

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.

ROBERT

OK. Today, I want to review a little bit of what we did last time. I think with all the

GALLAGER:

details, some of you sort of lost the main pattern of what was going on. So let me try to talk about the main pattern. I don't want to talk about the details anymore.

I think, for the most part in this course, the best way to understand the details of proofs is to read the notes where you can read them at your own rate, unless there's something wrong in the notes. I will typically avoid that from now on.

The main story that we want to go through, first is the idea of what convergence with probability 1 means. This is a very peculiar concept. And I have to keep going through it, have to keep talking about it with different notation so that you can see it coming any way you look at it.

So what the theorem is-- and it's a very useful theorem-- it says let Y_n be a set of random variables, which satisfy the condition that the sum of the expectations of the absolute value of each random variable is less than infinity.

First thing I want to point out is what that means, because it's not entirely clear what is the means to start with. When you talk about a limit from n to infinity, from n equals 1 to infinity, what it remains is this sum here really means the limit as m goes to infinity of a finite sum. Anytime you talk about a limit of a set of numbers, that's exactly what you mean by it.

So if we're saying that this quantity is less than infinity, it says two things. It says that these finite sums are less than infinity for all a m . And it also says that these finite sums go to a limit. And the fact that these finite sums are going to a limit as m gets big says what's a more intuitive thing, which is that the limit as m goes to infinity.

And here, instead of going from 1 thing m , I'm going from m plus 1 to infinity, and it doesn't make any difference whether I go from m plus 1 to infinity or m to infinity. And what has to happen for this limit to exist is the difference between this and this, which is this, has to go to 0. So that what we're really saying here is that the tail sum has to go to 0 here.

Now, this is a much stronger requirement than just saying that the expected value of the magnitudes of these random variables has to go to 0. If, for example, the expected value of y sub n is $1/n$, then $1/n$, the limit of that as n goes to infinity, is 0. So this is satisfied. But when you sum $1/n$ here, you don't get 0. And in fact, you get infinity. Yes, you get infinity. So this requires more than this.

This limit here, the requirement that that equals 0, implies that the sequence y sub n actually converges to 0 in probability, rather than with probability 1. And the first problem in the homework for this coming week is to actually show that. And when you show it, I hope you find out that it is absolutely trivial to show it. It takes two lines to show it, and that's really all you need here.

The stronger requirement here, let's us say something about the entire sample path. You see, this requirement really is focusing on each individual value of n , but it's not focusing on the sequences. This stronger quantity here is really focusing on these entire sample paths.

OK. So let's go on and review what the strong law of large numbers says. And another note about the theorem on convergence, how useful that theorem is depends on how you choose these random variable, Y_1 , Y_2 , and so forth.

When we were proving the strong law of large numbers in class last time, and in the notes, we started off by assuming that the mean of x is equal to 0. In fact, we'll see that that's just for convenience. It's not something that has anything to do with anything. It's just to get rid of a lot of extra numbers. So we assume this.

We also assume that the expected value of x to the fourth was less than infinity. In other words, we assume that this random variable that we were adding, these IID

random variables, had a fourth moment. Now, an awful lot of random variables have a fourth moment. The real strong law of large numbers, all that assumes is that the expected value of the magnitude of x is less than infinity. So it has a much weaker set of conditions.

Most of the problems you run into, doesn't make any difference whether you assume this or you make the stronger assumption that the mean is equal to 0. When you start applying this theorem, it doesn't make any difference at all, because there's no way you can tell, in a physical situation, whether it is reasonable to assume that the fourth moment is finite and the first moment isn't. Because that question has to do only with the very far tails of the distribution.

I can take any distribution x and I can truncate it at 10 to the google. And if I truncate it at 10 to the google, it has a finite fourth moment. If I don't truncate it, it might not have a fourth moment. There's no way you can tell from looking at physical situations.

So this question here is primarily a question of modeling and what you're going to do with the models. It's not something which is crucial.

But anyway, after we did that, what we said was when we assume that this is less than infinity, we can look at s_n , which is x_1, \dots, x_n all IID. We then took this sum here, took to the fourth moment of it, and we looked at all the cross terms and what could happen. And we found that the only cross terms that worked that were non-0 was where either these quantities were paired together, $x_1 x_1, x_2, x_2$, or where they were all the same.

And then when we looked at that, we very quickly realized that the expected value of s_n to the fourth was proportional to n squared. It was upper banded the by three times n squared times the fourth moment of x .

So when we look at s_n to the fourth divided by n to the fourth, that quantity, summed over n , goes as 1 over n squared, which has a finite sum over n . And therefore, the probability of the limit as n approaches infinity of s_n to the fourth over

n fourth equals 0.

The probability of the sample path, the sample paths converge. The probability of that is equal to 1, which says all sample paths converge with probability 1. So this is enough, not quite enough to prove the strong law of large numbers. Because what we're interested in is not s_n to the fourth over n fourth. We're interested in $s_{\text{sub } n}$ over n . So we have to go one step further.

This is why it's tricky to figure out what random variables you want to use when you're trying to go from this theorem about convergence with probability 1 to the strong law, or the strong law for renewals, or any other kind of strong law doing anything else. It's a fairly tricky matter to choose what random variables you want to talk about there.

But here, it doesn't make any difference, as we've said. If you let s_n of ω over n be a $\text{sub } n$ to the $1/4$. In other words, what I'm doing now is I'm focusing on just one sample path. If I focus on one sample path, ω , and then each one of these terms has some value, a $\text{sub } n$, and then $s_{\text{sub } n}$ of ω over n is going to be equal to a $\text{sub } n$ to the $1/4$ power, the $1/4$ power of this quantity over here.

Now, the question is if the limit of these numbers-- And now remember, now we're talking about a single sample path, but all of these sample paths behave the same way. So if this limit here for one sample path is equal to 0, then the limit of a $\text{sub } n$ to the $1/4$ is also equal to 0.

Why is that true? That's just a result about the real number system. It's a result about convergence of real numbers. If you take a bunch of real numbers, which are getting very, very small, and you take the fourth root of those numbers, which are getting very small, the fourth root is a lot bigger than the number itself.

But nonetheless, the fourth root is being driven to 0 also, or at least the absolute value of the fourth root is being driven to 0 also. You can see this intuitively without even proving any theorems. Except, it is a standard result, just talking about the real number system.

OK. So what that says is the probability that the limit of s_n/n equals 0. This is now talking about sample paths for each sample path. This limit s_n/n either exists or it doesn't exist. If it does exist, it's either equal to 0 or equal to something else. It says that the probability that it exists and that it's equal to 0 is equal to 1.

OK. Now, remember last time we talked about something a little bit funny. We talked about the Bernoulli process. And we talked about the Bernoulli process using one value of p for the probability that x is equal to 1. And what we found is that the probability of the sample path where s_n/n approach to p , the probability of that set was equal to 1.

If we change the probability that x is equal to 1 to some other value, that is still a perfectly well-defined event. The event that a sample path, that s_n/n for a sample path approaches p . But that sample path, that event, becomes 0 as soon as you change p to some other value.

So what we're talking about here is really a probability measure. We're not using any kind of measure theory here, but you really have to be careful about the fact that you're not talking about the number of sequences for which this limit is equal to 0. You're really talking about the probability of it. And you can't think of it in terms of number of sequences.

What's the most probable sequence for a Bernoulli process where p is, say, 0.317? Who knows what the most probable sequence of length n is? What? There is none? Yes, there is. It's all 0's, yes. The all 0's sequence is more likely than anything else.

So why don't these sequences converge to 0? I'm In this case, these sequences actually converge to 0.317, if that's the value of p .

So what's going on? What's going on is a trade-off between the number of sequences and the probability of those sequences. You look at a particular value of n , there's only one sequence which is all 0's. It's much more likely than any of the other sequences.

There's an enormous number of sequences where the relative frequency of them is close to 0.317. They're very improbable, but because of the very large number of them, those are the ones that turn out to have all the probability here.

So that's what's going on in this strong law of large numbers. You have all of these effects playing off against each other, and it's kind of phenomenal that you wind up with an extraordinarily strong theorem like this. When you call this the strong law of large numbers, it, in fact, is an incredibly strong theorem, which is not at all intuitively obvious. It's very, very far from intuitively obvious. If you think it's intuitively obvious, and you haven't studied it for a very long time, go back and think about it again, because there's something wrong in the way you're thinking about it. Because this is an absolutely incredible theorem.

OK. So I want to be a little more general now and talk about sequences converging to a constant α with probability 1. If the probability of the set of ω such as the limit as n goes to infinity is z_n of ω . In other words, we will look at a sample path for a given ω .

Let's look at the probability that that's equal to α rather than equal to 0. That's the case of the Bernoulli process. Bernoulli process with probability p , if you're looking at s_n of ω as s_n of ω over n , then this converges to α . You're looking at s_n to the fourth over n to the fourth, it converges to p to the fourth power. And all those are equal to 1.

Now, note that z_n converges to α if, and only if, $z_n - \alpha$ converges to 0. In other words, we're talking about something that's relatively trivial here. It's not very important. Any time I have a sequence of random variables that converges to some non-0 quantity, p , or α , or whatever, I can also talk about $z_n - \alpha$. And that's another sequence of random variables. And if this converges to α , this converges to 0.

So all I was doing when I was talking about this convergence theorem of everything converging to 0, what I was doing was really taking these random variables and looking at their variation around the mean, at their fluctuation around the mean,

rather than the actual random variable itself. You can always do that. And by doing it, you need to introduce a little more terminology, and you get rid of a lot of mess, because then the mean doesn't appear anymore.

So when we start talking about renewal processes, which we're going to do here, the inter-renewal intervals are positive. It's important, the the fact that they're positive, and that they never go negative. And because of that, we don't really want to subtract the mean off them, because then we would have a sequence of random variables that weren't positive anymore.

So instead of taking away the mean to avoid a couple of extra symbols, we're going to leave the mean in from now on, so these random variables are going to converge to some constant, generally, rather than converge to 0. And it doesn't make any difference.

OK. So now, the next thing I want to do is talk about the strong law for renewal processes. In other words, I want to talk about what happens when you have a renewal counting process where n of t is the number of arrivals up until time t . And we'd like to see if there's any kind of law of large numbers about what happens to n of t over t as t gets very large.

And there is such a law, and that's the kind of thing that we want to focus on here. And that's what we're now starting to talk about.

What was it that made us want to talk about the strong law of large numbers instead of the weak law of large numbers? It was really the fact that these sample paths converge. All these other kinds of convergence, it's the distribution function that converges, it's some probability of something that converges.

It's always something gross that you can look at at every value of n , and then you can find the limit of the distribution function, or find the limit of the mean, or find the limit of the relative frequency, or find the limit of something else. When we talk about the strong law of large numbers, we are really talking about these sample paths.

And the fact that we could go from a convergence theorem saying that s_n to the fourth over n to the fourth approached the limit, this was the convergence theorem. And from that, we could show the $s_{sub n}$ over n also approached the limit, that really is the key to why the strong law of large numbers gets used so much, and particularly gets used when we were talking about renewal processes.

What you will find when we study we study renewal processes is that there's a small part of renewal processes-- there's a small part of the theory which really says 80% of what's important. And it's almost trivially simple, and it's built on the strong law for renewal processes.

Then there's a bunch of other things which are not built on the strong law, they're built on the weak law or something else, which are quite tedious, and quite difficult, and quite messy. We go through them because they get used in a lot of other places, and they let us learn about a lot of things that are very important.

But they still are more difficult. And they're more difficult because we're not talking about sample paths anymore. And you're going to see that at the end of the lecture today. You'll see, I think, what is probably the best illustration of why the strong law of large numbers makes your life simple.

OK. So we had this fact here. This theorem is what's going to generalize this. Assume that $z_{sub n}$ and greater than or equal to 1, this is a sequence of random variables. And assume that this sequence of random variables converges to some number, α , with probability 1.

In other words, you take sample paths of this sequence of a random variables. And those sample paths, there two sets of sample paths. One set of sample paths converged to α , and that has probability 1. There's another set of sample paths, some of them converge to something else, some of them converge to nothing, some of them don't converge at all. Well, they converge to nothing. They don't converge at all.

But that set has probability 0, so we don't worry about it. All we're worrying about is

this good set, which is the set which converges. And then what the theorem says is if we have a function f of x , if it's a real-valued function of a real variable, what does that mean?

As an engineer, it means it's a function. When you're an engineer and you talk about functions, you don't talk about things that aren't continuous. You talk about things that are continuous. So all that's saying is it gives us a nice, respectable function of a variable. It belongs to the national academy of real variables that people like to use.

Then what the theorem says is that the sequence of random variables f of z sub n -- OK, we have a real-valued function of a real variable. It maps, then, sample values of z sub n into f of those sample values. And because of that, just as we've done a dozen times already when you take a real-valued function of a random variable, you have a two-step mapping. You map from ω into z sub n of ω . And then you map from z sub n of ω into f of z sub n of ω . That's a simple-minded idea.

OK. Example one of this, suppose that f of x is x plus β . All this is just a translation, simple-minded function, president of the academy. And supposes that the sequence of random variables converges to α . Then this new set of random variable u sub n equals z sub n plus β . The translated version converges to α plus β . Well, you don't even need a theorem to see that. I mean, you can just look at it and say, of course.

Example two is the one we've already used. This one you do have to sweat over a little bit, but we've already sweated over it, and then we're not going to worry about it anymore. If f of x is equal to x to the $1/4$ for x greater than or equal to 0, and z sub n , random variable, the sequence of random variables, converges to 0 with probability 1 and these random variables are non-negative, then f of z sub n converges to f of 0.

That's the one that's a little less obvious, because if you look at this function, when x is very close to 0, but when x is equal to 0, it's 0, it goes like this. When x is 1, you're up to 1, I guess.

But it really goes up with an infinite slope here. It's still continuous at 0 if you're looking only at the non-negative values.

That's what we use to prove the strong law of large numbers. None of you complained about it last time, so you can't complain about it now. It's just part of what this theorem is saying.

This is what the theorem says. I'm just rewriting it much more briefly. Here I'm going to give a "Pf." For each ω such that $\lim_{n \rightarrow \infty} z_n(\omega) = \alpha$, we use the result for a sequence of numbers that says the limit of $f(z_n(\omega))$, this limit of a set of sequence of numbers, is equal to the function at the limit value.

Let me give you a little diagram which shows you why that has to be true. Suppose you look at this function f . This is $f(x)$. And what we're doing now is we're looking at a_1 here. a_1 is going to be $f(z_1(\omega))$. a_2 , I'm just drawing random numbers here. a_3, a_4, a_5, a_6, a_7 . And then I draw $f(a_1)$.

So what I'm saying here, in terms of real numbers, is this quite trivial thing. If this function is continuous at this point, as n gets large, these numbers get compressed into that limiting value there. And as these numbers get compressed into that limiting value, these values up here get compressed into that limiting value also.

This is not a proof, but the way I construct proofs is not to look for them someplace in a book, but it's to draw a picture which shows me what the idea of the proof is, then I prove it. This has an advantage in research, because if you ever want to get a result in research which you can publish, it has to be something that you can't find in a book. If you do find it in a book later, then in fact your result was not new. And you're not supposed to publish results that aren't new.

So the idea of drawing a picture and then proving it from the picture is really a very valuable aid in doing research. And if you draw this picture, then you can easily construct a proof of the picture. But I'm not going to do it here.

Now, let's go onto renewal processes. Each and inter-renewal interval, $x_{sub i}$, is positive. That was what we said in starting to talk about renewal processes.

Assuming that the expected value of x exists, the expected value of x is then strictly greater than 0. You're going to prove that in the homework this time, too.

When I was trying to get this lecture ready, I didn't want to prove anything in detail, so I had to follow the strategy of assigning problems, and the problem set where you would, in fact, prove these things, which are not difficult but which require a little bit of thought.

And since the expected value of x is greater than or equal to 0, the expected value of s_1 , which is expected value of x , the expected value of x_1 plus x_2 , which is s_2 , and so forth, all of these quantities are greater than 0 also. And for each finite n , the expected value of s_n over n is greater than 0 also. so we're talking about a whole bunch of positive quantities.

So this strong law of large numbers is then going to apply here. The probability of the sample paths, s_n of ω over n , the probability that that sample path converges to the mean of x , that's just 1. Then I use the theorem on the last page about-- It's this theorem.

When I use f of x equals $1/x$ that's continuous at this positive value. That's why it's important to have a positive value for the expected value of x . The expected value of x is equal to 0, $1/x$ is not continuous at x equals 0, and you're in deep trouble. That's one of the reasons why you really want to assume renewal theory and not allow any inter-renewal intervals that are equal to 0. It just louses up the whole theory, makes things much more difficult, and you gain nothing by it.

So we get this statement then, the probability of sample points such that the limit of n over s_n of ω is equal to $1/\bar{x}$. That limit is equal to 1. What does that mean? I look at that and it doesn't mean anything to me. I can't see what it means until I draw a picture. I'm really into pictures today.

This was the statement that we said the probability of this set of ω such that the limit of n over s_n of ω is equal to $1/\bar{x}$ is equal to 1. This is valid whenever you have a renewal process for which \bar{x} exists, namely, the expected

value of the magnitude of x exists. That was one of the assumptions we had in here.

And now, this is going to imply the strong law renewal processes. Here's the picture, which lets us interpret what this means and let's just go further with it. The picture now is you have this counting process, which also amounts to a picture of any set of inter-arrival instance, x_1, x_2, x_3, x_4 , and so forth, and any set of arrival epochs, s_1, s_2 , and so forth.

We look at a particular value of t . And what I'm interested in is n of t over t . I have a theorem about n over s_n of ω . That's not what I'm interested in. I'm interested in n of t over t . And this picture shows me what the relationship is.

So I start out with a given value of t . For a given value of t , there's a well-defined random variable, which is the number of arrivals up to and including time t . From n of t , I get a well-defined random variable, which is the arrival epoch of the latest arrival less than or equal to time t .

Now, this is a very funny kind of random variable. I mean, we've talked about random variables which are functions of other random variables. And in a sense, that's what this. But it's a little more awful than that.

Because here we have this well-defined set of arrival epochs, and now we're taking a particular arrival, which is determined by this t we're looking at. So t defines n of t , and n of t defines $s_{n \text{ of } t}$, if we have this entire sample function. So this is well-defined.

We will find as we proceed with this that this random variable, the time of the arrival most recently before t , it's in fact a very, very strange random variable. There are strange things associated with it. When I look at t minus $s_{n \text{ of } t}$, or when I look at the arrival after t , $s_{n \text{ of } t} + 1$ minus t , those random variables are peculiar. And we're going to explain why they are peculiar and use the strong law for renewal processes to look at them in a kind of a simple way.

But now, the thing we have here is if we look at the slope of b of t , the slope of this line here at each value of t , this slope is n of t divided by t . That's this slope. This

slope here is n of t over s sub n of t . Namely, this is the slope up to the point of the arrival right before t .

This slope is then going to decrease as we move across here. And at this value here, it's going to pop up again. So we have a family of slopes, which is going to look like--

What's it going to do? I don't know where it's going to start out, so I won't even worry about that. I'll just start someplace here. It's going to be decreasing. Then there's going to be an arrival. At that point, it's going to increase a little bit. It's going to be decreasing. There's another arrival. So this is s sub n , s sub n plus 1, and so forth.

So the slope is slowly decreasing. And then it changes discontinuously every time you have an arrival. That's the way this behaves. You start out here, it decreases slowly, it jumps up, then it decreases slowly until the next arrival, and so forth. So that's the kind of thing we're looking at.

But one thing we know is that n of t over t , that's the slope in the middle here, is less than or equal to n of t over s sub n of t . Why is that? Well, n of t is equal to n of t , but t is greater than or equal to the time of the most recent arrival. So we have n of t over t is less than or equal to n of t over s sub n of t .

The other thing that's important to observe is that now we want to look at what happens as t gets larger and larger. And what happens to this ratio, n of t over t ? Well, this ratio n of t over t is this thing we were looking at here, which is kind of a mess. It jumps up, it goes down a little bit, jumps up, goes down a little bit, jumps up.

But the set of values that it goes through-- the set of values that this goes through, namely, the set right before each of these jumps-- is the same set of values as n over s sub n . As I look through this sequence, I look at this n of t over s sub n of t . That's this point here, and then it's this point there.

Anyway, n of t over s sub n of t is going to stay constant as t goes from here over to

there. That's the way I've drawn the picture. I start out with any t in this interval here. This slope keeps changing as t goes from there to there.

This slope does not change. This is determined just by which particular integer value of n we're talking about.

So n of t over s sub n of t jumps at each value of n . So this now becomes just the sequence of numbers. And that sequence of numbers is the sequence n divided by s sub b .

Why is that important? That's the thing we have some control over. That's what appears up here. So we know how to deal with that. That's what this result about convergence added to this result about functions of converging functions tells us.

So redrawing the same figure, we've observed that n of t over t is less than or equal to n of t over s sub n of t . It goes through the same set of values as n over s sub n , and therefore the limit as t goes to infinity of n over t over s sub n of t is the same as the limit as n goes to infinity of n over s sub n . And that limit, with probability 1, is 1 over x bar. That's the thing that this theorem, we just "pfed" said to us.

There's a bit of a pf in here too, because you really ought to show that as t goes to infinity, n of t goes to infinity. And that's not hard to do. It's done in the notes. You need to do it. You can almost see intuitively that it has to happen. And in fact, it does have to happen.

OK. So this, in fact, is a limit. It does exist. Now, we go on the other y , and we look at n of t over t , which is now greater than or equal to n of t over s sub n of t plus 1 . s sub b of t plus 1 is the arrival epoch which is just larger than t .

Now, n of t over s sub n of t plus 1 goes through the same set of values as n over s sub n plus 1 . Namely, each time n increases, this goes up by 1 . So the limit as t goes to infinity of n of t over s sub n of t plus 1 is the same as the limit as n goes to infinity of n over the epoch right after n of t .

This you can rewrite as n plus 1 over s sub n plus 1 times n over n plus 1 . Why do I

want to rewrite it this way? Because this quantity I have a handle on. This is the same as the limit of n over s sub n . I know what that limit is.

This quantity, I have an even better handle on, because this n over n plus 1 just moves-- it's something that starts out low. And as n gets bigger, it just moves up towards 1. And therefore, when you look at this limit, this has to be 1 over x bar also. Since n of t over t is between these two quantities, they both have the same limit. The limit of n of t over t is equal to 1 over the expected value of x .

STUDENT: Professor Gallager?

ROBERT Yeah?

GALLAGER:

STUDENT: Excuse me if this is a dumb question, but in the previous slide it said the limit as t goes to infinity of the accounting process n of t , would equal infinity. We've also been talking a lot over the last week about the defectiveness and non-defectiveness of these counting processes. So we can still find an n that's sufficiently high, such that the probability of n of t being greater than that n is 0, so it's not defective. But I don't know. How do you--

ROBERT n of t is either a random variable or a defective random variable of each value of t .

GALLAGER: And what I'm claiming here, which is not-- This is something you have to prove.

But what I would like to show is that for each value of t , n of t is not defective. In other words, these arrivals have to come sometime or other. OK. Well, let's backtrack from that a little bit.

For n of t to be defective, I would have to have an infinite number of arrivals come in in some finite time. You did a problem in the homework where that could happen, because the x sub i 's you were looking at were not identically distributed, so that as t increased, the number of arrivals you had were increasing very, very rapidly.

Here, that can't happen. And the reason it can't happen is because we've started out with a renewal process where by definition the inter-arrival intervals all have the

same distribution. So the rate of arrivals, in a sense, is staying constant forever.

Now, that's not a proof. If you look at the notes, the notes have a proof of this. After you go through the proof, you say, that's a little bit tedious. But you either have to go through the tedious proof to see what is after you go through it is obvious, or you have to say it's obvious, which is a subject to some question.

So yes, it was not a stupid question. It was a very good question. And in fact, you do have to trace that out. And that's involved here in this in what we've done.

I want to talk a little bit about the central limit theorem for renewals. The notes don't prove the central limit theorem for renewals. I'm not going to prove it here. All I'm going to do is give you an argument why you can see that it sort of has to be true, if you don't look at any of the weird special cases you might want to look at. So there is a reference given in the text for where you can find it.

I mean, I like to give proofs of very important things. I didn't give a proof of this because the amount of work to prove it was far greater than the importance of the result, which means it's a very, very tricky and very difficult thing to prove, even when you're only talking about things like Bernoulli.

OK. But here's the picture. And the picture, I think, will make it sort of clear what's going on. We're talking now about an underlying random variable x . We assume it has a second moment, which we need to make the central limit theorem true. The probability that $s_{\text{sub } n}$ is less than or equal to t for n very large and for the difference between t --

Let's look at the whole statement. What it's saying is if you look at values of t which are equal to the mean for $s_{\text{sub } n}$, which is $n \bar{x}$, plus some quantity α times σ times the square root of n , then as n gets large and t gets correspondingly large, this probability is approximately equal to the normal distribution function.

In other words, what that's saying is as I'm looking at the random variable, $s_{\text{sub } n}$, and taking n very, very large. The expected value of $s_{\text{sub } n}$ is equal to n times \bar{x} , so I'm moving way out where this number is very, very large. As n gets larger

and larger, n increases and \bar{x} increases, and they increase on a slope 1 over \bar{x} . So this is n , this is n over \bar{x} . The slope is n over n over \bar{x} , which is this slope here.

Now, when you look at this picture, what it sort of involves is you can choose any n you want. We will assume the \bar{x} is fixed. You can choose any t that you want to. Let's first hold b fixed and look at a third dimension now, where for this particular value of n , I want to look at the-- And instead of looking at the distribution function, let me look at a probability density, which makes the argument easier to see.

As I look at this for a particular value of n , what I'm going to get is $b \bar{x}$. This will be the probability density of s sub n of x . And that's going to look like, when n is large enough, it's going to look like a Gaussian probability density. And the mean of that Gaussian probability density will be mean $n \bar{x}$.

And the variance of this probability density, now, is going to be the square root of n times σ . What else do I need? I guess that's it. This is the standard deviation.

OK. So you can visualize what happens, now. As you start letting n get bigger and bigger, you have this Gaussian density for each value of n . Think of drawing this again for some larger value of n . The mean will shift out corresponding to a linear increase in n .

The standard deviation will shift out, but only according to the square root of n . So what's happening is the same thing that always happens in the central limit theorem, is that as n gets large, this density here is moving out with n , and it's getting wider with the square root of n . So it's getting wider much more slowly than it's getting bigger. Than It's getting wider much more slowly than it's moving out.

So if I try to look at what happens, what's the probability that n of t is greater than or equal to n ? I now want to look at the same curve here, but instead of looking at it here for a fixed value of n , I want to look at the probability density out here at some fixed value of t . So what's going to happen? The probability density here is going to be the probability density that we we're talking about here, but for this value up

here. The probability density here is going to correspond to the probability density here, and so forth as we move down.

So what's happening to this probability density is that as we move up, the standard deviation is getting a little bit wider. As we move down, standard deviation is getting a little bit smaller. And as n gets bigger and bigger, this shouldn't make any difference.

So therefore, if you buy for the moment the fact that this doesn't make any difference, you have a Gaussian density going this way, you have a Gaussian density centered here. Up here you have a Gaussian density centered here at this point. And all those Gaussian densities are the same, which means you have a Gaussian density going this way, which is centered here. Here's the upper tail of that Gaussian density. Here's the lower tail of that Gaussian density.

Now, to put that analytically, it's saying that the probability that n of t is greater than or equal to n , that's the same as the probability that s sub n is less than or equal to t . So that is the distribution function of s sub n less than or equal to t . When we go from n to t , what we find is that n is equal to t over \bar{x} -- that's the mean we have here-- minus α times σ times the square root of n over \bar{x} .

In other words, what is happening is the following thing. We have a density going this way, which has variance, which has standard deviation proportional to the square root of n . When we look at that same density going this way, ignoring the fact that this distance here that we're looking at is very small, this density here is going to be compressed by this slope here.

In other words, what we have is the probability that n of t greater than or equal to n is approximately equal to Φ of $\alpha \cdot n$ is equal to t over \bar{x} minus this α here times σ times the square root of n over \bar{x} . Nasty equation, because we have an n on both sides of the equation. So we will try to solve this equation. And this is approximately equal to t over \bar{x} minus α times σ times the square root of n over \bar{x} times the square root of x .

Why is that true? Because it's approximately equal to the square root-- Well, it is equal to the square root of t over \bar{x} , which is this quantity here. And since this quantity here is small relative to this quantity here, when you solve this equation for t , you're going to ignore this term and just get this small correction term here.

That's exactly the same thing that I said when I was looking at this graphically, when I was saying that if you look at the density at larger values than n , you get a standard deviation which is larger. When you look at a smaller value of n , you get a standard deviation which is smaller.

Which means that when you look at it along here, you're going to get what looks like a Gaussian density, except the standard deviation is a little expanded up there and little shrunk down here. But that doesn't make any difference as n gets very large, because that shrinking factor is proportional to the square root of n rather than n .

Now beyond that, you just have to look at this and live with it. Or else you have to look up a proof of it, which I don't particularly recommend.

So this is the central limit theorem for renewal processes. n of t tends to Gaussian with a mean t over \bar{x} and a standard deviation σ times square root of t over \bar{x} times 1 over square root of x . And now you sort of understand why that is, I hope.

OK. Next thing I want to go to is the time average residual life. You were probably somewhat bothered when you saw with Poisson processes that if you arrived to wait for a bus, the expected time between buses turned out to be twice the expected time from one bus to the next.

Namely, whenever you arrive to look for a bus, the time until the next bus was an exponential random variable. The time back to the last bus, if you're far enough in, was an exponential random variable. The sum of two, the expected value from the time before until the time later, was twice what it should be.

And we went through some kind of song and dance saying that's because you come in at a given point and you're more likely to come in during one of these longer

inter-arrival periods than you are to come in during a short inter-arrival. And it has to be a song and a dance, and it didn't really explain anything very well, because we were locked into the exponential density.

Now we have an advantage. We can explain things like that, because we can look at any old distribution we want to look at, and that will let us see what this thing which is called the paradox of residual life really amounts to. It's what tells us why we sometimes have to wait a very much longer time than we think we should if we understand some particular kind of process.

So here's where we're going to start. What happened? I lost a slide. Ah. There we are.

Residual life, y of t , of a renewal process at times t , is the remaining time until the next renewal. Namely, we have this counting process for any given renewal process. We have this random variable, which is the time of the first arrival after t , which is $s_{n \text{ of } t} + 1$. And that difference is the duration until the next arrival.

Starting at time t , there's a random variable, which is the time from t until the next arrival after t . That is specifically the arrival epoch of the arrival after time t , which is $s_{n \text{ of } t} + 1$ minus the number of arrivals that have occurred up until time t .

You take any sample path of this renewal process, and y of t will have some value in that sample path. As I say here, this is how long you have to wait for a bus if the bus arrivals were renewal processes.

STUDENT: Should it also be $s_{n \text{ of } t}$, where there is minus sign on that?

ROBERT GALLAGER: No, because just by definition, a residual life, the residual life starting at time t is the time for the next arrival. There's also something called age that we'll talk about later, which is how long is it back to the last arrival. In other words, that age is the age of the particular inter-arrival interval that you happen to be in. Yes?

STUDENT: It should be $s_{n \text{ of } t} + 1$ minus t instead of minus N , because it's the time from t to--

ROBERT

Yes, I agree with you. There's something wrong there.

GALLAGER:

I'm sorry. That I should be $s_{n+1} - t$. Good. That's what happens when you make up slides too late at night. And as I said, we'll talk about something called age, which is $a(t) = t - s_n$.

So this is a random variable defined at every value of t . What we'd like to look at now is what does that look like as a sample function as a sample path. The residual life is a function of t --

Nicest way to view residual life is that it's a reward function on a renewal process. A renewal process just consists of these-- Well, you can look at in three ways. It's a sequence of inter-arrival times, all identically distributed. It's the sequence of arrival epochs. Or it's this unaccountably infinite number of random variables, n of t .

Given that process, you can define whatever kind of reward you want to, which is the same kind of reward we were talking about with Markov chains, where you just define some kind of reward that you achieve at each value of t . But that reward-- we'll talk about reward on renewal processes-- is restricted to be a reward which is a function only of the particular inter-arrival interval that you happen to be in.

Now, I don't want to talk about that too much right now, because it is easier to understand residual life than it is to understand the general idea of these renewal reward functions. So we'll just talk about residual life to start with, and then get back to the more general thing.

We would like, sometimes, to look at the time-average value of residual life, which is you take the residual life at time τ , you integrate it at the time t , and then you divide by time t . This is the time average residual life from 0 up to time t . We will now ask the question, does this have a limit as t goes to infinity? And we will see that, in fact, it does.

So let's draw a picture. Here is a picture of some arbitrary renewal process. I've

given the inter-arrival times, x_1, x_2 , so forth, the arrival epochs, s_1, s_2 , so forth, and n of t . Now, let's ask, for this particular sample function what is the residual life? Namely, at each value of t , what's the time until the next arrival occurs?

Well, this is a perfectly specific function of this individual sample function here. This is a sample function, now in the interval from 0 to s_1 , the time until the next arrival. It starts out as x_1 , drops down to 0.

Now, don't ask the question, what is my residual life if I don't know what the rest of the sample function is? That's not the question we're asking here. The question we're asking is somebody gives you a picture of this entire sample path, and you want to find out, for that particular picture, what is the residual life at every value of t .

And for a value of t very close to 0, the residual life is the time up to s_1 . So it's decaying linearly down to 0 at s_1 . At s_1 , it jumps up immediately to x_2 , which is the time from any time after s_1 to s_2 . I have a little circle down there. And from x_2 , it decays down to 0.

So we have a whole bunch here of triangles. So for any sample function, we have this sample function of residual life, which is, in fact, just decaying triangles. It's nothing more than that. For every t in here, the amount of time until the next arrival is simply s_2 minus t , which is that value there. This decay is with slope minus 1, so there's nothing to finding out what this is if you know this. This is a very simple function of that.

So a residual-life sample function is a sequence of isosceles triangles, one starting at each arrival epoch. The time average for a given sample function is, how do I find the time average starting from 0 going up to some large value t ? Well, I simply integrate these isosceles triangles. And I can integrate these, and you can integrate these, and anybody who's had a high school education can integrate these, because it's just the sum of the areas of all of these triangles.

So this area here is $\frac{1}{2}$ times x sub i squared, then we divide by t . So it's $\frac{1}{2}$

over t times this integral. This integral here is the area of the first triangle plus the area of the second triangle plus $1/3$ plus $1/4$, plus this little runt thing at the end, which is, if I pick t in here, this little runt thing is going to be that little trapezoid, which we could figure out if we wanted to, but we don't want to.

The main thing is we get this sum of squares here, there that's easy enough to deal with. So this is what we found here.

It is easier to bound this quantity, instead of having that little runt at the end, to drop the runt to this side and to extend the runt on this side to the entire isosceles triangles. So this time average residual life at the time t is between this and this.

The limit of this as t goes to infinity is what? Well, it's just a limit of a sequence of IID random variables. No, excuse me. We are dealing here with sample function. So what we have is a limit as t goes to infinity. And I want to rewrite this here as x sub n squared divided by n of t times n has to t over $2t$. I want to separate it, and just divide it and multiply by n of t .

I want to look at this term. What happens to this term as t gets large? Well, as t gets large, n of t gets large. This quantity here just goes through the same set of values as the sum up to some finite limit divided by that limit goes through. So the limit of this quantity here is just the expected value of x squared.

What is this quantity here? Well, this is what the renewal theorem deals with. This limit here is 1 over 2 times the expected value of x . That's what we showed before. This goes to a limit, this goes to a limit, the whole thing goes to a limit. And it goes to limit with probability 1 for all sample functions. So this time average residual life has the expected value of x squared divided by 2 times the expected value of x .

Now if you look at this, you'll see that what we've done is something which is very simple, because of the fact we have renewal theory at this point. If we had to look at the probabilities of where all of these arrival epochs occur, and then deal with all of those random variables, and go through some enormously complex calculation to find the expected value of this residual life at the time t , it would be an incredibly

hard problem. But looking at it in terms of sample paths for random variables, it's an incredibly simple problem.

Want to look at one example here, because when we look at this, well, first thing is just a couple of examples to work out. The time average residual life has expected value of x squared over 2 times the expected value of x . If x is almost deterministic, then the expected value of x squared is just a square of the expected value of x .

So we wind up with the expected value of x over 2, which is sort of what you would expect if you look from time 0 to time infinity, and these arrivals come along regularly, then the expected time you have to wait for the next arrival varies from 0 up to x . And the average of it is $x/2$. So no problem there.

If x is exponential, we've already found out that the expected time we have to wait until the next arrival is the expected value of x , because these arrivals are memoryless. So I start looking at this Poisson process at a given value of t , and the time until the next arrival is exponential, and it's the same as the expected time from one arrival to the next arrival. So we have that quantity there, which looks a little strange.

This one, this is a very peculiar random variable. But this really explains what's going on with this kind of paradoxical thing, which we found with a Poisson process, where if you arrive to wait for a bus, you're waiting time is not any less because of the fact that you've just arrived between two arrivals, and it ought to be the same distance back to the last one at a distance of first one. That was always a little surprising.

This, I think, explains what's going on better than most things. Look at a binary random variable, x , where the probability that x is equal to ϵ is $1 - \epsilon$, and the probability that x is equal to $1/\epsilon$ is ϵ . And think of ϵ as being very large.

So what happens? You got a whole bunch of little, tiny inter-renewal intervals, which are ϵ apart. And then with very small probability, you get an enormous one.

And you wait for $1/\epsilon$ for that one to be finished. Then you've got a bunch of little ones which are all ϵ apart. Then you get an enormous one, which is $1/\epsilon$ long.

And now you can see perfectly well that if you arrive to wait for a bus and the buses are distributed this way, this is sort of what happens when you have a bus system which is perfectly regular but subject to failures. Whenever you have a failure, you have an incredibly long wait. Otherwise, you have very small waits.

So what happens here? The duration of this whole interval here of these little, tiny inter-arrival times, the distance between failure in a sense, is $1 - \epsilon$, as it turns out. It's very close to 1. This distance here is $1/\epsilon$. And this quantity here, if you work it out--

Let's see. What is it? We take this distribution here, we look for the expected value of x^2 . Let's see. $1/\epsilon^2$ times ϵ , which is $1/\epsilon$ plus $1 - \epsilon$ times something. So the expected time that you have to wait if you arrive somewhere along here is $1/2\epsilon$. If ϵ is very small, you have a very, very long waiting time because of these very long distributions in here. You normally don't tend to arrive in any of these periods or any of these periods.

But however you want to interpret it, this theorem about renewals tells you precisely that the time average residual life is, in fact, this quantity $1/2\epsilon$. That's this paradox of residual life. Your residual life is much larger than it looks like it ought to be, because it's not by any means the same as the expected interval between successive arrivals, which in this case is very small.

OK I think I'll stop there. And we will talk more about this next time.