So now that we have defined what a Markov chain is, what can we do with it? Well, we usually build models in order to predict some phenomenon of interest. In the case of a Markov chain, there is randomness. And so it is natural to think about making probabilistic predictions.

For example, going back again to our checkout counter example, you have arrived at 6:45 PM. There are two customers in a queue. And you want to predict the number of customers in the queue at 7:00 PM. Assuming time steps are in seconds, that corresponds to 900 times steps later.

There is no way to know exactly where the system will be. But you may be able to give probabilistic prediction. That is, to give the probability for the system to be in a given state 900 time steps later. Our main purpose will be to calculate such probabilities.

And notation will be useful here. Suppose that the Markov chain of interest starts in a given state, i, and that it runs for n transitions. Let us introduce the notation rijn to represent the n step transition probability of ending in state j.

This is the initial state, i. And this is the definition, rij of n. First note that these are probabilities. Given that you started in i, after n steps, you will end up in some state with probability 1. So the summation of all rij of n for all possible states j, will be 1.

And this is true for all i, all initial state, and for any time step n. Also, because we have a time invariant Markov chain, rijn is also given by this formula here for any possible value of s. Going from here to here, plus s, plus s. In other words, if currently you are in state i at time steps s, and you're interested in knowing what is the probability of being in state j at time n plus s, which mean n steps later, you will still have the same expression-- rijn.

So how to calculate rijn. For some particular n, this is easy. For example, for rij of zero, that means that there are no transition, it will be either 1 if i equal j, and zero otherwise. In one step, in other words, when n equals 1, rij of 1 will be the probability transition given by the Markov chain. And that is true for all i and all j.

Now let us calculate rijn for n greater than or equal to 2. We are going to apply the total probably theorem, and break up the calculation of that quantity by considering the different ways that the event

of interest can happen. Again, the event of interest is to go from i, state i, to state j in n times steps.

There are many ways for that event to happen. Let's group these many different ways, as follows. Let us consider the first n minus 1 steps. And group together all possible ways of going from i to a state k, a given state k, in n minus 1 steps. And this wiggle path here represent all possible ways of doing that.

Using the definition above, that probability of going from i to k in n minus 1 steps will be rik of n minus 1. Now assuming that you ended up in state k in n minus 1 transitions, the probability that you end up in state j in the next transition is simply the one-step transition probability, pkj. So altogether, the probability of going from state i to state j in n steps, and in being in state k after n minus 1 steps is simply rik n minus 1 times pkj. Note that state k can be any of the finite number of possible states of our system. In summary, all such paths can be represented by the following diagram.

So again, from time zero, you are in state i. And you want to be at time n in state j. And you break down all the possible ways by first looking at what would happen after step n minus 1. You can be in state 1, state k, all the way to state m.

We have calculated these expression here. This is what we have done here. This is for state 1 and state m. So the overall probability of reaching node j is obtained by an application of the total probability theorem.

It gives the following formula. So this is corresponding to the total probability theorem. Here, this is the calculation that we have done here.

And we sum over all possibilities for k. Where did we use the Markov property in this calculation? Well, the key step here was when we said that this probability here was pkj.

Going back to the calculation that we had here, this was in fact the probability of being in state j at times n, given you started in state i, and you were in state n minus 1 in k. And that probability being equals to the probability of xn equals j given the last time. That is due to Markov. And this is pkj.

This is a recursion in the following sense. Assume that you have calculated rik n minus 1 for all possible values of i, and all possible value of k. And you have stored all these values, m square of them.

For any pair, ij, you can now calculate rijn using that formula. And you can do it in, essentially m

multiplication, and m minus 1 additions. That is, in a number of steps or number of elementary steps proportional to m.

You do this for all m square pair of ij at the time step n. And then you repeat for n plus 1, et cetera. So this is the essence of the recursion.

Here is a variation that is another recursion for computing rij of n. You start at i. And suppose that in the one time step, you find yourself in state k.

The probability here is the one-step transition, pik. And then, given that you are in state k, what is the probability that you will end up in state j in n minus 1 step? Will be, again, looking at this formula that we had here, rkj of n minus 1.

Again, you have to consider all possible values for k here. And the application of the total probability theorem gives the following alternative recursion. rij of n, is the sum for all k equals 1 to m of pik times rkj of n minus 1. These two recursions-- this one and this one-- are different. They are both valid, and could be useful, depending on the specific questions you may want to answer.

Finally note, that if the initial state is itself random, that is given by a random distribution-- this is the initial distribution on the state, than the state probability distribution after n steps will be given by this formula. This is the state after n step. It's simply that. And this is, yet again, an application of the total probability theorem.