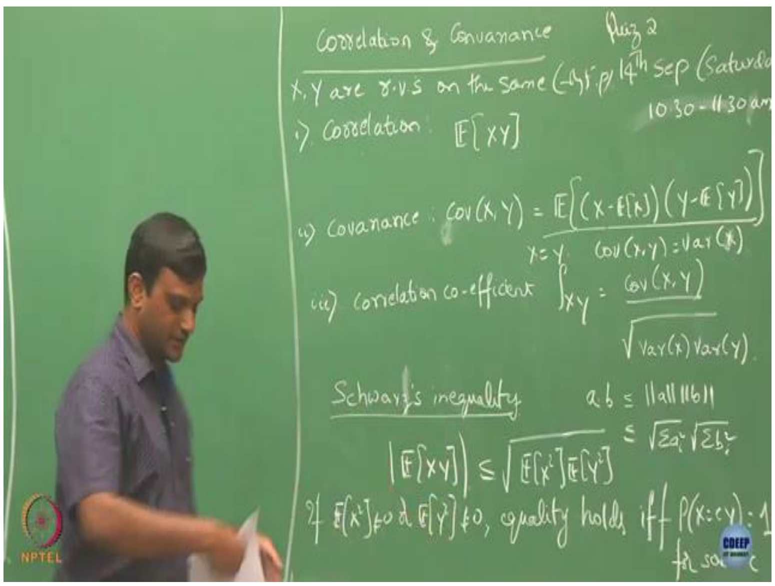
Introduction to Stochastic Processes
Professor Manjesh Hanawal
Department of Industrial Engineering & Operational Research
Indian Institute of Technology, Bombay
Lecture No. 13
Correlation and Covariance

So, today moving on that, we will further focus on some more properties of this set of random of variables.

(Refer Slide Time: 0:33)



So, today we will discuss the notion called Covariance. So, when we have a set of random variable, we know how to characterize the distributions. We already said that we are going to characterize that through that joint distributions, and we also want to understand some of these properties of this joint distribution. For example, if you recall for a single kind of variable, we had defined the notion of expectation, and variance. But, when we have set of random variables, what kind of properties that we would be interested in?

One thing that easily comes to mind is we have to basically understand, how together they behave like both of them are moving, suppose you have two random variable X and Y, you want to understand whether both of them are taking in the values in the same direction or they take value in opposite direction, or some things of that nature. So, let see if we can try to capture that through some of this notions.

We are going to call correlation. So, let us say, we have let us say X, and Y are random variables, on the same probability space, we are going to define correlation as simply expectation of the product. Then we are going to define their covariance as, and the third thing I want to define as correlation coefficient.

So, also like I am inherently assuming that this random variable X, and Y are such that their second moments are finite. So, that their variance is also finite. So, if the variance is not finite then this relation here, this definition is not so clear, I mean not well defined. So, that does not make sense, so that is why we just assume that their second moments or the variances are finite.

So, if you have two random variables, you can always try to characterize these values. If I already know their joint distributions. If I know the joint to be distribution I can just find the product and the (corres) using the associated joint distributions I can compute all this things.

Now, something that immediately comes to mind is suppose X is same as Y, or let us say X and Y. So, Y is just X. So, this is going to be what? Correlation in this case, expectation of XY that is basically second moment. And now, in this case if Y is same as X, what is this going to be?

Students: (())(4:57)

Professor: And what is that? So, if X is equal to Y, you know this is nothing but covariance of, X, Y is nothing but variance of X. So, again like, we are just like generalizing these notions when we have multiple set of random variables. Now, to understand bit more about these properties, we have this relation what is called as Schwarz's inequality, you people know what is the Schwarz's inequality on a set of, on a pair of vectors. So, let us say a is a vector, b is a vector, their inner product is going to be?

Student is answering: (())(6:06)

Professor: Form of a vector and, is there a square root? Their or No? No square root. So, if I am going to write it as this is going to be what? This is going to be summation a i square and this is going to be summation b i square. So, what is a is the nothing but a vector with components a i similarly. So, there is an analog portion of this for our vectors also.

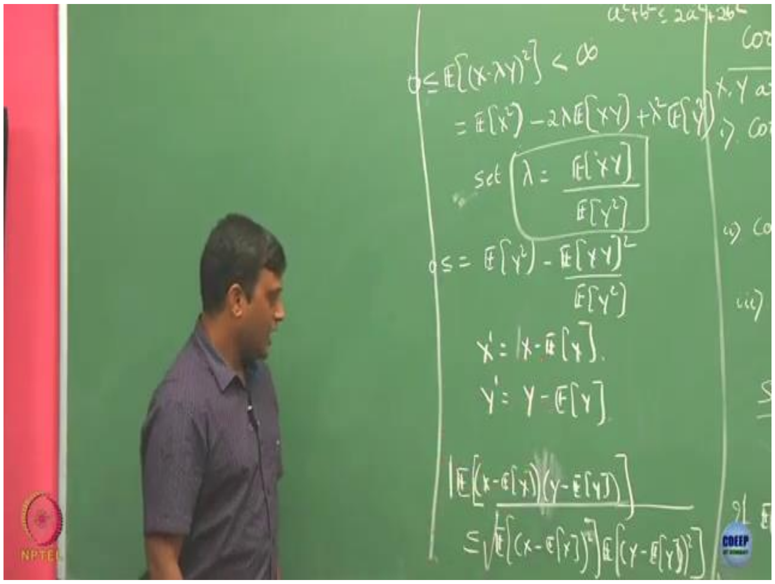
So, that says if you are going to look at the expectation of this product then this is going to be upper bonded expectation of X square Y square. So, if you look at the correlation of the random variables X, Y and if you look at their absolute value that is going to be upper bonded in this session. And further, thus, F can show that equality holds if and only if.

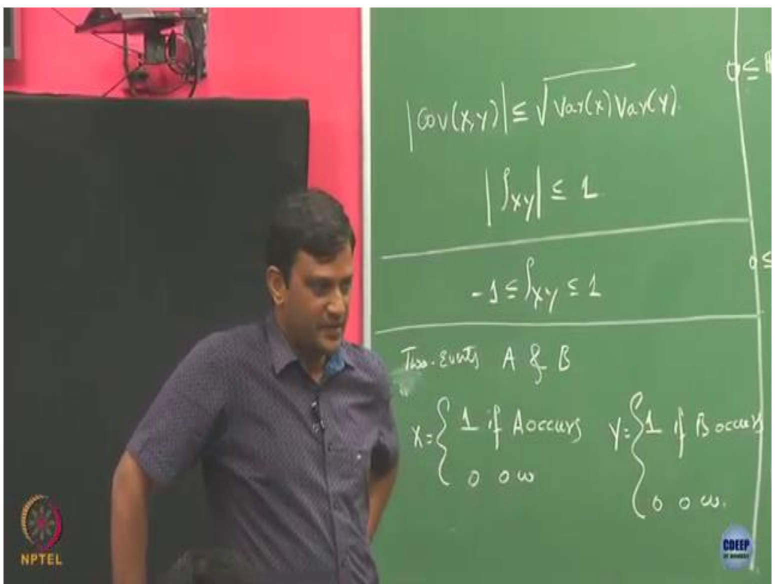
So, (())(7:53) their name directly indicate what they are, at least this meaning of correlation is clear like what is, we are calling it correlation and what we are doing is taking their product and taking their expectation. How they are correlated? And similarly, covariance, this is, at the first time we can take it as an extensions of other variance definition, variance we defined for single random variable, but now when we have two random variable and if you want to centralize them by removing the mean and then look at their correlation.

So, for example, if this covariance is what? This is nothing but correlations of this centered random variables. Where we have already removed the mean values. So, we will try to interpret what this correlation coefficient means in a moment. So, for that we need to understand, we need to have these bonds, the Schwarz's inequality.

The Schwarz's inequality says this we can have a upper bond like this and suppose at least one of this random variable's second moment is not zero, then their inequality holds if and only if one random variable can be expressed as a linear, is just a scaled version of the other random variable with probability 1, for some C, I do not know whatever that C, but as long as random variable Y can be expressed, so X can be expressed as Y by scaling it with some C. This is, if this holds then it is fine. Now, a quick look into why this relation should hold.

(Refer Slide Time: 9:46)





So, let us the take a case of, let us try to understand this. So, this can be expressed as what? Expectation of X square minus 2 lambda expectation of X Y plus lambda square expectation of Y square. I just expanded it. So, just check that if both X and Y they have finite second moment, this quantity is always going to be bounded and further this quantity is into greater than or equals to zero. It is going to be greater than or equal to zero, it is clear because we are scoring and taking expectation.

But, this is also be going to true because you can always bond that like, for any a and b this relation holds because of that you just expand this and then apply this result in terms and we know the second moments are fine, you can also argue that this is going to be finite. So, just verify that and now this is true irrespective of what lambda I choose, as long as lambda is finite, this relation, the second relation is true whatever lambda I am going to set.

Now, sets specifically, expectation of X Y divided by expectation Y square and now if you are going to simplify, so for time being assume that expectation of Y square is not zero and then I can, then this is defined properly, then I am going take that value as lambda, if I am going to take that lambda and plug in here, this is going to be equal to expectation of X square minus expectation of X Y divided by expectation of Y square, just by the plugging that lambda value and just simplifying you get it. And this, we already know this is going to be greater than or equals to zero. Now, if you just going to.

Student: (())(12:45) equals to 0.

Professor: So, I am saying that at least assume one of them is not equal to zero. I will just first consider a case expectation of Y square is not equal to zero and then define lambda in that case like this and then plug in. If you want to consider the case expectation of X is not equals to zero then just do the opposite, you just take Y minus lambda X and then define it, then replace expectation of Y square by X square, you can do similarly.

Now, if you just do this cross multiplication here, and simplify you get exactly what is what, what we are claiming as Schwarz's inequality. Now, to look into this, again it is easy to see that if suppose, let us say first assume the case X is equal to C Y for some C. If you now just go back and plug Y equals to C X here, you see that both right hand side and left hand side match, in that case they hold with equality.

Now, assume that this relation holds with equality then we can see that now you have to come up what is the value of C that works out here, in this case you just going to take that C to be this value lambda whatever you have. In that case also then if the equality holds then that equality satisfied this value of lambda. You can just check that. So, with this we have this relation expectation of this correlation, the mod of this correlation is upper bonded by this quantity.

Now, I have this. Now, replace X Y, define a new random variable. Let us say X prime which is going to be X minus expectation of X and then Y prime equals to Y expectation of Y, do this. Just I have defined two new random variables, and if they have finite second moments X and Y so, thus X prime and Y prime will have finite second moments. Now, you just plug in here, if I just plug in here what I am going to get, expectation of…

So, if you replace X by X prime and Y by Y prime and then here also X get replaced by X prime, Y by Y prime, then what you are going to get is basically, I have just replaced X by X prime and then introduce their definitions we have. So, now if you look into this, what is the quantity on the left hand side I have? This is expectation of centered X and Y. This is what we call, this is according to our definition, this is nothing but covariance.

And what is this quantity over here? Expectation of X minus expectation of X whole square this is nothing but variance of X and then we have variance of Y. So, now with this relation, what we can right away say about this row X Y?

Student: less than (())(17:27).

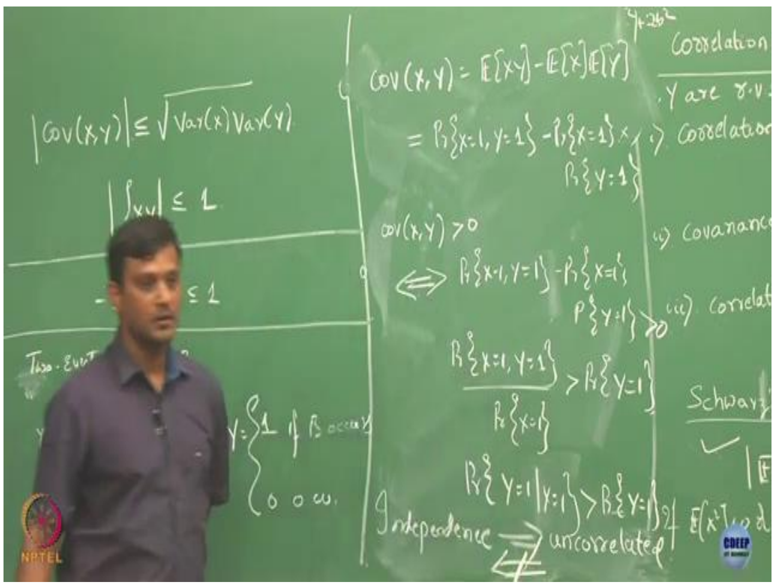
Professor: So, I know that rho X Y is going to be less than or equals to 1. Now, let us try to understand, what is this correlation coefficient means? So, what we are trying to, like as I said we have a two random variables in this case and we have all these properties, what we want to understand is how their behavior like is there anything we can infer when we look them jointly.

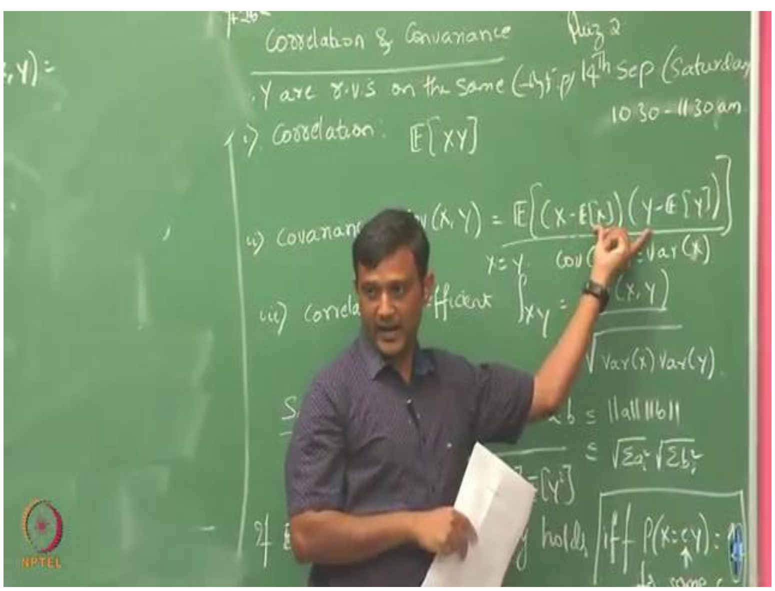
So, as of now our correlation coefficient the way we have defined we have shown that. That is going to be less than or equals to 1 and now let us understand, so this is like mod here because I had a mod here and so this is what I am saying is absolute value of rhos X Y is less than or equals to 1 that means my rho X Y will be between 1 and minus 1. Suppose, if I find out that this correlation coefficient happens to the more than, it is going to be less than 1 but also know that that it is going to be more than 0 that is my rho X Y happens to be positive. What does that mean?

So, let us take a simple, I am going to take two events A and B, two events that are coming from my script F. Now, I am going to define a random variable X which is going to be 1 if A occurs 0 otherwise. I can define a random variable like this, whenever it happens I am going to take the outcome to be 1, if anything other than this happens I simply say 0 and I am going to define another random variable if B occurs and 0 otherwise.

Now, I want to understand suppose is there anything like if B happens, already happened, subsequently or whatever like I want to understand given that B has happened does it have any bearing on happening of A, does it make happening of A more likely or less likely? Or that what kind of information it reveals to me?

(Refer Slide Time: 20:35)





So, let us try to understand by evaluating this correlation coefficient here. So, covariance of variance of X Y, I have written the formula here, but if you are just going to cross multiply the terms within the square bracket and expand them it simplifies to simply, just simplify, so just do the multiplication and expand. Now, I am going to take this, I am going to apply this formula on these two specific random variable I have defined here. What are the possible values the product X Y can take?

Student: 1, 0.

Professor: It can only take 1 or 0, it is going to take 1, 1.

Student: (())(21:28)

Professor: 1 occurs and then what is going to be the expected value of X and Y?

Student: Probability of 1 by 2.

Professor: Probability that X is equal to 1 and that is the only case when the problem is one in all the cases rather it is going to be 0 that is why the expectation of X Y is nothing but this probability into 1, but I am skipping that 1 here. Now, what about this? What is the expectation of X? This is going to be simply X equals to 1 and Y equals to 1, because I am talking about expectation. So, how you are going to compute the expectation? Expectation is value of this and the corresponding probability.

Student: Probability we have written.

Professor: Probability we have written the associated value of this random variable with this probability is 1, that is exactly the expectation value all other terms are going to be 0. Now, suppose, let us say assume, assume the covariance is going to be positive. If this is positive, let us say, if and only if because this is an equality, this guy is going to be positive and this is positive and this guy is positive implies and now focus on this part.

So, this is nothing but probability that X equals to 1, Y equals to 1, probability what I want here is X equals to 1 is greater than probability that Y equals to 1. And now if you are apply the conditional probability definition here, what is this? Y given X. Now, go back and, let us go back and plug in the meaning of this, X equals to 1 means A occurs, Y equals to 1 means B occurs.

Now, what are they saying is, suppose let us say X equals to 1 that is A has occurred, now what is saying that the probability that B has occurred is more than just like probability of even B itself occurring. That means, if A has, if you know that A has occurred that B is also occurring seems to be higher than just if you are going to ask the question whether B has occurred.

So, what it is telling? The covariance is telling that if the covariance is positive that means one event happening already gives, it says that the other events happening conditioned on this is more likely than the unconditional probability and this is what? If you are going to take this to be positive.

If you are going to take this to be negative, what is that?

Student: (())(25:24)

It is going to be opposite that means condition that A event has occurred, the probability that B is occurring is now going to be less than the unconditional probability of B happening itself. So, inner way what is happening if covariance is positive that means one event happening has increased the likelihood of other event also happening together. So, that is the meaning of covariance here.

Student: (())(25:57) it does not contain equal to sign, it is greater than or equal to…

Professor: Yeah, so, if you want to.

Student: Independent, Y and X are independent.

Professor: So, if X and Y are independent this is going to be 0. So, if independence means.

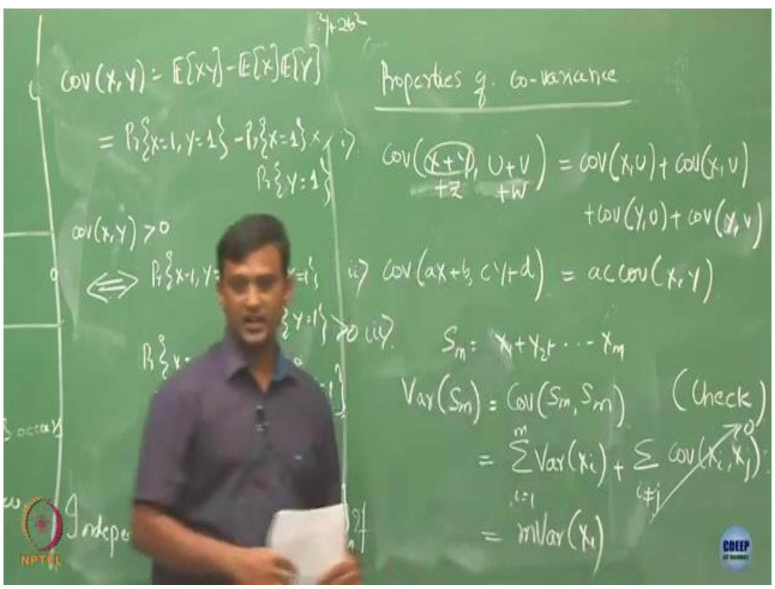
Student: X Y is greater.

Professor: So if it, so, now let us come to the case like equals to 0, covariance X Y equals to 0. If covariance X Y is equals to 0 means what? Expectation of X Y is equals to expectation of X into expectation Y. When these expectation of X Y is going to be equal expectation of X into expectation of Y, when X and Y are independent.

So, one, so in that case you, if X and Y are independent you already understand that happening of 1 will not reveal any information about the other, only when they are other than independent maybe 1 will give you information about happening or not happening about the other thing. So, as you already see, independence implies. So, we are going to say that when the covariance are going to be 0, if X and Y we are going to say X and Y are uncorrelated then their covariance is 0.

So, independence implies uncorrelated, but is, it is in general not true that the other direction is true, uncorrelation, if two random variables are uncorrelated it not mean that they are independent. So, you think about examples where that is going to happen, I will just leave it as an exercise for you.

(Refer Slide Time: 28:06)



We already said that when covariance is 0 we are going to call uncorrelated independence implies uncorrelated and then we already said that when X is equals to Y covariance implies variance of the random variable X. Now, I am going to quickly list some properties you can verify yourself they are just like, say, if you recall when we had expectation of, defined expectation we have defined various properties for expectations. Like expectation is linear, expectation of a random variable if you scale, expectation also just scales, all this properties.

So, similarly, we can write properties of covariance. So, if you have covariance of X plus Y and U plus V. So, now you are looking at two random variables which are themselves expressed as sum of other two random variables. So, this one random variable which is expressed as sum of these two random variables. This can be expressed as, in terms of the covariance of this pair of random variables like this X, U plus covariance of X comma V, plus covariance of Y comma U, plus covariance of X comma… Y comma V. And similarly, covariance of a X plus b C Y plus d.

So, here again I am looking at covariance of two random variables but each random variable is now affine function of another random variable. Now, how, what will be the covariance of this? A covariance of this is going to be simply a into c covariance of X, comma Y. So, it does not matter what is the value of B and D here the intercept values in this affine function did not matter, all you need to do is what value your scaling this value, the constant offset is not going to affect our covariance matrix.

So, often you will be, you will end up with sum of random variables. So, for example, think of 5 courses, and the number you are going to get in each course is like a random variable. Let us call X 1, X 2, X 3, X 4, X 5 and you want to decide what is the total sum of the marks I am going to get across these 5 courses. So, in that case you are going to define some random variable like this let us say X 1 plus X 2 all the way up to Xm and for this you want to basically calculate let say the variance of the sum across 5 courses. So, how you are going to compute this? One thing is, yeah?

Student: (())(31:58).

Professor: Is it? Is it variance of X 1 plus X 2 is equals to variance of X plus variance of X 2? We know that expectation of X 1 plus X 2 is expectation of X 1 plus expectation X 2, but we never said variance is also linear, expectation is a linear operator but not variance. Now, how you are going to calculate the variance of this? We know by definition variance is nothing but covariance of Sm, comma Sn.

Now does this help? If I am going to write like this does this help? Why it helps? Because then I can go and expand this property and if I can expand then I know to only worry about a pair of random variable at any time. So, if this is the case, you can just go and expand all these things, what you will end up is basically variance of Xi I equals to 1 to M then equals to J of Xi and Xj. You can just, I am just like simplifying this after plugging this covariance term with Sm defined like the sum of random variable.

So, here I have defined only for two random variables in the sum, but you can expand if you have more than two here. Suppose, let say you have 3 here, how you are going to apply this formula to get it worked for this sum of 3 random variables? So, you can initially treat these two as one random variable, then we have only sum of two random variable, apply the formula and then expands on this, I mean each pair, you can group so that you are dealing with only sum of two random variables and expand that group.

So, by doing this you will end up with this formula, please verify this. Suppose, I say that the scores we are going to get in this 5 courses are independent. It does not, one course is going to square upon. So, then how does this formula simplifies?

Student: (())(34:55).

Professor: So, then if all of them are independent. This is going to be, so, is it necessary that in this case I need to tell you that the scores are independent or is it sufficient if I say that they are pair wise independent?

Student: Pair wise.

Professor: So, we know about like pair wise independence is a weaker notion than independence across the set of random variable. So, what I need here? I need a independence or just a pair wise independence?

Student: Pair wise independence.

Professor: Pair wise independence will make this term 0, in this case it is simply going to be sum of the variance of each of your random variables. And if I say further that all of them are identically distributed that means they have the same X but they are just like multiple copies. If I say they are identically distributed then you can just write it as like m times maybe variance of X 1 because these are all going to be the same value.