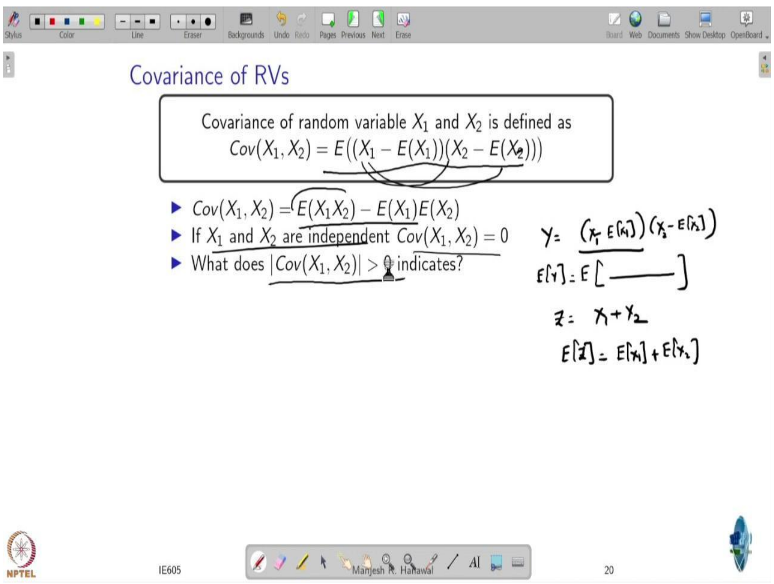
Engineering Statistics
Professor Manjesh Hanawal
Department of Industrial Engineering and Operations Research
Indian Institute of Technology Bombay
Week 3
Lecture 13
Covariance of Random variables

Okay, let us get started again. Covariance. We talked about expectation. We talked about variance, Expectation of a random variable, we talked about variance of a random variable we talked about. And that time we were all dealing with 1 random variable, but now we have extended our scope. Now, we are talking about multiple random variables. And then we talked about joint cdf, we talked about joint pdf.

When we are going to talk about multiple things, we said that they could be dependent. We talked about the independent if we are talking about independent, there has to be some dependency or notion of dependency also. Now how to capture this notion of dependency. 1 notion to capture that, to what extent some things are dependent. If something is not dependent, yes, there is a dependency, we need to now characterize how much they are dependent.

(Refer Slide Time: 1:32)



For that, we use a notion of covariance and covariance for simplicity, we will define it between 2 random variables. And it is defined like this by definition, covariance between 2 random variables is notice that 2 things here, I am defining basically a new random variable, which is a product of 2 random variables. This is 1 random variable. And this is another random variable, this random variable I obtained by subtracting mean from the X1 and the second one obtained from X2 after subtracting its mean value.

And then I am taking the product and this is I am calling a random variable Y, and now the covariance is between X and Y actually the expectation of this which is expectation of this quantity, which I have written here. So, what it is basically doing is it is taking the product of these 2 random variables after centralizing them. What I mean by centralizing is that I am subtracting the mean from the random variables.

Whenever we subtract the mean from a random variable, they call it centralizing them. So, this is basically we have centralized the random variables taking their product and looking at their expectation. And we will argue now that this in a way captures how much they are dependent on each other if at all they are dependent. One obvious thing that will come from simple algebraic manipulation is if you just expand this product and apply the definition of expectation, you will get this.

So, how, by the way, how many of you know that expectation operator is linear operator? In IE621 was this discussed? So, what do you mean by let us say, I have 2 random variables, and I am taking their product X1 plus X2 and if I take expectation of Z what is this value is?

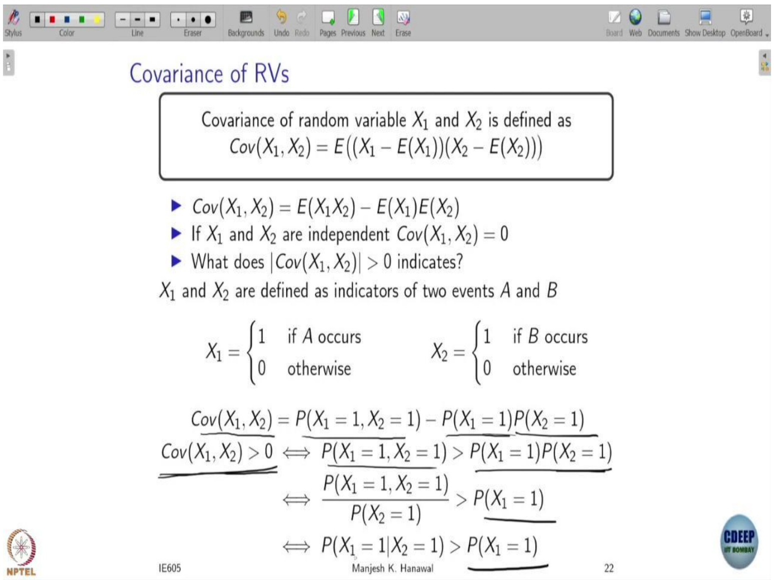
Students: Expectation of X1 plus Expectation of X2.

Professor: Does this required X1 and X2 be to be independent? No, it does not matter what the distribution is that is why this is called expectation is a linear operator.

So, you need to use that linearity property of expectation and you take their product here inside multiply X1 with X2, X1 with this like that, you get 4 terms and after simplify, you will get this term. Now, one immediate observation is suppose if X1 and X2 are independent, what is going to happen to this term? This term is nothing but expectation of X1 into expectation of X2.

So, if X1 and X2 are independent, then you see that right away the covariance of X1 and X2 is going to be 0. So, if 2 random variables are independent, we are saying their value covariance is going to be 0, then what is the meaning that if their value is not 0 then it cannot be independent. There is some relation between them.

(Refer Slide Time: 5:26)



So, let us try to understand this. What does this mean? let us take 2 events A and B on some sample space. Now I am going to define 2 random variables. Random variable one, I am going to say assign to value 1 if that event A occurs otherwise I will take its value a 0. Whenever event A occurs, I will take random variable X1 to be 1 otherwise I will take 0.

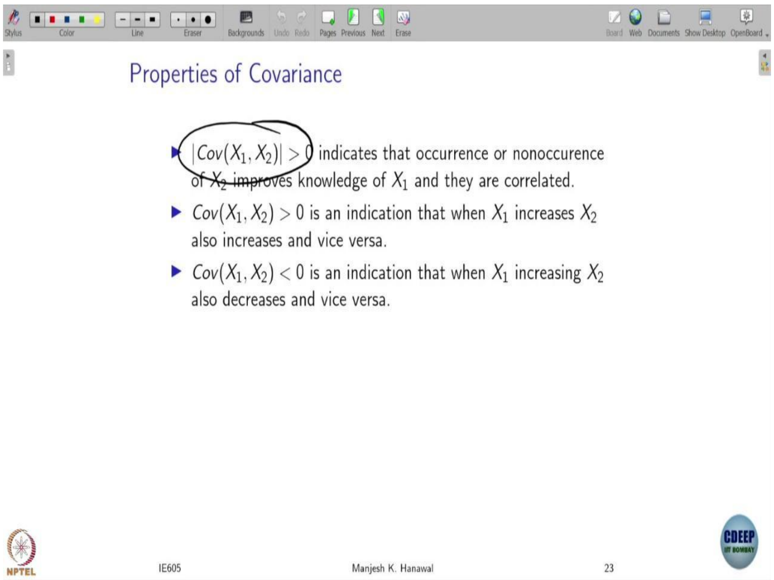
Similarly, for X2, X2 is based on event B. If B2 occurs, I will take X1, X2 to B1 otherwise I will take it to be 0. Now let us try to understand what does covariance of this means now, you can quick calculations if you apply this definition of expectations the covariance between X1 and X2 you can verify that this will simplify to probability that X1 is 1, X2 is 1 minus probability that X1 equals 1 times probability that X2 equal to 1, this will it will work out like this.

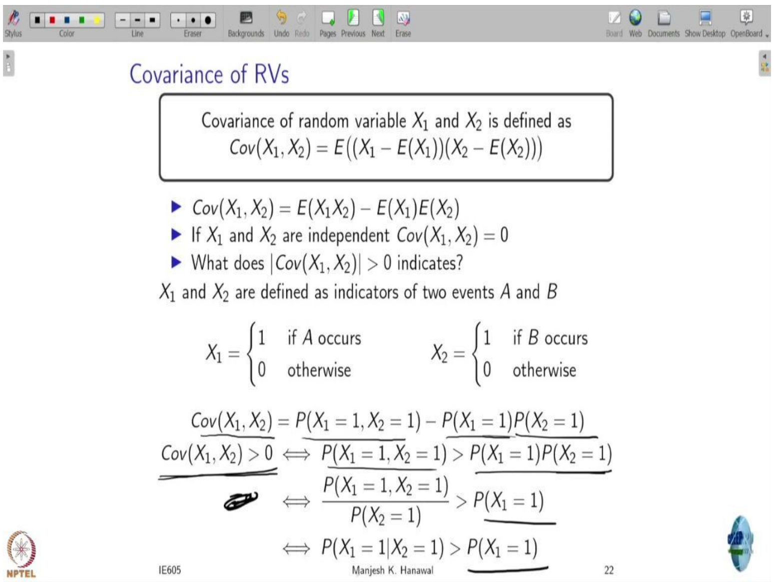
Now, assume that X1 and X2 the covariance is positive, you assume this? What does this indicate? Let us see, what does this indicate? If this has to be positive, then the right side has to also be positive. This is going to imply that this term has to be larger than this term. Everybody agrees. Now, I will do further simplification. This X2 equals to 1. I will bring in the denominator that this ratio has to be larger than this quantity.

Now that by the definition of our conditional probability, what it is saying is this probability that X1 equals to 1 given that X2 equals 1. Everybody agree with this definition of conditional probability? Now, this conditional probability is greater than probability that X1 equals to 1. What it is saying is, this is a if and only if conditions all of them if the covariance between X1 and X2 is positive, that means that if X2 equals to 1 has happened, then the conditional probability that X1 equals to 1 is higher than the unconditional probability of X1 taking value 1.

That means, if X2 is going to have happened, 1, then X1 is also more likely to take the value of 1. That is what this is saying. That means, it is kind of already telling that somehow if I know something about X2, I am able to infer something about X1 not only that, but in what directions they are moving in a way that if X2 is 1, it is likely that X1 is also going to be the same, there are only 2 values set 0, 1. It is saying that if X1 is 1, X1 is also is going to 1 it is not going towards 0 it is going towards 1.

(Refer Slide Time: 8:50)



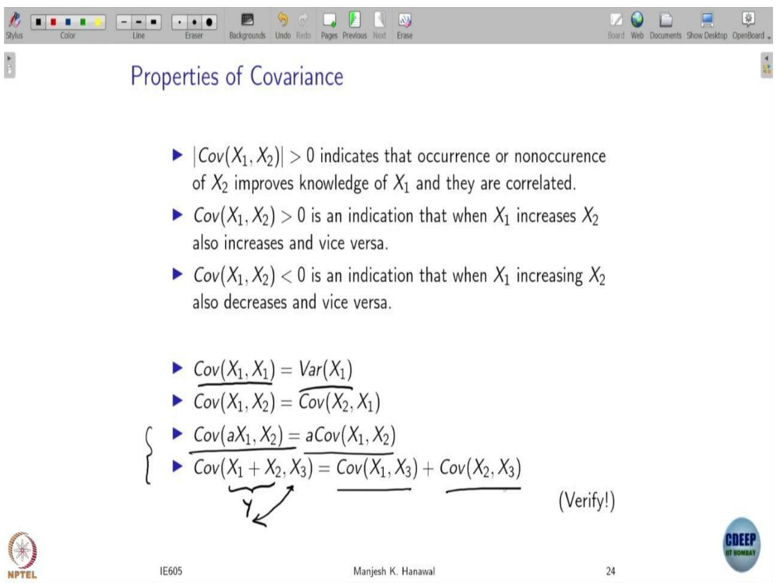


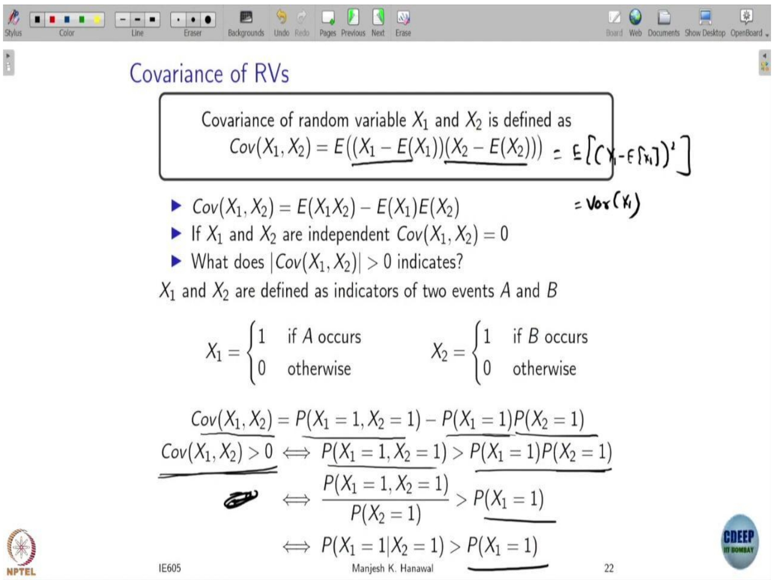
So, what it is going to indicate is if this absolute value of the covariance, if it is positive, it indicates that positive is strictly positive, it indicates that occurrence or non-occurrence of 1 random variable provide some knowledge about the other. So, there is some dependency. If this value has been 0, then we know that by definition, this unconditional probabilities equals to conditional probability.

If they are independent. That is when this is equals to 0, but now that we are assuming this is greater than 0, that means X2 equals to 1 implying that X1 equals to 1 has a higher probability now. Now, in general, the covariance can take positive or negative value, positive or negative value. And whenever it is a positive value, it is an indication that when 1 random variable is increasing, other is also increasing on an average sense.

All we are talking in average sense nothing is in terms of 1 realization, it is average across all realization. And similarly, whenever this covariance is going to be less than 0, this is an indication that when X1 increases X2 is decreasing, let they are not in the same direction. And again this is an average sense nothing like absolute realizations.

(Refer Slide Time: 10:33)





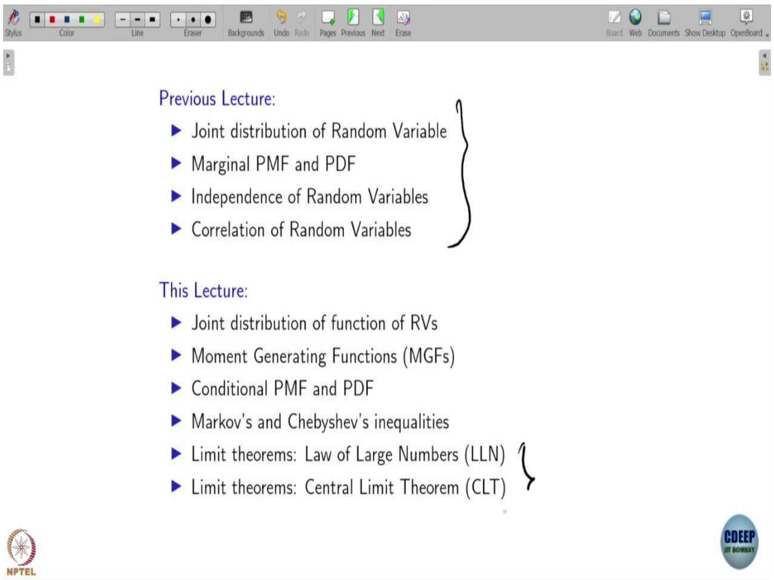
Now, this definition of the covariance has straightforward properties. Now, you can always ask for a covariance of a random variable with itself the covariance of a random variable X1 with itself. And by definition, this is exactly variance of X1. So, here, if you make it X2 is also X1, this is going to be expectation of X1 minus expectation of X1, whole square. And this is exactly the definition of variance of X1.

The second one covariance is not dependent on the ordering of these random variables, whenever you are talking about 2 random variables, either you take covariance of X1 and X2 or covariance of X2 and X1 they are the same. Because you multiply whatever like you multiply this term first, this term first and then this, this one first, and that one does not matter.

And now, if you scale 1 of the random variables, let us say X1 by factor A, then whole of this covariance gets scaled by a factor of A. Again, this I would like you to verify this last 2 properties. And if I have 1 random variable, let us say I have now 3 random variables, and 1 random variable Y, which is nothing but the sum of 2 random variable X1 and X2.

And now I want to look at the covariance between Y and X3 this can be written as covariance of X1 and X3 and covariance of X2 and X3. Again, this follows from properties of your covariance, basically, by their definition of covariance, please check all these things.

(Refer Slide Time: 12:56)

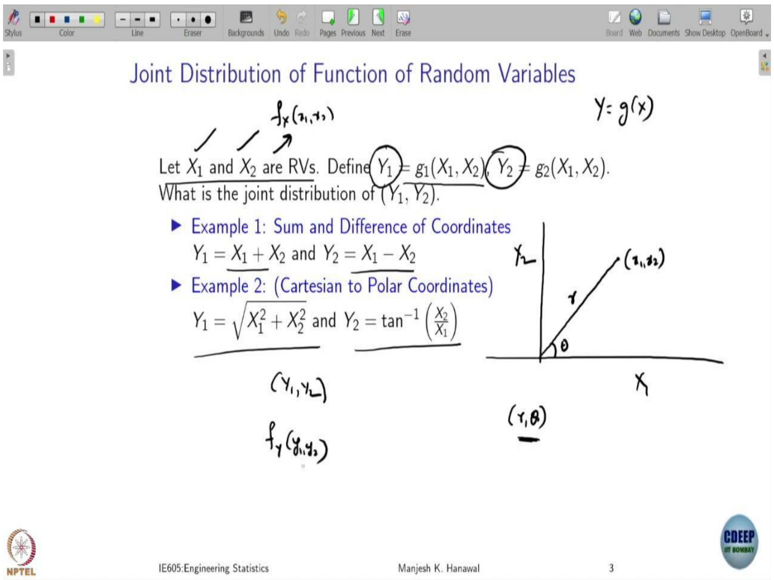


Now, I am going to switch to next slides. So meanwhile, if you have any questions about this covariance, independence of the random variable you should ask now, let us continue. So, is the moment generating functions are covered in IE621?

Student: No, No

Professor: Let us go through it, so far, we covered all the stuffs. Independence of random variable and correlation of random variable. So, what we will do is now we will quickly start migrating to the statistics. And these 2 theorems, the limit theorems law of large numbers and central theorem limits. These are our bridging bridges between probability and statistics, so once we conclude these 2 theorems, we are ready to get into statistics.

(Refer Slide Time: 13:41)



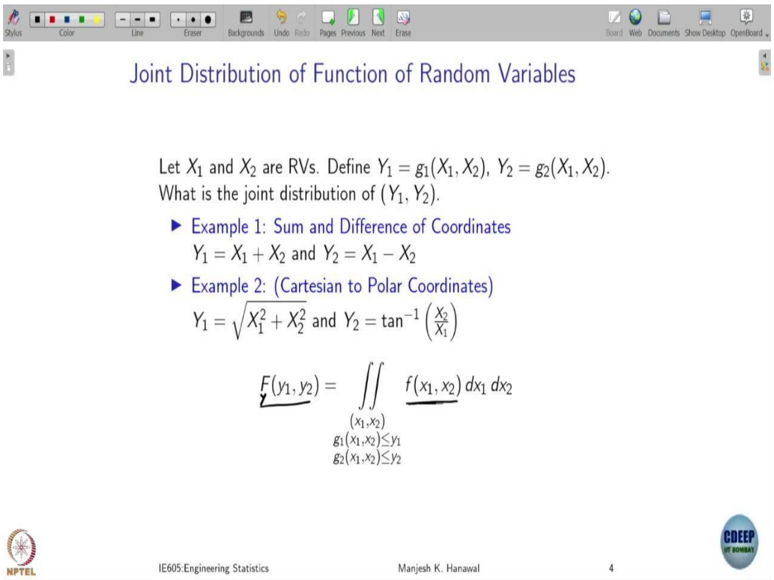
But before that, let us understand some more concepts of probability. Let us say you have 2 random variables. And now, you are trying to get new random variables out of this and you may have multiple of them. So, one of them I get is by applying some function g1 and I got a new random variable Y1 which depends on both X1 and X2 and I have another random variable Y2 which is again applied on obtained by X1 and X2 after applying this function g2.

So, now we are expanding right earlier, I had just 1 random variable and I did this like one function of one random variable, now I have multiple random variables and I get I am applying function which is applied on both these random variables. Now the question is how to derive joint distributions of these 2 random variables Y1 and Y2 where you may end up with such situation.

Let us say have 2 random variable X1 and X2. You are interested in their sum and their difference. Now Y1 is the sum and Y2 is the difference. And now, you want to understand what is the joint distribution of Y1 and Y2. And another thing you have let us say Cartesian to polar coordinates, let us say you have this X1 and another is X2, you maybe take some point here X1 and X2 one thing is you will be interested in 1 is your magnitude and so, you have this Cartesian form r and θ.

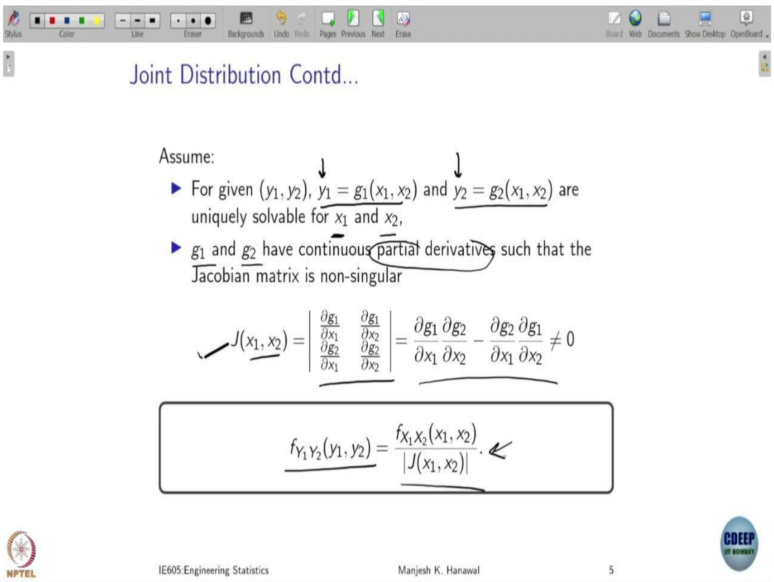
So, you are r the radius or the distance is given based on X1 and X2 and your angle is also depends on X1 and X2. Now, you have these 2 random variables you have based on that you have new random variables. Now, you may want to know how this radius and angles jointly behave. So, that is you are interested in finding that joint distribution of Y1 and Y2 how to do this and what is your given you know, well X1 and X2 and you know their joint distribution let us say their joint distribution X1 and X2 is available to you. But, you from that you need to find out the joint distribution of Y1 and Y2 how to do this.

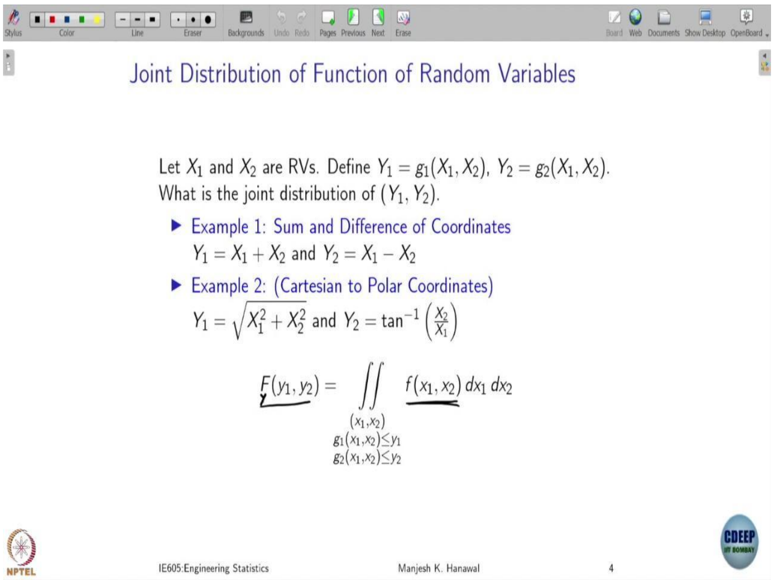
(Refer Slide Time: 16:38)



One obvious way is, if I want to find out cdf of my random variable, let us call this y at point y1 and y2 is simply integrate it over all X1 and X2 such that g1 of X1 and X2 is less than equal to y1 and g2 of X1 and X2 equals to y2, you are basically interested in that region and then this is given to you that the joint PDF of X1 and X2 is given to you. So, integrate it over you are region of interest.

(Refer Slide Time: 17:20)





But this is often tedious, so, what we will look into is some simple methods to compute this. So, for that we will assume that this we have 2 equations this is one equation this is another equation for given X1 and X2 value you get y1 and for the same X1 and X2 depending on your function g2 you will get y2 and let us say that these are such that you are going to get uniquely for a given value of y1 and y2 you can get what is the corresponding x1 and x2 that results in this value uniquely.

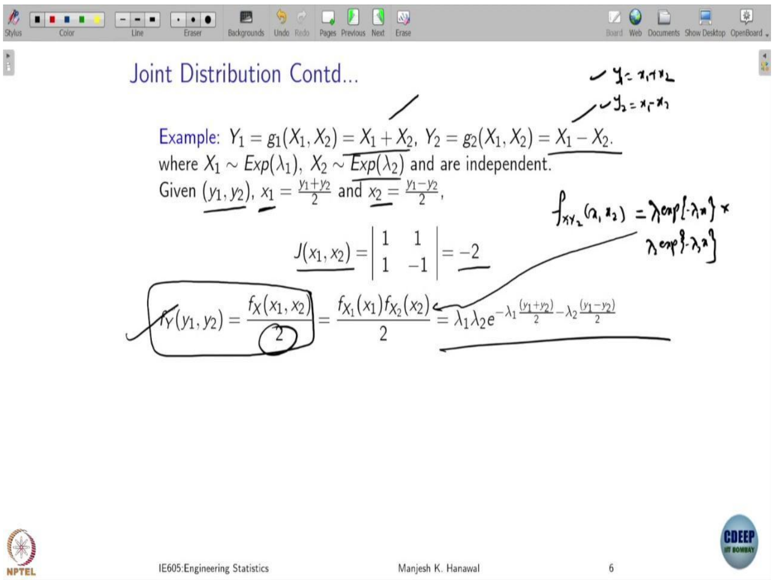
Now, also assume that you are functions g1 and g2 how continuous partial derivatives I am talking partial derivatives here, because g function is functions of multiple variables here. Now, we are going to define something called a Jacobin matrix at the point x1 and x2, which is defined like this actually, I am taking the determinant of this Jacobian matrix here and which is computed like this.

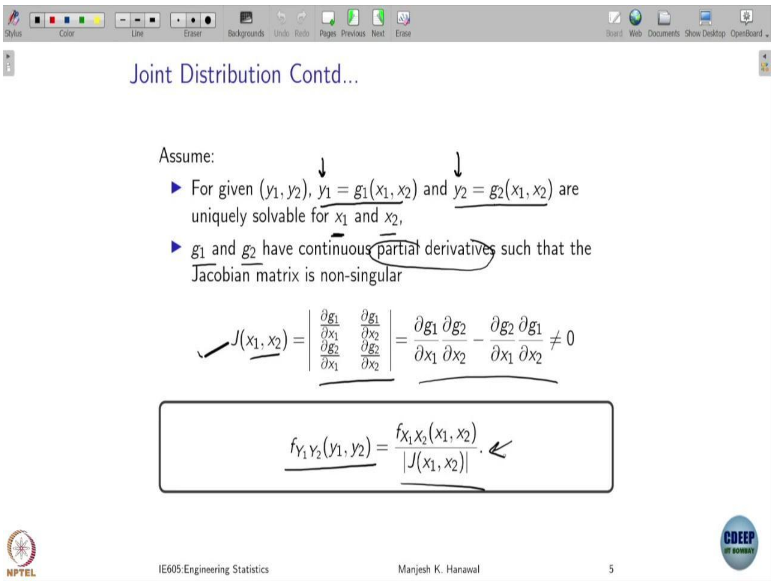
So, first row corresponding to the function g1 and the second row corresponding to the function g2. Assume that this Jacobian function is not 0 at any point x1 and x2. Now, it so happens that the joint PDF of your new random variable at point y1 and y2 you can obtain simply by dividing joint PDF of X1 and X2 by this determinant of your Jacobin metrics.

You do not need to go and do complex math here like this complex integration here to get this all you need to do is like first of all this is this was only giving you the CDF if you need to get a PDF, you need to differentiate it before you get a PDF. But here we are expressing these things directly in terms of the PDF. How it comes will not get into the proof this is just for our use.

And we are going to use it later actually. I want you to be aware of this, this relation, we are simply saying that the method is, if I have given 2 random variables and 2 functions and I have Y1 and Y2 as new random variable to compute the joint distribution and Y1 and Y2, I need to first find out the Jacobian matrix and find out that X1 and X2, which uniquely solve for that given y1 and y2 and for a given f function, I use this relation.

(Refer Slide Time: 20:33)





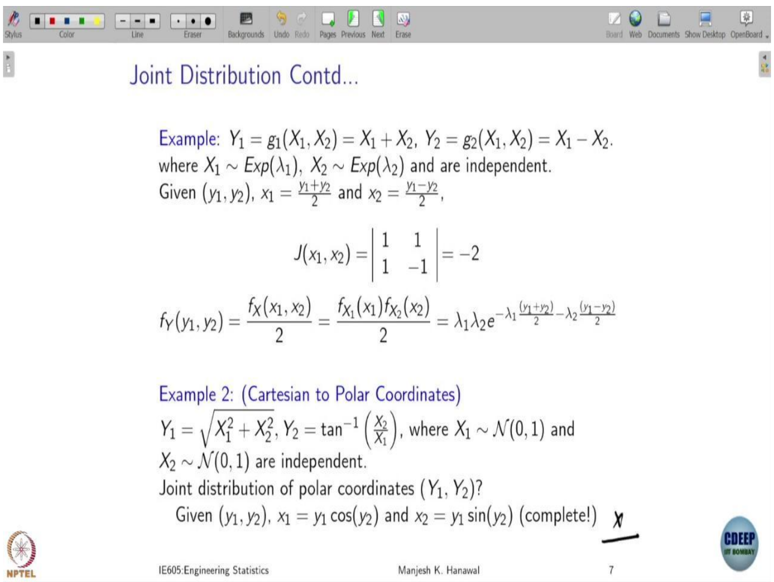
An example is here. Like the example I said, let us take g1 function to be a sum of random variables and g2 be difference of these 2 random variables and assume that X1 here is exponential with parameter lambda 1 and X2 is exponential with parameter lambda 2 and also assume that they are independent. So, then what is the value of this the joint distribution of x1 and x2 it is going to be a product of 2 exponential one is $\lambda_1 x$ and other is into $\lambda_2 x$.

Now another thing we said, our method applies if we are going to for a given x1 and x2, I will uniquely obtain x1 and x2 which solves that equations. So, in this case, if you are given y1 and y2, you will see that x1 corresponds to y1 plus y2 by 2 and x2 is y1 minus y2 by 2, all you need to do is solve these equations. So what I have is y1 equals to x1 plus x2 and Y2 is X1 minus X2.

For a given y1 and y2, I compute and find out what is x1 and x2 and x1 happens to be this and x2 happens to be this. Now, if you go and compute the determinant of the Jacobian matrix, this is exactly this value. Now, I am going to simply apply the formula we discussed, that denominator is -2. But why did I take 2 here, notice that this is absolute value of the determinant value in the denominator.

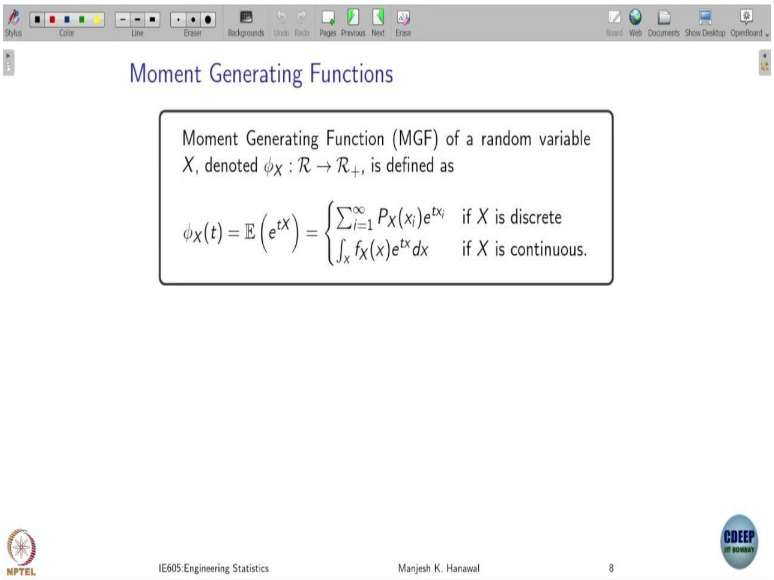
So, even though I got the determinant as -2 here, I have taken this absolute value, that is why it is 2 here. And the numerator, the joint pdf, it has split into the marginal because they are independent. And now all I have done is use this formula here and replaced X1 by y1 plus y2 by 2 and X2 by y1 minus y2 by 2 now this is entirely in terms of my Y.

(Refer Slide Time: 23:26)



So, there is another example here. This is based on the Cartesian things. And 1 more thing for tomorrow's quizzes, this question may or may not come tomorrow's quiz, you better workout it may or may not come if it comes, you have worked out you already have the solution.

(Refer Slide Time: 23:46)



So, this moment generating function maybe we will take it up in the next class. This is a new topic and we will stop here.