CS 70 Discrete Mathematics and Probability Theory
Fall 2025 Sabin, Hug **Discussion 14A**

1 Markov Chains Intro I

Note 25

A **Markov chain** models an experiment with states, transitioning between states with some probability. A Markov chain is uniquely defined with the following variables:

- $\mathscr{X}$ is the set of possible states in the Markov chain. For this course, we'll only be working with Markov chains with a finite state space.
- X_n is a random variable denoting the state of the Markov chain at timestep n.
- P is the transition matrix. The element row i and column j in the matrix is defined as
$$P(i,j) = \mathbb{P}[X_{n+1} = j \mid X_n = i].$$
In particular, this is the probability that we transition from state i to state j.
- π_0 is the initial distribution; it is a row vector, where $\pi_0(i) = \mathbb{P}[X_0 = i]$. (Similarly, π_n is the distribution of states at timestep n; we have $\pi_n(i) = \mathbb{P}[X_n = i]$.)

Markov chains also have the **Markov property**:
$$\mathbb{P}[X_{n+1} = j \mid X_n = i, X_{n-1} = a_{n-1}, \ldots, X_0 = a_0] = \mathbb{P}[X_{n+1} = j \mid X_n = i].$$
That is, the next state depends only on the current state, and not on any prior states (this is also known as the memoryless property of Markov chains).

The **stationary distribution** (or the **invariant distribution**) of a Markov chain is the row vector π such that $\pi P = \pi$. (That is, transitioning does not change the distribution of states.)

A before B: Suppose we want to compute the probability of reaching state A before reaching state B. To compute this quantity, let $\alpha(i) = \mathbb{P}[A \text{ before } B \mid \text{at } i]$. Then, we have:
$$\begin{aligned}
\alpha(A) &= 1 \\
\alpha(B) &= 0 \\
\alpha(i) &= \sum_j P(i,j)\alpha(j)
\end{aligned}$$
Here, we use the law of total probability when computing $\alpha(i)$; we consider all possible transitions *out of* state i. These are called the **first step equations (FSE)**.

Hitting time: Suppose we want to compute the expected number of steps until you reach state A. To compute this quantity, let $\beta(i) = \mathbb{E}[\text{steps until } A \mid \text{at } i]$. Then, the first step equations become:
$$\begin{aligned}
\beta(A) &= 0 \\
\beta(i) &= 1 + \sum_j P(i,j)\beta(j)
\end{aligned}$$

Here, we use the law of total expectation when computing $\beta(i)$; we consider all possible transitions *out of* state i.

(a) Consider the transition matrix P of a Markov chain.

  (i) Is it always true that every *row* of P sums to the same value? If so, state this value and briefly explain why this makes sense. If not, briefly explain why.

  (ii) Is it always true that every *column* of P sums to the same value? If so, state this value and briefly explain why this makes sense. If not, briefly explain why.

(b) Compute $\mathbb{P}[X_1 = j]$ in terms of π_0 and P. Then, express your answer in matrix notation—that is, give an expression for the row vector π_1, where $\pi_1(j) = \mathbb{P}[X_1 = j]$. Generalize your answer to express π_n in matrix form in terms of n, π_0, and P.

(c) Note that we only need to provide $\mathscr{X}$, P, and π_0 in order to uniquely define a Markov chain; the random variables X_n are implicitly defined.

  (i) Explain how you can compute the distributions of the random variables X_n for $n \geq 0$ using only these parameters. (*Hint*: Part (b) can be helpful.)

  (ii) The Markov property is also implicit in this definition of a Markov chain. If the Markov property *does not hold*, are $\mathscr{X}$, P, and π_0 sufficient to compute the distributions of X_n for $n \geq 0$? Justify your answer.

Solution:

(a) (i) Yes, every row must sum to 1. Note that the element at row i and column j gives the probability of transitioning from state i to state j; the sum of the elements of a row gives us the sum of all transition probabilities *out* of state i. The fact that this must sum to 1 means that we will always transition from any given state to *some* next state—we must do something at every timestep.

  (ii) No, there are no restrictions on the sum of each column. The sum here would represent the sum of all transition probabilities *into* a state j, which has no inherent restrictions; these probabilities depend on the starting state, not the ending state.

(b) 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.

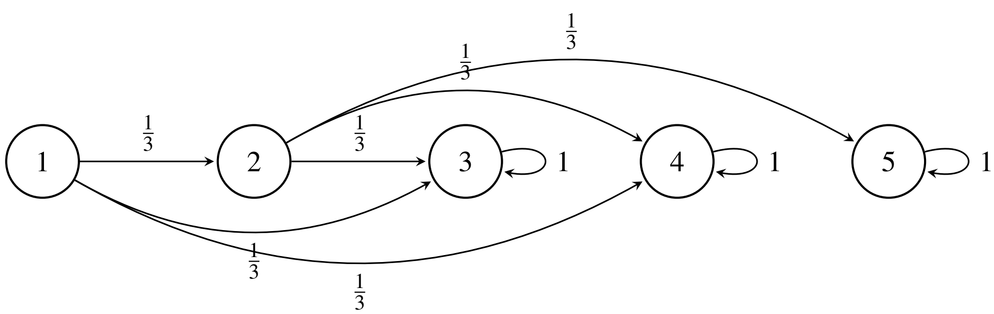
(c) (i) The important insight here is that π_n is exactly the distribution of X_n, where X_n takes on values in $\mathscr{X}$. We can compute the distribution of X_n as $\mathbb{P}[X_n = i] = \pi_n(i)$, where $\pi_n = \pi_0 P^n$.

(ii) If the Markov property does not hold, then P would not be sufficient to determine π_n from π_{n-1}; we'd need to know additional information about how the transition probabilities depend on the entire history of states.

2 Skipping Stones

Note 25 We consider a simple Markov chain model for skipping stones on a river, but with a twist: instead of trying to make the stone travel as far as possible, you want the stone to hit a target. Let the set of states be $\mathscr{X} = \{1,2,3,4,5\}$. State 3 represents the target, while states 4 and 5 indicate that you have overshot your target. Assume that from states 1 and 2, the stone is equally likely to skip forward one, two, or three steps forward. If the stone starts from state 1, compute the probability of reaching our target before overshooting, i.e. the probability of $\{3\}$ before $\{4,5\}$.

Solution: Here is the Markov Chain we are working with:



Let $\alpha(i)$ denote the probability of reaching the target before overshooting, starting at state i. Then:

$$
\begin{aligned}
\alpha(5) &= 0 \\
\alpha(4) &= 0 \\
\alpha(3) &= 1 \\
\alpha(2) &= \frac{1}{3}\alpha(3) + \frac{1}{3}\alpha(4) + \frac{1}{3}\alpha(5) = \frac{1}{3} \\
\alpha(1) &= \frac{1}{3}\alpha(2) + \frac{1}{3}\alpha(3) + \frac{1}{3}\alpha(4) = \frac{1}{9} + \frac{1}{3}
\end{aligned}
$$

Therefore, $\alpha(1) = 1/9 + 1/3 = 4/9$.

3 Consecutive Flips

Note 25 Suppose you are flipping a fair coin (one Head and one Tail) until you get the same side 3 times (Heads, Heads, Heads) or (Tails, Tails, Tails) in a row.

(a) Construct an Markov chain that describes the situation with a start state and end state.

(b) Given that you have flipped a (Tails, Heads) so far, what is the expected number of flips to see the same side three times?

(c) What is the expected number of flips to see the same side three times, beginning at the start state?

Solution:

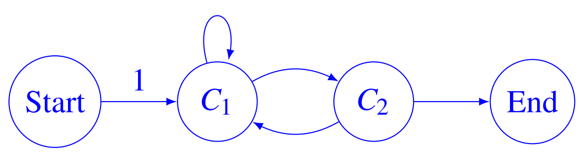
There are two approaches to this question, both of which are equivalent. The first approach, with 4 states, is significantly easier to work with, but may be less intuitive to come up with. The second approach, with 6 states, may be easier to come up with intuitively, but the increased number of states makes computations a little more tedious.

Approach 1: 4 states

(a) We can model this process as a Markov chain with 4 states: Start, C_1, C_2, and End, denoting the number of consecutive flips so far.

- When starting out, we can either flip a heads or a tails, so with probability 1 we transition to C_1, where we have one consecutive flip so far.
- If we're at C_1, we've flipped only one of either heads or tails consecutively, so there are equal transitions to C_2 (if we flip the same as before) or to itself (if we flip something different).
- If we're at C_2, we've flipped two heads or tails consecutively, so there are equal transitions to End (if we flip the same as before) or to C_1 (if we flip something different).

The Markov chain is illustrated below; all transitions have probability $\frac{1}{2}$ unless otherwise marked. (The self loop at the End state is omitted for clarity.)



(b) If we flipped a tails and then a heads, we are currently in the C_1 state (since we last flipped only one head in a row). Thus, we must calculate the expected number of flips to get to End from C_1.

Let $\beta(i)$ denote the expected number of flips to reach End, given that we start from state i. The first step equations give us the following system:

$$\begin{aligned}
\beta(C_1) &= 1 + \frac{1}{2}\beta(C_1) + \frac{1}{2}\beta(C_2) \\
\beta(C_2) &= 1 + \frac{1}{2}\beta(C_1) + \frac{1}{2}\beta(\text{End}) \\
\beta(\text{End}) &= 0
\end{aligned}$$

If we solve this system of equations, we get $\beta(C_1) = 6$ and $\beta(C_2) = 4$.

We wanted to find $\beta(C_1)$, which gives us an answer of 6 flips.

(c) Here, we want to find $\beta(\text{Start})$, which can be computed as

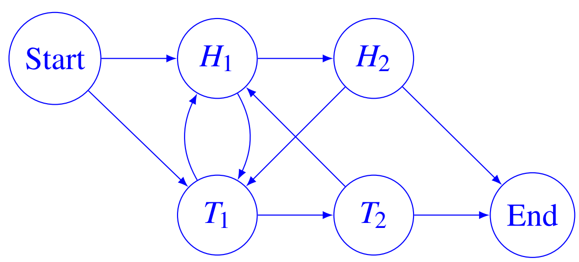
$$\beta(\text{Start}) = 1 + \beta(C_1) = 1 + 6 = 7.$$

Approach 2: 6 states

(a) We can model this process as a Markov chain with 6 states: Start, H_1, H_2, T_1, T_2, and End, denoting the number of times we've flipped either heads or tails so far.

- When starting out, we can either flip a heads or tails, so there are equal transitions to H_1 and T_1.
- If we're at H_1, we've flipped one heads so far, so there are equal transitions to H_2 (if heads) and T_1 (if tails).
- If we're at H_2, we've flipped two heads so far, so there are equal transitions to End (if heads) and T_1 (if tails).
- If we're at T_1, we've flipped one tails so far, so there are equal transitions to H_1 (if heads) and T_2 (if tails).
- If we're at T_2, we've flipped two tails so far, so there are equal transitions to H_1 (if heads) and End (if tails).

The Markov chain is illustrated below; all transitions have probability $\frac{1}{2}$. (The self loop at the End state is omitted for clarity.)



(b) If we flipped a tails and then a heads, we are currently in the H_1 state (since we last flipped only one head in a row). Thus, we must calculate the expected number of flips to get to End from H_1.

Let $\beta(i)$ denote the expected number of flips to reach End, given that we start from state i.

The first step equations give us the following system:

$$\begin{aligned}
\beta(H_1) &= 1 + \frac{1}{2}\beta(T_1) + \frac{1}{2}\beta(H_2) \\
\beta(H_2) &= 1 + \frac{1}{2}\beta(\text{End}) + \frac{1}{2}\beta(T_1) \\
\beta(T_1) &= 1 + \frac{1}{2}\beta(T_2) + \frac{1}{2}\beta(H_1) \\
\beta(T_2) &= 1 + \frac{1}{2}\beta(\text{End}) + \frac{1}{2}\beta(H_1) \\
\beta(\text{End}) &= 0
\end{aligned}$$

If we solve this system of equations, we get $\beta(H_1) = 6, \beta(H_2) = 4, \beta(T_1) = 6, \beta(T_2) = 4$.

We wanted to find $\beta(H_1)$, which gives us an answer of 6 flips.

(c) Here, we want to find $\beta(\text{Start})$, which can be computed as

$$\beta(\text{Start}) = 1 + \frac{1}{2}\beta(H_1) + \frac{1}{2}\beta(T_1) = 1 + \frac{1}{2}\cdot 6 + \frac{1}{2}\cdot 6 = 7.$$

4 Can it be a Markov Chain?

Note 25

(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.

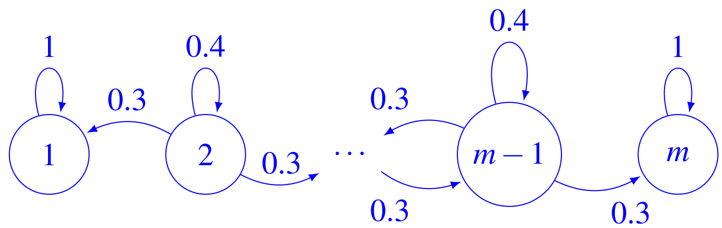
Given that the fly starts at state i, where $1 < i < m$, model this process as a Markov Chain. (Don't forget to specify the initial distribution!)

(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.

Provide the state space for Y_n. Is the process Y_n a Markov chain?

Solution:

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



The initial distribution is $\pi_0(i) = 1$, and $\pi_0(j) = 0$ for $j \neq i$.

(b) The state space is $\{0,1\}$, the set of possible values that Y_n can take on.

Y_n cannot be a Markov chain 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.