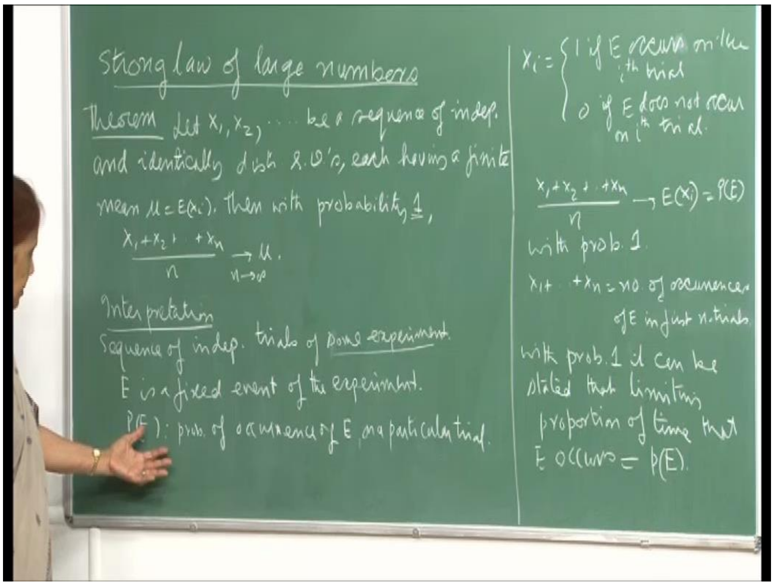
Introduction to Probability Theory and its Applications

Prof. Prabha Sharma
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

Lecture - 23
Strong law of large Numbers Joint MGF

(Refer Slide Time: 00:15)



So, now I will talk after having discussed the weak law of large numbers, we will talk about strong law of large numbers. And I will first just take the theorem, this is theorem simply says that if x 1, x 2, x n is a sequence of independent and identically distributed random variables. Each having a finite mean mu equal to expected x i then with probability 1, see this is the important thing.

Now, we are saying that the probability 1, this average of the sample values x 1 plus x 2 plus x n upon n will converts to mu as n goes to infinity. So that means, this is the sure event. So, therefore you can immediately see the difference between the weak law of large numbers, there it sets the probability such an probability x bar converges to or x bar n converges to mu here we are saying that with probability 1, x bar n will converges to mu, so that means, this is a sure event. Provided the expectation of each of the x i is finite.

So, before we start proving the theorem, let us just interpret what does it mean and what we are saying is that if you conduct sequence of independent trails of some experiment

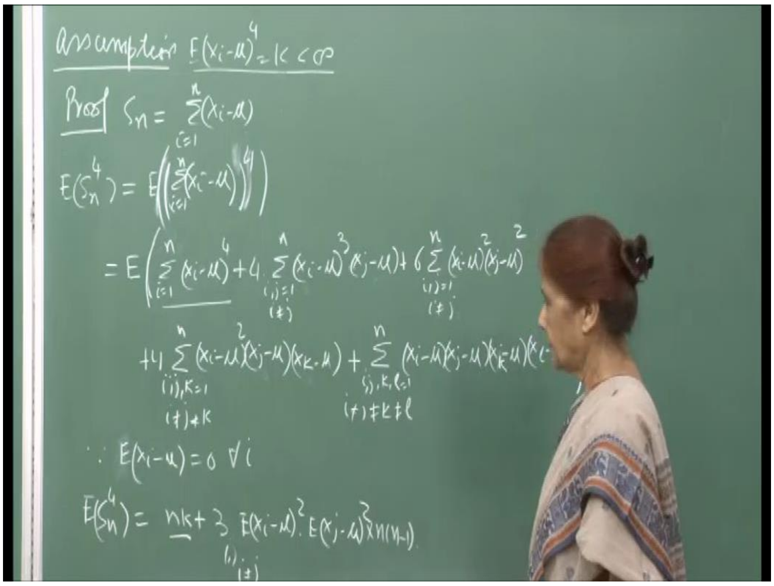
E, some experiment. Suppose, you conduct independent trials of an experiment, if say for example, test tossing 2 coins. So, you go on doing that and then E is the fixed event of the experiment. So, you decide that you just decide one of the events that will occur when you are conducting this experiment say for example, you are tossing 2 coins and you want 2 heads to appear three times, see you know one after another. Suppose E is that event, so you go on tossing the coin or the 2 coins and you do the experiment till you are ok. In this case I am talking of the occurrence of this thing. So, maybe we can say that I toss 2 coins ten times.

And then I want to see how many times I get 2 heads; that means, both the coins show head that would be event E for example. So, E is a fixed event of the experiment then and let P E denote the probability of the occurrence of E on a particular trial. So, this is probability of occurrence of E on a particular trial right. Now, define x i as 1 if E occurs on the i th trial. So, I am defining an indicator variable just to show you that how we can interpret this strong law of large numbers. So, it say that if x i is 1, if E occurs on the i th trial and 0 if E does not occur on the i th trial. So, this will be the indictor variable of the event E; that means, if E occurs on the i th trial will say x i takes the value 1, otherwise x i takes the value 0 right. So, then what the strong law of large numbers is saying that see this sequence x 1 plus x 2 plus x n upon n this is converging to mu as n goes to infinity with probability 1.

So that means, and what is this count x 1 plus x 2 plus x n. x 1 plus x 2 x n is the number of occurrences of E in the first n trials right, because x i is 1 E occurs in the i th trial. So, when you add up x 1 plus x 2 plus x n that will be the total number of times E has occurred, when you have conducted the first n trails. You just started and then you started counting, you started your trials and you started to count the number of times E occurs and that is given by x 1 plus x 2 plus x n right. So, number of and strong law of large number is saying that this ratio; that means, the number of times E has occurred divided by the total number of trials that will converge to your expected value of x i, which is equal to P E right. So, this is, if we are denoting the probability of the occurrence of E by P E. So, this is the probability of E right, I mean; I have denoted P E by the probability of occurrence of E.

And so this ratio will converge to E x i, which is the probability of E right. And this is probability 1. So, this is certain event right. So, if you interesting interpretation and therefore, this, the strong law of large numbers reinforces our concept of the way we had defined probability through relative frequency.

(Refer Slide Time: 05:30)



Now, let us prove the result. So, we have assumed that expectation of x i minus mu raise to 4 is equal to k is less than infinity. So, we are assuming that the fourth movement about the mean is finite and then we show that. So, let us define S n as sigma i varying from 1 to n, x i minus mu then we want to compute expectation of S n 4 right, which would means that sigma x i minus mu this whole thing, i varying from 1 to n this whole thing raise to 4 expectation of this. So now, if you expand this so I should have a mu here also summation.

So, this should be sigma x i minus mu, I am writing sigma this. So, therefore, this whole thing is 4. So, this should be this right and then this whole thing is raise to 4. So, sigma x i minus mu i varying from 1 to n and I am saying S n 4. So, this whole thing raise to 4 and then this expectation. So, when your x i taking the fourth power, sigma x i minus mu raise to, this whole thing raise to 4. So, therefore, I am now, expanding this by the binomial theorem. So, this will be summation i varying from 1 to n, x i minus mu raise to 4 right. So, this is your up to n terms each raise to forth power then you will take 2 at a time, product of 2 at a time. So, it will be 4 times sigma i j varying from 1 to n, x i minus mu cube into x j minus mu, where i is different from j right.

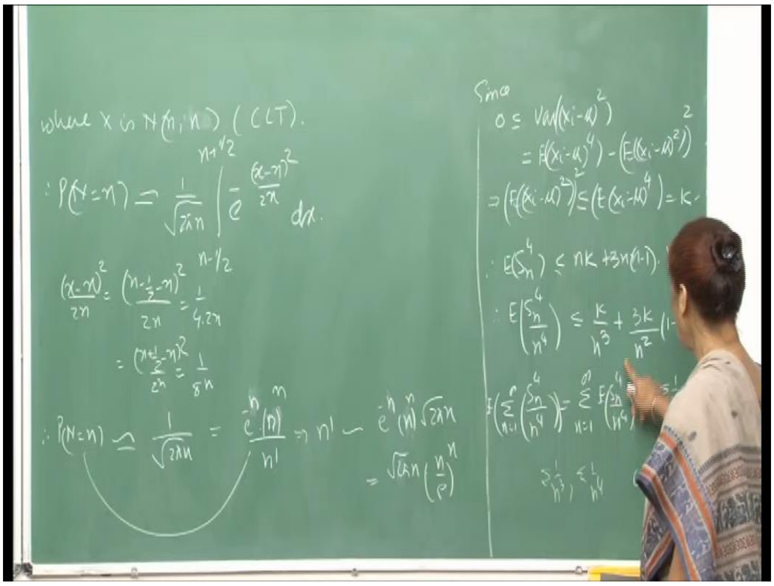
And similarly then you will again take 2 at a time i and j. And this will be six times x i minus mu whole square into x j minus mu whole square summation i j from 1 to n, i again not equal to j. Then you will take three terms at a time so i j k and that will be plus four times summation i j k all varying from 1 to n, but i is not equal to j is not equal to k. So, all three in this have to be different and this will be x i minus mu whole square into x j minus mu into x k minus mu. And finally, product of four terms, where again i j k l are all different, I should have said here I not equal to j not equal to k not equal to l right and this is also varying from 1 to n right. So, this is x i minus mu into x j minus mu into x k minus mu into x l minus mu. So, this is the expansion. S n raise to 4 and so the expectation is all outside. So, this is the big bracket and the expectation of this. Now, of course, expectation can go inside so linear function.

So, in the sense that yes, so expectation can be taken inside then I have assumed independence of the random variables x 1, x 2, x n. So, then expectation of the product is product of 2 random variables is the product of the expectations. So, E can also go inside here now inside the summation sign and since expectation of x i minus mu 0 for all i. So, you see that the expectation of this will be 0 and similarly this will not be 0, but here again you have linear terms.

So, expectation of this and expectation of this will also be 0 and here of course, all the four expectation will be 0, because these are independent. So, this will be summation expectation of x i minus mu into expectation of x j minus mu and so on. So, these terms will disappear. So, you are only be left with sigma i varying from 1 to n, x i minus mu raise to 4 and then 6 times summation i j varying from 1 to n, i not equal to j, x i minus mu whole square x j minus mu whole square.

Now, we have already assumed that this is equal to k and this is less than infinity. So, here you have n such terms, again independence tells you that you can just add up these numbers, you can add up k n times. So, this will be n k and then here, you are saying i is not equal to j. So, the choices you can have is n into n minus 1 by 2. So, this many pairs you can have such i j. So, that i is not equal to j. So, therefore, this will be n into n minus 1 by 2. So, that will cancels out with 6 or will be 3. 3 times this is what you will get.

(Refer Slide Time: 10:23)



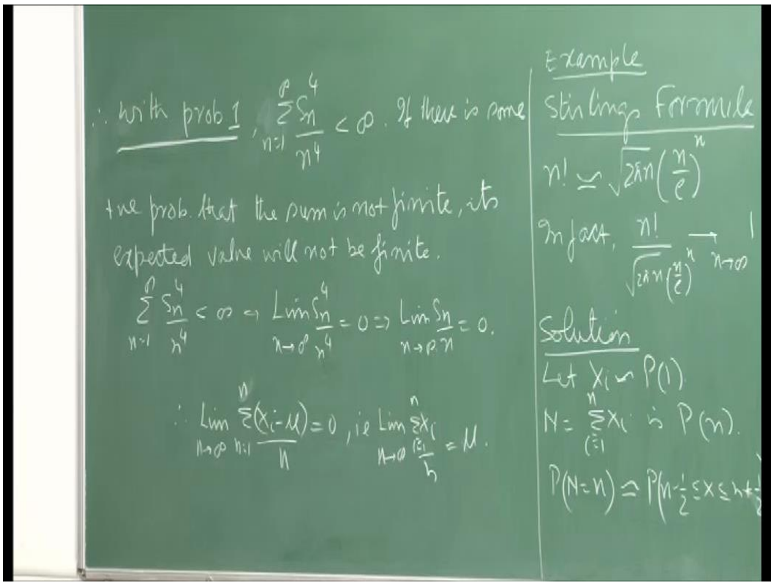
What we are saying is that since variance of x i minus mu whole square, because I am assuming that this always non negative. So, this is equal to expectation of x i minus mu raise to 4 right. If I give write the expression for this. This is the fourth movement expectation of, fourth movement about mu minus expectation of x i minus mu whole square the variance of x i minus mu whole square.

So, that will be expectation of the square of square of this. So, just say to 4 minus expectation of x i minus mu whole square then whole square right. This of course, is your variance of x i so anyway. So, then since this is non negative therefore, this is from it follows that your expectation of x i minus mu square whole square is less than or equal to expectation of x i minus mu raise to 4, which we are taking to be k. So, therefore, this is also finite right. So, therefore, everything is finite here right. These things are also finite, because this square is finite so therefore, both of these are finite. So, therefore, expectation of S n 4 is less than or equal to if you want to write n k plus 3 n into n minus 1 into k, because each of them is less than or equal to root k. So, that becomes k here right.

And therefore, when you divide the whole thing, both the sides by n raise to 4 expectation of S n 4 divided by n 4. So, this becomes k by n cube 3 k upon n square into 1 minus 1 by n. So, this you can utilize for large values of n. So, this will become 1 right. Now, since 1 upon n cube sigma 1 upon n cube n sigma 1 upon n 4 are both convergent series right. Remember, because sigma 1 by n cube n sigma 1 by n 4. So, n goes to infinity 1 to infinity these are convergent. So now, I can take the; that means, when I take the summation here, this is convergent series, because both these series are convergent. And so I write expectation of sigma and varying from 1 to infinity S n 4 upon n 4 is equal to this, because since this is convergent series, I can take expectation inside. And so I get this here right and yes. So, see this is what we have shown is that this is finite right.

Expectation of S n 4 upon n raise to 4 summation n varying from 1 to infinity, this is a finite series.

(Refer Slide Time: 13:10)



So, therefore, with probability 1, this summation n raise to 1 to infinity S n 4 upon n 4 should be less than infinity. I mean see actually, we have shown that each of this is, because each of this is k upon n cube plus 3 k minus n square into 1 minus 1 by n. So, therefore, this summation, when I take the summation here, sigma 1 by n cube sigma 1 by n square they are both convergent. So, therefore, this is a converge series, but because this is we can take E outside right, because of linearity. So, then this is finite expectation of sigma and varying from 1 to infinity S n 4 upon n raise to 4 is finite. And so we are saying that the inside thing the, this expression or this series must be finite, because if there is some positive probability that the sum is not finite.

If this sum is not finite then its expectation will not be finite and we have shown that the expectation is less than sigma expectation this thing and therefore, that thing is finite. So, this must be finite, because if there was any positive probability that this is not finite then the expectation would not be finite. So, therefore, I am saying with probability 1. So, this is the main point right. I will repeat the argument that, we have said that this is a finite series, but this I can rewrite as expectation this right. And why we are saying this, because this whole is finite, because of this was not finite then expectation would not be finite, but here we have this is, this whole thing is finite. This is equal to this and this is finite. So, therefore, sigma n varying from 1 to infinity S n 4 upon n raise to 4 is finite.

And if a series infinite series has a finite sum it is a convergent series then the n th term must got to 0. Otherwise again from your convergence of series, you know that this is the necessary condition that the n th term must go to 0, if the series is convergent. So, therefore, sigma S n 4 upon n 4, n varying from 1 to infinity less than infinity implies that the n th term must go to 0, as n goes to infinity and if the now this goes to 0 then the fourth power 1 one fourth root of this will also go to zero. So, therefore, limit s n upon n as n goes to infinity is 0 right. And so just replacing the value of S n here, this is sigma x i minus mu by n, n varying from 1 to such a i varying sorry, i varying from 1 to n limit S n goes to infinity is 0 right. And so you can just take summation inside here. So, sigma x i by n i varying from 1 to n limit n goes to infinity is mu. So, this is with probability 1.

So, essentially here I just needed the a fact that to prove this strong law of large numbers; that means, first of all let us just we clear. So, what we are saying is that this will happen with probability 1 so; that means, it is a sure event. And so as n goes to becomes larger and larger what we are saying is that this x n bar, your x n bar will converge to mu. So, little get closer and closer to mu and this is a sure event this is happening with probability 1. In the weak law of large numbers I be just simply said that the probability of x n bar minus mu see this value greater than delta probability of this could be shown to be less than epsilon and then of course. So, therefore, this was only in terms of probability now here we are this is the sure event that x n bar must go to mu as n goes to infinity ok.

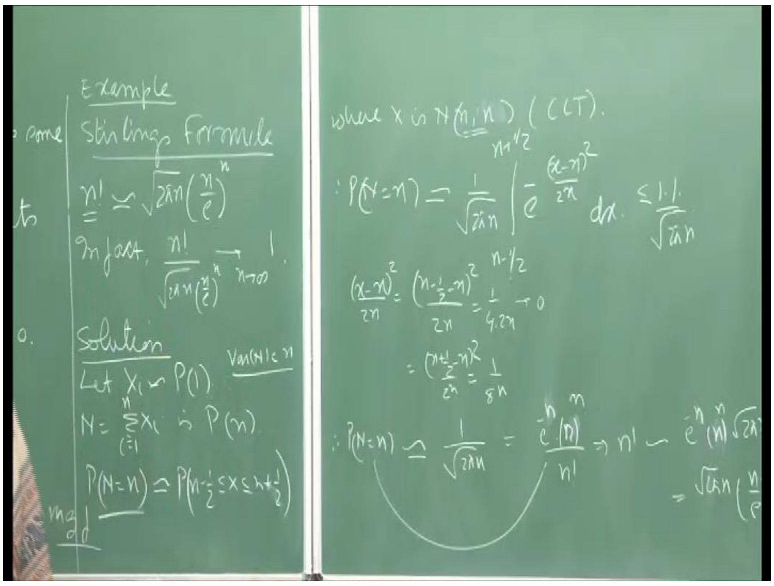
Now, the thing is n of course, here I just needed the fact that expectation of x i minus mu raise to 4, this thing is less than infinity right. So, what I want to say is that if the kind of distribution that we have discussed in this course all of them I could show you the existence of m g f and I have not taken any distribution for which the m g f did not exist of which the mean and the variance did not exist. So in fact, all the distribution that we have considered here so therefore, you can see that for all of them this condition will also be satisfied, because if the m g f exists then the force movement will also be finite. In fact, the movement m g f you can what we mean that m g f exists when you expanded you get different powers of t raise to n upon n factorial would give you the n th movement or about the origin.

So, if that is finite then you can see that this will also finite right. And so therefore, the strong law of large numbers also holds for all this distributions as so the weak law of large numbers and strong law of large numbers both hold. And so essentially if only when you have situations where you are well actually yeah, maybe I should not really worry about that part, but essentially the proof has been, this proof has been given under the condition that expectation of x i minus mu raise to 4 is less than infinity fine. And that this is the show event; that means, here this will converge the x n bar will converges to mu as n goes to infinity with probability 1. Now, just want to look at Stirling formula here and see all of you know that n factorial can be approximated by under root of 2 pi n into n by E raise to n. So, many times this is a very useful way of approximating the E factorial right.

And many limiting situations and so on, we it is very helpful to be able to replace n factorial by this and then you can get a good results. So, in other words what we are saying is that n factorial upon under root 2 pi n, n by E raise to n goes to 1 as n goes to infinity, this is the idea right. Now, the solution what we are doing is here is, here we are saying that, lets x i be poison 1; that means, the lambda is 1. So, I mean this thing the parameter for the poison distribution is 1. So, let me take x i this then take n to be sigma x i, i varying from 1 to n.

So, this will be poison n right. And for poison n your variance; that means, variance of n is also n remember for poison lambda mean and variance are the same and they are both equal to parameter lambda right. So now, if you want to estimate this probability n equal to n this using the central limit theorem, using the central limit theorem I will say that x this can be approximated using the continuity factor the x lies between n minus half and n plus half where x is your normal n n ok, so applying the central limit theorem.

(Refer Slide Time: 21:05)

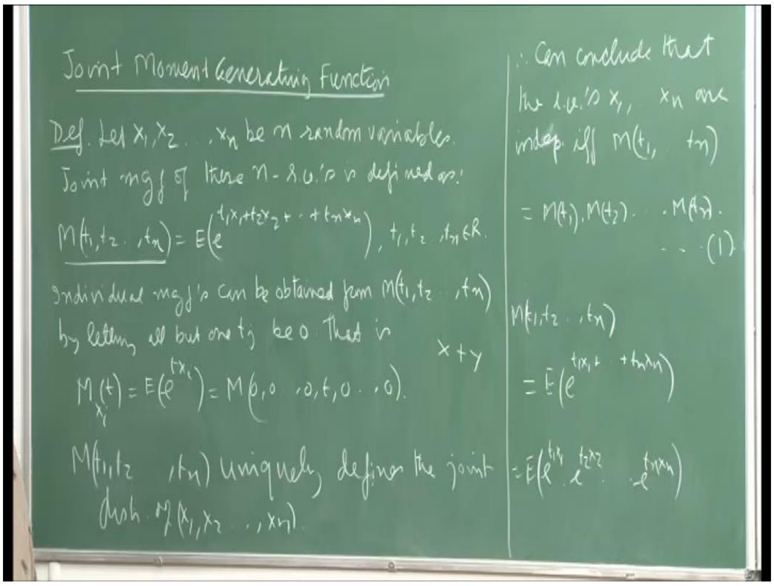


Let approximate this probability by saying that the corresponding normal so for, large n we will say that n behave like normal variable with mean n and variance n right. So, this is what you want to compute and therefore, in terms of. So, I want to write this probability. So, this 1 because x is normal I will write 1 upon under root 2 pi n, because our variance is n. So, standard evasion will be root n and this will be n minus half to n plus half of E minus x minus n whole square to n d x. So, this is my probability using, because I have used the central limit approximation fine. Now, just look at this integrant see what I am saying here is that x minus n whole square upon 2 n at the lower limit n minus half is n minus half minus n by 2 n whole square, which is 1 by 4 into 2 n.

So, this goes to 0 as n goes to infinity and so E raise to minus something going to 0 is 1 right. And similarly when you substitute n plus half for x then again this will be 1 by 8 n right. So, you see the in the limiting case as n becomes large the two limit come close right and the value of the integrant is close to 1 right, because for n large this is always 1. So, therefore, we can always say that this integral is you can, you take the maximum value of the integrant which is 1 into the length of the interval which is also 1. So, this is this upon root 2 pi n. So, just apply this approximation, because the theorem from integral calculus to this integrant is throughout then you multiply that by the length of the interval. So, you get 1 upon under root 2 pi n right.

As so this probability is approximated by 1 upon under root 2 pi n and but since this is we said this is poison random variable, because we started their option that n is sigma x i. So, then this probability in terms of poison probability can be written as E raise to minus n, n raise to n divided by n factorial. And so from when you equate this 2 and you get that n factorial, I mean you equate with this and then approximate by 1 upon under root 2 pi n. So, your n factorial is under root 2 pi n, n by E raise to n. So, you know see the interesting application I mean, I just came across that this application what I taught discussed with you about central limit theorem. So, the strong law of large numbers we have sort of established, but as we saw that for us actually there will be no difference and we will continue to approximate mu by x n bar and for reasonable large values of n.

(Refer Slide Time: 24:24)



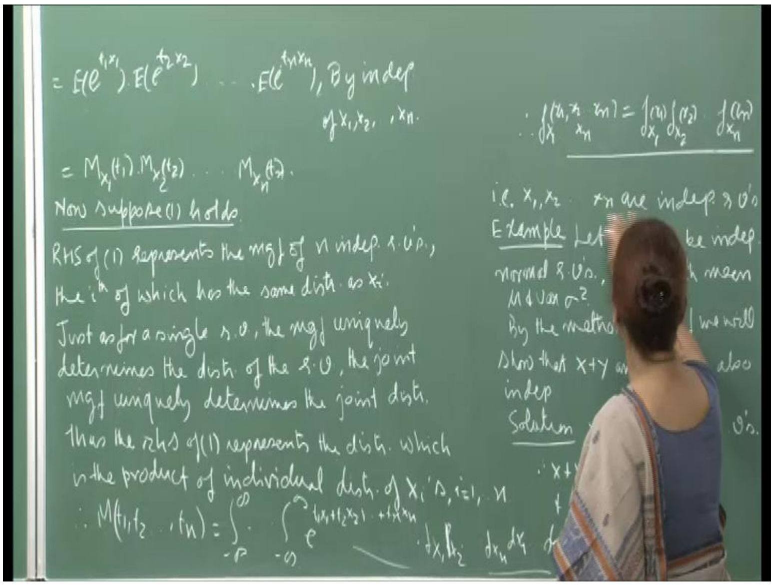
So now, I will want to talk about joint movement generating functions we talk about the movement generating function for a single random variable. And then we talked of we know we could compute for independent random variables when you talk of sum of independent random variable like two random variables x and y are independent. Then I could also you know, because of two independence, we could define the movement generating function of x plus y, because it was just the product of the movement generating function of x and y, but there should be a general definition of movement generating function of more than 1 variable when they are not independent. So, therefore, just completing this ah this part of the theory. So, what we saying is. So, the definition simply says that if x 1, x 2, x n are n random variables and then the joint movement generating function of these.

So, I mean these n random variables. So, we are given the joint density function of the n random variables then we can define the movement generating function of these n random variables as of course, right now I am not listed this simply the expectation; that means, so you now need n real numbers t 1, t 2, t n. So, we will say that the movement generating function of x 1, x 2, x n is actually and I wrote the m g f by m of t 1, t 2, t n this is expectation of E raise to t 1 x 1 plus t 2 x 2 up to t n x n for all real numbers t 1, t 2, t n for which this expectation is defined. And the individual m g f can be obtained from this by putting all but one of the t i s equal to 0 and then getting the corresponding function from here, because then it will be say for example, for the i th you want to compute the m g f of or obtain the m g f of the i th random variable here.

Then I will put all other t i s equal to 0. So, m of x i t would be E t x i right, expectation of E raise to t x I, which will be in terms of the function n here will be simply 0 0 and then t i you write as t and all other as zeros. So, therefore, when you defined the joint m g f you can get the individual m g f also and just as in one variable case we had we did not proved the result, but we stated it and said that if we movement generating function uniquely defines all distribution functions. So, once you have obtained the movement generating function of a random variable then you know what is distribution function and also be and of course, it is unique right. So, here also joint case we will again just assume this result that the movement generating function uniquely defines the joint distribution of x 1, x 2, x n.

So, yeah, and now, what we want we said uniquely defines this and now under independence yeah. So, therefore, if the joint density function is uniquely defined then we can conclude that, if the random variables x 1, x 2, x n are independent. Then I mean this is the condition if and only if your m t 1 t 2 t n can be written as the product of individual this thing. So, here if you want to write you can into this also in; that means, you can decompose your joint movement generating function into the product of individual m g f. So, I mean assuming that if this result we have sought of accepting that the m g f will define the distribution function uniquely and so now, we can yeah so you want let us show the if and only part. So, now, if they are independent then of course, the things follow immediately, because you will write expectation of E t 1 x 1 plus t n x n and that will be and this you can then write as product and because x 1 x 2 x n are independent. The expectation I can take inside and so this whole thing this can be written as this. This is because x 1 x 2 x n are independent right, product of the expectations and so it immediately follows that this is your m g f of x 1, this is m g f of x n. And so you can write this.

(Refer Slide Time: 29:17)



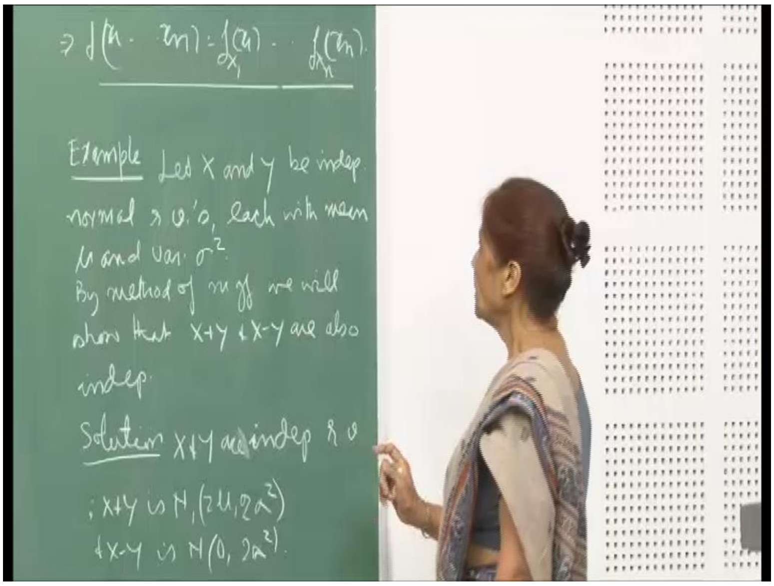
Now, the other way, now let us show the converse that is now suppose one holds. So, we want to show that, this relationship will also we can conclude from here that x 1 x 2 x n are independent random variables. So, we can see if you look at the right hand side of one. So, this part then this represents the m g f of n independent random variable, because its product of n m g f.

So therefore, and which we know, we have said that if two random variables are independent then the m g f of 2 random variables will be the product of individual random variables. So, just extending that rule this represents the product of this represents the m g f of n independent random variables. Now, the i th of this random variable, the i th term here, m x i t i of which has the same distribution is the x i right. Because, so here each one of them for example, m x 1 t 1. So, this is the movement generating function of x 1 and as we have been saying that the movement generating functions characterize the p d f uniquely. So, therefore, each of the terms here, each of the m g f here determine uniquely the corresponding distribution p d f or so which as of the i th variable right. So, just as for a single random variable the m g f uniquely determines the distribution of the random variable.

The joint m g f uniquely determines the joint distribution. So therefore, from here we can say that the product of the. So, that the joint m g f this will give me, because this is the joint m g f of x 1 x 2 x n. So, this will determine the joint m g f of x 1 x 2 x n, but then that is we have shown is the product of the individual p d f. And this is how we have defined independence of the random variables x 1 x 2 x n that is if the joint p d f which I have written down here. So, the right hand side of one represents the distribution which is the product of individual distribution of x i s and therefore, this is the and so here and therefore, you can expression wise also write that m of t 1 t 2 t n is equal to this expectation of t 1 x 1 plus t 2 x 2 plus t n x n into f x 1 f x 2 f x n; that means, the joint l c d function of x 1 x 2 x n joint p d f is the product of individual p d f.

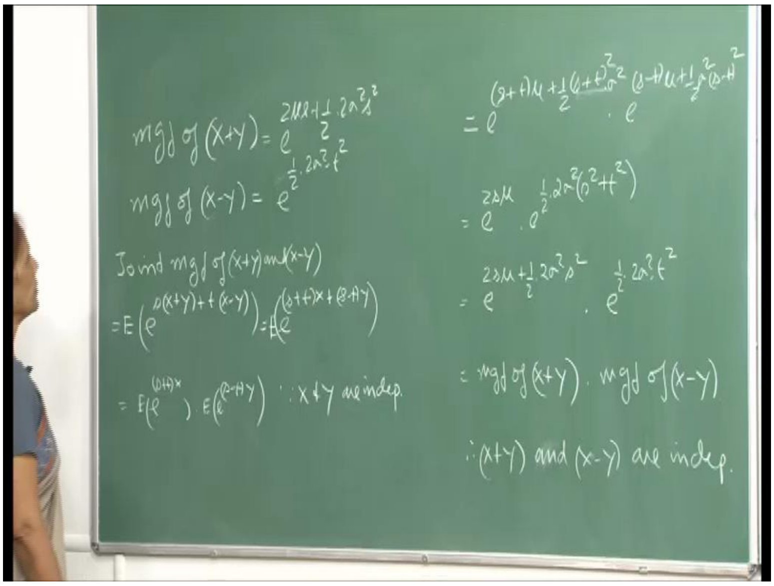
So this, what we are concluding, we can immediately conclude from here right, because the m g f uniquely characterize your p d f. So, therefore, just using that fact I can conclude that the joint p d f is this and therefore, x 1 x 2 x n are independent random variables. So, we need proof of the fact that if you can write the joint movement generating function as a product of individual this thing then it implies that the random variables are independent. And if they are independent then you can also write the m g f joint m d f, m g f as the product of the individual m g f s. So, we have been using some of these results, but now I have some sought of new supported it by theory.

(Refer Slide Time: 33:07)



See in this example, I am just trying to demonstrate the use of you know joint m g f. So, even though you know x and y are independent random, normal random variables each with mean mu and variance sigma square. So, if you start with that then we have already shown that you know by using the method of transformation that x plus y and x minus y are also independent random variables. And in fact, they are normal random variables, but now, you want to use the method of m g f to show that x plus y and x minus y are independent. And then of course, once we have shown once we obtain the individual m g f then as I have been saying that once you know the m g f you can also determine the distribution function or density function of the random variable. So, we will do that. So, just as an illustration of what we have just discussed, I want to go through this example. So, since x and y are independent and they are normal independent random variables and they have both mu and sigma square right.

(Refer Slide Time: 34:34)



So therefore, x plus y will be normal 2 mu and then variances will get added, because they are independent. So, 2 sigma square and for x minus y the mean will be 0 and the variance will be again 2 sigma square. So therefore, if you want to write m g f of x plus y, because its normal with mean 2 mu and variance 2 sigma square therefore, it will be E raise to 2 mu s plus half into 2 sigma square s square right this is. And similarly m g f of x minus y will be because mu is 0 the mean obviously is 0. So, it will be E half into 2 sigma square into t square, this is simply t square right. Now by our formula we will write the joint m g f of x plus y and x minus y. So, this will be expectation of E raise to s times x plus y plus t times x minus y for s and t real numbers right s and t belonging to R right, which I can by rewriting this right. So, now, I collect the x terms and the y terms.

So, this is s plus t is the coefficient of x and s minus t is the coefficient of y. So, this is what you have right. Now, we will use the independence of x and y, because this is simply some s plus t times x which can be your t 1 and s minus t which can be your t two. So, this is E raise to t 1 x plus t 2 y, but x and y are independent random variables. So, therefore, I can decompose this m g f into the individual m g f. So, this will become expectation of E raise to s plus t into x into expectation of E raise to s minus t y right. So now, I can use the independence of x and y, because this is written as this way and so s plus t can be treated as another real number and s minus t can be treated as different real number right. And so because of independence of x and y I can decompose into this right.

Now, let me right the m g f of, because x is again normal with mean sigma and variance sigma square and this also is mean mu and sigma is the variance. So therefore, when I write the m g f s plus t E raise to s plus t into mu plus half s plus t whole square sigma square and the other will be E raise to s minus t mu plus half s minus t whole square into sigma square right. And then you see we just rearrange the terms simplify the expression. So, s plus t into mu and s minus t into mu will become 2 s mu right. And here the product terms will cancel out the 2 s t here and the minus 2 s t it will cancel out and it will be E raise to half into 2 sigma square s square plus t square right. So now, again I collect the s terms. So, this is E raise to 2 s mu plus half into 2 sigma square s square and this is E half to sigma square t square.

And this is what exactly see this is the m g f of x plus y, because this is and that is what I am saying. So, this is m g f of x plus y, because this is 2 mu and 2 sigma square and you know, you can also say that these are x plus y is normally distributed with mean 2 mu and 2 sigma square and this is the m g f of x minus y. So, there you see that mu is 0 and the variance is 2 sigma square right. And so since from the theorem that I had just stated and proved to you, this see that if you are joint m g f can be written as the product of the individual m g f then the variables must be independent right. And so we conclude that x plus y and x minus y are independent and also we can conclude that x plus y is normal 2 mu 2 sigma square and x minus y is normal 0, 2 sigma square. So, you know with the series through series of examples, I will tried to revisit the results.

Which we have already I will try to revisit the results, which we have already you know obtained especially; I will apply this concept of joint m g f to sums of random variables. And then try to show you that sometimes this method is easier and we can get the results faster. So, it depends on the situation and of course, lot of experience, but this is also another important tool and I taught that this course we must define this and you know give you the results. So, that you can sometimes when other methods do not work this will proved to be quite.