to compute the probability of winning the series given that we won the first game, let's define the events here. So  $A$  be the event that we win the series.  $B$  will be the event that we win the first game. And I want to compute the probability of  $A$  given  $B$ .

And we use our formula. Where's the formula for that? It's way back over there. The probability of  $A$  given  $B$  is the probability of both happening, the probability of  $A$  and  $B$  divided by the probability of  $B$ .

So now I just have to compute these probabilities. So to do that I got to figure out which sample points are in  $A$  and  $B$  here. So let's write that down. There's  $A$ ,  $B$ ,  $A$  and  $B$ . All right. So  $A$  is the event that we win the series. Now this sample point qualifies, that one does, and this one.

B is the event we won the first game. And that's these three sample points. And then A and B intersect B is these two. All right.

So for each event that I care about I figure out which sample points are in that event. And now I just add the probabilities up. So what's the probability of A and B?  $7/18$ .  $1/3$  plus  $1/18$ .

What's the probability of B? Yeah.  $1/2$ ,  $9/18$ . I got these three points. So this'll be  $1/3$  third plus  $1/18$  plus the extra one,  $1/9$ . So I've got  $7/18$  over  $9/18$ .  $7/9$  is the answer. So the probability we win the Series given we won the first game is  $7/9$ . Any questions?

We're going to do this same thing about 10 different times. OK? And it will look a little different each time maybe. But it's the same idea. And the beauty here is it's really easy to do. I'm going to give you a lot of confusing examples. But really, if you just do this is it's going to be very easy.

All right. Somebody talked about the series lasting three games. What's the probability the series lasts three games? Can anybody look at that and tell me?  $1/3$  because what you would do is add up these three sample points. And it's the opposite of these two. So it's  $2/3$  chance of two games, a  $1/3$  chance of three games. So it's not likely to go three games.

All right. So to this point, we've seen examples of a conditional probability where it's A given B where A follows B, like, we're told B happened. Now what's the chance of A. And A is coming later. The probability of winning today's game given that you won yesterday's game, the probability of winning the series given you already won the first game.

Next, we're going to look at the opposite scenario where the events are reversed in order. The probability that you won the first game given that you won the series. All right. Now, this is inherently confusing because if you're trying to figure-- if you know you the series, well, you already know what happened in the first game because it's been played. So how could there be any probability there? It happened.

Well, so what the meaning is is over all the times where the series was played, sort of what fraction of the time did the team that won the series win the first game is one way you could think about it. Or, maybe you just don't know. The game was played. You know you won the series. But you don't know who won the first game. And so you could think of a probability still being there.

Now when you think about it, it gets me confused still. But just think about it like the math. It's the same formula. OK. It doesn't matter which happened first in time. You use the same mathematics. In fact, they give a special name these kinds of things. They're called a posteriori conditional probabilities.

It's a fancy name for just saying that things are out of order in time. All right? So it's a probability of B given A where B precedes A in time. All right? So it's the same math. It's just they're out of order.

So let's figure out the probability that you won the first game given that you want the series. Let's figure it out. So I want probability of B given A now for this example. Well, it's just the probability of B and A over the probability of A.

We already computed the probability of A and B. That's  $\frac{1}{3}$  plus  $\frac{1}{18}$ . What's the probability of A, the probability of winning the first game?  $\frac{1}{2}$ . It's those three sample points and they better add up to  $\frac{1}{2}$  because we sort of said, the probability of the first game's  $\frac{1}{2}$ . So that's over  $\frac{1}{2}$ , which is  $\frac{9}{18}$ .

Well this was  $\frac{7}{18}$  over  $\frac{9}{18}$ . It's  $\frac{7}{9}$ . So the probability of winning the first game given that you won series is  $\frac{7}{9}$ . Anybody notice anything unusual about that answer here?

It's the same as the answer over there. Is that a theorem? No. The probability of A given B is not always the probability of B given A. It was in this case. It is not always true. In fact, we could make a simple example to see why that's not always the case.

All right. So say here's your sample space. And say that this is B and this is A. What's the probability of A given B in this case? 1. If you're in B-- wait. No. It's not 1. What's the probability of A given B if I got some-- probably less than 1. Might be I've drawn it as  $\frac{1}{3}$  third if it was uniform.

But in this case, the probability of A given B is less than 1. What's the probability of B given A? 1, because if I'm in A I'm definitely in B. All right. So that's an example where they would be different. And that's the generic case is they're different. All right?

When are they equal because they were equal in this case? What makes them equal? Let's see. When does the probability of A given B equal a probability of B given A?

Let's see. Well, if I plug-in the formula, this equals the probability of A and B over the probability of B. That equals the probability of B and A over a probability of A. So when are those equal? Yeah. When probability A equals probability B. All right. So that's one case.

What's the other case? Yeah-- when it's 0. Probability-- there's no intersection. Probability of A intersect B is 0. That's the other case. All right. But usually these conditions won't apply-- just happened to in this example by coincidence.

Any questions about that? All right. Yeah. So the math is the same with a posteriori probabilities. It's really, really easy.

All right. So let's do another simple example that'll start to maybe be a little more confusing. Say we've got two coins.

One of them is a fair coin. And by that, I mean the probability comes up heads is the same as the probability comes up tails is  $1/2$ . The other one is an unfair coin. And in this case, that means it's always heads.

The probability of heads is 1. The probability of tails is 0. All right? I've got two such coins here.

All right. Here is the unfair coin-- heads and heads. Actually, they make these things look like quarters sometimes. Here's the fair coin-- heads and tails. All right.

Now suppose I pick one of these at random, 50-50, I pick one of these things, and I flip it, which I'm doing behind my back, and lo and behold, it comes out and, you see a heads. What's the probability I'm holding the fair coin? I picked the coin, 50-50, behind my back. So one answer is, I picked the fair coin with 50% probability.

But then I flipped it behind my back and I showed you the result. And you see heads. Of course, if I'd have shown you tails, you would have known for sure it was the fair coin because that's the only one with the tails.

But you don't know for sure now. You see a heads. What's the probability this is the fair coin given that you saw a heads after the flip?

How many people think  $1/2$ ? After all, I picked it with probability  $1/2$ . How many people think it's less than  $1/2$ ? Good. OK.

Somebody even said  $1/3$ . Does that sound right? A couple people like  $1/3$ . OK. All right.

Now, part of what makes this tricky is I told you I picked the coin with 50% probability. But then I gave you information. So I've conditioned the problem.

And so this is one of those things you could have an ask Marilyn about. Is it  $1/2$  or is it  $1/3$ ? Because I picked it with 50% chance, what does the information do for you? Now, I'll give you a clue.

Bobo might have written in and said it's  $1/2$ . And his proof is that three other mathematicians agreed with him.

[LAUGHTER]

All right? OK. So let's figure it out. And really it's very simple. It's just drawing out the tree and computing the conditional probability.

So we're going to do the same thing over and over again because it just works for every problem. Of course, you could imagine debating this for awhile, arguing with somebody. Is it  $1/2$  or  $1/3$ ? Much simpler just to do it.

So the first thing is we have, which coin is picked? So it could be fair-- and I told you that happens with probability  $1/2$ -- or unfair, which is also  $1/2$ . Then we have the flip. The fair coin is equally likely to be heads or tails, each with  $1/2$ . The unfair coin, guaranteed to be heads, probability 1.

All right. Now we get the sample point outcomes. It's fair in heads with the probability  $1/4$ , fair in tails, probability  $1/4$ , unfair in heads, probability  $1/2$ . Now we define the events of interest. A is going to be that we chose the fair coin. And B is at the result, is heads. And of course what I want to know is the probability that I chose the fair coin given that I saw a heads.

So to do that we plug in our formula. That's just the probability of A and B over the probability of B. And to compute that I got to figure out the probability of A and B and the probability of B.

So I'll make my diagram. A here, B here, A and B. A is the event I chose the fair coin. That's these guys. B is the event the result is heads. That's this one and this one. And A intersect B, That's the only point. So this is really easy to compute now.

What's the probability of A and B?  $1/4$ . It's just that sample point. What's the probability of B?  $3/4$ ,  $1/4$  plus  $1/2$ . So the probability of A given B is  $1/3$ .

Really simple to answer this question. Just don't even think about it. Just write down the tree when you get these things. So much easier just to write the tree down.

All right. Now the key here is we knew the probability of picking the fair coin in the first place. Maybe it's worth writing down what happens if that's a variable-- sum variable  $P$ . Let's do that.

For example, what if I hadn't told you the probability that I picked the fair coin? I just picked one and flipped it. Think that'll change the answer? It should because you got to plug something in there for the  $1/2$  for this to work.

So let's see what happens. Say I picked the fair coin with probability  $P$  and the unfair coin with  $1$  minus  $P$ . And this is the same heads and tails,  $1/2$ ,  $1/2$ . Heads, the probability  $1$ . Well now, instead of  $1/4$  I get  $P$  over  $2$  up here. And this is now  $1$  minus  $P$  instead of  $1/2$ .

So the probability of A given B is the probability of A and B is  $p$  over  $2$ . And the probability of B is  $P$  over  $2$  plus  $1$  minus  $P$ . That's  $P$  over  $2$  up top,  $1$  minus  $P$  over  $2$ , and that is all multiplied by-- what am I going to multiply--  $2$  here. I'll get  $P$  over  $2$  minus  $P$ .

So the probability with which I picked the coin to start with impacts the answer here. For example, what if I picked the unfair coin for sure? That would be  $P$  being  $0$ . Well, the probability that I picked the fair coin is  $0$  over  $2$ , which is  $0$ . All right though-- even know I showed you the heads, there's no chance it was the fair coin because I picked the unfair coin for sure.

Same thing if I picked the fair coin for sure, better be the case this is  $1$ . So I get  $1$  over  $2$  minus  $1$ . It's  $1$ . Any questions? So it's important you know the probability I picked the fair coin to start with. Otherwise, you can't go anywhere.

All right. What if I do the same game? Pick a coin with probability  $p$ . But now I flip it  $K$  times. Say I flip it  $100$  times. And every time it comes up heads. I mean you're pretty sure you got the unfair coin because you never saw a tails. Right?

So let's do that. Let's compute that scenario. So instead of a single heads I get  $K$  straight heads and no tails. This would happen with  $1$  over  $2$  to the  $K$ . This would happen with  $1$  minus  $1$  over  $2$  to the  $K$ . So this is now  $p$  over  $2$  to the  $K$ . This is now  $P$   $1$  minus  $2$  to the minus  $K$ .

Let's recompute the probabilities. I'm going somewhere where this. Wait a minute.

So now we're looking at the event that B is K straight heads. Come up. And I want to know the probability that I picked the fair coin given that it just never comes up tails. The math is the same.

The probability now that I picked the fair coin and got k straight heads is just  $p$  times  $2$  to the minus  $K$ . The probability that I got  $K$  straight heads is  $P$  times  $2$  to the minus  $K$  plus the chance I picked the unfair coin, which is  $1$  minus  $P$ . And if I multiply top and bottom by  $2$  to the  $K$ , I get  $P$  over  $P$  plus to the  $K$   $1$  minus  $B$ .

All right. So it gets very unlikely that I've got the fair coin here as  $K$  gets big. Like if  $K$  is  $100$  I got a big number down here. And basically it's  $0$  chance-- close to  $0$  chance of the fair coin.

But now say I do the following experiment. I don't tell you  $P$ . But I pull a coin out and  $100$  flips in a row it's heads. Which coin do you think I have? I flipped it  $100$  straight times and it's heads every time.

Yeah. There's not enough information. You don't know. What do you want to say? You want to say it's the unfair coin but you have no idea because I might have picked the fair coin with probability  $1$ , in which case it is the fair coin and it just was unlucky that it came up heads  $100$  times in a row. But it could be.

So you could say nothing if you don't know the probability  $P$ . Because sure enough, if I plug in  $P$  being  $1$  here, that wipes out the  $2$  to the  $K$  and I just get probability  $1$ . OK?

All right. Now when this comes up in practice is with things like polling. Like, we just had an election. And people do poles ahead of time. And they sample thousands of voters from  $1\%$  of the population.

And they say, OK, that  $60\%$  of the people are going to vote Republican. And they might have a margin of error, three points, whatever that means. And we'll figure that out next week. What does that tell you about the electorate as a whole-- the population if they sample  $1\%$  at random,  $60\%$  are Republican. Yeah?

**AUDIENCE:** [INAUDIBLE] The options you have, is it all heads or is it all tails? It should be one option all heads and another option at least one tails.

**PROFESSOR:** You're right. Oops. All right. At least one tail for this one. Yeah. Good. That is true. OK. Any questions about that example?

OK. Now we're back to the election and there's a poll that says they sampled 1% of the population at random and 60% said they're going to vote Republican. And the margin of error is 3% or something. What does that tell you about the population of the country?

Nothing. That's right. It is what it is. All you can conclude is that either the population is close to 60% Republican or you were unlucky in the 1% you sample. That's what you can conclude because the population really is fixed in this case.

It is what it is. There's no randomness in the population. All right? So you have next week for recitation. You're going to design a poll and work through how to calculate the margin of error and work through what that really means in terms of what the population is like.

Now of course, if it comes out 100 straight times heads, you've got to be really unlucky to have the fair coin. And the same thing with designing the poll if you're way off. Any questions about that?

OK. The next example comes up all the time in practice. And that's with medical testing. Maybe I'll leave-- no. I'll take that down. We know that now.

Now in this case-- in fact, this is a question we had on the final exam a few years ago. And there's a good chance this kind of question's going to be on the final this year. There's a disease out there. And you can have a test for it.

But like most medical tests, they're not perfect. Sometimes when it says you've got the disease you really don't. And if it says you don't have it, you really do.

So in this case, we're going to assume that 10% of the population has the disease, whatever it is. You don't get symptoms right away. So you have this test.

But if you have the disease there is a 10% chance that the test is negative. And this is called a false negative, because the test comes back negative but it's wrong, because you have the disease.

And similarly, if you have the disease-- or sorry-- if you don't have the disease, there's a 30%

chance that the test comes back positive. And it's called a false positive because it came back positive, but you don't have it. So the test is pretty good. Right? It's 10% false negative right, 30% false positive right.

Now say you select a random person and they test positive. What you want to know is the probability they have the disease given that it's a random person. So actually, this came up in my personal life.

Many years ago when my wife was pregnant with Alex, she was exposed to somebody with TB here at MIT. And she took the test. And it came back positive.

Now the bad thing-- TB's a bad thing. You don't want to get it. But the medicine for it you take for six months. And she was worried about taking medicine for six months when she's pregnant because who knows what the TB medicine does kind of thing if you have a baby.

So she asked the doc, what's the probability I really have the disease? The doc doesn't know. The doc maybe could give you some of these steps, 10% false negative, 30% false positive. But it tested positive. So they just normally give you the medicine.

So say this was the story. What would you say? What do you think? How many people think that it's a least a 70% chance you got the disease? She tested positive and it's only got a 30% false positive rate. Anybody? So you don't think she's likely to have it.

How many people think it's better than 50-50 you have the disease? A few. How many people think less than 50%. A bunch. Yeah. You're right, in fact.

Let's figure out the answer. It's easy to do. So A is the event the person has the disease. And B is the event that the person tests positive.

And of course what we want to know is the probability you have the disease given that you tested positive. And that's just the probability of both events divided by the probability of testing positive. So let's figure that out by drawing the tree.

So first, do you have the disease? And it's yes or no. And let's see. The probability of having the disease, what is that for a random person? 10%. that the stat. So it's-- actually, we'll call it 0.1. And 9.9 you don't have it.

And then there's the test. Well, you can be positive or negative. Now if you have the disease,

there is a-- the chance you test negative is 10%, 0.1. Therefore there's a 90% chance you test positive.

Now if, you don't have the disease, you could test either way. If you don't have the disease there's a 30% chance you test positive. 30 here and 70% percent chance you're negative.

Now we can compute each sample point probability. This one is 0.1 times 0.9 is 0.09. 0.1 times 1 is 0.01. 0.9 and 0.3 is 0.27. 0.9 and 0.7 is 0.63.

So all sample points are figured out. Now we figure out which sample points are in which sets.

So we have event A, event B, and A intersect B. Let's see. A is the event you have the disease. That's these guys. B is the event you test positive. That's this one and this one. A intersect B is just this one.

All right. We're almost done. Let's just figure out the probability you have the disease. What's the probability of A intersect B? 0.09. It's just that one sample point.

What's the probability that you tested positive? 0.36. Yeah. 0.09 plus 0.27, which is 0.36. So I got 0.09 over 0.36 is 1/4. Wow. That seems bizarre.

Right? You've got a test, 10% percent false negative, 30% false positive. Yet, when you test positive there's only a 25% chance you have the disease. So maybe you don't take the medicine. So if there's risk both ways, probably don't have the disease. Yeah?

**AUDIENCE:** [INAUDIBLE] disease change because you've already been exposed to somebody that has it?

**PROFESSOR:** That's a great point, great point, because there's additional information conditioning this in the personal example I cited. You were exposed to somebody. So we need to condition on that as well, which raises the chance you have the disease. That's a great point. Yeah.

Just like in the-- well, we haven't got to that example. Do another example with that exact kind of thing is very important. All right. So this is sort of paradoxical that it looks like a pretty good test-- low false positive, full false negatives, but likely be wrong, at least if it tells you have the disease.

In fact, let's figure out. What's the probability that the test is correct? What's the probability the test is right in general? 72%. Let's see. So it would be 0.09 plus 0.63. 72%.

So it's likely to be right. But if it tells you you have the disease it's likely to be wrong. It's hard.

Why is this happening? Why does it come out that way? Yeah?

**AUDIENCE:** Then there is only a 1 in 64 chance that you have the disease. So if it comes back negative, then it's a pretty good indication that you're OK.

**PROFESSOR:** Yeah. If it comes back negative than it really is doing very well. That's right. But why is it when it comes back positive that you're unlikely to have the disease if it's a good test. Yeah.

**AUDIENCE:** The disease is so rare.

**PROFESSOR:** The disease is so rare. Absolutely. This number here is so small. And that's what's doing it. Because if you look at how many people have the disease and test positive, it's 0.09. So many people don't have the disease that even with a small false positive rate, this number swamps out that number.

In fact, imagine nobody had the disease. You'd have a 0 here. All right? And then you would always be wrong if you said you had it. OK? That's good.

OK. This comes up in weather prediction, the same paradox. For example, say you're trying to predict the weather for Seattle. Sometimes it seems like this in Boston. And you just say, it's going to rain.

Forget all the fancy weather forecasting stuff, the radar, and all the rest. Just say it's going to rain tomorrow. You're going to be right almost all the time. All right? And in fact, if you try to do fancy stuff, you're probably going to be wrong more of the time.

All right. For example, in this case, if you just say the person does not have the disease, forget the lab test. Just come back with negative. How often are you right? 90% of the time you're right. Much better than the test you paid a lot of money for.

I see. You've got to be careful what you're looking for, how you measure the value of a test or a prediction. Because presumably the one you paid for is better, even though accurate less of the time. Any questions about that?

OK. So For the rest of today we're going to do three more paradoxes. And in each case they're going to expose a flaw in our intuition about probability. But the good news is in each case it's easy to get the right answer. Just stick with the math and try not to think about it. Now the first example is a game involving dice that's called carnival dice that you can find in

carnivals and you can also find in casinos.

It's a pretty popular game, actually. So the way it works is as follows. The player picks a number from 1 to 6-- we'll call it  $N$ -- and then rolls three dice. And let's say they're fair and mutually independent.

We haven't talked about independent. So they're fair dice. For now, normal dice-- nothing fishy. And the player wins if and only if the number he picked comes up on at least one of the dice. So you either win or you lose the game depending on if your lucky number came up at least once.

Now you've got three dice, each of which has a 1 in 6 chance of coming up a winner for you. So how many people think this is a fair game-- you got a 50-50 chance of winning-- three dice, each  $1/6$  chance of winning? Anybody think it's not a fair game? A bunch of you. How many people think it is a fair game-- 50-50? A few. All right.

Well, let's figure it out. And instead of doing the tree method, which we know we're supposed to do, we're just going to wing it, which is always seems easier to do. If you're in the Casino you want to just wing it instead of taking your napkin out and drawing a tree.

So the claim, question mark, is the probability you win is  $1/2$ . And the proof, question mark, is you let  $A_i$  be the event that the  $i$ -th die comes up  $N$ . And  $i$  is 1 to 3 here.

So then you say, OK. The probability I win is the probability of  $A_1$ -- I could win that way-- or  $A_2$ , or  $A_3$ . All I need is one of the die to come up my way. And that is the probability of  $A_1$  plus the probability of  $A_2$  plus the probability of  $A_3$ . And each die wins for me with probability  $1/6$ . And that is then  $1/2$ .

So that's a proof that we win with probability of  $1/2$ . What do you think? Any problems with that proof?

**AUDIENCE:** [INAUDIBLE]

**PROFESSOR:** Well that's a great point. Yeah. So if I extended this nice proof technique I couldn't have probability of  $7/6$  of winning with seven die. Yeah?

**AUDIENCE:** [INAUDIBLE]

**PROFESSOR:** Yeah. You're very close. I didn't technically assume that.

**AUDIENCE:** [INAUDIBLE]

**PROFESSOR:** They could double up. Yeah. There's no intersection in the events. In fact, there is intersection because there's a chance I rolled all six-- all Ns. Say N is 6. I could roll all sixes and then each of these would be a winner. But I don't get to count them separately. Then I only win once in that case.

In other words, all of these could turned on at the same time. There's an intersection here. So this rule does not hold. I need the  $A_i$  to be disjointed for this to be true-- the events to be disjointed.

And they're not disjointed because there's a sample point were two or more of the die could come up the same being a winner, which means the same sample point, namely all die are N, comes up in each of these three. So they're not disjointed.

Now what's the principal you used two weeks ago when you did cardinality of a set-- cardinality of a union of sets? Inclusion, exclusion. And the same thing needs to be done here.

So let's do that. And then we'll figure out the actual probability. So this is a fact based on the inclusion, exclusion principle. The probability of  $A_1$ , union  $A_2$ , union  $A_3$ , is just what you think it would be from inclusion, exclusion.

It's a probability of  $A_1$  plus a probability of  $A_2$  plus the probability of  $A_3$  minus the pairwise intersections.  $A_1$  intersect  $A_3$  minus probability of  $A_2$  intersect  $A_3$ . And is there anything else? Plus, the probably of all of them matching.

OK. So the proof is really the same proof you use for inclusion, exclusion with sets. The only difference is that in a probability space, we have weights on the elements. And the weight corresponds to the probability.

So in fact, if you were drawing the sample space, say here's  $A_1$  and here's  $A_2$ , and here's  $A_3$ . Well, you need to add the probabilities here, here, and here. Then you subtract off the double counting from here, from here, and from here. And then you add back again what you subtracted off too much there. Same proof, it's just your have weights on the elements of probabilities.

All right. So let's figure out the right probability. That's  $1/6$ ,  $1/6$ ,  $1/6$ . What's the probability of

the first two die matching-- both of them?  $1/36$ . We'll talk more about why that is next time. But there's a 6 for A1 then given that  $1/6$  for the second die matching. So it's  $1/6$  times  $1/6$  minus the  $1/36$ .  $1/36$ , the chance that all three match is  $1/216$  or 6 cubed. So when you add all that up you get the 0.421 and some more.

So the chance of winning this game is 41% which makes it a worst game in the casino. It is hard to find a worse game than this. Roulette, much better. We'll study Roulette in the last lecture-- much better game. And even that's a terrible game to play.

So it looks like an easy game. There's a quick proof that it's 50-50. But it's horrible odds against the house. Now, this is a nice example because it shows how a rule you had for computing the cardinality of a set gives you the probability. All right.

In fact, all the set laws you learned a couple weeks ago work for probability spaces the same way. And there were several of those in homework that you just had the last problem set. Any questions about that?

OK. Now in addition, all those set laws you did also work for conditional probabilities. For example, this is true. The probability of A union B given C-- whoops-- given C, is the probability of A given C plus the probability of B given C minus the intersection, A intersect B given C. In other words, take any probability rule you have and condition everything on an event, C, and it still works.

And the proof is not hard. You can go through each individual law but it all comes out to be fine. All right. You have to be a little careful though because you got to remember which side you're doing, which what you're putting on either side of the bar here. For example, what about this one? Is this true?

Claim. Let's take-- say C and D are disjoint. Is this true? Then the probability of A conditioned on C union D. So given that either C or D is true, does that equal the probability of A given C plus probability of A given D?

We know that if I swapped all these, it's true. The probability of C union D when C and D are disjoint is the probability that C given A plus the probability of D given A. That I just claimed.

And what about this way? Can I swap things around? Yeah?

**AUDIENCE:** [INAUDIBLE] would C union D be 0?

**PROFESSOR:** If C and D are disjoint,  $C \cup D$  would just be  $C \cup D$ . But you're not a good point. What if C and D are disjoint? That's a good example. Let's draw that. Let's look at that case.

So we've got a sample space here. And you've got C here and D here. And just for fun, let's make A be here-- include all of them. What's the probability-- is this going to do what I want? Yeah. What's the probability of A given C? 1. If I'm in C I'm in A. A is everything here. So the probability of A given C is one. What's the probability of A given D? 1.

All right. Well, this is a problem because I can't have the probability of-- what's the probability of A given  $C \cup D$ ? Well, it can't be 2. Right? It's 1. They are not equal.

So you cannot do those set rules on the right side of the conditioning bar. You can do them on the left, not on the right. All right. So this is not true. Now nobody would do this. Right? I mean, the probability of-- not that it's-- see this example?

This you just would never make this mistake again seeing that example. Everybody understand the example, how it's clearly not always the case that probability of A given  $C \cup D$  is a probability of A given C plus probability of A given D?

Because now I'm going to show you an example where you're going to swear it's true. All right? And this is a real life example. Many years ago now there was a sex discrimination suit at Berkeley.

There was a female professor in the math department. And she was denied tenure. And she filed a lawsuit against Berkeley alleging sex discrimination. Said she wasn't tenured because she's a woman.

Now, unfortunately sex discrimination is a problem in math departments. It's historically been a difficult area. But it's always hard to prove. It's a nebulous kind of thing.

They don't say, hey, you can't have tenure because you're a woman. They'd get sued and get killed for that. So she had to get some math to back her up. So what she did is she looked into Berkeley's practices and she found that in all 22 departments, every single department, the percentage of male PhD applicants that were accepted was higher than the percentage of female PhD applicants that were accepted.

Now you could understand some of the departments accepting more male PhDs than female PhDs. But all 22? What are the odds of that? I mean, so the immediate conclusion is, well,

that's clearly there's sex discrimination going on at Berkeley. OK?

Well Berkeley took a look at that and said, nothing good. That doesn't look good for them. But they did their own study of PhD applicants. And they said that if the university as a whole-- look at the University as a whole, actually, the women, the females have a higher acceptance rate for the PhD Program than the men.

So look. Berkeley said, we're accepting more women than men percentage-wise. So how could we be discriminating against women? And this is where the same argument the female faculty member's making, But they're saying as a university as a whole, when you add up all 22 departments. Well, that sounds pretty good. How could they be discriminating?

OK. So the question for you guys, is it possible that both sides we're telling the truth, that in every single department the women have a lower acceptance rate than men, but on the university as a whole the women are higher percentage? It sounds like it's-- and just to avoid any confusion here, people only apply to one department and they're only one sex. So you can't-- Carroll didn't apply.

[LAUGHTER]

How many people think that one of the sides, actually, when they look at the studies was wrong, that they're contradictory? Nobody? You've been in 6 over 2 too long. How many people think it's possible that both sides were right? Yeah. All right. So let's see how this works.

And to make it simple I'm going to get down to just two departments rather than try to do data for all 22. And I'm going to do not the actual data but something that's represents what's going on. OK. So we're going to look at the following events. A is the event that the applicant is admitted. FCS is the event that the applicant is female and applying to CS.

FEE is the event that the applicant is female and applying to EE. MCS is the event the applicant is a male and CS. And then finally we have MEE is the event the applicant is male and in EE. So we're just going to look at two departments here and try to figure out if it can happen that in both departments the women are worse off but if you take the union they're better off.

So the female professor's argument effectively is, the probability of being admitted given that

you're a female in CS is less than the probability of being admitted given that you're a male at CS. And same thing in EE. Probability of being admitted in EE if you're a female is less than if you're a male.

OK? Now Berkeley is saying it's sort of the reverse. The probability that you're admitted given that you're a female in either department is bigger than the probability of being admitted if you're a male in either department. OK. So we've now expressed their arguments as conditional probabilities Any questions?

Can you sort of see why this seems contradictory? Not plus, union. Because this is sort of like- - these are just joined. This is the sum of those. And this is sort of the sum of those. And yet the inequality changed.

All right. In fact, this is the logic that we've just debunked over there-- exactly that claim. In fact, these are not equal as the sum. So let's do an example. Say that-- let's do it over here. I'll put the real values in over here.

Say that for women in computer science, 0 out of 1 were admitted compared to the men, were 50 out of 100 were admitted. And then in EE, 70 out of 100 women were admitted compared to the men, which had 1 out of 1. All right? So as ratios, 70% is less than 100%. 0% is less than 50.

Now if I look at the two departments as a whole, I get 70 over 101 is in fact bigger than 51 over 101. All right? And so as a whole women are a lot more likely to be admitted even though in each department they're less likely to be admitted. OK?

So what went wrong with the intuition, which you didn't fall victim to, but people often do, that it shouldn't have been possible given that? What's going on here that make it so that it's not a less than when you look at the union of the departments? Yeah?

**AUDIENCE:** [INAUDIBLE] they're weighted differently?

**PROFESSOR:** Yeah. They're weighted very differently. You got huge waves here. Right? So if I look at the average of the percentages here, well it's 35% for the women versus 75% for the men. So the average of the percentage is just what you'd think. 35 is less than 75.

But I've got huge weightings on these guys, which changes the numbers quite dramatically. So it all depends how you count it. Actually, who do you think had a better-- Yeah. Go ahead.

**AUDIENCE:** [INAUDIBLE]

**PROFESSOR:** Who won the lawsuit? Actually, the woman won the lawsuit. And which argument would you buy now? You've got two arguments. Which one would you believe if either? Which one? I mean, now if I look at exactly this data I might side-- I might side with Berkeley looking at these numbers.

Then again, when you think about all 22 departments and the fact they weren't this lopsided, not so good. So in the end Berkeley lost. I'm going to see another example in a minute where it's even more clear which side to believe in. But it really depends on the numbers as to which one you might, if you had to vote, which way you'd vote.

Here's another example. This is from a newspaper article on which airlines are best to fly because they have the best on-time rates. And in this case they were comparing American Airlines and America West, looking at on-time rates. And here's the data they showed for the two airlines.

Here's American Airlines. Here's America West. And they took five cities, LA, Phoenix, San Diego, San Francisco, and Seattle. And then you looked at the number on time, the number of flights, and then the rate, percentage on time. And then same thing here. Number on time, number of flights, and the rate.

So I'm just going to give you the numbers here. So they had 500 out of 560 for a rate of 89%, 220 over 230 for 95, 210 over 230 for 92%, 500 over 600 for 83%, and then Seattle. They had a lot of flights. That's where they're-- we have a hub of 2,200 for 86%.

And if you added them all up, they got 3,300 out of 3,820 for 87% on time. Now the data for American West looks something like the following. In LA it's 700 out of 800 for 87%. they're based in Phoenix. They got a zillion flights there. 4,900 out of 5,300 for 92%. And 400 over 450 for 89%, 320, over 450, 71%, 200 over 260 for 77%. And then you add all them up. And you've got 6,520 over 7,260 for 90%.

So the newspaper concluded and literally said that American West is the better airline to fly because they're on-time rate is much better. It's 90% versus 87%. What do you think? Which airline would you fly looking at that data?

**AUDIENCE:** [INAUDIBLE]

**PROFESSOR:**

I know which one I'd fly. It looks like America West is better. Every single city, American Airlines is better. 92 versus 89. Everywhere it's better by a bunch. 83 versus 71. 86 versus 77. Every single city, American Airlines is better. Yet, America West is better overall. And that's what the newspaper said. They went on this.

But of course, no matter where you're going you're better off with American Airlines. All right? Now what happened here? The waiting. In fact, America West flies out of Phoenix where the weather's great. So you get a higher on-time rate when in a good-weather city. And they got most of their flights there. American Airlines got a lot of flights in Seattle where the weather sucks and you're always delayed.

All right? And so they look worse on average because so many of their flights are in a bad city and so many of America West are in a good city. All right? So it makes America West look better when in fact, in this case, it's absolutely clear whose better.

American Airlines is better, every single city. All right. That's why Mark Twain said, "There's three kinds of lies-- lies, damned lies, and statistics." We'll see more examples next time.