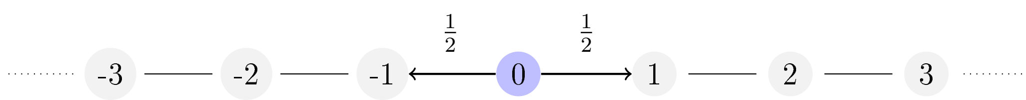
CS5120: Probability & Computing **Jan-May 2020**

Markov Chains 1

1 Random walks and Markov chains

Let's begin with some examples.

Example 1: As a simple example of a random walk, imagine starting at 0 on the real line. After every discrete time step, you move from your current position with probability 1/2 to the integer on the left and with probability 1/2 to the integer on the right.

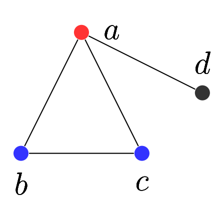


Exercise 1: Find the probability of being at 0 after two steps. What other positions can you be in after two steps, and with what probability?

More generally, the following are two problems that we would like to solve.

- Given a number t of steps, find the probability distribution of being at various positions after t steps.
- Find the expected number of time-steps to reach a given position.

Example 2: A random walk can be defined on any undirected graph, as follows: you start at an initial vertex v_0, and from any vertex v, you can move to each of the neighbors of v with equal probability, that is: $\dfrac{1}{deg(v)}$.



In the example above, let's suppose that the initial vertex in the random walk is a. We can then define a sequence of variables $X_0, X_1, X_2, \ldots$, where

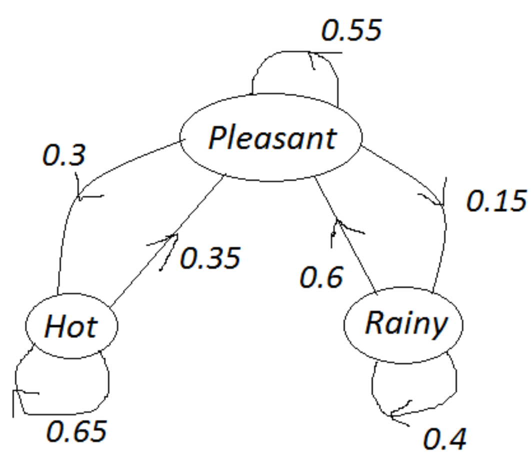
$X_0 = a$, and for $i \geq 1$, X_i is the vertex after i steps of the random walk, with the probability distribution of X_{i+1} being completely determined by the value of X_i.

Some example values are: $Pr[X_1 = b] = Pr[X_1 = c] = Pr[X_1 = d] = 1/3$, and $Pr[X_2 = a] = 1/3 \cdot 1/2 + 1/3 \cdot 1/2 + 1/3 \cdot 1 = 2/3$.

Example 3: A natural example of a random walk is surfing on the internet (or sites like Wikipedia and youtube), where an user may keep following links somewhat randomly. However, you'll notice that not all links are equally likely to be clicked in practice.

Thus we may consider a more general model of a random walk, where the probabilities of visiting two different neighbors from the current vertex are not necessarily equal. This generalization is called a Markov chain, and a classical example is the following.

Example 4: The following is a simplistic model of weather prediction, in which every day can be either hot, pleasant or rainy. The probability distribution of a given day's weather is determined by the previous day's. For example, if today is hot, then tomorrow is hot with probability 0.65, and pleasant with probability 0.35.



The probablities of transitions can be represented as a matrix, as shown below. The rows and columns are indexed by the states in the order Hot, Pleasant, Rainy.

2 The transition probability matrix

The following is the transition probability matrix for the above example.

$$M = \begin{bmatrix} 0.65 & 0.35 & 0 \\ 0.3 & 0.55 & 0.15 \\ 0 & 0.6 & 0.4 \end{bmatrix}$$

Our working definition of discrete, finite-state Markov chains will be the following.

Definition A discrete, finite-state Markov chain consists of a finite set of states: $\Omega = \{s_1, \ldots, s_n\}$, a transition probability matrix: $(M_{i,j})$, and a sequence $X_0, X_1, X_2, \ldots$ of random variables, satisfying the following condition.

$$Pr\left(X_{t+1} = s_j | X_t = s_i, X_{t-1} = s_{i-1}, \ldots, X_0 = s_0\right) = Pr(X_{t+1} = s_j | X_t = s_i) = M_{i,j}.$$

The condition above is called the memory-less property and it is what characterizes a Markov chain: the probability distribution of X_{t+1} is determined completely by the value of X_t. Note that X_{t+1} is not independent of the values $X_{t-1}, X_{t-2}, \ldots$; rather, all the dependence is captured by X_t, or formally, $X_{t+1}|X_t$ is independent of the values $X_{t-1}, X_{t-2}, \ldots$.

What about the initial state X_0? We will assume that it is chosen according to some initial probability distribution $\pi_0 = (\pi_0(1), \ldots \pi_0(n))$, where $\pi_0(i)$ is the initial probability of being at state i.

2.1 Calculation of probability distribution after t steps

Let $\pi_t = (\pi_t(1), \ldots, \pi_t(n))$ be the probability distribution after t steps. For Example 4, we assumed that $\pi_0 = (1, 0, 0)$.

We then have, for $i \geq 1$:

$$Pr[X_t = s_i] = \sum_j Pr\left(s_i \text{ was reached from } s_j \text{ in the last step}\right) Pr\left(X_{t-1} = s_j\right)$$

and thus:

$$\pi_t(i) = \sum_j \pi_{t-1}(j) M_{ji} \quad \forall i \geq 1. \tag{1}$$

Notice that the RHS is the inner (dot) product of the (row) vector π_{t-1} and the ith column of M. Thus, we can write the above set of equations as: $\pi_t = \pi_{t-1} \cdot M$.

Thus $\pi_t = \pi_0 M^t$. Let's see some examples.

$$\pi_1 = (1,0,0) \cdot \begin{bmatrix} 0.65 & 0.35 & 0 \\ 0.3 & 0.55 & 0.15 \\ 0 & 0.6 & 0.4 \end{bmatrix} = (0.65, 0.35, 0).$$

$$\pi_2 = (0.65, 0.35, 0) \cdot \begin{bmatrix} 0.65 & 0.35 & 0 \\ 0.3 & 0.55 & 0.15 \\ 0 & 0.6 & 0.4 \end{bmatrix} = (0.5275, 0.42, 0.0525).$$

To write the probability distribution as a column vector, left-multiply by **the transpose** of M.

$$\pi_2^T = \begin{bmatrix} 0.65 & 0.3 & 0 \\ 0.35 & 0.55 & 0.6 \\ 0 & 0.15 & 0.4 \end{bmatrix} \cdot \begin{pmatrix} 0.65 \\ 0.35 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.5275 \\ 0.42 \\ 0.0525 \end{pmatrix}$$

Exercise 2: Calculate π_{10} with $\pi_0 = (1,0,0)$, and with $\pi_0 = (1/3, 1/3, 1/3)$. You may use a program/software. Repeated squaring or diagonalizing the matrix also help. What do you observe?

Exercise 3: Suppose that we toss a biased coin (say $Pr[H] = p$) repeatedly, and the state is the result of the most recent toss (H/T). Model this as a Markov chain.

3 Expected time to reach a state

Example 5: Consider an undirected graph $G = (V, E)$, where $V = \{a, b, c\}$ and $E = \{ab, bc\}$. Suppose that we start a random walk at the vertex b. What is the expected number of steps to reach a?

To answer this, we define, for two vertices u, v the quantity T_{uv} as the expected number of steps to reach v from u. It turns out that we can formulate linear recurrences involving these values, and hence solve for them.

In the above example, we want to find T_{ba}. We have: $T_{ba} = \frac{1}{2}(1) + \frac{1}{2}(1 +$

T_{ca}), because with probability $1/2$, we reach a from b in one step, and with probability $1/2$, we move to c in one step, from which the expected time to reach a is T_{ca}.

Similarly, we have: $T_{ca} = 1+T_{ba}$. Substituting this in the previous recurrence, we get: $T_{ba} = 1 + \frac{1}{2}(1 + T_{ba})$, which yields $T_{ba} = 3$.

Exercise 4: In Example 4, write recurrences for T_{HR} and T_{PR}. [H, P, R stand for Hot, Pleasant, Rainy.]

Thus calculate T_{HR}, the expected number of days until Rain, assuming that today is a Hot day.

Exercise 5: Find T_{ac} for the graph below.