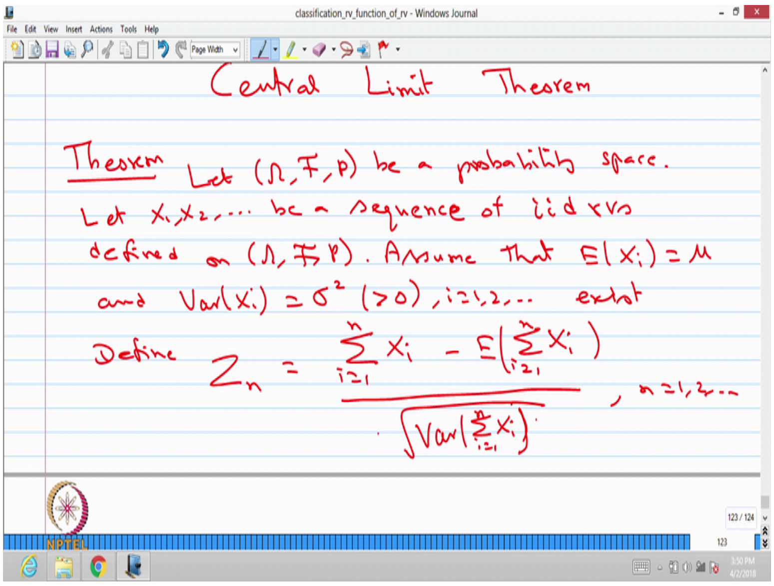
Introduction to Probability Theory and Stochastic Processes

Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture - 45

Now, we will move into the third lecture that is called the Central limit theorem.

(Refer Slide Time: 00:06)



So, this is a very important result in probability that is central limit theorem which has the wide application in many real world problems. Therefore, this theorem will be used again and again in many problems.

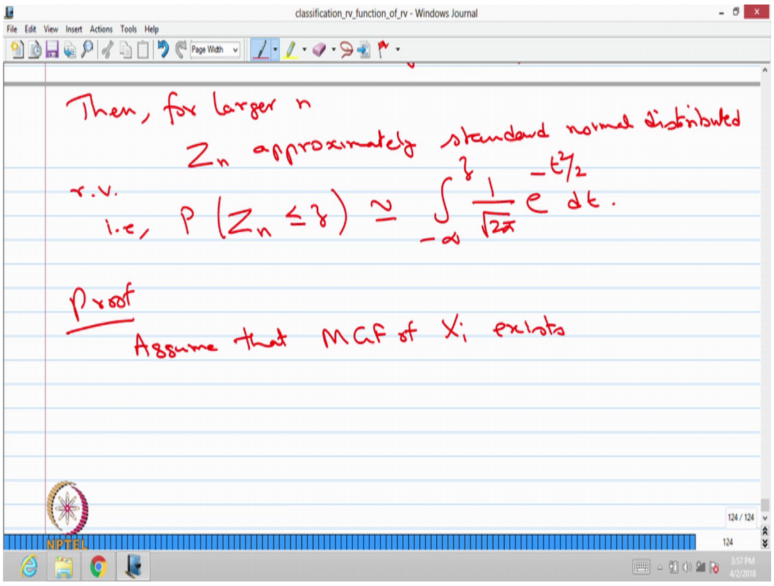
So, let me give the central limit theorem first, then I give the proof; then we will go for 1 or 2 examples of how to use the central limit theorem in the real world problems. Let me give the theorem first. Even though, there are many versions over the central limit theorem, first we will get the easiest version; because it is a introduction to the probability theory and stochastic process course. If the course is advance probability theory course, then we can go for 2 3 levels of a central limit theorem.

So, here we will present only the simplest version of the central limit theorem; whereas, we will discuss how the complicated version in the central limit theorem after I give the proof of the simplest one. So, we will give the simplest version of the central limit theorem. Let omega F capital P be a probability space, let X 1, X 2 and so on be a sequence of iid random variables defined on omega F capital P.

Assume that assume that expectation of Xi that is equal to mu and variance of Xi that is equal to sigma square which is greater than 0 for i is equal to 1, 2 exist; that means, we make sure that this sequence of random variables are iid as well as at least second order moment exist and the variance of each random variable is greater than 0.

Since I made it iid random variable, the sigma square is greater than 0 and also the finite quantity. And defining, defining the new sequence of random variable I call it as a Z suffix n that is nothing but sum of n random variables minus expectation of this sum of random variables divided by square root of variance of sum of these n random variables. I am defining a sequence of random variable for n is equal to 1 2 and so on.

(Refer Slide Time: 04:15)



What the central limit theorem says then, what the central limit theorem says, then for larger n, Zn approximately standard normal distributed random variable. Then for larger n, Zn approximately standard a standard normal distributed random variable; that means, that is the probability of Z n less than or equal to small z; approximately minus infinity to z, 1 divided by square root of 2 pi e power minus t square by 2 d t.

This is valid only for larger n that is very important and that to the cdf, CDF further a random variables Z approximately the integration from minus infinity to Z, 1 divided by square root of 2 pi e power minus t square by 2 d t that is nothing but the cdf of standard normal distribution. This is valid as long as X is or as long as X is or iid random variables defined and a probability space with at least second order moment exist and variance is greater than 0. And then, making a sum of random variables there my there minus there mean divided by the standard deviation that is approximately a standard normal distributed random variable for larger n.

Indirectly whenever you have a normal distribution with the parameters mu and sigma square; by subtracting the mean divided by the standard deviation that becomes standard normal distribution. So, the same thing we are applying in the Z n. The random variable is a sum of random variable that is a 1 random variable for fixed n minus their mean divided by the standard deviation; that means, this transformation is the transformation from normal distribution to the standard normal.

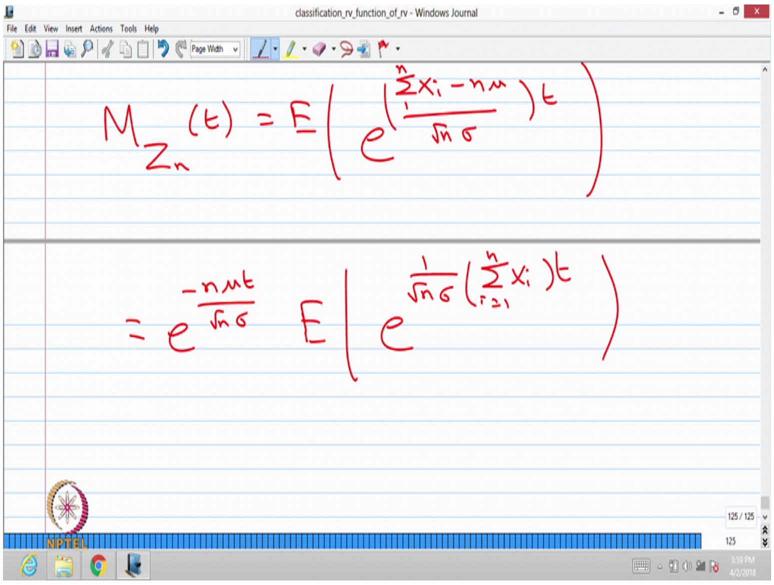
That means, indirectly when we say when we say Z n approximately a standard normal distributed random variable, indirectly what we are saying the sum of n random variable approximately a normal distributed random variable with mean expectation of that random variable with the variance, variance of sum of random variable for larger n.

That is a meaning of a Z n approximately a standard normal distributed normal variable that is equivalent of a sum of random variable is approximately a normally distributed random variable with the mean is expectation of a sum of random variable and the variance is variance of sum of random variables. And here the assumptions are very important it should be iid random variables with the at least a second order moment exist.

Now, we will go for proof of this theorem. For the proof we will make the assumption that m g f of each Xi exist; even though for some random variable m g f may not exist and here we made the assumptions only at least second order moment exist, that does not mean that m g f for moment generating function of each Xi is exist.

We make the additional assumption of m g f exist, then later we will relax the m g f exist then we can give the proof of it. So, without loss of generality we assume that m g f of Xi is exist for all the random variable because all are iid random variables. With that assumption will give the proof, later we can relax this also.

(Refer Slide Time: 09:46)

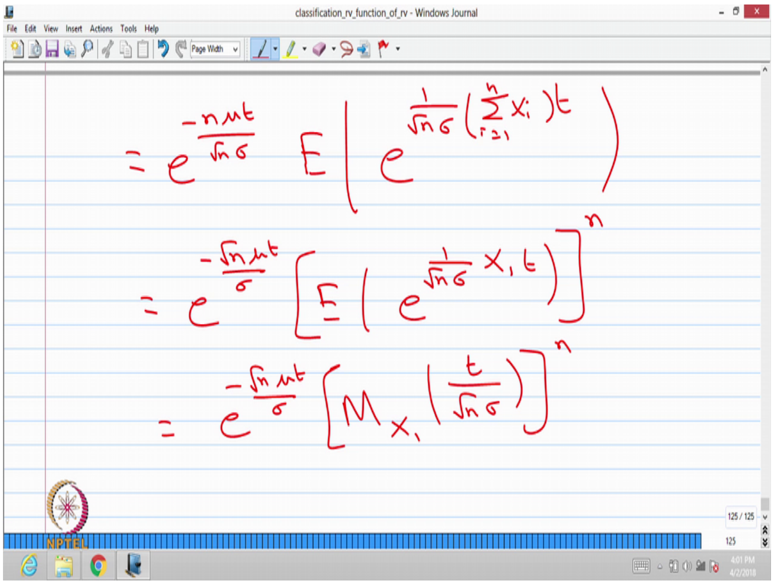


Let us go for finding out the m g f of Z n as a function of t moment generating function for the random variables Z n as a function of t that is nothing but expectation of e power sum of random variables 1 to n. Since we made a all are iid random variables, their mean is going to be mu n times mu divided by variance of sum of random variables; each random variable variance is sigma square. Therefore, sum of random variables is n sigma square. Here, you need a square root of variance; therefore, square root of n sigma as a function multiplied by t.

So, this quantity is going to be the m g f of the random variable Z n. This is possible as long as the m g f of Xi exist. Therefore, you made the assumptions m g f exist that is same as all the constant you can take it out. Therefore, it is going to be exponential of minus n times mu t divided by square root of n sigma multiplied by expectation of e power 1 divided by square root of n sigma.

Then the summation of Xi is 1 is equal to i is equal to 1 to n times t. This is same as e power minus square root of n mu t divided by sigma.

(Refer Slide Time: 11:39)

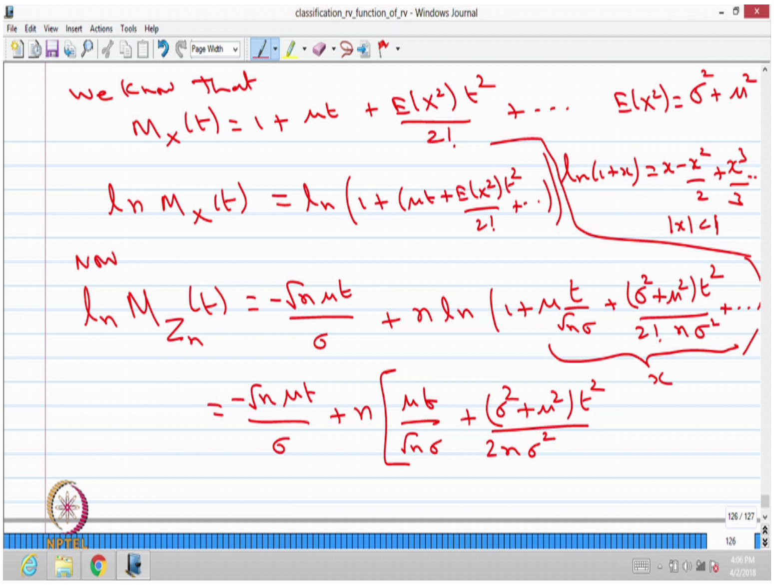


You can use the expectation of e power summation of X is t that is nothing but the all are iid random variable. Therefore, you can go for expectation of e power 1 divided by square root of n sigma for 1 random variable X 1 t.

After getting the expectation you can raise e to the power n because all are independent as well as identical that is same as e power minus square root of n mu times t by sigma. This is nothing but m g f of the random variable X 1 instead of t, you can write t divided by square root of n sigma, both are on the same; whether you write m g f of 1 divided by square root of n sigma X 1 of t or m g f of X 1 t is replaced by t divided by square root of n sigma, both are 1 and the same; this power n because of identical.

Now, we need the expansion of m g f for any random variable; then, we can substitute that, we know that.

(Refer Slide Time: 13:18)



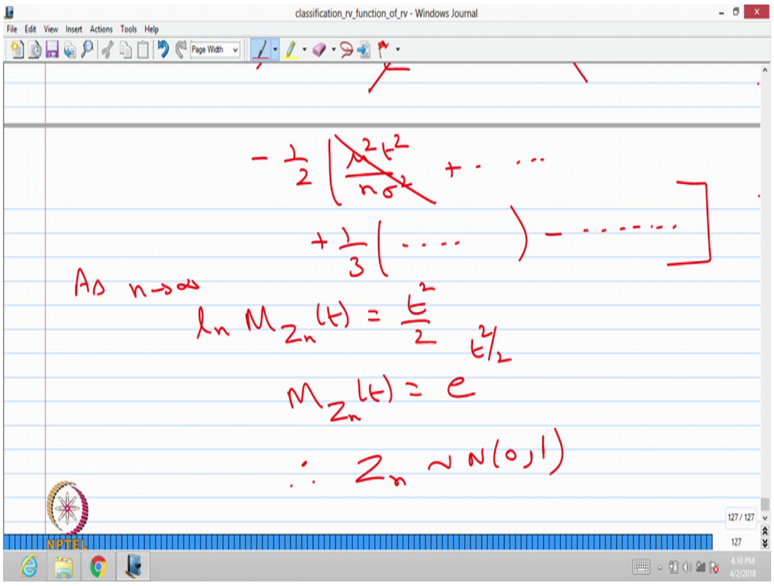
We know that m g f of any random variable X can be written as 1 plus mu t plus expectation of X square t square by 2 factorial and so on; again, you can write expectation of X square as variance of X. Suppose variance of X is sigma square plus mu whole square. So, one can write expectation of X square as a sigma square plus mu square; I am going to substitute little later by taking a logarithm of m g f of X t, I can use l n of 1 plus X as X minus X square by 2 plus X cube by 3 and so on provided mod X is less than 1. I can use this identity for the l n of m g f of X is the l n of 1 plus mu t plus expectation of X square t square by 2 factorial and so on.

So, I can make it as the l n of 1 plus all the other term, I can make it as the sort of X mu t plus expectation of X square t square by 2 factorial and so on. This I can keep it as a 1 plus X 4. So, I have not substituted l n of 1 plus X now, I am just writing l l n of the whole series as the 1 plus remaining terms as the X.

Now, I am going to apply the same logic for the m g f of Z n; that means, now l n of m g f of the random variable Z n of t that is going to be when you take a logarithm, it becomes minus square root of n mu t by sigma. Then, the remaining terms with the power; therefore, it becomes n power n becomes n times l n of 1 plus mu. Here, t is replaced by t by square root of a n sigma plus expectation of X square is sigma square plus mu square times t square by 2 factorial n sigma square and so on. This is going to be minus square root of n mu t plus sorry divided by sigma plus now I am going to apply l n of 1 plus X that is n times it is X minus X square by 2 plus X cube by 3. So, X is going to be this.

So, the first terms in the X that is mu t divided by divided by square root of n sigma plus sigma square plus mu square t square divided by n sigma square 2 factorial is 2. I am not going to write other terms of X limit as it is, whereas now I am going to write minus X square by 2 terms that is minus 1 by 2 times.

(Refer Slide Time: 17:45)



In the X square also I am not going to write X square of all the terms, I am going to write the X square of only first term that is mu square t square by n sigma square; all the other terms I leave it as it is. There is a reason behind that; I am not going to write other terms of X square.

Similarly, I am not going to write any terms for the X cubes, only I write 1 by 3 all the other term as it is. Like that there are some more terms some more terms for the expansion of l n of 1 plus X. This is going to be close bracket. The reason is as n tends to infinity, even though I use a word for larger n here, we are going for as n tends to infinity. The n in the numerator and many terms in the n in the denominator that canceled and the all the other terms will be in the form of 1 divided by n; not only that this one and this one cancel. Whereas, the sigma square plus mu square t square this one with the first term here that cancels.

So, the left out is sigma square t square divided by 2 n sigma square that will be cancelled with n in the numerator. So, you will have a only sigma square t square by 2 sigma square sigma square also cancel. So, you will left out with t square 2. Even though we have many more terms as n tends to infinity all the other terms vanish. So, you will have a as n tends to infinity l n of m g f of Z n is going to be t square by 2; all the other term vanishes as n tends to infinity. Now, I am taking a exponential both side; that means, m g f of Z n that is going to be e power t square by 2. If you recall the generating function for the standard distributions waves discuss for many discrete type random variables.

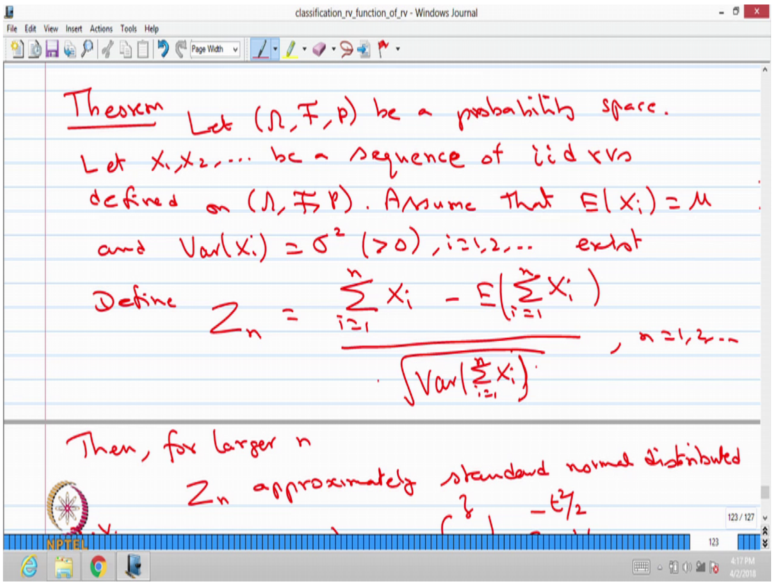
Similarly, we have discussed continuous type random variables m g f. So, if you compare the m g f of this with the m g f of standard distribution, you can conclude the you can conclude by using the uniqueness theorem of 2 different m g f s are same for all t, then both the random variables are identically distributed. So, you can conclude the Z n is standard normal distribution. So, this is valid for n tends to infinity; that means, for larger n the Z n approximately a standard normal distribution that is a proof. In this proof we have made assumption of m g f of exist.

Now, we can see what could be the proof or how the proof goes when you do not have a assumption of m g f. The similar derivation I can go for characteristic function. So, the characteristic function of Z n of t that is going to be expectation of e power the whole expression t, where t is replaced by i times t, where i is square root of minus 1. For that I do not need any assumption because the characteristic function exist for all the random variables; therefore, the characteristic function for Z n exist. So, I can directly compute the characteristic function of Z n.

In this result wherever the t I have to replace by I times t that is going to be the derivation of characteristic function. So, if I do the same derivation everything goes in the same fashion because I keep iid random variables mean is mu variance is sigma square and so on. Therefore, wherever there is a t it will be replaced by i times t. So, that will be cancelled wherever there is a t square that is going to be minus t square because it is going to be i square t square i square is minus 1. After you do the simplification till the as n tend to infinity, you will get the answer minus t square by 2 for the l n of characteristic function of Z n; that means, the characteristic function of Z n is going to be e power minus t square by 2 that is there result for the characteristic function for standard normal distribution, then we can conclude also Z n is approximately a standard normal distribution.

So, whether we made the assumption m g f or not the derivation is almost similar way to conclude that it is approximately a standard normal distribution. I said I am going to discuss the little higher versions of the central limit theorem.

(Refer Slide Time: 23:42)



Yes, see the theorem carefully I have made iid random variable. Suppose, if it is not identical distributed then you can find what are all the changes; that means, if each X is or not a identically distributed, then their mean will be mu is variance will be sigma i squares; that means, each one may have a different means. Still you can apply the theorem because Z n is going to be sum of random variable minus their mu.

So, whatever the mean mu i's, you add all them mu i's, find out the summation of mu i's; that is going to be the expectation. In this theorem, when they are identical it becomes n times mu if they are not identical. Then it becomes mu 1 plus mu 2 plus so on mu n. Similarly, the denominator here it is a square root of square root of n sigma, but if they are not identical, then you will have a sigma 1 square plus sigma 2 square and so on square root of that.

Still the derivation goes, but we cannot apply the power n. We cannot apply the power n the way we have done it here because of identical we got power n. So, when you go for derivation for non identical distributed random variable you have a individual m g f in the product form.

So, when you take a logarithm and so on, the expression will be huge. The process of a derivation may be tedious, but still as n tends to infinity you can conclude the same result. The derivation may be very complicated when they are non identically distributed, still we can go for it the same derivation.

One more observation, here we have use the independent random variable in finding the square root of variance of sum of random variables. Since all the random variables are independent the variance of sum of random variable is nothing, but the individual variances summation. If they are not independent, then you have to go for adding the covariance of any 2 random variables.

So, since we mean the assumption there independent random variable we are finding the individual variance, then we are sum it up; that is going to be the variance of sum of random variables. Otherwise you have to co use the covariance of any 2 random variables; that means, we can relax instead of they are independent random variable you can make the assumptions all the random variables covariance of any 2 random variables 0, that is enough.

You do not need a independent assumption. Independent is a strongest assumption comparing to the covariance of any 2 random variables are going to be 0 because the covariance of any 2 random variable 0, that does not imply they are independent. But if 2 random variables are of some random variables are mutually independent, then the covariance of any 2 random variables are going to be 0.

So, here in this theorem, I made a strongest condition; therefore, this is the simplest version of central limit theorem. Whereas, we can go for covariance of any 2 random variables are 0 that is enough to use the central limit theorem. One more observation over this central limit theorem, why this is a used in many situations?.

You see the theorem very carefully, we have not used any distribution for random variables Xi's and we have used the only the mean and variance of random variables and assumption of independent nothing else. Because of that this theorem is used in many real world problems; that means, many real world problems many random variables which we have created, those random variables we may not know the distribution of that. We may not know the distribution of those random variables, but we may know the mean and variance as a numbers.

We may know mean and variance of those random variables, even they are dependent or the dependency maybe very very minimal or we can ignore the dependency or we can make the usage of those random variables or independent or in the lighter sense we can use the concept of covariance of those 2 random variables are 0 with that assumption we can use this theorem. So, the big advantage of this theorem is there is no assumption over the distribution or we do not need the distribution of each Xi's; we need only the mean and variance.

Therefore, we can use this theorem to find out the probability of event using a standard normal distribution by approximating this random variable as a standard normal distribution. That means, whatever be the distribution of those random variables. Once we sum it up by subtracting there mean divided by the standard deviation for larger n we can always approximate in material of whether it is a discrete type random variable or continuous type random variables.

As long as there independent random variable, that can be approximated with a normal distribution; by normalizing it can be approximated with a standard normal distribution. Therefore, we use this theorem quite a lot in many real world problems.

Now, let us go for a few examples how 1 can use the central limit theorem.