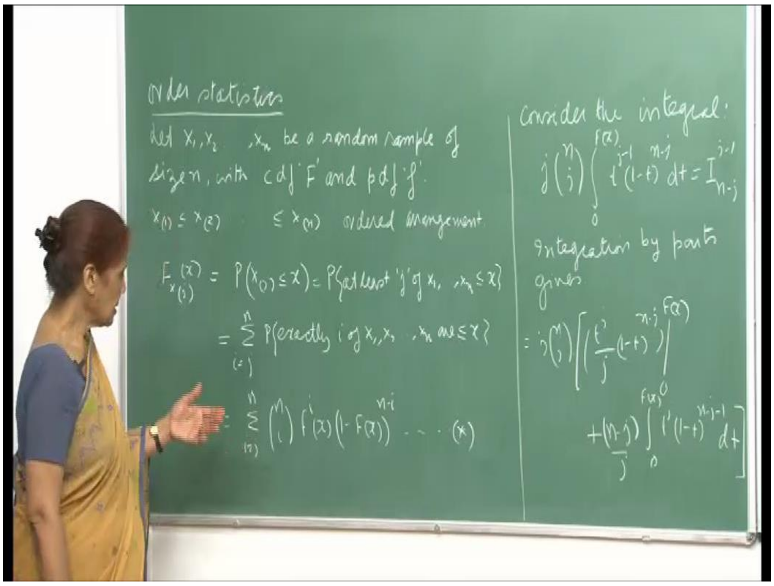
**Introduction to Probability Theory and its Applications**
**Prof. Prabha Sharma**
**Department of Mathematics and Statistics**
**Indian Institute of Technology, Kanpur**

**Lecture - 17**
**Covariance Correlation Cauchy-Schwartz Inequalities Conditional Expectation**

(Refer Slide Time: 00:15)



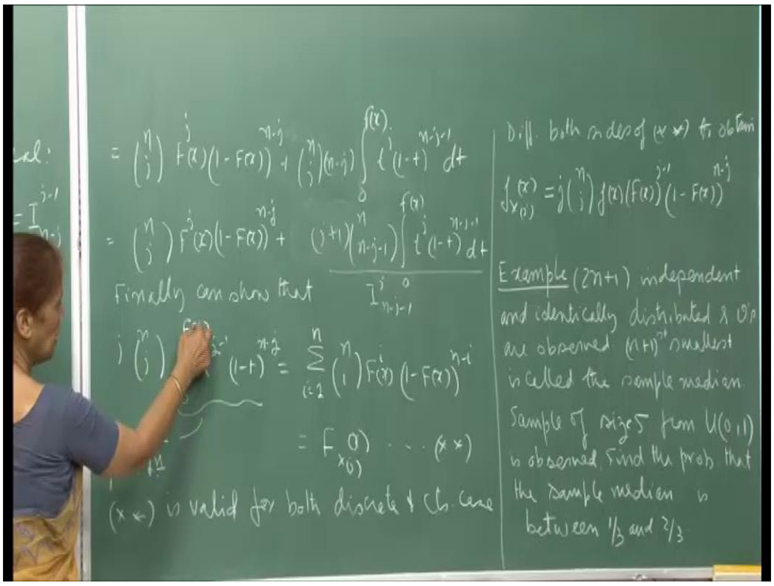

There order statistics. So, let us just begin with the first one. So, if you want to find the density function. So, cumulative, we will start with c d f, because once you have obtain this, we can obtain the p d f also. So, here f x j is the c d f for the j x order statistic. So, which means and... So, the value f x j x means that this is the probability of x j being less than or equal to x. And what does this mean, if this is the j th order statistic; that means, up to 1 2 x 1 x 2 up to x j, they should all be less than or equal to x at least. So, at least j of the x 1 x 2 x n should be less than or equal to x. More can be less than x, but at least I am saying, I am co computing the probability that x j should be less than or equal to x; that means, at least j of x 1 x 2 x n should be less than or equal to x. So, now I can write, since it is at least, so it can be j j plus 1 j plus 2 which are less than or equal to x. So, therefore, this probability can be written as summation i varying from j 2 n though probability of exactly i of x 1 x 2 x n are less or equal to x. So, should be clear.

So, therefore, this implies that. I can write this probability as n c i, because out of the n, you are choosing i, any of the i can be less than or equal to x n c i, and then this is f i x, because i of them you want to be less than or equal to x. So, this is this into 1 minus f x.

So, the remaining n minus i are greater than or equal to x. So, because exactly i of them are less than x, so therefore, remaining n minus i are greater than or equal to x. So, therefore, this will be the probability of that. So, you are summing this up. And now let us give it more concise form. So, because this is of course, very unwieldy you cannot. So, now consider the integral, and this is way you look see how we can relate, you know summations with integrals and so on.

So, consider the integral j n c j 0 to f x t raise to j minus 1 1 minus t raise to n minus j d t, and let me call this integral I of j minus 1. So, this index and n minis j index of 1 minus the power of 1 minus t. Now if you integrate by parts, so integration by parts will give. Here let me treat this as the first function. So, integral of this will be j upon t j upon j then 1 minus t is to j minus j, this computed from 0 to f x plus n minus j upon j 0 to f x t j and derivative of the second function. So, the derivative would be n minus j 1 minus t raise to n minus j minus 1, so this is what you get by integration by parts. So, here of course, at 0 this is 0 at f x it will be. So, here I have written also the j part. See this j cancels out, because both the terms have j in the denominator.

(Refer Slide Time: 03:32)




So, this cancels out, and you are left with n c j f j raise to x 1 minus f x is to 1 minus j plus when you come to the integral this will be n minus j into n c j n minus j i of j n minus j minus 1, because j here this was j minus 1 this is minus 1 this was n minus j this is n minus j minus 1. Now, very nicely you can simply just write down this expression and this, and manipulate the terms and you can immediately see that this whole this can be written like this; j plus 1, and therefore, this will then become. Well the whole thing why did I write, because my i j is together. So, therefore, I do not have to write this term this is note there because I am trying to say that, you manipulate this, even this is not correct I should have written the integral here only. So, let me just say this is not correct. let me write continue writing the integral and j. So, what are we getting here; 0 to f x 0 to f x t raise to j 1 minus t raise to n minus j minus 1 d t, and then I am saying that you manipulate this and you can write this as. So, let me rewrite this, I am writing as, because now you need j plus 1 n n minus j n minus j minus 1, 0 to f x t raise to j 1 minus t raise to n minus j minus 1 d t.

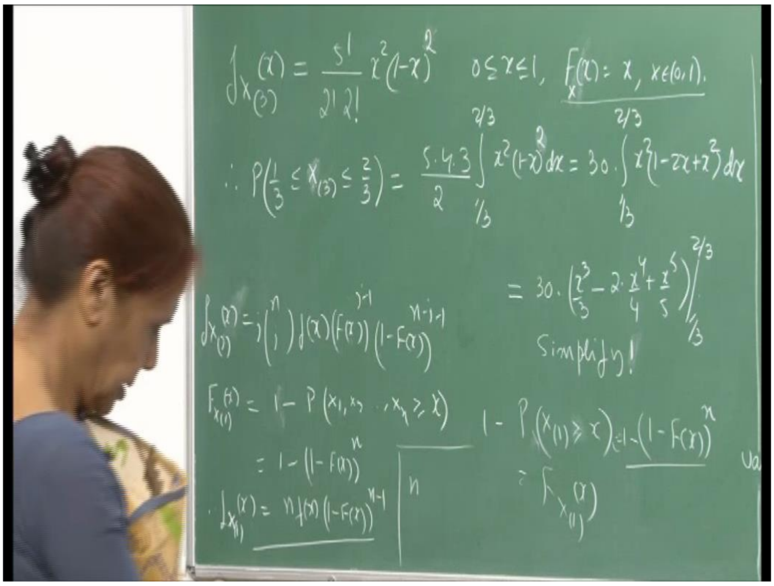
So, this whole thing can be written as i of j n minus j minus 1. So, therefore, you see, the iterative relationship is there. So now, this plus integral, where the power of the term 1 minus t raise n minus j has now become n minus j minus 1, power of t is going up right j minus 1 to j and so on. So, now, iteratively when I write, I will again get a term when I integrate this by part, I will get it term here plus than this actually j plus 1 f x is raise to j plus 1 then n minus j minus 1, and then another integral which will be i j plus 1 n minus j minus 2. So, this way iteratively when you do it, this power finally, becomes 0 and this will become your n. So, therefore, you can show, this summation is equal into the integral, and this what I have said that is. So, finally, you can show that this integral that I write down in a beginning, is equal to this sum, which is equal to your cumulative density function of x j. And this you should have recognized by now, because this is.

See this is the beta integrant, together with the, or beta function when you want to make it p d f. And since the limits are from 0 to f x therefore this is called incomplete beta function. So, finally, I have been to able to replace, get this probability, the cumulative density functions, in terms of this integral. So, when you differentiate both sides of a double star, you will get f x j from here. It will be the p d f of x j, and this is you know differentiation under integral sign. So, since this is function of x, this will become f x here, and otherwise you just substitute for t f x and so you get the same, that for special cases; say for example, when you are sample values of from the uniform distributions, or I think from, may be from normal distribution. We will see through a examples, then it is easier to get this explicit expression, for your c d f and for your p d f. So, we will go through this example, to see how. And of course, the question arises as to why we are doing this, and you will see that.

Let us just go through this example and you will know why we are talking about obtaining p d f for these order statistics. So, if you have a sample of size 2 and plus 1 independent and identical distributor and variables are observed, then the n plus first. See n observation from this side and on this side. So, n plus first is on the center, smallest is called the sample median. So, when you arrange them order them, and then the n plus first 1 smallest, is called the sample median.

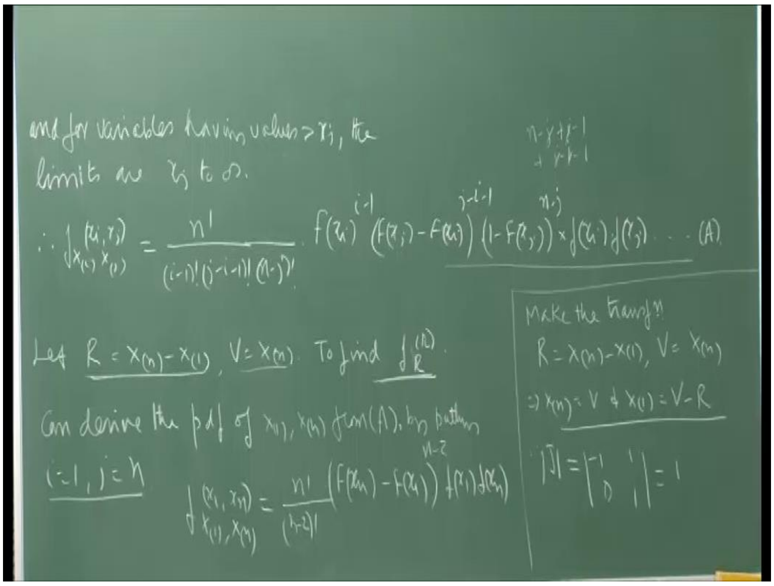
So, now let us say we want to find out the. So, we have a sample of size 5 from uniform 0 1, is observed, find the probability that the sample median is between 1 by 3 and 2 by 3, and this you know when your handling data's, large data's, sometimes you are only interested in what the median of these sample size is. So, we will go about, now obtaining the expression, because you want to find the probability of the sample median between 1 3 and 2 by 3. So, here you are actually talking about x 3; your j is 3 here, because sample size is 5. So, the median will be determined by the third ordered, third smallest statistic.

(Refer Slide Time: 09:36)




So, x 3 is the median, when you of the sample of size 5. So, x 3 will represent the median of the sample, so by our formula. See you remember the formula was j n c j f x F X raise to j minus 1 1 minus f x n minus j minus 1 for f x j. So, put j equal to 3 here, and this will give you 5 c 3, and 3 times this, and then f x. Now for a uniform distribution, your p d f is just 1 the interval 0 to 1. So, this is 1, and this is given by, you know for a uniform distribution.

(Refer Slide Time: 10:17)




So, proof of Cauchy Schwartz inequality, now expected value of y square can be greater than or equal to 0, it cannot be negative, because this is y square. So, therefore, when you integrate y square from minus infinity to infinity or whatever it is into f x which is a p d f non negative. So, therefore, this must be non negative. But then y expected value of y square equal to 0 would imply. So, two things are possible; either expected value of y square is 0, or expected value of y square is greater than 0.

So, if it is 0, then this imply that probability of y equal to 0 is 1, because this expected values this, that needs, see you will write value of y square, whatever possible values is y square takes, into the probability of y square taking a particular value and so on, and this will imply that probability of y equal to 0 is 1. And hence, probability of x y equal to 0 is also 1, yes because this is a certain event, y taking the value 0 is the certain event; hence probability x y equal to 0 is also a certain event and so the this probability 1, and therefore, this implies, that expected value of x y is 0, because this takes the value 0 with probability 1.

So, 1 into 0 plus, or you integrate or whatever it is, whichever you want to write it down, expected value of x y will be 0, and the inequality therefore, will be satisfy because this is 0 and this is 0. So, the inequality is satisfied. So, therefore, we will now proof for the case, when then expected value of y square is positive. x 1 less than or equal to x. So, I will look at the opposite event, which is greater than or equal to x. So, if first order statistic is greater than x; this implies that all the sample values must be greater than or equal to x.

So, this is equal to what, 1 minus f x raise to n, because if the first order statistic is greater than x, since it is a smallest, all other values of bigger than x 1, so all of them must satisfy this inequality, and therefore, this probability is 1 minus f x raise to n, and we are interested in finding out the F x c d f. So, that will become 1 minus of this. So, 1 minus of this, which is equal to 1 minus of this, will give me my f x 1 x, this is the whole idea, so therefore, this is what you have, and which I can write as 1 minus 1 minus f x raise to n. So, when you differentiate with respect to x, you get the p d f here, and this will be simply minus minus becomes plus. So, n times derivative of this is small f x. So, n f x into 1 minus f x raise to n minus 1. So, this will be a general expression.

So, now what I am trying to say is that, with this expression also and now. Let us just substitute j equal to 1 here. So, what do you get; this is 1 and so if you substitute in this formula, it will be n c 1 which is n, then f x and this is 1 minus 1. So, this is 1 this is 1 minus f x raise to n minus, this is minus j minus 1, how am I getting. So, this is coming out to be j is 1. So, this is coming out to be n minus 2, accordingly, say for us it should be 1 minus f x raise to n, and at least from here it appears that this should be this, and I do it n times, I differentiate, and then I take this so it should be n minus 1.

So, where is this other 1 missing, because you taking j to be 1. So, are you sure this is n minus j or minus n minus j minus 1. Let us just verify that, what is the formula, correct formula; it has to be n minus j. It should be n minus j. Let us just make sure so that we do not make the mistake of n minus j. So, see that helps to verify. So, you see this is n minus j and therefore, j is 1 so it will be n minus 1, so both things match. You can obtain it directly or you can do it through the formula, the formula that we have obtained. Now, just a simple example to show you, now let us see it also helps to write down the joint p d f of all the other statistics, that you see.

Actually, what is happening is this is some arrangement of the sample values x 1 x 2 x n, and the possible arrangements of these n sample values is n factorial. So, one of them will match this order and so you can you can do the thing through the regress mathematics, by showing that the, your r n region can be divided into n factorial regions, and each factorial, in each 1 of these factorial regions the, one arrangement of the sample values is there, and when you do the transformation, because you have to do it over whole of r n. So, the Jacobian will be one of the permutation matrices, and the value of the permutation matrices is always 1. I mean you take the positive part, otherwise the value of the permutation matrices plus minus 1.

So, without going into all that, we can simply say that, the joint adjective function would be n factorial into, you see since the variables are independent, the joint density function of the sample values x 1 x 2 x n is nothing, but the product of the individual density functions, and may be if you want to feel good you can, but I am not writing this I am simply saying this is x 1 and this is x n, but they are the same.

So, therefore, I am not writing these indices. So, all of them are the same p d f and therefore, this would be n factorial into f x 1 into f x n, this is the whole idea, so the general expression, where x 1 x 2 x n are varying from minus infinity to infinity. Because after all the order statistics is only one of the arrangements, and there are n factorial possible arrangements of the sample values of the n sample values. So, now to find joint p d f, x will be equal to expected value of x y upon expected value of y square into y.

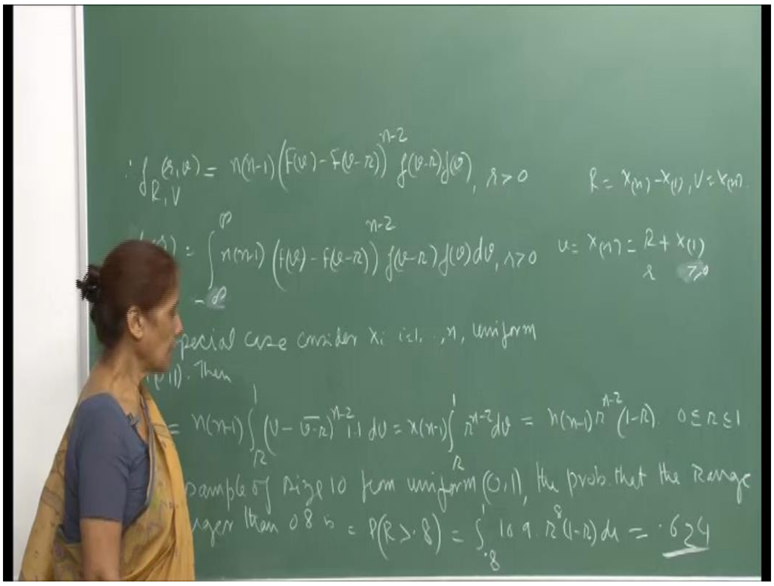
So, the minus sign is not there, see it gets cancel out. So, x I x j then what is happening in this, I will integrate this. So, for i minus 1 sample values, the limits of integration will be from minus infinity to x i, because i minus one of them have to be less than or equal to x i. For variables between x i and x j order statistics x i x j, the limits are x i to x j, and for variables having values greater than x j, the limits of from x j to infinity. So, once I do this integration; that means, I will be integrating for i minus 1 then for variables between x i and x j, and then for all the variables having values greater than x j.

Let me write it this way. So, then once you do this, you will get the joint density function of. Remember, because for marginal when you had the joint density function, to obtain p d f of one of them, you would integrate respect to the other one, and then get the marginal p d f for the first variable. So, here also we have done the same, and therefore, these remain intact, and for the remaining; see this is f x i raise to i minus 1, because you are integrating from 0 to from minus infinity to x i and for variables between x i and x j, you are integrating from f x i to f x j. So, this is j minus i minus 1, and this is 1 minus f x j and minus 1, so this minus 2. So, these add up to n minus 2, and then you have the remaining 2 x i and x j. So, this will give you the joint density function. So, out ultimate aim say therefore, see the range of the sample values is also of lot of interest, in many situations, so we want to ultimately find out the range of the sample values. So, let me just define 2 random variables here, which are r is x n minus x 1. So, this is the range, and v is the largest sample value, and here of course, you should try to see that you can compute the p d f of x n directly, and then again verify from this formula.

So, for when you want to find out the p d f of v of x n, when you say probability x n less than or equal to x, which would mean that all the sample values are less than or equal to x. So, you will immediately get, this thing to be f x raise to n. So, the cumulative density function of x n will immediately come out to be f x raise to n, and you would differentiate. So, n times f x small f x into F x, so raise is to n minus 1. So, here you can directly get this also. Anyway, so we have to find out the p d f of capital r. So, can derived the p d f of; see x 1 x n. Now first I need to know the joint p d f of x 1 and x n, once I obtain that, then these are functions of x 1 and x n, so I will use my transformation formula, and get the joint p d f of r and v, and from the joint p d f of r and v, I will then finally get the p d f of r, this is the whole idea. So, for to derive the p d f of x 1 and x n from your this formula, I will simply write i as 1 and j as n. So, i 1 this term is gone, and here also this term is gone, this term is gone. So, you are left with n factorial upon n minus 2 factorial.

So, n factorial upon minus 2 factorial, then this f x n minus f x 1 raise to n minus 2, and then this is out, because n minus n is 0 and so this is f x 1 f x n. So, this is your joint p d f of x 1 and x n. Once I have this, then I will make the transformation, that I will write r as x n minus x 1 and v as x n. So, then from here you will go get the relationship for. So, x n comes out to be capital V and x 1 comes out to be v minus r, and then you write the Jacobian. So, r is my first variable. So, this will be minus 1 1 and then here it will be 0 1. So, the Jacobian absolute value is 1. So, now, I can get the joint density function of r n v. So, with Jacobian as 1, absolute variable Jacobian is 1, and these this transformation; that is your x n is v and x 1 is v minus r, in this 1 we just substitute for x 1 x n multiply by the Jacobian.

(Refer Slide Time: 23:02)



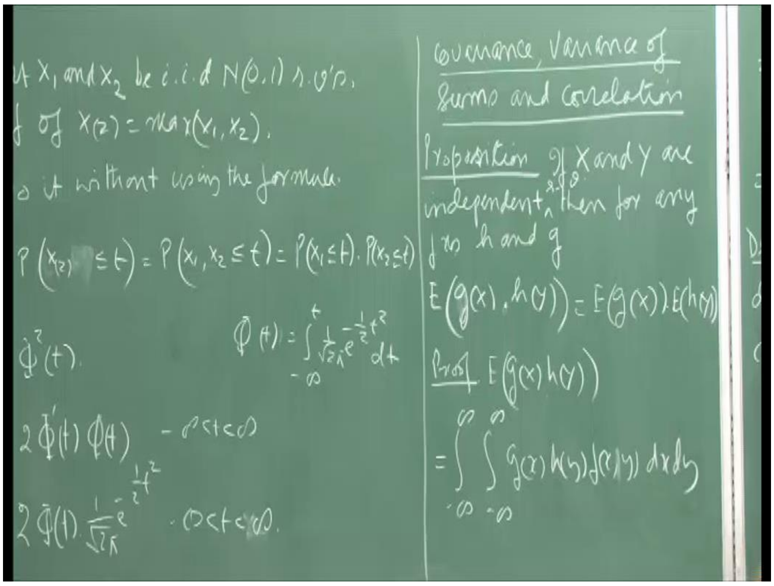
So, then your f of f of r v; that means, capital R is equal to small r, and v then this function this p d f of r and v can be obtain from this p d f n into n minus 1 f v minus f v minus r which is your x 1 this is n minus 2, and this is f of v minus r f v, and here of course, it is understood that r is greater than 0, because r represent the range and x n is greater than x 1, greater than or equal to x 1, so this is the case, and of course, we have range is 0 there is no point. So, we are taking r to be some positive number. So, therefore, once I get the joint p d f or r and v, now my interest is in the getting the p d f of capital r.

So, I will integrate this, and in general you will integrate from minus infinity to infinity. Well why should I say from minus infinity to infinity, it should always be from, it should be r, because you see here your x n is r plus x 1. So, since this is non-negative then I mean, ok fine. In general, it would be minus iterative, because x 1. In general we are allowing x 1 to take vary from minus infinity to, this sample size is from minus for a population, which is from minus infinity to infinity. So, in that case this is fine. In general we can write as minus infinity to infinity. Now, as a special case consider, the case 1 x i r from uniform 0 1.

So, as a special case consider x i i varying from 1 to n, from uniform on 0 1. Then this function this p d f will reduce to whatever I written here the f r r. So, I am writing this. So, the p d f of the range variable would reduce to n into n minus 1, and now in this case it will be r, because as I am saying not that this is non-negative, if f all of the sample values are coming from uniform 0 1, all values are non negative, and therefore, this x n has to be greater than or equal to r plus. So, in x n is your v. So, when this values v and this is r. So, then v has to be greater than or equal to r. So, now in the joint density function, you are integrating with respect to v, to get the p d f of r. So, then that will be the range will be from r to 1, because variables are from 0 to 1. So, then this will become v, this will be v minus r raise to n minus 2, and both the p d f are 1 1. So, 1 1 d v, this is your this thing, and you can see the simplification v cancels out this r raise to n minus 2 d v. So, the integral here would be v, which will be 1 to r 1 minus r.

So, therefore, this is your p d f for the range, and now you can find out the possible. So, for a sample of size 10 from uniform 0 1, the probability that the range is larger than 0.8. So, these questions are of full out of interest. So, you want to find out the range of values, the sample that you have observed. So, if you are saying that the range is larger than 0.8 then you want to compute the probability that capital R is greater than 0.8 therefore, you will integrate this function from 0.8 to 1, and if you simplify here you get the answer is 0.6 to 4, which is a pretty large probability, of range values being more than 0.8, the range of the sample being more than 0.

(Refer Slide Time: 27:04)



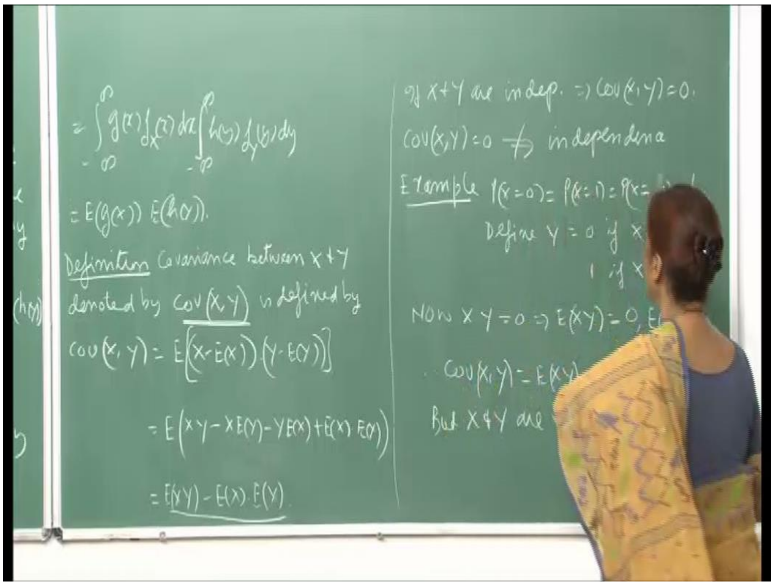

So, another example; now from the normal distribution, because I thought we had had enough cases for uniform. So, if x 1 and x 2 are identically independently distributed, from a normal 0 one; that means, the mean is 0, and the variance is 1. So, this is a sample, x 1 x 2 is a sample from normal 0 1. So, find the p d f of x 2, which will represent the max of x 1 and x 2. So, we will again obtain this, without using any formula. So, here again as I explain it to you, that if you want to find the c d f of cumulative density function for x 2, second the largest one, then this will be probability x 2 less than or equal to t, which will imply that both the values x 1 and x 2 should be less than or equal to t, and since they are independent, this is equal into probability x 1 less than or equal to t into probability x 2 less than or equal to t, and coming from t and 0 1 will be root 2 pie e raise 2 minus 1 by 2 t square d t. So, this is our notation for phi t for a normal distribution. So, this is standard normal distribution, so phi t.

So, therefore, this will be phi square t. So, the cumulative density function for the max of 2 sample values x 1 and x 2 coming from normal 0 1, is given by phi square t. So, then if you want to find the p d f, just this differentiate this, which would be twice phi prime t f t. So, phi prime theta will be nothing, but the normal p d f, which is given by this, so it will be twice phi t into 1 by root 2 pi e raise to minus half t square, so minus infinity to t. So, this one can, then you know integrate find out, whatever probability you are interested in. So, it looks like in that at least the, normal if your sample is from a normal distribution, or from uniform distribution, you can you know easily obtain p d f of the order statistics. in other cases also, one can see of course, there method for computing difficult integrals, by many other ways, by numerical methods.

Now continuing with our joint distribution functions, and the other important parameters that we need to look at, and define here is; covariance variance of sums, and correlation. So, this will also have a lot of implication, and see here of the purpose of before I talk about, define the covariance and the variance, and then the correlation, simple proposition, which in fact, there was no need to prove it also, but I have written it down for completeness sack; x and y are independent, so if x n y r independent random variables.

This is understood random variables, then for any function h n g. For any functions h n g, expectation of g x into h y, is expectation of g x into expectation of h y; that means, the independence carries over to, the function g x and h also. So, here this is the proof is simple, because if you want to write the expectation of g x h y, it will be minus infinity to infinity, minus infinity to infinity g x h y f x, y d x d y, but since x and y are independent, the joint density function can be written as the, product of the marginal densities. So, here when you write this as; f x and into f y, then I can even separate out the integrals, because it will minus infinity to infinity g x marginal of x into minus infinity to infinity h y into marginal of y into d y, and so by definition this is e g x into e h y. So, once we have this behind us, then we can talk of, define the, first of all the covariance.

(Refer Slide Time: 31:30)




So, the idea of covariance between 2 variables x and y, is denoted by, this is the notation, and is defined by the covariance expectation of x minus e x into y minus e y, and if you open up this, if you expand this expression, this will be x y minus x e y minus y e x plus e x into e y, when I take the expectation inside, it will be expectation of x y. Then this will be expectation x into expectation y. Then this will be minus expectation y into expectation x plus this. So, one of them the 2 of these will cancel out each other, and you will be left with. So, this is a simpler expression to handle, when you are talking of covariance. So, it is expectation x y minus e x into e y, this is the definition of covariance, and let us see what does it indicate, or why do we. So, now if x and y are independent; see if x and y are independent, then e of x y will be written as e x into e y. So, then e x into e y minus e x into e y will be 0, so covariance. So, therefore, if x and y are independent, this implies that covariance x y is 0.

But, unfortunately the converse is not true; that is, covariance equal to 0, does not always imply independence of the random variable. Very simple, I will tell you, that the converse of this result is not true. So, independence always implies that the covariance is

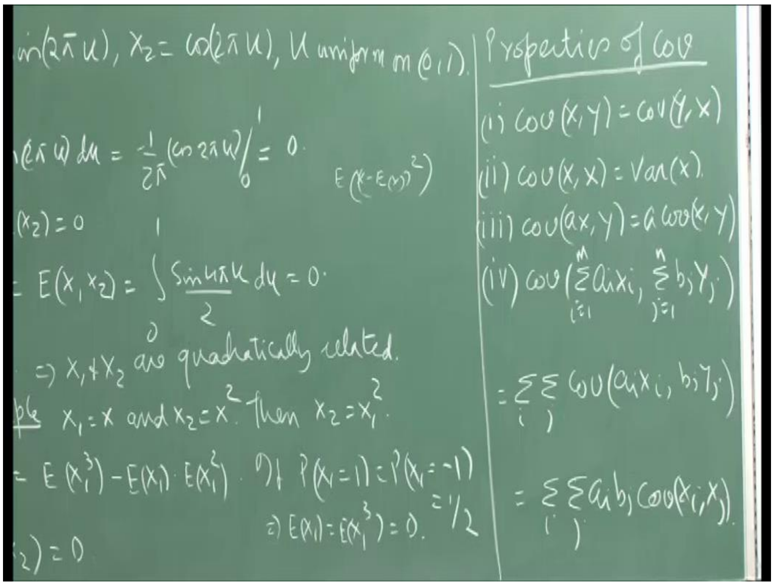
0, but if the covariance is 0, it need not imply that the variables are independent. So, let us see, we defining random variable x, which takes 3 value. So, probability x equal to 0, and probability x equal to 1, and probability x equal to minus 1 is equal to 1 by 3, so all 3 are equally likely. Then I am defining a random variable y, which is totally dependent on x. So, y 0 if x is not 0 and y is 1 if x is 0.

So, now if you look at the values of this product, this will always be 0, because y 0 when x is not 0 and y is 1 if x is 0, so this product will always be 0. If the product is 0, so the random variable just takes only 0 values. So, then this expectation will be 0, because variable is taking all possible values as 0. So, this is expectation of x y 0, and you see from here expectation of x is 0. see x and z having the same p d f and c d f, does not imply the that x and z are dependent, but we see here that when given x, z can only take the values x and minus x, we have just see this, and therefore, x and z are completely dependent, because what will be the expectation of x.

So, expectation x will be, so this 0. So, expectation of x is 0; therefore, from the covariance formula, this is 0, this is 0, so the whole thing is 0. So, covariance 0, but we know that x and y are not independent, yes x and y are not independent, if you want you can do this way, what was the. I mean, what will you use you will use, you can show that probability x y, because they are discrete random variable, so all possible values.

So, in fact, x y takes all 0 values. So, therefore, here you have to show that. How would you want to go about doing it, normally for a discrete thing you want to show that for all possible values, of this product, the probability. So, for all of them, is not equal to the product of individual probabilities. But here you will have to yeah you will have to write out in detail, but anyway you can, as it is there is not much to really prove, because the way you are defining your y, it is totally dependent on x. So, that gives you the. So, therefore, covariance 0, does not in imply independents of the random variables.

(Refer Slide Time: 36:21)

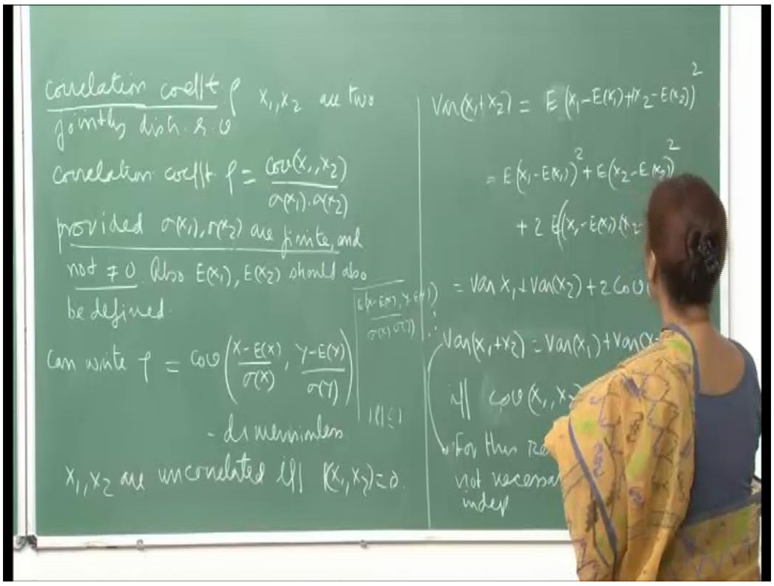



So, continue with this, let’s take another example here; let x 1 be sine 2 pi u and x 2 is cos 2 pi u, so 2 different functions of a uniform random variable 0 1. So, u is your uniform random variable on 0 1. We consider random variables, obtained by taking functions sine 2 pi u and cos 2 pi u. Now, let us see, if functions sin 2 pi u and cos 2 pi u right now let us see if you compute expectation x 1, this will be 0 to 1 sin 2 pi u d u, which will be minus 1 by 2 pi cos 2 pi u from 0 to 1 which is 0, because cos 2 pi minus cos pi cos 0 both are 1. So, this becomes 0. Similarly, you can show that expectation of x 12 will also be 0, and when you compute the. So, therefore, the covariance of x 1 x 2, will reduce to, just expectation of x 1 x 2, but expectation of x 1 x 2 will be, you see sin 2 pi u into cos 2 pi u will be sin of 4 pi u divided by 2. So, again this is same kind of function from 0 to 1. So, that value will also be 0, but then x 1 plus, because x 2 will be 1 minus x 1 square under root.

So, therefore, the covariance is 0, it does not suddenly imply independence of x 1 and x 2, which you can see otherwise also, because x 1 square plus x 2 square is 1. I will come back to this example in a while. Now, properties of some which we can immediately show; properties of covariance function. So, this is, first of all it does not matter, what order you write, covariance x comma y, is same as covariance y comma x, because this expectation of x minus e x into y minus e y. So, the order is not important. Then when you take both x and y to be the same, then co covariance x x, because that is expectation of x minus e x. So, square this will covariance x.

So, therefore, this is equal to variance x, and if you take covariance e x comma y, then again by definition, because a will be here, a will be here also, you will be able to take it out, and it will be a times covariance x comma y. And then you can apply this principle in general, because we have already shown it for this, and then since because of this. So, you can show that if you take summation; sigma a i x i i varying from 1 to n sigma j varying from 1 to n v j y j. Then again, taking all possible products here, and covariance you can take it, because its expectation function which is linear. I can take it inside or the summation sign so this can be written as this, and then this is summation that will be outside, and then here this is a i b j will come outside, and this will be covariance of x i x j. So, this is the general expression, and I will show you nice application of this, after a while. How you can use this formulae to simplify some computations.

(Refer Slide Time: 39:35)




Now, the moments you define the covariance function, you will be immediately have define the correlation coefficient rho, and we will see the implication and the usefulness of this parameter, so if x 1 and x 2 are two jointly distributed random variables. Then the correlation coefficient rho, is covariance x 1, x 2 divided by variance of x 1 into variance of x 2. Now, of course, this definition is valid, only when sigma x 1 and sigma x 2 are finite and. In fact, there should not be zeros, because if you are x 1 and x 2 or any of them is a constant variable, taking only constant values; that means, no randomness about it. Then your sigma x 1 will be 0, so you cannot divide by 0. So, this quantities is defined, only when sigma x 1 and sigma x 2 are finite, and they are not 0. In fact, this applies to your covariance function also, but expectation x 1 expectation x 2, should also be defined. So, in fact, I should when I defined the covariance function, I should have spelled out that the definition is valid, or as long as your expectation functions exists, are defined.

Now, you can see, you can immediately see that here, the covariance function, the correlation coefficient, can be define nicely in this way, and once you define it this way, then it becomes dimension less, because I have standardize the way root x minus e x divided by define divided by the various standard deviation, and y minus e y divided by stranded deviation. So, this becomes. If you remember how this standardize your normal variate. So, the same thing we are doing, and once you do this then this becomes dimension less. Now, if you want you can try it out here, because see the covariance you are defining as expectation of x minus e x and y minus e y. So, this you are defining it is this. So, here also it is, and then for the covariance you simply taking rho x and then rho y. So, by this definition, I can take this inside, so there is no big deal. I mean I am not doing anything manipulation here, simply taking this inside, because we have already seen; that the constant can be take outside or inside, does not matter.

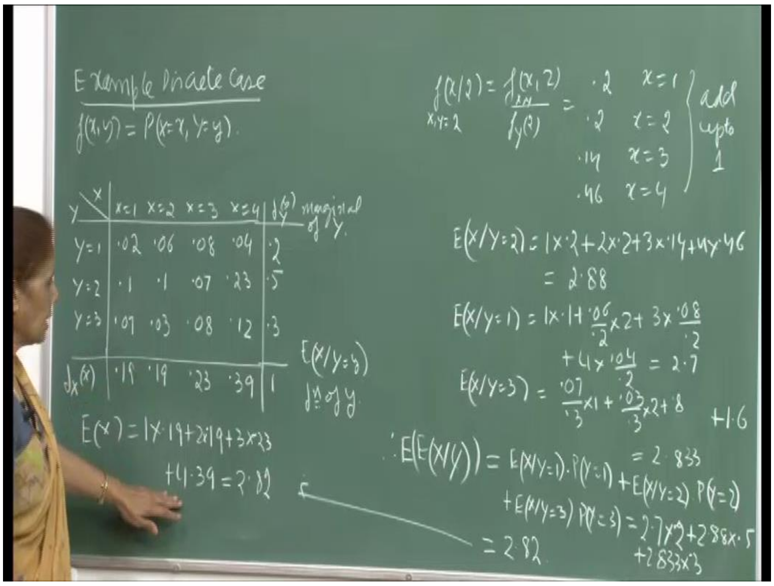
So, therefore, now becomes standardize (( )). So, this is dimension less definition of the correlation. Now, we used the word. So, if see rho x 1 x 2 0; obviously, is possible when the covariance is 0. So, essentially when the covariance is 0, we use the word; that x 1 and x 2 are uncorrelated, and you have already seen, that 2 variables been uncorrelated, does not mean independence. So, therefore, we have coin this word; that 2 variables are uncorrelated, if and only if the correlation coefficient as we call it rho is 0.

So, this is our terminology that x 1 and x 2 are uncorrelated, provided the covariance is 0 between the 2 random variables. And we will now through Schwartz inequality and so on I will show you, that the number rho, measures the relationship again between, it tries to show; co covariance simply showed you that, whether I mean if the variable are independent then the covariance is 0. Now here rho gives you much more information than that. It will show you that, see we will first of all show that rho is less than or equal to 1 always, because we have standardize the thing, divided by the stranded deviations, and then we will show that rho is equal to 1, then they are perfectly related the 2 variables.

And this actually measures the relationship, but again here we will try to show you that, it may not always measure the. It may it may predict linear relationship very well, but not non-linear relationship, but so we will come to that. Anyway, so this is a very useful parameter, and here also I think the same example, I was trying to take, is that if your x 1 is x and x 2 is x square, so it is the square of x then. See this is the relationship between the 2 variables x 2 is equal to x 1 square, and the covariance will come out to be. So, covariance will be expectation of x 1 cube minus expectation x 1 into expectation x 1 square. Now if I take x 1 to be a variable, which takes 2 values x 1 is 1 and x 1 x 1 is minus 1, both the values it takes it probability half.

Then you see expectation x 1 is also 0, and expectation of x 1 cube is also 0, because this will also be 1 into half, then minus 1 into half, and this will also be 1 into half and half n minus 1 into half, and this will be 0. So, therefore, your covariance is 0, so this will imply that your rho is 0. So, the variables you are saying are uncorrelated, but certainly they are not independent. Now, another immediate use of the word uncorrelated, we can show here, while computing the variance of sum of 2 variables, and this can might be extended to many more, you know when you have sums of more than 2 variables. So, here for example, the variance of x 1 plus x 2, you will define as x 1 minus e x 1.

(Refer Slide Time: 45:50)



So, now we will compute the expected value of x given Y, and therefore, the values of Y will vary. So, when you write this expression, computing it through conditional expectation of x, for different values of y. So, then this will be see conditional expectation of x given y equal to 1. So, that into probability y equal to 1 plus conditional expectation of x given y equal to 2 into probability y equal to 2 plus conditional expectation of x given y equal to 3 into probability y equal to 3, and this, because we have made these computation, so you see that 2.7 into probability of y equal to 1, which is 0.2 from here, and then plus 2.88 is the conditional expectation of x given y equal to 2. So, 2.88 into 0.5, which is 2.88 into 0.5 plus conditional expectation of x given y equal to 3, which is 2.833. So, that into 0.3, this is 0.3, and this adds up to 2.82 which is the same as this, which we computed from here.

So, this is what I want to show you, and therefore, here remember even if somewhere in the text sometimes, you find that capital letter is missing, whichever the conditional. So, whenever we talk of expectation, then the whole idea is that; this expectation of x given Y, when I write the random variable here, then this is the random variable. And so I can talk in terms of expectation of this random variable, and this will be of course, the probability; that means, the value of e x given particular value of y.

So, you compute this expectation, for a particular value of y multiplied by the probability of that particular value y, and then you add up for all possible values of y, and then you get this. So, therefore, you can break up the expectation x, also in this way. So, expectation x in other words we are saying, is expectation and then again expectation conditional y. e x 1 plus x 2 minus e x 2 whole square, and when you open up the square it will be x 1 minus e x 1 whole square plus x 2 minus e x 2 whole square plus twice expectation of x 1 minus e x 1 into x 2 minus e x 2.

So, this is variance x 1, this variance x 2, and this is covariance x 1 x 2. So, now from here it follows, that variance of x 1 plus x 2, is variance x 1 plus variance x 2, if and only if covariance x 1 x 2 is 0. So, if and only if; like if covariance x 1 x 2 is 0 then you get this, and if you saying this variance is equal to this, then covariance must be zero. So, this is if and only if relationship, and for this result to be true, it is not necessary x 1 x 2 to be independent. See earlier we had talked of independence, then I talked of some of 2 independent random variable. I had shown you that this will be equal to this. But now we are saying, since we have different find this term uncorrelated. So, what we are saying is, that for variance of x 1 plus x 2, to be equal to variance x 1 plus variance x 2. It is enough that the covariance is 0 or the variables are uncorrelated. It is not necessary for x 1 x 2 to be independent. It is enough, if x 1 and x 2 are uncorrelated. I cannot write it here, but this is uncorrelated. So, this is one advantage, one use of this function. We will talk about this some more in the next lecture.