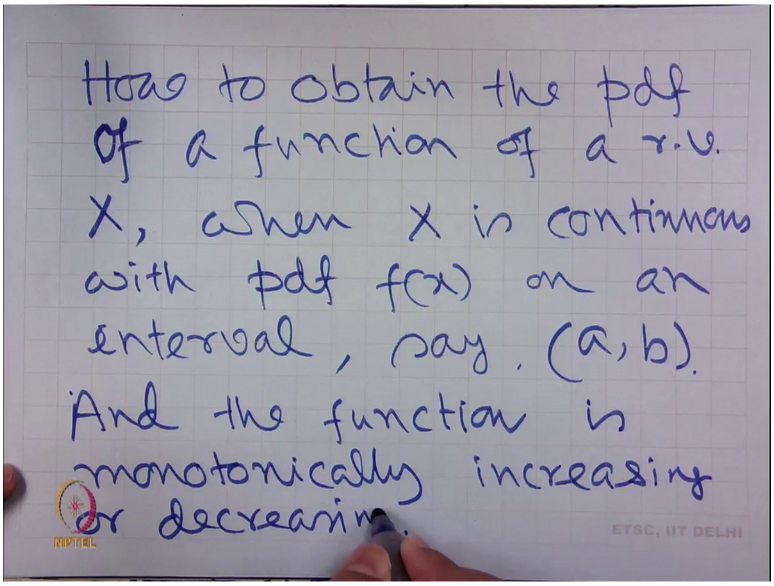
Statistical Inference
Prof. Niladri Chatterjee
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 05
Statistical Inference

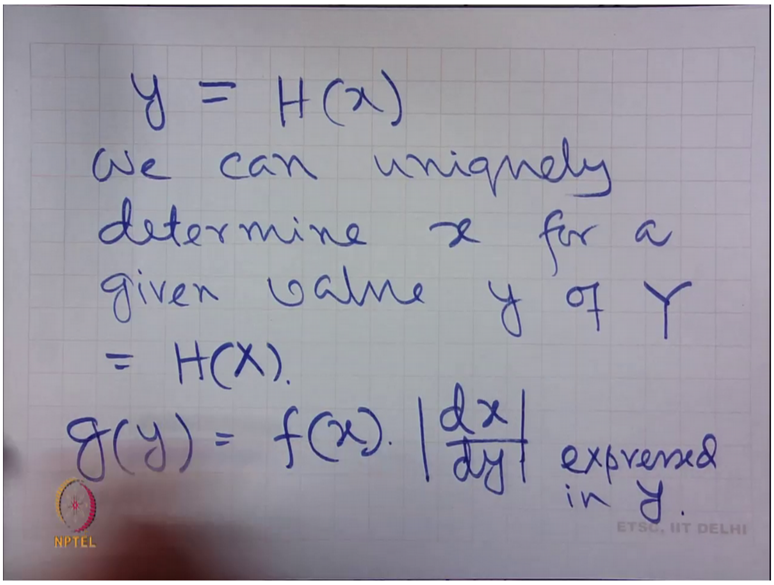
Welcome friends to my MOOC's series of lectures on Statistical Inference. This is lecture number 5. If you remember in the last lecture I finished with how to obtain the pdf of a function of a random variable X; when X is continuous with pdf fx on an interval say a to b.

(Refer Slide Time: 00:41)



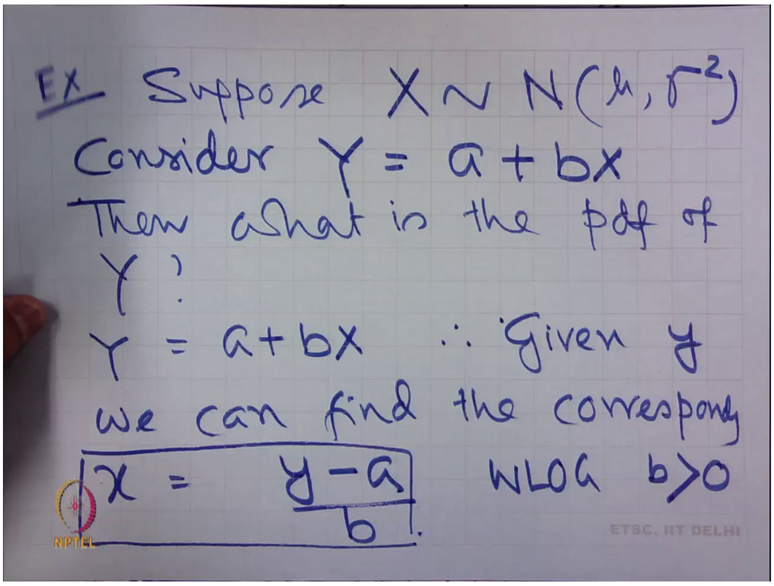
And the function is monotonically increasing or decreasing, as I discussed that the monotonicity is important, because then for any given y is equal to say H of x we can uniquely determine x for given y.

(Refer Slide Time: 02:14)



And the result was that pdf of y, if we call it gy then what we obtained is that g y is equal to f at x into modulus of dx dy expressed in y. So, with that result let us now look at some problems.

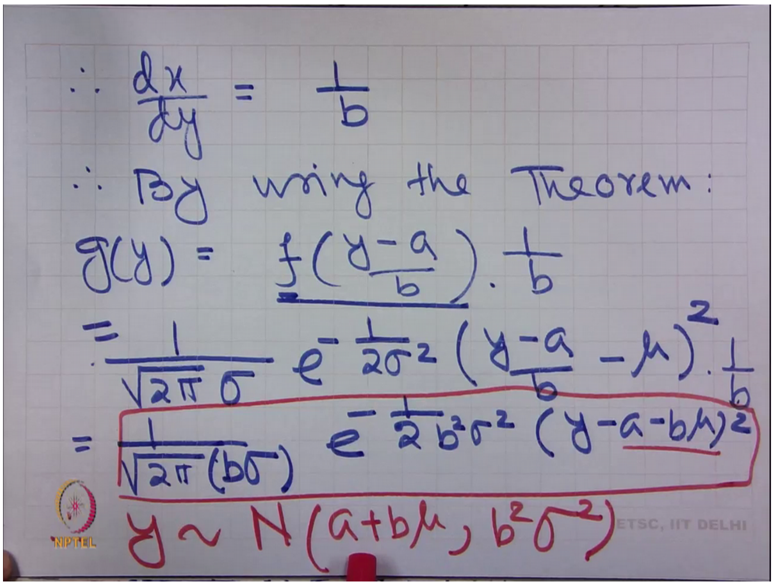
(Refer Slide Time: 03:31)



Suppose X is a normal variate with mu and sigma square, X is a normal variable with expected value is equal to mu and variance of X is equal to sigma square. Consider Y is equal to a plus b X, then what is the pdf of Y?

So, we proceeded in the following way Y is equal to a plus bX therefore, given value of y we can find the corresponding x is equal to y minus a divided by b, without loss of generality let b be greater than 0.

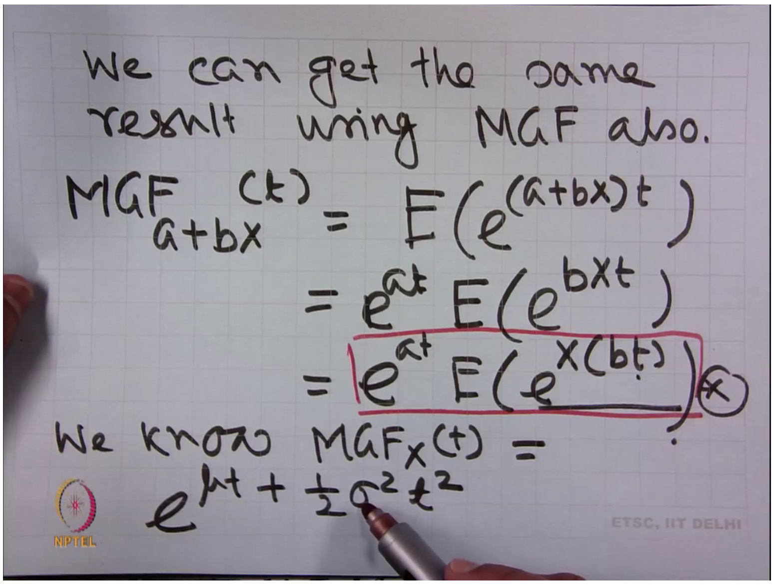
(Refer Slide Time: 05:17)



Since x is equal to y minus a by b therefore dx of dy is equal to 1 by b; therefore, by using the theorem g of y is equal to f at y minus a by b multiplied by 1 by b, it is the dx dy, and the modulus sign is not needed if because we are using b to be positive. Now, we know that f is the density of a standard density of a normal variable with mean mu variance sigma square. Therefore, this term is 1 over root over 2 pi sigma into e to the power 1 upon 2 sigma square into y minus a upon b minus mu whole square, multiplied by 1 by b. This is equal to 1 over root over 2 pi into b sigma this becomes here e to the power minus 1 upon 2 b square sigma square into y minus a minus b mu whole square.

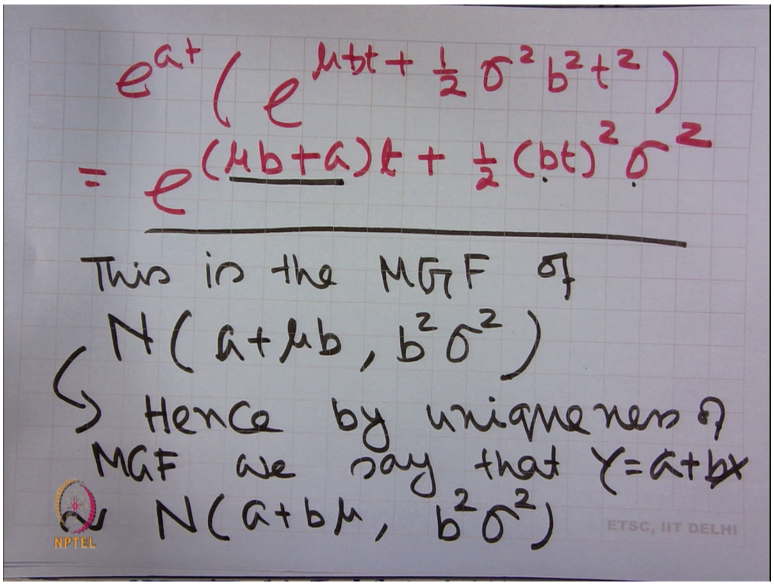
So, we obtain the pdf of y is this. What can we say from here? We can see that therefore, y is normal with mean is equal to a plus b mu and variance is equal to b square sigma square. Therefore, we find that if we make a linear transformation on a normal variable x, then the resulting variable is also a normal distribution with appropriate adjustment in the mean and in the variance.

(Refer Slide Time: 08:44)



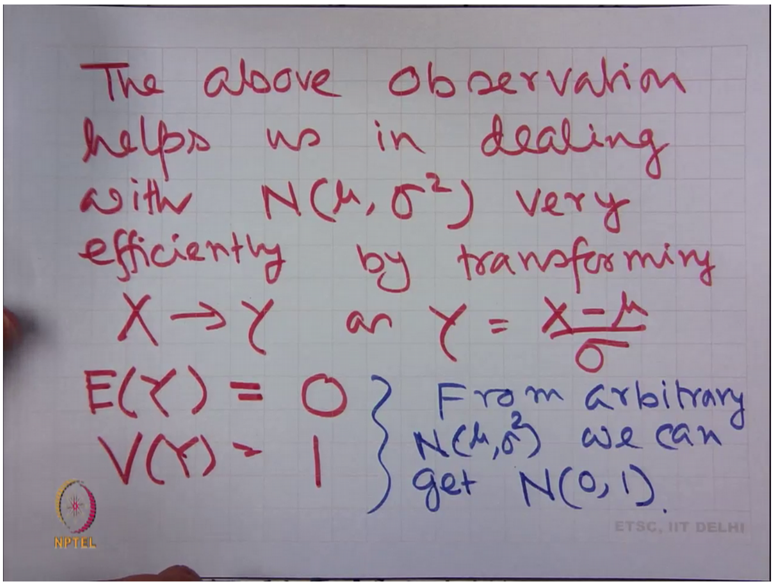
We can get the same result using moment generating function also. What is the moment generating function of a plus bX? Is equal to expected value of e to the power a plus bX into t; This is equal to e to the power a t multiplied by expected value of e to the power b X t; is equal to e to the power a t into expected value of e to the power x into bt. Now, we know that the MGF of e to the power mu t plus half sigma square t square. If we look at the term, we can find that it is basically the same term with t replaced by bt, also we can see that there is a multiplier e to the power t.

(Refer Slide Time: 10:34)



Therefore, this term we can write by comparing with this as e to the power a t into multiplied by e to the power mu t plus mu bt plus half sigma square b square t square; is equal to e to the power mu b plus a t plus half bt square into sigma square. Now, we know that this is the MGF of normal variable with mean is equal to a plus mu b, and variance is equal to b square sigma square. Hence, by uniqueness of moment generating function we can say that y is equal to a plus bx is distributed as normal with mean a plus b mu and variance b square sigma square. So, the same result we can get using moment generating function, but the above theorem helps us to get it in a very simple way.

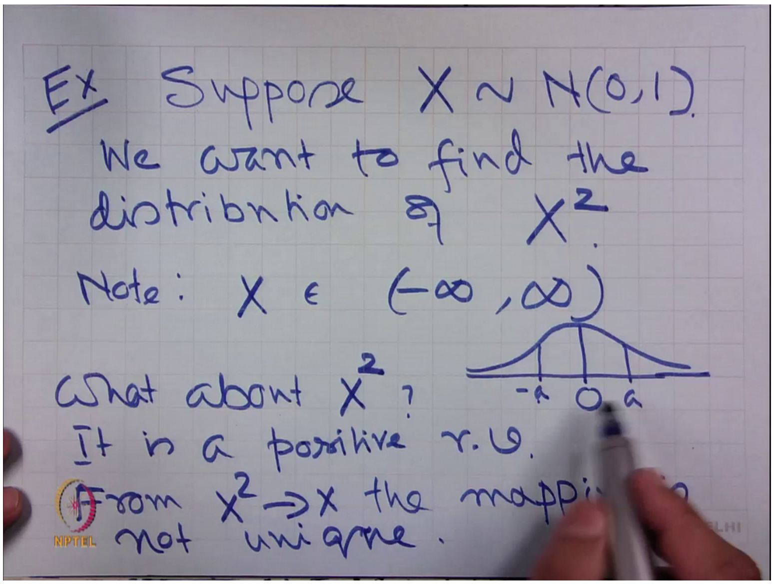
(Refer Slide Time: 12:56)



The above observation helps us in dealing with normal mu sigma square very efficiently by transforming x to y as y is equal to x minus mu by sigma. What is the advantage? The advantage is expectation of Y is equal to 0, and variance of Y is equal to 1. Therefore, from arbitrary normal variate with mu and sigma square, we can get standard normal distribution by doing this linear transformation.

And why we use normal 0 1? Because that makes our life simple; Even if you look at it from moment generating function for standard normal the moment generating function is e to the power t square by 2, but for arbitrary mu and sigma the moment generating function becomes e to the power mu t plus half sigma square t square. And therefore, dealing with that mathematically becomes more complicated. Let us now look at slightly more complicated problem.

(Refer Slide Time: 15:15)

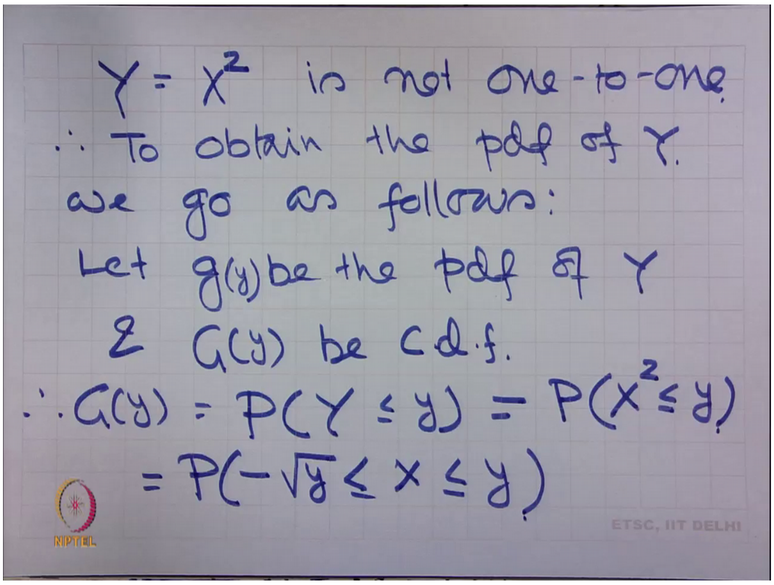


Suppose X is a variate with normal 0 1. We want to know the distribution of x square. Note X belongs to minus infinity to plus infinity. Because it is a standard normal variable and it is minus infinity to plus infinity and all of us know that it is symmetric around 0. What about X square? X square as you can see is a positive random variable.

And another problem with respect to this transformation is that this mapping is from x squared to x is not unique. Because 4 minus a and a for both of them, since for minus a and a for both of them the value of x square is equal to s square. Therefore, from x square when I go back to x this mapping is not unique, as I have explained in my previous lecture with respect to a discrete random variable if you remember I have taken minus 2 minus 1 0 1 2, and from there I explained that the inverse mapping is not there.

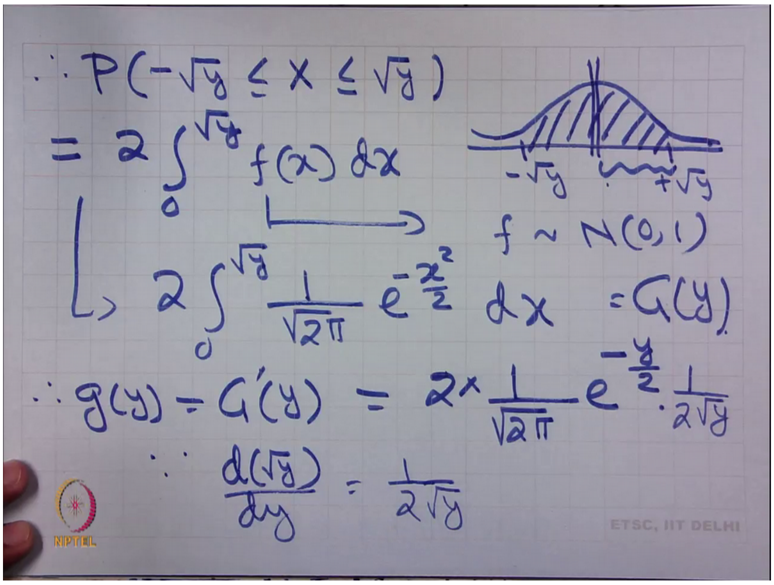
Or in other words, we can see that in this case y is equal to x square, the function is not monotonically increasing. And therefore, this theorem does not hold as such. So, what we do? We make a small adjustment.

(Refer Slide Time: 18:18)



Y is equal to X square is not one to one therefore, to obtain the pdf of Y, we go as follows. Let g be the pdf of Y, maybe I write it as g y and G y be the cumulative distribution function; Therefore, G of y is equal to probability Y less than equal to y which is equal to probability X square less than equal to y. And because of the symmetricity around 0, we can find this is equal to probability minus root y less than equal to x less than equal to y.

(Refer Slide Time: 19:56)

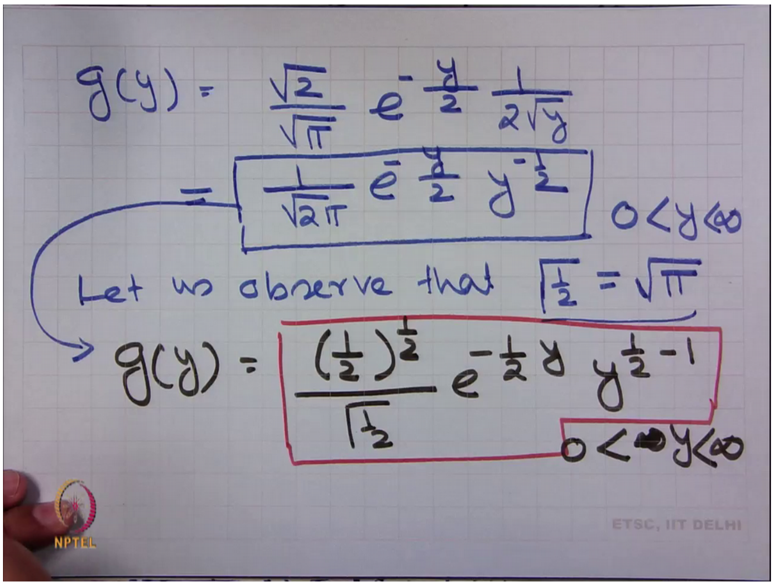


Now, since normal is symmetric, and if this is minus root y and this is plus root y, then this probability is actually 2 times the probability that x lies between 0 to root y. Therefore, probability minus root y less than equal to x less than equal to root y is equal to 2 times 0 to root y fx dx. And since f is normal 0 1 so, this becomes 2 times 0 to root y 1 over root over 2 pi e to the power minus x square by 2 dx.

So, this is the value of Gy that is the cumulative distribution function for the random variable y which is nothing but x square. Therefore, gy is equal to G prime y; that means, I am differentiating this with respect to y. So, we know that that we get by first replacing x with root y in this formula, multiplied by the derivative of root y with respect to y.

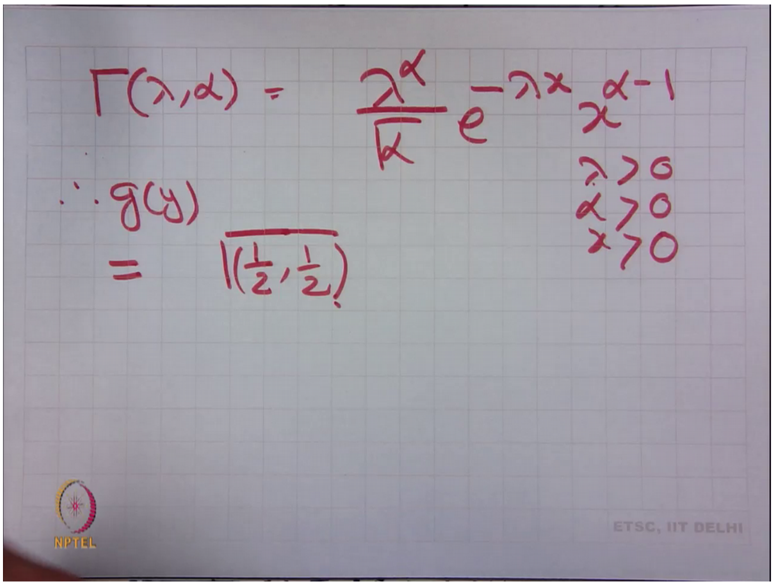
So, this is 2 times 1 over root 2 pi e to the power minus root y square is equal to y by 2 multiplied by d root y d y. Because we know that it is dx dy expressed in terms of y and x is equal to root y. Therefore, we can write it as half root y since; so, we replace this value here. Therefore, what we get?

(Refer Slide Time: 23:00)



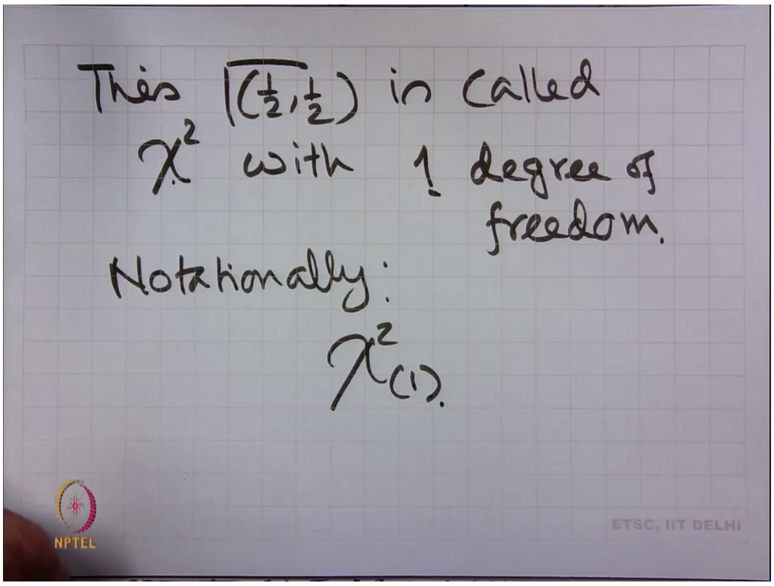
We get g of y is equal to root 2 upon root pi e to the power minus y by 2 into 1 upon 2 root y, which on simplification becomes 1 over root 2 pi e to the power minus y by 2 y to the power minus half. So, we get a new type of density function for y; where y lies in the interval 0 to infinity. Now is it a density that is completely unknown to us. Perhaps most of you will say, yes. So, let us observe one thing, gamma of half is equal to root over pi. I am sure you have come across this in your mathematics course. So, I am utilizing this property, and writing this as g y is equal to half to the power half 1 upon root 2 is equal to half to the power half upon gamma half e to the power minus half y, y to the power half minus 1 0 less than y less than infinity.

(Refer Slide Time: 25:21)



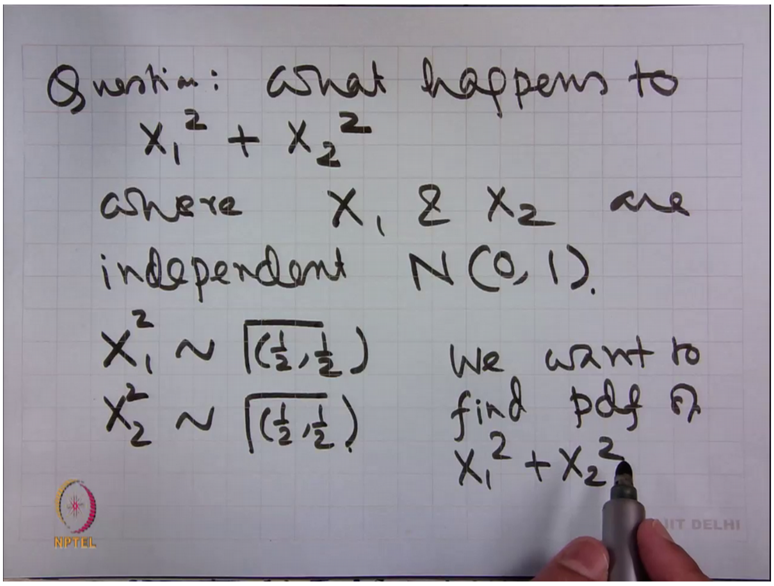
Do you remember this density? I am sure you can because, this is of the form gamma lambda alpha which is lambda power alpha upon gamma alpha e to the power minus lambda x, x to the power alpha minus 1, lambda greater than 0 alpha greater than 0 and x greater than 0. Therefore, we can say is equal to actually gamma half comma half.

(Refer Slide Time: 25:56)



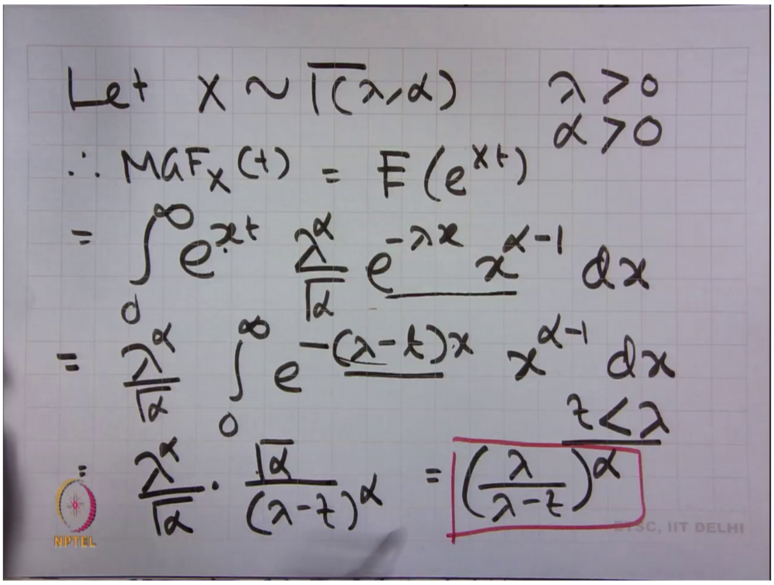
This gamma half half is called chi square with one degrees of freedom. Why it is called one degree of freedom? Because this chi square has come from one normal distribution, that is why there is one independent random variable x which is giving rise to this chi square distribution. And therefore, we call it chi square with one degrees of freedom and notationally chi square with one degrees of freedom.

(Refer Slide Time: 27:01)



Now, the question is what happens to X 1 square plus X 2 square where X 1 and X 2 are independent normal 0 1. We know that X 1 will become gamma half, half. X 2 will become gamma half, half. And therefore, given 2 different random variables which are independent we want to find pdf of X 1 square plus X 2 square. This we can do in many ways, let me first do it using moment generating function. In fact, I prove something more general than just half half. In fact, I prove it for general gamma distribution.

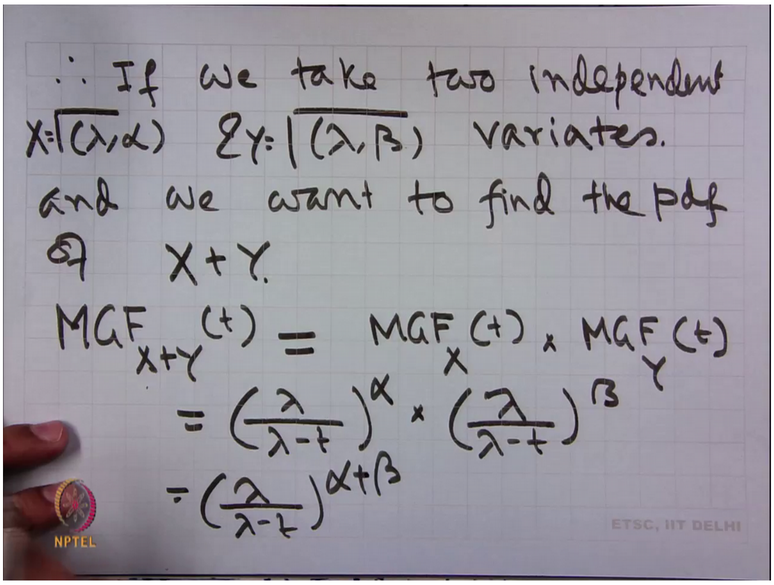
(Refer Slide Time: 28:51)



Let x be a gamma lambda alpha lambda greater than 0 alpha greater than 0. Therefore, moment generating function of x is equal to expected value of e to the power X t; is equal to 0 to infinity e to the power x t multiplied by the pdf of x, which is lambda power alpha upon gamma alpha e to the power minus lambda x, x to the power alpha minus 1 dx, right? Is equal to so, I have used this and this together multiplied by x to the power alpha minus 1 dx and this will converge for t less than lambda; Because in order to converge this part has to be positive. So, that along with this minus it becomes negative.

So, this is this will hold good for t less than lambda. And we can easily find out this integral, because we know that when we are integrating this part only. This is the pdf of gamma distribution. Therefore, this integrates to 1, and therefore, the integration of this part has to be gamma alpha upon lambda power alpha so that it cancels out. Therefore, by comparing we can easily write that this part is going to be instead of gamma alpha upon lambda power alpha it is going to be gamma alpha upon lambda minus t power alpha, which says that the MGF is equal to lambda upon lambda minus t whole to the power alpha. So, this is the moment generating function for normal for gamma lambda alpha.

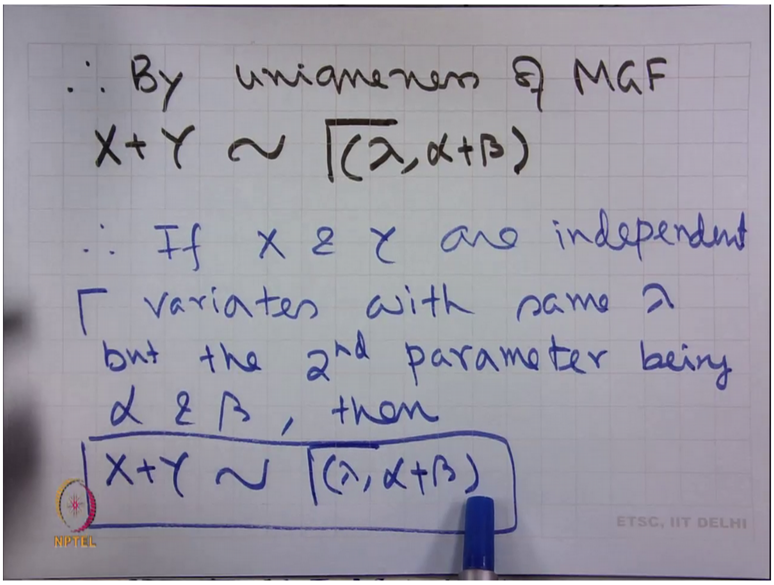
(Refer Slide Time: 31:35)



Therefore, if we take 2 independent gamma random variables, gamma lambda alpha and gamma lambda beta variates; And we want to know the pdf of so, say this is called X and this is called Y X plus Y then MGF of X plus Y t, we have already seen that if 2 random variables are independent, then the MGF of their addition of their sum is product of their individual moment generating functions.

And just now we have found out that this moment generating function is lambda upon lambda minus t whole to the power alpha. This is similarly going to be lambda upon lambda minus t whole to the power beta. Therefore, the product becomes lambda upon lambda minus t whole to the power alpha plus beta.

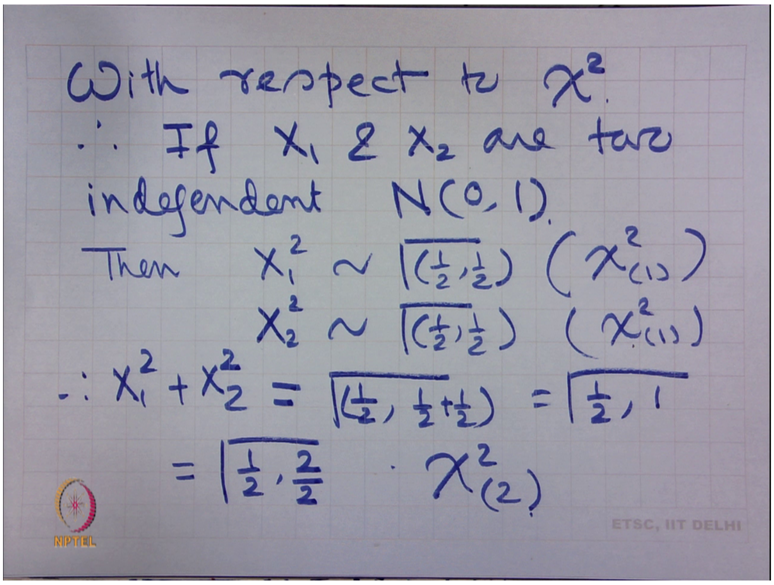
(Refer Slide Time: 33:45)



And therefore, by uniqueness theorem by uniqueness of moment generating function, we can say that x plus y is distributed as gamma with lambda same, but this parameter becomes alpha plus beta. So, what we obtain is that if X and Y are independent gamma variates with same lambda, but the second parameter being alpha and beta, then X plus Y is also a gamma variate with lambda alpha plus beta.

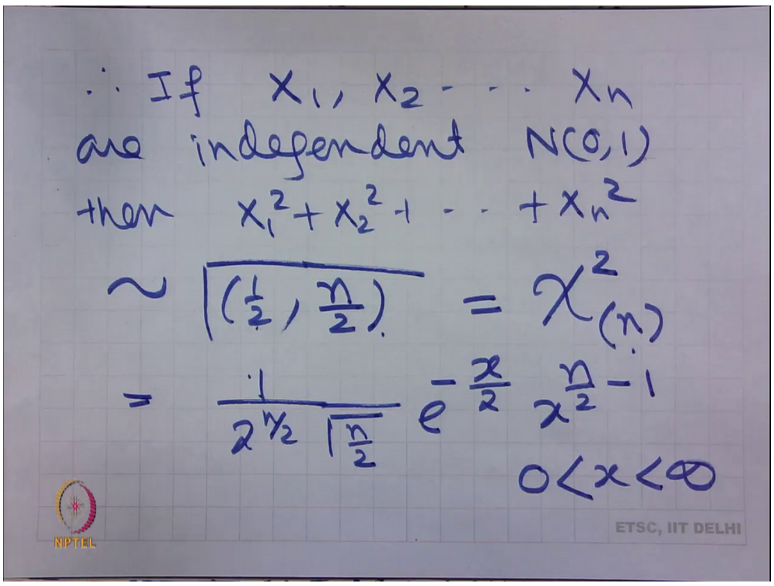
Therefore, we find an interesting result with respect to gamma random variable. That as we keep on adding independent random gamma variables, with the same lambda then the summation of these random variables is also gamma with the same parameter lambda, but the second parameter being the sum of the individual parameters.

(Refer Slide Time: 35:35)



The advantages now with respect to chi square therefore, if X 1 and X 2 are 2 independent normal 0 1, then X 1 square is basically a gamma variate with half and half. X 2 square is also a gamma variate with half and half, therefore, X 1 square plus X 2 square is it also a gamma variate with parameters half and half plus half is equal to gamma with half comma one; which we write as gamma with half 2 by 2. And this is called chi square with 2 degrees of freedom. Why 2 degrees of freedom? Because we have used 2 independent normal 0 1; Can we therefore, generalize from there?

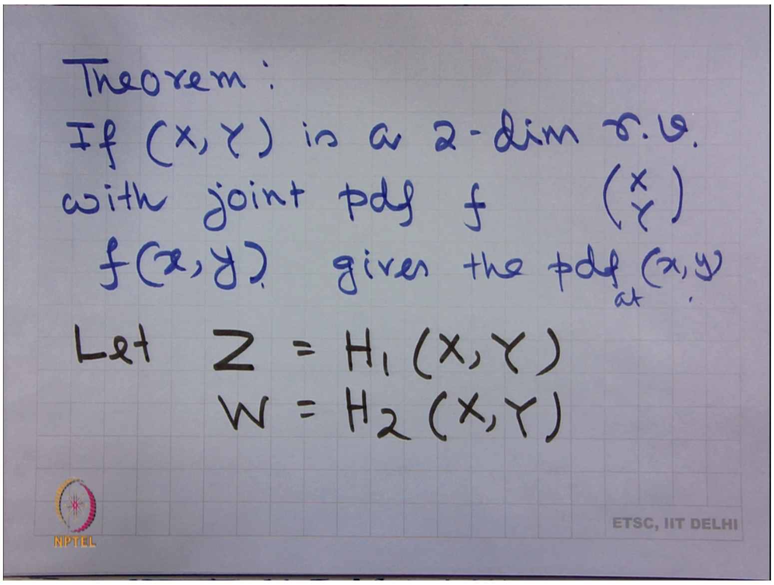
(Refer Slide Time: 37:21)



We can, in fact, X 1 X 2 Xn are independent normal 0 1, then X 1 square plus X 2 square plus Xn square will be gamma with half. And the second parameter it is half for each on each one of them therefore, when we add them up will get n by 2. Therefore, a gamma half with n by 2 is same as a chi square distribution with degrees of freedom n. So, sum of square of in independent random variables is chi square with n degrees of freedom, and we can write it is pdf very simply, which is equal to lambda power alpha upon gamma alpha into e to the power minus lambda x x to the power n by 2 minus 1 for 0 less than x less than infinity.

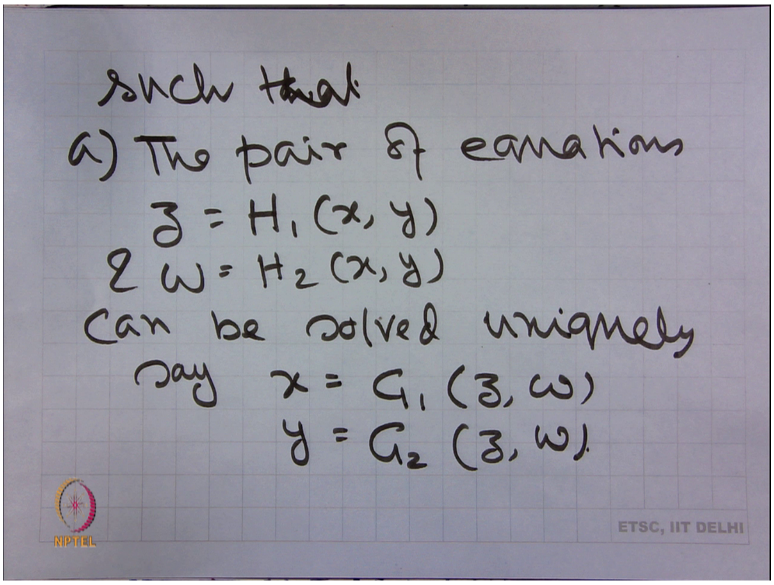
In this case, we could easily find the sum of 2 independent random variables. In general, how do you find the sum of 2 arbitrary random variables or why some why not weighted sum of 2 random variables, the difference between 2 random variables or even the product of 2 random variables or division of x by y, when y not equal to 0, that is also a random variable. So, is there any way to find the pdf of a function of 2 random variables. So, the following theorem helps us in adding the following theorem helps us to find the pdf of function of 2 random variables under certain conditions. So, I am stating the theorem, but I am not going to prove it, because the mathematics for proving that is beyond the scope of this lecture.

(Refer Slide Time: 40:43)



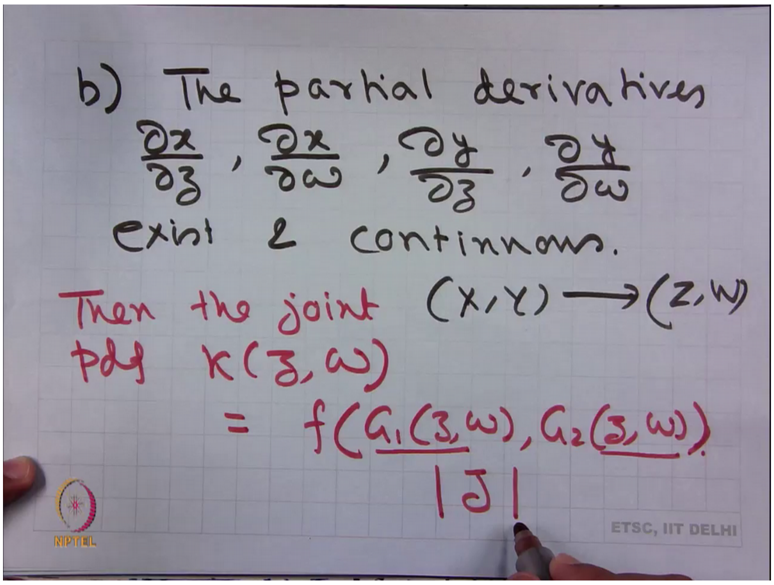
So, X and Y 2 random variables which we can write it as X Y. And therefore, I am writing as a 2 dimensional random variable with joint pdf f; that is, f of x y gives the pdf at x comma y. Now, consider two functions H 1 and H 2; such that Z is equal to H 1 of X Y. So, Z is a function of X and Y, W is a function of X Y. Therefore, basically from XY plane we are transforming it into another plane of z w; such that the pair of equations z is equal to H 1 x y and w is equal to H 2 x y can be solved uniquely.

(Refer Slide Time: 42:05)



Say, x is equal to G 1 of z w, and y is equal to g 2 of z w. So, you notice the similarity with when we are talking about function of a single random variable. We wanted the function to be monotonically increasing or decreasing so that we can get the inverse unity. Similarly, given z and w we want to obtain that from the values z and w we can identify the x and y uniquely. So, this condition ensures that also.

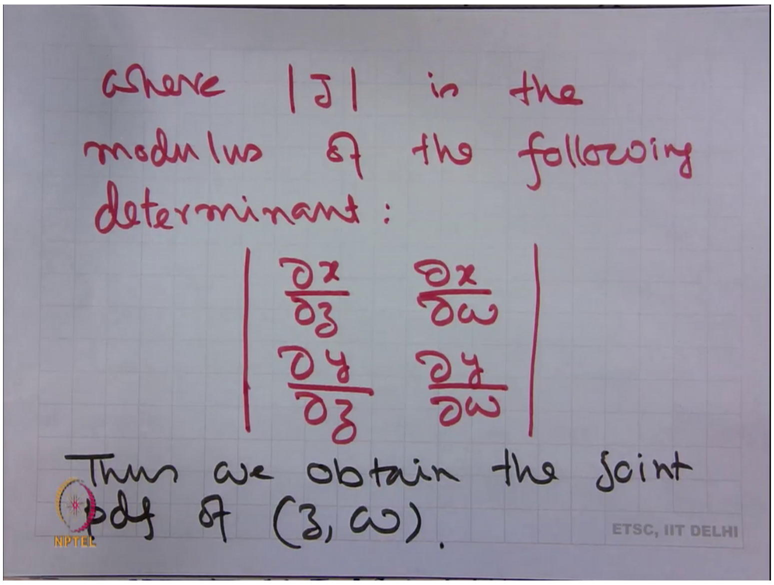
(Refer Slide Time: 43:39)



Another condition is that the partial derivatives del x, del z, del x, del w, del y, del z, del y, del w exists and continuous. So, we have from XY plane a mapping to Z W plane such that it is one to one. So, that given any pair here we can uniquely determine the corresponding X Y. And also the partial derivatives of X and Y with respect to both Z and W they exist and continuous.

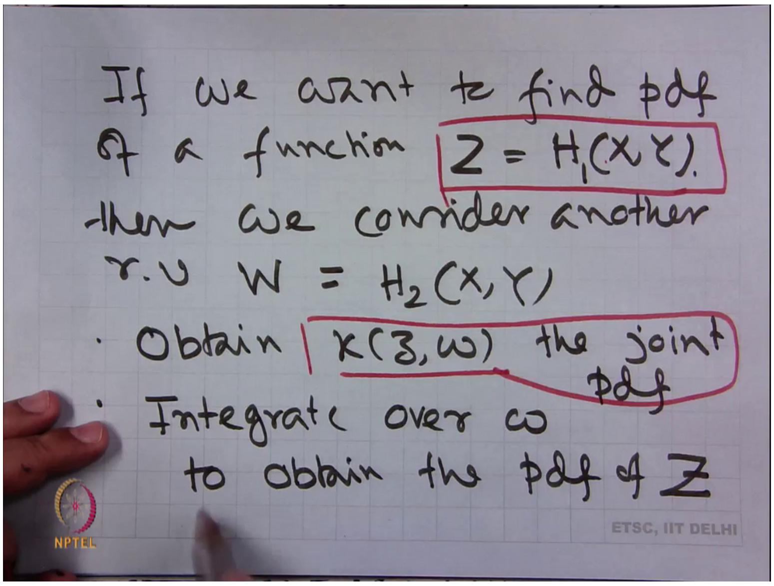
Then the joint pdf k z w so, we are looking at the joint probability density function of these 2 D random variables z w is f at G 1 z w, G 2 z w; that means, we are looking at the pdf of original random variable x y, but expressed in terms of z w multiplied by something which is called the jacobian. What is Jacobean?

(Refer Slide Time: 45:51)



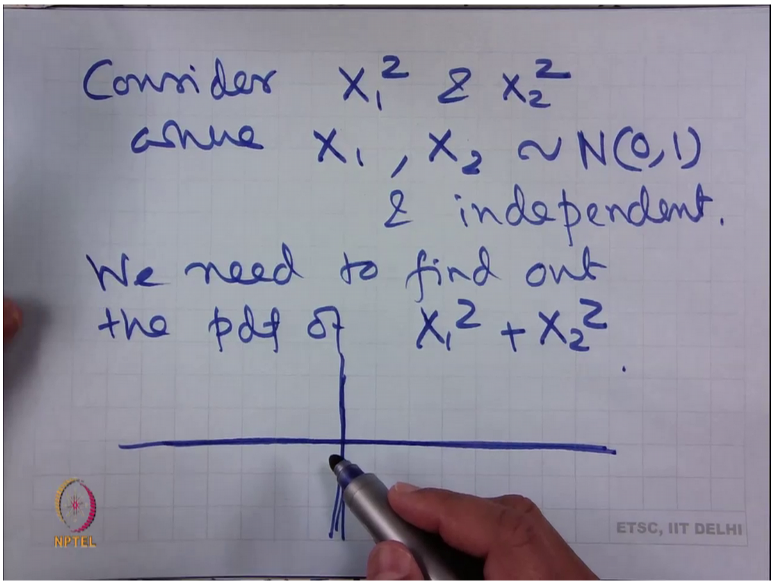
Jacobian is the modulus of the following determinant del x, del z, del x, del w, del y, del z, del y, del w. So, we compute the determinant and the pdf f expressed in terms of z w multiplied by the determinant gives us the pdf for z w the joint pdf of z and w. What is the advantage?

(Refer Slide Time: 47:07)



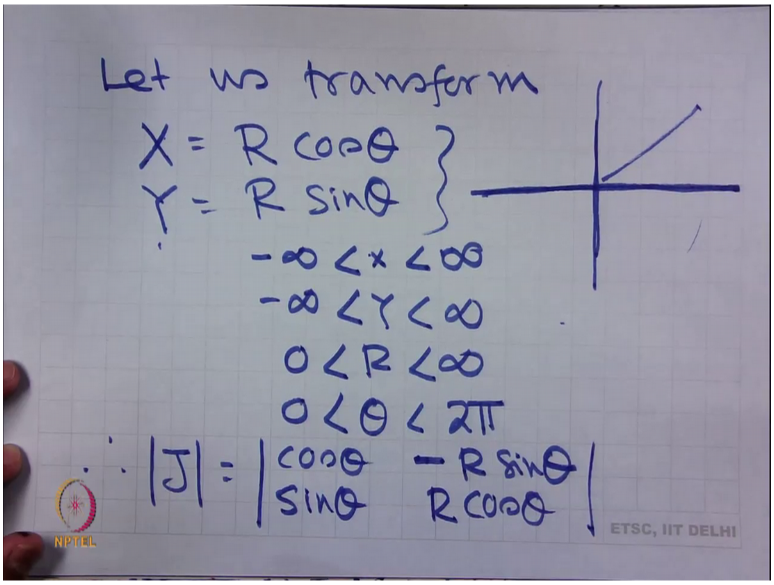
The advantages if we want to find pdf of a function z is equal to H of x y, then we consider another random variable w is equal to say z is equal to H 1. So, let us call it H 2 of X Y obtain k z w the joint pdf, and just now I have shown the formula for obtaining it integrate over w to obtain the pdf of z. Since our interest was only in this. In order to get it is pdf, we need to find out first the joint pdf by appropriately defining w. So, that this integration becomes easier, and from there we obtain the pdf of z. Therefore, let us now look at the same problem of summation of 2 chi square distribution.

(Refer Slide Time: 49:30)



We need to find out the pdf of X 1 square plus X 2 square. Since X 1 is normal 0 1 so, we can consider X 1 along this axis, and X 2 along the y axis so that X 1 X 2 together can cover the entire 2 D plane, where you are doing that? Because that guides us to take a very meaningful transformation; what is that?

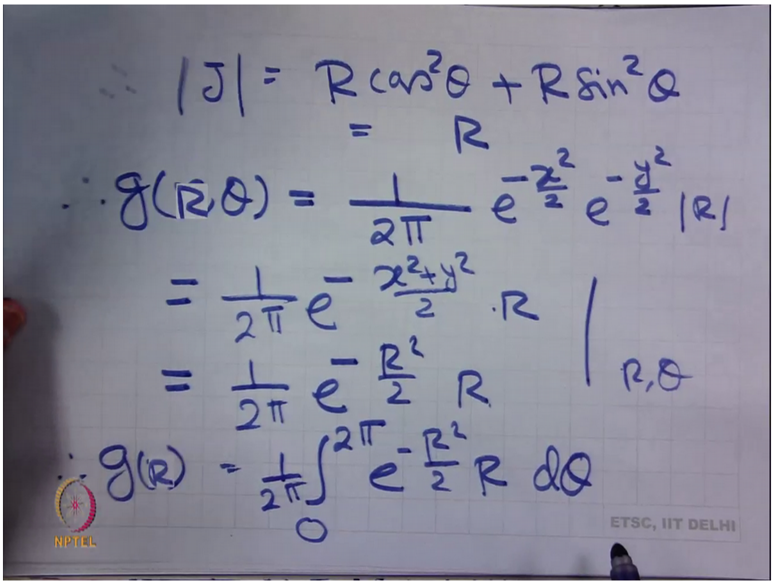
(Refer Slide Time: 50:45)



So, we transform X is equal to R cos theta Y is equal to R sin theta. So, instead of X 1 and X 2, let me call them X and Y, therefore, we are covering the entire 2 D plane therefore, minus infinity less than x less than infinity, minus infinity less than y less than infinity or is the radius of transformation, therefore, R is going to be from 0 to infinity, and theta is covering the entire plane. So, theta will belong to 0 to 2 pi.

Therefore, Jacobean of transformation is equal to dX dR which is cos theta. DX d theta which is minus R sin theta, dY d R which is sin theta and dY d theta is equal to R cos theta.

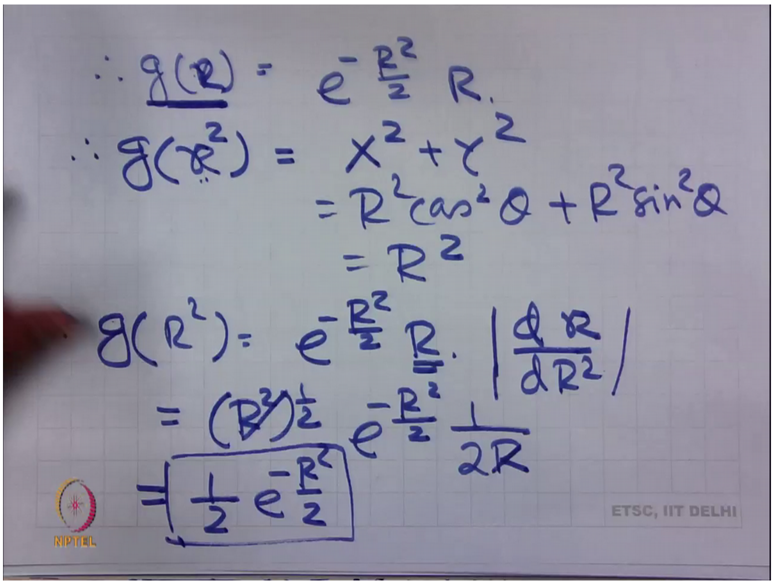
(Refer Slide Time: 52:16)



Therefore determinant of J is equal to R cos square theta plus R sin square theta is equal to R. Therefore, that joint pdf R theta is equal to since x and y are independent, we can write the distribution as the product of their individual is equal to 1 over 2 pi e to the power minus x square plus y square by 2 into R. Now this has to be expressed in terms of R theta, therefore, this is 1 over 2 pi into e to the power minus R square by 2 into R.

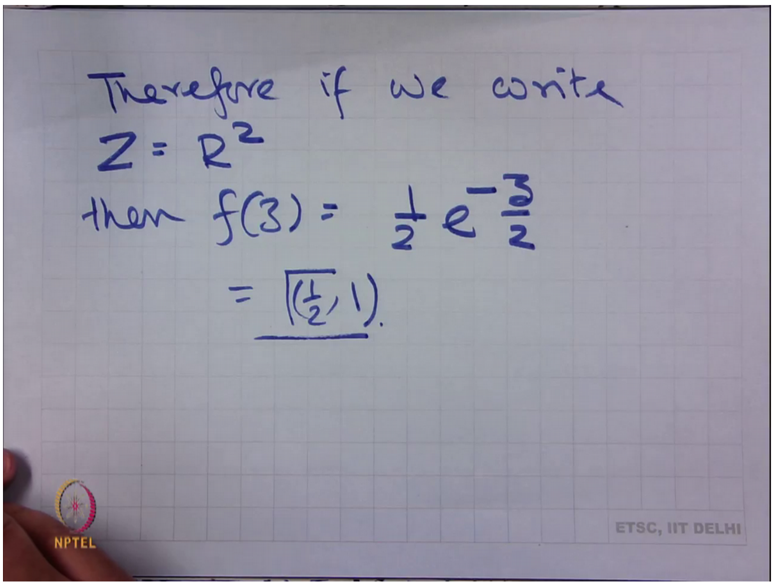
Therefore, g of R is equal to now I have to integrate out theta from 0 to 2 pi e to the power minus R square by 2 into R into d theta. And therefore, these 2 pi is being cancelled therefore, what we get is, but we need to find out the density of R square.

(Refer Slide Time: 53:58)



Because we need x square plus y square is equal to R square cos square theta plus R square sin square theta is equal to R square. But so far we obtained the pdf of R, now we need to find out the pdf of R square, therefore, we use the transformation of the first one, let me write it in capital. Therefore, g of R square is equal to e to the power minus R square by 2 into R multiplied by d R upon d R square; is equal to these are therefore, we write as R square to the power half into e to the power minus R square by 2 and dr upon dr square is equal to 1 upon 2 R. So, this cancels with this, and what we are get it half e to the power minus R square by 2.

(Refer Slide Time: 55:31)



Therefore if we write Z is equal to R square, then f at z is equal to half into e to the power minus z by 2; which is equal to gamma half gamma 1. So, this is the result that we have already obtained using moment generating function. With that I stopped here, in the next class I will be talking about some more different types of random variables, and we will try to obtain their pdf's in a similar way.

Thank you.