

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:** So this is the outline of the lecture for today. So we just need to talk about the conditional densities for Poisson process. So these are the things that we've seen before in previous sections, so I'm just going to give a simple proof for them and some intuition. So if you remember the Poisson process, this was the famous figure that we had.

So our time is at zero and afterwards zero and then we had some arrivals. So our time is then  $t$ . We had  $N$  of  $t$  arrivals, which is equal to  $n$ . This is  $S_n$ . This is the figure that we had. So for the first equation, we're looking at-- We want to find the conditional distribution of interval times condition and last arrival times.

So first of all, we want to find the joint distribution of the interval arrival times and the last arrival, this equation. So you know that interval arrival times are independent in the Poisson process, so just looking at this thing, you know that these things are independent of each other so we will have  $\lambda$  to the power of  $N$   $e^{-\lambda}$  to the power minus  $\lambda$ .

And then, this last one corresponds to the fact that  $x_n + 1$  is equal to  $n + 1$  minus summation of  $x$ . These two events are equivalent to each other, so this is going to correspond to  $\lambda e^{-\lambda}$  to the power of minus  $\lambda$ . So we just did it the other way, so it's going to be like-- So you can just get rid of these terms and it's going to be that equation.

So it's very simple, it's just the independence of inter-arrival times. So you look at the last term as the inter-arrival time for the last arrival, and then they're going to be independent, and the terms kind of cancel out and you have this equation. So this conditional density will be the joint distribution-- or the distribution of  $S_n + 1$ ,

which is like these. So the joint decision will be like this.

So what does the intuition behind this mean? This means that-- you've seen this term before. This means that we have uniform distribution over the interval of zero to one, condition on something. So previously, what we had was the conditional distribution of arrival times condition on the last arrival, so  $F$  of  $s_1$  and condition on  $s_{n+1}$ . But here we have distribution of  $x_1$  to  $x_n$  and condition on  $s_{n+1}$ . And previously, the constraint that we had was like we should have ordering over the arrival times. So because of this constraint we had this  $n$  factorial, if you remember, that order of statistics things that you talked about.

Now here, what we have is something like this. So all the arrival times should be positive, and summation of them should be less than  $t$ . Because well, the summation of them to  $n+1$  is equal to  $s_{n+1}$ . And the last term is positive, so this thing should be less than  $t$ . So these two constraints are sort of dual to each other. So because of this constraint, I had some  $n$  factorial over the conditional distribution of arrival times, condition on the last arrival time. Now, I have inter-arrival conditional distribution of inter-arrival time condition on last arrival time and we should have this  $n$  factorial here.

The other interesting thing is that if I condition it under  $n$  of  $t$ , so the number of arrivals at time  $t$ . The time is going to be very similar to the previous one. Just to prove this thing, it's very simple again. I just wanted to show you--

So what does this thing mean? This thing means that these are the distributions of the inter-arrival times and the last event corresponds to the fact that  $x_{n+1}$  is bigger than  $t$  minus summation of  $x_i$ . Because we have an arrival time at  $s_{n+1}$  and then after that there's no arrival. Because at time  $t$  we have  $n$  arrivals. So this corresponds to do is thing. So again, we'll have something like  $\lambda^n e^{-\lambda t}$ . This is for the first event and you see that these are independent because of the Poisson properties of the Poisson process.

And the last term will be-- property of the Poisson and a variable bigger than something is-- Yes, this cancel out. And then we'll have this term, this first  $n$   $\lambda^n$

to the power of  $n$ , just forget about  $n + 1$ . And property distribution of  $n$  of  $t$  equal to  $n$  is this term without  $n + 1$ . So  $\lambda^n$  to the  $1/2$ , and  $e$  to the power minus  $\lambda t$  over  $n$  factorial. So these terms cancel out and we'll have this thing.

So again, we should have these constraints that summation of  $x_k$ ,  $k$  equal to  $1$ ,  $n$  should be less than  $t$ . No matter if they are conditioning of  $n$  of  $t$  or  $s_{n+1}$ , the last arrival time or the number of arrivals at some time instant, the condition of distribution is uniform subject to some constraints. Is there any questions?

So if you look at the limit, and so again, looking at this figure, you see that this corresponds to  $s_{n+1}$  and you see that  $s_{n+1}$  is bigger than  $t$ . So if you take the limit and take the  $t_1$  going to  $t$ , these two distributions are going to be the same. So  $t_1$  corresponds to  $s_{n+1}$ .  $t_1$ . So basically what's it's saying is that conditional distribution of arrival times are uniform, independent of what is going to happen in future.

So in future, there might be some, I mean, the knowledge that we have about future could be the fact that knowing what happens in this interval or the next arrival is going to happen at some point, I mean, condition on these things the arrival times are going to be uniform.

The other fact is that this distribution that I showed you, they are all symmetric with respect to interval arrival times. So it doesn't matter, I mean, no arrival has more priority to any other one or they're not different to each other. So if you look at the permutation of them, they're going to be similar.

So now, I want to look at the distribution of  $x_1$  or  $s_1$ , they are equal to each other. I can easily calculate the probability of  $x_1$  bigger than some  $\tau$  condition, for example,  $n$  of  $t$  is equal to  $n$ . So what is this conditional distribution?

So condition on  $n$  of  $t$  equal to  $n$ , I know that in this interval I have  $n$  arrivals and  $n$  inter-arrival times, and I want to know if  $x_1$  is in this interval, or  $s_1$ .  $s_1$  is equal to  $x_1$ . And you know that  $s_i$ 's are also uniform. So you know that these are independent of each other, and so you can say that the probability of this thing is like this one. So

all of them should be bigger than  $\tau$ .

And since this is symmetric-- so this corresponds to  $x_1$ . Since this was-- everything was symmetric for all inter-arrival times. This is also the complimentary distribution function for  $x_k$ . Is there any questions? So the other think that are not in the slides that I want to tell you is the conditional distribution of  $s_i$  given  $n$  of  $t$ .

So I know that there  $n$  arrivals in this interval, and I want to see if the  $i$  arrival is something like this. So the probability that-- so this even corresponds to the fact that there is one arrival in this interval and  $i$  minus 1 arrival was here and  $n$  minus  $i$  minus 1 arrival was here. Oh, no, sorry,  $n$  minus  $i$  arrival is here.

So the probability of having one arrival here corresponds to-- since arrival times are uniform, and I want to have one arrival in this interval like this, and I want to have  $i$  minus arrivals here  $i$  minus one arrivals here, and  $n$  minus  $i$  arrivals here. So it's like a binomial probability distribution so out of the  $n$  arrival--  $n$  minus 1 remaining arrivals, I want to have some of them in this interval and some of them here. So this is the distribution of this thing.

Yeah, so just as a check we know that the last inter-arrival-- last arrival time, the distribution of that thing is-- Here, you just cancel these things out, too-- is-- We need to get back to this equation, but it's just the same as this one but you let  $i$  equal to  $n$ . This is OK? Just to think  $i$  equal to  $n$ , you get this one.

**AUDIENCE:** Does this equation also apply to  $i$  equals [INAUDIBLE]?

**PROFESSOR:** Yeah, sure. Yeah,  $i$  equal to 1, we will have  $n$  minus  $\tau$ . And if you look at this thing, so this is a complimentary distribution function. You just take to zero till you get that. This is just a check of this formulation. Yeah. So is there any other questions? Did anybody get why we have this kind of formulation? It was very simple, it's just the combination of a uniform distribution and a binomial distribution.

**AUDIENCE:** What is  $\tau$ ?

**PROFESSOR:**  $\tau$  is this thing.  $\tau$  is the  $s_i$ . Change of notation.

**AUDIENCE:** [INAUDIBLE]?

**PROFESSOR:** I said that I want one arrival in this interval and  $i - 1$  arrival was here and  $n - i$  arrival is here. So I define it, the Bernoulli process. So probability of falling in this interval is  $\tau$  over  $t - \tau$ . And I say that--

**AUDIENCE:** [INAUDIBLE]?

**PROFESSOR:** OK, so falling in this interval in a uniform distribution corresponds to success. I want, among  $n - 1$  arrivals,  $i - 1$  of them to be successful and the rest of them to be [INAUDIBLE]. [INAUDIBLE] that one of them should fall here and the rest of them should fall here. Richard?

**AUDIENCE:** So that's the [INAUDIBLE]?

**PROFESSOR:** Yeah, you just get rid of this  $dt$ , that's it.

[? AUDIENCE: Why do ?] those stay,  $n$  over  $t$ .

**PROFESSOR:** Which one?

**AUDIENCE:** Can you explain again the first term? The first term there, the  $n$  over  $t$ ?

**PROFESSOR:** Oh, OK, so I want to have one interval definitely here but among  $n$  arrivals I want them to fall here. So any one of them should cancel. But then  $i - 1$  of them should fall here and  $n - i$  intervals should fall here. Any other questions? It's beautiful, isn't it?

There's a problem. I studied  $x_1$  to  $x_n$ , but what is the distribution of this remaining part? So if I condition on  $n$ ,  $n_t$  or condition on  $s_n + 1$ , what is the distribution of this one or this one or doing distribution off all these? So this is what we're going to look. So first of all, I do the conditioning over  $s_n + 1$ , and I find the distribution of-- the distribution of  $x_1$  to  $x_n$  condition on  $s_n + 1$ , is there any randomness in distribution of  $x_n + 1$ ? So this is going to be  $s_x + 1$ .

So this is going to be  $x_n + 1$ . So if I find a distribution of  $x_1$  to  $x_n$  condition on this

condition is  $s_n + 1$ , it's very easy to show that  $x_n + 1$  is equal to  $s_n + 1$ . So there's no randomness here. But looking at the figure, you see that-- and the distributions, you see that everything is symmetric. So I've found the distribution  $x_1$  to  $x_n$  and I can find  $x_n + 1$  easily from them.

But what if I find the distribution of  $x_2$  to  $x_n + 1$ , condition on this thing, and then find  $x_1$  from them? So you can say that  $x_1$  is equal to  $x_n - 1$  to  $n + 1$ . And then we have this solution. So it seems that there's no difference between them and actually, there's not. So condition on  $s_n + 1$ , we have the same distribution for  $x_1$  to  $x_n$  or  $x_2$  to  $x_n + 1$ . But there are only  $n$  free parameters because the  $n + 1$ 's can be calculated in your equations.

If you take any  $n$  random variables out of this  $n + 1$  random variables, we have the same story. So it's going to be uniform, our  $0$  to  $t$  for this interval. So it's symmetric. Now, the very nice thing is about  $n$  of  $t$ , equal  $t - [? \text{ so } ?] n$  must be equal to  $n$ .

So this is easy because we had these equations, but what if I call this thing  $x_n^*$  of  $n + 1$  and I do the conditioning over  $n$  of  $t$ ? So I don't know the arrival time for  $n + 1$ 's arrival but I do know the number of arrivals at time [INAUDIBLE]  $t$ . Condition on that, what is the distribution for example, for  $x_n^*$  of  $n + 1$  or  $x_1$  to  $x_n$ ? Again, easily we can say that it's the same story, so it's uniform. So you can take out of  $x_1$  to  $x_n$  and  $x_n^*$  of  $n + 1$ , you can take  $n$  off them and find the distribution of that.

OK, so the other way to look at this is that if I find the distribution of  $x_1$  to  $x_n$  condition on  $n$  of  $t$ , I will have the expect that  $x_n^*$  of  $n + 1$  is equal to  $t - s_n$ , which is  $t - \sum_{i=1}^n x_i$ . Can anyone of you see this? Hope so. Oh, OK. So again, if I find the distribution of  $x_1$  to  $x_n$ , I can find  $x_n^*$  of  $n + 1$  deterministically. Just as a signage check, you can find that  $f$  of the star of  $x$ -- oh, sorry,  $f$  of  $x_n^*$  of  $n + 1$  condition on--

So looking at this formulation and just replacing everything, we can't find this thing, and you can see that this is similar to this one again. So this is the distribution for--

well, this is the distribution for  $x_1$ . So you can find the density and this is going to be density. So what I'm saying is that  $x_1$  has the same marginal distribution as  $x_n$  plus 1 star. So you can get it with this kind of reasoning, saying that the marginal distribution should be the same or looking at this kind of formulation. When you have this distribution you can find the distribution of summation of  $x_i$ ,  $i$  equal to 1 to  $n$ . And then you can find-- So  $t$  is determined, so you can just replace everything and find it. And you can see that these two are the same.

**AUDIENCE:** [INAUDIBLE] the same distribution that is one?

**PROFESSOR:** Yeah. Condition on  $n$  [INAUDIBLE]. So everything is symmetric and uniform. But you can only choose  $n$  for your parameters because the  $n$  plus 1 is a combination of the-- Right? So this is the end of the chapter [INAUDIBLE] for Poisson process, is there any problems or any questions about it?

So let's start the Markov chain. So in this chapter we only talk about the finite-state of Markov chains. Markov chains are some processes that changing integer-time sense. So it's like we're quantizing time and any kind of change can happen only in an integer-time sense. And it's different in this way from Poisson processes because we can have the definition of Poisson processes for any continuous time, but Markov chains are only defined for integer-times.

And finite-state Markov chains is a Markov chain where states can be only in a finite set, so we usually call it 1 to  $M$ , but you can name it in any way you like. What we are looking for in finite-state Markov chains is the probability distribution of the next state conditioned on whatever [? history ?] that I have. So I know that at integer times there can be some change.

So what is this change? We model it with the probability distribution of this change corresponding to the history. So we can model any discrete integer time processing this way. Now, the nice thing about Markov chains or homogeneous Markov chains is that this distribution is equal to these things.

So you see  $i$  and  $j$ , so  $i$  is the previous state and  $j$  is the state that I want to go in so

the distribution of the next state condition and all the history only depends on the last step. So even the previous state, the new state is independent of all their earlier things that might happen.

So there's a very nice way to show it. So in general, we say that  $x_{n+1}$  is independent of  $x_{n+1}$ . So if I know the periods-- and these things are true for all the possible states.

So whatever the history in the earlier process that I had, I only care about the state that I was in in previous time. And this is going to show me what is the distribution of next state. So the two important things that you can see this in this formulation is that, first of all, this kind of independence, it is true for all  $i, j, k, m$ , and so on. The other thing is that it's not time-dependent. So if I'm in state, the probability distribution of my next state is going to be the same if the same thing happens in 100 years. So it's not time-dependent. That's why we call it homogeneous Markov chains.

And we can have non-homogeneous Markov chains, but there are not many nice results about them. I mean, we cannot find many good things. And all the information that we know to characterize the Markov chain is this kind of distribution. So we call these things transition probabilities. And something else which is called initial probabilities, which tell me that, at time instance 0, what is the probability distribution of the states. So knowing Markov chains, you can find that probability distribution of  $x_1$  is equal to probability of  $x_1$  condition and  $x_0$  [INAUDIBLE]. So knowing this thing and these transition probabilities, I can find the probability distribution of  $x_1$ .

Well, very easily, the probability of  $x_n$  is equal to-- sorry. so you can do this thing iteratively. And just knowing the transition probabilities and initial probabilities, you can find the probability distribution of the states at any time instant to forever. Do you see it's very easy? You just do this thing iteratively. So I talked about the independence and initial. Yeah, so what we do is, we characterize the Markov chains with these set of transition probabilities and the initial state. And you see that

these transition probabilities, we have, well,  $m$  times  $n$  minus 1, free parameters in them, because the summation of each of them should be equal to 1.

So for each initial state, I can have a distribution over the next step. So I'm going to talk about it in the matrix form. And what we're doing in practice usually, we assume the initial state to be a constant state. So we usually just define a state, call it initial state. Usually, we call it  $x_0$ . I mean, it's what's you're going to usually see if you want to study the behavior of Markov chains. So these are the two ways that we can visualize the transition probabilities.

So matrix form is very easy. So you just look at the  $n$  probability distributions that you have. And with Markov chain with  $n$  number of states, you will have  $n$  distributions, each of them corresponding to the conditional distribution of next step, condition of the present step equal to  $i$ . So  $i$  can be anything. So this was the thing I was saying, the number of free parameters. So I have  $n$  distribution, and each distribution has  $n$  minus 1 free parameters, because the summation of them should be equal to 1.

So this is the number of free parameters that I have. So is there any problem with the matrix form? I'm assuming that you've all been seeing this sometime before. Fine? So the nice thing about matrix form is that you can do many nice stuff with them. We can look at the notes for them, and we will see these kinds of properties later. But you can imagine that just looking at the matrix form and doing some algebra over them, we can get very nice results about the Markov chains. The other representation of the Markov chains is using a graph, a directed graph in which each node corresponds to a state and each arc, or each edge, corresponds to a transition probability.

And the very important thing about graphical model is that there's a very clear difference between 0 transition probabilities and non-0 transition probabilities. So if there's any possibility of going from one state to another state, we have an edge or an arc here. So probability of 10 to the power of minus 5 is different from 0. We have an arc in one case and no arc in another case, because there's a chance of

going from this state to the next state. And even then, the probability is 10 to the power of minus 5.

And OK, so we can do a lot of inference by just looking at the graphical model. And it's easier to see some properties of the Markov chain by looking at the graphical models. We will see these properties that we can find by looking at the graphical model. So this is not really classification of states, actually. So there are some definitions, very intuitive definitions here. So there's something called a walk, which says that this is an ordered string of the nodes, where the probability of going from each node to the next one is non-0.

So for example, looking at this walk, you can have the probability of going from 4 to 4 is positive, 4 to 1, 1 to 2, 2 to 3, and 3, to 2. So there is no kind of constraints in the definition of the walk. There should be just positive probabilities in going from one state to the next state. So we can have repetition. We can have whatever we like. And the number of-- well, we can find the number of states for sure.

So what is the maximum number of steps in a walk? Or minimum number? Minimum number of states is two-- [? one. ?] So what is the maximum number of steps? Well, there's no constraint, so it can be anything. So the next thing that we look is a path. A path is a walk where there is no repeated nodes. So I never go through one node twice. So for example, I have this path, 4, 1, 2, 3. We can see here, 4, 1, 2, 3. And now we can say, what is the maximum number of steps in a path?

**AUDIENCE:** [INAUDIBLE].

**PROFESSOR:** Steps, not states.

**AUDIENCE:** [INAUDIBLE].

**PROFESSOR:**  $n$  minus 1, yeah. Because you cannot go through one state twice. So you have a maximum number of steps in  $n$  minus 1. And the cycle is a walk in which the first and last node is repeated. So first of all, there's not a really great difference between these two cycles, so it doesn't matter where I start. And we don't care

about the repetition and definition of cycle. Oh, OK. Yeah, sorry. No node is repeated. So sorry, something was wrong. Yeah, we shouldn't have any repetition except for the first and last nodes. So the first and last node should be the same, but there shouldn't be any repetition.

So what is the maximum number of steps in this case?

**AUDIENCE:** [INAUDIBLE].

**PROFESSOR:** n, yeah, because we have an additional step. Yeah?

**AUDIENCE:** You said in a path, the maximum number of steps is n minus 1?

**PROFESSOR:** Yeah.

**AUDIENCE:** I mean, if you have n equals, like, 6, couldn't there be 1, 2, 3, 4, 5, 6?

**PROFESSOR:** No, it's the steps, not the states. So if I have path from 1 to 2, and this is the path. So there's one step. Just whenever you're confused, make a simple model. It's going to make everything very, very easy. Any other questions? So this is another definition. We say that a node  $j$  is accessible from  $i$  if there is a walk from  $i$  to  $j$ . And by just looking at the graphical model, you can verify the existence of this walk very easily. But the nice thing, again, about this thing is what I emphasized in the graphical model. It talks about if there's any positive probability of going from  $i$  to  $j$ .

So if  $j$  is accessible from  $i$  and you start at node  $i$ , there's a positive probability that you will end in node  $j$  sometime in the future. There's nothing said about the number of steps needed to get there, or the probability of that, except that this probability is non-0. It's positive. So for example, if there's a state like  $k$ , that we have  $p$  of  $ik$  is positive and  $p$  of  $kj$  is positive, then we will have  $p$  of  $ij$  [ $? too ?$ ] is positive. So we have this kind of notation. I don't have a pointer. Weird. So yeah, we have this kind of my notation.  $p_{ijn}$  means that the probability of going from state  $i$  to state  $j$  in  $n$  steps.

And this is exactly like  $n$  steps. Actually, this value could be different for different

number of  $n$ . For example, if we have a Markov chain like this,  $p$  of 131 is equal to 0, but  $p$  of 132 is non-0. So node  $j$  is accessible from  $i$  if  $p_{ijn}$  is positive for any  $n$ . So we say that  $j$  is not accessible from  $i$  if this is 0 for all possible  $n$ 's. Mercedes?

**AUDIENCE:** Are those actually greater than or equal to?

**PROFESSOR:** No, greater. They are always greater than or equal to 0. Probability is always non-negative. So what we want is positive probability, meaning that there's a chance that I will get there. I don't care how small this chance is, but it shouldn't be 0.

**AUDIENCE:** [INAUDIBLE]. So  $p$  to  $ij$ , though, couldn't it be equal to  $p_{ik}$ , [ $? p_{kj}$ ? ?]

**PROFESSOR:** Yeah, exactly. So in this case,  $p_{132}$  is equal to  $p_{12}$ ,  $p_{23}$ ,

**AUDIENCE:** So I guess I'm asking if  $p_{2ij}$  should be greater than or equal to  $p_{fk}$ .

**PROFESSOR:** Oh, OK. No, actually. You know why? Because there can exist some other state like here.

**AUDIENCE:** Right. But if that doesn't exist and there's only one path.

**PROFESSOR:** Yeah, sure. But there's no guarantee. What I'm saying is that  $p_{ik}$  is positive,  $p_{kj}$  is positive. So this thing is positive, but it can be bigger. In this case, it can be bigger. I don't care about the quantity, about the amount of the probability. I care about its positiveness, it's greater than 0. I just want it to be non-0. The other thing is that when we look at  $p_{ijn}$ , I don't care what kind of walk I have between  $i$  and  $j$ . It's a walk, so it can have repetition, it can have cycles, or it can have anything. I don't care. So if you really want to calculate  $p_{ij}$  for  $n$  equal to [INAUDIBLE] 1,000, you should really find all possible walks from  $i$  to  $j$  and add the probabilities to find this value. And all the walks within steps.

Not here. Let's look at some examples. So is node 3 accessible from node 1? So you see that there's a walk like this, 1, 2, 3. So there's a positive probability of going to state 3 from node 1. So node 3 is accessible from 1. But if you really want to calculate the probability, you should also look at the fact that we can have cycles. So 1, 2, 3 is a walk. But 1, 1, 2, 3 is also a walk. 1, 2, 3, 2, 3 is also a walk. You see?

So you have to count all these things to find the probability of  $p_{13n}$  for any  $n$ .

What about state 5? So is the node 3 accessible from node 5? You see that it's not. Actually, if you're going to state 5, you never go out. With [INAUDIBLE] probability of 1, you stay there forever. So actually, no state except state 5 is accessible from 5. So is node 2 accessible from itself? So accessible means that I should have a walk from 2 to 2 in some number of steps. So you can see that we can have 2, 3, 2, or many other walks. So it's accessible.

But as you see, node 6 is not accessible from itself. If you are in node 6, you always go out. So it's not accessible. So let's go back to these definitions. Yeah, this is what I said, and I'm emphasizing again. If you want to say that  $j$  is not accessible from  $i$ , I should have a  $p_{ijn}$  equal to 0 for all  $n$ . The other thing is that, if there's a walk from  $i$  to  $j$  and from  $j$  to  $k$ , we can prove easily that there's a walk from  $i$  to  $k$ . So having a walk from  $i$  to  $j$  means that for some  $n$ , this thing is positive. So this is  $i$  to  $j$ . From  $j$  to  $k$ , this means that  $p_{jk}$ , for some  $n$ , this is positive.

So looking at  $i$  to  $k$ , I can say that  $p_{ik, m+n}$  is greater than. And I have this thing here because of the reason that I explained right now, because there might be other paths from  $i$  to  $k$ , except the paths going through  $j$ , or except the walks that have  $n$  steps to get to  $j$  and  $n$  steps to get to  $k$ . So we know that, well, we can concatenate the walks from-- so if there's a walk from  $i$  to  $j$  and  $j$  to  $k$ , we have a walk from  $i$  to  $k$ . So we say that states  $i$  and  $j$  communicate if there's a walk from  $i$  to  $j$  and from  $j$  to  $i$ . So it can go back and forth. So it means that there's a cycle from  $i$  to  $i$  or  $j$  to  $j$ . This is the implication.

So it's, again, very simple to prove that if  $i$  communicates with  $j$  and  $j$  communicates with  $k$ , then  $i$  communicates with  $k$ . In order to prove that, I should assume that  $i$  communicates to  $j$ , I need to prove that there is a walk from  $i$  to  $k$ , and I need to prove that there is a walk from  $k$  to  $i$ . So this means that  $i$  communicates with  $k$ . So these two things can be proved easily from the concatenation of the walks and the fact that  $i$  and  $j$  communicate and  $j$  and  $k$  communicate. Now, what I can define is something called a class of states. So a class is called a non-empty set of states,

where all the pairs of states in a class communicated with each other, and none of them communicate with any other state in the Markov chain. So this is the definition of class.

So I just group the pairs that can communicate with the state and just get rid of all those who do not communicate to the [INAUDIBLE]. So for defining a class or for finding a class, or for naming a class, we can have a representative state. So I want to find all the states that communicate with each other in a class. So I can just pick one of the states in this class and find all the states that communicate with this single state, because if two states communicate with one state, then these two states communicate with each other. And if there's a state that doesn't communicate with me, it doesn't communicate with anybody else whom I'm communicating with. I'm going to prove it now, just in a few moments.

Just, I want to look at this figure again. So first I take state 2 and find a class that has this state in itself. So you see that in this class, we have state 2 on 3, because there's only a state 3 that communicates with 2. And correspondingly, we can have class having state 4 and 5. And you see that state 1 is communicating with itself, so it's a class by itself. But it doesn't communicate with anyone else. So we have this class also.

So next question, why do we call C4 a class? So it doesn't communicate with itself. If you're in state 6, you go out of it with probability of 1, eventually. So why do we call it a class? Actually, so this is the definition of the classes. But we want to have some very nice property of other classes which says that we can partition the states in a Markov chain by using the classes. So if I don't count this case as a class, I cannot partition it, because partitioning means that I should cover the whole states in the classes. What I to do is to do some kind of partitioning in a Markov chain that I have, by using the classes, so that I can have a representative state for each class. And this is one way to partition the Markov chains.

And why do I say it's partitioning? Well, it's covering the whole finite space of the states. But I need to prove that there's no intersection between classes also. Why is

it like that? So meaning that I cannot have two classes where there is an intersection between them, because if there's an intersection-- for example,  $i$  belongs to  $C1$  and  $i$  belongs to  $C2$ -- it means that  $i$  communicates with all the states in  $C1$  and  $i$  communicates with all states in  $C2$ . So you can say that all states in  $C1$  communicate with all states in  $C2$ . So actually, they should be the same.

And there's only these states that communicate with each other. We have this exclusivity that's conditional. So we can have this kind of partitioning. Is there any question? Everybody's fine? So another definition which is going to be very, very important in the future for us is the recurrency. And it's actually very simple. So a state is called recurrent if for all the states that we have  $j$  is accessible to  $i$ , we also have  $i$  is accessible to  $j$ . So if from some state  $i$ , in some number of states, I can go to some state  $j$ , I can get back to  $i$  from this state for sure. And this should be true for all the states in a Markov chain.

So if there's a walk from  $i$  to  $j$ , there should be a walk from  $j$  to  $i$ , too. If this is true, then  $i$  is recurrent. This is the definition of recurrence. Is it OK? And if it's not recurrent, we call it transient. And why do we call transient? I mean, the intuition is very nice. So what is a transient state? A transient state says that there might be some kind of walk from  $i$  to some  $k$ . And then I can't go back. So there's a positive probability that I go out of state  $i$  in some way, in some walk, and never come back to it. So it's kind of transitional behavior.

**AUDIENCE:** [INAUDIBLE]?

**PROFESSOR:** No, no, with some probability. So there exists some probability. There is a positive probability that I go out of it and never come back. It's enough for definition of transience. So you know why? Because I have the definition of recurrence for all the  $j$ 's that are accessible from  $i$ .

So with probability 1-- oh, OK, so I cannot say with probability 1. Yeah, I cannot say probability of 1 for recurrency. But for transient behavior, there exists some probability. There's a positive probability that I go out and never come back. State 1 is transient. And state 3, is it recurrent or transient? Some idea?

**AUDIENCE:** [INAUDIBLE].

**PROFESSOR:** Transient? What about state 2? It's wrong, because there should be something going on [INAUDIBLE]. So now here, it's recurrent. Good? Oh yeah, we have examples here. So states 2 and 3 are recurrent, because the only state that I can go out from this state is to themselves. So they are only accessible from themselves. And there's a positive probability of going from one to the other. So they're recurrent. But states 4 and 5 are transient. And the reason is here, because there is a positive probability that I go out of state 4 in that direction and never come back. And there's no way to come back.

And state 6 and 1 are also transient. State 6 is very, very transient because, well, I go out of it in one step and never come back. And state 1 is also transient, because I can go out of it in this direction and never be able to come back. Do you see? So this is a very, very, very important theorem, which says that if we have some classes, the states in the class are all recurrent or all transient. This is what we are going to use a lot in the future, so you should remember it. And the proof is very easy. So let's assume that state  $i$  is recurrent. And let's define  $S_i$ , all the states that communicate with  $i$ , that are accessible from  $i$ .

And you know that since  $i$  is recurrent, being accessible from  $i$  means that  $i$  is accessible from them. So if  $j$  is accessible from  $i$ ,  $i$  is also accessible from  $j$ . So we know that  $i$  and  $j$  communicate with each other if and only if  $j$  is in this set. So this is a class. This is the class that contains  $i$ . So this is like the class that I told you, that we can have a class where this state  $i$  is representative. And actually, any state in a class can be representative. It doesn't matter.

So by just looking at state  $i$ , I can define a class. The states that are accessible from  $i$ — and actually, this is the class that contains  $i$ . So let's assume that there's a state called  $k$  which is accessible from some  $j$  which is in this set. So  $k$  is accessible from  $j$ , and  $j$  is accessible from  $i$ . So  $k$  is accessible from  $i$ . But  $k$  accessible from  $i$  implies that  $i$  is also accessible from  $k$ , because  $i$  is recurrent. So you think  $i$  is accessible from  $k$ ? And you know that  $j$  is also accessible from  $i$  because, well,  $S_i$  was also a

class, recurrent class.

So you see that  $j$  is also recurrent. So if  $k$  is accessible from  $j$ , then  $j$  is also accessible from  $k$  for any  $k$ . So this is the definition of recurrency. So what I want to say is that if  $i$  is recurrent and it's in the same class as  $j$ , then  $j$  is recurrent for sure. And we didn't prove it here, but if one of them is transient, then all of them are transient too. So the proof is very simple. I proved that if there is any state like  $k$  that's accessible from  $j$ , I need to prove that  $j$  is also accessible from  $k$ . And I proved it like this. It's easy.

So the next definition that we have is the definition of the periodic states and classes. So I told you that the number of steps in a walk-- so when I say that there's a walk from some state to the other state, I didn't specify the number of steps needed to get from the state to another state. So assuming that there is a walk from state  $i$  to  $i$ , meaning that  $i$  is accessible from  $i$ , it's not always true. You might go out of  $i$  and never come back to it. So assuming that there's a positive probability that we can go from state  $i$  to  $i$ , you can look at the number of steps needed for this walk. And we can find the greatest common divisor of this number of steps. And it's called the period of  $i$ .

And if this number is greater than 1, then  $i$  is called to be periodic. And if it's not, it's aperiodic. So the very simple example of this thing is this Markov chain. So probability of 1,1 in step is 0. probability of 1, 1 in two steps is positive. And actually, probability of 1, 1 for all even number of steps is positive. But for all odd number of steps, it's 0. And actually, you can prove that 2 is the greatest common divisor. So this is the very easy example to show that it's periodic.

So there is a very simple thing to check if something is aperiodic. But it doesn't tell me-- I mean, if it doesn't exist, we don't know its periodicity. So if there is a walk from state 1 or state  $i$  to itself, and in this walk, we go from a state called, like,  $j$ , and there is a loop here-- so  $P_{jj}$  is positive. What can we say about the periodicity of  $i$ ? No, the period is 1. Yeah, exactly.

So in this case, state  $i$  is always aperiodic. So if there's a walk, and in this walk,

there is a loop, we always say that. But if it doesn't exist, can we say anything about the periodicity? No. It might be periodic. It might be aperiodic. It's just a check. So whenever you see a loop, it's aperiodic. Whenever you see a loop in the walk from  $i$  to  $i$ , if there's a loop in the other side of the Markov chain, we don't care.

So the definition is fine. So just looking at this example, so if we're going from state 4 to 4, the number of states that we need is, like, 4, 6, 8. So 4, 1, 2, 3, 4 is a cycle, or is a walk from 4, to 4, and the number of steps is 4. 4, 5, 6, 7, 8, 9, 4 is another walk. So this corresponds to  $n$  equal to 6. And 4, 1, 2, 3, 4, 1, 2, 3, 4 is another walk which corresponds to  $n$  equal to 8.

So you see that we can go like this or this or this. So these are different  $n$ 's. But we see that the greatest common divisor is 2. So the period of state 4 is 2. For state 7, we have this thing. So the minimum number of steps to get from 7 to itself is 6, and then you can get from 7 to 7 in 10 steps. And I hope you see it. So it's going to be like this. And so again, the greatest common divisor is 2.

So we proved that if one state in a class is recurrent, and then all the states are recurrent. And we said that recurrency corresponds to having a cycle, having a walk from each state to itself. So this is the result, very similar to that one, which says that all the states in the same class have the same period. And in this example, you see it too. The proof is not very complicated, but it takes time to do that. So I'm not going to do it. But it's all nice. So it's good if you look at the text, you'll see that.

**AUDIENCE:** [INAUDIBLE] only for recurrent states?

**PROFESSOR:** Yeah. For non-recurrent ones, it's--

**AUDIENCE:** [INAUDIBLE].

**PROFESSOR:** It's 1. It's aperiodic. OK, yeah. Periodicity is only defined for recurrent states, yeah. We have another example here. I just want to show you. So I have two recurrent classes in this example. Actually, three. So one of them is this class. What is the period for a class corresponding to state 1? It's 1, because I have a loop. It's very simple. For any  $n$ , there is a positive probability. What is the period for this class,

states-- containing states 4 and 5?

Look at it. There is definitely a loop in this class. So it's 1. No, I said that whenever there is a loop, the greatest common divisor is 1.

So for going from a state 5 to 5, we can go it in one step, two step, three step, four step, five. So the greatest common divisor is 1. So here, I showed you that if there is a loop, then it's definitely aperiodic, meaning that the greatest common divisor is 1.

**AUDIENCE:** [INAUDIBLE] have self-transitions [INAUDIBLE] 2?

**PROFESSOR:** No, I'm talking about 4 and 5. In what?

**AUDIENCE:** If they didn't have self-transitions? [INAUDIBLE].

**PROFESSOR:** Oh yeah. It would be like this one?

**AUDIENCE:** Yeah, [INAUDIBLE].

**PROFESSOR:** Yeah, definitely. So the class containing states 2 and 3, they are periodic.

**AUDIENCE:** Oh, wait. You just said [INAUDIBLE].

**PROFESSOR:** Oh, OK.

**AUDIENCE:** [INAUDIBLE].

**PROFESSOR:** Yeah, actually, we can define for transient states. Yeah, for transient classes, we can define periodicity, like this case. Yeah, why not? This is another very important thing that we can do in Markov chains. So I have a class of states in a Markov chain, and they are periodic. So the period of each state in a class is greater than 1. And you know that all the states in a class have the same period. So it means that I can partition the class into these subclasses, where there is only-- OK, so-- do I have that? I don't know.

So let's assume that  $d$  is equal to 3. And the whole team is the class of states that I'm talking about. I can partition into three classes, in which I only have transitions

from one of these to the other one. So there's no transition in this subclass to itself. And the only transitions are from one class to the next one. So this is sort of intuitive. So just looking at three states, if the period is 3, I can partition it in this way. Or I can have it like two of these in this case, where I have like this. You see?

So there's a transition from this to these two states, and from these two states to here, and from here to here. But I cannot have any transition from here to itself, or from here to here, or here to here. Just look at the text for the proof and illustration. But there exists this kind of partitioning. You should know that. Yeah, so if I am in a class, in a subclass, the next state will be in the next subclass, for sure. So I know this subclass in [INAUDIBLE]. And in two steps, I will be in the other [INAUDIBLE]. So you know that the set of classes that I can be in state,  $nd + m$ .

So the other definition that you can say is that, again, you can choose one of the states in a class that I talked about. So let's say that I choose a state 1. And I define  $S_m$ --  $S_m$  corresponds to that subclass-- are all the  $j$ 's where  $S_{1j}$ . So I told you that I will have  $d$  classes, I can define each class like this. So this is all the possible states that I can be in  $nd + m$  step. So starting from state 1, in  $nd + m$  step, I can be in set of classes, for some  $m$ . So  $m$  can be big. But anyway, I call this the subclass number  $m$ .

So let's just talk about something. So I said that in order to characterize the Markov chains, I need to show you the initial state or initial state distribution. So it can be deterministic, like I start from some specific state all the time, or there can be some distribution, like  $p_{x0}$ , meaning that this is the distribution that I would have at my initial state. I don't have it here, but using, I think, chain rules, we can easily find a distribution of states at each time instant by having the transitional probabilities and the initial state distribution. I just wrote it for you.

So for characterizing the Markov chain, I just need to tell you the transition probabilities and the initial state. So a very good question is that, is there any kind of stable behavior when time goes, like in very far future? So I had an example here, just looking at here. So this is a very simple thing that I can say about this Markov

chain. I know that in the very far future, there is a 0 probability that I'm in state 6. And you know it why? Because any time that I am in state 6, I will go out of it with probability 1. So there's no chance that I will be there.

Or I can say that, if I start from state 4, in very, very far future, I will not be in state 1, 4, or 5, for sure. You know why? Because there is a positive probability of going out of these three states, like here. And then if I go out of it, I can never come back. So there's a chance of going from state 4 to state 2. And if I ever go there, I will never come back.

So these are the statements that I can say about the state of behavior or the very, very future behavior of the Markov chain. So the question is that, can we always say these kind of statements? And what kind of statements, actually, we can have? And so you see that, for example, for a state six [INAUDIBLE] example, I could say that the probability of being in state 6 as  $n$  goes to infinity is 0. So can I always have some kind of probability distribution over the states in the future as  $n$  goes to infinity? This is a very good question.

And actually, it's related to a lot of applications that we have for Markov chains, like the queuing theory. And you can have queues for almost anything. So one of the most fundamental and interesting classes of states are the states that are called ergodic, meaning that they are recurrent and aperiodic. So if I have a Markov chain that has only one class, and this class is recurrent and aperiodic, then we call it a ergodic Markov chain. So we had two theorems saying that if a state in a class is recurrent, then all the states are recurrent. And if a state in a class is aperiodic, then all the states are aperiodic.

So we can say that some classes are ergodic and some are not ergodic. So if the Markov chain has only one class, and this class is aperiodic and recurrent, then we call it a ergodic Markov chain. And the very important and nice property of ergodic Markov chains is that they lose memory as  $n$  goes to infinity, meaning that whatever initial distribution that I have for the initial state, I will lose memory of that. So whatever state that I start in, or whatever distribution that I start in, after a while, for

a large enough  $n$ , the distribution of states does not depend on that.

So again, looking at that chain rule, I could say that I can find the probability distribution of  $x$  by looking at this thing. And looking at these recursively, it all depends on  $P$  of the initial distribution. And the ergodic Markov chains have this property that, after a while, this distribution doesn't depend on initial distribution anymore. So they lose memory. And actually, usually we can calculate the stable distribution. So this thing goes to a limit which is called  $p_j$ . So the important thing is that it doesn't depend on  $i$ . It doesn't depend where  $i$  starts. Eventually, I will converge the distribution. And then this distribution doesn't change.

So for a large enough  $n$ , I have this property for ergodic Markov chains. We will have a lot to do with these properties in the future. So I was saying that this  $p_j$ ,  $p_i$ , should be positive. So being in each state has a non-0 probability. The first thing that I need to prove is that  $p_{ij} > 0$  and is nonzero for a large enough  $n$  for all  $j$  and all the initial distributions. And I want to prove that this is true for ergodic Markov chains. This is not true, generally. Well, this is more a combinatorial issue, but there is a theorem here which says that for an ergodic Markov chain, for all  $n$  greater than this value, I have non-0 probability of going from  $i$  to  $j$ .

So the thing that you should be careful here is that, for all  $n$  greater than this value. So for going from a state 1 to 1, I can go it in six steps and 12 steps and so on. But I cannot go it in 24 steps, I think. I cannot go from 1 to 1 in 25 steps. But I can go from 1 to 1 in 26, 27, 28, 29, to 30. So for  $n$  greater than  $n - 1$  squared plus 1, I can go from any state to any state. So I think this [? bond ?] is tied for state 4. Sorry, maybe what I said is true for state 4. Yeah.

So I'm not going to prove this theorem. Well, actually you don't have the proof in the notes, either. But yeah, you can look at the example and the cases that are discussed in the notes. Is there any questions? Fine?