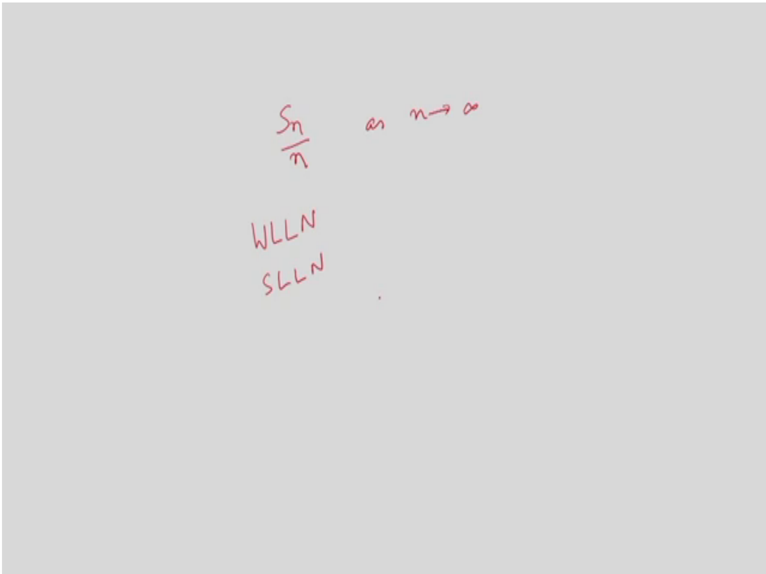
Advanced Topics in Probability and Random Processes
Prof. P. K. Bora
Department of Electrical Engineering
Indian Institute of Technology, Guwahati

Lecture - 11
Central Limit Theorem

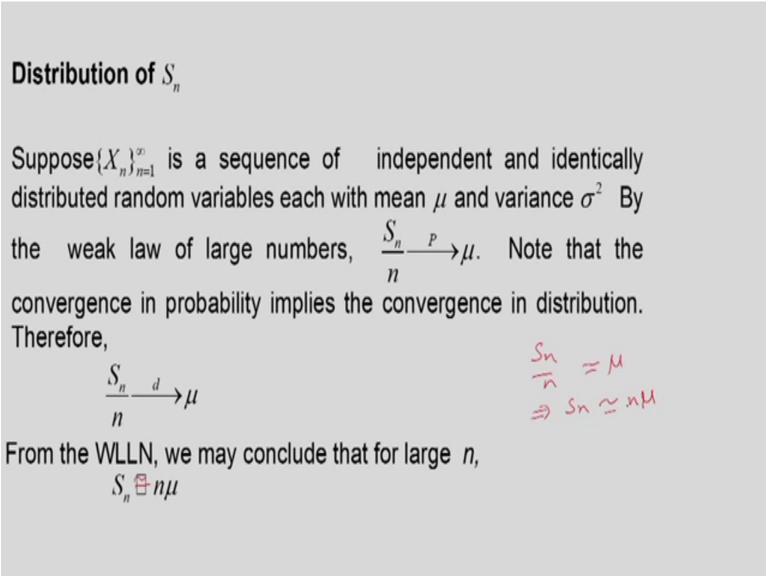
Central Limit Theorem, in this lecture we will talk about central limit theorem. This is one of the most remarkable result in probability theory and it has theoretical important as well as lot of applications in different areas of science, engineering and social sciences.

(Refer Slide Time: 01:00)



So, we will recall what we have discussed so, far we have interested to find out the behaviour of S n by n that is the sample average as n tends to infinity. So, this is our goal and in the last class we discussed about two important result that is WLLN and SL. So, weak law large number and strong law of large number this say about the behaviour of S n by n as n tends to infinity.

(Refer Slide Time: 01:36)

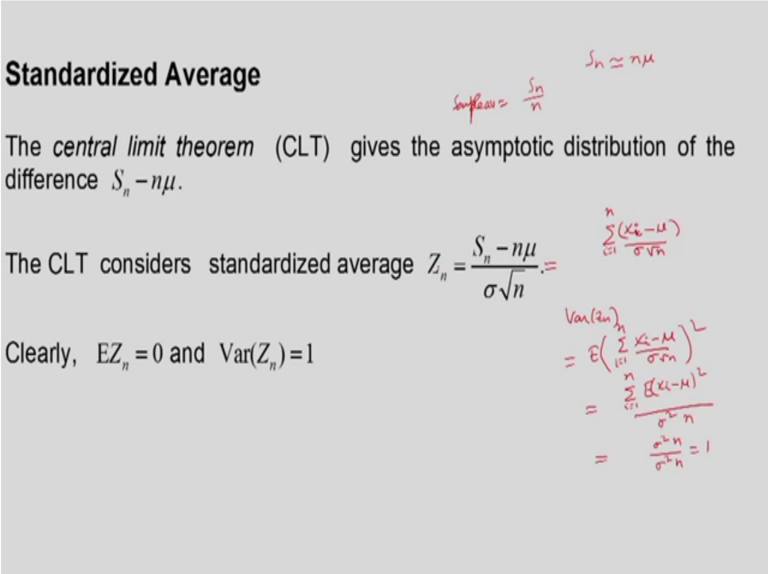


So, let us consider this special case that is suppose X n to be sequence of independent and identically distributed random variables each with mean mu and variance sigma square. Now, according to weak law of large number S n by n converges to mu in probability that is weak law of large number and strong law of large number, similarly it converges almost sure to mu in this case because it is iid.

Our goal is what is the distribution of S n, but it says that S n by n as n tends to infinity it converges to a fixed quantity mu, but is there any deviation from that mu or how does S n behaves when n is large. So, it will be approximately equal to suppose when n is large because, S n by n will become mu implies that S n will be approximately equal to n mu. But, how much deviation will be there that we do not know from this two laws, weak law of large numbers and strong law of large numbers.

Also, we know that S n by n converges in probability to mu. Now, convergence in probability implies convergence distribution therefore, S n by n convergence in distribution to mu. So that means, the distribution of S n so, S n will become approximately equal to n mu and as a consequence of weak law of large number S n will be fixed number. It is the approximately equal to n mu, but there will be no deviation, there will be no distribution for S n.

(Refer Slide Time: 03:45)



But, actually there will be when we consider S n for large n S n will be approximately equal to S n is approximately equal to n mu; that means, it will deviate from n mu and now what you how we can find out this deviations. So, what is the probability of any deviations? So, those type of answer we have to give and that is that is a great practical importance. So, how we can proceed? So, in weak law of large numbers we considered S n by n so, that is the sample average sample average.

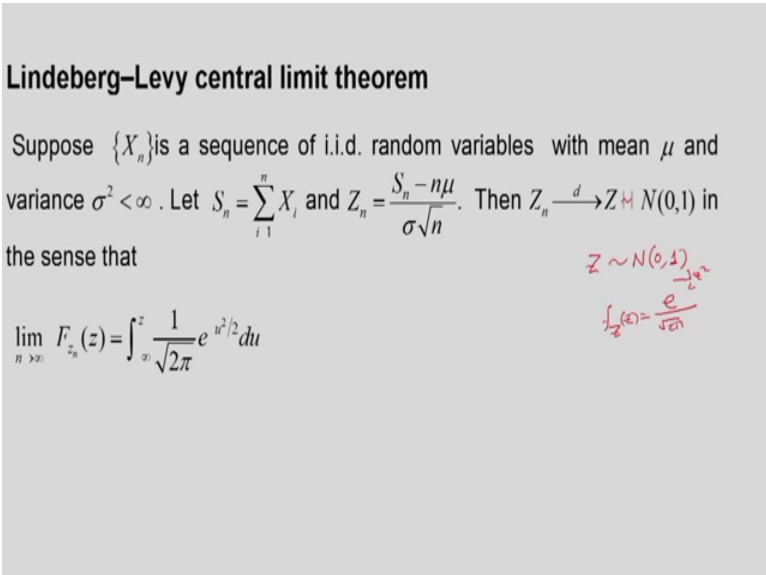
So, sample average convergence to true average that is the large numbers, but in the case of central limit theorem we will find out the asymptotic distribution of suppose, S n minus n mu that is the deviation. Now, here we consider the standardized average unlike sample average here we standardize the average. So, instead of dividing by n so, S n minus n mu this is the deviation divided by sigma we are doing sigma into root n.

So, instead of dividing by n we are dividing by root n. So, this is equal to we can write this is equal to that is summation X n X i minus mu, i is equal to 1 to n divided by sigma into root n. So this is the quantity we are considering. Now, this random variable now, Z n is a standardized random variable in these sense that E of Z n E of Z n, if I find out E of Z n then there will be here E of X i will be mu. So, mu minus mu will become 0. So, that way you have Z n is equal to 0, it is a 0 mean random variable.

Similarly, variance of Z n will be equal to how do I find variance of Z n is equal to E of summation X i minus mu divided by sigma into root n, i going from 1 to n whole square so, variance of Z n. Now, these are these are independent random variable. So, because of that this variance of this sum will be equal to the sum of the individual variances. So, that way we can get that this is equal to summation E of X i minus mu whole square, i is equal to 1 to n divided by because it is square sigma square and root n square is n and this will become n sigma square again.

So, that is equal to sigma square into n divided by sigma square into n that is equal to 1. So; that means, variance of Z n is equal to 1. So, we are now considering the asymptotic behaviour of Z n, how Z n is distributed when n is very large.

(Refer Slide Time: 07:39)

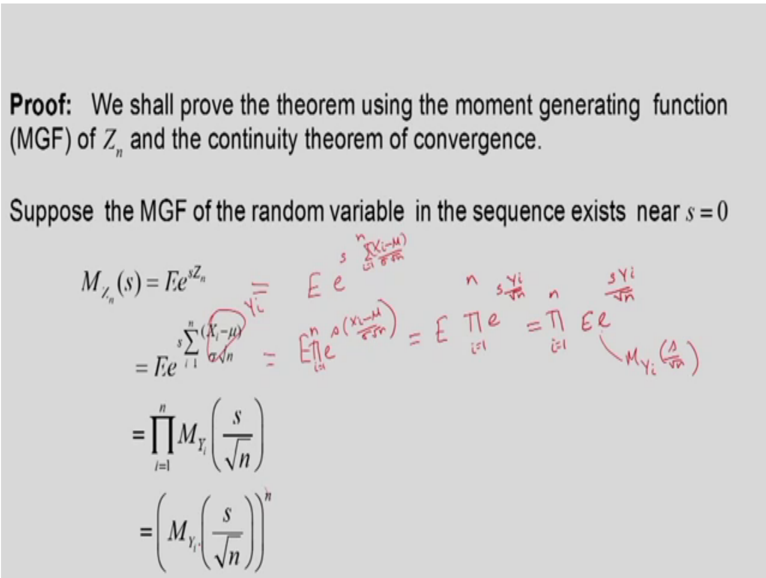


So, the central limit theorem gives answer to this question, what is the distribution of Z n and there are different statements. We are considering a simple statement, the simplest case that is the Lindeberg-Levy central limit theorem. Here we consider X n to be a sequence of i.i.d. random variables. So, here we assume that X n is i.i.d. Independent and Identically Distributed, but that is not very critical. We need only i.i.d. condition independent condition, but for simplicity we consider the case of i.i.d. independent and identically distributed sequence.

Now, as earlier we define S n is equal to summation X i, i is equal to 1 to n Z n is the standardised average. So, Z n is distributed as that is we have to prove that according to the central limit theorem, Z n converges in distribution to Z; where Z is normal random variable standard normal random variable. What does it means? N is so, this Z is standard normal 0 1; that means, it has mean 0 and variance 1. So, that way f z of z is equal to e to the power half of z square divided by root over 2 pi into 1 root over 2 pi. So, that this is the PDF for z.

So, what we are telling that if X n is a sequence of i.i.d. random variables with mean mu and variance variance sigma square which is finite of course, let S n is the sum of the sequence and Z n is the standardised average. Then Z n in distribution converges to standard normal random variable. So, what does it mean that limit of the CDF F z n of z that is the CDF of Z n that standardized average, the limit as n tends to infinity will be the CDF of standard Gaussian. So, this is the standard Gaussian if I integrate from minus infinity to z that will be will be the limit of F z n of z. So, that is the Lindeberg-Levy central limit theorem.

(Refer Slide Time: 10:43)



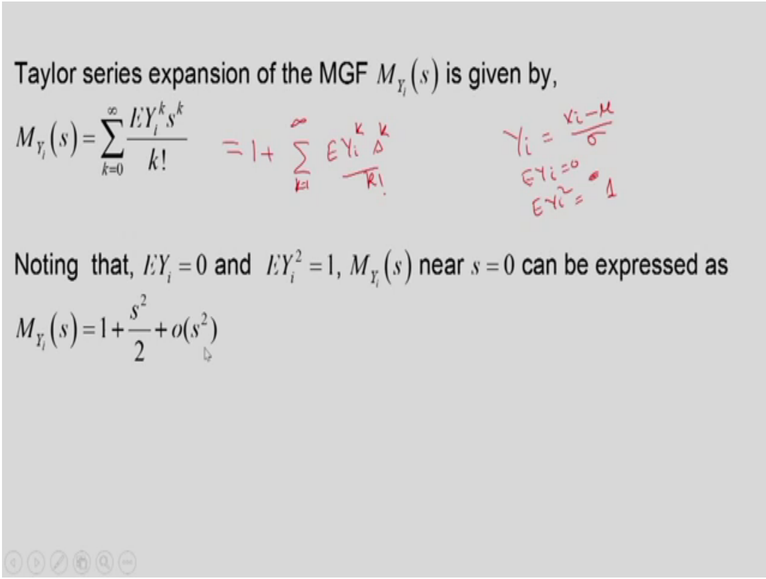
We will try to put this using moment generating function of Z n and the continuity theorem of convergence. That means, if moment generating functions sequence converges then corresponding distribution functions will also converge. That is the continuity theorem, we discussed earlier. Now, we will prove by a moment generating function we know what is moment generating function, moment generating function of Z n will be equal to E of expected value of e to the power s Z n ok. So, that is equal to same thing I can write E of e to the power s, now Z n as I have defined that is equal to X i minus mu divided by sigma into root n i going from 1 to n. So, this is the moment generating function.

And now I needed this quantity suppose because, it is exponential we can write it as a product E of e to the power so, this is the sum is there. So, here I can write a product; product i is equal to 1 to n e to power s into X i minus mu divided by sigma into root n. So, let me call this part X i minus sigma let me call this random variable as Y i ok. So, then we can write in terms of this MGF of Y i that is our goal ok. Now, at this stage so, what we have got is that this is equal to E of product of i is equal to 1 to n e to the power s Y i by root n, we can write like this.

But this X is and equivalently Y is are independent random variable so, this expectation operation I can take inside. So, this will be product i going from 1 to n E of e to the power s Y i by root n ok. So, this is the moment generating function, now this we can write as moment generating function of Y i at point s divided by root n. That is why this part we can write as M Y i this part that is M of Y i at which point at point s by root over n. So, that way we get this expression.

And, since all this random variables Y 1 Y 2 up to Y n they are identically distributed therefore, this probability will be there will be only one moment generating function MGF. So, we can write it as M Y i s by root n to the power n. So, this is the result we got. So, f m Z n of s is equal to M Y i f point s by root over n to the power n. Now, this Y i is the 0 mean unity variance random variable.

(Refer Slide Time: 14:39)

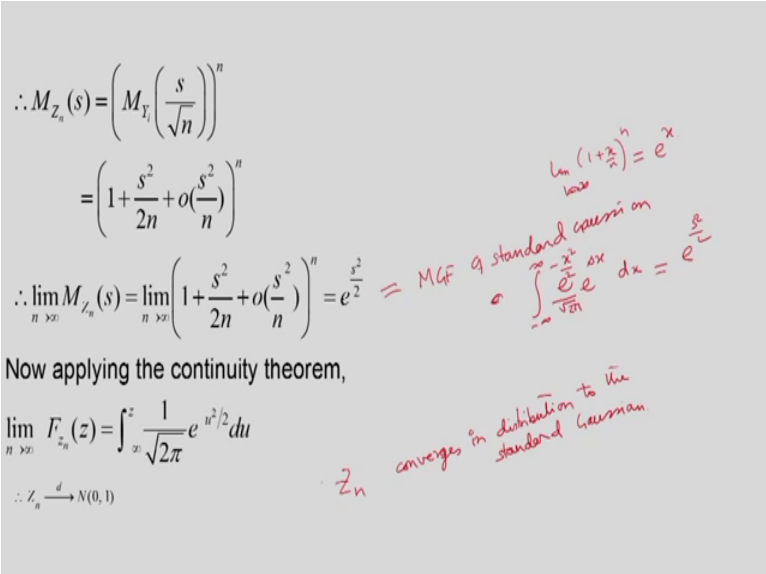


Now, moment generating function because it generates the moment, that is why it is called moment generating function and its power series expansion is given by this. This is the summation of some coefficient to the power into s to the power k divided by factorial k and this coefficients are the moments. So, that way M Y i s it is the summation of the power series. What is power series? This is s to the power k divide by factorial k with coefficient, coefficient is given by E of Y i to the power k. So, for this series first term will be always equal to E of Y i to the power 0 will be always equal to 1 that is equal to 1. So, it will start with 1; that means, we can write it also as 1 plus summation. So, summation k is equal to 1 to infinity E of Y i to the power k s to the power k divide by factorial k.

Now, Y i; what is our Y i? Y i is equal to X i minus mu divided by sigma. So, E of Y i is equal to 0 and E of Y i square will be equal to E of that is equal to E of X i minus mu whole square divide by sigma square. So, that way that will be equal to 1. So, therefore this part is expansion we can write near s is equal to 0. This expression is for suppose reason of convergence, but here we have particularly consider in near s is equal to 0.

So, that we can write M Y i of s is as 1 1 is this one plus then E of y square is equal to 1. Therefore, k is equal to 1 it will be 0 when k is equal to 2 I will get this 2 s square by 2 and then this E of X Y i square is equal to 1. And, the remaining term we can write in small o s square this notation so, that this term will go down to 0 as s tends to 0, when s tends to 0 this term will go down quickly to 0 so, this is the M Y i of s.

(Refer Slide Time: 17:33)

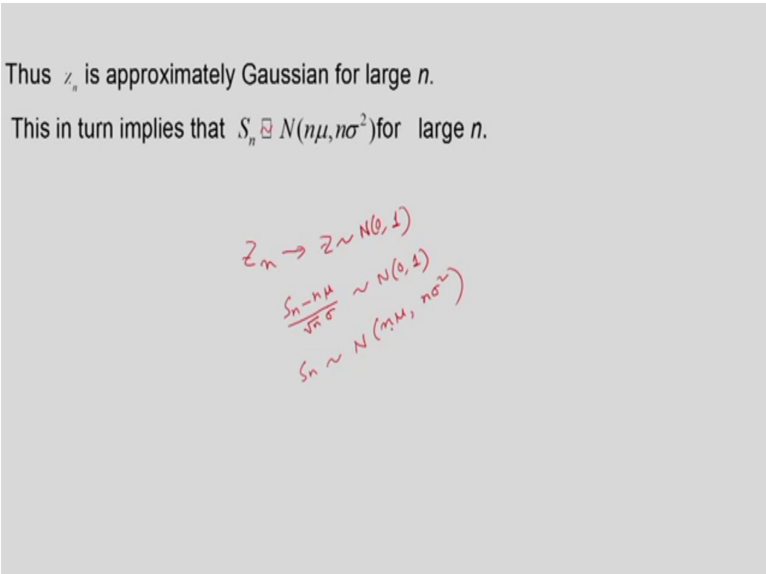


Therefore, our goal is to find out M Z n of s. So, that will be M Y i of s divided by root n to the power n. So, to the power n therefore, we can write as 1 plus s square by 2 n because it is s by root n. So, square when I make it square s square by n. So, that 1 plus s square by 2 n plus o of s square by root n square is n whole to the power n. So, this is the M Z n of s. Now, question is what will happen as n tends to infinity. So, limit of M Z n of s as n tends to infinity. So, this we can apply standard result 1 plus suppose, x by n to the power n limit n tends to infinity so, that will be equal to e to the power x. So, that result because this term will quickly go down to 0 as n tends to infinity.

So, we will have the power 1 plus s square divided by 2 n to the power n as n tends to infinity. So, that will be equal to e to the power s square by 2. Now, this e to the power s square by 2 is the MGF of standard Gaussian, MGF of standard Gaussian so, that Gaussian. So, what does it mean? Suppose e of suppose if I consider e to the power x square by minus x square by 2 root over 2 pi, this is the density function into e to the power s x d x. If I carry out this integration that, then I will get this e to the power s square by 2. Moment generating function of the standard Gaussian is this, this integration is also the by using the property of the Gaussian itself we can complete this integral.

So, what we have observed that limit of M then Z n of s as n tends to infinity is equal to e to the power s square by 2 which is the MGF of standard Gaussian. So, now we recall the continuity theorem, if MGF converges then corresponding CDFs also will converge. So, that we get that limit of F z n of z as n tends to infinity will be equal to the CDF of standard Gaussian that is given by this. So, what we have established Z n converges in distribution to the standard Gaussian. So, this is the remarkable result we have proved in the case of i.i.d. case, but we have used the strong assumption that moment generating function exist. But, more rigorous proof can be obtained, but we consider the simplest proof.

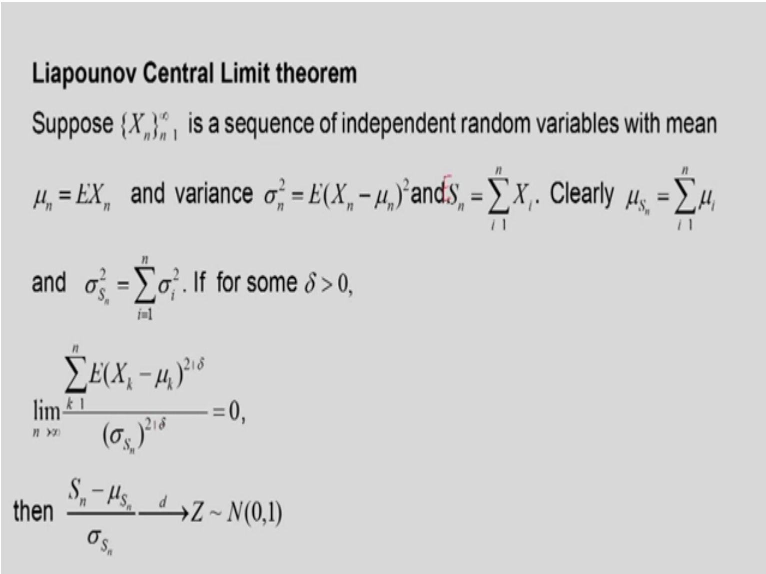
(Refer Slide Time: 21:35)



So, what is the implication of this that Z n it is says that Z n converges to Z which is standard Gaussian, 0 mean unity variance Gaussian. So, that means, what is Z n? That is S n minus n mu divided by root over n into sigma. So, that is approximately normal 0 1. So, therefore from this we can say that this S n itself when n is large because, this is Gaussian. So, from that we can find out that this S n itself will be distributed as S n will be distributed as normal it is normal. So, what will be the mean of S n? That is equal to n mu and variance of S n is n sigma square.

So, this is the distribution for S n approximate distribution now, it is normal with mean n mu and variance n sigma square. Earlier we establish that S n is approximately equal to n mu there is no deviation, but now we have learnt how to characterise this deviation and that is normally distributed. Now in our derivation that independent and identically distributed that was the assumptions, but that need not be the case; there are different statements and the more important is that S n should be a sequence of independent random variables.

(Refer Slide Time: 23:35)

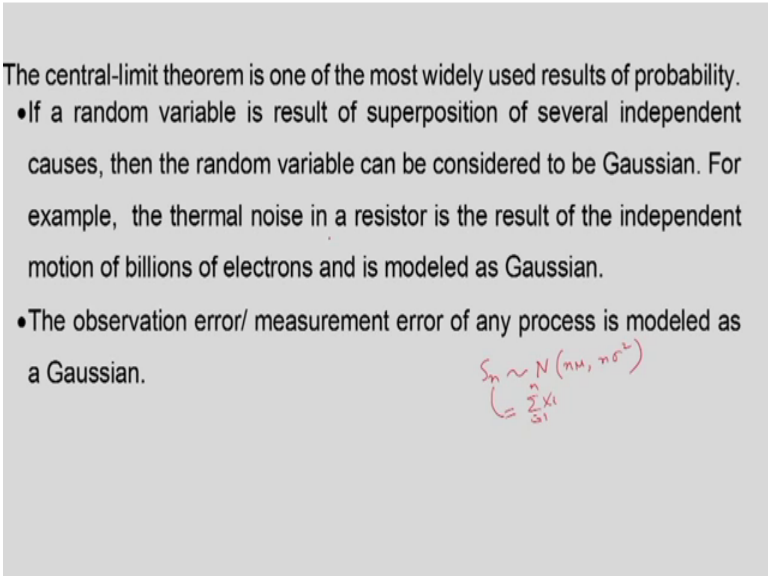


For example, there is a Liapounov central limit theorem. So, according to that suppose this is a sequence of independent random variables. Now, this is knows i.i.d. because is X n will have different mean mu n and different variance sigma n square. And, S n you define as summation of X i as earlier and clearly if I have to find out E of S S n that is mu S n E of S n, if I have to find out that is equal to mu S n is equal to summation of the individual means. So, that way mu of S n is equal to summation of mu i, i going from 1 to n. Similarly, sigma S n square, because of independence it is some of the variances that is sigma i square; i going from 1 to n.

So, we have a sequence of random variable, independent random variable with mu n and sigma n define. Now, suppose for some delta; delta may be 1 2 etcetera, this limiting condition is satisfied E of X k minus mu k to the power 2 plus delta. So, this is the central moment of order 2 plus delta, if we take the summation from k is equal to 1 to n and normalise by corresponding variance sigma S n square variance is this, that also we will consider some power 2 plus delta. If this limit they limit of this ratio goes down to 0 as n tends to infinity then this quantity S n minus mu S n mu S n is given by this divided by sigma S n, sigma S n is given by this that will be distributed as normal distribution.

So, this is also known as the Liapounov central limit theorem. So, according to that only independence condition is that required and some restriction on the moments. But, essentially we have to consider a sequence of independent random variables with finite mean and variance then we can apply the central limit theorem.

(Refer Slide Time: 26:09)

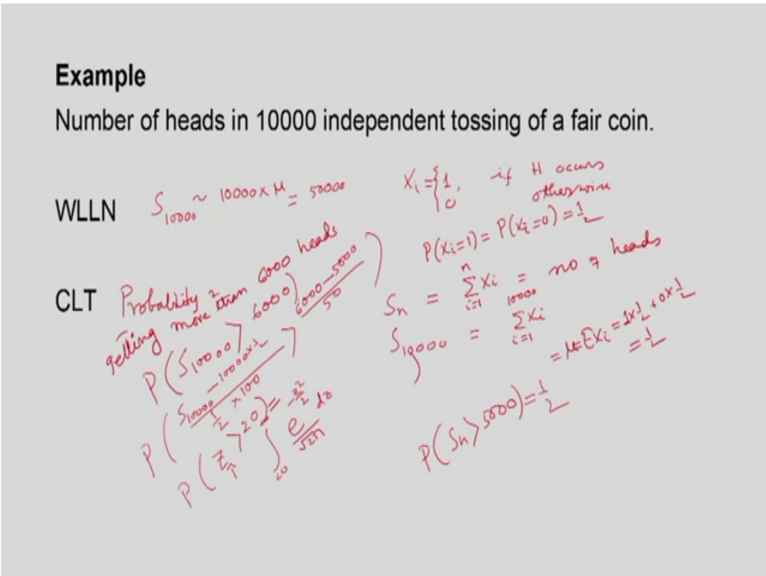


Now, let us try to interpret this result our S n, what we have established that S n is approximately Gaussian with mean n mu variance for i.i.d. case for example, variance n sigma square so, this is the distribution of S n. Now, when it happens when S n is equal to that is this S n is equal to summation X i, i is equal to 1 to n, S n is a summation of independent random variables that is important. So, if a random variable is a result of superposition of several independent causes this I can consider to be causes.

So, S n is a result of several independent causes, it is a superposition of several independent random variables. Then the random variable can be consider to be Gaussian, a because it will be approximately Gaussian. For example, you consider the thermal noise in a conductor. Now, this noise happens because of the random motion of electrons. Now, there are millions of electrons in a small conductor. Now, the noise voltage is result of the motion of so many electrons. So that means, it is the result of superposition of the voltage generated by is electron because, of the thermal motion and they are they are motions independent. Therefore, these are independent random variable therefore, this thermal noise whatever thermal noise is generated in that way that can be modelled as a Gaussian.

Similarly, observation error measurement error etcetera in is always modelled as a Gaussian. So, that way Gaussian plays an important role in not only in signal analysis, but modelling and then filtering etcetera various applications. So, that way this central limit theorem as its important in engineering, in science, in social sciences and whole about probability is applied central limit theorem has a role to play. So, that way we have seen the central limit theorem.

(Refer Slide Time: 29:10)



Now, we can consider one example suppose number of heads in 10000 independent tossing of a fair coin. So, these that means, if I consider the result of one tossing of a coin that is X i is equal to 1 suppose if head occurs 0 otherwise, if tail occurs if tail occurs then it will become 0. So, that way X i is a Bernoulli random variables.

So, in this case it is fair coin. So, P of X i is equal to 1 that is equal to P of X i is equal to 0 is equal to half. Now, number of heads in 10000 independent tossing of fair coin that means, S n in this case of course, S 1000 S n is equal to that is summation X i, i is equal to 1 to n. So, this will give the number of heads because, for single head if it is head we will get 1. So, if we count all once then we will get S n that is the number of heads. So, in this case S 10000 so, that is equal to summation X i, i is equal to 1 to 10000. Now, according to weak law of large number I know what is E of X i, E of X i that is equal to mu E is equal to E of X i that is equal to now, X i can take 2 value 1 into with probability half plus 0 with probability half, that is equal to half mu is equal to half.

So, according to weak law of large number this S 10000 will be approximately equal to 10000 n that is the n into mu that is equal to mu is equal to half. So, this will be approximately equal to 5000. So, this is the weak law of large number, but within the weak law of large number we cannot consider any deviation. But, practically there will be deviation we may get clearly we may get near 5000 heads nearly 5000 number of heads, but it may vary to 5500 or 4500 like that. How to tackle that? Now, CLT gives me a way to tackle that. So, suppose if I ask the question, what is the probability?

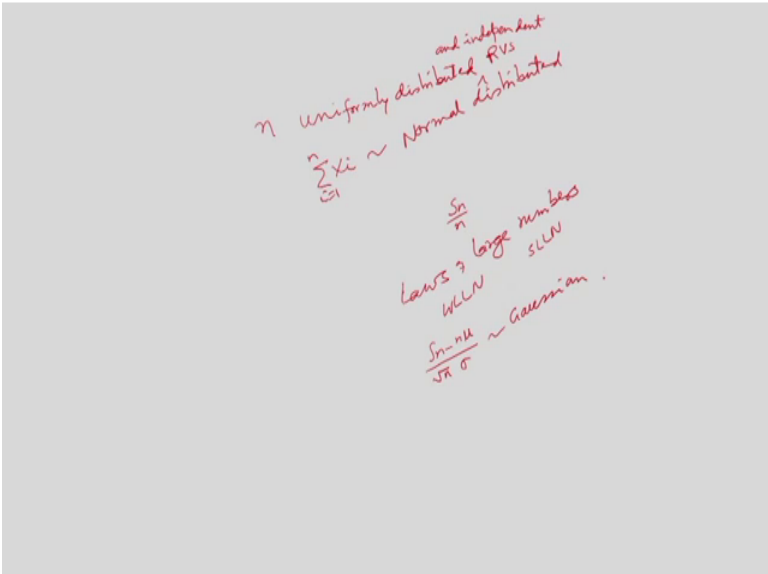
So, probability of getting more than; suppose 6000 heads. So, what we want to we want to find out what is the probability there S 1000 S 10000 that is greater than 6000. So, this probability if I have to find out now, I know that this S was 10000 this is same as probability that S 10000 minus n mu, n is here 10000 into mu is equal to half divided by now root over n sigma is half into root over n; root over n is root over 10000 is 100.

So, this is greater than now, same thing we have to do the right hand side also 6000 minus this is 5000, 5000 divided by this is may be 50 ok. So, this probability we have to find out. Now, I know that this is standard Gaussian so probability that this is Z, Z is standard Gaussian. So, this Z is greater than now this is 1000 divided by 50, 1000 divided by 50 that is 20, Z is greater than 20. So, this quantity will be equal to it because, it is a standard Gaussian so, it will be integration from 20 to infinity it require minus Z square by 2 d z divided by root over 2 pi.

So, this integral we can evaluate using the table or standard q function type of tool and we can evaluate this probability so, that way we can find out what is the probability that Z n or S n will deviate from the expected value that is 5000 by some amount. So, this is the utility of the central limit theorem. For example, if I say what is the probability that this probability that S n is greater than 5000. So, we can because 5000 is the mean, this is a standard Gaussian. So, what is the; this is a Gaussian so, for the Gaussian what is the probability that the random variable is greater than the mean. So, this we can apply the CLT and we can so, that this is equal to half.

So, that way we discuss the utility of the central limit theorem. So, in this lecture we covered the central limit theorem. This central limit theorem says that sum of independent random variables is approximately Gaussian, when it is large suppose n is large the sum of independent random variables. So, it is required that sum of independent random variables their PDF, their PDF will be if it is a continuous random variable all random variables concerned continuous then that sum will be PDF will be approximately Gaussian.

(Refer Slide Time: 37:18)



For example, if we consider suppose n uniformly distributed random variables. So, now in this case that means, summation X i, i is equal to 1 to n. So, there n uniformly distributed random variable. Now, because of this sum of independent random variables so, uniformly distributed and independent we have to write independent. So, what will happen that this sum will be always Gaussian distributed this will be normal distributed ok. So, normal distributed and this what will be the mean of this distribution. So, each random variable mean we have to find out, then sum of these means will be the means of this normal distribution.

Now, this n can be suppose if n binomial random variables, suppose we if we consider n binomial random variables and they are independent suppose. In that case also this sum the CDF of this sum will converge to the CDF of the Gaussian because, this is a discrete sum now sum of discrete random variables, it cannot be continuous random variable.

But, its CDF will converge to the CDF of the standard CDF of the Gaussian random variables. So, that way this is the importance of the central limit theorem, that if we have independent random variables and if we get the sum then sum will be always Gaussian distributed.

So, that way we have considered now, three results that is we started with laws of large number laws of large number laws of large numbers; we have weak law of large numbers. So, basically we are interested how does S n by n behaves. So, if I consider S n by n then we have laws of large number and then within laws of large number, we have WLN WLLN and SLLN, Strong Law of Large Numbers.

So, this result had derived when we consider S n by n, but to derive how this S n for example, is distributed as n is large to get the tensor we consider not S n, but we consider normalization suppose S n minus n mu divided by root n this is the importance root n into sigma. So, this quantity is Gaussian according to CLD. So, that way in laws of large number we are interested in S n by n, here we consider another sequence S n minus n mu divided by root of n because, of that we got the Gaussian.

Thank you.