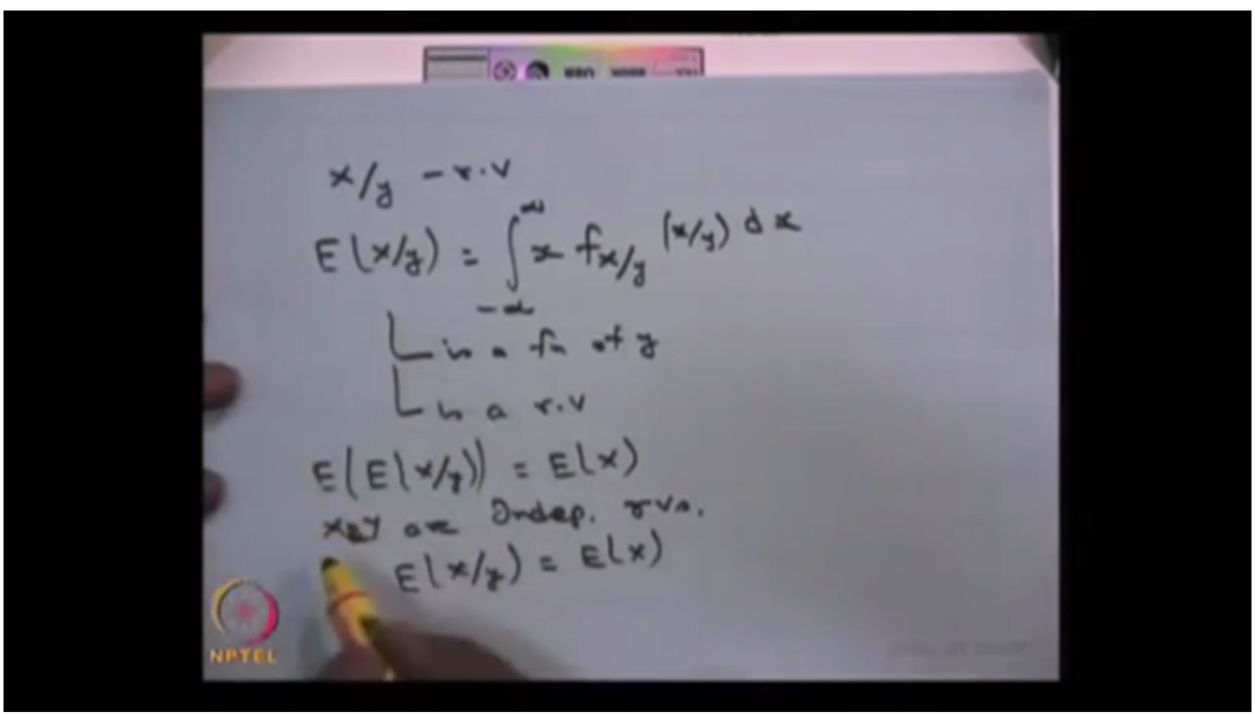
So one more thing that is a conditional expectation. So since I said X given Y is random variable I can go for finding out what is the expectation of X given Y. So this is called the conditional expectation, that means the X given Y is the still it is a random variable but it is a conditional distribution. Therefore finding out the expectation for that that is called the conditional expectation.

Suppose I treat both the random variables are continuous case then the conditional expectation is nothing but minus infinity to infinity X times f(X) given Y of X, X given Y integration with respect to X. That means by treating X and y are continuous random variable I can able to define the conditional expectation is this provided this expectation exist. That means in absolute sense if this integration converges then without absolute whatever the value you are going to get that is going to be the conditional expectation of the random variable. And if you note that since the Y also can take any value therefore this is a function of Y; not only this is a function of Y the conditional expectation is a random variable also. That means X given Y is a random variable. The expectation of X given Y is a function of Y and Y is a random variable it takes a different value small y therefore expectation of X given Y is also a random variable. That means you can able to find out what is the expectation of expectation X given Y.

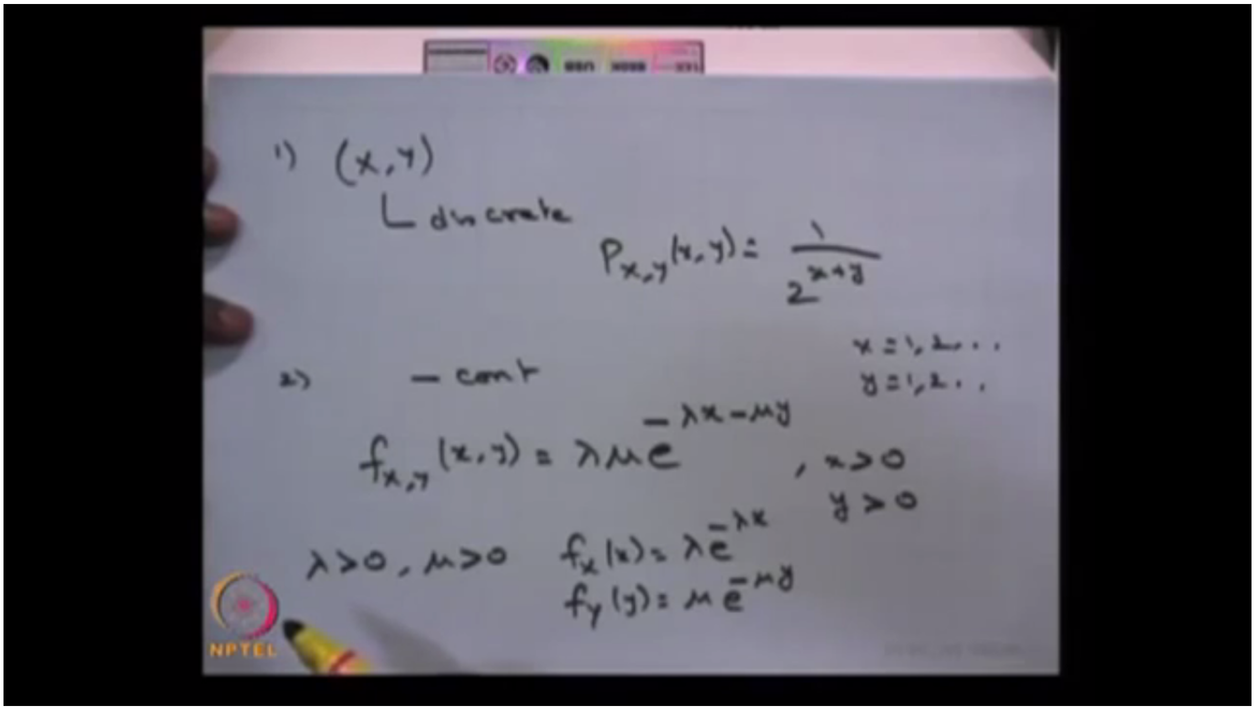
if you compute that it is going to be expectation of X this is a very important property in which you are relating two different random variable with the conditional sense and if you are trying to find out the expectation of that that is going to be the original expectation. That means the usage of this concept instead of finding out the expectation of one time random variable if it is easy to find out the conditional expectation then you find out the expectation of conditional expectation that is same as the original expectation. Suppose you have two random variables or independent random variables then that there is no dependency over the random variable X and y therefore the expectation of X given Y that is same as the expectation of X.



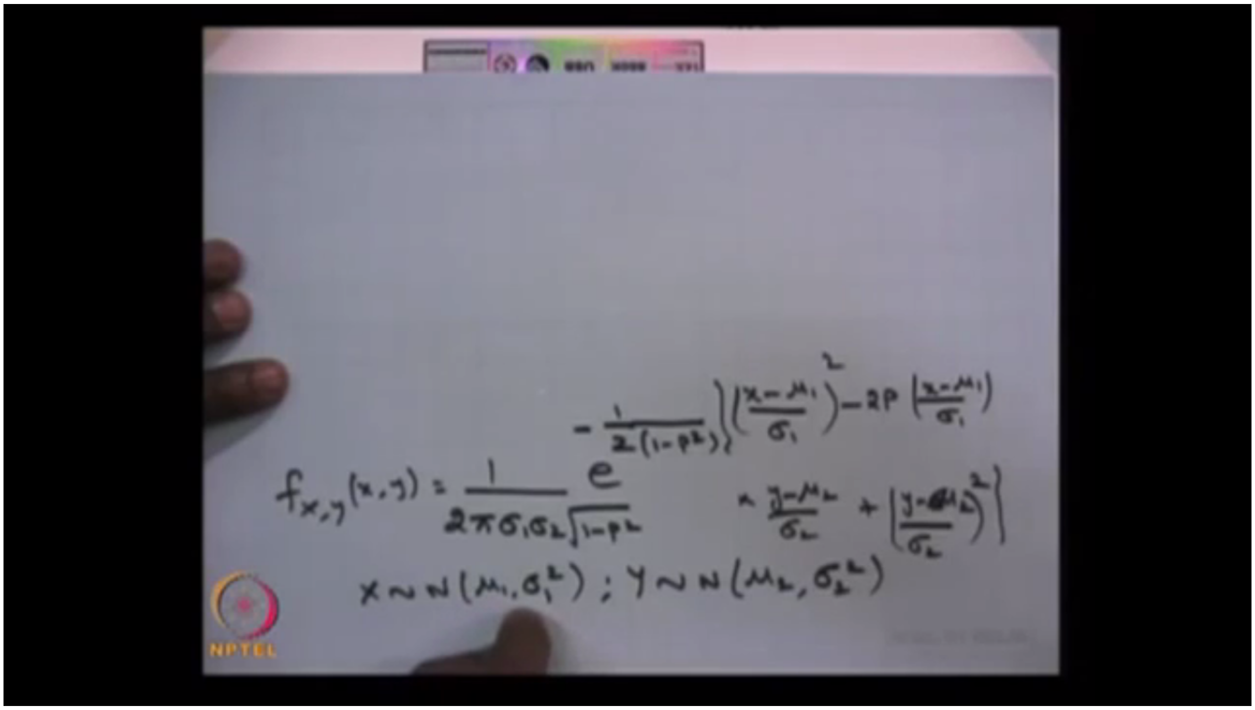
So this can be validated here also because this expectation of X given Y is going to be expectation of X. X and the expectation of X is a constant and the expectation of our constant is a constant that is same as the constant. So that can be cross-checked. So here I have given expectation of X given Y in the integration form. If both the random variables are continuous then accordingly you have to use initially the joint probability mass function then conditional probability mass function to get that conditional expectation. And this conditional expectation is a very much important to give one important property called the martingale property in the stochastic process in which you are going to discuss not only two random variables, you are going to discuss you have a n random variables and you can try to find out what is the conditional expectation of one random variable given that the other random variable takes some value already.

So there we are going to find out what is the conditional expectation of n dimensional random variable with the given that remaining n minus 1 random variable takes already some value. So here I have given only with the two random variables how to compute the conditional expectation but as such a you are going to find out the conditional expectation of n random variables with the n minus 1 random variables already taken some value. So before I go to the another concept let me just give a few examples in which I have already given if both the random variables are of a discrete type I have given an example of a joint probability mass function has 1 divided by 2 power X plus y and the X takes a value 1, 2 and so on and Y takes the value 1, 2 so this is a joint probability mass function example. And suppose you have a random variables are of the continuous type then I can give one simple example of the joint probability density function of a two dimensional continuous type random variable has a joint probability density function lambda times mu E power minus lambda X minus mu Y where X can take the value greater than zero Y can take the value greater than zero and lambda is strictly greater than zero as well as mu greater than zero. So this is going to be the joint probability density function of two dimensional continuous type random variable. You can cross-check this is going to be joined because it is going to be always take great not equal to zero values for all X and y and if you make a double integration over minus infinity to infinity over X and y then that is going to be Y. And you can verify the other one if you find out the marginal distribution of this random variable you may land up the marginal distribution of this random variable is going to be lambda times E power minus lambda X and similarly if you find out the marginal distribution of the same one you will get mu times minus mu Y and if you cross check the product is going to be the joint probability density function then you can conclude this both the variables are independent random variable.

Similarly you can find out what is the marginal distribution of the random variable X. Similarly marginal distribution of Y. If you cross check the similar independent so property of independent then that is satisfied. Therefore, you can conclude here the random variables X and y both are discrete as well as both are independent random variable also. So the advantage with the independent random variable always you can find out from the joint you can find out the marginals but if you have a marginal you cannot find out the joint unless otherwise they are the independent random variable. Therefore, the independent random variable makes easier to find out the joint distribution with the provided marginal distribution.



And here is the one simple example of – here is a simple example of bivariate normal distribution in which the both are under variables X and y are normally distributed therefore the together joint distribution is going to be of the form. Let me write the joint probability density function of two dimensional normal distribution random variable as 1 divided by 2 pi sigma 1 sigma 2 multiplied by square root of 1 minus rho square into e power minus half times of 1 minus rho square multiplied by X minus mu 1 by sigma 1 whole square minus 2 times Rho minus 2 times Rho into X minus mu 1 by Sigma 1 that is multiplied by y minus mu 2 by Sigma 2 plus y minus mu2 by sigma2 whole square. So here if you find out the marginal distribution of the random variable X and the marginal distribution of Y you can conclude X is going to be normally distributed with the mean mu 1 and the variance Sigma 1 square and similarly you can come to the conclusion y is also normally distributed with the mean mu 2 and the variance Sigma 2 square. That means if you make the plot for the joint probability density function that will be of this shape one is the X and one is the y and this is going to be the joint probability density function for fixed values of mu 1 and mu 2 and Sigma 1 and the Sigma 2 and this is going to be the joint probability density function and here Rho is nothing but the correlation coefficient. That means what is the way the random variable X and y are correlated that comes into the picture when you are giving a joint probability density function of this random variable and they are not independent random variable unless otherwise the Rho is going to be 0. So if the Rho is going to be 0 then it gets simplified and you can able to verify the joint probability density function will be the product of probability density function and each one is going to be a probability density function of a normal distribution with the mean mu 1 and the variance Sigma 1 square and mu 2 and the Sigma 2 square.



So this bivariate normal distribution is very important one, when you discuss the multinomial normal distribution. So only we can able to give the joint probability density function of the bivariate so the multivariate you can able to visualize how the joint probability density function will look like and what is the way the other factors will come into the picture.

So other than correlation and the coefficient – other than covariance correlation and correlation coefficient we need the other called covariance matrix also because in the stochastic process we are going to consider n dimensional random variable as well the sequence of random variables. So you should know how to define the covariance matrix of n dimensional random variable. That means if suppose you have a n random variables X1 to Xn then you can define the covariance matrix as you just make a row wise X1 to Xn and the column also you make X1 to Xn. Now we can fill up this is going to be a n cross n matrix in which each entity is going to be covariance of so that means the matrix entity of i, j is nothing but what is the covariance of that random variable Xi with the Xj. You know that the way I have given the definition covariance of Xi and Xj if I and J are same then that is nothing but e of X square minus e of X whole square therefore that is nothing but the variance of that random variable. Therefore, this is going to be variance of X1 and this is going to be the variance of X2. Therefore, all the diagonal elements are going to be variance of Xi's whereas other than the diagonal elements we can fill it up this is going to be a covariance of X1 with X2 and the last like that the last element will be covariance of X1 with the Xn.

Similarly second row first column will be covariance of X2 with X1. And you can use the other property the covariance of XI, XJ same as covariance of XJ with the XI also because you are trying to find out expectation of X into y minus expectation of X into expectation of Y. Therefore, both the covariance of X2 with X1 is same as X1 with X2 so it is going to be whatever the value you are going to get it is going to be the symmetric matrix and all the diagonal elements are going to be the variance.

So the way I have given the two dimensional normal distribution that is a bivariate normal suppose you have a n dimensional random vector in which each random variable is a norm distribution then you need what is the covariance matrix for that then only you can find out what is it – then only you can able to write what is a joint probability density function of n dimensional random variable.