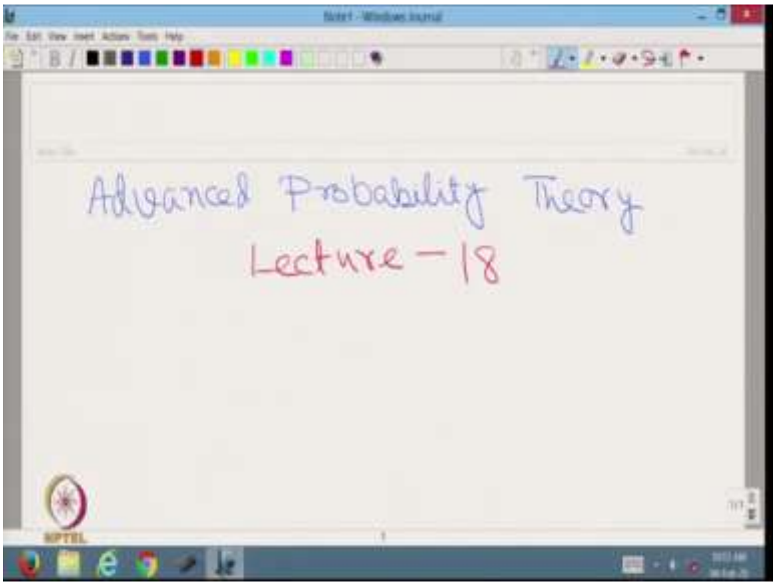
Advanced Probability Theory

Professor Niladri Chatterjee
Department of Mathematics
Indian Institute of Technology, Delhi

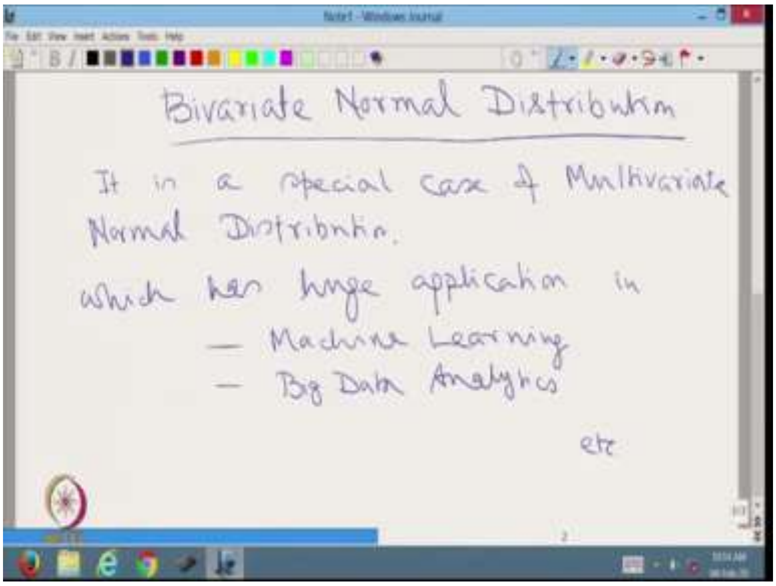
Lecture 18

(Refer Slide Time: 00:26)



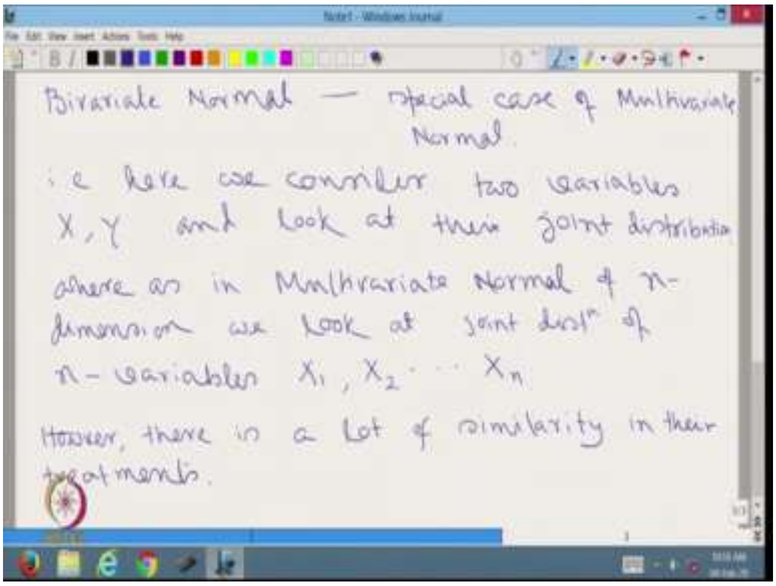
Welcome students to the MOOCs lecture series on Advanced Probability Theory, this is lecture number 18.

(Refer Slide Time: 00:33)



As I said in the last class, that today we shall discuss Bivariate Normal Distribution. In fact it is a special case of multivariate normal distribution and multivariate has huge application in machine learning, big data analytics, etc.

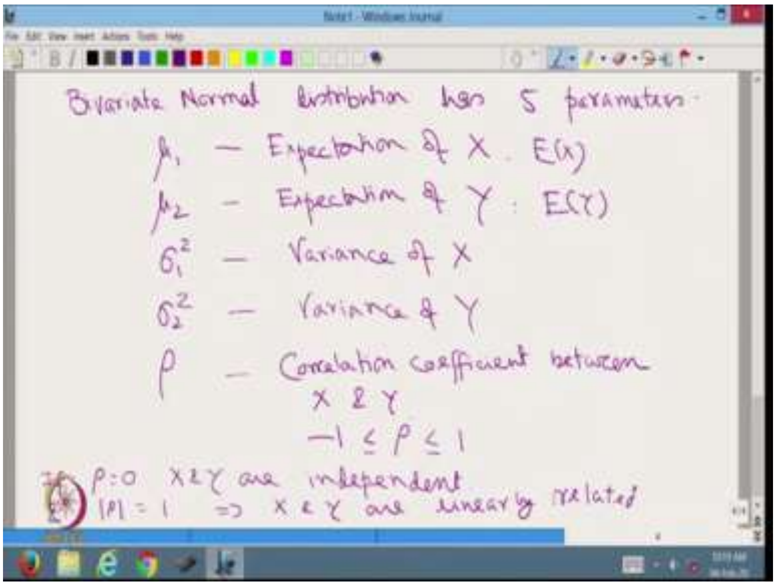
(Refer Slide Time: 01:49)



So bivariate normal as I said the special case of multivariate normal that is, here we consider 2 variables x and y and look at their joint distribution whereas in multivariate normal of n dimension, we look at joint distribution of n variables x1, x2, xn. However, there is lot of similarity in their treatments.

Hence if we study bivariate in great detail we will get a good insight of what happens in multivariate normal case.

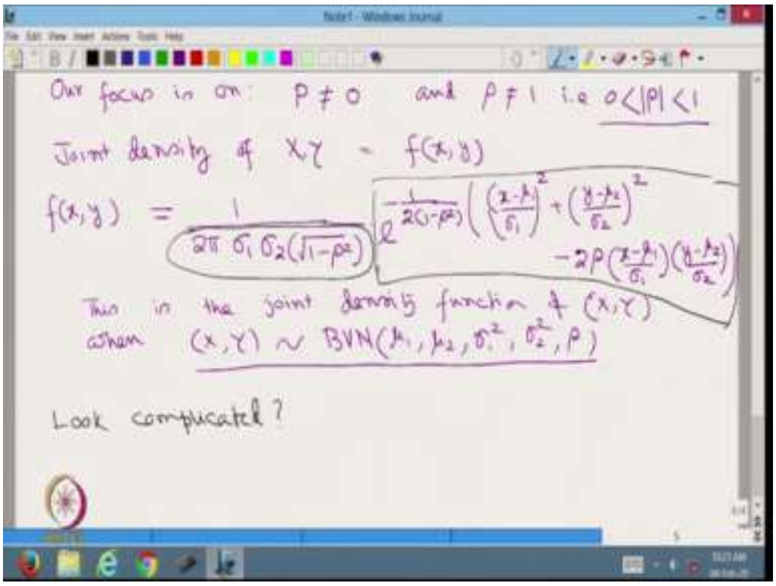
(Refer Slide Time: 04:08)



So bivariate normal distribution has 5 parameters, what are they? Mu1,which is expectation of x that is E x, mu2 expectation of y, that is E of y, sigma 1 square, which is variance of x, sigma 2 square, which is variance of y and rho, which is correlation coefficient between x and y.

And we know that minus 1 less than equal to rho less than or equal to 1. If rho is equal to 0, x and y are independent and mod rho is equal to 1 implies x and y are linearly related.

(Refer Slide Time: 06:13)



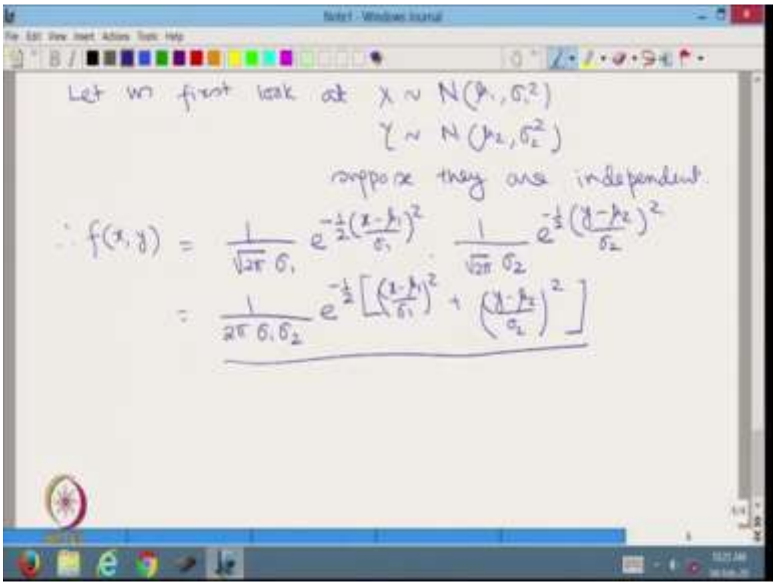
So, our focus is on rho not equal to 0 and rho not equal to 1 that is 0 strictly less than mod rho strictly less than 1.

So, what is therefore the joint density of x y is equal to f x, y and for bivariate normal f x, y is the following 1 upon 2 Pi sigma 1 sigma 2 root over 1 minus rho square into e to the power minus 1 upon 2 into 1 minus rho square into x minus mu 1 upon sigma 1 whole square plus y minus mu 2 upon sigma 2 whole square minus 2 rho x minus mu 1 upon sigma 1 into y minus mu 2 upon sigma 2.

So, this is the joint density function of x, y when x, y is distributed as bivariate normal with mu 1, mu 2, sigma 1 square, sigma 2 square, rho. Now, does it look complicated? Perhaps many of you will find that it is a very complicated expression. But if you understand the mathematics, it will be simple to remember how such expression has been arrived at.

So, let us first divide it into 2 parts, one is this factor and other is the exponential thing. So, I will explain these two now.

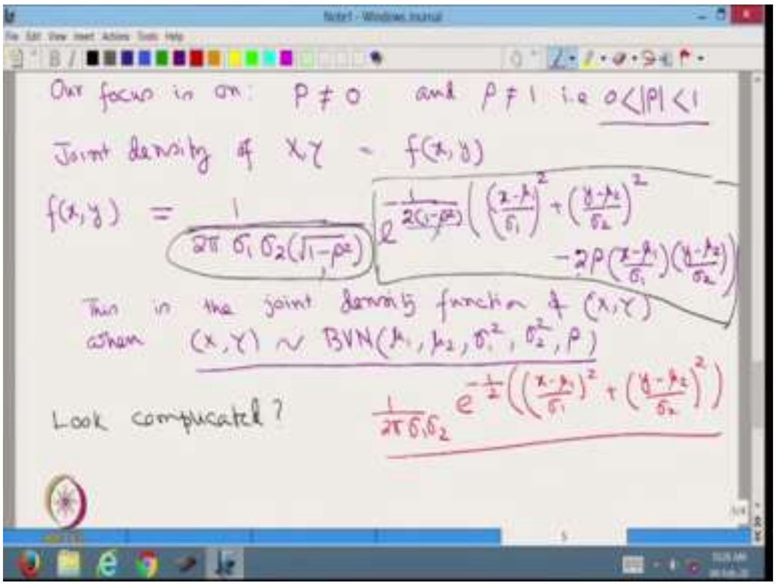
(Refer Slide Time: 09:51)



Now, in order to understand let us first look at x which is normal with mu 1 and sigma 1 square and y is normal with mu 2 and sigma 2 square and suppose they are independent.

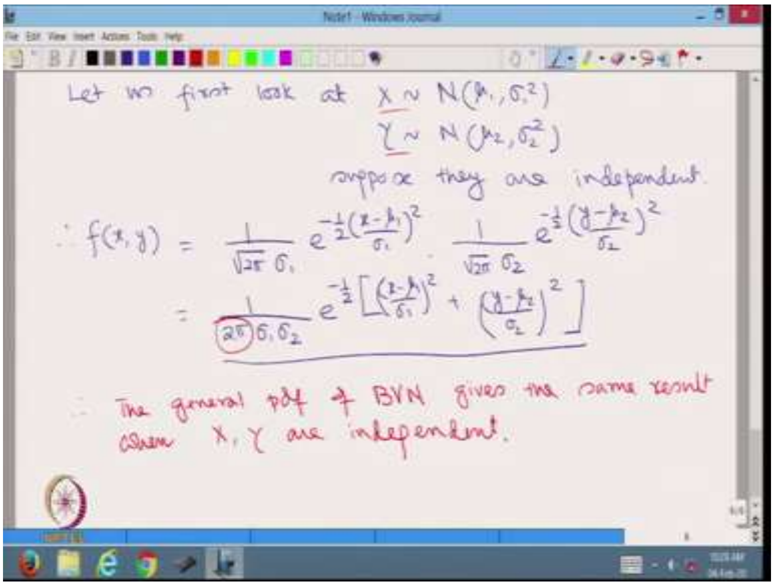
Therefore, f x, y we know is going to be 1 over root over 2 Pi sigma 1 e to the power minus half x minus mu 1 upon sigma 1 whole square multiplied by 1 over root over 2 Pi sigma 2 e to the power minus half y minus mu 2 upon sigma 2 whole square which we can write as 1 over 2 Pi sigma 1 sigma 2 e to the power minus half into x minus mu 1 upon sigma 1 whole square plus y minus mu 2 upon sigma 2 whole square. So, this is the joint density if X and Y are independent.

(Refer Slide Time: 11:36)



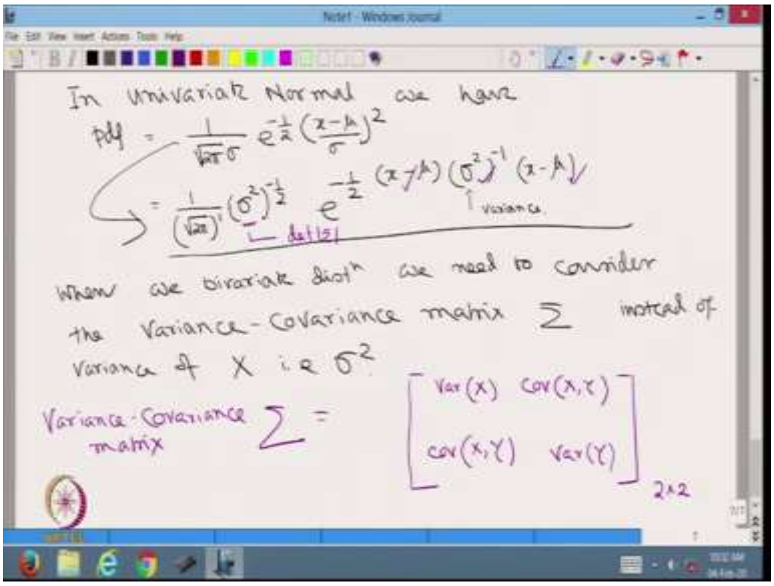
Now, let us look at this formula. We know that, if x and y are independent, then correlation between them is 0. Therefore, this term becomes 1, this becomes 2 into 1 is equal to 2 and this becomes 0. Therefore, we are left with 1 over 2 Pi sigma 1 sigma 2 e to the power minus half into x minus mu 1 upon sigma 1 whole square plus y minus mu 2 upon sigma 2 whole square.

(Refer Slide Time: 12:24)



Now, let us compare this with this, we find that they are identical. Therefore, the general pdf of bivariate normal gives the same result when x and y are independent. But this gives us one more insight that this term has become 2 Pi. Because for one normal, we get 1 root over 2 Pi, for another, we get another root over 2 Pi. Therefore, root over 2 pi square, that is 2 Pi comes into picture.

(Refer Slide Time: 13:26)

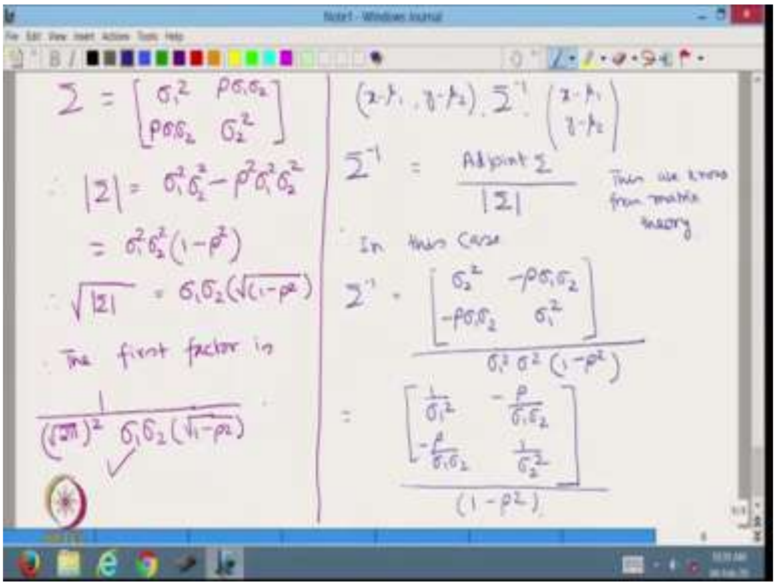


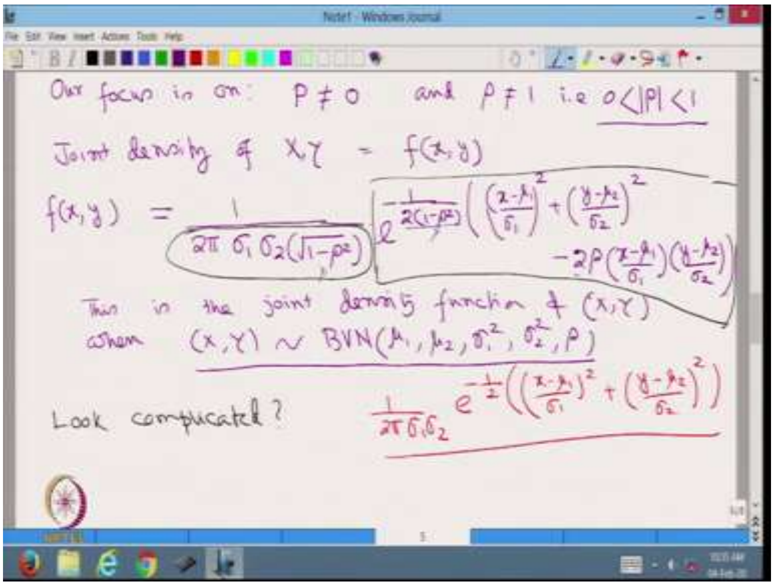
Now, in univariate normal we have pdf is equal to 1 over root over 2 Pi sigma e to the power minus half x minus mu by sigma whole square is equal to let us write it as 1 over root over 2 Pi to the power 1 sigma square to the power minus half into e to the power minus half into x minus mu sigma square to the power minus 1 into x minus mu.

We can decompose this pdf in this form. Here sigma square is the variance. When we have bivariate distribution we need to consider the variance-covariance matrix sigma instead of variance of x, that is sigma square. Now variance-covariance matrix sigma is equal to variance of X covariance between X and Y, covariance between X and Y into and variance of y which is of dimension 2 cross 2.

So, the variance sigma square need to be replaced by the inverse of this 2 by 2 matrix and here because it is a scalar we need to use here, determinant of sigma. So, this is how we need to change and ofcourse, this x minus mu which was a scalar now becomes 2D, what that is 2 dimensional variable.

(Refer Slide Time: 16:53)





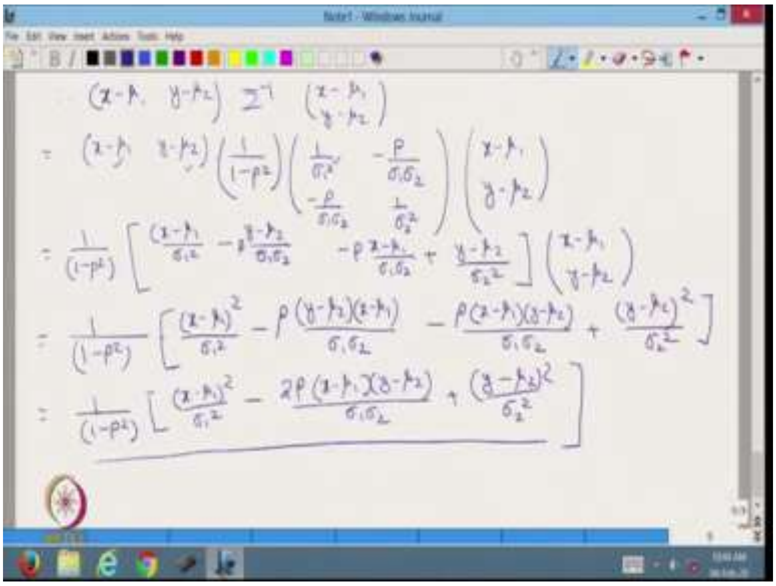
With this understanding, now, let us look at what is sigma. Sigma is equal to sigma 1 square sigma 2 square rho sigma 1 sigma 2 that we know rho sigma 1 sigma 2. Therefore, determinant of sigma is equal to sigma 1 square sigma 2 square minus rho square sigma 1 square sigma 2 square is equal to sigma 1 square sigma 2 square into 1 minus rho square.

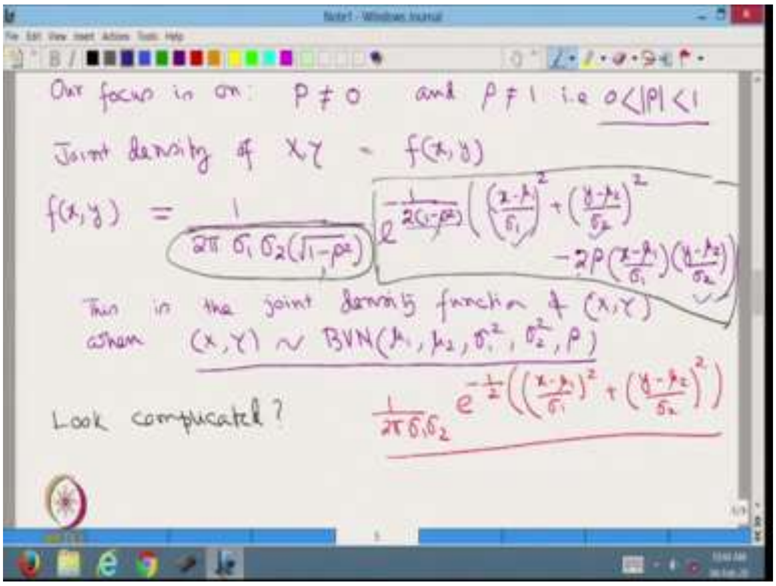
Therefore, square root of determinant of sigma is equal to sigma 1 sigma 2 into root over 1 minus rho square. Therefore, the first factor of the joint pdf 1 over 2 Pi, because it is root over 2 Pi whole square into determinant of sigma to the power minus half. Therefore, we are writing sigma 1 sigma 2 into root over 1 minus rho square. And if we compare with this, it is 2 Pi sigma 1 sigma 2 root over 1 minus rho square. Therefore, the first component is achieved.

Now, let us look at the second component, which is we are writing now x minus mu 1, y minus mu 2 sigma inverse. Now, we are writing it as a column vector, x minus mu 1, y minus mu 2. What is sigma inverse? Sigma inverse is equal to adjoint of sigma from our divided by determinant of sigma, this we know from matrix theory. I hope you remember this.

Therefore, in this case sigma inverse is equal to sigma 2 square minus rho Sigma 1 sigma 2 minus rho sigma 1 sigma 2, sigma 1 square divided by determinant of sigma, which is is equal to sigma 1 square sigma 2 square multiplied by 1 minus rho square is equal to 1 upon sigma 1 square minus rho upon sigma 1 sigma 2 minus rho upon sigma 1 sigma 2 and 1 upon sigma 2 square divided by 1 minus rho square.

(Refer Slide Time: 20:53)





So, therefore, x minus mu 1 y minus mu 2 sigma inverse, x minus mu 1 y minus mu 2 is equal to x minus mu 1 y minus mu 2 into 1 upon 1 minus rho square into 1 upon sigma 1 square minus rho upon sigma 1 sigma 2 minus rho upon sigma 1 sigma 2 1 upon sigma 2 square into x minus mu 1 y minus mu 2 is equal to 1 upon 1 minus rho square into now we are multiplying this with this, this with this and adding up.

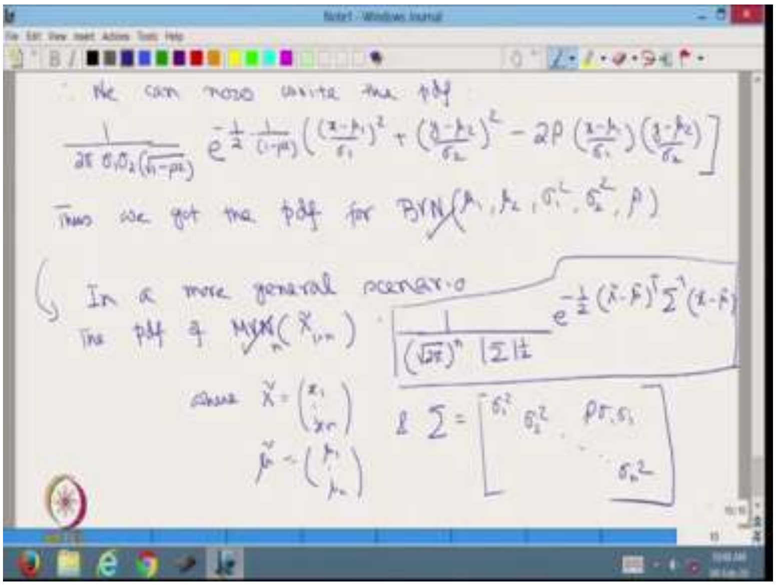
Therefore, we are getting x minus mu 1 upon sigma 1 square minus y minus mu 2 upon sigma 1 sigma 2 and the second term is minus rho x minus mu 1 sigma 1 sigma 2 plus y minus mu 2 sigma 2 square multiplied by x minus mu 1 y minus mu 2 is equal to 1 upon 1 minus rho square.

Now, we are multiplying the taking the dot product is equal to x minus mu 1 whole square upon sigma 1 square minus rho y minus mu 2 into x minus mu 1 upon sigma 1 sigma 2 minus rho, this multiplied by this, x minus mu 1 y minus mu 2 upon sigma 1 sigma 2 plus y minus mu 2 whole square upon sigma 2 square.

Is equal to 1 upon 1 minus rho square into x minus mu 1 whole square upon sigma 1 square minus 2 rho x minus mu 1 into y minus mu 2 upon sigma 1 sigma 2 plus y minus mu 2 upon whole square upon sigma 2 square.

Now, let us compare with the term that we have written before. It is minus 1 upon this quantity which we have obtained as x minus mu 1 upon sigma 1 whole square y minus mu 2 upon sigma 2 whole square minus 2 rho x minus mu 1 upon sigma 1 y minus mu 2 upon sigma 2. So, precisely the same term that we get here.

(Refer Slide Time: 24:57)

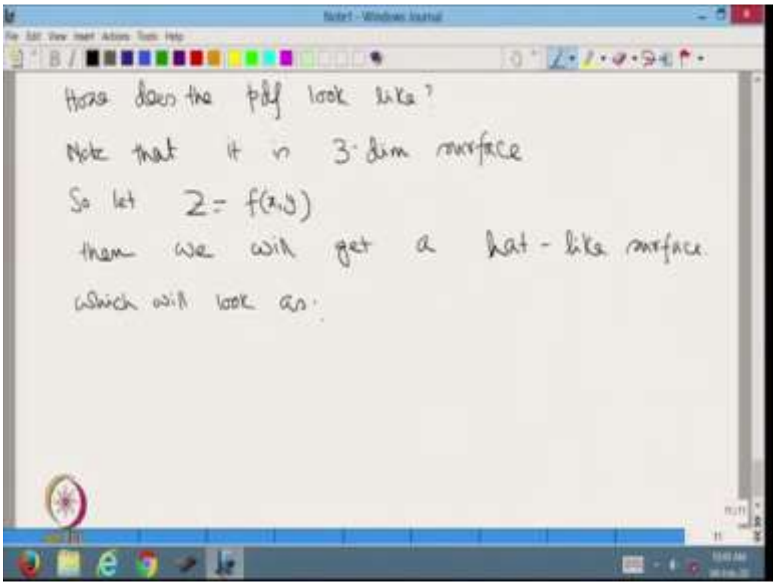


Therefore, we can now write the pdf in a very similar way as 1 over 2 Pi sigma 1 sigma 2 into root over 1 minus rho square e to the power minus half. Now, we are multiplying it with the power that we have just obtained into 1 minus rho square, in the bracket we have x minus mu 1 upon sigma 1 whole square plus y minus mu 2 upon sigma 2 whole square minus 2 rho x minus mu 1 upon sigma 1 y minus mu 2 upon sigma 2.

Thus, we got the pdf for bivariate normal with mu 1, mu 2, sigma 1 square, sigma 2 square and rho. So, very easily, we can now understand how this formula has been arrived at. In a more general scenario, pdf of multivariate normal of dimension n with x vector which is equal to 1 cross n will be 1 over root over 2 Pi whole to the power n determinant of sigma to the power half e to the power minus half into x minus mu transpose sigma inverse x minus mu.

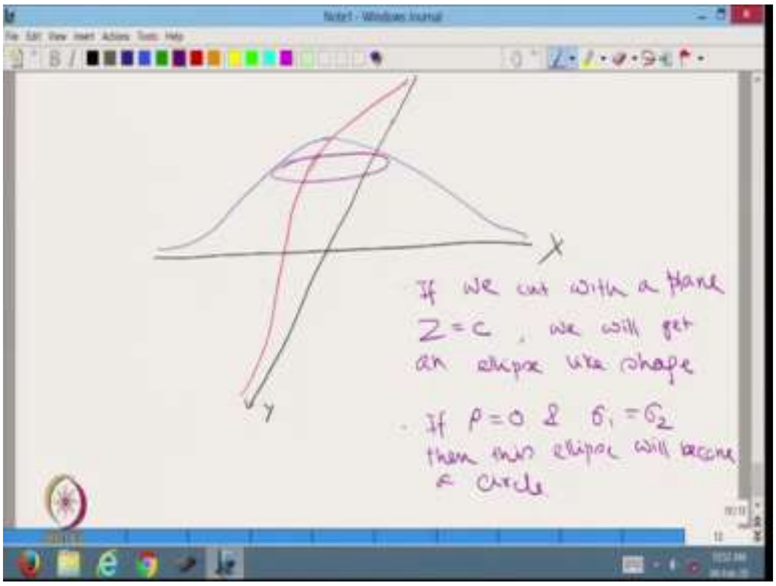
So, this is going to be the formula for the pdf of multivariate normal where x is x1, x2, xn, mu is equal to mu, mu 2, mu n and sigma is equal to sigma 1 square, sigma 2 square, sigma n square and the ijth element here is going to be rho sigma i sigma j. So, this is the way you can generalize multivariate normal, but this is not within the scope of this class. So, we just stop here with that and let us focus on bivariate normal.

(Refer Slide Time: 28:23)



Question is how does the pdf look like? Note that it is a 3 dimensional surface. So, let Z denote f of x, y then we will get a hat-like surface which is in 2D let me draw it.

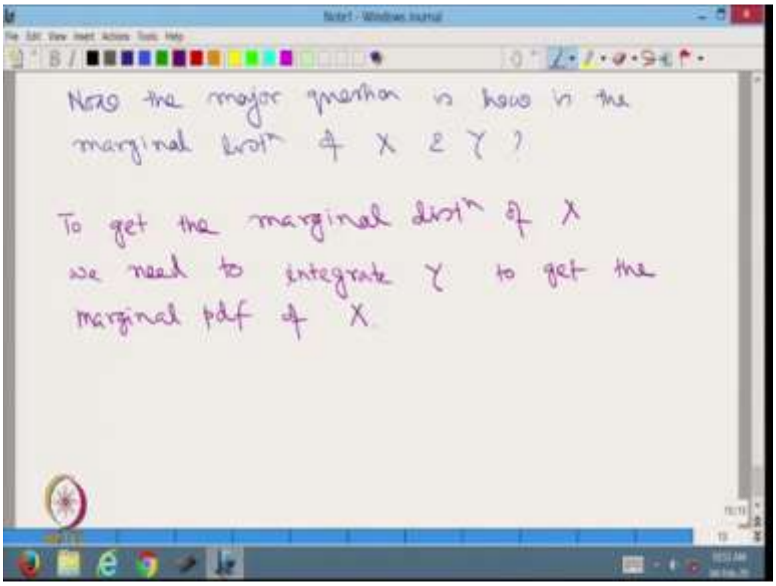
(Refer Slide Time: 29:25)



So, suppose this is my X axis and this is my Y axis, I am trying to draw the curve for x. It will look like this. For y, it will look like this and when we cut with a plane Z is equal to C we will get an ellipse like shape like this. Moreover, if rho is equal to 0 and sigma 1 is equal to sigma 2, then this ellipse will become a circle.

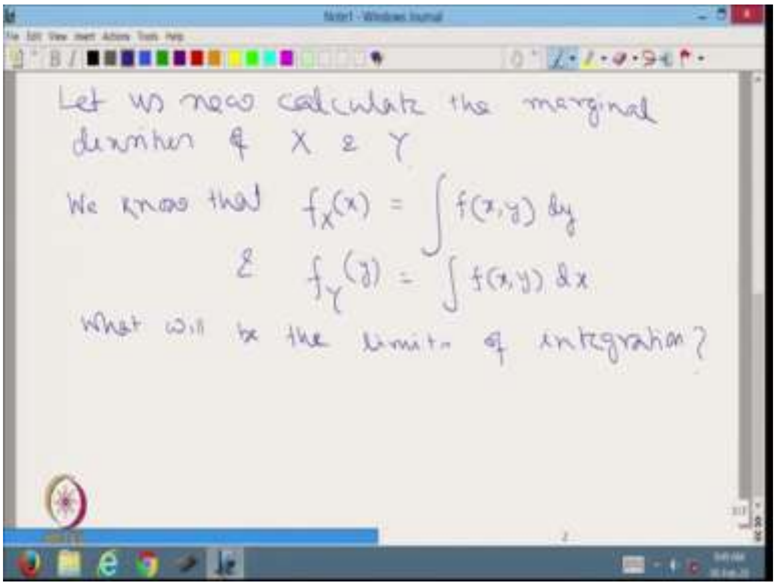
Hope I could make it clear because a hat-like shape is difficult to draw on a 2D plane, but you can visualize it in a similar way.

(Refer Slide Time 31:10)



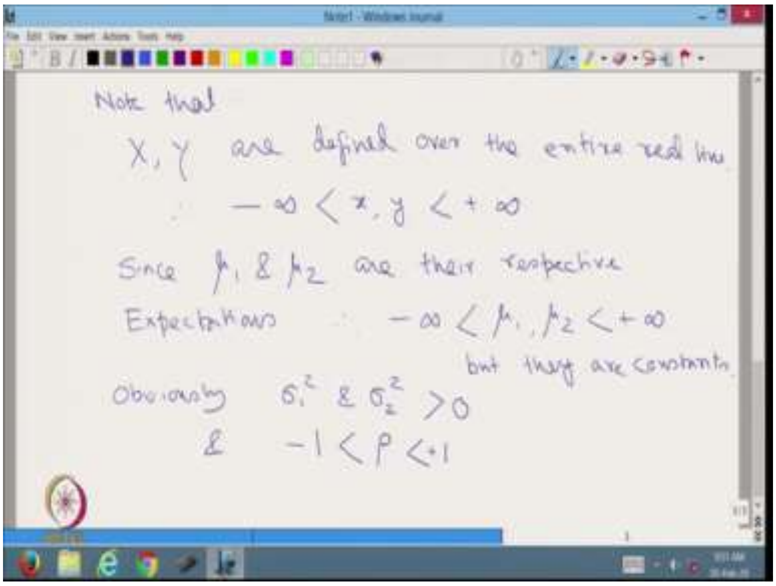
Now, the major question is, how is the marginal distribution of x and y? So, to get the marginal distribution of x, we need to integrate y to get the marginal pdf of X.

(Refer Slide Time: 32:15)



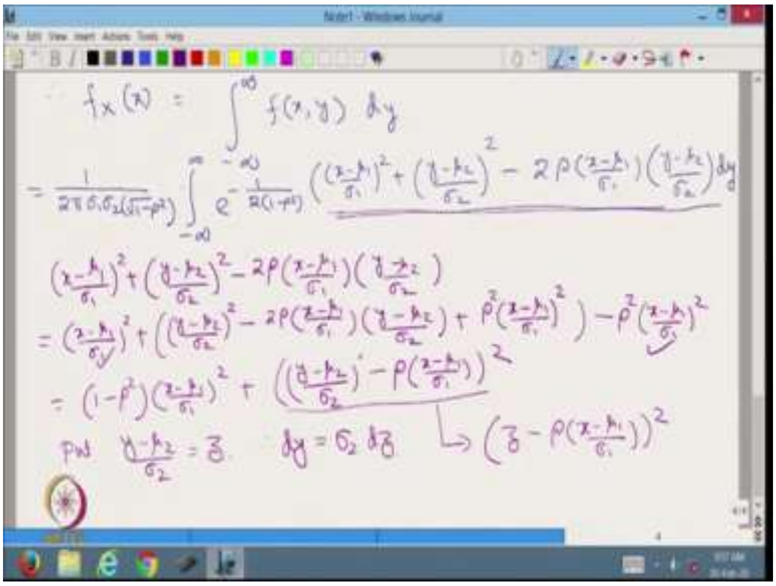
Okay so, let us now calculate the marginal densities of X and Y. We know that fx of x is equal to integration of f x, y dy and fy of y is equal to integration of a f x, y dx. Question is what will be the limit of integration?

(Refer Slide Time: 33:25)



Note that X and Y are defined over the entire real line therefore, minus infinity less than x, y less than plus infinity. Since mu 1 and Mu2 are their respective expectations, therefore, minus infinity less than mu 1 comma mu 2 less than plus infinity. But they are constants. Obviously, sigma 1 square and sigma 2 square are greater than 0 and minus 1 less than rho less than plus 1.

(Refer Slide Time: 35:00)



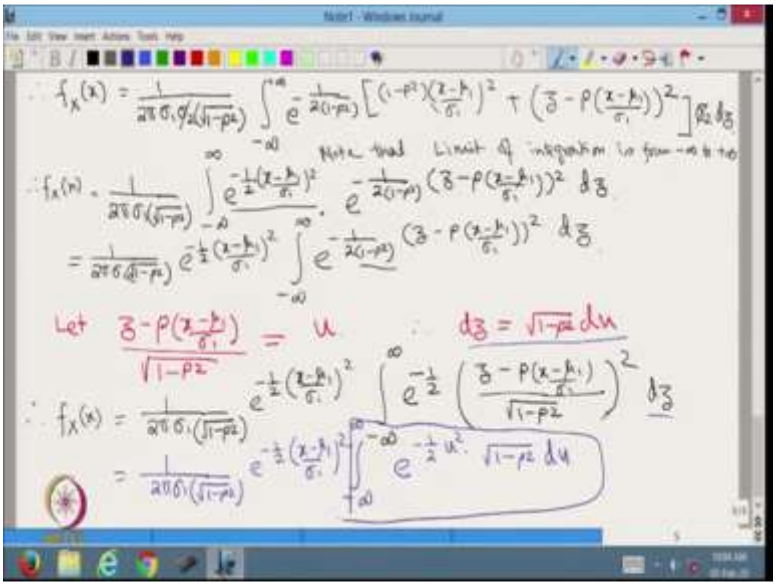
Therefore fx of x is equal to minus infinity to infinity f x, y dy is equal to we take the constant term out, into integration minus infinity to plus infinity e to the power minus 1 upon 2 into 1 minus rho square into x minus mu 1 upon sigma 1 whole square plus y minus mu 2 upon sigma 2 whole square minus 2 rho x minus mu 1 upon sigma 1 y minus mu 2 upon sigma 2 dy.

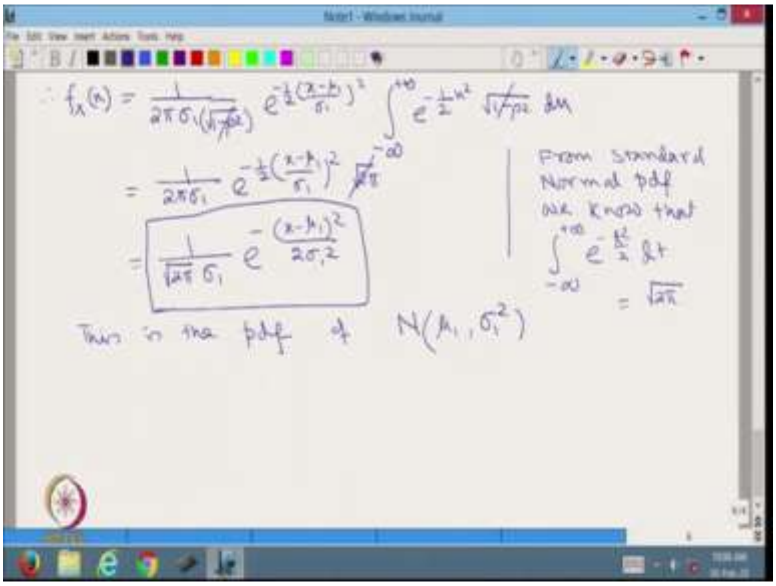
So, this is a very long expression, but we can simplify it. How to do that? Let us just consider this part, x minus mu 1 upon sigma 1 whole square plus y minus mu 2 upon sigma 2 whole square minus 2 rho x minus mu 1 upon sigma 1 into y minus mu 2 upon sigma 2, we can write it as x minus mu 1 upon sigma 1 square plus y minus mu 2 upon sigma 2 square minus 2 rho x minus mu 1 upon sigma 1 into y minus mu 2 upon sigma 2 plus rho square x minus mu 1 upon sigma 1 whole square.

And since we have added this term, we subtract it as well. This is equal to now we can put these 2 terms together, 1 minus rho square into x minus mu 1 upon sigma 1 whole square plus now we can write this term as y minus mu 2 y minus mu 2 upon sigma 2 minus rho x minus mu 1 upon sigma 1 whole square.

Put y minus mu 2 upon sigma 2 is equal to Z. Therefore, d y is equal to sigma 2 dz and this term becomes Z minus rho is x minus mu 1 upon sigma 1 whole square. Let us make this substitution here.

(Refer Slide Time 39:10)





Therefore fx of x, we can write it as 1 upon 2 Pi sigma 1 sigma 2 into root over 1 minus rho square integration of minus infinity to plus infinity e to the power minus 1 upon 2 into 1 minus rho square into 1 minus rho square into x minus mu 1 upon sigma 1 whole square plus Z minus rho into x minus mu 1 upon sigma 1 whole square into sigma 2 into dz.

Note that limit of integration is from minus infinity to plus infinity. Therefore, fx of x is equal to we can cancel this Sigma 2 and therefore, we have 1 over 2 pi sigma 1 into root over 1 minus rho square minus infinity to infinity e to the power minus 2 into x minus mu 1 upon sigma 1 whole square multiplied by e to the power minus 1 upon 2 into 1 minus rho square Z minus rho into x minus mu 1 upon sigma 1 whole square into dz.

Now, this part has nothing to do with Z, therefore, this we can write it as 1 upon 2 Pi sigma 1 into root over 1 minus rho square e to the power minus half into x minus mu 1 upon sigma 1 whole square integration of minus infinity to plus infinity e to the power minus 2, 1 minus rho square Z minus rho into x minus mu 1 upon sigma 1 whole square dz. Now let us make another transformation.

Let Z minus rho x minus mu 1 upon sigma 1 upon root over 1 minus rho square is equal to u. Therefore, dz is equal to root over 1 minus rho square du. Therefore, fx of x, now we can write it is 1 over 2 Pi sigma 1 into root over 1 minus rho square e to the power minus half into x minus mu 1 upon sigma 1 whole square integration of minus infinity to plus infinity e to the power minus half Z minus rho x minus mu 1 upon sigma 1 upon the root over 1 minus rho square whole square dz.

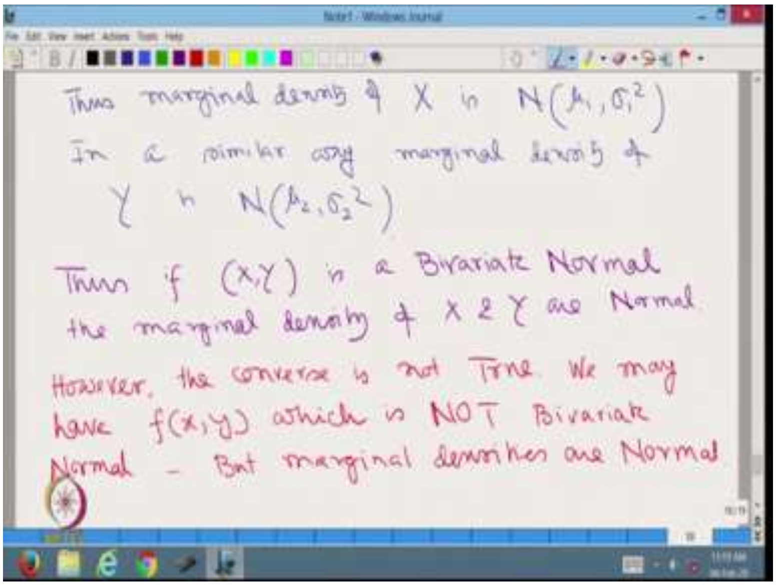
So, what we have done? We have taken this inside the square which we can write it as 1 upon 2 Pi sigma 1 root over 1 minus rho square e to the power minus half into x minus mu 1 upon sigma 1 whole square. Now, here we are making the substitution and therefore we are getting integration minus infinity to plus infinity e to the power minus half u square into root over 1 minus rho square du. We have used this here.

Now, if we look at this portion, therefore, fx of x is equal to 1 over 2 Pi sigma 1 root over 1 minus rho square into e to the power minus half x minus mu 1 upon sigma 1 whole square into integration of minus infinity to plus infinity e to the power minus half u square into root over 1 minus rho square du.

Now, we cancel this. Therefore, this is coming out to be 1 over 2 pi sigma 1 e to the power minus half x minus mu 1 upon sigma 1 whole square and this integration is equal to root over 2 Pi. From standard normal pdf we know that integration minus infinity to plus infinity e to the power minus t square by 2 dt is equal to root over 2 Pi.

So, we cancel it with this and therefore, this is coming out to be 1 over root over 2 Pi sigma 1 e to the power minus x minus mu 1 whole square upon 2 sigma 1 square. This is the pdf of normal with mean mu 1 and variance sigma 1 square.

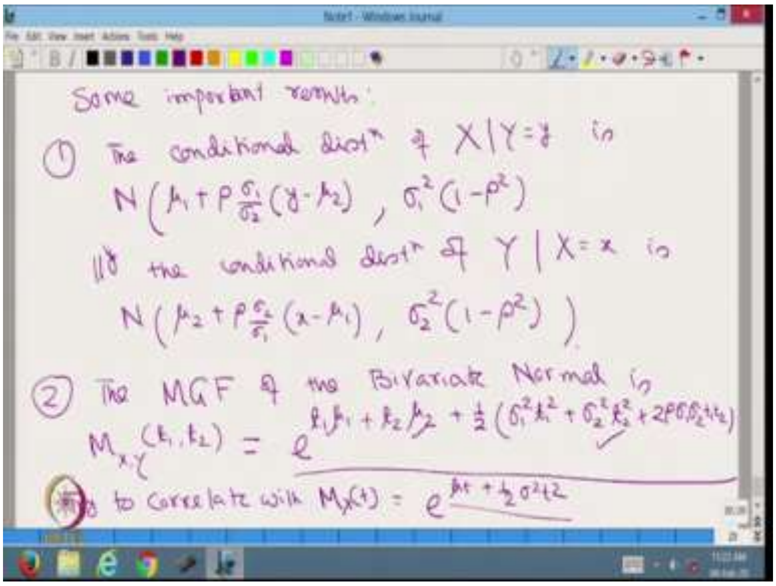
(Refer Slide Time: 47:37)



Thus marginal density of x is normal with mu 1, sigma 1 square. In a similar way, marginal density of y is normal with mu 2, sigma 2 square. Thus, if x, y is a bivariate normal then marginal density of x and y are normal. However, the converse is not true we may have f x, y which is not bivariate normal, but marginal densities are normal.

So, this is a very important concept that from bivariate normal we can get marginals to be normal, but it does not mean that whenever the two marginals are normal, the joint distribution is going to be bivariate normal. So, you need to remember that.

(Refer Slide Time: 50:12)

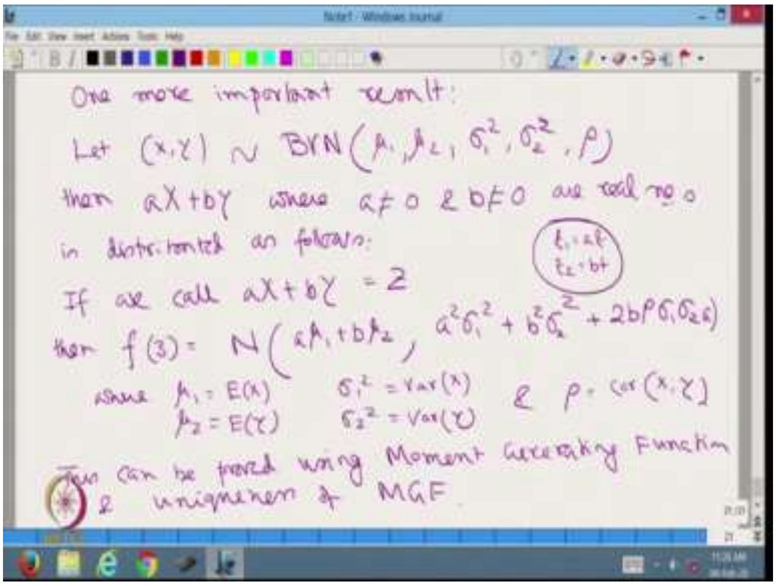


Now, let me give you some important results. One, the conditional distribution of X given Y is equal to y is normal with mean is equal to mu 1 plus rho sigma 1 upon sigma 2 y minus mu 2 and its variance is equal to sigma 1 square into 1 minus rho square. Similarly, the conditional distribution of Y given X is equal to X is normal with mean is equal to mu 2 plus rho sigma 2 upon sigma 1 x minus mu 1 and its variance is going to be sigma 2 square into 1 minus rho square.

The second result is the moment generating function of the bivariate normal is M xy t1, t2 is equal to e to the power t1, mu 1 plus T2 mu 2 plus half into sigma 1 square t1 square plus sigma 2 square t2 square plus 2 rho sigma 1 sigma 2 t1t2. It is apparently slightly complicated but try to correlate with moment generating function of x is equal to e to the power mu t plus half sigma square t square, then you can see that most of the terms are very similar.

Because there are 2 random variables therefore, their respective means are coming and this is coming from the variance-covariance matrix.

(Refer Slide Time: 53:31)



So, one more important result is the following. Let x, y be distributed as bivariate normal with mu 1, mu 2, sigma 1 square, sigma 2 square and rho, then aX plus bY, where a not equal to 0 and b not equal to 0 are real numbers is distributed as follows.

If we call aX plus bY is equal to z, then fz is equal to normal, with mean is equal to a mu 1 plus b mu2, and variance is equal to a square sigma 1 square plus b square sigma 2 square plus 2b rho sigma 1 sigma 2 a, where mu 1 is equal to expectation of x, mu 2 is equal to expectation of y, sigma 1 square is equal to variance of x, sigma 2 square is equal to variance of y and rho is equal to correlation between x and y.

This can be proved using moment generating function and uniqueness of that. I suggest that you try this with aX plus bY, where a and b are not equal to 0 and you have to take, you will find that t1 is equal to at and t2 is equal to bt, try this, That will give you a lot of insight into the mathematics of bivariate normal. Okay friend, I stop here today. In the next class we shall start functions of random variables and their distributions. Thank you.