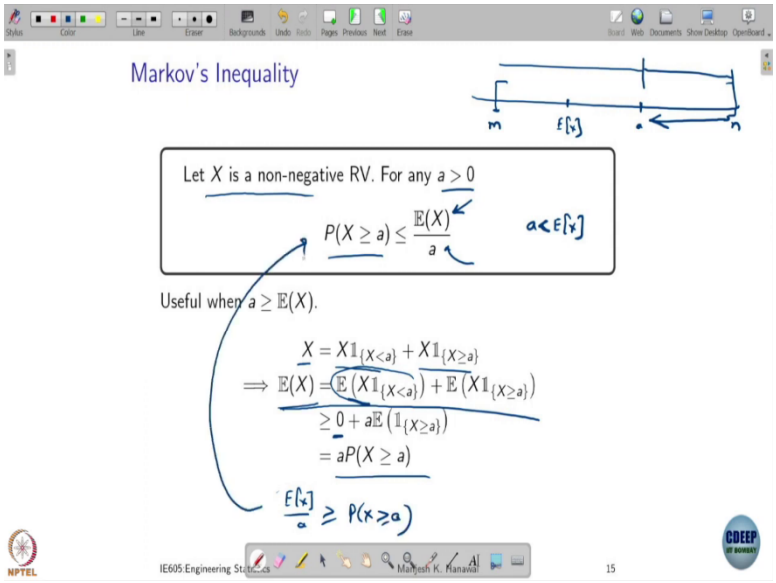
Engineering Statistics
Professor Manjesh Hanawal
Industrial Engineering and Operations Research,
Indian Institute of Technology, Bombay
Lecture-16
Laws of Larger numbers, Central Limit Theorem

(Refer Slide Time: 01:18)



Now, most of the times when we start looking into the system models, exact probabilities, maybe we will not be able to compute, but we are just happy if we get some bound on it. This often happens. If you, if I give you a very complex function to integrate, maybe you may not be able to integrate and get to what is the finite integral of that. But if I say that just give me a bound on that. Maybe it is easier.

So, like that in some situations you may just end up dealing with complex systems where you may just want to know what is the worst case, and in that case, maybe just bounds are enough. So, for that we will start looking into some bounds on the probabilities that we have studied. And the first one is something called Markov's inequality.

So, what Markov's inequality say is, suppose if you have a X, which happens to be non-negative random variables, and you take any value a. Now the probability that my X random variable is going to take value larger than a, that will be upward bounded by expectation of x divided by a.

Suppose. If you know the expected value of your random variable X, then you can easily compute this value. So let us say your random variable is going to take some value in the interval [a, b] and its expected value somewhere here. Now it can take any value in the interval a b. Now you are asking what, let us say some, maybe instead of this, I will use different version maybe let say. Let me call this m and n, and its mean value is somewhere here.

And now you want to ask the question, my random variable always, let us say this is a, always going to take value in this range, that is my X is larger than a. What we are now going to say, that, that probability is upper bounded by expectation of X by a. And this is true only for non-negative random variables. Now, naturally this is going to be useful when a is going to be greater than expectation. In case a is less than expectation of x, what is going to happen?

This ratio is going to be something like greater than 1, that becomes a trivial upper bound for. Like probability we any way, we know that cannot be greater than 1. So, this becomes useful only when you have a greater than expectation of X. This, this is like one of the looks very simple, the proof is also simple, but this is like a, becomes like a building block for many of the advanced probability.

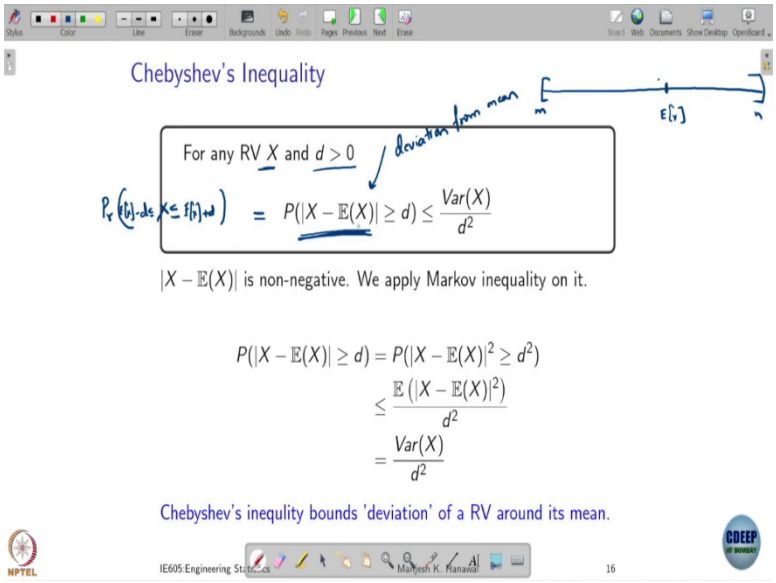
You study where we will derive many, many different varieties of bounds like this. But ultimately, they all emerge from the ideas that is used in the Markov's inequality, and something called Chebyshev's inequality, which we are going to see in the next slide. Now let us see how the proof line goes on. A random variable x, I can split it into 2 parts. One part when X is less than a, and another part X is greater than equals to a.

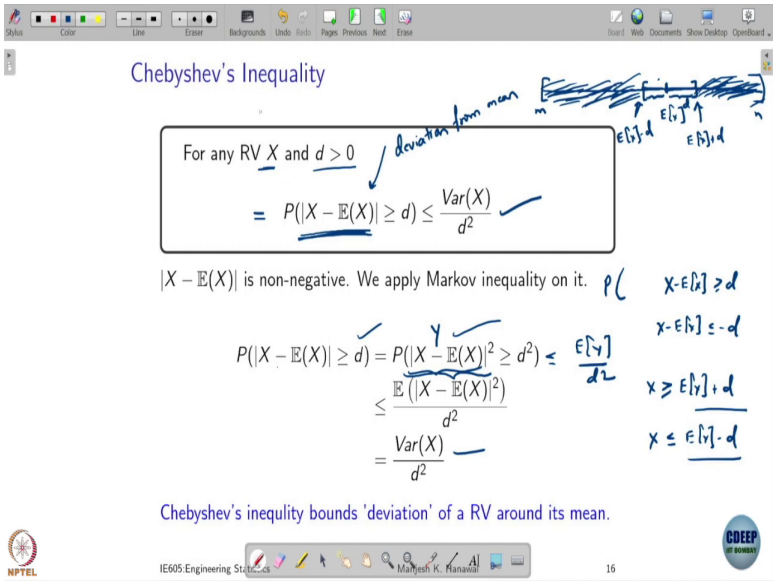
You understand the meaning of this splitting part? So, I am basically split into this range and this part. Now if I look into the expectation, we know the expectation is linear operator so I can write this expectation of the sum of these 2 expectations. Now X being a non-negative value, can this expectation be anytime lower than 0? No. It has to be greater than 0. So that is why I will actually take a lower bound on it as 0.

And then I will end up expectation of X is simply a times probability that X is greater than a. Then I am done. If I just, I take simply a in the denominator, then I get probability x greater than is by expectation of X divided by a. So, all I am now doing is the next step, I did is

expectation of X by a, is I just took this a in the denominator, this is X greater than equals to a, that is exactly the claim.

(Refer Slide Time: 05:37)





Now, suppose you know something more, instead of mean you also know variance, then we can have something more. And that is called something Chebyshev's inequality. What Chebyshev's inequality says is, if you have a random variable X, notice that X here need not been non-negative, any random variable. And you are given some value d > 0. Now I am basically asking what is the probability that the absolute value of the difference between X and its mean value is larger than d.

So, this we can call it deviation from mean. That deviation from the mean is being larger than d. What is that probability? And Chebyshev says that this probability is upper bounded by variance of X divided by d square. What it is saying like suppose you take a, let us say X is taking in this range value, and its expectation is somewhere here, it is taking its expected value is somewhere here.

Now you are asking, see what is the alternate way of writing this? Alternate way of writing this is, probability that X is upper bounded by expectation of X plus d, and lower bonded by expectation of X minus d. Is this fine? So that is what X is greater than expectation of minus d and upper bonded by expectation of D, is exactly means that, X minus expectation of e the modulus is going to be greater than d. Am I correct here? Maybe, let us see, I am not sure I am correct. Let me rewrite this.

Student: 1 minus.

Professor: Let us take the case where this quantity is positive. Then it is simply X minus expectation of X is greater d. And if it is negative then it is X, expectation of X minus d is less than or equals to minus d. Now, let us combine these 2, so we will get expectation of X plus d, and X is less than or equals to expectation of X minus d. So basically, I am by writing like this, I am basically asking X is greater than this, and X is also less than minus d.

So, if you take expectation of X, you look into this like you add, this is like a d, and another thing lets like, this is like also d. So, this point is expectation of X plus d, and this point is expectation of X minus d. What you are asking is, you do not want X to be in the d neighbourhood of your mean. So this is like a d neighbourhood. This region is the d neighbourhood, like this is within the d range of your mean.

What you are asking is, X takes value outside this, this entire thing. You are just asking is X does not lie within d range of your mean, it is just lying outside. What is that probability and that is given by this bound. Is this clear? Now why this bound is true? This bond again simply follows from Markov's inequality. First, I will start, I am interested in this quantity. Expectation of minus, X minus expectation of X is absolute value being larger than d.

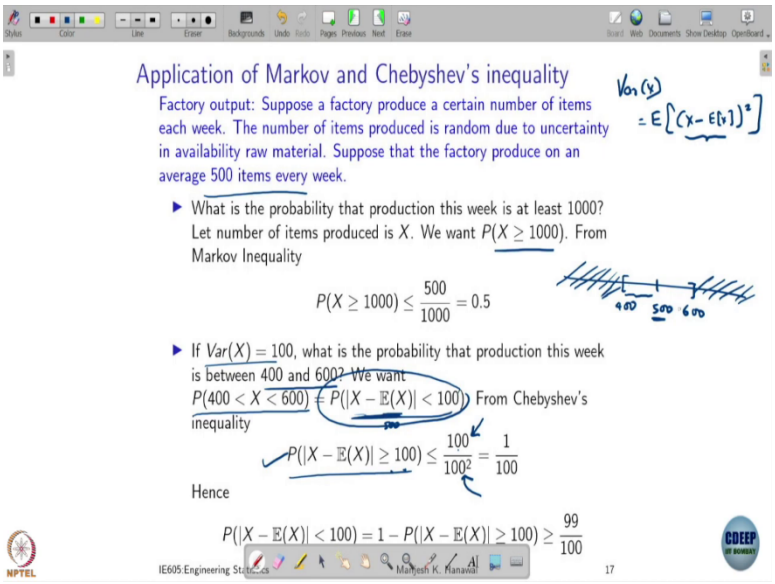
Now what I did is I took the square on both sides. Is these 2 probabilities are going to remain the same? They are going to remain. See, notice that because of the absolute value, left side is also positive, and the d is assumed to be positive value. So, if I square them, the sector of omegas which will satisfy the, remain the same. So, the event set remains the same in both the cases. So, because of this, these 2 probabilities are the same, and if you are not sure about it, please look into this.

Now once I have this, I am going to now go back to my Markov's inequality. This portion I am going to treat it as a new random variable, which is anyway positive. Because of the square, this entire thing I am going to call it as a new random variable. Let us call this Y this entire thing. Now I am basically asking the question, what is the probability that Y is going to be greater than equals to d square?

While Y being? Non-negative valued. By Chebyshev this is nothing, but expectation of Y divided by d square. And expect Y is one by definition X minus expectation of X whole square, and by definition expectation of X minus its means square whole squared. What is this? This is variance. So that is why we get variance divided by d square. So that is why, or like as I said earlier, Chebyshev is trying to bound the deviation of your random variable from its mean value.

What did Markov's inequality give? Markov's inequality telling your random variable taking value larger than its mean value. And we just obtain distribution from the Markov's inequality itself. And you can obtain actually many more advanced inequalities what often call as concentration inequalities building on them. Now let us quickly look into one example that I put here, how this could be useful?

(Refer Slide Time: 13:15)



So I will let all of you to read this factory output example here. So, a factory is producing certain items, certain number of items every day and of course the product that is being done, it depends on the raw material. Depending on how much raw material is available that day, the number of output is going to vary and let us say we know that on an average the factory produces 500 items every week.

Now I want to ask the question, what is that probability that the number of items produced by the factory is going to be larger than 1000? So now basically I am asking, mean value was about it was producing 500 every week. Now I am asking it is going to be producing more than thousand can Markov equality come to our help in this case? Now, we are basically asking what is the probability that X is going to be greater than equals to 1000 that is nothing but expected value effects divided by 1000, that is 0.5. So we know that the probability that more than 1000 units will be produced is going to be less than half.

The next question, suppose we know that the variance of the factory is like about 100. So, what is variance? Variance is capturing about the deviation about the mean value. So, by definition variance of X is expected value of X minus expectation of X, so this is like again as we said this is kind of deviation from the mean like mean say how much the variation is happening, this is capturing basically variations around the mean value.

Suppose you know that variation, variation is noticed to be 100. Now what is the probability that the production this week is going to be between 400 and 600? Now I am asking more refined question. It is between 400 and 600. So, I am asking this question, what is the probability that output is going to be between 400 and 600.

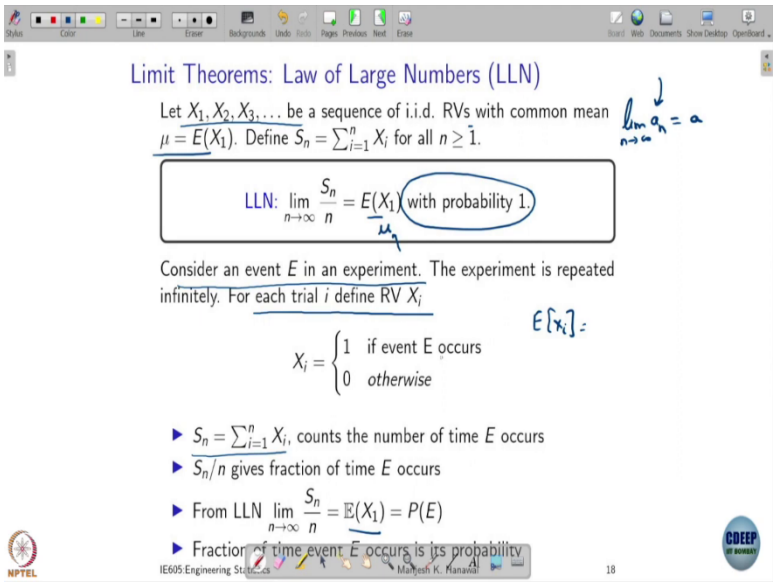
Now since I know that mean value is 500, I can pose this question like this, that is like mean value something 500, I am asking between 400 and 600, that means the deviation from the mean is like about a hundred both left side and right side. So, the difference, the deviation from the mean is being less than a 100, it should be 100.

Now the Chebyshev inequality when I have this immediately that strikes to our mind is Chebyshev equality. Chebyshev inequality gives bounds on such deviations, but Chebyshev inequality give us this quantity being larger than or equal to 100. Let us compute that. What is this quantity is going to be?

Variance divided by D square, variance is 100 and D is also 100, 100 by 100 square that is 1 by 100. But now how to get this value from this. That is compliment. Now instead of asking what I got is outside this value and if I subtract this value from one, then I will get something inside between 400, so that is why now probability that X minus expectation of less than 100 is 1 minus this which is now 99 by 100.

What it is saying that under this given randomness I will with 99 percent of the chance that I will rely within 400 to 600 range, my output will be between 400 to 600. So finally, through this example you see that how Markov equality and Chebyshev equality can come to help, and this is like a very toyish example like. When you analyse more complex system, you need to first properly define the, what is the random variable there and need to define what is the deviation we are talking about and then apply these results, next.

(Refer Slide Time: 18:09)



Limit Theorems. So, these are the some of the fundamental results in probability. The first one of them is called law of large numbers. What last large numbers says, says is if you have the sequence of i.i.d random variables and let us say its mean value is mu and I denote the running sum as Sn.

Sn is sum of my end random variables here and now this running sum, I am going to now interested in the running average. If I look into the limit of this running average, this limit is always going to converge to mean value of those random variables. Notice what is the expectation of X? That is exactly mean mu. Because all of them have the same mean.

Right now, ignore this time what is mean by probability 1, this will be made more precise in IE 621 course. So, this is called something called convergence in probability. Notice that you people know convergence, limit you know all of you. How many of you do not know what is limit here? Limit of a sequence. When we talk about limit is going to a, whenever it exists, all these an's are some deterministic quantities but whereas here Sn by n is it a deterministic quantity?

Sn is a sum of n delta variables, so it is not the deterministic quantity, this is a random quantity. What is law of large number saying is even when I take limit of this running average, the limit is always going to be deterministic. So, in a way what it is saying is if you are going to add lot

of random variables and take their average, it is going to be essentially going to be approaching to be a constant. And you will see that this is one of the fundamental theorem which connects basically probability and statistics.

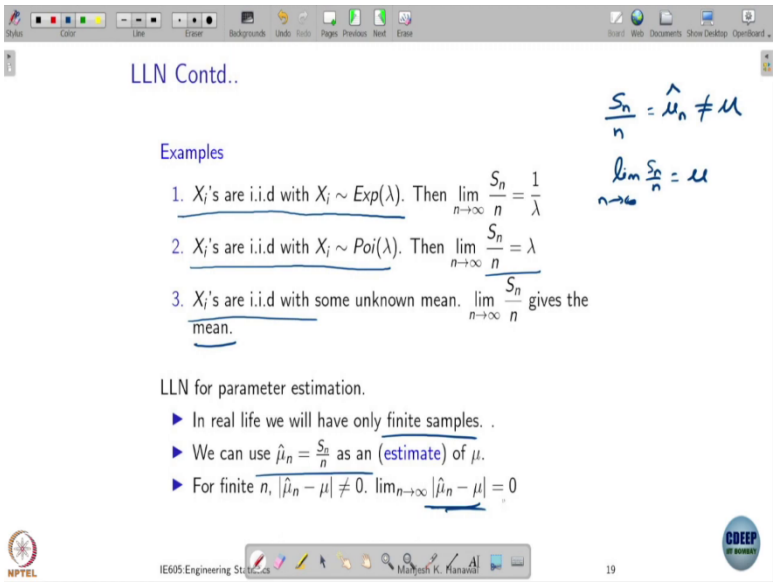
The expected value is like of mu in this case. These are the parameters of our distributions and now it is trying to connect this parameter with the data we are going to observe, that data is the sequence of random variable we are going to observe. So, what it basically connects is your data with the underlying parameters and that is why this is a crucial link between your parameter estimation and data, and parameter estimation we are going to see that is going to be the critical part in our statistics.

Now just to understand this law of large number intuitively, suppose let us say you conduct, you are interested in event E and in your experiment if that E event E happens, you are going to assign value 1 to a random variable, otherwise you will assign value 0 and you keep on repeating that experiment again and again, again, again and you are going to call for the ith trial the random variable you are going to call it as Xi and in the Xi they are going to take value 1 and 0 is Xi's are they identically distributed? They are going to be identity because this is the same trial repeated. Do they have the identical distribution? Yes, and they are independent anyway because they have been repeated without having any influence on each other.

Now if you took into this sum of this random variable and Sn by n like this, what is law of large number is saying is this is nothing but Sn by n is going to do expectation of X1 but in this case what is the expectation of X1 here Xi here probability that event E occurring. So naturally what basically counting is how many times the event has occurred when you repeated that experiments multiple times.

And that naturally is going to be, has to be probability of that event E occurring in an i.i.d trials. And the fraction of the time that even E occurs is basically it is probability we talked about this when we talked about the frequentist notion of probability. If you recall in the first lecture or second, we talked about frequency notion that is what. So, this is a simplest question like the Xi's need not be always like this Xi's can be arbitrary and this theorem holds for any arbitrary exercise as long as they have some finite means.

(Refer Slide Time: 23:55)



Let us look into some examples now. Suppose I have a sequence of random variables which are i.i.d and all of them have exponential with parameter lambda. So, if you look into their sum and take the average, law of large number says that that should go to exponential mean, mean of that exponential distribution. What does the mean of that exponential distribution 1 by lambda.

Similarly, if this Xi's are all poisson with parameter lambda, what this value should go to lambda because lambda is mean of the poisson distribution and similarly suppose if you are going to deal with some Xi's i.i.d random variable, whose mean you do not know a prior but you know that if you take this limit it should be converging to its mean.

Now can you tell me a mechanism if I tell you, I am going to give samples drawn from exponential distribution, but I will not tell you what is the parameter lambda. I will not tell you what is the, I will just tell you they are exponentially dis samples are drawn exponentially distributed. Can you tell me what will be that parameter lambda of that exponential distribution? So, you take the sum take the running average and I will give as many samples you want. So just take that and go to infinity, that value is what you are going to get actually 1 by lambda and the reciprocal of that will going to give you the parameter of your exponential distribution.

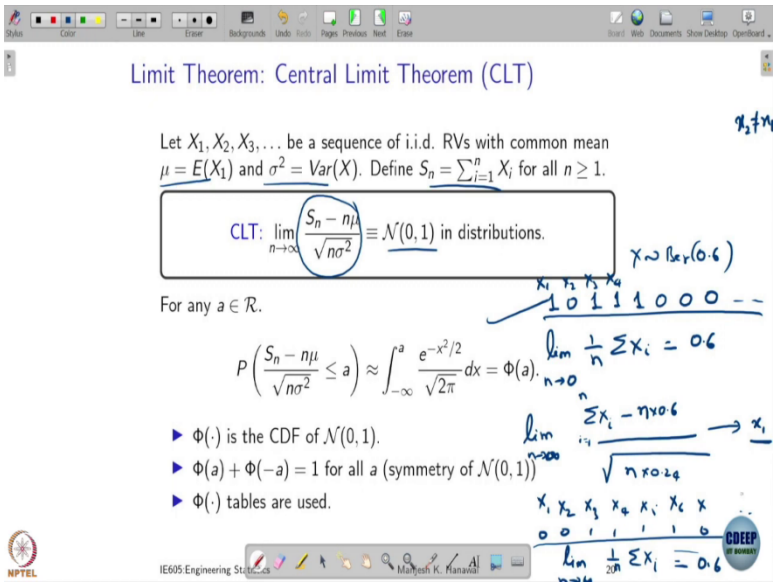
So, you see that like if you know that data is coming from certain distribution, but you do not know the parameter, law of large number provides one means to generate those parameters, fine parameters, next.

Suppose let us say I in even though I said that I will give you as many samples and you will let n go to infinity, that is the hypothetical thing. In reality it is never happens that we keep on generating an n goes to infinity. All I can do is if you keep on asking maybe I will keep on giving a 100 if you ask maybe 200 after that I will say get lost, I cannot do more.

So, let us say if we had to deal with finite samples. Now whatever let us say Sn by n, let us compute for that particular n and let us denote this by mu n hat. Now do you expect this mu n to be same mu, if the mean value is mu, it need not be. The limit value, we know that as n tends to infinity this is mu, but for particular n this need not be mu. There is going to be some difference between mu hat n and mu.

But as limit n goes to infinity, what is going to happen to this difference, that difference goes to 0 as n goes to 0. But for some particular n this difference need not go to 0. And since in real life you have to deal with only finite n, we may want to understand if I have to deal with certain n number of samples, how much will be this difference? We know that that is going to go to 0, but how fast it decays to zero?

(Refer Slide Time: 27:47)



So for that another celebrated theorem of probability comes into picture called central limit theorem. What central limits theorem says is like if you have a sequence of random variables and it has a certain mean and variance each one of them and as usual, I define the sum as summation of Xi. But now if I look into this quantity, Sn minus n mu, mu is the mean and then divided by square root of n sigma square, this limit is going to converge to not a number, but what it is going to converge, what it is saying is whatever value it is going to converge those points to which it converge, they will have normal distribution. You understand this point? Let us quickly do one example.

Let us say take a coin toss, I did a first coin toss, and I got a number 1, same coin I did second times 0, I did thrice, fourth, fifth, sixth, seventh, eighth, like that and let us say that coin is Bernoulli 0.6. And now if I add this, these are like this was X1 was this, X2 was this, X3 was this, X4 was like this, like this and if I add 1 by N and I take limit, this is going to be what? This is going to be 0.6.

Now what I am going to do, instead of taking this, I am going to take Xi minus and n. What is the value of mu? 0.6 divided by n, what is a sigma square of a Bernoulli with 0.6 is? What is the formula? that is going to be 0.024, and if I take as n tends to 0 for this realization, this value may converse to some value call that as x1, I do not know what is that x1.

This is for this particular sequence and now let us say I generated another sequence where X1 was 0, X2 was 0, X3 was 1, X4 was 1, X5 was 1, X6 was 1, and like that. Now even on this sequence I can do the same thing summation Xi 1 by n and I let limit n goes to infinity. What this value is going to be on this sequence? This is going to be 0.6 again. So, notice that irrespective what is the sequence we always got as 0.6, that is what law of large numbers said. As long as you take enough number of i.i.d samples and average them, the limit is going to be 0.6.

But if you now do this value like this, now if you do the same thing for on this new sequence, it need not be X1 again, you are following this. Now, if I use these new values of X1, X2 instead of this and plug in here, this value need not be same as X1, that could be something called X2, which can be different from X1. But what we are now saying is this points X1 and X2, like they appear like as if they are following Gaussian random variable with means 0 and 1. That is the

difference between the law of large numbers and the central limit theorem fine. Let us stop here.