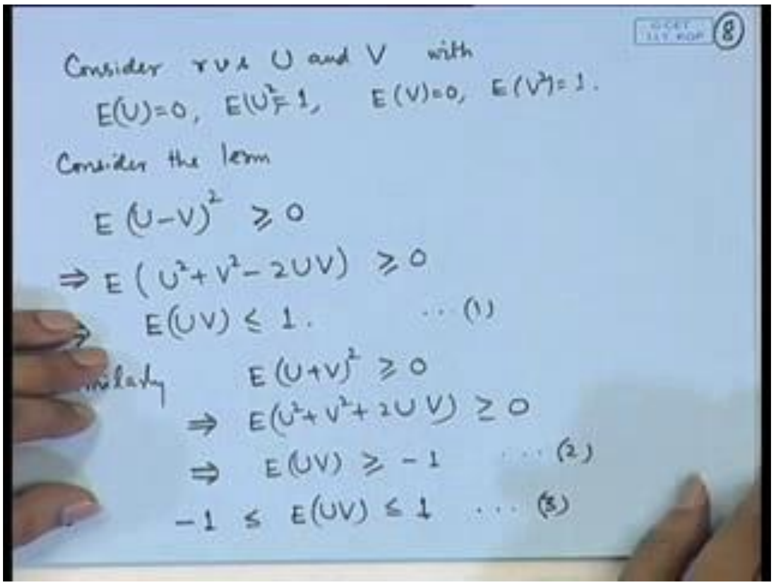
Probability and Statistics
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture – 38
Linearity Property of Correlation and Examples

(Refer Slide Time: 00:21)

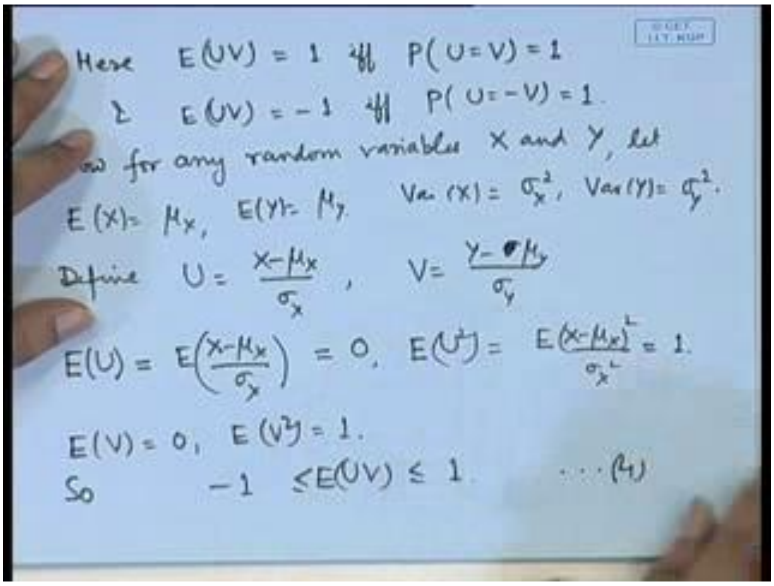


Consider say random variables U and V with say expectation of U is 0, expectation of U square is equal to 1, expectation of V is equal to 0, expectation of V square is equal to 1. If we consider these 2 random variables, consider the term expectation of U minus V whole square. Now naturally this is greater than or equal to 0 being the average value of a non negative term, now this will imply expectation of U square plus V square minus 2 U V is greater than or equal to 0, substituting the value of expectation U square and expectation of V square as 1, this relationship is reducing to expectation of U V is less than or equal to 1.

Similarly, expectation of U plus V whole square is greater than or equal to 0, this yields expectation of U square plus V square plus 2 U V is greater than or equal to 0. Once again substituting the values of expectation U square and expectation V square as 1, we get expectation of U V greater than or equal to minus 1. So, we have got that expectation of U into V lies between plus 1 and minus 1 provided expectation of U and expectation of V is 0, and expectation of U square and expectation of V square is 1.

Now, we can consider when the equality will be attained. So, here equality at 1 will be attained when expectation of U minus V whole square is equal to 0. Now expectation of a non negative random variable is 0 if and only if the random variable itself is 0; that means, U must be equal to V with probability 1. In a similar way the equality at minus 1 will be attained, if U is equal to minus V with probability 1, let us consider the expressions for this.

(Refer Slide Time: 03:30)

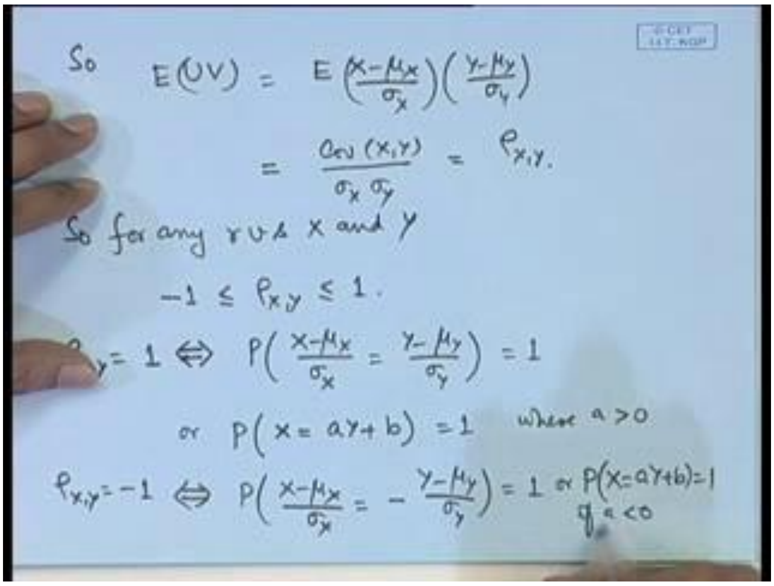


Here, expectation of U V is equal to 1 if and only if probability of U is equal to V is equal to 1 and expectation of U into V is equal to minus 1. So, now, let us take for any random variables X and Y let us use the notation that expectation of X is equal to mu x, expectation of Y is equal to mu y, variance of X is equal to say sigma x square, and variance of Y is equal to sigma y square. Define U is equal to X minus mu x by sigma x V is equal to y minus mu y by sigma y.

So, if we take expectation of U, this is equal to expectation of X minus mu x by sigma x, and by the linearity property of expectation, it is expectation x minus mu x by sigma x that is simply 0. If we consider expectation of U square that is equal to expectation of X minus mu x square by sigma x square; now the numerator here is simply variance of x, that is sigma x square so it is 1.

So, in a similar way you can see that expectation of V is 0, and expectation of V square is 1. So, if we make use of the inequality that we have proved here for U V random variables with the property that expectations are 0 and the expectations of the squares are 1. So, we get expectation of U V between minus 1 to 1; now for any random variables x and y when U and V are defined like this, what is the expectation of U V representing?

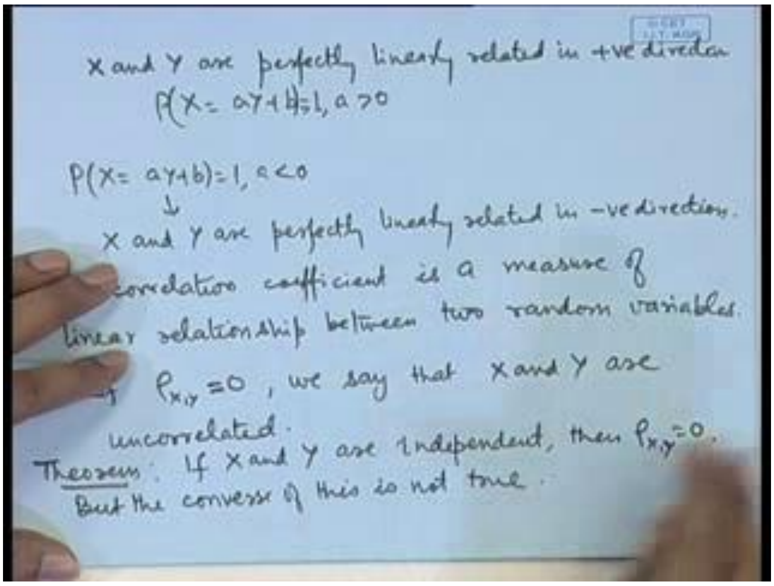
(Refer Slide Time: 06:13)



So, expectation of U V is equal to expectation of X minus mu x by sigma x into y minus mu y by sigma y. So, the numerator here is simply the covariance term between x, y divided by the standard deviations of the x and y that is the correlation coefficient between. So, for any random variables x and y, the correlation coefficient lies between plus 1 and minus 1; the least value is minus 1, and the maximum value is plus 1.

Now, rho x, y is equal to 1. So, we look at the conditions for attaining the equality, expectation of U V was 1 if and only if probability of U is equal to V is 1. So, this will be satisfied if and only if probability that X minus mu x by sigma x is equal to Y minus mu y by sigma y is equal to 1; or you can say probability that x is a linear function of y; that is x is equal to some a times y, plus b where a is a positive number, because sigma x by sigma y. Similarly rho x, y is equal to minus 1, if and only if x minus mu x by sigma x is equal to minus y minus mu y by sigma y is equal to 1 or probability that x is equal to a y plus b is equal to 1, if a is negative. So, this condition that x is equal to a y plus b, where a is positive this is known as that x and y.

(Refer Slide Time: 08:35)



So, we can write it that x and y are perfectly linearly related in positive direction; that is x is equal to a y plus b, for a positive and probability of this statement is 1, and if we say probability of x equal to a y plus b is equal to 1 where a is negative, then we say that x and y are perfectly linearly related in negative direction.

Now, this gives a interpretation for the coefficient of correlation. So, we can see that in general coefficient of correlation lies between minus 1 to plus 1, the bounds minus 1 and 1 are attained, so minus 1 is attained when there is a perfect linear relationship in a negative direction or you can say perfectly negatively linearly related and the equality at 1 is attained when it is perfectly linearly related in the positive direction or perfectly positively linearly related.

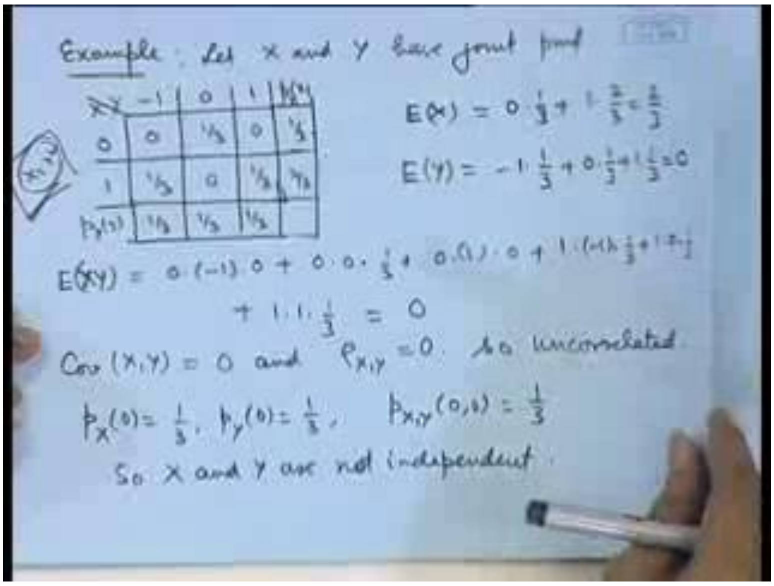
So, in general any value between minus 1 to 1 denotes the degree of the linear relationship between random variables x and y. Suppose I say the correlation coefficient is equal to 0.7, that shows that there is a good positive correlation between x and y and here the relationship is of the linear type. If we say rho x y is equal to minus 0.3 it shows that there is a lower degree of negative linear relationship between the random variables x and y; when rho x y is equal to 0, we say that the random variable x and y are uncorrelated. Now here uncorrelated means that the linear relationship is not existent between random variables x and y.

So, we can interpret this statement as. So, correlation coefficient is a measure of linear relationship between two random variables, if rho x y is 0, we say that x and y are uncorrelated. At this point it is important to understand the difference between uncorrelatedness and independence; if we say that the random variables are uncorrelated it does not mean that they are independent of course, if x and y are independent it will imply uncorrelatedness, because if x and y are independent then covariance term is 0 therefore, correlation term will also be 0.

So, we have the following result if X and Y are independent, then rho x y is 0, but the converse of this is not true. So, proof is of course, obvious that if X and Y are independent then covariance term is 0, and therefore the correlation term is 0.

Let us look at the converse of it.

(Refer Slide Time: 13:07)



Let x and y have joint pmf; as x is taking value 0 1, y is taking value minus 1, 0 and 1, the probabilities are 0, 1 by 3, 0, 1 by 3, 0, 1 by 3. So, here we consider say expectation of X. So, here you can calculate the marginal distributions by adding the rows and columns. So, the marginal distribution of x is 1 by 3, 2 by 3, the marginal distribution of y is obtained as 1 by 3, 1 by 3 1 by 3. So, expectation of X is equal to 0 into 1 by 3, plus 1 into 2 by 3, that is equal to 2 by 3.

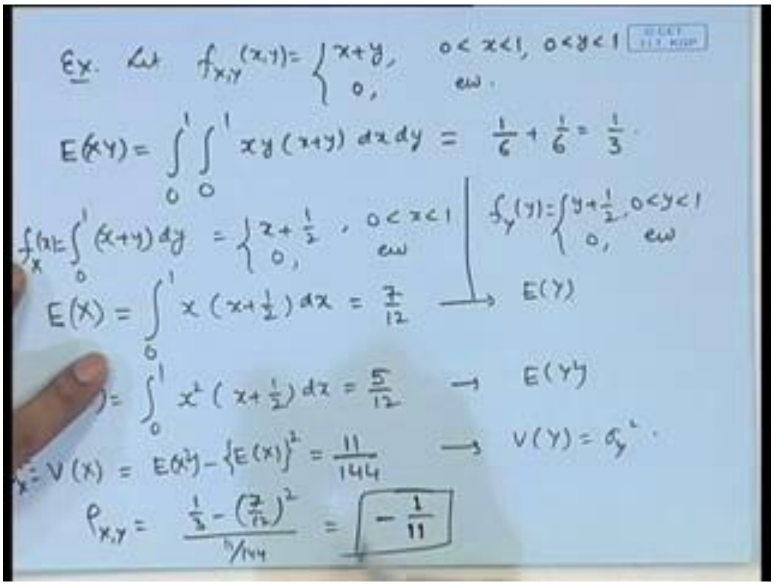
If we look at expectation of Y that is equal to minus 1 into 1 by 3, plus 0 into 1 by 3, plus 1 into 1 by 3 that is equal to 0. If we look at expectation of X into Y; so we look at all the possibilities of the x and y values here. So, X is 0, Y is minus 1 with probability 0, plus x is 0 y is 0 with probability 1 by 3, X is 0, Y is 1 with probability 0, X is 1, Y is minus 1 with probability 1 by 3, X is 1 Y is 0 with probability 1 by 3, plus X is 1 Y is 1 with probability 1 by 3.

You can see these terms vanish we are getting minus 1 by 3 and plus 1 by 3 so it is 0 therefore, covariance term is 0 and consequently rho x, y is 0, but we can see here that the product of the marginal distributions for example, P x 0 and P y 0 both are 1 by 3, but P x y. So, let us write here P x 0 is 1 by 3, P y 0 is 1 by 3, but if we consider P x y 0 0 that is also 1 by 3. So, X and Y are not independent. So, they are uncorrelated, but not independent.

Now, this further brings out the contrast between the concept of independence and correlatedness. So, independence simply means then that the random variables or you can say the observance of the 2 phenomena has nothing to do with each other, that is one phenomena which yields the random variables x, and the phenomena which yields the random variable y they are totally independent.

Whereas the correlation gives a degree of linear relationship between the random variables; so if they do not have a linear relationship, the random variables may become uncorrelated, but that does not mean that they are independent. For example even in this problem, it may happen that x is actually y square, because the probability that x equal to 0 is same as probability y is equal to 0, and probability x equal to 1 is sum of probability y is equal to minus 1 and probability y is equal to 1.

(Refer Slide Time: 18:00)

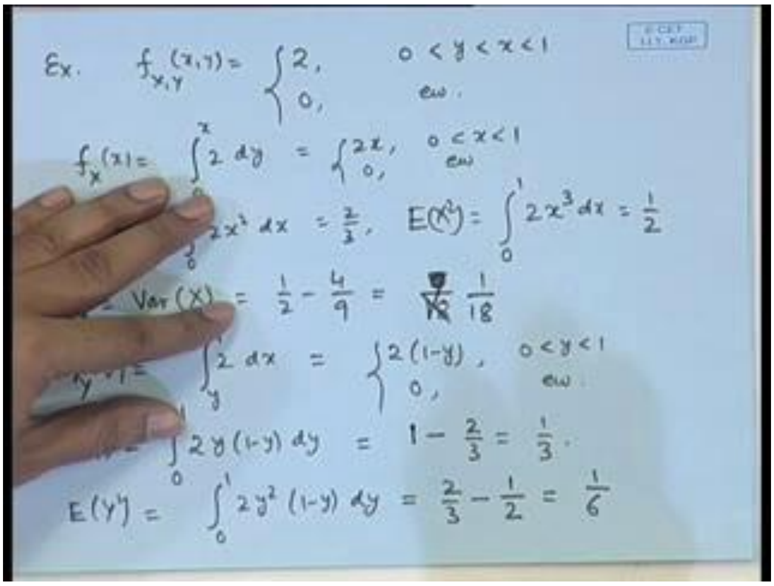


So, this could be a non-linear relationship. So, let us take up few examples for calculation of the covariance's and correlation term. So, let f x, y be equal to x plus y for 0 less than x less than 1, 0 less than y less than 1 and 0 elsewhere. In order to calculate the coefficient of correlation, we need the first and second moments of x and y and also the first product moment of the joint distribution. So, expectation of X into y that is equal to double integral x y into x plus y, d x d y, 0 to 1. So, this is simply integral of x square y plus x y square, which we can easily evaluate and it is equal to 1 by 6 plus 1 by 6 that is equal to 1 by 3.

In order to calculate the moments of x and y separately, we can make use of the marginal distributions. So, you can see here that this will be equal to x plus half for 0 less than x less than 1 and similarly the marginal distribution of y will become y plus half for 0 less than y less than 1 and 0 elsewhere.

The distributions of x and y are same. So, it is enough if we calculate the moments for one of them. So, expectation of X becomes the integral of x into x plus half from 0 to 1, which is obviously, 7 by 12. So, the same value will be expectation of Y and expectation of X square likewise can be calculated that is equal to 5 by 12 same as expectation Y square and therefore, variance of X that is equal to expectation of X square minus expectation of X whole square that is equal to 11 by 144, that is sigma x square and sigma y square.

(Refer Slide Time: 21:15)



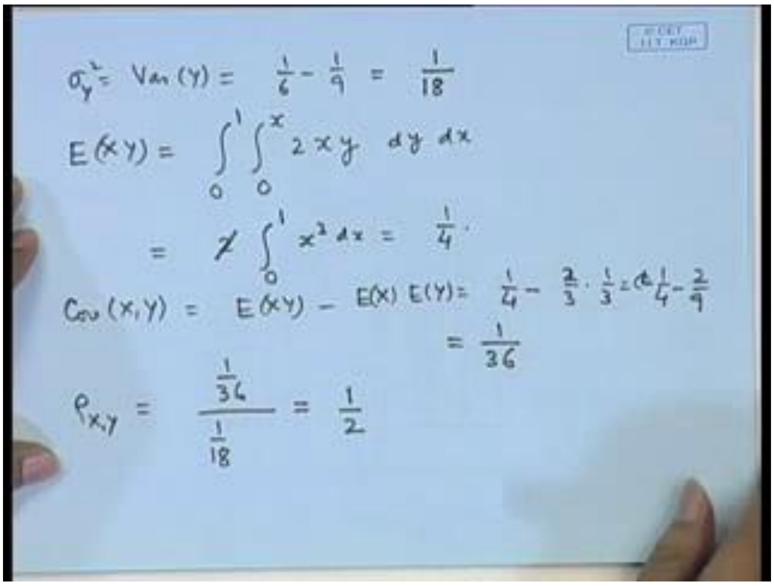
So, correlation coefficient less than equal to 1 by 3 minus 7 by 12 square divided by eleven by 144, which after simplification is minus 1 by 11. So, this means that there is a low degree of negative linear relationship between the random variables X and Y. We can take up some example which we did yesterday, so let us consider say f x y is equal to say 2, 0 less than y less than x less than 1, 0 elsewhere. Here suppose we want the distribution of x then it is equal to integral with respect to y, 2 d y from 0 to x that gives 2 x.

So, expectation of X is equal to 2 x square d x from 0 to 1 that is 2 by 3, expectation of X square is equal to integral 2 x cube d x 0 to 1 that is equal to half therefore, sigma X square that is variance of X is equal to half minus 4 by 9 that is equal to 5 by 18, I am sorry this is not 5 by 18 it is 1 by 18.

So, we are able to calculate the mean and variance of the distribution of x, similarly let us calculate the marginal distribution of y, that is integral 2 d x from y to 1 that is equal to 2 times 1 minus y for 0 less than y less than 1, 0 elsewhere.

So, expectation of Y is equal to integral 2 y into 1 minus y dy from 0 to 1, that is equal to twice y. So, that is y square that is 1 minus 2 y square. So, that is 2 by 3, that is equal to 1 by 3. Expectation of Y square will be equal to 2 y square 1 minus y, d y integral from 0 to 1. So, once again 2 y square; so y square is integrated to y cube by 3, so this is 2 by 3, minus 2 y cube. So, y cube is integrated to y to the power 4 by 4. So, that is 2 by 4that is half so it is equal to 1 minus 4 minus 3 by 6 that is 1 by 6.

(Refer Slide Time: 24:16)

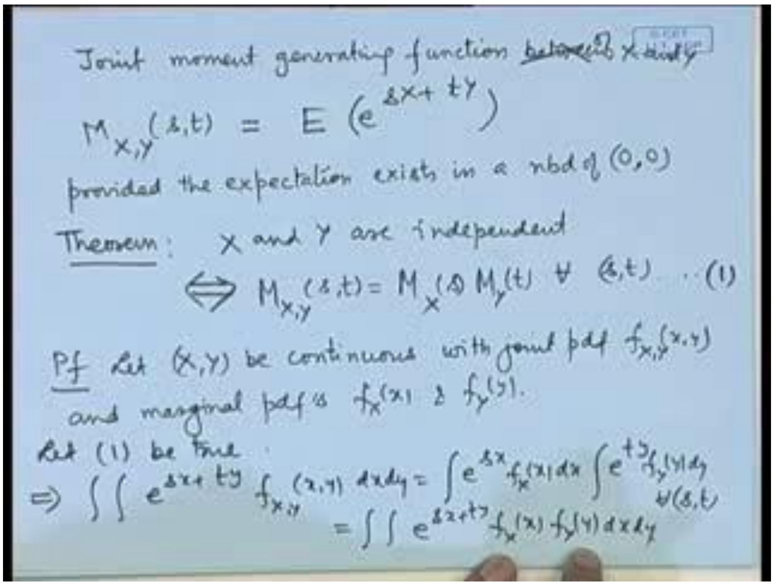


So, we can calculate the variance of y that is equal to 1 by 6 minus 1 by 9. So, that is equal to 1 by 18 now we have calculated the means and variances of the random variables x and y, in order to get the correlation coefficient we need the product moment also.

So, expectation of X into Y that is equal to 2 x y, now here we can choose the order of integration we may do d y d x. The range of y is from 0 to x, and the range of x is from 0 to 1 so this is equal to 2 times integral 0 to 1. Now firstly, we are integrating with respect to y. So, we get y square by 2, so this term cancels out and that gives us x square. So, we are left with x cube d x that is equal to 1 by 4.

So, covariance term between x and y that is equal to expectation of X Y minus expectation X into expectation of Y; that is equal to 1 by 4, minus 2 by 3 into 1 by 3 that is equal to 1 by 4 minus 2 by 9. So, once again it is equal to 1 by; so 9 minus 8 by 36, so 1 by 36. So, the coefficient of correlation is equal to 1 by 36 divided by 1 by 18 both variances of x and y are 1 by y 18. So, sigma x into sigma y will be 1 by 18, it is equal to half. So, this shows that there is a moderate degree of relationship, moderate degree of positive linear relationship between the random variables x and y, we also introduce the concept of the joint moment generating function between the random variables x and y as follows.

(Refer Slide Time: 27:07)



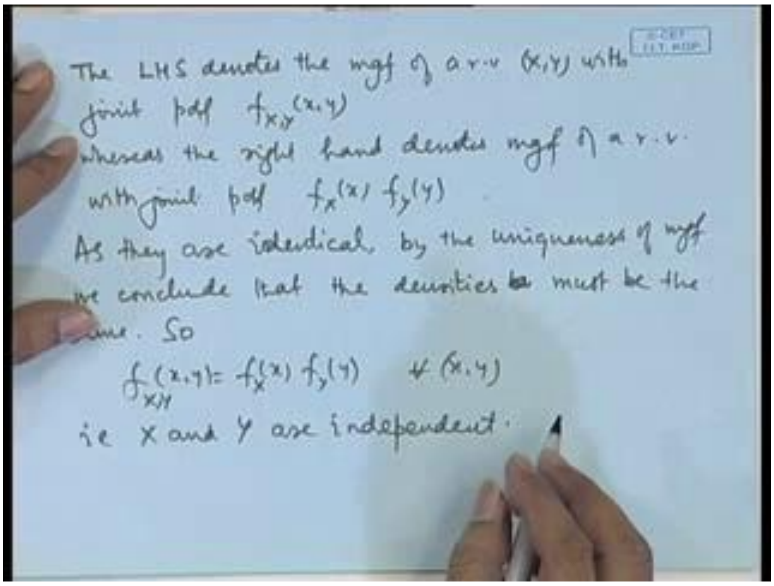
Joint moment generating function between X and of X and Y; so we use the notation M X, Y at the point s, t, it is expectation of e to the power s x plus t y provided the expectation exists in a neighborhood of 0; at 0, 0 this always exist and it is equal to 1, so in a neighborhood of 0, 0 this should exist. Now the nature of this term is nice, it suggests certain properties for example, it is equal to expectation of e to the power s x into e to the power t y; so the first thing that we observe that if the random variables are independent, it will become product of the expectations of e to the power s x and e to the power t y, which are nothing but the individual moment generating functions of x and y.

However, because of the uniqueness property of the mgf's we have a stronger result here; X and Y are independent, this implies and implied by that M X, Y s t is equal to M X s M Y t for all s, t. In order to prove this let us consider say the case of continuous random variable with joint pdf say f x, y and marginal pdf's say f x and f y. Now, here let us notice that if the random variables are independent then this expectation of a product will be equal to the product of the expectations, so the joint mgf will be equal to the product of the individual or marginal mgf's, let us look at the converse.

Let this 1 relation be true, then this implies that integral e to the power s x plus t y, f x, x y, d x d y is equal to integral e to the power s x, f x d x into integral e to the power t y, f y d y for all s t. Now the right hand side we can write as product of the 2 integrals can be written as a combine integral e to the power s x plus t y, f x into f y d x d y.

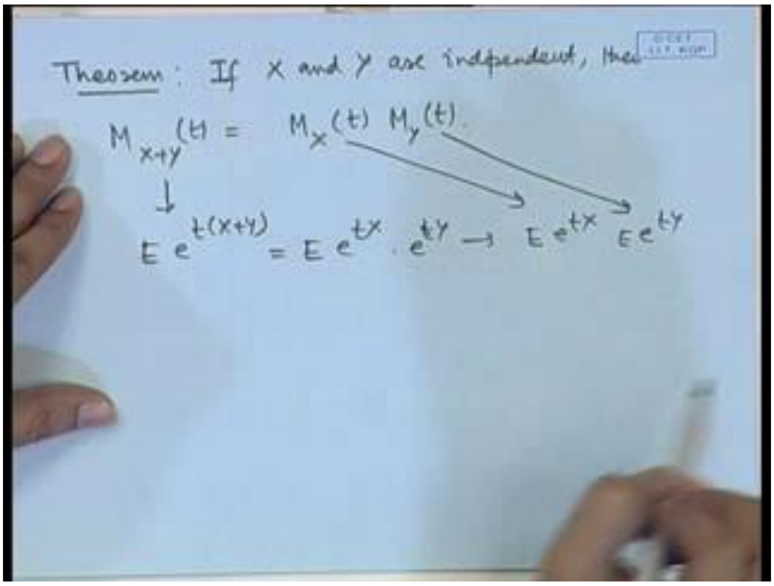
Now, note here the left hand side denotes the joint mgf of random variables x and y when the pdf is f x, y. The right hand side denotes the joint distribution the joint mgf of the random variables x and y, when their joint distribution is given by f x into f y and this statement is true for all s t. So, by the uniqueness of the mgf the 2 densities must be same.

(Refer Slide Time: 31:40)



So, the left hand side denotes the moment generating function of a random variable X,Y with joint pdf f x,y x, y whereas, the right hand side denotes mgf of a random variable with joint pdf f x into f y. As they are identical by the uniqueness of mgf we conclude that the distributions must be the same. So, we should have f x y, x, y is equal to f x into f y for all x and y that is X and Y are independent. So, this is quite a strong property and that is true because of the uniqueness property of the moment generating functions. So, independence of random variables can also be proved through the consideration of the joint mgf.

(Refer Slide Time: 33:36)



We end this one with another result, if X and Y are independent then M x plus y that is the moment generating function of random variable x plus y, which is equal to the product of M X and M Y. The proof of this is almost trivial, the left hand side is expectation of E to the power t x plus y and so if independence is there this can be written as t x into e the power t y, which will become expectation of e to the power t x into expectation of e to the power t y, which will be this term and this term respectively.

This result is extremely useful in obtaining the distributions of sums of various random variables, because if we know the mgf's of those random variables then we can identify the distribution of the sum by identification of the mgf of the sum as the mgf of certain random variables. So, we will consider the bivariate normal distribution in the next lecture.