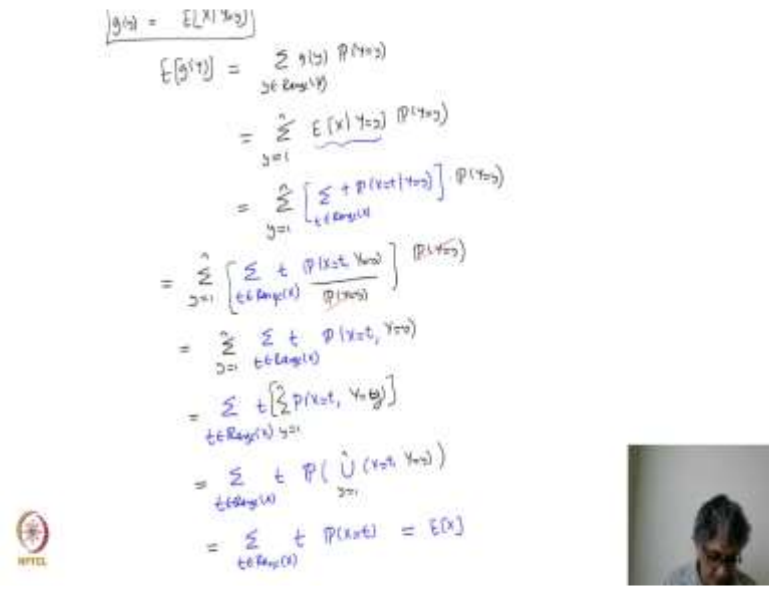
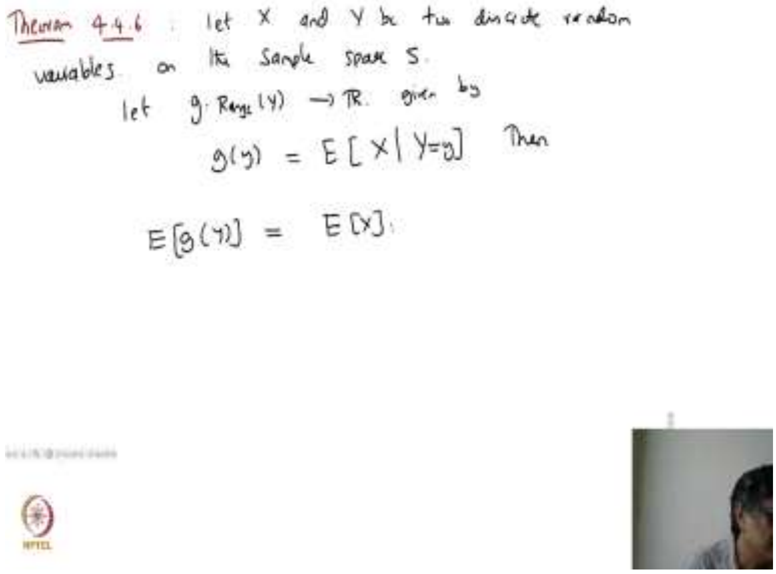
Introduction to Probability – With Examples Using R
Professor. Siva Atherya
Theoretical Statistics & Mathematics Division
Indian Statistical Institute, Bangalore
Lecture No. 26
Conditional Expectation and Covariance

(Refer Slide Time: 00:16)



So, now in the last class we were doing this idea that if I if I take E of X given Y equal to Y and I average it out then I get essentially that expectation of X. So, so let me just formally write it down of the theorem so that we understand it properly. So, let me do the next page.

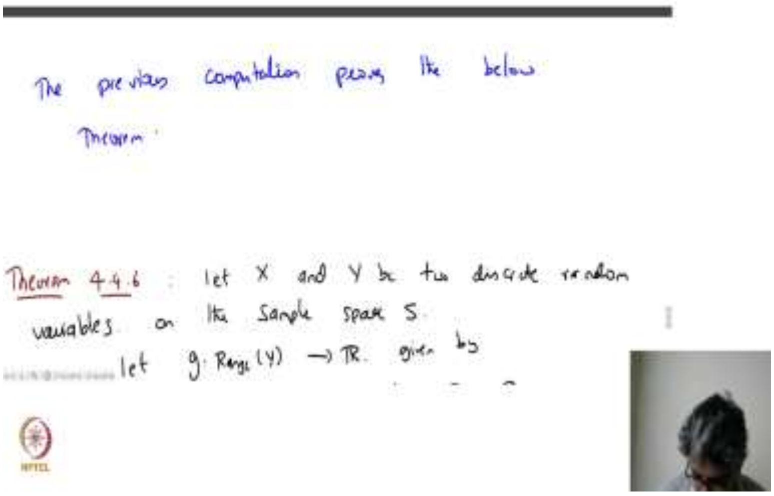
(Refer Slide Time: 00:41)



So, here is the theorem. Theorem is 4.4.6. So, let X and Y be two discrete random variables random variables. Then on the same sample space as S so it is important. And what you do is let us say Y g is a function from the range of Y to real line. Defined by let g be a function given by g of Y is the conditional expectation of X given Y equal to 1. Then what we have shown the previous computation is E of g of Y that is this E is over the exponential Y is going to be just E of X. So, this is a very useful tool.

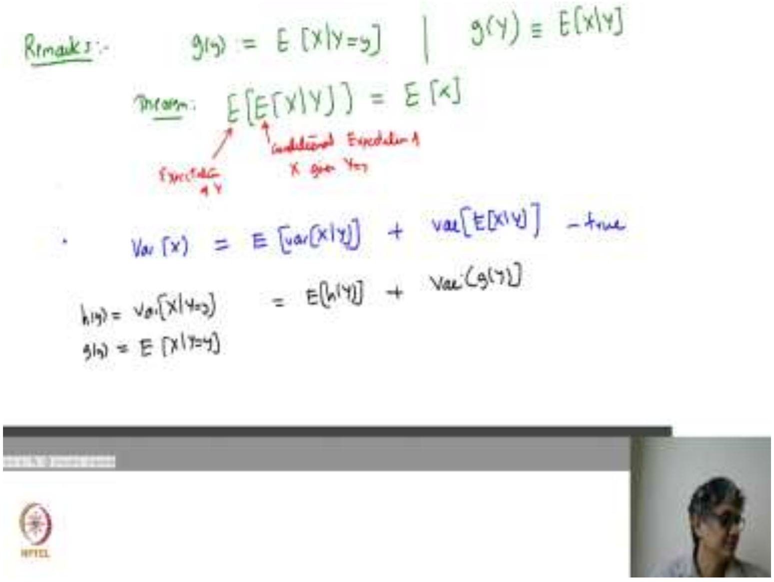
So, one this is what we showed in the previous in this previous in this previous page that is exactly. What we should do if I take the average value of g of Y with all possible value of Y I will get E of X by this. First I write down definition of E of g of Y then I write the conditional mean what means and then I just identify the fact that if you expand it out the Y average is out and you just get E of X. So, when in probability they they cannot or in statistics they do this I find it a little confusing but I will just tell you this. So, what they do is this theorem so this is what we showed in the previous class about.

(Refer Slide Time: 03:11)



So, the previous computation proves the below theorem. So, now the the couple of things before I sort of wind up this whole thing so one extension is the formula so this is one remark.

(Refer Slide Time: 03:36)



So, let us say one remark. So, one remark is the following that so usually you think of g of little y is E of X given Y equal to y that is just a definition of the function. So, one notation people use this if you think of g of capital Y it is a random variable you denote it by E of X given Y. So, that is this is one thing people do a lot as g of X Y. If you want to denote g of Y you think of it as E of X given random 1.

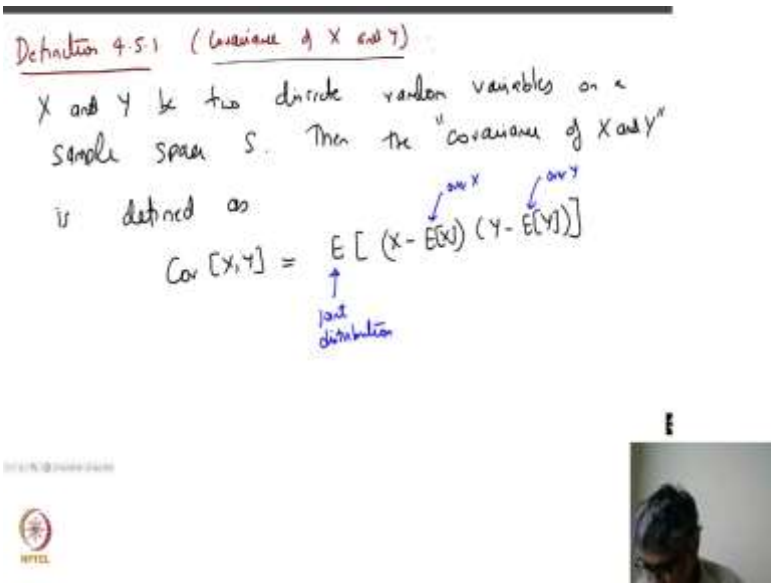
Then what the theorem shows is that the theorem shows is if I take E of X given Y equal to Y that is the same as saying if I take the random E of X given Y that is g of capital Y I average it out I get E of X. That is another way of thinking about the theorem. Of course, this is you might be a little careful this E is is the conditional expectation of X given Y and this E is the is the 1 over Y.

This is the short form of the theorem but if this confuses you then I would just allow you to do the theorem extension. There is sort of an extension to the variance but I will do that a little later. So, let me just take it here without without proof so the claim for the variance is also true holds. So, you can also do the following that the variance of X is something that I will not show but I will allow you to sort of check it out. Is essentially, again going to be the expected value of variance of X given Y plus the variance of expected value of X given Y. So, this is something that is that is also doable as a claim it is also true.

So, here the the way to understand this is that this is so let us call it h of y is the variance of X given Y equal to y we compute that first. Here again you define that g of y as the expected value of X given Y equal to y. Then you understand this formula as the expected value of h of capital Y plus the expected value plus the variance of g of capital Y. That is how you understand the formula.

But, this can also be shown I will not do it in class but maybe I will give it as a assignment. This is a very very very powerful formula in statistics the analysis of variance formula that is the variance of X is the average of the conditional variance plus variance of the conditional average useful formula (())(07:19). So, now we have seen one level of dependence between X and Y so now I want to sort of understand precisely. So, I know how if I give you Y I know how it affects X in terms of the conditional expression. What I want to do is I just want to understand if I can summarize this how in average the terms how X affects Y. So, that is called covariance. So, I will just write that down the definition.

(Refer Slide Time: 08:04)



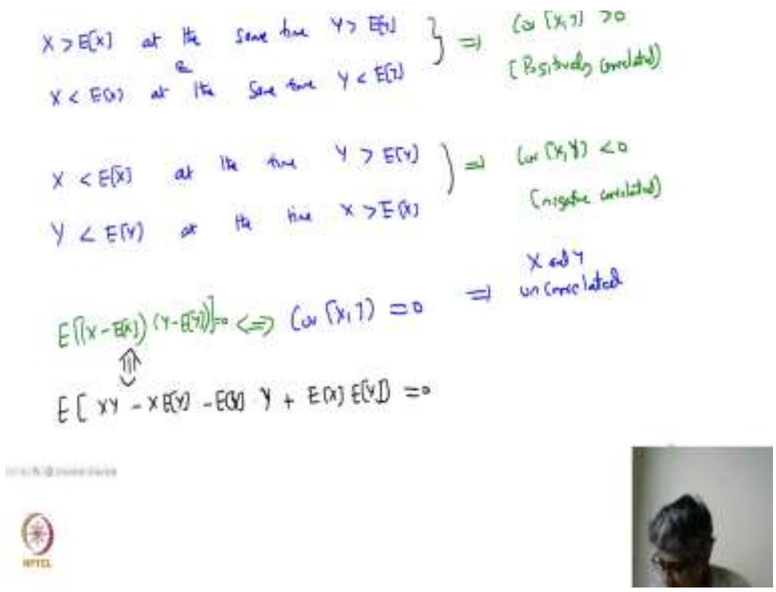
So, here is the here is definition of 4.5.1. So, just as we understood X value of X and standard deviation for a single random variable. So, covariance so does it for two variables X and Y. So, so this this this sort of signifies the variation among X and Y together at the same time. So, here is it let us say X and Y be two discrete random variables. Two discrete random variables, variables on the same sample space S on the same sample space S is S. Then the covariance of X and Y is defined as so here is the covariance of X comma Y that is just.

So, what I do is if I just a if I just the variance I would do E of X minus E of X the whole square I would just take E of X X minus E of X times the Y minus E of Y. And this expectation is the joint one. So, this is the joint expectations and this is just over the over X and this is just over Y. That is the of course this is this is again the same way you have this random variable so this the sum can be finite infinite or does not converge. So, what does it tell you so if you look at a little bit if you start that form a little bit what does this formula tell you?

So, this formula tells you that so you look at X minus E of X and Y minus E of Y. So, X is suppose X and X is larger than E of X at the same time that Y is larger than E of Y then the covariance will be larger. And so in some sense the the or let me put I do not mean to be larger it just may be positive or negative. So, if X is larger than E of X at the same time that Y is larger than E of Y, then the plot is going to be positive. And if X is smaller than E of X at the same time Y is smaller than E of Y then also plot is going to be positive.

So, covariance positive means that both behave the same way. But, the reverse is true. Let us say suppose when X is smaller than E of X Y is larger than E of Y then the product will be negative and vice versa. So, it means covariance kind of measures typically how X and Y both behave with respect to them. So, let me just write this down properly.

(Refer Slide Time: 11:48)



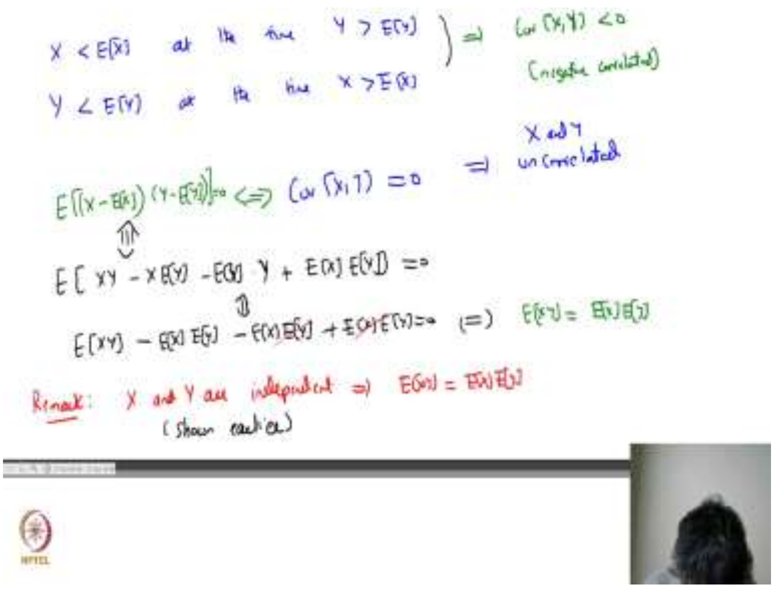
So, suppose X is bigger than E of X at the same time Y is bigger than E of Y. And let us say X is less than E of X at the same time X is less than Y is less than E of Y. That means they both behave the same way the same sample problems. So, this will imply that the covariance of X Y is positive. That means X is larger than E of X at the same time Y larger E of X. Suppose, the reverse is true that means if if X is less than E of X at the time Y is above E of Y and vice versa Y is less than E of Y at the time X is bigger than E of X.

Suppose, this is true in a schematic way sample by sample point. Then this implies that in some sense that the covariance would be negative. So, it means the random variables behave in some opposite fashion. So, this in this in in probability this implies that this implies that covariance of X Y is negative. So, now this implies that so typically in statistics you tell them that these are positively correlated. If the covariance is positive you say they are positively correlated. If they are negative you say they are negatively correlated.

And this comes to the last one if they flip flop enough that that that covariance of X Y is 0. So, covariance of X Y 0. You know then you say that the random variables are not correlated. So, they are uncorrelated. Say X and Y are uncorrelated. And this you should be a little careful we already seen an idea before so this is the same as by definition is the same as E of X Y. So, let us list this little computation the same as E of X minus E of X times Y minus E of Y is 0 that is by definition.

And if you work it out that is the same as E of if you multiply it out what I get I get E of X Y minus X times E of Y minus E of X times Y times Y and minus E of X times plus E of X times E of Y is equal to 0.

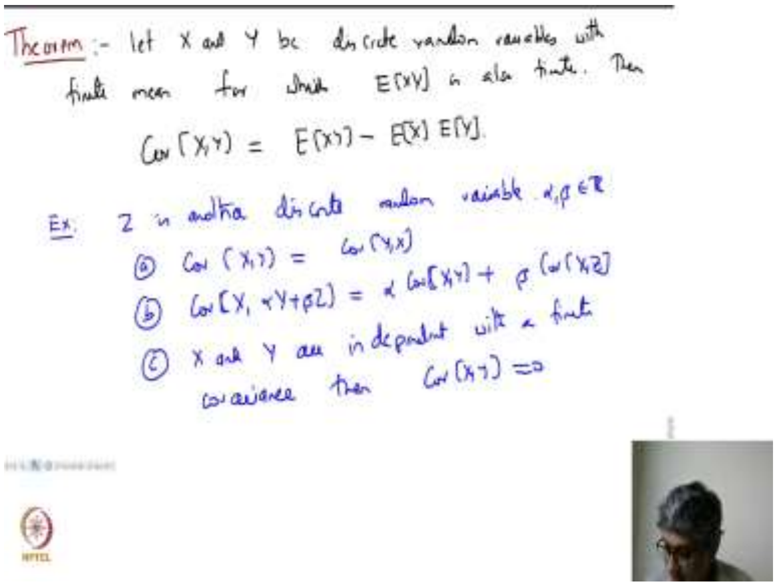
(Refer Slide Time: 15:11)



Now, if I just keep going a little bit I use expectation as linear. So, the first guy gives me E of X Y second guy gives me E of X times E of Y because E of Y is a constant. The third guy gives me E of X the constant and then I have an E of Y and the last guy is both the constant so I get E of X times E of Y is equal to 0. And that gives me the fact that if I cancel off one guy with this guy and I just get all I get is that that is the same as E of X Y is E of X times E of Y.

So, random variable are uncorrelated if E of X Y is equal to E of X times E of Y is 0. You have seen this expression before if you think a little bit it is just we shown before long time back that if X and Y are independent then this implies that E of X Y is is equal to E of X times E of Y. The converse is not true so the converse is just saying that E of X Y is equal to E of X times E of Y the random variables are uncorrelated that is all. That is something we should keep in mind this we had shown earlier. Yes, so here is a let me just summarize these ideas into one theorem so that we sort of understand everything together. So, here is the theorem that I will I will leave as an exercise to prove.

(Refer Slide Time: 17:24)



So, theorem is the following that let X and Y let X and Y be discrete random variables. So, here in the above computation I assume they are finite mean type. So, that finite mean for which E of X Y is finite is also finite, then the covariance of X Y is going to be equal to E of X Y minus E of X that is what we showed this is what we have shown. And the next thing we showed was that we can choose to show the following this I leave the exercises.

Let us say Z is another discrete random variable random variable. Then we have covariance of X first thing we know is covariance of X Y is let us say A comma B R covariance of X Y is the same as covariance of Y X that is the same that is easy to observe. And the part is that so here A and B actually alphabet. Covariance of X and alpha Y plus beta Z is the same as alpha times the covariance of X Y plus the beta times covariance of X Z. So, this one can show easily from definition.

And the idea is that this also we have seen is that if X and Y are independent independent with the finite covariance then the covariance of X comma Y is 0. That also we have seen.

So, let me just do a review and sort of summarize the discussion a little bit so let me go back up. So, what I wanted to say was that for a single random variable you had mean and standard deviation. So, that that gave you one understands the two numbers it gave you some understanding of whether random was centered at and how the spread was.

Two random variables you can define what is called a covariance. The covariance of X comma Y is the average joint average of X minus E of X times Y minus E of Y. So, now the what you observe is that if X is large and same time Y also is large and vice versa then the covariance tends to be positive. If X is smaller than E of Y E of X and Y is larger than E of X same of E of Y same time and reverse then the covariance will become negative. Because, the plot is negative.

And then we observe that we call it uncorrelated if covariance of X and Y are 0. But, then in this computation on the bottom side of the left hand bottom part of the left hand side we observe that covariance of X Y is equal to 0 means that if you do the algebra properly you get E of X Y is same as E of X times E of Y. And you recall earlier we had shown that X and Y are independent E of X Y is E of X times E of Y.

So, once we do this whole, whole scenario I know that the following fact that if X and Y are random variables first thing I know is that covariance of X Y is just E X Y minus E of X times E of Y. That is just by definition and algebra. Then if I have another number with Z and alpha and beta are real numbers the first thing I observe is the covariance X Y is same as covariance of Y X is symmetric that is easy to see for the definition. The other one is that it is linear in the in the covariance. The third is that if they are independent then the covariance is 0 and converse is not true, just be careful.