CS 547 Lecture 34: Markov Chains

Daniel Myers

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, as opposed to the mean value analysis approach we've used for most of this course. Understanding Markov chains will allow us to derive some new results that would be difficult to get using MVA alone.

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 constant. In a *continuous time Markov chain* (CTMC), changes can happen at any instant.

As you might expect, finite and discrete models are easier to analyze, so we'll use a DTMC for our examples in this lecture. Queueing models rely on CTMCs with infinite states, which we'll cover in the next note.

The Markovian Property

Suppose we have a DTMC. Let X_n be a random variable denoting the state the model is in at time step n. One thing we might be interested in reasoning about is the probability of being in a given state, say state j, at time step n. In general, this probability could depend on the entire time history of the chain; that is, the probability of being in state j at time n is influenced by the state the model was in at every previous time step. Suppose the model was in state i at time step $n-1$ and some state s_t at each time step $0 \leq t < n-1$. The probability we're interested in could be written as

$$P[X_n = j \mid X_{n-1} = i \text{ and } X_{n-2} = s_{n-2} \text{ and} \ldots X_0 = s_0]$$

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

$$P[X_n = j \mid X_{n-1} = i \text{ and } X_{n-2} = s_{n-2} \text{ and} \ldots X_0 = s_0] = P[X_n = j \mid X_{n-1} = i]$$

For any pair of states i and j, let P_{ij} denote the *one-step transition probability* of moving from state i to state j, independent of the time step or previous history. Specifiying the one-step transition probabilities is a compact way of describing a Markov chain.

An Example Model

Consider a model with just two states, called state 0 and state 1, and transition probabilities given by

- $P_{00} = 1 - p$
- $P_{01} = p$

- $P_{10} = q$
- $P_{11} = 1 - q$

So, the probability of moving from state 0 to state 1 is p, and the probability of moving from 1 to 0 is q, etc.

It's convenient to collect the transition probabilities into a matrix.

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

Now, let $\pi^{(n)}$ be a vector denoting the probability of being in each state after n transitions. Given a starting distribution, $\pi^{(0)}$, we can use total probability to calculate the probability to calculate the chance of being in each state after one transition step.

If the model is state 0 after one transition, there are two ways it could have gotten there: by starting in state 0, then staying in state 0 with probability $1 - p$, or by starting in state 1 and transitioning to 0 with probability q. Similar arguments apply to being in state 1.

$$\pi_0^{(1)} = \pi_0^{(0)}(1-p) + \pi_1^{(0)} q$$
$$\pi_1^{(1)} = \pi_0^{(0)} p + \pi_1^{(0)}(1-q)$$

Writing these equation in matrix form,

$$\begin{pmatrix} \pi_0^{(1)} & \pi_1^{(1)} \end{pmatrix} = \begin{pmatrix} \pi_0^{(0)} & \pi_1^{(0)} \end{pmatrix} \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$
$$\pi^{(1)} = \pi^{(0)} P$$

Using row vectors for the π values is somewhat atypical (most linear algebra books use column vectors and place the matrix on the left), but is necessary because of how we defined the matrix P.

The Limiting Probabilities

If we know $\pi^{(0)}$ we can easily calculate any value of $\pi^{(n)}$, just by repeatedly multiplying by P.

$$\pi^{(n)} = \pi^{(0)} P^n$$

The matrix elements P_{ij}^n represent the probability of being in state j after n transitions, given that we started in state i.

What happens as $n \to \infty$? It isn't obvious, but it can be shown using induction that the matrix P will converge to a stable matrix with a single value in each column, and identical rows.[1]

$$\lim_{n\to\infty} P^n = \begin{pmatrix} \frac{q}{p+q} & \frac{p}{p+q} \\ \frac{q}{p+q} & \frac{p}{p+q} \end{pmatrix}$$

This is a surprising result! It turns out – at least for this simple example – that the probability of being in state 0 converges to a constant value of $\frac{q}{p+q}$, regardless of the starting state. Similarly, the probability of being in state 1 converges to $\frac{p}{p+q}$. Also note that these two probabilities sum to 1, so this is an actual distribution.

Let $\pi = \lim_{n\to\infty} \pi^{(n)}$ be a vector denoting the *limiting probability* of being in each state. If we want to calculate π, we can simply multiply the matrix P by itself until the values converge, then read the values from any one of the rows.

[1] M. Harchol-Balter, *Performance Modeling and Design of Computer Systems*, Ch. 8.

The Stationary Equations

If we don't want to perform repeated matrix multiplication, there is another way of obtaining the limiting probabilities. The vector $\pi = (\pi_0, \pi_1, \ldots, \pi_{N-1})$ is a *stationary distribution* for the Markov chain if it satisfies

$$\pi = \pi P$$
$$\sum_{j=0}^{N-1} \pi_j = 1$$

It's possible to prove theorems demonstrating that the solution to the stationary equations will also be the limiting distribution.

It's necessary to add the requirement that the elements of π sum to 1 to obtain a solution. Without the extra constraint, the system $\pi = \pi P$ is underdetermined – the equations are not linearly independent – and has infinitely many solutions.

Questions of Existence and Convergence

At several points thus far, we've taken limits without guaranteeing that the limit actually existed. In particular, there's no obvious reason why P^n should converge as $n \to \infty$, or that the long-run probability of being in each state should be stable.

In fact, there are a set of "good" properties for Markov chains. Any model satisfying these good properties will have a limiting distribution, which can be calculated by iteratively multiplying the P matrix, or by solving the stationary equations. For now, we'll assume that all of our chains have these properties, and we'll look at them in more detail in a future class.