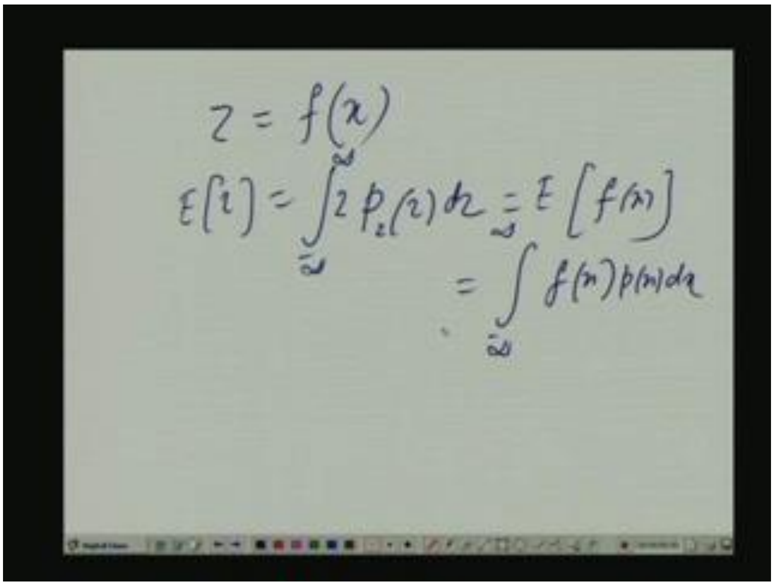
Probability and Random Variables
Prof. Dr. M. Chakraborty
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 15
Correlation Covariance and Related Innver

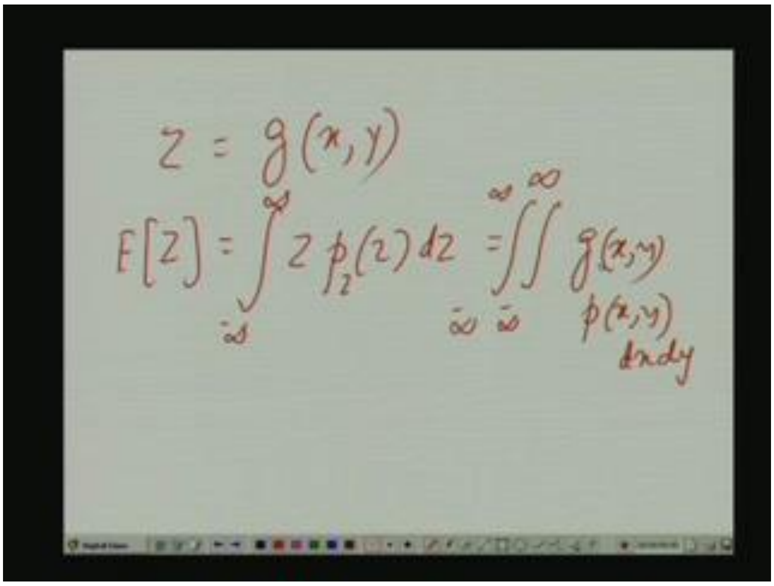
So, as you know, we have been considering the case of joint statistics involving two random variables. We found out the joint probability density and joint probability distribution of the functions of two random variables also to such cases.

(Refer Slide Time: 01:08)



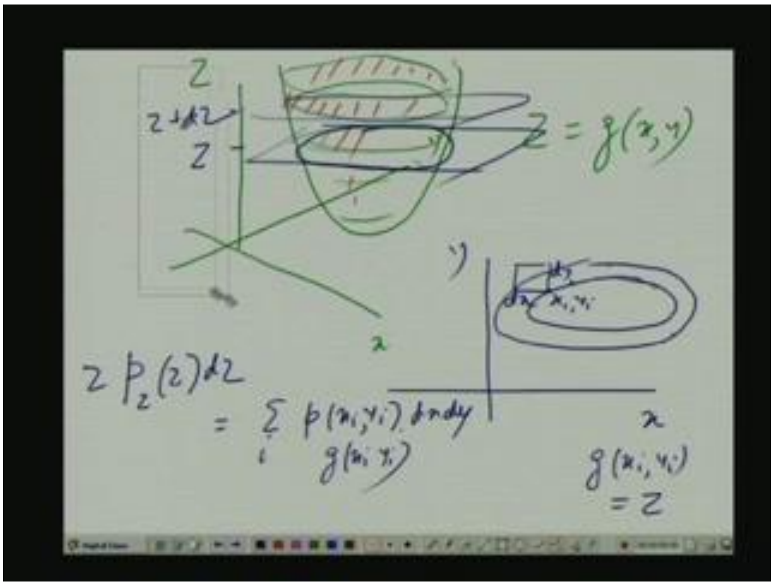
More cases or more examples are possible. More examples are possible, but I did not consider. I think you have got enough idea about how to proceed. Now, we have to do something similar to what we did in the case of function of single random variable. One of them was things like this. If z was given to be a function of random variable x; then you are shown that, E of z; which is nothing but… We have shown that, this is nothing but E of f x, which is multiplied by p x dx. We have also proved it. Now, here it was a function of single variable. We will now consider the case for two variables. A similar result will come up and we will prove it. I will not say that I will prove it, but I will just develop argument in favor of the result along analogous lines, and that will be enough.

(Refer Slide Time: 02:25)



So, we have given this; maybe… where g is a function of two variables – two random variables: x and y. Then, this will be nothing but this will prove… The proof actually that is, this is the expected value of g of x, y. The proof actually is purely along analogous lines. I will just show it by an example.

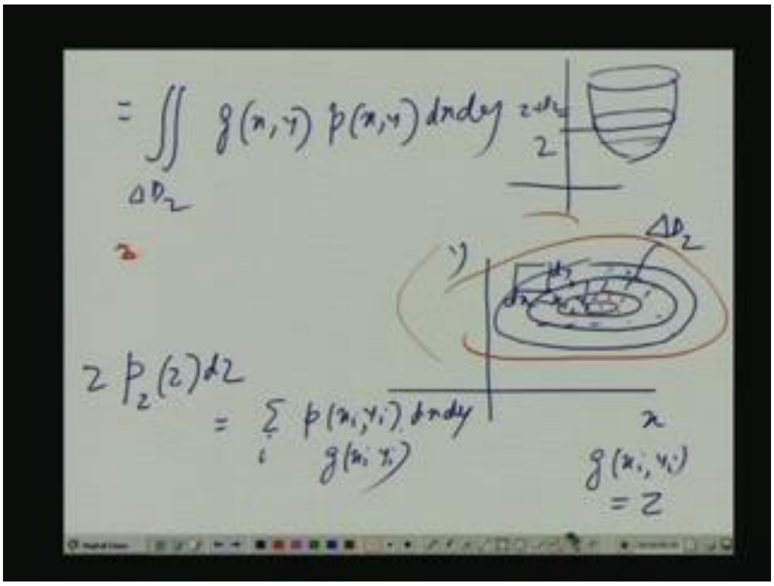
(Refer Slide Time: 03:40)



Suppose these are: z axis; this is x; this is y. And the function is like this. It is basically a surface. It is a bowl-shaped curve. These are the contours actually. So, the surface is given by z; that is, z, which is g of x, y – actually, is nothing but a function; where, for any xy, the corresponding value of Z lies on the surface of this bowl-shaped surface; that is, I can indicate the surface also like this along this side like this. So, in this case, take a particular z. If you pass a plane parallel to the xy plane through this z, it will cut this surface or a contour like this. So, this contour will be – if you just have xy plane, there will be just a contour. And then I take from z; I go up – z plus dz. Again I pass a plane through it. So, it will be another contour. On this contour, at any point, I have got a pair x comma y. For that, the function g x, y will give the constant value z. And for any point on this contour, if I evaluate z, I will get a constant value, which is z plus dz. So, probability of z lying between capital Z at z plus dz is what? That is same as the probability of this x comma y pair lying in this region.

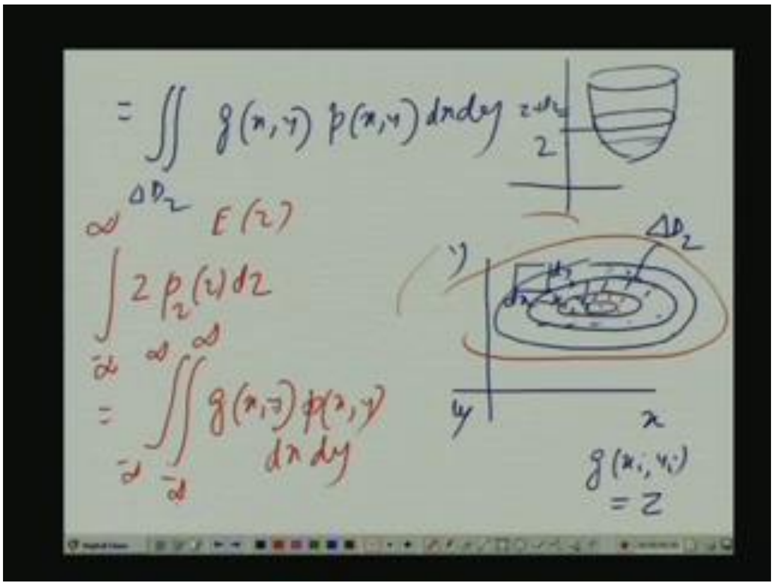
What is that probability? That is, you can take various points very close – maybe x i, y i; you take a dx here and a dy here. So, an elementary area at x i, y i. Then, take another point very near to it; again find out an elementary area. So, find the probabilities of xy pair falling within these elementary areas; take dx and dy to be infinitesimally small and sum them. What we will get? Basically a sum i p x i, y i dx dy. I will let dx dy tend to 0. And these points will come infinitesimally close to each other, so that this will become an integral – definite integral; that is, we all know definite integral is nothing but element of a sum – discrete sum. And if I multiply this by z; if I am simply multiplying this side also by z; but z is same as… z is what you get if you evaluate this function g at any point on this contour. So, g x i, y i is equal to z for any i whether you evaluate it here or here or here or here or here, like that. So; that means this gets multiplied by g x i, y i. And now, I will let these points – x i y i for various i’s come close to each other – infinitely close; and dx dy are infinitesimally small. So, this summation then becomes an integral. I erase this now.

(Refer Slide Time: 08:44)



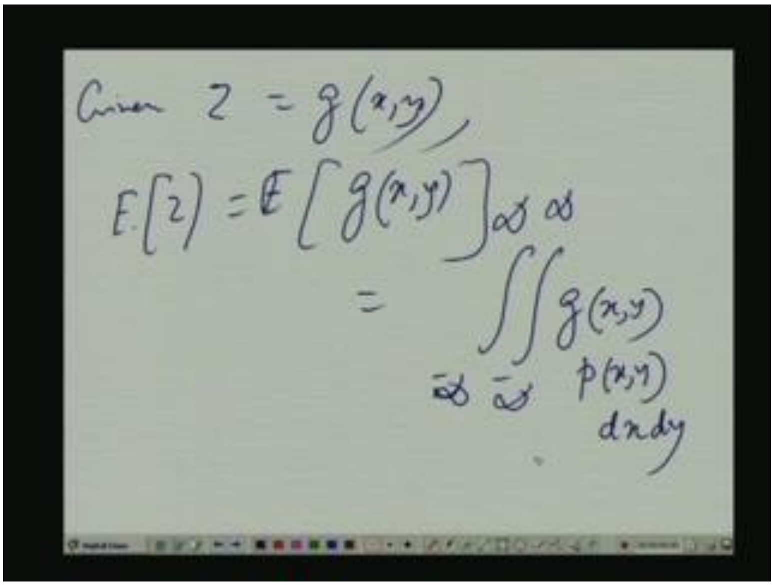
So, it leads to… If I now denote this area by delta dz; for a given z, I got this contour; and on the contour, I have got an infinitesimal area ((Refer Time: 09:01)) region. So, delta dz. So, these are… This summation where these points are infinitely close to each other and dx dy are infinitely small; this becomes an integral g x, y p x, y dx dy. And then I had said that, contour like this corresponding to some fixed z; and I took z as z plus dz here. Then, I take z – another z very close to this previous one. Again do the same thing; get another contour; another one very close that, and like that. So, I keep varying z continuously from minus infinity to infinity. In this case, this will become what? This will become an integral over all the analogous regions. I have got one. Then, if I bring z further down, I will get another one between these two. Then, I will get another one – this one and like that. Finally, it will come to a point, because z cannot go below this. And as I move z upward, I will get bigger and bigger contours. So, entire xy plane will be covered. So, virtually, this with give rise to this thing.

(Refer Slide Time: 10:59)



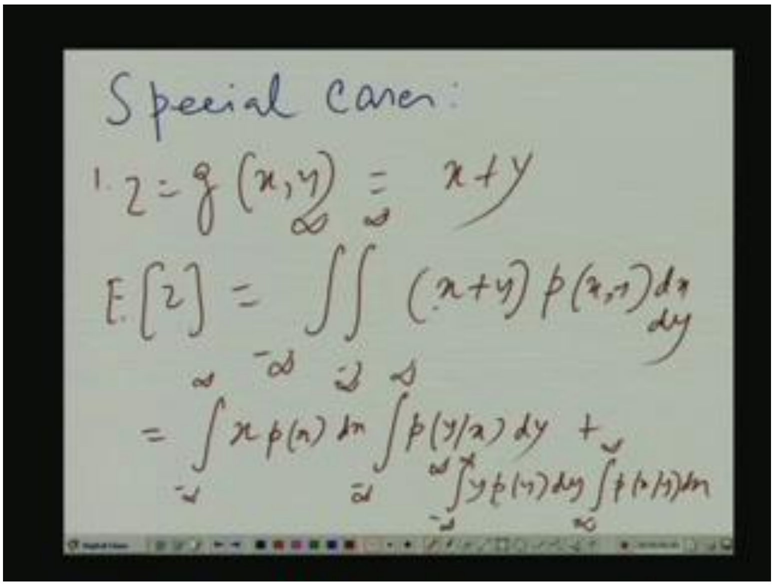
So, if I find out z… Taking z over the entire range, which is incidentally is nothing but E of z. These amounts to doing this integration about the entire xy plane, because these contours will cover the entire xy plane. So, xy plane means… – which is nothing but expected value of g of x, y.

(Refer Slide Time: 11:53)



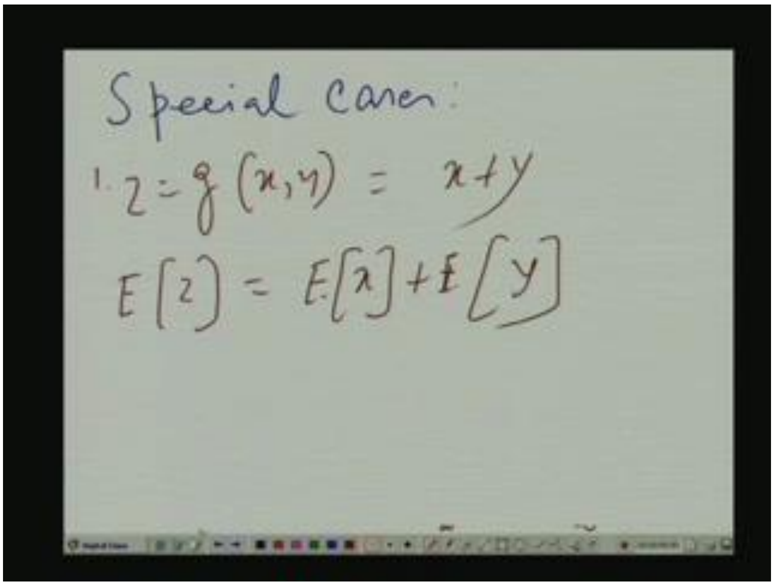
So, I rewrite the result, given…

(Refer Slide Time: 12:40)



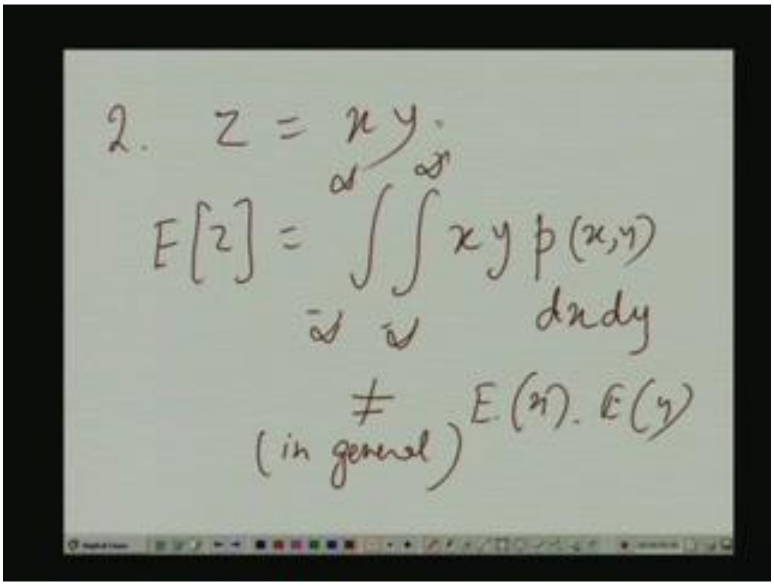
It is nothing but… ((Refer Time: 12:34)) of special cases g x, y; which is z is nothing but just a summation of x and y. In that case, what is E z? And you all know I can write it; take x times p x, y; which all of us can write like this. p x comma y is nothing but p of y by x; that is, conditional probability density given x; then the probability of y – that times p x. So, p x comma y is broken here as p x times p y by x. This integral is with respect to x. So, x p x – they are brought within this integral. This integral is with respect to y. So, just this comes out. This is from the first term – x times p x comma y. Similarly, I can also break it as a – p x comma y can be broken as a product of p y and p x by y. So, that leads to… But, this integral is 1; whereas, given x, what is the probability? I mean this is the probability of getting particular y; and then y is moved toward the entire range. So, total probability will be definitely 1, because some value rather must come, is a certain event. So, I am left with 1 times this integral. And this is nothing but expected value of x. And similarly, here p x by y dx and x is moved from minus infinity to infinity. So, one value at least must… I mean x will take at least one value. So, if it is moved from minus infinity to infinity – total probability, that should be equal to 1, because it is certain event. So, I am left with nothing but expected value of y. So, this gives rise to the following.

(Refer Slide Time: 15:29)



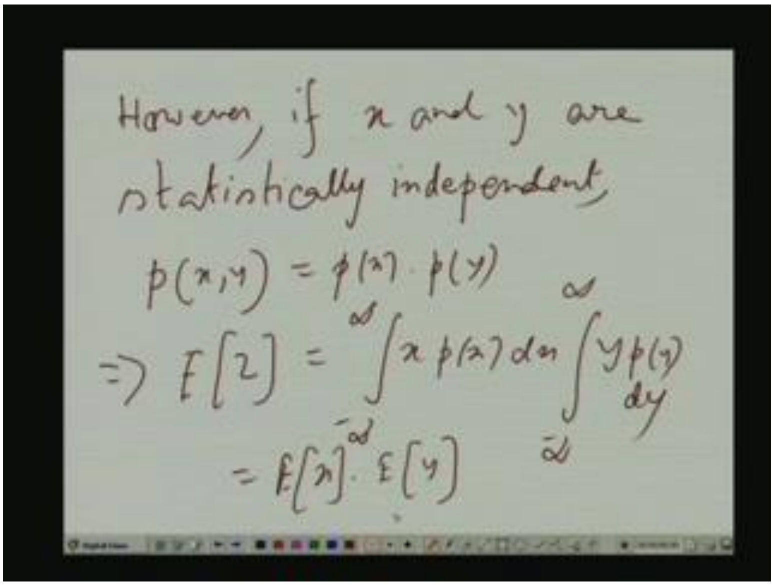
Ok.

(Refer Slide Time: 15:45)



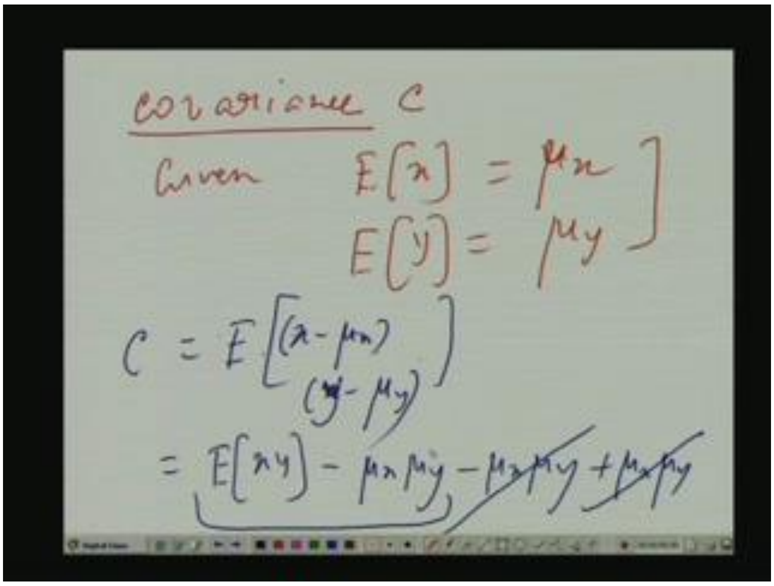
However, if z is xy, then what is E z? Here we cannot; this is not in general – in general, E x E y. If it is x plus y, E z is E x plus E y. But, if it is x into y, E z does not mean this is equal to E of x times E of y in general. But, what if x and y – they are statistically independent? If they are statistically independent, there is a joint density. p of x comma y is nothing but p x into p y.

(Refer Slide Time: 16:49)



That is, however… This is very simple. We replace p x comma y as this product. We take x p x under one integral. In that case, E z is nothing but product of E of x and E of y.

(Refer Slide Time: 18:20)



Then, I come to a very important topic… Given E of x; that is, mean of x is mu x. Then, covariance C – C is nothing but expected value of one function of x and y. What is that function? Actually, you are dealing with real-valued variable; otherwise, I would have got to put a conjugate here. Now, what is the significance of this? I call it covariance. But, why it is so important? Actually, if a covariance is high, then this means that, two variables: x and y – they are highly correlated with each other. And if it is less; so if it is close to 0; then that means they are not correlated to each other. How?

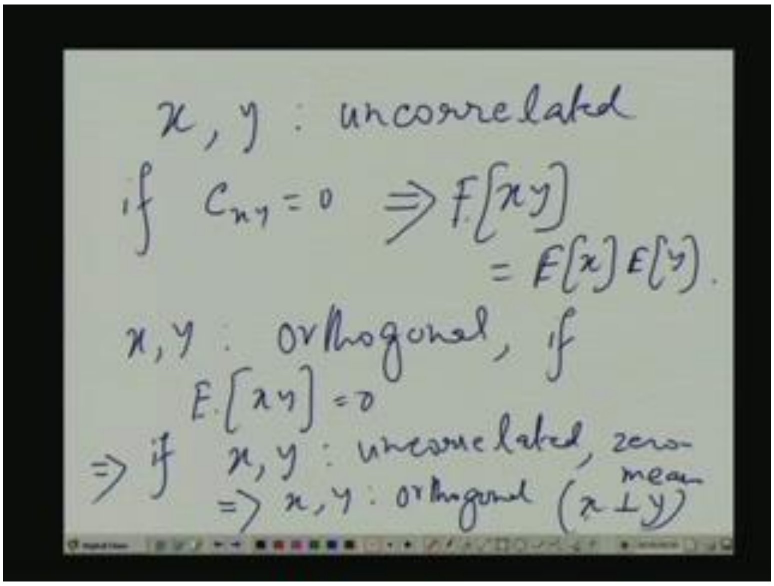
Suppose x and y – they are two such physical parameters or physical variables, which are not related to each other at all. Then, the variation of x is independent of variation of y and vice versa. So, in each experiment, you find out this value and find out this value. This is nothing but deviation of x across the mean and deviation of y across the mean. Sometimes this deviation can be positive; this deviation can be positive; so product is positive. Or, this deviation can be negative; this also negative; again, product is positive. But, sometimes this can be positive; this is negative. When this is negative, this is positive. This is so because these two variables – x and y – they are uncorrelated. So, sometimes x can go higher than mu. But, y does not have to go above its mean; y then can go down the mean; go below the mean and vice versa. x can take values less than mean in some experiments, but y can take values much above the mean – its mean – mu i. Or sometimes both of them go above the mean; sometimes both of them go down the mean – all are possible. So, positive-positive – product is positive; negative-negative – product is positive.

But, we also have cases, where one can be positive; another can be negative. Or, this deviation can be negative; this can be positive; in which cases the product is negative. And if you really take up many such experimental observations; carry out these products – whatever values you observed; based on this, carry out these products and then sum them up and average; then obviously, you should get values close to 0, because x and y – they are not correlated with each other. So, sometimes positive values came up, sometimes negative values came up for this product; and average gives rise to 0. On the other hand, if x and y – they are highly correlated, then either they go together up above their respective means – meaning the product is positive; or, they go below their respective means again together; again the product is positive. And on only seldom occasions or on rare occasions or even few occasions only, one can be positive; one can be negative.

So, largely, if you average, you will get a larger value of the C; that means, if C is high, x and y – there is some kind of handshaking between them, some kind of relation between them. So, they move together in the same direction or move together in the same sense. It is also possible that, x can be positive. And that time y is such – y takes negative value. And when x takes negative value, y takes positive value. Even then you will get a high value; negative value of covariance, but its magnitude will be high. So, whenever there is a relation between them, then you tend to get higher value for this product. But, when they are not having any relation between them, then all possibilities are in this thing; there is both positive, both negative; one deviation positive and another deviation negative, and vice versa. And you sum up all cases – average over a large number of trials; you will get a zero value or close to zero value. That is why this quantity is important.

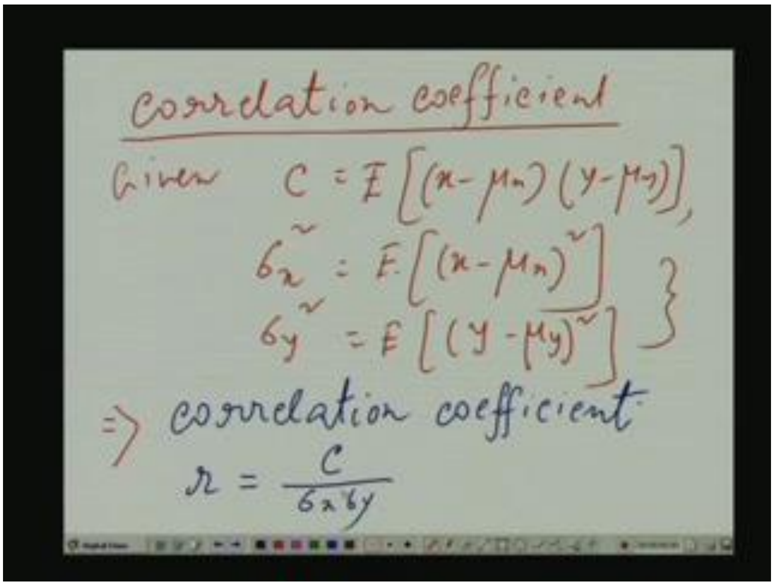
Now, if you expand it or multiply E x, y. And as we know, expectation is a linear operator; you can apply it on each of the terms. xy – that comes here, then mu x, E y; E of y, which is nothing but mu y minus E of x mu y. But, E of x is mu x again. So, mu x mu y, and then minus minus plus mu x mu y, which is a constant. So, expected value of that is itself. This and this cancels; you are left with this. So, if the covariance is 0, we say that, two variables are uncorrelated; obviously, I have given you the physical meaning of it. In that case, E of xy – it simply turns out to be product of the respective means.

(Refer Slide Time: 24:01)



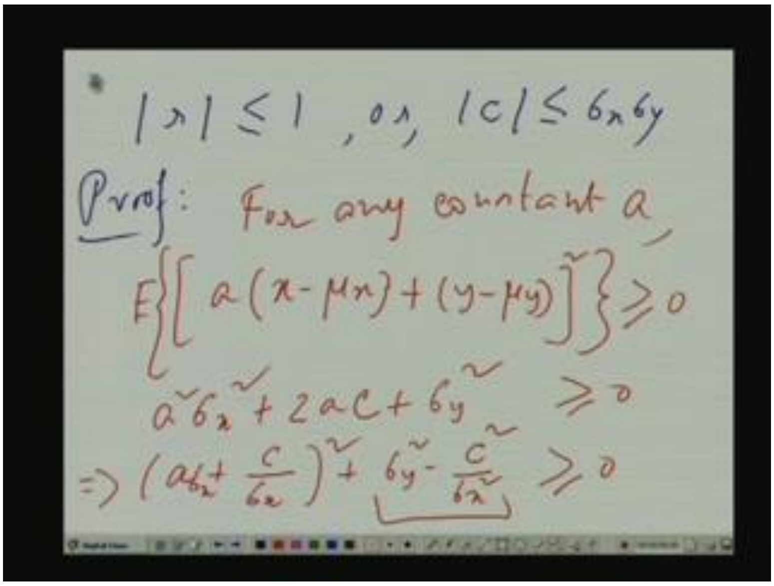
So, two variables: x and y – they are uncorrelated. C – I put xy subscript, so that it means that, covariance related to x and y. This is zero implying. This is another term, which was also important. x, y – they are orthogonal, if… – means if x, y – uncorrelated, zero mean; either both or at least one. In that case, what do we have? If since they are uncorrelated, E xy is Ex Ey. And at least one mean is 0, if not for both. So, we have got E xy equal to 0. Then, this would lead to x, y – orthogonal. We write as x orthogonal y.

(Refer Slide Time: 25:45)



We define something called correlation coefficient. Also, the variances are given. So, actually, it is a normalized version of the covariance. The covariance is C. This is normalized by this product of individual variances. Sigma x actually is called as standard deviation; sigma x square is the variance; this is a correlation coefficient. Obviously, if they are uncorrelated, r is 0, because C is 0. Sigma x and sigma y can never be 0.

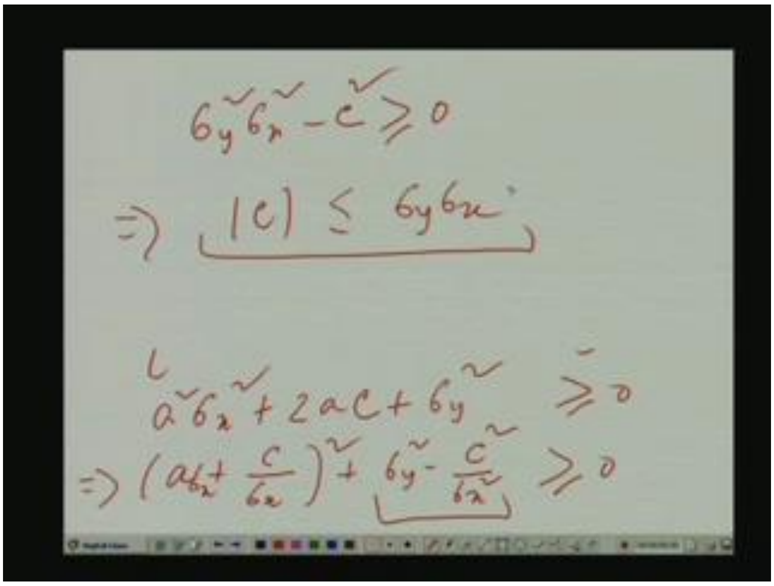
(Refer Slide Time: 27:56)



We will now show that… See C can be positive or negative; but sigma x sigma y – they are positive. How to show this? There are various ways of proving it. Proof – for any constant – you take any constant; let us say any number – real number a. If I take E of… Now, this is the expected value of the square of a random variable. After all, x is a random variable. So, x minus mu – a random variable; a times – that is a random variable; y minus mu is a random variable. Total summation is a random variable. And E of the square of that is related to the power. So, obviously, as we know, this is greater than equal to 0 always; for any constant a, mu is important; it is true for any a.

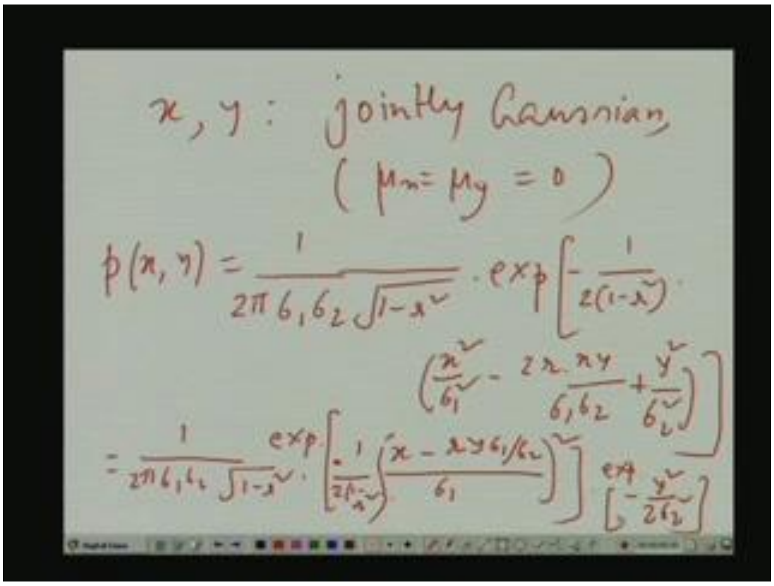
If we expand it now, we get a square – square of this term – expected value over this, because a is constant. So, we have got a square and expected value of x minus mu x whole square is nothing but sigma x square, then twice a times x minus mu x y minus mu y – expected value of that. So, twice a and expected value of x minus mu x y minus mu y – the product is nothing but the covariance C, and expected value of y minus mu y whole square; which is nothing but sigma y square. This is greater than equal to 0. But, this we can write as a sigma x. So, a square sigma x square comes here; twice a sigma x by C sigma is sigma x. So, twice ac comes. So, you bring in another extra term – C square by sigma x square. So, that has to be canceled. So, you have got… This must be greater than equal to 0. But, this is true for any a. So, if I choose a, so that this quantity is 0 means if a is taken to be C by… minus C by sigma x square; C is a constant; sigma x is a constant; I can always choose a like that. a is minus of C by sigma x square; then this entire quantity becomes 0. So, this inequality is true for that a also; which means this fellow must always be greater than equal to 0.

(Refer Slide Time: 32:04)



That means… So, what is this? Minus C. In the denominator, you have got sigma x square, which is positive. So, forget that. So, this must be greater than equal to 0 – C square; that means, obviously, mod C – if you take C square to the right; take the square root and take the positive value; so that is always less than equal to this product. Actually, C then can be either in the range plus sigma y sigma x to minus sigma y sigma x. This is what you wanted to prove.

(Refer Slide Time: 33:00)



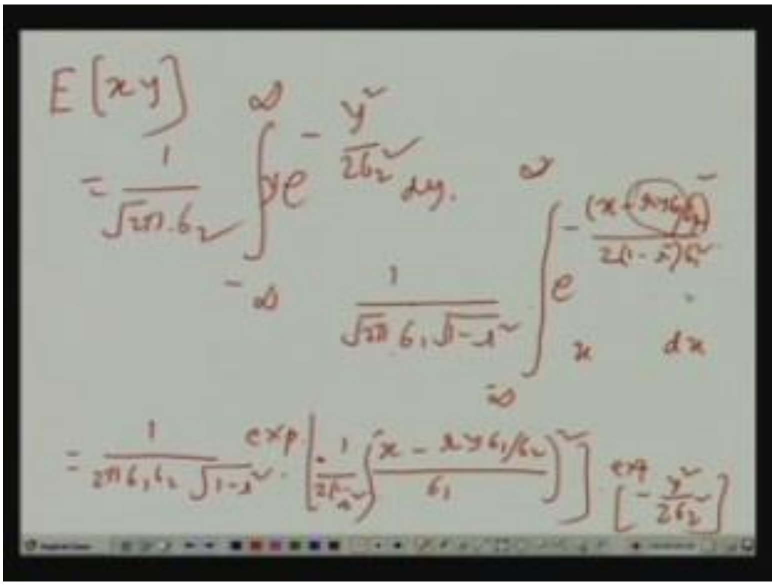
Let us consider a case of the… Suppose x, y – they are given to be jointly… This is an example – jointly Gaussian or also called normal. Assume mu x, mu y equal to 0. So, x, y – both are 0 means jointly Gaussian random variable. I have got sigma x square – the variance of x and sigma y square as the variance of y. In that case, first, let us write down what are the probability density. We have seen it earlier. It was… Let me just consider the notes, because it is basically a big expression with means becoming 0. This is nothing but… That time we said r is something called… r is a constant, but it is called correlation coefficient.

Now, indeed we will find out – we will show that, that was a correct statement that, for this case, the correlation coefficient will turn out to be r and nothing else… – times… So, let us first consider this expression. This expression can be written as… See this is a school level stuff. Whole square of this minus twice x by sigma 1 – twice x by sigma 1; then r times y by sigma 2. So, I want to bring an extra term – r square y square by sigma

2 square; cancel it by bringing a negative of that. So, y square by sigma 2 square is common – 1 minus r square within bracket. That comes up. And that 1 minus r squares cancels with these. So, you can write it like this.

In one case, we have got x by sigma 1 minus r y by sigma 2 whole square. This I can even write as ry sigma 1 by sigma 2. The whole thing is divided by sigma 1 – whole square. In fact, I have to put this term also. So, let me… This is one term. ((Refer Slide Time: 37:56)) another term y square by sigma 2 square – that I canceled the extra term that was brought; that is, y square r square by sigma 2 square. So, y square by sigma 2 square. Then, within bracket, 1 minus r square; where, that 1 minus r square cancels with this 1 minus r square. So, essentially, left with another term – another exponential – simply y square by twice sigma 2 square; that is all.

(Refer Slide Time: 39:10)

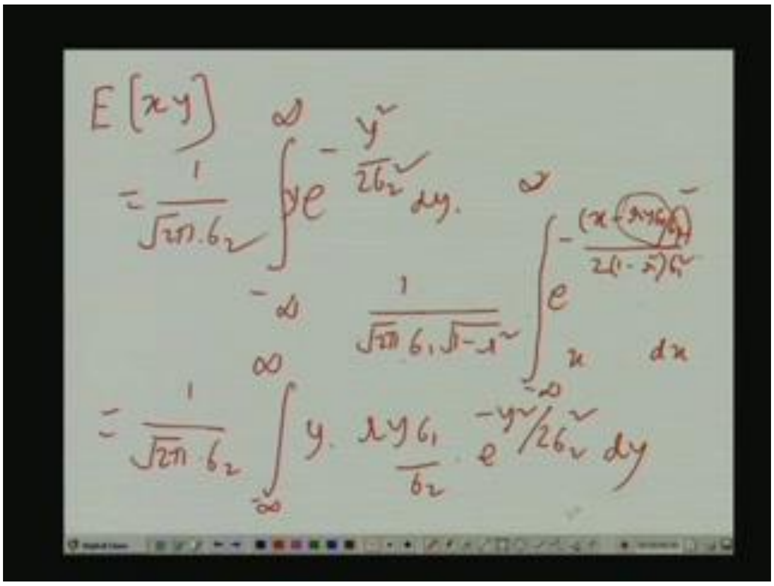


So, let us find out what is E of xy. What is E of xy? This is the joint density. So, multiply this by xy; integrate from minus infinity to infinity. But, it is a product of two exponentials. So, I can work like this. One this exponential involves only y. So, this will be an outer integral e to the power… And from here I take out… I write 2 pi as square root 2 pi multiplied by again square root 2 pi. So, here I take out this term – 1 by square root 2 pi and this sigma 2… dy here. And the rest of the terms come here – 1 by root 2 pi sigma 1 square root. And then integral twice 1 minus r square sigma 1 square. I made a mistake; this is xy. So, y this should be multiplying y here; and this should be multiplying x dx.

Now, consider this outer integral. Firstly, this is nothing but e to the power minus y square by twice sigma 2 square and 1 by root 2 by sigma 2; it is nothing but the expression for the Gaussian probability density function for the variable y. And we are multiplying y by the density integrating. So, it will give rise to nothing but average value of y – e of y. But, how about here? This also has the form of the Gaussian density. Here x is a variable. This is a Gaussian density for a variable x; whose mean is this quantity – this quantity. And variance is square root of this product – 1 minus r square sigma 1 square. That comes here also. And this density is multiplied by the corresponding variable – x integrated. So, what is this? This will give rise to expected value of x.

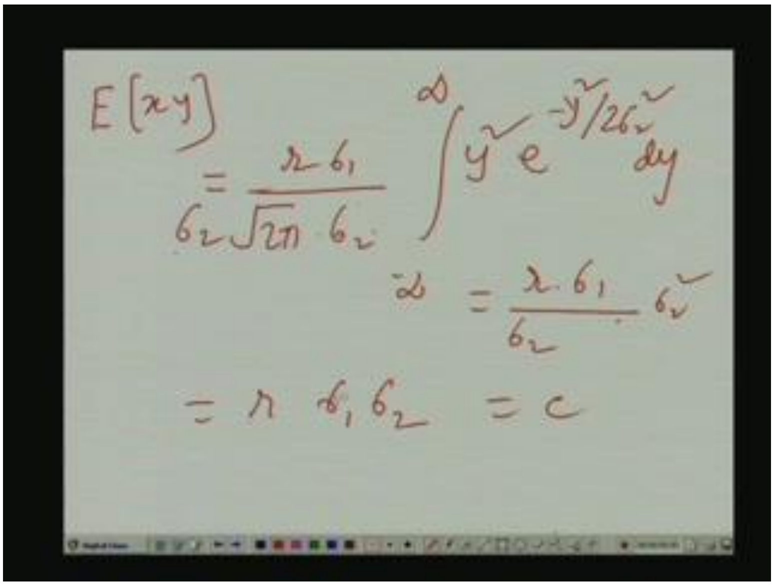
What is the expected value of x? That is nothing but this mean in this case. This will give rise to the expected value of x. But, the expected value of x in such case – in fact, if you want, you can replace this x by x prime dx prime. And you can call it x prime minus this whole square divided by this thing. So, what is x prime? x prime is a random variable; whose mean is given by this r y sigma 1 by sigma 2. And variance is square root of 1 minus r square sigma 1 square. So, this density is multiplied by x or x prime, rather integrated. So, I will get the expected value or the mean value – mean value of x prime; which is nothing but this quantity. So, that will emerge.

(Refer Slide Time: 44:13)



So, I get E y out of the first one; no, not yet. From this, I get… y and this function. But, from here I get the expected value of that variable x, which is r y sigma 1 by sigma 2, so that y is coming again. So, ry sigma 1 by sigma 2, then this exponential.

(Refer Slide Time: 45:27)



This gives rise to r sigma 1 sigma 2. And there is y square. So, r goes out; sigma 1 goes out. And now, I have a y square. Instead of calling it sigma 2 square, let me write it as sigma 2 times this. So, now consider this function – 1 by square root by sigma 2 e to the power minus y square by twice sigma 2 square. That is the probability density of this Gaussian random variable y. That is multiplied by y square integrated. So, that will give rise to the variance of y. After all, mean is 0. So, it gives rise to the variance. And that variance is nothing but sigma 2 square; that variance is nothing but sigma 2 square. In fact, I forgot to mention – sigma 1 is corresponding to the sigma 1 square is the probability, is the variance of random variable x; sigma 2 square is the variance of random variable y. So, this will give rise to sigma 2 square. So, r sigma 1 as it is; and one sigma 2 left out; and sigma 2 square comes up. So, you get r times sigma 1 sigma 2.

And, what is E xy? It is same as C also, because mean of x and mean of y were taken to be 0. So, covariance is nothing but as we know E of x minus mu x; but mu x is 0 times y minus mu y; mu y is 0. So, it turns out to be E of x times y only; which is C. So, C is equal to r sigma 1 sigma 2. So, what is r? It is nothing but C by sigma 1 sigma 2; which is the definition of the correlation coefficient. So, we now justify our previous assumption; that is, constant r, which occurs in the joint density function for jointly Gaussian random variables; it is nothing but the correlation coefficient between those.

So, today, we will not proceed further. We will stop here. But, in the next class, we will consider joint moments. Like earlier we have taken moments for a single random variable; we are moving – generating functions; then you had characteristic functions and all that. Similarly, here also we will take the joint moment involving two variables. And then from that, we will proceed to joint characteristic functions. This will… In the end, take us to what is called central limit theorem and all that. So, that is all for today.

Thank you very much.

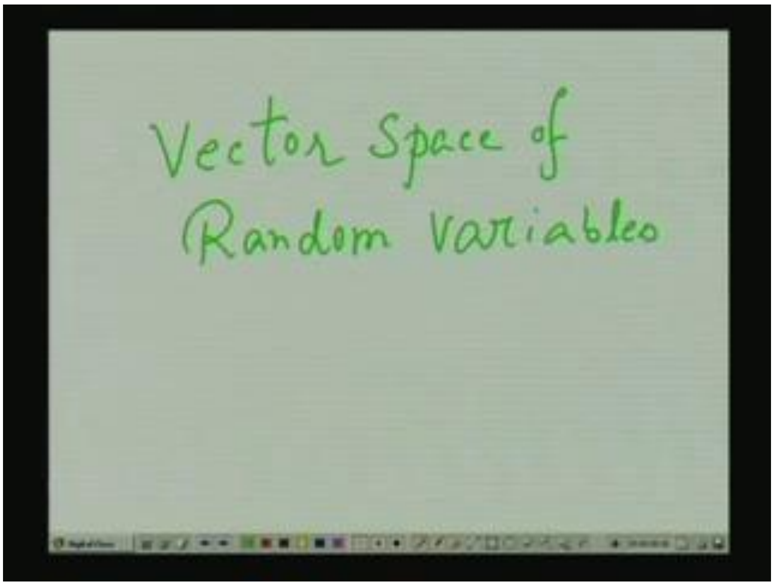
Preview of next lecture

Lecture – 16

Vector Space of Random Variables

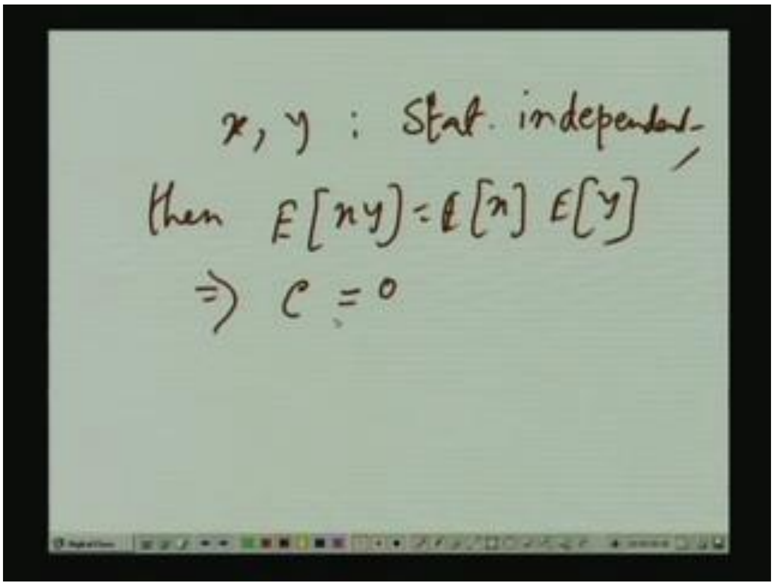
So, in today's class, we will be concentrating on a new topic and a very interesting topic.

(Refer Slide Time: 49:41)



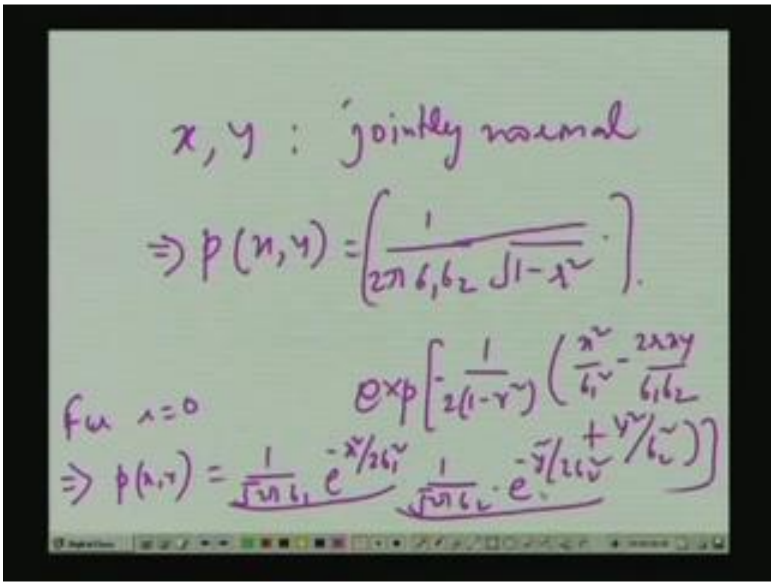
It is called vector space of random variables. In fact, this is a huge topic that cannot be covered just in one lecture. So, we will not make any attempt for that. But before I am going to that let us just… We will go back to what we were discussing last time, because little bit of that was left.

(Refer Slide Time: 50:32)



Last time we had said that, if two – x and y are statistically independent; then E xy is nothing but Ex Ey. This leads to the fact that, covariance C or correlation coefficient, which is C by sigma x sigma y – that is equal to 0. So, if they are statistically independent, it is always true; that is trivially seen. But, if this is given; that is, x and y are two random variables, which are uncorrelated; then it does not necessarily mean that, x and y are also statistically independent; that is, joint density of x and y, p xy is not in general a product of p of x and p of y. But, for a particular class of random variables, mostly actually the Gaussian random variables, that is what x and y are mutually jointly Gaussian; then statistical independence means uncorrelatedness as before; but also uncorrelatedness means statistical independence; that you can see. So, that works for Gaussian random variables.

(Refer Slide Time: 51:59)

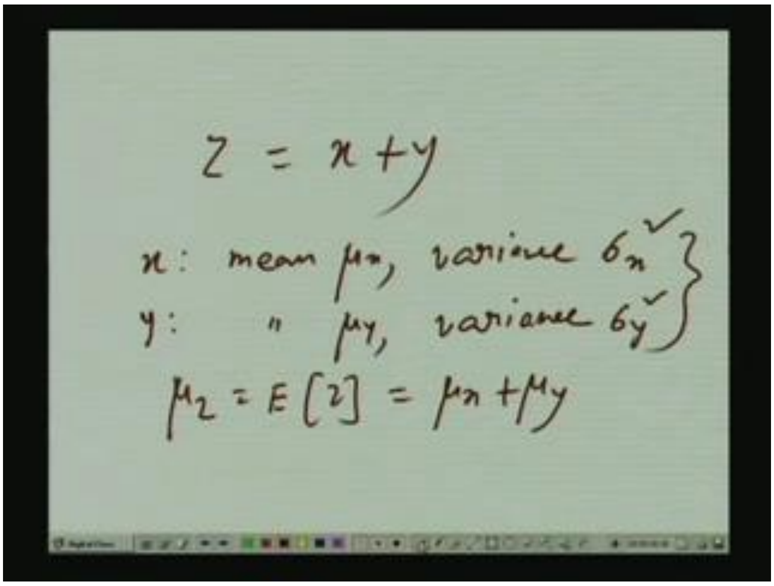


That means suppose jointly Gaussian or sometimes we call jointly normal. Then, we know what is the probability density. What is the probability density? Just a minute; this thing times exponential minus 1 by… Now, we have seen earlier. In fact, we have proved that, this r is nothing but the correlation coefficient. So, suppose it is given that, x and y are uncorrelated; in that case, r is 0. So, if r is 0, this term is 0 – 1 minus r square plus 1. So, which is nothing but… So, this becomes nothing but minus 1 by 2; and then x square by sigma whole square plus y square by sigma 2 square. And this simply becomes 1 by 2 pi sigma 1 sigma 2. So, for r equal to 0, this leads to a very simple thing. We can write this as a product. One of them is root 2 pi sigma 1 e to the power minus x square by twice sigma whole square; and again 1 by root 2 pi sigma 2 e to the power minus y square by twice sigma 2 square.

So, this is p x; this is p y; which means the joint density is nothing but the probability density of x multiplied by the probability density of y; that is, x and y are statistically independent. So, you see as I said that, if x and y are statistically independent, they are always uncorrelated. That is true in all cases. But, if x and y are given to be jointly Gaussian or jointly normal; that is equivalent of saying; then if they are uncorrelated, then again the reverse is true; that is, they are statistically independent too. But, that is not true for other cases. So, it is not true in general. But, in the case of Gaussian random variables, statistically independent leads to uncorrelatedness; and uncorrelatedness leads to statistical independence.

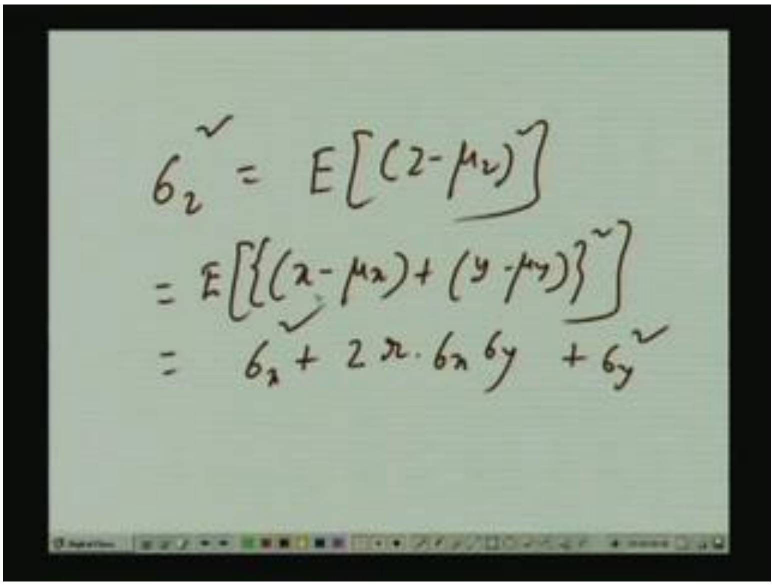
Now, we come to this important topic called vector space, rather Hilbert space of random variables. And this cannot be covered just in one lecture, because this topic of vector space as such is a semester-long topic. So, I will not try; but I will just tell you the motivation.

(Refer Slide Time: 55:28)



Actually, the random variables – variables and z is another random variable obtained by summing x and y. x has mean mu x; y has mean mu y… So, what is the mean of z? Obviously, expected value of z is nothing but expected value of x plus y. And you can use the linearity of expectation operator. So, that is nothing but expected value of x plus expected value of y. So, obviously, mu z, which is nothing but E of z is same as mu x plus mu y. That comes trivially. And what happens to the variance?

(Refer Slide Time: 56:33)



Sigma z square; which is nothing but E of z minus mu z whole square. Now, this you can write as… You replace z by x plus y; mu z by mu x plus mu y. So, this becomes nothing but whole square. So, square of this, expected value of that; that will give you the variance of x. Square of this, expected value of that; that will give you the variance of y. And then we wrote twice E of x minus mu x y minus mu y; which is nothing but the covariance. Covariance – either you can write by C or you can write it as a product of correlation coefficient r times sigma x sigma y. So, if r is given, sigma x is given; sigma y is given. You can find out sigma z square.

So, I will stop here today. But, let me tell you in the nutshell, what we are going to do soon. From here we will move to moments; like we remember in the case of single random variable, we defined moments. And moments were used for what? We derived some properties of the moments; and from moments, we marched to characteristic functions; here also we marched to what is called joint characteristic function. But, before going from moment to joint characteristic function, I will again come back to one topic, which I had left out thinking that I may not need it. But, that topic I find is also is important; that is, the two functions of two random variables. So far I considered only one function of two random variables; that is, g of x comma y. But, I will be now considering two such functions; maybe one is f of x comma y – you call it z; another is g of x comma y – you can call it maybe v. So, there are two functions. And obviously, z and v are in general jointly related. So, I have to find out the joint distribution function and joint density of z and y, so that we will be doing in the next class.

Thank you very much.