As a warm-up, just to see how to use steady-state probabilities, let us look at our familiar example.

This is a two-state Markov chain, and we did write down the complete balance equations for this chain, and found the steady-state probabilities before.

Notice that we can find these by using the trick which we introduced for birth and death processes.

You cut the chain along this line, and argue that the frequency of transition of this type has to be the same as the frequency of transition of this type.

So if you have pi 1 here and pi 2 here, what it means is pi 1 times 0.5, which represent the frequency of these kind of transitions, has to be equal to pi 2 times 0.2, which is this kind of transition, plus pi 1 plus pi 2, the normalization equation.

And by solving the system of equations, you obtain the same thing as what we obtained before, which is the steady-state probabilities of state 1 and of state 2.

Let us now try to calculate some related quantities.

Suppose that you start at state 1, and you want to calculate this particular probability.

So you start at state 1, and you want to know what is the probability that next time you will be in state 1, and 100 times step later, you are still in state 1.

Now, the conditional probabilities of two things happening is the conditional probability of the first one happening, x1 equals 1, given x0 equals 1, and given that the first one happens, the probability that the second one happened-- x100 equals 1, given x1 equals 1, and x0 equals 1.

So what is this?

This first one is the transition probability from state 1 to state 1, p 1 1.

How about the second probability?

Because of the Markov property, that information is irrelevant, and so that probability is r 1 1 in 99 steps.

Now, 99 is possibly a big number, and so we approximate this quantity, and we are going to use the steady-state probability of state 1 for doing that.

And that gives us an approximation, p 1 1, times pi of 1, which, then, is 0.5 times 2 over 7.

Now, how about this expression?

Given that you start in state 1, what is the probability that at time step 100, you are in 1, and time step 101 you are in 2.

By doing the same technique gives the conditional probability of the first thing happening, and then, given that thing happening, the probability that the second one happen.

x101 equals 2, given x100 equals 1, and x0 equals 1.

And again, here what we have is r 1 1 of 100, and times-- again, the Markov property tells us that we can forget about this one-- and this is the probability transition from 1 to 2, pi 1 2.

And again, here, if n equals 100 is large enough, we can approximate that by pi 1 times p 1 2.

And this, then, 2 over 7 times 0.5 again.

Finally, let's calculate this third expression, where, again, we start at state 1, and now we are asking, what is probability that after time step 100, you are in 1, and 100 steps later, you are again in 1?

We use the same trick as before, this probability that the first thing happened, and given that, the probability of the second one happened.

Again, this is r 1 1 of 100, and this one, for the same reason as before, we can forget this term.

From 100 to 200, you have 100 time steps, r 1 1 of 100.

And 100 is big enough, so we're going to approximate both by the same values, and we get pi 1 square, which is 2 over 7 square.

Now, in this calculation, we assume that n equals 99, or n equals 100 were big enough-- big enough so that the limit has taken effect.

But how do we know that our approximation is good?

In other words, is n equals 99 or 100 large enough?

Well, this has something to do with the mixing time scale of our Markov chain, and by mixing time scale, I mean, how long does it take for the initial states to be forgotten?

So how can you see that here?

Well, you can first try a simulation.

So using any of your favorite software, simulate.

If you calculate, and you look at, as a function of n, and you draw r 1 1 of n, at n equals 0, this is 1.

n equals 1, then you have 5, et cetera.

At 1, this is going to be p 1 1, and p 1 1 was 0.5, so already it's here.

So initially it was here, here.

And 2 over 7, so this is 0.5.

2 over 7 is here.

And what we know is that when n goes to infinity, these things goes to 2 over 7.

And if you simulate, you will see that it goes very fast.

So with you, I'm just joining points, and we will see that n equals 5 already.

You are very, very close to 2 over 7.

So you have really an exponential decrease here.

In fact, if you look at the simulation result, and you look at n equals 5 already, here you have two correct decimal already.

And for n equals 10, it's correct up to five decimals.

Or, if you do not want to have simulation, but simply think in terms of order of magnitude-- so here it would be another approach would be order of magnitude type of argument-- and in order to do that, starting here, on average, how many trials, or how many time steps would it take in order for you to observe such a transition here?

Well, you use the geometric, random variable, and this is the amount of time, on average, until you have success.

And it is 1 over the probability here, so it takes an average two time steps to go from here to here.

And to go from here to here, it takes, on average, 1 over 0.2, which is about five time steps.

So as an order of magnitude, given you started here in state 1, after, on average, about 10 iterations, there will be some randomness.

There is a high likelihood, on average, that you will go there and come here.

And then if you do n equals 100, which is 10 times that, in terms of order of magnitude, it looks like n is large enough.

So that would be a back to the envelope calculation.

Now, this kind of calculation is useful in general, not just here.

So for example, let's do it again.

Assume that instead of 0.5 here, the probability that you had was 0.999.

And maybe here, instead of 0.8, it was 0.998.

Now, in order to observe such a transition, it will take, on average, since this number here would become 0.001, it would take, on average, about 1,000 time steps in order to observe one such transition.

So if you look at time steps of that order, it will not be enough if your chain were of this type.

After n equals 100, the likelihood would be that you will still be here, so the initial state, or the initial condition, would matter still.

So you would take about 1,000 to get there, and then here it would have 0.002.

That means that, on average, from here you would take about 500 iterations before you observe that for the first time, so the same order of magnitude.

So in order to get enough randomness here, a good rule would be to multiply this 1,000 by 10.

So maybe with n equals 10,000, you would feel confident enough, in that specific case, in order to use this kind of approximation that we have used here.

And finally, for those interested, you can study that by theory.

And here there is an entire field that try to study how fast a Markov chain converges to steady state, and the so-called mixing time.

And it turns out that for these Markov chain, you can say that the convergence, or the rate of convergence, is of the order c at the power n, where c is a number that is between 0 and 1.

And the closer c is to 1, the slower the convergence is.

And the closer c is to 0, the faster the convergence is.

And for example, here, for our initial case, the c was 0.3, so that was the first case for the second chain.

With these kind of probabilities, 0.99 and 0.998, the c would be 0.997.