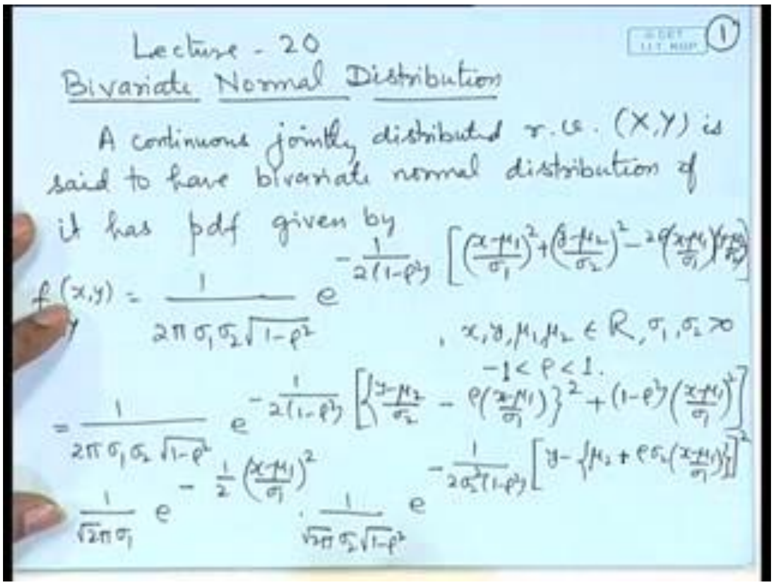
Probability and Statistics
Prof. Somesh Kumar
Department of Mathematics
Indian Institute Technology, Kharagpur

Lecture – 39
Bivariate Normal Distribution – I

In the last 2 lectures, we have discussed the distributions of a Bivariate Random Variables. So, we looked at how to derive the marginal distributions and the conditional distributions. We also discussed various characteristics of the joint distributions such as the moments, product moments, covariance and the coefficient of correlation and we also looked at some of the features of these characteristics.

Today I will introduce a particular joint distribution, it is known as bivariate normal distribution.

(Refer Slide Time: 00:59)



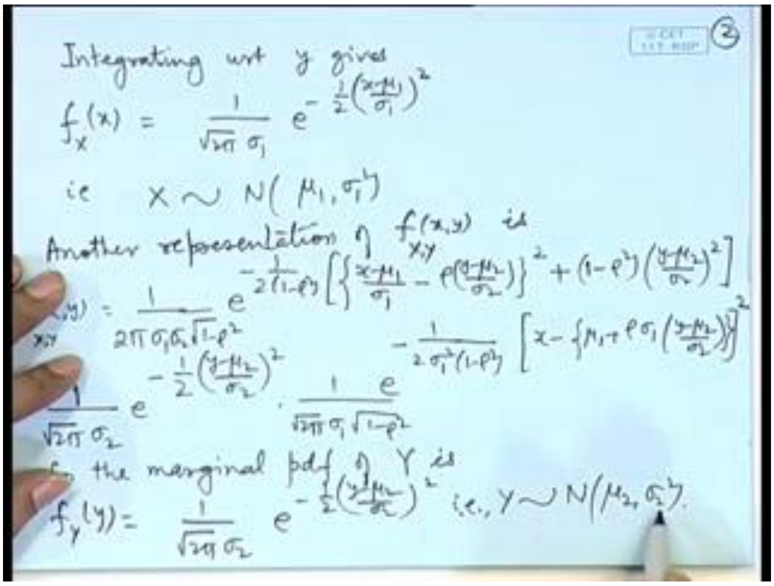
A continuous jointly distributed random variable X Y is said to have bivariate normal distribution if it has the probability density function given by f x y equal to 1 by 2 pi sigma 1 sigma 2 root 1 minus rho square e to the power minus 1 by 2 1 minus rho square x minus mu 1 by sigma 1 square plus y minus mu 2 by sigma 2 square minus 2 rho x minus mu 1 by sigma 1 y minus mu 2 by sigma 2. Here the range of x y mu 1 mu 2 is the whole real line and sigma 1 sigma 2 are positive and rho is between minus 1 and plus 1.

First of all we look at that what are the marginal distributions and the conditional distributions, and overall a structure of this bivariate normal distribution we will likely study. Suppose we want to find out the marginal distribution of x in that case we need to integrate this joint distribution with respect to y. A closure examination of the density function reveals that in the exponent we have a term which is a term like which appear in the exponent of the normal distribution. So if you want to integrate with respect to y we can convert it into a density with respect to y. So, that suggests that we make a perfect square in y. So, we can factorize it as 1 by 2 pi sigma 1 sigma 2 root 1 minus rho square e to the power minus 1 by 2 1 minus rho square.

Here if we make a square in y then we have y minus mu 2 by sigma 2 minus rho x minus mu 1 by sigma 1 whole square. Now this square corresponds to this and the cross product 1 corresponds to this so; that means, I have added rho square into x minus mu 1 by sigma 1 whole square. So, if we subtracts this I will get the term as 1 minus rho square x minus mu 1 by sigma 1 square which we can write as 1 by root 2 pi sigma 1 e to the power minus 1 by 2 x minus mu 1 by sigma 1 square and 1 by root 2 pi sigma 2 root 1 minus rho square e to the power minus 1 by 2 sigma 2 square 1 minus rho square y minus mu 2 minus rho sigma 2 x minus mu 1 by sigma 1. This I can write as plus and arrange that in the bracket square.

You can see here that the first term is a normal density for x and the second term is a normal density for y. So, if we want to find out the marginal distribution of x we can integrate this with respect to y and we notice here that this entire term denotes a distribution which is normal with mean mu 2 plus rho sigma 2 x minus mu 1 by sigma 1 and variance sigma 2 square into 1 minus rho square. So, if we integrate with respect to y, this term will give us unity and we will get only this term.

(Refer Slide Time: 05:42)



Integration with respect to y gives f x x equal to 1 by root 2 pi sigma 1 e to the power minus 1 by 2 x minus mu 1 by sigma 1 square. That is the marginal distribution of x is normal mu 1 sigma 1 square. In a similar way we can a split this term when we want to integrate with respect to x then I make it as a perfect to square in x, so we will write x minus mu 1 by sigma 1 minus rho y minus mu 2 by sigma 2.

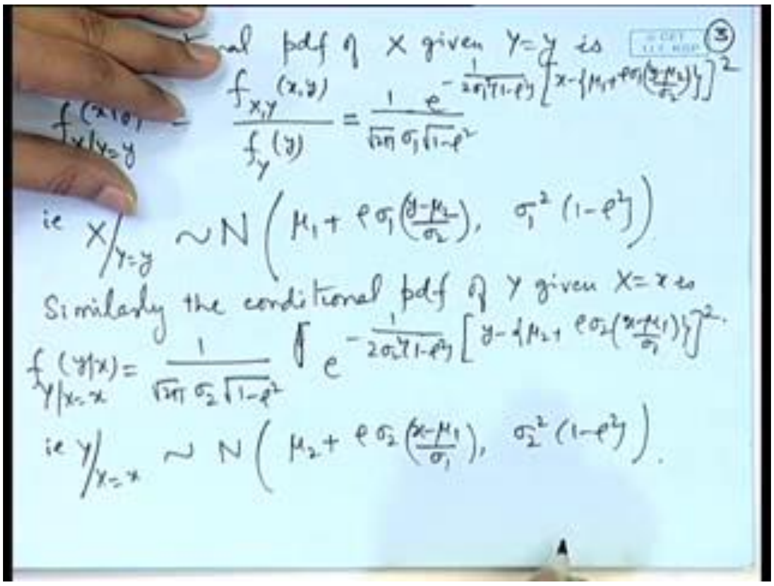
Another way of writing is; another representation of f x y can be 1 by 2 pi sigma 1 sigma 2 root 1 minus rho square e to the power minus 1 by 2 1 minus rho square, and now I make a square with respect to x so x minus mu 1 by sigma 1 minus rho y minus mu 2 by sigma 2 whole square. So comparing with the joint density, we can see here that x minus mu 1 by sigma 1 whole square that is coming here and the cross product term is minus 2 rho x minus mu 1 by sigma 1 into y minus mu 2 by sigma 2 which is the term appearing here. So, we have added the term rho square y minus mu 2 by sigma 2 square. So, subtracting this we get 1 minus rho square y minus mu 2 by sigma 2 whole square.

We can write it as 1 by root 2 pi sigma 2 e to the power minus 1 by 2 y minus mu 2 by sigma 2 whole square 1 by root 2 pi sigma 1 root 1 minus rho square e to the power minus 1 by 2 sigma 1 square 1 minus rho square and x minus mu 1 plus rho sigma 1 y minus mu 2 by sigma 2 whole square. So, notice here that the second term is a density of normal random variable with mean mu 1 plus rho sigma 1 into y minus mu 2 by sigma 2 and variance sigma 1 square into 1 minus rho square.

If we integrate this joint density with respect to x the term, this term integrates to 1 and we are left with a normal density. So, the marginal pdf of y is f y y that is equal to 1 by root 2 pi sigma 2 e to the power minus 1 by 2 y minus mu 2 by sigma 2 square; that is the marginal distribution of y is normal mu 2 sigma 2 square. So, we come across these interesting phenomena that if x y follow a joint bivariate normal distribution then the marginal distributions of x is normal mu 1 sigma square and the marginal distribution of y is normal mu 2 sigma 2 square. That means, given a joint bivariate normal distribution the marginal distributions are univariate normal.

Now, we also calculate the conditional distributions of x given y and y given x. Now if you look at the conditional distribution of x given y, then we have to divide the joint distribution of x y by the marginal distribution of y. Now from this breakup, we can see that this joint distribution if we divide by the marginal of y this term gets canceled out and we are left with this particular term which is nothing but the normal distribution. This proves that the conditional distribution of x given y is normal, and the mean and the variance as specified here.

(Refer Slide Time: 10:41)

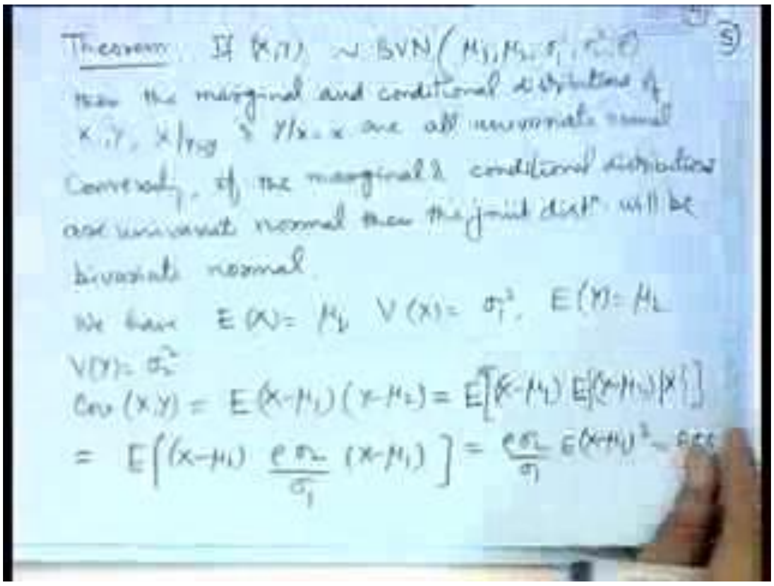


So we have the conditional probability density function of x given y is equal to y that is obtained as the joint distribution divided by the marginal distribution of y. So, after simplification it is equal to 1 by root 2 pi sigma 1 root 1 minus rho square e to the power minus 1 by 2 sigma 1 square 1 minus rho square x minus mu 1 plus rho sigma 1 y minus mu 2 by sigma 2 whole square. That is we can say that x given y is equal to y follows a normal with mean mu 1 plus rho sigma 1 y minus mu 2 by sigma 2 and variance sigma 1 square into 1 minus rho square.

In a similar way notice here that the joint distribution of x y were earlier factorized like this and if we divide by the marginal distribution of x then this term gets canceled out and we are left with this term which is again a normal distribution with a certain mean and a certain variance. This proves that the conditional distribution of; similarly the conditional pdf of y given x that is obtained as 1 by root 2 pi sigma 2 root 1 minus rho square e to the power minus 1 by 2 sigma 2 square 1 minus rho square y minus mu 2 plus rho sigma 2 x minus mu 1 by sigma 1 whole square. That is y given x is equal to x follows normal with mean mu 2 plus rho sigma 2 x minus mu 1 by sigma 1 sigma 2 square 1 minus whole square.

We conclude that if the joint distribution is bivariate normal the marginal as well as the conditional distributions are univariate normal. Now the converse of this is also true if the conditionals and the marginals are univariate normal the joint distribution will be bivariate normal. So, this is also a characterizing property of the bivariate normal distribution.

(Refer Slide Time: 13:52)



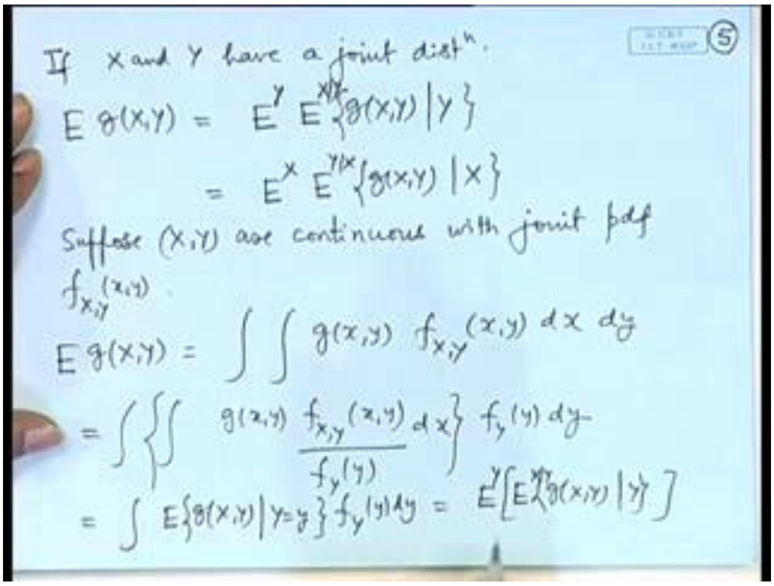
We can state it as a theorem: if x y follows bivariate normal with some parameters say mu 1, mu 2, sigma 1 square, sigma 2 square and rho then the marginal and conditional distributions of x y x given y and y given x are all univariate normal.

Conversely, if the marginal and conditional distributions are univariate normal then the joint distribution will be bivariate normal. So, this is quite useful in obtaining any probability related to marginal are the conditional distributions of the x and y, because we can make use of the standard normal distribution by making a suitable transformation. Any joint probability statement about bivariate normal distribution will need the tables of a standard bivariate normal distribution. By a standard bivariate normal distribution, we mean mu 1 is equal to 0, mu 2 is equal to 0, sigma 1 square and sigma 2 square is equal to 1; but rho will is still be there and therefore the several tables will be required with respect to which will be related to the joint probabilities of the bivariate normal distribution.

Since the marginal distributions are identified, we have expectation of x is equal to mu 1 variance of x is equal to sigma 1 square expectation of y is equal to mu 2 and variance of y is equal to sigma 2 square. Now we also considered the covariance term between x and y. So, the covariance term between x and y is expectation of x minus mu 1 into y minus mu 2. Now this product centre product moment can be calculated by the joint integration of the density function multiplied by this function.

However at this stage, we introduce some formula for revaluation of the joint moments.

(Refer Slide Time: 17:18)



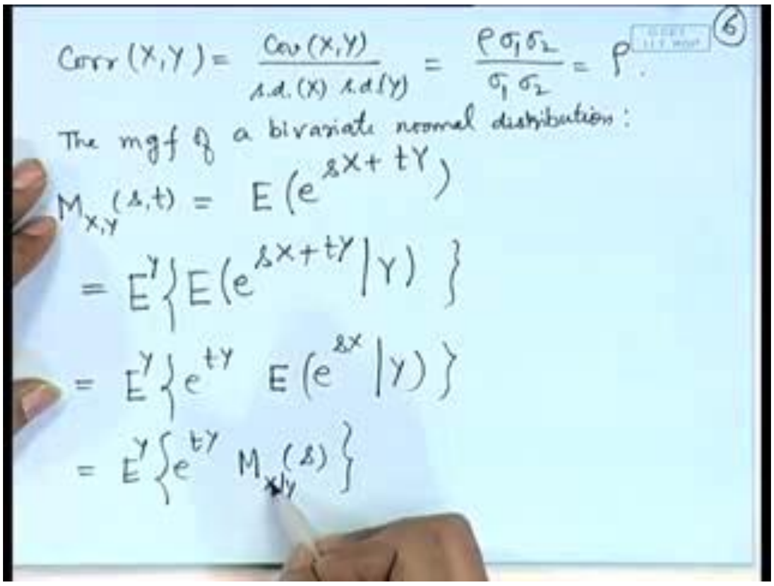
If x and y have a joint distribution then in general expectation of a function can be calculated in a stages. We may calculate firstly the conditional and then with respect to marginal or alternatively we may considered it as a expectation of g x y given x y given x provided of course the expectations do exist.

Let me give a roughly sketch of the proof: suppose x and y are continuous with joint pdf say f x y. So, expectation of g x y you can express as integral g x y f x y x y, suppose we keep the order of integration as dx, dy. Then this we can express as g x y f x y divided by f y multiplied by f y dy. So this quantity, inner quantity is nothing but the expectation of g x y given y is equal to y multiply by the density of y which is nothing but expectation of expectation g x y given y. That means the joint expectations can be calculated in a stages; firstly with respect to a conditional distribution and then with respect to marginal distribution in either order.

If we make use of this then expectation of x minus mu 1 into y minus mu 2 we can write it as x minus mu 1 into expectation of y minus mu 2 given x . So, inner expectation is the conditional expectation with respect to the distribution of y given x and the outer is with respect to x. The conditional distribution of y given x was calculated to be a univariate normal distribution and the mean was mu 2 plus certain term. So, expectation of y given x will be mu 2 plus rho sigma 2 x minus mu 1 by sigma 1. Therefore, expectation of y minus mu 2 given x will be equal to x minus mu 1 rho sigma 2 by sigma 1 x minus mu 1 which is nothing but rho sigma 2 by sigma 1 expectation of x minus mu 1 square which is sigma 1 square. So, it is rho sigma 1 sigma 2.

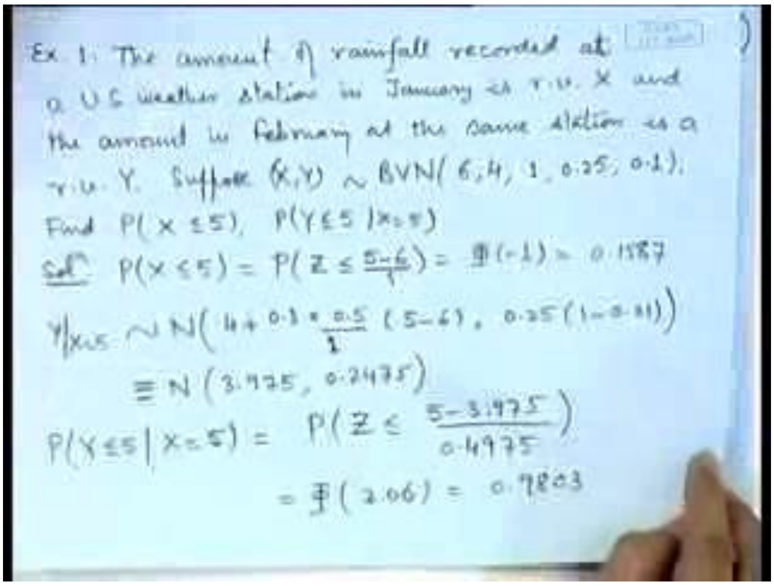
So, we conclude that the covariance of the x y in a bivariate normal distribution is given by rho sigma 1 sigma 2.

(Refer Slide Time: 21:10)



Therefore, we can calculate the coefficient of correlation between x y as covariance between x y divided by a standard deviation of x standard deviation of y; that is equal to rho sigma 1 sigma 2 by sigma 1 sigma 2 that is equal to. So, the parameter rho of a bivariate normal distribution denotes the correlation coefficient between the random variables x and y.

(Refer Slide Time: 21:47)



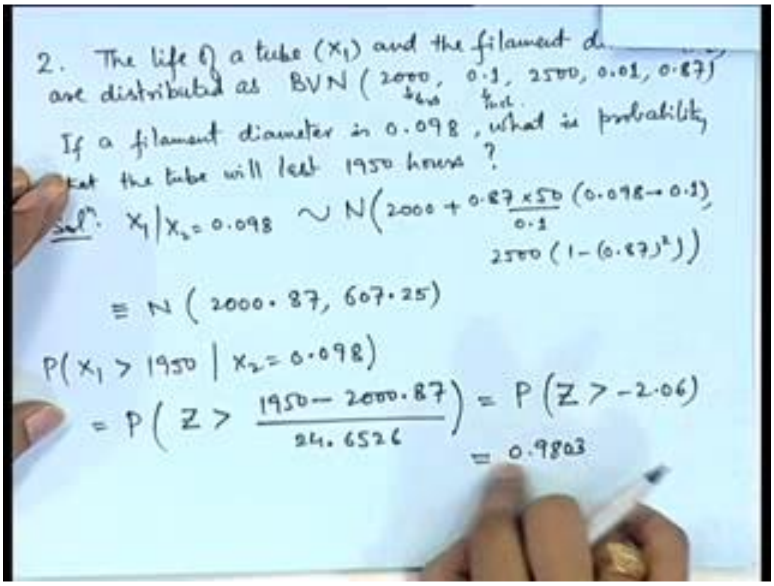
Let us look at a problem here: the amount of rainfall recorded at a US weather station in January is a random variable X and the amount of rainfall recorded in February at the same station is a random variable Y. Suppose the distribution of x and y is observed to be a bivariate normal distribution with mean 6. So, the mean of the random variable x is 6 the mean of the random variable y is 4. So, suppose it is in measured in inches because it is the amount of rainfall as centimeters the variances are 1 and 0.25 and rho is equal to 0.1. We are interested to calculate what is the probability that x is less than or equal to 5 or what is the probability of y being less than or equal to 5 given that x is equal to 5.

Notice here probability of x less than or equal to 5 can be calculated from the marginal distribution of x which is having mean 6 and variance unity. So, it is simply transform to the standard normal probability as Z less than or equal to 5 minus 6 by 1, here Z denotes the standard normal random variable. So, from 5 here subtracted the mean of stand divided by the standard deviation which is equivalent to the CDF value of the standard normal variate at minus 1 which we take see from the tables of normal distribution as 0.1587.

Suppose we are interested in the probability of y less than or equal to 5 given that in January the rainfall is 5. So, we need the conditional probability of y less than or equal to 5 given x is equal to 5. For this we will firstly calculate the conditional distribution of y given x is equal to 5. Now making use of the conditional distribution of y given x which is given by normal with mean mu 2 plus rho sigma 2 x minus mu 1 by sigma 1. So, here mu 2 is 4 rho is 0.1, sigma 2 is 0.5, sigma 2 is 1 and the point x is small; a small x is 5. So, x is 5 and mu 1 is 6.

This is the mean of the conditional distribution of y given x. So, after simplification this turns out to be 3.975. The variance of the conditional distribution is sigma 2 square into 1 minus rho square which is 0.25 into 1 minus 0.01, so it is evaluated to be 0.2475. So, the conditional probability of y less than or equal to 5 given x is equal to 5 can be calculated from this distribution. So, we transform it to the standard normal distribution. So, this Z less than or equal to 5 minus 3.975 divided by square root of this that is 0.4975 so, after simplification it turns out to be phi of 2.06, if it says 0.9803 which is quite high probability. But that is understandable because, in general there is more rain so since the variables are correlated it is affecting the probability of y also.

(Refer Slide Time: 25:23)



Let us take up another example of a similar nature. The life of a tube which is measured as random variable x 1 and the filament diameter which is measured as a random variable x 2, the life is measured in say hours and the diameter is measured in inches. They are distributed as a bivariate normal distribution with mu 1 is equal to 2000 hours, mu 2 is 0.1 inches, the sigma 1 square is 2500, sigma 2 square is 0.01, and the coefficient of correlation is 0.87. So, the manufacturer may use the filament diameter length which can be measured to estimate the life of the tube.

If a filament diameter is 0.0981, what is the probability that the tube will last 1950 hours? So, we are interested to calculate, what is the probability of surviving till 1950 hours given that the diameter is 0.098 inches? For this we need the conditional distribution of x 1 given x 2 is equal to point 098. So, we make use of the formula for the conditional distribution of x given y here. So, that is mu 1 that is 2000 plus rho 0.87 sigma 1 is 50 divided by sigma 2 is 0.1 y minus mu 2, so y is the point at which are conditioning that is 0.098 minus mu 2 that is 0.1. So, after simplification this turns out to be 2000.87, and the variance here is sigma 1 square into one minus rho square which is equal to 607.25.

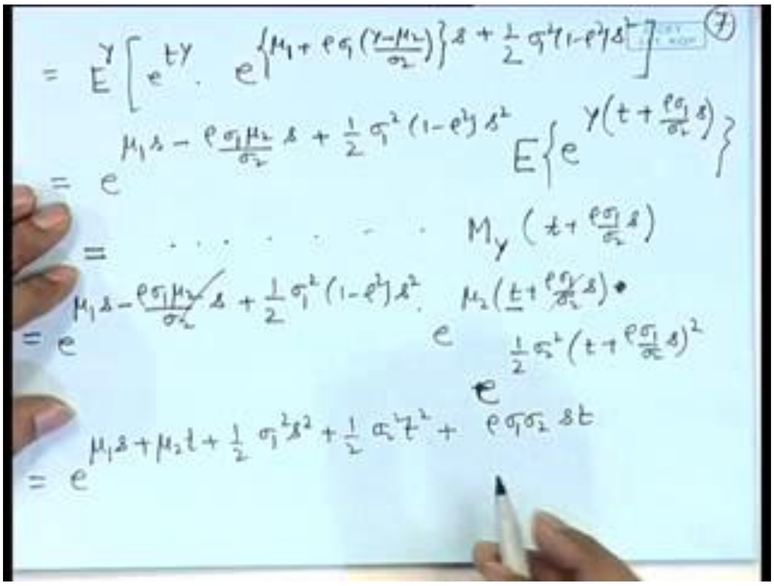
The conditional probability of x 1 greater than 1950 given that x 2 is equal to 0.098 can be calculated using this univariate normal distribution. So, after transformation 2 standard normal we get it as probability of Z greater than minus 2.06 which is evaluated as 0.9803. So, likewise any probability statement related to the marginal distributions or the conditional distributions of x or y or x given y or y given x can be calculated using the univariate normal properties.

We also look at the moment generating function of a bivariate normal distribution; the moment generating function of a bivariate normal distribution. So, it is defined as M x y s t that is equal to expectation of e to the power s x plus t y. Now again you consider there is some function g of x y. So, the joint expectation we can calculate easily in terms of conditional and the marginal expectations. So, we will use that; we can write as expectation of expectation e to the power sx plus ty given say in the previous one we have done the calculation using conditional distribution of x so we can use the conditional distributions of y here. Now, given y this e to the power ty terms is fixed so we can separate it out and we are left with expectation of e to the power sx given y.

Now, notice here that this inner expectation is nothing but the moment generating function of the conditional distribution of x given y. So, this is equal to expectation of e to the power t y into the moment generating function of the conditional distribution of x given y at the point s. Now here the conditional distribution of x given y is univariate normal, we already know the form of the moment generating function of a univariate normal distributions; suppose the normal mu sigma square distribution is there then we have seen that the mgf is represented as e to the power mu t plus half sigma square t square.

Here, the point is s in place of t and x given y the distribution has the parameters mu 1 plus rho sigma 1 y minus mu 2 by sigma 2 and sigma 1 square 1 minus rho square. So, we make use of this.

(Refer Slide Time: 30:50)



This can be expressed as expectation of y e to the power ty e to the power mu 1 plus rho sigma 1 y minus mu 2 by sigma 2 into s plus half sigma 1 square 1 minus rho square s square. So, this is the value coming after substituting the value of the moment generating function of the conditional distribution of x given y which is univariate normal and therefore, the form is known to us.

Now here there are certain constant terms and we can separate it out e to the power mu 1 s minus rho sigma 1 mu 2 by sigma 2 s plus half sigma 1 square 1 minus rho square s square. We have expectation of e to the power y t plus rho sigma 1 by sigma 2 s. So, we notice here that this is nothing but the moment generating function of y at the point t plus rho sigma 1 by sigma 2 s, so this term is there.

So, notice here that the distribution of y is again univariate normal with parameters mu 2 and sigma 2 square, therefore the moment generating function has a known form. In place of the point t we substitute t plus rho sigma 1 by sigma 2 s. So, we write it as e to the power mu 1 s minus rho sigma 1 mu 2 by sigma 2 s plus half sigma 1 square 1 minus rho square s square e to the power mu 2 t plus rho sigma 1 by sigma 2 s plus. So, I will write it as e to the power half sigma 2 square t plus rho sigma 1 by sigma 2 s whole square.

We have e to the power mu 1 s plus mu 2 t; that is this term. Now we note here minus rho by rho sigma 1 sigma by sigma 2 mu 2 s, this term is coming here also as a plus sign plus rho sigma 1 by sigma 2 mu 2 s. So, this term gets cancel with this term. Then we have half sigma 1 square s square and half sigma 2 square t square. Now when we take a square here it is becoming twice rho sigma 1 by sigma 2 st; so sigma 2 and sigma 2 square so, you will get it as plus rho sigma 1 sigma 2 st. And the square term here rho square sigma 1 square s square with a half here we will get canceled with minus half sigma 1 square rho square s square. We are left with this term as the mgf of the bivariate normal distribution.

So, notice here that e to the power mu 1 s plus half sigma 1 s square denotes the mgf of the normal distribution with parameter mu 1 and a sigma 1 square that is the mgf of x. Similarly, e to the power mu 2 t plus half sigma square t square denotes the mgf of y. So, we have these terms and an additional term coming here. So, using this we can prove certain more properties regarding the bivariate normal distribution.