Introduction to Markov Chains

CMS 380 Simulation and Stochastic Modeling

State Transition Models

A Markov chain is a model consisting of a group of *states* and specified *transitions* between the states. Older texts on queueing theory prefer to derive most of their results using Markov models, although there are other techniques we'll see later in the course that can be more useful for particular situations.

Kinds of Markov Chains

A Markov chain can have a finite or infinite number of states. In a *discrete time Markov chain* (DTMC) each state change takes place at a fixed decision point and the time between changes is (conceptually) constant. In a *continuous time Markov chain* (CTMC), changes can happen at any instant. Finite and discrete models are often easier to analyze, but queueing models are expressed as continuous-time Markov chains.

A First Example

Let's imagine a factory that makes two kinds of products:

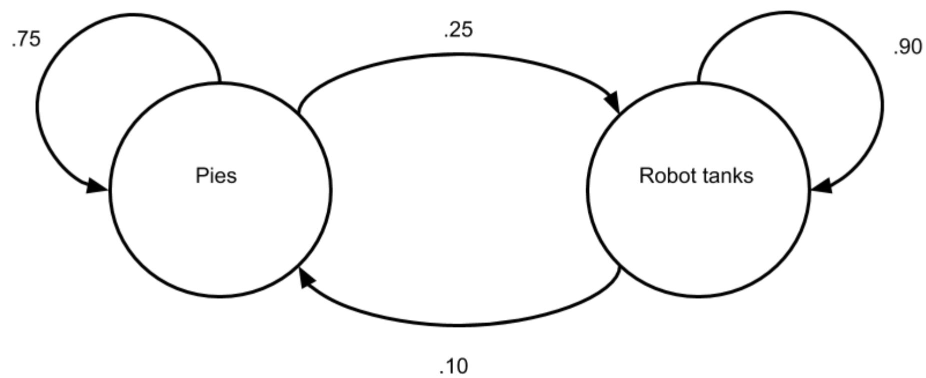
1. Delicious pies

2. Robot battle tanks

The factory makes only one kind of product (pies or tanks) each week. At the end of each week, the plant managers decide whether to keep making the same thing or switch to a different product for the next week's production.

- If the factory has just finished making pies, there is a 25% chance it will switch to making robot tanks and a 75% chance it continues making pies.
- If the factory has just made robot tanks, there is a 10% chance of switching back to pies and a 90% chance of continuing to make robot tanks.

Graphically, the state transition model looks like this:



This is a discrete-time state transition model. At each step, the model is in exactly one state (pies or robot tanks) and transitions randomly to another state according to the given probabilities.

Here are a few common questions we can ask about a model like this:

- Suppose the model runs for N transitions. What is the probability of being in the $Pies$ state at time step N? What about the probability of being in the $Tanks$ state at times step N?
- What are the long-run fractions of time spent in each state? In other words, what fraction of weeks, over the long run, would we expect the factory to be making pies?

Let's consider the first question. In general, the probability of being in some state j at time step N could depend on the entire time history of the chain. That is, if there is a long-run dependence between states, the state at time N might depend on all of the previous $N - 1$ time steps.

Formally, let X_N be a random variable representing the model's state at time step N. Let j be a state of interest and s_{N-1}, s_{N-2}, etc. be the history of previous time steps. The probability could be written as

$$P(X_N = j \mid X_{N-1} = s_{N-1} \text{ and } X_{N-2} = s_{N-2} \text{ and } X_{N-3} = s_{N-3} \text{ and } \ldots)$$

The *Markovian Property* says that transitions in a Markov chain **depend on only the current state, and not on any history of previous states**. Transitions in a Markov model are **memoryless**.

$$P(X_N = j | X_{N-1} = s_{N-1})$$

The only transitions that matter in a Markov chain are *one-step transition probabilities*.

Let P_{ij} represent the one-step transition probability of moving from state i to state j.

The Transition Matrix

Graphical models are infeasible for complex chains, so it's common to collect the one-step transition probabilities into a matrix. The starting state is on the left and the destination state is on the top. For the previous example,

```
                Pies      Robot Tanks


Pies            .75          .25


Robot Tanks     .10          .90
```

The Stationary Equations

Let's return to one of our original motivating questions: what is the long-run fraction of time spent in each state?

Let π_{pies} denote the fraction of time steps where the factory is producing pies and π_{tanks} denote the fraction of time steps where the factory produces robot battle tanks.

Our goal is to derive expressions for the π values using the one-step transition probabilities of the Markov chain. The key to doing this is to exploit the concept of **global balance**.

- Intuitively, the number of transitions *out* of each state must match the number of transitions *into* the state.
- Note that the model *always* makes a transition on every time step. Even if we leave one state and return back to it—like leaving state $Pies$ and then immediately returning—that still counts as a transition.

Therefore, we can write down equations that describe the balance conditions in terms of π_{pies} and π_{tanks}.

The first equation says that the probability of transitioning out of state $Pies$ must match the probability of transitioning into it:

$$(.75 + .25)\,\pi_{pies} \; = .75\,\pi_{pies} + .10\,\pi_{tanks}$$

The left-hand side sums up the probabilities of the arcs leaving state $Pies$. These probabilities must sum to 1, so the left-hand side reduces to π_{pies}. The right-hand side sums up the probabilities for entering state $Pies$.

A similar equation exists for state $Tanks$:

$$(.90 + .10)\,\pi_{tanks} = .25\,\pi_{pies} + .90\,\pi_{tanks}$$

Simplifying the two equations will show that there's actually only one relationship between the two unknowns

$$.25\,\pi_{pies} = .10\,\pi_{tanks}$$

Solving the model requires one more equation. Since the model must always be in one of the two states, total probability guarantees that

$$\pi_{pies} + \pi_{tanks} = 1$$

Now armed with two equations for the two unknowns, solving the system is straightforward.

$$\pi_{pies} = .28572$$
$$\pi_{tanks} = .71428$$

One of the deliverable problems asks you to solve the stationary equations for a more general form of this two-state model.

M/M/1

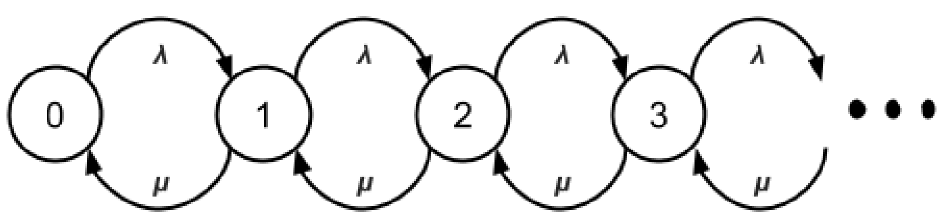
It's time: we're going to use a Markov chain model to derive a queueing equation.

Our previous example focused on a discrete time Markov chain with a finite number of states. Queueing models, by contrast, may have an infinite number of states (because the buffer may contain any number of customers), and allow transitions in continuous time.

States in the M/M/1 Markov model correspond to the number of customers in the system. The arcs correspond to the *rate* of transitions between states

- Arrivals come into the queue at constant rate λ, so all of the forward arcs are labeled with λ.
- When there are customers in the system, the customer in service completes at rate μ (recall that the average service time is $\overline{s} = 1/\mu$). Therefore, the backwards arcs are labeled with μ.
- State 0 corresponds to an empty queue. It only has a forward transition arc, because the only way to transition out of an empty queue is for a new customer to arrive.

I'm going to glide over some of the mathematical complexities of this model so we can focus on building intuitive understanding.



Let π_k denote the long-run fraction of time that the queue contains k customers. This is equal to the probability that a newly arriving customer finds

k already in the system. Our goal is to derive an expression for π_k by writing down the balance equations until a general pattern emerges.

First, analyze state 0. The rate of leaving state 0 due to arrivals is $\pi_0 \lambda$. The rate of entering state 0 from state 1 is $\pi_1 \mu$. The two rates must be equal, so we can solve for π_1 in terms of π_0.

$$\pi_1 = \frac{\lambda}{\mu}\pi_0$$

Now, analyze state 1. The total rate of leaving state 1 due to both arrivals and departures is $(\lambda + \mu)\pi_1$. The rate of entering state 1 depends on arrivals from state 0 and departures from state 2.

$$(\lambda + \mu)\pi_1 = \lambda\pi_0 + \mu\pi_2$$

Substituting $\pi_1 = \frac{\lambda}{\mu}\pi_0$:

$$\frac{\lambda^2}{\mu}\pi_0 + \lambda\pi_0 = \lambda\pi_0 + \mu\pi_2$$

And solving for π_2 in terms of π_0:

$$\pi_2 = \left(\frac{\lambda}{\mu}\right)^2 \pi_0$$

Continuing to solve the balance equations for higher states will reveal the general pattern:

$$\pi_k = \left(\frac{\lambda}{\mu}\right)^k \pi_0$$

This is almost the complete solution: all we need is an expression for π_0. Recognizing that π_0 is the probability that the queue is empty, it must be the case that

$$\pi_0 = 1 - U$$

Also recall that Utilization Law, which required

$$U = \frac{\lambda}{\mu}$$

Therefore, the probability of having k customers in the M/M/1 queue is

$$\pi_k = U^k\left(1 - U\right)$$

There is a more general way to solve for π_0: sum up all of the π_k values and set them equal to 1, then solve for π_0:

$$\sum_{k=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^k \pi_0 = 1$$

The summation is a geometric series, which simplifies to

$$\pi_0 \frac{1}{1 - \frac{\lambda}{\mu}} = 1$$

Solving for π_0 and using the Utilization Law gives the same result:

$$\pi_0 = 1 - U$$