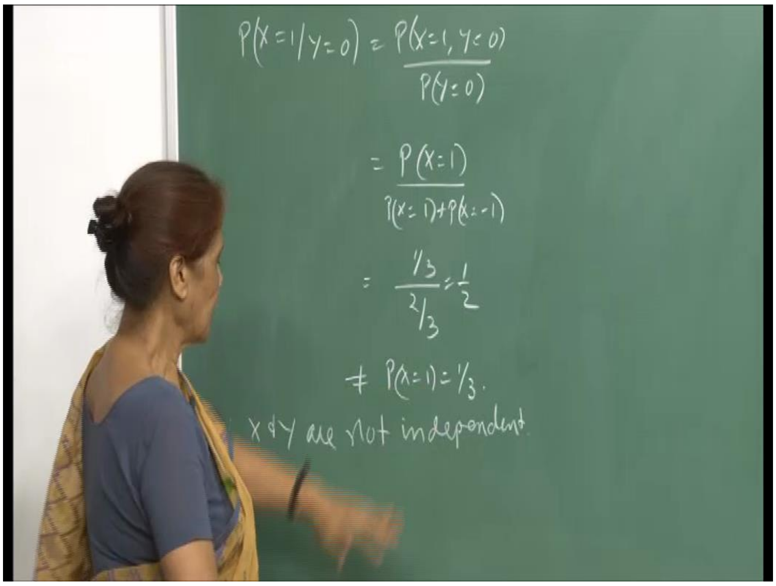
Introduction to Probability Theory and its Applications

Prof. Prabha Sharma
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

Lecture No - 16
Order Statistics Covariance and Correlation

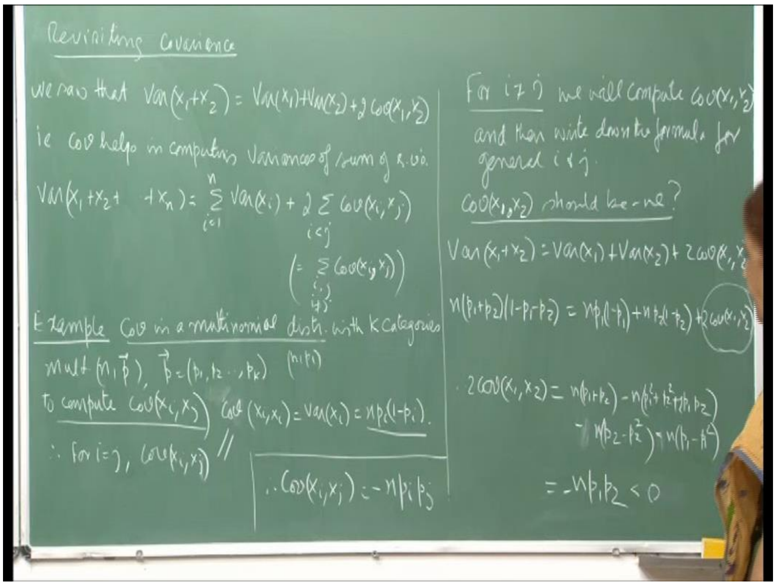
(Refer Slide Time: 00:15)



So, we showed that covariance of x and y is 0, but then and of course by definition also it was clear that x and y are not independent, but we can also show analytically; and that is see just consider this conditional probability x equal to 1 and y equal to 0, given y equal to 0. So, probability, conditional probability x equal to 1, given y equal to 0. So, it will be equal to probability x equal to 1, y equal to 0 divided by probability y 0, y equal to 0, but this is by definition it is equal to probability x equal to 1. And then here it will be probability x equal to 1 plus, probability a cumulative distribution function because probability y is 0; so should be not. Yes. So, this is y 0 when x is not 0. So, x is 0, x is not 0; that means, x is value 1, and x is value minus 1.

So, therefore, this is 1 by 3 divided by 2 by 3 which equals half, but this is not equal to probability x equal to 1. See, if x and y are independent this conditional probability should have been equal to probability x equal to 1; so therefore, x and y are not independent.

(Refer Slide Time: 01:32)



In the last lecture, I defined covariance and then I also defined the correlation, and just tried to show you that correlation is nothing but the dimensionless version of covariance; and now here, a few more examples of or uses of a covariance. So, you saw that when you wrote variance x 1 plus x 2, the formula was variance x 1 plus variance x 2 plus twice covariance (x 1, x 2). So, in general, if x 1 and x 2 are not independent then I afford to compute the variance of x 1 plus x 2, I need covariance (x 1, x 2), right.

And, in general, if you take it, if you take sum of n variables and you want to compute the variance, then by the formula it would be because again the property that we defined for the variance, you can just apply them iteratively; and this will be sigma i varying from 1 to n, variance x i, right. Because it will be covariance of x 1, x 1, x 2 comma x tends to 1 and then the product terms where x i and x j are different; i and j are different.

So, this will be equal to twice sigma covariance (x i, x j) if you put i less than j, because remember for the covariance also we said that covariance (x i, x j) is the same as covariance (x j, x i). So, if you impose this, i less than j, then it will become twice because covariance (x 2, x 1) I will write as covariance (x 1, x 2). So, then it will become twice this; or you can write this as summation i j where of course, you have to say that, i is not equal to j, then it will be covariance. So, whichever formula suits you, you can use that. So, this will be covariance (x i, x j) simply; without the two if you simply summing over all possible values of i and j, so that, i is not equal to j, right.

Now, interesting, again an example to show you that how you can make use of these properties that we have enunciated for covariance. So, consider the multinomial distribution, remember they were n objects and then they were k categories, multinomial distribution with k categories. So, then probability of, success in 1 categories p 1, p 2, and p k; and we have already discussed this distribution and we saw that the probabilities for successes will behave like binomial.

So, for example, in category 1 it will be binomial n p 1, it will be binomial n p 2, and so on. So, this is how you defined the (multinomial distribution, p vector) where p vector is p 1, p 2, p k. So, these are the probabilities of being a particular category which means, success in the first category, so on.

So, now if you want to compute covariance (x i, x j) for any i, j, then for, i equal to j, it will be the covariance (x i, x j) will be covariance (x i, x i), which is we know is variance x i; and since each, x i, is binomial and n p i, x i, is binomial n p i. So, we know from the binomial distribution formula that the variances, n p i into 1 minus p i, right.

Now, for, i not equal to j, we want to compute covariance (x i, x j). So, let just do it for x covariance (x 1, x 2). So, the question asked is what do you expect; should this be negative, covariance (x 1, x 2); see, the idea is that now after having defined correlation also, we see that this measures the linear relationship between; and of course, we still have to talk about kosis law of equality and so on.

So, anyway the expectation is that this will be negative, why? Because, you see if the total number of objects is fixed which is n, right, now if large number of people or objects are in x 1, then accordingly x 2 the number will not be that large. So, there will be negative correlation or negative relationship between x 1 and x 2, and that is what covariance will measure here. So, and we will see after computation that it actually comes out to be a negative number. So, variance (x 1 plus x 2) is variance (x 1) plus variance (x 2) plus twice covariance (x 1, x 2) which we wrote down earlier.

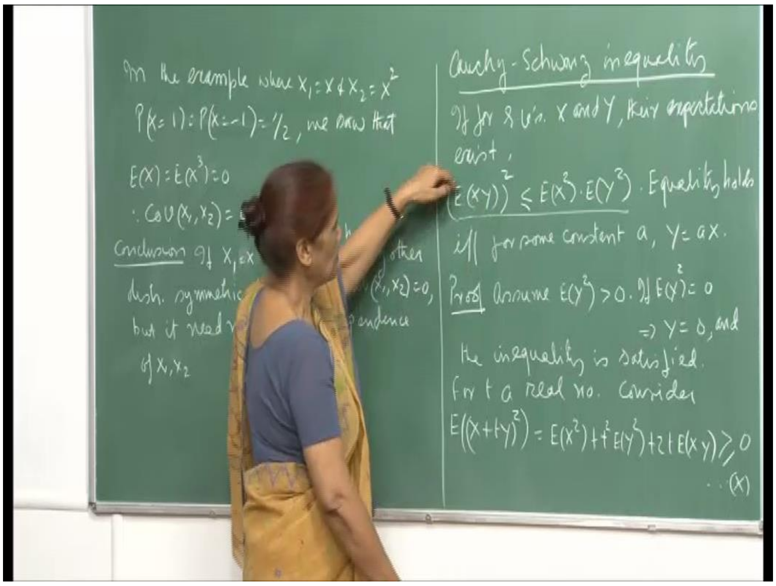
Now, it turns out that we can compute these 3 things because variance x 1 is n p 1 into 1 minus p 1 plus, variance x 2 is n p 2 1 minus p 2; and x 1 plus x 2, again we had seen that when you merging 2 binomials like n p 1 and n p 2 then the sum will behave like a binomial (n, p 1 plus p 2). So, there must have, you must have done some exercises also or you can just sit down and obtain this result for yourself.

So, therefore, variance for x 1 plus x 2 would be n times p 1 plus p 2 into 1 minus p 1 minus p 2; and so from this relationship it is this quantity that we want to compute. So, therefore, 2 covariance (x 1, x 2), will be, just take this to this side, so n p 1 plus p 2, I am just opening up the bracket. So, n (p 1 plus p 2) minus n p 1 plus p 2 whole square which you can write as, I do not why I have written it as. So, this is plus, right; plus 2 p 1 p 2.

And then, you see the terms will cancel out because p 1 square plus p 2 square; you see here this is n p 1 plus p 2 and this is minus n p 1 plus p 2 which cancels out, right; and then plus n p 2 square plus n p 1 square and minus n p 1 square plus p 2 square. So, everything cancels out; you are left with, minus 12 p 1 p 2. So, therefore, this is less than 0 because p 1 p 2 and n all are positive cumulative distribution function 0. So, the covariance; and so once you know that the covariance (x 1, x 2) is minus n p 1 p 2, you can immediately conclude that covariance (x i, x j) will give minus n p i, p j.

So, this was made possible because of this formula which was again written down using covariance. And now for this multinomial distribution you can immediately write down the formula for a covariance (x i, x j) as this, cumulative distribution function; relation coefficient also which will again turn out to be negative. Because the correlation coefficient will be simply this divided by standard deviation of x i, and this will be standard deviation x j, which we already know, right, because it will be n into p i 1 minus p i under root, and this will be n into p j 1 minus p j under root. So, therefore, this computation has become so simple.

(Refer Slide Time: 08:57)



In the example where we took x 1 to be x, and x 2 equal to x square, and I said that probability x equal to 1 is equal to probability x equal to minus 1 is half. Then we saw that expectation of x and it was equal to expectation x cube was 0 and therefore, you could see show that covariance (x 1, x 2) where x 1 is x and x 2 that x square is 0.

Now, actually what you can, you can always show this if whenever x 1 that is x is normal 0 sigma square; that means, expectation of x is 0; and or x 1 has any other distribution which is symmetric about 0 because you saw that x here is symmetric about 0. So, probability (x equal to 1) is equal to probability (x equal to minus 1) is half. So, this is symmetric about 0.

So, now if I take, instead of this if I have taken a distribution of x to be normal 0 sigma square or any other distribution which is symmetric about the origin, then you can show that covariance (x 1, x 2) is 0. So, therefore, you can construct so many examples where covariance is 0, between 2 random variables, but they are not independent because there is a definite quadratic relationship between the 2.

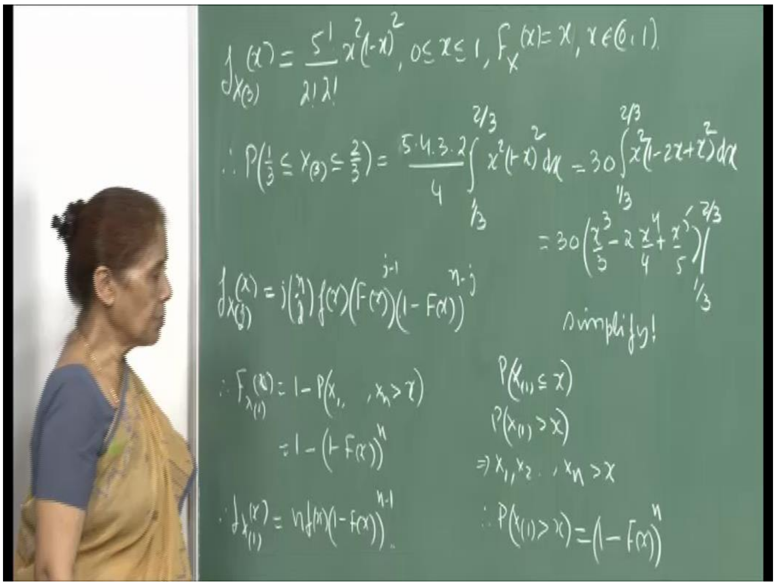
And so knowing one value, knowing value of one you can predict the value of 2, second variable exactly. And therefore, this is what you want to, sort of through this example I thought we can show you and emphasize this fact again that covariance 0 just says that the 2 variables are uncorrelated, but their need, they need not be independence. So, independence goes much deeper than that.

Now, very interesting inquality and very powerful one which we can show, see right now; if their expectations exist then this is the inequality; that means, expectation of x y square, expectation of x y square, sorry, expectation x y whole square 1 minus of 1 minus r 1 t into 1 minus r 2 t. So, the same principle will be used and you can show that expectation y square. And, equality holds, if I know only if for some constant, a, y is a x; that means, there is a linear relationship between x and y. So, if x and y are linearly related then the kosis words inequality would be satisfied as equality, otherwise it will be restrict inequality.

So, now we are, in this we are assuming, let us see, y square, is a positive value random variable. So, expectation y square can be either 0 or positive, right. Now, if it is 0; so we are assuming that expectation y square is positive because if it is 0, so when can expectation y square be 0 because y square again is a positive random variable. So, it will take only positive values, and so when you write the expectation it will be possible values of y square into the probability with which it takes its values.

So, therefore, that will be positive sum. So, that cannot be 0 unless y is 0. So, that is clear. So, therefore, if expectation y square is 0, it will imply that y is a 0 variable, y takes only 0 value, and then this inequality will be satisfied because if y is 0 this is 0 and this is 0. So, both sides you have 0, and so the inequality satisfied is equality. So, therefore, it is safe to assume; I mean, there is no harm in, no loss of generality I take it to be.

(Refer Slide Time: 12:43)



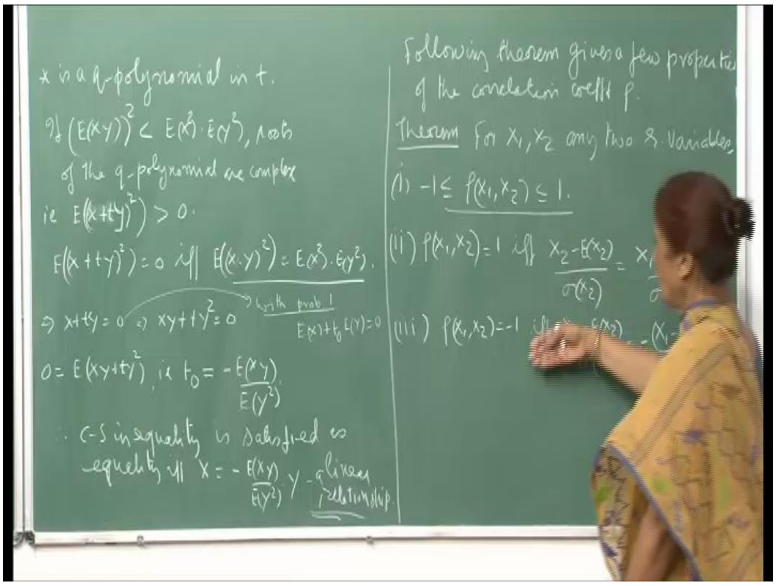
So, f x j x will be j n c 1, f x, F x raise to j minus 1, 1 minus F x raise to n minus j. Now, we can compute the p d F for x 1, the first order statistics independently, and then we can confirm that the, what we have obtained follows by this formula also. So, let us consider the probability of x 1 less than or equal to x. So, then the compliment of this event will be probability x 1 greater than x, right; this is less than or equal to x, so then here it will be, x 1 greater than x, the compliment.

Now, if this smallest statistic, the smallest or the statistic is greater than 1 it implies that all the statistics must be greater than x. So, x 1, x 2, and x n, and all must be greater than x. So, the 2 events are equivalent, and therefore, I can say that the probability x 1 greater than x is equal to 1 minus of F x raise to n, right; because the probability for this is 1 minus or for any x, for any x i greater than x, the probability is 1 minus F x and therefore, since all of them have to be greater than n. So, this is 1 minus F x raise to n, right.

And then, so therefore, I can write the cumulative distribution function for x 1; and this should be, x; F x 1(x) will be 1 minus of this; 1 minus of x 1, x 2, and x n, all greater than x, right. And therefore, this will be 1 minus of 1 minus F x raise to n because this is what you have here; and this is equivalent to, or for x 1 greater than small x, and then implies all these are greater than small x; and therefore, this is the event, right.

And so when you differentiate both the sides you get F x 1 (x) as n; then minus minus becomes plus and the derivative of this is f x, and this is 1 minus F x raise to n minus 1. So, if you substitute, j equal to 1, here this will be 1, and; this should have been n c j. So, n c 1 which is n, then F x; and this of course, j is 1, so 1 minus 1 is 0, no contribution, then 1 minus F x raise to n minus 1; so the 2 match 0, because t is a real number, so therefore, this is satisfied.

(Refer Slide Time: 15:22)



Now, if it is equal to 0 then E x plus t y whole square is equal to 0, if and only if E x y because, that means, that this equation is satisfied as equality, equal to 0. Then the discrepant must be equal to 0. So, this is it, right. So, this is one part. And, now we have to, we want to show that under, for what value of t this will happen. So, you see from here because x plus, expectation of x plus t y whole square is 0, this as we have argued earlier, the random variable itself must be 0 with probability 1, right; x plus t y has to be 0 because otherwise its expectation cannot be 0, right.

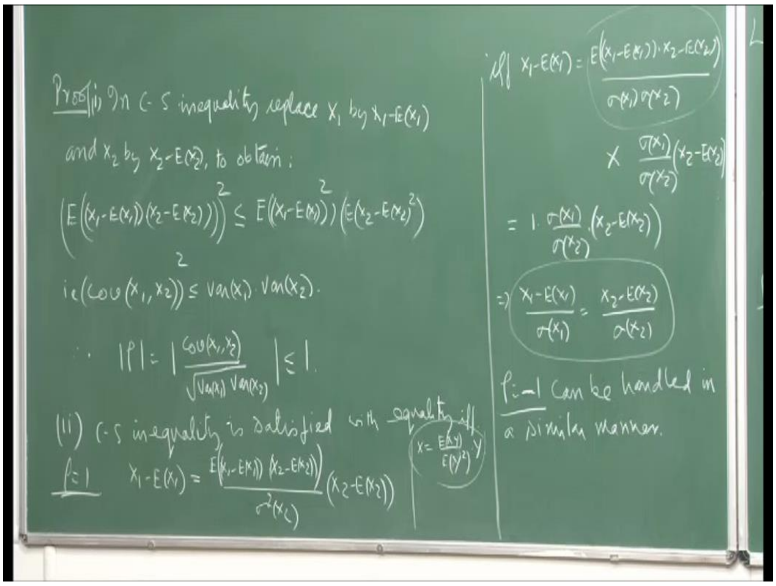
So, now, the thing is that from here itself you can say that we can compute the value of t which makes this happened, right; which makes the discrepant equal to 0. But, as you see, if I do it here, if I take the expectation, here it will be E x plus t; and let us say the value of t naught, t which is a t naught we are looking for. So, this is equal to 0. So, from here we say that y cannot be compute the value of t naught, but you see I cannot guarantee about E y being non 0, right. And so therefore, I cannot compute the value of t naught from here.

So, what we do is if we multiply by this y, then again x y plus t y square is a 0 random variable, right, because this is 0. And so now, if I take the expectation, so this will be 0 equal to expectation of x y plus t y whole square. And so then from here when you are doing expectation inside this will be t naught is expectation x y upon p y square.

So, therefore, kosis words inequality is satisfied as equality, if and only if, right; if and only if x can be written as, this is t naught. So, from, since this is now 0, I have computed the required t naught. So, this will be x is equal to minus E x y upon E y square into y. Now, this is a linear relationship between x and y. So, kosis words inequality is satisfied, whenever x and y are related linearly; and this is the constant which relates x and y, right. And, we will see the implications of this.

Now, using kosis words inequality we can prove following properties of the correlation coefficient rho; and so for any, x 1, x 2, any 2 random variables we will first show that your value after correlation coefficient lies between minus 1 and 1; and this is what we meant by standardization. And, because covariance, the only difference between correlation coefficient and covariance is that you divide covariance by the standard deviations of x 1 and x 2; and then you get a standardized quantity. And so this will be between minus 1 and 1; and if it is 1 then they are positively related for x 1 and x 2; and if it is minus 1 then they are negatively related. So, we will just go through these properties in a few minutes.

(Refer Slide Time: 18:59)



In the kosis words inequality, replace x 1 by x 1 minus expected x 1, and x 2 by x 2 minus expected x 2. So, then the inequality would look like expectation of x 1 minus E x 1 into, sorry, expectation of the product, x 1 minus E x 1 into x 2 minus E x 2; this whole square, is less than or equal to, expectation of x 1 minus E x 1 whole square and expectation x 2 minus E x 2 whole square, right. Because the kosis words inequality we obtained for any 2 random variables x, y.

So, here I can replace the variables x by x 1 minus E x 1 and y by x 2 minus E x 2. So, this is valid, right. And, so which means that, which reduces to covariance (x 1, x 2) whole square is less than or equal to variance x 1 into variance x 2. So, kosis words inequality really simplifies proving these properties of the correlation coefficient. So, but this is nothing but if you divide this by this then it says that covariance (x 1, x 2) whole square divided by variance x 1, variance x 2, is less than or equal to 1.

So, if I take the square root then, a positive past of the square root then this will be less than or equal to 1, right. So, with the, so for absolute value of the correlation coefficient is less than or equal to 1. So, the first property is easily proved using the kosis words inequality. And now, you want to show that if it is satisfied as it quality; and remember in the, when we proved the kosis words of inequality we showed that x will be equal to expected x y upon E y square into y. So, this was it.

Now, here I have replaced x by x 1 minus E x 1 and y by x 2 minus E x 2, and here to so this becomes. So, therefore, this will be because of our transformed variables. This is x 1 minus E x 1 is E of, expected value of x 1 minus E x 1 into x 2 minus E x 2 upon sigma square x 2, and this is x 2 minus E x 2. So, this is the linear relationship between x 1 and x 2. But, this quantity you can see is the covariance, and then you write, you need variance of standard deviation x 1 into standard deviation x 2. So, this will come here.

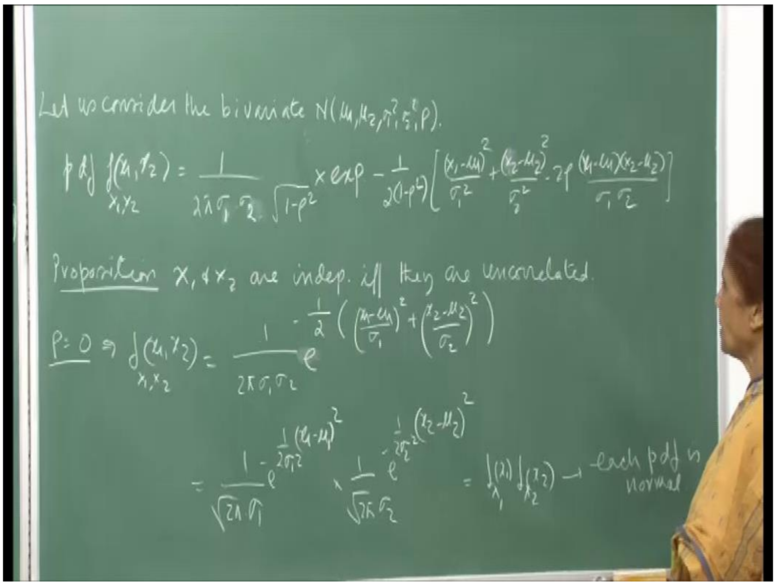
So, you know, just rewrite this expression. So, 1 sigma x square I keep here, the other is here; now here I am dividing by sigma x 1; so I multiply; and therefore, this is what I get. Now, this quantity; I did not say here. So, the 2, part 2 was we had to show that rho is equal to 1. So, we start with this. So, if I start with this then in the kosis words we said that if it is satisfied as quality then we get this relationship which then the minus E x 2.

And so just divide by sigma x 1 here. So, therefore, this would be; so we said that if; so that means, you are just getting the specialized linear relationship; you can write it little differently; the same thing here we can write in this way because we are saying that rho is equal to 1, right. So, then you can predict the actual relationship, till actual linear relationship between x 1 and x 2, if you.

And, similarly if rho is equal to minus 1 then there will be a minus sign here; the same analysis will be done, right. So, this is what we are trying to say is that, you know, your quantity rho is correlation coefficient; is measures the linear relationship nicely; it captures the relationship, linear relationship. But it fails to show you the relationship when it is a quadratic or it is non-linear, and rather I should just say that when the relationship between 2 variables is a non-linear, then it fails to. So, being 0 does not help you, right. So, therefore, that means, you cannot conclude that other variables are independent if the covariance is 0.

So, now let us take this special case, and show you that in, when the 2 variables are normally distributed then you can show that the variables being uncorrelated implies independence, and the other way. Of course, the other way you know, if 2 variables are independent then certainly the correlation coefficient will be 0, but we will show it the other way; that is if the correlation coefficient is 0 then the variables are independent; and this is valid true for a normal distribution only.

(Refer Slide Time: 24:14)



So, here just look at the bivariate normal distribution. So, bivariate normal distribution you have the means are mu 1, mu 2; variances are sigma 1 square, sigma 2 square; and the correlation coefficient is rho. So, the expression for the p d F for a bivariate normal distribution is 1 upon 2 pi sigma 1 sigma 2, under root 1 minus rho square. And therefore, you see this is valid because rho we have just shown is between minus 1 and 1; the absolute value rho is less than or equal to 1.

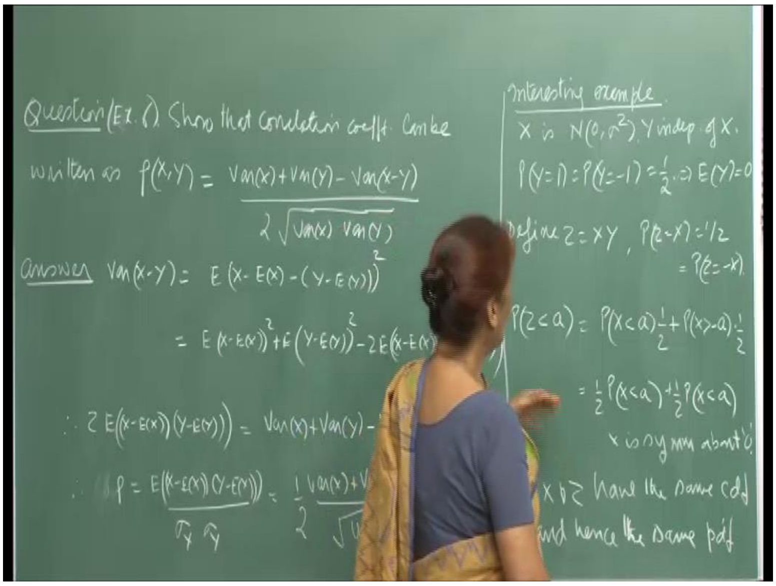
So, and then exponential E raise to minus 1 upon 2, 1 minus rho square; then x 1 minus mu 1 whole square upon sigma 1 square plus, x 2 minus mu 2 whole square upon sigma 2 square minus, 2 rho into the product term divided by sigma 1 sigma 2. So, this is the expression for a bivariate normal distribution, right. So, the proposition is that if x 1 and x 2 are independent or x 1 and x 2 are independent, if and only if they are uncorrelated. So, this is what we can finally establish after giving you so many examples where un-correlation or uncorrelated did not mean independence.

So, if rho is 0 then you can see immediately from here this expression simplifies, this becomes 1, this is also 1. So, it will be 1 upon 2 pi sigma 1 sigma 2; then E raise to minus 1 by 2, right; and x 1 minus mu 1 upon sigma 1 whole square plus, x 2 minus mu 2 upon sigma 2 whole square; this term is not there anymore. So, now you can immediately decompose this E raise to this. So, therefore, you can write this as product of 2.

So, here it will be 1 upon root 2, say, 2 pi i I can write 1 upon root 2 pi sigma 1; then the x 1 term, you know, put together here; and the x 2 term is this. And, you can see that these are 2 p d fs; and each of them say, this is normal mu 1 sigma 1 square, right; p d F separate p d fs, and each is normal. So, therefore, in fact, so much simplification here, right. The moment you say that they are uncorrelated, then they are also independent by our definition, right.

If the product of, if the joint p d F can be written as the product of individual p d fs or the marginal p d fs then it will be said that the variables are independent. So, therefore, and so if and only if part gets proved because rho 0 implies independence, and of course, independence implies that rho is 0. So, therefore, the proposition is established; that is if x 1 and x 2 are independent then they are, if and only if they are uncorrelated, provided x 1 and the joint p d F of x 1 and x 2 is a bivariate normal distribution. So, you can see how we are relating the result that we are getting; and then of course, all these simply, finally gets used, and you know, estimating lot of probabilities that are useful to you.

(Refer Slide Time: 27:41)



So, in the sum I am just discussing the exercises 6 which I will be discussing at the end of this lecture. So, there is a question that I have posed there, and I have asked you to show that correlation coefficient can be written as, rho x y, variance x plus, variance y minus, variance x minus y, upon twice under root of variance x into variance y. So, essentially what I am saying is that the covariance (x, y) can be written as variance x plus, variance y minus, variance x minus y divided by 2, because, this anyway figures in the definition of rho (x, y).

So, this answer is straightforward. You start with variance x minus y, and so that will be x minus E (x) minus, y of minus E (y) whole square, expectation of this whole square, right. And, open up the brackets; so this will be expectation of x minus E (x) whole square plus, expectation y minus E (y) whole square, and minus twice product term expectation of x minus E (x) into y minus E (y), right.

And this can be; so this I can bring to this side; so therefore, immediately you have variance x plus variance y, minus variance x minus y is equal to this; but this is nothing but your covariance. So, in fact, now you can divide this by, rho x rho y, and 2 you can take to the other side. So, it just because see, what happens is that all these different expressions for the same thing that you keep using are handy, sometimes it helps to, because you know these where values, because you, from the known standard distribution of these variables; then you can immediately write down the correlation coefficient.

Now, again see, by theme now here has been to show you as many examples as possible about, you know, the covariance other or the correlation coefficient being 0, but variables are not independent. And, you can see how, you know, contrived them a look at these examples, but they make a point. So, now, here x is normal 0 sigma square, and suppose y is independent of x.

So, x is normal distributed 0 sigma square, and y is independent of x; and the probability y equal to 1, equal to y minus 1, is half. So, therefore, and this imply that your expectation y is 0, right; if probability y equal to 1 and y equal to minus y is half then your expectation y is 0.

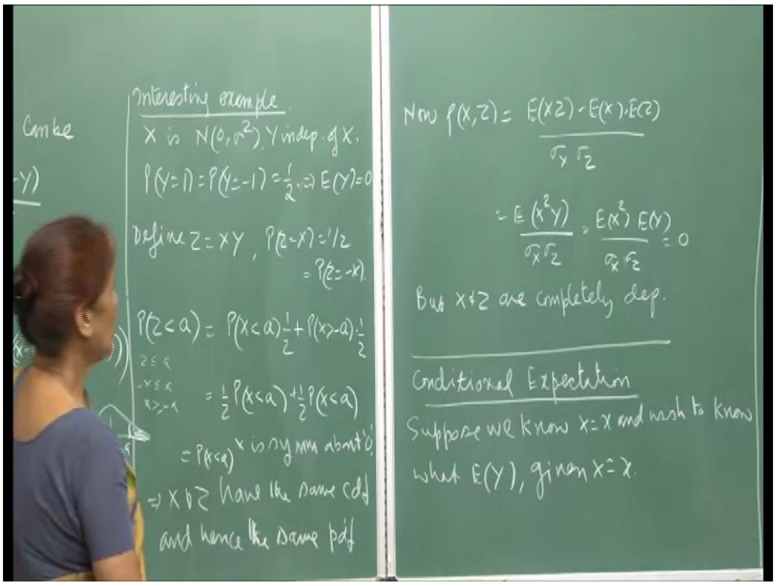
Now, define another variable z which is equal to x, y. So, you see, immediately from here probability z equal to x is half, and probability z equal to minus x is also half because y is either 1 or minus 1, right. Now, if you compute, probability z less than a, then this will be x less than a; and that is with probability half, right; because z is equal to x with probability half, and then x is equal to minus x.

So, if you are writing; so you will be writing, z less than or equal to a, which is minus x less than or equal to a. So, this is equivalent to, x greater than or equal to minus a, right. So, that is what I have written, probability x greater than minus a, into half. But remember, x is normal, normally distributed; and if x is normally distributed it is a symmetric about the origin, and so x less than a, and x greater than minus a, are the same probabilities.

Now, if you can carefully see, you see, in the normal because 0 sigma square. So, therefore, if you take this thing here, so let us say, this is, take a to be positive and that is the same thing; and this is minus a. So, x less than a, is, you see, from the normal thing, this area and this area are the same, right. So, x less than a, is this all probability; and x greater than minus a, is this which are the same, right; because a tails these values are the same; therefore, this area and this area are the same, right.

Therefore, this event is the same as, x less than a; and though therefore, this follows from the x being symmetric about the origin. And so again it is not necessary here that they should be normally distributed because I think anything which is symmetric about the origin would have done the job, right. So, therefore, this is that; and so this is equal to; I should have put it here. So, from here it follows- this is probability x less than a. So, that means, z and a, z and x, have the same cumulative density function from the same c d F which implies that they have the same p d F also. So, x and z have the same c d F, and they have the same p d F, right.

(Refer Slide Time: 32:51)

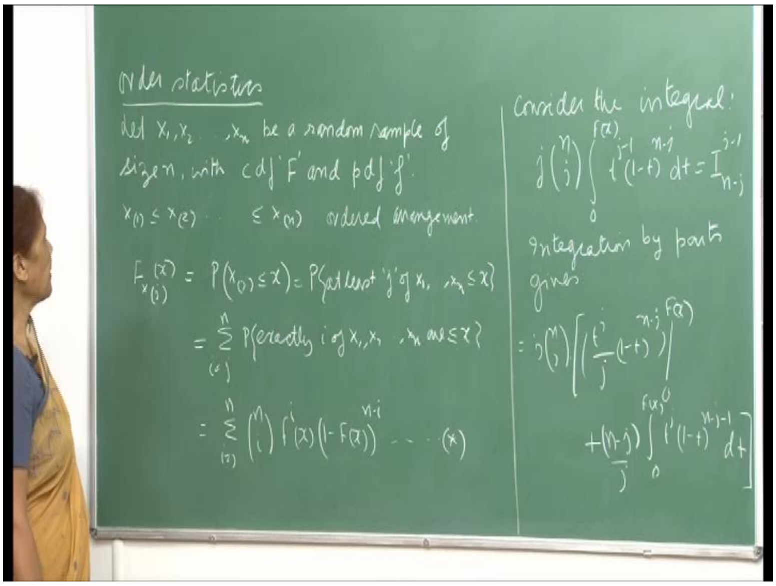


Now, if you compute the correlation coefficient between x and z, that should be expectation (x z) minus E (x) into E (z) upon sigma x into sigma z. But, E (x z) is expectation x square into y, and your E x and E z. So, E x is 0, and therefore, E z will also be 0 because this is normal. So, that means, you need a distribution which is symmetric about the origin. So, therefore, then its expectation will also be 0. So, you therefore, I do not think you need this x to be normally distributed, fine.

So, then this part is 0, and this is E x square y, because x, y; and x and y being independent this is E x square into equal to E (y), right. But, E y is also 0, remember; y is again symmetric; y is 1, and y is minus 1. So, with probability half, both the values have equal probability. So, E (y) is 0. So, E (y) being 0, you get this as 0. So, therefore, the correlation coefficient i 0, but x and z are completely dependent by definition as we saw, right; x and z are completely different- these things because they have the same c d F, they have the same p d f, but still they are uncorrelated.

So, this is; again, you know, I am just, wherever I get these kind of examples I just thought I will bring them to you to show you the, ok. And, so right now we have said reasonably good amount joint probability density functions which was the, so more than 1 variable; then we talked about how we can obtain joint density functions of more than 1 variable.

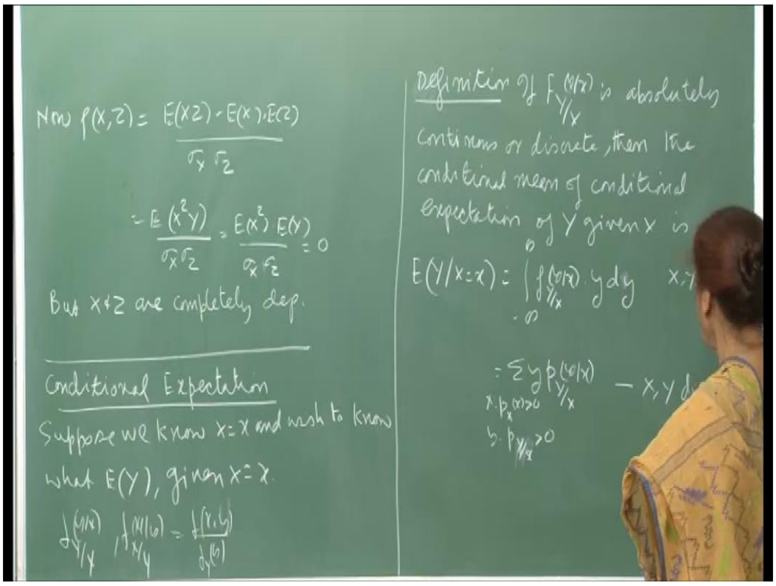
(Refer Slide Time: 34:25)



Now, let me talk of order statistics; so you know further application of the same concept. So, see if you have a sample of size n, random samples, so x 1, x 2, x n, are the observed values; and the c d f, they are coming from the same distribution, so you can say these are also identically independently distributed random variables because it is a random sample. So, c d F is; that means, the cumulative density function is denoted by F, and the probability density function is denoted by small f.

So, when you order the observation, so this will be smallest one. So, therefore, this will be the notation; so x 1 less than or equal to x 2, less than or equal to x n. So, this is the order arrangement of the n sample values that you obtained, ok. Now, so the question arises, can we find of course, one would want to talk about the joint density function of all x 1, x 2, x n, and in particular you would want to find out the density function for the p d f. So, either both of them are continuous or both are discrete, this is when we are defining the conditional expectation.

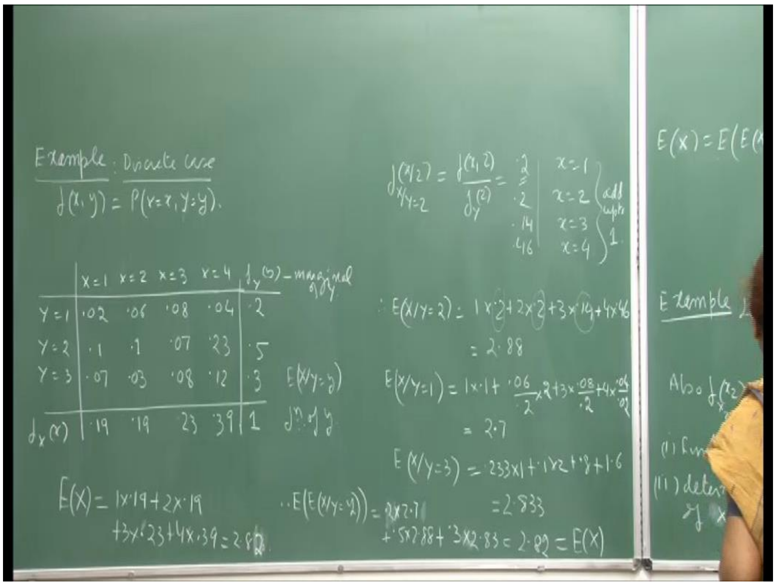
(Refer Slide Time: 35:47)



So, the nature of the 2 variables should be same in the sense that either both are continuous or both are discrete, right. So, then the definition is of course, straightforward because now x is equal to x, so this is fixed; this is given to you. So, now, you have to find the expectation of y, given x equal to x, would be from minus infinity to infinity, y times f conditional distribution of y, given x. So, this would be the definition. This is the case when, x and y, both are continuous.

And, in the discrete case, it will be, the summation will be for all x for which p x, x is greater than 0, because remember this conditional p d F will have p x x in the denominator. So, therefore, we will only consider summation to those x for which this is positive; and then of course, probability y, given x, for all y for which this is positive because otherwise the product will be 0. So, under this condition you can for the discrete case, when x and y are both are discrete, you can define the expected, conditional expectation by this formula.

(Refer Slide Time: 37:00)



So, we will start from the, take this example of a discrete case, where the joint density function is given as probability x equal to x, and y equal to y. So, even from this table you can immediately see, now x equal to 1 and y equal to 1. So, this is the probability. So, you can read the table. So, this is 1 2, 1 3, 1 4, and so on, right. And, you see when you add up these probabilities, they give you what?

They are the values of x equal to 1, x equal to 2, x equal to 3, x equal to 4. So, you immediately get the probability for y equal to 1. So, therefore, when you add up these rows, the numbers give you the marginal p d F of or probability mass function for y, right. So, this point 2 is the probability when y is equal to 1. Similarly, y equal to 2 because the possible values of x are 1, 2, 3, 4.

So, when you add up these probabilities 1 2, 2 2, 3 2, and 4 2, you get the probability of y equal to 2, so that adds upto 0.5. And this is, similarly probability y equal to 3 is 0.3, and these 3 must be add up to 1. Similarly, here when you add up the probabilities of, the conditional probabilities, x equal to 1, and y varies from 1, 2 and 3, they will give you the marginal for x. So, this will be the probability x equal to 1, this will be the probability x equal to 2, x equal to 3, and x equal to 4; and they also add up to 1, right.

So, now from our definition, see I am writing f where it should be ps, but does not matter because the discrete case we are used to habit of writing the p in terms of ps, the probabilities, so does not matter. But, you see now here you can immediately find out probability, conditional probability of x when y is 2. So, conditional probability x 1, y is equal to 2. So, for example, here when you want to compute conditional probability of x equal to 1 given y is equal to 2. So, calculations are simple; y is equal to 2 is given; you are given by this, right; and so conditional probability. So you will divide by a probability y equal to 2 which is divided by the standard deviation of x 1 and standard deviation of x 2; y is equal to 2 is 0.1 divided by 0.5 which turns out to be 0.2.

Similarly, conditional probability of x equal to 2, given that y is equal to 2, will be this joint density function of x equal to 2, y equal to 2, divided by the probability of y equal to 2, which is 0.5. So, again 0.1 upon 0.5 is point 2; and similarly the other 2 computations. And, if you remember, I have not analytically proved it, but we should be able to, maybe that is what we should do next time.

So, here in any case you see these probabilities also add up to 1, as they should because this is now the, you have got the conditional which is also a probability mass function; and therefore, the probabilities here should add up to 1. So, it is 0.4 which is then 0.54; and so 0.54 plus 0.46 is 1. So, you just do verification, right.

So, now we want to define the, compute the expectation value of x given y is equal to 2. So, I mean you just take the definition that we wrote down. So, here the marginal's are given to you; point, now this is be 1.4. So, the expectation here would be when x is equal to 1 then the probability that you obtained; I am computing it for y equal to 2, sorry. So, we have computed these probabilities. So, when x is equal to 1, so when you are computing this expectation y is equal to 2, so then it will be value of x equal to 1 into, the probability that you get; the probability match function when y is 2 and x is 1, right. Is it ok?

So, the expression that I wrote down, see here it will be, you are computing see y is fixed. So, you are computing the expectation of x, given y is equal to 2. So, as x takes different values, given y; so you will multiply by the corresponding probability when x is; for example, x is 1 and y is 2. So, x is 1 and y is 2; this is the probability when x is 2, and given y is equal to 2 then this is 0.2. So, we will take those probabilities, the conditional probabilities, and multiply by the corresponding values that x takes, right.

So, the conditional probability are here; this is this; 0.14 is this and 0.46 is this. So, I multiply by the corresponding values that x takes; and therefore, this is 2.28. In fact, I am going to talk some more in terms of the functional aspect of expected value of; in other words, in fact, we can sit here when I am talking of expectation of x given y equal to, Y equal to small y.

So, you see, because you are taking expectation with respect to x, so then you will be, this will turn out to be a function of y. So, I start giving important example, but we will discuss this in detail in the next lecture. So, this will be a function of y because you have taken expectation with respect to x. So, that part is gone; x part is gone; it is no longer a function of x, but it will continue to be a function of y, right.

And then we will see what kind of relationships we can predict on what, how we can use this. So, but initially through this example I just want to show you how you go about computing these conditional expectations; this is the whole idea. So, similarly you can compute the expectation of x, given y is equal to 1. So, now I did not do this detail calculation here, but you can see that when you wanting to compute.

For example, here x is 1 and then given y is equal to 1, so x is 1 you will be writing that probability. So, I will divide 0.02 by 0.2 because y is equal to 1. So, y is equal to 1 is this. See, you simply have to, just as we computed the probabilities for, conditional probabilities for x equal to 1, given y is equal to 2, I simply divided these numbers by the corresponding probability, y equal to 2. So, here also, when y is equal to 1, you divide these probabilities by this, and you get the conditional probabilities of x equal to 1, y is equal to 1.

And here, 0.06 divided by 0.2 will give you the probability that x is to, conditional probability x is equal to 2 and y is equal to 1. So, this way you can show. So, this know, so that is what all I have done; I have divided this by 0.2. So, then I have written it as 0.1 into 1; so computing the conditional expectation; so multiply by 1.

Then, similarly 2 times 0.06 divided by 0.2. And then, 3; so 0.08 divided by 0.2; and then 0.04 divided by 0.02, and 4 into that, right. So, that number comes out to be 2.7, right. And, in the same way, you compute the expected value of x, given y is equal to 3. So, here I will take 0.07 divided by 0.3 into 1, and 0.03 divided by 0.3 into 2, and so on. So, you will compute those expectations.

And now, as I am saying that if you take this expectation; and yes, this I have just now written down this expression; we will spend lot of time on it, trying to show you. But, computationally you see, if I now want to compute the expected value here, as I told you, this is a function of y, right. So, when you want to compute expectation through conditional expectation of x, y, given equal to y, then all I have to do is to multiply by this corresponding. For example, 2.7, I will multiply by the probability that y is equal to 1 because this is the conditional expectation of x, given y equal to 1. So, I will multiply by; so you can treat this as a function of y. So, this into the probabilities that y takes the value 1. So, that will be 0.2 into 2.7.

Similarly, this will be the condition; this is the conditional expectation of x, given y equal to 2. So, this is again will be 2.88 into probability that y is equal to 2 which is 0.5; so 0.5 into 2.88. And similarly, here a probability that y takes the value 3, and that into the expectation, here conditional expectation. So, this number comes out to be 2.82. And, we can verify that this is actually equal to E raise to x because this is the marginal density of x. So, to compute the expectation of x, I will multiply 0.19 into 1 plus 0.19 into 2 plus 3 times 0.23 plus 4 times 0.39 which again gives me the number 2.82. So, the 2 numbers are equal.