Introduction To Adaptive Signal Processing

Prof. Mrityunjoy Chakraborty
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture No # 05
Multivariate Gaussian Density

Ok. So, today we will be discussing Gaussian density, but not just for one variable for many real continuous random variables together. So, that will be called joint, the jointly Gaussian. Gaussian is sometimes also called normal, jointly Gaussian random variables. This jointly is sometimes also called multivariate, there are different names. So, here we are given capital N, real continuous random variables say $x_1, x_2, ... x_N$.

They are jointly Gaussian. So, every experiment I will observe them together. They may have a relation between them one can be temperature, another can be humidity, another can be anything. They will be called jointly Gaussian if they are joint probability density.

We have discussed what is joint probability density for two variables X Y then for three variables. So, in general the joint probability density if it takes a particular formula form, then it will be called that they will be called jointly Gaussian. Before that let us see, suppose x is the vector that is all the random variables in stack, they are put in a stack in a vector. We have already discussed it yesterday towards the end, same thing I am saying, and a mean vector is a vector of μ_1 is a mean of expected value of x_1, μ_2 expected value of x_2 μ_N then we know we can take the deviation the increment vector $x - \mu$, that is every random variable, the incremental value it takes around its mean $x_1 - \mu_1$. So, sometimes can be positive, sometimes can be negative and for all other variables.

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_N \end{bmatrix}$$

$$\underline{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \cdot \\ \cdot \\ \mu_N \end{bmatrix}$$

So, they are in a stack $x_1 - \mu_1$, $x_2 - \mu_2$, $x_N - \mu_N$ that is what this is. We have seen that this vector if I take expected value of this is a column vector and its transpose, we have seen in the last class that is the covariance matrix.

$$C_{XX} = E[(\underline{x} - \underline{\mu})(\underline{x} - \underline{\mu})^t]$$

See maybe you can say x, x because 1 x minus mu comes here, another x minus mu row vector comes here, you multiply 1 element here 1 element here. So, this has x, this has x. So, 2 elements of x get multiplied, maybe $x_1 - \mu_1$, $x_3 - \mu_2$.

So, $x_1 x_3$ terms come that is why we indicate it by $2x$. This is second order thing; 2 random variables are multiplied. It is called the covariance matrix for this random vector. It is a random vector; everybody is a random variable and they are jointly random. So, this we discussed last time mu is the mean alright, then this random variable $x_1, \dots x_N$: jointly Gaussian, if they are probably joint probability density that is p capital X1 capital X2 dot dot dot capital Xn and the values x1 x2 dot dot dot dot xn that takes this form.

$$P_{X_1, X_2, \dots, X_N}(x_1, x_2, \dots x_N)$$

First 1 by 2 pi n by 2 and then determinant of this matrix C_{XX} its square root, positive square root into e to the power into. I do not have space here. So, I am writing here. e to the power minus half this vector this difference vector x minus mu. So, you are choosing x1 x2 xn.

So, you are putting them in a vector subtracting the mean vector from them that is x minus, you are putting those values. So, for those x1 x2 xn. So, those values of x1 x2 xn you have to put here in the x vector, corresponding mu values mu 1 mu 2 mu n in this vector the column vector. So, in the formula they come as row, column transpose is row, then C_{XX} and I am assuming C_{XX} matrix is invertible that is inverse exists. Not all matrices are invertible, but we are assuming assumed when we can assume that that we will discuss later assumed to be invertible.

$$P_{X_1,X_2,\ldots,X_N}(x_1, x_2, \ldots x_N) = \frac{1}{(2\pi)^{\frac{N}{2}}\sqrt{\det\,(C_{XX})}}\; e^{\frac{-1}{2}(\underline{x}-\underline{\mu})^t}\, {C_{XX}}^{-1}(\underline{x} - \underline{\mu})$$

That is C_{XX} inverse exists. So, again C_{XX} was a matrix is inverse is a matrix, matrix then again, the same vector x minus mu. Now, this is a column vector, this is a matrix, a matrix times a column vector is a column vector. So, this is a column vector, this is a row vector. So, row into column is a scalar.

So, e to the power minus half into some scalar. So, it is a scalar, 2 pi to the power N by 2 scalars, determined is a scalar. Because probability cannot be a vector, probability is a scalar. So, whole thing is a scalar. So, in the case of jointly Gaussian random variables, this should be the joint probability density.

If capital N is 1 there is only one random variable, then you will see this boils down to or well-known Gaussian density function for one variable. That is if you take N to be 1 only then obviously, 2 pi to the power N by 2, N is 1. So, you have got I am writing here N equal to 1, you see 1 by 2 pi to the power 1 by 2. So, 1 by root 2 pi then C_{XX}, C_{XX} x has only one element x1 and mu is mu 1 whole vector. So, C_{XX} means what it was a column vector earlier into row vector, but now column has only one element x1 minus mu 1, corresponding row also has only one element x1 minus mu 1.

So, it is nothing but x1 minus mu 1 whole square, expected value of that and that is called the variance. sigma x square sigma x1 square and now square root. So, it will be just sigma x1, x1 is a variable, sigma x1. Then e to the power minus half again x minus mu transpose has only one value x1 minus mu 1. Matrix is a scalar now that was as we told the variance cxx is a variance sigma x square.

So, inverse of that is 1 by sigma x square. So, you have got minus half, So, 2 I bring here, sigma x square I bring here, the scalar, and x1 minus mu 1 again x1 minus mu 1 only one element one element.

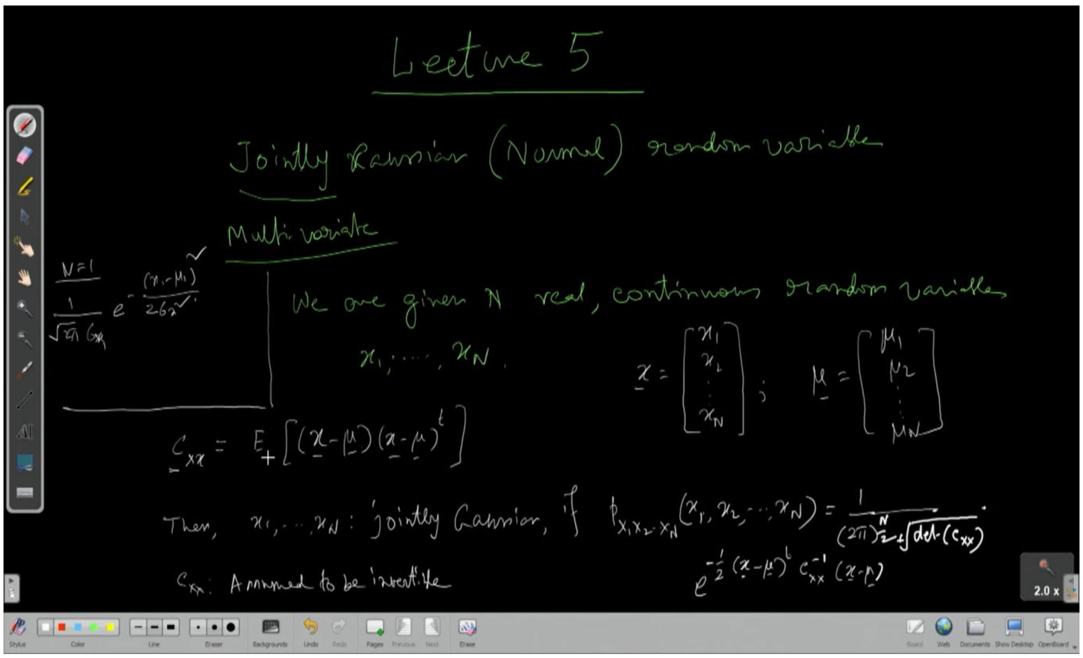
$$= \frac{1}{\sqrt{2\pi}\, C_{XX}}\, e^{\frac{-(x_1 - \mu_1)^2}{2\sigma_x{}^2}}$$

So, this is what you have for the normal ordinary Gaussian density. we all know general form of the Gaussian density function for one variable, but this is more general. If this happens then they are of course, they are real and continuous in if they are complex then the formula will be more general, I mean more modified that we will not consider in this course.

If you are interested you can consult books. Now, we will see a very interesting thing. Earlier I had said that if a set of random variables they are statistically independent. That is if the joint density is product of individual marginal densities, then they are also uncorrelated. Their covariance is I mean, between any two-element covariance is 0, between any two elements.

Covariance between any two element 0 means what will be this matrix it will be a diagonal matrix because if you take ith element, in the previous class we have seen if you take ith element and jth element, ith comma jth element that will be xi minus mu i times xj minus mu j, product and their expectation which is a correlation covariance between xi and xj. If i and j are not same then if they are uncorrelated, this correlation is 0 which means if i and j are not same that is I am not considering diagonal entries because for diagonal entries same i same j, 1 1, 2 2, 3 3, 4 4, but if that does not happen then other elements are 0. So, for a if the variables are uncorrelated then the covariance matrix is a diagonal matrix.

(Refer Slide Time: 11:18)



Ok, that we have seen, let me again elaborate this. We have seen this I did in the last class though. C_{XX} was alright.

$$C_{XX} = E[(\underline{x} - \underline{\mu})(\underline{x} - \underline{\mu})^t]$$

So, this is a vector the column vector, this is a row vector. So, multiply you get a matrix. i comma jth element of this matrix C_{XX} i comma j, this is a symmetric matrix that we have seen at that time only I showed, i comma jth element will be what? ith ith guy ith element of this vector and jth element of this row vector their product and expectation. So, E with respect to xi xj only they matter here in that product.

$$[C_{XX}]_{ij} = E_{X_iX_j}[(X_i - \mu_i)(X_j - \mu_j)]$$

So, xi minus mu i and xj minus mu j and i and j they are not same, they are two different random variables.

$$= r_{X_iX_j}$$

So, you are taking increment of this around the mean, the increment of this around its mean and multiplying and expecting. So, that is correlation $r_{X_iX_j}$. So, if xi xj are uncorrelated this will be 0 that means, as long as then i and j i comma jth element and i and j are not same because they are different random variables then only, I can talk of their correlation.

$$[C_{XX}]_{ij} = 0; i \neq j$$

If i and j are not same and they are uncorrelated then that element is 0 that is i comma jth element is 0 if i and j are not same. If i and j are same it will be xi minus mu i, again xi minus mu i because i and j are same.

So, xi minus mu i whole square expected value which is variance. So, i comma ith element which is ith diagonal element will be sigma xi square variance of the ith element xi. So, diagonal element will be the variances, but on off diagonal elements will be 0 when the elements are uncorrelated. Now, I had told earlier that if a set of elements are statistically independent, we can show they are uncorrelated that is a correlation between every two elements there, covariance between every two elements there will be 0. But the converse is not in general true that is if a number of variables they are given to be uncorrelated that is covariance between any two is 0.

It does not mean in general that they are statistically independent that is overall joint density will be a product of individual marginal densities, except in the case of jointly Gaussian random variables then one means the other and vice versa both are equivalent that is what we will see here. Now, suppose that is suppose we are given the fact that x1 to xn are uncorrelated, meaning C_{XX} i comma j 0, if as I told here if i not equal to j, and C_{XX} ith diagonal into i comma i would be the variance, sigma xi square, variance of xi fine.

$$[C_{XX}]_{i,i} = {\sigma_{X_i}}^2$$

So, suppose this is given and of course, they are jointly Gaussian then what happens determinant, number 1, determinant of C_{XX}, now C_{XX} is a diagonal matrix, all other elements are 0 only, diagonal elements are ${\sigma_{X_1}}^2$, ${\sigma_{X_2}}^2$, ${\sigma_{X_2}}^2$ dot dot dot their product. So, determinant is ${\sigma_{X_1}}^2$, ${\sigma_{X_2}}^2$ all right.

$$\det[C_{XX}] = {\sigma_{X_1}}^2, {\sigma_{X_2}}^2, \dots {\sigma_{X_N}}^2$$

So, in that Gaussian formula p x1 x2 dot dot dot xn we remember x1 x2 dot dot dot xn, we had 1 by root 1 by 2 pi whole to the power N by 2.

ok, and then in the denominator had positive square root of the determinant. So, if you take positive square root, it will become $\sigma_{X_1}, \sigma_{X_2}, \dots \sigma_{X_N}$ like that. So, what I do the whole probability density function I write as a product of some terms, one term, similar terms, one is just root 2 pi, but I had 2 pi to the power N by 2. So, I should have 1 by root 2 pi, again 1 by root 2 pi, again 1 by root 2 pi dot dot dot capital n times. So, I will have maybe 1 by root 2 pi here dot dot dot lastly again another 1 by root 2 pi capital N number.

Then $\sigma_{X_1}, \sigma_{X_2}, \dots \sigma_{X_N}$ there is I am taking positive square root this product. So, I take σ_{X_1} here, I take σ_{X_2} here, I take σ_{X_N} here. So, you multiply the denominators you get back what we had earlier then e to the power is something that we have to see. what we had, e to the power, this is where I do the calculation, we had e to the power minus half C_{XX} inverse x minus mu. Now, cxx inverse means what, C_{XX} is a diagonal matrix, all diagonal entries are positive because they are variances.

So, inverse means it will be again diagonal matrix, but it will be a sigma 1 by sigma x1 square dot dot dot 1 by sigma xn square and 0 on this side 0 on this side this is cxx inverse this matrix. And then you have x1 minus mu 1 top guy then x is x2 minus mu 2 dot dot xn minus mu n and of course, this term. So, when you multiply this with this, the diagonal matrix, So, just this multiplied by this, this multiplied by the next one, this multiplied by this. So, you get a row vector same, but they are scaled this is multiplied by this this is multiplied by like that with that if I multiply x1 minus mu 1 dot dot dot xn minus mu n. So, this after this product if I do the multiplication, you will have x1 minus mu 1 here and here whole square divided by this plus really plus because a row vector and this matrix times this a column vector.

So, row into column means term by term multiply and add plus. So, x1 minus mu 1 into x1 minus mu 1 by sigma x square x1 square plus x2 minus mu 2 here also. So, x2 minus mu 2 whole square by sigma x2 square plus dot dot dot dot. So, e to the power some terms which are summed that mean, I can separate them out while e to the power minus half is common one term will be just x1 minus mu 1 whole square by this another will be dot dot dot another will be you can just multiply this exponential.

$$P_{X_1,X_2,\dots X_N}(X_1,X_2,\dots X_N)$$

$$= \frac{1}{\sqrt{2\pi}\sigma_{x_1}} e^{\frac{-1}{2}\frac{(x_1-\mu_1)^2}{\sigma_{x_1}{}^2}} \frac{1}{\sqrt{2\pi}\sigma_{x_2}} e^{\frac{-1}{2}\frac{(x_2-\mu_2)^2}{\sigma_{x_2}{}^2}} \dots \frac{1}{\sqrt{2\pi}\sigma_{x_N}} e^{\frac{-1}{2}\frac{(x_N-\mu_N)^2}{\sigma_{x_N}{}^2}}$$

So, they will add up the exponent's powers will get add up I will get that, but this is what this is the Gaussian density of the single random variable x1.

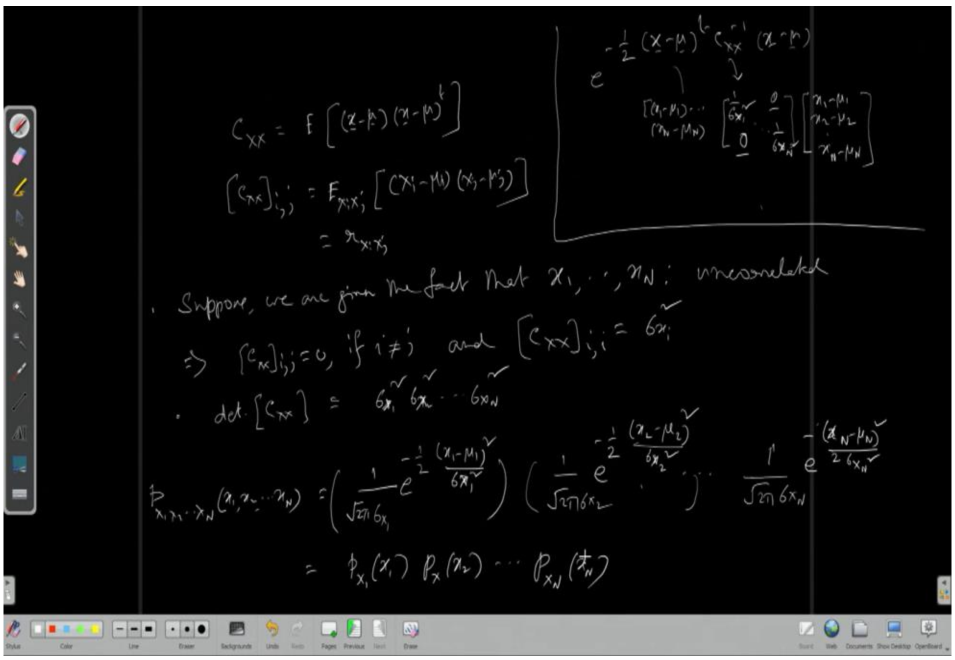
This is the Gaussian probability density of a single random variable x2 when they are jointly Gaussian, they are individually Gaussian also ok. There is a meaning I forgot to tell you if they are jointly Gaussian, they are individually Gaussian also.

$$= P_{X_1}(x_1)P_{X_2}(x_2)\dots P_{X_N}(x_N)$$

So, that means, this is nothing, but the probability density of this guy capital X1, this is nothing, but probability density of this guy dot dot dot is nothing, but probability density of xn. So, you see now in this case overall joint density is a product of the individual marginal densities.

So, they are statistically independent. So, in the Gaussian case if you are giving the variables to be uncorrelated then they are statistically independent also. converse is always true, but this is not always true, this is happening in the case of Gaussian. I mean in it does not happen in the case of other well-known probability densities, but here all right.

(Refer Slide Time: 21:22)



This is one important thing; another thing is complex random variables. All right complex random variables, here, I will say Z is a complex random variable if there are two real variables, two real continuous random variables capital X and Y.

$$Z = X + j\,Y$$

So, I conducted experiment, I observed two real valued continuous valued random variables, say capital X and Y and I take their values and write in this form a real variable I call, a complex variable Z, then obviously, Z is also random, but it is complex, then Z will be called a complex random variable, it has got two real random variables. So, therefore, even though Z is a single random variable it has actually two. So, I should have one X is capital X capital Y. So, Z will be, maybe if you take this to be small x and this to be small y. So, this point is Z small z, as small x plus small j y, this point.

$$Z = X + jY$$

And then if I go by d x and this is by d y then this is, you understand one thing d z will be d x plus j d y right. So, d x plus j d y. So, this point is nothing, but small z plus d z because x and d z means d x. So, x plus d x and take j common j here and in d z and j here. So, j into y plus d y here, y plus d y is this much and x plus d x is this much.

So, x plus d x plus j times y plus d y this is d z. So, d z is d x plus j d y, d x plus j d y and originally z at the two phasors. So, they basically x plus j y and it is d x plus j d z. So, x plus d x, z plus d z and j common y plus d y to this point.

$$dz = dx + jdy$$

That means, probability of capital Z lying between small z and z plus d z is equivalent to what will happen that is capital Z will lie between these two this means it will lie within this box.

It will fall anywhere in this box there is a meaning of this. That means, the same as probability of capital X lying between here to here simultaneously with capital Y lying between these which we know is nothing, but p x y joint density small x small y d x d y.

$$\equiv prob\ (x \le X \le x + dx, y \le Y \le y + dy) \equiv P_{XY}(x, y)dx\ dy$$

So, when expressed in terms of x y the real and individual component real and imaginary components then their joint density that will be same as the probability of capital Z falling here. Event is same capital Z lying here means capital X lying between here-to-here capital Y lying between here to here. So, the two things are two events are same capital Z falling in the box is same as capital X lying from here to here and capital Y lying from here to here.

So, this probability and this probability they are same all right. Now, given a function which could be complex or real $f(Z)$ then it takes z, it can give you a real number or it can give you a complex number, does not matter $f(Z)$, every time you find a z there is some x and y, you put that value x plus j y here and get a value. Since that is random this function also takes value randomly. So, what is the expectation e z $f(Z)$, it is same as if you write working in terms of z, if you write it as x plus j y then what is the meaning of this e z? there is $f(Z)$ taking a value from with the I mean when capital Z lies falls anywhere in the box since the widths are infinitely small we assume the value of $f(Z)$ does not change it will have the same value as here at this small z or at this value at this pair ok, but chance of that I mean it will not always happen.

$$E_Z[f(Z)] = \iint f(X + jY)P_{XY}(x, y)dx\, dy$$

$$= E_{XY}[f(X + jY)]$$

So, I have to multiply it by the chance. Chance means you have to multiply this f of z value that is f of this value by the probability of capital Z occurring here or equivalently this capital X occurring capital X Y occurring in this box. So, that is I have to when I express z in terms of x y, I have to multiply this by the joint density. And then average ok, multiply everywhere and take average then as we have done many times in the past weighted average. So, this to be divided by summation of the weights. weights are this part, but their average their integral is 1, certainty probability of certainty is 1.

So, you get this which is nothing, but e x y of this function f now you write in terms of x y x plus j y all right. Then one thing e z suppose this is given and I suddenly and f of z is a complex function. So, it takes z as argument and gives you a complex number f of z. Now instead of f of z I take the conjugate that is whatever value f of z takes I take a complex conjugate of that. So, then it becomes a function f star z f star z what will happen to that.

$$E_Z[f^*(Z)] = \iint f^*(X + jY)P_{XY}(x, y)dx\,dy$$

So, again just same thing I have to do, but instead of f it will be f star small x plus these are small within the integral that the variables of integral. So, now you see probability density is real is never complex, d x d y real and this is star. So, I can as well take the star out I mean this product I can take put the star outside, the star of a product means star will be put on everybody, but they are real. So, no change. So, this product and star and then integral is a summation.

$$= [\iint f(x + jy)P_{XY}(x, y)dx\,dy]^*$$

$$= [E_Z[f(Z)]]^*$$

So, summation of complex quantities with the conjugate you will get the same thing if you first sum and then conjugate that is z 1 star plus z 2 star that is first conjugate z 1 star z 2 star then add that is integral you will have the same thing if you do first addition z 1 plus z 2 then star that is z 1 star plus z 2 star we all know is z 1 plus z 2 star first add. So, first you integrate and then star that is first you integrate and this product but this is what I have here. So, that means, this will be nothing, but E z of the original function f z and then star. So, if you are taking f star z and then expectation value with respect to z what you have to do you have to first take the expectation value of the original f z no star and then put star on that and implication of that is suppose f z is z. So, that means, E z f z star is z star and z is suppose mu plus sorry 1 minute z is suppose x plus j y, E z z star will be what first you have to take E z of z and then star E z of z means expected value of this plus j times expected value of this.

$$f(Z) = Z = X + jY$$

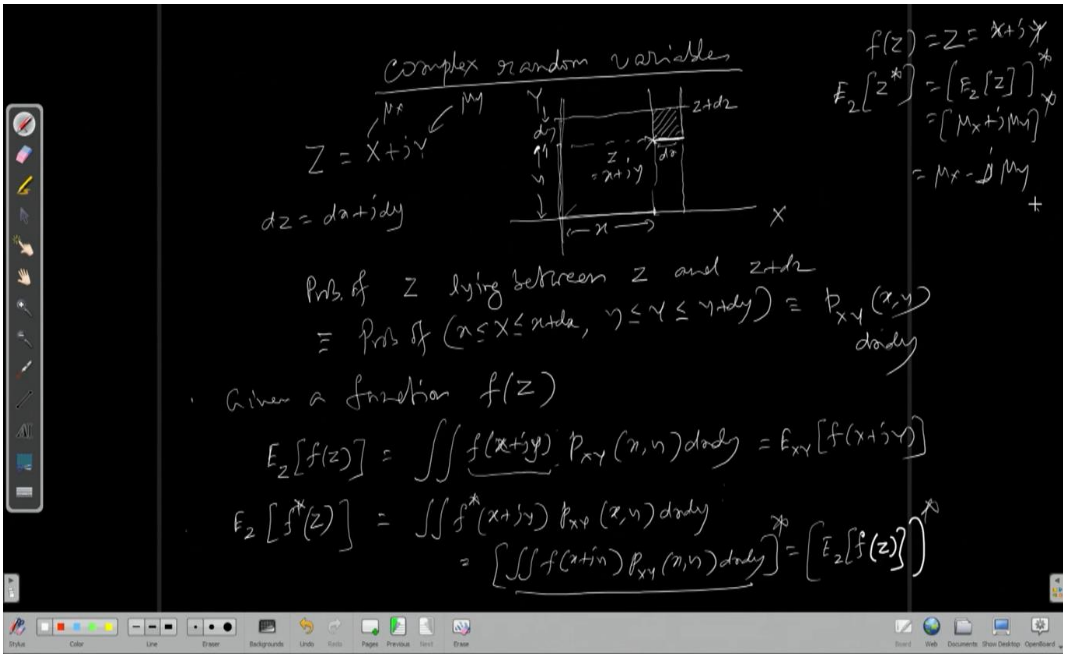
$$E_Z[Z^*] = [E_Z[Z]]^*$$

$$= [\mu_X + j\mu_Y]^*$$

$$= \mu_x - j\,\mu_y$$

So, that was mu x it was mu y expected value. So, mu x plus this is mu x plus j mu y star which is same as mu x minus j mu y. So, this is very important if you have z and you are taking expectation of z star which is suppose z it will be original expected value and then conjugate.

(Refer Slide Time: 32:08)



So, mu x minus j mu y. So, I stop here and we will build up on these complex random variables and in particular we will consider complex covariance and complex covariance matrices in the next class. Thank you very much.