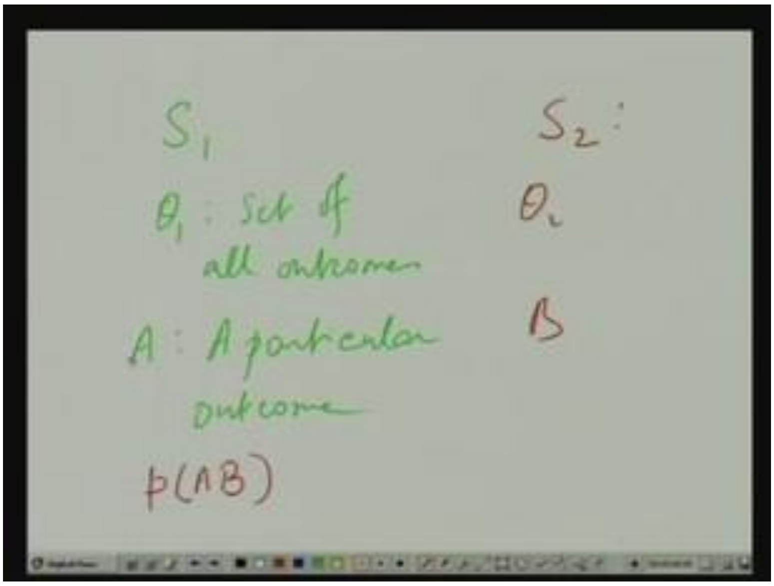
Probability and Random Variables
Prof. M. Chakraborty
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 13
Function of Two Random Variables

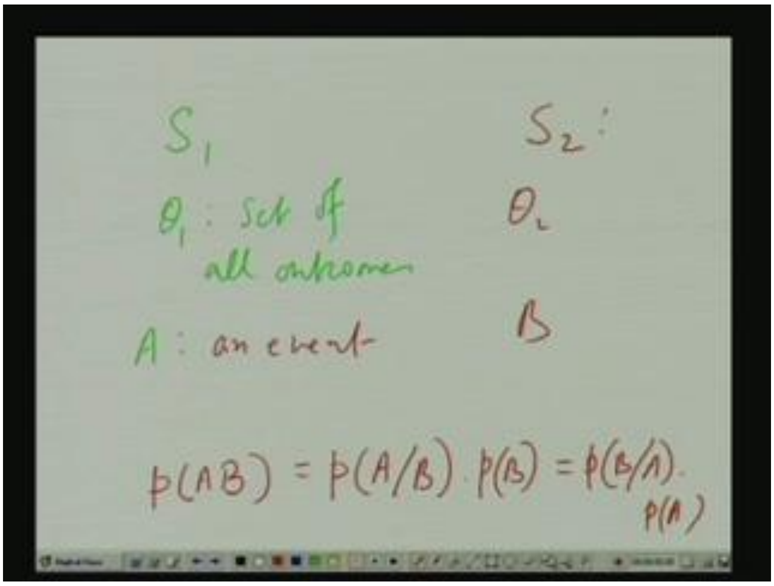
So, last time, we were discussing two random variables, that is, the joint statistics, joint probability density, joint probability distribution and things like that. Then, a question that… A very important question, a very important point of property is rather is statistical independence of two random variables; that is when can we say that, the other two random variables in these variables x and y – they are statistically independent. Now, to answer that let us go back to some of our earlier discussions.

(Refer Slide Time: 01:32)



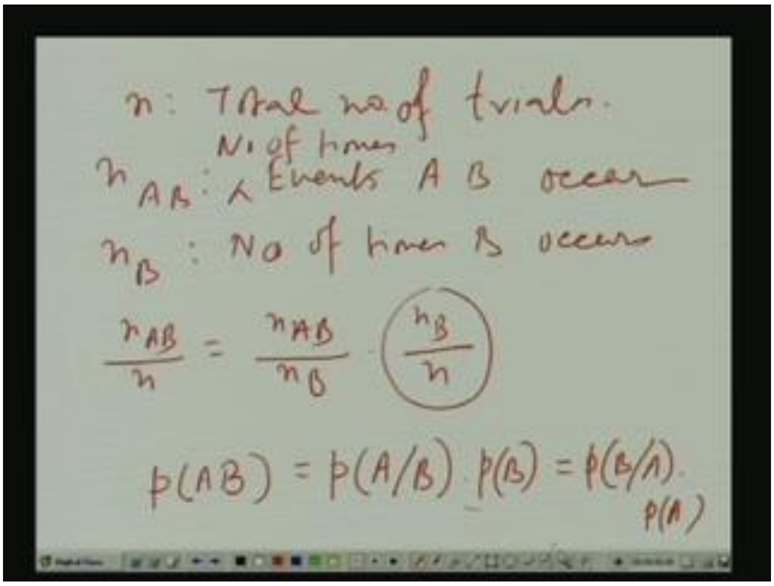
Suppose we have got two probability spaces. One is S 1; that has the outcomes say theta 1's, that is, set of all outcomes. And A be a particular outcome. Similarly, suppose S 2; given theta 2 and B; theta 2 is the set of all outcomes here. This is another probability space and B is a particular outcome. In that case, we have seen one result that, the joint probability of this event or maybe instead of calling A to be an outcome, it will be better or more generally if we say is an event, that is a set of outcomes, and B also is an event. So, let me erase it.

(Refer Slide Time: 02:52)



So, what is AB? This is a joint event; that is, we have got both these experiments going on; theta 1 is the set of all outcomes; theta 2 is set of outcomes; A – a particular event B – a particular event. AB means joint event; A occurring and B occurring simultaneously. What was this probability of this joint event? That you had seen was given by either like this p A by B into p B, was equal to also p B by A into p A. In case you have forgotten, we can also see how it comes.

(Refer Slide Time: 03:42)

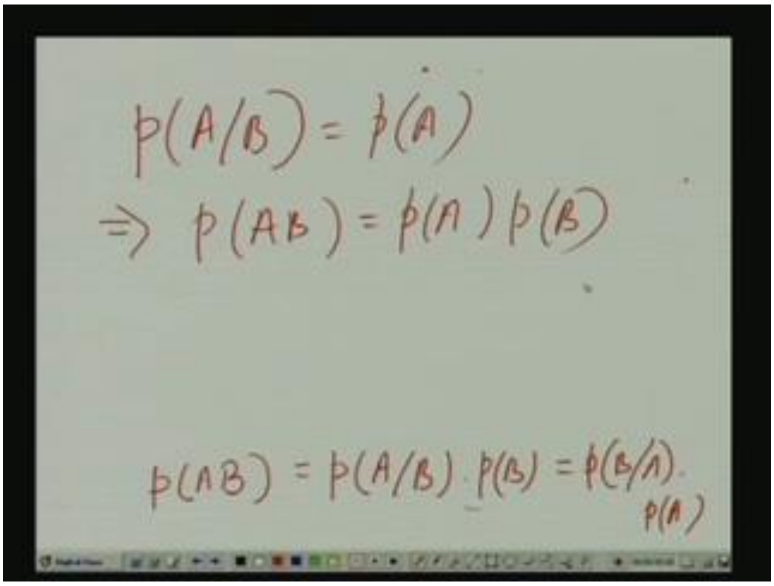


That is, suppose n is the total number of trials; out of these total trials, actually we are observing both outcomes from or events from S 1 and S 2 simultaneously parallely.

Suppose on n AB occasions, events A B occur; and on n A, what is an n B occasions? Number of times – these also number of times; number of times, the events AB occur; that is, A and B occurring simultaneously; that is n A B. And n B is the number of times B occurs. Obviously, n B should be greater than equal to n A B because number of times B occurs means the event from the other space could be A or may not be A. But, as long as the event from the space is B, it is incorporated, it is added here as part of this. Then, you have seen that, the joint probability of the event AB is nothing n A B by n; which we can write as n A B by n B to n B by n.

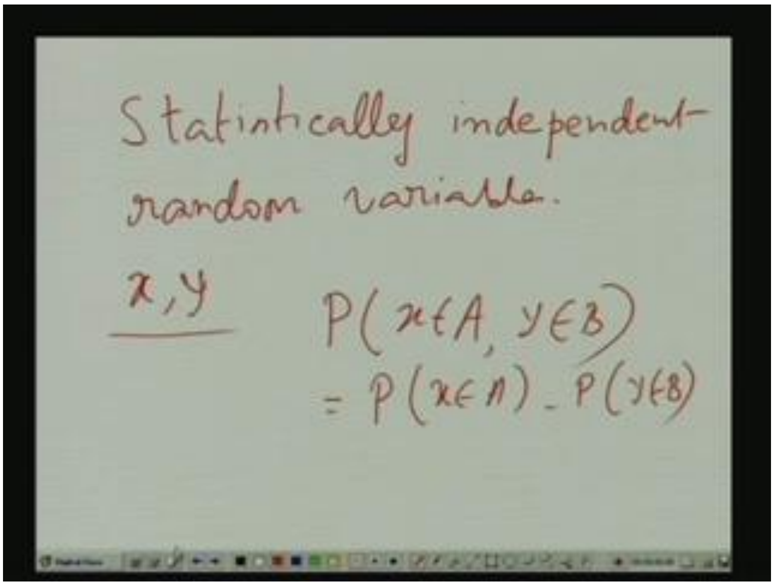
Now, if n is large, n B by n is simply the probability of B, which is here. And n A B by n B – What is it? n B is the total number of trials for which the event B is fixed; that is, the event, which is occurring from the other set, other space, that is, B. And it is fixed at B. Under that condition, how many times we get A? n A B times. So, under the condition that B is fixed, what is the probability of AB occurring? That is the probability. It is the conditional probability – p A by B. By the same logic, you can have p B by A into p A. That you have seen. These works well; these works for the events; but here there is no notion of the random variable; only events. So, in the context of events, we have already defined what is joint probability?

(Refer Slide Time: 06:31)



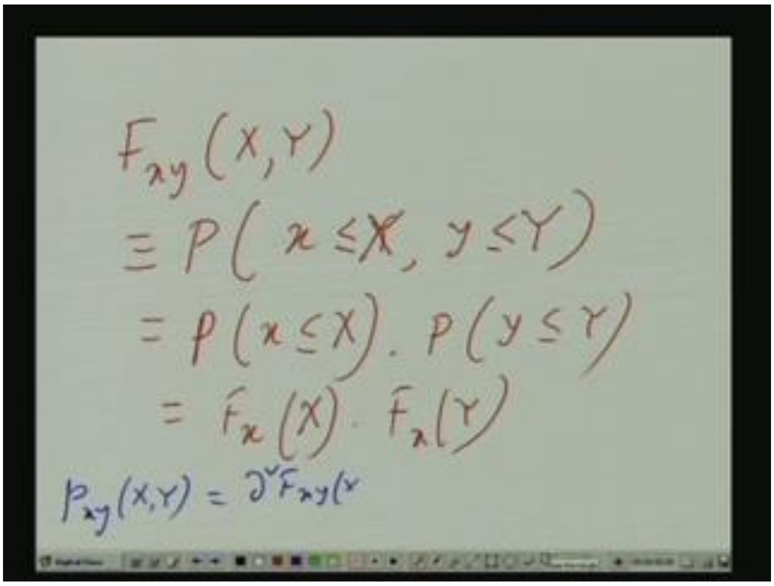
Then, we have said that, two events – A B – they are statistical independent if p A by B or independent rather is simply p A. And this is obvious. Suppose A and B are two such events that, B really… I mean this probability space S 1 and S 2 – I mean they are such or the corresponding experiments or the contexts are such; then they have no mutual influence. Then, absolutely, they are independent of each other. In that case, whether you find the probability of A by restricting the event from space to B or not, you should get the same probability, because the other space has no effect on the outcome in the space S 1. So, p A by B should be equal to p A. And in this case… In that case, the joint probability becomes… That is the product of the individual probabilities. So, joint probability or probability of the joint event AB is nothing but the probability of the event A multiplied by the probability of event B. That is what we have seen that time. But, now, we do not have the events as such; we are speaking in the language of random variables. So, you have to define similarly, what is meant by statistical independent random variables.

(Refer Slide Time: 07:46)



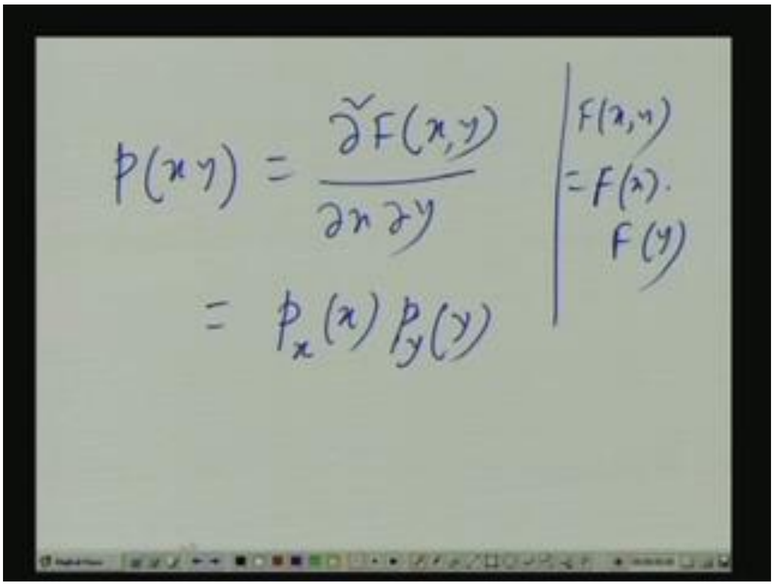
Here we are giving x, y – two random variables. So, x and y. Then, we say that, they would be statistically independent if the probability of the event say x belonging to A – a particular set A along the x-axis; and y – element of B – the joint probability; x taken from some set A and y taken from some set B. This is on the y-axis; this is on the x-axis. The joint probability is nothing but the probability of x belonging to A multiplied by the probability of y belonging to B. But, this should be true for any set A on the x-axis and any set B on the y-axis. Remember the word, notice the word – any. So, if it is true; that if I take any set on the x-axis and call it A; and any set on the y-axis and call it B, that the joint probability of x belonging to A and y belonging to B; if the joint probability is simply the product of the individual probabilities; that is, the probability of x belonging to A multiplied by probability of y belonging to B; then we will say that, x and y – they are statistically independent random variables. This is true for any set A. This should be valid for any set A from the x-axis and any set B on the y-axis. If that be so; then what happens to the probability distribution – joint probability distribution?

(Refer Slide Time: 10:16)



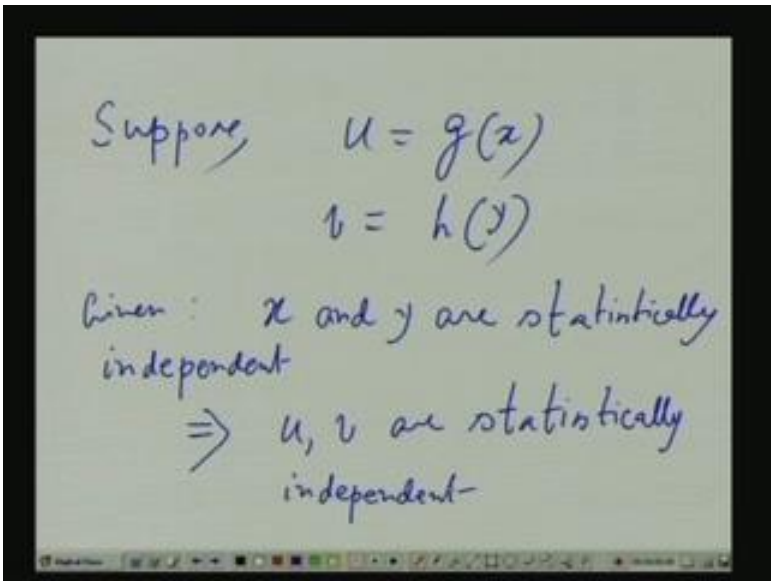
F x y of X, Y. What happens here? So, this means… What is this? After all, this is nothing but the probability of x. We can easily extend the previous argument. This means x less than equal to capital X means what? That on the x-axis, I am going from minus infinity up to x. So, that subset of the real axis, so x belonging to that subset. And here… So, here capital A x is nothing but that subset of real axis… What subset? From minus infinity up to capital X. And x less than equal to capital X means x belonging to that subset A. Similarly, y less than equal capital Y means we take a subset B on the y-axis. B goes from minus infinity up to capital Y; and y less than equal to capital Y means y belonging to B. So, if x and y are independent, then using our previous argument, this probability is nothing but product of the individual probabilities – probability of x less than equal to X multiplied by probability of y less than equal to capital Y. But, the first one is nothing but the probability distribution of x up to capital X. Second one is the probability distribution of y up to capital Y; which means if two random variables are statistically independent, then their joint distribution is nothing but the product of individual distributions. Same would hold for the probability density also. After all, what is probability density? p x y at capital XY is nothing but… Or, let me just simplify the notation instead of writing so many… Writing in terms of…

(Refer Slide Time: 13:00)



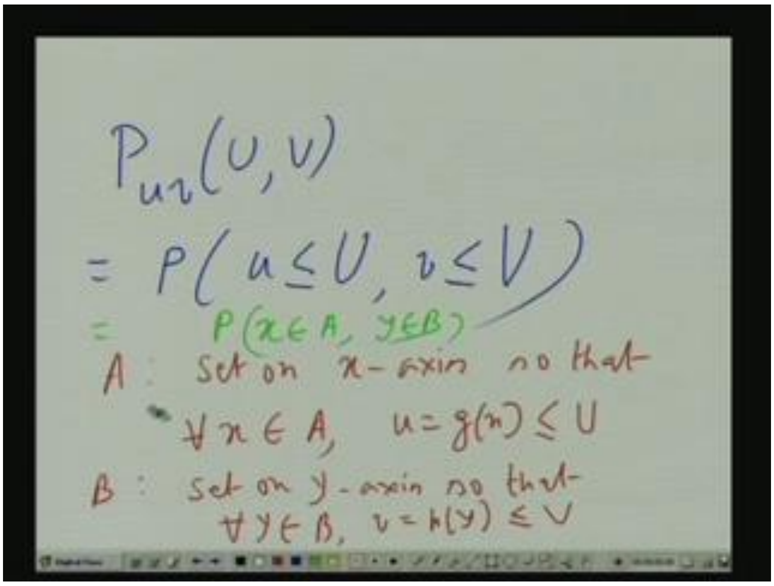
This is nothing but summed on using the subscripts and directly putting x – small x and y the variables inside, because that makes just a simple notations; that is all. Is the double derivative of these with respect to x and y. But, we have seen that, if x and y – they are statically independent; then this F x comma y is nothing F x multiplied by F y. And if we replace F of x, y; that is, as… If we put it back here, then this simply becomes p x x. Here I am bringing back again x y subscript, because otherwise… Using the same notation p here, p here; but actually this is one probability density; this is another probability density. These two functions are not same. That is why I am just putting a subscript x here and y here just to indicate that, these two are two defined functions. These are density function associated with x. This is the density function associated with y. Just for that, I am putting the two subscripts.

(Refer Slide Time: 14:41)



Then, suppose u is a random variable, which is equal to nothing but some function g of x. And v is another random variable, which is nothing but some function h of y; where, x and y – they are jointly random variables; they are random variables and they have got some joint density. For that, it is given that, x and y – they are statistically dependent. Then, we say that, u and v – these two random variables also will be statistically independent. Why and how? So, I repeat the statement. Given – x and y are statistically independent; it should imply u, v also are… It is not very difficult to show; that we can easily show it. What is the joint probability density? Or maybe instead of density, it is better to work with distributions. So, let me erase this.

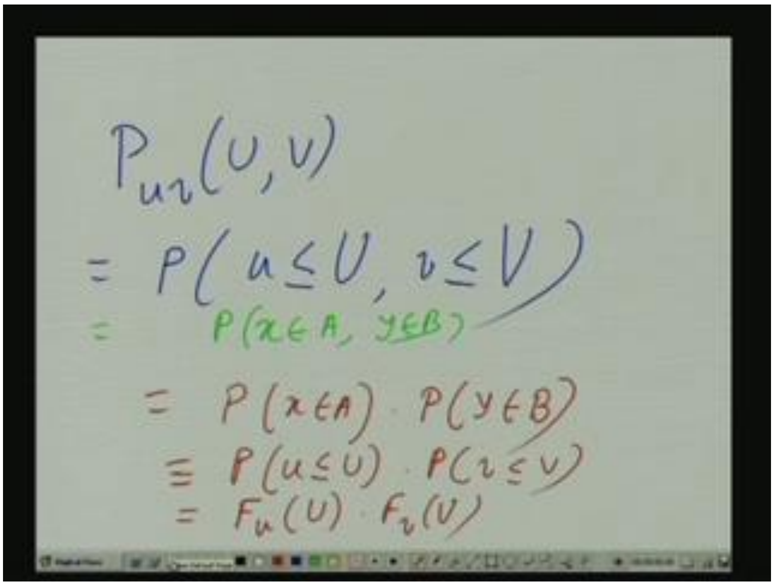
(Refer Slide Time: 16:41)



Joint probability distribution of u v, what is this? This is nothing but the probability of u taking value less than equal to sum number capital U; v taking value less than equal to sum number capital V. Now, u is coming from x; x is a random variable; it can take any value on the real axis. And since we have given a function g, which work on x and gives you u. So, let A be a set on x-axis so that, for all x belonging to A… Let me g x; that is, u equal to g x less than equal to capital U. And another thing is implicit here; which I am not mentioning; that is, any x not belonging to A… For any x not belonging to A, corresponding u equal to g x is greater than capital U. So, I am basically looking at the entire x-axis. So, I am collecting all regions for which g of x is less than equal to this capital U, that collection of all such regions is called A here. So, outside A, on the x-axis… In all regions, there is outside A or not belonging to A; corresponding u will be greater than capital U.

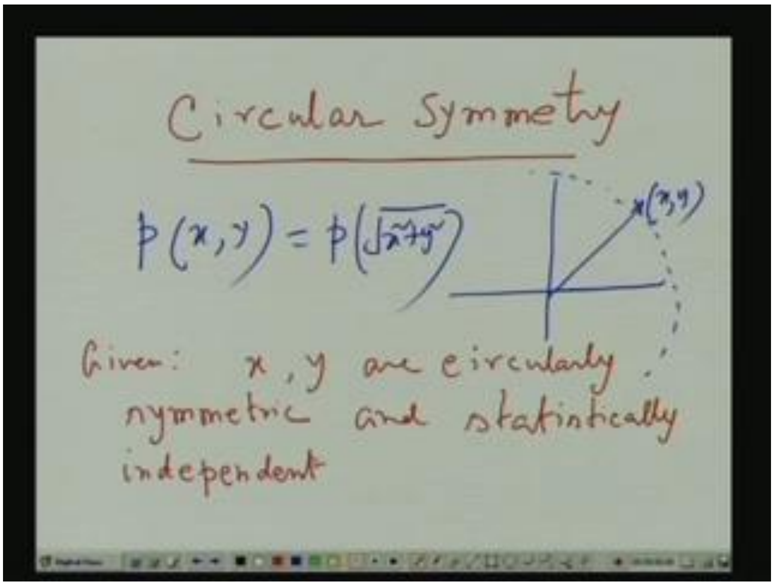
Similarly, let B be… That is, whenever I am looking at this y-axis actually, I am collecting all regions for which… I mean if you take y from those regions v of… that is, h of y is less than equal capital V. Further… And collection of these regions is called B. And further if you do not take y from B; but from any other portion – means y-axis; then the corresponding h of y; that is, v should be greater than capital V. In that case, this probability – this probability of u less than equal to capital U and v less than equal to capital V. This event – this is equivalent to this event; x belonging to capital A, and simultaneously, y belonging to capital B. So, probability of u less than equal to U, v less than equal to capital V same as probability of x belonging to capital A y belonging to capital B.

(Refer Slide Time: 20:34)



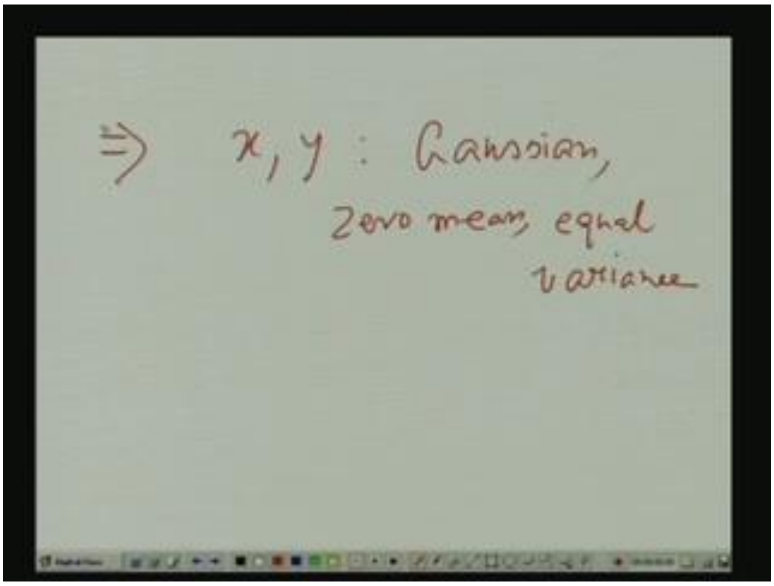
But, we have been given this condition that x and y – they are statistically independent; that means, this probability is nothing but product of these two probabilities. And x element of A means it is equivalent to small u less than equal to capital U, because only when small u less than equal to capital U means, that is, if and only if x is an element of A. So, these events are same. So, this probability is nothing but is equivalent to probability of… that is, by same logic, this is equivalent to the probability of v less than equal to capital V. And by definition, this is the probability distribution of u. And these are probability distribution of V. So, this means that, u and v – they are also then statistically independent. This is why I wanted to… This is what I wanted to prove. So, I repeat if x and y – they are statistically independent random variables; they are function of x and function of y also are statistically independent random variables.

(Refer Slide Time: 22:01)



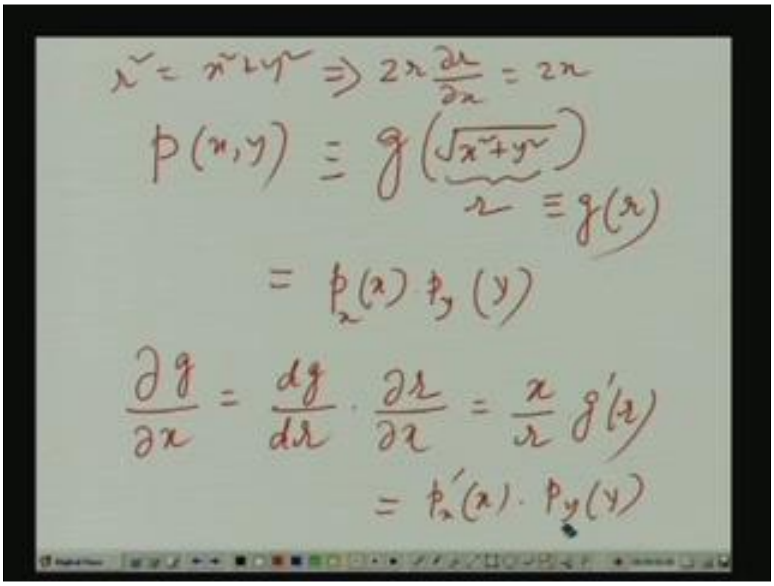
Then, I move to this topic. Two random variables are given x and y. So, they can take any value; that xy pair can take any value along the xy plane. But, they will be called circularly symmetric random variables. If the joint density, that is, joint density – it either depends on x l or not depends on y alone. But, it simply depends on the distance of the point; there is a point here say x, y. The distance of this point from origin; if this density depends on the distance, that is, square root x square plus y square; then x and y will be called circularly symmetric random variables. In this case, if I draw a circle to the same radius; where, at all points on the circumference of the circle have got the same density. So, it is circularly symmetric. Now, suppose it is given… And they are also statistically independent.

(Refer Slide Time: 24:24)



Then, this should mean x, y they can have any… They cannot have any other density – other probability density. But, they have to be Gaussian; that is, they are Gaussian. Sometimes with this Gaussian is also called normal distribution. So, we can see Gaussian or normal, zero mean, equal variance. That is the joint density. Since they are statistical independent, after all they will be… The joint density will be product of individual densities; individual densities will be Gaussian with zero mean. And variance for both of them will be same. That is what we will prove now. But, for that, x y should not only be statistically independent. They must be circularly symmetric. Then, their joint density or other individual densities for that matter are Gaussian. They have to be Gaussian random variable. There cannot be any other probability density possible. How?

(Refer Slide Time: 25:51)

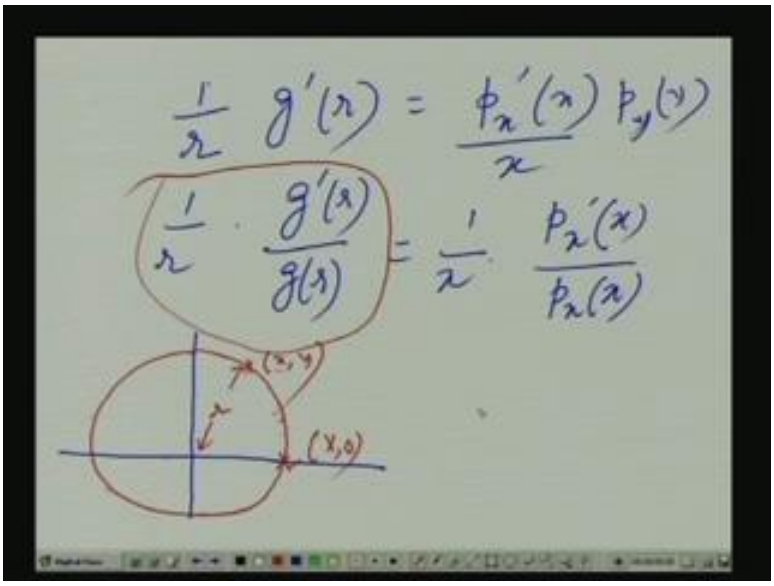


So, the joint density p x comma y – when you write it in terms of x and y, density is given as this function. But, as I told you, x and y – they are circularly symmetric; that means this density depends only on square root of x square plus y square or r. So, if you express it as a function of r, it takes a defined form – this function. It is no longer in terms of p; you can express it as a defined function. That function I call g. That is equivalent to g r. This is r. But, we have given this fact that, x and y – they are statistically independent. So, g r is also product of these two. Again I am putting the subscript x and y. Just differentiate between the probability density for x and probability density for y, because I am using the same symbol p for both.

Now, see one thing; if I want to carry out this differentiation g of r; and therefore, it depends on x and y both. So, the partial derivative from g with respect to x. I can write it as dg dr; it is d, because when you write it as a function of r, it is a function of single variable. So, no question of partial derivative, but usual derivative. Then, del r; here del comes, because r is a function of x and y both; it is a function of two variables. So, del r del x. And what is del r del x? We have seen that, r square is x square plus y square. This means twice r… If you differentiate this set with respect to x twice r, then del r del x; that is equal to twice x dx; which means del r del x is nothing but… No dx; which means del r del x is nothing but x by r. So, what I get here is x by r dg dr; which I call it as g prime r. That is a derivative from g r with respect to r; that is what we get.

Now, we are giving this result. g r is equal to… So long I took g r and just differentiate it noting that, r is this. I saw the differentiation can be… I mean the partial derivative of g with respect to x can be written down like this. dg dr, del r del x; and del r del x is x by r. What we are also giving is this side; g r is equal to p x x p y y. So, if I take the partial derivative of g r with respective x, then I do the next same thing on the right-hand side. So, I get del g del x; which is on one side equal to this; on one hand, equal x by r g prime r. This is also equal to p x prime x; that is, the derivative of p x x with respect to x multiplied by p y y. So, let me write it down again clearly here.

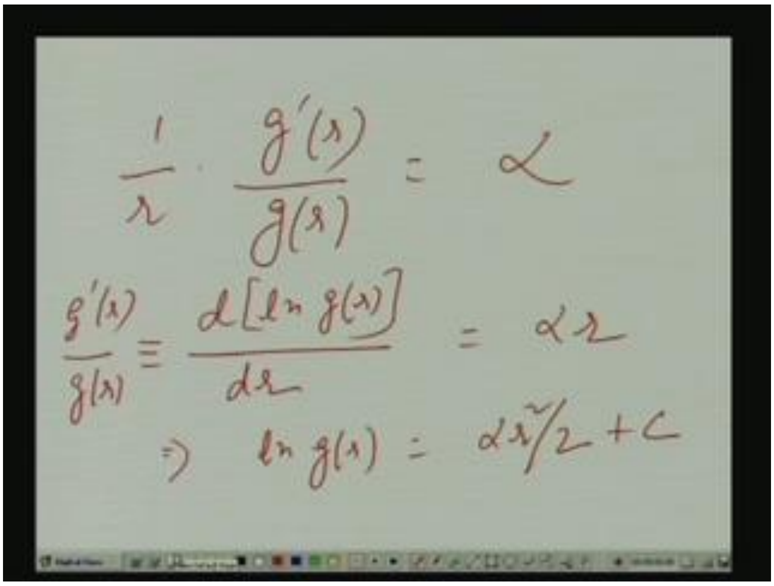
(Refer Slide Time: 30:14)



Or, for that matter, you can put a 1 here and bring x here. Divide both sides by g r; divide both sides by g r. So, this becomes 1 by r. And this becomes… What is g r? g r is p x x multiplied by p y y. So, p y y will cancel. So, this side becomes 1 by x. Then, comes something very interesting; some interesting interpretation that comes out of this equation. Consider this. This point I call capital X, 0. And this was the starting r. x, y; this x was coming here. Now, you see… Look at the left-hand side – this value. This is independent… This is just a function of r. And the radius of this circle is r; this is r. So, at any point on the circle, left-hand side retains the same value; which means if I am at this point, where x equal to capital X; where x equal to capital X, I get the same value. So, at x equal to capital X, this function; and therefore this side. So, this function gives a value; which is same as the value I was getting here or here or here or here. So, let that value be something.

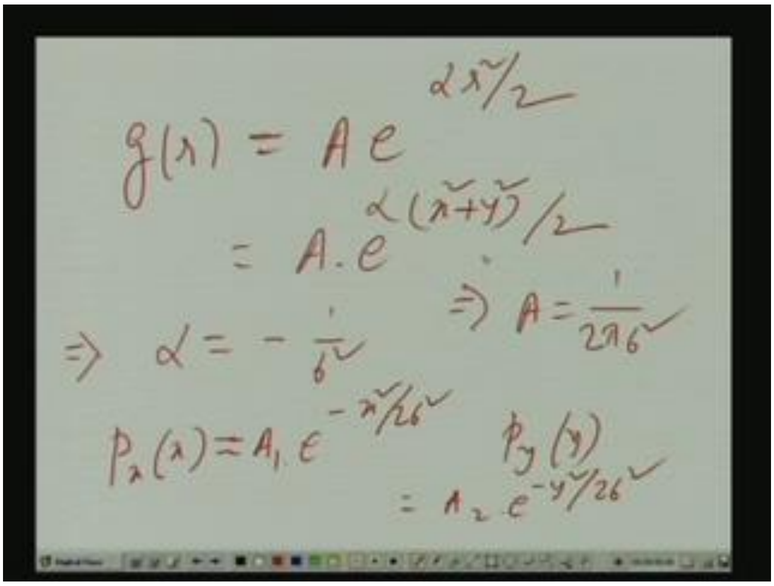
And now, for this x; at this particular point, for this x, right-hand side was giving a value. If I now go on moving on the circle, I get other values for x also. But, whether I am having x here or here; I mean whether I am taking the x for this point or for this point of or this or this point, this ratio has to be same, because left-hand side is not changing. And what is left-hand side? Left-hand side is given by its value at this point; which means this ratio does not change. Whether I am here, here, here, here, here – all these points; that is, whether I am moving on the circle like this, its value does not change. So, that means this is a constant.

(Refer Slide Time: 34:26)



If this be a constant, that means… where, constant equal to alpha. In fact alpha should be negative; that we will see; that means g prime r by g r equal to alpha r. g prime r by g r can be written as… that is, g prime r by g r is actually this. When this from our definition is alpha r. Obviously, d dr of ln gr means first we will differentiate ln gr with respect to g. So, you get 1 by g r, then g prime r. And that is equal to alpha r. So, ln g r will be the integral of this. g r will be exponential of that. So, what happens? In fact, I should have put the integral limits here – alpha r square by 2 say plus some constant say c. So, integral goes from some initial point to r; which means g r will be u to the power this entire thing.

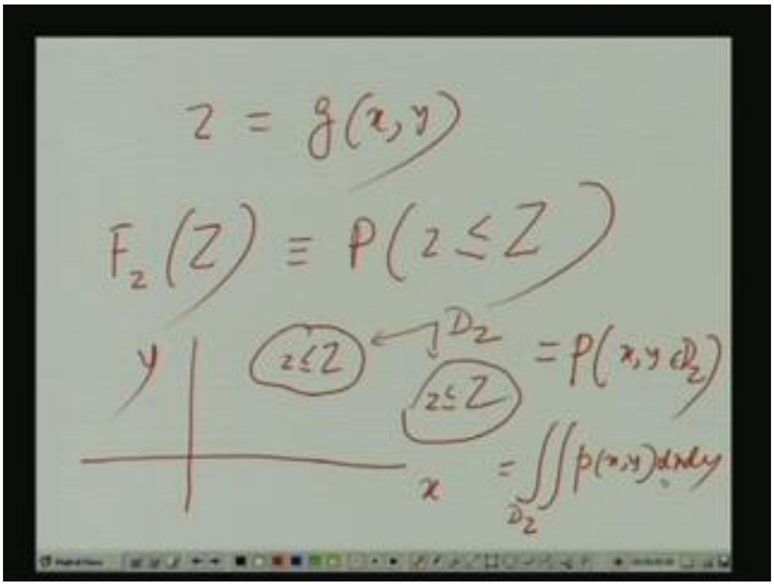
(Refer Slide Time: 36:52)



. And u to the power c will give us a constant; you can call that constant capital A; which means, g r is nothing but some constant A times e to the power alpha r square by 2. And we know r square is x square plus y square. So, it is nothing but A times e to the power x square plus y square by 2. So, you can easily see that, x and y – they are Gaussian, because the dependence is in this form – exponential… Then, power of 2 comes; and exponential of that is coming… So, alpha is obviously, minus 1 by sigma square. So, alpha actually is a negative quantity – 1 by sigma square; and individual densities. And what is capital A? Capital A should be such… that is, we have now got… – into this constant say A 1. A 1 we know what it is; 1 by root 2 by sigma. Similarly, p y y – say A 2. A 1 A 2 – product should be equal to capital A. So, A 1 is 1 by root 2 by sigma. And this is 1 by root 2 by sigma. So, it should be 1 by 2 pi sigma square; otherwise, it does not become a probability density function.

So, you see this form immediately suggests, that is, a product of two Gaussian probability density functions, zero mean. That is why x square is coming; no x minus mu whole square; just x square. And alpha is common; which means same variance. And the product of two Gaussian densities means joint density is nothing but product of individual densities; which means they are statistically independent. And that is true. That is already given. So, what I wanted to prove that, each is actually Gaussian random variable with zero mean. And both have equal variance. That is proved. So, this is an interesting property for jointly Gaussian variables.

(Refer Slide Time: 40:06)



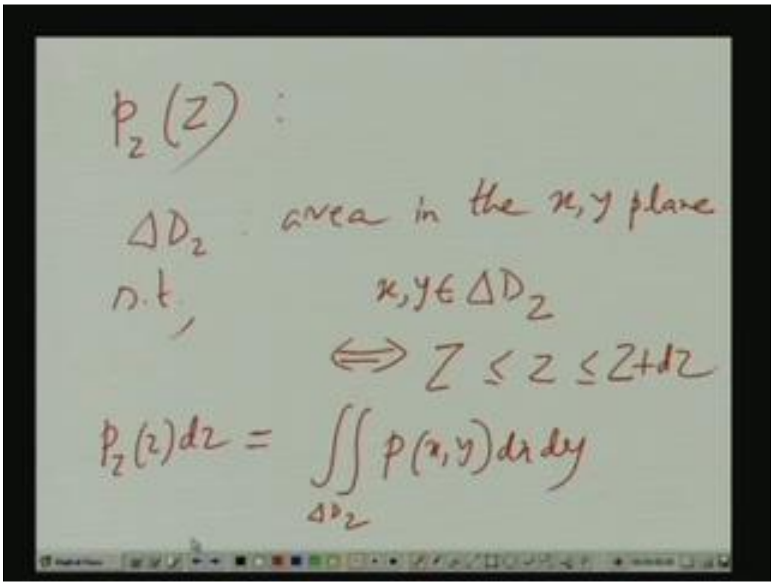
Now, when you are discussing just a single random variable, we consider an important topic – that was function of a random variable; that is, given x and its probability density are function g x. How to find y equal to g x? How to find out the probability density of y? So, we worked out various examples; we found out some relations and all that. We want to extend the same concept to the function of two random variables now that is, given a function z; given a variable z as function of two random variables g x, y. That time, previously, I took y-axis – one vertical axis for y and vertical axis for x. And I could easily plot the function. Here there are two variables.

So, I would suggest that, you consider one axis for x; another axis for y; and take the vertical axis. There is one which is going vertically up from this plane. Consider that to be z. So, it is a three-dimensional plot – three dimensional plot. So, I have to find out the probability density of z or maybe the probability distribution of z say to start with; that is, I have to find out F z capital Z. So, capital Z is some number given to us. But, what is F z capital Z? That is probability of small z taking value less than equal to capital Z. So, what I will do; in the xy plane…

Suppose this is the xy plane. I will find out the various regions; which has such that if you take x and y – this pair from within these regions, then corresponding Z as given by the formula – this function becomes less than equal to capital Z. This regions may be connected with each other, may not be connected; maybe one region is this; another here. Together I call it D z. These two together – I call it D z. This is x-axis; this is y-axis. So, within this, z is less than equal to capital Z; within this zone, z is less than equal to Z; together is D z. So, obviously, this event – that small z less than equal to capital Z – this is equivalent to this event; that x and y – this pair is belonging to D z. Remember I have collected all such regions in the xy plane – all such regions; for which their corresponding xy pair give rise to z satisfying z less than equal capital Z. So, that is why you can make this statement that, the event z less than equal to capital Z is equivalent to the event that, x comma y – this pair belongs to this region D z.

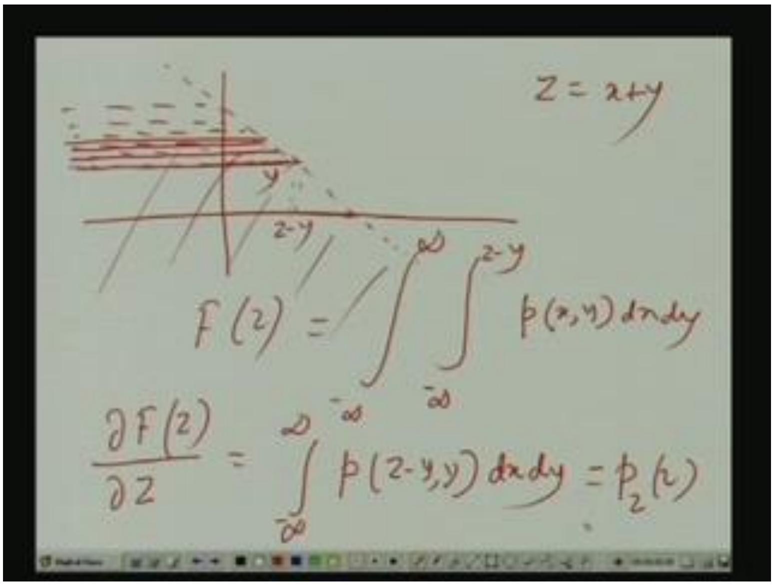
So obviously, the probability of z less than equal to capital Z is same as the probability of x comma y belonging to D z; which means this is nothing but p of x, y belonging to D z ((Refer Time: 43:09)) nothing but you can take the double integral about D z p x, y dx dy; small p x, y is the probability density. This is the probability density. So, you get the probability distribution of z from the joint probability density of x, y by this manner. And how about the probability density of z? This was probability distribution of z. How about the probability density of z?

(Refer Slide Time: 43:56)



That is how about… This should mean for this we call… It is delta D z be an area in the xy plane such that wherever x, y belongs to delta Dz, this gives rise to – this is equivalent to small z taking values between capital Z and Z plus dz. After all, what is this probability density for this random variable z? It is nothing but the probability that z takes values between capital Z and Z plus dz. That probability will be given by this probability density times dz; that is given that we have already seen. But, since z is coming from xy pair, I find out in this xy plane and area; we call it differential area – delta D z, so that whenever xy falls in delta D z, we have got this ((Refer Time: 46:05)) satisfied; that means what is this probability? It is nothing but same integral; but this time integral over delta D z p x, y dx dy. Let us take an example. If we cannot complete it, we will continue with the example in the next class.

(Refer Slide Time: 46:49)

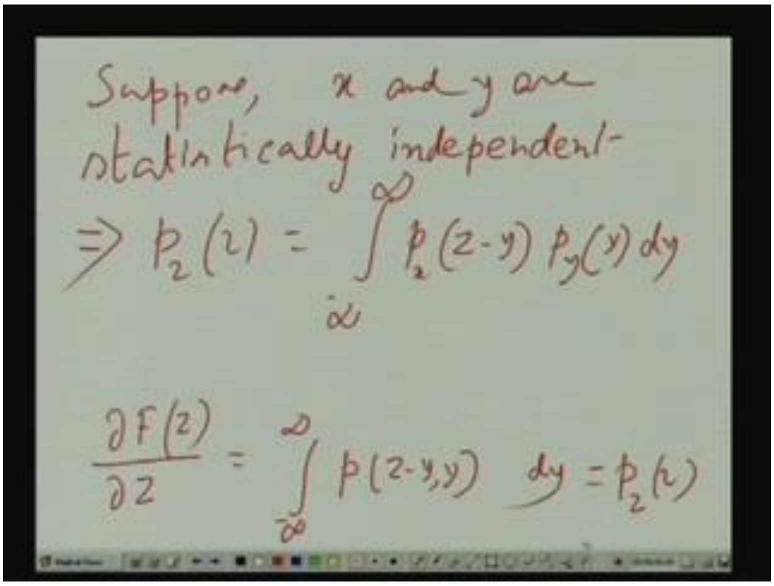


Consider a very simply case – straight line; z is equal to x plus y. Consider line x plus y equal to 1 or x plus y equal to z. You can see at y equal to 0, this is z; and at x equal to 0, this is z. So, as the value of z increases; that is, as this line moves further outward, value of z also increases. So, it is a plane going further up. It is a plane going further up, because as this intercept increases, the value of z also increase. That is a simple case. So, how to find out the probability distribution of z? So, what we do; we take a particular line say; if this value is y, then what is this value? z minus y.

So, probability of xy line within this is what? We integrate the density; say with respect to x, from minus infinity up to z minus y. And then we move y like this; that is, first, along this line – along a particular line from minus infinity to which point? z minus y dx along this line. But, this is only one particular line. Now, I have to move this line from minus infinity up to infinity. So, I will be here on this line up to this; next line, up to this; next line, up to this and likewise up to this, because of the limit. As y moves, this limit also moves leftward or expand this entire area – this entire area. So, there will be the other integral now – minus infinity to infinity; y from minus infinity to infinity. And this integral is a function of z alone. So, for a particular z, this will be the probability distribution.

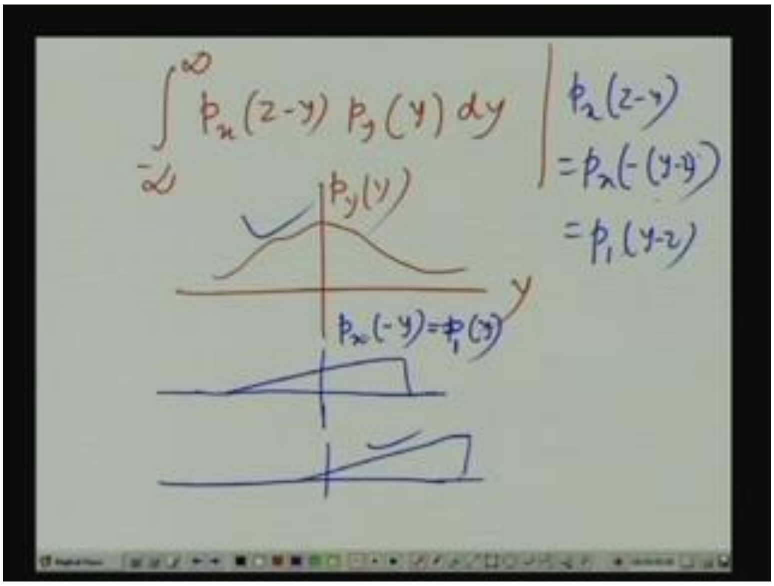
Next, how about the probability density? One way is just to differentiate this with respect to z. I do not know whether you know it or not; but if you differentiate it with respect to z, what you get? See the entire thing is a function of z alone. This integral given as it is; this integral goes, because this function when integrated with respect to x, it will be z put; you will get something with z. So, if you differentiate, this integral will go of course and x is to be replaced by z minus y. This integral will remain, because this is with respect to y. And if you differentiate, the outer integral remains. This inner integral will go. I repeat again – you are integrating this with respect to x and then putting the limit z or z minus y. So, if you now differentiate it back with respect to z, you get back the same thing. Only thing, x has to be replaced by z minus y. So, you basically get… So, this is a function of z alone. So, this gives you…

(Refer Slide Time: 52:20)



Now, suppose x and y are statistically independent; that means this probability density p x comma y is nothing but p x into p y. So, in this integral, it becomes p of… So, you can say p x just to differentiate now – p y y. And this integral… There will be no dx. Rather I told you, you are differentiating with respect to z. So, you should get back p of x comma y with x replaced by z minus y, so that integral with respect to x gone means dx gone. So, no dx here, only single integral. Look at this integral. This is a famous integral called convolution integral. There are two functions: p x and p y. p y – y comes here. In this case, this function y comes with the minus sign. And it is evaluated at z. So, z comes here. This is a famous convolution integral, which I think you are familiar with, because convolution comes very frequently in the discussion of linear time and invariant systems – both continuous and discrete. These integrals – it is corresponding to a continuous system. Actually, this integral means what? How to carry out this integral graphically?

(Refer Slide Time: 55:04)



In case you have forgotten, I rewrite the integral first again. This is for given z. And then p y. How to carry out this integral graphically? You plot say p y y – some function. And remember p x z minus y – you can write it as p x minus of y minus z. So, first, plot this function – p x minus y. You can call it – give it even a name p say 1 y. So, you call it say p 1 y; you get some function. Maybe I should not draw the same shape for both the functions; maybe suppose it is like this. But, we want to evaluate this function at z. So, it should be p 1 y. This is… So, it will be p 1 y minus z. So, you have to shift it to the right by z if z is positive or shift it to the left by z, if z is negative. So, you shift it by z. And then you multiply this function and this function. See the overlapping area; multiply it to an integral – integrate. That will give rise to this integral. This is a typical way of… That is a very standard way rather of carrying out convolution integral.

So, in the next class, we will carry out some examples of this; that is, two random variables: x and y. And they are having particular probability densities for each; you form a random variable z as a summation x and y. What sort of shape? What sort probability density function you have for z? That we will evaluate by carrying out this convolution integral. So, that is all for today. We start from here in the next class.

Thank you very much.

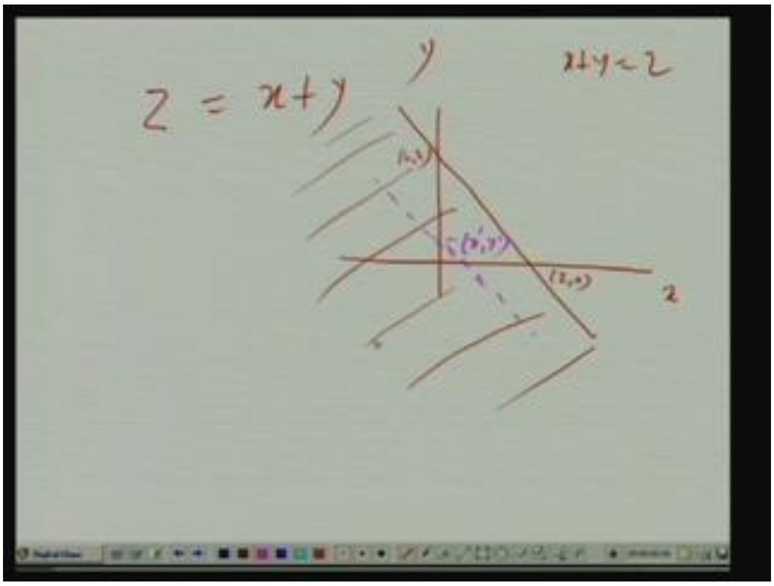
Preview of next lecture

Lecture – 14

Function of Two Random Variables (Contd.)

In the last class, we started with this topic – function of two random variables. We continue from there. In fact, I take up the same example, which I ((Refer Time: 58:48)) upon in the last class for some reason.

(Refer Slide Time: 58:52)



That is, suppose if it is given that, where x and y are two random variables and there is a summation; you have to find out… So, Z is a function of x and y. So, you have to find out the probability distribution and also probability density of this variable z. Now, how to do that? We recall…