CS 547 Lecture 35: Markov Chains and Queues

Daniel Myers

If you read older texts on queueing theory, they tend to derive their major results with Markov chains. In this framework, each state of the chain corresponds to the number of customers in the queue, and state transitions occur when new customers arrive to the queue or customers complete their service and depart.

Continuous Time Markov Chains

Our previous examples focused on *discrete time* Markov chains 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*. Our previous solution approaches relied on writing down the transition probability matrix P and either multiplying it by itself or solving the stationary equations to obtain the long run probability of being in each state.

To solve continuous time models, we need a new strategy that doesn't require writing down the entire transition matrix. Our new method will be to set up and solve the *balance equations* associated with each state.

Intuitively, the number of transitions into a state must be balanced by the number of transitions out of the state – in other words, *what goes in must come out*. If we transition into a state at some time, we must transition back out at some time in the future (note that this includes transitions from a state back to itself, which are treated the same as any other transition in a Markov chain).

Consider a queueing model, and let π_0 denote the probability of being in state 0 (that is, the probability of having zero customers in the queue) and π_1 denote the probability of being in state 1. Let the queue receive Poisson arrivals at rate λ and have exponentially distributed service times with rate μ.

Suppose we measure the queue for some long, statistically representative time period of length T. The total amount of time spent in state 0 will be $\pi_0 T$ and the total amount of time spent in state 1 will be $\pi_1 T$.

Arrivals occur at constant rate, so the expected number of times we go from having an empty queue to a queue with one customer must be $\pi_0 T\lambda$ (the total amount of time that the queue is empty times the arrival rate). Similarly, departures from the queue occur at constant rate μ, so the total number of times we transition from having a queue with one customer to a queue with zero customers is $\pi_1 T\mu$. The balance condition requires that these two values must be equal.

$$\pi_0 T\lambda = \pi_1 T\mu$$

Dividing each side by thge length of the time period T gives the *rate* of transition between the two states, with the LHS representing the rate of transition from state 0 to state 1 due to arrivals and the RHS representing the rate of transition from state 1 to state 0 due to departures.

$$\pi_0\lambda = \pi_1\mu$$

M/M/1

Let's apply the balance concept to the M/M/1 queue. The figure shows a graphical version of the model, with forward transitions corresponding to Poisson arrivals at rate λ and backwards transitions corresponding to completions at rate μ.

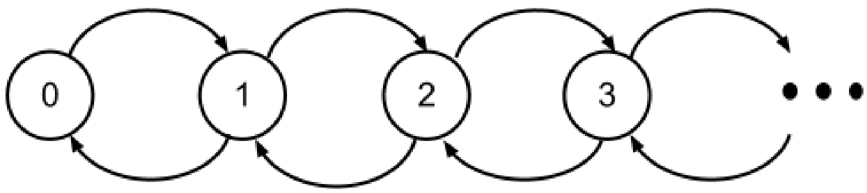

Figure 1: Markov chain model for the M/M/1 queue

Our overall goal is to derive an expression for π_k, the probability of having k customers in the queue. To do so, we'll start solving 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 $\pi_1 = \frac{\lambda}{\mu}\pi_0$,

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

Solving for π_2,

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

If we continue solving the balance equations for higher states, a general pattern emerges.

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

This is almost the solution we need – we just need a value for the initial condition π_0.[1] Recall that the π values represent *probabilities* of being in each state. Therefore, the total probability found by adding up all the π values must equal 1, to make a valid distribution.

$$\sum_{k=0}^{\infty} \pi_k = 1$$

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

[1] If you keep in mind that π_0 is the probability of having an empty queue, it's obvious that $\pi_0 = 1 - U$, but it's nice to show the general solution technique.

The summation is a geometric series, which converges when $\frac{\lambda}{\mu} < 1$.

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

By the utilization law, $\frac{\lambda}{\mu} = U$, so $\pi_0 = 1 - U$, exactly as expected.

The final formula is
$$\pi_k = U^k(1 - U)$$
which we already derived as the queue length distribution for M/M/1 using Little's result.