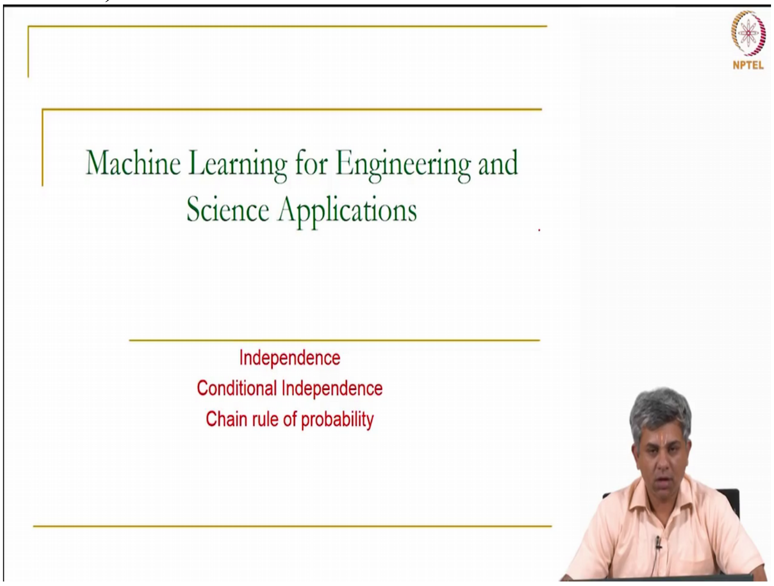
Machine Learning for Engineering and Science Applications
Professor Dr. Balaji Srinivasan
Department of Mechanical Engineering
Indian Institute of Technology Madras
Independence Conditional Independence Chain Rule Of Probability

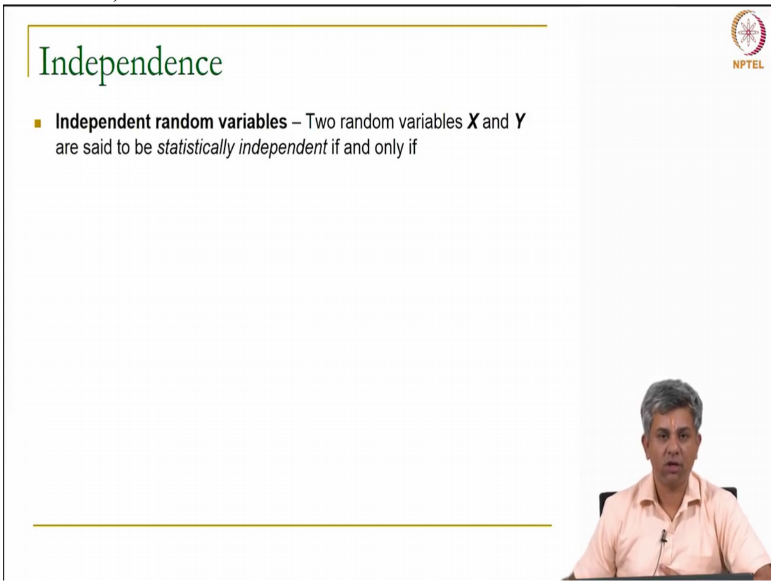
(Refer Slide Time: 00:13)



In this video we will be looking at some further ideas on probability. Specifically, we will be looking at three simple ideas, that of independence which you will be already familiar with from school probability.

We will also be looking at conditional independence which you might or might not be familiar with. And finally we will be looking at chain rule of probability.

(Refer Slide Time: 00:37)

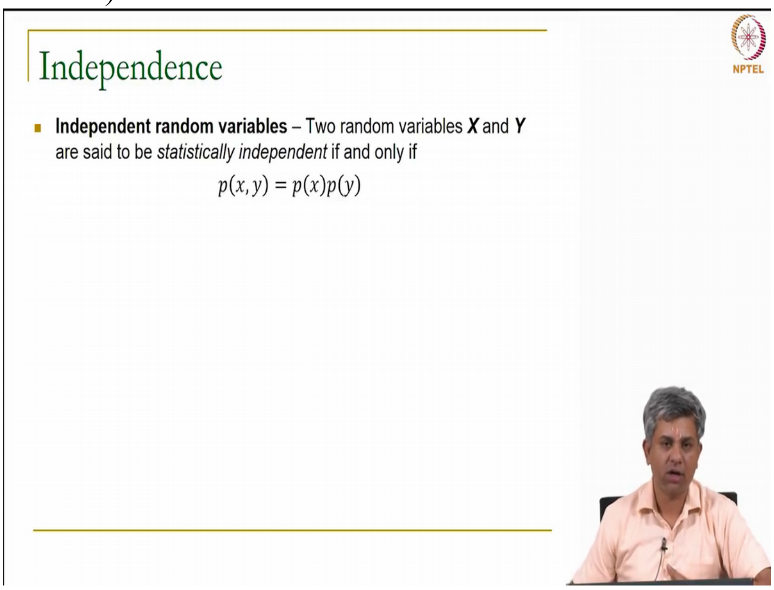


So independence as a notion, all of us know that two random variables are independent means when an event X happens, its, it has no bearing on whether event Y happens or not.

A simple example would be, suppose I toss a coin and it gives heads or tails. This has no bearing on me tossing another coin and what outcome will come out of it.

So such random variables are called independent random variables and the mathematical condition

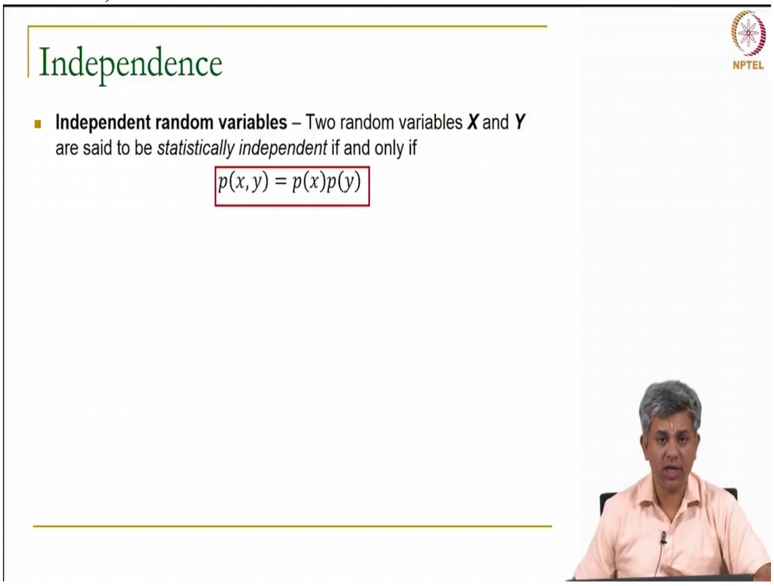
(Refer Slide Time: 01:06)



under which two variables x and y, we would say they are statistically independent if p of x y is equal to p of x multiplied by p of y.

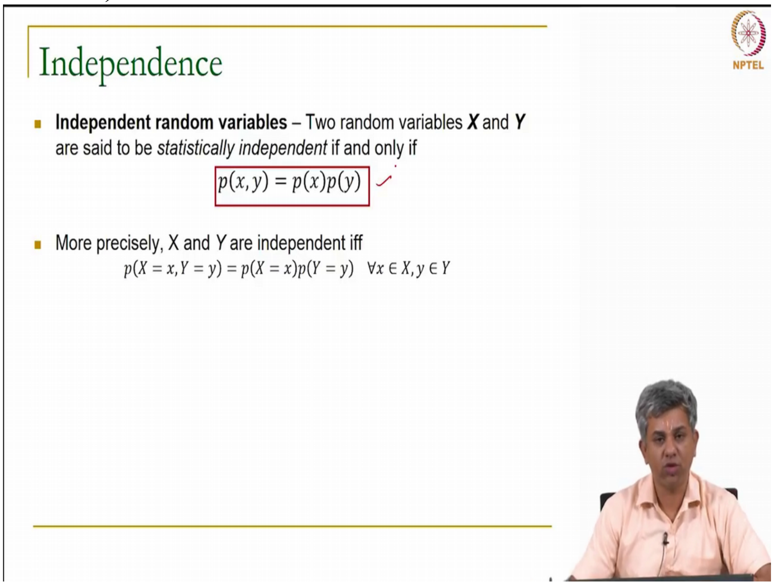
To be a little bit more precise,

(Refer Slide Time: 01:18)



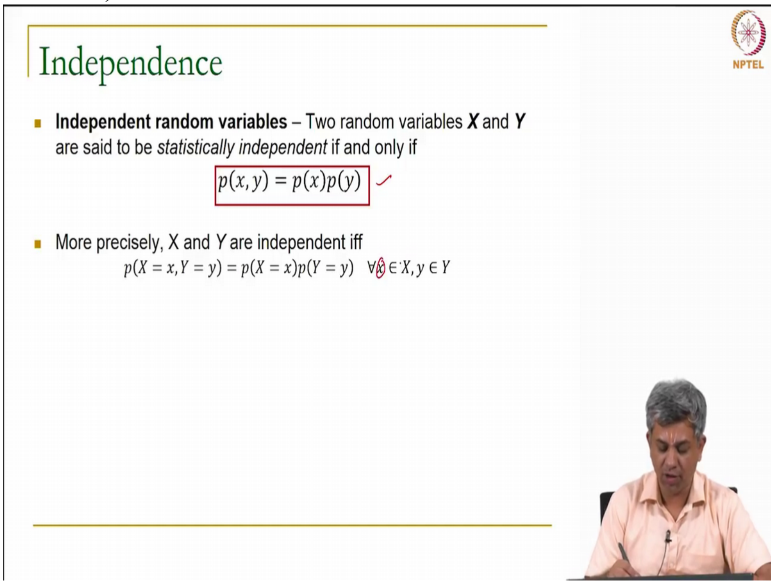
we have to use a little bit more formal notation. Though the calculation is actually this,

(Refer Slide Time: 01:25)



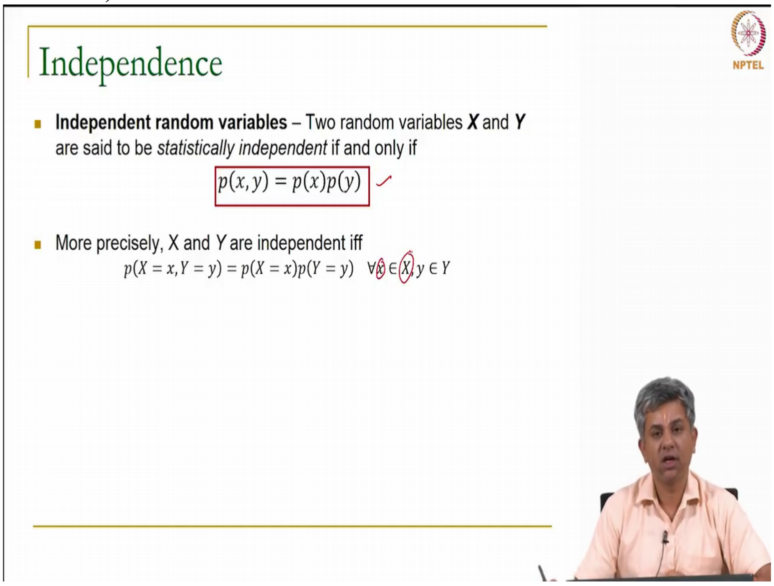
you should say that for every possible x which belongs to

(Refer Slide Time: 01:29)



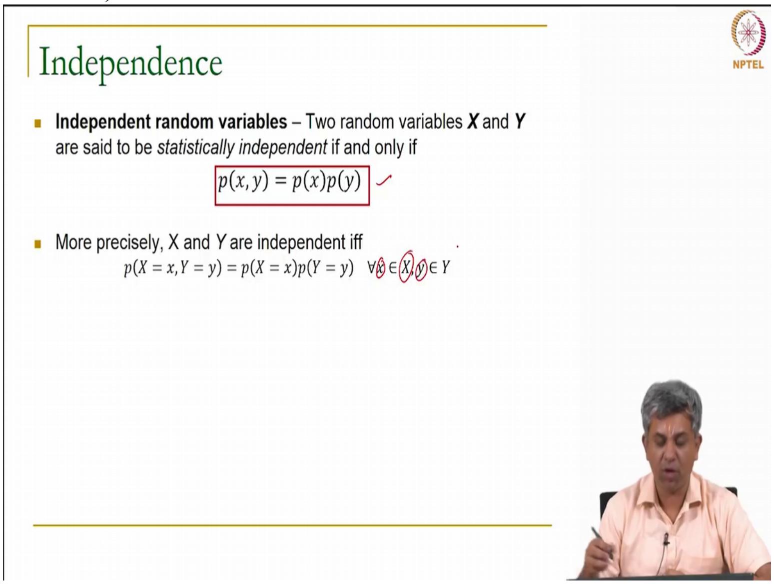
event X

(Refer Slide Time: 01:30)



and every possible value y which belongs to event Y.

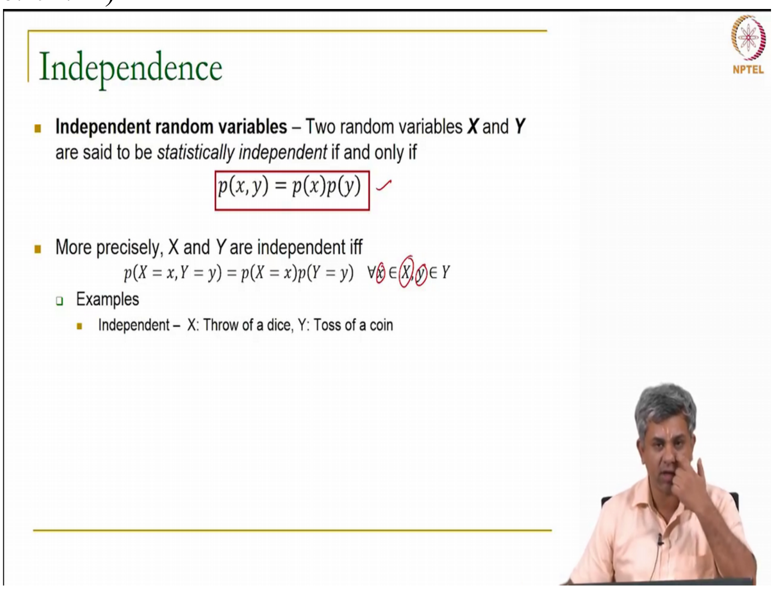
(Refer Slide Time: 01:33)



For example, I am tossing one coin on one hand and another coin on another hand, the event X would be the outcome of the left hand coin which could be either heads or tails. Similarly, the event on the right hand would be either heads or tails.

Now for each of these events, heads heads, heads tails, tails heads and tails tails, all of them should work out such that p of x comma y which is a joint probability of x and y should be the product of the individual probability, or product of the marginal probabilities; p of X equal to x multiplied by p of Y equal to y.

(Refer Slide Time: 02:12)



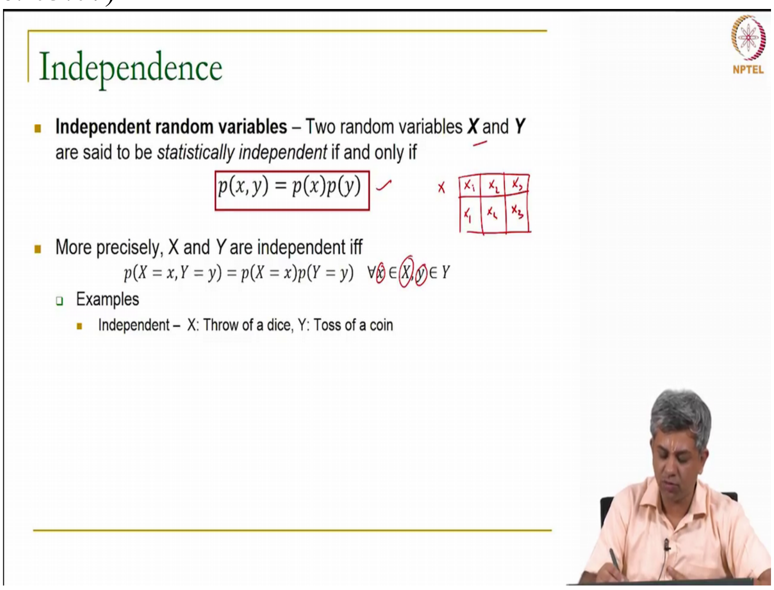
Now we have a simple example, the example I gave or you could say X is the event of throw of a dice, and Y is the event of toss of a coin. So, of course X can take 6 independent values, 1, 2, 3,4,5,6 and Y has two possible values heads and tails.

Now each of these combinations, you know 1 and heads, 2 and heads, 3 and heads etc, etc, the 12 individual probabilities, all of them should obey this law, Ok.

So all those joint probabilities should obey this law and we know that they will because notionally, physically we have an idea obviously that the event X and event Y are independent.

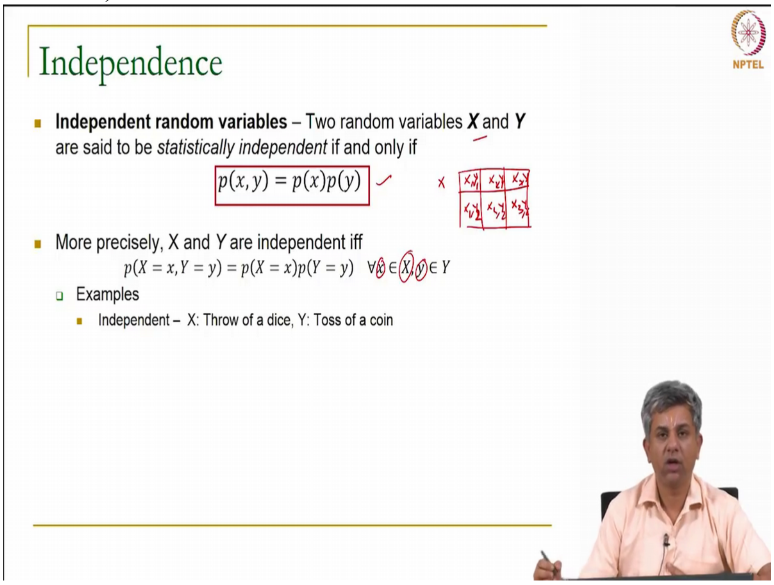
This definition is also a good idea for you to find out whether two events are independent or not. Suppose, you know like we saw the previous video, suppose we have a joint probability table known as say X has values x 1, x 2, x 3 and Y takes values y 1 and y 2.

(Refer Slide Time: 03:07)



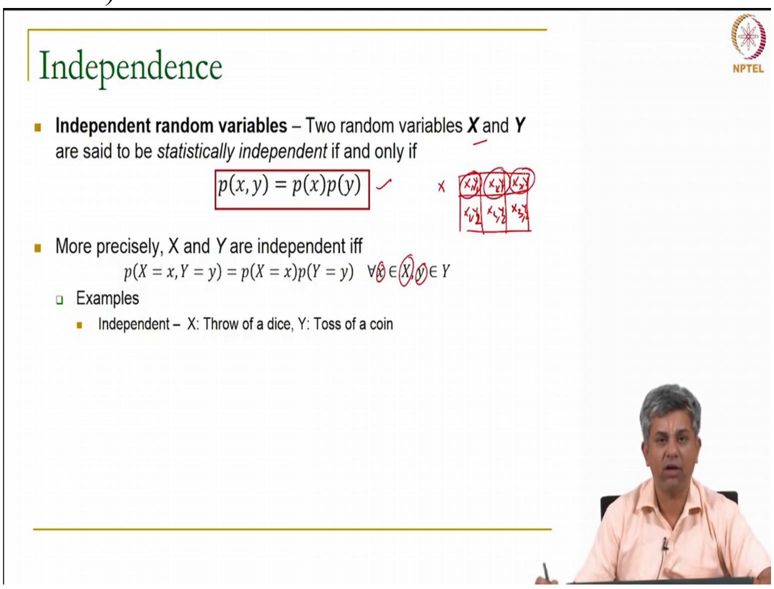
So we have all these joint events y 2, x 2 y 2, x 3 y 2,

(Refer Slide Time: 03:17)



so if you draw this and suppose all you have are all these joint probability

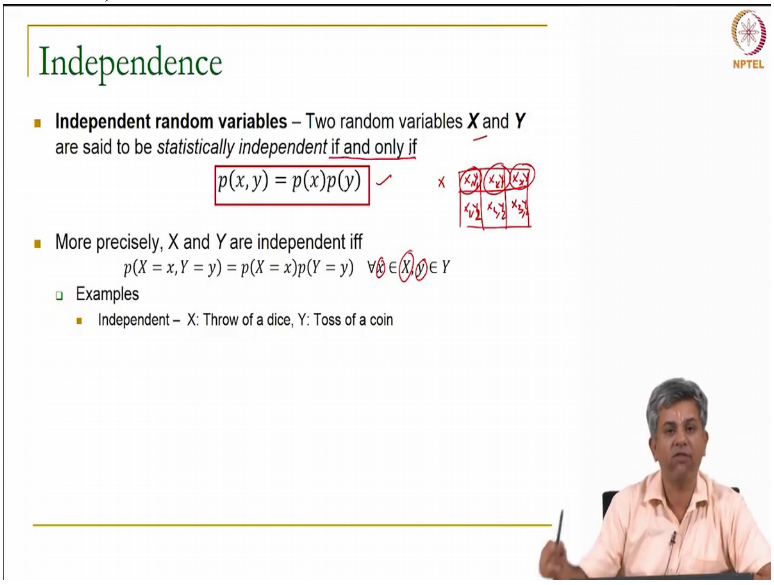
(Refer Slide Time: 03:22)



values. Then you can find out the marginal values of probabilities X and Y by simply adding these up.

And then check whether p of x y is a product of p of x and p of y, and that gives you a check for whether the event is, two events are independent or not. Remember that the condition given is the

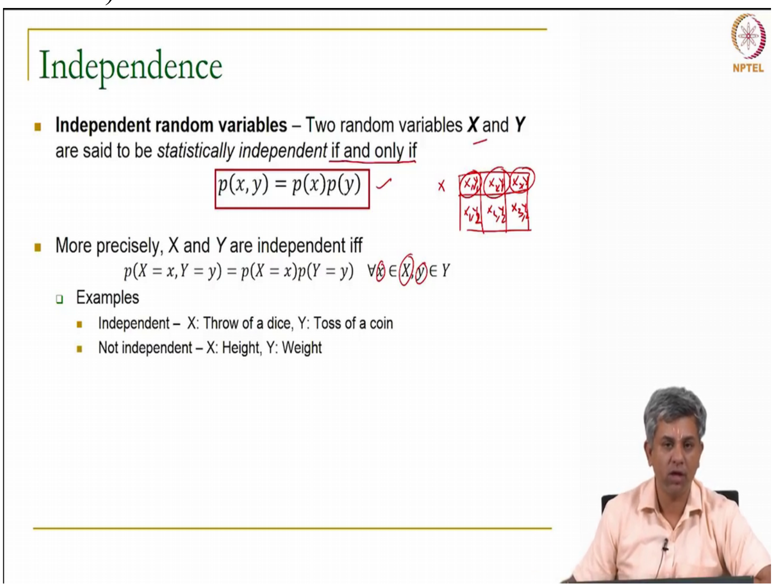
(Refer Slide Time: 03:43)



if and only if condition, Ok.

So if two events are independent then you will get p of x y is equal to p of x multiplied by p of y. And only if this is true do we consider the two events to be

(Refer Slide Time: 03:53)

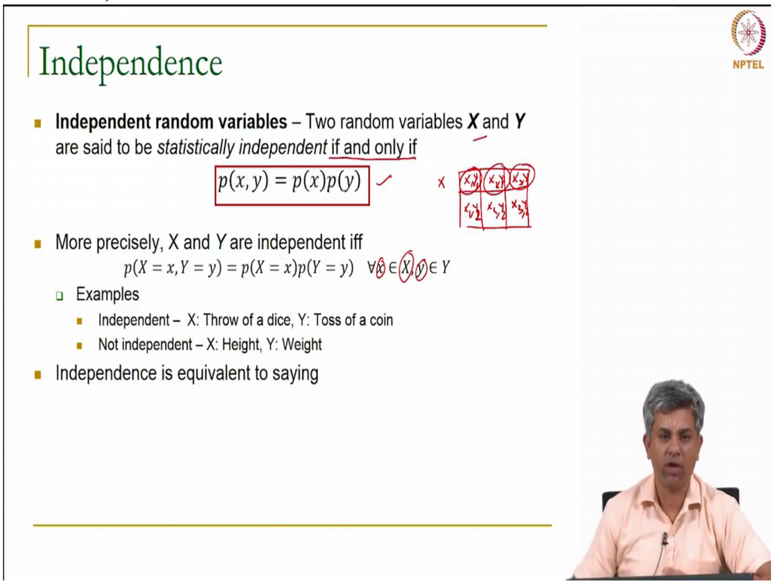


independent.

An example of two measurements which are not independent are height and weight. Or two random variables which are not independent are height and weight. So even though, you know you could have a very tall person who is very, very slim and you could have a short person who is very, very stout, in general as height increases weight will increase.

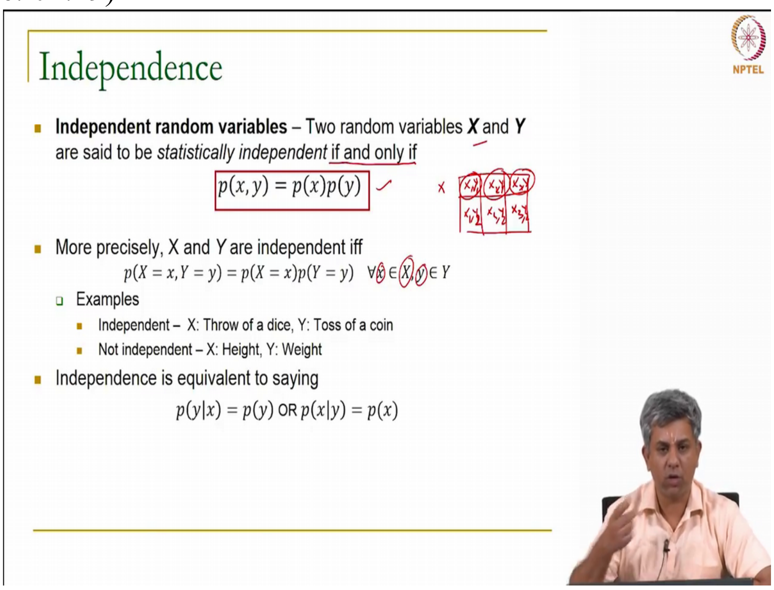
So the random variables x and y are actually correlated random variables. They are not independent random

(Refer Slide Time: 04:24)



variables. This definition of independence is equivalent

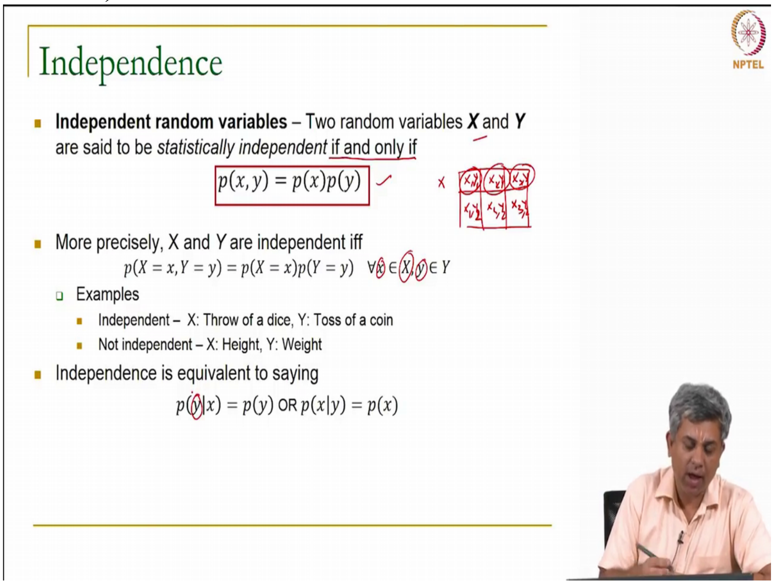
(Refer Slide Time: 04:29)



to giving a conditional probability statement also.

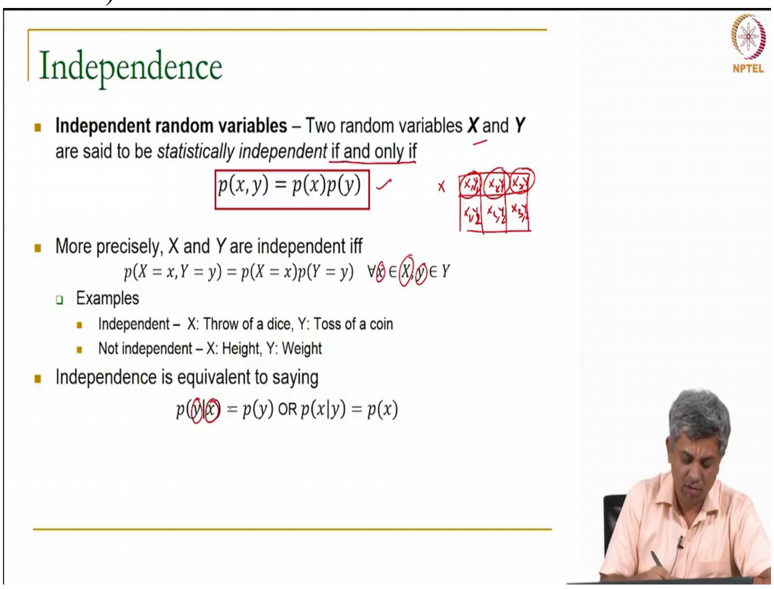
So for example if two events X and Y are independent then I will say that p of y given x is the same as p of y. What does it mean? Physically it means that the dependence of y on x here, you know whether

(Refer Slide Time: 04:50)



y happens or not has no bearing

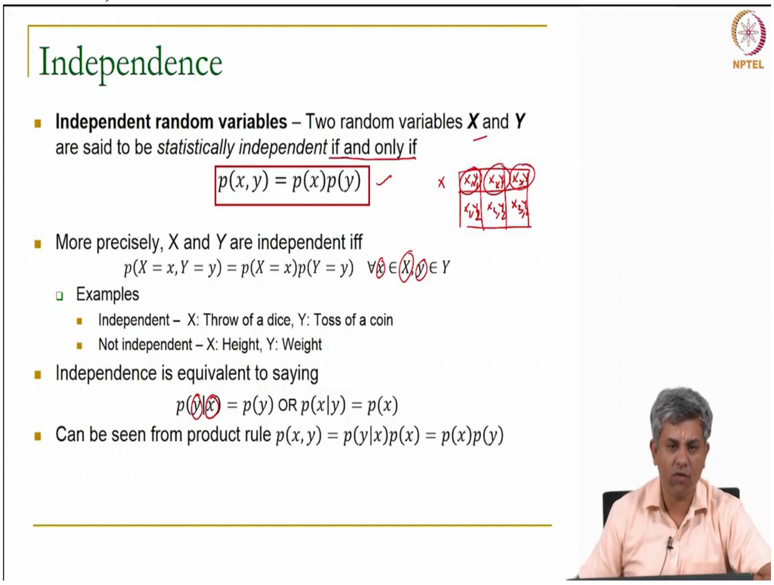
(Refer Slide Time: 04:52)



on whether x happens. So therefore p of y given x has to be the same as p of y.

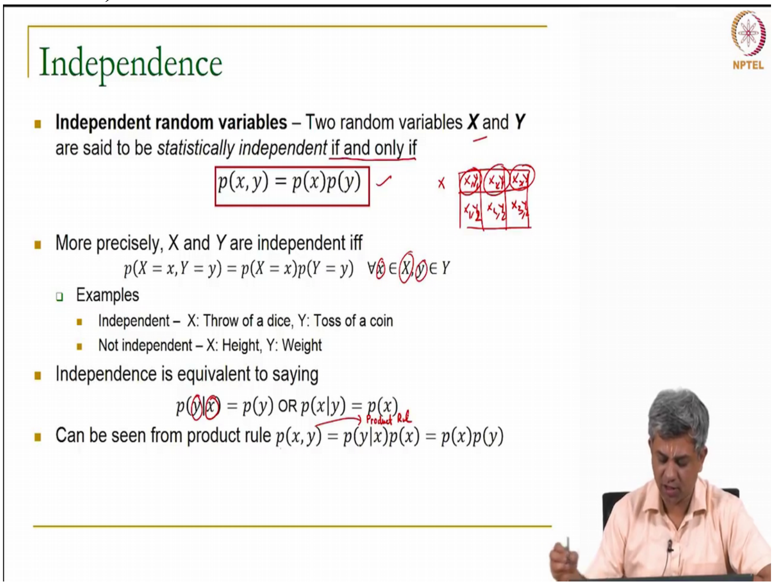
Equivalently you could also say that p of x given y is equal to p of x. You can prove this,

(Refer Slide Time: 05:02)



the equivalent of these two statements very simply, remember that p of x comma y by the product rule is

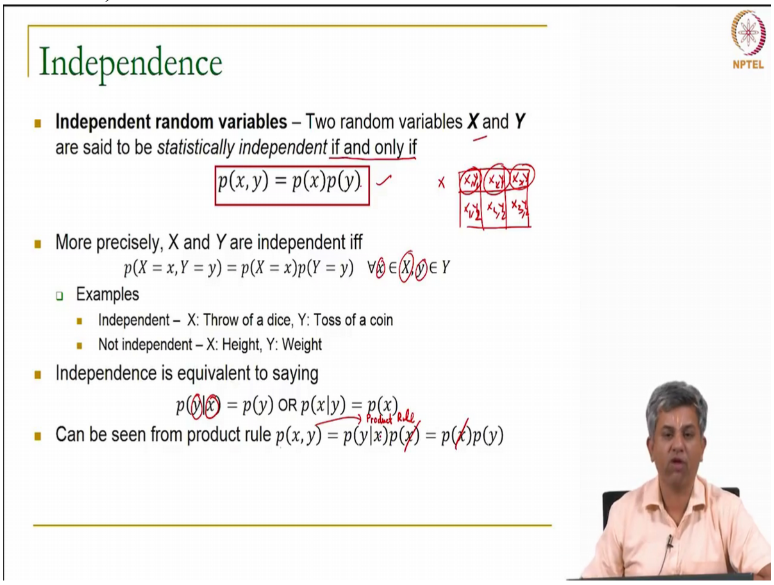
(Refer Slide Time: 05:14)



p of y given x multiplied by p x.

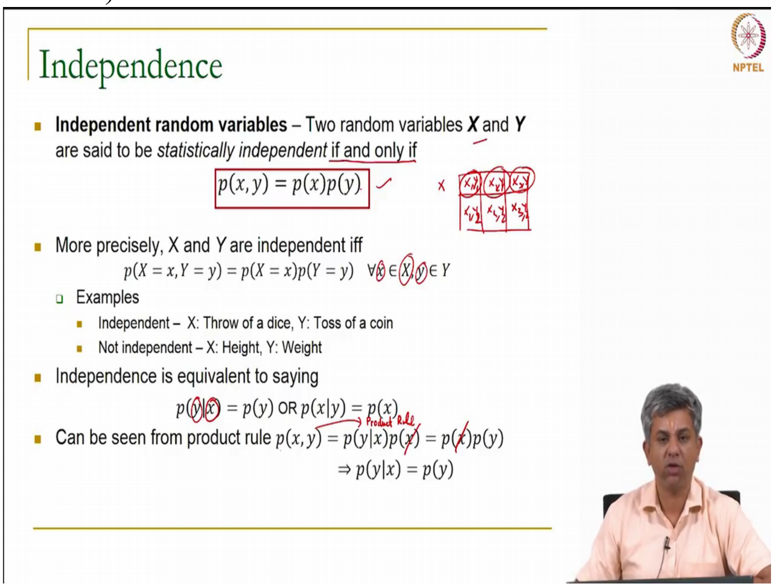
Now suppose we take this definition of independence. So p of x y is equal to p of x multiplied by p of y. And then equate the two, you immediately get

(Refer Slide Time: 05:29)



that p of y given x is equal

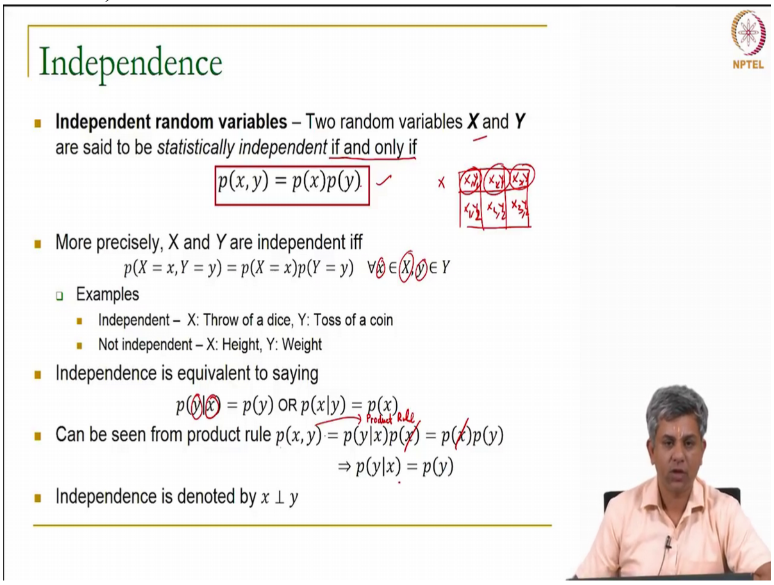
(Refer Slide Time: 05:31)



to p of y. So the two statements at least you can derive this from here.

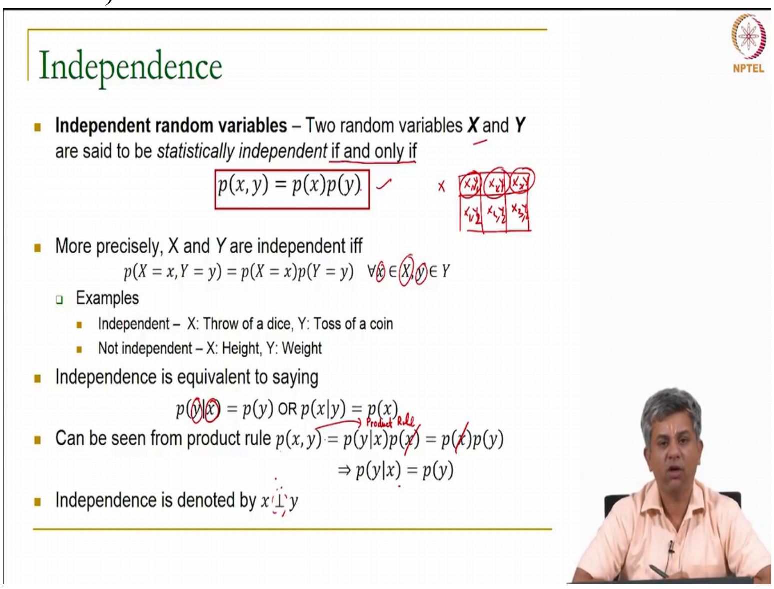
And similarly if you go the reverse direction you can derive the other one from this. So the two statements are actually equivalent.

(Refer Slide Time: 05:42)



Typically, we denote independence. The notation is x and this is a perpendicular

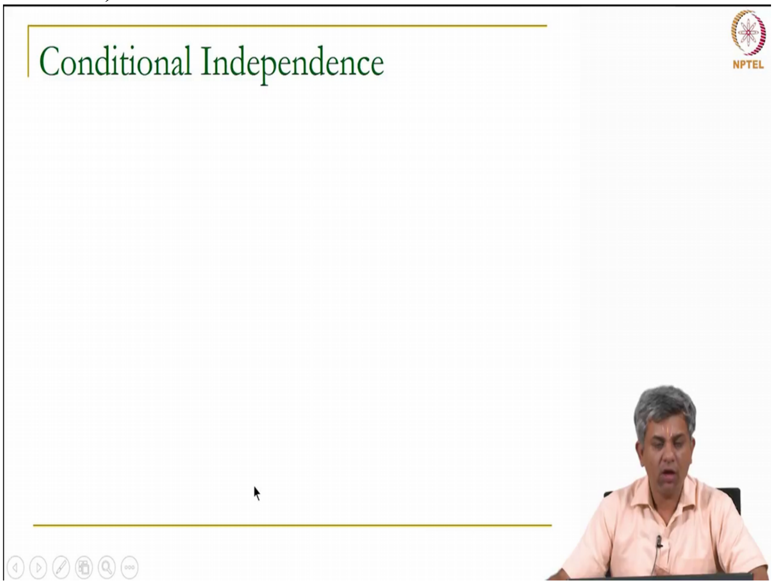
(Refer Slide Time: 05:49)



sign, y because we kind of take orthogonal variables as if they were independent variables, Ok.

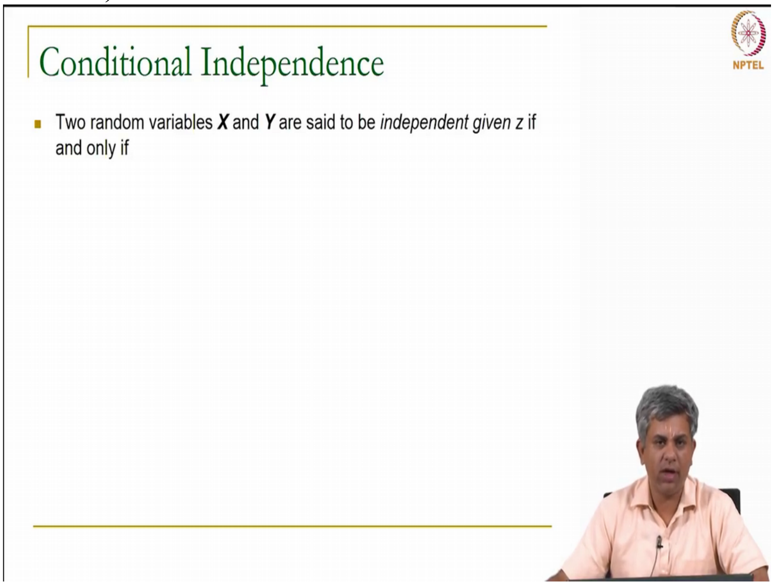
So please remember this notation.

(Refer Slide Time: 05:58)



Now we continue on and look at a slightly more

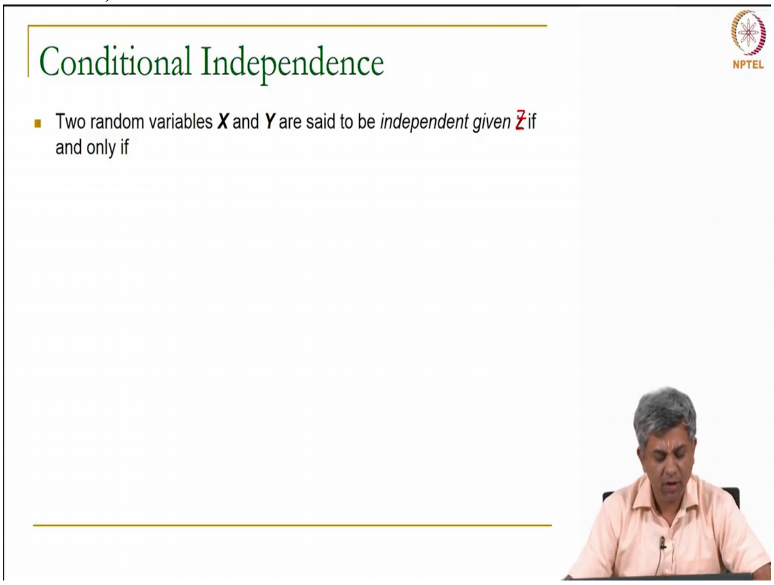
(Refer Slide Time: 06:02)



involved notation that of conditional independence. This is not simple independence. It is conditional independence. The definition is kind of obvious.

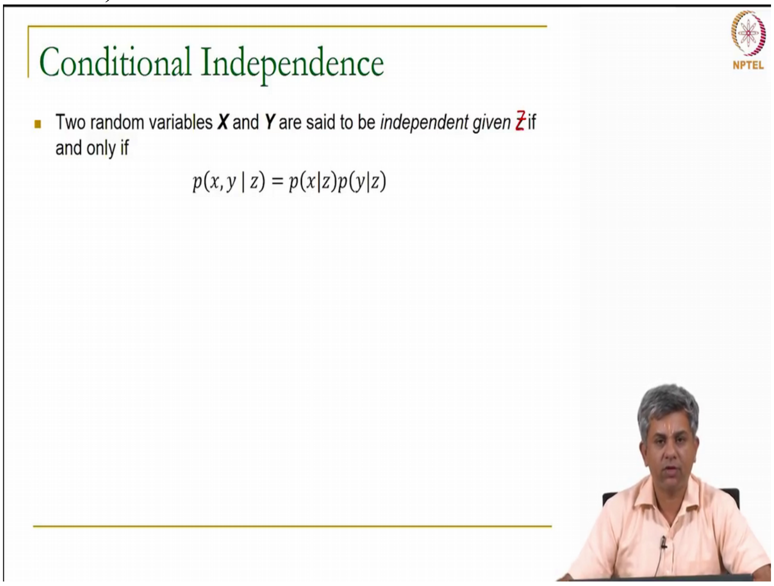
It is the simple extension of the previous definition that we had which is that 2 random variables are, X and Y are, said to be independent given Z, Ok, given a third variable or third event, it should be capital Z

(Refer Slide Time: 06:24)



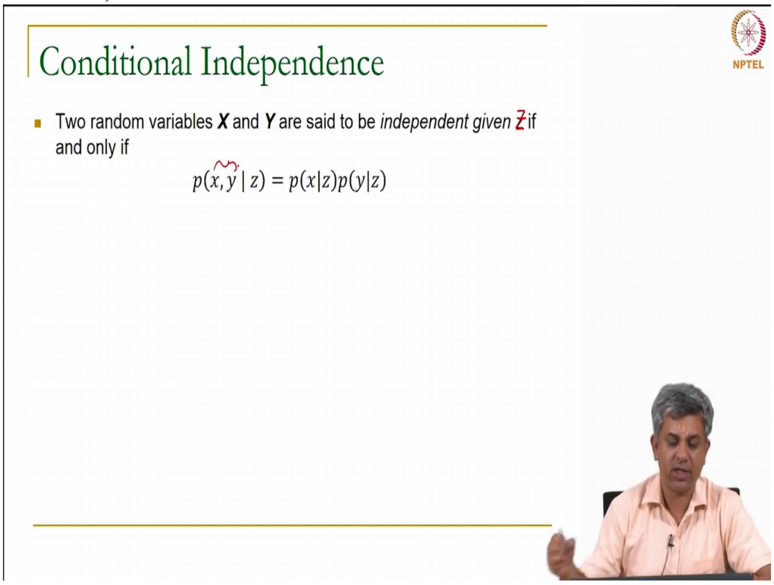
if and only if

(Refer Slide Time: 06:27)



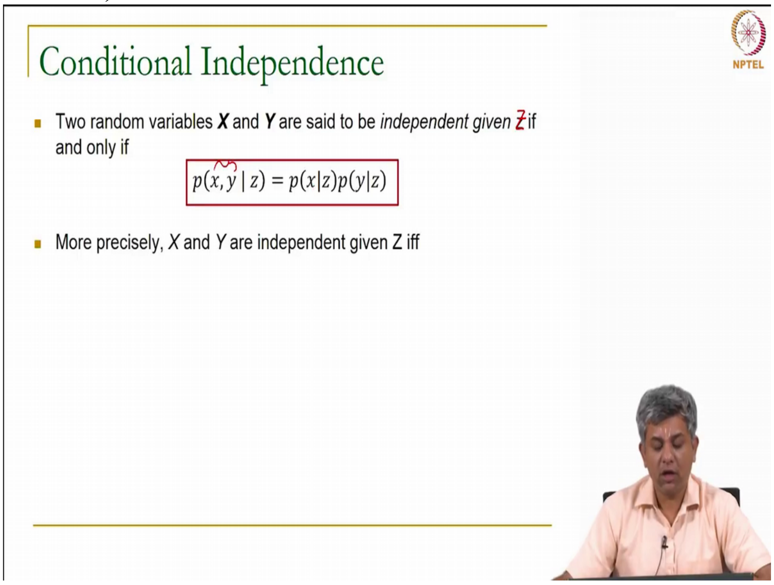
we say, it is a simple extension as I said, p of x and y, joint probability

(Refer Slide Time: 06:33)



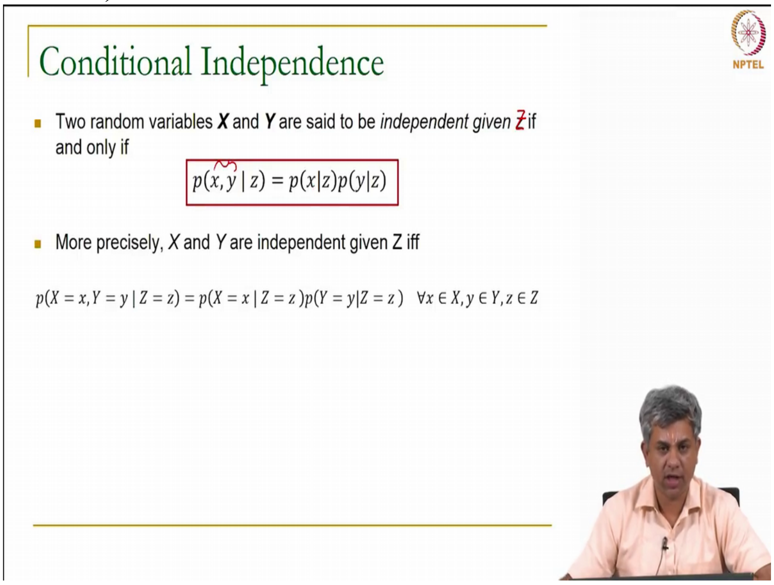
given z is equal to p of x given z multiplied by p of y given z.

(Refer Slide Time: 06:39)



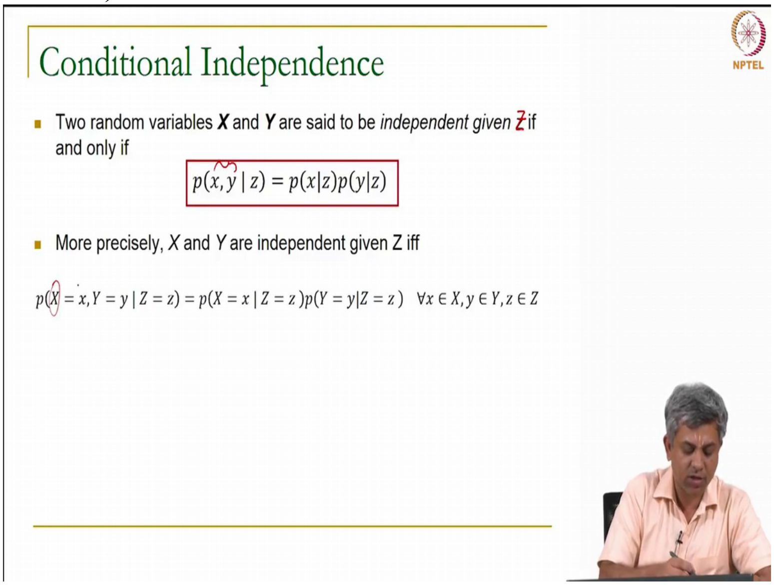
Let us give a more precise definition like we did last time.

(Refer Slide Time: 06:43)



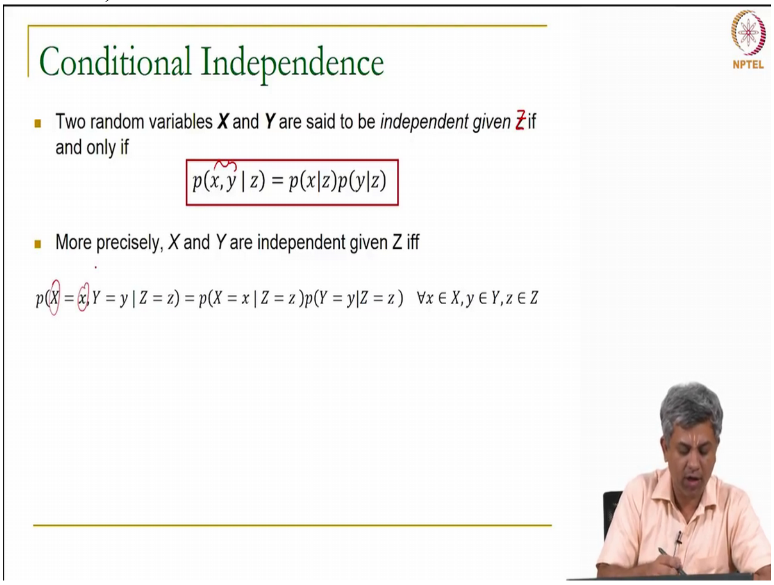
It is simply an extension. Instead of simply saying p of x comma y given z, we say p that the event X

(Refer Slide Time: 06:51)



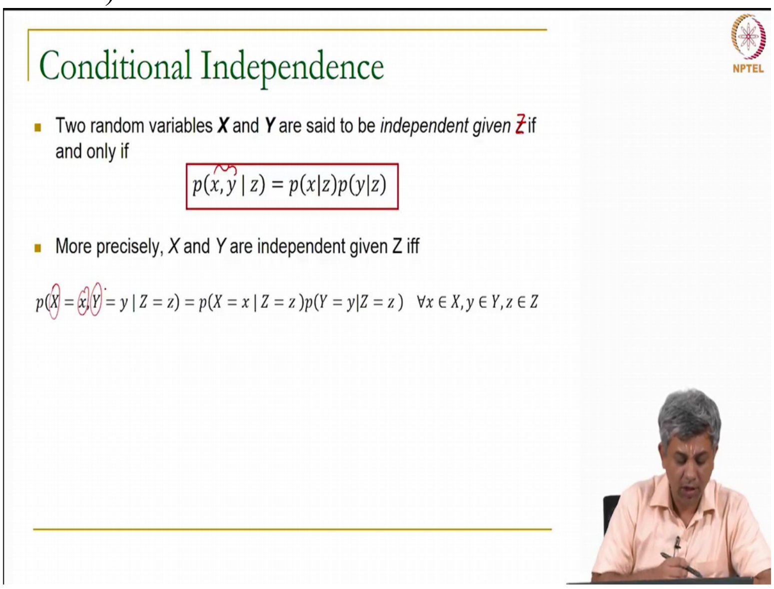
or the random variable X takes the value small x,

(Refer Slide Time: 06:53)



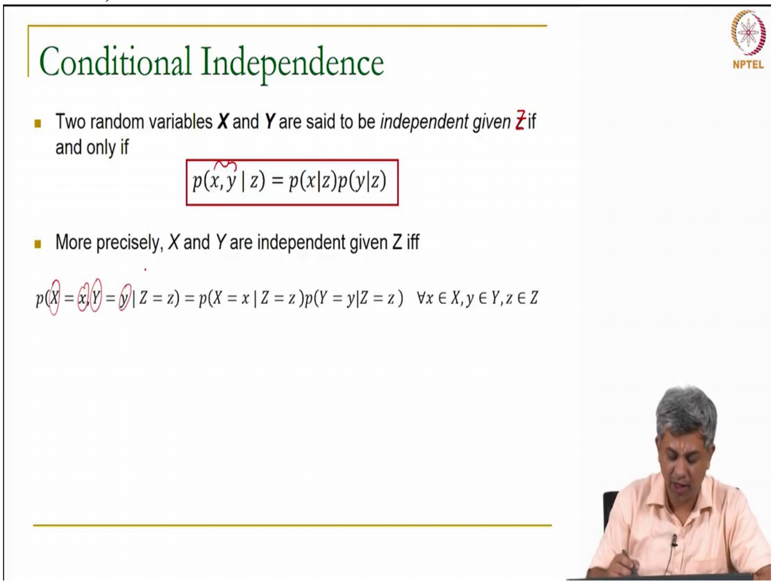
the random variable

(Refer Slide Time: 06:55)



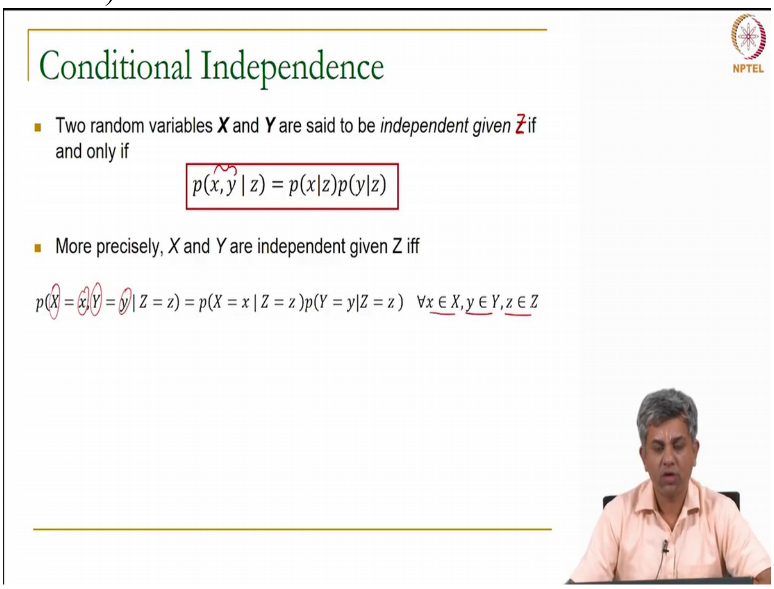
Y takes the value small y

(Refer Slide Time: 06:57)



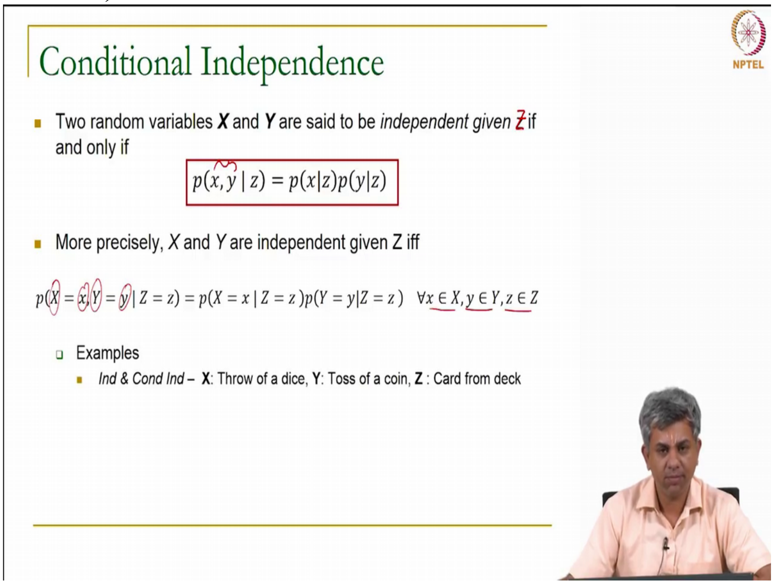
etc, etc is like this for all values of X, Y and Z.

(Refer Slide Time: 07:04)



So let us take a few examples in order to clarify this notion of conditional independence. So here is the first example.

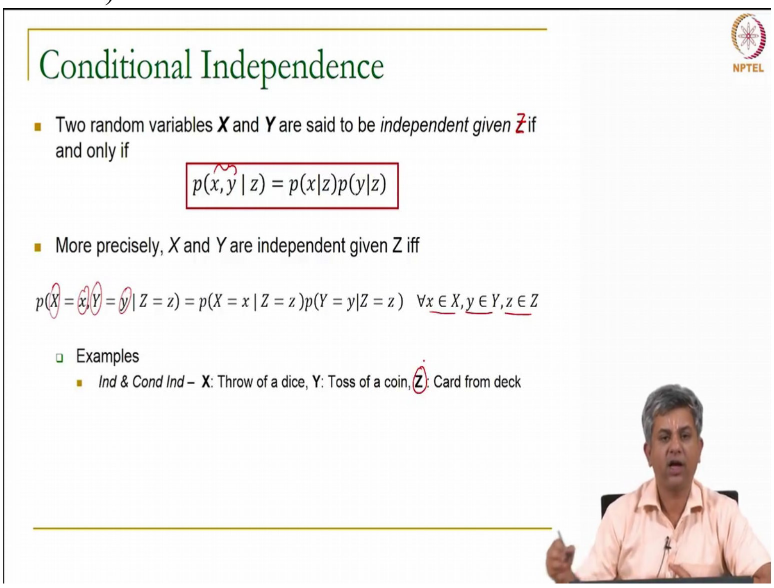
(Refer Slide Time: 07:12)



So let us take the case where X is, you know a dice throw. Y is a toss of a coin and Z is drawing a card from a deck.

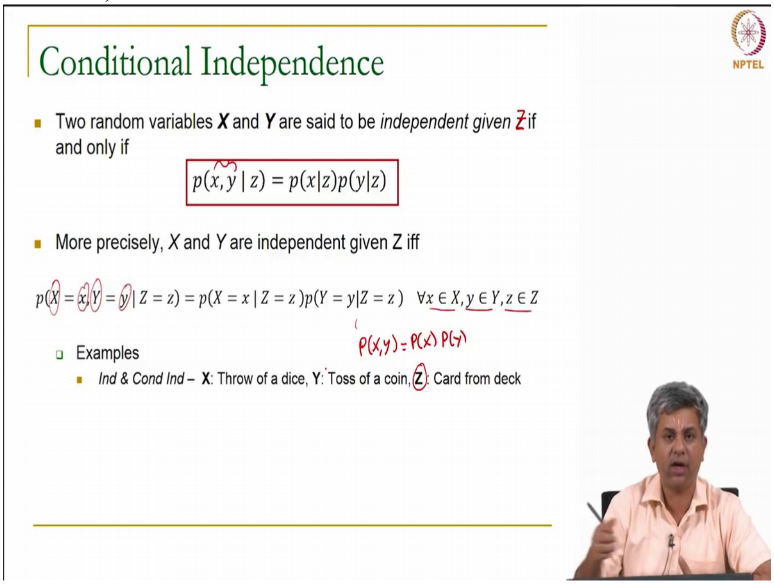
In this case you can see that obviously drawing the card from the deck

(Refer Slide Time: 07:30)



has no bearing on X or Y. Actually Z is, if you take pair wise, Y is independent from Z, X is also independent from Z, Ok. You will also see that automatically p of x comma y is equal to p of x p of y, because x and y are also independent

(Refer Slide Time: 07:48)

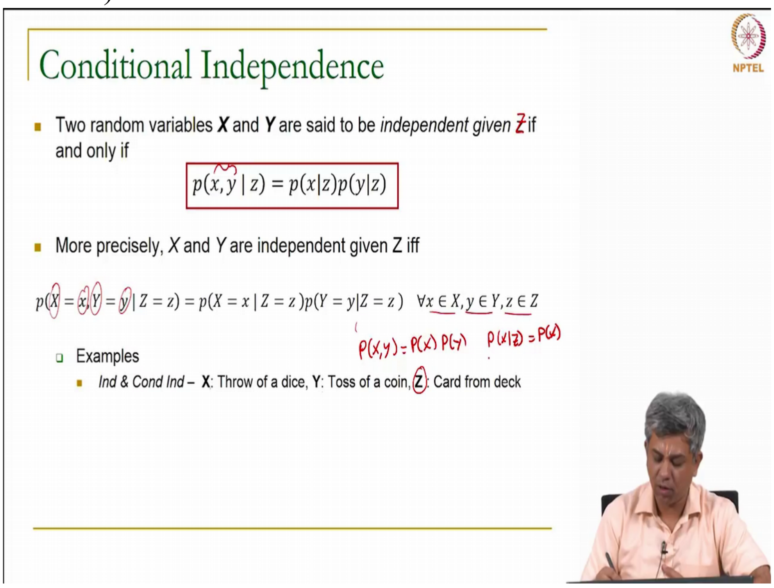


as we saw earlier.

The throw of a dice has nothing to do a toss of a coin. So these events X and Y are of course independent. They are also conditionally independent in the sense that x comma y given z is the same as x given z and y given z.

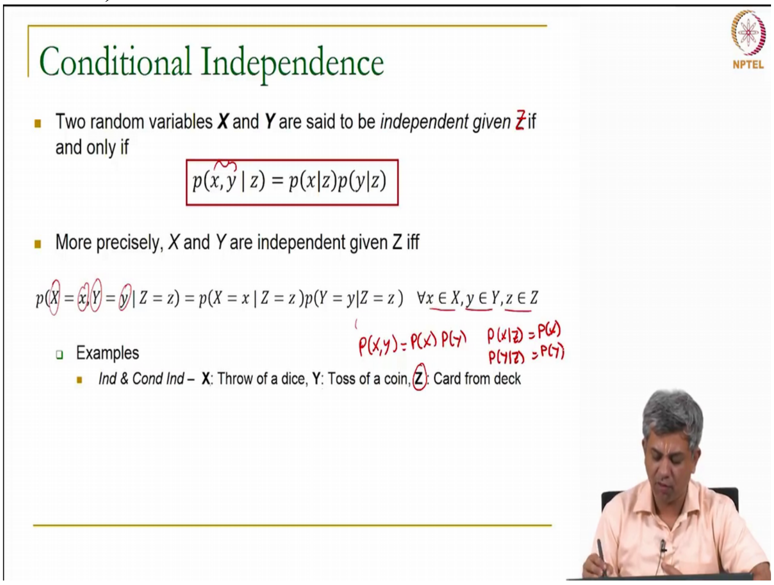
Why? Because since x and z are independent then probability of x given z

(Refer Slide Time: 08:11)



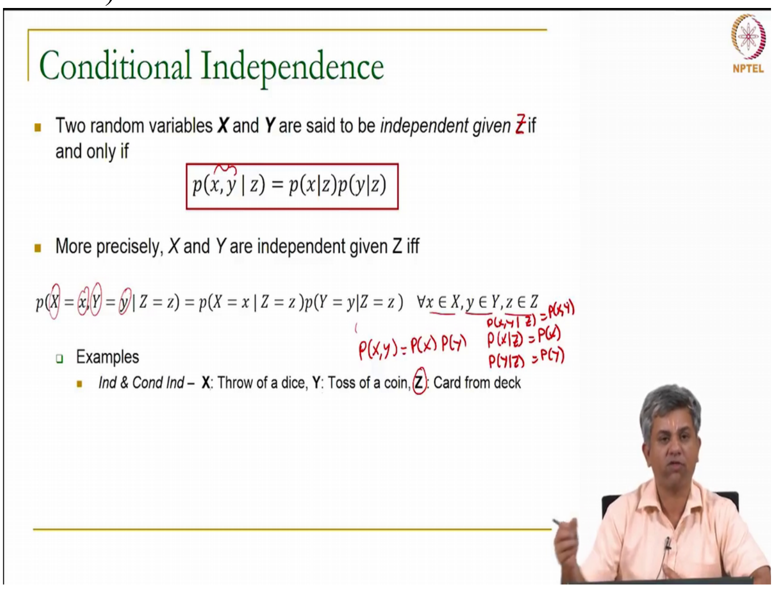
is probability of x. Probability of y given z is

(Refer Slide Time: 08:15)



probability of y. And similarly probability of x comma y given z, probability of x comma y.

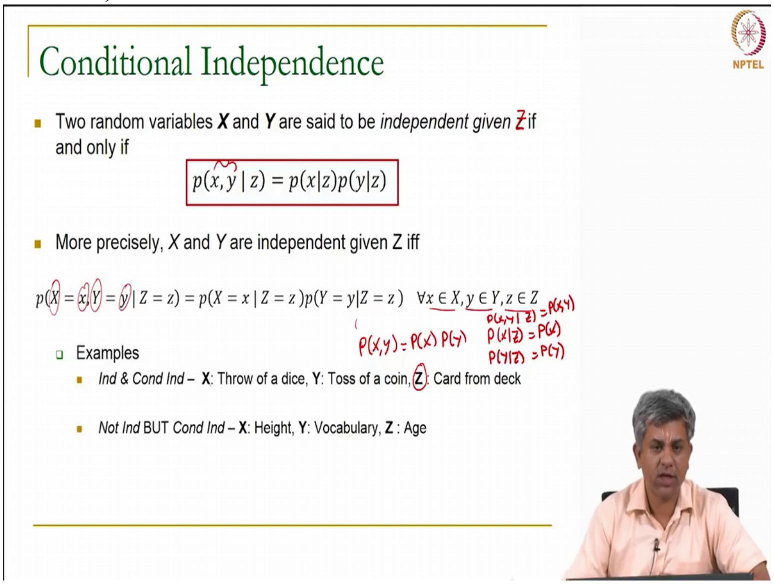
(Refer Slide Time: 08:22)



And the rest of it follows from the fact that x and y are already independent.

Ok so this is an example where x and y are not only independent but they are also conditionally independent.

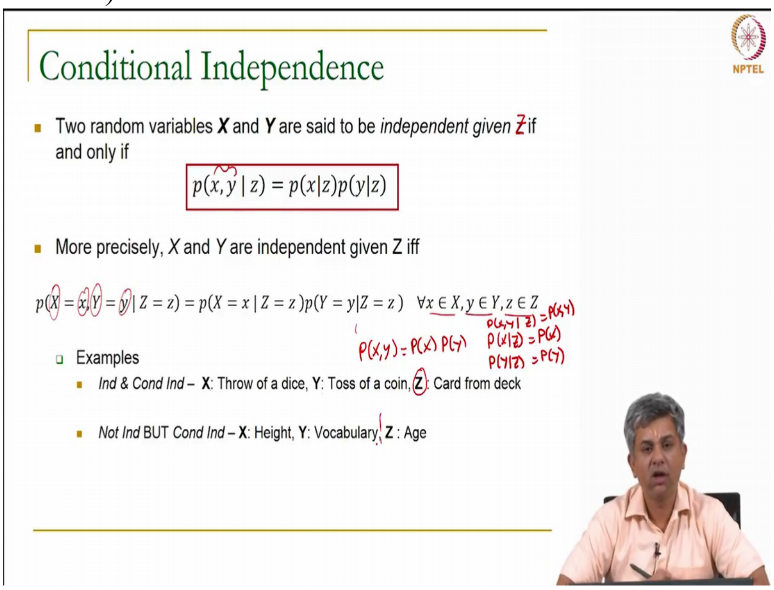
(Refer Slide Time: 08:33)



Let us take the different example. Let us take the example where X denotes height, Y denotes vocabulary of this person of whose height we are measuring, Ok.

In this case let us first ask the question, let us ignore the fact of Z, whatever definition I gave for Z here

(Refer Slide Time: 08:50)

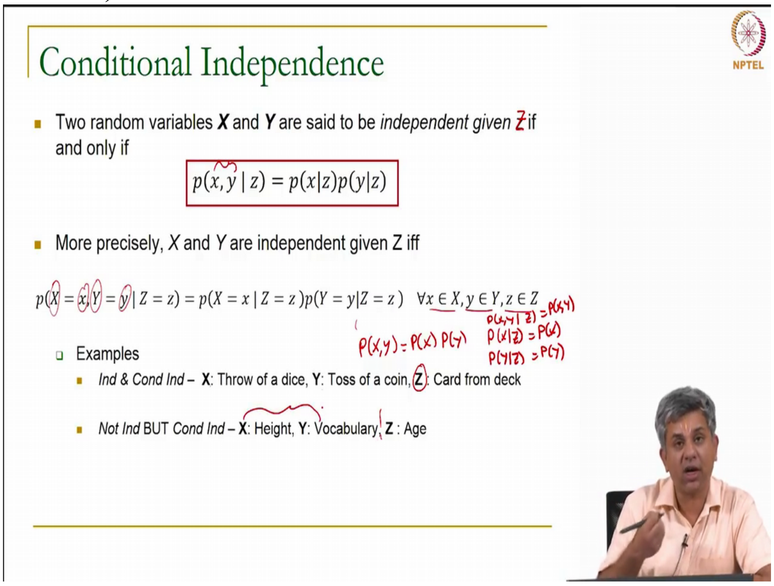


and let us look at his height independent of vocabulary.

Now a priori it looks like it should not matter what a person's height is, you know the vocabulary is independent of the person's height.

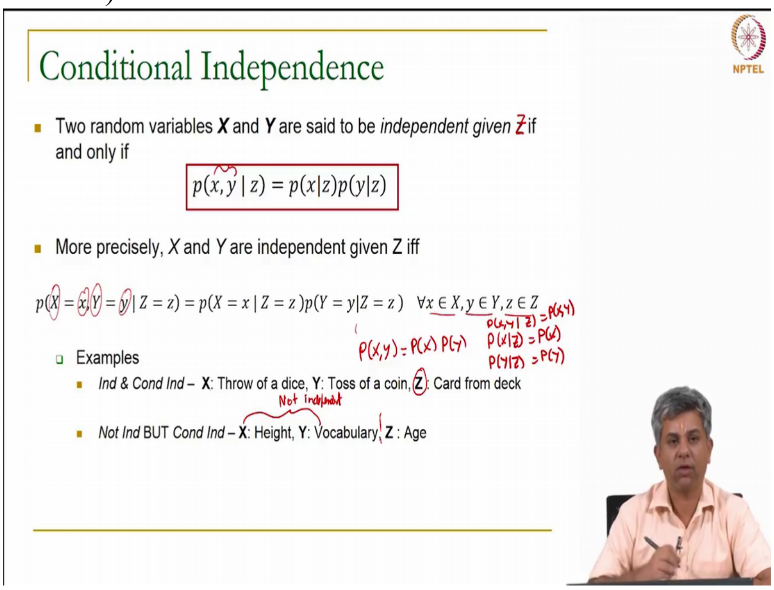
However if I tell you this person is just 2 feet tall, it automatically means, or it is most probable that this person therefore must be a child and therefore must have low vocabulary. So X and Y by themselves, unless

(Refer Slide Time: 09:21)



I gave some further conditions are not independent.

(Refer Slide Time: 09:28)



So please remember this. In this case X and Y are not really independent variables. They are only independent suppose I give a particular condition.

So let us give such a condition. One such condition would be age. So suppose I say that if I look at all people of age 12, would it matter what the person's height is for the vocabulary.

And the answer would be, at least common sense says, and our common observation says that it is not true.

Similarly, if I fix the age at 30, people of the age of 30, regardless of height will have vocabularies that are whatever they are. They do not at least depend on height.

Similarly, if I fix the age at 2, it will not matter what the person's age is, or the baby's age is, you know the height and vocabulary would be independent. It could be a slightly taller child with lesser vocabulary or it could be a shorter child with more vocabulary etc. Ok.

So this is an example of a case where the two variables are not independent but they are conditionally independent, Ok. So if you give a particular condition they actually become independent.

So let us look at a third case where the two variables were originally independent but they actually, after you apply condition, they are no longer independent.

So here is a simple example. I have 2 dice. I throw one dice and the value I denote as X or that is the event X. The second dice throw, that has value as Y, Ok and now these two events as I know are independent.

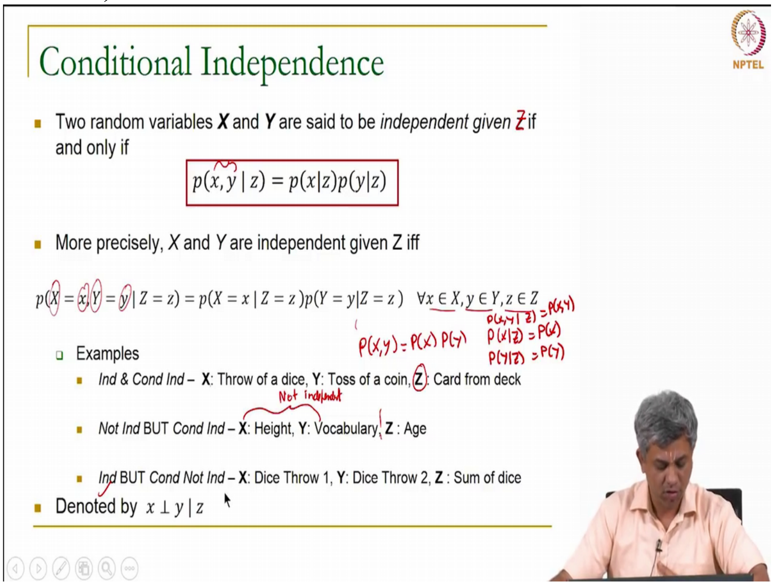
If I have two different independent dice, I throw them; the value I get on one has no bearing on the value I get on the other, Ok.

But suppose I fix the some of the dice, Ok and that is the variable Z, then if I find out x comma y given z, the moment I give you the value of x and I give the value of z, the value of y is fixed. Therefore, these events are no longer conditionally independent, Ok.

So here is a case where the events are independent, but after you add a condition they are no longer independent, Ok.

So conditional independence as you can see is a separate idea from that of independence. You can have all sorts of combination, independent, conditionally not independent etc, etc. as you just saw

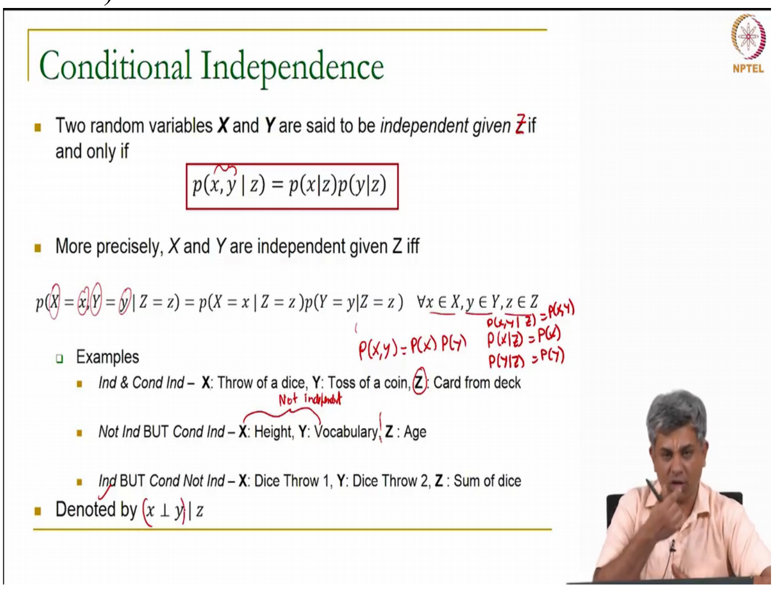
(Refer Slide Time: 11:35)



on the slide.

The notation that we use for conditional independence is x

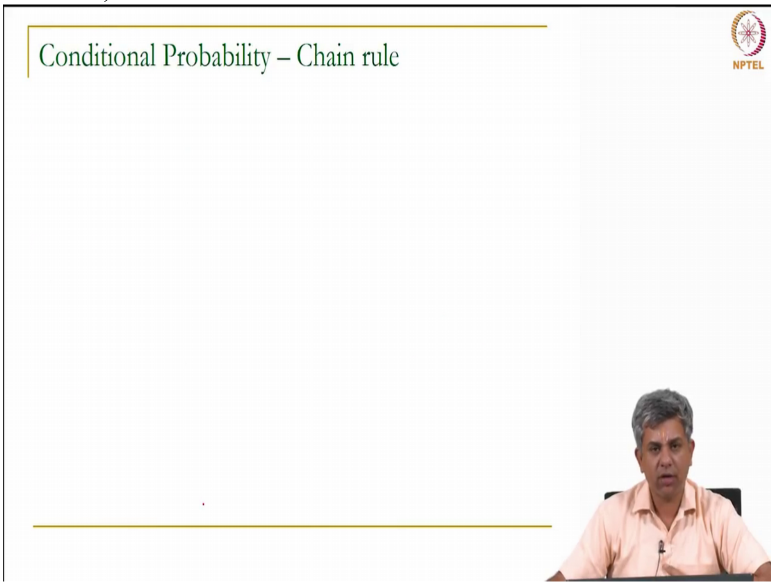
(Refer Slide Time: 11:42)



orthogonal or x perpendicular, that means x independent of y given z, Ok. So this follows naturally from what we saw in the previous slide.

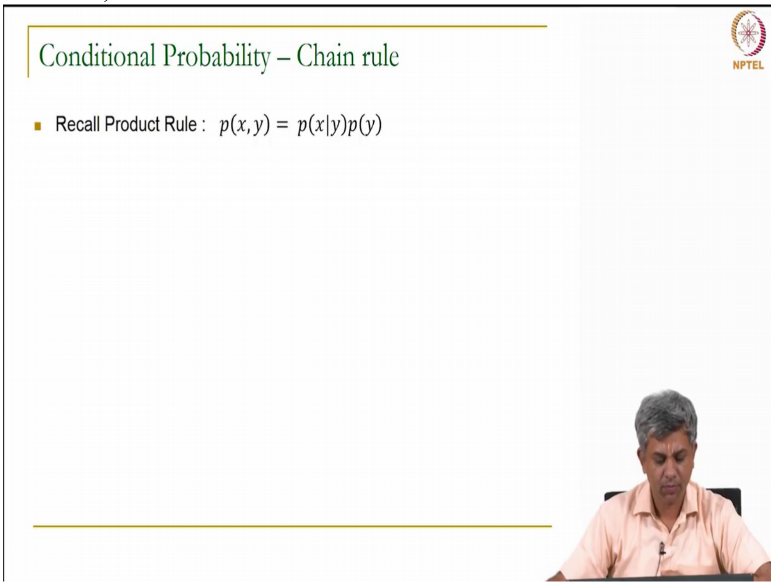
So let us

(Refer Slide Time: 11:53)



look at the chain rule of conditional probability.

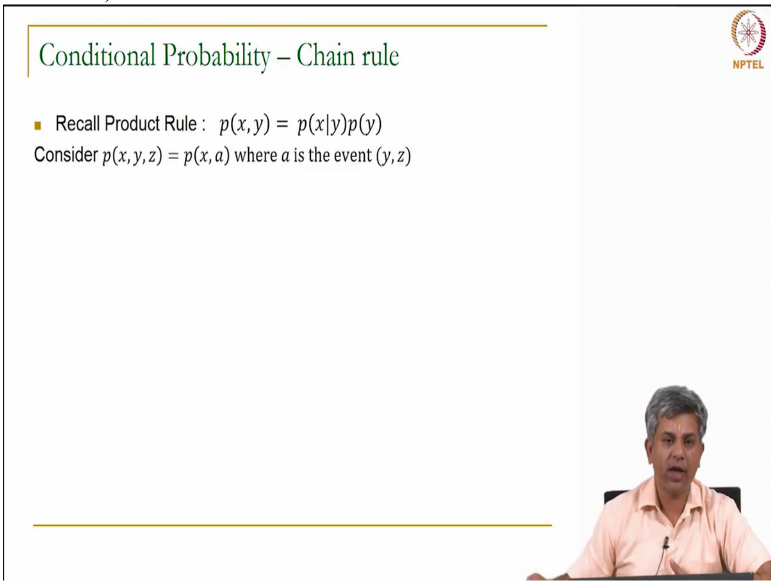
(Refer Slide Time: 11:57)



So remember that the product rule that we had for the joint probability p of x given y. This was equal to p of x comma x given y multiplied by p of y, Ok.

Now

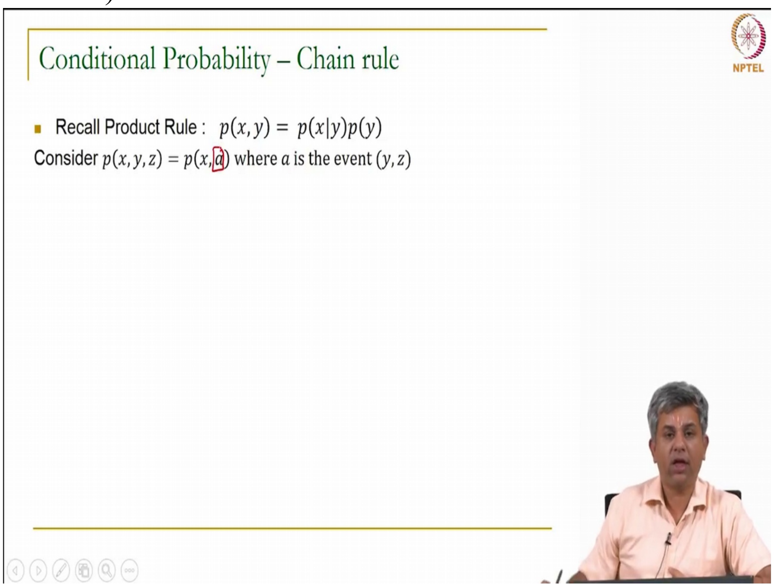
(Refer Slide Time: 12:12)



we might try and extend this to three variables, the joint probability of three variables. So let us say you have probability of x comma y comma z.

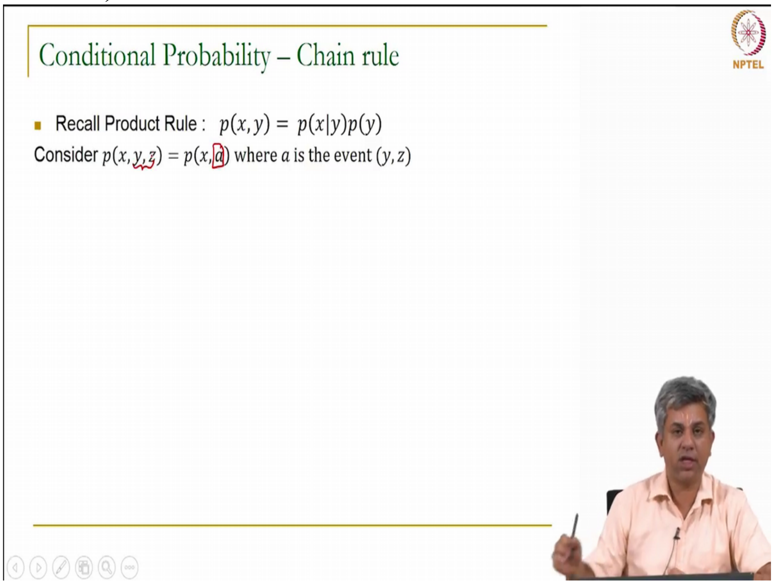
A good way to sort of find out the expression for this is to reuse this idea. For this we denote

(Refer Slide Time: 12:31)



a as the event y comma z

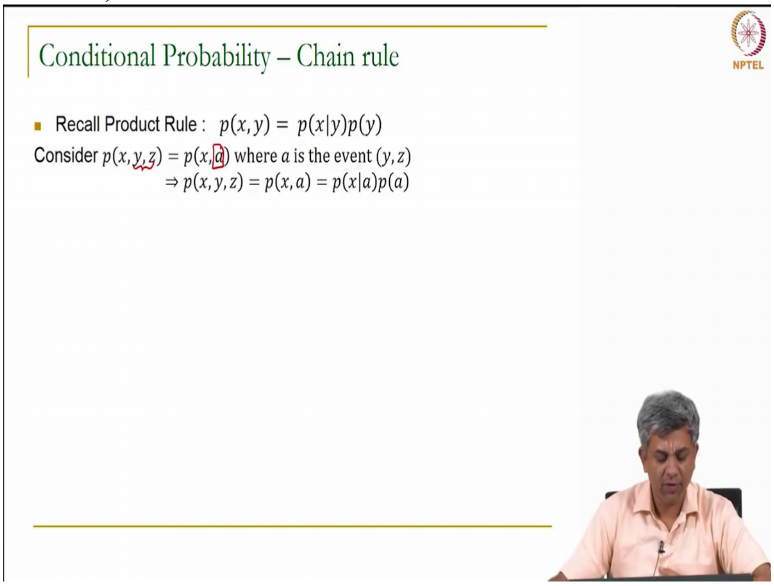
(Refer Slide Time: 12:34)



that means a is the event of y and z occurring together, the joint event y comma z, Ok.

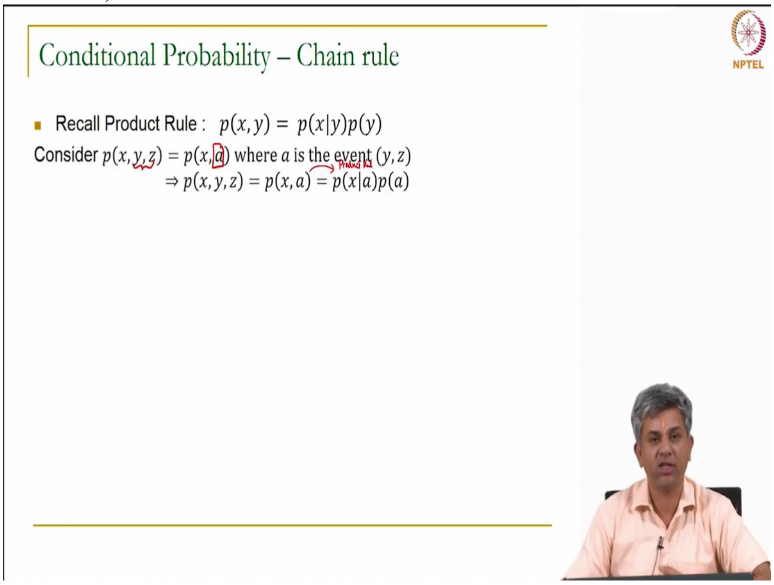
So if we do that,

(Refer Slide Time: 12:42)



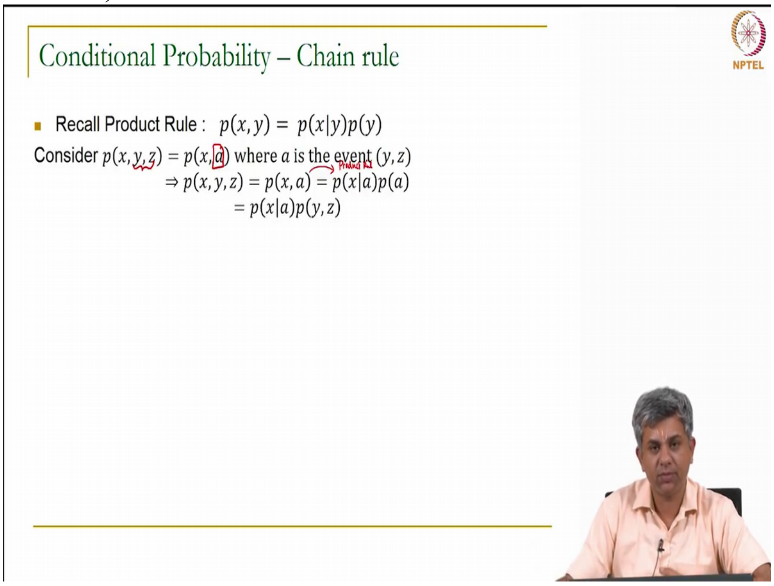
so we can now write p of x comma y comma z is p of x comma a, which by the product rule is simply p of x

(Refer Slide Time: 12:54)



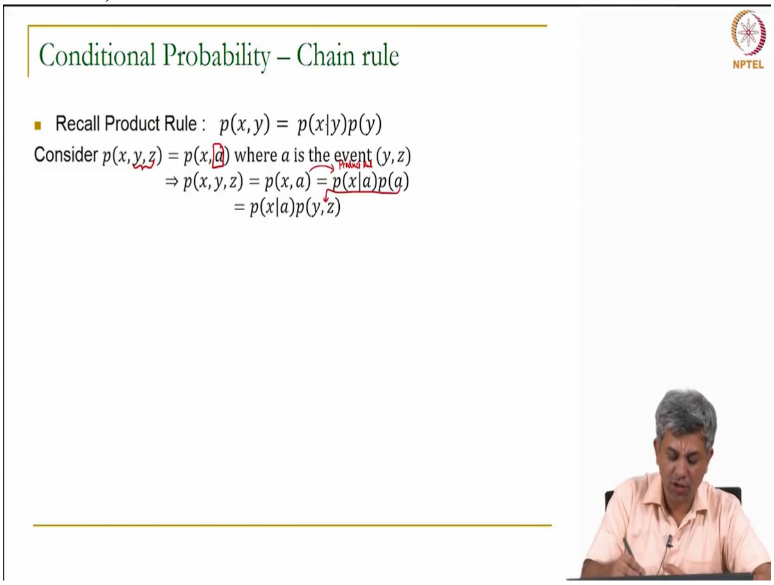
given a multiplied by p of a.

(Refer Slide Time: 12:57)



We can open this up, remember a was

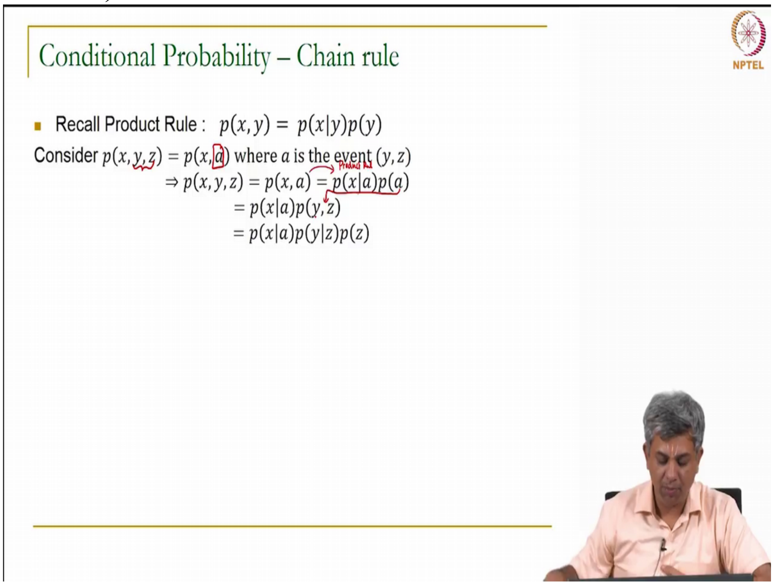
(Refer Slide Time: 13:02)



y comma z. So p of x given a multiplying p of y comma z.

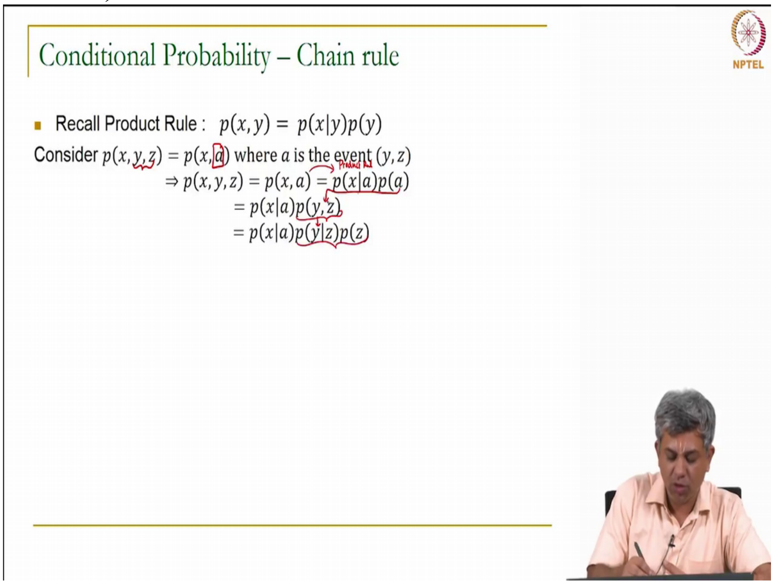
We can open

(Refer Slide Time: 13:08)



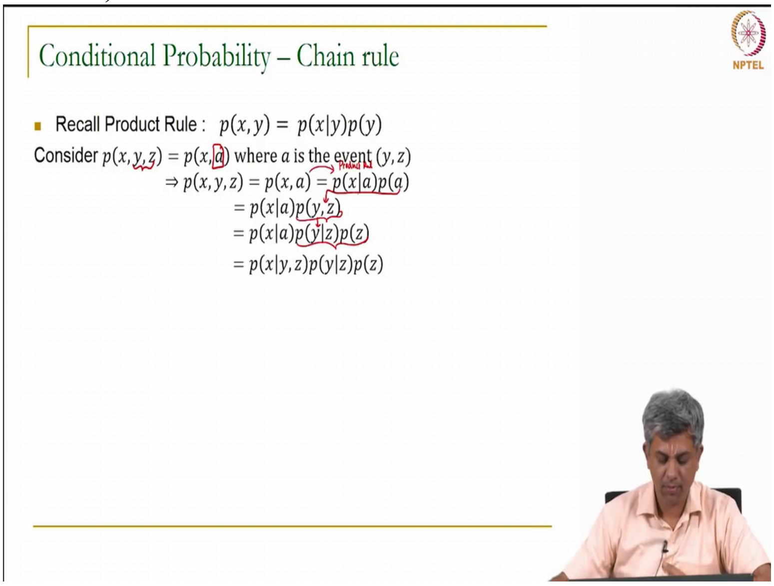
up p of y comma z to this once again using the product rule.

(Refer Slide Time: 13:14)



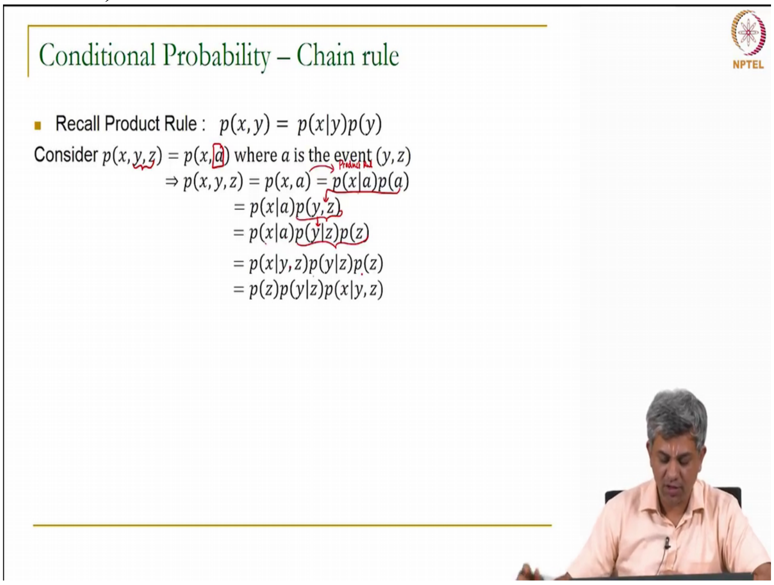
So p of y comma z is p of y given z multiplied by p z.

(Refer Slide Time: 13:19)



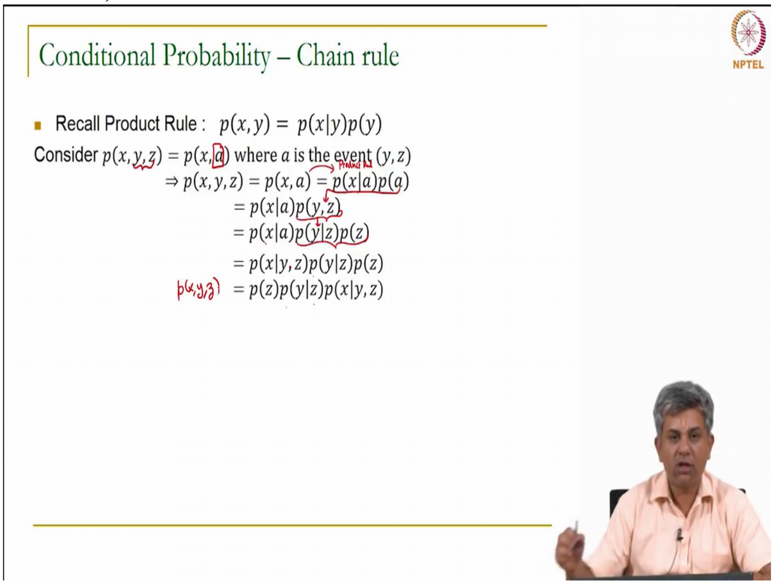
So notice this result. I am just flipping the whole thing. p of x comma x given y comma z multiplied by p of y given z multiplied by p z. If we reverse this,

(Refer Slide Time: 13:32)



we can write this as p of z multiplied by p of y given z multiplied by p of x given y and z. You can see that this is actually a natural interpretation of the

(Refer Slide Time: 13:44)

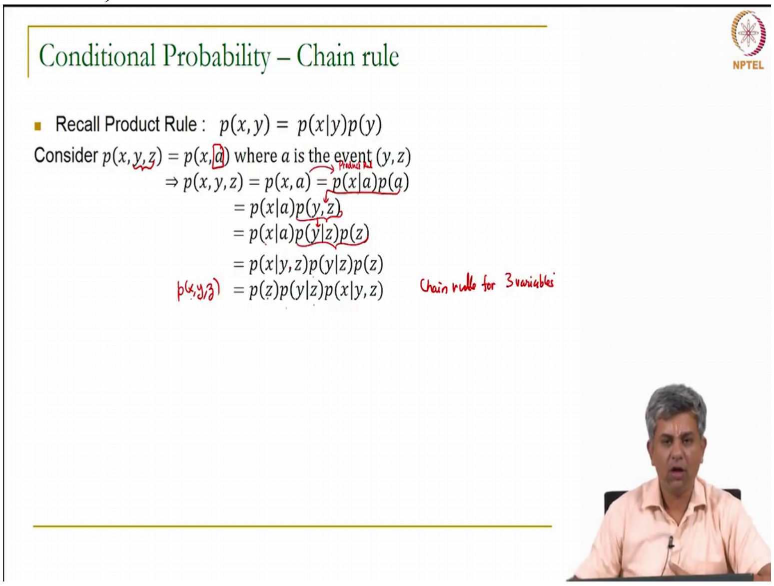


probability of all 3 events, x and y and z happening. So what does this say?

This says the probability that x y z occur is the probability that z occurs multiplied by the probability of y given that z occurs then the probability that x happens given that y and z have also occurred, Ok.

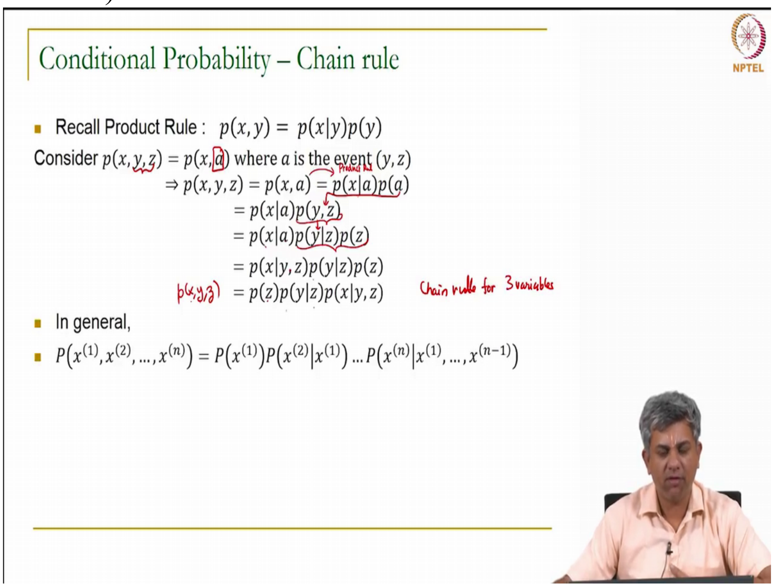
So this is a simple interpretation of what happens and this is why this is an example of the chain rule for 3 variables.

(Refer Slide Time: 14:24)



Of course now we can extend it to n

(Refer Slide Time: 14:28)

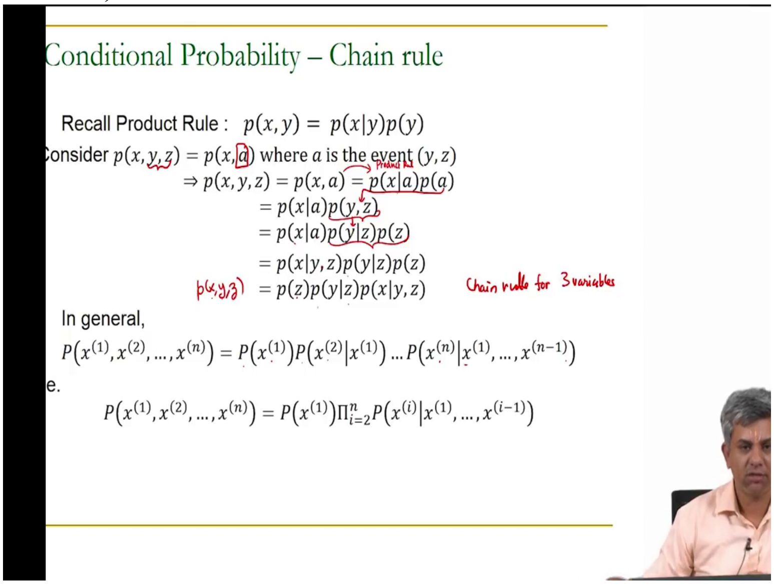


variables in general.

So, instead of just 3, if you have n events, x 1 through x n which occur jointly you can now write it as the probability that the first event occurs multiplied by the probability that the second event occurs given that the first has occurred, so on and so forth until the probability that the nth event occurs given that the first n minus 1 occur.

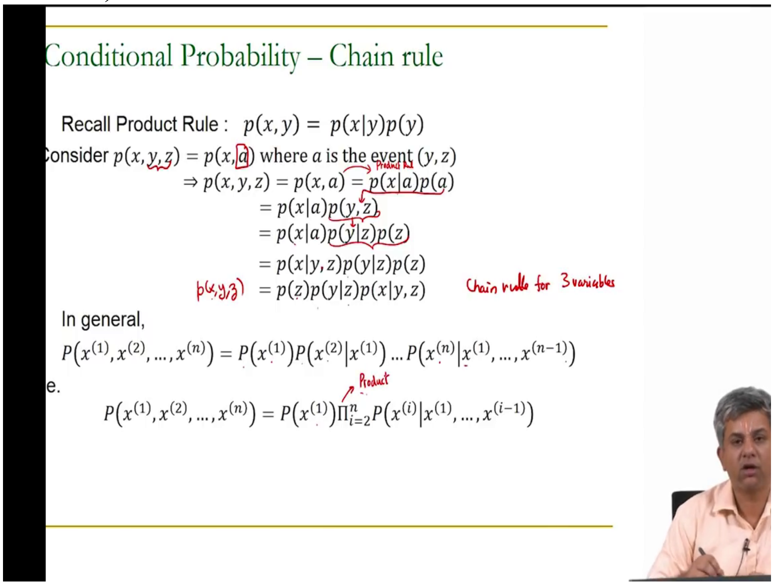
In compact notation,

(Refer Slide Time: 14:52)



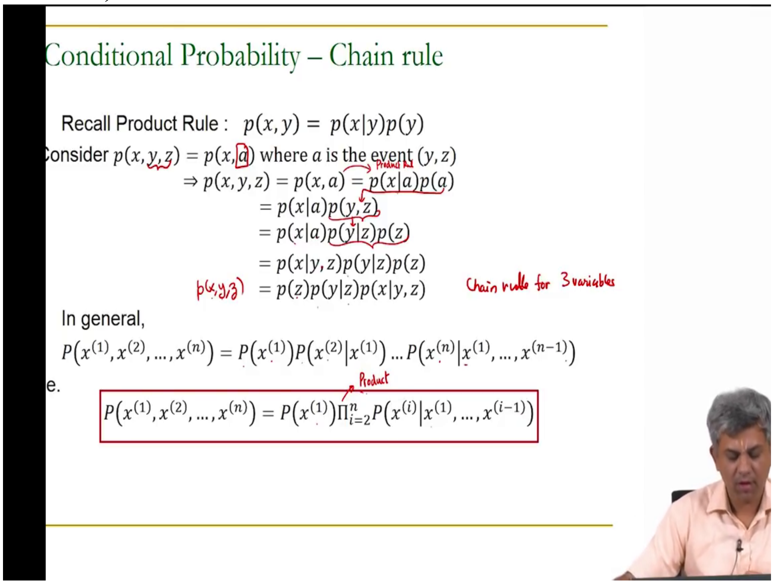
you can write this as the probability the first event occurs multiply by the product, remember pi denotes

(Refer Slide Time: 15:03)



product, product of i equals to 2 to n of this event, x i given that x 1 through x i minus 1 occur,

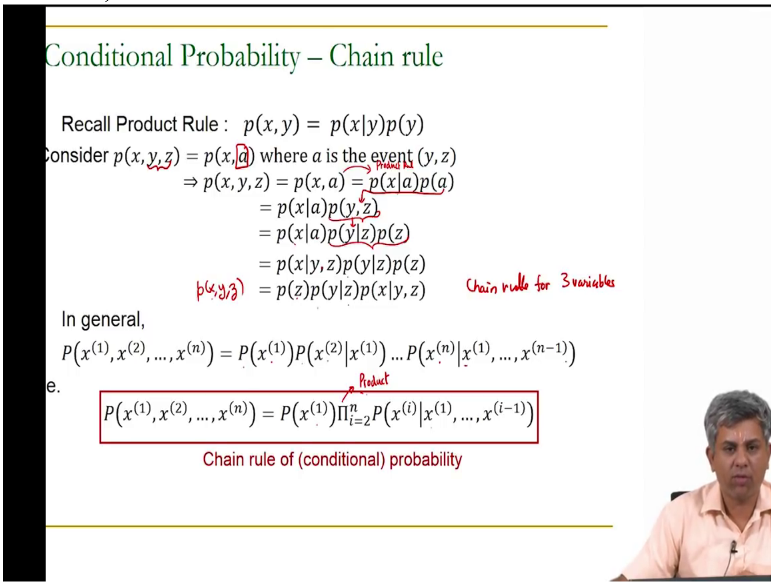
(Refer Slide Time: 15:15)



Ok.

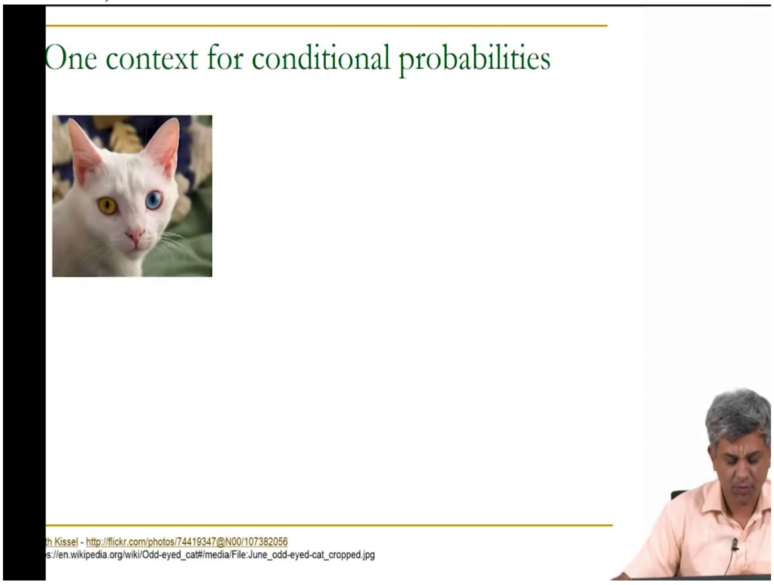
This is the chain rule of probability,

(Refer Slide Time: 15:18)



Ok.

(Refer Slide Time: 15:20)

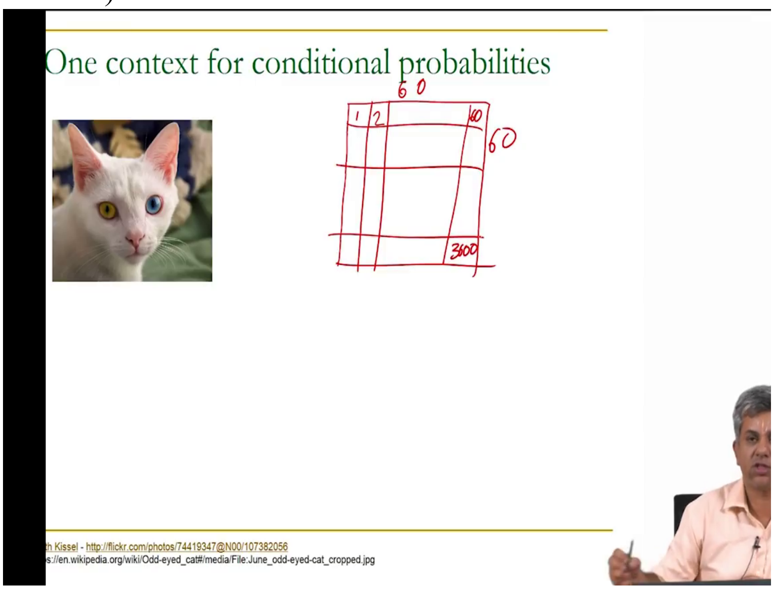


So here is the simple context where we will be using conditional probabilities later. This is just to give you some, you know look ahead what will be happening.

Let us say you have some such image. This is the image of something called the odd-eyed cat. It does occur naturally even though it looks like a fake image. You have some eye, cat with two different eyes. But suppose we ask what is the probability that such and such image occurs.

Now remember our idea from the linear algebra videos that let us say this is a 60 cross 60 image. It has 3600 pixels. First pixel, you know sixtieth pixel, so on and so forth uptil the

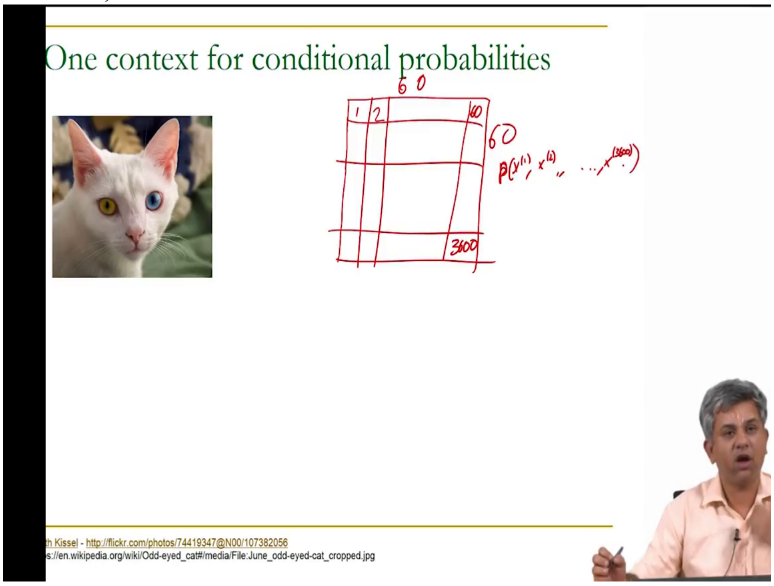
(Refer Slide Time: 16:04)



three thousand six hundredth pixel.

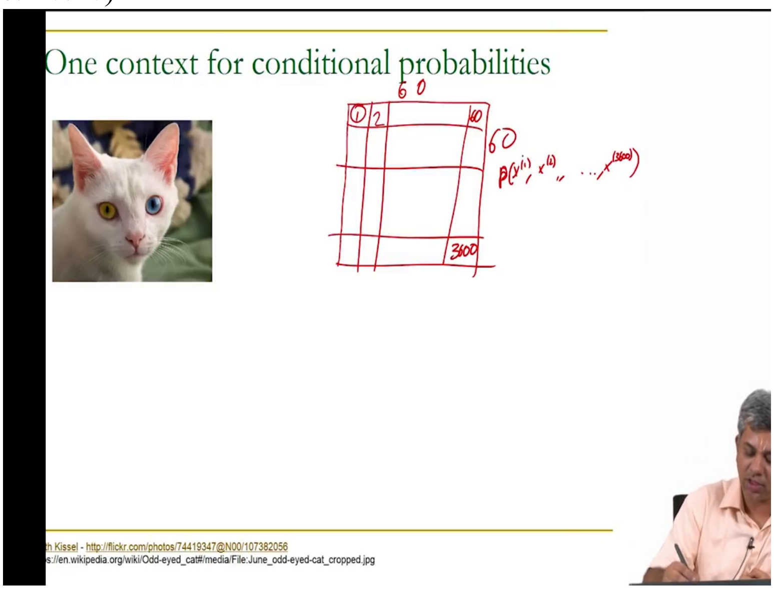
So what I want is basically the probability that all these pixels take the given intensities that they have taken. So we

(Refer Slide Time: 16:21)



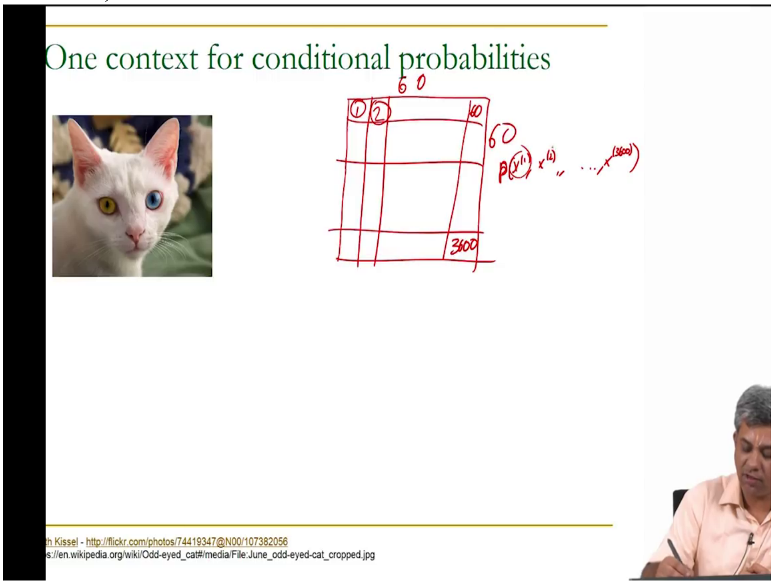
can think of this image now as a single event of the first pixel taking the value x 1,

(Refer Slide Time: 16:26)



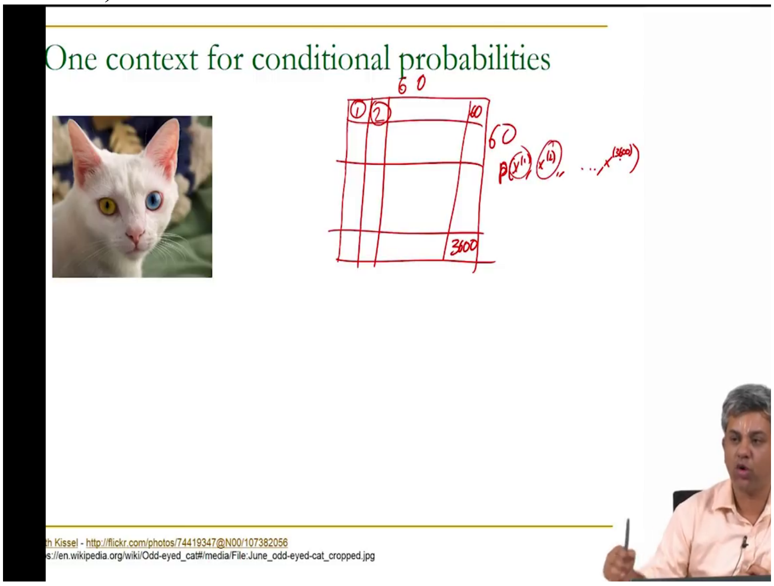
second pixel

(Refer Slide Time: 16:27)



taking the value x 2, so on

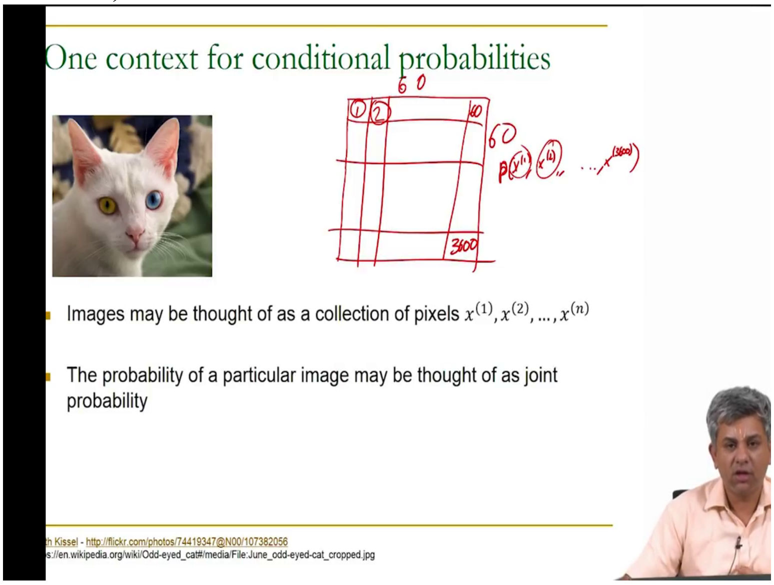
(Refer Slide Time: 16:29)



and so forth till the three thousand six hundredth pixel taking the value x 3600, Ok.

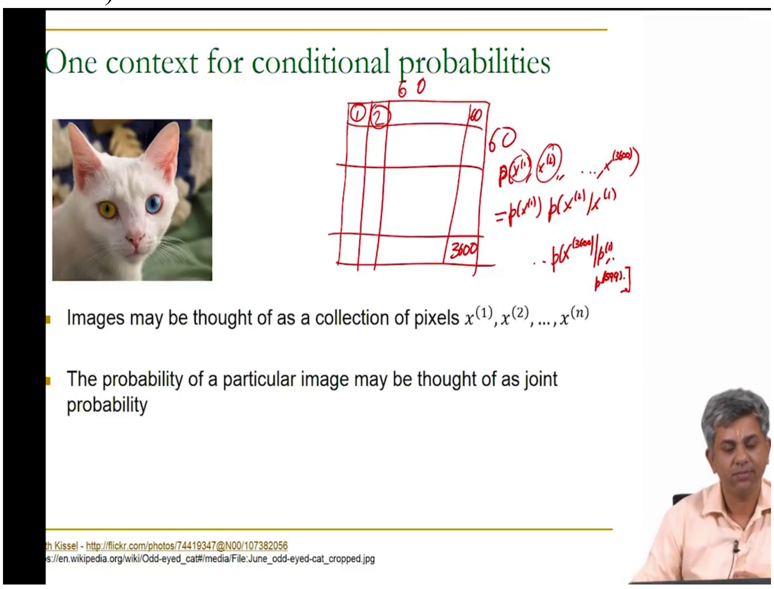
And how do we find this out?

(Refer Slide Time: 16:40)



You can basically think of this as the joint probability and now you can rewrite using the chain rule that we just wrote, you know what the probability would be. This probability is equivalent to p of x 1 multiplied by p of x 2 given x 1 so on and so forth, p of x 3600 given 1, 2 up to 3599,

(Refer Slide Time: 17:07)



Ok.

Now in addition to this, if we somehow got to use independence of, conditional independence which we will also do later using some kind of machine learning models, you will see that this whole expression can be simplified tremendously.

So this is one context, this is not the only context where we will be using conditional probabilities and the chain rule but this is one context where we can use this very, very conveniently.