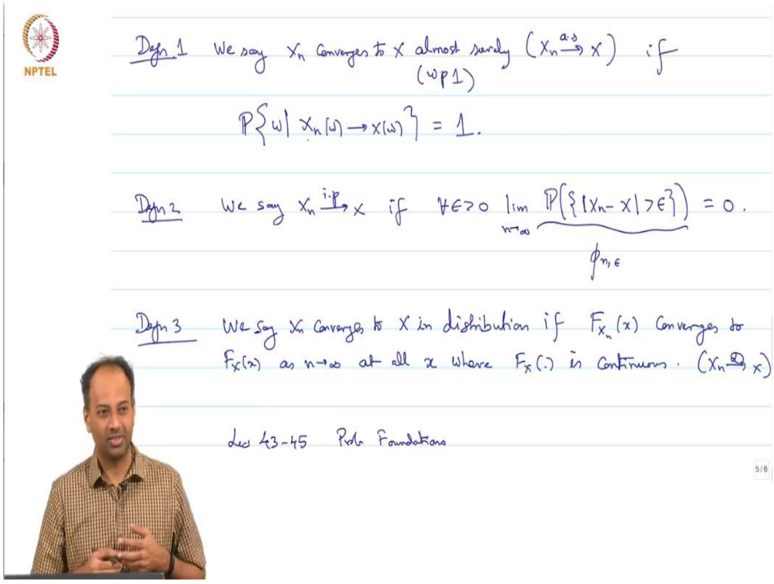
Stochastic Modeling and the Theory of Queues
Prof. Krishna Jagannathan
Department of Electrical Engineering
Indian Institute of Technology - Madras

Module - 1
Lecture - 3
Laws of Large Numbers and Central Limit Theorem

Welcome back. We will continue our discussion, a brief overview of Convergence of Random Variables.

(Refer Slide Time: 00:22)



Yesterday we defined 3 different notions of convergence of random variables. The first is convergence with probability 1 or convergence almost surely, which is the same thing, which says that the sequence X n of random variables. You look at the set of omegas for which the sequence X n of omega converges to X of omega. So, for each little omega, X n of omega is a sequence of real numbers and X of omega is a real number.

So, you just test whether X n of omega, for that particular omega, converges to X of omega. So, if it does converge, you put that omega into one bucket; and if it does not converge to X of omega, you put the omega into some other bucket. Now, X n; you say that this X n converges to X almost surely, if the bucket where the convergence does take place has probability 1. The bucket where the convergence does not take place has probability 0.

So, that is almost sure convergence. Convergence in probability, which is definition 2 here, is a little weaker. You may not immediately see that it is weaker. It only concerns itself about the behaviour of the probability at n. So, you look at the probability that the difference between X n and X, the absolute difference between X n and X exceeds some epsilon. You choose any epsilon, 10 to the -6, 10 to the -8, whatever you want, and you look at the probability that X n - X exceeds epsilon.

So, this is some P n, epsilon. So, for a fixed epsilon, this is a sequence of probabilities in n, and you want this sequence of probabilities to go to 0. So, in the second definition, it is the sequence of probabilities that is converging to 0, as opposed to a sequence of random variables themselves which are converging in the definition 1. And finally, definition 3 talks about convergence in distribution.

So, you look at that CDF of the nth random variable F X n, and you send n to infinity, and you see if this converges to F X. F X; X is the proposed limit. And you look at the; let us say, you fix a little x for one thing. So, you fix a little x and you send n to infinity and look at whether F X n of x converges to F X of F X. And if this happens for all little x wherever F X is continuous; where the limiting distribution is continuous, then you say, X n converges to X in distribution.

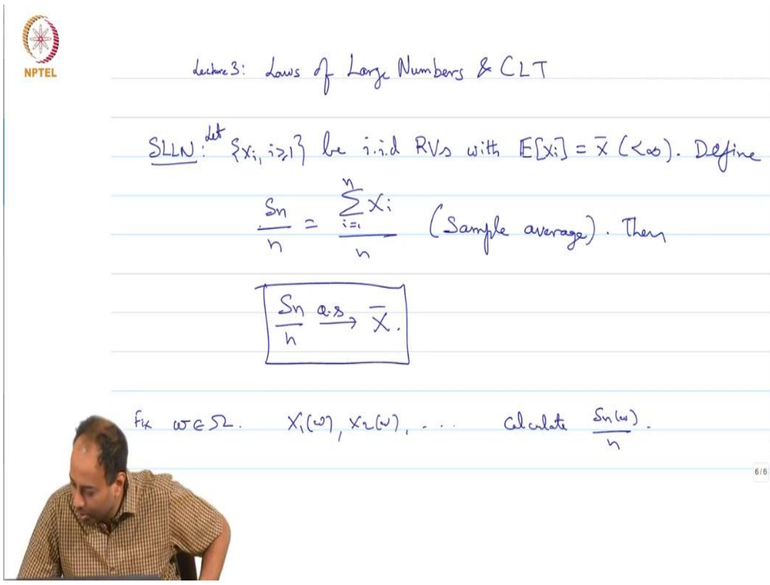
So, these 3 are the most commonly used and they are progressively weaker. This can be shown. So, convergence almost surely is the strongest notion followed by convergence in probability, then followed by convergence in distribution. So, convergence almost surely implies convergence in probability, which implies convergence in distribution, but none of the reverse implications are true.

So, just if you have not seen this thoroughly before, there are, you can view it. So, there is a; so, the lectures 43 through 45 of the course Probability Foundations, which I recorded about 5 years ago. This has this in much more detail. So, this is also on YouTube. And you can look at these 3 lectures if you want to learn more about these notions of convergence. In fact, you can watch it. I suggest you watch it anyway, because it is good review for you.

Now, so, maybe I should give you some examples to illustrate these convergence notions. See, the most well-known result, perhaps the most important result of almost sure convergence is, you know what it is? Is this strong law of large numbers. There is a strong law of large numbers, which basically says that sample averages converges to the expected value almost surely. So, that is an example of almost sure convergence.

There is also a weak law, which says the sample average converges in probability to the sample mean. Convergence in distribution. You know what is the most famous result in convergence in distribution? Central limit theorem. Central limit theorem, you might have studied, right? So, let me just recap this.

(Refer Slide Time: 05:26)



So, this is Lecture 3; this is Module 3. **"Professor - student conversation starts"** Yeah, okay, Lecture 3. Is that okay? No? Bigger? Large numbers; I will write large. Laws of Large Numbers and Central Limit Theorem, let us say, right? **"Professor - student conversation ends"** There is a strong law of large numbers. You basically take a sequence of random variables X i.

Let this be; let independent and identically distributed random variables with expected value of X i equal to some X bar. Let us say this is presumed to be finite. Define the sample average S n over n as sum over X i, i is equal to 1 to n over n. This is the sample average. So, you are, so you know what IID is independent identically distributed. So, you have a sequence of random variables, which all come from the same distribution.

The marginal distributions are the same, and they are independent. So, all finite order joint CDFs will factorise into product of the marginals. So, you know what that means. And you are assuming that the, each of these expected of value of X i is equal to some X bar, and this has to be finite. And then, you would generate a million IID samples of these random variables X i, and you take the sample average.

You sum all of them and divide by the number of samples. So, this laws of large number generally deal with the sample average getting close in some sense to the so called unsampled average, which is the expected value. So, strong law of large numbers says that; so say, then, S n over n converges to X bar almost surely. This is the strong law of large numbers. Sample average converges to the expected value of each of these random variables almost surely.

So, again, what does this mean? So, whenever; so, let us say, for some; so, for a particular omega, let us say you fix omega in Omega. So, you have a real sequence X 1of omega, X 2 of omega, ... And from this you can calculate S n of omega over n. So, each of these X i of omega, you can add till X n of omega; sum by n, divided by n; S n of omega or omega. So, for each omega, you get this sequence S n over n of omega.

Again, you will, you make different buckets. You test whether S n of omega; for that particular omega, you test whether S n of omega over n converges to the number X bar. X bar is just a number. S n of omega over n is a sequence of real numbers, because you are fixing little omega. So, if S n of omega over n converges to X bar, you put that omega in one bucket.

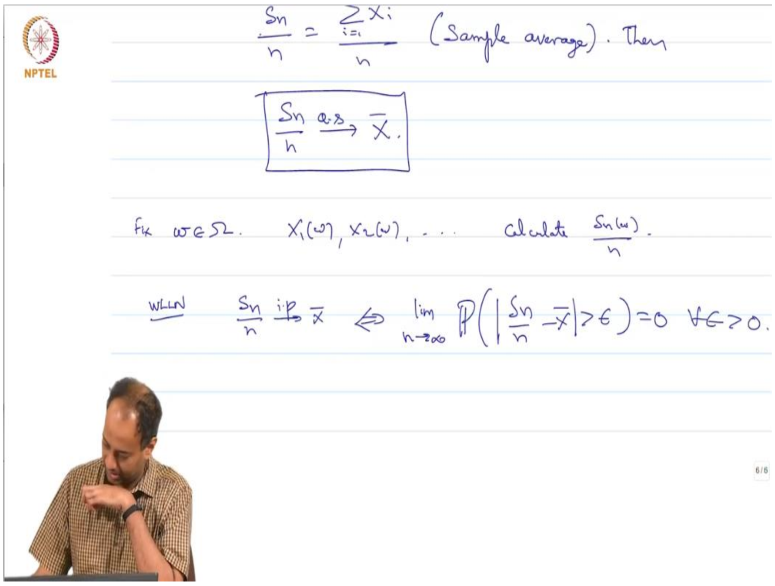
"Professor - student conversation starts" Sorry. X bar is a number. See, each of these X i's are IID. So, expected value of X i is some number, some finite number we are assuming. If it is infinity, this is not true. The expected value of a random variable can be infinite or even undefined. In those cases, law of large number does not hold. Strong law holds if the expected; actually, it holds; necessary and sufficient condition for strong law to hold is, if expected value of absolute X i is finite.

So, it is not; so, you can look at it as X bar of omega. So, normally, the limit is also a random variable. But here the limit is a number. You can view the limit also as a random variable which is always constant; the constant random variable, so to speak. Maybe that is your confusion. Usually, X n converges almost surely to some other random variable. Here the limit random variable is just a number, which is X bar. **"Professor - student conversation ends"**

So, you collect those omegas where this convergence to X bar happens, and you put, collect those omegas where; this S n of omega over n may not even converge, may not converge to anything. Or it may converge to something which is not X bar. So, in those cases, you put it in a different bucket. What strong law is saying is that the probability measure of the omegas were S n of omega over n converges to the expected value X bar is 1.

That is what this is saying. So, this is slightly non-trivial; it is not a very trivial result to prove. And this result is altogether only about 100 years old. This was proved much after the measure theoretic missionary was brought into probability theory. Whereas the weak law which;

(Refer Slide Time: 12:25)

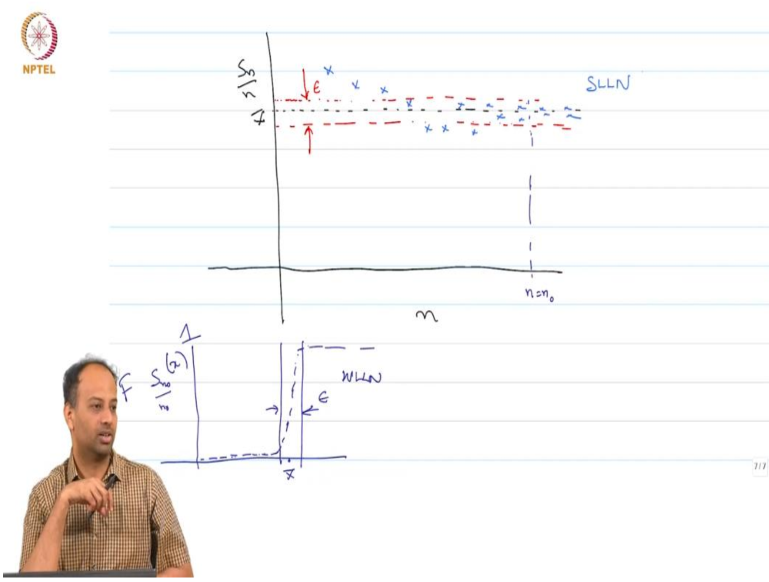


Weak law of large numbers under the same conditions, it says that S n over n goes to X bar in probability. So, meaning that; so, this is just the same as saying; if you say limit n tending to infinity, probability of; if you look at that guy; S n over n - X bar > epsilon. This limit is equal to 0 for all epsilon > 0. So, what this mean is that, you take a very large n and look at S n over n.

The probability that the sample average deviates from X bar by more than epsilon goes to 0 as n becomes larger and larger. Of course, you already know that almost sure convergence implies convergence in probability. So, you may ask what is special. There is nothing really special about the weak law; in fact, it is contained in the strong law. But the weak law is about 300 years old. It is easier to prove and easier to visualise.

And even people with; when undergraduate students, for example, who do not have measure theoretic background can understand what this means. But, strong law is a little harder to digest. It is a stronger result, but it is harder for students to digest and it is, to people figure, longer time to figure out that it is even true. I hope you appreciate the difference between these 2. So, maybe I should illustrate with one picture.

(Refer Slide Time: 14:20)



So, perhaps, I will just draw a picture to illustrate what I am talking about. So, let us say this is my n. And for a particular omega, I am going to plot S n over n. So, what we are; so, for some large n, let us say, let me give myself; let us say, this is my; let us say, that is my X bar. And I am going to give myself an epsilon band. So, I am going to look at; so, that is my epsilon band. Around, so, that width is epsilon.

So, I am giving myself an epsilon band around the true mean X bar. And I am looking at what S n over n is actually doing. So, weak law is basically has to do with fixing some large n. You will fix some large, let us say, you fix some large n = n nought. Let us say it is large. And you look at the probability that S n over n falls outside this band at n, at n nought. You look at the probability that S n over n falls outside the epsilon band.

So, what it is saying is that, what weak law is saying is that the probability of S n over n falling outside this epsilon band goes, becomes smaller and smaller as n becomes large. Is that clear? So, if you were to plot the CDF of S n over n for large n, you fix an n nought; you plot the CDF of S n over n; it will be, much of the probability will be concentrated around X bar. So, the CDF will rise very sharply at X bar.

So, if I were to plot; so, this is what weak law is saying. So, if I were to plot; this is a different plot. So, you look at whether S n over n is falling within the epsilon band or not, for that n nought. So, you fix n nought and look at S n nought over n nought. Sorry, this is the CDF of x. You are fixing n nought and looking at the CDF of S n nought over S n nought over n nought. So, that is X bar.

So, the CDF will look; so, it will remain quite flat and then suddenly take off very quickly and then reach 1. So, essentially, if you look at that epsilon band around; so, this is your epsilon band around your sample mean. Much of the probability will be concentrated around this sample mean. And the probability outside this band will be very small. And as n becomes larger and larger, it becomes smaller and smaller and goes to 0.

This is what weak law is saying. This is an illustration for weak law. Whereas strong law concerns itself with the convergence of the entire sequence S n over n. So, weak law only looks at a particular n. Fix an n and you look at the CDF of S n over n, in the probability that it deviates from X bar by epsilon. Whereas, you look at; for strong law, you are looking at the entire sequence. So, for different values of n, so, let me plot with, let us say blue colour.

So, S n for n = 10, it may be here. So, whatever, 100. So, this sequence jumps around. Let us say this is a particular omega, this corresponds to a particular omega. So, it may jump around a bit; sort of gets closer; they jump around; but at some point, the entire sequence; so, if you look at any n nought, there exists an n nought beyond which the entire sequence gets constrained to an epsilon band. You fix whatever epsilon you want.

The sequence S n; this is some particular omega, this blue crosses correspond to the sample averages of some particular little omega. Of course, if you have a different little omega, you will get a different places for these crosses. However, what you are saying is that, on a set of probability 1, these little omegas, on a set of probability 1, these crosses will be constrained to an epsilon band beyond some n nought, never to ever get out.

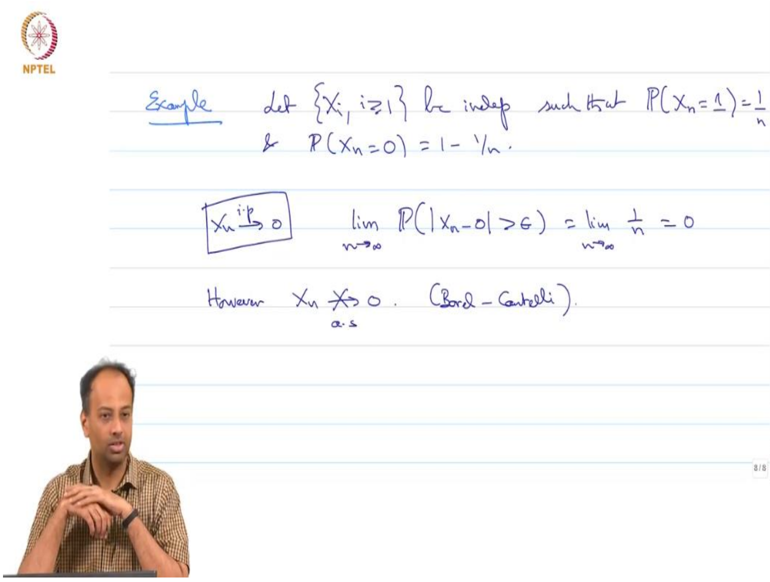
It will never get out beyond some n nought. Weak law is only looking at a particular n, and looking at the probability that you go outside this band. But strong law is looking at the behaviour of the entire sequence S n over n, for n greater than or equal to some n nought. So, our entire sequence converges to X bar; that is what this is saying. So, the entire sequence gets within an epsilon, never to get out.

So, in some sense, you can; so, this is a illustration for strong law. So, in some sense, you can say that strong law has to do with the entire behaviour for large n. The sequence itself constraints itself to its; it will always be within the epsilon band around X bar for large enough n, beyond some n. Whereas, for weak law, we are only looking at a particular n. **"Professor - student conversation starts"** Well, yeah.

I mean, the weak law holds for a particular n nought. The probability that your sample mean deviates from X bar by more than any epsilon goes to 0. But strong law says that, not just at n nought, at n nought and beyond n nought, if the entire sequence of your sample mean will lie within an epsilon band. So, it will, for any epsilon. So, this is convergence of the sequence itself. **"Professor - student conversation ends"**

So, maybe I will just give you; see, this strong law and weak law, I mean, they are both true. But maybe it is a little, it is perhaps a better idea to illustrate with a sequence of random variables which converges almost surely, which converges in probability but not converge almost surely. So, there, you can find an example like that. So, here, both convergences are true. So, let me give you an example.

(Refer Slide Time: 22:44)



So, you take another example. Let X i be independent. So, you have independent random variables, such that probability that X n = 1 is 1 over n. And probability that X n = 0 is equal to 1 minus 1 over n. So, X n's are a sequence of independent random variables, such that the nth random variable is 1 with probability 1 by n. So, this is different from strong. I am done talking about laws of large numbers, okay?

I am giving you a different example, just to illustrate the conceptual difference. So, in this example, so, the hundredth random variable is 1 with probability 1 over 100. So, as you go farther and farther, it is more and more likely to be 0. So, the millionth random variable will be 1 with probability 10 power -6 and 0 with probability 1 - 10 power -6. So, here, this is an example where X n converges to 0 in probability.

But you can argue that X n does not converge to 0 almost surely; does not happen. So, to see that; **"Professor - student conversation starts"** sorry? Yeah. So, that is where I am coming to. **"Professor - student conversation ends"** So, you can argue that X n converges to 0 in probability. Why? It is because, if you; you look at this; limit n tends to infinity, probability that absolute value of X n minus; limit is 0, the proposed limit is 0; > epsilon.

This is what you have to look at. **"Professor - student conversation starts"** What is the probability that absolute X n > epsilon? Say, what values does X n take? 1 or 0, right? So, if it is bigger than epsilon, then? Yeah. So, it has to be 1, right? So, this probability simply limit n tending to infinity 1 over n, which is 0. **"Professor - student conversation ends"** So, you have convergence in probability.

But you can argue that X n does not converge to 0 almost surely. However, X n does not converge to 0 almost surely. To see this; well, one easy way to see this is to invoke Borel-Cantelli lemma, which I suppose some of you have studied. If you have not studied, again, I would suggest, you go and look it up somewhere. So, this comes from; that the fact that X n does not converge to 0 almost surely comes from Borel-Cantelli lemma.

One easy way to see it is Borel-Cantelli lemma. But you can also make some first principal argument. It is not difficult. So, the essence is that, although there are; this X n is overwhelmingly likely to be 0. What happens in this case is that, if you fix some large n, you are guaranteed to have some occasional 1 beyond that n. With probability 1, there will be infinitely many ones popping up.

Even though they are becoming increasingly rare, you can show using Borel-Cantelli lemma that there are infinitely many ones in the sequence, the sequence of X n's. Since there are infinitely many ones in probability 1, this X n cannot go to 0 almost surely, because this 1 keeps popping up once, very rarely, but it still keeps popping up. So, if X n, probability X n = 1 was 1 by n square, then you can argue using Borel-Cantelli lemma that X n will also go to 0 almost surely.

So, if you have not seen Borel-Cantelli lemma, I suggest, you go take a look at it. It is a very, it is a cute result. So, this is an example to really appreciate the difference between these two. So, this almost sure convergence looks at the convergence of the entire sequence. And this sequence, although most of the entries are 0, there will be occasionally ones popping up. So, the sequence does not converge to 0. So, this is a good illustration of this phenomena. So, this module is finished.