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

PROFESSOR: OK. I guess we might as well get started. We talked a little bit about Markov processes last week. I want to review a little bit of what we did then, and then move on pretty quickly. This is a rather strange chapter, because it's full of notation. When I first started reviewing it myself after not having looked at it for about a year, I was horrified by the amount of notation.

And then I realized, what we're doing in this chapter is putting together all the stuff we've learned from Poisson processes. There are Poisson processes all through this Markov chains renewals. And all three of these, with all their notation, are all sitting here. And then we get into new things, also, so I apologize for the notation. I'm going to try to keep it down to the minimal amount this time.

And see if we can sort of get to the bottom line as easily as we can, so that when you do the exercises and read the notes, you can go back in fill in some of the things that are missing. OK. So accountable state Markov process. The easiest way to define it, and the way we're defining it is as an extension of accountable state Markov chain.

So you start out with accountable state Markov chain, and then along with each step, say from state x sub n to x sub n plus 1, there is an exponential holding time, u sub n plus 1. We said that it was a little strange associating the holding time of the final state rather than the initial state. But you almost had to do that, because of many things that would get very confusing otherwise.

So we're just doing that. We start out in state x0, which is usually given. If it's not given, it can be random. And then we go to state x1 after a waiting time, u sub 1. Go from x1 to x2 with a waiting time u sub 2. The waiting time u sub i is a function of the

1

state we're coming from, in the sense that it's an exponential random variable who's rate is given by the state we're coming from.

So this diagram here gives you, really, what the dependence of this set of random variables is. You have a sequence of random variables here. A sequence of random variables here. Each random variable here conditional on the initial state that it's coming from, is independent of everything else. And that's what this dependence diagram means.

It's a generalization of what we talked about before. When we talked about Markov chains before, we just had a string of states. And each state x sub n is dependent only on the prior state x sub n minus 1. And given the prior state x sub n minus 1, it's statistically independent of all states before that. Here we have this more general situation of a tree.

I thought I'd better illustrate this a little more. If you have a directed tree, the dependencies, each random variable conditional as parent is statistically independent of all earlier random variables. In other words, this random variable here conditional on this random variable is statistically independent of this, this, this, and this.

And so you can move forward in this way, defining each random variable as being statistically independent of everything in the past, conditional on just one previous random variable. As an example of this, if you look at the probability of x0, x1, x2, and u2, these five random variables here. Probability of x0, probability of x1 given x0. Probability of x2 given x1.

Probability of u2 given x1. You can write that out in that way. From this, you can rewrite this in any way you want to. You take these two equations, and you can rewrite them as a probability of x1 times the probability of x0 given x1, times the probability of x2 given x1, times the probability u2 given x1.

In other words, what's happening here is if you condition everything on x1, this random variable here, this stuff is statistically independent of this. Is statistically

independent of all of this. Given any one node in this three, given the value of that node, everything on every set of branches coming out from it is statistically independent. This is a remarkably useful property.

This is the Markov property in general. I mean, Markov chains, we only use the fact that it's on a chain. In general you use the fact that it's on a tree. And all of this stuff can be used in remarkable ways. I didn't know this until probably five years ago. And suddenly when I realized it, I think because somebody was pointing it out in a research paper they were writing.

Suddenly all sorts of things became much, much easier. Because everything like the fact we pointed out before, that the past is independent of the future, given the present, that's one example of this. But this is far more general than that. It says that anything on this tree, if you can start at any point on the tree, and everything going out from there is statistically independent, given this node on the tree.

So that's a valuable thing. And so you get your condition on any node, breaks a tree into independent sub trees. You can then go on from there and break it down further and further and further, until you get out to the leaves. So this independence property is really the general thing that we refer to when we say a set of random variables are Markov. OK.

The evolution in time with a Markov process, this diagram I find very helpful to see what's going on in a Markov process. You have a set of states. Initially, you're in a state x0, the state at time 0 is some given value i. This is a sample path here. The next state we'll say is j, the next state is a k. When you're in state i, there's some holding time, which has rate new 1, new sub i.

It's an exponential random variable which tells you how long it takes until this transition. This transition occurs at time s1, which is u1 is equal to s1. The next transition is at s2. Equals u1 plus u2. The next transition is at x3, which is u1 plus u2 plus u3. Now you start to see why we've numbered these holding times the way we have, so we can talk about the times that each of these transitions take place.

We usually assume that the embedded Markov chain for a Markov process, remember, the embedded Markov chain now is just the Markov chain itself without these holding times on it. We assume it has no self transitions because if you're sitting in a state x of t equals i, and suddenly there's a transition back to i again, and you look at the process in terms of x of t, t greater than or equal to 0.

The state that you're in at each time t, what happens? You don't see it. It's a totally invisible transition, because you're sitting in state i. You suddenly have a transition back to i that takes 0 time. So you stay in state i. You can put that transition in or you can take it out. It doesn't make any difference. It won't affect the process at all. OK.

Aside from that issue of these self transitions, a sample path of both x sub n, each of these x sub 0 equals i, x sub 1 equals j. x sub 2 equals k. A sample path as those plus the holding times specifies what x of t is at each instant of time. And if you know what x of t is at each unit of time, that tells you when the transitions are occurring.

When you know when the transitions are occurring, you know what the these u's are. And when you see what the transition is into, you know what the state is. So the description of a Markov process in terms of the process, what we call a process x sub t for all t greater than or equal to 0, and the set of random variables, the embedded Markov chain, and the holding times are both equivalent to each other.

This shouldn't be any surprise to you by now, because every process we've talked to, we've talked about. We've described in the same way. We described the Poisson process in multiple ways. We described Markov chains in multiple ways. We described renewal processes in multiple ways. And this is just another example of that.

You use whatever description you want to after you've shown they're all equivalent. So there's really nothing new here. Or is there? Who can see what there is about this relationship, which is different from what we've just been talking about? It's using one extra property. This is not just a consequence of this, it also uses something else. And what else does it use?

What I'm doing is saying x of t at this instance of time here is dependent on given x of t is some previous time here. The state here, given the state here is independent of everything in the past. So what else am I using there? I'm using the memory looseness of the Poisson process. I'm using the memory looseness of the exponential random variable.

If I'm given the state here, and I'm conditioning on the state here, this is an exponential random variable in here. The time the next transition is exponential given this time here. And it doesn't matter when the previous transition took place. So if I'm given the state at this time here, the time to the next transition is an exponential random variable, the same distribution as u2.

So what this says is I'm using the initial description in terms of an embedded Markov chain plus holding times, and I'm adding to that the fact that the holding times are exponential, and therefore they're memory less. OK, that clear to everybody? It's vitally important for all of this.

Because it's hard to do anything with Markov processes without realizing explicitly that you're using the fact that these random variables are memory less. At the end of this chapter, there's something called semi Markov processes. The description is that semi Markov processes are exactly the same.

Semi Markov chains are exactly the same as markup processes except, these holding times are not exponential. They can be anything. And as soon as the holding times can be anything, the process gets so complicated that you hardly want to talk about it anymore.

So the fact that we have these exponential holding times is really important in terms of getting this condition here, which lets you talk directly about the process, instead of the embedded Markov chain. You're going to represent a Markov process by a graph for the embedded Markov chain, and then you give the rates on top of the nodes.

So if you're in state 0, the holding time until you enter the next state is given as

some, the rate of that exponential is given as new 0. The rate here is given as new 1, so forth. Ultimately, we're usually interested in this process, x of t, t greater than or equal to 0, which is the Markov process itself. x of t is equal to xn for t in sub n, between sub n and s sub n plus 1.

What does that mean? Well, it means at this point, we're taking the Markov process as the fundamental thing, and we're describing it in terms of the nth state transition. But we know that the nth state transition takes place at time s sub n, namely it takes place at the sum of all of these exponential holding times up until that point. And that state stays there until the next exponential holding time.

So this really gives you the linkage between the Markov process in this expression, and the markup process in terms of this graphical expression here with the embedded chain, and the exponential holding times. OK. You can visualize a transition from one state to another in tree very convenient ways.

And these are ways that we've, I hope we've really learned to think about from looking at Poisson processes. You can visualize this transition by first using the next state by these transition probabilities in the embedded Markov chain. And then you choose a transition time, which is exponential with rate new sub i.

Equivalently, because it's a Poisson process, you can choose the-- well, no, because these are independent given the state. You can choose the transition time first, and then you can choose the state, because these are independent of each other conditional on the state that you're in.

And finally, equivalently, which is where the Poisson process comes in, a really neat way to think of Poisson processes is to have an enormously large number of Poisson processes running all the time. There's one Poisson process for every transition in this Markov chain. So you have accountably infinite number of Poisson processes, which sounds a little complicated at first.

But you visualize a Poisson process for each state pair i to j, which has a rate q sub ij, which is the rate at which transitions occur out of state i times p sub ij. This is the

rate which, when you're in state i, you will go to state j.

And this makes use of all this stuff about splitting and combining Poisson processes. If you have a Poisson process which has rate new sub i and you split it into a number of Poisson processes, for each next state you might go to, you're splitting it into Poisson processes of rate new sub i times p sub ij. And what's happening there is there's a little switch.

The little switch has probabilities p sub ij, and that switch is telling you which state to go to next. All of this is totally artificial. And I hope by this time, you are comfortable about looking at physical things in a totally artificial way, because that's the magic of mathematics. If you didn't have mathematics, you couldn't look at real things in artificial ways.

And all the science would suddenly disappear. So what we're doing here is defining this Markov process in this artificial way of all of these little Poisson processes, and we now know how they all work. When the entry to state i, the next state is the j with an x Poisson arrival, according to q sub ij. So all these Poisson processes are waiting to have an arrival come out.

To have a race, one of them wins, and you go off to that state. OK, question. What's the conditional distribution of u1 given that x0 i, and x1 equals j? And to imagine this, suppose that there are only two places you can go from state 0. You can go into state 1 with some very large probability, say 0.999. Or you can go into state 2 with some very, very small probability.

And what that means is this exponential random variable going from state 0 into state 2 is a very, very slow, random variable. It has a very small rate. And the exponential random variable going into state 1 is a very, very large random variable. So you would think that if you go from state 0 the state x2, it means it must take a very long time to get there. Well, that's absolutely wrong.

The time that it takes to go from x0 to x2 is this random variable, u sub i. Where u sub i is the state you happen to be in at this point when you're in-- x0 is the state

that we start in. x0 is a random variable. It has some value i. With this value i, there's an exponential random variable that determines how long it takes you to get to the next state.

This random variable conditional on x0 equals i is independent of which state you happen to go to. And what that means is that the conditional distribution of u1, given x sub 0 is equal to i, and x sub 1 equals j. If you've had your ears at all open for the last 10 minutes, it is exponential with rate i. With rate new sub i.

These holding times and the next states you go to are independent of each other, conditional on where you happen to be. It's the same thing we saw back in Poisson processes. It was confusing as hell back then. It is still confusing as hell.

If you didn't get it sorted out in your mind then, go back and think about it again now, and try to get your common sense, which tells you when you go to state 2, it must take a long time to get there, because that exponential random variable has a very long holding time. And that just isn't true. And it wasn't true when we were dealing with a Poisson process, which got split either.

These two things are independent of each other. Intuitively, why that is, I almost hesitate to try to say why it is, because it's such a tricky statement. If you happen to go to state 2 instead of to state 1, what's happening is that all of this time that you're waiting to have a state transition, when you finally have this state transition, you then flip a switch to see which state you're going to go to.

And the fact that it's taken you a long time to get there says nothing whatsoever about what this switch is doing, because that switch is independent of how long it takes you for the switch to operate. And I know. It's not entirely intuitive, and you just have to beat yourself on the head until it becomes intuitive. I've beaten myself on the head until I can hardly think straight.

It still isn't intuitive to me, but maybe it will become intuitive to you. I hope so. So anyway, this gives the conditional distribution of u1 given that x0 is equal to i, and x1 is equal to-- no. This says that the exponential rate out of state i is equal to the sum

of the exponential rates to each of the states we might be going to. We have to go to some other state.

We have no self transitions as far as we're concerned here. Even if we had self transitions, this formula would still be correct, but it's easier to think of it without self transitions. p sub ij, this is the switch probability. It's q sub ij divided by new sub i. This is the probability you're going to go to j given that you were in state i.

The matrix of all of these cues specifies the matrix of all of these p's, and it specifies new. That's what this formula says. If I know what q is, I know what new is, I know what p is. And we've already said that if you know what p is and you know what new is, you know what q sub ij is. So these are completely equivalent representations of the same thing.

You can work with either one you want to. Sometimes one is useful, sometimes the other is useful. If you look at an mm1q, mm1q, you remember, is exponential arrivals. Exponential service time. The time of the service is independent of when it happens, or who it happens to. These service times are just individual exponential random variables of some rate mu.

The arrivals are the inter-arrival intervals are exponential random variables of rate lambda. So when you're sitting in state 0, the only place you can is to state one. You're sitting there, and if you're in state 0, the server is idle. You're waiting for the first arrival to occur. The next thing that happens has to be an arrival because it can't be a service.

So the transition here is with probability 1. All these other transitions are with probability lambda divided by lambda plus mu, because in all of these other states, you can get an arrival or a departure. Each of them are exponential random variables. The switch probability that we're talking about is in lambda over lambda plus mu to go up.

Mu over lambda plus mu to go down for all states other than 0. If you write this, this is in terms of the embedded Markov chain. And if you write this in terms of the

transition probabilities, it looks like this. Which is simpler? Which is more transparent? Well this, really, is what the mm1q is. That's where we started when we started talking about mm1q's.

This in this situation is certainly far simpler. You're giving these transition probabilities. But don't forget that we still have this embedded Markov chain in the background. Both these graphs have the same information. Both these graphs have the same information for every Markov process you want to talk about

OK. Let's look at sample time approximations to Markov processes. And we already did it in the last chapter. We just didn't talk about it quite so much.

We quantized time to increments of delta. We viewed all Poisson processes in a Markov process. Remember, we can view all of these transitions as independent Poisson processes all running away at the same time. We can view all of these Poisson processes as Bernoulli processes with probability of a transition from i to j in the increment delta, given as a delta times qij, to first order in delta.

So we can take this Markov process, turn it into a rather strange kind of Bernoulli process. For the M/M/1 queue, all that's doing is turning into a sample time, M/M/1 process. And we can sort of think the same way about general Markov processes. We'll see when we can and when we can't.

Since shrinking Bernoulli goes to Poisson, we would conjecture the limiting Markov chain as delta goes to 0 goes through a Markov process. In a sense that X of t is approximately equal to X prime. This is X prime as in the Bernoulli domain at delta times n.

You have to put self-transition into a sample time approximation. Because if you have a very small delta, there aren't big enough transition probabilities going out of the chain to fill up the probability space. So in most transition, you're going to just have a self-transition. So you need a self-transition, which is 1 minus delta times nu sub i, and these transitions to other states, which are delta times q sub ij. This has the advantage that if you believe this, you can ignore everything we're saying about

Poisson processes because you already know all of it.

We already talked about sample time processes. You can do this for any old process almost. And when you do this for any old process, you're turning it into a Markov change instead of a Markov process. This is the same argument you tried to use when you were a senior in high school or freshman in college when you said, I don't have to learn calculus, because all it is is just taking increments to be very small and looking at a limit. So I will just ignore all that stuff. It didn't work there. It doesn't work here. But it's a good thing to go every time you get confused, because this you can sort out for yourself.

There is one problem here. When you start making shrieking delta more and more, if you want to get a sample time approximation, delta has to be smaller than 1 over the maximum of the nu sub i's. If it's not smaller than the maximum of the nu sub i's, the self-transition probability here is unfortunately negative. And we don't like negative probabilities. So you can't do that.

If you have a Markov process, has a countably infinite number of states, each of these nu sub i's are positive. But they can approach 0 as a limit. As if they approach 0 as a limit, you cannot describe a sample time Markov chain to go with the Markov process. All you can do is truncate the chain also and then see what happens. And that's often a good way to do it.

OK, so we can always do this sample time approximation. What is nice about the sample time approximation is that we will find it in general, if you could use the sample time approximation, it always gives you the exact steady state probabilities. No matter how crude you are in this approximation, you always wind up with the exact rate values when you're all done. I don't know why.

We will essentially prove today that that happens. But that's a nice thing. But it doesn't work with the nu sub i's approach 0, because then you can't get a sample time approximation to start with.

OK, let's look at the embedded chain model in the sample time model of M/M/1

queue. I hate to keep coming back to the M/M/1 queue. But for Markov processes, they're all so similar to each other that you might as well get very familiar with one particular model of them, because that one particular model tells you most of the distinctions that you have to be careful about.

Let's see. What is this? This is the embedded chain model that we've talked about before. When you're in state 0, the only place you can go is to state 1.

When you're in state 1, you can go down with some probability. You can go up with some probability. And since these probabilities have to add up to 1, it's mu over lambda plus mu and lambda over lambda plus mu, and the same thing forever after.

If we're dealing with the sample time model, what we wind up with is we start out with qij, which is lambda here. The time it takes to get from state 0 to make a transition, the only place you can make a transition to is state 1. You make those transitions at rate lambda. So the sample time model has this transition and discrete time with probability lambda delta, this transition with probability mu delta and so forth up. You need these self-transitions in order to make things add up correctly.

The steady state for the embedded chain is pi sub 0 equals 1 minus rho over 2. How do I know that? You just have to use algebra for that. But it's very easy.

I'm going to have to get three of these things. OK, any time you have a birth-death chain, you can find the steady state probabilities. The probability of going this way is equal to the probability of going this way.

If you add in the probability of the steady state that you're concerned with, steady state transition this way, the same as the probability of a steady state transition this way. And you remember, the reason for this is in a birth-death chain, the total number of transitions from here to here has to be within 1 of the number of transitions from here to there. So that if steady state means anything-- and if these of long-term sample space probabilities with probability 1 mean anything-- this has to be true.

So when you do that, this is what you get here. This is a strange 1 minus rho over 2. It's strange because of this strange probability one here, and this strange probability mu over lambda plus mu here. And otherwise, everything is symmetric, so it looks the same as this one here.

For this one, the steady state for the sample time doesn't depend on delta. And it's pi sub i prime equals 1 minus rho times rho to the i where rho equals lambda over mu. This is what we did before. And what we found is since transitions this way have to equal transitions this way, these self-transitions don't make any difference here.

And you get the same answer no matter what delta is. And therefore you have pretty much a conviction, which can't totally rely on, that you can go to the limit as delta goes to 0 and find out what is going on in the actual Markov process itself. You'll be very surprised with this result if this were not the result for steady state probabilities in some sense for the Markov process.

However, the embedded chain probabilities and these probabilities down here are not the same. What's the difference between them? For the embedded chain, what you're talking about is the ratio of transitions that go from one state to another state. When you're dealing with the process, what you're talking about is the probability that you will be in one state.

If when you get in one state you stay there for a long time, because the rate of transitions out of that state is very small, so you're going to stay there for a long time. That enhances the probability of being in that state.

You see that right here, because pi 0 is 1 minus rho over 2. And pi 0 prime is 1 minus rho. It's bigger. And it's bigger because you're going to stay there longer, because the rate of getting out of there is not as big as it was before.

So the steady state probabilities in the embedded chain and the steady state probabilities and the sample time approximation are different. And the steady state probabilities and the sample time approximation are the same, when you go to the limit of infinitely fine sample time. Now what we have to do is go back and look at

renewal theory and all those that and actually convince ourselves that this works.

OK, so here we have renewals for Markov processes. And what have we done so far? We've been looking at the Poisson process. We've been looking at Markov chains. And we've been trying to refer to this new kind of process. Now we bring in the last actor, renewal theory.

And as usual, Poisson processes gives you the easy way to look at a problem. Markov chains gives you a way to look at the problem, when you'd rather write equations and think about it. And we renewal theory gives you the way to look at the problem when you really are a glutton for punishment, and you want to spend a lot of time thinking about it. And you don't want to write any equations, or you don't want to write many equation.

OK, an irreducible Markov process is a Markov process for which the embedded Markov chain is irreducible. Remember that an irreducible a Markov chain is one where all states are in the same class. We saw that irreducible Markov chains when we had a countably infinite number of states, that they could be transient, the state simply wanders off with high probability, never to return.

If you have an M/M/1 queue, and the expect the service time is bigger than the expected time between arrivals, then gradually the queue builds up. The queue keeps getting longer and longer as time goes on. There isn't any steady state. Looked at another way, the steady state probabilities are always 0, if you want to just calculate them.

So we're going to see the irreducible Markov processes can have even more bizarre behavior than these Markov chains can. And part of that more bizarre behavior is infinitely many transitions in a finite time. I mean, how do you talk about steady state when you have an infinite number of transitions and a finite time?

I mean, essentially, the Markov process is blowing up on you. Transitions get more and more frequent. They go off to infinity.

What do you do after that? I don't know. I can write these equations. I can solve

these equations. But they don't mean anything.

In other words sometimes, talking about steady state, we usually write equations for steady state. But as we saw with countable state Markov chains, steady state doesn't always exist there. There it evidenced itself with steady state probabilities that were equal to 0, which said that as time went on, things just got very diffused or things wandered off to infinity or something like that. Here it's this much worse thing, where in fact you get an infinite number of transitions very fast. And we'll see how that happens a little later.

You might have a transition rate which goes down to 0. The process is chugging along and gets slower, and slower, and slower, and pretty soon nothing happens anymore. Well, always something happens if you wait long enough, but as you wait longer and longer, things happen more and more slowly. So we'll see all of these things, and we'll see how this comes out.

OK, let's review briefly accountable state Markov chains-- an irreducible, that means everything can talk to everything else; is positive recurrent if and only if the steady state equations, pi sub j equals the sum of pi sub i, p sub ij. Remember what this is. If you're in steady state, the probability of being in a state j is supposed to be pi sub j.

The probability of being in a state is supposed to be equal then to the sum of the probabilities of being in another state and going to that state. That's the way it has to be if you're going to have a steady state. So this is necessary.

The pi sub j's have to be greater than or equal to 0. And the sum of the pi sub j's is equal to 1. It has a solution. If this has a solution, it's unique, and if pi sub i is greater than 0 for all i, if it's positive recurrent.

We saw that if it wasn't positive recurrent, other things could happen. Also the number of visits, n sub ij of n, remember in a Markov chain, what we talked about when we used renewal theory was the number of visits over a particular number transitions from i to j. n sub ij of n is the number of times we hit j's in the first n trials.

15

I always do this. Please take that n sub ij of n and write 1 over n times n sub ij of n is equal to pi j. You all know that. I know it too. I don't know why it always gets left off of my slides.

Now, we guessed for a Markov process the fraction of time in state j should be p sub j equals pi sub j over a nu sub j divided by the sum over i of pi sub i over nu sub i. Perhaps I should say I guessed that because I already know it. I want to indicate to you why if you didn't know anything and if you weren't suspicious by this time, you would make that guess, OK?

We had this embedded Markov chain. Over a very long period of time, the number of transitions into state i is going to be the number of transitions into state i divided by n is going to be pi sub i. That's what this equation here says, or what it would say if I had written it correctly.

Now, each time we get to pi sub i, we're going to stay there for a while. The holding time in state i is proportional to 1 over nu sub i. The rate of the next transition is nu sub i. So the expected holding time is going to be 1 over nu sub i, which says that the fraction of time that we're actually in state i should be proportional to the number of times we go into state j times the expected holding time in state j.

Now when you write p sub j equals pi sub j over nu sub j, you have a constant there which is missing. Because what we're doing is we're amortizing this over some long period of time. And we don't know what the constant of amortization is.

But these probability should add up to 1. If life is at all fair to us, the fraction of time that we spent in each state j should be some number which adds up to 1 as we sum over j. So this is just a normalization factor that you need to make the p sub j's sum up to 1.

Now what this means physically, and why it appears here, is something we have to go through some more analysis. But this is what we would guess if we didn't know any better. And in fact, it's pretty much true. It's not always true, but it's pretty much true.

So now let's use renewal theory to actually see what's going on. And here's where we need a little more notation even. Let n sub i of t be the number of transitions between 0 and t for a Markov process starting in state i. I can't talk about the number of transitions if I don't say what state we start in, because then I don't really have a random variable.

I could say let's start and steady state, and that seems very, very appealing. I've tried to do that many times, because it would simplify all these theorems. And it just doesn't work, believe me. So let's take the extra pain of saying let's start in some state i. We don't know what it is, but we'll just assume there is some state.

And the theorem says that the limit of M sub i of t is equal to infinity. Here I don't have a 1 over t in front of it. I've written incorrectly. And this is a very technical theorem.

We proved the same kind of technical theorem when we were talking about Markov chains, if you remember. We were talking about Markov chains. We said that in some sense, an infinite number of transitions into each one of the states had to occur.

The same kind of proof is the proof here. It had the same kind of proof when we were talking about renewal theory. What is going on is given any state, given the state the transition has to occur within finite time, because there some exponential holding time there. So the expected amount of time for the next transition is 1 over nu sub i. And that's finite for every i in the chain.

And therefore, as you go from one state to another, as the frog those jumping from one lily pad to another, and each lily pad that it jumps on there's some expected time before it moves, and therefore assuming that it keeps moving forever and doesn't die, which is what we assume with these Markov chains, it will eventually go through an infinite number of steps. And the proof of that is in the text. But it's exactly the same proof as you've seen several times before for renewal process in countable state Markov chains.

Next theorem is to say let M sub ij of t be the number of transitions to j, starting in state i. We can't get rid of the starting state. Somehow we have to keep it in there.

We have some confidence that it's not important, that it shouldn't be there. And we're going to see it disappear very shortly. But we have to keep it there for the time being.

So if the embedded chain is recurrent, then n sub ij of t is a delayed renewal process. And we sort of know that. Essentially, transitions keep occurring. So renewals in the state j must keep occurring. And therefore, any time you go to state j, the amount of time that it takes until you get there again, is finite.

We're not selling it to expect at time is finite. Expect time might be infinite. We'll see lots of cases where it is. But you've got to get there eventually. That's the same kind of thing we saw for renewal theory when we had renewals and the things that could happen eventually they did happen.

I don't know whether any of you are old enough to have heard about Murphy's Law. Murphy was an Irish American to whom awful things kept happening. And Murphy's Law says that if something awful can happen, it will.

This says if this can happen, eventually it will happen. It doesn't say it will happen immediately. But it says it will happen eventually. You can think of this as Murphy's Law, if you want to, if you're familiar with that.

So we want to talk about steady state for irreducible Markov processes. Now, let p sub j of i be the time average fraction of time in state j for the delayed renewal process. Remember we talked about p sub j in terms of these sample time Markov change.

And we talked about them a little bit in terms of imagining how long you would stay in state j if you were in some kind of steady state. Here we want to talk about p sub j of i. In terms of strong law of large numbers kinds of results, we want to look at the sample path average and see the convergence with probability one.

OK, so p sub j of i is a time average fraction of time in state j for the delayed renewal process. Remember we said that delay renewal processes were really the same as renewal processes. You just had this first renewal, which really didn't make any difference.

And so p sub j of i is going to be the limit as t approaches infinity of the reward that we pick up forever of being in state j. You get one unit of reward whenever you're in state j, 0 units when you're anywhere else. So this is the time average fraction of time you're in state j.

This is divided by t. And the assumption is you start in time i. So that affects this a little bit.

The picture here says whenever you go to state j, you're going to stay in state j for some holding time U sub n. Then you go back to 0 reward until the next time you went to state j. Then you jump up to reward of 1. You stay there for your holding time until you get into some other state, a and that keeps going on forever and ever.

What does the delayed renewal reward theorem say? It says that the expected reward over time is going to be the expect to reward in one renewal divided by the expected length of the renewal path. It says it's going to be expected value of U sub n divided by the expected time that you stay in state j. So it's 1 over nu sub j times the expected time you're in a state j.

That's a really neat result that connects this steady state probability. Excuse my impolite computer. This relates to fraction of time you're in state i to the expected delay in state j. It's one of those maddening things where you say that's great, but I don't know how to find either of those things.

So we go on. We will find them. And what we will find is W sub j. If we can find W sub j, we'll also know p sub j.

M sub ij of t is delayed renewal process. The strong law for renewal says the limit as t approaches infinity of Mi j of t over t is 1 over this waiting time. This is the number of renewals you have as t goes to infinity. This is equal to 1 over the expected

19

length of that renewal period.

Great. So we take the limit as t goes to infinity. Of mij of t over mi of t. This is the number of transitions overall up to time t starting in state i. This is the number of those transitions which go into state j. How do I talk about that? Well, this quantity up here is the number of transitions into state j over the number of transition that take place.

n sub ij of t is the number of transitions out of total transitions. mi of t is the total number of transitions. mi of t goes to infinity. So this limit here goes to the limit of n sub ij of n over n. Remember, we even proved this very carefully in class for the last application of it. This is something we've done many times already in different situations.

And this particular time we're doing it, we won't go through any fuss about it. It's just a limit of n sub ij of n over n. And what is that? That's the fraction, long term fraction of transitions that go into state j. We know that for accountable state Markov chain, which is recurrent, which is positive recurrent, this is equal to pi sub j. So we have that as equal to pi sub j.

We then have one over w sub j, that's what we're trying to find is equal to the limit of mij, of t over t, which is the limit of mij of t over mi of t times mi of t over t. We break this into a limit of two terms, which we've done very carefully before. And this limit here is pi sub j. This limit here is the limit of m sub i of t. And we have already shown that the limit of n sub i of t is equal to--

Somewhere we showed that. Yeah. 1 over w sub j is equal to pj times new sub j. That's what we proved right down here. p sub j of i is equal to 1 over new sub j times w sub j. Except now we're just calling it p sub j because we've already seen it. Doesn't depend on i at all. So one over w sub j is equal to p sub j times new sub j.

This says if we know what w sub j is, we know what p sub j is. So it looks like we're not making any progress. So what's going on? OK, well let's look at this a little more carefully. This quantity here is a function only of i. 1 over w sub j, over p sub j and

new sub j is a function only of j. Everything else in this equation is a function only of j.

That says that this quantity here is independent of i, and it's also independent of j. And what is it? It's the rate at which transitions occur. Overall transitions. If you're in steady state, this says that looking at any state j, the total number of transitions that occur is equal to, well, I do it on the next page. So let's goes there. It says that p sub j is equal to 1 over new sub j.

w sub j is equal to pi j over new j times this limit here. OK. So in fact, we now have a way of finding p sub j for all j if we can just find this one number here. This is independent of i, so this is just one number, which we now know is something that approaches some limit with probability one. So we only have one unknown instead of this countably infinite number of unknowns.

Seems like we haven't really made any progress, because before, what we did was to normalize these p sub js. We said they have to add up to 1, and let's normalize them. And here we're doing the same thing. We're saying p sub j is equal to pi sub j over new sub j with this normalization factor in. And we're saying here that this normalization factor has to be equal to 1 over the sum of pi k over new k.

In other words, if the p sub j's add to 1, then the only value this didn't have is 1 over the sum of pi k times new sub k. Unfortunately, there are examples where the sum of pi sub k over new sub k is equal to infinity. That's very awkward. I'm going to give you an example of that in just a minute, and you'll see what's going on.

But if pi sub k over new sub k is equal to infinity, and the theorem is true, it says that the limit of mi of t over t is equal to 1 over infinity, which says it's zero. So what this is telling us is what we sort of visualize before, but we couldn't quite visualize it. It was saying that either these probabilities add up to 1, or if they don't add up to 1, this quantity here doesn't make any sense.

This is not approaching a limit. The only way this can approach, well, this can approach a limit where the limit is 0. Because this theorem holds whether it

approaches the limit or not. So it is possible for this limit to the 0. In this case, these p sub js are all 0. And we've seen this kind of thing before.

We've seen that on a Markov chain, all the steady state probabilities can be equal to zero, and that's a sign that we're either in a transient condition, or in a null recurrent position. Namely, the state just wonders away, and over the long term, each state has 0 probability. And that looks like the same kind of thing which is happening here. This looks trivial.

There's a fairly long proof in the notes doing this. The only way I can find to do this is to truncate the chain, and then go to the limit as the number of states gets larger and larger. And when you do that, this theorem becomes clear. OK, let's look at an example where the sum of pi k over new k is equal to infinity, and see what's going on.

Visualize something like an mm1 queue. We have arrivals, and we have a server. But as soon as the queue starts building up, the server starts to get very rattled. And as the server gets more and more rattled, it starts to make more and more mistakes. And as the queue builds up also, customers come in and look at the queue, and say I'll come back tomorrow when the queue isn't so long.

So we both have this situation where as the queue is building up, service is getting slower and the arrival rate is getting slower. And we're assuming here to make a nice example that the two of them build up in exactly the same way. So that's what's happening here. The service rate when there's one customer being served is 2 to the minus 1.

The service right rate when there are two customers in the system is 2 to the minus 2. The service rate when there are three customers in the system is 2 to the minus 3. For each of these states, we still have these transition probabilities for the embedded chain. And the embedded chain, the only way to get from here to here is with probability 1, because that's the only transition possible here.

We assume that from state 1, you go to states 0 with probability 0.6. You go to state

two with probability 0.4. With state 1, from state 2, you go up with probability 0.4, you go down with probability 0.6. The embedded chain looks great. There's nothing wrong with that. That's a perfectly stable mm1 queue type of situation. It's these damned holding times which become very disturbing.

Because if you look at pi sub j, which is supposed to be 1 minus rho times rho to the j. Rho is 2/3, it's lambda over mu. So it's lambda over lambda plus mu over rho plus lambda plus mu. It's 0.4 divided by 0.6, which is 2/3. If we look at pi sub j over new sub j, it's equal to 2 to the j times 1 minus rho, times rho to the j. It's 1 minus rho times 4/3 to the j.

This quantity gets bigger and bigger as j increases. So when you try to sum pi i over new sub j, you get infinity. So what's going on? None of the states here have an infinite holding time associated with them. It's just that the expected holding time is going to be infinite.

The expected number of transitions over a long period of time, according to this equation here, expected transitions per unit time is going to 0. As time goes on, you keep floating back to state 0, as far as the embedded chain is concerned. But you're eventually going to a steady state distribution, which is laid out over all the states. That steady state distribution looks very nice.

That's 1 minus rho times rho to the j. Rho is 2/3, so the probability of being in state j is going down exponentially as j gets big. But the time that you spend there is going up exponentially even faster. And therefore, when we sum all of these things up, the overall expected rate is equal to zero, because the sum of the pi j over new j is equal to infinity. OK.

So this is one of the awful things that are going to happen with Markov processes. We still have an embedded chain. The embedded chain can be stable. It can have a steady state, but we've already found that the fraction of time in a state is not equal to the fraction of transitions that go into that state. pi sub j is not in general equal to p sub j.

And for this example here with the rattled server and the discouraged customers, the amount of time that it takes, the expected amount of time that it takes for customers to get served is going to zero. Even though the queue was saying stable. Does mathematics lie? I don't know. I don't think so. I've looked at this often enough with great frustration, but I believe it at this point.

If you don't believe it, take this chain here and truncate it, and solve the problem for the truncated chain, and then look at what happens as you start adding states on one by one. What happens as you start adding states on one by one is that the rate at which this Markov process is serving things is going to zero.

So the dilemma as the number of states becomes infinite, the rate at which things happen is equal to 0. It's not pleasant. It's not intuitive. But that's what it is. And that can happen. Again, let's go back to the typical case of a positive recurrent embedded chain, where this funny sum here is less than infinity.

If the sum here is less than infinity, then you can certainly express p sub j as pi sub j over new sub j divided by the sum over k of p sub k over new sub k. Why can I do that? Because that's what the formula says. I don't like to live with formulas, but sometimes things get so dirty, the mathematics play such awful tricks with you that you have to live with the formulas, and just try to explain what they're doing.

p sub j is equal to this. Limit of the service rate, if this quantity is non infinite, then things get churned out of this queueing system at some rate. And the p sub js can be solved for. And this is the way to solve for them. OK, so that's pretty neat. It says that if you can solve the embedded chain, then you have a nice formula for finding the steady state probabilities.

And you have a theorem which says that so long as this quantity is less than infinity with probability one, the fraction of time that you stay in state j is equal to this quantity. Well, that's not good enough. Because for the mm1 queue, we saw that it was really a pain in the neck to solve for the steady state equations for the embedded chain. Things looked simpler for the process itself.

So let's see if we can get those equations back also. What we would like to do is to solve for the p sub j's directly by using the steady state embedded equation. Embedded equations say that pi sub j is equal to the sum over i, pi sub i times p sub ij. The probability of going into a state is equal to the probability of going out of a state.

If I use this formula here, pi sub j over new sub j divided by some constant is what p sub j is. So pi sub j is equal to p sub j times new sub j times that constant. Here we have the p sub j times the new sub j divided by that constant. And that's equal to this sum here over all i divided by the same constant. So the constant cancels out.

Namely, we left out that term here, but that term is on this side, and it's on this side. So we have this equation here. If you remember, I can't remember, but you being younger can perhaps remember. But when we were dealing with a sample time approximation, if you leave the deltas out, this is exactly the equation that you got. It's a nice equation.

It says that the rate at which transitions occur out of state i, the rate at which transitions occur out of state i, out of state j, excuse me, is p sub j times new sub j. Here's the holding time, and there's also the probability of being there. Excuse me.

Let's be more clear about that. I'm not talking about given you're in state j, the rate at which you get out of state j. I'm talking about the rate at which you're in state j and you're getting out of state j. If you could make sense out of that. That's what this is. This quantity here is the overall rate at which you're entering state j.

So these equations sort of have the same intuitive meaning as these equations here do. And then if you solve this equation in the same way, what's that doing? Oh. This gets you pi sub j in terms of the p sub j's. So there's a very nice symmetry about this set of equations. The p sub j's are found for the pi sub j's in this way.

The pi sub j's are found from the p sub j's by this same sort of expression. The theorem then says if the embedded chain is positive recurrent, and the sum of pi i over new i is less than infinity, then this equation has a unique solution. In other

25

words, there is a solution to the steady state process equations. And pi sub j and p sub j are related by this, and by this.

If you know the pi sub j's, you can find the p sub j's. If you know the p sub j's, you can find the pi sub j's. There's a fudge factor, and the sum of pi sub i over new sub i is equal to sum of p sub j times new sub j to the minus 1. This equation just falls out of looking at this equation and this equation. I'm not going to do that here, but if you just fiddle with these equations a little bit, that's what you find.

I think graduate students love to push equations. And if you push these equations, you will rapidly find that out. So there's no point to doing it here. OK. You can do the opposite thing, also. If the steady state process equations are satisfied, and the p sub j's are all greater than zero, and the sum of the p sub j's are equal to 1.

And if these equations are less than infinity, this is just by symmetry with what we already did. Then pi sub j has to be equal to p sub j times new sub j divided by this sum. And this gives the steady state equations for the embedded chain. And this shows that the embedded chain has to be positive recurrent, and says that you have to be able to go both ways.

So we already know that if you can solve the steady state equations for the embedded chain, they have to be unique, the probabilities all have to be positive. All those neat things for accountable state and Markov chains hold true. This says that if you can solve that virtually identical set of equations for the process, and you get a solution, and the sum here is finite, then in fact you can go back the other way.

And going back the other way, you know from what we know about embedded chains that there's a unique solution. So there has to be a unique solution both ways. There has to be positive recurrence both ways. So we have a complete story at this point. I mean, you'll have to spend a little more time putting it together, but it really is there.

If new sub j is bounded over j, then the sum over j of p sub j, new sub j is less than infinity. Also, the sample time chain exists because this is bounded. It has the same

steady state solution as the Markov process solution. In other words, go back and look at the solution for the sample time chain. Drop out the delta, and what you will get is this set of equations here.

If you have a birth death process, it's a birth death process for both the Markov process, and also for the embedded chain. For the embedded chain, you know that what you have to do to get-- for a birth death chain, you know an easy way to solve the steady state equations for the chain are to say transitions this way equal transitions this way. It's the same for the process.

The amount of time that you spend going this way is equal to the average amount of time you spent going this way. So it says for a birth death process, p sub i times q sub i, i plus one. That's an i there, is equal to p sub i plus 1 times the transition probability from i plus 1 back to i. So this symmetry exists almost everywhere.

And then if the sum of p sub j, new sub j is equal to infinity, that's the bad case. These equations say that pi sub j is equal to 0 everywhere. This is sort of the dual of the situation we already looked at. In the case we already looked at of the lazy or rattled server and the discouraged customers, eventually the rate at which service occurred went to 0.

This situation is a situation where as far as the embedded chain is concerned, it thinks it's transient. As far as the process is concerned, the process thinks it's just fine. But then you look at the process, and you find out that what's happening is that an infinite number of transitions are taking place in a finite time. Markov process people call these irregular processes.

Here's a picture of it on the next slide. Essentially, the embedded chain for a hyperactive birth death chain. As you go to higher states, the rate at which transitions take place gets higher and higher. And for this particular example here, you can solve those process equations. The process equations look just fine. This is where you have to be careful with Markov processes.

Because you can solve these Markov process equations, and you get things that

look fine, when actually there isn't any steady state behavior at all. It's even worse than the rate going to zero. When the rate goes to infinity, you can have an infinite number of transitions taking place in a finite time. And then nothing happens. Well, I don't know whether something happens or not.

I can't visualize what happens after the thing has exploded. Except essentially, at that point, there's nothing nice going on. And you have to say that even though the process equations have a steady state solution, there is no steady state in terms of over the long period of time, this is the fraction of time you spend in state j. Because that's not the way it's behaving. OK.

I mean, you can see this when you look at the embedded chain. The embedded chain is transient. You're moving up with probability 0.6. You're moving down with probability 0.4. So you keep moving up. When you look at the process in terms of the process with the transition rates q sub ij, the rates going up are always less than the rates going down.

This is because as you move up in state, you act so much faster. The transition rates are higher in higher states, and therefore the transition rates down are higher, and the transition rate down from state 2 to state 1 is bigger than the transition right up from i to i plus 1. And it looks stable, as far as the process is concerned.

This is one example where you can't look at the process and find out anything from it without also looking at the embedded chain, and looking at how many transitions you're getting per unit time. OK, so that's it. Thank you.