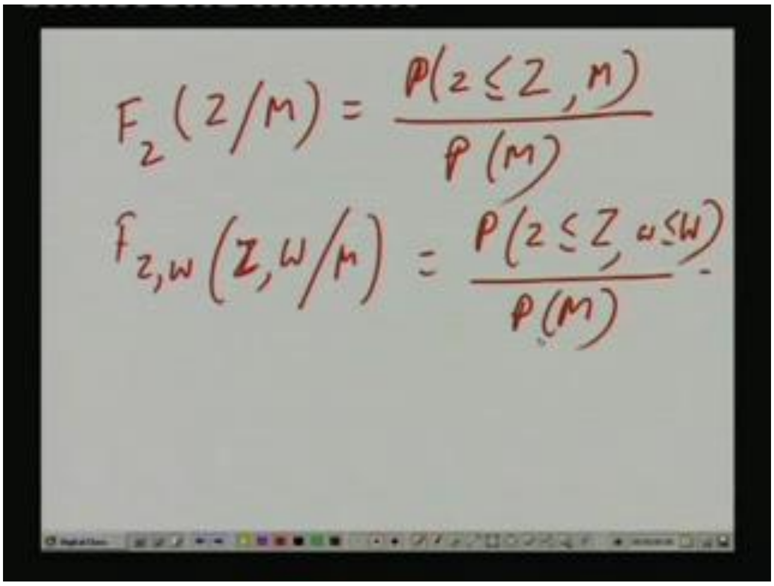
Probability & Random Variables
Prof. M. Chakraborty
Department of Electronics & Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 20
Joint Conditional Densities (Contd.)

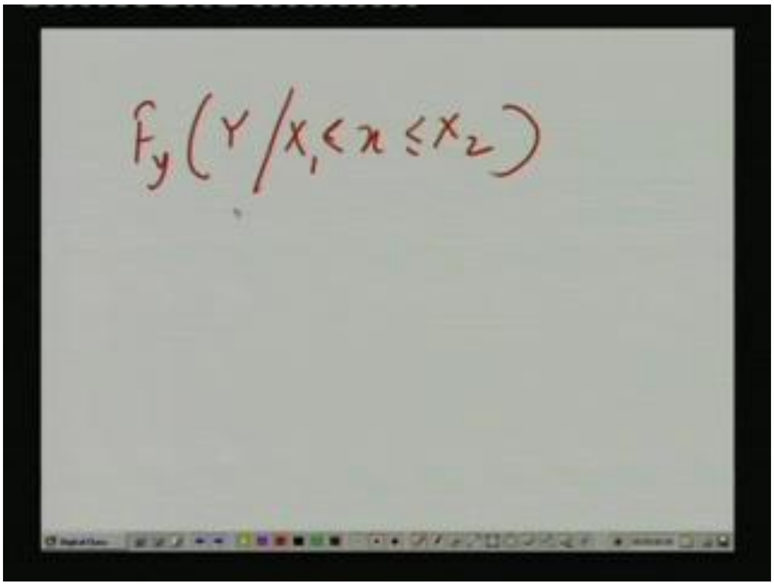
So, in the last class, we were discussing this conditional distributions and densities. Just to recap, what we did last time were something like this.

(Refer Slide Time: 01:03)



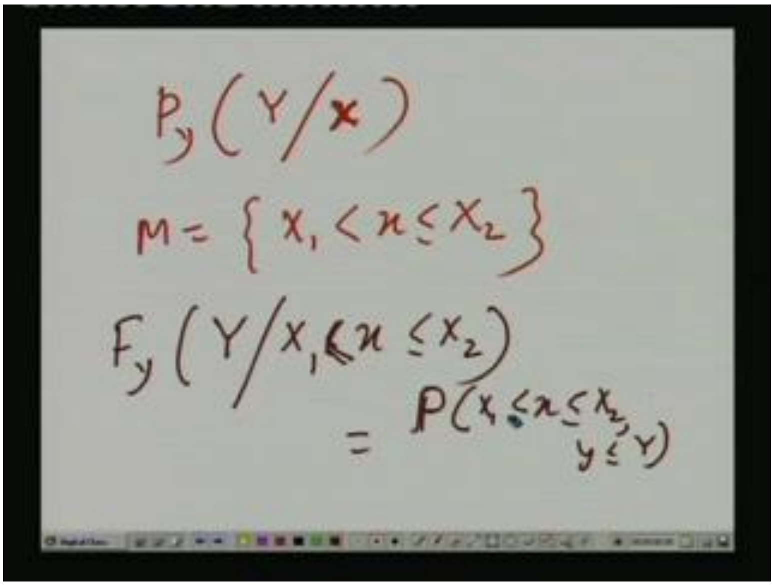
We said that, this distribution giving z as a function of x and y, we said that, by M; M is an event. This was nothing but P of… And also, if instead of z, we have got two functions of two random variables Z and W. Then also the same thing – same logic capital Z capital W by M was… These are the thing.

(Refer Slide Time: 01:59)



And then what we did? We took some special case like we found out this distribution. F y; we took z to be y; z is a general function of x and y. But, in that example, we took z to be y. So, this subject to this thing – x less than equal to X 2 greater than Y by X 1. This we found out. And once you find it out, you can differentiate it with respect to y. In fact, it should be differentiated with respect to capital Y, because it will be a function of capital Y. That will give you the probability density – conditional probability density. There is condition to this condition – x less than equal to X 2 greater than X 1. We did that.

(Refer Slide Time: 03:00)

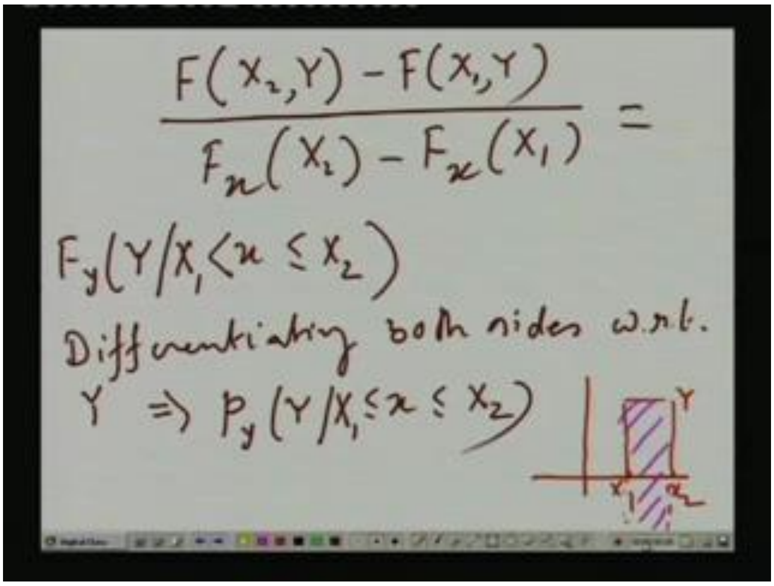


Now, today, we will be taking another special case, which is a very important case actually of this type, that is, how to find out, we say conditional probability density P y – capital Y given x. A particular value x is given; maybe you consider capital X. X is given; what is the probability density; what is the conditional density of capital Y. Earlier, I had a range of x; x is lying say within x 2 and x 1 or x is less than equal to x 2 up to minus infinity. But, now, in this case, x has a particular specific value – capital X. And subject to that, what is the probability of Y taking capital Y, where x and y both are random variables. You have to do that. We will be doing this in a limiting way. We define the event M to be this first.

Then, what is this distribution? First, you find out the distribution; distribution of y; that is, y taking value up to capital Y subject to the same condition, that is, x less than equal to X 2 greater than equal to X 1. But, this we know. After all, this is a probability… There is a probability; that is, what is a probability? See the problem is densities are not probabilities as such; they are the derivatives of probability distributions; but probability distributions are probabilities. After all, what is this conditional probability? What is this distribution? This distribution means what is the probability of small y taking values less than equal to capital Y subject to some condition. Now, what is a condition? x less than equal to capital x 2 greater than X 1 – greater than X 1. So, these are probabilities.

And, we know very well that a conditional probability can be written like a joint probability first divided by probability of the condition; that is, nothing but F y – not Fy; In fact, you can write not F; it is a probability actually. What is a joint probability of x lying within the same range and y less than equal to cap… not much space here; I rewrite y less than equal to capital Y. So, joint probability of x falling within this range between X 2 and X 1. This should be actually strictly greater than… And y less than equal to capital Y divided by the probability of x less than equal to X 2 greater than X 1. This comes from our very basic notion of conditional probability that a conditional probability is a ratio of the joint probability divided by the probability of the condition. This works for the distribution. But, density as such is not a probability function; it is the derivative of a probability function. What is that probability function? Distribution function. That is why to use this result, we start with the distribution function – probability distribution. It is a conditional probability distribution; we write it like this as a ratio of two probabilities. From here we will be going to the density. Now, this you know. Let me erase some part.

(Refer Slide Time: 07:19)

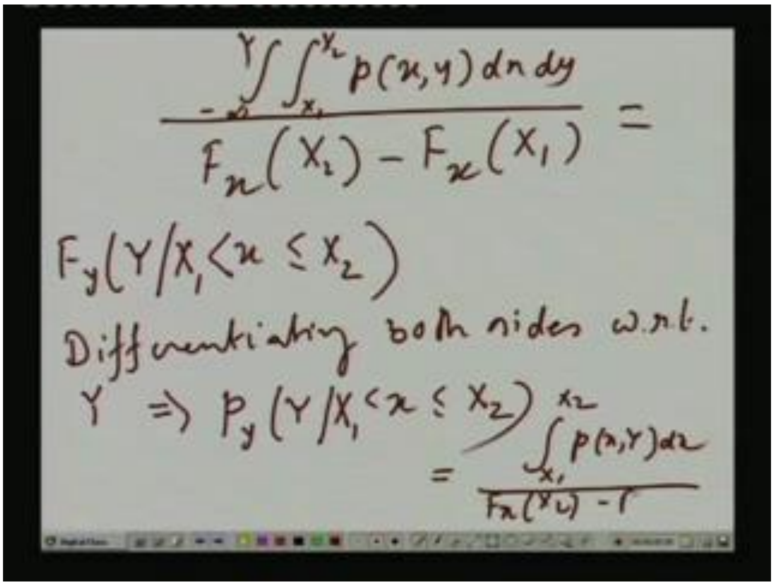


Now, the numerator can be written like this – F; F is the joint probability distribution of x and y. I can as well put a subscript x y here; but I am not doing it; it is assumed; it is implicit. So, just let me draw it and then you see the correctness. After all, what is this difference? In both the cases, small y is taking the value capital Y. So, here it means the total probability of y taking value less than equal to capital Y jointly with x taking value less than equal to capital X 2; same for here; instead of X 2, we have X 1. So, this difference will give you the joint probability of y taking value less than equal to capital Y, but x falling within X 2 and X 1. So, that is how we replace the numerator by this. And denominator is very simple; it is available in terms of the probability distribution of x. And what was it? This was equal to… And this was equal to F y capital Y divided by the condition; that is, x less than equal to X 2 greater than X 1. This was equal to this. Maybe you can put this equal to here; this was equal to this.

But, as I told you, I am interested in the density. I have done with the distribution. I am interested with the density. And right now, this is available subject to this condition that, x is lying within a band given by the limits X 2 and X 1. But, finally, we have to make it… I mean the band width has to be 0; that is, small x should be equal to some particular value capital X. And subject to that condition, I have to find out the density. But, now, with this condition only, that is, x less than equal to X 2 greater than X 1, we have got a conditional distribution. From this, we first derive the conditional density by deriving it by – differentiating it with respect to Y; that is, what you get here. Differentiate both sides Y. On this side, I get p y capital Y subject to the same condition.

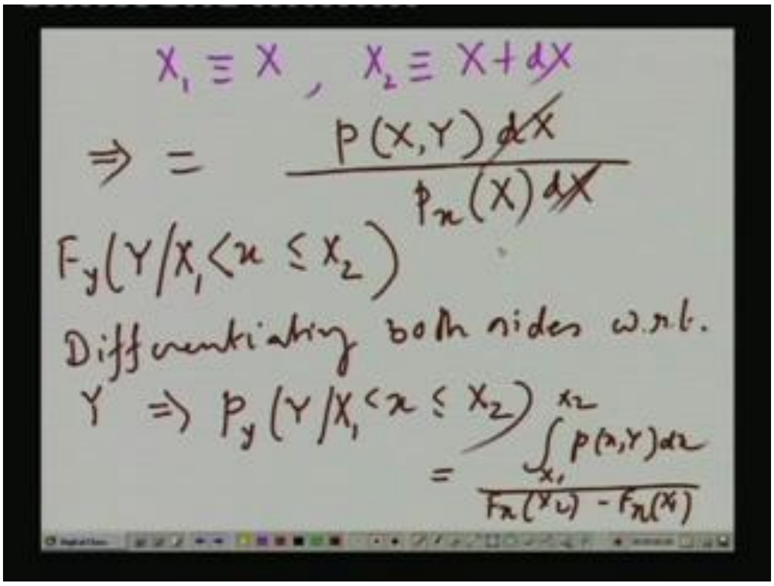
Now, on this side… On this side, what I get? There is a question. Now, there is not enough space here. Firstly, you see denominator remains as it is, because we are differentiating with respect to Y – capital Y. So, denominator does not change. Only problem is the numerator. What is the numerator after all? You can see what the numerator means; I can just draw here small part. There is suppose there are two limits: one is X 2; another is X 1; and this is this – capital Y. It goes like this. Then, the numerator here means the probability of the pair x, y falling within this zone; that is, small y ranging from minus infinity below up to capital Y; and small x starting at X 2, but going down up to X 1; of course, without touching X 1. It is less than equal to X 2 greater than X 1. Within this range, as given by this shaded rectangle with one side going up to infinity minus infinity, what is the probability? Now, what is the probability here? It is nothing but the joint density of x, y integrated over this area; that means you can say that, I can replace the numerator. This numerator I can replace by this.

(Refer Slide Time: 13:14)



dx dy… In one case, it goes up to from X 1 to X 2; in another case, minus infinity to Y. Now, we are differentiating; let me erase this figure now. So, I get some space. So, we differentiated both sides with respect to capital Y. For the left-hand side, by definition, we get the conditional density of Y – I mean capital Y given by this condition – x lying within less than equal to X 2 or greater than X 1. On the right-hand side, if you differentiate with respect to capital Y, you see only this integral – outer integral has the limit from minus infinity to capital Y. So, if you differentiate, from our knowledge of calculus, we know what we get is nothing but the inner integral itself; only thing is just variable y – inner integral with respect to x, that is dx; dy gone. And instead of small y, we will have capital Y. If you differentiate with respect to capital Y, you will get the same thing – the inner integral; and small y takes the capital Y. So, we can then write that, this will give you X 1 by X 2 p x capital Y dx divided by of course what we had earlier – F x X 2 minus F x X 1. Now, let us confine ourselves to this range.

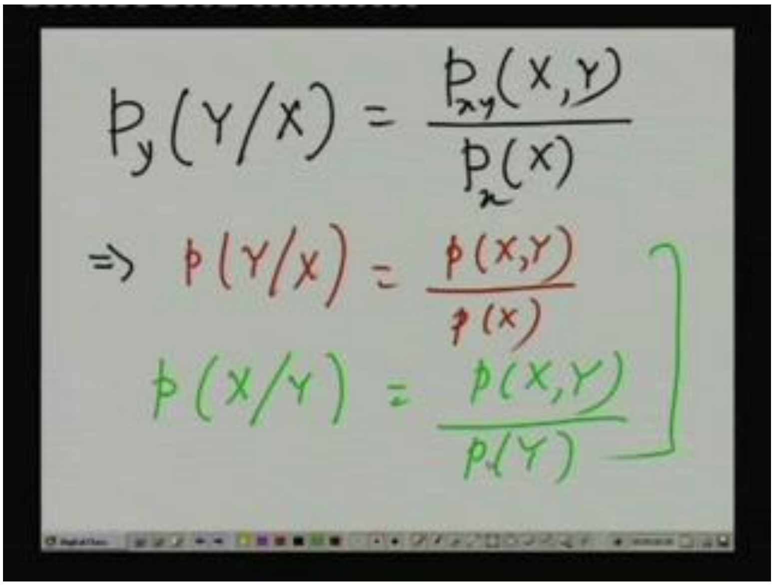
(Refer Slide Time: 15:46)



Let X 1 be equal to some capital X and X 2 is very close to X 1 – X plus dx. If dx is very small; that if they are very close to each other — infinitesimally close to each other, then you know this integral that can be approximated by p X comma Y dx – p capital X rather – p capital X comma capital Y into dx, and how about this denominator? Denominator is nothing but… After all, what is the denominator meaning? It actually was meaning – the total probability of x lying within the range less than equal to X 2 greater than X 1. Now, if X 2 and X 1 are very close to each other, then this difference will be nothing but probability density of x, that is, p x X – capital X because X 1 is capital X now – times dx; that means this leads to equal to… That this is equal to – carries here p capital X, capital Y dx. That happens to the integral. Limits are capital X and capital X plus dx. And they are very close to each other. So, this integral is nothing but p of capital X comma capital Y times dx. And the denominator we have got – the probability density of X at capital X times dx. When X 1 and X 2 are very close to each other and the difference is only dx, the dead probability of x lying within that range between capital X and capital X plus dx, is nothing but p x X; that is, the probability density of x evaluated at capital X times this dx. And the 2 dx – they cancel; which means we get just this ratio.

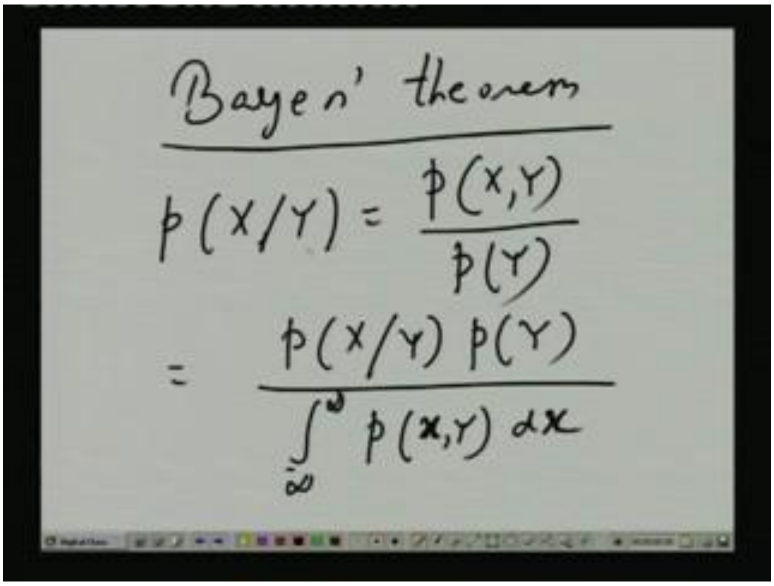
Now, if you make dx tend to 0; then X 1 and X 2 coincide. We have only capital X. X 1 and x 2 coincide; we have only capital X. But, this ratio is independent of dx; that means we can now write that, p of y within bracket capital Y by capital X; that is, small x taking the value capital X subject to that condition y capital Y, that density is nothing but ratio of these two probabilities; where, the ratio of these two probability densities. This is a joint density between X, Y. You can also put… If you want, you can put it X, Y here. Joint density between XY evaluated at capital X comma capital Y divided by the density of X evaluated at capital X. So, now, I write this result neatly.

(Refer Slide Time: 19:03)



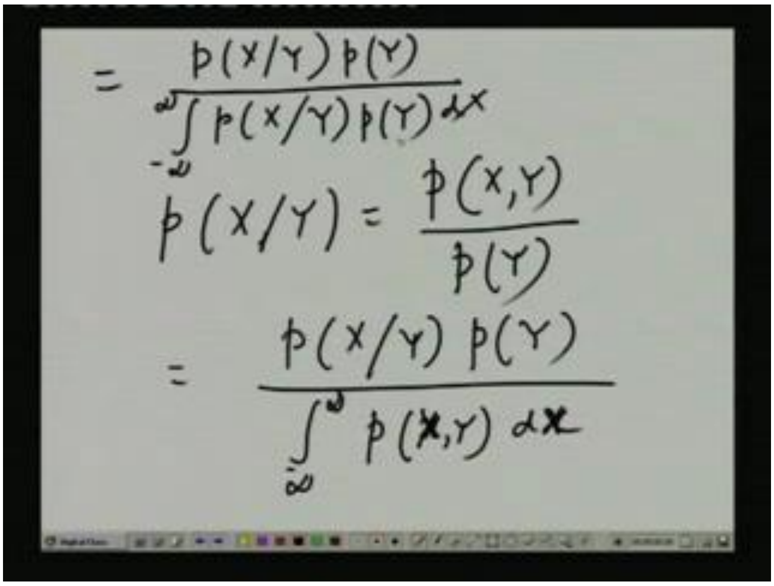
We write like this actually. I should put xy here; I should put x here. But, I suggest that, if I just do not use subscripts – p of X comma Y, p of X and P Y by X; even then we will only be confused thinking that the three p's – they are same, because here p X comma Y obviously means this is a joint density function. p of X means it is just a density function of x. And p of Y by X means it is a conditional density. So, obviously, it will imply that, these are three different probability functions – probability density functions. In that case, you can simply drop the subscript and just write p Y by X… By the same token then you can also write…Further here also, we can extend the Bayes' theorem concept.

(Refer Slide Time: 20:45)



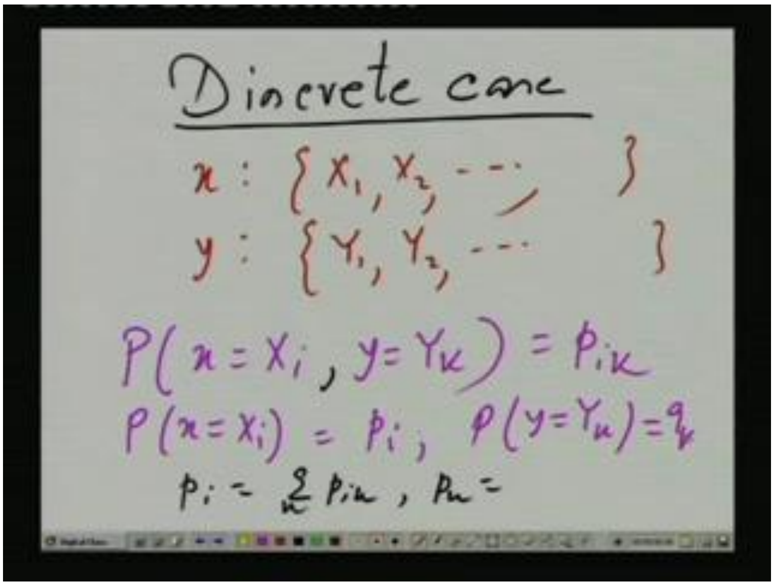
We have seen p X by Y – this is what – p X comma Y divided by p Y. We have seen this – p X by Y; this is p X comma Y; this is the joint density function between X, Y. And this is a density function for Y; this you have seen. And then this you can write as p of X by Y. This we have seen just a while back. Numerator can be written using the just previous result. And p Y – it is a joint density function of Y. So, you can also write it as. This also you have seen earlier – integral of p X, Y dx. And p X comma Y as we know, we can write as p X by Y into p Y; that means you can write it as… Let me erase this part.

(Refer Slide Time: 22:54)



You can write as… And again p X, Y – we write in the same manner – dx. In the discrete case, you can easily extend this result. So, maybe we cannot… We do not need to… Or maybe we will just write down for the sake of completeness, then discrete case.

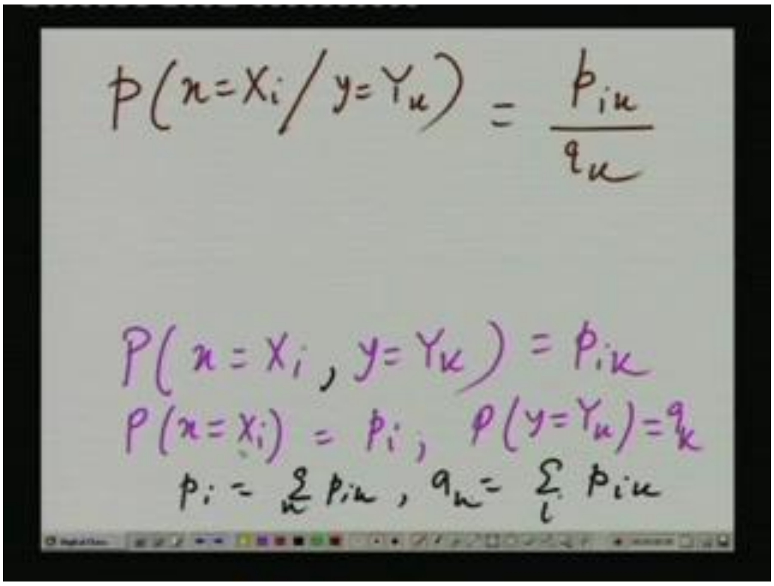
(Refer Slide Time: 23:52)



In the discrete case, actually the same argument holds. Here x and y – they are random variables – jointly random; but they are taking some discrete values. x can take some discrete values from this say X 1, X 2 up to say… It can go up to infinity like this and… Or, it can go up to some finite number of terms. Similarly, y – it can take values from Y

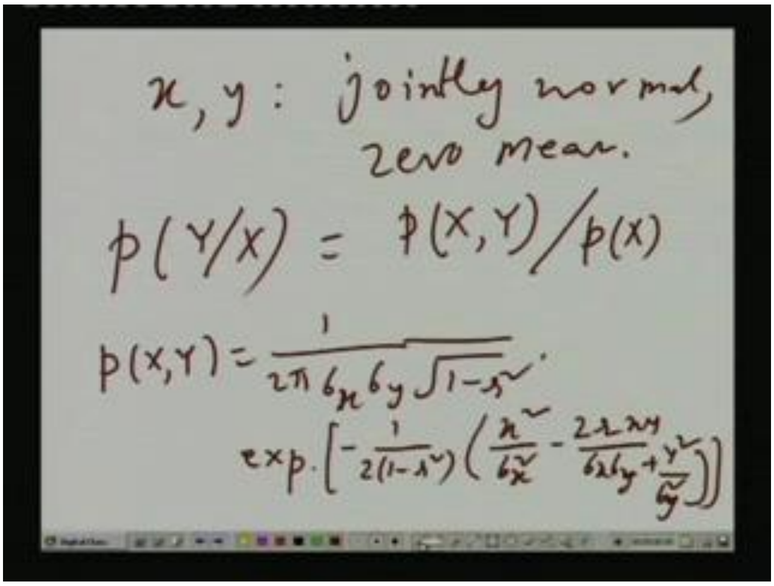
1, Y 2 – either going up to infinity or some finite number of terms; in this case, because it is very simple here now. The probability… There is no question of density here because this is not continuous. Say X i given y equal to Y k. This we call p i k. p of x equal to X i is p I; p of y equal to Y k, is p k. And obviously, you know that, p i is nothing but summation of p i k over all k; that is, holding i fixed, take all possible cases – y taking Y 1, Y 2, Y 3 up to… Or, I mean whatever be the terms we have at the end. Add up all the probabilities; that will be the probability for x taking X i and p i; that means you have seen that, p i is nothing but summed over k; p k is nothing but… Instead of p, it is better we replace it by some other symbols say q, because p is already used here with one subscript. With two subscripts, p – no problem; we understand that… Just a minute; I just made a mistake in the notation; it is not conditional density, it is a joint density – joint probability; there is a comma. And let us call it q k. So, q k gives again summation of p i k over i.

(Refer Slide Time: 27:08)



In this case… In this case, p of x taking a particular value say X i subject to y taking Y j, Y k will be what? Again, it is a probability. So, it is a conditional probability. So, it will be a ratio of joint probability, that is, x taking X i by y taking Y k – divided by the probability of y taking Y k. So, we are not going into that detail; it is simply becoming p i k divided by q k, likewise. So, that is just discrete case; it is nothing but whatever we had in the continuous case; it is just the same logic. But, it is lot simpler here because there is no integral; they are discrete summations; no dx, no density and all that; just probabilities. Problem comes only when it is continuous, because then you first define the probability distribution, which is a probability function. But, then we take the derivative of that, which is the density function. So, density is not actually a probability; but it is a derivative of the probability function.

(Refer Slide Time: 28:28)



We just take an example now. Suppose x and y – they are jointly Gaussian – jointly normal, zero mean. I have to find out what is Y by X; that is, given small x equal to some value capital X, what is the probability density of Y for capital Y? This we have to find out. For that, we use the same formula; that is, this is nothing but… We will be using the same formula this joint density is Gaussian we know. So, let us do some manipulation over that. And we can easily see what it will be. And of course, if they are jointly Gaussian, we have seen earlier that, X and Y – both are individually Gaussian also.

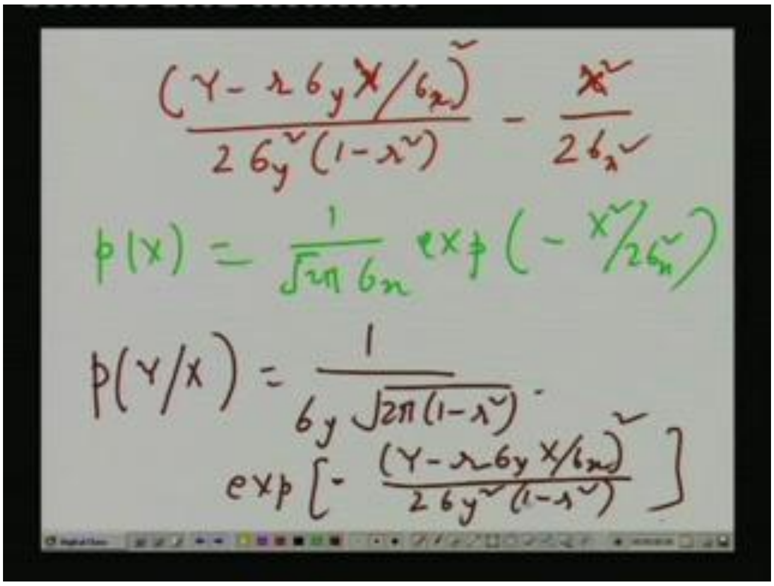
Now, since they are jointly Gaussian, we just go through their – the joint density function. Even though we have distributed complicated functions, we will just write down. For the zero mean case, what was P X, Y. With zero mean, we know that, it was 2 pi sigma 1 – sigma one is the variance of x or you can write sigma x – sigma x sigma y square root 1 minus r square. This we have seen already – times exponential what? Minus 1 by twice 1 minus r square; r is the correlation coefficient times x square by sigma 1 square minus twice r xy by sigma 1 sigma 2 plus y square by sigma 2 square.

This was the function – joint density function for two random variables: x and y, which are jointly normal and have zero mean.

Now, here let us consider this term; x – what is sigma… Instead of sigma 1, I should write sigma x; this is mistake I am making. What is sigma x? Sigma x square – it is the variance of x; x itself is Gaussian with zero mean; y itself is Gaussian with zero mean; variance of x is sigma x square; variance of y is sigma y square. And when they are… you take the joint density; they are jointly normal also with r being the correlation coefficient. Then, the joint density takes this form; this we have already seen, is just a repetition of that ((Refer Time: 31:50)) or previous observation or previous results. So, this is p X comma Y. And in this p Y by X, this is the expression.

Now, here we just need to do some algebraic manipulation of this quantity inside this argument; that is, inside this exponential thing, we have got this expression – minus 1 by 2 – 1 minus r square; and then within this bracket, this quantity – here we will be doing some manipulation. Since it is simple algebraic manipulation, I will not do the derivation; I will show the result and then you can verify.

(Refer Slide Time: 32:56)



This argument ((Refer Slide Time: 33:02)) you can write like Y minus r sigma x x by sigma y whole square divided by twice sigma y square – 1 minus r square minus x square by twice sigma x square. Now, let us verify; this is capital X. You can verify. And we will verify that, this is nothing but what we have in this argument. Actually, I am telling you what we are trying to do; we are trying to extract… From this entire thing, we are trying to extract out this term – x square by twice sigma x square. This is our purpose. How do we do that? You can just verify. If you add… If you subtract this from this; if you carry out this algebraic subtraction, you will indeed get this; how? The denominator 2 comma twice sigma y square sigma x square; which comes here also – 1 minus r square; that comes, then if you expand it y minus… If you take this term y minus r sigma x capital X by sigma y whole square multiplied by sigma x square, then what you get? Just a minute.

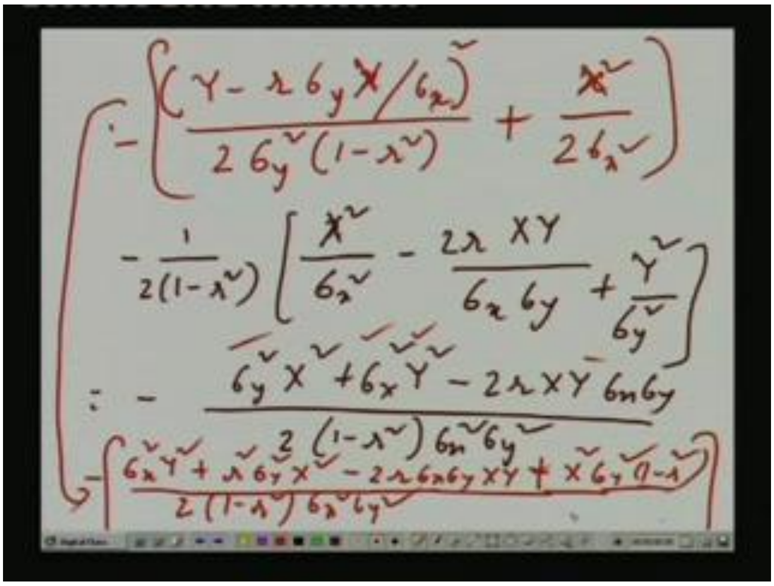
So, you get firstly, Y square sigma x square. So, that term will be here. If we add Y square… If you find Y square sigma x square; so that term comes. Then, here minus twice Y r sigma x X by sigma y – and this multiplied by sigma x square. Actually, this is a small exercise; I would say that, I would leave it to you for finding out, for checking this. And since we are not giving homework, let us verify that, this argument of exponential actually gives rise to this. If you cannot do, I will do in the next class. This is just some algebraic ((Refer Time: 36:16)) this and nothing else. Now, what we do? If this is true, what we do from here is this that, we have to divide this by the probability density of X.

Now, the probability of density of X – again the exponential function comes; and it is the argument. This term is there – minus x square by twice sigma x square. That goes; that is, what is probability density of X? 1 by root 2 pi sigma x exponential minus X square by twice sigma x square. And you have to divide p X comma Y by p X. If you make this division and you have already seen that, the argument of exponential can be written like this; we are assuming this is true; then this term cancels. It is minus X square by twice sigma X square is present here also, here also. So, that cancels; which means… And what happens to the outside? 1 by root 2 pi and 1 by 2 pi. So, actually, we get 1 by root 2 pi term. Then, 1 by square root 1 minus r square remain; but sigma x and sigma x cancels; only sigma y remains. So, now, it is very simple; I can erase this part.

Under this case, is very simple now – sigma y square root 2 pi 1 minus r square into exponential… Here this will come up – this first term – minus Y minus r sigma x capital X by sigma y whole square divided by twice sigma y square 1 minus r square. I think I made a mistake here. This is why it was not coming. This should be sigma y… So, I make the correction here also – sigma y by sigma x. We will verify this result today only instead of… because this is small result. But, look at this function. What is this function? I say that, it is once again a Gaussian density function. So, conditional density also is Gaussian. What is the mu then? What is the mean? Mean is this quantity – r sigma y x sigma x. And what is the new variance? That is sigma y square times 1 minus r square. So, on the outside, we have 1 by root 2 pi; then square root of the variance – sigma y; and square root 1 minus r square – that is coming. So, it is Gaussian. This is Gaussian with mean; I repeat r sigma y capital X divided by sigma x; and variance – sigma y square times 1 minus r square.

And now, if it were not zero mean; then what happens – instead of capital X, instead of capital Y, in that overall expression, we would be having X minus mu x and Y minus mu y. But, the rest from the calculations remain same. So, in this case, what we would had? Instead of Y, it was Y minus mu y; instead of X, X minus mu x. That is the only difference. Now, we just for the sake of completeness, we verify this result. I said that, you can take it as a homework; but this no need. So, we just have to write down that expression again; that is, just the quantity, which I had earlier within the… I mean as an argument of exponent.

(Refer Slide Time: 42:26)



That was 1 minus 2 1 minus r square – within bracket, x square by sigma X square – this is capital X – twice r XY divided by sigma x sigma y plus Y square by sigma y square. Now, this is this. It is actually algebra – simple algebra ((Refer Slide Time: 43:07))

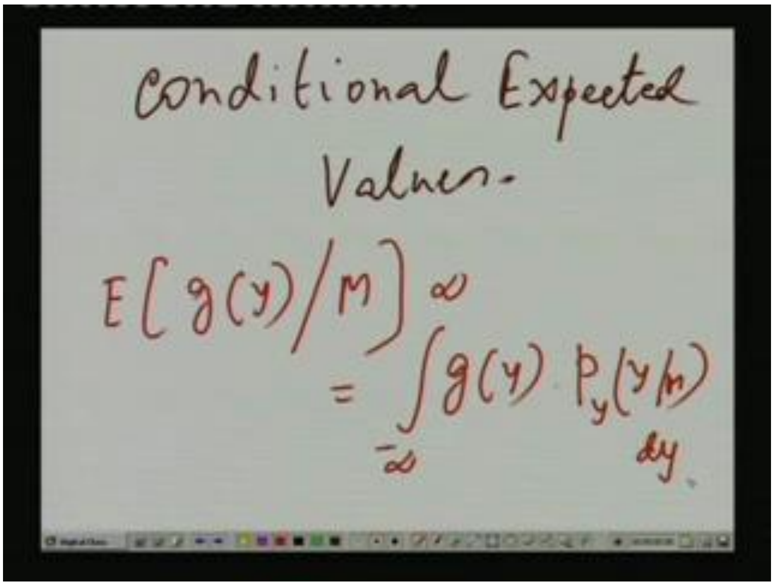
Verify; you can verify either starting from this side and going to that side or the other way. Y square sigma x square; I mean what happens to these terms here actually? Minus twice… because in one term, you get sigma y square x square; another term sigma x square y square, and in another case, twice r XY sigma x sigma y. On the other hand, if you start with this, you break it up. Let us verify what you get minus capital X square… Here it will come.

First, consider this part, this side. Here what you are getting; you can easily see denominator remain same in both cases – sigma x square sigma y square. If you break it up; square it up and multiply by sigma x square; you get… This minus sign… Actually, there will be a minus before this. So, do not bother about the minus sign here. I will come to that. Sigma x square y square – one term; this is present here. So, this is present – this term – sigma x square y square, then another square of this – r square sigma y square x square by sigma x square. So, that cancels with this sigma x square. So, you will get r square sigma y square x square divided by sigma x square. But, that cancels with this sigma x square. That minus twice r sigma y xy by sigma x multiplied by sigma x square. So, you will get twice r sigma x sigma y xy. This is what you get from this side. And from this side, you get just minus x square sigma y square.

Now, you can see that, the expressions – this is present – sigma x square y square present. We have already seen then twice r x y sigma x y; that is present – twice r xy sigma x sigma y. And this is another term here – minus x square sigma x square comma 1 minus r square. Actually, there should have been x square sigma y square 1 minus r square. So, 1 minus r square should be present. x square sigma y square 1 minus r square. This will be present. So, you can now see – this term and this term cancels; x square sigma x square r square. And here you take…

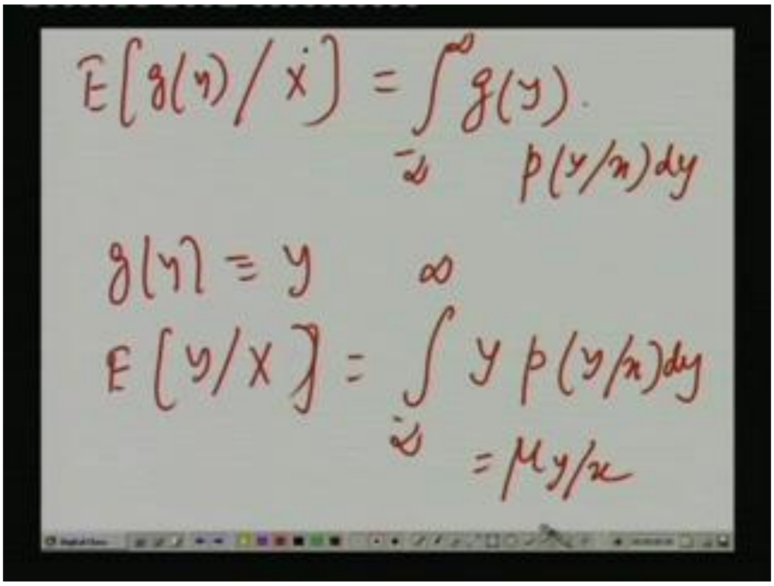
According to me, this should be plus; yes, this should be plus here. This will present. So, this cancels now. Easily seen, this minus r square x square sigma x square – that cancels with this and you are left with x square sigma x square, which was present here. So, all the terms are present. They are same. Only thing is minus… I took the minus sign already. So, this… There should have been minus sign here. So, these two are same; out of which, minus of x square by twice sigma x square is kept aside. That cancels with the denominator. But then in the denominator also, we get another probability density function, that is, for x; which is Gaussian. So, within the argument for the exponent there, we will have minus of x square by twice sigma x square. That cancels. So, this is seen now. So, this is done.

(Refer Slide Time: 48:51)



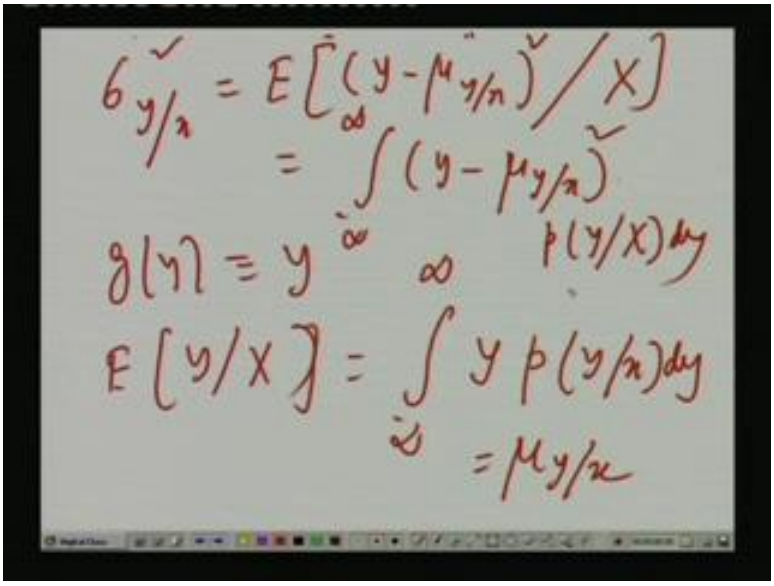
Following the similar lines, we can also define conditional expected values. Absolutely similar line of argument; that is, suppose we consider instead of random variable y – a function of y g y subject to a condition m, this is nothing but… This is nothing but by our previous definition… What is the probability y? I mean this g y multiplied by the probability of… You can write this y by M; that is, even though we are taking expected value of not y, but function of y; to carry out the expected value, you have seen earlier; you have to just take the function multiplied by the corresponding probability density. The probability density is for y subject to M, which is this. And as a special case…

(Refer Slide Time: 50:28)



As a special case, E of g y; given X that is, x equal to X. This will be nothing but… This is the condition is this M; M is now x equal to capital X. By the same logic, it will be g y multiplied by p of y by x dy. As a particular case of g y, you can take g y to be y itself. When g y is equal to y; then in fact, we get E of y by X. This is small y; this is capital X; X is subjected to a particular value, that is, capital X; and this y is variable here. So, this is nothing but… This we call conditional mean of y. Similarly, you can define conditional variance also.

(Refer Slide Time: 52:22)



That is, you can define… This is nothing but expected value of y minus mu y by x whole square subject to the condition capital X; that is, x is taking the value capital X. And once again, this will be nothing but… So, that completes this topic of functions of random variables. I think we have covered this in great detail. From here… I mean so far we considered function of one random variable – first one random variable, then one function of one random variable; then two functions; then two random variables; then one function of two random variables; then two functions of two random variables. So, this has generated… This has produced the background for extending this treatment to a random sequence; where we have got just not two random variables, but in the arbitrary number of random variables forming the sequence. So, this is the topic, which we will take up next time.

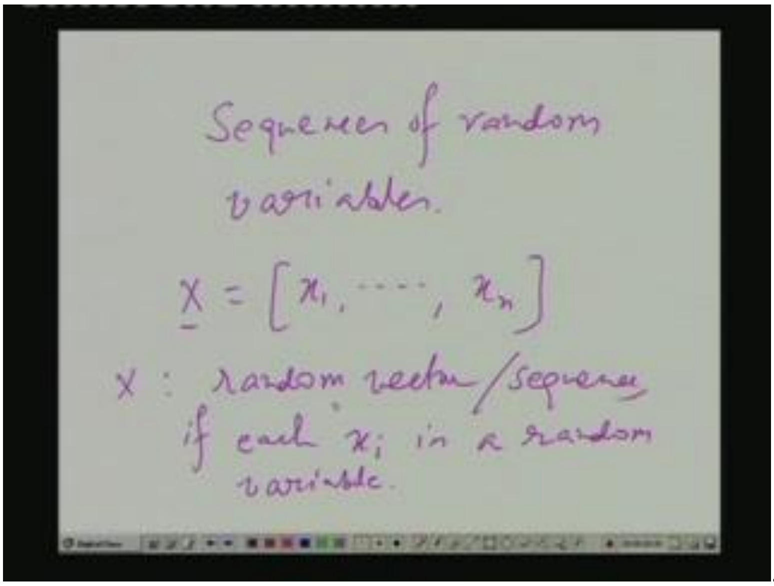
Thank you very much.

Preview of Next Lecture

Lecture – 21

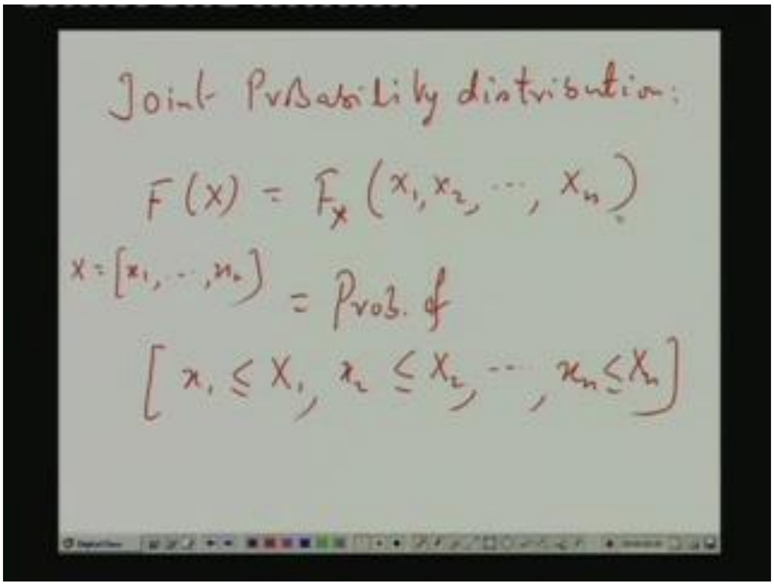
Sequences of Random Variables

(Refer Slide Time: 54:08)



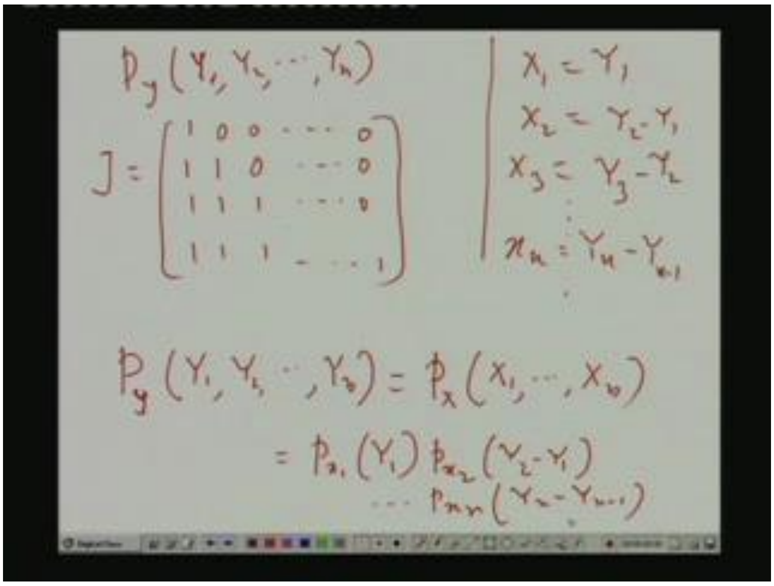
So, today we discussed an important topic, that is… In fact, we start this topic – sequences of random variables. You see we started with first, a single random variable and we considered a function of a single random variable. Then, we generalized it to two random variables. So, you can then say that, it is a sequence just of two elements – two random variables – two-element sequence; where, we considered a function of two random variables; then two functions of two random variables and all that. That whole treatment will now be generalized; where we will be considering a vector say X as x 1 dot dot dot dot x n. This X will be called a random vector if each element x 1 to x n is a random variable; that is, X will be called a random vector or say sequence if each x i is a random variable. So, that you see; earlier we considered two random variables. So, we had a vector of two elements: x 1 and x 2; we call it xy, because this is a different notation. But, now, it is more general in such random variables. So, from the ((Refer Time: 55:48)) side you can see that, if you talk of the joint density or joint probability distribution of X, it will be basically a function of n variables – x 1 to x n.

(Refer Slide Time: 56:06)



So, we can define joint probability distribution… Actually, I should write like F X; where, X say takes values x 1 – just specifying values – x 2 dot dot dot x n. And this capital X actually is a vector positive of the variables – small x 1 dot dot dot small x n. Here capital X 1 is a particular value for small x 1; x 1 is a variable; capital X 2 is a particular value for the variable small x 2 and likewise. What does this mean? It means the probability of this event that, x 1 is less than equal to capital X 1; x 2 less than equal to capital X 2; dot dot dot dot; x n less than equal to capital X n. This should occur jointly. There are n joint events; that is, one is small x 1 less than equal to capital X 1; another is small x 2 less than equal to capital X 2; so on and so forth. This should occur jointly – simultaneously. That is why it is called joint distribution. This is denoted by F X x 1 up to X n. Sometimes when I do not need, I may not… I may skip this subscript X; and that will be put x 1 to x n. I think you can easily understand sometimes.

(Refer Slide Time: 57:56)



But, when there is confusion ((Refer Time: 57:57)) And capital X 1 is Y 1 p x 2; which is Y 2 minus Y 1 dot dot dot p x n Y n minus Y n minus 1. So, you can easily see that, the concepts involved here – they are nothing new; they are simply generalizations of the concepts that were valid or that were introduced rather in the case of two random variables. That is why I am not giving any proof and all that; you can argue about this on your own.

So, I stop here today. In the next class, I will consider this issue further and go into things like mean, covariance, correlation; in fact we will have a correlation matrix now, covariance matrix and things. Again the characteristic function issue has relevant here. And that takes us through a very important theorem called central limit theorem. So, that is all for today.

Thank you very much.