CS 70 Spring 2017	Discrete Mathematics and Probability Theory Rao	DIS 12a

1 Markov Chain Basics

A Markov chain is a sequence of random variables X_n, $n = 0,1,2,\ldots$. Here is one interpretation of a Markov chain: X_n is the state of a particle at time n. At each time step, the particle can jump to another state. Formally, a Markov chain satisfies the Markov property:

$$\Pr(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \Pr(X_{n+1} = j \mid X_n = i), \tag{1}$$

for all n, and for all sequences of states $i_0, \ldots, i_{n-1}, i, j$. In other words, the Markov chain does not have any memory; the transition probability only depends on the current state, and not the history of states that have been visited in the past.

(a) In lecture, we learned that we can specify Markov chains by providing three ingredients: $\mathscr{X}$, P, and π_0. What do these represent, and what properties must they satisfy?

(b) If we specify $\mathscr{X}$, P, and π_0, we are implicitly defining a sequence of random variables X_n, $n = 0,1,2,\ldots$, that satisfies (1). Explain why this is true.

(c) Calculate $\Pr(X_1 = j)$ in terms of π_0 and P. Then, express your answer in matrix notation. What is the formula for $\Pr(X_n = j)$ in matrix form?

Solution:

(a) $\mathscr{X}$ is the set of states, which is the range of possible values for X_n. In this course, we only consider finite $\mathscr{X}$.

P contains the transition probabilities. $P(i,j)$ is the probability of transitioning from state i to state j. It must satisfy $\sum_{j\in\mathscr{X}} P(i,j) = 1\ \forall i \in \mathscr{X}$, which says that the probability that *some* transition occurs must be 1. Also, the entries must be non-negative: $P(i,j) \geq 0\ \forall i,j \in \mathscr{X}$. A matrix satisfying these two properties is called a stochastic matrix.

Note that we allow states to transition to themselves, i.e. it is possible for $P(i,i) > 0$.

π_0 is the initial distribution, that is, $\pi_0(i) = \Pr(X_0 = i)$. Similarly, we let π_n be the distribution of X_n. Since π_0 is a probability distribution, its entries must be non-negative and $\sum_{i\in\mathscr{X}} \pi_0(i) = 1$.

(b) The sequence of random variables X_n, $n = 0,1,2,\ldots$, is defined in the following way:

- X_0 has distribution π_0, i.e. $\Pr(X_0 = i) = \pi_0(i)$.
- X_{n+1} has distribution given by

$$\Pr(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \Pr(X_{n+1} = j \mid X_n = i) = P(i,j),$$

for all $n = 0,1,2,\ldots$.

It is important to realize the connection between the Markov property (1) and the transition matrix P. P contains information about the transition probabilities in one step. If the Markov property did not hold, then P would not be enough to specify the distribution of X_{n+1}. Conversely, if we only specify P, then we are implicitly assuming that the transition probabilities do not depend on anything other than the current state.

(c) By the Law of Total Probability,

$$\Pr(X_1 = j) = \sum_{i \in \mathscr{X}} \Pr(X_1 = j, X_0 = i) = \sum_{i \in \mathscr{X}} \Pr(X_0 = i) \Pr(X_1 = j \mid X_0 = i) = \sum_{i \in \mathscr{X}} \pi_0(i) P(i, j).$$

If we write $\pi_1(j) = \Pr(X_1 = j)$ and π_0 as row vectors, then in matrix notation we have

$$\pi_1 = \pi_0 P.$$

The effect of a transition is right-multiplication by P. After n time steps, we have

$$\pi_n = \pi_0 P^n.$$

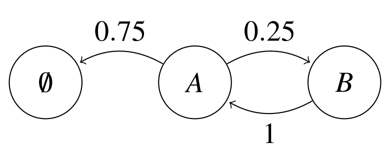
At this point, it should be mentioned that many calculations involving Markov chains are very naturally expressed with the language of matrices. Consequently, Markov chains are very well-suited for computers, which is one of the reasons why Markov chain models are so popular in practice.

2 Markov Conversation

Alice is hosting a party. As she's talking to her guests, she notices that conversations naturally transition between casual and more interesting topics. Consider the following simple model of conversations: Each type of topic takes a certain amount of time, and can transition to different topics as specified.

1. A Topics: These take 5 minutes. At the end, they can transition into a B topic (w.p. 25%), or the conversation can terminate (w.p. 75%).
2. B Topics: These take 16 minutes. At the end, they are always followed by an A Topic.

The following diagram illustrates the conversation flow, where "$\emptyset$" means the conversation has terminated, and "A", "B" correspond to the conversation topics.

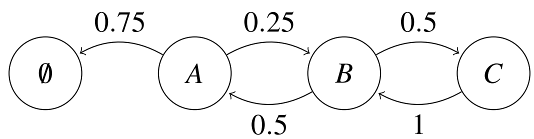


Using the above model:

(a) What is the expected length of a conversation that starts on an A topic?

As the party progresses, Alice revises her model of conversation to include three types of topics:

1. A Topics: These take 5 minutes. At the end, they can transition into a B topic (w.p. 25%), or the conversation can terminate (w.p. 75%).
2. B Topics: These take 15 minutes. At the end, they can be followed by a C topic (w.p. 50%), or can go back to an A topic (w.p. 50%).
3. C Topics: These take 25 minutes. They are always followed by a B topic.



Alice starts to wonder how the expected length of her conversations depend on who she talks to. Assume the following model: If she talks to acquaintances, they start on an A topic. With close friends, they start on a B topic (w.p. 50%), or on a C topic (w.p. 50%). Using the revised model:

(b) Alice starts talking to her acquaintance Bob. What is the expected length of their conversation?

(c) Alice starts talking to her close friend Charlie. What is the expected length of their conversation?

(d) Assume people at the party are equally likely to be close friends or acquaintances. But Eve noticed that Alice and Dave talked for 45 minutes (i.e. they reached state $\emptyset$ after 45 minutes)! What is the probability that Dave is a close friend of Alice?

Solution:

(a) Let $T_\emptyset$ denote the first time that we reach $\emptyset$. Formally, $T_\emptyset = \min\{n \in \mathbb{N} : X_n = \emptyset\}$. Write down the first-step equations. Let $\mathbf{E}_i[\cdot]$ denote the expectation if we start the Markov chain from state i, that is,

$$\mathbf{E}_i[\cdot] = \mathbf{E}[\cdot \mid X_0 = i].$$

Then, the first-step equations are

$$\begin{aligned}
\mathbf{E}_\emptyset[T_\emptyset] &= 0, \\
\mathbf{E}_A[T_\emptyset] &= 5 + \frac{3}{4}\mathbf{E}_\emptyset[T_\emptyset] + \frac{1}{4}\mathbf{E}_B[T_\emptyset], \\
\mathbf{E}_B[T_\emptyset] &= 16 + \mathbf{E}_A[T_\emptyset].
\end{aligned}$$

Substitute the third equation into the second equation.

$$\mathbf{E}_A[T_\emptyset] = 5 + \frac{1}{4}(16 + \mathbf{E}_A[T_\emptyset]) \implies \frac{3}{4}\mathbf{E}_A[T_\emptyset] = 9 \implies \mathbf{E}_A[T_\emptyset] = 12.$$

(b) We are looking for $\mathbf{E}_A[T_\emptyset]$, as before. The first-step equations are

$$\begin{aligned}
\mathbf{E}_\emptyset[T_\emptyset] &= 0, \\
\mathbf{E}_A[T_\emptyset] &= 5 + \frac{3}{4}\mathbf{E}_\emptyset[T_\emptyset] + \frac{1}{4}\mathbf{E}_B[T_\emptyset], \\
\mathbf{E}_B[T_\emptyset] &= 15 + \frac{1}{2}\mathbf{E}_A[T_\emptyset] + \frac{1}{2}\mathbf{E}_C[T_\emptyset], \\
\mathbf{E}_C[T_\emptyset] &= 25 + \mathbf{E}_B[T_\emptyset].
\end{aligned}$$

Substitute the last equation into the third equation.

$$\begin{aligned}
\mathbf{E}_B[T_\emptyset] = 15 + \frac{1}{2}\mathbf{E}_A[T_\emptyset] + \frac{1}{2}(25 + \mathbf{E}_B[T_\emptyset]) &\implies \frac{1}{2}\mathbf{E}_B[T_\emptyset] = 15 + \frac{25}{2} + \frac{1}{2}\mathbf{E}_A[T_\emptyset] \\
&\implies \mathbf{E}_B[T_\emptyset] = 55 + \mathbf{E}_A[T_\emptyset].
\end{aligned}$$

Now, substitute this equation into the equation for $\mathbf{E}_A[T_\emptyset]$.

$$\mathbf{E}_A[T_\emptyset] = 5 + \frac{1}{4}(55 + \mathbf{E}_A[T_\emptyset]) \implies \frac{3}{4}\mathbf{E}_A[T_\emptyset] = 5 + \frac{55}{4} \implies \mathbf{E}_A[T_\emptyset] = 25.$$

(c) From the previous part, we can substitute our solution for $\mathbf{E}_A[T_\emptyset]$ into the equation for $\mathbf{E}_B[T_\emptyset]$ to obtain

$$\mathbf{E}_B[T_\emptyset] = 55 + 25 = 80.$$

Then, we can substitute our answer for $\mathbf{E}_B[T_\emptyset]$ into $\mathbf{E}_C[T_\emptyset]$ to obtain

$$\mathbf{E}_C[T_\emptyset] = 25 + 80 = 105.$$

Since the conversation with a friend begins at state B with probability $1/2$ and at state C with probability $1/2$, the expected length of the conversation is

$$\frac{1}{2}\mathbf{E}_B[T_\emptyset] + \frac{1}{2}\mathbf{E}_C[T_\emptyset] = 92.5.$$

(d) There are only two possible conversation paths that take 45 minutes:

- Path 1: $A \to B \to A \to B \to A \to \emptyset$
- Path 2: $C \to B \to A \to \emptyset$

So the event that the conversation lasts 45 minutes is exactly the event that the conversation took one of these two paths. Further, Path 1 starts on a Casual topic (which only acquaintances would start on), and Path 2 starts on a Deep topic (which only close friends would start on). Therefore, letting the event "Path 1" be the event that the conversation took Path 1, the probability we are interested in is:

$$\Pr(\text{Path 2} \mid \text{Path 1 or Path 2})$$

It is easy to find $\Pr(\text{Path 1})$ if we first condition on the conversation starting on topic A:

$$\Pr(\text{Path 1}) = \Pr(\text{Path 1} \cap X_0 = A) = \Pr(\text{Path 1} \mid X_0 = A)\Pr(X_0 = A).$$

From the transition probabilities, we can find $p_1 = \Pr(\text{Path 1} \mid X_0 = A) = (1/4)^2(1/2)^2(3/4)$, since each $A \to B$ transition occurs with probability $1/4$, each $B \to A$ transition occurs with probability $1/2$, and $A \to \emptyset$ occurs with probability $3/4$.

Similarly, $p_2 = \Pr(\text{Path 2} \mid X_0 = C) = (1/2)(3/4)$.

So,

$$\begin{aligned}\Pr(\text{Path 2} \mid \text{Path 1 or Path 2}) &= \frac{p_2 \Pr(X_0 = C)}{p_1 \Pr(X_0 = A) + p_2 \Pr(X_0 = C)} = \frac{p_2(1/4)}{p_1(1/2) + p_2(1/4)} \\ &= \frac{p_2}{2p_1 + p_2} = \frac{16}{17} \approx 0.94.\end{aligned}$$