

## MITOCW | Lec-21

---

PATRICK WINSTON: Here we are, down to the final sprint.

Three to go.

And we're going to take some of the last three, maybe two of the last three, to talk a little bit about stuff having to do with probabilistic approaches-- use of probability in artificial intelligence.

Now, for many of you, this will be kind of a review, because I know many of you learned about probability over the [? sand ?] table and every year since then.

But maybe we'll put another little twist into it, especially toward the end of the hour when we get into a discussion of that which has come to be called belief nets.

But first, I was driving in this morning, and I was quite astonished to see, as I drove in, this thing here.

And my first reaction was, oh my god, it's the world's greatest hack.

And then I decided, well, maybe it's a piece of art.

So I'd like to address the question of how I could come to grips with that issue.

There's a distinct possibility that this thing is a consequence of a hat, possibly the result of some kind of art show.

And in any event, some sort of statue appeared, and statues don't usually appear like that.

So I got the possibility of thinking about how all these things might occur together or not occur together.

So the natural thing is to build myself some sort of table to keep track of my observations.

So I have three columns in my table.

I've got the possibility of a statue appearing, a hack having occurred, and some sort of art show.

And so I can make a table of all the combinations of those things that might appear.

And I happen to have already guessed that there are going to be eight rows in my table.

So it's going to look like this.

And this is the set of combinations in this row where none of that occurs at all.

And down here is the situation where all of those things occur.

After all, it's possible that we can have an art show and have a hack be a legitimate participant in the art show.

That's why we have that final row.

So we have all manner of combinations in between.

So those are those combinations.

Then we have F, F, T, T, F, F, T, T, F, T, F, T, F, T, F, T.

So it's plain that the number of rows in the table, or these binary possibilities, is 2 to the number of variables.

And that could be a big number.

In fact, I'd love to do a bigger example, but I don't have the patience to do it.

But anyhow, what we might do is in order to figure out how likely any of these combinations are, is we might have observed the area outside the student center and rest of campus over a long period of time and keep track of what happens on 1,000 days.

Or maybe 1,000 months or 1,000 years.

I don't know.

The trouble is, these events don't happen very often.

So the period of time that I use for measurement needs to be fairly long.

Probably a day is not short enough.

But in any case, I can keep a tally of how often I see these various combinations.

So this one might be, for example, 405, this one might be 45, this one might be 225, this one might be 40, and so on.

And so having done all those measurements, kept track of all that data, then I could say, well, the probability that at any given time period one of these things occurs will just be the frequency-- the number of tallies divided by the total number of tallies.

So that would be a number between 0 and 1.

So that's the probability for each of these events.

And it's readily calculated from my data.

And once I do that, then I can say that I got myself a joint probability table, and I could perform all manner of miracles using that joint probability table.

So let me perform a few of those miracles, while we're at it.

There's the table.

And now, what I want to do is I want to count up the probability in all the rows where the statue appears.

So that's going to be the probability of the statue appearing.

So I'll just check off those four boxes there.

And it looks like the probability of the statue appearing is about 0.355 in my model.

I don't think it's quite that frequent, but this is a classroom exercise, right?

So I can make up whatever numbers I want.

Now, I could say, well, what's the probability of a statue occurring given that there's an art show?

Well, I can limit my tallies to those in which art show is true, like so.

And in that case, the probability has just zoomed up.

So if I know there's an art show, there's a much higher probability that a statue will appear.

And if I know there's a hack as well as an art show going on, it goes up higher still to 0.9.

We can also do other kinds of things.

For example, we can go back to the original table.

And instead of counting up the probability we've got a statue, as we just did, we're going to calculate the probability that there is an art show.

I guess that would be that one and that one, not that one, but that one.

So the probability there's an art show is one chance in 10.

Or we can do the same thing with a hack.

In that case, we get that one off, that one on, that one off, that one on, that one off, that one on, that one off.

So the probability of a hack on any given time period is about 50-50.

So I've cooked up this little demo so it does the "ands" of all these things.

It could do "ors," too, with a little more work.

But these are just the "ands" of these various combinations.

Then you can ask more complicated questions, like for example, you could say, what is the probability of a hack given that there's a statue?

And that would be limiting the calculations to those rows in which the statue thing is true.

And then what I get is 0.781.

Now, what would happen to the probability that it's a hack if I know that there's an art show?

Will that number go up or down?

Well, let's try it.

Ah, it went down.

So that's sort of because the existence of the art show sort of explains why the statue might be there.

Now, just for fun, I'm going to switch to another situation, very similar.

And the situation here is that a neighbor's dog often barks.

It might be because of a burglar.

It might be because of a raccoon.

Sometimes, there's a burglar and a raccoon.

Sometimes, the damn dog just barks.

So let's do some calculations there and calculate the probability that a raccoon is true, similar to what we did last

time.

Looks like on any given night-- it's kind of a wooded area-- there's a high probability of a raccoon showing up.

And then we can ask, well, what is the probability of the dog barking given that a raccoon shows up?

Well, in that case, we want to just limit the number of rows to those where a raccoon-- or where the dog is barking.

Looks like the probability of the dog barking, knowing nothing else, is about  $[\frac{3}{7}]$ . But now we want to know the probability of the raccoon-- that's these guys here need to get checked.

These are off.

So that's the probability of a raccoon.

Did I get that right?

Oh, that's probability of a burglar.

Sorry, that was too hard.

So let me go back and calculate-- I want to get the probability of a raccoon.

That's true, false, true, false, true, false, true.

So the probability of a raccoon, as I said before is 0.5.

Now, what happens to that probability if I know the dog is barking?

Well, all I need to do is limit my rows to those where the dog is barking, those bottom four.

And I'll click that there, and you'll notice all these tallies up above the midpoint have gone to zero, because we're only considering those cases where the dog is barking.

In that case, the probability that there's a raccoon-- just the number of tallies over the total number of tallies-- gee, I guess it's 225 plus 50 divided by 370.

That turns out to be 0.743.

So about 75% of the time, the dog barking is accounted for-- well, the probability of a raccoon under those conditions is pretty high.

And now, once again, I'm going to ask, well, what is the probability of a raccoon, given that the dog is barking and there's a burglar?

Any guess what will happen there?

We did this once before with the statue.

Probability first went up when we saw the statue and then went down when we saw another explanation.

Here's this one here.

Wow, look at that.

It went back to its original condition, its a priori probability.

So somehow, the existence of the burglar and the dog barking means that the probability of a raccoon is just what it was before we started this game.

So those are kind of interesting questions, and there's a lot we can do when we have this table by way of those kinds of calculations.

And in fact, the whole miracle of probabilistic inference is right in front of us.

It's the table.

So why don't we go home?

Well, because there's a little problem with this table-- with these two tables that I've shown you by way of illustration.

And the problem is that there are a lot of rows.

And I had a hard time making up those numbers.

I didn't have the patience to wait and make observations.

That would take too long.

So I had to kind of make some guesses.

And I could kind of manage it with eight rows-- those up there.

I could put in some tallies.

It wasn't that big of a deal.

So I got myself all those eight numbers up there like that.

And similarly, for the art show calculations, produced eight numbers.

But what if I added something else to the mix?

What if I added the day of the week or what I had for breakfast?

Each of those things would double the number of rows of their binary variables.

So if I have to consider 10 influences all working together, then I'd have 2 to the 10th.

I'd have 1,000 numbers to deal with.

And that would be hard.

But if I had a joint probability table, then I can do these kinds of miracles.

But Dave, if I could have this little projector now, please.

I just want to emphasize that although we're talking about probabilistic inference, and it's a very powerful tool, it's not the only tool we need in our bag.

Trouble with most ideas in artificial intelligence is that their hardcore proponents think that they're the only thing to do.

And probabilistic inference has a role to play in developing a theory of human intelligence.

And it certainly has a practical value, but it's not the only thing.

And to illustrate that point, I'd like to imagine for a few moments that MIT were founded in 1861 BC instead of 1861 AD.

And if that were so, then it might be the case that there would be a research program on what floats.

And this, of course, would be a problem in experimental physics, and we could imagine that those people back there in that early MIT would, being experimentally minded, try some things.

Oh, I didn't know that's what happened.

It looks like chalk floats.

Here's a rock.

No, it didn't float.

Here's some money.

Doesn't float.

Here's a pencil.

No, it doesn't float.

Here's a pen.

Here's a piece of tin foil I got from Kendra.

That floats.

That's a metal.

The other stuff's metal, too.

Now I'm really getting confused.

Here's a little wad of paper.

Here's a cell ph-- no, actually, I've tried that before.

They don't float.

And they also don't work afterward, either.

I don't need to do any of that in the MIT of 1861 AD and beyond, because I know that Archimedes worked this all out.

And all I have to do is measure the volume of the stuff, divide that by the weight, and if that ratio is big enough, then the thing will float.

But back in the old days, I would have to try a lot of stuff and make a big table, taking into account such factors as

how hard it is, how big it is, how heavy it is, whether it's alive or not.

Most things that are alive float.

Some don't.

Fish don't, for instance.

So it would be foolhardy to do that.

That's sort of a probabilistic inference.

On the other hand, there are lots of things where I don't know all the stuff I need to know in order to make the calculation.

I know all the stuff I need to know in order to decide if something floats, but not all the stuff I need to know in order, for example, to decide if the child of a Republican is likely to be a Republican.

There are a lot of subtle influences there, and it is the case that the children of Republicans and the children of Democrats are more likely to share the political party of their parents.

But I don't have any direct way of calculating whether that will be true or not.

All I can do in that case is what I've done over here, is do some measurements, get some frequencies, take some snapshots of the way the world is and incorporate that into a set of probabilities that can help me determine if any given parent is a Republican, given that I've observed the voting behavior their children.

So probability has a place, but it's not the only tool we need.

And that is an important preamble to all the stuff we're going to do today.

Now, we're really through, because this joint probability table is all that there is to it, except for the fact we can't either record all those numbers, and it becomes quickly a pain to guess at them.

There are two ways to think about all this.

We can think about these probabilities as probabilities that come out of looking at some data.

That's a frequentist view of the probabilities.

Or we could say, well, we can't do those measurements.

So I can just make them up.

That's sort of the subjective view of where these probabilities come from.

And in some cases, some people like to talk about natural propensities, like in quantum mechanics.

But for our purposes, we either make them up, or we do some tallying.

Trouble is, we can't deal with this kind of table.

So as a consequence of not being able to deal with this kind of table, a gigantic industry has emerged for dealing with probabilities without the need to work up this full table.

And that's where we're going to go for the rest of the hour.

And here's the path we're going to take.

We're going to talk about some basic overview of basic probability.

Then we're going to move ourselves step by step toward the so-called belief networks, which make it possible to make this a practical tool.

So let us begin.

The first thing is basic probability.

Let us say basic.

And basic probability-- all probability flows from a small number of axioms.

We have the probability of some event  $a$  has got to be greater than 0 and less than 1.

That's axiom number one.

In a binary world, things have a probability of being true.

Some have a probability of being false.

But the true event doesn't have any possibility of being anything other than true, so the probability of true is equal to 1, and the probability of false-- the false event, the false condition-- has no possibility of being true, so that's 0.

Then the third of the axioms of probability is that the probability of  $a$  plus the probability of  $b$  minus the probability

of  $a$  and  $b$  is equal to the probability of  $a$  or  $b$ .

Yeah, that makes sense, right?

I guess it would make more sense if I didn't switch my notation in midstream--  $a$  and  $b$ .

So those are the axioms that mathematicians love to start up that way, and they can derive everything there is to derive from that.

But I never can deal with stuff that way.

I have to draw a picture and think of this stuff in a more intuitionist type of way.

So that's the formal approach to dealing with probability, and it's mirrored by intuitions that have to do with discussions of spaces, like so, in which we have circles, or areas, representing  $a$  and  $b$ .

And to keep my notation consistent, I'll make those lowercase.

So you can think of those as spaces of all possible worlds in which these things might occur.

Or you can think of them as sample spaces.

But in any event, you associate with the probability of  $a$  the size of this area here relative to the total area in the rectangle-- the universe.

So the probability of  $a$  is the size of this circle divided by the size of this rectangle in this picture.

So now all these axioms make sense.

The probability that  $a$  is certain is just when that fills up the whole thing, and there's no other place for a sample to be, that means it has to be  $a$ .

So that probability goes all the way up to 1.

On the other hand, if the size of  $a$  is just an infinitesimal dot, then the chances of landing in that world is 0.

That's the bound on the other end.

So this-- axiom number one-- makes sense in terms of that picture over there.

Likewise, axiom number two.

What about axiom number three?

Does that make sense in terms of all this stuff?

And the answer is, sure, because we can just look at those areas with a little bit of colored chalk.

And so the probability of a is just this area here.

The probability of b is this area here.

And if we want to know the probability that we're in either a or b, then we just have to add up those areas.

But when we add up those areas, this intersection part is added in twice.

So we've got to subtract that off in order to make this thing make a rational equation, so that makes sense.

And axiom three makes sense, just as axioms one and two did.

So that's all there is to basic probability.

And now you could do all sorts of algebra on that, and it's elegant, because it's like circuit theory or electromagnetism, because from a very small number of axioms-- in this case three-- you can build an elegant mathematical system.

And that's what probability subjects do.

But we're not going to go there, because we're sort of focused on getting down to a point where we can deal with that joint probability table that we currently can't deal with.

So we're not going to go into a whole lot of algebra with these things.

Just what we need in order to go through that network.

So the next thing we need to deal with is conditional probability.

And whereas those are axioms, this is a definition.

We say that the probability of a given b is equal to, by definition, the probability of a and b.

I'm using that common notation to mean [INAUDIBLE] as is conventional in the field.

And then we're going to divide that by the probability of B.

You can take that as a definition, and then it's just a little bit of mysterious algebra.

Or you could do like we did up there and take an intuitionist approach and ask what that stuff means in terms of a circle diagram and some sort of space.

And let's see, what does that mean?

It means that we're trying to restrict the probability of a to those circumstances where b is known to be so.

And we're going to say that-- we've got this part here, and then we've got the intersection of a with b.

And so it does make sense as a definition, because it says that if you've got b, then the probability that you're going to get a is the size of that intersection-- the pink and orange stuff-- divided by the whole of b.

So it's as if we restricted the universe of consideration to just that part of the original universe as covered by b.

So that makes sense as a definition.

And we can rewrite that, of course, as  $P(a \text{ and } b)$  is equal to the probability of a given b times the probability of b.

That's all basic stuff.

Now, we do want to do a little bit of algebra here, because I want to consider not just two cases, but what if we divide this space up into three parts?

Then we'll say that the probability of a, b, and c is equal to what?

Well, there are lots of ways to think about that.

But one way to think about it is that we are restricting the universe to that part of the world where b and c are both true.

So let's say that y is equal to b and c-- the intersection of b and c, where a and b are both true.

Then we can use this formula over here to say that probability of a, b, and c is equal to the probability of a and y, which is equal to the probability of a given y times the probability of y.

And then we can expand that back out and say that  $P(a \text{ given } b \text{ and } c)$  is equal to the probability-- sorry, times the probability of y, but y is equal to the probability of b and c, like so.

Ah, but wait-- we can run this idea over that one, too, and we can say that this whole works is equal to the probability of a given b and c times the probability of b given c times the probability of c.

And now, when we stand back and let that sing to us, we can see that some magic is beginning to happen here, because we've taken this probability of all things being so, and we've broken up into a product of three probabilities.

The first two are conditional probabilities, so they're really all conditional probabilities.

The last one's conditional on nothing.

But look what happens as we go from left to right.

a is dependent on two things.

b is only dependent on one thing and nothing to the left.

c is dependent on nothing and nothing to the left.

So you can sense a generalization coming.

So let's write it down.

So let's go from here over to here and say that the probability of a whole bunch of things--  $x_1$  through  $x_{10}$ -- is equal to some product of probabilities.

We'll let the index  $i$  run from  $n$  to 1.

Probability of  $x$  to the last one in the series, conditioned on all the other ones-- sorry, that's probability of  $i$ ,  $i$  minus 1 down to  $x_1$  like so.

And for the first one in this product,  $i$  will be equal to  $n$ .

For the second one,  $i$  will be equal to  $n$  minus 1.

But you'll notice that as I go from  $n$  toward 1, these conditionals get smaller-- the number of things on condition get smaller, and none of these things are on the left.

They're only stuff that I have on the right.

So what I mean to say is all of these things have an index that's smaller than this index.

None of the ones that have a higher index are appearing in that conditional.

So it's a way of taking a probability of the end of a whole bunch of things and writing it as a product of conditional probabilities.

So we're making good progress.

We've done one.

We've done two.

And now we've done three, because this is the chain rule.

And we're about halfway through our diagram, halfway to the point where we can do something fun.

But we still have a couple more concepts to deal with, and the next concept is the concept of conditional probability.

So that's all this stuff up here-- oops.

All this stuff here is the definition of conditional probability.

And now I want to go to the definition of independence.

So that's another definitional deal.

But it's another definitional deal that makes some sense with a diagram as well.

So the definition goes like this.

We say that  $P$  of  $a$  given  $b$  is equal to  $P$  of  $a$  if  $a$  is independent of  $b$ .

So that says that the probability of  $a$  doesn't depend on what's going on with  $b$ .

It's the same either way.

So it's independent.

$b$  doesn't matter.

So what does that look like if we try to do an intuitionist diagram?

Well, let's see.

Here's a.

Here's b.

Now, the probability of a given b-- well, let's see.

That must be this part here divided by this part here.

So the ratio of those areas is the probability of a given b.

So that's the probability of this way divided by the probability of both ways.

So what's the probability of a in terms of these areas?

Well, probability of a in terms of these areas is the probability-- let's see, have I got this right?

I've got this upside down.

The probability of a given b is the probability of the stuff in the intersection-- so that's both ways-- divided by the probability of the stuff in b, which is going this way.

And let's see, the probability of a not conditioned on anything except being in this universe is all these hash marks, like so, divided by the universe.

So when we say that something's independent, it means that those two ratios are the same.

That's all it means in the intuitionist's point of view.

So it says that this little area here divided by this whole area is the same as this whole area for a divided by the size of the universe.

So that's what independence means.

Now, that's quite a lot of work.

But we're not done with independence, because we've got conditional independence to deal with.

And that, too, can be viewed as a definition.

And what we're going to say is that the probability of a given b and z is equal to the probability of a given z.

What's that mean?

That means that if you know that we're dealing with  $z$ , then the probability of  $a$  doesn't depend on  $b$ .

$b$  doesn't matter anymore once you're restricted to being in  $z$ .

So you can look at that this way.

Here's  $a$ , and here's  $b$ , and here is  $z$ .

So what we're saying is that we're restricting the world to being in this part of the universe where  $z$  is.

So the probability of  $a$  given  $b$  and  $z$  is this piece in here.

$a$  given  $b$  and  $z$  is that part there.

And the probability of  $a$  given  $z$  is this part here divided by all of  $z$ .

So we're saying that the ratio of this little piece here to this part, which I'll mark that way, ratio of this to this is the same as the ratio of that to that.

So that's conditional independence.

So you can infer from these things, with a little bit of algebra, that  $P$  of  $a$  and  $b$  given  $z$  is equal to  $P$  of  $a$  given  $z$  times  $P$  of  $b$  in  $z$ .

Boy, that's been quite a journey, but we got all the way through one, two, three, four, and five.

And now the next thing is belief nets, and I'm going to ask you to forget everything I've said for a minute or two.

And we'll come back to it.

I want to talk about the dog and the burglar and the raccoon again.

And now, forgetting about probability, I can say, look, the dog barks if a raccoon shows up.

The dog barks if a burglar shows up.

A burglar doesn't show up because the dog is barking.

A raccoon doesn't show up because the dog is barking.

So the causality flows from the burglar and the raccoon to the barking.

So we can make a diagram of that.

And our diagram will look like this.

Here is the burglar, and here is the raccoon.

And these have causal relations to the dog barking.

So that's an interesting idea, because now I can say that-- well, I can't say anything yet, because I want to add a little more complexity to it.

I'm going to add two more variables.

You might call the police, depending on how vigorous the dog is barking, I guess.

And the raccoon has a propensity to knocking over the trash can.

So now, I've got five variables.

How big a joint probability table am I going to need to keep my tallies straight?

Well, it'll be 2 to the 5th.

That's 32.

But what I'm going to say is that this diagram is a statement, that every node in it depends on its parents and nothing else that's not a descendant.

Now, I need to say that about 50 times, because you've got to say it right.

Every node there is independent of every non-descendant other than its parents.

No, that's not quite right.

Given its parents, every node is independent of all other non-descendants.

Well, what does that mean?

Here's the deal with calling the police.

Here's its one and only parent.

So given this parent, the probability that they were going to call the police doesn't depend on anything like B, R, or T. It's because all of the causality is flowing through this dog barking.

I'm not going to call the police in a way that's dependent on anything else other than whether the dog is barking or not.

Because this guy has this as a parent, and these are not descendants of calling the police, so this is independent of B, R, and T.

So let's go walk through the others.

Here's the dog.

The dog's parents are burger appearing and raccoon appearing.

So the probability that the dog appears is independent of that trash can over there, because that's not a descendant.

It is dependent on these parents.

How about the trash can?

It depends only on the raccoon.

It doesn't depend on any other non-descendant, so therefore, it doesn't depend on D, B, or P. How about B?

It has no parents.

So it depends on nothing else, because everything else is either a non-descendant, because B does not depend on R and T, because they're not descendants.

It's interesting that B might depend on D and P, because those are descendants.

So it's important to understand that there's this business of independence given the parents of all other non-descendants.

And you'll see why that funny, strange language is important in a minute.

But now, let's see-- I want to make a model of what's going to happen here.

So let me see what kind of probabilities I'm going to have to figure out.

This guy doesn't depend on anything upstream.

So we could just say that all we need there is the probability that a burglar is going to appear.

Let's say it's a fairly high-crime neighborhood-- 1 chance in 10-- 1 day in 10, a burglar appears.

The raccoon doesn't depend on anything other than its own propensity, so its probability, we'll say, is 0.5.

Raccoons love the place, so it shows up about 1 day in 2.

So what about the dog barking?

That depends on whether there's a burglar, and the other parent is whether there's a raccoon.

So we need to keep track of the probability that the dog will bark for all four combinations.

So this will be the burglar, and this will be the raccoon.

This will be false, false, true, true-- oops-- false, false, true, false, false, true, true, true.

So let's say it's a wonderful dog, and it always barks if there's a burglar.

So that would say that the probability here is 1.0, and the probability here is 1.0.

And if there's neither a burglar nor a raccoon, the dog still likes to bark just for fun.

So we'll say that's a chance of 1 in 10.

And then in case there's a burglar, let's say this.

There's no burglar, but there is a raccoon-- he's tired of the raccoons, so he only barks half the time.

Do these numbers, by the way, have to add up to 1?

They clearly don't.

These numbers don't add up to one.

What adds up to 1 is this is the probability that the dog barks.

And then the other phantom probability is out here.

And these have to add up to 1.

So that would be 0.9, that would be 0.0, that would be 0.5, and this would be 0.0.

So because those are just 1 minus the numbers in these columns, I don't bother to write them down.

Well, we still have a couple more things to do.

The probability that we'll call the police depends only on the dog.

So we'll have a column for the dog, and then we'll have a probability of calling the police.

There's a probability for that being false and a probability for that being true.

So if the dog doesn't bark, there's really hardly any chance we'll call the police.

So make that 0, 0, 1.

If the dog is barking, if he barks vigorously enough, maybe 1 chance in 10.

Here, we have the trash can-- the final thing we have to think about.

There's the trash can; rather, the raccoon.

And here's the trash can probability.

Depends on the raccoon being either present or not present.

If the raccoon is not present, the probability the trash can is knocked over by, say, the wind is 1 in 1,000.

If the raccoon is there, oh man, that guy always likes to go in there, so that's 0.8.

So now I'm done specifying this model.

And the question is, how many numbers did I have to specify?

Well, let's see.

I have to specify that one, that one, that one, that one, that one, that one-- that's 6, 7, 8, 9, 10.

So I had to specify 10 numbers.

If I just try to build myself a joint probability table straightaway, how many numbers would I have to supply?

Well, it's 2 to the n.

So it's 2 to the 5th, that's 32.

Considerable saving.

By the way, how do you suppose I made that table?

Not by doing all those numbers.

By making this belief network and then using the belief network to calculate those numbers.

And that's why this is a miracle, because with these numbers, I can calculate those numbers instead of making them up or making a whole lot of tally-type measurements.

So I'd like to make sure that that's true.

And I can use this stuff here to calculate the full joint probability table.

So here's how this works.

I have the probability of some combination-- let's say the police, the dog, the burglar, the trash can, and the raccoon.

All the combinations that are possible there will give me an entry in the table-- one row.

But let's see-- there's some miracle here.

Oh, this chain rule.

Let's use the chain rule.

We can write that as a probability that we call the police given d, b, t, and r.

And then the next one in my chain is probability of d given b, t, and r.

Then the next one in the chain is the probability of b given t and r.

And the next one in my chain is P of t given r.

And the final one in my chain is p of r.

Now, we have some conditional independence knowledge, too, don't we?

We know that this probability here depends only on  $d$  because there are no descendants.

So therefore, we don't have to think about that, and all the numbers we need here are produced by this table.

How about this one here?

Probability that the dog barks depends only on its parents,  $b$  and  $r$ , so it doesn't depend on  $t$ .

So  $b$ , in turn, depends on-- what does it depend on?

It doesn't depend on anything.

So we can scratch those.

Probability of  $t$  given  $r$ , yeah, there's a probability there, but we can get that from the table.

And finally,  $P$  or  $r$ .

So that's why I went through all that probability junk, because if we arrange things in the expansion of this, from bottom to top, then we arrange things so that none of these guys depends on a descendant in this formula.

And we have a limited number of things that it depends on above it.

So that's the way we can calculate back the full joint probability table.

And that brings us to the end of the discussion today.

But the thing we're going to think about is, how much saving do we really get out of this?

In this particular case, we only had to devise 10 numbers out of 32.

What if we had 10 properties or 100 properties?

How much saving would we get then?

That's what we'll take up next time, after the quiz on Wednesday.