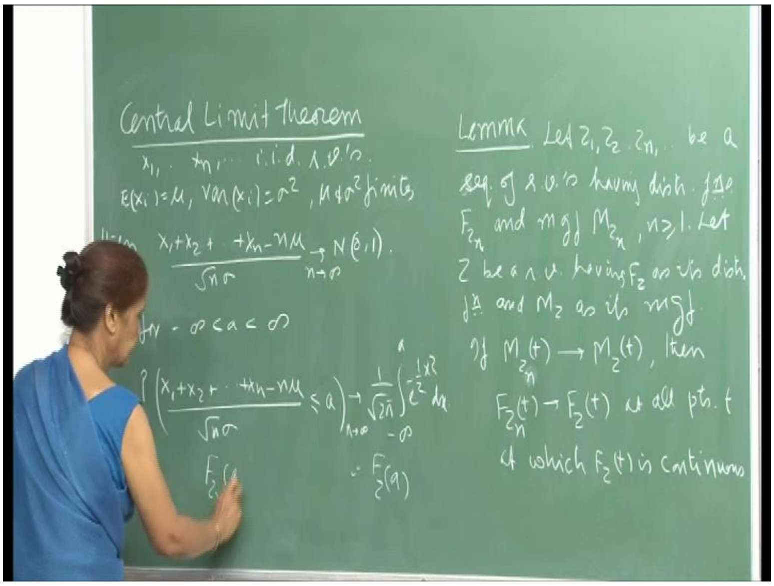
Introduction to Probability Theory and its Applications

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

Lecture - 21
Central Limit Theorem

(Refer Slide Time: 00:14)



So, we will as I said in the last lecture, we will now after having stated the central limit theorem, talk to you about its importance; we will actually see now once I prove the theorem and then I give you applications, you will see how important the theorem is, and how widely used the theorem is. So, as we said that the central limit theorem say that if you have the sequence of identically independently distributed random variables with mean and variance finite, then we say that this, sum of these random variables x 1 plus x 2 plus x n, it will have mean mu and variance n sigma square.

So, this will go to, this will converge to N (0, 1) as n goes to infinity; that means the distribution no matter what the original distribution of the exercise was; now when you take the sum and you let n go to infinity, then this random variable will converge to the standard normal variant. So, this is convergence in distribution, right. Or in other words, as we can state it in other way also, this is for any, a which is from minus infinity to infinity of finite number. Then, the probability of this number being less than or equal to this random variable, less than or equal to, a, will converge to the standard normal; that means, this is your F z a, fine; and this is your F z n a. So, this converges to this as n goes to infinity. So, in distribution the convergences they are, right.

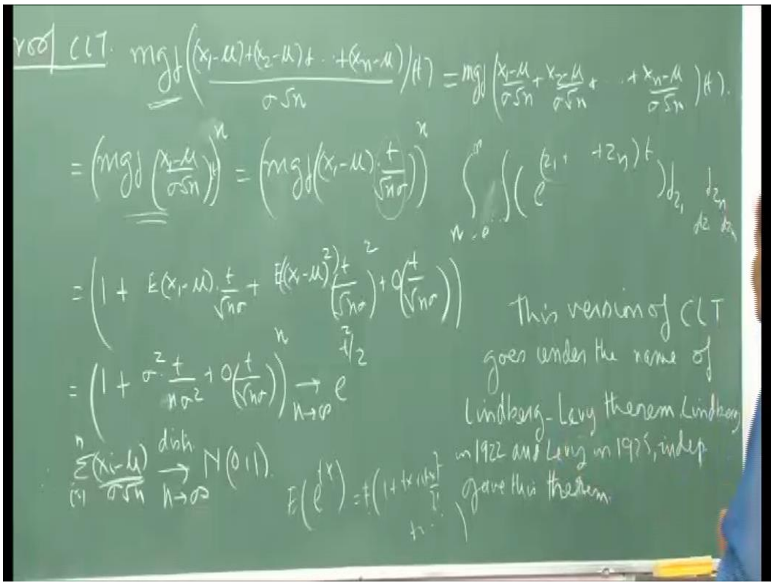
Now, in order to prove this I need to use this lemma which talks about uniqueness of the m g f, and I will not give a proof for this. We will just except the lemma as it is. So, this says that if z 1, z 2, z n, again is a sequence of random variables having distribution functions F z n, and m g f M z n, right; and greater than or equal to 1. So, the distribution function is F z n, and the m g f of z n would be M z n, and greater than or equal to 1. Let z be a random variable having F z as its distribution function and M z add its m g f, right.

So, then we, if M z n t converges to M z t that means as n goes to infinity this movement generating function converges to the movement generating function of the variant F z, then we say that the corresponding distribution functions will also converge to the distribution function of z; that means, if, so this is, this talks about the uniqueness of the m g f. That means, if the m g f, M z n t converge to M z t for the variant z, the m g f of z is M z t; as n goes to infinity then the corresponding distribution function of z n which is F z n t will converge to the distribution function of z, for at all points t at which F z t is continuous which means that it is defined, right.

So, this is the idea; that means, m g f uniquely, and while discussing m g f, I also tried to tell you that m g f uniquely give you your density function or the distribution function because the parameters you can compute; and actually not only the parameter, but the distributions are, is the same. I mean, once you get a m g f you can uniquely fix the distribution, right, of their form of the m g f function.

So, this is what we are stating here, which we have also been using otherwise. So, now, let us; so the proof is not very difficult to straight forward. So, I am just rewriting this function, this random variable, mu is getting attaching to each of the exercise.

(Refer Slide Time: 04:22)



So, therefore, the same thing can; so now, I am writing the m g f of the random variable which we want to show well converge to standard normal variant and distribution. So, this can be, at a point t, I am defining the m g f. So, this is in fact, e raise to this whole thing into t, right. The m g f of this would be e raise to whatever you want to call it, y n or something, or z n, then e rise to z n t.

So, m g f excepted value of e raise to z n t you can say, right; which I can write as x 1 minus mu upon sigma root n plus, x 2 root n plus x 2 minus mu, and so on. Now, since x 1, x 2, x n, are identically distributed and independent random variables, so the m g f here, by again by the property of the m g f; well, the thing is that, yes, I should have; so yes, the order has been a little, because I will be talking of the m g f for the independent random variables, you know, more than 1 variable.

So, that should have come, the lecture should have come before that; ok, anyway we can talk about it. So, what I am saying here is that this m g f can be written as the product of the m g f of x 1 minus mu upon sigma root n raise to n because of the independence. This anyway also follows from independence because you see, when you write this let me say, I was saying z 1 plus z n, and you are taking t. So, when you write expectation of this, right, because the p d fs are.

So, this will be z 1 into F z n because the join density function would be this, right. So, your expectation when you write of this would be whatever given, n th order integral from minus infinity to infinity, depending on if the variants are defined; I am taking the general definition into d z 1, d z n, right. So, then you see you can separate out the integrals, and so each integral will be the m g f of this; and since they are identical it will be the same, right. I am writing f z 1, f z n, where they all the same, f z 1, f z 1. So, therefore, you can immediately see from here that this will be, that this m g f can be written as this because of independence and identically distributed random variables.

Now, this I can write as, this should have been at t; m g f at t, right. And so let me just separate out this. So, this will be m g f of x 1 minus mu, and I am taking the variant to be t upon root n sigma; so this raise to n. Now you expand this; remember this is, when you writing this, so I am just expanding my t x, and this would be, this will be expectation of 1 plus t x plus t x whole square by factorial 2, and so on, and the expansion of e raise to t x. So, that is what I am doing.

And then, I am taking expectation inside because; so now, if you, in the central limit theorem when I assume that they have finite variance; so another assumption I should have made is that because I am using the m g f way of, I am using the moment generating functions to prove the theorem I should have also said here that m g f x i exist. And so obviously, we will be talking about all those, ts, at which the m g f exist, right, at which the m g f is defined.

So, therefore, I take this expectation since the m g f exists. So, I can take the expectation inside because the series is convergent series, and so I can take the expectation inside. So, this is what you have raise to n. And, you see here, I am not considering higher powers of t because see this is square then it will be t q, t 4, and so on. So, I am just writing this whole as a, you know, higher order terms of t, powers of t, right.

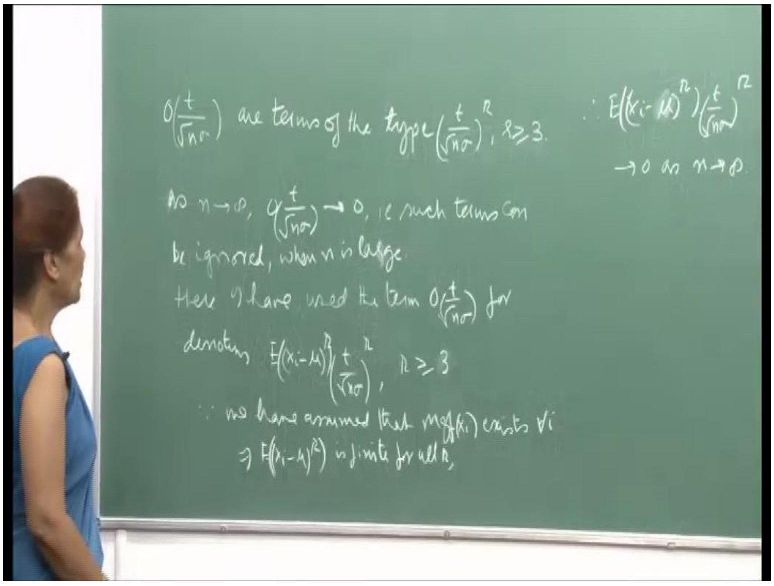
So, then see, expectation of x 1 minus mu is 0 because we have taken, assumed that each x i has mean mu. So, this term is 0, and therefore you will be left with this thing. So, and expectation of x and minus mu whole square would be sigma square. So, this is 1 plus sigma square t upon, should be t square and sigma square; there should have been a 2 also; sorry, this will be 2 factorial, and so on. So, therefore, there will be a 2 here, right, because t x square upon 2 factorial, and so on; and then higher terms, higher order terms of t, right; this raise to n.

Now, so as n goes to infinity, you see, because n raise to half is in the denominator. So, then when you say take the third power it would be n raise to 3 by 2, and so on. So, these things will go to 0 as n goes to infinity. So, but here; so I will ignore this; and then you see what happens to this; 1 plus sigma square. So, the sigma square cancels out, t square by 2 n, right; and so if you write this as t square by 2 into n, and then raise to n. So, I hope, you know, that most of you that this will converge to e raise to t square by 2 because n in the denominator and then n power. So, as n goes to infinity I can safely ignore these terms.

So, then this will be, this will converge to e raise to t square by 2 because sigma square sigma square cancel out, and you are left with t square by 2 into 1 upon n raise to n. So, this converges to t square by 2 as n goes to infinity, right. And, you know that this is the m g f of random variant which is normal 0, 1, right.

And so we have shown that using the lemma because I have shown that the m g f of this random variable, sigma x i minus n mu upon root n sigma, that converges to the m g f of a standard normal variant. Therefore, by the lemma I can assume that this random variable converges to N 0 1 in distribution as n goes to infinity.

(Refer Slide Time: 11:30)



So, you know, using the m g f the proof really simplifies. And so in the proof I have used expression, small o of t upon root n sigma. So, the understanding is that this denotes terms of the type, t upon root n sigma raise to r, r greater than or equal to 3; say in that proof proving the central limit theorem I wrote down the terms upto t square, and then I wrote down that the later terms will all be having higher power of t upon root n sigma. So, this is the expression.

And see, the understanding is that as n goes to infinity, this is for this expression; that means, because r is greater than or equal to 3. So, any way this will go to 0. So, that means, as n becomes larger and larger the terms that become very small. So, their contribution is negligible. Therefore, we ignore them, right. This is the idea.

But now, for your convenience I have expressed, I have used this term for, this notation for also including the term expectation of x i minus mu raise to r. So, that means, I am taking that this now denotes for me in that proof, this into t upon root n sigma raise to r for r greater than or equal to 3. But then, since we have assumed that the m g f exists for all x i, x i s are all identically distributed. So, the m g f exists; that means, all movements exists.

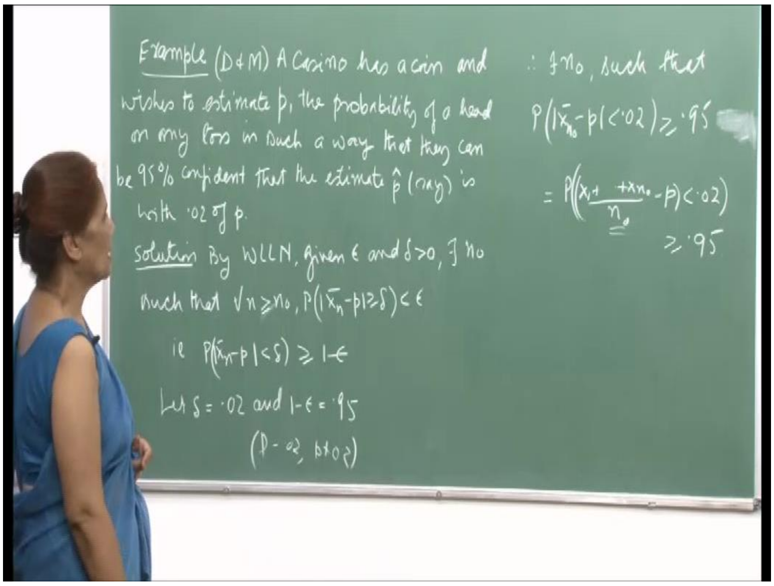
So, all movements are finite, and therefore, these numbers are finite for all r; hence the same thing apply; that means, if this becomes small then n becomes larger and larger; this whole thing also becomes very small and goes to 0, right. So, this is the idea that as n goes to infinity this will go to 0. So, therefore, we can neglect the term. So, this is, I have used this expression elsewhere also.

And, so this is, the understanding is that when you write small o then it means that these are higher order terms whatever you have written down beyond that all higher order terms; that means, of power higher than 2, here for us the way I am using it r greater than or equal to 3. So, therefore, for large n I can ignore such terms in my sum.

This version of c l t, central limit theorem goes under the name of Lindberg Levy theorem also. Lindberg in 1922 and Levy in 1925 independently gave this theorem; we showed, proved this result; so independent of each other in 3 years gap. So, therefore, this is also sometimes known as the Lindberg Levy theorem, but most commonly it is referred to as the central limit theorem.

So, the proof is simple. It just using the independence identically distributed random variables and the properties of the m g f. So, through the property of m g f you could show that this sum of the random variables which are independent identically distributed random variables will, minus n mu upon root n sigma, converges to a standard normal variant.

(Refer Slide Time: 14:36)



So, now let us look at this interesting example. This is from Mishra, and I will give you the references at the end of the course. So, a casino has a coin and wishes you know remember casino is people bet; and so game is tossing a coin to show a head. So, casino has a coin and wishes to estimate p the probability of the head on any toss in such a way that they can be 95 percent confident that the estimate p hat is within 0.02 of p.

So, obviously, they want to have, an idea is to how what is the probability that the coin will throw, show a head when it is tossed. It is important to them because every, whenever a head is tossed then it will be person who is playing the game wins. So, the casino has to pay. So, therefore, they want to be confident that whatever their estimate is that is within 0.02 of p, the actual p, right.

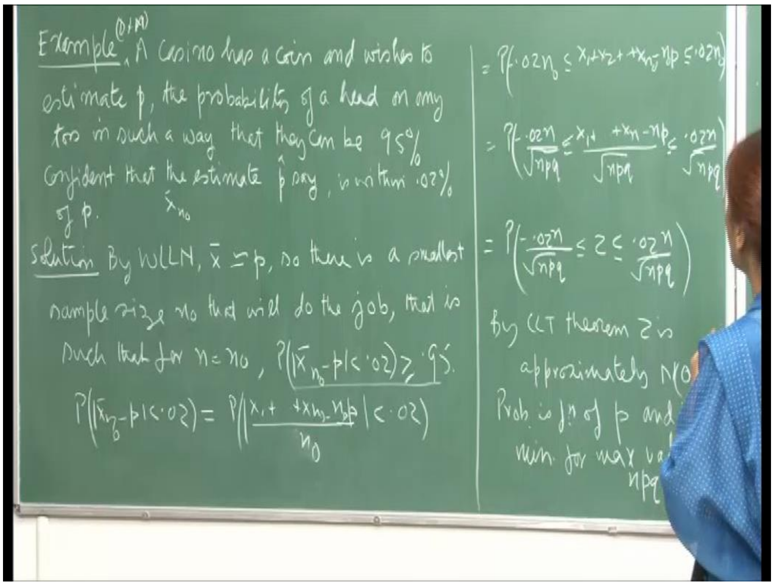
And so weak law of large numbers helps you out here. So, given an epsilon and delta greater than 0, we know that there exists a n naught; this smallest value of n such that for all n greater than or equal to n naught, probability of x n bar minus p greater than or equal to delta is less than epsilon, right. So, for all n greater than or equal to n naught, this difference is greater than or equal to delta is less than epsilon, right.

And so the complimentary, the event would be that probability x n bar minus p in absolute value is less than delta. So, this probability is greater than or equal to 1 minus epsilon. So, the casino, for the casino problem your delta is 0.02. So, that your x n bar is within 0.02 of p, right. So, it can be either little less than p with; that means, it can be p minus 0.02, and p plus 0.02. So, this is what you want.

So, your x n bar should be in this interval. So, delta is 0.02; and here 1 minus epsilon is 0.95. So, this probability that your x n bar is within, is in this interval should be, the probability should be greater than or equal to 1 minus epsilon. So, that means, you would be 95 percent confident; this is the idea. So, therefore, just write out this. So, this is exactly what; so once you give the values of delta and epsilon you get that probability x n naught minus p in absolute value, should be less than 0.02. So, this probability should be greater than or equal to 0.95.

This I do not need to write this because now I am using the value n naught. And, so when you expand this x 1 plus, x 2 plus, x n naught, upon n naught minus p, so this should less than 0.02 is greater than or equal to 0.95. So, this will; so therefore, you can find such an n naught; and therefore, the casino can, by tossing the coin that many times they can find out the estimate p hat for the probability of p.

(Refer Slide Time: 18:17)



Same event can be rewritten as, this has to be or it was not really necessary because since we have said that there is an n, n naught which will do the job. So, therefore, I could have carried it as n only, and then I found it out, but anyway. So, therefore, this is therefore, this everything is n naught here, right. So, now, I have this event, and if I divide all the numbers by, under root n naught p q, because the variance of each x i is p q.

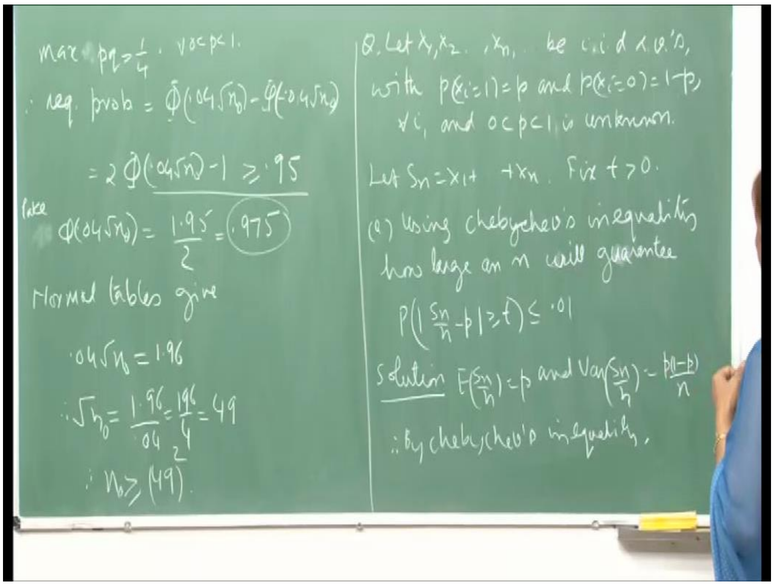
And therefore, the variance of x 1 plus, x 2 plus, x n naught, will be n naught p q; and so divide by the standard deviation; and then therefore, this event is the same as this. So, the probabilities are the same, right; 0.02 n naught, divided by under root of n naught p q. So, I divide throughout by under root n naught p q to, you know, standardize. And, therefore, this I am, this is my random variant now which is the standardized variant; and by c l t theorem, this is the standard normal variant, approximately, of course, right; approximately this is standard normal variant.

And, so probability is a function of, this probability would come out to be a function of because your numbers on the 2 end are, the interval in which you are founding, wanting z to be, that probability will depend on t. Now, the thing is that you are, see, if I find; so here again you have to do n naught, n naught.

So, the whole idea is that if I put this equal to, if I put t maximum value of n naught p q here, then I will get the, because its denominator if I put the maximum value then this will be the smallest. So, if; that means, this interval will be the smallest; for the maximum value of this number it will be smallest interval; and so probability of this smaller interval, if this is greater than 0.95 then for any other p this probability would be greater than 0.95; that is the idea.

So, my event what I am doing here is by writing the maximum value for this, this event will be subset of all other events whatever the value of p because q is 1 minus p, right. So, therefore, if this probability can be made to be equal to 0.95 or greater than or equal to 0.95, then for all values of p this will be greater than or equal to 0.95; this is idea, fine.

(Refer Slide Time: 21:02)



So, therefore, the and we know that the maximum n p q is 1 by 4 for all 0 less than p less than 1, we know this; I am sorry, for I am writing this p q. So, maximum value of p q is 1 by 4, and therefore, the required probability, since by c l t theorem this is standard normal, so this will be; now let we have started doing it. So, I will write n naught everywhere. So, this will be 5 of which is the, for the cumulative probability.

This is 0.04. So, if I am writing 1 by 4, so this will be 1 by 2; and so you take it here, and therefore, the 0.04 n naught, minus phi of minus 0.04 n naught, right; from here which by symmetry of the normal, standard normal variant this is 1 minus of this 5.04 under root n naught minus 1 minus this. So, this becomes twice 5.04 root n minus 1.

And, now we want this to be greater than or equal to 0.95, and so when I compute the value of n for, by putting it equal to 0.95, then again for all values of n greater than or equal to n naught this inequality will be satisfied, right. So, the normal tables give me that; so if therefore, this probability becomes 1 plus 0.95 divided by 2. So, now, the normal tables tell me that corresponding to this value 0.975 that means the area under the normal curve is 0.975 that the corresponding value is equal to 1.96. That means, 0.04 under root n naught is equal to 1.96.

So, from the normal tables corresponding to this probability I get at this number must correspond to 1.96, and therefore, your root n naught is equal to this which is 49; and therefore, your n naught greater than or equal to 49 square. So, that means, for that many sample value or that many trials you have. So, this, with this, that this number n naught which is greater than or equal to 49 square your estimate of p which will be obtained by p hat; so that p hat will be essentially your; so what we are saying is p hat is summation or you can say x 4 9 square bar.

So, this estimate of your p will be within 0.02 of your original p with probability 0.95. So, this is, you know, interesting application of your central limit theorem. And, because we could reduce the whole thing to computing the standard normal probability we got the answer here, right. Now, again I will just try to have a variety of examples to show use of this central limit theorem.

So, now, here another question that is asked is if suppose you have x 1, x 2, x n, again a sequence of identical independent distributed random variables, and this random variables, right; probability x i equal to 1 is p, and probability x i equal to 0 is 1 minus p for all i; and again p is unknown, right. So, you want to estimate this probability. And, as

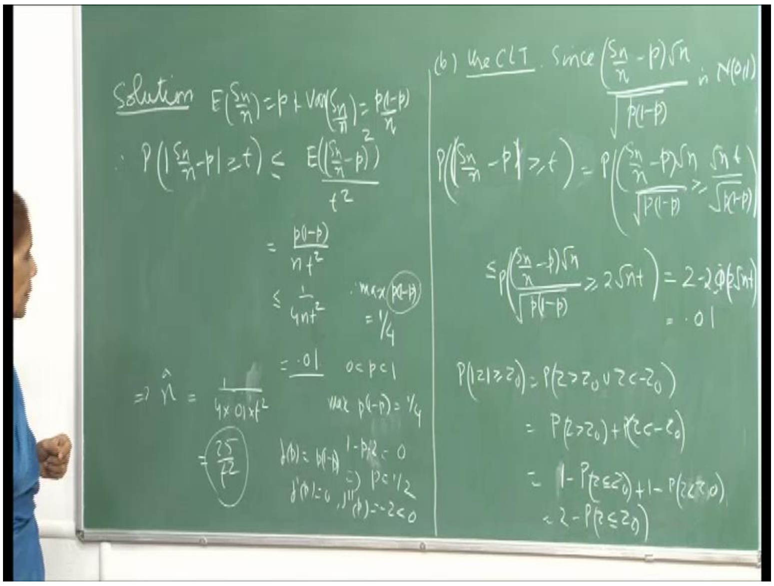
I told you at, of course, the Greek law of large numbers tells you that x bar will, x n bar will be a good estimate, provided n is large enough, right.

So, now, let us see, if you put; so now, let us define s n as x 1 plus, x 2 plus, x n; and let us fix t. So, here in this example I am trying to show you the, you know, the accuracy of central limit theorem. And, obviously, we expect better answer from the central limit theorem then if I just use the chebycheu's inequality. So, this is the whole idea, now by doing, now wanting to do this exercise.

And, so here let us see that; so the question is using chebycheu's inequality how large an n will guarantee, that this s n which is the sum of these random variables x 1 to x n. So, this divided by n minus p in absolute value is greater than or equal to t is; so if I have fixed the t; yes; so given a, t, then you want this probability to be less than or equal to 0.01, right.

And, we will now here compare because how large an n. So, chebycheu's inequality will also give me the answer, give me a value of n; and then central limit theorem will also give me a value of n. So, we will compare the 2 values, right. So, here expected s n upon n is p, and variance s n upon n is p into 1 minus p by n, right; because they are independent random variables identical. So, therefore, we will first use the chebycheu's inequality, and then compute through the, an estimate n; and then through the central limit theorem also.

(Refer Slide Time: 26:42)



So, applying the chebycheu's inequality because expect to tally of s n upon n is p and variance of s n upon n is p into 1 minus p by n. So, therefore, by the chebycheu's inequality probability that s n minus n minus p in absolute value is greater than or equal to t. So, this is less than or equal to expectation of this square divided by t square. In other words, the variance of s n upon n; the variance of s n upon n is p into 1 minus p by n.

So, therefore, by chebycheu's inequality we get this estimate, and here again we are applying the same logic. So, I hope that you can easily see that the maximum value because if you have 0 less than p less than 1, we are, earlier this thing also I used the fact that max of p 1 minus p is equal to 1 by 4; you can easily check, I mean let us just spend a minute and try to see; that means, this is the function of p, and I can find out the derivative.

So, the derivative would be 1 minus p by 2, right; and so this is I put this equal to 0. So, that gives me p equal to, I am sorry; this is 2 p because minus p square. So, the derivate is minus 2 p. So, this implies, let p is a half; certainly p cannot be 2 because p is lying between. So, this is a critical value, and to make sure that this gives you the maximum value the second derivate.

That means, the second if I call, if I am saying f p is p into 1 minus p, then f prime p when you put 0 gives you p equal to half, and f double prime p is equal to minus 2 which is less than 0; so; that means, the critical value that we have obtained will be maximizing value. And, you see, p equal to half gives you the maximum of this, and you can also prove this by concavity, and so on, anyway.

So, therefore, this is less than this; I mean here; and if I am putting the maximum value the, obviously, this equality will convert to inequality, is less than or equal to this. So, 1 by 4 n t square, right; and this should be equal to 0.01. So, we get the estimate for n; maybe I should put something like this, 1 upon 4. So, your n will be equal to 1 upon 4 t square into 0.01 which is 25 by t square. So, this is your chebycheu's estimate of this probability; I mean for value of n for which this probability would be less than or equal to 0.01, fine.

Now let us try to apply the central limit theorem. And, the central limit theorem says that this variant when you take as s n by n minus p divided by, the standard deviation which is p into 1 minus p by n, right. So, you divide by, that is the root n go upstairs, and this is approximately normal, standard normal, right; for lodging of n i, you can say that this is approximately this. So, now, you want to compute this probability greater than or equal to t.

And here again I will write this as; so I am standardizing it and therefore, dividing it by, dividing the whole thing by the standard deviation. So, that gives me the right hand side as root n t upon under root p 1 minus p. So, here again because this is greater, remember. So, then if I put the maximum value as again I do earlier this becomes smaller interval; and therefore, I should have said I think we have list out of the this things, sorry; the absolute value, right; and here also it should be the absolute value, fine.

So, the interval, I said if you put the maximum value here then this becomes smallest, and so the interval is smallest. So, if the probability is; and then we are wanting the; so for larger interval the probability would be higher. So, therefore, if I am taking the smaller interval, the probability I am wanting to be what was our, this thing. The problem that first, the problem stated was that this should be less than or equal to 0.01.

So, if I am saying that this should be greater, and so here if I am writing the maximum value, so what will it be; that means, I am wanting the mod z to be greater than or equal to z naught. And, if I am taking the smallest value here, so that means, if this is your 0 then you are asking for, yes. So, you are asking for this probability and this probability to be less than or equal to this.

Now, if I am taking the minimum value here; that means, I am putting the maximum value here, then, obviously, you will be taking larger area. So, therefore, the value of n which satisfies for maximum value here, will satisfy for all values of p, right. So, this is the idea. Now, let us see; so this greater than or equal to; so therefore, I am substituting 1 by 4 for p into 1 minus p which gives me half, and so that may comes, makes it 2 root n t, right.

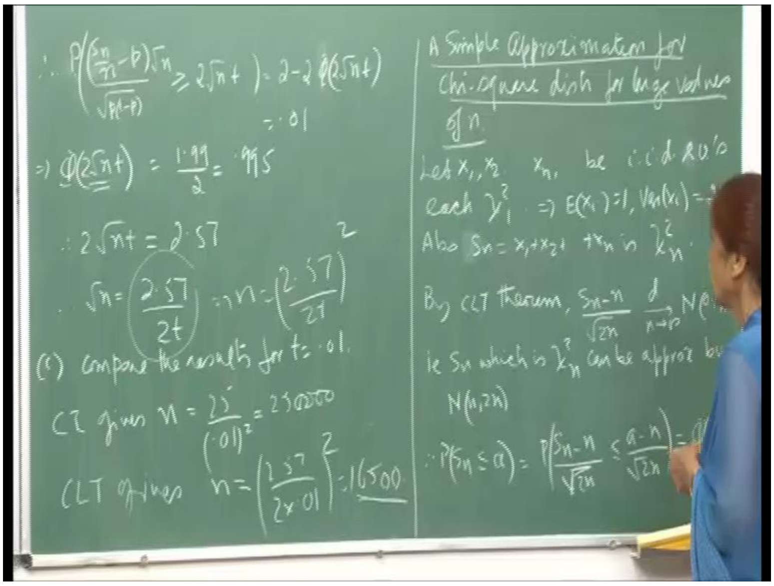
So, it is just the same argument, you would draw the figure and you can verify yourself. So, now, you see this is greater than or equal to 2 root n t. So, let me just show you the details here; that means, we are asking for probability z greater than or equal to z naught, right; which is the same as probability z greater than z naught union, probability z less than minus z naught, right. This will be the absolute value; as I said z greater than z naught and z less than minus z naught, right.

Now, these are disjoint events. So, therefore, I can write the probability of the union as the sum of the probabilities. So, this will be probability z greater than z naught plus, probability z less than minus z naught. And so this becomes 1 minus probability z less than or equal to z naught, and probability z less than minus z naught; see here again I can just show you the; so that means, if you have they are.

So, if you are wanting z less than this is minus z naught; you are wanting this probability; but that is the same if you write z naught here, this is the same as this probability, right. So, therefore, probability z less than minus z naught is 1 minus of probability z less than z naught. So, z less than z naught is this whole probability. So, from 1 minus I will get this which is equal to probability z less than minus z naught, and therefore, I get this.

So, this is 2 minus twice probability z less than or equal to z naught, where z is your standard normal variant. And, so the same thing I have used here, right. This is your z naught. So, this whole probability is 2 minus twice phi 2 under root n t which should be equal to 0.01.

(Refer Slide Time: 34:37)



So, like you are finding the value of n for which this will be equal, and then higher values of n this will always be less than 0.01. And so this is 0.01, therefore, this tells me that 5 2 root n t; if you bring this to this side and this here, something wrong. So, you bring this to this side, and take this to this side. So, it will be 1.99 divided by 2 which is 0.995.

And so you look up these standard normal tables, and corresponding to this probability the corresponding value of the variant is 2.57. So, from the normal tables I get that when this number is 2.57 the corresponding normal probability from minus infinity to 2.57 is 0.995. So, therefore, this gives me root n as 2.57 upon 2 t. So, the number by this central limit theorem; and therefore, this imply that your n is 2.57 upon 2 t whole square, right.

So, this is our central limit estimate, and that is our chebycheu's inequality estimate; now if you, third part of the question is compare the results for t equal to 0.01. So, for t equal to 0.01 the chebycheu's inequality gives you a number value of n which is 250000, that means 250000. Whereas, the central limit theorem will only ask for this many sample values.

So, if you want to estimate the, get the sample size say that s n upon n defers from p in absolute value by; you know, this difference, so that means, the probability that this is greater than or equal to t is less than 0.01. So, that number for central limit theorem is much much smaller compared to the chebycheu's inequality. So, this was another aspect of central limit theorem which I thought we should have a look at.

Then, another usage of central limit theorem is to how we approximate the chi square distribution for large values of n because you see that somebody is already done the calculations for the central limit theorem for the, sorry, for the normal variant, standard normal variant. So, we can make use of those tables to compute chi square distribution large values; now, for when the n is large.

So, the idea here is that if you take x 1, x 2, x n, are again identically independently distributed random variables, each is chi square 1. So, this implies that expectation of x i is 1, and variance of x i is 2, right. And then, also we know from the reproductive property of chi square, in fact, through the joint m g f also we can show; and even otherwise we have seen that because each of them is independent identically distributed chi square. So, x 1 plus, x 2 plus, x n, will be chi square n, right.

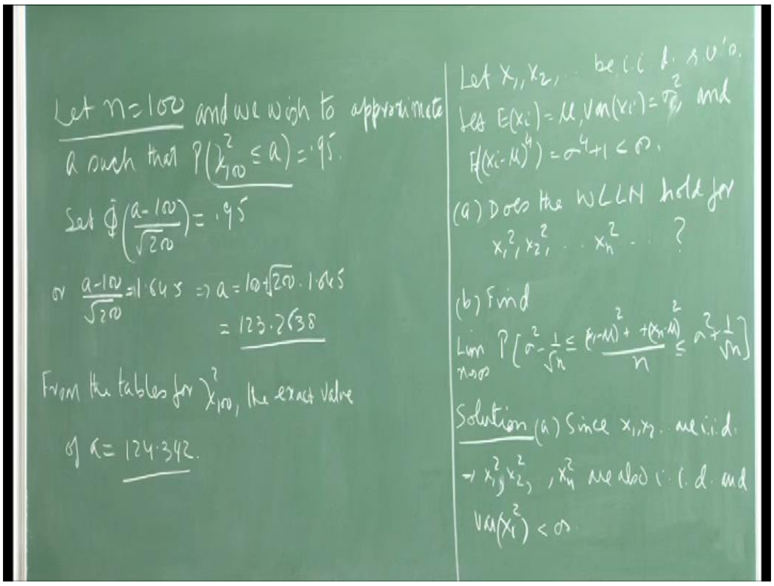
Through, this is, we have already seen it in one way that this when we sum of this independent identically distributed random variable. So, chi square has the reproductive property. So, this sum s n is chi square n. And later on we will also show through use of m g f how quickly you can say that the sum will be chi square n, if each is chi square 1, right, so anyway.

Now, by c l t theorem, s n minus n upon root 2 n, in distribution because we have standardized it, right, subtracting the mean of s n. So, s n will be, the mean of s n will be n because each is chi square 1. So, therefore, this is minus n upon root 2 n that will go into distribution to 0, normal 0 1, right, as n goes to infinity. So, or in other words, chi square is, the mean n normal; approximately you can say, mean n and variance 2 n.

So, this is the normal whichever the way you want to put it. So, therefore, you want to compute this probability, remember. For large n you want to be able to compute this probabilities using the normal tables. So, therefore, when I standardize this will be s n minus n under root 2 n. So, this will be become a minus upon under root 2 n, right, which is the value 5 of, a minus n upon root 2 n.

So, therefore, when n is large, I can, the central limit theorem will give me a good approximation. And so for large n this will be standard normal, close to a standard normal variant, and therefore, I can compute this probability for large chi square n by the normal, standard normal table.

(Refer Slide Time: 39:56)



So, now I said I will show you how we can approximate the chi square probabilities using central limit theorem. So, let n be 100, and we wish to; so I shown you that, you know, the formula; I give you the formula; that means, you can standardize the chi square random variable, and used, by using the central limit theorem you can get the probability. So, let us compare the values for n equal to 100, right.

And we will, so we want to approximate, a, so that chi square 100 less than or equal to the probability that chi square 100 is less than or equal to, a, is equal to 0.95, right, by using the central limit theorem, right. So, central limit theorem said that this can be converted to you see that phi chi square 100 minus, mean is 100 divided by, variance is 200, so under root 200. So, that becomes phi of, a minus 100 upon under root 200.

So, this will be approximately because we are saying, let us see how if n equal to 100, how good, is it large enough for a good approximation. So, if you put this equal to 0.95 then by the standard normal tables the 0.95 probability corresponds to this value b equal to 1.645. So, therefore, you equate these 2 numbers, and that gives you, a equal to 100 plus under root 200 into 1.645; and that comes out to be 123.2638.

So, using the central limit theorem, and then we just standardize this variant, right. And then, we said that if n is large enough this must be approximately normal 0 1. And so from there I get this probability. Now, from the tables for chi square 100 if you compute the exact value, then exact value of, a, that comes out to be 124. 342. So, you can see that the approximation is really very, very good; and this is for n equal to 100.

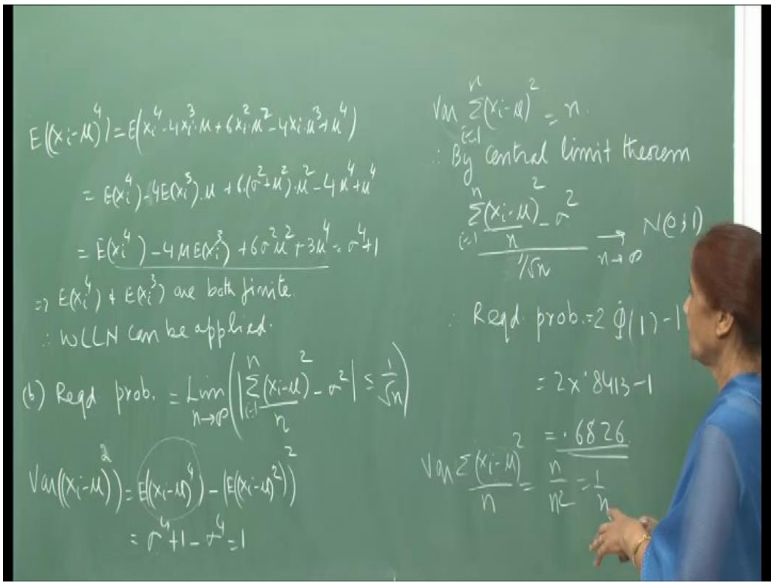
So, the point we are making is that we know if your n is larger you will get a better approximation, better than this; even the difference will be only at the decimal places. And so you can do, you can use your standard normal tables for computing these probabilities. So, this was one another point that I wanted to make about using central limit theorem results.

So, now, another application of central limit theorem; the question is if x 1, x 2, x n, again are identically independently distributed random variables, and you are given that expectation of each variable is mu and the variance is sigma square, and also that the expectation of x i minus mu raise to power 4 is sigma square plus 1, which is, that is an infinity because sigma you are taking to be a finite number.

So, the first question is does the weak law of large numbers hold for x 1 square, x 2 square, x n, I mean this sequence of square of the random variables. And, the second question is to find limit of this probability where x n minus mu whole square plus x n minus mu whole square divided by n. So, we are talking in terms of the squared random variables. So, because this central limit, the weak law of large numbers holds for x 1, x 2, x n, because your variance and mean are finite.

So, now, the question is that is the, does the weak law of large number holds for x 1 square, x 2 square, and sequence of squared random variables. So, therefore, we need to for the, for answering part a I need to say that the variance of each is x i square is finite, right. And, for this I have just made this calculation that, you know, if you open up this expectation of x i minus mu raise to 4 because that is what you given, as sigma 4 plus 1.

(Refer Slide Time: 44:21)



So, yes; and we miss this point, that for the weak law of large numbers I need to say that this x i squares are identically independently distributed, and they have finite variance. So, now, we have done enough of this thing to say that if x 1, x 2, x n, are identically distributed then obviously, x i squares are also identically distributed, right; this much of probability theory we have done so far. And then, that they would be independent can also, will also follow, right; because if x 1, x 2, x n, are independent then these will also be independent. So, I am sure you can work it out by all that we have done in the course by now.

And now, to show that the variances of this x i squares are finite which will be the same. So, I have just done this exercise, and surely there may be other ways of doing it also, showing that the variance is finite. So, any way I just opened up this expression; you know, then I took expectation inside. And, you can see that, here for example, expectation x i cube, then into q, and then when you can get 6 x i square; now expectation of x i square I know which is sigma square plus mu square, right; because that is already given to me.

And then this expectation x i’s mu. So, therefore, this is what you get- mu 4 minus 4, mu 4 plus mu 4; and here it is 6 times sigma square plus mu square into, mu square, right. And, then, you can write down this thing here. And, this then tells me, see if example here, this is 3 mu 4, because this 6 mu 4 minus, 4 mu 4 plus, mu raise to 4. So, that is 3 mu 4; and this is c sigma square mu square.

So, you get something like me these 2 expectation of x i raise to 4 minus, 4 times mu x i cube; this is equal to a finite number. And therefore, we can conclude that both are finite, right; and so weak law of large numbers can be applied. So, therefore, I needed this condition. If the x i’s are identically independently distributed, and if the fourth power expectation about the mean is finite then the weak law of large numbers can be applied.

And now, this probability we just need to again standardize our, this thing. So, summation, i varying from 1 to n, sigma x i square whole square. Now, variance of x i minus mu whole square summation, variance of each x i; I mean when you want to compute the variance of x i minus mu whole square then you are looking for; I should have said this is variance of x i minus, that is ok; variance of x i minus mu whole square that expectation of x i minus mu raise to 4 minus, expectation of x i minus mu whole square, whole square, right; this expectation is squared, right.

So, which now, since we are given this number which is sigma 4 plus 1, and this is expectation x i minus mu whole square is variance of x i which is sigma square; so raise to 2 will be minus sigma 4. So, the variance of each of this x i minus mu whole square is 1, right. And therefore, variance of summation i varying from 1 to n, x i minus mu; maybe I can rewrite this nicely so that it is readable.

So, then what we saying is that variance summation x i minus mu whole square, i varying from 1 to n, this is equal to n because each of them has variance 1. So, by central limit theorem this summation i varying from 1 to n minus sigma square divided by root n because each had variance sigma square, mean of x i’s minus mu whole square is sigma square; so then n sigma square divided by, n mean sigma square divided by, root n.

So, this goes to n 0 1, and distribution as n goes to infinity, and therefore, they required probability. So, therefore, I have divided by, here we are doing this; and yes, root n; if you want to write this as; I can just multiply by root n throughout; something is, you are wanting this to be less than or equal to 1. I am making it x i minus mu whole square, n minus sigma square; and then you are dividing by n, by root n; the required, this thing, we are looking for 1 by root n. So, variance x i minus; this is summation; variance summation x i minus mu whole square divided by n will be n upon n square which is 1 by n.

So, therefore, you will divide by 1 by root n, and so that, you know, if I bring 1 by root n here in the denominator then this becomes 1. So, therefore, because this goes to standard normal, this whole thing as n goes to infinity. So, this required probability, see absolute this less than or equal to; I have divided by 1 by root n; so this becomes less than or equal to 1. So, this is twice phi 1 minus 1. We have already done this so many times in the absolute values.

So, this is 2 phi of 1 minus 1; and again from the standard normal tables this is 0.8413; so this number. So, therefore, 2 into 0.8413 minus 1, and that come out to be this. So, integrate probabilities and so on, there is no n 2 in the results; and I will try to, I mean I think I will continue with the discussion on the central limit theorem in the next lecture also.