6.041SC Probabilistic Systems Analysis and Applied Probability, Fall 2013 Transcript – Lecture 18

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.

JOHN TSITSIKLIS: So what we're going to do is to review what we have discussed last time. Then we're going to talk about the classic application of Markov chains to analyze how do you dimension a phone system. And finally, there will be two new things today. We will see how we can calculate certain interesting quantities that have to do with Markov chains. So let us start.

We've got our Markov chain and let's make the assumption that our chain is kind of nice. And by nice we mean that we've got maybe some transient states. And then we've got a single recurrent class of recurrent states. So this is a single recurrent class in the sense that from any state in that class you can get to any other state. So once you're in here you're going to circulate and keep visiting all of those states. Those states appear transient. The trajectory may move around here, but eventually one of these transitions will happen and you're going to end up in this lump.

Let's make the assumption that the single recurrent class is not periodic. These are the nicest kind of Markov chains. And they're nicest because they have the following property, the probability that you find yourself at some particular state j at the time n when that time is very large. That probability settles to a steady state value that we denote by pi sub j. And there are two parts in the statement. One part is that this limit exists. So the probability of state j settles to something, and furthermore that probability is not affected by i. It doesn't matter where you started, no matter where you started, the probability of state j is going to be the same in the long run. Maybe a clearer notation could be of this form. The probability of being at state j given the initial state being i is equal to pi(j) in the limit.

Now, if I don't tell you where you started and you look at the unconditional probability of being at state i, you can average over the initial states, use the total expectation theorem and you're going to get the same answer pi(j) in the limit. So this tells you that to the conditional probability given the initial state in the limit is the same as the unconditional probability. And that's a situation that we recognize as being one where we have independence. So what this result tells us is that Xn and Xi are approximately independent. They become independent in the limit as n goes to infinity. So that's what the steady state theorem tells us. The initial conditions don't matter, so your state at some large time n has nothing to do, is not affected by what your initial state was. Knowing the initial state doesn't tell you anything about your state at time n, therefore the states at the times-- sorry that should be a 1, or it should be a 0 -- so the state is not affected by where the process started.

So if the Markov chain is to operate for a long time and we're interested in the question where is the state, then your answer would be, I don't know, it's random. But it's going to be a particular j with this particular probability. So the steady state probabilities are interesting to us and that

raises the question of how do we compute them. The way we compute them is by solving a linear system of equations, which are called the balance equations, together with an extra equation, the normalization equation that has to be satisfied by probability, because probabilities must always add up to 1.

We talked about the interpretation of this equation last time. It's basically a conservation of probability flow in some sense. What comes in must get out. The probability of finding yourself at state j at a particular time is the total probability of the last transition taking me into state j. The last transition takes me into state j in various ways. It could be that the previous time I was at the particular state, j and i made a transition from k into j. So this number here, we interpret as the frequency with which transitions of these particular type k to j, occur. And then by adding over all k's we consider transitions of all types that lead us inside state j. So the probability of being at the j is the sum total of the probabilities of getting into j.

What if we had multiple recurrent classes? So if we take this picture and change it to this. So here we got a secondary recurrent class. If you're here, you cannot get there. If you are here, you cannot get there. What happens in the long run? Well, in the long run, if you start from here you're going to make a transition eventually, either of this type and you would end up here, or you will make a transitional of that type and you will end up there. If you end up here, the long term statistics of your chain, that is, the probabilities of the different states, will be the steady state probabilities of this chain regarded in isolation. So you go ahead and you solve this system of equations just for this chain, and these will be your steady state probabilities. If you happened to get in here.

If, on the other hand, it happens that you went there, given that event, then what happens in the long run has to do with just this chain running by itself. So you find the steady state probabilities inside that sub chain. So you solve the linear system, the steady state equations, for this chain separately and for that chain separately. If you happen to start inside here then the steady state probabilities for this sub chain are going to apply.

Now of course this raises the question, if I start here, how do I know whether I'm going to get here or there? Well, you don't know, it's random. It may turn out that you get to here, it may turn out that you get there. So we will be interested in calculating the probabilities that eventually you end up here versus the probability that eventually you end up there. This is something that we're going to do towards the end of today's lecture.

So, as a warm up, just to see how we interpret those steady state probabilities, let us look at our familiar example. This is a 2-state Markov chain. Last time we did write down the balance equations for this chain and we found the steady state probabilities to be 2/7 and 5/7 respectively. So let us try to calculate some quantities.

Suppose that you start at state 1, and you want to calculate to this particular probability. So since we're assuming that we're starting at state 1, essentially here we are conditioning on the initial state being equal to 1. Now the conditional probability of two things happening is the probability that the first thing happens. But we're living in the world where we said that the initial state was

1. And then given that this thing happened, the probability that the second thing happens. But again, we're talking about conditional probabilities given that the initial state was 1.

So what is this quantity? This one is the transition probability from state 1 to state 1, so it's P11. How about the second probability? So given that you started at 1 and the next time you were at 1, what's the probability that at the time 100 you are at 1? Now because of the Markov property, if I tell you that at this time you are at 1, it doesn't matter how you get there. So this part of the conditioning doesn't matter. And what we have is the 99 step transition probability from state 1 to state 1.

So the probability that you get to 1 and then 99 steps later you find yourself again at one is the probability that the first transition takes you to 1 times the probability that over the next 99 transitions starting from 1, after 99 steps you end up again at state 1. Now, 99 is possibly a big number, and so we approximate this quantity. We're using the steady state probability of state 1. And that gives us an approximation for this particular expression.

We can do the same thing to calculate something of the same kind. So you start at state 1. What's the probability that 100 steps later you are again at state 1? So that's going to be P11-- not P -- R11. The 100 step transition probability that starting from 1 you get to 1, and then after you get to 1 at time 100 what's the probability that the next time you find yourself at state 2? This is going to be the probability P12. And approximately, since 100 is a large number, this is approximately pi(1) times P12.

OK. So that's how we can use steady state probabilities to make approximations. Or you could, for example, if you continue doing examples of this kind, you could ask for what's the probability that X at time 100 is 1, and also X at time 200 is equal to 1. Then this is going to be the transition probability from 1 to 1 in 100 steps, and then over the next 100 steps from 1 you get again to 1. And this is going to be approximately pi(1) times pi(1).

So we approximate multi-step transition probabilities by the steady state probabilities when the number n that's involved in here is big. Now I said that's 99 or 100 is big. How do we know that it's big enough so that the limit has taken effect, and that our approximation is good? This has something to do with the time scale of our Markov chain, and by time scale, I mean how long does it take for the initial states to be forgotten. How long does it take for there to be enough randomness so that things sort of mix and it doesn't matter where you started?

So if you look at this chain, it takes on the average, let's say 5 tries to make a transition of this kind. It takes on the average 2 tries for a transition of that kind to take place. So every 10 time steps or so there's a little bit of randomness. Over 100 times steps there's a lot of randomness, so you expect that the initial state will have been forgotten. It doesn't matter. There's enough mixing and randomness that happens over 100 time steps. And so this approximation is good.

On the other hand, if the numbers were different, the story would have been different. Suppose that this number is 0.999 and that number is something like 0.998, so that this number becomes 0.002, and that number becomes 0.001. Suppose that the numbers were of this kind. How long does it take to forget the initial state? If I start here, there's a probability of 1 in 1,000 that next

time I'm going to be there. So on the average it's going to take me about a thousand tries just to leave that state. So, over roughly a thousand time steps my initial state really does matter.

If I tell you that you started here, you're pretty certain that, let's say over the next 100 time steps, you will still be here. So the initial state has a big effect. In this case we say that this Markov chain has a much slower time scale. It takes a much longer time to mix, it takes a much longer time for the initial state to be forgotten, and this means that we cannot do this kind of approximation if the number of steps is just 99. Here we might need n to be as large as, let's say, 10,000 or so before we can start using the approximation.

So when one uses that approximation, one needs to have some sense of how quickly does the state move around and take that into account. So there's a whole sub-field that deals with estimating or figuring out how quickly different Markov chains mix, and that's the question of when can you apply those steady state approximations.

So now let's get a little closer to the real world. We're going to talk about a famous problem that was posed, started, and solved by a Danish engineer by the name of Erlang. This is the same person whose name is given to the Erlang distribution that we saw in the context of the Poisson processes. So this was more than 100 years ago, when phones had just started existing. And he was trying to figure out what it would take to set up a phone system that how many lines should you set up for a community to be able to communicate to the outside world.

So here's the story. You've got a village, and that village has a certain population, and you want to set up phone lines. So you want to set up a number of phone lines, let's say that number is B, to the outside world. And how do you want to do that? Well, you want B to be kind of small. You don't want to set up too many wires because that's expensive. On the other hand, you want to have enough wires so that if a reasonable number of people place phone calls simultaneously, they will all get a line and they will be able to talk.

So if B is 10 and 12 people want to talk at the same time, then 2 of these people would get a busy signal, and that's not something that we like. We would like B to be large enough so that there's a substantial probability, that there's almost certainty that, under reasonable conditions, no one is going to get a busy signal.

So how do we go about modeling a situation like this? Well, to set up a model you need two pieces, one is to describe how do phone calls get initiated, and once a phone call gets started, how long does it take until the phone call is terminated? So we're going to make the simplest assumptions possible.

Let's assume that phone calls originate as a Poisson process. That is, out of that population people do not really coordinate. At completely random times, different people with decide to pick up the phone. There's no dependencies between different people, there's nothing special about different times, different times are independent. So a Poisson model is a reasonable way of modeling this situation. And it's going to be a Poisson process with some rate lambda.

Now, the rate lambda would be easy to estimate in practice. You observe what happens in that village just over a couple of days, and you figure out what's the rate at which people attempt to place phone calls. Now, about phone calls themselves, we're going to make the assumption that the duration of a phone call is a random variable that has an exponential distribution with a certain parameter mu. So 1/mu is the mean duration of a phone call. So the mean duration, , again, is easy to estimate. You just observe what's happening, see on the average how long these phone calls are.

Is the exponential assumption a good assumption? Well, it's means that most phone calls will be kind of short, but there's going to be a fraction of phone calls that are going to be larger, and then a very small fraction that are going to be even larger. So it sounds plausible. It's not exactly realistic, that is, phone calls that last short of 15 seconds are not that common. So either nothing happens or you have to say a few sentences and so on.

Also, back into the days when people used to connect to the internet using dial up modems, that assumption was completely destroyed, because people would dial up and then keep their phone line busy for a few hours, if the phone call was a free one. So at those times the exponential assumption for the phone call duration was completely destroyed. But leaving that detail aside, it's sort of a reasonable assumption to just get started with this problem.

All right, so now that we have those assumptions, let's try to come up with the model. And we're going to set up a Markov process model. Now the Poisson process runs in continuous time, and call durations being exponential random variables also are continuous random variables, so it seems that we are in a continuous time universe. But we have only started Markov chains for the discrete time case.

What are we going to do? We can either develop the theory of continuous time Markov chains, which is possible. But we are not going to do that in this class. Or we can discretize time and work with a discrete time model. So we're going to discretize time in the familiar way, the way we did it when we started the Poisson process. We're going to take the time axis and split it into little discrete mini slots, where every mini slot has a duration delta. So this delta is supposed to be a very small number.

So what is the state of the system? So, you look at the situation in the system at some particular time and I ask you what is going on right now, what's the information you would tell me? Well, you would tell me that right now out of these capital B lines, 10 of them are busy, or 12 of them are busy. That describes the state of the system, that tells me what's happening at this point. So we set up our states base by being the numbers from 0 to B. 0 corresponds to a state in which all the phone lines are free, no one is talking. Capital B corresponds to a case where all the phone lines are busy. And then you've got states in between.

And now let's look at the transition probabilities. Suppose that right so now we have i-1 lines that are busy. Or maybe, let me look here. Suppose that there's i lines that are busy. What can happen the next time? What can happen is that the new phone call gets placed, in which case my state moves up by 1, or an existing call terminates, in which case my state goes down by 1, or none of the two happens, in which case I stay at the same state. Well, it's also possible that the phone call

gets terminated and a new phone call gets placed sort of simultaneously. But when you take your time slots to be very, very small, this is going to have a negligible probability order of delta squared, so we ignore this.

So what's the probability of an upwards transition? That's the probability that the Poisson process records an arrival during a mini slot of duration delta. By the definition of the Poisson process, the probability of this happening is just lambda delta. So each one of these upwards transitions has the same probability of lambda delta. So you've got lambda deltas everywhere in this diagram.

How about, now, phone call terminations? If you had the single call that was active, so if you were here, what's the probability that the phone call terminates? So the phone call has an exponential duration with parameter mu. And we discussed before that an exponential random variable can be thought of as the first arrival time in a Poisson process. So the probability that you get this event to happen over a delta time interval is just mu times delta.

So if you have a single phone call that's happening right now, with probability mu times delta, that call is going to terminate. But suppose that we have i phone calls that are currently active. Each one of them has a probability of mu delta, of terminating, but collectively the probability that one of them terminates becomes i times mu delta. So that's because you get the mu delta contribution -- the probability of termination from each one of the different phone calls.

OK, now this is an approximate calculation, because it ignores the possibility that two phone calls terminate at the same time. Again, the way to think of why this is the correct rate, when you have i phone calls that are simultaneously running and waiting for one of them to terminate, this is like having i separate Poisson processes that are running in parallel, and you ask for the probability that one of those processes records an event. Now when you put all those process together, it's like having a Poisson process with total rate i times mu, and so i times mu delta is the overall probability that something happens in terms of phone call terminations at those times.

So in any case, this is the transition probability for downwards transitions. Now that we've got this, we can analyze this chain. This chain has the birth death form that we discussed towards the end of last lecture. And for birth death chains, it's easy to write it out to find the steady state probabilities.

Instead of writing down the balance equations in the general form, we think in terms of a conservation of probabilities or of transitions by looking at what happens across a particular cut in this diagram. Number of transitions in the chain that cross from here to here has to be approximately equal to the number of transitions from here to there because whatever comes up must come down and then come up and so on. So the frequency with which transitions of this kind are observed has to be the same as the frequency of transitions of this kind.

What's the frequency of how often the transitions of this kind happen? And by frequency I mean quite percentage of the mini slots involve a transition of this kind? Well, for a transition of that kind to happen we need to be at states i-1, which happens this much of the time. And then the probability lambda delta that the transition is of this kind.

So the frequency of transitions of with which this kind of transition is observed is lambda delta times pi(i-1). This is the fraction of time steps at which a transition from specifically this state to specifically that state are observed. This has to be the same as the frequency with which transitions of that kind are observed, and that frequency is going to be i mu delta times pi(i), and then we cancel the deltas, and we are left with this equation here.

So this equation expresses pi(i) in terms of pi(i-1). So if we knew pi(0) we can use that equation to determine pi(1). Once we know pi(1), we can use that equation to determine pi(2), and so on, you keep going. And the general formula that comes out of this, I will not do the algebra, it's a straightforward substitution, you find that pi(i), the steady state probability of state i is given by this expression, which involves the pi(0) from which we started.

Now what is pi(0)? Well, we don't know yet, but we can find it by using the normalization equation. The sum of pi(i) has to be equal to 1. So the sum of all of those numbers has to be equal to 1. And the only way that this can happen is by setting pi(0) to be equal to that particular number. So if I tell you the value of capital B, you can set up this Markov chain, you can calculate pi(0), and then you can calculate pi(i), and so you know what fraction, you know the steady state probabilities of this chain, so you can answer the question. If I drop in at a random time, how likely is it that I'm going to find the states to be here, or the states to be there? So the steady state probabilities are probabilities, but we also interpret them as frequencies. So once I find pi(i), it also tells me what fraction of the time is the state equal to i. And you can answer that question for every possible i.

Now, why did we do this exercise? We're interested in the probability of the system is busy. So if a person, a new phone call gets placed, it just drops out of the sky. According to that Poisson process, that new phone call is going to find the system at a random state. That random state is described in steady state by the probabilities pi(i)'s.

And the probability that you find the system to be busy is the probability that when you drop in the state happens to be that particular number B. So i sub b is the probability of being busy. And this is the probability that you would like to be small in a well engineered system. So you ask the question, how should, given my lambda and mu, my design question is to determine capital B the number of phone lines so that this number is small.

Could we have done, could we figure out a good value for B by doing a back of the envelope calculation? Let's suppose that lambda is 30 and that mu is 1/3. So I guess that's, let us these rates to be calls per minute. And this mu, again, is a rate per minute. Again, the units of mu are going to be calls per minute. So since our time unit is minutes, the mean duration of calls is 1/mu minutes.

So a typical call, or on the average a call lasts for 3 minutes. So you get 30 calls per minute. Each call lasts for 3 minutes on the average. So on the average, if B was infinite, every call goes through. How many calls would be active on the average? So you get 30 per minute. If a call lasted exactly 1 minute, then at any time you would have 30 calls being active. Now a call lasts on the average for 3 minutes. So during each minute you generate 90 minutes of talking time.

So by thinking in terms of averages you would expect that at any time there would be about 90 calls that are active. And if 90 calls are active on the average, you could say OK, I'm going to set up my capital B to be 90. But that's not very good, because if the average number of phone calls that want to happen is if the average number is 90, sometimes you're going to have 85, sometimes you will have 95. And to be sure that the phone calls will go through you probably want to choose your capital B to be a number a little larger than 90. How much larger than 90? Well, this is a question that you can answer numerically.

So you go through the following procedure. I tried different values of capital B. For any given value of capital B, I do this numerical calculation, I find the probability that the system is busy, and then I ask what's the value of B that makes my probability of being busy to be, let's say, roughly 1 %. And if you do that calculation with the parameters that they gave you, you find that B would be something like 106.

So with the parameters they gave where you have, on the average, 90 phone calls being active, you actually need some margin to protect against the [?] fluctuation, if suddenly by chance more people want to talk, and if you want to have a good guarantee that an incoming person will have a very small probability of finding a busy system, then you will need about 106 phone lines.

So that's the calculation and the argument that the Erlang went through a long time ago. It's actually interesting that Erlang did this calculation before Markov chains were invented. So Markov's work, and the beginning of work on Markov chains, happens about 10-15 years after Erlang. So obviously he didn't call that a Markov chain. But it was something that he could study from first principles. So this is a pretty useful thing.

These probabilities that come out of that model, at least in the old days, they would all be very well tabulated in handbooks that every decent phone company engineer would sort of have with them. So this is about as practical as it gets. It's one of the sort of standard real world applications of Markov chains.

So now to close our subjects, we're going to consider a couple of new skills and see how we can calculate the few additional interesting quantities that have to do with the Markov chain. So the problem we're going to deal with here is the one I hinted that when I was talking about this picture. You start at a transient state, you're going to eventually end up here or there. We want to find the probabilities of one option of the two happening or the other happening.

So in this picture we have a class of states that's are transient. These are transient because you're going to move around those states, but there's a transition that you can make, and you go to a state from which you cannot escape afterwards. Are you going to end up here or are you going to end up there? You don't know. It's random. Let's try to calculate the probability that you end up at state 4.

Now, the probability that you end up at state 4 will depend on where you start. Because if you start here, you probably have more chances of getting to 4 because you get that chance immediately, whereas if you start here there's more chances that you're going to escape that way because it kind of takes you time to get there. It's more likely that you exit right away.

So the probability of exiting and ending up at state 4 will depend on the initial state. That's why when we talk about these absorption probability we include an index i that tells us what the initial state is. And we want to find this absorption probability, the probability that we end up here for the different initial states.

Now for some initial states this is very easy to answer. If you start at state 4, what's the probability that eventually you end up in this part of the chain? It's 1. You're certain to be there, that's where you started. If you start at state 5, what's the probability that you end up eventually at state 4? It's probability 0, there's no way to get there. Now, how about if you start at a state like state 2? If you start at state 2 then there's a few different things that can happen. Either you end up at state 4 right away and this happens with probability 0.2, or you end up at state 1, and this happens with probability 0.6. So if you end up at state 4, you are done. We are there.

If you end up at state 1, then what? Starting from state 1 there's two possibilities. Either eventually you're going to end up at state 4, or eventually you're going to end up at state 5. What's the probability of this happening? We don't know what it is, but it's what we defined to be a1. This is the probability -- a1 is the probability -- that eventually you settle in state 4 given that the initial state was 1. So this probability is a1.

So our event of interest can happen in two ways. Either I go there directly, or I go here with probability 0.6. And given that I go there, eventually I end up at state 4, which happens with probability a1. So the total probability of ending up at state 4 is going to be the sum of the probabilities of the different ways that this event can happen. So our equation, in this case, is going to be, that's a2, is going to be 0.2 (that's the probability of going there directly) plus with probability 0.8 I end up at state 1, and then from state 1 I will end up at state 4 with probability a1. So this is one particular equation that we've got for what happens if we start from this state.

We can do a similar argument starting from any other state. Starting from state i the probability that eventually I end up at state 4 is, we consider the different possible scenarios of where do I go next, which is my state j, with probability Pij. Next time I go to j, and given that I started at j, this is the probability that I end up at state 4.

So this equation that we have here is just an abstract version in symbols of what we wrote down for the particular case where the initial state was 2. So you write down an equation of this type for every state inside here. You'll have a separate equation for a1, a2, and a3. And that's going to be a system of 3 equations with 3 unknowns, the a's inside the transient states. So you can solve that 3 by 3 system of equations. Fortunately, it turns out to have a unique solution, and so once you solve it you have found the probabilities of absorption and the probability that eventually you get absorbed at state 4.

Now, in the picture that we had here, this was a single state, and that one was a single state. How do things change if our recurrent, or trapping sets consist of multiple states? Well, it doesn't really matter that we have multiple states. All that matters is that this is one lump and once we get there we are stuck in there.

So if the picture was, let's say, like this, 0.1 and 0.2, that basically means that whenever you are in that state there's a total probability of 0.3 of ending in that lump and getting stuck inside that lump. So you would take that picture and change it and make it instead a total probability of 0.3, of ending somewhere inside that lump. And similarly, you take this lump and you view it as just one entity, and from any state you record the total probability that given that I'm here I end up in that entity. So basically, if the only thing you care is the probability that you're going to end up in this lump, you can replace that lump with a single state, view it as a single state, and calculate probabilities using this formula.

All right, so now we know where the chain is going to get to. At least we know probabilistically. We know with what probability it is going to go here, and that also tells us the probability that eventually it's going to get there. Other question, how long is it going to take until we get to either this state or that state? We can call that event absorption, meaning that the state got somewhere into a recurrent class from which it could not get out.

Okay. Let's deal with that question for the case where we have only 1 absorbing state. So here our Markov chain is a little simpler than the one in the previous slide. We've got our transient states, we've got our recurrent state, and once you get into the recurrent state you just stay there. So here we're certain that no matter where we start we're going to end up here. How long is it going to take? Well, we don't know. It's a random variable. The expected value of that random variable, let's call it mu. But how long it takes to get there certainly depends on where we start. So let's put in our notation again this index i that indicates where we started from.

And now the argument is going to be of the same type as the one we used before. We can think in terms of a tree once more, that considers all the possible options. So suppose that you start at state 1. Starting from state 1, the expected time until you end up dropping states is mu1. Now, starting from state 1, what are the possibilities? You make your first transition, and that first transition is going to take you either to state 2 or to state 3. It takes you to state 2 with probability 0.6, it takes you to state 3 with probability 0.4.

Starting from state 2, eventually you're going to get to state 4. How long does it take? We don't know, it's a random variable. But the expected time until this happens is mu2. Starting from state 2, how long does it take you to get to state 4. And similarly starting from state 3, it's going to take you on the average mu3 time steps until you get to state 4.

So what's the expected value of the time until I end at state 4? So with probability 0.6, I'm going to end up at state 2 and from there on it's going to be expected time mu2, and with probability 0.4 I'm going to end up at state 3, and from there it's going to take me so much time. So this is the expected time it's going to take me after the first transition. But we also spent 1 time step for the first transition. The total time to get there is the time of the first transition, which is 1, plus the expected time starting from the next state. This expression here is the expected time starting from the next state, but we also need to account for the first transition, so we add 1. And this is going to be our mu1.

So once more we have a linear equation that ties together the different mu's. And the equation starting from state 4 in this case, of course is going to be simple, starting from that state the

expected number of steps it takes you to get there for the first time is of course, 0 because you're already there. So for that state this is fine, and for all the other states you get an equation of this form.

Now we're going to have an equation for every state. It's a system of linear equations, once more we can solve them, and this gives us the expected times until our chain gets absorbed in this absorbing state. And it's nice to know that this system of equations always has a unique solution. OK so this was the expected time to absorption. For this case where we had this scene absorbing state. Suppose that we have our transient states and that we have multiple recurrent classes, or multiple absorbing states.

Suppose you've got the picture like this. And we want to calculate the expected time until we get here or there. Expected time until we get to an absorbing state. What's the trick? Well, we can lump both of these states together and think of them as just one bad state, one place for which we're interested in how long it takes us to get there. So lump them as one state, and accordingly kind of merge all of those probabilities.

So starting from here, my probability that the next I end up in this lump and they get absorbed is going to be this probability plus that probability. So we would change that picture. Think of this as being just one big state. And sort of add those two probabilities together to come up with a single probability, which is the probability that starting from here next time I find myself at some absorbing state. So once you know how to deal with a situation like this, you can also find expected times to absorption for the case where you've got multiple absorbing states. You just lump all of those multiple absorbing states into a single one.

Finally, there's a kind of related quantity that's of interest. The question is almost the same as in the previous slide, except that here we do not have any absorbing states. Rather, we have a single recurrent class of states. You start at some state i. You have a special state, that is state s. And you ask the question, how long is it going to take me until I get to s for the first time? It's a single recurrent class of states. So you know that the state keeps circulating here and it keeps visiting all of the possible states. So eventually this state will be visited. How long does it take for this to happen?

Ok. So we're interested in how long it takes for this to happen, how long it takes until we get to s for the first time. And we don't care about what happens afterwards. So we might as well change this picture and remove the transitions out of s and to make them self transitions. Is the answer going to change? No. The only thing that we changed was what happens after you get to s. But what happens after you get to s doesn't matter. The question we're dealing with is how long does it take us to get to s. So essentially, it's after we do this transformation -- it's the same question as before, what's the time it takes until eventually we hit this state. And it's now in this new picture, this state is an absorbing state.

Or you can just think from first principles. Starting from the state itself, s, it takes you 0 time steps until you get to s. Starting from anywhere else, you need one transition and then after the first transition you find yourself at state j with probability Pij and from then on you are going to

take expected time Tj until you get to that terminal state s. So once more these equations have a unique solution, you can solve them and find the answer.

And finally, there's a related question, which is the mean recurrence time of s. In that question you start at s, the chain will move randomly, and you ask how long is it going to take until I come back to s for the next time. So notice the difference. Here we're talking the first time after time 0, whereas here it's just the first time anywhere. So here if you start from s, Ts* is not 0. You want to do at least one transition and that's how long it's going to take me until it gets back to s.

Well, how long does it take me until I get back to s? I do my first transition, and then after my first transition I calculate the expected time from the next state how long it's going to take me until I come back to s. So all of these equations that I wrote down, they all kind of look the same. But they are different. So you can either memorize all of these equations, or instead what's better is to just to get the basic idea. That is, to calculate probabilities or expected values you use the total probability or total expectation theorem and conditional the first transition and take it from there.

So you're going to get a little bit of practice with these skills in recitation tomorrow, and of course it's in your problem set as well.

6.041SC Probabilistic Systems Analysis and Applied Probability
Fall 2013