**Probability and Stochastics for finance**
**Prof. Joydeep Dutta**
**Humanities and Social Sciences**
**IIT Kanpur**

**Lecture - 05**
# Law of Large Numbers and Central Limit Theorem

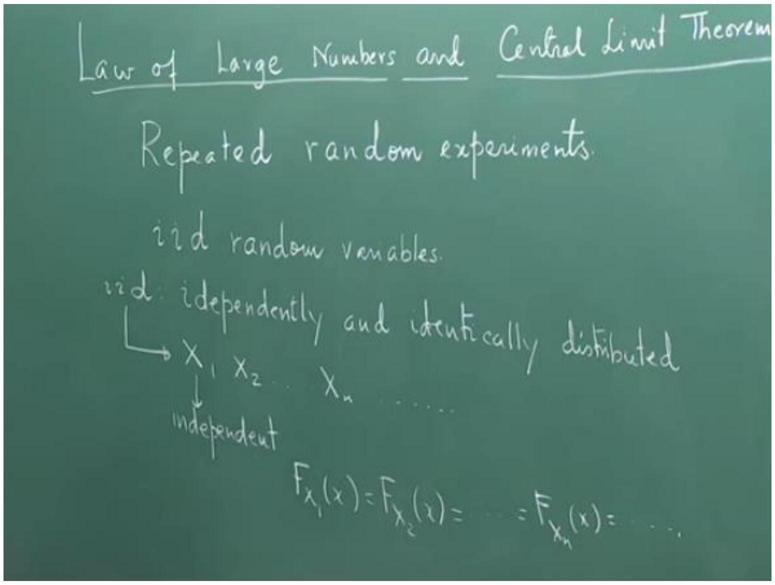
Today we are going to study something called law of large numbers and Central Limit Theorem. These are parts of any formal study of probability theory, but these are essentially about repeated experiments. We generate what is called random sample. Suppose we want to study something about the population of some you know population of population of the heights of the male members of a city.

You cannot really go and check the height of each and every male member. So what you really draw is a random sample so you really check the heights of certain number of male members of a city and then using that information, that data, you come to certain conclusion about the actual or the average height of the male members of that city or some more information about their heights.

So this information that you gather is essentially by this process called random sampling. So you do not look at the whole population, but you look at certain members of the population. So taking a random sample is same as taking a making a repeated experiment. So for example when you are going to measure the height of people, so you are making a repeated experiment. You are measuring the height of person 1, then person2, then person3, and so and so forth.

So your experiment is the same. You are measuring height and you do not know what is the outcome, but outcome is random so that is a random experiment. So this lecture is going to talk about repeated random experiments.

(Refer Slide Time: 02:05)

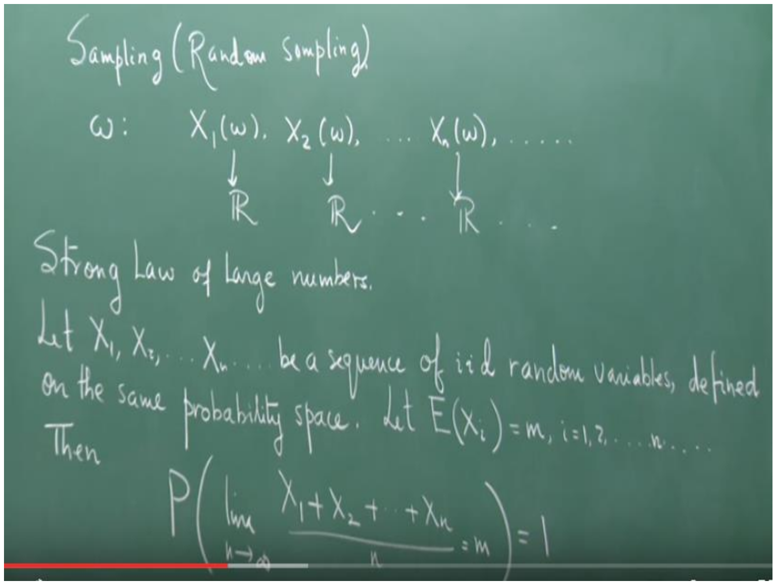



So to understand that we will first begin with the notion of iid random variables. So what does iid mean? IID essentially means independently and identically distributed. So the iid is a short form of independently and identically distributed. So the formal definition is that if you have a sequence of random variables, could be a finite sequence, could be an infinite sequence it does not matter.

If you have a sequence of random variable x1, x2, xn then iid means that these are independent, so iid means that iid this means that these are independent, so these are all independent and identically distributed means that distribution functions are same, they are the same functional format. IID random variables are the bread and butter of parametric estimation in statistics and so and so forth.

That is the meaning of iid. So the key idea of this repeated experiments leads to the idea of sampling or more precisely random sampling if you want to be more erudite.

(Refer Slide Time: 04:35)




So what do we do? Given a scenario omega, I am taking the value of the random variables each one of them at omega. So I am creating the sample path. So this observations are numerical observations. These are real numbers. Each of them are real numbers. This is real, this is real, and this is real; so all are real numbers.

So these are observations and if you take a finite number of them then I say that I have taken a finite sample from the population. So there are say 5000 people whose heights I have I have to know that this city has say 5000 people and I want to know what is the average height of the people in the city. So I do not go and measure 5000 people, that is a waste of time. The whole key idea of statistics is to do stuff by taking less observations and giving almost a nice answer.

So, if you want measure the average height you take a random sample. Of course, there is a whole study of how to take random sample. So you take a random sample and then you look at the heights of those random samples and take an average of those guys and that if the random sample is taken properly then that would actually be a fairly good representation of the actual average height of the population.

Of course, if you take a random sample quite large then you would be giving much better approximation to the original actual random average height which you would not know. So this if you take say only finite number of them then you are essentially doing random sampling. So associated with this is first is the strong law of large numbers. You would ask me is there anything called weak law of large numbers, there is something called weak law of large numbers which will be a part of the exercise so here we are taking about strong law of large numbers.

So what does strong law of large numbers tell us. The strong law of large number tell us that if the size of your random sample increases then given a scenario omega, the average of the sample, so if you take a finite number of sample. So suppose you have taken 500 samples then you calculate the average height of those people, then this will be an approximation to the average height of the total population.

But if you keep on increasing the number of people, number of people in the sample then your sample average goes towards or tend towards the population average with probability 1. That is the idea. So sample average tends towards the population average with probability 1. So let me write down the strong law of large numbers.

Let X1, X2, Xn be a sequence of iid random variables okay, random variables defined on the same probability space. So because they are identically distributed they have the same average, they have the same mean. Assume that they have some distribution. So once you take this you are assuming that they have some density function and you can calculate the mean so you have to make this more simplifying assumption.

So that is say n then the probability that the sample average leads to the population average which is m is 1 that is limit which is quite natural. It actually collaborates with our intuition so this is a very natural result. Of course, this simply means you are calculating this at every omega. So given a scenario, so what is this event? So it is a set of all omega such that whenever I take the limit of this as n tends to infinity this becomes m.

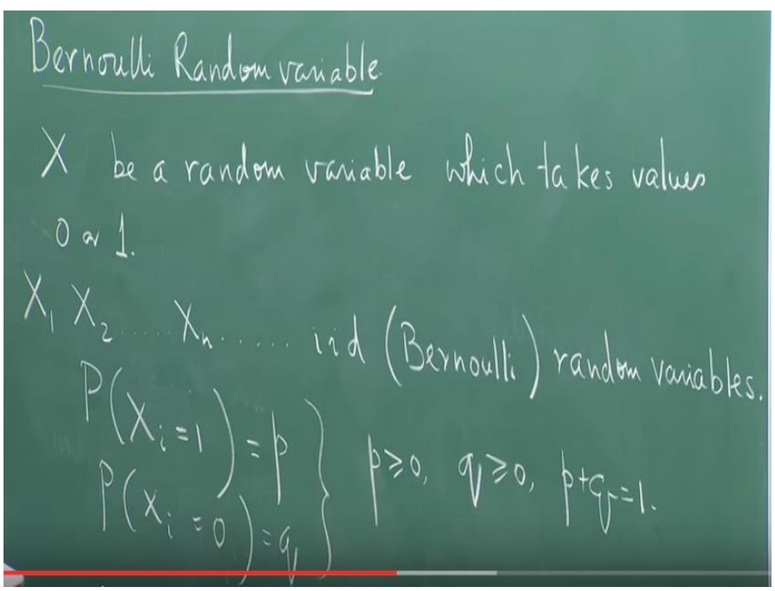
That set is a sure event. So this probability this event is a set of all omega in a sample space, set of small omega such that limit n tends to infinity x1 omega plus x2 omega plus dot dot dot xn omega by n is equal to m as n tends to infinity. Such a event is a sure event and that is exactly what the theorem states.

I do not want to prove the theorem right now unless I have time at the end. You can of course ask for the proof in the forum if you want (()) 11:58 whether proof can be given in the forum but the proof is slightly involved. It is better to have at the first stage many people would be possibly seeing this for the first time and so it is better to have a clear idea what it says than just going into the technical proof.

The technical proof here depend on 2 things. One is the Cheybyshev's inequality and one is the Borel-Cantelli lemma, which we have done in the last section, in the last lecture. Now once I do this we have to move gradually to the notion of the strong law of large numbers. So in order to do that we introduce what is called the Bernoulli variables. The Bernoulli variables are random variables which takes either the value 1 or 0 means it denotes success or failure.

So if I toss a head so if X is a random variable which denotes the number of head then X of h is 1 and X of t is 0 that is by by success if I mean that I have head and if by failure I mean I have tail then x of h the random variable when evaluated at h should give me 1 and random variable which is evaluated at t should give me 0. Such a random variable which actually (())13:25 success and failure that is called a Bernoulli random variable.

(Refer Slide Time: 13:35)

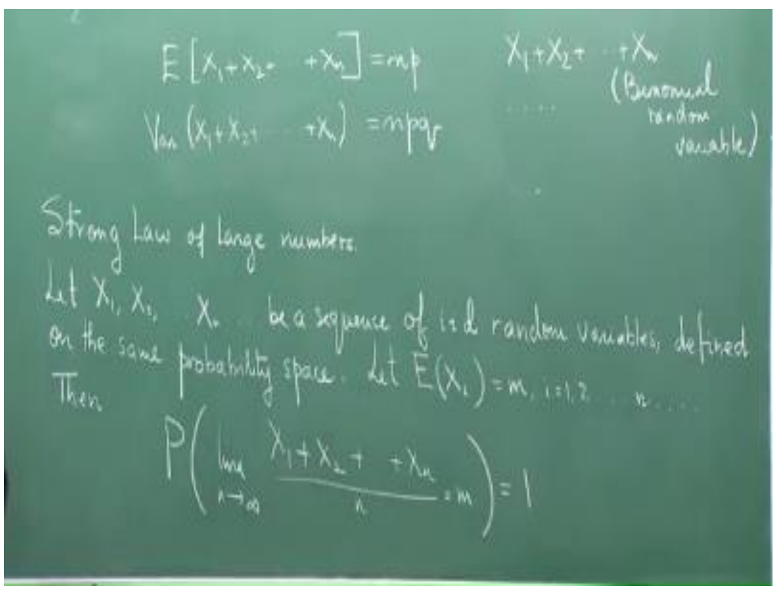



The Bernoulli random variables are random variables that take values either 0 or 1. So for every point in the sample space when you chose that point if it is considered as successful, will give you the value 1 or will consider as a failure and give you the value 0. So let X be a random variable which takes value 0 or 1. So these are called Bernoulli random variables, right. Now take a sequence of so these are iid random variables which are Bernoulli, iid Bernoulli random variables and they either take value 0 or 1.

So for every i I have given that if there is a success the probability of success is p, if there is a failure the probability of failure which is 0 is q and p and q both are greater than 0 and their sum is of course 1 the total probability and any one of the two has happened it will surely happen either it is success or it is failure. P greater than 0, q greater than 0, p+q=1, okay.

So once I know this I can of course compute its expectation and compute its variance. So this will go as a homework to you. We will not compute this but we will write down the results. So here we are now continuing from this part of the board and going up to this part.

(Refer Slide Time: 16:58)



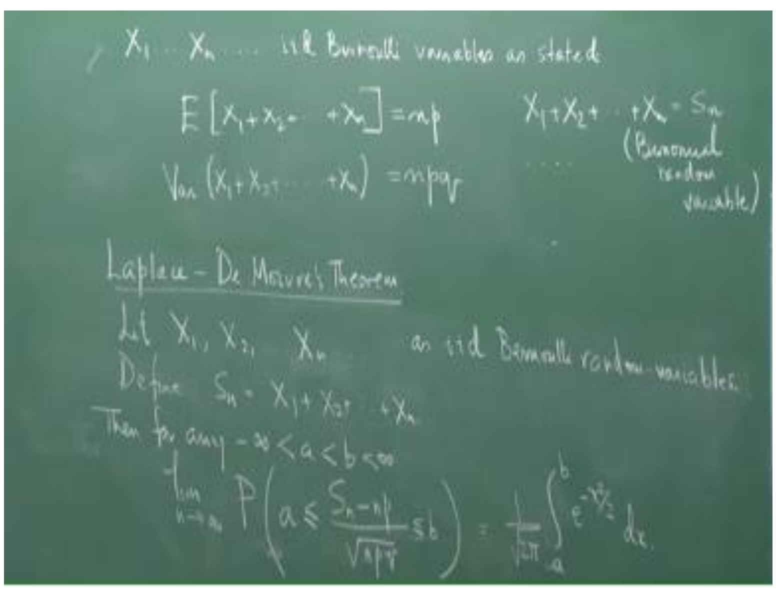

So if you have iid Bernoulli variables as stated since as we have stated the probabilities of success and failure and the expectation of the sum, so this sum actually means how many success I am having in a n trial n repeated trials of the random experiment, how many are successes. So if toss a coin toss head and tail so say up to say 50 tosses how many coins are actually head because for those the random variable take the value 1.

So such variables which measures which gives you the number of successes in a given number of trials such variables are called binomial random variable so this x1, x2, xn are called binomial random variables. People who have studied binomial distribution would immediately catch this fact. So this is called a binomial random variable and the variance because these are all independent.

Of course I expect you to know this very basic way of calculating variance as an expectations as I have again repeatedly told you that this is not just the first course in probability  theory then we would not have reached this part within this few lectures. Because of independence this is nothing but npq and those who know binomial distribution they perfectly know that these are the mean and variances.

So we will now provide some sort of a limit theorem which is called the Laplace de Moivre theorem for Bernoulli random variables only. Once that is settled we can, that is one of the precursors to what is called a central limit theorem. This is sometimes also written as S n the sum of 2 n variables so I am removing the strong law of large numbers writing and write down the Laplace de Moivre theorem which we state for binary random variables which could be actually better.

(Refer Slide Time: 20:25)



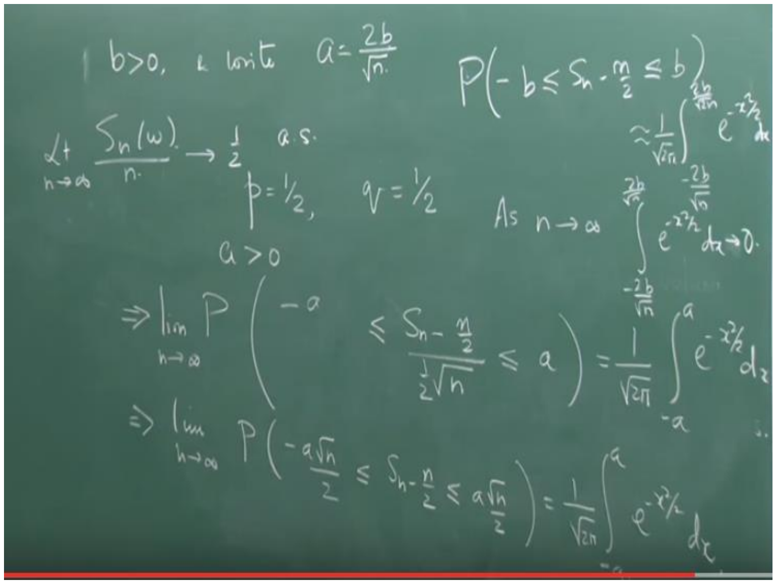
So this is also one form of central limit theorem but starting with iid random variables which are Bernoulli. So we consider X1, X2, Xn as iid random variables iid Bernoulli random variables. So now we will define the way we have defined S n so I am just defining for the heck of it, we could have just picked it up from there. Then for any a and b finite a and b what you have is the following.

So here I am normalizing the binomial random variable by subtracting the mean and dividing by the square root of the variance which is the standard deviation which is called the standard deviation. So whatever a and b you chose this limit as n tends to infinity will always provide you the probability that a standard normal variable is lying between a and b because here you see it is the density of the standard normal distribution, standard normal density function. So what does it mean? What is the conclusion from we can draw?

The rough conclusion that comes in is that if you increase your number of experiments repeated trials, as you increase the number of repeated trials, the Bernoulli and the normalized binomial variable this is called the normalized binomial variable; the normalized binomial variable starts looking or rather starts behaving like a standard normal variable which is very important because binomial variables are discrete random variables while that is a continuous random variable standard normal distribution.

So there is a deep link between the 2 and that is something which need to be seen. Now this is the beauty that your limit theorems are connecting the discrete and the continuous random variables and which is a very very important thing. So what are the interesting conclusions we can draw from the Laplace de Moivre theorem? We will actually later on modify the Laplace de Moivre theorem and write down the central limit theorem. The Laplace de Moivre theorem has the following.

(Refer Slide Time: 24:41)



Suppose I have scenario where my probability of success is half just like a coin toss this is basically the coin tossing scenario. Now for any a greater than 0 it implies using the Laplace de Moivre theorem that limit n tends to infinity probability of so np p is half here you have root npq which we have taken on this side, this has come one fourth, so I have taken this here.

So we are basically looking at the whole thing between a and minus a so I would write it with 1 more step so basically it is this one and this okay. Now this can be put in like this so this can be written also as limit n tends to infinity p of minus a root n by 2 less than equal to S n minus n by 2 less than equal to a root n by 2. This probability is in the limit is same as a standard normal variable lying between - a and + a.

Now write down take b to be strictly greater than 0 take b to be strictly greater than 0 and write a equal to 2 b by root n then what I get from here is so what I get from here is that probability of course there is a limit thing I am not writing the limit I am just writing in a very rough sense this is very roughly so I am just plugging in a with this.

So from here I get a is equal to 2b by root n so I get this is almost similar. So there is a limit so for very large n this is almost very very good crude approximation of the probability not very crude but alright, acceptable. So what does this show. As n tends to infinity this goes to 0 this goes to 0 so they tend to 0. So as n tends to infinity this whole thing it is a well-shaped curve it finally just becomes a line and then it goes to 0.

This goes to 0. As n tends to infinity not equal to 0 just goes to 0. So essentially what I am telling that even though I know that the average that the average S n omega by 1 by n so if take the average this goes to half in this case using the strong law of large numbers with probability 1 that is almost omega almost surely. So means for every omega except a few this goes to half. So probability of this event is equal to 1.

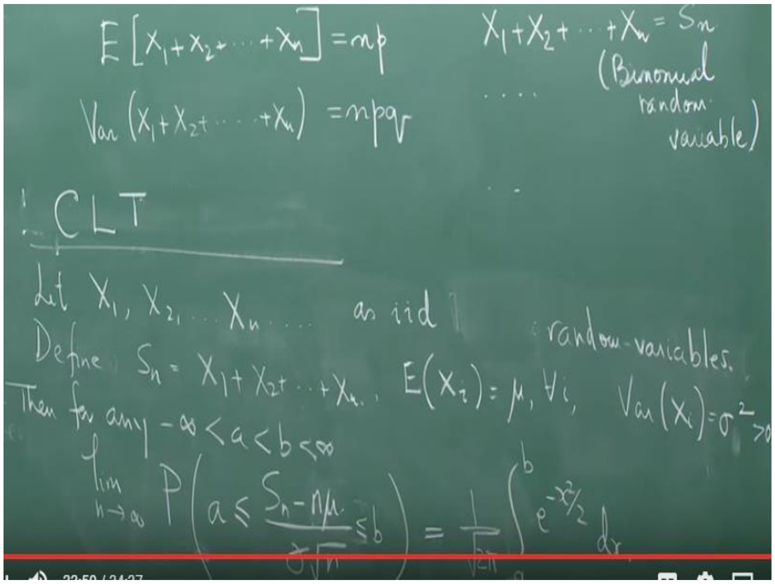
That is what we get from a strong law of large number, but what we are seeing here is something very interesting. It says that you cannot bound this difference. The fact that S n, this actually means probability this one. Means probability of mod S n minus n by 2 is less than equal to b this is somehow going to 0 is decreasing. So the fact that you can bind S n minus the Bernoulli the binomial random variable the deviation of the binomial random variable from its mean that you can bind it within certain value is that probability is less.

So it can fluctuate outside the value infinitely often or almost everywhere. That is for almost every omega S n omega minus n by 2 this should be strictly bigger than b which is a very very, so this can fluctuate with probability 1 of any value. This deviation can go over any value with probability 1. This is a very strange conclusion of the Laplace de Moivre theorem.

Now we are going to remove the Laplace de Moivre theorem the title and end it with the central limit theorem. So what is the central limit theorem? Central limit theorem which I write for short as CLT considers iid random variables but does not consider a Bernoulli random variable so you forget the Bernoulli.

So you define S n which is no longer a binomial variable just but a random variable which is a sum of n random variables where E of x i is meu for all i and the variance because they are identically distributed so they have same mean variance we assume that they have mean variance. Of course there could be random variables which are distributed by some distribution it has some density but does not have a mean like Cauchy distribution but we are here assuming all those things. So, it is sorry sigma square. So then the CLT says that then for any n s we have to replace this by meu and replace this by sorry replace this not by meu but you have sum of it will be n meu and replace this by n sigma square. So it will be sigma root n.

(Refer Slide Time: 33:59)




So this will be n means sigma root n so if you take any iid random variable it does not matter and if you take the sum and normalize it in this fashion in terms of the mean and variable so n meu is the expectation of S n and the standard deviation of S n is sigma root n right. So if you take a random variable subtract from its mean and divide by standard deviation then when n becomes very large that is the sample size is very large it starts behaving like a standard normal distribution.

So when you take large samples you assume their population is behaving like a standard normal population. Thank you very much.