

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

PROFESSOR:

So let's get started. With the quiz out of the way we can now move ahead again. I want to talk a little bit about the major renewal theorem. Partly review, but partly it's something that we need because we're going on to talk about countable-state Markov chains, and almost all the analysis there is based on what happens when you're dealing with renewal processes. Especially discrete renewal processes that are the renewals when you have recurrences from one state to the same state at some later point. So, in order to do that, we also have to talk a little bit about age and duration at a given time instead of in terms of the sample average, because that's sort of essential also. And it gives you a nice interpretation of why these peculiar things happen about duration being so much longer than it seems like it should be, and things of that sort. We'll probably spend close to half the lecture dealing with that, and the other half dealing with countable-state Markov chains.

We really have three major theorems dealing with renewal processes. One of them is a sample path time average, which we've called the strong law for renewals. It says that if you look at individual sample paths, it says that the limit as an individual sample path, m as the number of arrivals over that sample path if you look at the limit from 0 to infinity of the number of arrivals divided by time, that's 1 over \bar{x} . And the set of sample paths for which that's true has probability 1. That's what that statement is supposed to say.

The next one is the elementary renewal theorem. When you look at it carefully, that doesn't say anything, or hardly anything. All it says is that if you look at the limit as t goes to infinity, and you also take the expected value, the expected value is the number of renewals over a period t divided by t . In other words, the rate of renewals expected value of that goes to 1 over \bar{x} . Also, this is also leading to the point of view that the rate of renewals is 1 over the expected value of the expected in a

renewal time. Now, why doesn't that mean anything just by itself? Well, if you look at a set of random variables, which is, for example, [? zero ?] most of the time and a very large value with a very small probability.

Suppose we look at a set of non IID random variables where $x_{sub i}$ is equal to 0, the probability $1 - p$, and is equal to some very large value, say $1/p$, with probability p . Let's make it $1/p^2$. Then, in that case, this doesn't tell you anything the expected value of $x_{sub i}$ if we look at $x_{sub i}$ as being n of t over t . The expected value of $x_{sub i}$ gets very large. There's a worse thing that can happen here though, which is suppose you have this situation where n of t fluctuates within. And n of t typically looks like t over \bar{x} plus the square root of t . Now, if you look at this, I mean this is expected value of n of t . And as t wanders around, expected value of n of t goes up linearly the way it's supposed to, but it fluctuates around that point with something like square root of t . If you divide by t , everything is fine, but if you look at n of t over some large period of time, it's going to move all over the place. So knowing that the expected value of n of t over t is equal to $1/\bar{x}$ really doesn't tell you everything you'd like to know about this kind of process.

And finally, we have Blackwell's theorem, which is getting closer to what we'd like to know. What I'm trying to argue here is that, in a sense, the major theorems, the things that you really want to know, there's a strong law and the strong law says with probability 1 all these paths behave in the same way. And Blackwell's theorem, if you take m of t as the expected value of n of t , that's the thing which might wander all over the place here. Blackwell says that if the inter-renewal interval is an arithmetic random variable, namely if it only takes on values that integers time some span called λ then the limit of the expected value of n of t plus λ , namely 1 unit beyond where we start, minus the expected value of t is λ/\bar{x} . It says you only move up each unit of time by a constant here divided by \bar{x} . If you look at this long term behavior, you're still moving up at a rate of $1/\bar{x}$. But since you can only have jumps at intervals of λ , that's what causes that λ there. Most of the time when we try to look at what's going on here, and for most of the examples that we want to look at, we'll just set λ equal to 1.

Especially when we look at Markov chains. For Markov chains you only get changes every step of the Markov chain, and you might as well visualize steps of a Markov chain as being at unit times

AUDIENCE: I didn't quite catch the point of the first example. What was the point of that? You said the elementary renewal theorem [INAUDIBLE].

PROFESSOR: Oh, the point in this example is, you might not even have an expectation, but at the same time this kind of situation is a situation where n of t effectively moves up. Well, n of t over t effectively moves up at a nice regular way, and you have a strong law there, you have a weak law there, but you don't have the situation you want. Namely, looking at an expected value does not always tell you everything you'd like to know. I'm saying there's more to life than expected values. And this one says the other thing that if you look over time you're going to have things wobble around quite a bit, and Blackwell's theorem says, yes that wobbling around can occur over time, but it doesn't happen very fast.

The second one here is kind of funny. This is probably why this is called somebody's theorem instead of some lemma that everybody's known since the 17th century. Blackwell was still doing research. About 10 years ago, I heard him give a lecture in the math department here. He was not ancient at that time, so this was probably then sometime around the '50s or '60s, back when Blackwell did a lot of work on stochastic processes was being done.

The reason why this result is not trivial is because of this part here. If you have a renewal process where some renewals occur, say with interval 1, and some renewals occur with interval π is a good example of it. Then as t gets larger and larger, the set of times at which renewals can occur becomes more and more dense. But along with it becoming more and more dense, the jumps you get at each one of those times gets smaller and smaller, and pretty soon n of t , this expected value of n of t , is looking like something which, if you don't have your glasses on, it looks like it's going up exactly the way it should be going up, but if you put your glasses on you see an enormous amount of fine structure there. And the fine

structure never goes away. We'll talk more about that in the next slide.

You can really look at this in a much simpler way, it's just that people like to look at the expected number of renewals at different times. Since renewals can only occur at time separated by λ , and since you can't have two renewals at a time, the only question is, do you get a renewal at $m\lambda$ or don't you get a renewal at $m\lambda$? And therefore this limit here, a limit of m of t plus λ minus m of t , is really the question of whether you've gotten a renewal at t plus λ . So, you can rewrite this condition as the limit of the probability of a renewal at $m\lambda$ as equal to λ over \bar{x} .

What happens with the scaling here? I mean, is this scaling right? Well, the expected time between renewals is 1 over \bar{x} . If I take this renewal process and I scale it, measuring it in milliseconds instead of seconds, what's going to happen? 1 over \bar{x} is going to change by a factor of $1,000$. You really want λ to change by a factor of $1,000$, because the probability of a jump at one of these possible places for a jump is still the same. So you need a λ over the \bar{x} here.

If you model an arithmetic renewal process as a Markov chain, starting in renewal state 0 , this essentially says that $p_{0,0}$ the probability of going from 0 to 0 and in steps is some constant. I mean, calling π_0 , which is just what we've always called it, and π_0 has to be, in this case, λ over \bar{x} or 1 over \bar{x} . But what it's saying is that this reaches a constant. You know, that's the hard thing to prove. It's not hard to find this number. It's hard to prove that it does reach a limit, and that's the same thing that you're trying to prove here, so this and this are really saying the same thing. That the probability of renewal at $m\lambda$ is λ over \bar{x} , and expected value of renewals at a given time is also 1 over \bar{x} . So, that's really the best you could hope for.

If you look at the non-arithmetic case, I think one way of understanding it is to take the results that we had before, which we stated as part of Blackwell's theorem, which is this one. Divide both sides by δ and see what happens. If you divide m

of t plus δ minus m of t by δ , it looks like you're trying to go to a limit and get a derivative. We know we can't get a derivative. So, what's going on then? It says for any δ that you want the limit as t goes to infinity of this ratio has to be 1 over \bar{x} . So, it says that if you take the limit as δ goes to 0 of this quantity here, you still get 1 over \bar{x} . But this is a good example of the case where you really can't interchange these two limits. This is not a mathematical fine point. I mean, this is something that really cuts to the grain of what renewal processes are all about. So, you really ought to think through on your own why this makes sense when you take the limit as δ goes to 0 after you take this limit, whereas if you try to interchange the limits then you'd be trying to say you're taking the limit as t approaches infinity of a derivative here, and there isn't any derivative, so there isn't any limit, so nothing works.

Let's look a little bit at age and duration at a particular value of t . Because we looked at age and duration only in terms of sample paths, and we went to the limit and we got very surprising results. We found out that the expected age was the expected value of the inter-renewal interval squared divided by the expected value of the inter-renewal interval all divided by 2 , which didn't seem to make any sense. Because the inter-renewal time was just this random variable x , the expected inter-renewal time is \bar{x} . And yet you have this enormous duration. And we sort of motivated this in a number of ways. We drew some pictures and all of that, but it didn't really come together. When we look at this way I think it will come together for you.

So let's assume an arithmetic renewal process will span 1 . I partly want to look at an arithmetic process because it's much easier mathematically, and partly because that's the kind of thing we'll be interested in going to Markov chains with a countable number of states. You're looking at an integer value of t , z of t which is the age of time t is how long it's been since the last renewal occurred. So the age at time t is t minus s of 2 , in this case, and in general it's t minus s sub n . So that's what the age is. The duration \tilde{x} of t , is going to be the interval from this last renewal up to the next renewal after t . So, I've written this out here. \tilde{x} of t is the renewal time for the n of t -th renewal. n of t is this value here. So what we're doing is we're looking at

how long it takes to get from here to here. If I tell you that this interval starts here, this distribution here, we'll have the distribution of x of t . If all I tell you is we're looking around t for the previous interval in the next interval, then it's something different. How do we make sense out of this? Well, this is the picture that will make sense out of it for you, so I hope this makes sense.

If you look at an integer value of t , the z of t this age is going to be some integer greater than or equal to 0. Is it possible for the age to be 0? Yes, of course it is, because t is some integer value. An arrival could have just come in at time t , and then the age is 0. The time from 1 renewal until the next has to be at least 1, because you only get renewals in at integer times, and you can have two renewals at the same time. So x tilde of t has to be bigger than 1.

How do we express this in a different way? Well, let's let q sub j be the probability this is an arrival at time j . If you want to write that down with an equation, it's the sum over all n greater than or equal to 1 of the probability that the n -th arrival occurs at time j . In other words, q sub j is the probability that the first arrival occurs at time j , plus the probability that the second arrival occurs at time j , plus the probability a third arrival occurs at time j . Those are all disjoint events, you can't have two of them be the same. So this q sub j is just the probability that there is an arrival at time j . What's that a function of? It's a function of the arrivals that occur before time j . And it's independent. That's how long the next arrival takes. If I tell you, yes, there was an arrival here conditional on the fact that there was an arrival here, which arrival is it? I'm not talking about t or anything else, I'm just saying, suppose we know there's an arrival here. What you would look at would be the previous arrivals, the previous inter-renewal intervals, and in terms of that you would sort out what that probability is.

Then if there's an arrival here, what's the probability that x tilde of t is equal to this particular value here? Well, now this is where the argument gets tricky. q sub j is a probability of arrival of time j . What I maintain is the joint probability that z of t is equal to i and this duration here is equal to k . In other words, this is equal to i here, this is equal to k . The probability of that this q of t minus i , in other words, the

probability that this is an arrival at time t minus i , and the probability that this inter-renewal interval here has duration k where the restriction is that k has to be bigger than i . And that's what's fishy about this. But it's perfectly correct, just so long as I stick to values of i and k , where k is bigger than i . i is the age here and x hat is t , x tilde of t is the duration. And I can rewrite that joint probability as a probability that we get an arrival here, and that the probability the next arrival takes this time x tilde of t .

Now, what we've done with this is to replace the idea of looking at duration with the idea of looking at an inter-renewal interval. In other words, this probability here, this is not the pmf of x tilde, this is the pmf of x itself. This is the thing we hope we understand at this point. This is what you learned about on the first day of probability theory, when you started taking 6041 or whatever you took. You started learning about random variables, and if they were discrete random variables they had pmf's, and if you had a sequence of IID random variables, this was the pmf you had. So this joint probability here is really these very simple things that you already understand. But it's only equal to this for i less than or equal to t , 0 less than i , less or equal to t , and k greater than i . This restriction here that how far back you go to the last arrival can't be any more than t , it's really sort of a technical restriction. I mean, you need it to be accurate, but it's not very important. The important thing is that k has to be bigger than i because otherwise you get this arrival here and it's not beyond t and therefore it's not the interval that surrounds t , it's some other interval. So, that tells you what you need to know about this.

So, the joint probability is z of t , and x tilde of t is equal to i and k , is this conventional probability here. So what we know, there's a q sub i as the probability of an arrival of j that's equal to the expected value of an arrival at j and is equal to the expected number of arrivals that have occurred. Excuse me, that i there should be a j and this i should be a j . Oh wait, this is q sub i . This should be an i and this should be an i . If I'm looking at it on my computer and I don't have my glasses cleaned I can't tell the difference between them. q sub i is the probability of an arrival at i . The expectation of the number of arrivals at i is equal to the expected

number of arrivals at i minus expected number of arrivals at i minus 1.

So Blackwell says, what i is asymptotically when i gets large. He says that the limit as t goes to infinity. I think this thing's a little weak. In fact, it's terribly weak. So the thing that Blackwell says is that $q_{sub\ i}$ goes to a constant when i gets very large. Namely he says that the probability that you get an arrival at time i when i is very large is just some constant. It doesn't depend on i anymore. It wobbles around for a while, and then it becomes constant. This limit here is then 1 over \bar{x} , which is what $q_{sub\ i}$ is, times the probability of k .

Now, that's very weird because what it says is that this probability does not depend on the age at all. It's just a function of the duration. As a function of the duration, it's just $p_{sub\ x}$ of k divided by \bar{x} . If I go back to what we were looking at before, what this says is if I take this interval here and shift it around, those pairs of points will have exactly the same probability asymptotically.

Now, if you remember what we talked about this very vague idea called random incidents, if you look at a sample path and then you throw a dart at the sample path and you say, what's the duration beyond t , what's the duration before t , this is doing the same thing. But it's doing it in a very exact and precise way. So, this is really the mathematical way to look at this random instance idea. And what it's saying, if we go back to where we were, that's telling us the joint probability of age and duration is just a function of the inter-renewal interval. It doesn't depend on what i is at all. It doesn't depend on how long it's been since the last renewal, it only depends on the size of the inter-renewal interval.

So then we say, OK why don't we try to find the pmf of age? How do we do that? Well, we have this joint distribution of z and x with these constraints on it. k has to be bigger than i to make sure the inter-renewal interval covers t . And you look at that formula there, and you say, OK i travels from 0 up to k minus 1 , and k is going to travel from i plus 1 all the way up to infinity. So, if I try to look at what the probability of the age is, it's going to be the sum of k equals i plus 1 up to infinity of the joint probability. Because if I fix what the age is I'm just looking at all possible

durations from $i + 1$ all the way up to infinity.

So, the marginal for z of t is this complimentary distribution function evaluated at i divided by \bar{x} . That's a little hard to visualize. But if we look at the duration, it's a whole lot easier to visualize and see what's going on. If you want to take the marginal of the duration the pmf the duration is equal to k , what is it? We have to average out over z of t , which is the age. The age can be anything from 0 up to $k - 1$ if we're looking at a particular value of k . What that is again is this diagram here, and \hat{x} of t will have a particular value of k here, here, here, here, all the way up to here. So what we're doing is adding up all those values, knowing exactly what the idea of random instance was doing, but doing it in a nice clean way.

Back to the ranch, it says that the limit is t goes to infinity of \tilde{x} is going to be k times pmf of x divided by \bar{x} . You need to divide by \bar{x} so that this is a probability mass function, and we actually had the \bar{x} there all along. But if we sum this up over k , what do we get? We get k times the pmf of x , which is the expected value of x . So, that's all very nice. But what happens if we try to find the expected value of this duration function here? That's really no harder, the expected value of the duration. Here's one of the few places where we use pmf's to do everything. It's a sum from $k = 1$ to infinity of k , which is the k we had before, k times p_x of k over \bar{x} .

So it's k times $k p_x$ of k divided by \bar{x} . That's k^2 times the probability mass function of x divided by \bar{x} , which is the expected value of x^2 divided by the expected value of x . Now, that sounds a little weird also, but I think it's a whole lot less weird than the argument using sample paths. I mean, you can look at this and you can track down every little bit of it, and you can be very sure of what it's saying. You can find the expected age in sort of the same way, and it's done in the text, and it comes out to be expected value of x^2 divided by 2 times the expected value of x , which is exactly what the argument was for if you're looking at the sample path point. But you lose a $1/2$.

AUDIENCE: When you look at this [INAUDIBLE] it was for non-arithmetic distributions. And here

it's for arithmetic. So it's always true that the [INAUDIBLE], even for arithmetic?

PROFESSOR: Yes. The previous argument we went through using sample paths was for arithmetic distribution, as non-arithmetic was for any distribution at all. This argument here, you have to distinguish between arithmetic and non-arithmetic, and you have to distinguish because they act in very different ways. I mean, if you read the text and you read the section on non-arithmetic random variables, you wind up with a lot of very tedious stuff that's going on, a lot of really having to understand what [INAUDIBLE] integration is all about. I mean, if you put in densities it's all fine, but if you try to do it for these weird distributions which are discrete but non-arithmetic, you really have a lot of sweating to do to make sure that any of this makes sense. Which is why I'm doing it this way.

But, when we do it this way we get this extra factor of $1/2$ here. And where does the factor of $1/2$ come from? Well, the factor of $1/2$ comes from the fact that all of this argument was looking at a t , which has an integer value. Because we've made t the integer, the age could be 0. If we make t non-integer, then the age is going to be something between 0 and 1. And in fact, if we look at what's happening, what we're going to find is that the age of your integer values into the 6th, into the 6th plus one. And you know from the homework that this might have to be 10 to the 20th, but it doesn't make any difference. And now the age here can be 0, so the average age is going to be some value here. And as you look at larger and larger t 's, as you go from this integer to this integer, the age is going to increase. And then at this point there might be an arrival at this point. This is the probability of an arrival at this point. Then it goes up again. It goes up exactly the same way. It goes down. And the value it has at the integer is its lowest value, because we're assuming that when you look at the age you're looking at an age which could be 0 if you look at this particular integer value.

So, where's the problem? We take the sample average over time, and what we're doing is we're averaging over time here. So, you wind up with this point, then you crawl up, then you go down again, you crawl up, you go down again. And this average, which is the thing we found before for the sample path average, is exactly

1/2 larger than what we found here. So, miraculously these two numbers jive, which I find amazing after all this work. But they do. And the fact that you have the 1/2 there is sort of a check on the fact that people have done the work right, because nobody would ever imagine that was there until they actually went through it and found it. That, I think, explains why you get these peculiar results for duration and for age. At least, I hope it does.

Let's go on to countable-state Markov chains. And the big change that occurs when you go to countable-state chains is what you mean by a recurring class. Before with finite state Markov chains, we just blandly defined a recurrent state of the state which had the property that wherever you could go from that state, there was always some way to get back. And since you had a finite number of states, if there was some place you could go, you're always going to get back to eventually, because there was always some path that had some probability, and you keep repeating the possibility of doing that. So, you didn't have to worry about decision. Here you do have to worry about it. It's particularly funny if you look at a Markov chain model of a Bernoulli process.

So, a Markov chain model of a Bernoulli process, this is a countable-state process, you start out at state 0. You flip your coin, your loaded coin, which comes up 1, so the probability p and tails with probability q , and if it comes up heads you go to state 1. If it comes up tails, you go to state minus 1. So this state here is really the sum of the x of i 's. In other words, it's the number of successes minus the number of failures. So, we're looking at that sum and we're looking at what happens as time gets larger and larger. So, you go wandering around here, moving up with probability p , moving down the probability q , and you sort of get the idea that this is going to diffuse after a long time. How do you do that mathematically? And the text writes this out carefully. I don't want to do that here, because I think it's important for you in doing exercises and things like that to start to see these arguments automatically.

So, the thing that's going to happen as n gets large is that the variance of this state is going to be n times the variance of a single up or down random variable. You

calculate that as $1 - p - q^2$, and if p and q are both strictly between 0 and 1, that variance is always positive. It can't be negative, you can't have negative variance. But the important thing is it keeps increasing within. So, it says that if you try to draw a diagram, if you try to draw pmf of the probability that s_n is equal to 0, 1, 2, 3, 4, and so forth, what it's going to do is n gets very large. Because this pmf is going to keep scaling outward, and it's going to scale outwards with the square root of n .

And we already know that it starts to look Gaussian. And in particular what I mean by looking Gaussian is that it's going to be a quantized version of the Gaussian, because each time you increase by 1 there's a probability that s_n equals that increased value. So you're spreading out the individual values at each integer, have to spread, have to come down. They can't do anything else. If you're spreading a distribution out and it's an integer distribution, you can't keep these values the same as they were before or you would have a total probability, which would be growing with the square root of n . And the probability that s_n is something is equal to 1. So as you spread out you have this Gaussian distribution where the variance is growing within and where the probability of each individual value is going down as 1 over the square root of the n .

Now, this is the sort of argument which I hope can become automatic for you, because this is the kind of thing you see in problems and it's a big mess to analyze it. You have to go through a lot of work, you have to be very careful about how you're scaling things. And if you just look at this saying, what's going to happen here as n gets large is that this going to be like a Gaussian distribution, like a quantized Gaussian distribution, and if it goes out this way, it's got to go down. You can't go out without going down. So, it says that no matter what p is and no matter what q is, so long as neither of them are 0, this thing is going to spread out, and the probability that you're in any particular state after a long time is going to 0.

Now, that's not like the behavior of finite state Markov chains. Because in the finite state Markov chains, you have recurring classes, you have steady state probabilities for those recurring classes. Here you don't have any steady state probability

distribution for this situation. Because in a steady state every state has probability 0, and those probabilities don't add to 1, so you just can't deal with this in any sensible way. So, here we have a countable-state Markov chain, which doesn't behave at all like the finite state Markov chains we've looked at before.

What's the period of this chain, by the way? I said, here it's equal to 2, why is it equal to 2? Well, you start out at an even number, s_n is equal to 0, after one transition s_n is odd, after two transitions it's even again, so it keeps oscillating between even and odd, which means there's a period of two. So, we have a situation where all states communicate. The definition of communicate is the same as it was before. There's a path they get from here to there, and there's a path to get back again. And classes are the same as they were before, a set of states which all communicate with each other, are all in the same class, and if i and j communicate, then if j and k communicate, then there's a path to get from i to j and a path to get from j to k , to the path to get from i to k , to the path to get back again. So if i and j communicate and j and k communicate, then i and k communicate also, which is why you get classes out of this.

So, we have classes. What we don't have is anything relating to steady state probabilities, necessarily. So what we do? Well, another example, that's called a birth-death chain, we will see this much more often than the ugly thing I had in the last slide. The ugly thing I had in the last slide is really much simpler, but it's harder to analyze. It's gruesome.

Well, as a matter of fact, we probably ought to talk about this a little further. Because this does something. If you pick p equal to $1/2$, this thing is going to expand the center of it for any n is going to stay at 0, so the probability of each state is going to get smaller and smaller. What's the probability when you start at 0 that you ever get back to 0 again? Now, that's not an easy question, it's a question that takes a good deal of analysis and a good deal of head scratching. But the answer is 1. There's a 0 steady state probability of being in state 0, but you wander away and you eventually get back. There's always a path to come back, and no matter how far away you get it's just as easy to get back as it was to get out there, so it's

certainly plausible that you ought to get back with probability 1. And we will probably prove that at some point. I'm not going to prove it today because it would just be too much to prove for today.

And, also, you will need to have done a few of the exercises. The homework set for this week is kind of easy, because you're probably all exhausted after studying for the quiz. I hope the quiz was easy, but even if it was easy you still had to study for it, and that's the thing that takes the time. So this week we won't do too much, but you will get some experience working with a set of ideas.

We want to go to this next thing, which is called a birth-death chain. There are only non-negative states here, 0, 1, 2. It looks exactly the same as the previous chain, except you can't go negative. Whenever you go down to 0 you bounce around and you then can go up again and come back again as sort of like an accordion. If p is less than $1/2$, you can sort of imagine what's going to happen here. Every time you go up there's a force pulling you back that's bigger than the force going up, so you can imagine that you're going to stay clustered pretty close to 0. If p is greater than $1/2$, you're going to go off into the wild blue yonder, and you're never going to come back eventually, but if p is equal to $1/2$ then it's kind of hard to see what's going to happen again. It's the same as the situation before. You will come back with probability 1 if p is $1/2$, but you won't have any steady state probability, which is a strange case, which we'll talk about as we go on.

You looked at the truncated case of this in the homework, namely you looked at a case of what happened if you just truncated this chain of some value n and you found the steady state probabilities. And what you found was that if p is equal to $1/2$, the steady state probabilities are uniform. If p is greater than $1/2$, the states over in this end have high probabilities, they travel down geometrically to the states at this end. If p is less than $1/2$, the states are highly probable and these state go down geometrically as you go out here. What do you think happens if you take this truncated chain and then start moving n further and further out? Well, if you have the nice case where the probabilities are mostly clustered around 0, then as you keep moving it out it doesn't make any difference.

In fact, you look at the answer you've got and see what happens in the limit as you take into infinity. But, if you look at the case where p is greater than $1/2$, then everything is clustered up at the right n and every time you increase the number of states by 1, blah. Everything goes to hell. Everything goes to hell unless you realize that you just move everything up by 1, and then you have the same case that you had before. So, you just keep moving things up. But there isn't any steady state, and as advertised things go to infinity. If you look at the case where p is equal to $1/2$, as you increase the number of states what happens is all the states get less and less likely. You keep wandering around in an aimless fashion, and nothing very interesting happens.

What we want to do, just to be able to define recurrence, to mean that given that you start off in some state i , there's a future return to state i with probability 1. That's what recurrence should mean, that's what recurrence means in English. Recurrence means you come back. Since it's probabilistic, you have to say, well we don't know when we're going to get back, but we are going to get back eventually. It's what you say to a friend you don't want to see. I'll see you sometime. And then it might be an infinite expected time. But at least you said you're going to make it back. We will see the birth-death chain above is recurrent in this sense if p is less than $1/2$, and it's not recurrent if p is greater than $1/2$, and we're clearly going to have to struggle a little bit to find out what it is that p is equal to $1/2$. And that's a strange case and we'll call non-recurrent when we get to that point.

We're going to use renewal theory to study these recurrent chains, which is why we did renewal theory first. But first we have to understand first passage times a little better. We looked at first passage times a little bit when we were dealing with Markov chains. We looked at first passage time by looking at the expected first passage time to get from one state to another state, and we found a nice clean way of doing that. We'll see how that relates to this, but here instead of looking at the expected value we want to find the probability that you have a return after some particular period of time.

AUDIENCE: So why did [INAUDIBLE]?

PROFESSOR:

If p is greater than $1/2$, what's going to happen is you keep wandering further and further off to the right. You can come back. There's a certain probability that no matter how far you get out there, there's a probability that you can get back. But there's also a probability that you won't get back, that you keep getting bigger and bigger. And this is not obvious. But it's something that we're going to sort out as we go on. But it certainly is plausible that there's a probability, a positive probability that you'll never return. Because the further you go, the harder it is to get back. And the drift is always to the right.

And since the drift is always to the right, the further you get away, the less probable it gets that you ever get back. And we will analyze this. I mean, for this case, it's not very hard to analyze. Where were we? OK. The first pass each time probability, we're going to call it f_{ij} of n . This is the probability given that you're in state i at time zero.

It's the probability that you reach state j for the first time at time n . It's not the probability that you're in state j at time n , which is what we call p_{ij} to the n . It's just the probability that you get there for the first time. At time n . If you remember, quick comment about notation. We called p_{ij} is the probability that x_j, x_n equals j , given that x_0 equals i .

And f_{ij} . And now the n is in parentheses as the probability that x_n is equal to j , given that x_0 equals i , and x_1, x_2, \dots so forth are unequal to j . Now you see why I get confused with i 's and j 's.

The reason I'm using parentheses here and using a subscript here is when we're dealing with finite state Markov chains, it was just so convenient to view this as the ij component of the matrix p taken to the n -th power. And this is to remind you that this is a matrix taken to the n -th power. And then you take the ij element. There isn't any matrix multiplication here.

This is partly because we're dealing with countable state and Markov chains here. But partly also because this is an uglier thing that we're dealing with. But you can

still work this out. What's the probability that you will be in state j for the first time at n , given you're back in i ? It's really the probability that the sum of the probability that on the first transition, you move from i to k .

Some k unequal to j , because if we're equal to j , we'd already be there, and we wouldn't be looking at anything after that. So it's the probability we move from i to k . And then we have n minus 1 transitions to get from k back to j for the first time. If you think you understand this, most things like this, you can write them either way.

You can look at the first transition followed by n minus 1 transitions after it, or you can look at the n minus 1 transitions first, followed by the last transition. Here you can't do that. And you should look at it, and figure out why. Yes?

AUDIENCE: I'm sorry. I just was wondering if it's the same to say that all of the [INAUDIBLE] are given, like you would in [INAUDIBLE]? And [INAUDIBLE] all of them given just [INAUDIBLE]?

PROFESSOR: It's the probability that all of the x_1, x_2, x_3 and so forth are not equal to j .

AUDIENCE: Because here and there, you did [INAUDIBLE] exactly the same question. In here, [INAUDIBLE] given [INAUDIBLE]. And they're [INAUDIBLE] the same?

PROFESSOR: This is the same as that, yes. Except it's not totally obvious why it is, and I'm trying to explain why it is. I mean, you look at the first transition you take. You start out in state i . The next place you go, if it's state j , you're all through. And then you've gotten to state j in one step, and that's the end of it. But if you don't get to state j in one step, you get to some other state k .

And then the question you ask is what's the probability starting from state k that you will reach j for the first time in n minus 1 steps? If you try to match up these two equations, it's not at all obvious. If you look at what this first transition probability means, then it's easy to get this from that. And it's easy to get this from that. But anyway, that's what it is.

$f_{ij} > 1$ is equal to p_{ij} , and therefore this is equal to n greater than one. And you

can use this recursion, and you can go through and calculate all of these n-th order probabilities, as some [INAUDIBLE] is going to point out in just a minute. It takes a while to do that for an infinite number of states. But we assume there's some nice formula for doing it.

I mean, this is a formula you could in principal, compute it. OK, so here's the same formula again. I want to relate that to the probabilities of being in state j at time n , given the you were in state i at time 0 . Which by Chapman and Kolmogorov is this equation here. That's the same. You go first to state k , and then from kj , n minus one steps.

Here you're summing over all k , and you're summing over all k because in fact, if you get to state j in the first step, you can still get to state j again after n minus one more steps. Here you're only interested in the first time you get to state j . Here you're interested in any time you get to state j .

I bring this up because remember what you did when you solve that problem related to rewards with a Markov chain? If you can think back to that. The way to solve the reward problem, trying to find the first pass each time from sum i to sum j was to just take all the outputs from state j and remove them.

If you look at this formula, you'll see that that's exactly what we've done mathematically by summing only over k unequal to j . Every time we get to j , we terminate the whole thing, and we don't proceed any further. So this is just a mathematical way of saying just a Markov chain by ripping all of the outputs out of state j , and putting a self loop in state j .

OK, so this is really saying the same sort of thing that that was saying, except that was only giving us expected value, and this is giving us the whole thing. Now, the next thing is we would like to find what looks like a distribution function for the same thing.

This is the probability of reaching j by time n or before. And the probability that you reach state j by time n or before it's just the sum of the probabilities that you reach

state j for the first time at some time m less than or equal to n . It's a probability of reaching j by time n or before. If this limit now is equal to 1, it means that with probability 1, you eventually get there. Yes?

AUDIENCE:

I'm sorry, Professor Gallager. I'm [INAUDIBLE] definition of f_{ij} of n . So is it the thing you wrote on the board, or is it the thing in the notes? Like, I don't see why they're the same. Because the thing in the notes, you're only given that x_0 is equal to i . But the thing on the board, you're given that x_0 is equal to i , and you're given that x_1 , x_2 , none of them are equal to j .

PROFESSOR:

Oh, I'm sorry. This is-- you are absolutely right. And then given x_0 equals i . Does that make sense now? I can't see in my mind if that's equal to this. But if I sit there quietly and look at it for five minutes, I realize why this is equal to that. And that's why I write it down wrong half the time.

So, if this limit is equal to 1, it means I can define a random variable, which is the amount of time that it takes to get to state j for the first time. And that random variable is a non defective random variable, because I always get to state j eventually, starting from state i .

Now, what was awkward about this was the fact that I had to go through all these probabilities before I could say let t_{ij} be a random variable, and let that random variable be the number of steps that it takes to get to state j , starting in state i . And I couldn't do that because it wasn't clear there was a random variable.

If I said it was as effective random variable, then it would be clear that there had to be some sort of defective random variable. But then I'd have to deal with the question of what do I really mean by defective random variable? Incidentally, the notes does not define what a defective random variable is.

If you have a non-negative thing that might be a random variable, the definition of a defective random variable is that all these probabilities exist, but the limit is not equal to 1. In other words, sometimes the thing never happens. So that you can either look at this as you have a thing like a random variable, but it matched a lot of

sample points into infinity.

Or you can view it as it maps a lot of sample points into nothing. But you still have a distribution function for it. OK. But anyway, now we can talk about a random variable, which is the time to get from state i to state j , if in fact, it's certain that we're going to get there.

Now, if you start out with a definition of this distribution function here, and you play with this formula here, you play with that formula, and that formula a little bit, you can rewrite the formula for this distribution function like thing in the following way. This is only different from the thing we wrote before by the presence of p_{ij} here. Otherwise, f_{ij} of n is equal to just a sum over here without that.

With this, you keep adding up, and it keeps getting bigger. Why I want to talk about this, it's sort of a detail. But this equation is always satisfied by these distribution function like things. But these equations do not necessarily solve for these quantities. How do I see that? Well, if I plug one in for x_{ij} of n , and f_{ij} of n minus 1 for all i and all j , what do I get?

I get one that's equal to p_{ij} plus the sum over k unequal to j of p_{ik} times 1. So I get 1 equals 1. So a solution to this equation is that all of the f_{ij} 's are equal to 1. Is this disturbing? Well, no, it shouldn't be, because all the time we can write equations for things, and the equations don't have a unique solution. And these equations don't have a unique solution.

We never said they did. But there is a theorem in the notes, which says that if you look at all the solutions to this equation, and you take the smallest solution, that the smallest solution is the right one. In other words, the smallest solution is the solution you get from doing it this other way. I mean, this solution always works, because you can always solve for these quantities.

And you don't have to-- I mean, you're just using iteration, so all these quantities exist. OK. Now finally, these equations for going from state i to state j also work for going from state j to state j . If the probability of going from state j to state j

eventually is equal to 1, that's what we said recurrent ought to mean. And now we have a precise way of saying it.

If f_{jj} to infinity is equal to 1, an eventual return from state j occurs with probability 1, and the sequence of returns is a sequence of renewal epochs in a renewal process. Nice, huh? I mean, when we looked at finite state in Markov chains, we just sort of said this and we're done with it. Because with finite state and Markov chains, what else can happen? You start in a particular state.

If it's a recurrent state, you keep hitting that state, and you keep coming back. And then you hit it again. And the amount of time from one hit to the next hit is independent, as the next hit to the next yet hit. It was clear that you had a renewal process there. Here it's still clear that you have a renewal process, if you can define this random variable. Yes?

AUDIENCE: When you say the smallest set you mean sum across all terms of the set, and whichever gives you the smallest.

PROFESSOR: No, I mean each of the values being as small as possible. I mean, it turns out that the solutions are monotonic in that sense. You can find some solutions where these are big and these are small, and others where these are little, and these are big. OK. So now we know what this distribution function is. We know what a recurrent state is. And what do we do with it?

Well, we say there's a random variable with this distribution function. We keep doing things like this. We keep getting more and more abstract. I mean, instead of saying here's a random variable and here's what its distribution function is, we say if this was a random variable, then state j is recurrent. It has this distribution function.

The renewal process of returns to j , then, has inter renewal intervals with this distribution function. As soon as we have this renewal process, we can state this lemma, which are things that we proved when we were talking about renewal theory. And let me try to explain why each of them is true. Let's start out by assuming that state j is recurrent.

In other words, you have this random variable, which is the amount of time it takes to get back to j . If you get back to j with probability one, then ask the question, how long does it take to get back to j for the second time? Well, you have a random variable, which is the time that it takes to get from j to j for the first time.

You add this to a random variable, which is the amount of time to get from j to j the second time. You add two random variables together, and you get a random variable. In other words, the second return is sure if the first one is. The second return occurs with probability 1 if the first return occurs with probability-- I mean, you have a very long wait for the first return.

But after that very long wait, you just have a very long wait again. But it's going to happen eventually. And the third return happens eventually, and the fourth return happens eventually. So this says that as t goes to infinity, the number of returns up to time t is going to be infinite. Very small rate, perhaps.

If I look at the expected value of the number of returns up until time t , that's going to be equal to infinity, also. I can't think of any easy way of arguing that. If I look at the sum over all n of the probability that I'm in this state j at time n , and I add up all those probabilities, what do I get? When I add up all those probabilities, I'm adding up the expectations of having a renewal at time n .

By adding that up for all n . And this, in fact, is exactly the same thing as this. So if you believe that, you have to believe this also. And then we just go back and repeat the thing if these things are not random variables. Yes?

AUDIENCE: I don't understand what you mean [INAUDIBLE]? I thought that we put [INAUDIBLE] 2 and 3? [INAUDIBLE]?

PROFESSOR: And given 3, we proof 4.

AUDIENCE: Oh, [INAUDIBLE].

PROFESSOR: Yeah. And we don't have time to prove that--

AUDIENCE: [INAUDIBLE].

PROFESSOR: Yeah. OK. But in fact, it's not hard to say suppose one doesn't occur, then, should the other stuff occur. But none of these imply that the expected value of t_{jj} is finite, or infinite. You can always have random variables, which are random variables, namely this is an integer value random variable. It always takes on some integer value.

But the expected value of that value might be infinite. I mean, you've seen all sorts of random variables like that. You have a random variable where the probability of j is some constant divided by j squared. Now, you multiply j by $1/2$ squared, you sum it over j , and you've got infinity. OK, so you might have these random variables, which have an expected return time, which is infinite.

That's exactly what happens when you look at these back right at the beginning of today. 1, 2, 4. I'm going to look at this kind of chain here, and I set p equal to one half. What we said was that you just disperse here the probability that you're going to be in any state after a long time goes to zero, but you have to get back eventually.

And if you look at that condition carefully, and you put it together with all the things we found out about Markov chains, you realize that the expected time to get back has to be infinite for this case. And the same for the next example we looked at. If you look at p equals $1/2$, you don't have a steady state probability.

Because you don't have a steady state probability, the expected time to get back, the expected recurrence time has an infinite expected value. The expected recurrence time has to be 1 over the probability of the state. So the probability of the state is 0 . The expected return time has to be infinite. So, where were we?

Well, two states are in the same class if they communicate. Same as for finite state chains. That's the same argument we gave, and you get from there to there, and there to there. And you get from here to there, and here to here, then you get from here to there, and here to there. OK. If states i and j are in the same class, then

either both are recurrent, or both are transient.

Which means not recurrent. If j is recurrent, then we've already found that this sum is equal to infinity. And then, oh, I have to explain this a little bit. What I'm trying to show you is that if j is recurrent, then i has to be recurrent. And I know that j is recurrent if this sum is equal to infinity. And if this sum is equal to infinity, let's look at how we can get from i back to i in n steps.

Since i and j communicate, there's some path of some number of steps, say m , which guesses from i to j . There's also some path of say, length l which gets us from j back to k . And if there's this path, and there's this path, and then I sum this over all k , that I get is a lower bound to that. I first go to j , and then I spend an infinite number of states constantly coming back to j .

And then I finally go back to i . And what that says is that this sum of p_{ii} , i to the n , I mean, this is just one set of paths which gets us from i to i in n steps. If state j is recurrent, then t_{jj} might or might not have a finite expectation.

All I want to say here is that if t_{jj} , the time to get from state j back to state j has an infinite expectation, then you call state j . No recurrent instead of regular recurrent.

As we were just saying, if that expected recurrence time is infinite, then the steady state probability is going to be zero, which says that if something is no recurrent, it's going to be even a very, very different way from when it's positive recurrent, which is when the return time has a finite value.

AUDIENCE: [INAUDIBLE]?

PROFESSOR: Yes. Which means there isn't a steady state probability. To have a steady state probability, you want them all to add up to 1. So yes, they're all zero. And formally, there isn't a steady state probability. OK, thank you. We will--