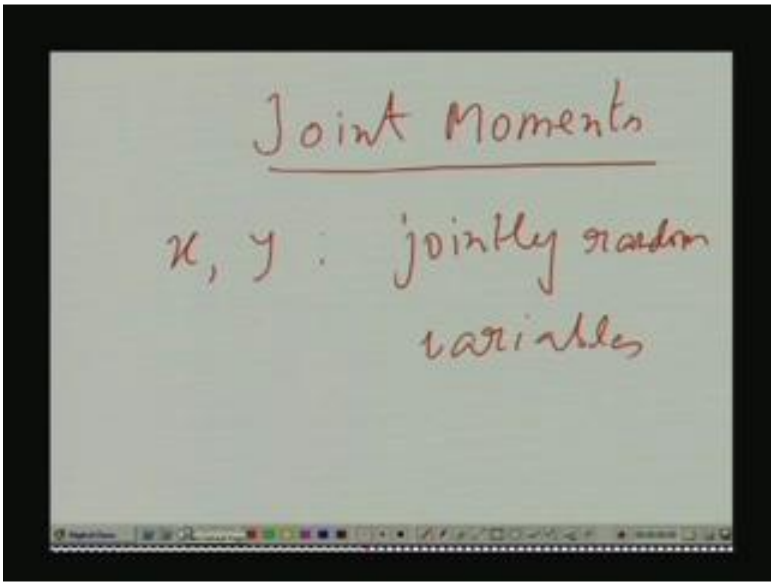
Probability and Random Variables
Prof. M. Chakraborty
Department of Electronics & Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

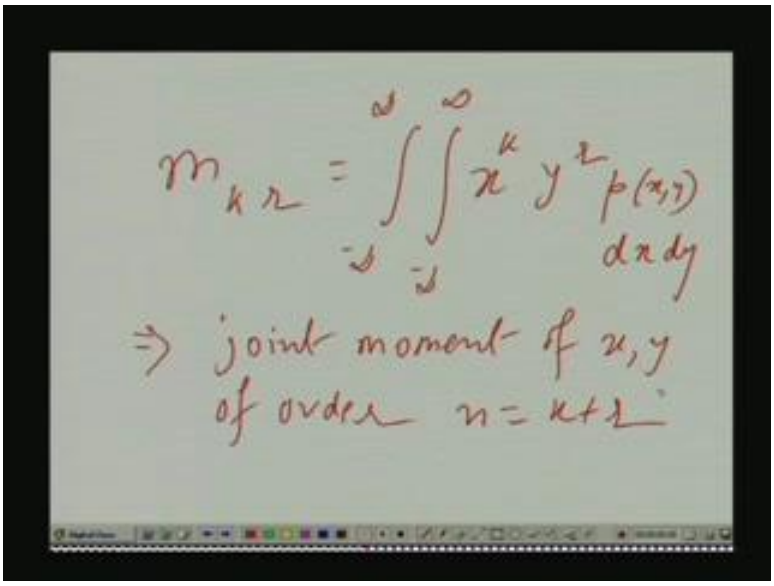
Lecture - 17
Joint Moments

(Refer Slide Time: 00:59)



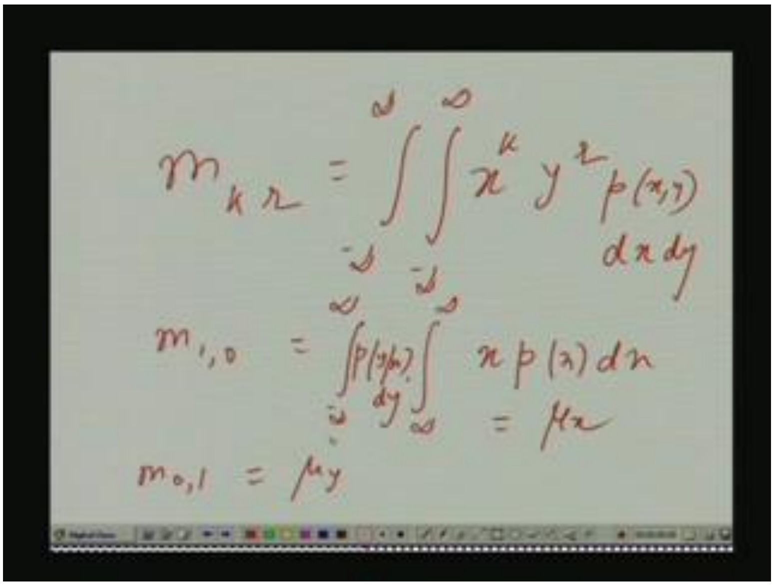
So, today we discuss… We discuss this first and then we will go up to joint characteristic functions. We are given jointly random variable.

(Refer Slide Time: 01:36)



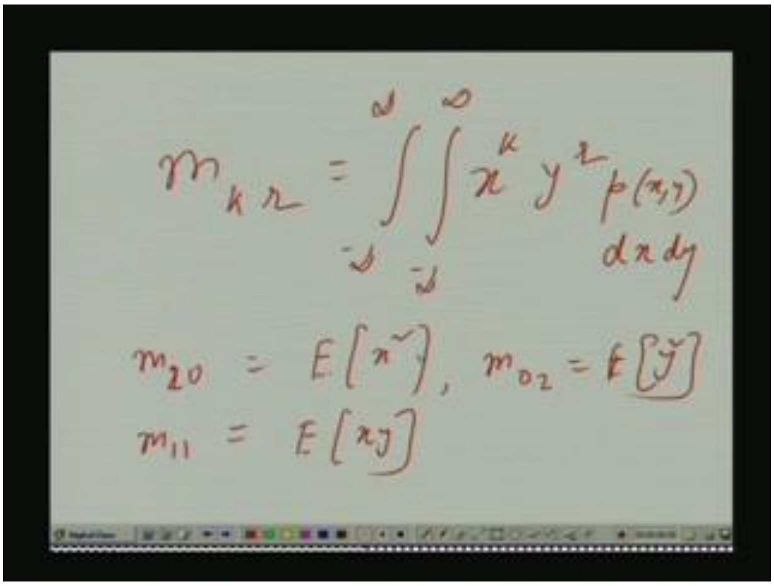
Then, we define the moment m k r as… where p x, y is the joint probability density of x, y – x to the power k, y to the power r. And this will be called joint moment of x, y of order n is equal to k plus r. So, it is very much similar to the moment that we dealt with for a single random variable case; it is a generalisation of that to two variables. Certain things follow easily

(Refer Slide Time: 02:47)



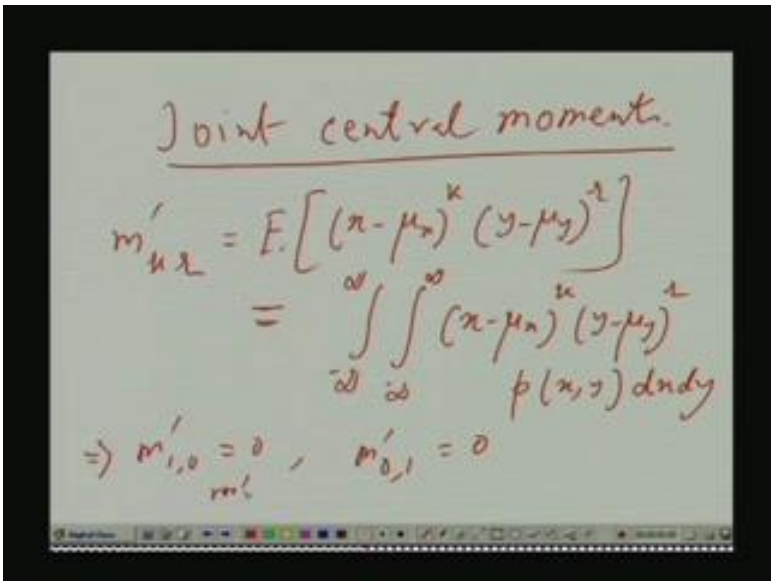
What is m 1, 0; that means x to the power 1, that is, x; y to the power 0, that is, 1; x, p x comma y dx dy. And p x comma y can be written as, that is… p x comma y can be written as… I mean you can write the entire thing like this. x p x dx; p x comma y will be written like… that is, p x comma y is p x times p of y given x. So, this integral is 1; whereas, condition theory x; total probability of y – taking values within minus infinity to infinity; that is equal to 1. And this is the mean value of x. So, m 1 0 is mu x. Similarly, m 0, 1 will be mu y; mu y means the mean of y. Then…

(Refer Slide Time: 04:20)



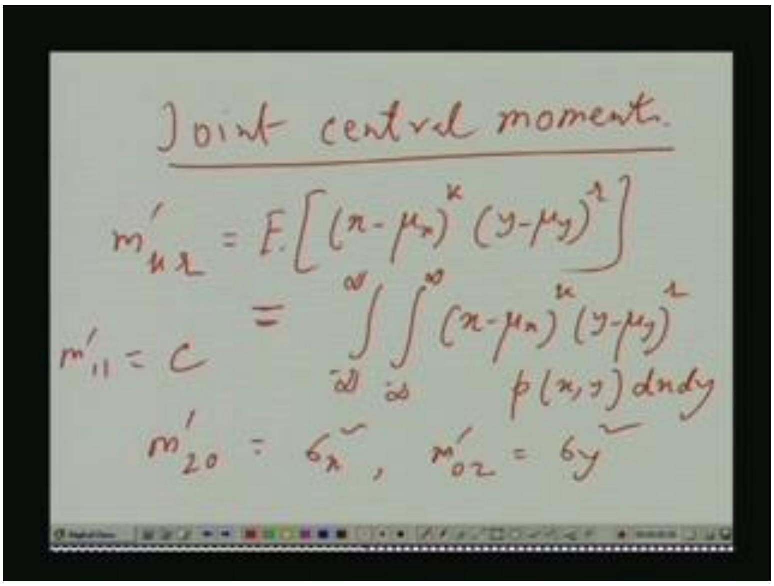
Then, what is m 1 1? Or say let us start with m 2 0. Clearly, it will be expected value of x square. Why? y 0 is 1 x square p x comma y dx dy; p x comma y will be written as before like p of y by x and then p x. This integral will be 1. And we get expected value of x square. Similarly, m 0 2 is E of y square. And m 1 1; that is, expected value of x y; that is, these are all second order moments – second order joint moments.

(Refer Slide Time: 05:19)



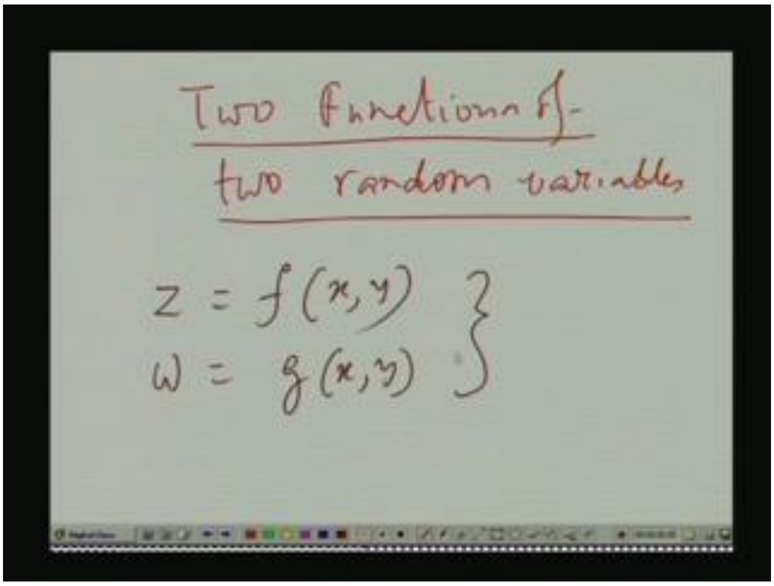
Similarly, you can have joint central moments. Here you can call it m prime k r. It is actually expected value of… So, earlier it was x; now, it is not x, but x minus its mean mu x raised to the power k – y minus its mean mu y raised to the power r or equivalently… So, you ((Refer Time: 06:30)) What is m prime 1 comma 0? y minus mu y to the power 0 is 1; x minus mu x p x comma y – that we again as before write; write as before – p x and y by x; p y by x integrated with respect to y from minus infinity to 1 because… And this integral will be x minus mu x time p of x. Since mean of x is mu x, that will be 0; So, this will be equal to 0. Similarly m prime 0 comma 1 also will be 0. How about m prime? I will just erase it.

(Refer Slide Time: 07:23)



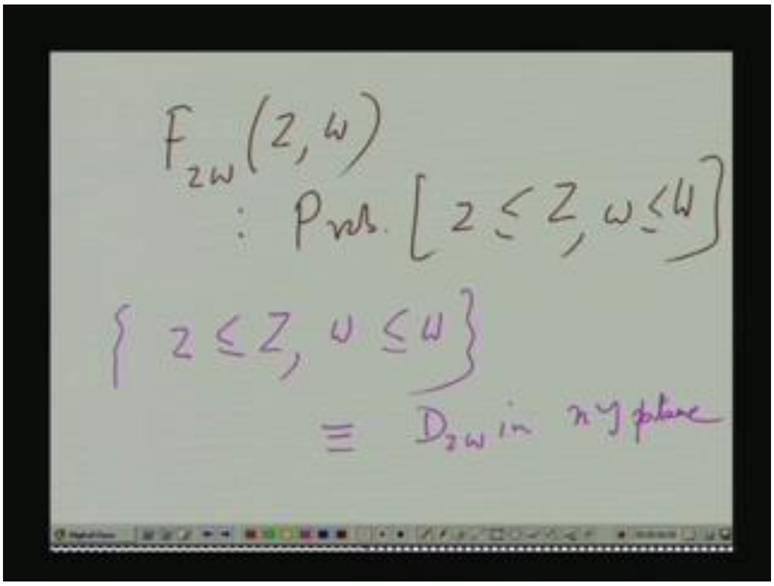
How about m prime 2 0? That is nothing but expected value of x minus mu x whole square. So, that is the variance sigma x square; and m prime 0 2 – this by the same reasoning, sigma y square. And m prime 1 1 will be what? Expected value of x minus mu x times y minus mu y; which is nothing but the covariance; so that is equal to C. We will take up some example. But, before that we come back to some topic, which I had left out; but, I need the results from the topic for subsequent treatment here in the context of moments and characteristic functions. So, I will come to that topic first; that is, earlier we had considered joint random variables say x and y at a function – a single function of this random variables; that is, maybe z is equal to f of x comma y; but, now, we will be considering two such functions: one is z; another is w; two defined functions. But, both are joint functions of x and y. So, both z and w will be random variables.

(Refer Slide Time: 08:47)



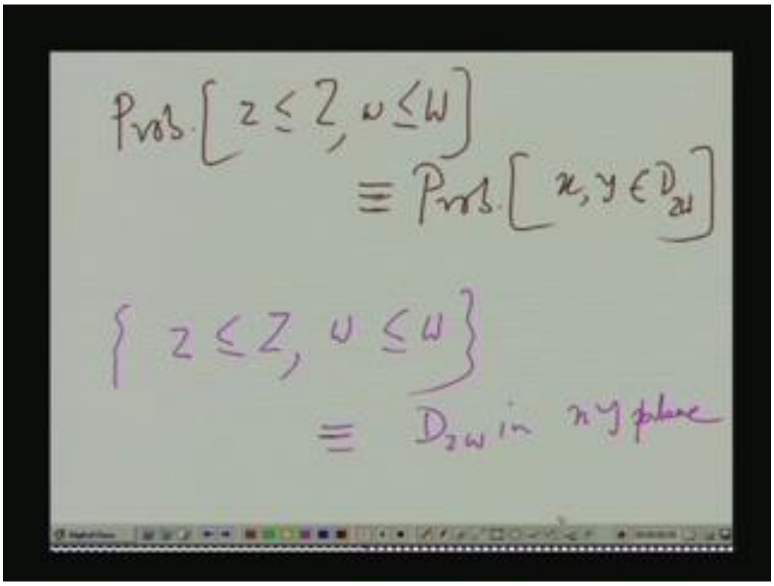
So, we are considering two functions of two random variables; that is, we are given z is equal to f of x comma y; and w is equal to g of x comma y. Obviously, z and w – since they are functions of x and y, in general, they are also jointly random. So, I can have joint probability distribution and joint probability density functions for w and z.

(Refer Slide Time: 10:09)



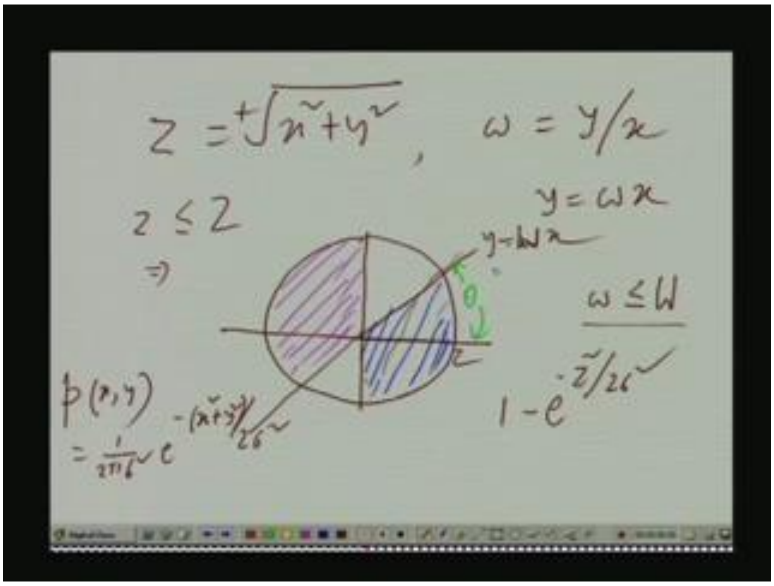
So, I am suppose interested in finding out F z w of z comma w; that is, the joint distribution of z, w; that is, probability of z less than equal to capital Z; w less than equal to capital W – total probability This is the probability – joint probability distribution. We have to evaluate these in terms of the given joint density of x and y or given joint probability distribution of x and y. So, this set we consider… So, this covers… This actually describes an area in the zw plane; where, for any point, z comma w, we have z less than equal to capital Z; w less than equal to capital W. This is equivalent. This corresponds to some area D z w in xy plane; that is, it means that, for any x comma y belonging to this area – D z w, the corresponding small z and small w – they satisfy this relation; which means the probability of z and w simultaneously satisfying these inequalities is same as the joint probability of x and y falling within or remaining within this area D z w in the xy plane; that is…

(Refer Slide Time: 12:50)



That is… Now, let us take an example.

(Refer Slide Time: 13:52)



Suppose it is given that, z is equal to positive square root of this and w is as an example y by x. So, we have got one function z – a function of x, y; and another variable w, which is a function of x, y again; but, two different functions. What does these corresponds to? That is, here if small z is less than equal to some capital Z, these amounts to a circle of radius capital Z. So, the whole area inside the circle corresponds to z remaining less than equal to capital Z. That is very obvious. But, how about this? This means y is equal to wx. For our example, suppose w is positive; just because w can be positive or negative; but, just for our example, suppose w is positive; that means we just draw a straight line y equal to wx. So, that means, w capital Wx; that means, w less than equal to capital W means what?

First, consider this side. When x is positive and small w is less than equal to capital W, then y is any point below this line, because on this line, y is capital W times x. But, for any small w less than equal to capital W; say less than capital W, the corresponding wx will be below this. So, that means, for positive x, any y here below this line will give raise to ratio y by x equal to small w; that is, less than capital W. If we include this line also in this area, then it will be less than equal to… So, we have got this area for positive x. I am taking the overlap between this circle, because the circle corresponds to this inequalities – small z less than equal to capital Z. And the whole region below this straight line was giving raise to small w less than equal to capital W. But, I am only taking the overlap between the two: the circle and that region. So, this shaded region comes.

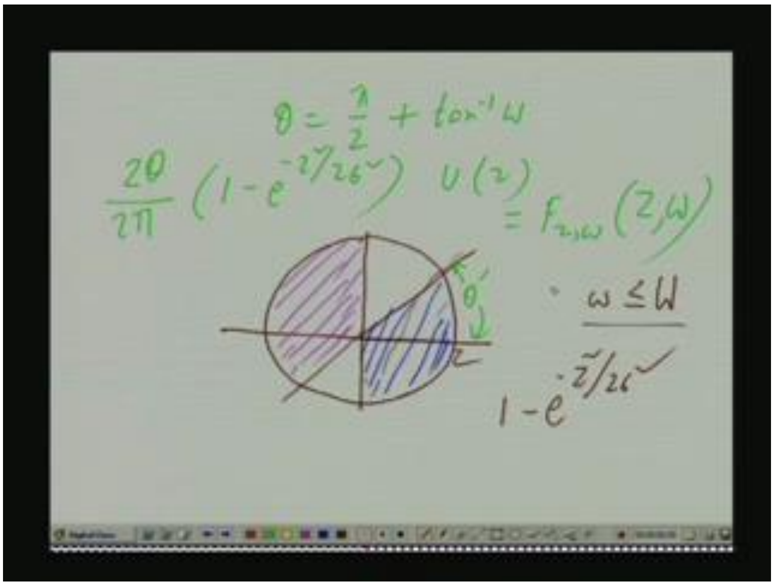
Now, you go to the negative x. You get x is negative; then wx also is negative, because w i for example sake, I took w to be positive. So, wx also is negative. So, minus of y on this straight line – on this straight line, I mean, minus of some negative y divided by negative x till the ratio is positive, you get w. But, if you go up, then even if y is negative, its magnitude is less than earlier and x remains same – negative, but same; which means the value of the ratio becomes less. And if you go to this quadrant, I mean the second quadrant, then y has already become positive; and y by x; where, x is negative, is a negative number; which is less than w. So, in this case, as we go above this line, we come across points; where, the ratio y by x is again less than capital W. So, joint probability of z less than equal to capital Z and small w less than equal to capital W is same as the joint probability of x comma y falling in this region and falling in this region – falling in the two-sided regions.

Now, up to this, it is okay; but, now, to compute that probability, I now need the probability density function – joint probability density function for x and y. Suppose it is given that, x, y – they are circularly symmetric and Gaussian; or, equivalently called normal; that is, p of x comma y – suppose is given to be the circularly symmetric jointly Gaussian below mean… In this case, first, what is the probability of x, y lying within this entire circle? We have worked out this earlier. That is why we dealt with a single function of two random variables. So, Z equal to square root of x square plus y square; and x, y are given to be circularly symmetric jointly random variables. In that case, we found out the total probability of x and y pair falling within this circle.

What is that expression? That expression is… That was 1 minus e to the power minus z square by twice sigma square. That was the probability of the pair x comma y lying within a circle of radius capital Z; but, now, I am not interested in the entire circle; rather, I am only interested in the shaded area. So, how much is the shaded area? For that, I consider this angle – call it theta. And this angle is pi by 2. And from symmetry, the two areas are same obviously. If this angle is theta, this is also theta. This one-fourth of the circle; this is also one-fourth of the circle. These two areas are same of symmetry. So, whatever will be the area of this shaded portion, which is shaded by blue; just twice that will be this area of overlap. And what is this area? First, how much is this angle?

This angle is pi by 2 plus theta. What is theta? Theta is tan inverse capital W. Obviously, theta is tan inverse capital W, because tan theta is the slope; slope is capital W; slope of this straight line is capital W. So, what is theta? Theta is nothing but tan inverse capital W. So, this – a total angle is… Let me erase some portion.

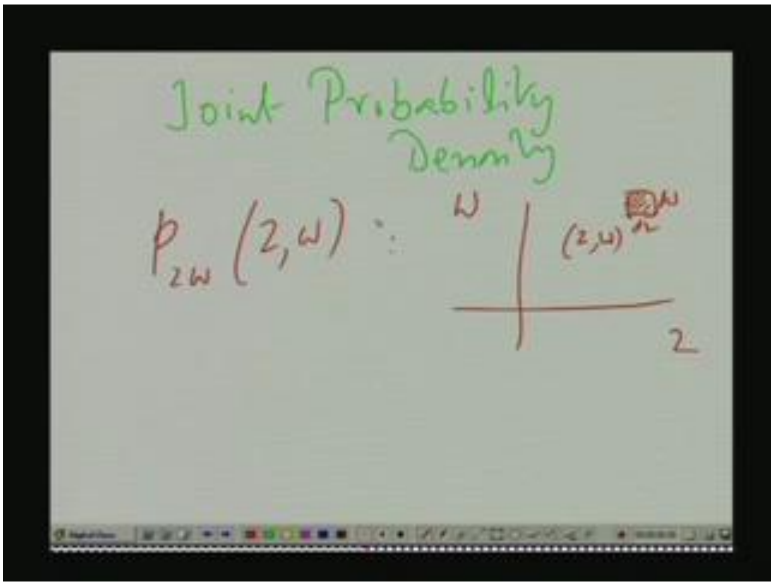
(Refer Slide Time: 22:18)



So, this angle is nothing but pi by 2 plus tan inverse w. Let me call it theta prime. So, theta prime is tan inverse w. This total angle – I call theta. And total angle around origin is 2 pi; out of which, theta here and theta here. So, twice theta. So, that means what is the probability of x comma y falling in these two shaded regions? It is nothing but twice theta: one theta from the blue-shaded region, another theta from the pink-shaded region. So, 2 theta divided by twice pi.

That is the ratio times 1 minus e to the power minus z square by twice sigma square. You can also put U z because z can be 0 or positive; z cannot be negative. That is the radius of the circle. Why 2 theta by 2 pi? Because 1 minus e to the power minus z square by twice sigma square; that is, the total probability of x, y remaining with the entire circle. But, I am now not interested in the entire circle, but just a segment of it. What is the proportion of that segment vis-a-vis the total area? That is nothing but twice theta by twice pi, because two angles are theta and theta – 2 theta. But, the total angle is 2 pi. So, the ratio is just twice theta by 2 or twice pi; that is, theta by pi. So, that is your… You can write that is equal to the F z w. So, that is an example for computing the joint probability distribution function. How about the joint probability density?

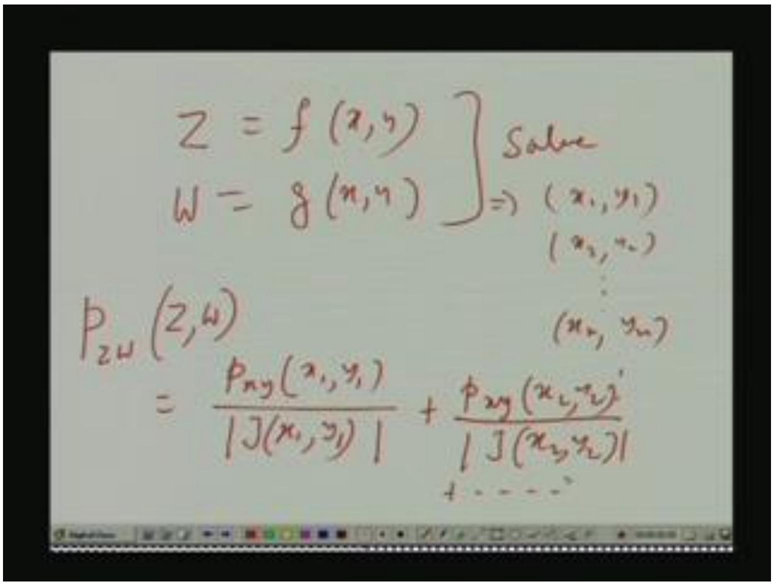
(Refer Slide Time: 24:58)



That is, p zw capital Z, w. What is the physical meaning of it? In the zw plane, there is a point capital Z, capital W; around that, there will be a small area. So, the probability of Z comma W falling in this area will be equal to this function p z w of Z comma W times dz dw. This is dz; this side is dw. So, dz dw. This function is then called the joint probability density; that is as per definition. But, as I told you, z is function of x comma y – f of x comma y. And w is a function of again x comma y – say g of x comma y. And therefore, Z and W are jointly related. And therefore, we consider the joint density.

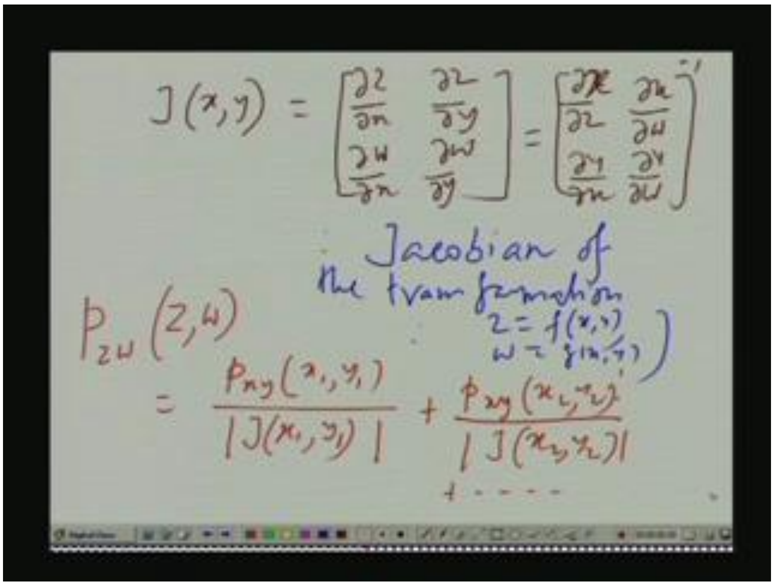
Now, what is the joint density? Given the joint probability density expression for x, y; and also the two functions are of course given; that is, Z is equal to f of x, y; where, f is given; W equal to g of x, y; where, g is given. And the joint probability density of x, y; that is, p x, y is given. Then, we have to evaluate this. Now, while the procedure remains same as before; before means when we had just one variable at hand. I will not go into the derivation, because this derivation is little more complicated than earlier. I will only quote the result; and so an application of it.

(Refer Slide Time: 26:57)



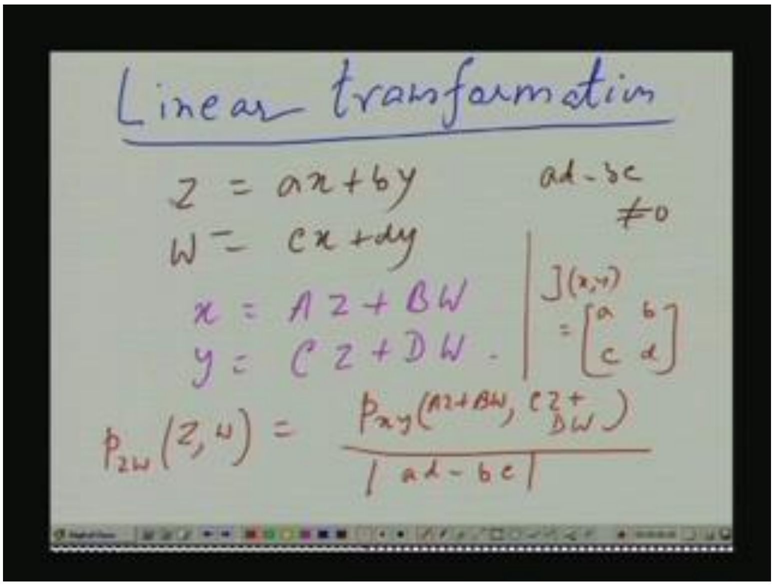
Here suppose capital Z is given and capital W is given; you solve them; you get solutions like x 1, y 1; x 2, y 2; dot dot dot; x n, y n; dot dot dot dot. So, for x equal to x 1 and y equal to y 1; so this is satisfied. f of x 1, y 1 is Z; g of x 1, y 1 is W. Similarly, f of x 2, y 2 is Z; g of x 2, y 2 is W and likewise; so on and so forth. In that case… So, for what… I repeat again; what we first do? You started with some known capital Z and known capital W. I first put them in this equation; solve them. Suppose I get some discrete solutions for x, y; there is either x 1, y 1 or x 2, y 2 or dot dot dot dot x n, y n; and so on and so forth. So, these values are known. They are all available in terms of given Z and given W. Then, what I do; the formula is this – is equal to p x y at x 1, y 1. Mind you x 1 and y 1 – they are now available in terms of capital Z and W. So, this indeed actually becomes a function of z and w. This divided by determinant of a matrix. So, that matrix is called the Jacobian. So, I will tell you what is a Jacobian. Then, again put the second solution divided by the corresponding Jacobian and so on and so forth. This is the result; where, let me give the expression for the Jacobian.

(Refer Slide Time: 29:20)



We can either write it like del z del x, del z del y, del w del x, del w del y. It can be shown that, this is also the inverse of… This is called Jacobian of the transformation. Which transformation? That is Z equal to f x, y and W equal to g x, y. This transformation; it takes x, y pair; gives you z, w pair. So, this transformation is a Jacobian with the matrix actually defined like this. So, this probability density means evaluate p x, y at one solution – x 1 comma y 1; which is available in terms of Z and W divided by the determinant of the Jacobian at that x 1 and y 1. So, again it is a function of Z and W. Then, again do the same for point x 2, y 2 and so on and so forth. This is a result. We will not prove it, but we will just see some examples, where we use this result.

(Refer Slide Time: 31:29)



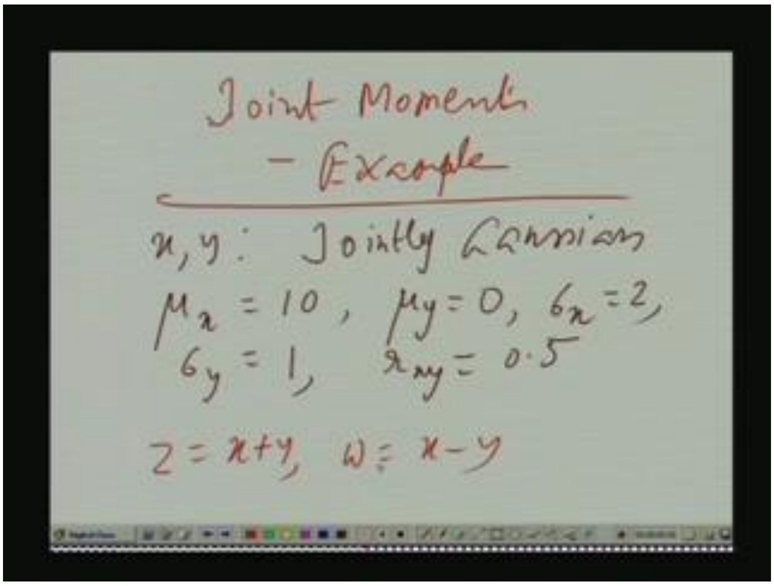
Suppose we consider a thing like this; that is, suppose it is given that, Z is equal to some ax plus by – linear and W is cx plus dy. And it is given that, the determinant of this matrix ad minus bc not equal to 0. So, you can easily solve this; you can easily solve this solution; x also will be available as a linear combination of Z and W. Y will be available as a linear combination of Z and W. So, you can even write the solutions like… And how many solutions? For a given determinant Z and capital W, these are linear equations; only one solution. No such x 1, y 1 and x 2, y 2 and x 3, y 3 and all that. So, using our previous result, it will become obvious that… Is what? This is equal to… What is the Jacobian by the way? Del z del x is a; del z del y is b; J x comma y. Del z del x, that is, a; Del z del y, that is, b. Then, del w del x – c; and del w del y – d. So, the determinant of that is ad minus bc. And we take the mod of that by the way; not only just determinant, mod of that.

I must qualify my previous statement. It is after taking the determinant; determinant can be negative or positive; but, take the mod. Obviously, it cannot be negative, because probability cannot be negative. So, mod of that… So, it is then p x y – only one solution; one is A, Z plus BW; another is CZ; C means capital C – plus DW divided by mod ad minus bc. Further, suppose it is given that x and y – they are jointly normal; then our claim is Z and W also are jointly normal. Why? Obviously, p x y – this is given to be jointly normal. So, it will have an exponential form; where, this will be squared and this will be squared. Of course, the corresponding variances and other things will come up;

but, the square term involving AZ plus BW and again a square term involving CZ and DW represent. And when you square them up, then you call it the terms z with power z square. So, you get one expression for that. Again, you collect the terms involving W square; get another expression. And then there is a cross term involving the product ZW with some coefficient. So, that is… And the whole thing has to be a probability density. So, that is the typical form of a Gaussian density function; where, you have got an exponential form as a joint density. There will be a square term involving one term; square term will be another variable; and a product involving the two variables.

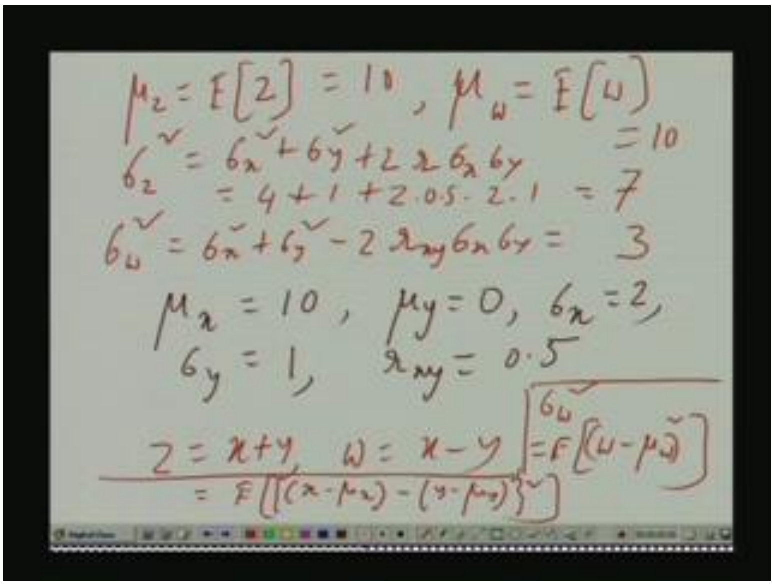
Of course these are ((Refer Slide Time: 35:49)) some coefficients. They will determine the new variance and new mean. So, obviously, it follows that, under such linear transformation, if x and y are given to be jointly Gaussian in the beginning, then Z, W also are jointly Gaussian. Obviously, what is the mean of Z? That is a times mean of x plus b times mean of y. What is a mean of W? C times mean of x plus D times mean of y. Similarly, you can find out the individual variances of Z, W. You can find out the correlation – I mean the expected value of the product ZW; just multiply the two. All these are possible. So, those will be appearing in the joint density expression. And we will see that, it corresponds to the joint Gaussian function.

(Refer Slide Time: 36:57)



So, then I will come back to our previous example involving moments. Again back to joint moments. Suppose as before, x and y – they are given; they are jointly normal; they are jointly normal. We consider the jointly Gaussian. It is given that, mu x is equal to 10; mu y is equal to 0; sigma x… Not sigma x square, but sigma x equal to 2; sigma y is equal to 1; and the correlation coefficient r x y is equal to 0.5. And you obtain two random variables: z and y – Z and W as a function of x and y; in fact, as just linear combinations of x and y; ((Refer Slide Time: 38:21)) given that, Z is equal to x plus y; W is equal to x minus y. What we have to find out is the moments for Z and W. Obviously, mean of Z and mean of W – very easy to find out, then variances of Z and W and the correlation coefficient involving Z and W. This we have to find out.

(Refer Slide Time: 39:03)

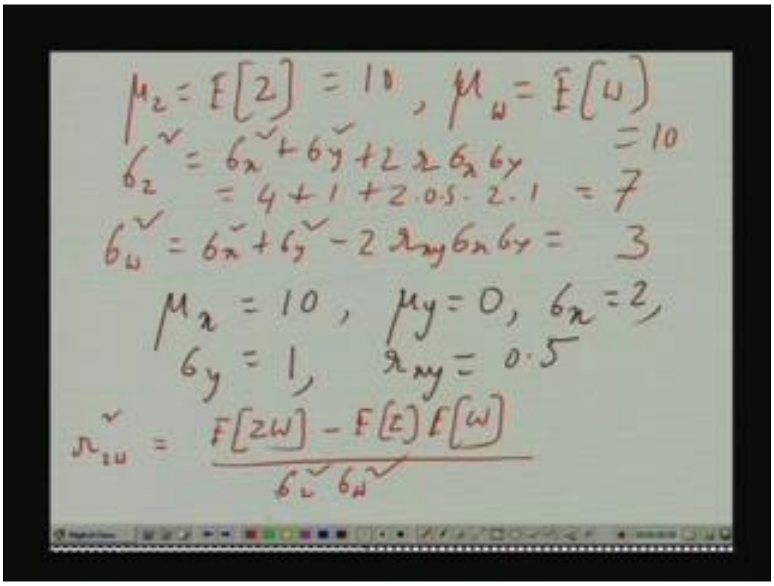


So, what is mu z? That is, E of Z – that is nothing but E of x plus E of y. E of x is 10; E of y is 0. So, this is 10. Similarly, again 10 obviously, because E of x, which is 10 and E of y is 0. So, 10 minus 0 – 10. So, these two moments are done. Then, what is sigma z square? We had already seen earlier that, if x and y are added, the corresponding variance is what? That was sigma x square plus sigma y square plus twice r sigma x sigma y. Sigma x square is 4; sigma y square is 1 and twice r; r is r xy. So, 2 into 0.5 into sigma x 2 sigma y 1 – 4 and 5; and then 2. So, it is 7.

Similarly, what is sigma w square? E of whole square of this. So, that is sigma x square plus sigma y square minus E of x, y. What is E of x, y? What is sigma w square? Sigma w square is expected value of w minus its mean whole square; that is, you can say that, sigma w square is expected value of w minus mu w square. Now, w minus mu w; w is x minus y. And what is mu w? mu w is mu x minus mu y. So, you can proceed like the way we proceeded in the case of z equal to x plus y. So, one term is you can write like this – x minus mu x minus y minus mu y whole square. Mu w is nothing but mu x minus mu y; w is nothing but x minus y. So, I will just substitute it here – squared up. So, one term will be coming from the first term; which will gives raise to sigma x square. Another will be coming from this term – square of this term expected. So, that will give raise to sigma y square. And there will be one more term – minus twice expected value of the product of these two, which is the covariance between x and y.

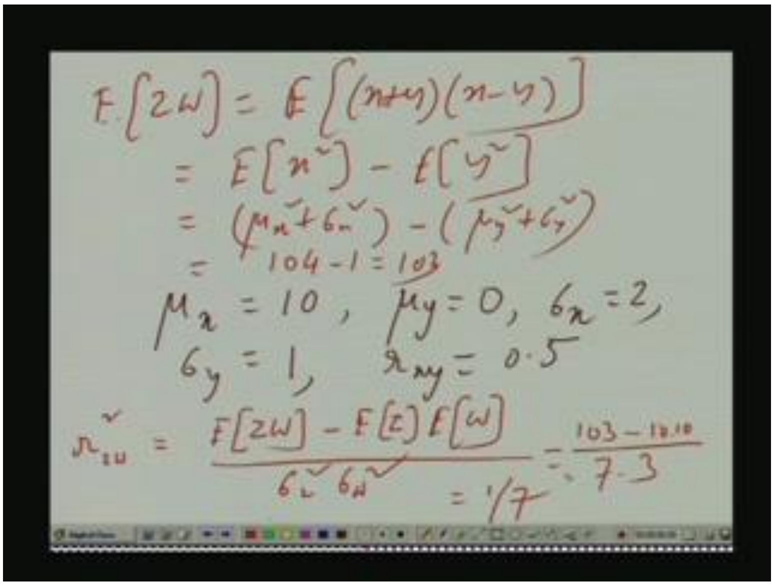
And, what is covariance? Covariance is nothing but r x y sigma x sigma y. So, if you put them back here, what you get? What is sigma x square? That is 4; sigma x square is 1 – 4 plus 1 – 5 minus twice 0.5; that is, 1; sigma x is 2; this is 1. So, 2... So, 5 minus 2, which is 3. And lastly, we have to find out the correlation coefficient between Z and W.

(Refer Slide Time: 44:20)



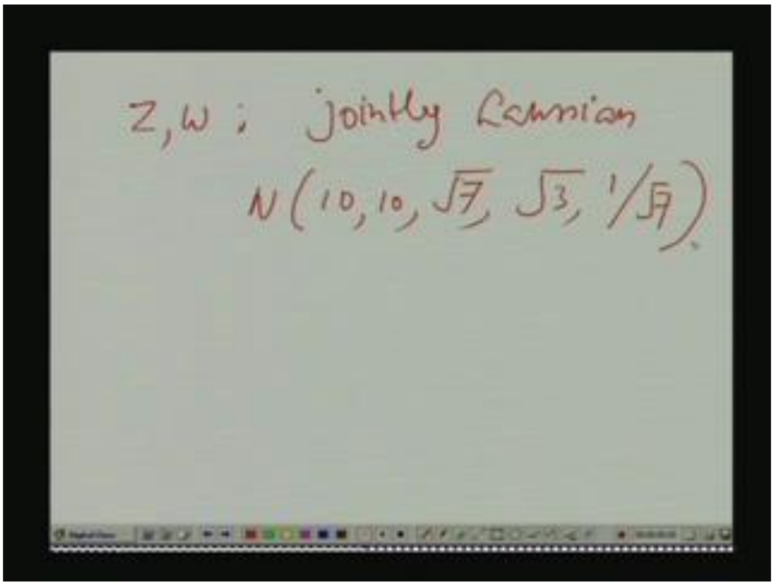
Now, what is the correlation coefficient? That is covariance divided by respective variances. That is, I am taking the square on both sides. This was coming to be the square of covariance, so E of Z minus mu Z times W minus mu W. You can expand it and you will get this thing – minus E Z, that is, mean of Z E of W divided by sigma Z square sigma W square. What is E Z W? This we have to find out. What is E Z W?

(Refer Slide Time: 45:36)



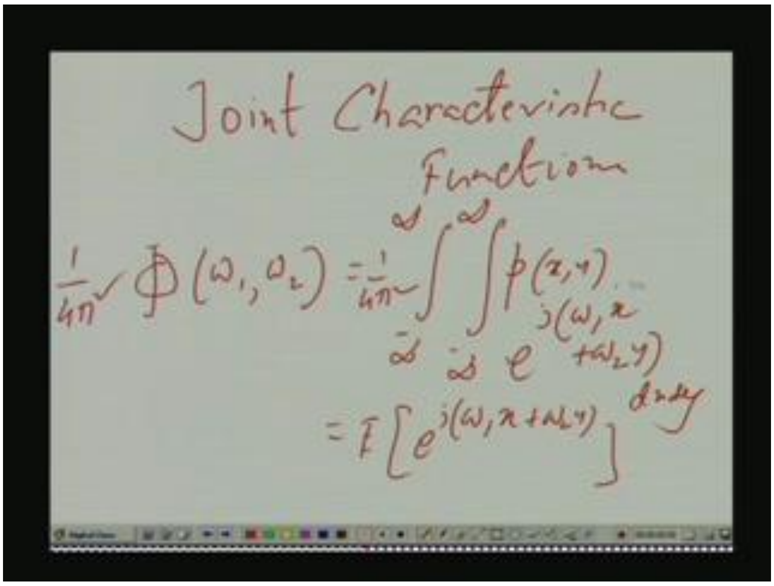
Now… Break it up; x square and minus y square So, E of x square minus E of y square. What is E of x square? It is nothing but we have seen mu x square plus sigma x square. Similarly, mu y square plus sigma y square. What is mu x? 10. So, 100 plus 4; so 104 minus – mu y we have seen is… If a mu y is given to be 0; sigma y square is given to be 1. So, 104 minus 1; which is 103. So, we put those here – 103 minus E z, E w; E z we have already find out. E of z was 10; E w was 10 divided by sigma z square and sigma w square; sigma z square is 7; sigma w square is 3. So, 103 minus 100; that is 3; 3 by 7 into 3. So, ((Refer Slide Time: 47:08)) be 1 by 7; that means, in short…

(Refer Slide Time: 47:25)



In short, we can say they are jointly Gaussian like this. N stands for normal; and normal is same as Gaussian ((Refer Time: 47:47)) is the mean. So, mean of x is… mean of z is 10; mean of w is 10; then the variances – not variances, standard derivation to square root of the variance – positive square root; square root 7 for sigma z; square root 3 for sigma w; and the correlation coefficient, which is 1 by square root 7. Square of correlation coefficient was 1 by 7. So, this is 1 by root 7.

(Refer Slide Time: 48:35)

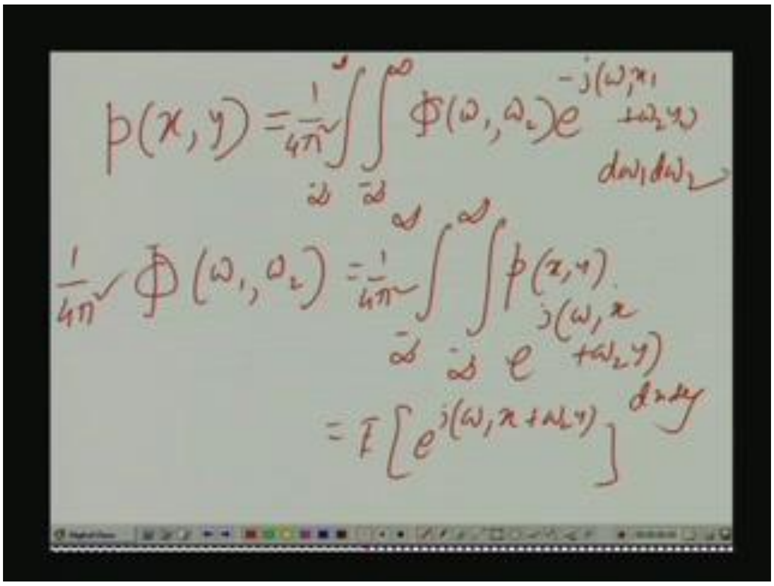


Now, we go to the next topic, that is, joint characteristic function. We have already seen what is a characteristic function in the case of single variable. But, now, we have got two random variables say x and y and we have got their joint density function P of x comma y. So, obviously, we will have a Fourier transform with two frequencies: omega 1 and omega 2. So, we define the joint characteristic function as… e to the power j omega 1 x plus omega 2 y dx dy. So, given p of x comma y… Or, you can equivalently see also; that is nothing but expected value of e to the power j omega 1 x plus omega 2 y. So, given p x comma y, we can find out phi by this formula.

And, given the characteristic function, we can find out p of x comma y by the inverse formula. That is again obtained by recalling the forward and backward Fourier transform relations; that is, direct and indirect inverse – direct and inverse Fourier transform relations.; that is, you can see that, if you multiply both sides by 1 by 4 pi square, then this is nothing but inverse Fourier transform of p x comma y and omega 1 comma omega

2. See there is a plus sign here; no minus sign; no minus sign. We recall in the one variable case, the inverse relation was 1 by 2 pi integral – some function of say x e to the power j – positive – e to the power plus j omega x dx. Here we just have two variables: x comma y. So, p x comma y e to the power this – 1 by 4 pi square. So, this is again inverse Fourier transform of x comma y at omega 1 comma omega 2. So, p x comma y is nothing but direct Fourier transform of this quantity.

(Refer Slide Time: 51:35)



That is, p x comma y; p x comma y is 1 by 4 pi square… So, this is the definition. And from the given phi omega 1, omega 2, we can also find out the marginal characteristic functions, that is, just phi of omega 1 and phi of omega 2; phi of omega 1 being the characteristic function of x for the random variable x alone; and phi of omega 2 being the characteristic functions from the random variable y alone. And then we will see that, when two variables are independent – statistically independent, then this joint characteristic function becomes just a product of the marginal characteristic functions and vise versa; that is, if the joint characteristic function is a product of two marginal characteristic functions, then x and y – they are statistically independent. And then we will relate that to convolution and all that. So, that will be for the next class. So, that is all for today.

Thank you very much.

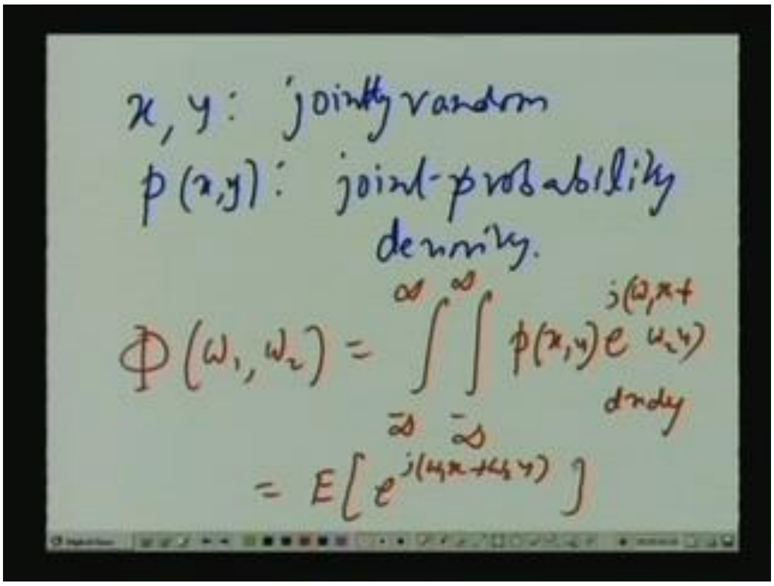
Preview of Next Lecture

Lecture – 18

Joint Characteristic Functions

In the previous class, we ended with just a brief description of what is called joint characteristic functions. So, today we start from there; maybe there will be a little repetition, but that one will be helpful.

(Refer Slide Time: 54:01)

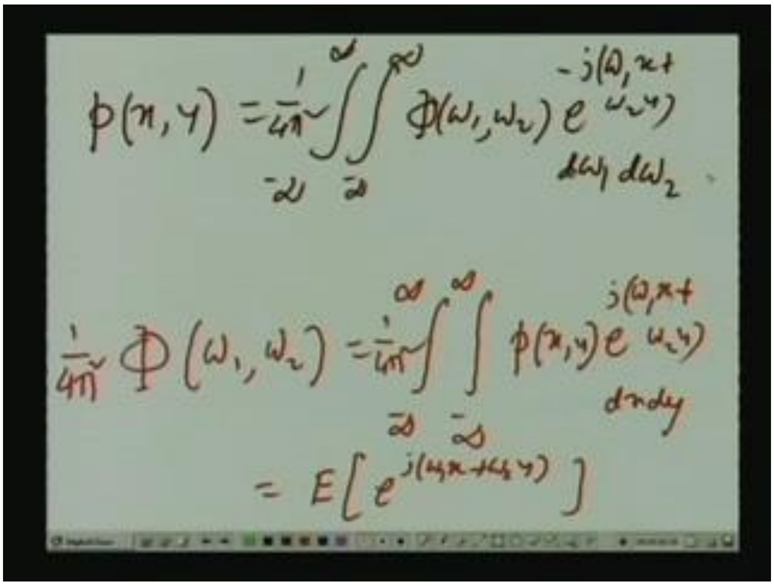


So, here we are given two random variables say x and y – jointly random. Then, we have already seen what is a characteristic functions in the case of a single random variable. So, here we will be simply extending that to the case of two variables. So, here the joint characteristic functions earlier was a function of only one frequency – omega; now, since there are two random variables: x and y involved, there will be two frequency variables: omega 1 and omega 2; and it will be defined like this – omega 1 x plus omega 2 y dx dy. You can also see that, this is nothing but the expected value of this exponential. After all, this is a function of x and y. I wanted to… If I want to find out its expected value, I will simply multiply it by the joint probability density; integrate from minus infinity to infinity both with x and y. So, essentially, joint characteristic functions – phi omega 1 omega 2 is nothing but the expected value of e to the power j omega 1 x plus omega 2 y.

Then, you can also see that… I can write this; if this is given, I can write this as this. Then, multiply left-hand side by 1 by 4 pi square; this is also 1 by 4 pi square. Then, you can easily see that, the right-hand side is nothing but inverse Fourier transform of this function p x comma y – inverse Fourier transform of this function of two variables. So, we have got a plus sign here; e to the power plus j. So, it is not minus, because it is inverse Fourier transform; it is plus j omega 1 x plus omega 2 y. So, there are two frequency variables; integral as usual is from minus infinity to infinity.

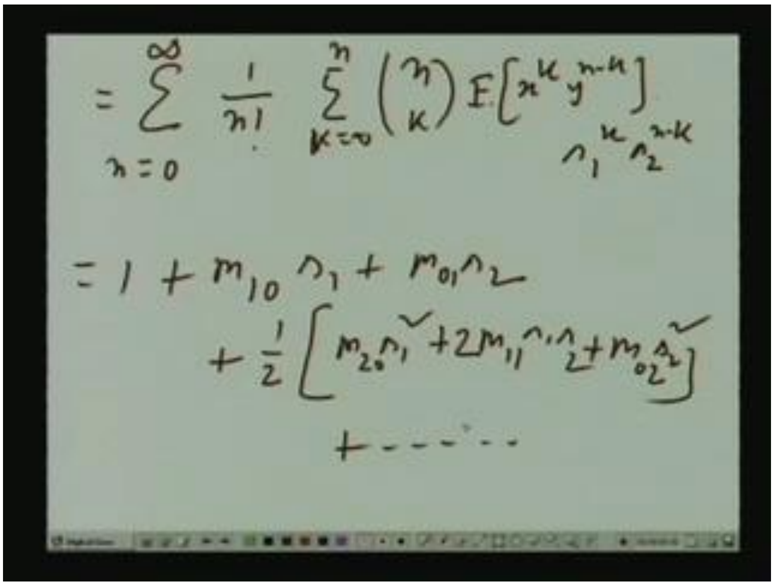
In the case of inverse Fourier transform involving only one variable, we have 1 by 2 pi. But, since there are two variables, it becomes 1 by 4 pi square. If that be the case, then we know that, p x comma y also can be viewed as the direct Fourier transform of this quantity on the left – 1 by 4 pi square of… 1 by 4 pi square times phi omega 1 omega 2.

(Refer Slide Time: 57:19)



So, that means this is a inverse formula; that is, given the characteristic functions – joint characteristic function, now, the minus sign will come – e to the power minus j omega 1 x plus omega 2 y. But, the integral will be with respect to omega 1 and omega 2. Certain things we can see now. Also, one more definition… Let me remove this 1 by 4 pi square now. This was just for explanation purpose. Along with phi omega 1 and omega 2, there is another definition, which also comes out to be useful sometimes. Actually, often phi omega 1 omega 2 is seen to be – in practical cases, seen to be an exponential function. So, instead of dealing with phi as such, it is sometimes better to take logarithm of this.

(Refer Slide Time: 58:33)



And then twice… where, k is 1; ((Refer Slide Time: 58:39)) this 2 1 will become 2; we have factorial 2 divide by factorial 1 – twice E, and now x to the power 1 y to the power 1 s 1 to the power 1 s 2 to the power 1. So, m 1 1 s 1 s 2, so on and so forth. We stop here today. And from this, we will derive some new results and take up some examples in the next class.

Thank you very much.