

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](http://ocw.mit.edu)

**ROBERT**

**GALLAGER:**

OK so today we're going to review a little bit Little's theorem-- we're going to review it a little bit, but say a few new things about it. I want to say something about Markov chains and renewal processes, because one of the most valuable things about understanding both is that you can use renewal theory to solve an extraordinary number of Markov chain problems, and you can use Markov chains to solve an awful lot of renewal problems.

And I want to make that clear today because it's a trick that you have perhaps seen in the homework, or perhaps you've missed it. If you missed it you've probably done a lot of extra work that you wouldn't have had to do otherwise. But it's a very, useful thing. So it's worth understanding it.

Finally we'll talk a little bit about delayed renewal processes at the end. What we will say essentially is there's a long section on delayed renewal processes, which goes through and does everything we did for ordinary renewal processes, and as far as almost all the asymptotic properties are concerned, they're exactly the same.

It's just modifying a few of the ideas a little bit. The essential thing there is that when you're looking at something asymptotically, and the limit as  $t$  goes to infinity, what happens in that first little burst of time doesn't really make much difference anymore.

So we will talk about that if we get that far. OK, one of the main reasons why convergence with probability 1 is so important, you've probably wondered why we're spending so much time talking about this, why we use it so often?

I would like you get some idea of why it is often much easier to use than the convergence of probability. So I'm going to give you two examples of that. One of

them is this initial thing which we talked about in the notes also and I talked about on lecture before.

There's this nice theorem which says that if the sequence of random variables converges to some number--  $\alpha$ -- with probability 1, and if  $f$  of  $x$  is a real valued function of a real variable, that's continuous, that  $x$  equals  $\alpha$ . In other words, as you start converging, as you get close to  $\alpha$ , this function is continuous there. And since it's continuous there, you have to get closer and closer. That's the essence of that.

So It says that then it's function of  $z_n$ , a function of a random variable, a real valued function of a random variable is also a random variable. It converges with probability 1 to  $f$  of  $\alpha$ . That was the thing we use to get the strong law for renewals, which says it's a probability that the limit as  $t$  goes to infinity of  $n$  of  $t$  of  $\omega$  over  $t$  is equal to  $1$  over  $\bar{x}$  with probability 1.

In other words, the probability of this set of sample points for which this limit exists is equal to 1 anytime you get confused by one of these statements, that says with probability 1, and by now you're probably writing this just as an add-on at the end, and you often forget that there's an awful lot tucked into that statement.

And I tried to put a little of it there. Initially when we talked about it we put more in, saying the probability of the set of  $\omega$  such that this limit exists is equal to 1. We state it in all sorts of different ways.

But always go back, and think a little bit about what it's really saying. Random variables are not like numbers. Random variables are far more complicated things. Because of that they have many more ways they can approach limits. They have many more peculiar features about them.

But anyway, the fact that this theorem holds true it is a result of a little bit of monkeying around with  $n$  divided by the sum of  $n$  random variables, and associating that with an  $n$  of  $t$  over  $t$ . But it's also associated with this function here. So you have the two things. The thing which is difficult conceptually is this one here. So that's

one place where we used the strong law, where if we try to state a weak law of large numbers for renewals, without being able to go from this strong law to the weak law, it'd really be quite hard to prove it.

You can sit down and try to prove it if you want to, and I think you'll see that it really isn't very easy. So strong law of renewals also holds if the expected value of  $x$  is equal to infinity.

In this case, understanding why this is true really requires you to think pretty deeply about random variables-- and have an infinite expectation of what that means. But the idea here is since  $x$  is a random variable-- in other words, it can't take on infinite values-- except with probability 0. So it's always finite.

So when you add a bunch of them you get something which is still finite. So that  $s_n$  is a random variable. In other words, if you look at the probability that  $s_n$  is less than or equal to  $t$ , and then you let  $t$  go off to infinity, the fact that  $s_n$  is a random variable means that the probability that  $s_n$  is less than or equal to  $t$  goes to 1.

And it does that for sample values with probability 1 is a better way to say it. Here I'm actually stating the convergence in probability 2, because it follows from the convergence with probability 1. Since you have convergence with probability 1,  $n$  of  $t$  over  $t$  also converges in probability, which says that the probability as  $t$  goes to infinity that  $n$  of  $t$  over  $t$  minus 1 over  $\bar{x}$  magnitude is greater than  $\epsilon$ , is equal to 0.

That's this funny theorem we proved about convergence with probability 1. Set of random variables implies convergence in probability.

Here's another theorem about convergence which is called the Elementary Renewal Theorem. We talked about that, we proved half of it. After we talked about talked about Wald's equality. And we said the other half really wasn't very interesting. And I hope some of you at least looked at that. After you look at it, it's a bunch of mathematics, and a bunch of equations.

And it says this-- so we have three limits theorems about  $n$  of  $t$ . About what happens when  $t$  gets large there're this number of renewals that occur, and the time  $n$  of  $t$ . One of them is a strong law, which is really a sample path average. I'm trying to start using the words sample path average instead of time average because I think it gives you a better idea of what's actually going on.

But the strong law is really a sample path argument. The weak law-- this thing here about convergence and probability-- it still tells you quite a bit, because it tells you that as  $t$  it gets large, the probability that  $n$  of  $t$  over  $t$  can be significantly different than one over  $\bar{x}$  is going to 0.

This in a sense tells you even less. I mean why does this tell you less than this does? What's significant thing does this tell you that this doesn't tell you?

Suppose we had a situation where half the time with probability  $1/2$   $n$  of  $t$  over  $t$  is equal to  $2$  over  $\bar{x}$ , and the other half of the time it's equal to 0. That can't happen, But according to this it could happen. The expected value of  $n$  of  $t$  over  $t$  would still be  $1$  over  $\bar{x}$ .

But this statement doesn't tell you when you think about whether  $n$  of  $t$  over  $t$  is really squeezing down on  $1$  of  $\bar{x}$ , it just tells you that the expected value of it is squeezing down on  $1$  over  $\bar{x}$ . So this is really a pretty weak theorem, and you wonder why people spend so much time analyzing it?

I'll tell you why in just a minute. And it's not it's not a pretty story. We talked about residual life. I want to use this which I think you all understand-- I mean for residual life, and for duration and age you draw this picture, and that is perfectly clear what's going on.

So I don't think there's any possibility of confusion with that. Here's the original picture of a sample path picture of a rival apex, of the number of arrivals up until time  $t$  climbing up. And then we look at residual life, the amount of time at any time until the next arrival comes.

This is strictly a sample path idea, for a particular sample path, from 0 to infinity, you

look at the whole thing. In other words, think of setting up an experiment. And this experiment you view with the entire sample path for this particular sample point that you're talking about.

You don't stop at any time, you just keep on going. Obviously you can't keep on going forever, but you keep on going long enough that you get totally bored, and say well I'm not interested in anything after 20 years. And nobody will be interested in my results if I wait more than 20 years, and I'll be dead if I wait much longer than that.

So you say we will take this sample path for a very long time. This is the sample path that we get. We then argue that the integral of  $y$  of  $t$  over  $t$  as a sum of terms, this is a random variable here.  $y$  of  $t$  is a random variable. It's a number. If I put in a particular sample point, each of these terms here are random variables. The sum of them is a random variable.

And if I put in a particular sample point, it's a sum of numbers. Now, we did the following thing with that-- I think it was pretty straightforward. You look at what the sum is up to  $n$  of  $t$ . In other words, for a particular time that you're looking at, the experiment that you do is you integrate  $y$  of  $t$ -- which is this residual life function-- from 0 to  $t$ .

At the same time, at time  $t$  there's a certain number of renewals that have occurred, and you look at  $1$  over  $2t$  times the sum of  $x$  of  $n$  squared, up to that point, not counting this last little bit of stuff here, and then you upper bound it by this sum, counting this little bit of extra stuff here.

And we pretty much proved in class and in the notes that this little extra thing at the end doesn't make any difference even if it's very big. Because you're summing over such a long period of time. That's one argument involved there.

The other argument involved is really very hidden, and you don't see it unless you write things down very carefully. But it tells you why the strong law of numbers are so important. So I wanted to talk about here a little bit.

What that says-- I mean the thing that is kind of fishy here, is here we're summing up to  $n$  of  $t$ , and we don't really know what  $n$  of  $t$  is. It depends on how many arrivals have occurred. And if you write this out carefully as a sample path statement what is it saying? Let's go into the next slide, and we'll see what it's saying.

For the sample point  $\omega$ -- let's assume for the moment that this limit exists-- what you're talking about is the sum from  $n$  equals 1 up to the number of arrivals that have taken place up until time  $t$  for this particular sample point.

And it's the sum of these squares of these inner renewal times, and we're dividing by  $2t$ , because we want to find the rate at which this is all going on. We write it out then as a limit. of  $x$  of  $n$  squared over  $\omega$  divided by  $n$  of  $t$  of  $\omega$ , times  $n$  of  $t$  of  $\omega$ , divided by  $2t$ .

In other words, we simply multiply and divide by  $n$  of  $t$  in  $\omega$ . Why do we want to do that? Because this expression here looks very familiar. It's a sum of random variables. With the sum of  $n$  of  $t$  of  $\omega$  random variables, but we know that as  $t$  gets large,  $n$  of  $t$  of  $\omega$  gets large also.

So that we know that with probability 1, as  $t$  approaches infinity, this sum here-- if we forget about that term-- this sum here by the strong law of large numbers is equal to-- well in fact by the weak law of large numbers-- it is equal to the expected value of  $x$  squared.

If we take the limit of this term, we get the limit of  $n$  of  $t$  of  $\omega$  over  $2t$ . The strong law for renewals tells us that is equal to  $1$  over the expected value of  $x$ . So that gives us our answer.

Now why can we take a limit over this times this, and say that that's equal to the limit of this times the limit of that? If you're dealing with random variables, that's not correct in general.

But here what we're dealing with-- as soon as we put this  $\omega$  in, we're dealing with a sum of numbers. And here we're dealing with a number also. For every value

of  $t$ , this is the number.

In other words, this is a function of  $t$  numerical function of  $t$ , a real valued function of  $t$ . This is a real valued function of  $t$ . And what do you know about the limit of a product of two sequences of real numbers?

If you know a little bit of analysis, then you know that you're going to take the limit of a product, and take the sequence of that limit, then the answer that you get is the first limit times the second limit. OK?

I mean you might not recognize the statement in that generality, but if I ask you what it is the sum of  $a^n$  times  $b^n$ ? If we know that the limit of  $a^n$  is equal to say,  $a$ ? And the limit of  $b^n$  is equal to  $b$ ? Then we know that this sum here, in the limit the limit is  $n$  to infinity, it's just going to be  $a$  times  $b$ . And you're going to sit down and argue that for yourselves looking at the definition of what a limit is. So it's not a complicated thing.

But you can't do that if you're not dealing with a sample path notion of convergence here. You can't make that connection. If you only want to deal with the weak law of large numbers, if you want to say infinity doesn't really make any sense because it doesn't exist, I can't wait that long. And therefore the strong law of large numbers doesn't make any sense. You can't go through that argument, and you can't get this very useful result which says-- What was I trying to prove? I was trying to prove that the expected value of residual life as a time average is equal to the expected value of  $x$  squared divided by 2 times the expected value of  $x$ .

This makes sense over finite times. Yes?

**AUDIENCE:** [INAUDIBLE]?

**ROBERT** Yeah. No, I want to divide by  $n$ . And then I think that makes it all right, OK?

**GALLAGER:** Something like that. Let's see, is that right the way I have it now?

**AUDIENCE:** [INAUDIBLE].

**ROBERT** Yeah I want to do that, it's right too, but if I want to take the summation-- oh no, the

**GALLAGER:** summation makes it messier, you're right.

The thing I'm trying to state is the limit as  $n$  goes to infinity of  $a_n$  times  $b_n$ , is  $a$  times  $b$ . And there are restrictions there, you can't have limit of  $a_n$  going to 0 or something, and the limit of  $b_n$  going to infinity, or strange things like that.

But what I'm arguing here-- I don't want you getting involved with this because I haven't really thought about it. What I want you to think about is the fact that you can use the laws of analysis for real numbers and whether you've studied analysis or not, you're all familiar with those because you all use them all the time.

And when you're dealing with the strong law of large numbers, you can convert everything down to a sample path notion, and then you're simply dealing with limits of real numbers at that point. So you don't have to do anything fancy.

So this result would be hard to do in terms of ensemble averages. If you look at the end of Chapter 4 in the notes, you see that the arguments there get very tricky, and very involved. It does the same thing eventually, but in a much harder way.

OK, for the sample point  $\omega$ , oh, we did that.

OK, Residual life and duration are examples of renewal reward functions, so this is just saying what we've already said so let's not dwell on it anymore.

OK stopping trials. Stopping trials will be on the quiz. The Wald equality will be on the quiz. We will get solutions to problem set seven back to you, although we won't get your graded solutions back to you before the quiz. I hope we'll get them out tomorrow. I hope. Yes, we will. They will be on the web.

Stopping trial is a positive integer valued random variable such as for each  $n$ , the indicator random variable indicator of  $j$  equals  $n$ . In other words, the random variable which takes the value 1 if this random variable  $j$  is equal to  $n$ , takes the value of 0 otherwise. If that random variable is a function of  $x_1$  up to  $x_n$ . If you look at  $x_1$  to  $x_n$ , and you can tell from just looking at that and not looking at the future at all whether you're going to stop at time  $n$ . Then you call that a stopping

trial.

And we generalize that to look at a sequence  $x_1, x_2, \dots, x_n$  and some other set of random variables--  $v_1$  to  $v_n$ . And the same argument. If the rule to stop is a rule which is based on only what you've seen up until time  $n$ , then it's called the stopping trial.

Possibly the effect of stopping trial it's the same except that  $j$  might be a defective random variable. In other words, there might be some small probability that you never stop, that you just keep on going forever. When you look at one of these problems, you use stopping rules on, it's not immediately evident before you start to analyze it what do you ever stop or not. So you have to analyze it somewhat before you know whether you're going to stop.

So it's nice to do things in terms of defective stopping rules, because what you can do there holds true whether or not stop. Wald's equality then says if these random variables or a sequence, IID sequence they each have a mean  $\bar{x}$ , and if  $j$  is a stopping trial, and if the expected value of  $j$  is less than infinity-- in other words, if it exists-- then the sum at the stopping trial satisfied-- the expected value of the sum equals expected value of  $\bar{x}$  times the expected value of  $j$ .

Those of you who did the homework this week noticed three examples of where this is used not to find the expected value of  $s_j$ , but where it's used to find the expected value of  $j$ .

And I guess 90% percent of the examples I've seen do exactly that you can find the expected value of  $s_j$  very easily. I mean you have an experiment where you keep going until something happens. Something happens, this is the sum of these random variables reaches some limit. And when they reach the limit you stop.

If when they reach the limit you stop, you know what that limit is. You know what the expected value of  $s_j$  is, because that's where you stop. And from that, if what  $\bar{x}$  is you then know what the expected value of  $j$  is. So we should really state it as expected value of  $j$  is equal to the expected value of  $s_j$  divided by the expected

value of  $x$ , because that's where you usually use it.

OK this question of whether the expected value of  $j$  has to be less than infinity or not. If the random variable  $x$  you're dealing with, is a positive random variable, then you don't need to worry about that restriction. The only time when you have to worry about this restriction is where  $x$  can be both positive and negative. And then you have to worry about it a little bit. If you don't understand what I just said, go back and look at that example of stop when you're ahead. Because the example of stop when you're ahead, you can't use Wald's equality there, it doesn't apply. Because the expected amount of time until you stop is equal to infinity, and the random variable has both positive and negative values. And because of that, the whole thing breaks down.

OK. Let's talk a little bit about Little's theorem. As we said last time, Little's theorem is essentially an accounting trick. I should tell you something about how I got into teaching this course. I got into teaching it because I was working on networks at the time. Queuing was essential in networks. And mathematicians have taken over the queuing field.

And the results were so complicated, I couldn't understand them. So I started teaching it as a way of trying to understand them. And I looked at Little's theorem, and like any engineer, I said, aha. What's going on here is that the sum of these waiting times is equal to the integral of  $L$  of  $t$ , the difference between  $A$  of  $t$  and  $D$  of  $t$ , as you proceed.

So there's this equality here. If I look at this next busy period, I have that same equality. If I look at the next busy period, I have the same in equality. And anybody with any smidgen of common sense knows that that little amount of business at the end, about that final period, can't make any difference. And because of that, Little's theorem is just this accounting equality.

It says that the sum of the  $w$ 's is equal to the integral of  $L$ . And that's all there is to it. When you look at this more, and you look at funny queuing situations, you start to realize that these busy periods can take very long periods of time. They might be

infinite. All sorts of strange things could happen. So you would like to be able to prove something.

Now, what happens when you try to prove it, this is the other reason why I ignored it when I started teaching this course. Because I didn't understand the strong law of large numbers. I didn't understand what it was. Nobody had ever told me that this was a theorem about sample values.

So I tried to prove it. And I said the following thing. The expected value of  $L$  by definition is  $1$  over  $t$  times the limit.  $L$  is the number of customers in the system at time  $t$ . So we're going to integrate the number of customers in a system over all this period  $t$ . And I'm going to divide by  $1$  over  $t$ .

Oh my God. Would you please interchange that limit in the  $1$  over  $t$ ? I mean it's obvious when you look at it, that has to be what it is. And by this accounting identity, this is equal to the limit of this sum from  $l$  equals zero to  $N$  of  $t$  or  $w$  sub  $l$  over  $t$ . With this question of having omitted this last little busy period, whatever part of it you're in when you get the  $t$ .

I mean, that's the part that's common sense. You know you can do that. Now  $\lambda$  is equal to the limit of  $1$  over  $t$  times  $A$  of  $t$ .  $A$  of  $t$  is just the number of arrivals up until time  $t$ .  $A$  of  $t$ , when we're doing Little's theorem, counts this fictitious arrival at time zero. Or I should say that the renewal theory omits as fictitious the real arrival at time zero, which is what's going on in Little's theorem.

So then the expected value of  $w$  is going to be the limit as  $t$  approaches infinity of  $1$  over  $a$  of  $t$  times the sum of  $w$  sub  $l$ . I'm going to break that up in the same way I broke this thing here up. It's the limit of  $t$  over  $a$  of  $t$  times the limit of  $w$  sub  $l$  from  $l$  equals  $1$  to  $a$  of  $t$ ,  $1$  over  $t$ . Breaking up this limit here requires taking this sample pass view again.

In other words, you look at a particular sample point  $\omega$ . And for that particular sample point  $\omega$ , you simply had the same thing as I was saying here. And I don't know whether I said it right or not. But anyway, that's what we're using. It does

work for real numbers. And therefore, what we wind up with is this, which is 1 over lambda, and this, which is the expected value of L, from this, from the accounting identity.

OK. So again, you're using the strong law as a way to get from random variables to numbers. And you understand how numbers work. OK, one more example of this same idea. One of the problems in the homework, problem set six are problems I thought-- yes.

**AUDIENCE:** About the previous slide, that's from [INAUDIBLE] right?

**PROFESSOR:** Yes, yes. Sorry.

**AUDIENCE:** Then how-- is there an easy way to go from there to the [INAUDIBLE] distributions?

**PROFESSOR:** Oh, to the ensemble average. Yes, there is, but not if you don't read the rest of Chapter 4. And it's not something I'm going to dwell on in class. It's--- it's something which is mathematically messy and fairly intricate. And in terms of common sense, you realize it has to be true.

I mean, if the ensemble average, up until time  $t$ , is approaching the limit, then you must have the situation that that limit is equal to the sample path average. The question is whether it's approaching a limit or not. And that's not too hard to prove. But then you have all this mathematics of going through the details of it, which in fact is tricky.

OK. So back to mark up change and renewal processes. You remember you went through a rather tedious problem where you were supposed to use Chebyshev's inequality to prove something. And maybe half of you recognized that it would be far easier to use Markov's inequality. And if you did that, this is what I'm trying to do here.

The question is, if you look at the expected amount of time, from state  $I$  in a Markov chain, until you return to state  $I$  in the Markov chain again, and you can do that by this theory of attaching rewards to states in a Markov chain. What you wind up with

is this result that the expected renewal time in an ergodic Markov chain, is exactly equal to 1 over the steady state probability of that state.

And if any of you find a simple and obvious way to prove that from the theory of Markov chains, I would be delighted to find it. Because I've never been able to find any way of doing that without going into renewal theory. And renewal theory lets you do it almost immediately. And it's a very useful resolve.

So the argument is the following. You're going to let  $Y_1, Y_2$ , and so forth be the inter-renewal periods. Here we're looking at a sample path point of view again. No, we're not.  $Y_1, Y_2$  are the random variables that are the inter-renewal periods. The elementary renewal theorem is something we talked about. It says that the expected value of  $n \text{ sub } I \text{ of } t \text{ divided by } t$  is equal to 1 over the expected value of  $y$ .

That's the elementary renewal theorem for renewal theory. So we've stopped talking about Markov chains now. We've said, for this Markov chain, you can look at recurrences from successive this is to state  $I$ . That forms a renewal process. And according to the elementary renewal theorem for renewal processes, this is equal to 1 over  $\bar{y}$ .

Now we go back to Markov chains again. Let's look at the probability of being in state  $I$  at time  $t$ , given that we were in state  $I$  at time zero. That's the probability that  $n \text{ sub } I \text{ of } t$ , minus  $n \text{ sub } I \text{ of } t \text{ minus } 1$  is equal to 1. That's the probability that there was an arrival at time  $t$ , which in terms of this renewal process means there was a visit to state  $I$  at time  $t$ .

Every time you get to state  $I$ , you call it a renewal. We've defined a renewal process which gives you a reward of 1 every time you hit state  $I$ , and reward of zero all the rest of the time. So this is equal to probability of  $N_i \text{ of } t \text{ minus } N_i \text{ of } t \text{ minus } 1$ . The probability that that's equal to 1. So its expected value of  $N \text{ sub } I \text{ of } t \text{ minus } N \text{ sub } I \text{ of } t \text{ minus } 1$ . That's the expected value.

This is always greater than or equal to this. This is either 1 or it's zero. You can't have two arrivals. You can't have two visits at the same time. So you add up all of

these things. You sum this from  $n = 1$  up to  $t$ , and what do you get? You sum this, and it's a telescoping series.

So you add expected value of  $n$ ,  $1$  of  $t$  minus  $n = 0$  of  $t$ , which is  $0$ , plus  $n = 2$  of  $t$  minus  $n = 1$  of  $t$ , plus  $n = 3$  of  $t$  minus  $n = 2$  of  $t$ . And everything cancels out except the  $n = I$  of  $t$ . The  $I$  here is just the state we're looking at. We could've left it out. So  $p_{sub I, I}$  of  $t$ , approaches  $\pi_{sub I}$  exponentially. Because down there, very fast, it stays very close. If we sum up over  $n$  of them, and these are quantities which are approaching this limit,  $\pi_{sub I}$ , exponentially fast, then the sum divided by the number of terms we're summing over is just  $\pi_{sub I}$  also.

So what we have is a  $\pi_{sub I}$  that's equal to this limit. That's equal to the expected value of  $n = I$  over  $t$ . The elementary renewal theorem reads out, which is equal to  $1$  over  $\bar{y}$ . So the expected recurrence time for state  $I$  is equal to  $\pi_{sub I}$ .

If you look carefully at what I've done there, I have assumed that-- well, when I did it, I was assuming it was an ergodic Markov chain. I don't think I have to assume that. I think it can be periodic, and this is still true. You can sort that out for yourselves.

That is the first slide I'll use which says, whenever you see a problem trying to prove something about Markov chains and you don't know how to prove it right away, think about renewal theory. The other one will say, whenever you're thinking about a problem in renewal theory, and you don't see how to deal with it immediately, think about Markov chains. You can go back and forth between the two of them.

It's just like when we were dealing with Poisson processes. Again, in terms of solving problems with Poisson processes, what was most useful? It was this idea that you could look at a Poisson process in three different ways. You could look at it as a sum of exponential [INAUDIBLE] arrival times. You could look at it as somebody throwing darts on a line. And you could look at it as a Bernoulli process which is shrunk down to time zero with more and more arrivals.

By being able to look at each of the three ways, you can solve pieces of the problem

using whichever one of these things is most convenient. This is the same thing too. For renewal processes and Markov chains, you can go back and forth between what you know about each of them, and find things about the other. So it's a useful thing.

Expected number of renewals. Expected number of renewals is so important in renewal theory that most people call it  $m$  of  $t$ , which is the expected value, of  $n$  of  $t$ .  $n$  of  $t$ , by definition, is the number of renewals that have occurred by time  $t$ . The elementary renewal theorem says the limit, as  $t$  goes to infinity, of expected value of  $n$  of  $t$  over  $t$  is equal to  $1$  over  $\bar{x}$ .

Now, what happens here? That's a very nice limits theorem. But if you look at trying to calculate the expected value of  $n$  of  $t$  for finite  $t$ , there are situations where you get a real bloody mess. And one example of that is supposedly in our arrival interval, is  $1$  or the square root of  $2$ .

Now, these are not rationally related. So you start looking at the times at which renewals can occur. And it's any integer times  $1$ , plus any integer times the square root of  $2$ . So the number of possibilities within a particular range is growing with the square of that range. So what you find, as  $t$  gets very large, is possible arrival instance are getting more and more dense. There are more and more times when possible arrivals can occur.

There's less and less structure to the time between those possible arrivals. And the magnitude of how much the jump is at that possible time, there's no nice structure to that. Sometimes it's big, sometime it's little. So  $m$  of  $t$  is going to look like an enormously jagged function.

When you get out to some large  $t$ , it's increasing. And you know by the elementary renewal theory that this is going to get close to a straight line here.  $m$  of  $t$  over  $t$  is going to be constant. But you have no idea of what the fine structure of this is.

The fine structure can be extraordinarily complicated. And this bothered people, of course, because it's always bothersome when you start out with a problem that

looks very simple, and you try to ask a very simple question about it. And you get an enormously complicated answer.

So an enormous amount of work has been done on this problem, much more than it's worth. But we can't ignore it all because it impacts on a lot of other things. So if we forget about this kind of situation here, where the inter-arrival interval is either 1 or something which is irrational. If you look at an inter-arrival interval, which is continuous where you have a probability density.

And as a probability density is very nicely defined, then you can do things much more easily. And what you do to do that is you invent something called the renewal equation. Or else you look in a textbook and you find the renewal equation. And you see how that's derived. It's not hard. I'm not going to go through it. Because it has very little to do with what we're trying to accomplish here.

But what the renewal equation says is that the expected number of renewals up until time  $t$  satisfies the equation. But it's the probability that  $x$  is less than or equal to  $t$ . Plus a convolution here of  $m$  of  $t$  minus  $x$  times  $d f$  of  $x$  of  $x$ .

If you state this in terms of densities, it's easier to make sense out of it. It's the interval from zero to  $t$ , of  $1$  plus  $m$  of  $t$  minus  $x$ , all times the density of  $x$   $dx$  This first term here just goes to  $1$  after  $t$  gets very large. So it's not very important. It's a transient term The important part is this convolution here, which tells you how  $m$  of  $t$  is increasing.

You look at that, and you say, that looks very familiar. For electrical engineering students who have ever studied linear systems, that kind of equation is what you spend your life studying. At least it's what I used to spend my life studying back when they didn't know so much about electric engineering. Now, there are too many things to learn. So you might not know very much about this.

But this is a very common linear equation in  $m$  of  $t$ . It's a differential. It's an integral equation from which you can find out what  $m$  of  $t$  is. You can think conceptually of starting out at  $m$  of  $0$ , which you know. And then you use this equation to build

yourself up gradually.

So you started  $m(t) = 0$ . Then you look at  $m(t)$  with  $\epsilon$ , where you get this term here a little bit. And you break this all up into intervals. And pretty soon, you find something very messy happening, which is why people were interested in it.

But it can be solved if this density has a rational or plus transform. Now I don't care about how to solve that equation. That has nothing to do with the rest of the course. But the solution has the following form. The expected value of  $n(t)$  is going up with  $t$  as  $1/\bar{x}$ . I mean, you know it has to be doing that.

Because we know from the elementary renewal theorem that eventually the expected value of  $n(t)/t$  has to look like  $1/\bar{x}$ . So that's that term here,  $m(t)/t$  looks like  $1/\bar{x}$ . There's this next term, which is a constant term, which is  $\sigma^2/\bar{x}^2$ .

That looks sort of satisfying because it's dimensionless. Minus  $1/2$ , plus some function here which is just a transient, which goes away as  $t$  gets large. Now you have to go through some work to get this. But it's worthwhile to try to interpret what this is saying a little bit.

This  $\epsilon$  of  $t$ , it's all this mess we were visualizing here. Except of course, this result doesn't apply to messy things like this. That only applies to those very simple functions that circuit theory people used to study. Because they were relevant for inductors and capacitors and resistors and that kind of stuff.

So we had this most important term. We have this term, which asymptotically goes away. And we have this term here, which looks rather strange. Because what this is saying is that this asymptotic form here,  $m(t)$ , as a function of  $t$ , has this slope here, which is  $1/\bar{x}$  as the slope. Then it has something added to it.

This is  $1/\bar{x}$  plus some constant. And in this case, it's  $\sigma^2/2\bar{x}^2 - 1/2$ . If you notice what that  $\sigma^2/2\bar{x}^2 - 1/2$  is for an exponential random variable, it's zero. So for an

exponential random variable which has no memory, you're right back to this curve here, which makes a certain amount of sense. Because this is some kind of transient at zero, which says where you start out makes a difference.

Now, if you look at a-- I mean, I'm using this answer for simple random variables that I can understand just to get some insight about this. But suppose you look at a random variable, which is deterministic.  $x$  equals 1 with probability 1. What are these renewals look like then?

Start out here. No renewals for a while. Then you go up here. You're always underneath this curve here. You look at a very heavy tailed distribution like this thing we-- you remember this distribution? Where you have  $x$  is equal to  $\epsilon$  with probability  $1 - \epsilon$ , and that's equal to  $1/\epsilon$  with probability  $\epsilon$ .

So that the sample functions look like a whole bunch of very quick. And then there's this very long one. And then a lot of little quick ones. We can look at what that's doing as far as expected value of  $n$  of  $t$  is concerned. You're going to get an enormous number of arrivals. right at the beginning. And then you're going to go for this long period of time with nothing. So you're going to have this term here, which is sticking-- let's put you way up here. OK? So that now you're starting out because of this transient that you're starting out at a particular instant when you don't know whether you're going to get an  $\epsilon$  or a  $1 - \epsilon$ . You're not in one of these periods where you're waiting for this  $1/\epsilon$  period to end. Then you get that term there. Then you get this  $\epsilon$  of  $t$ , which is just a term that goes with  $y$ .

OK so that's what this formula is telling you. Want to talk a little bit about Blackwell's theorem also. Mostly talking about things today that we're not going to prove. I mean, it's not necessary to prove everything in your life. It's important to prove those things which really give you insight about the result. I'm actually going to prove one form of Blackwell's theorem today. And I'll prove it-- guess how I'm going to prove it? I mean, the theme of the lecture today is if something is puzzling in renewal theory, use Markov chain theory. That's what I'm going to do. OK.

What does Blackwell's theorem say? It says that the expected renewal rate for large  $t$  is  $1/\bar{x}$ . OK, it says that if I look at a little tiny increment here instead of this crazy curve here, if I look way out after years and years have gone by, it says that I'm going to be increasing at a rate which is very, very close to nothing. I might be lifted up a little bit or I might be lifted down a little bit, but the amount of change in some very tiny increment here of  $t$  to  $t + \epsilon$ .

Blackwell's theorem tries to say that the expected change in this very tiny increment here of size  $\epsilon$  is equal to  $1/\bar{x}$  times  $\epsilon$ . The expectation equals  $\epsilon$  over expected value of  $x$ . OK?

So it's saying what the elementary renewal theorem says, but something a lot more. The elementary renewal theorem says you take  $n$  of  $t$ , you divide it by  $t$ . When you divide it by  $t$ , you lose all the structure. That's why the elementary renewal theorem is a fairly simple statement to understand. Blackwell's theorem is not getting rid of all the structure, he's just looking at what happens in this very tiny increment here and saying, it behaves in this way.

Now, you think about this and you say, that's not possible. And it's not possible because I've only stated half of Blackwell's theorem. The other part of Blackwell's theorem says if you have a process that can only take jumps at, say, integer times, then this can only change at integer times. And since it can only change at integer time, I can't look at things very close too, I can't look at  $\epsilon$  very small. I can only look at intervals which are multiples of that change time.

OK. So that's what he's trying to say. But the other thing that he's not saying, when I look at this very tiny interval here between  $t$  and  $t + \epsilon$ , it looks like he's saying that  $m$  of  $t$  has a density and that this density is  $1/\bar{x}$ . And it can't have a density, either.

If I had any discrete random variable at all, that discrete random variable can only take jumps at discrete times. So you can never have a density here. If you have a density to start with, then maybe you have a density after you're through. But you can't claim it. So all you can claim is that for very small intervals, you have this kind

of change. You'll see that that's exactly what his theorem says.

But when you make this distinction between densities and discrete, we still haven't captured the whole thing. Because if the interarrival interval is an integer random variable, namely it can only change at integer times, then you know what has to happen here. You can only have changes at integer times.

You can generalize that a little bit by saying if every possible value of the inter-renewal interval is a multiple of some constant, then you just scale the integers to be less or greater. So the same thing happens. When you have these random variables like we'll take one value 1 or value square root of 2, then that's where this thing gets very ugly and nothing very nice happens.

So Blackwell said fundamentally, there are two kinds of distribution functions-- arithmetic and non-arithmetic. I would say there are two kinds-- discrete and continuous. But he was a better mathematician than that and he thought this problem through more. So he knew that he wanted to lump all of the non-arithmetic things together.

A random variable has an arithmetic distribution if its set of possible sample values are integer multiples of some number, say  $\lambda$ . In other words, if it's an integer valued distribution with some scaling on it. That's what he's saying there. All the values are integers, but you can scale the integers bigger or less by multiplying them by some number  $\lambda$ .

And the largest such choice of  $\lambda$  is called the "span of the distribution." So that when you look at  $m$  of  $t$ , what you're going to find is when  $t$  gets larger, only going to get changes at this span value and nothing else. So each time you have an integer times  $\lambda$ , you will get a jump. And each time you don't have an integer times  $\lambda$ , it has to stay constant. OK.

So if  $x$  is arithmetic with span  $\lambda$  greater than 0, then every sum of random variables has to be arithmetic with a span either  $\lambda$  or an integer multiple of  $\lambda$ . So  $n$  of  $t$  can increase only at multiples of  $\lambda$ . If you have a non-

arithmetic discrete distribution like 1 and  $\pi$ , the points at which  $n$  of  $t$  can increase become dense as  $t$  approaches infinity.

Well, what we're doing here is separating life into three different kinds of things. Into arithmetic distributions, which are like integer distributions. These awful things which are like two possible values as discrete, but they take values which are not rational compared with each other. Points at which  $n$  of  $t$  can increase become dense. And finally, the third one is if you have a density and then you have something very nice again. So what Blackwell's theorem says is that limit as  $n$  goes to infinity of  $m$  of  $t$  plus  $\lambda$  minus  $m$  of  $t$  is equal to  $\lambda$  divided by  $\bar{x}$ . OK?

This is if you have an arithmetic  $x$  and a span of  $\lambda$ . This is what we were saying should be the thing that happens and Blackwell proved that that is what happens. And he said that as  $t$  gets very large, in fact, what happens here is that this becomes very, very regular. Every span you get a little jump which is equal to the span size times the expected increase. And then you go level for that span interval and you go up a little more.

So that what you get in a limit is a staircase function. And this is  $\lambda$  here. And this is  $\lambda$  times  $1$  over  $\bar{x}$  here. OK? And this might have this added or subtracted thing that we were talking about before. But in this interval, that's the way it behaves.

What he was saying for non-arithmetic random variables, even for this awful thing like  $\pi$  with probability half and  $1$  with probability half, he was saying pick any  $\delta$  that you want,  $10^{-6}$ , if the limit is  $t$  goes to infinity of  $m$  of  $t$  plus  $\delta$  minus  $m$  of  $t$  is equal to  $\delta$  over  $\bar{x}$ . At density has this behavior, but these awful things, like this example we're looking at, which gets very-- I mean, you have these points of increase which are getting more and more dense and increases which are more and more random. Yes.

**AUDIENCE:** Is this true only for  $\bar{x}$  finite or is this true even if  $\bar{x}$  is infinite?

**PROFESSOR:** If  $\bar{x}$  is infinite? I would guess it's true if  $\bar{x}$  is infinite because then you're

saying that you hardly ever get increases. But I have to look at it more carefully. I mean, if  $x$  bar is infinite, it says that the limit is  $t$  goes to infinity if this difference here goes to 0. And it does. So yes, it should hold true then, but I certainly wouldn't know how to prove it. And if you ask me what odds, I would bet you \$10 to \$1 that it's true. OK? [CHUCKLE].

OK. Blackwell's theorem uses very difficult analysis and doesn't lead to much insight. If you could code that properly, I tried to read proofs of Blackwell's theorem. I've tried to read Blackwell's proof, and Blackwell is a guy who writes extraordinarily well, I've tried to read other people's proof of it, and I've never managed to get through one of those proofs and say, yes, I agree with that. But maybe you can go through them. It's hard to know. But I wouldn't recommend it to anyone except my worst enemies.

The hard case here is this non-arithmetic but discrete distributions. What I'm going to do is prove it for you now as returns to a given state in a Markov chain. In other words, if you have a renewal interval which is integer, you can only get renewals at times 1, 2, 3, 4, 5, up to some finite limit. Then I claim you can always draw a Markov chain for this. And if I can draw a Markov chain for it, then I can solve the problem. And the answer that I get is a surprisingly familiar result for Markov chains and it proves Blackwell's theorem for that special case.

OK, so for any renewal process with inter-renewals at a finite set of integer times, there's a corresponding Markov chain which models returns to state 0. I'm just going to pick an arbitrary state 0. I want to find the intervals between successive returns to state 0. And what I'm doing here in the Markov chain is I start off at state 0. The next thing that happens is I might come back to state 0 in the next interval.

One thinks it got passed out to you, the self-loop is not there. The self-loop really corresponds to returns in time 1, so it should be there. If you don't return in time 1, then you're going to go off to state 1, as we'll call it. From state 1, you can return in one more time interval, which means you get back in time 2. Here you get back in time 1, here you get back in time 2, here you get back in time 3, here you get back

in time 4, and so forth. So you can always draw a chain like this.

And the transition probabilities-- nice homework problem would be to show that the probability of starting at  $i$  and going to  $i + 1$  is exactly this. Why is it this? Well, multiply this probability by this probability by this probability by this probability, and what you get is the probability that the return takes five steps or more. You multiply this by this by this by this. Multiply this thing for different values of  $i$ , and what happens? Successive terms-- this all cancels out. So you wind up with  $1 - x^i$  when you're all done. Or  $1 - x^{i+1}$ . OK?

Now here's the interesting thing. For a lazy person like me, it doesn't make any difference whether I've gotten this formula right or not. I think I have it right, but I don't care. I've only done it right because I know that some of you would be worried about it and some of you would think I was ignorant if I didn't show it to you. But it doesn't make any difference.

When I get done writing down that this is the Markov chain that I'm interested in, I look at this and I say, this is ergodic. I can get from any state here to any other state. I will also assume that it's aperiodic because if it weren't aperiodic, I would just leave out the states 1, 3, 5, and so forth for that period 2 and so forth. OK.

So then we know that the limit as  $n$  goes to infinity of  $p_{00}^n$ . In other words, the probability of being in state 0 at time  $n$  given that you were in state 0 at time 0 is  $\pi_0$ .  $\pi_0$  I can calculate. If I'm careful enough calculating this, I can also calculate the steady state probabilities here. Whether I'm careful here or not, I know that after I get rid of the periodicity here, that I have something which is ergodic here. So I know I can find those  $\pi$ 's.

Now, so we know that  $\pi_0$ , we already saw earlier today, is equal to  $1 / \bar{x}$  over the expected renewal time between visits to state 0. So with  $\pi_0$  equal to  $1 / \bar{x}$  and this equal to  $\pi_0$ , the expected difference between the probability renewal of time  $n$  and the probability renewal of time  $n - 1$  is exactly  $1 / \bar{x}$ , which is exactly what Blackwell said. Yes.

**AUDIENCE:** Can you please explain why this proves the Blackwell theorem? I don't really see it.

**PROFESSOR:** Oh, I proved the Blackwell theorem because what I've shown here is that as  $n$  gets large, the probability that you will be in state 0 at time  $t$  given that you're in state 0 at time 0-- in other words, I'm starting off this renewal process in state 0. So the probability of being in state 0 at time  $n$  is really exactly this thing that Blackwell was talking about. OK? Blackwell was saying-- I mean,  $\lambda$  here is 1 because I've just gotten rid of that. So the limit of  $m$  of  $t$  plus  $t$  plus 1 minus  $m$  of  $t$  is the expectation of a renewal at time  $t$  plus 1. OK? And that's  $1$  over  $\bar{x}$ .

**AUDIENCE:** So why is this renewal process-- why is this Markov chain model exactly what we have in renewal? So you're claiming that we have a renewal if and only if we return to state 0 in this Markov chain.

**PROFESSOR:** Yeah.

**AUDIENCE:** So that's the thing I don't see. Is it supposed to be obvious?

**PROFESSOR:** Oh, you don't see why that's true? Let me try to do that. I thought that at least was obvious, but as I found as I try to develop this course, things which are obvious are the things which are often not obvious.

If I have a random variable, let's say, which takes on the value 1 with probability  $1/2$  and the probability 2 with probability  $1/2$  and I use that as the inter-renewal time for a renewal process, then starting off in time 0 with probability  $1/2$ , I will have a renewal in time 1 and I will have a renewal in time 2 with probability  $1/2$  also.

**AUDIENCE:** Right.

**PROFESSOR:** That's exactly what this says if I draw it for-- I don't need that.

**AUDIENCE:** I see.

**PROFESSOR:** OK?

**AUDIENCE:** OK. Do you mind doing it for a slightly more complicated example just so it's easier

to see in full generality? So it looks like [INAUDIBLE] values or something.

**PROFESSOR:** OK. And then this won't be-- let's make this  $1/2$ . And this  $1/2$ . OK. And this-- what is this going to be? I mean, this has to be 1 at this point.

**AUDIENCE:** So then this would take on 1 with probability  $1/2$ , 2 with probability  $1/4$ , and 3 with probability [INAUDIBLE]?

**PROFESSOR:** I think so, yes.

**AUDIENCE:** Good. Thanks.

**PROFESSOR:** I mean, this is a question of whether I've calculated these numbers right or not. And looking at this example, I'm not at all sure I have. But as I say, it doesn't make any difference. I mean, so long as you buy the fact that if I don't return in time 0, then I'm in some situation where it's already taken me one unit of time, I'm not through, I have to continue and I keep continuing and that's-- OK?

OK. And there's-- oh. I already explained delayed renewal processes. I will explain it again. A delayed renewal process is a modification of a renewal process for which the first inter-renewal interval  $x_1$  has a different distribution than the others. And the intervals are all independent of each other. So that the first interarrival period might do anything. After that, they all do the same thing.

And the argument here is if you're looking at a limit theorem that how long any-- if you're looking at the limit of how many arrivals occur over a very long period of time, the amount of time it takes this first arrival to occur doesn't make any difference. It occurs at some time. And after that, it gets amortized over an enormously long time which is going to infinity.

So if it takes a year for the first arrival to occur, I look at 1,000 years. If it only takes me six months for the first arrival to occur, well, I still look at 1,000 years, but I mean, you see the point. This first interval becomes unimportant compared with everything else. And because of that, the strong law still is going to hold. That says that convergencing probability also occurs. All it does as far as  $m$  of  $t$  is concerned,

the expected value of  $m$  of  $t$ , it moves it up or moves it down, but doesn't change the slope of it and so forth.

Even if the expected time for the first renewal is infinite. And that sounds very strange, but that still is true and it's true by essentially the same argument. You wait until the first arrival occurs. It has to occur at some point. And after that, you can amortize that over as long as you want, you're just looking at a limit. When you look at a limit, you can take as long as you want to, and you take long enough that you wash out the stuff at the beginning.

I mean, if you told me that, I would say you're waving your arms. But if you read the last section of the notes and you summarize it, that's exactly how you will summarize it. OK.