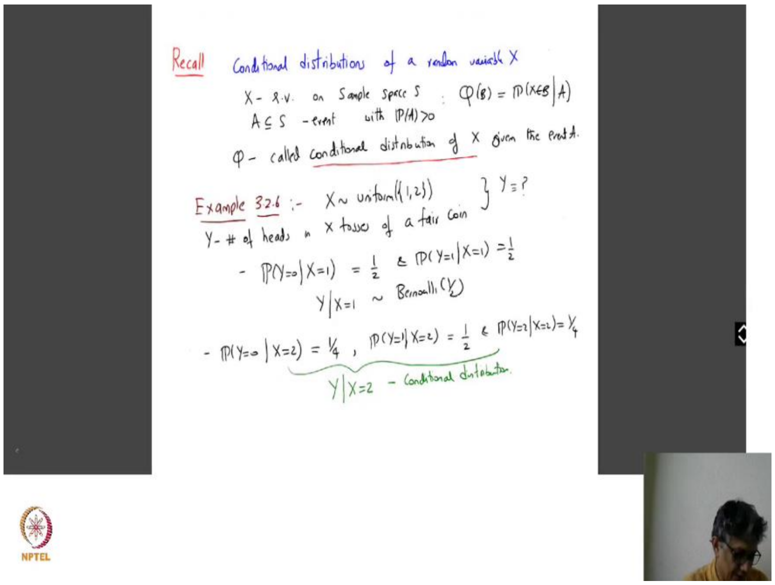
Introduction to Probability – With Using R

Professor Siva Athreya
Theoretical Statistics and Mathematics Division
Indian Statistical Institute, Bangalore

Lecture 17

Memoryless property of Geometric Distribution

(Refer Slide Time: 0:33)



So, we just wanted to recall what I did last time, so I had said that I was doing this idea of conditional distributions so, of a random variable X and the idea was that if I give you a X random variable on a sample space S and A is some subset of S event, the probability of A is positive if these two things then what you could do is you could define a conditional probability Q of any event B as the chance of X and B given the event and Q is called the conditional distribution, distribution of X given, X given the event A, that is the key concept which will under study last time so this is the conditional distribution of X given.

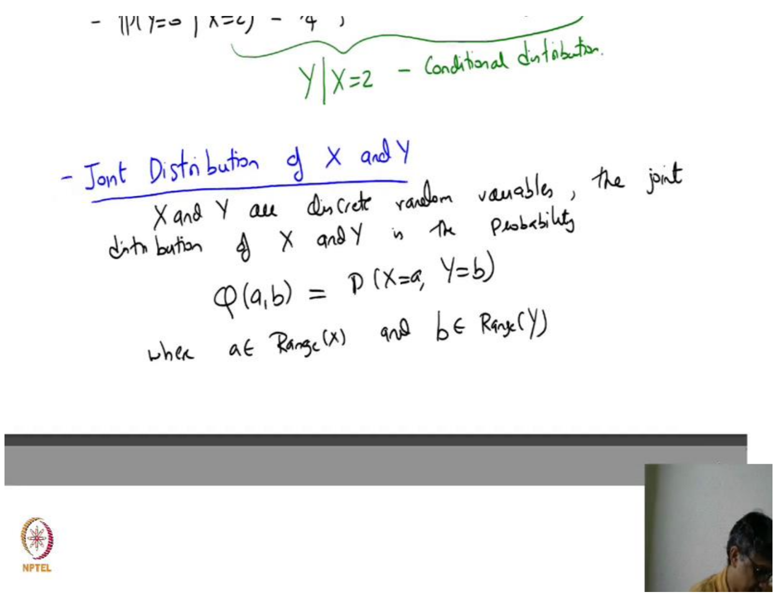
So, then we studied this with regard to an example so we said that first we have this example, so let me I just call it I will just continue the example again so I will start with an example 3.2.6, I will keep playing example again and again. So, the idea was that we took we chose a number between 1 and 2 so that is like saying that we took X to be uniform 1, 2 and then we said Y is the number of tosses of a fair coin so Y is the, so Y we described Y let me just rewrite it a little bit on the side, so Y is the number of tosses so number of heads in X tosses of the (())(3:31) of a fair coin… and so we try understand the distribution of X from these two things distribution of Y.

So, what is how does Y behaves, X is clear and how does Y behave. So, one of the things we calculated last time was that so we said that so if you look at Y is equal to 0 given X is equal to 1 that means you toss a coin once and you want to know how many heads you get you get 0 heads that happens with probability half and we also found out that probability Y equal to 1 given X equal to 1 (())(4:20) so therefore one would say that Y given X equal to 1 is distributed like Bernoulli half.

Then we also went into the next step we said fine if X has a value of 2 then we saw that the chance that Y equal to 0 given X equal to 2 that means you get no heads in 2 tosses that is the same as 1 water and chance that you get 1 head in 2 tosses if you do the calculation of Bernoulli a binomial 2 comma half and you have only one success that happens with probability R and the last one was the chance that Y equal to 2 given X equal to 2 that happened with the chance last quarter.

So, this specified the distribution… (())(5:25) so these 3 numbers specified the distribution of Y given X equal to 2. So, that is where we were so then we said this is one way of understanding X and Y so let me just I have this new trick now let me see if it works go down a little bit you can see the whole page, so here this is where we were at, so then after this I said fine now, I want to go back.

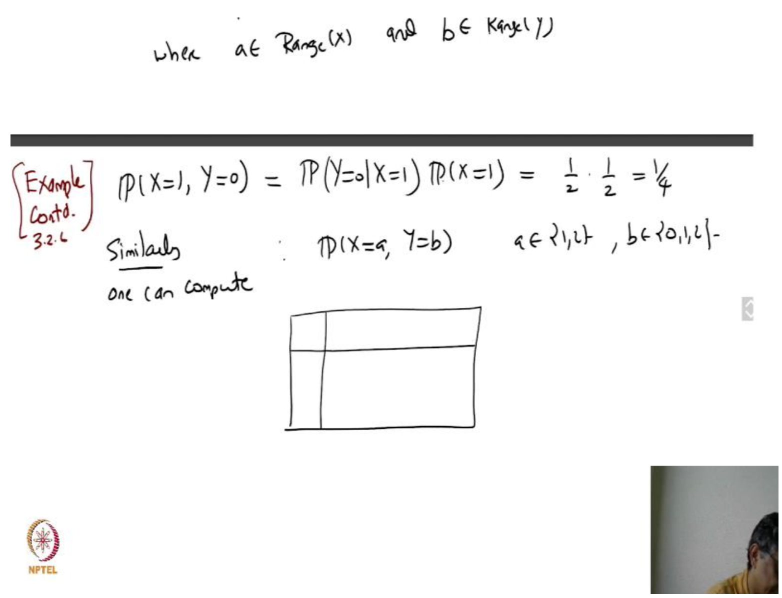
(Refer Slide Time: 6:28)



So, now the next thing I did was I said fine this is one way of understanding X and Y, the other way to understand X and Y is what is called the joint distributions, of X and Y two random variables X and Y, just keep this example in mind example 3.2.6 and here I said the following.

I said if X and Y are discrete random variables then the joint distribution of X and Y so is given by is the probability Q, so Q of a comma b is the chance that X takes the value a and Y takes the value b where a and b are where a is in the range of X and Y is in the range of b, what am I saying I just need to say that b is in the range of Y. So, this is the probability cube given by this, so this action that pairs the value of X and Y so if you go to the same example that we did before so continue example let me go to page 2 so it is easier for us to follow.

(Refer Slide Time: 8:25)



Page 2 we did we also noticed that the chance that X equal to 1 and Y equal to 0 so just let me rewrite that down, so the example continued, so it is 3.2.6. So here we notice the fact that if you want to compute this guy you can do it you can say that Y equal to 0 given X equal to 1 and the chance that X equal to 1 that is just by definition of conditional of probability.

This we know is a half the first one and second term also we know the half because you choose two numbers between 1 and 2 and this gives you a quarter and similarly you could compute the other ones. The chance that X equal to 1, Y equal to 1 is something and all the other numbers. So, maybe I will write here X equals similar to compute the chance that X equal to a, Y equal to b for a and in 1 comma 2 and b and 0 1and because that is what Y takes.

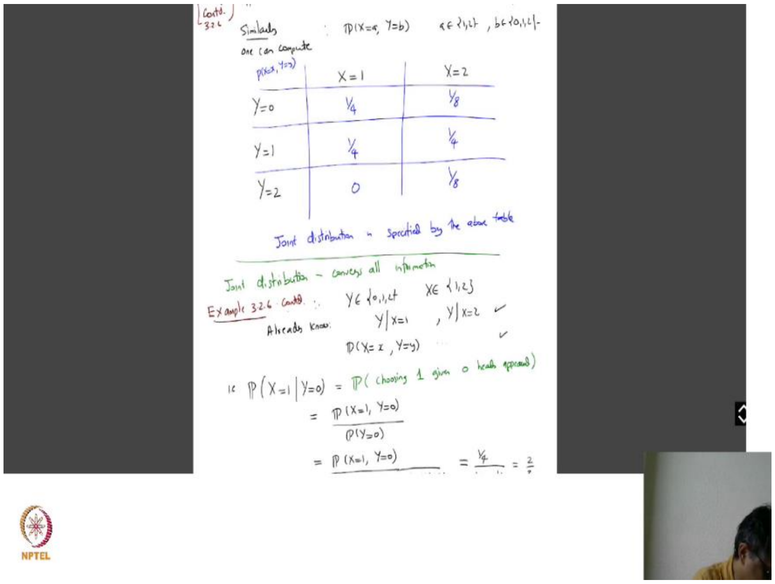
So, then once you do this you end up with this table. So, a table like this so it is maybe a lot on the table first so you have X equal to 1 and you have Y equal to 1, so X equal to 0 and X equal to 1 (())(10:22) make a little mess of it so X equal to 1 and Y equal to 2 is equal to 2 sorry and Y can take the values 0 Y can take the value 1 Y can take the value of 2 and then we know that in this let us say we put the probability of X equal to Y, X equal to 1 and Y equal to 0 so what I will get here I will get one quarter I will get one eight to do the computation I will get one quarter, 0 I get a quarter here and one eight.

So, I am filling in the probability of X equal to, a and Y equal to b in this table, that is all I am doing. So, this table is probability X equal to X and Y equals to 2. So, now this gives you a complete picture of the joint distribution this is the entire distribution, so joint distribution is specified by the above table. So, this is all sort of where we were at last time, we sort of told you how to do this thing.

So, you can go back and forth between conditional and joint distribution by this (())(11:53) right here. So, let me now try and see if I have another trick I learned last night let us see if I can implement it today let us go to the split screen, split view I want to see the same book again, I want to split view again, like this then let me just make this little bit less, let me move the first page across over here so we can see both the things in one side.

So, this is what I was discussing last time, and this is where we were at so we had discussed the condition distribution and then the joint distribution of X and Y and this is the competition we had come. Today I want to build on this a little bit and then see where we are going. Let me go back to full screen now.

(Refer Slide Time: 13:45)



Let us start today, so today what I want to do is the following today what I want to do is I want to understand several things from this about condition, so one easy trick is that suppose I know the joint distribution then you can also compute the condition. So, if you look at the about table right here right that is easy to see, so if I know the joint, I also know the conditional joint gives you everything so one thing is that so one observation is that the joint distribution conveys all information that is one thing to understand.

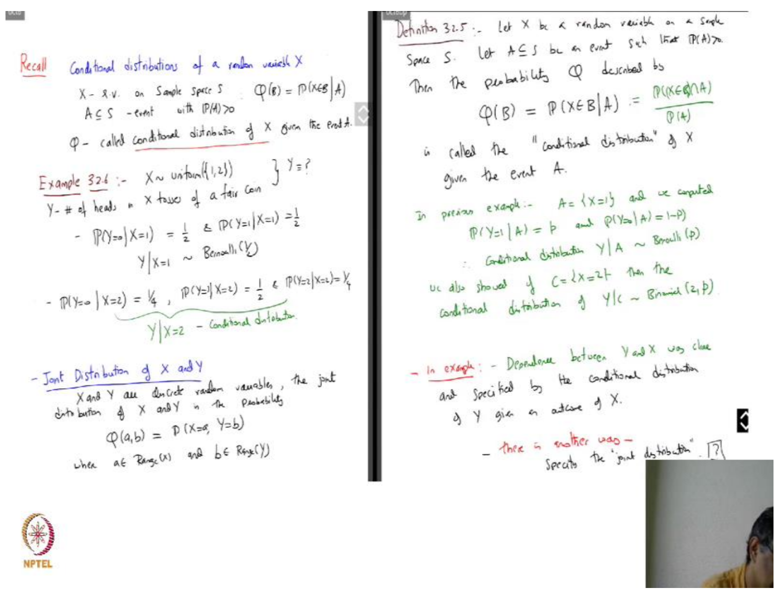
So, one may when we do the following competition so let us do an example again continue, so the example continues. So, the idea is that we know that Y takes the value 0, 1 and 2 right and we know X takes the value 1 and 2 this much thing. So, now there is a sort of a can I from the joint can I get other information so for example we already know how does Y behave when X is equal to 1, how does Y behave when X is equal to 2 this, we know from the model directly from that we also know the joint distribution of X, the probability of Y equal to X equal to X and Y equal to 1 is also we know.

So, now can I go the reverse way? Can I do this simpler thing? Can I know suppose I observe no heads at the end of my experiment, can I go and check what was X. So, that is I observe no heads that is Y is equal to 0 and I want to understand if X was 1 did, I pick 1 the first, so the same as saying let me write this down in green.

So, it is the probability of choosing 1 given 0 heads appear. All you are told is that 0 heads up but that is the probability I am trying to understand, so it is a little reverse like I am trying to do but this also you can do it in exactly the same way, so you just do is equal to the chance that X is equal to 1 and Y is equal to 0 at the top divided by probability that Y is equal to 0.

Now, Y equal to 0 you can do the same thing you can do a little trick you can do probability X equal to 1, Y equal to 0 this we know from the table and the bottom divided as probability Y equal to 0 X equal to 1 and probability Y equal to 0, X equals to 2. These are two things and then you can go back to the above table and you can just see Y equal to 0, X equal to 1 was a one quarter so I can put Y equal to X one quarter so I put a one quarter at the top and at the bottom I have Y equal to 0 is one quarter plus Y equal to 0 X equal to 2 is one eight. So, I can put a one eight so this becomes I do it carefully I will get 2 times. So, that means the chance of X equal to 1 given Y equal to 0 is two thirds.

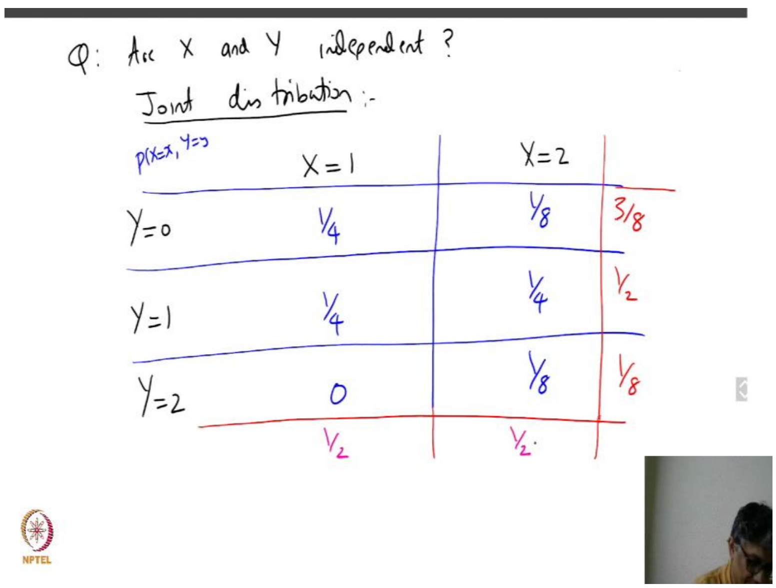
(Refer Slide Time: 18:08)

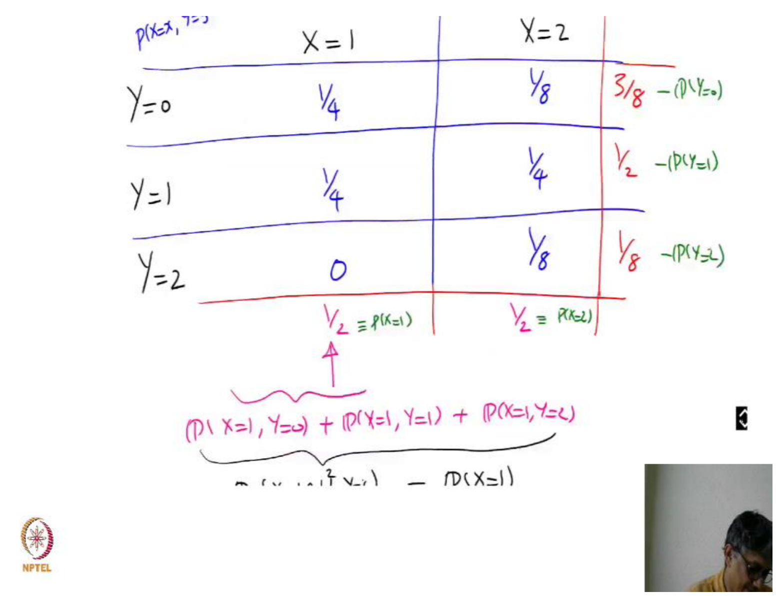


So, what did we do for this let me go back to the split screen view again so you can just view it properly I still do not know how to get it in the first shot correctly but we will do twice let us wait. So, now here like we discussed this is the second page so let us go back to the first page here, this one and here let me again go back to the first let me get the first page here, so here the idea was that I had to come over to open a wrong page let me just do this this way this, so the idea was that I looked at this joint probability table and then I was able to show you that using the joint probability table I can compute the distribution of X equal to 1 given Y equal to 0.

So, we can compute other conditional distributions given the joint distribution. So, join distributes much more information so next competition I want to do the following so let us say I tell you that I give you this table and I say that are X and Y independent? So, what you can check is you can go to the table and see let us look at something simple let us look at X equal to 1 and Y equal to 0.

(Refer Slide Time: 19:29)





So, are they are X and Y we know this this for a fact that X and Y are not independent suppose I ask you are X and Y independent? So, what you could do is the first thing you could do is let me just make it smaller here itself let us just see so the idea is that you go to the table see in this table you can observe that let us do a simple let me just copy this table let me come here, let me duplicate it, let me bring it down, so this was the joint distribution.

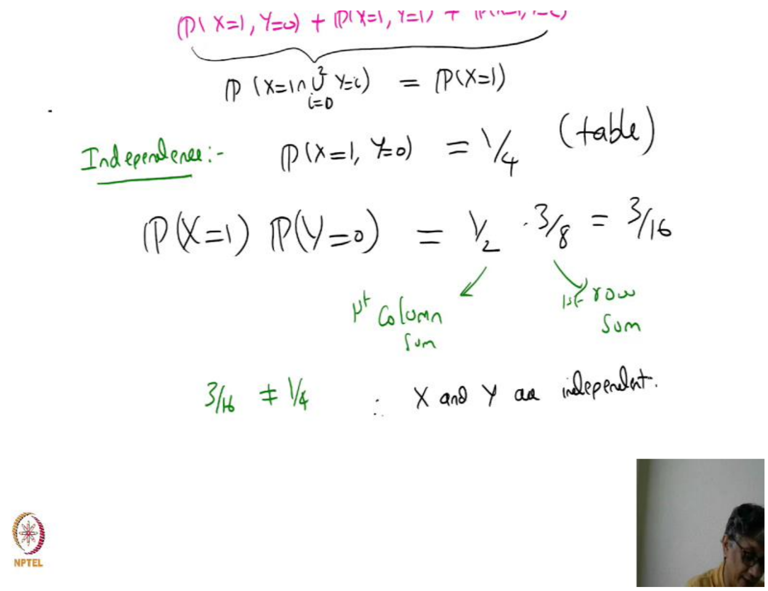
Let me go back up to something so the joint distribution what we found out. So, one observation is immediately the following so suppose if I go to this last column if I go to this last one and the rows on this side if I add things up if I add the row substances so you add one half and one eight what will I get? I will get one half plus one eight is three eighths that is one thing I will get one half plus one fourth is row sums one half and then this is one eight and I come on this side let me do a different color let me just let us say from pink let us say let me go here if I add the column sum is what I get I get one half here, I get if I do one eight one quarter and it is one half here.

So, now if you just think a little bit what do these numbers represent? So, this number represents the chance that what am I adding I am adding these 3 numbers, so this number represents I am adding the chance that Y is equal to 0, X is equal to 1 let me write to the reverse way it is a little bit easier to understand. I am adding the number X equal to 1, Y equal to 0 I am adding the next one is chance that Y equal to 1 or X equal to 1 and Y equal to 1 and I am also adding the chance that X equal to 1 and Y.

So, if you think a little bit this is just going to be equal to 1, let us go back to that this is just going this whole thing is just they are all independent, mutually exclusive events so you can write this as the chance that X equal to 1 union Y I equal to 0 to 2 of Y equal to I but this is the whole space and that is the same as the chance that X is equal to 1.

So, it means in all the row sums I am getting the chance that X is equal to 1 let me add that in green check so along the row sums I am getting the chance that X is equal to 1, this is the chance that X is equal to 2 to add these things up and if I look at the row sums this is the same as the chance that Y is equal to 0, chance that Y equal to 1 and the chance that Y is equal to 2. So, the joint distribution table gives you even the individual ones for X and Y which you found before that is one thing.

(Refer Slide Time: 23:20)

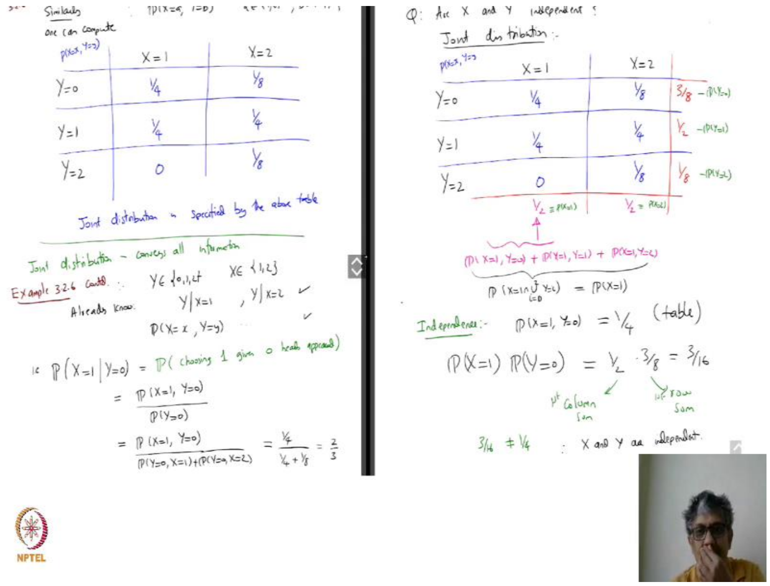


So, now what I want to say I want to see if they are independent, so if they are independent let us do a simple check. So, for independence what you have to check if for independence you have to check let us do a simple thing, let us take X equal to 1 and Y equal to 0. So, from the table what do we see? From the table we see that X equal to 1, Y equal to 0 is one quarter, this from the table.

Then we go to the appropriate row sum and check what happens to X equal to 1 and chance that X, Y equal to 0. So, for X equal to 1 I go down to the first column I get 1 half, so let me write that is right and then X equal to 2 Y equal to 0 I go to the first row I get 3 8 and I get 3 over 16.

So, this is from the row sum, this from the column sum and this middle somewhere this is from the column sum and this one row sum. So, even that is the first column and the first row, so once you do this you clearly see 316 is not equal to one quarter. So, therefore X and Y are not independent, of course we knew this in the beginning itself, but this is a verification (())(25:03). So, from the joint distribution table you can get many many things you can check your independence property, you can check conditional distribution, so this particular table gives you everything.

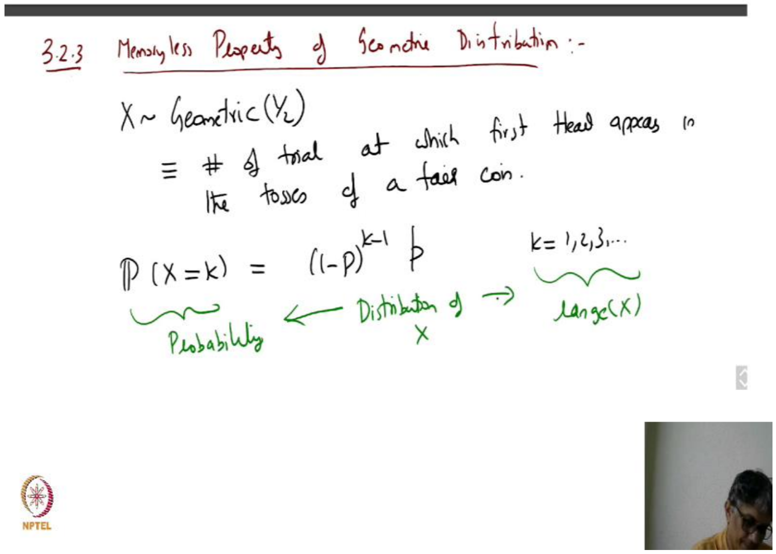
(Refer Slide Time: 25:53)

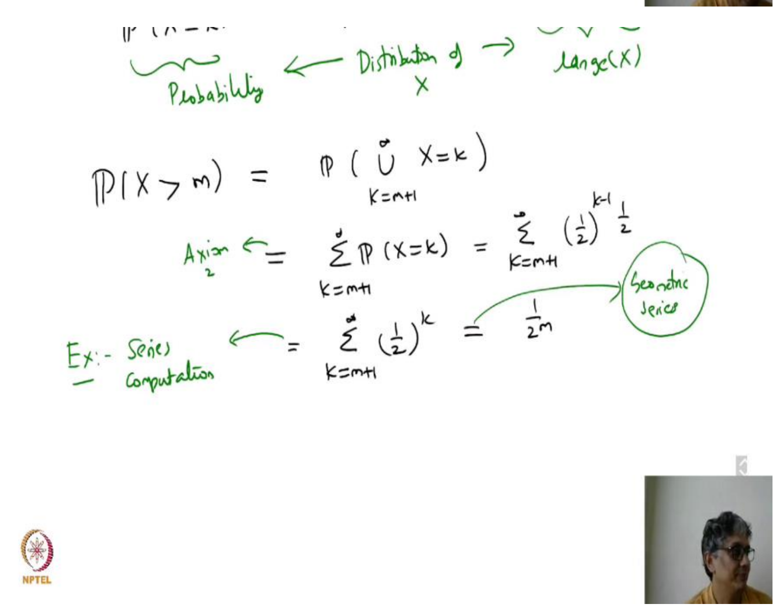


So, let me just again let me try the split screen again see if it works, (())(25:29) eight again, split again (())(25:38) small… so this is the last page I did and this is the second page. So, this is where we are at, we can see both sides that the distribution of X given Y can be given can we also write down distribution table and so can you can check things like independence, you can also check the individual distributions of X and Y by doing the column sum of the row sum and so on.

So, here let me see if I can have the mark here this is the thing and what is the distribution of X or 1 so this both these things can be obtained from this. So, now I just want to close this class with one simple example or a couple of simple examples, so one is that a random variable can behave differently under some conditioning by itself, so the conditioning event can be quite interesting to understand. So, let us try and see if I can observe that.

(Refer Slide Time: 27:23)





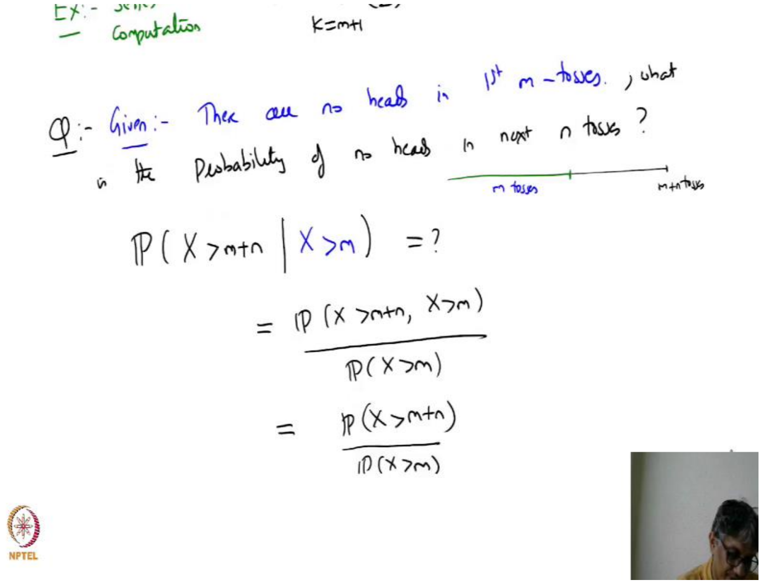
So, let us say this is called the memoryless property of geometric it is very very crucial property geometry I want to establish it for you, this is called section 3.2.3, this is called memory less property of geometry distribution. See here what happens so here let us go for the nice interesting property, so the first thing we know is that let us say X is a let us say geometric half, that is what is X represents in the as an experiment X represents the number of trial at which so to think of it as first you can think of it as first success or first head appears in the tosses of a fair coin that can be the understanding, So, now one thing we observed long time back was the chance that let us say X equal to K the distribution was going to be you have failures for the first K minus 1 tosses and your success is the Kth top so it was K equal to 1,2,3 for the distribution of X. So, the range of X was this on this side and this is the probability and together this gave you what we would always call as the distribution of X.

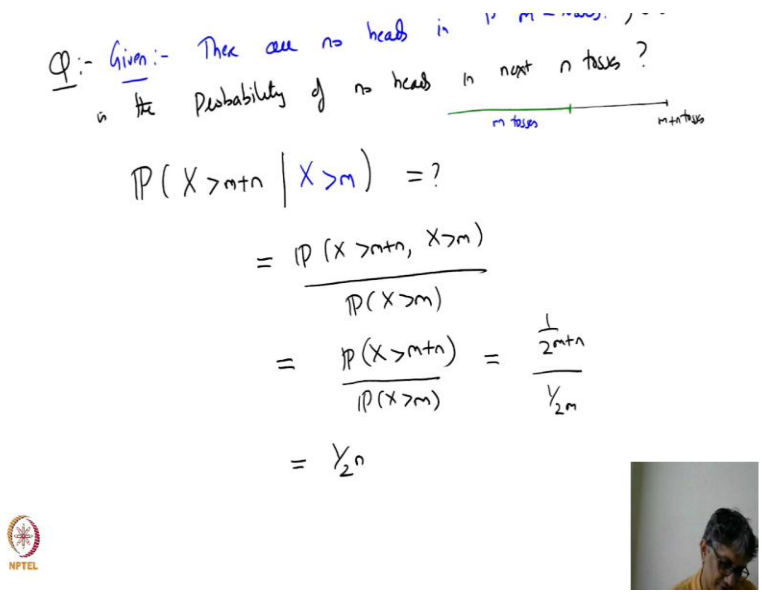
So, now I know this so what do I want I said memory less let me just understand what memory lessons let me try and justify that a little bit, so what I try to do is I try to first understand the chance that X is bigger than m that means you take at least m tosses to get the first head. So, what does that mean? That means you look at the event when K is equal to m plus 1 to infinity and such that X is equal to K. So, X is any one of the values K from m plus 1 to infinity, so that is the same as if you use the maximum probability 2.

So, what do you get, you get the same as the sum from K equal to M plus 1 to infinity the chance that X is equal to Q. So, all this is well defined we know that X equal to K in our case is just so let me just backpack a little bit let me just say since I had fair coin my P was a half in my example P was equal to half.

So, what I would get is I would just get the same as sum from K equal to m plus 1 to infinity, one half to the power K minus one half times one half of one half and of course I am writing it in a little sloppy way to justify it as a limiting object so that is same as submission K equal to 1 m plus one to infinity there is one half to the power K and that if you do a geometric sum you would just get one half to the power M. So, these are these two things I will leave the exercises for you to try so one is that this is a series computation, and this is a fact about geometric series. (())(31:40) So, chance of X bigger than M is equal to 1 over 2 3.

(Refer Slide Time: 31:59)





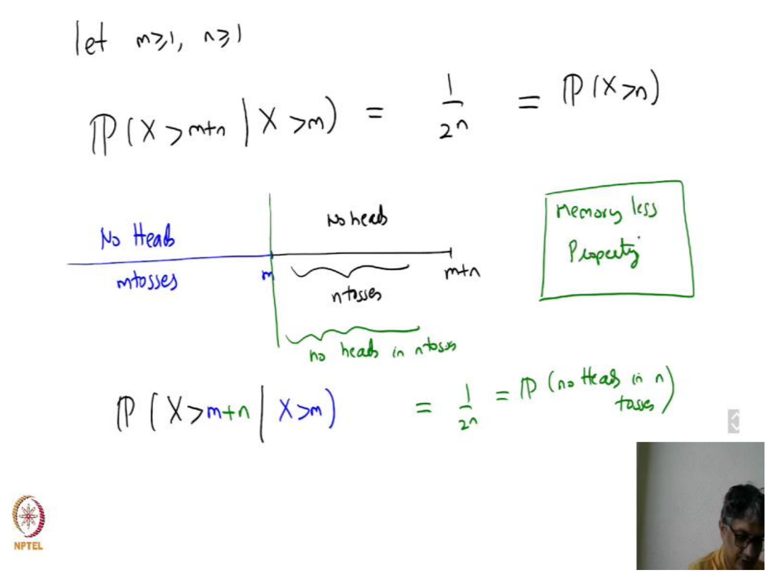
Now, let us do a simple thing so let us say I have waited till M trials so let us say so we are given so let us go back to the black here so let us say I am given the following fact. I am given that there are no heads in the first M tosses, so that means there are no (())(32:14) and then I am asked to condition I must understand the question is the probability what is the probability in words the probability of no heads so I am given this in the next N tosses.

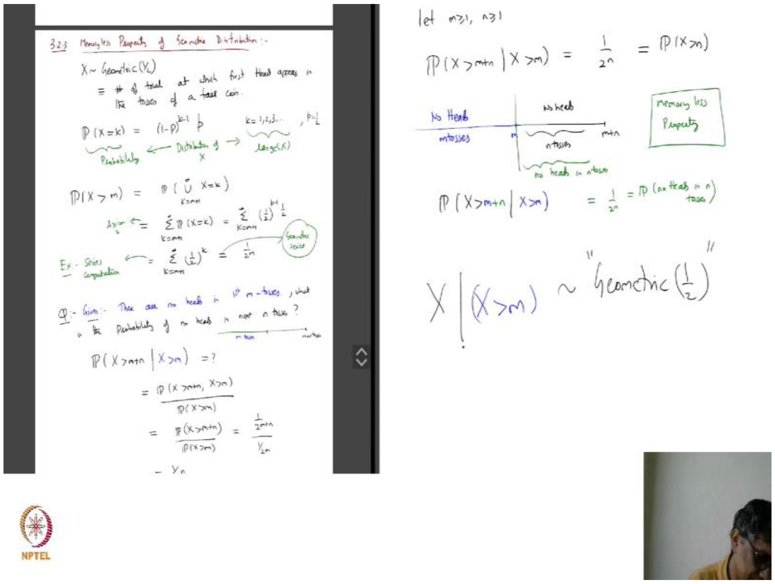
Given there are no heads now what is the probability of no heads in the next N toss this is the question what I have asked. So, now how do I write this question that means I want to understand the chance that I am given there are no heads in the first N tosses that is the same as saying X is bigger than that is the conditioning given, X is bigger than m and then what I want to understand there are no heads in the next n tosses that means I have already waited till m so m until m there are no heads that is m and then I want to meet another so this is given m tosses no heads and I want to wait for another m, so let me another m that will go this way so another m if I have to add I will go up to m plus n tosses.

So, I want no heads in next m means I want no m plus n as well, so I want to understand chance that X is bigger than m plus n given that X is bigger than m. So, how does one do this you do the same trick like before the chance that X is bigger than m plus n and X is bigger than m the whole thing divided by the probability that X is bigger than m and that is the same as so the bottom is easy to do you already understood it the top is it this both these happen so that is the same as saying so X is bigger than m plus n and X greater than M that is already included so it is the same as X is greater than m plus n and the bottom is just X is bigger than m.

So, if you go back to the first computation that we had X bigger than m was 1 over 2 to the n so we plug that in here so the top we get 1 over 2 to the m plus n and the bottom we get as 1 over 2 to the m. So, we work this computation out you will get the same as 1 over 2 to the n. So, what have we observed? What do we observe?

(Refer Slide Time: 34:59)





So, let us just what do we observe we go to the next page so what we observed is that so let us say let m bigger than equal to 1 n between equal to 1 we observe that the chance that X is bigger than m plus m given X bigger than m is the same as 1 over 2 to the n N but this we know from the same computation is the same as the chance that X is bigger than n.

So, this is the key memoryless property so for example in a geometric you wait till m tosses you wait till end tosses and then you look at that you are given there are no heads here, so there are no heads given some information that there are no heads happening then you go and see fine in the next n ones that is if I wait till next n tosses I want no heads again. So, it means in the m plus this is the nth trial I have no hands and then this is the m plus nth trial and I have come to here let me write this in black again.

So, whatever what have I seen here so if I know that there are no heads in the first n tosses and I want to observe there are no heads in the further n tosses that is just the same as just forgetting about this part so you just can just forget about this part this does not matter this does not really matter, so all you have to look at is that look at the next n tosses are fresh and just look at the fact that there are no heads. That is all you have to look at so let me write it down what I have to look at is no heads in n tosses.

So, what I am trying to say I will write a little bit more in more picture so on this side I have the chance that the probability of no heads in n tosses so that was 1 over 2 to the n on this side and on this side what do I have? I have the chance that I want to look at the fact that X has no heads in the next n tosses given that in the first m I have no heads. So, in some sense that if I just write properly write this as in blue X is m and X is n so idea is that it does not depend on what happened the first time it only depends on what happens the next time this is called memory less, it has no memory.

So, this is something very crucial that we will develop in a let me put a switch preview again let me see if that works now let us see hopefully it works I have to go again and do this whole thing so that is what our standard last computation here make it small bigger perhaps, so that is what I was trying to complete that is the joint distribution the memoryless properties if I know X is geometric I can do this computations that I can observe that X bigger than m plus n given X greater than n is the same as X bigger than m, here I think I thought this below so let me make a little bit smaller same as X bigger than n and the idea was that gives you this sort of very quantifiable idea of memory less but the geometry random variable does not keep memory.

If I have no head still m tosses the future is the same as the past, so that means how do you write this down in a probabilistic way you can write down as X given X is bigger than n m if I use a distribution that is the same as let me write this down as in blue so that it is clear, so in if you write it in a little fancy notation all I can say is X given the event X bigger than m is also distributed as geometric is also distributed as so I mean I use this notation for distribution it is also distributed as geometry class.

That is all. Let us take some shortcut there is a little bit of pitch of salt in the sense that this random variable takes values and m plus n to infinity but (())(39:51) this is not quite geometric half of the geometry half takes values from 1 to infinity this takes values from n plus 1, n plus 2, n plus 3 and so on.