

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

**PROFESSOR:** OK, so I guess we're ready to start. Clearly, you can talk to us about the exams later if you want to. And we will both be around after class for a bit-- or I'll be around after class for a bit and regular office hours.

Wanted to finish talking about [INAUDIBLE] state Markov change today and go on to talk about Markov processes. And the first thing we want to talk about is what does reversibility mean. I think reversibility is one of these very, very tricky concepts that you think you understand about five times, and then you realize you don't understand it about five times. And hopefully by the sixth time, and we will see at about six times, so hopefully by the end of the term, it will look almost obvious to you.

And then, we're going to talk about branching processes. I said and what got passed out to you that I'd be talking about round robin and processor sharing. I decided not to do that. It was too much complexity for this part of the course. I will talk about it just for a few minutes. And then we'll go into Markov processes. And we will see most of the things we saw in Markov change again but in a different context and in a slightly more complicated context.

So for any Markov chain, we have these equations. Typically, you just state the equation of the probability of  $X_{n+1}$  given all the previous terms is equal to probability of  $X_{n+1}$  given  $X_n$ . It's an easy extension to write it this way. The probability of lots of things in the future given everything in the past is equal to lots of things in the future just given the most recent thing in the past.

So what we did last time was to say let's let  $A_+$  be any function of all of these things here, and let's  $A_-$  be any function of all of these things here except for  $X$

sub  $n$ , namely  $X_{n-1}$  down to  $X_0$ . And then what this says more generally is the probability of all these future things condition on  $X_n$ , and all the past things is equal to probability of the future given just  $X_n$ .

Then, we wrote that by multiplying by the probability of  $A$  minus given  $X_n$ . And you can write it in this nice symmetric form here. I'm hoping that these two laser pointers, one of them will keep working. And as soon as you write it in this symmetric form, it's clear that you can again turn it around and write in the past given the present and the future is equal to the past given the present.

So this formula really is the most symmetric form [? for it ?] and it really shows the symmetry of past future, at least as far as Markov chains are concerned. Yeah?

**AUDIENCE:** I don't understand [INAUDIBLE] write that down though. I feel like I'm missing a step. For example, let's say I [INAUDIBLE], I can't infer where I came from?

**PROFESSOR:** No, that's not what this says. I mean all it says is a probabilistic statement. It says everything you can say about  $X_{n+1}$  which was the first way we stated. Everything you know about  $X_{n+1}$ , you can find out by just looking at  $X_n$ . And knowing the things before that doesn't help you at all.

When you write it out a Markov chain in terms of a graph, you can see this because you see transitions going from one state to the next state. And you don't remember what the past is. The only part of the past you remember is just that last state. It look you're still puzzled.

So it's not how it's saying we can't tell anything about the past and the future. In fact, if you don't condition on  $X_n$ , this stuff back here has a great deal to do with the stuff up here. I mean it's only when you do this conditioning, it is saying that the conditioning at the present is the only linkage you have between past and future. If you know where you are now, you don't have to know anything about the past and know what's going to happen in the future.

That's not the way life is. I mean life is not a Markov chain. It's just the way these Markov chains are. But this very symmetric statement says that as far as Markov

chains are concerned, past and future look the same. And that's the idea that we're trying to use when we get into reversibility. This isn't saying anything about reversibility, yet this is just giving a general property that Markov chains have.

And when you write this out, it says the probability of this past state given  $X_n$  and everything in the future is equal to the probability of the past state given  $X_{n-1}$ . So this is really the Markov condition running from future down to past.

And it's saying that if you want to evaluate these probabilities of where you were given anything now and further on, or put it in a more sensible way, if you know everything over the past year, and from knowing everything over the past year, you want to decide what can you tell about what happens the year before, what it's saying is the probability of what happened the year before is statistically a function only on the last day on the first day of this year that you're conditioning on.

So Markov condition works in both directions. You need to study state and forward change to be there in order to have homogeneity in a backward chain. In other words, usually, we define a Markov chain by starting off at time zero and then evolving from there. So when you go backwards, that fact that you started at time zero and said something about time zero destroys the symmetry between past and future. But if you start off in steady state, then everything is as it should be.

So if you have a positive-recurrent Markov chain in steady state, it can't be in steady state unless it's positive-recurrent, because otherwise, you can't evaluate the steady-state probabilities. The steady-state probabilities don't exist. And the backward probabilities are probability that  $X_{n-1} = j$  given that  $X_n = i$  is the transition probability from  $i$  to  $j$  times the steady-state probability  $\pi_j$  over  $\pi_i$ .

This looks more sensible if you bring the  $\pi_i$  over there,  $\pi_i$  times probability of  $X_{n-1} = j$  given  $X_n = i$  is really the probability of  $X_n = i$  and  $X_{n-1} = j$ . So what this statement is really saying is it's  $\pi_i$  times the probability of  $X_{n-1} = j$  given  $X_n = i$  is really the probability of being in state  $j$  at time  $n-1$  and state  $i$  at time  $n$ . And we're

just writing that in two different ways. It's the base law way of writing things in two different ways.

If we define this backward probability, which we said you can find by base law if you want to work at it, if we define this to be the backward transition probability,  $p_{ij}^*$ , in other words,  $p_{ij}^*$  is in this world where things are moving backwards, it corresponds to  $p_{ji}$  in the world where things are moving forward.

$P_{ij}^*$  is then the probability of being in state  $j$  at the next time back given that you're in state  $i$  at this time if you're in state  $i$  at the present. In other words, if you're visualizing moving from future time back to backward time, that's what your Markov chain is doing now. These star transition probabilities are the probabilities of moving backward by one step, conditional going where you were at time  $n$ , where you're going to be at time  $n - 1$ , if you will.

As I said, these things are much easier to deal with if you view them on a line, and you have a right moving chain which is what we usually think of as the chain moving from past to future. And then you have a left moving chain, which is what you view as moving from future down to past.

OK, we define a chain as reversible if these backward probabilities are equal to the forward transition probabilities. So if a chain is reversible, it's says that  $p_{ij} = P_{ji}^*$ , this is the probability that you are at a time  $n - 1$ , you were in state  $i$ , and then you move to state  $j$ . So it's the probability of being in one state at one time, the next state at the next time. It's the probability that  $X_{n-1}$  and  $X_n$  are  $ij$ .

And this probability here is-- this equation is moving forward in time. So this equation here is the probability that you were in state  $j$ , and you move to state  $i$ . So what we're saying is the probability of being in  $i$  moving to  $j$  is the same as the probability of being in  $j$  and moving to  $i$ .

It's the condition you have on any birth-death chain. We said that on any birth-death chain, the fraction of transitions from  $i$  to  $j$  has to be equal to the total number of transitions from  $j$  to  $i$ . It's not that the probability of moving up given  $i$  is the same as

that of moving back. That's not what it's saying. It's saying that the probability of having a transition over time is  $\pi_i$  times  $P_{ij}$ .

Reversibility says that you make as many up transitions over time as you make down transitions over the same pair of states. I think that's the simplest way to state the idea of reversibility. The fraction of time that you move from state  $i$  to state  $j$  is the same as the fraction of time in which you move from state  $j$  to state  $i$ .

It's what always happens on a birth-death chain, because every time you go up, if you ever get back to the lower part of the chain, you have to move back over that same path. You can easily visualize other situations where you have the same condition if you have enough symmetry between the various probabilities involved. But the simplest way is to have this sort of-- well, not only the simplest, but also the most common way is to have a birth-death chain.

OK, so this leads us to the statement, all positive-recurrent birth-death chains are reversible, and that's the theorem. Now the question is what do you do with that? Let's have a more general example than a birth-death chain. Suppose the non-zero transition of a positive-recurrent Markov chain form a tree. Before we had the states going on a line, and from each state to the next state, there were transition probabilities, you could only go up or down on this line.

What I'm saying now is if you make a tree you have the same sort of condition that you had before if the transitions on the states look like a tree. So these are the only transitions that exist in this Markov chain. These are the states here. Again, you have this condition. The only way to get from this state out to this state is to move through here so that the number of transitions that go from here to there must be within one of a number of transitions that go from here back to there.

So you have this reversibility condition again on any tree. And these birth-death chains are just very, very skinny trees where everything is laid out on a line. But this is the more general case. And you'll see cases of this as we move along.

The following theorem is one of these things that you use all the time in solving

problems. And it's extraordinarily useful. It says for a Markov chain with transition probabilities  $P_{ij}$ , if a set of numbers  $p_i$  exists so that all of them are positive, they sum to one. If you can find such a set of numbers, and if they satisfy this equation here, then you know that the chain is reversible, and you know that those numbers are the steady-state probability. So you get everything at once.

It's sort of like a guessing theorem. And I usually call it a guessing theorem, because starting out, it's not obvious that these equations have to be satisfied. They're only satisfied if you have a chain which is reversible. But if you can find a solution to these equations, then, in fact, you know it's reversible, and you know you found steady-state probabilities.

It's a whole lot easier to solve this equation usually than to solve the usual equation we have for steady-state probabilities. But the proof of the theorem-- I just restated the theorem here, leaving out all of the boiler plate. If we take this equation for fixed  $j$ , and we sum over  $i$ , what happens? When you sum over  $i$  on this side, you get the sum over  $i$  of  $p_i P_{ij}$ . When you sum over  $i$  on this side, you get  $p_j$ , because when you sum  $P_{ji}$  over  $i$ , you have to get one.

When you're in state  $j$ , you have to go someplace. And you can only go one place, each with different probabilities. So that gives you the usual steady-state conditions. If you can solve those steady state conditions, then you know from what we did before that the chain is positive-recurrent. You know there are steady-state probabilities. You know there's probabilities are all greater than zero. So if there's any solution to these steady-state equations, then you know the chain has to be positive-recurrent. And you know it has to be reversible in this case.

OK, here are a bunch of sanity checks for reversibility. In other words, if you're going to guess at something that's reversible and try to solve these equations, you might as well do a sanity check first.

The simplest and most useful sanity check is if you want it be reversible, and there's a transition from  $i$  to  $j$ , then there has to be a transition from  $j$  to  $i$  also. Mainly the number of transitions going from  $i$  to  $j$  has to be the same over the long term to

number of transitions going from  $j$  to  $l$ .

If there's a zero transition probability one way and not the other way, you can't satisfy that equation. If the chain is periodic, the period has to be too. Why is that? Well, it's a long proof in the notes. And if you write everything down in algebra, it looks a little long. If you just think about it, it's a lot shorter. If you're going around on a cycle of, say, length three, if the chain is periodic, and it's periodic with some period other than two, then you know that the set of states has to partition into a set of subsets.

And you have to move from one subset, to the next subset, to the next subset, and so forth. When you go backwards, you're moving around that cycle in the opposite direction. Now, the only way that moving around a cycle one way and moving around it the other way works out is when the cycle only has two states [? set in, ?] because then you're moving, and you're moving right back again.

OK, so the period has to be two if it's periodic. If there's any set of transitions  $i$  to  $j$ ,  $j$  to  $k$ , and  $k$  to  $l$ , namely if you can move around this way with some probability, then the probability of moving back again has to be the same thing. And that's what this is saying. This is moving around this cycle of length three one way. This is the forward probabilities for moving around a cycle, the opposite way and to have reversibility. The probability of going one way has to be the same as the probability going the other way.

Now, that sounds peculiar, and it gives me a good excuse to point out one of the main things that's going on here. When you say something is reversible, it doesn't usually mean that  $P_{ij}$  is equal to  $P_{ji}$ . What it means is that  $\pi_i P_{ij}$  equals  $\pi_j P_{ji}$ . Namely, the fraction of transitions here is the same as the fraction of transitions here.

Why is it that here I'm only using the probabilities, and I'm not saying anything about the initial probability? It's because both of these cycles start with state  $i$ . So what you really want to do is say  $\pi_i P_{ij} P_{jk} P_{ki}$  is the same as  $\pi_i P_{ji}$  [INAUDIBLE] And then you cancel out the  $\pi_i$ . So when you have a cycle, you don't

need that initial steady-state probability in there.

There's a nice generalization of the guessing theorem to non-reversible change and that generalization and it's proved the same way that this is proved. If you can find a set of transition probabilities,  $P_{ij}^*$ , and to be a set of transition probabilities, they have to be non-negative. When you sum this over  $j$ , you have to get one. That's what you need to have a set of transition probabilities. Then, all you need is  $\pi_i \sum_j P_{ij} = \sum_j \pi_j P_{ji}^*$ .

In other words, when you look at this backward transition probability for an arbitrary Markov chain which is positive-recurrent, this has to equal this. This is one of the conditions that you have on a Markov chain. The interesting thing here is this is enough. If you can guess a set of backward transition probabilities to satisfy this equation for all  $i$  and  $j$ , then you know you must have a set of steady-state probabilities where the steady-state probabilities are [INAUDIBLE].

And the way to prove this is the same as before. Namely, you sum this over  $j$ . And when you sum this over  $j$ , you get the backward transition probability. So I'm not going to prove it. I mean the proof is in the notes, and it's really the same proof as we went through before.

And incidentally, if you read the section on round robin, you will find the key to finding out what's going on there is, in fact, that theorem. It's that way of solving for what the steady-state probabilities have to be. While I'm at it, let me pause for just a second, because we're not going to go through that section on round robin. Let me talk about what it is, what processor sharing is, and why that result is pretty important.

If you're at all interested in-- well, let's see. First pack of communication is something important. Second, computer systems of all types is important. There was an enormous transition probably 20 years ago from computer systems solving one job at a time, and then it went to the system of solving many jobs concurrently. It would work a little bit on one job, a little bit in another, a little bit in another, and so forth. And it turns out to be a very good idea for doing that.

Or if you're interested in killing systems, what happens if you have a killing system-- suppose it's a GG1 queue. So you have a different service time for each customer. Or let's make it an MG1 queue. Makes the argument cleaner. Different customers have different service times. We've seen that in an MG1 queue, everybody can be held up by one slow customer. And if the customers have an enormously, widely varied service time, some of them have required enormously long service time, that causes an enormous amount of queuing.

What happens if you use processor sharing, you have one server. And it's simultaneously allocating service to every customer which is in queue. So it takes a service capability, and it splits it up end-wise. And when you talk about processor sharing, you assume that there's no overhead for doing the splitting. And if there's no overhead for doing the splitting, you can see intuitively what happens.

The customers that don't need much service are going to be held up a little bit by these customers who require enormous amounts of service, but not too much. Because this customer that requires enormous service is getting the same rate of service as you are. If that customer requires 100 hours of service, and you only require one second of service, you're going to get out very much faster than they do.

What happens when you analyze all of this? It turns out that you've turned the MG1 queue into an MM1 queue. In other words, if you're doing processor sharing, it takes the same expected amount of time for you to get out as it would if all of the service times were exponential. Now, that is why people went to time sharing a long time ago.

Most of the arguments for it, especially in the computer science fraternity, were all sorts of other things. But there's this very simple queuing argument that led to that. Unfortunately, it's a fairly complicated queuing argument, which is why we're not going through it. But it's a very important argument.

Why, at the same time, did we go to packet communication? Well, there are all sorts

of reasons for going to packet communication instead of sending messages, long messages, one at a time. But one of them, and one of them is very important, is the same process of sharing resell. If you split things up into small pieces, then what it means is effectively things are being served in a process of sharing matter.

So again, you get this losing the slow truck effect. And everybody gets through effectively in a fair amount of time. OK. I probably said just the wrong amount about that, so you can't understand what I was saying. But I think if you read it, you will get the idea of what's going on.

OK. Let's look at an MM1 queue now. An MM1 queue, you remember, is a queue where you have customers coming in. In a [Poisson] manner, the interval between customers is exponential. That you couldn't model for a lot of things. The service time is exponential. And what we're going to do to try to analyze this in terms of mark of change, is to say well, let's look at sampling the state of the MM1 queue at some very finely spaced interval of time.

And we'll make the interval of time,  $\delta$ , so small that there's a negligible probability of having two customers come in in the same interval. And so there's a negligible probability of having a customer come in and a customer go out in the same interval. It's effectively the same argument that we use to say that a Poisson process is effectively the same as a Bernoulli process, if you make the time interval the step size for the Bernoulli process very, very small, and the probability of success very, very small. As you make that time interval smaller and smaller, it goes in to a Poisson process as we showed a long time ago.

This is the same argument here. And what we get then is this system, which now has a state. And the state is the number of customers that are in the system. As one customer is being served, rest of the customers are sitting in a queue. The transitions over some very small time,  $\delta$ , there's a probability  $\lambda \delta$  that a new customer comes in. So there's a transition to the right.

There's a probability  $\mu \delta$ , if there's a server being served, that that service gets finished in this time  $\delta$ . And if you're in state zero, then of course, you can get a

new arrival coming in, but you can't get any service being done. So it's saying, as you're all familiar with, you have this system where any time there are customers in the system, they're getting served at rate  $\mu$ .

$\mu$  has to be bigger than  $\lambda$  to make this thing stable. And you can see that intuitively, I think. And when you're in state zero, then the server isn't doing anything. So the server is resting, because the server is faster than the arrival process. And then the only thing that can happen is a new arrival comes in, and then the server starts to work again, and you're back in state 1.

So this is just a time sampled version of the MM1 queue. And if you analyze this either from the guessing theorem that I just was talking about or the general result for birth death change that we talked about last time. You see that  $\pi_{n-1} \lambda \Delta$  is equal to  $\pi_n \mu \Delta$ . The fraction of transitions going up is equal to the fraction of transitions going down.

You take the steady state probability of being in state  $n-1$ . You multiply it by the probability of an up transition. And you get the same thing, as you take the probability of being in state  $n$  and multiply it by a down transition. If you define  $\rho$  as being  $\lambda/\mu$ , then what this equation says is a steady state probability of being in state  $n$  is  $\rho$  times the steady state probability of being in a state  $n-1$ .

This is the same as the general birth death result, except that  $\rho$  is a constant overall state rather than state 1.  $\pi_n$  is then equal to  $\rho^n \pi_0$ . And  $\pi_n$  is then equal to, if you re-curse on this, you get this. Then you use the condition that the  $\pi_i$ 's have to add up to 1. And you get  $\pi_n$  has to be equal to  $1 - \rho^n$ .

OK. This is all very simple and straightforward. What's curious about this is it doesn't depend on  $\Delta$  at all. You can make  $\Delta$  anything you want to. And we know that if we shrink  $\Delta$  enough, it's going to look very much like an MM1 queue. But it looks like an MM1 queue no matter what  $\Delta$  is. Just so long as  $\lambda + \mu \Delta$  is less than or equal to 1.

You don't want transition probabilities to add up to more than 1. And you have these self loops here which take up the slack. And we saw before that the steady state probabilities didn't have anything to do with these self transitions. And that will turn out to be sort of useful later on.

So we get these nice probabilities which are independent of the time increment that we're taking. So we think that this is probably pretty much operating like an actual MM1 queue would operate. OK. Now here's this diagram that I showed you last time, and I told you was going to be confusing. And I hope it's a little less confusing at this point.

We've now talked about reversibility. We know what reversibility means. We know that we have reversibility here. And what's going on? We have this diagram on the top, which is the usual diagram for the way that an MM1 queue operates. You start out in state zero. The only thing that can happen from state zero is at some point you get an arrival.

So the arrival takes you up there. You have no more arrivals for a while. Some later time, you get another arrival. [INAUDIBLE] So this is just the arrival process here. This is the number of arrivals up until time  $T$ . The same time, when you have arrivals, eventually since the server is working now, at some point there can be a departure. So we go over to here in the sample sequence. There's eventually a departure there. There's a departure there.

And then you're back in state zero again. You go along until there's another arrival. Corresponding to this sample path of arrivals and departures, we can say what the state is. The state is just the difference between the arrivals and the departures for this sample path. So the state here start out at time 1.  $x_1$  is equal to 0. Then at time  $x_2$ , suddenly an arrival comes in.  $x_2$  is equal to 1,  $x_3$  is equal to 1,  $x_4$  is equal to 1,  $x_5$  is equal to 1.

Another arrival comes in. So we have a queue of 1. We have the server operating on one customer. Then in the sample path, we suppose there's a departure. And

we suppose that the second arrival required hardly any service. So there's a very fast departure there.

Now, what we're going to do is to look at what happens. This is the picture that we have for the Markov chain. This with the picture we had for the sample path of arrivals and departures for what we thought was the real life thing that was going on. We now have the state diagram. And now what we're going to do is say, let's look at this backwards.

And since looking at it backwards in time is complicated, let's look at it coming in this way. So we have the state diagram, and we try to figure out what, going backwards, is going on here from these state transitions. Well in going backwards, the state is increasing by 1. So that looks like something that we would call an arrival.

Now why am I calling these arrivals and departures? It's because the probability of any sample path along here is going to be the same as a backward sample path, the same sample path, going backwards. That's what we've already established. And the probabilities going backwards are going to be the same as the probabilities going forward.

Since we can interpret this going forward as arrivals causing up transitions, departures causing down transitions, going backwards we can say this is an arrival in this backward going chain. This is an arrival in a backward going chain. This is a departure in the backward going chain. We go along here. Finally, there's another departure in the backward going chain.

This state diagram-- with two of them, we might make it. Yes, OK. The state diagram here determines this diagram here. If I tell you what this is, you can draw this. You can draw every up transition as an arrival, every down transition as a departure. So this diagram is specified by this diagram. This diagram is also specified by this diagram. So this and this each specify each other.

Now if we interpret this as arrivals and this is departures, and we have the probabilities of an MM1 chain, then we say the statistics of these arrivals here are

the same as a Bernoulli process, which is coming along the other way and leading to arrivals. What that says is the departure process here is a Bernoulli process.

Now you really have to wrap your head around that a little bit. Because we know that departures only occur when you're in states greater than or equal to 1. So what's going on? When you're looking at it in forward time, a departure can only occur from a non-negative state to some other state. Namely from some non-negative state to some smaller state.

Now, when I look at it backwards in time, what do I find? I can be in state zero. And there could have been a departure which may-- if I'm in state zero at time zero, and I say there was a departure between  $n - 1$  and  $n$ , that just says that the state at time  $n - 1$  was equal to 1. Not that the state at time  $n$  was equal to 1.

Because I'm running along here looking at these arrivals going this way, departures going this way. When I'm in state zero, I can get an arrival. I can't when I'm in state 1. If I were here, I couldn't get a departure in the next unit of time. Because the state is equal to zero. But I can be coming from a departure in the previous state. Because in the previous state, the state was 1.

I mean, you really have to say this to yourself a dozen times. And you have to reason about it. You have to look at the diagram, read the notes, talk to your friends about it. And after you do all of this, it will start to make sense to you. But I hope I'm at least making it seem plausible to you.

So each sample path corresponds to both a right and left moving chain. And each of them are MM1. So we have Burke's theorem. And Burke's theorem says given an MM1 sample time Markov chain in steady state, first, the departure processes Bernoulli at rate  $\lambda$ . OK.

Let me put it another way now. When we look at it in the customary way, we're looking at things moving upward in time. We know there can't be a departure when you're in state zero. That's because we're looking at departures after you're in time zero. When we look at time coming in backwards, we're not being dependent on the

state to the left of that departure.

We're only dependent on the state after the departure. The state after departure can be anything. OK? And therefore, after departure you can be in any state at all. And therefore, you can always have a departure, which leaves you in state zero. That's exactly what this theorem is saying. It's saying-- yes.

**AUDIENCE:** [INAUDIBLE] departure process?

**PROFESSOR:** Well, a couple of reasons. If you had a Bernoulli process and departure rate was  $\mu$ , over a long period of time, you'll have more departures than you have arrivals. But the other, better reason is that now you're amortizing those departures over all time. And before, you were amortizing them only over times when the state of the chain was greater than what?

The probability of the state of the chain is greater than  $1/\rho$ . And that's the difference between  $\lambda$  and  $\mu$ . OK? It's not nice, but that's the way it is. Well actually, it is nice when you're solving problems with these things. I mean some of you might have noticed when you were looking at the quiz problem dealing with Poisson processes, that it was very, very sticky to say things about what happens at some time in the past, given what's going on in the future.

Those are nasty problems to deal with. This makes those problems very easy to deal with. Because it's saying, if you go backward in time, you reverse the role of departures and arrivals. Yes.

**AUDIENCE:** Can you explain that one more time, why it's  $\lambda$  and not  $\mu$ ? Just the last thing you said [INAUDIBLE].

**PROFESSOR:** OK. Last thing I said was that the probability that the state is bigger than zero is  $\rho$ . Because the probability of the state is zero is  $1 - \rho$ . I mean that's not obvious, but it's just the way it is. So that if you're trying to find the probability of a departure and you don't know what the state is, and you just look in at any old time, it's sort of like a random incidence problem.

I mean, you're looking into this process, and all you know is you're in steady state. And you don't know what the state is. I mean you can talk about the earlier state. You can't talk about-- I mean usually when we talk about these Markov chains, we're talking about state of time  $n$ , transition from time  $n$  to  $n + 1$ . And in that case, you can't have a departure if you're in state zero at time  $n$ .

Now the transition from time  $n$  to  $n + 1$ , if we're moving the other way in time, we're starting out at time  $n + 1$ . And we're starting out at time  $n + 1$ . If you're in state zero there, you can still be coming out of a departure from time  $n$ . I mean suppose at time  $n$  the state is 1, and at time  $n + 1$  the state is zero. That means there was a departure between  $n$  and  $n + 1$ .

But when you're looking at it from the right, what you see is the state at time  $n + 1$  is zero. And there's a probability of a departure. And it's exactly the same as the probability of a departure given any other state. OK? If you're just doing this as mathematicians, you could look at these formulas and say yes, I agree with that. It's all very simple.

Since we're struggling here to get some insight as to what's going on and some understanding of it, it's very tricky. Now, the other part of Burke's theorem says the state at  $n + \delta$  is independent of departures prior to  $n + \delta$ . And that seems even worse. It says you're looking at this Markov chain at a particular time. And you're saying the state of it is independent of all those departures which happened before that.

That's really saying something. But if you use this reversibility condition that says, when you look at things going from right to left, arrivals become departures and departures become arrivals. Then that statement there is exactly the same as saying the state of a forward going chain at time  $n$  is independent of the arrivals that come after time  $n$ .

Now, you all know that to be true. Because you're all used to looking at these things moving forward in time. So whenever you see a statement like that, just in your head reverse time, or turn your head around so that right becomes left and left

becomes right. And then departures become arrivals and arrivals become departures. You can't do one without the other. You've got to do both of them together.

OK. So everything we know about the MM1 sample time chain has a corresponding statement with time reversed and arrivals and departure switched. So it's not only Burke's theorem. I mean, you can write down 100 theorems now. And they're all the same idea.

But the critical idea is the question that two of you asked. And that is, why is the departure rate going to be  $\lambda$  when you look at things coming in backwards.? And the answer again is that it's  $\lambda$  because we're not conditioning it on knowing what the prior state was. And everything else you know about these things, you always condition things on the prior state. So now we're getting used to conditioning them on the later state.

OK. Let's talk about branching processes. Branching processes have nothing to do with reversibility. Again, these are just very curious kinds of processes. They have a lot to do with all kinds of genetic kinds of things, with lots of physics kinds of experiments.

I don't think a branching process corresponds very closely to any one of those things. This is the same kind of modeling issue that we come up against all the time. What we do is, we pick very, very simple models to try to understand one aspect of a physical problem. And if you try to ask for a model that understands all aspects of that physical problem, you've got a model that's too complicated to say anything about.

But here's a model that says if you believe that one generation to the next, if the only thing that's happening is the individuals in one generation are spawning children or are spawning whatever it is in that next generation, and every individual does this in an independent way, then this is what you have to live with.

I mean that's what the mathematics says. The model is no good, but the

mathematics is fine. So the mathematics is we suppose that  $x$  of  $n$  is the state of the Markov chain at time  $n$ , and the Markov chain is described in the following way.  $x$  of  $n$ , we think of as being the number of elements in generation  $n$ , and for each element  $k$ , out of that  $x$  of  $n$ , each element gives rise to a number of new elements.

And the number of new elements it gives rise to we call it  $y$  sub  $kn$ . The  $n$  at the end is for the generation, the  $k$  is for the particular element in the case generation. So  $y$  sub  $kn$  is the number of offspring of the element  $k$  in the  $n$ -th generation.

After the element in the  $n$ -th generation gives birth, it dies. So it's kind of a cruel world, but that's this particular kind of model. So the number of elements in the  $n$  plus first generation then, is the sum of the number of offspring of the elements in the  $n$ -th generation. So it says  $x$  of  $n$  plus 1 is equal to this  $y$  sub  $kn$  is the number of offspring of element  $k$ , and we sum that number of offspring from 1 to  $x$  of  $n$ , and that's the equation we get.

The assumption we make is that the non-negative integer random variable  $y$  sub  $kn$ -- these random variables-- are independent, and identically distributed over both  $n$  and  $k$ . There's this usual peculiar problem that we have where we're defining random variables that might not exist, but we should be used to that by now.

I mean we just have the random variable there and we pick them out when we need them is the best way to think about that. The initial generation  $x$  of 0 can be an arbitrary positive random variable, but it's usually taken to be  $y$ .

So you start out with one element, and this thing goes on from one generation to the next. It might all die out, or it might continue, it might blow up explosively, and we want to find out which it does. OK so here's the critical equation. Let's look at a couple of examples of why  $y$  sub  $kn$  is deterministic, and equals 1, and  $x_n$  is equal to  $x_{n-1}$  is equal to  $x_0$  for all  $n$  greater than or equal to 1. So this example isn't very interesting.

If  $y$  sub  $kn$  is equal to 2, then each generation has twice as many elements as the previous generation. Each element has two offspring. So you have something that

looks like a tree, which is where the name branching process comes from because people think of these things in terms of trees.

Each one element here two offspring, two elements here, and now if you visualize this kind of chain, you can think of this as being random. So the perhaps in this first generation there are two offspring. Perhaps this one has no offspring the next time, so this dies out. This one has two offspring.

This one has two offspring. this one has no offspring, this one has two. Four, we're up to four. And then all of them die out. So we're talking about that kind of process, which you can visualize as a tree just as easily as you can visualize it this way.

Personally I find it easier to do this as a tree. Because that's personal preference. OK, so just talked about this third kind of animal here. If the probability of no offspring is  $1/2$ , and the probability of twins is  $1/2$ , then  $x_n$ , it's a rather peculiar Markov chain. It can grow explosively, or it can die out.

Who would guess that it's going to grow explosively on the average? And who would guess that it will die out on the average? I mean would anybody hazard to make a guess that this will die out with probability one?

Well it will, and we'll see that today. It can grow for quite a while, but eventually it gets killed. When we look at this process now, the state 0 is trapping state. The states 0 was always a trapping state for branching processes. Because once you get to state 0, there's nothing to have offspring anymore.

So state 0 is always a trapping state. But in other states you can have rather explosive growth. For this particular thing here the even numbered states all communicate, but there are transient. Each odd numbered state doesn't communicate with any other state. As you see from this diagram here, you're always dealing with an even number of states here.

Because each offspring each element has either two or 0 offspring. So you're summing up a bunch of even numbers, and you never get anything odd, except this initial state of one, which you get out of right away.

OK so how do we analyze these things? We want to find the probability for the general case that the process dies out. So let's simplify our notation a little bit. We're going to let the pmf on  $y$ -- we have a pmf on  $y$  because  $y$  is an integer random variable. It's 0, or one, or 2, or so forth.

We'll call that  $p$  sub of  $k$ . For the Markov chain namely  $x_0, x_1, x_2$ , and so forth, we're going to let piece of  $ij$  as usual, be the transition probabilities in the Markov chain. And here it's very useful to talk about the probability the state  $j$  has reached on or before step  $n$ , starting from state  $i$ .

Remember we talked about that-- I forget whether we talked about last time, or the time before-- but you can look it up, and what it is. The thing we derive for it is the probability that you will have touched state  $j$  in one of the  $n$  previous tries starting in state  $i$ . It's  $p$  sub of  $ij$ .

That's the probability you reach it right away so you're successful. And then for everything else you might reach on the first try, there's the probability of going to that state, initially. So this is what happens in the first trial. And then there's the probability you will have gone from state  $k$  to state  $j$ , in any one of the  $n$  minus 1 states after that.

So  $f_{ij}$  of one is  $p_{ij}$ . And now what we're interested in is we start with one element, and we're interested in the probability that it dies out, before it explodes. Or just the probability it dies out. So  $f_{1,0}$  of  $n$  is the probability starting in state 1 that you're going to reach state 0 after  $n$  steps.

So it's  $p_0$  plus sum here of  $p$  sub  $k$  probability you go immediately to state  $k$ , and now here's the only hard thing about this. What I claim now is if we go to state  $k$ , and state  $k$  we have  $k$  elements in this first generation. Now what's the probability that starting with  $k$  elements we're going to be dead after  $n$  minus 1 transitions? Well to be dead after  $n$  minus 1 transitions every one of these elements has to die out.

And they're independent. Everything that's going on from each element is

independent of everything from each other element. So the probability that this first one dies out is  $f_{10}$  over  $n$  minus 1 steps. Probability the second one dies out-- same thing. you take the product of them.

So this is the probability that we die out initially, this sum from  $k$  equals 1 to infinity. Is the probability that each of the  $k$  descendants dies out within time  $n$  minus 1. We can take a  $p_0$  into the sum here, we sum from 0 up to infinity, because  $s_{10}$  to the 0 is just equal to 1. So we get this nice looking formula here.

Let me do this quickly, and then we'll go back and talk about it. Let's talk about the  $z$  transform of this birth process. OK so we have this discrete random variable  $y$  with pmf  $p_{k|h}$  of  $z$  is the sum over  $k$  if  $p_{k|h}$  times  $z$  to the  $k$ .

It's just another kind of transform, we have all kinds of transforms in this course. And this is one transform. Given the state of a pmf, you can define a function into  $z$  in this way.

So  $f_{10}$  of  $n$  is then equal to  $h$  of  $f_{10}$  of  $n$  minus 1. It's amazing that all this mess turns into something that looks so simple. So this will be the probability that we will die out by time end, if in fact we know what the probability of dying out at time  $n$  minus 1 is.

So let's try to solve this equation. And it's not hard to solve as you would think. There's this  $z$  transform  $h$  of  $z$ ,  $h$  of  $z$  is given there. What do I know about  $h$  of  $z$ ?

I know its value, it's  $z$  equal to 1. Because it's  $z$  equal to 1, I'm just summing  $p_{k|h}$  times 1. So  $h$  of 1 is equal to 1. That's what this is in both cases here. What else do I know about it?

$h$  of 0 is equal to  $p_0$ . And if you take the second derivative of this, you find out immediately that the second derivative is positive. So this curve, this convex, it goes like that. As it's been drawn here. The other thing we know is that this derivative at 1 the derivative of  $h$  of  $z$  is equal to the sum of  $k$  times  $p$  to the  $k$ , times  $k$ , times  $z$  to the  $k$  minus 1.

I set  $z$  equal to 1, and what is that? It's the sum of  $p_k$  times  $k$ . So this derivative here is  $\bar{y}$ . In this case, I'm looking at a case where  $\bar{y}$  is equal to 1, in this case, I'm looking at a case where  $\bar{y}$  is bigger than 1.

Everybody with me so far?

**AUDIENCE:** So what is the  $\bar{y}$ ?

**PROFESSOR:**  $\bar{y}$  is the expected value of the random variable  $y$ . And the random variable  $y$  is the number of offspring than any one element will have. The whole thing is defined in terms of this  $y$ . I mean it's the only thing that I've given you except all these independence conditions.

It's like the Poisson process. There's only one element in it, which is  $\lambda$ . A complicated process, but it's defined in terms of everything being independent of everything else. And that's the same thing kind of thing we have here.

Well what I've drawn here is a graphical way of finding what  $f_1(0)$  of 1,  $f_1(0)$  of 2,  $f_1(0)$  of 3 is and so forth. And we start out with one element here, and I want to find  $h$  of  $f_1(0)$  of one is  $h$  of  $f_1(0)$  of 0.

What is  $f_1(0)$  of 0? It has to be 1. So  $f_1(0)$  of  $n$  is just equal to  $h$  of  $p_0$ . So I trace from here, from  $p_0$  over to this slope of one. So this is  $p_0$  down here, and this point here is  $f_1(0)$  of 1, at this point.

Starting here I go over to here, and down here this is  $f_1(0)$  of one, as advertised. I can move up to the curve and I get  $h$  of 2.  $h$  of  $f_1(0)$   $h$  of  $z$ , where  $z$  is equal to  $f_1(0)$  of 1. And so forth along here. I'm not going to spend a lot of time explaining the graphical procedure, because this is something that you look at on your own, and you sort it out in two minutes, and if I explained it, I mean you'll be looking at it at a different speed and I'm explaining it at, so it won't work.

But what happens is starting out with  $\sum p_0$ , you just move along each of these points are  $f_1(0)$  of 1,  $f_1(0)$  of 2,  $f_1(0)$  of 3, up to  $f_1(0)$  of infinity. This is the probability that the process will die out eventually.

So it's the point at which  $h$  of  $z$  equals  $z$ . That's the root of the equation,  $h$  of  $z$  equals  $z$ . We already know what that is pretty much, because we know that we're looking at it case here where  $\bar{y}$  is greater than 1.

So it means the slope here is bigger than 1. We have a convex curve which starts on this side of this line, that ends on the other side of this line. There's got to be a root in the middle. And there can only be one root, so we eventually get to that point. And that's the probability of dying out.

Now, over in this case,  $\bar{y}$  is equal to 1. Or I could look at a case  $\bar{y}$  is less than 1, And what happens then? Keep moving around the same way, I got up at that point, and in fact  $h$  1 0 of infinity is equal to 1. Which says the probability of dying out is equal to 1.

These things that I'm calculating here are in fact the probability of dying out by time 1, the probability of dying out by time 2, and so forth all the way up. In here we start out on this side of the curve. We keep getting crunched in.

We wind up at that point, and in this case, we keep getting crunched up, and we wind up at that point. So the general behavior of these branching processes is so long as there's a possibility of an element having no children, there's a possibility that the whole process will die out.

But if the expected number of offspring is greater than 1, then that probability of dying out is less than 1. Unless the expected number of offspring is less than or equal to 1, then the probability of dying out is in fact equal to 1.

So that was just this graphical picture, and that does the whole thing, and if you think about it for 10 minutes in a quiet room, I think it will be obvious to you, because there's no rocket science here. It's just a simple graphical argument. I have to think about it every time I do it, because it always looks implausible.

So it says the process can explode if the expected number of elements from each element is larger than 1. But it doesn't have to explode. There's an interesting

theorem that we'll talk about when we start talking about martingales. And that is that the number of elements in generation  $n$  divided by the expected value of  $y$  to the  $n$ -th power  $x_{sub n}$  divided by  $y\text{-bar}$  to the  $n$ -th power.

This is something that looks like it ought to be kind of stable. And it says that this approach is a random variable. Namely with probability 1, this has some random value that you can calculate. With a certain probability, this is equal to 0. With a certain probability this is some larger constant. And it can be any old constant at all with different probabilities.

And you can sort of see why this is happening. Suppose you have this process. Suppose  $y\text{-bar}$  is bigger than 1. Suppose it's equal to 2 for example. So the expected number of offspring of each of these elements is two, so the number of offspring of 10 to the 6 elements is 2 times 10 to the sixth.

What this is doing is dividing by that multiplying factor. What's going to happen then is after a certain amount of time, you have so many elements, and each one of them is doing something independently, so the number of offspring in each generation divided by an extra  $y\text{-bar}$  is almost constant. And that's what this theorem is saying.

So that after a while it says the growth rate becomes fixed. And that sort of obvious intuitively.

That's enough for that. Should've not been so talkative about the earlier things. But Markov processes turn out to be pretty simple, given what we know about Markov chains. There's not a lot of new things to be learned here. Just a few. Accountable state Markov process is most easily viewed as a simple extension of accountable state Markov chain.

And along with each state in the Markov chain, there's a holding time. So what happens in this process is it goes along. At a certain point there's a state change. The state change is according to the Markov chain, and amount of time that it takes is an exponential random variable which depends on the state you are in.

So in some states you move quickly. In some states you move slowly. But the only thing that's going on is you have a Markov chain, and each state of the Markov chain, there's some rate which determines how long it's going to take to get to the next state change.

So that you can visualize what the process looks like-- This is the state at times 0. This determines some holding time  $u_1$ . It also determines some state at time 1. The state you go to is independent of how long it takes you to get there. This then determines the rate, so it tells you the rate of this exponential random variable.

And we have this plus some process leading off plus we have this Markov process leading along here, and for each state of the Markov process, you have this holding time. You will ask-- as I do every time I look at this-- why did I make this  $u_1$  instead of  $u_0$ ? It's because of the next slide. OK? Here's the next slide, which shows what's going on.

So we start off at time 0,  $x$  of 0 is in some state  $i$ . We stay in state  $i$  until some time  $u_1$ , at which the state changes. The state change now is some state  $j$ . We stay in that same state  $j$  until the next state change. We stay in that state until the next state change, and it is since we want to make the first state change time  $s$  of one, we sort of have to make the first interval between 0 on the state  $u_1$ .

So these things are off base from the  $u$ 's. And this is the way that a way that a Markov process evolves. You simply have what looks like a Poisson process, a variable rate, and the variable rate is varying according to the state of a Markov chain every time you have an arrival in the variable rate Poisson process, you change the rate according to this Markov process.

So it's everything about Markov chains, plus Poisson processes all put together. OK, I think I'll stop there, and we will continue next time.