1 Markov Chain Basics

Note 22

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:

$$\mathbb{P}[X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0] = \mathbb{P}[X_{n+1} = j \mid X_n = i], \tag{1}$$

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 $\mathbb{P}[X_1 = j]$ in terms of π_0 and P. Then, express your answer in matrix notation. What is the formula for $\mathbb{P}[X_n = j]$ 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, $\pi_0(i) = \mathbb{P}[X_0 = i]$. 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. $\mathbb{P}[X_0 = i] = \pi_0(i)$.
- X_{n+1} has distribution given by

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

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. Note that this convention is different from what EE16A uses, if you have taken that class/are taking it right now.

(c) By the Law of Total Probability,

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

If we write $\pi_1(j) = \mathbb{P}[X_1 = j]$ 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 Can it be a Markov Chain?

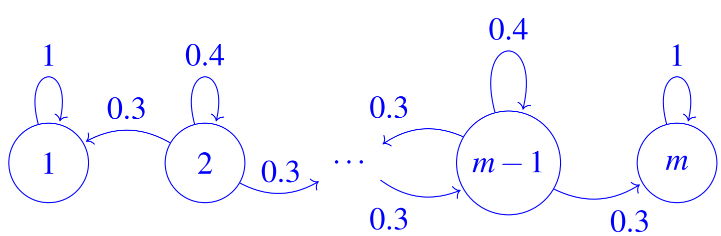
Note 22

(a) A fly flies in a straight line in unit-length increments. Each second it moves to the left with probability 0.3, right with probability 0.3, and stays put with probability 0.4. There are two spiders at positions 1 and m and if the fly lands in either of those positions it is captured. Given that the fly starts between positions 1 and m, model this process as a Markov Chain.

(b) Take the same scenario as in the previous part with $m = 4$. Let $Y_n = 0$ if at time n the fly is in position 1 or 2 and let $Y_n = 1$ if at time n the fly is in position 3 or 4. Is the process Y_n a Markov chain?

Solution:

(a) We can draw the Markov chain as such:



(b) No, because the memoryless property is violated.

For example, say $\mathbb{P}[X_0 = 2] = \mathbb{P}[X_0 = 3] = 1/2$ and $\mathbb{P}[X_0 = 1] = \mathbb{P}[X_0 = 4] = 0$. Then

$$\begin{aligned}
\mathbb{P}[Y_2 = 0 \mid Y_1 = 1, Y_0 = 0] &= \mathbb{P}[X_2 \in \{1,2\} \mid X_1 = 3, X_0 = 2] \\
&= \mathbb{P}[X_2 = 2 \mid X_1 = 3] = 0.3 \\
\mathbb{P}[Y_2 = 0 \mid Y_1 = 1, Y_0 = 1] &= \mathbb{P}[Y_2 = 0, Y_1 = 1, Y_0 = 1] / \mathbb{P}[Y_1 = 1, Y_0 = 1] \\
&= \mathbb{P}[X_2 = 2, X_1 = 3, X_0 = 3] / (\mathbb{P}[X_1 = 3, X_0 = 3] + \mathbb{P}[X_1 = 4, X_0 = 3]) \\
&= \frac{0.5 \cdot 0.4 \cdot 0.3}{0.5 \cdot 0.4 + 0.5 \cdot 0.3} = \frac{6}{35}
\end{aligned}$$

If Y was Markov, then $\mathbb{P}[Y_2 = 0 \mid Y_1 = 1, Y_0 = 0] = \mathbb{P}[Y_2 = 0 \mid Y_1 = 1] = \mathbb{P}[Y_2 = 0 \mid Y_1 = 1, Y_0 = 1]$. However, $0.3 > 6/35$, and so Y cannot be Markov.

3 Allen's Umbrella Setup

Note 22

Every morning, Allen walks from his home to Soda, and every evening, Allen walks from Soda to his home. Suppose that Allen has two umbrellas in his possession, but he sometimes leaves his umbrellas behind. Specifically, before leaving from his home or Soda, he checks the weather. If it is raining outside, he will bring exactly one umbrella (that is, if there is an umbrella where he currently is). If it is not raining outside, he will forget to bring his umbrella. Assume that the probability of rain is p.

(a) Model this as a Markov chain. What is $\mathscr{X}$? Write down the transition matrix.

(b) What is the transition matrix after 2 trips? n trips? Determine if the distribution of X_n converges to the invariant distribution, and compute the invariant distribution.

Solution:

(a) Let state i represent the situation that Allen has i umbrellas at his current location, for $i = 0$, 1, or 2.

Suppose Allen is in state 0. Then, Allen has no umbrellas to bring, so with probability 1 Allen arrives at a location with 2 umbrellas. That is,

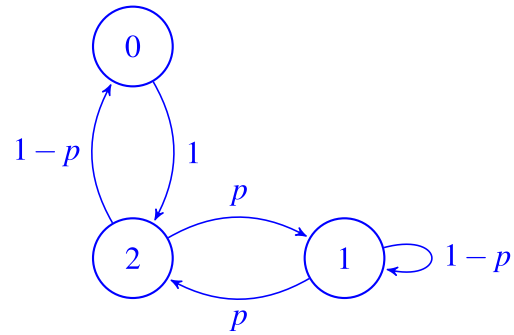
$$\mathbb{P}[X_{n+1} = 2 \mid X_n = 0] = 1.$$

Suppose Allen is in state 1. With probability p, it rains and Allen brings the umbrella, arriving at state 2. With probability $1 - p$, Allen forgets the umbrella, so Allen arrives at state 1.

$$\mathbb{P}[X_{n+1} = 2 \mid X_n = 1] = p, \qquad \mathbb{P}[X_{n+1} = 1 \mid X_n = 1] = 1 - p$$

Suppose Allen is in state 2. With probability p, it rains and Allen brings the umbrella, arriving at state 1. With probability $1 - p$, Allen forgets the umbrella, so Allen arrives at state 0.

$$\mathbb{P}[X_{n+1} = 1 \mid X_n = 2] = p, \qquad \mathbb{P}[X_{n+1} = 0 \mid X_n = 2] = 1 - p$$



We summarize this with the transition matrix

$$P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1-p & p \\ 1-p & p & 0 \end{bmatrix}.$$

(b) The transition matrices would be expressed as P^2 and P^n. Below we find the stationary distribution.

Observe that the transition matrix has non-zero element in its diagonal, which means the minimum number of steps to transit to state 1 from itself is one. Thus this transition matrix is irreducible and aperiodic, so it converges to its invariant distribution. To solve for the distribution, we set $\pi P = \pi$, or $\pi(P-I) = 0$. This yields the balance equations

$$\begin{bmatrix} \pi(0) & \pi(1) & \pi(2) \end{bmatrix} \begin{bmatrix} -1 & 0 & 1 \\ 0 & -p & p \\ 1-p & p & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}.$$

As usual, one of the equations is redundant. We replace the last column by the normalization condition $\pi(0) + \pi(1) + \pi(2) = 1$.

$$\begin{bmatrix} \pi(0) & \pi(1) & \pi(2) \end{bmatrix} \begin{bmatrix} -1 & 0 & 1 \\ 0 & -p & 1 \\ 1-p & p & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$$

Now solve for the distribution:

$$\begin{bmatrix} \pi(0) & \pi(1) & \pi(2) \end{bmatrix} = \frac{1}{3-p} \begin{bmatrix} 1-p & 1 & 1 \end{bmatrix}$$