The following content is provided under a Creative Commons license. Your support will help MIT OpenCourseWare continue to offer high quality educational resources for free. To make a donation or view additional materials from hundreds of MIT courses, visit MIT OpenCourseWare at ocw.mit.edu.

PROFESSOR: OK. Today we're all through with Markov chains, or at least with finite state Markov chains. And we're going on to renewal processes. As part of that, we will spend a good deal of time talking about the strong law of large numbers, and convergence with probability one.

The idea of convergence with probability one, at least to me is by far the most difficult part of the course. It's very abstract mathematically. It looks like it's simple, and it's one of those things you start to think you understand it, and then at a certain point, you realize that you don't. And this has been happening to me for 20 years now.

I keep thinking I really understand this idea of convergence with probability one, and then I see some strange example again. And I say there's something very peculiar about this whole idea. And I'm going to illustrate that you at the end of the lecture today.

But for the most part, I will be talking not so much about renewal processes, but this set of mathematical issues that we have to understand in order to be able to look at renewal processes in the simplest way.

One of the funny things about the strong law of large numbers and how it gets applied to renewal processes is that although the idea of convergence with probability one is sticky and strange, once you understand it, it is one of the most easy things to use there is. And therefore, once you become comfortable with it, you can use it to do things which would be very hard to do in any other way.

And because of that, most people feel they understand it better than they actually do. And that's the reason why it sometimes crops up when you're least expecting it, and you find there's something very peculiar. OK, so let's start out by talking a little bit about renewal processes. And then talking about this convergence, and the strong law of large numbers, and what it does to all of this.

This is just review. We talked about arrival processes when we started talking about Poisson processes. Renewal processes are a special kind of arrival processes, and Poisson processes are a special kind of renewal process. So this is something you're already sort of familiar with.

All of arrival processes, we will tend to treat in one of three equivalent ways, which is the same thing we did with Poisson processes. A stochastic process, we said, is a family of random variables. But in this case, we always view it as three families of random variables, which are all related. And all of which define the other.

And you jump back and forth from looking at one to looking at the other, which is, as you saw with Poisson processes, you really want to do this, because if you stick to only one way of looking at it, you really only pick up about a quarter, or a half of the picture. OK.

So this one picture gives us a relationship between the arrival epochs of an arrival process, the inter-arrival intervals, the x1, x2, x3, and the counting process, n of t, and whichever one you use, you use the one which is easiest, for whatever you plan to do.

For defining a renewal process, the easy thing to do is to look at the inter-arrival intervals, because the definition of a renewal process is it's an arrival process for which the interrenewals are independent and identically distributed. So any process where the arrivals, inter-arrivals have that property are IID.

OK, renewal processes are characterized, and the name comes from the idea that you start over at each interval. This idea of starting over is something that we talk about more later on. And it's a little bit strange, and a little bit fishy. It's like with a Poisson process, you look at different intervals, and they're independent of each other.

And we sort of know what that means by now. OK, you look at the arrival epochs for a Poisson process. Are they independent of each other? Of course not. The arrival epochs are the sums of the inter-arrival intervals. The inter-arrival intervals are the things that are independent. And the arrival epochs are the sums of inter-arrival intervals.

If you know that the first arrival epoch takes 10 times longer than its mean, then that the second arrival epoch is going to be kind of long, too. It's got to be at least 10 times as long as the mean of the inter-arrival epochs, because each arrival epoch is a sum of these inter-arrival epochs. It's the inter-arrival epochs that are independent.

So when you say that the one interval is independent of the other, yes, you know exactly what you mean. And the idea is very simple. It's the same idea here. But then you start to think you understand this, and you start to use it in a funny way. And suddenly you're starting to say that the arrival epochs are independent from one time to the other, which they certainly aren't.

What renewal theory does is it lets you treat the gross characteristics of a process in a very simple and straightforward way. So you're breaking up the process into two sets of views about it. One is the long term behavior, which you treat by renewal theory, and you use this one exotic theory in a simple and straightforward way for every different process, for every different renewal process you look at.

And then you have this usually incredibly complicated kind of thing in the inside of each arrival epoch. And the nice thing about renewal theory is it lets you look at that complicated thing without worrying about what's going on outside. So the local characteristics can be studied without worrying about the long term interactions.

One example of this, and one of the reasons we are now looking at Markov chains before we look at renewal processes is that a Markov chain is one of the nicest examples there is of a renewal process, when you look at it in the right way.

If you have a recurrent Markov chain, then the interval from one time entering a particularly recurrent state until the next time you enter that recurrent state is a renewal. So we look at the sequence of times at which we enter this one given state. Enter state one over here. We enter state one again over here. We enter state one again, and so forth.

We're ignoring everything that goes on between entries to state one. But every time you enter state 1, you're in the same situation as you were the last time you entered state one. You're in the same situation, in the sense that the inter-arrivals from state one to state one again are independent of what they were before.

In other words, when you enter state one, your successive state transitions from there are the same as they were before. So it's the same situation as we saw with Poisson processes, and it's the same kind of renewal where when you talk about renewal, you have to be very careful about what it is that's a renewable. Once you're careful about it, it's clear what's going on.

One of the things we're going to find out now is one of the things that we failed to point out before when we talked about finite state and Markov chains. One of the most interesting characteristics is the expected amount of time from one entry to a recurrent state until the next time you enter that recurrent state is 1 over pi sub i, where pi sub i is a steady state probability of that steady state.

Namely, we didn't do that. It's a little tricky to do that in terms of Markov chains it's almost trivial to do it in terms of renewal processes. And what's more, when we do it in terms of renewal processes, you will see that it's obvious, and you will never forget it.

If we did it in terms of Markov chains, it would be some long, tedious derivation, and you'd get this nice answer, and you say, why did that nice answer occur? And you wouldn't have any idea. When you look at renewal processes, it's obvious why it happens. And we'll see why that is very soon.

Also, after we finish renewal processes, the next thing we're going to do is to talk about accountable state Markov chains. Markov chains with accountable, infinitely countable set of states. If you don't have a background in renewal theory when you start to look at that, you get very confused. So renewal theory will give us the right tool to look at those more complicated Markov chains.

OK. So we carry from Markov chains with accountably infinite state space comes largely from renewal process. So yes, we'll be interested in understanding that. OK. Another example is GTM queue. The text talked a little bit, and we might have talked in class a little bit about this strange notation a queueing theorist used. There are always at least three letters separated by slashes to talk about what kind of queue you're talking about. The first letter describes the arrival process for the queue.

G means it's a general rival process, which doesn't really mean it's a general arrival process. It means the arrival process is renewal. Namely, it says the arrival process is IID, inter-arrivals. But you don't know what their distribution is. You would call that M if you meant a Poisson process, which would mean memory lists, inter-arrivals.

The second G stands for the service time distribution. Again, we assume that no matter how many servers you have, no matter how the servers work, the time to serve one user is independent of the time to serve other users. But that the distribution of that time has a general distribution. It would be M as you meant a memory list distribution, which would mean exponential distribution.

Finally, the thing at the end says we're talking about IQ with M servers. So the point here is we're talking about a relatively complicated thing. Can you talk about this in terms of renewals? Yes, you can, but it's not quite obvious how to do it. You would think that the obvious way of viewing a complicated queue like this is to look at what happens from one busy period to the next busy period.

You would think the busy periods would be independent of each other. But they're not quite. Suppose you finish one busy period, and when you finish the one busy period, one customer has just finished being served. But at that point, you're in the middle of the waiting for the next customer to arrive.

And as that's a general distribution, the amount of time you have to wait for that next customer to arrive depends on a whole lot of things in the previous interval. So how can you talk about renewals here? You talk about renewables by waiting until that next arrival comes.

When that next arrival comes to terminate the idle period between busy periods, at that time you're in the same state that you were in when the whole thing started before. When you had the first arrival come in. And at that point, you had one a rival there being served you go through some long complicated thing. Eventually the busy period is over. Eventually, then, another arrival comes in.

And presto, at that point, you're statistically back where you started. You're statistically back where you started in terms of all inter-arrival times at that point. And we will have to, even though it's intuitively obvious that those things are independent of each other, we're really going to have to sort that out a little bit, because you come upon many situations where this is not obvious.

So if you don't know how to sort it out when it is obvious, you're not going to know how to sort it out when it's not obvious. But anyway, that's another example of where we have renewals. OK. We want to talk about convergence now. This idea of convergence with probability one. It's based on the idea of numbers converging to some limit.

And I'm always puzzled about how much to talk about this, because all of you, when you first study calculus, talk about limits. Most of you, if you're engineers, when you talk about calculus, it goes in one ear and it goes out the other ear, because you don't have to understand this very much. Because all the things you deal with, the limits exist very nicely, and there's no problem. So you can ignore it.

And then you hear about these epsilons and deltas, and I do the same thing. I can deal with an epsilon, but as soon as you have an epsilon and a delta, I go into orbit. I have no idea what's going on anymore until I sit down and think about it very, very carefully. Fortunately, when we have a sequence of numbers, we only have an epsilon. We don't have a delta. So things are a little bit simpler.

I should warn you, though, that you can't let this go in one ear and out the other ear, because at this point, we are using the convergence of numbers to be able to talk about convergence of random variables, and convergence of random variables is indeed not a simple topic. Convergence of numbers is a simple topic made complicated by mathematicians.

Any good mathematician, when they hear me say this will be furious. Because in fact, when you think about what they've done, they've taken something which is simple but looks complicated, and they've turned it into something which looks complicated in another way, but is really the simplest way to deal with it. So let's do that and be done with it, and then we can start using it for random variables.

A sequence, b1, b2, b3, so forth of real numbers. Real numbers are complex numbers that doesn't make any difference, is said to converge to a limit, b. If for each real epsilon greater than zero, is there an integer M such that bn minus b is less than or equal to epsilon for all n greater than or equal to n? Now, how many people can look at that and understand it? Be honest. Good.

Some of you can. How many people look at that, and their mind just, ah! How many people are in that category? I am. But if I'm the only one, that's good. OK. There's an equivalent way to talk about this. A sequence of numbers, real or complex, is said to converge to limit b.

If for each integer k greater than zero, there's an integer m of k, such that bn minus b is less than or equal to 1 over k for all n greater than or equal to m. OK. And the argument there is pick any epsilon you want to, no matter how small. And then you pick a k, such that 1 over k is less than or equal to epsilon.

According to this definition, bn minus b less than or equal to 1 over k ensures that you have this condition up here that we're talking about. When bn minus b is less than or equal to 1 over k, then also bn minus b is less than or equal to epsilon. In other words, when you look at this, you're starting to see what this definition really means. Here, you don't really care about all epsilon.

All you care about is that this holds true for small enough epsilon. And the trouble is there's no way to specify a small enough epsilon. So the only way we can do this is to say for all epsilon. But what the argument is is if you can assert this statement for a sequence of smaller and smaller values of epsilon, that's all you need.

Because as soon as this is true for one value of epsilon, it's true for all smaller values of epsilon. Now, let me show you a picture which, unfortunately, there's a kind of a complicated picture. It's the picture that says what that argument was really talking about. So if you don't understand the picture, you were kidding yourself when you said you thought you understood what the definition said.

So what the picture says, it's in terms of this 1 over k business. It says if you have a sequence of numbers, b1, b2, b3, excuse me for insulting you by talking about something so trivial. But believe me, as soon as we start talking about random variables, this trivial thing mixed with so many other things will start to become less trivial, and you really need to understand what this is saying.

So we're saying if we have a sequence, b1, b2, b3, b4, b5 and so forth, what that second idea of convergence says is that there's an M1 which says that for all n greater than or equal to M1, b4, b5, b6, b7 minus b all lies within this limit here between b plus 1 and b minus 1.

There's a number M2, which says that as soon as you get bigger than M of 2, all these numbers lie between these two limits. There's a number M3, which says all of these numbers lie between these limits. So it's saying that, it's essentially saying that you can a pipe, and as n increases, you squeeze this pipe gradually down.

You don't know how fast you can squeeze it down when you're talking about convergence. You might have something that converges very slowly, and then M1 will be way out here. M2 will be way over there. M3 will be off on the other side of Vassar Street, and so forth. But there always is such an M1, M2, and M3, which says these numbers are getting closer and closer to b.

And they're staying closer and closer to b. An example, which we'll come back to, where you don't have convergence is the following kind of thing. b1 is equal to 3/4, in this case. b5 is equal to 3/4. b25 is equal to 3/4. b5 to the third is equal to 1. And so forth. These values at which b sub n is equal to 3/4, b is equal to little b plus 3/4 get more and more rare.

So in some sense, this sequence here where b2 up to b4 is zero. b6 up to b24 is zero and so forth. This is some kind of convergence, also. But it's not what anyone would call convergence. I mean, as far as numbers are concerned, there's only one kind of convergence that people ever talk about, and it's this kind of convergence here.

This, although these numbers are getting close to b in some sense. That's not viewed as convergence. So here, even though almost all the numbers are close to b, they don't stay close to b, in a sense. They always pop up at some place in the future, and that destroys the whole idea of convergence. It destroys most theorems about convergence.

That's an example where you don't have convergence. OK, random variables are really a lot more complicated than numbers. I mean, a random variable is a function from the sample space to real numbers. All of you know that's not really what a random variable is.

All of you know that a random variable is a number that wiggles around a little bit, rather than being fixed at what you ordinarily think of a number as being, right? Since that's a very imprecise notion, and the precise notion is very complicated, to build up your intuition about this, you have to really think hard about what convergence of random variables means.

For convergence and distribution, it's not the random variables, but the distribution function of the random variables that converge. In other words, in the distribution function of z sub n, where you have a sequence of random variables, z1, z2, z3 and so forth, the distribution function evaluated at each real value z converges for each z in the case where the distribution function of this final convergent random variable is continuous.

We all studied that. We know what that means now. For convergence and probability, we talked about convergence and probability in two ways. One with an epsilon and a delta. And then saying for every epsilon and delta that isn't big enough, something happens. And then we saw that it was a little easier to describe. It was a little easier to drive describe by saying the convergence in probability.

These distribution functions have to converge to a unit step. And that's enough. They converge to a unit steps at every z except where the step is. We talked about that. For convergence with probability one, and this is the thing we want to talk about today, this is the one that sounds so easy, and which is really tricky. I don't want to scare you about this.

If you're not scared about it to start with, I don't want to scare you. But I would like to convince you that if you think you understand it and you haven't spent a lot of time thinking about it, you're probably due for a rude awakening at some point. So for convergence with probability one, the set of sample paths that converge has probability one.

In other words, the sequence Y1, Y2 converges to zero with probability one. And now I'm going to talk about converging to zero rather than converging to some random variable. Because if you're interested in a sequence of random variables Z1, Z2 that converge to some other random variable Z, you can get rid of a lot of the complication by just saying, let's define a random variable y sub n, which is equal to z sub n minus c.

And then what we're interested in is do these random variables y sub n converged to 0. We can forget about what it's converging to, and only worry about it converging to 0. OK, so when we do that, this sequence of random variables converges to 0 with probability 1. If the probability of the set of sample points for which the sample path converges to 0. If that set of sample paths has probability 1-- namely, for almost everything in the space, for almost everything in its peculiar sense of probability-- if that holds true, then you say you have convergence with probability 1.

Now, that looks straightforward, and I hope it is. You can memorize it or do whatever you want to do with it. We're going to go on now and prove an important theorem about convergence with probability 1. I'm going to give a proof here in class that's a little more detailed than the proof I give in the notes. I don't like to give proofs in class. I think it's a lousy idea because when you're studying a proof, you have to go at your own pace.

But the problem is, I know that students-- and I was once a student myself, and I'm still a student. If I see a proof, I will only look at enough of it to say, ah, I get the idea of it. And then I will stop. And for this one, you need a little more than the idea of it because it's something we're going to build on all the time. So I want to go through this proof carefully. And I hope that most of you will follow most of it. And the parts of it that you don't follow, I hope you'll go back and think about it, because this is really important.

OK, so the theorem says, let this sequence of random variables satisfy the expected value of the magnitude of Y sub n, the sum from n equals 1 to infinity of this is less than infinity.

As usual there's a misprint there. The sum, the expected value of Yn, the bracket should be there. It's supposed to be less than infinity. Let me write that down. The sum from n equals 1 to infinity of expected value of the magnitude of Y sub n is less than infinity. So it's a finite sum. So we're talking about these Yn's when we start talking about the strong law of large numbers. Yn is going to be something like the sum from n equals 1 to m divided by m. In other words, it's going to be the sample average, or something like that. And these sample averages, if you have a mean 0, are going to get small.

The question is, when you sum all of these things that are getting small, do you still get something which is small? When you're dealing with the weak law of large numbers, it's not necessary that that sum gets small. It's only necessary that each of the terms get small. Here we're saying, let's assume also that this sum gets small.

OK, so we want to prove that under this condition, all of these sequences with probability 1 converge to 0, the individual sequences converge.

OK, so let's go through the proof now. And as I say, I won't do this to you very often. But I think for this one, it's sort of necessary.

OK, so first we'll use the Markov inequality. And I'm dealing with a finite value of m here. The probability that the sum of a finite set of these Y sub n's is greater than alpha is less than or equal to the expected value of that random variable. Namely, this random variable here. Sum from n equals 1 to m of magnitude of Y sub n. That's just a random variable. And the probability that that random variable is greater than alpha is less than or equal to the expected value of that random variable divided by alpha.

OK, well now, this quantity here is increasing in Y sub n. The magnitude of Y sub n is a non-negative quantity. You take the expectation of a non-negative quantity, if it has an expectation, which we're assuming here for this to be less infinity, all of these things have to have expectations. So as we increase m, this gets bigger and bigger. So this quantity here is going to be less than or equal to the sum from n equals 1 to infinity of expected value of Y sub n divided by alpha.

What I'm being careful about here is all of the things that happen when you go from finite m to infinite m. And I'm using what you know about finite m, and then being very careful about going to infinite m. And I'm going to try to explain why as we do it. But here, it's straightforward. The expected value of a finite sum is equal to the finite sum of an expected value.

When you go to the limit, m goes to infinity, you don't know whether these expected values exist or not. You're sort of confused on both sides of this equation. So we're sticking to finite values here.

Then, we're taking this quantity, going to the limit as m goes to infinity. This quantity has to get bigger and bigger as m goes to infinity, so this quantity has to be less than or equal to this.

This now, for a given alpha, is just a number. It's nothing more than a number, so we can deal with this pretty easily as we make alpha big enough. But for most of the argument, we're going to view alpha as being fixed.

OK, so now the probability that this sum, finite sum is greater than alpha, is less than or equal to this. This was the thing we just finished proving on the other page. This is less than or equal to that. That's what I repeated, so I'm not cheating you at all here.

Now, it's a pain to write that down all the time. So let's let the set, A sub m, be the set of sample points such that its finite sum of Y sub n of omega is greater than alpha. This is a random-- for each value of omega, this is just a number. The sum of the magnitude of Y sub n is a random variable. It takes on a numerical value for every omega in the sample space. So A sub m is the set of points in the sample space for which this quantity here is bigger than alpha.

So we can rewrite this now as just the probability of A sub m. So this is equivalent to saying, the probability of A sub m is less than or equal to this number here. For a fixed alpha, this is a number. This is something which can vary with m.

Since these numbers here now, now we're dealing with a sample space, which is a little strange. We're talking about sample points and we're saying, this number here, this magnitude of Y sub n at a particular sample point omega is greater than or equal to 0.

Therefore, a sub m is a subset of A sub m plus 1. In other words, as m gets larger and larger here, m here gets larger and larger. Therefore, this sum here gets larger and larger. Therefore, the set of omega for which this increasing sum is greater than alpha gets bigger and bigger. And that's the thing that we're saying here, A sub m is included in A sub m plus 1 for m greater than or equal to 1.

OK, so the left side of this quantity here, as a function of m, is a non-decreasing bounded sequence of real numbers. Yes, the probability of something is just a real number. A probability is a number. So this quantity here is a real number. It's a real number which is non-decreasing, so it keeps moving upward.

What I'm trying to do now is now, I went to the limit over here. I want to go to the limit here. And so I have a sequence of numbers in m. This sequence of numbers is non-decreasing. So it's moving up. Every one of those quantities is bounded by this quantity here. So I have an increasing sequence of real numbers, which is bounded on the top. What happens?

When you have a sequence of real numbers which is bounded-- I have a slide to prove this, but I'm not going to prove it because it's tedious.

Here we have this probability which I'm calling A sub m probability of A sub m. Here I have the probability of A sub m plus 1, and so forth. Here I have this limit up here. All of this sequence of numbers, there's an infinite sequence of them. They're all non-decreasing. They're all bounded by this number here. And what happens?

Well, either we go up to there as a limit or else we stop sometime earlier as a limit. I should prove this, but it's something we use all the time. It's a sequence of increasing or non-decreasing numbers. If it's bounded by something, it has to have a finite limit. The limit is less than or equal to this quantity. It might be strictly less, but the limit has to exist. And the limit has to be less than or equal to b.

OK, that's what we're saying here. When we go to this limit, this limit of the probability of A sub m is less than or equal to this number here.

OK, if I use this property of nesting intervals, when you have A sub 1 nested inside of A sub 2, nested inside of A sub 3, what we'd like to go do is go to this limit. The limit, unfortunately, doesn't make any sense in general. With this property of the axioms, it's equation number 9 in chapter 1 says that we can do something that's almost as good.

What it says that as we go to this limit here, what we get is that this limit is the probability of this infinite union. That's equal to the limit as m goes to infinity of probability of A sub m.

OK, look up equation 9, and you'll see that's exactly what it says. If you think this is obvious, it's not. It ain't obvious at all because it's not even clear that this-- well, nothing very much about this union is clear. We know that this union must be a measurable set. It must have a probability. We don't know much more about it. But anyway, that property tells us that this is true.

OK, so where we are at this point. I don't think I've skip something, have I? Oh, no, that's the thing I didn't want to talk about.

OK, so A sub m is a set of omega which satisfy this for finite m. The probability of this union is then the union of all of these quantities over all m. And this is less than or equal to this bound that we had.

OK, so I even hate giving proofs of this sort because it's a set of simple ideas. To track down every one of them is difficult. The text doesn't track down every one of them. And that's what I'm trying to do here.

We have two possibilities here, and we're looking at this limit here. This limiting sum, which for each omega is just a sequence, a non-decreasing sequence of real numbers. So one possibility is that this sequence of real numbers is bigger than alpha. The other possibility is that it's less than or equal to alpha. If it's less than or equal to alpha, then every one of these numbers is less than or equal to alpha and omega has to be not in this union here.

If the sum is bigger than alpha, then one of the elements in this set is bigger than alpha and omega is in this set. So what all of that says, and you're just going to have to look at that because it's not-- it's one of these tedious arguments. So the probability of omega such that this sum is greater than alpha is less than or equal to this number here.

At this point, we have made a major change in what we're doing. Before we were talking about numbers like probabilities, numbers like expected values. Here, suddenly, we are talking about sample points. We're talking about the probability of a set of sample points, such that the sum is greater than alpha. Yes?

AUDIENCE: I understand how if the whole sum if less than or equal than alpha then every element is. But did you say that if it's greater than alpha, then at least one element is greater than alpha? Why is that?

PROFESSOR: Well, because either the sum is less than or equal to alpha or it's greater than alpha. And if it's less than or equal to alpha, then omega is not in this set. So the alternative is that omega has to be in this set. Except the other way of looking at it is if you have a sequence of numbers, which is approaching a limit, and that limit is bigger than alpha, then one of the terms has to be bigger than alpha. Yes?

AUDIENCE: I think the confusion is between the partial sums and the terms of the sum. That's what he's confusing. Does that make sense? He's saying instead of each partial sum, not each term in the sum.

PROFESSOR: Yes. Except I don't see how that answers the question. Except the point here is, if each partial sum is less than or equal to alpha, then the limit has to be less than or equal to alpha. That's what I was saying on the other page. If you have a sequence of numbers, which has an upper bound on them, then you have to have a limit. And that limit has to be less than or equal to alpha. So that's this case here. We have a sum of numbers as we're going to the limit as m gets larger and larger, these partial sums have to go to a limit. The partial sums are all less than or equal to alpha. Then the infinite sum is less than or equal to alpha, and omega is not in this set here. And otherwise, it is.

OK, if I talk more about it, I'll get more confused. So I think the slides are clear.

Now, if we look at the case where alpha is greater than or equal to this sum, and we take the complement of the set, the probability of the set of omega for which this sum is less than or equal to alpha has-- oh, let's forget about this for the moment.

If I take the complement of this set, the probability of the set of omega, such that the sum is less than or equal to alpha, is greater than 1 minus this expected value here.

Now I'm saying, let's look at the case where alpha is big enough that it's greater than this number here. So this probability is greater than 1 minus this number. So if the sum is less than or equal to alpha for any given omega, then this quantity here converges.

Now I'm talking about sample sequences. I'm saying I have an increasing sequence of numbers corresponding to one particular sample point. This increasing set of numbers is less than or equal. Each element of it is less than or equal to alpha, so the limit of it is less than or equal to alpha. And what that says is the limit of Y sub n of omega, this has to be equal to 0 for that sample point. This is all the sample point argument.

And what that says then is the probability of omega, such that this limit here is equal to 0, that's this quantity here, which is the same as this quantity, which has to be greater than this quantity. This implies this. Therefore, the probability of this has to be bigger than this probability here.

Now, if we let alpha go to infinity, what that says is this quantity goes to 0 and the probability of the set of omega, such that this limit is equal to 0, is equal to 1.

I think if I try to spend 20 more minutes talking about that in more detail, it won't get any clearer. It is one of these very tedious arguments where you have to sit down and follow it step by step.

I wrote the steps that's very carefully. And at this point, I have to leave it as it is. But the theorem has been proven, at least in what's written, if not in what I've said.

OK, let's look at an example of this now. Let's look at the example where these random variables Y sub n for n greater than or equal to 1, have this following property. It's almost the same as the sequence of numbers I talked about before. But what I'm going to do now is-- these are not IID random variables. If they're IID random variables, you're never going to talk about the sum being finite. Sum of the expected values being finite.

How they behave is that for one less than or equal to 5, you pick one of these random variables in here and make it equal to 1. And all the rest are equal to 0. From 5 to 25, you pick one of the random variables, make it equal to 1, and all the others are equal to 0. You choose randomly in here. From 25 to 125, you pick one random variable, set it equal to 1. All the other random variables, set it equal to 0, and so forth forever after.

OK, so what does that say for the sample points? If I look at any particular sample point, what I find is that there's one occurrence of a sample value equal to 1 from here to here. There's exactly one that's equal to 1 from here to here. There's exactly one that's equal to 1 from here to way out here at 125, and so forth.

This is not a sequence of sample values which converges because it keeps popping up to 1 at all these values. So for every omega, Yn of omega is 1 for some n in this interval. For every j and it's 0 elsewhere. This Yn of omega doesn't converge for omega. So the probability that that sequence converges is not 1, it's 0. So this is a particularly awful example. This is a sequence of random variables, which does not converge with probability 1.

At the same time, the expected value of Y sub n is 1 over 5 to the j plus 1 minus 5 to the j. That's the probability that you pick that particular n for a random variable to be equal to 1. It's equal to this for 5 to the j less than or equal to n, less than 5 to the j plus 1.

When you add up all of these things, when you add up expected value of Yn equal to that over this interval, you get 1. When you add it up over the next interval, which is much, much bigger, you get 1 again. When you add it up over the next interval, you get 1 again. So the expected value of the sum-- the sum of the expected value of the Y sub n's is equal to infinity. And what you wind up with then is that this sequence does not converge--

This says the theorem doesn't apply at all. This says that the Y sub n of omega does not converge for any sample function at all. This says that according to the theorem, it doesn't have to converge. I mean, when you look at an example after working very hard to prove a theorem, you would like to find that if the conditions of the theorem are satisfied what the theorem says is satisfied also.

Here, the conditions are not satisfied. And you also don't have convergence with probability 1. You do have convergence in probability, however. So this gives you a nice example of where you have a sequence of random variables that converges in probability. It converges in probability because as n gets larger and larger, the probability that Y sub n is going to be equal to anything other than 0 gets very, very small. So the limit as n goes to infinity of the probability that Y sub n is greater than epsilon-- for any epsilon greater than 0, this probability is equal to 0 for all epsilon. So this quantity does converge in probability. It does not converge with probability 1. It's the simplest example I know of where you don't have convergence with probability 1 and you do have convergence in probability.

How about if you're looking at a sequence of sample averages.

Suppose you're looking at S sub n over n where S sub n is a sum of IID random variables. Can you find an example there where when you have a-- can you find an example where this sequence S sub n over n does converge in probability, but does not converge with probability 1?

Unfortunately, that's very hard to do. And the reason is the main theorem, which we will never get around to proving here is that if you have a random variable x, and the expected value of the magnitude of x is finite, then the strong law of large numbers holds.

Also, the weak law of large numbers holds, which says that you're not going to find an example where one holds and the other doesn't hold. So you have to go to strange things like this in order to get these examples that you're looking at.

OK, let's now go from convergence in probability 1 to applying this to the sequence of random variables where Y sub n is now equal to the sum of n IID random variable divided by n. Namely, it's the sample average, and we're looking at the limit as n goes to infinity of this sample average.

What's the probability of the set of sample points for which this sample path converges to X bar? And the theorem says that this quantity is equal to 1 if the expected value of X is less than infinity. We're not going to prove that, but what we are going to prove is that if the expected value of the fourth moment of X is finite, then we're going to prove that this theorem is true.

OK, when we write this from now on, we will sometimes get more terse. And instead of writing the probability of an omega in the set of sample points such that this limit for a sample point is equal to X bar, this whole thing is equal to 1. We can sometimes write it as the probability that this limit, which is now a limit of Sn of omega over n is equal to X bar. But that's equal to 1.

Some people write it even more tersely as the limit of S sub n over n is equal to X bar with probability 1. This is a very strange statement here because this-- I mean, what you're saying with this statement is not that this limit is equal to X bar with probability 1. It's saying, with probability 1, this limit here exists for a sample point, and that limit is equal to X bar.

The thing which makes the strong law of large numbers difficult is not proving that the limit has a particular value. If there is a limit, it's always easy to find what the value is. The thing which is difficult is figuring out whether it has a limit or not. So this statement is fine for people who understand what it says, but it's kind of confusing otherwise. Still more tersely, people talk about it as Sn over n goes to limit X bar with probability 1. This is probably an even better way to say it than this is because this is-- I mean, this says that there's something strange in the limit here. But I would suggest that you write it this way until you get used to what it's saying. Because then, when you write it this way, you realize that what you're talking about is the limit over individual sample points rather than some kind of more general limit. And convergence with probability 1 is always that sort of convergence.

OK, this strong law and the idea of convergence with probability 1 is really pretty different from the other forms of convergence. In the sense that it focuses directly on sample paths. The other forms of convergence focus on things like the sequence of expected values, or where the sequence of probabilities, or sequences of numbers, which are the things you're used to dealing with. Here you're dealing directly with sample points, and it makes it more difficult to talk about the rate of convergence as n approaches infinity.

You can't talk about the rate of convergence here as n approaches infinity. If you have any n less than infinity, if you're only looking at a finite sequence, you have no way of saying whether any of the sample values over that sequence are going to converge or whether they're not going to converge, because you don't know what the rest of them are. So talking about a rate of convergence with respect to the strong law of large numbers doesn't make any sense.

It's connected directly to the standard notion of a convergence of a sequence of numbers when you look at those numbers applied to a sample path. This is what gives the strong law of large numbers its power, the fact that it's related to this standard idea of convergence.

The standard idea of convergence is what the whole theory of analysis is built on. And there are some very powerful things you can do with analysis. And it's because convergence is defined the way that it is.

When we talk about the strong law of large numbers, we are locked into that particular notion of convergence. And therefore, it's going to have a lot of power. We will see this as soon as we start talking about renewal theory. And in fact, we'll see it in the proof of the strong law that we're going to go through.

Most of the heavy lifting with the strong law of large numbers has been done by the analysis of convergence with probability 1. The hard thing is this theorem we've just proven. And that's tricky. And I apologize for getting a little confused about it as we went through it, and not explaining all the steps completely. But as I said, it's hard to follow proofs in real time anyway. But all of that is done now.

How do we go through the strong law of large numbers now if we accept this convergence with probability 1? Well, it turns out to be pretty easy. We're going to assume that the expected value of the fourth moment of this underlying random variable is less than infinity. So let's look at the expected value of the sum of n random variables taken to the fourth power.

OK, so what is that? It's the expected value of S sub n times S sub n times S sub n times S sub n. Sub n is the sum of Xi from 1 to n. It's also this. It's also this. It's also this. So the expected value of S to the n fourth is the expected value of this entire product here. I should have a big bracket around all of that.

If I multiply all of these terms out, each of these terms goes from 1 to n, what I'm going to get is the sum from 1 to n. Sum over j from 1 to n. Sum over k from 1 to n. And a sum over l from 1 to n. So I'm going to have the expected value of X sub i times X sub j times X sub k times X sub l. Let's review what this is.

X sub i is the random variable for the i-th of these X's. I have n X's-- X1, X2, X3, up to X sub n. What I'm trying to find is the expected value of this sum to the fourth power.

When you look at the sum of something, if I look at the sum of numbers, [INAUDIBLE] of a sub i, times the sum of b sub i, I write it as j.

If i just do this, what's it equal to? It's equal to the sum over i and j of a sub i times a sub j. I'm doing exactly the same thing here, but I'm taking the expected value of it. That's a finite sum. the expected value of the sum is equal to the sum of the expected values.

So if I look at any particular value of X-- of this first X here. Suppose I look at i equals 1 . I suppose I look at the expected value of X1 times-- and I'll make this anything other than 1. I'll make this anything other than 1, and this anything other than 1.

For example, suppose I'm trying to find the expected value of X1 times X2 times X10 times X3. OK, what is that?

Since X1, X2, X3 are all independent of each other, the expected value of X1 times the expected value of all these other things is the expected value of X1 conditional on the values of these other quantities. And then I average over all the other quantities.

Now, if these are independent random variables, the expected value of this given the values of these other quantities is just the expected value of X1. I'm dealing with a case where the expected value of X is equal to 0. Assuming X bar equals 0.

So when I pick i equal to 1 and all of these equal to something other than 1, this expected value is equal to 0. That's a whole bunch of expected values because that includes j equals 2 to n, k equals 2 to n, and X sub l equals 2 to n.

Now, I can do this for X sub i equals 2, X sub i equals 3, and so forth. If i is different from j, and k, and l, this expected value is equal to 0. And the same thing if X sub j is different than all the others. The expected value is equal to 0. So how can I get anything that's nonzero?

Well, if I look at X sub 1 times X sub 1 times X sub 1 times X sub 1, that gives me expected value of X to the fourth. That's not 0, presumably. And I have n terms like that.

Well, I'm getting down to here. What we have is two kinds of nonzero terms. One of them is where i is equal to j is equal to k is equal to l. And then we have X sub i to the fourth power. And we're assuming that's some finite quantity. That's the basic assumption we're using here, expected value of X fourth is less than infinity.

What other kinds of things can we have? Well, if i is equal to j, and if k is equal to l, then I have the expected value of Xi squared times expected value of Xi squared Xk squared. What is that?

Xi squared is independent of Xk squared because i is unequal to k. These are independent random variables. So I have the expected value of Xi squared is what? It's just a variance of X. This quantity here is the variance of X also. So I have the variance of Xi squared, which is squared. So I have sigma to the fourth power. So those are the only terms that I have for this second kind of nonzero term where Xi-- excuse me. not Xi is equal to Xj. That's not what we're talking about. Where i is equal to j. Namely, we have a sum where i runs from 1 to n, where j runs from 1 to n, k runs from 1 to n, and l runs from 1 to n.

What we're looking at is, for what values of i, j, k, and l is this quantity not equal to 0? We're saying that if i is equal to j is equal to k is equal to l, then for all of those terms, we have the expected value of X fourth. For all terms in which i is equal to j and k is equal to l, for all of those terms, we have the expected value of X sub i quantity squared.

Now, how many of those terms are there?

Well, x sub i can be any one of n terms. x sub j can be any one of how many terms? It can't be equal. i is equal to j, how many things can k be? It can't be equal to i because then we would wind up with X sub i to the fourth power. So we're looking at n minus 1 possible values for k, n possible values for i. So there are n times n minus 1 of those terms.

I can also have-- let me write in this way. Times Xk Xl equals k. I can have those terms. I can also have Xi Xj unequal to i. Xk equal to k and Xl equal to i. I can have terms like this. And that gives me a sigma fourth term also.

I can also have Xi Xj unequal to i. k can be equal to i and l can be equal to j. So I really have three kinds of terms. I have three times n times n minus 1 times the expected value of X squared, this quantity squared. So that's the total value of expected value of S sub n to the fourth. It's the n terms for which i is equal to j is equal to k is equal to l plus the 3n times n minus 1 terms in which we have two pairs of equal terms. So we have that quantity here.

Now, expected value of X fourth is the second moment of the random variable X squared. So the expected value of X squared squared is the mean of X squared squared. And that's less than or equal to the variance of X squared, which is this quantity. The expected value of Sn fourth is-- well, actually it's less than or equal to 3n squared times the expected value of X fourth. And blah, blah, blah, until we get to 3 times the expected value of X fourth times the sum from n equals 1 to infinity of

1 over n squared.

Now, is that quantity finite or is it infinite? Well, let's talk of three different ways of showing that this sum is going to be finite.

One of the ways is that this is an approximation, a crude approximation, of the integral from 1 to infinity of 1 over X squared. You know that that integral is finite.

Another way of doing it is you already know that if you take 1 over n times 1 over n plus 1, you know how to sum that. That sum is finite. You can bound this by that.

And the other way of doing it is simply to know that the sum of 1 over n squared is finite.

So what this says is that the expected value of S sub n fourth over n fourth is less than infinity. That says that the probability of the set of omega for which S to the fourth over n fourth is equal to 0 is equal to 1. in other words, it's saying that S to the fourth over omega over n fourth converges to 0. That's not quite what we want, is it?

But the set of sample points for which this quantity converges has probability 1. And here is where you see the real power of the strong law of large numbers. Because if these numbers converge to 0 with probability 1, what happens to the set of numbers is Sn to the fourth of omega divided by n to the fourth, this limit-- if this was equal to 0, then what is the limit as n approaches infinity of Sn of omega over n to the fourth?

If I take the fourth root of this, I get this. If this quantity is converging to 0, the fourth root of this also has to be converging to 0 on a sample path basis of the fact that this converges means that this converges also.

Now, you see if you were dealing with convergence in probability or something like that, you couldn't play this funny game. And the ability to play this game is really what makes convergence in probability a powerful concept. You can do all sorts of strange things with it. And we'll talk about that next time. But that's why all of this works. So that's what says that the probability of the set of omega for which the limits of Sn of omega over n equals 0 equals 1.

Now, let's look at the strange aspect of what we've just done. And this is where things get very peculiar. Let's look at the Bernoulli case, which by now we all understand. So we consider a Bernoulli process, all moments of X exist. Moment-generating functions of X exist. X is about as well-behaved as you can expect because it only has the values 1 or 0. So it's very nice.

The expected value of X is going to be equal to p in this case. The set of sample paths for which Sn of omega over n is equal to p has probability 1. In other words, with probability 1, when you look at a sample path and you look at the whole thing from n equals 1 off to infinity, and you take the limit of that sample path as n goes to infinity, what you get is p. And the probability that you get p is equal to 1.

Well, now, the thing that's disturbing is, if you look at another Bernoulli process where the probability of the 1 is p prime instead of p. What happens then? With probability 1, you get convergence of Sn of omega over n, but the convergence is to p prime instead of to p.

The events in these two spaces are exactly the same. We've changed the probability measure, but we've kept all the events the same. And by changing the probability measure, we have changed one set of probability 1 into a set of probability 0. And we changed another set of probability 0 into set of probability 1. So we have two different events here.

On one probability measure, this event has probability 1. On the other one, it has probability 0. They're both very nice, very well-behaved probabilistic situations. So that's a little disturbing.

But then you say, you can pick p in an uncountably infinite number of ways. And for each way you count p, you have uncountably many events. Excuse me, for each value of p, you have one event of probability 1 for that p. So as you go through this uncountable number of events, you go through this uncountable number of p's, you have an uncountable number of events, each of which has probability 1 for its own p.

And now the set of sequences that converge is, in fact, a rather peculiar sequence to start with. So if you look at all the other things that are going to happen, there are an awful lot of those events also. So what is happening here is that these events that we're talking about are indeed very, very peculiar events. I mean, all the mathematics works out. The mathematics is fine. There's no doubt about it.

In fact, the mathematics of probability theory was worked out. People like Kolmogorov went to great efforts to make sure that all of this worked out. But then he wound up with this peculiar kind of situation here. And that's what happens when you go to an infinite number of random variables. And it's ugly, but that's the way it is. So that what I'm arguing here is that when you go from finite m to infinite n, and you start interchanging limits, and you start taking limits without much care and you start doing all the things that you would like to do, thinking that infinite n is sort of the same as finite n.

In most places in probability, you can do that and you can away with it. As soon as you start dealing with the strong law of large numbers, you suddenly really have to start being careful about this. So from now on, we have to be just a little bit careful about interchanging limits, interchanging summation and integration, interchanging all sorts of things, as soon as we have an infinite number of random variables. So that's a care that we have to worry about from here on.

OK, thank you.