

## MITOCW | watch?v=j-DXRbhJ1DA

---

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](https://ocw.mit.edu).

- PROFESSOR:** OK. So first we're going to review last class. The first question, what does it mean if sets A, B, C are a partition of set D? If you know it, just raise your hand. Yeah.
- AUDIENCE:** A, B, and C are not in set D?
- PROFESSOR:** Hmm?
- AUDIENCE:** A, B, and C are not in set D?
- PROFESSOR:** No
- AUDIENCE:** A, B, and C make up D?
- PROFESSOR:** OK. And what are we assuming?
- AUDIENCE:** That the [INAUDIBLE]? Oh, that A, B, and C are set D.
- PROFESSOR:** Yeah. So they're disjoint. OK? Did everyone hear that? So A, B, and C are disjoint and they make up all of D. OK. So how do you calculate P, probability of A given B, using the formula for conditional probability? So essentially I'm just asking you what is the formula of conditional probability? Slide. C disjoint-- OK. So, two? Can anyone tell me? So probability of A given B. How do you calculate that?
- AUDIENCE:** The number of possible outcomes of A intersect B over--
- PROFESSOR:** Yup. Does everyone understand why we do this? So like we said with the universal set, we always do probability of A over the universal set, which is 1. But for a conditional probability, we change the universal set to B. So we do the event that both of these happen over our new set, which is B. OK? Does everyone understand that? Yeah?
- AUDIENCE:** Wait. What does the upside down U mean?
- PROFESSOR:** Oh, were you not here? OK. This intersect--
- AUDIENCE:** Sorry.
- PROFESSOR:** No, that's fine. So if you have A and B-- do you know anything about set theory?

**AUDIENCE:** A little bit.

**PROFESSOR:** A little bit. OK. So this is event A, right? And this is B. And the intersect is anything that both share.

**AUDIENCE:** OK.

**PROFESSOR:** Yup. OK. What is the difference between  $P(A|B)$  and  $P(B|A)$ ? It's almost kind of self explanatory.

**AUDIENCE:** So the first one is what's the probability of A happening if you already know B. And the second one is what's the probability of B happening if you already know A.

**PROFESSOR:** And are these two same? Always? No. OK, yeah. So you'll have to calculate it out, because your universal set is different. OK. For if B causes A, what is the conditional probability that P of A given B?

So if you know that B causes A, what's the probability that B is happening given that B happened?

**AUDIENCE:** 100%?

**PROFESSOR:** Yup. So for that one, does anyone understand why that-- that's the reason? So if B causes A, if A happened, and you know B happened, A has to happen. Right? But if B was a possible cause of A, then you're not-- and A is caused by many things, then this is not necessarily 1, because something else could cause A. Right? But if you know B definitely causes A, then-- OK?

So in the last one, does conditional probability require that B causes A?

**AUDIENCE:** No.

**PROFESSOR:** No? Can you give me an example?

**AUDIENCE:** So if A is there is a white dog in the room, and B is there's a dog in the room. And you know that 45% of dogs are white. Then the fact that there's a dog in the room doesn't cause it to be a white dog,

**PROFESSOR:** Yeah.

**AUDIENCE:** But it does mean that there's a dog in the room. So it's more likely that there's going to be a white dog than if you don't know [INAUDIBLE] at all.

**PROFESSOR:** Yeah. So B gives you information about A, but it doesn't necessarily have to cause A. OK? Does that make sense? OK. Any questions? Yeah.

**AUDIENCE:** So, wait. I'm certainly confused. What is the A [INAUDIBLE] B means?

**PROFESSOR:** OK. So from last class we learned that-- you know that, populated B. Right? You know this is, right? So this means given B. It's like--

**AUDIENCE:** Given.

**PROFESSOR:** Yeah. B is knowledge that you are given, and then you have to figure out the probability of A with that new knowledge.

**AUDIENCE:** OK.

**PROFESSOR:** OK? Anything else? OK. We're going to move on to that Monty Hall problem. Did anyone work on it? You did have to, but-- OK. So what did you get for your answer? Do you stay or switch?

**AUDIENCE:** Always switch.

**PROFESSOR:** OK. How about?

**AUDIENCE:** Switch.

**PROFESSOR:** Switch?

**AUDIENCE:** Switch. Switch.

**PROFESSOR:** OK. Do any of you feel comfortable explaining it up here? How you? I know that you've done it once, last class. So you want to come up and explain it?

**AUDIENCE:** I sort of know how to explain it.

**PROFESSOR:** OK.

**AUDIENCE:** I know that it's the [INAUDIBLE], but--

**PROFESSOR:** OK. All right. How about you want to tell me first?

**AUDIENCE:** OK. Well for the probability of door 1 is  $2/3$ , and then, wait, and then if he shows the Door 2 then you know that the probability, most people would think that now it's cut in half, but it's actually  $2/3$ . So there's more of a probability that it would be--

**PROFESSOR:** OK. So yours is more intuitive, actually.

**AUDIENCE:** I guess.

**PROFESSOR:** Yeah. So you don't really need to write anything. Did everyone hear her? Yes? No?

**AUDIENCE:** No.

**PROFESSOR:** No. OK. So what she said was-- what she's essentially doing is kind of combining the doors. Oh, wait. Sorry. For you guys who weren't here, do you know what the problem is? No. OK. I'll go over that first. Just bear with me, people who were here. OK.

So the Monty Hall problem on the sheet. So you're at a game show. There's three doors, and one of these doors has equal probability of getting 100 million prize behind it. So the first step is that you pick a random door. So let's say you pick this one. And then-- but you don't open it yet, so you don't know what's behind it. And then the host picks one of these doors and opens it to reveal nothing.

So let's say he picks this one. I can't draw. Whatever. And there's nothing, right? There's nothing there. So this is what the host picks. And this is what you originally picked. And now he says, before he opens your door you still have the chance of switching to this door. So the an-- the question was, do you want to stay with your original door, or do you want to switch? Because you want to maximize your probability of does this one still have the prize, or is it this one.

Does that problem make sense to everyone? Wait. OK. So what-- what was your name, again?

**AUDIENCE:** Priyanka.

**PROFESSOR:** Priyanka says that here you have originally one third chance, right?

**AUDIENCE:**  $2/3$ .

**PROFESSOR:** Oh, no what-- your original door.

**AUDIENCE:** Oh,  $1/3$

**PROFESSOR:** Yeah. So when you first choose you only have  $1/3$  of a chance. And what she's doing is kind of combining these two doors as one, as  $2/3$  chance. And since you already know that this one is nothing, it doesn't really affect this probability, even though he opened it. So this door actually has  $2/3$  of a chance.  $2/3$  of a chance to have the million dollar prize. So the fact that the host opened this does nothing to change this probability.

So you're kind of seeing it as one door with one third chance, and this door a  $2/3$  chance. Right? That's the intuitive explanation. Does everyone understand this? It will take a little time. I know it took me like, a long time to understand it intuitively. So OK.

Did anyone do it using conditional probability, mathematical way? None of you?

**AUDIENCE:** Well, you can just draw out all the possibilities, because there are only six, right?

**PROFESSOR:** Right. So do you want to explain it up here? If you don't, I can do it. I don't want to make you. No? OK. I'll do it. So does everyone understand this? Can I erase it? Or I should just erase this one. OK. If I need to erase better, let me know, too. Because I don't know what it looks like far away. OK.

So using conditional probability, we know about the trees, right? And how you have certain events, and every time a new event happens you branch out the tree. OK. So let's say we have three doors again. If I'm in your way, let me know. OK. All right. And you don't know which one has a million dollars. So let's just assume you pick door A. So we're just going to assume in this tree that you picked door A. OK? OK.

So your first event is, where's the prize? Right? So where is the prize? It has  $1/3$  chance, right, of being in door A, B, or C. OK. So this is your first step. Does everyone do that? OK. So the important thing to know is that your host knows where the door-- where the prize is, because if it doesn't, then it doesn't really add any additional information, right? So now we need to know the host picks which door, right?

So if the prize is behind door A, and you pick door A, then he has two choices, right? Because both are empty. He can choose either B or C, right? And there's a half chance that he'll pick B and C. Does that make sense to everyone? So if he wants to reveal an empty door, he has two choices. Because the actual prize is in the door you chose.

But if the prize is in B, and you picked A, which door can he open?

**AUDIENCE:** C.

**PROFESSOR:** C. Yeah. So the only one he can open is C, right? Same thing for C. He only can choose B. So does everyone see that? Yeah? OK. So you know with trees we multiply out the probabilities, right? So you know this has, what,  $1/6$ ,  $1/6$ ,  $1/3$   $1/3$ . OK?

So that's just the problem set up. If you want to answer the question, you need to figure out, what if I stay in door A, and what if I switch? So if you stay here, you win, right? So you get money. If you stay, you win. Right? And if you switch, you lose. So here you lose, right? Because it's actually in B. And here you lose because the prize is actually in C, right? And if you switch, you get it. Does this make sense? I know I'm going kind of fast. So Are there any questions so far?

OK. So you can figure out the probability of winning if you stay. So you add  $1/6$  plus  $1/6$ , is  $1/3$ . Right? And then the probability of winning if you switch is  $2/3$ . So this proves mathematically that switching will win-- will get you the better probability of winning. Is there any part that confuses anybody? Yeah?

**AUDIENCE:** Can you go over the stay and switch part?

**PROFESSOR:** OK. So this is all assuming that along each way that the prize is in A, right? So if the prize is in A-- that's all right here, right-- if the prize is in A, you stay in A, you're going to win, right? But if you switch, and you're assuming that the prize is in door A, then you're not going to win, right?

But for this row, you're assuming that prize is in B, and if you switch, you have to win. And symmetrically this is the same way, too. All right. That make sense? Anything else? OK. So that just proves, sometimes conditional probability is good for you. OK.

So we're going to move on to this class's stuff, which is Bayes' rule. Oh, crap. Did anyone need this? I can leave it up. No? OK. If you still need it after class, let me know. I'm sorry. I'll ask before I do that. OK.

OK. So I'm not going to write out the first thing. But Bayes' rule is basically finding out your reverse probability. So remember I was asking you what's the difference between this, and this. Right? So if you look at the slide that I handed you out. The first slide that says Bayes' rule. If we use the radar example that we showed from before. When we did do the problem

we figured out the probability that the radar registers, given that the plane is present. Right? But now we want to know if the plane is present, given that the radar registers. Does everyone see the difference in that?

So the first one is kind of saying how accurate the radar is, but the second one is what you really want to know, is how much you can rely on the radar. Because the first one is more of a mechanical thing, but the second is actually using the radar. Does that make sense to everyone? OK. So I'll write it up here.  $P$  is our event that the plane is present.

All right. If you can't read my handwriting, let me know that, too. Radar. OK. So what we want to know is the probability that the plane is there given that the radar registers. OK, so for you guys who weren't here, and just for you other guys, too, I'll write out the chart again. The tree chart. I'll just do it up here.

So if you guys first remember, if the plane was present, we have a 0.05 probability that the plane is present. Which means a 0.95 probability that it's not present. And the next thing we have was whether the radar picked up on it, right? So whether it registered. OK. So we were given, last time, the probabilities of this. Given whether the plane was there or not. So the radar registers 0.99 at the time the plane is there, which means 0.01 if it's there but it doesn't register. And then if it registers anyway, even if it's not there, that's a 0.1 chance, which means a 0.90 chance here.

For you guys who weren't here, do understand the problem setup? This is all you need to know to understand the next step. OK. And using the multiplication rule you can get 0.0455, 0.0005, 0.0950, and 0.8550. OK? Does everyone understand what these numbers are referring to?

So this is the probability that this part of the branch happened. The plane is there and the radar says yes. Et cetera, et cetera, et cetera. OK? Yeah?

**AUDIENCE:** So that's achieved by multiplying the two together, right?

**PROFESSOR:** Which you can do for a sequence of events like this. All right. OK. So if you see 0.99 is the probability that the radar registers given that the plane is there. Right? But we actually want to know this, the reverse of it. So you've seen the definition of probability. You have probability of  $P$  given  $R$  equals-- can anyone tell me?

**AUDIENCE:** The intersect R. over the probability of R.

**PROFESSOR:** Yup. OK. So first we can find probability of R. So given this thing, can anyone tell me what the probability that the radar registers is? Or have an idea of how you can figure out what this probability is?

OK. So probability of the radar registry-- I should have done this over there, but-- you have it here and here, right? So what you need to do-- so that's this branch plus this branch. Right? So you add this probability plus this probability. So you have 0.0495 plus 0.0950. And what's the probability of this? Can anyone see that in the branch? The probability that the plane is there and the radar registers.

**AUDIENCE:** 0.0495.

**PROFESSOR:** So it's only this branch. Right? What she said. So it's 0.0495, and you're left with 0.3426, which is about a 34% chance. Does everyone understand how we got from what we were given, the probability of the radar registering given that the plane is present, to the probability that the plane is present given that the radar says it's there. Does everyone understand?

So this probability, even though the radar is pretty accurate, right? It's 0.99% chance. You still have a 34% chance that the radar-- that the plane is actually there given that the radar says yes. So even though it seems like a very accurate radar, this probability is not really what you want. You want to be sure that the plane is there if the radar says yes. So in an ideal world you have 100%. Right?

So the thing that's throwing this off is-- the radar off, is probably this part. You don't want the radar to say yes if it's not actually there. So if you have an ideal radar, this would be zero. Right?

**AUDIENCE:** Can I go to the restroom.

**PROFESSOR:** Yes. If you need to pee, go. Yes?

**AUDIENCE:** So [INAUDIBLE] one of these problems is you kind of find out what the probability of P is in this example if R is true. All right.

**PROFESSOR:** Yup. Does this make sense to everyone? OK. So the next slide. There's an example of this, this military application. I'm actually working at Lincoln Labs, if you don't know about it. And



what they do a lot of military defense like this. So if you were trying to register if a plane was there, that plane could be an enemy aircraft, or it could be a commercial aircraft. And you want the radar to make sure it knows what it's seeing. Because you don't want to waste a missile and shoot something that's not actually what you're seeing. Right? Yeah.

So ideally in the military they have really good radars. It won't be this kind of probability. OK. Does that help figure out why we need this kind of stuff? OK. Oh, yeah, and this is an example of Bayesian probability that we mentioned before how Bayesian probability is a measure of how much you believe something will happen. So this is like that. Right? You can't repeat it over and over again like a coin, exact same experiment all by itself. This is dependent on whether the plane is there or not. OK. Any questions? OK.

In order for us to figure out Bayes' rule, we have to make several assumptions. And you can see it on the slide. We have the-- OK. So we have probability of A is what we know. Is whether the radar registers or not. So we have all that information. Right? If you don't know this information. Like, say you don't know how accurate the radar is, then you can't ever figure out the reverse. Right? So you have to know everything about R before you can figure out P of R.

So that's just saying that if you have a bunch of things like, this is event A1, this is event A2, and this is event A3. And you have probability of B happening. But you don't really know what probability of B is, but you know the probability of A2 and B, probability of A1 and B, etc. Then in some way or another you can get probability of B. And I write that out better. So this is used-- Bayes' rule uses the total probability theorem, which I have written out there. I'll write it out again.

Probability of B equals probability of A1 and B happening, plus everything all the way up to An and B. So for the radar example we only had two A's. Right? A is the probability that the radar registers, and A2 is whether-- wait, hold on. Sorry. Hold on. OK. It is there. OK.

So this was actually our probability that the radar registers given that the plane is there, union with the radar registers plus the probability that-- I'm sorry, this P is really confusing-- plus the probability that the plane isn't there. Union with R. Can you guys see that? So you can do that for multiple things, but for this case, we only had two. So probability that the plane was there, probability that it wasn't there. So that's this part, and this part.

And the way you can see this with a tree is that you have A1 happening, A2 happening, An happening. And then you know if B happens or if B does not happen. You know if B happens

or B doesn't happen. So that's kind of what we did, right? We know the plane is there or not, and then we know in each case whether the radar registers or not. And then to get probability of B you just do this branch, this branch, and this branch. So then you can get probability of B. So does that make sense? It's a little confusing, let me know.

So even though we don't know probability of B off hand, but you know everything that can lead to B, and all the probabilities, you can get B. Does that make sense? Any questions? OK.

So Bayes' rule is basically figuring out, you've seen the total probability theorem. You're reverse. So in the end, you have probability of  $A_i$ , whatever it is, for us it's  $A_1$ , given B equals probability of A-- and I use this to figure out the bottom half. To get this. This is just a generic way of doing what we did earlier. Does everyone see that? OK. Can I erase this? Yeah? OK.

So those are real examples of you seeing Bayes' rule. It's very useful. We use it a lot in artificial intelligence. In that class I use it a lot. I'm sure they use it a lot for other applications, too. So that's a very important concept. Make sure you get that straight. OK.

So we're going to use conditional probability to derive what independence means. So if I told you that the probability of A given B is actually equal to the probability of A, what does that indicate? That's kind of like saying, if you have B, it doesn't really matter. You still have the probability of A. So B doesn't affect A in any other way. They're independent. And that also works for the reverse, because these are just numbers. Right? So if A is independent of B, B is independent of A. OK? So that's how you define independence.

So in order to figure out if something is independent or not, we're going to use conditional probability. So you have probability of A equals  $P(A)$ , A given B. And B doesn't matter, but we're going to use this, anyway. So given the definition of conditional probability, we have this is  $\frac{P(A \cap B)}{P(B)}$ . And you can kind of erase that and put in  $P(A)$ . So if this is true, that means this has to be true. And since we don't like it when we divide by zero, we're just going to move this up. So you get this right. So this is how you test for independence. OK?

So if I told you that the probability of A given B still equals the probability of A, it means B doesn't matter, which means this has to hold true. Does everyone see how I got that? How that makes sense? OK. This is a question mark, if you don't know. So if you don't know, you try it out. If it's equal, means independent.

So if you have two disjoint events, are they independent? You have A, B. Are these two

independent?

**AUDIENCE:** So that means if you have one then you definitely can't have the other?

**PROFESSOR:** Right. So that's intuitive, although a lot of people think that these two are independent. Or not independent. Or independent. Sorry. But they actually aren't independent. So if you do the math, the probability of each of these has to be greater than zero, but their intersection is zero. So if you have probability, you have your test. And this holds true with independent. A union B? Zero. Because they are disjoint. But you can't really have an event unless you have greater than zero probability.

So this is a greater than zero thing. This is a greater than zero. And they obviously can't equal zero. So it's not independent. It's like saying, I flipped a coin and I got heads. That definitely means you can't have tails. So even though they're disjoint, they're dependent on each other, because if one happens, the other one can't for sure. Does that make sense to everyone? This trips a lot of people up. OK. We can't move this. I'll just work on this.

So we're going to use independence to prove that successive rolls are independent of each other. A lot of times they just tell you that rolls are independent of each other. But we're going to use this to prove it. So instead of a six sided die, we have a four sided die. And the sides, the numbers on the sides are 1, 2, 3, and 4. So instead of 1-6 we have 1-4. And the answer we're trying to get is, are successive rolls independent? So.

OK? And on the sheet I have for you, I did write out the sample space just in case. So you guys can see. Each one of those is a combination. So the first one, 1-1, is your first roll is a 1, your second roll is a 1, et cetera, et cetera. OK? So for our events we're going to do A equals first roll is  $i$ . And because this is a very generalized answer, we're not going to make it first roll is 1, or 2, or 3, or 4.  $i$  can be any of these numbers. Right? So does that event make sense? OK? We don't want to specify too much. And B is the same thing, except it's the second roll is, say,  $j$ . And  $j$  has to be in here, too. OK? So that's how we're going to define our problem.

So we have our test again. First I'm going to do  $P(A \text{ and } B)$ . So what this means is that the probability that the first roll is  $i$  and the second roll is  $j$ . And that has to be  $1/16$ , because there are 16 different combinations. And given an  $i$ , given a  $j$ , you only can have one. Does this make sense? So if we say  $i$  is 1,  $j$  is 4, there's only  $1/16$  chance.

And the next step is the left side. So I want to figure out, what's the probability of A with  $i$ . So if

we assume that  $i$  equals 1 again, how many different ways, in that sample space, is the first roll an  $i$ ? Well you have 1-1, 1-2, 1-3, 1-4. So it's four different combinations. Does that make sense? Should I write this out? So you have 1-1, 1-2-- this is assuming  $i$  equals 1. But because I want to generalize it, that's why I'm just saying it's  $i$ . But you can really just fill in any number, and it would be the same probability.

So you just have  $B$  equals  $j$  on it's own, keep in mind that this is the second roll, right? So if we assume again that  $j$  equals 4, it's still the same probability, because now you have 1-4-- thank you, by the way-- 1-4, 2-4, 3-4, 4-4. Everyone see that? So the second-- this is the probability that the second roll is a 4. But I actually generalized it to  $j$ . I'm just doing  $j$  equals 4. So you can see that better. So that's 4 out of 16.  $i$  and  $j$  can be any number, again.

So if we do that test for independence, we have probability of  $A$ , probability  $B_j$  equals probability of  $A_i$  union  $B$ . This is  $1/16$ . Does everyone see that? So if we assume  $A$  is 1 and  $B$  is  $j$ , you only can have  $1/4$  out of the 16 combinations. So that's  $1/16$ . And then this probability is  $4/16$ .  $4/16$ . 1, 4. 1, 4. They're equal. Right? So that means they are independent. Did everyone see how I did this? OK.