EE 381 Probability & Statistic with Applications to Computing (Fall 2020)

Lecture 13
Markov Chains

Duc H. Tran, PhD

Markov Chains

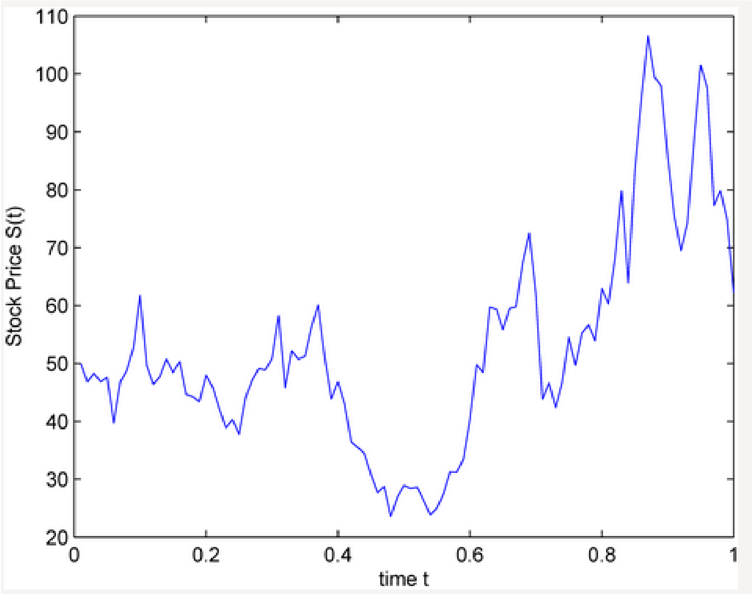
- Basic Concepts of Random Processes
- Markov Chains
  - ✓ Introduction
  - ✓ State Transition Matrix and Diagram
  - ✓ Probability Distributions
  - ✓ Classification of States
  - ✓ Stationary Distributions

Random Processes

In real-life applications, people are interested in multiple observations of random values over a period of time.

Example: observing the stock price of a company over the next few months.

Let $S(t)$ be the stock price at time $t \in [0, \infty)$, assuming $t=0$ refers to current time.



A possible realization of values of a stock observed as a function of time.

Random Processes

At any fixed time $t_1, t_2 \in [0, \infty)$, $S(t_1)$, $S(t_2)$ are a random variables.

When we consider the values of *S(t)* for $t \in [0, \infty)$, we say that *S(t)* is a random process or a stochastic process, defined by:

$$\{S(t), t \in [0, \infty)\}$$

Therefore, a random process is a collection of random variable usually indexed by time (or sometimes by space)

Random Processes

There are 2 types of random processes:

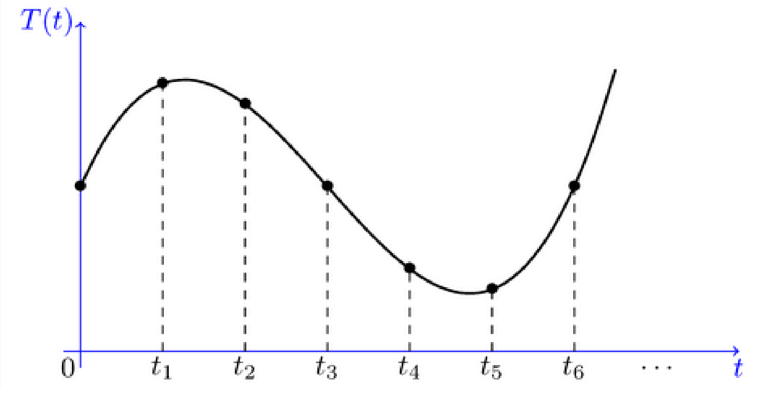
1) Continuous-time random process: is a random process $\{X(t), t \in J\}$, where J is an interval on the real line such as$[-1,1]$, $[0, \infty)$, $(-\infty, \infty)$, etc.

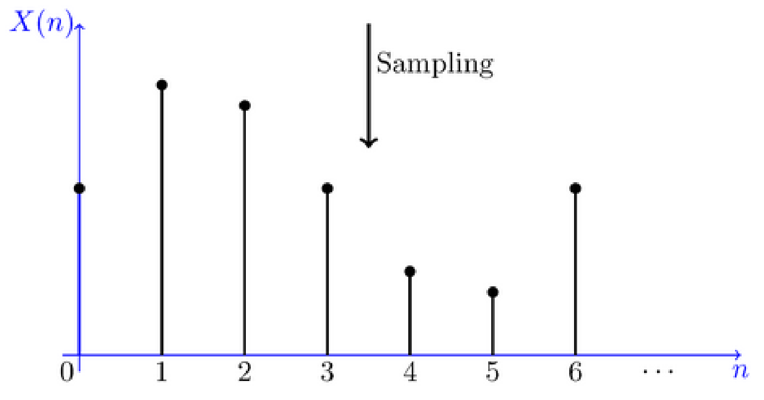
Ex.: Let *T(t)* be the temperature in NYC at time $t \in [0, \infty)$, assuming *t* is measured in hours and *t=0* refers to the time we start measuring the temperature.

2) Discrete-time random process: is a random process $\{X(n) = X_n, n \in J\}$, where *J* is a countable set such as *N* or *Z*.

Ex.: Let X_n for $n \in N$ be the amount of time the *i*th customer spends at the bank.

Discrete-time processes are sometimes obtained form continuous-time processes by discretizing time (sampling at specific time)





Random Processes as Random Functions

- Consider a random process $\{X(t), t \in J\}$, resulted from a random experiment (e.g., observing the stock prices of a company).

Recall: any random experiment is defined on a sample space *S*.

- After observing the values of *X(t)*, we obtain a function of time, which is just one of the many possible outcomes of this random experiment.
- These possible functions of *X(t)* is called **sample function** or **sample path**.

→A random process can be thought of as a random function of time.

- $\{X(t), t \in J\}$ will be equal to one of many possible sample functions after we are done with our random experiment.
- In engineering applications, random processes are often referred to as **random signals**.

Random Processes as Random Functions

Example: You have 1000 dollars to put in an account with interest rate R, compounded annually. That is, if X_n is the value of the account at year n, then

$$X_n = 1000(1+R)^n, \qquad for\ n = 0,1,2\ ...$$

The value of R is a random variable that is determined when you put the money in the bank, but it does not change after that. In particular, assume that $R \sim Uniform(0.04, 0.05)$

a) Find all possible sample functions for the random process $\{X(n) = X_n, n = 0,1,2\ ...\}$

Here, the randomness in X_n comes from the random variable R. If R=r, then

$$X_n = 1000(1+r)^n, \qquad for\ n = 0,1,2\ ...$$

Thus, the sample functions are of the form $\mathrm{f(n)} = 1000(1+r)^n$, n=0,1,2,… where r $\in$ [0.04,0.05].

In other words, for any r $\in$ [0.04,0.05], you obtain a sample function for the random process X_n.

Random Processes as Random Functions

Example (ctn.):

b) Find the expected value of your account at year three. That is, find $E[X_3]$.

The random process X_3 is given by: $X_3 = 1000(1+R)^3$

If you let Y=1+R, then $Y \sim Uniform(1.04, 1.05)$, so its PDF is:

$$f_Y(y) = \begin{cases} 100 & 1.04 \le y \le 1.05 \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} E[X_3] &= 1000E[Y^3] \\ &= 1000 \int_{1.04}^{1.05} 100y^3 \quad dy \\ &= \frac{10^5}{4}\left[y^4\right]_{1.04}^{1.05} \\ &= \frac{10^5}{4}\left[(1.05)^4 - (1.04)^4\right] \\ &\approx 1,141.2 \end{aligned}$$

Markov Chains: Introduction

- Consider a discrete-time random process $\{X_m,\ m = 0,1,2, \dots\}$. Simple case: X_m's are independent → analysis of this process if straightforward. There is no "memory" in the system, and each X_m can be looked at independently from previous ones.
- Independence assumption is not valid for many real-life processes → reasonable to assume that X_m's are dependent → models where value of X_m depends on the previous values is needed.
- In Markov Chains, X_{m+1} depends on X_m. It does not depend on the other previous values X_0 , $X_1, \dots , X_{m-1}$.

Markov Chains: Introduction

- Markov Chains are usually used to model the evolution of “states” in probabilistic system.

Example: Consider a system with the set of possible states $S = \{0,1,2, \dots\}$. If $X_n = i$, we say that the system is in state *i* at time *n*.

Example: suppose we are modeling a queue at a bank. The number of people in the queue is a non-negative integer. The state of the system is the number of people in the queue.

If X_n shows the number of people in the queue at time *n*, then $X_n \in S = \{0,1,2, \dots\}$. The set S is called the state space of the Markov Chain.

Markov Chains: Introduction

Discrete-time Markov Chains:

- Consider the random process $\{X_n, n = 0,1,2,\dots\}$, where $R_{X_i} = S \subset \{0,1,2,\dots\}$. We say that this process is a Markov Chain if:

$$P(X_{m+1} = j | X_m = i, X_{m-1} = i_{m-1}, \cdots, X_0 = i_0) \\ = P(X_{m+1} = j | X_m = i),$$

for all $m, j, i, i_0, i_1, \dots, i_{m-1}$. If the number of states is finite, e.g., $S = \{0,1,2,\dots,r\}$, we call it a finite Markov Chain.

- The numbers $P(X_{m+1} = j | X_m = i)$ are called the transition probabilities and do not depend on time. We can define it as:

$$p_{ij} = P(X_{m+1} = j | X_m = i)$$

In particular:

$$\begin{aligned} p_{ij} &= P(X_1 = j | X_0 = i) \\ &= P(X_2 = j | X_1 = i) \\ &= P(X_3 = j | X_2 = i) = \cdots. \end{aligned}$$

In other words, if the process is in state i, it will next make a transition to state j with probability p_{ij}.

Markov Chains: State Transition Matrix and Diagram

- We often list the transition probabilities in a matrix, which is called the state transition matrix or transition probability matrix. Assuming the states are $\{0,1,2,\ldots,r\}$, then the state transition matrix is given by:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1r} \\ p_{21} & p_{22} & \cdots & p_{2r} \\ \vdots & \vdots & \vdots & \vdots \\ p_{r1} & p_{r2} & \cdots & p_{rr} \end{bmatrix}$$

- Note that $p_{ij} \geq 0$, and for all i, we have:

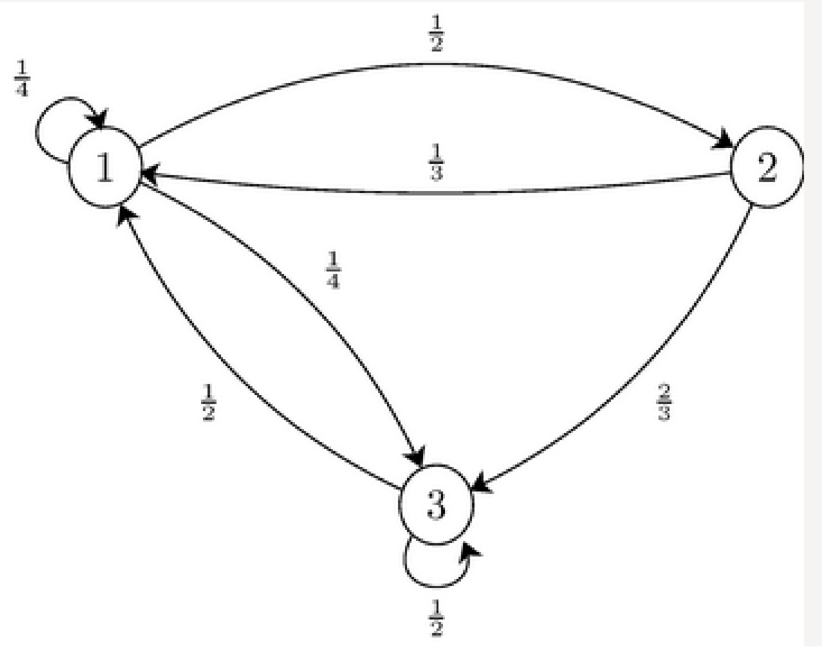
$$\sum_{k=1}^{r} p_{ik} = \sum_{k=1}^{r} P(X_{m+1} = k | X_m = i) \\ = 1.$$

- In other words, the rows of any state transition matrix must sum to 1.

Markov Chains: State Transition Matrix and Diagram

- A Markov Chain is usually shown by a state transition diagram.
- Consider a Markov chain with 3 possible states **1, 2, 3** and the following transition probabilities:

$$P = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

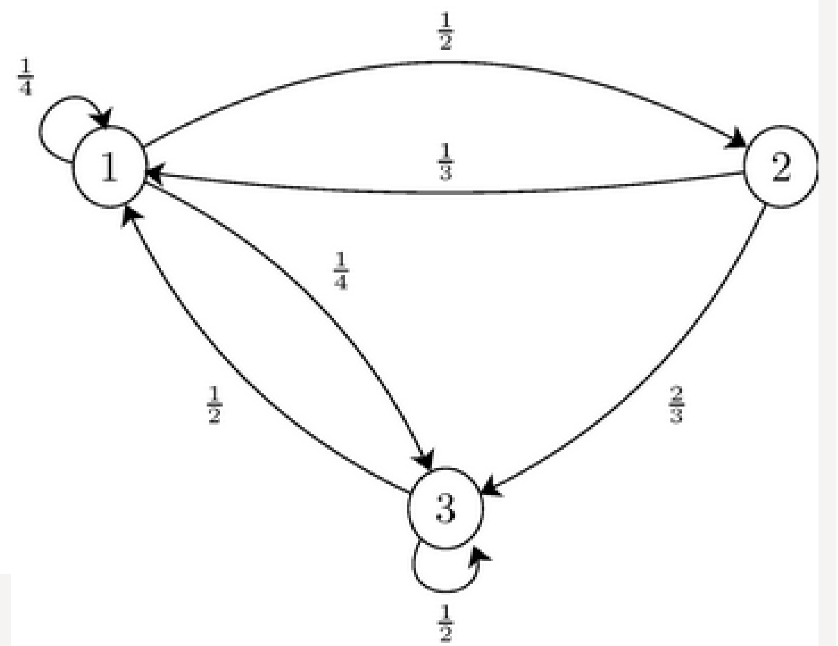


- When there is no arrow from state i to state j, it means that $p_{ij} = 0$.

Markov Chains: State Transition Matrix and Diagram

- **Example**:

Consider the following Markov Chain:



a) Find $P(X_4 = 3|X_3 = 2)$

$$P(X_4 = 3|X_3 = 2) = p_{23} = \frac{2}{3}.$$

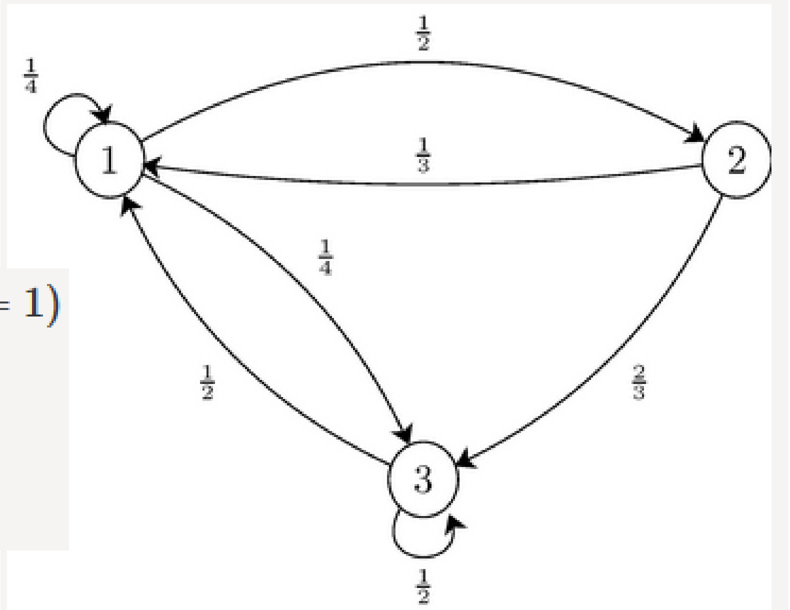
b) Find $P(X_3 = 1|X_2 = 1)$

$$P(X_3 = 1|X_2 = 1) = p_{11} = \frac{1}{4}.$$

Markov Chains: State Transition Matrix and Diagram

- **Example (ctn.)**:

c) If $P(X_0 = 1) = \frac{1}{3}$, find $P(X_0 = 1, X_1 = 2)$



$$\begin{aligned} P(X_0 = 1, X_1 = 2) &= P(X_0 = 1)P(X_1 = 2 | X_0 = 1) \\ &= \frac{1}{3} \cdot p_{12} \\ &= \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}. \end{aligned}$$

d) If $P(X_0 = 1) = \frac{1}{3}$, find $P(X_0 = 1, X_1 = 2, X_2 = 3\ )$

$$\begin{aligned} &P(X_0 = 1, X_1 = 2, X_2 = 3) \\ &\quad = P(X_0 = 1)P(X_1 = 2 | X_0 = 1)P(X_2 = 3 | X_1 = 2, X_0 = 1) \\ &\quad = P(X_0 = 1)P(X_1 = 2 | X_0 = 1)P(X_2 = 3 | X_1 = 2) \quad \text{(by Markov property)} \\ &\quad = \frac{1}{3} \cdot p_{12} \cdot p_{23} \\ &\quad = \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{2}{3} \\ &\quad = \frac{1}{9}. \end{aligned}$$

Markov Chains: State Probability Distributions

n-step transition probabilities: p_{ij} gives us the probability of going from state i to state j in one step. How about the probability of going from state i to state j in two steps, i.e.,

$$p_{ij}^{(2)} = P(X_2 = j | X_0 = i)$$

We can write:

$$\begin{aligned} p_{ij}^{(2)} = P(X_2 = j | X_0 = i) &= \sum_{k \in S} P(X_2 = j | X_1 = k, X_0 = i) P(X_1 = k | X_0 = i) \\ &= \sum_{k \in S} P(X_2 = j | X_1 = k) P(X_1 = k | X_0 = i) \quad \text{(by Markov property)} \\ &= \sum_{k \in S} p_{kj} p_{ik}. \end{aligned}$$

$$p_{ij}^{(2)} = P(X_2 = j | X_0 = i) = \sum_{k \in S} p_{ik} p_{kj}$$

Explanation: In order to get to state j, we need to pass through some intermediate state k. The probability of this event is $p_{ik}p_{kj}$

Markov Chains: State Probability Distributions

To obtain $p_{ij}^{(2)}$, we sum over all possible intermediate states. Accordingly, we can define the two-step transition matrix as figure 1.

In fact, we notice that $p_{ij}^{(2)}$ is the element in the *i*th row and *j*th column of the matrix P^2 (figure 2).

$$P^{(2)} = \begin{bmatrix} p_{11}^{(2)} & p_{12}^{(2)} & \dots & p_{1r}^{(2)} \\ p_{21}^{(2)} & p_{22}^{(2)} & \dots & p_{2r}^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ p_{r1}^{(2)} & p_{r2}^{(2)} & \dots & p_{rr}^{(2)} \end{bmatrix}$$

Figure 1: $P^{(2)}$

$$P^2 = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1r} \\ p_{21} & p_{22} & \dots & p_{2r} \\ \vdots & \vdots & \vdots & \vdots \\ p_{r1} & p_{r2} & \dots & p_{rr} \end{bmatrix} \cdot \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1r} \\ p_{21} & p_{22} & \dots & p_{2r} \\ \vdots & \vdots & \vdots & \vdots \\ p_{r1} & p_{r2} & \dots & p_{rr} \end{bmatrix}$$

Figure 2: P^2

We conclude that the two-step transition matrix can be obtained by squaring the state transition matrix: $P^{(2)} = P^2$ → very helpful for calculating the probability that the system is in state *i* at time *n*.

Markov Chains: State Probability Distributions

Consider a Markov chain $\{X_n, n = 0,1,2, \ldots\}$, where $X_n \in S = \{1,2, \ldots, r\}$. Suppose that we know the probability distribution of X_0. More specifically, define the row vector $\pi^{(0)}$ (initial state vector) as:

$$\pi^{(0)} = [\, P(X_0 = 1) \quad P(X_0 = 2) \quad \cdots \quad P(X_0 = r) \,]$$

We can use the law of total probability to obtain the probability distribution of X_1, X_2,…

$$\pi^{(n+1)} = \pi^{(n)} P, \;\; \text{for } n = 0, 1, 2, \cdots;$$
$$\pi^{(n)} = \pi^{(0)} P^n, \;\; \text{for } n = 0, 1, 2, \cdots.$$

Markov Chains: State Probability Distributions

Example: consider a system that can be one of two possible states, S={0,1}.

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

Suppose that the system is in state **0** at time *n=0*, i.e., $X_0 = 0$. Find the probability that the system is in state **1** at time *n=3*.

Solution: We know that:

$$\pi^{(0)} = \begin{bmatrix} P(X_0 = 0) & P(X_0 = 1) \end{bmatrix} \\ = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

Thus,

$$\pi^{(3)} = \pi^{(0)} P^3 \\ = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}^3 \\ = \begin{bmatrix} \frac{29}{72} & \frac{43}{72} \end{bmatrix}.$$

The probability that the system is in state **1** at time *n=3* is 43/72.

Markov Chains: Absorbing states

An absorbing state is a recurrent state from which it cannot escape once it get to it, $p_{kk} = 1$. This means you get absorbed by a state.

The probability to reach a given absorbing state *s* starting from any state *i* of a Markov Chain with *m* states is given by: $a_i = \sum_{j=1}^{m} p_{ij} a_j$

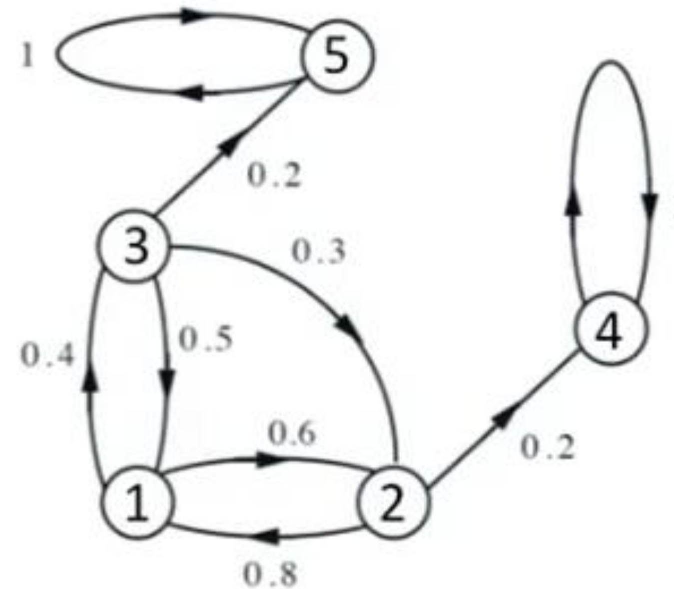
Note: $a_s = 1$ and $a_{s'} = 0$ for other absorbing states.

Example: determine the probability a_i that the chain eventually settles in state *4* given it started in state *i*?

$i = 4, a_4 = 1$

$i = 5, a_5 = 0$

$a_1 = \frac{18}{28}, a_2 = \frac{20}{28}, a_3 = \frac{15}{28}$



Markov Chains: Mean hitting and return times

Mean hitting times: The expected time until the chain/process (starting from a certain state) hits a certain state for the first time. It is also called the *mean first passage time*.

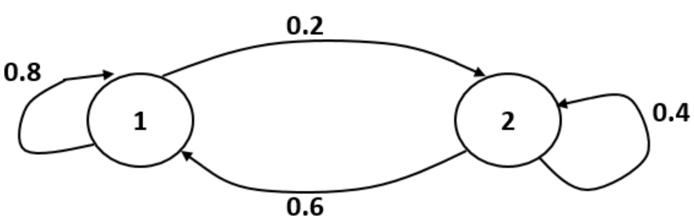
The mean hitting times is defined by the following equation:

$$t_i = 1 + \sum_j t_j p_{ij}$$

Example: Determine the mean hitting times of state 1 for this chain starting from state 2.

$$t_2 = 1 + \sum_j t_j p_{2j} = 1 + t_1 p_{21} + t_2 p_{22}$$

So $t_2 = \frac{5}{3}$



Markov Chains: Mean hitting and return times

Mean return times: The expected time for the chain/process returns to a certain state for the first time if it started from that state. It is also called the *mean recurrence time*.

The mean return times is defined by the following equation:

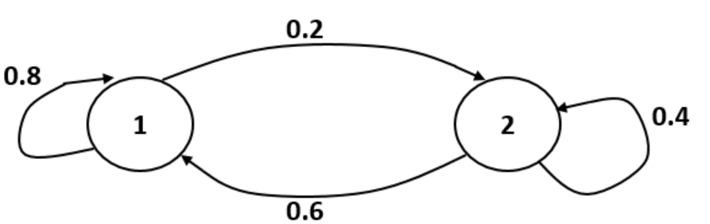
$$r_l = 1 + \sum_k t_k p_{lk}$$

where t_k is the expected time until the chain hits state *l*.

Example: Determine the mean return times of state 1 for this chain starting from state 1.

$$r_1 = 1 + \sum_k t_k p_{1k} = 1 + t_1 p_{11} + t_2 p_{12} = 1 + 0.2 * \frac{5}{3}$$

So $r_1 = \frac{4}{3}$



Markov Chains: Steady State Conditions

- Steady state conditions: the initial state vector that will have no effect when multiply it with the state transition matrix of the chain.

$$v * P = v$$

Example: $[x \quad y] * \begin{bmatrix} 0.6 & 0.4 \\ 0.15 & 0.85 \end{bmatrix} = [x \quad y]$

Solve for $[x \quad y] = \left[\frac{3}{11} \quad \frac{8}{11}\right]$.

- Limiting matrix: the *m-step* of the state transition matrix that will give the same outcomes after we reach that step.

Example:

$$[0.1 \quad 0.9] * \begin{bmatrix} 0.6 & 0.4 \\ 0.15 & 0.85 \end{bmatrix}^2 = [0.23775 \quad 0.76225]$$

$$[0.1 \quad 0.9] * \begin{bmatrix} 0.6 & 0.4 \\ 0.15 & 0.85 \end{bmatrix}^{10} = [0.27272 \quad 0.72727]$$

$$[0.1 \quad 0.9] * \begin{bmatrix} 0.6 & 0.4 \\ 0.15 & 0.85 \end{bmatrix}^{25} = [0.27272 \quad 0.72727]$$

The probabilities appear to remain steady after numerous steps. $\begin{bmatrix} 0.6 & 0.4 \\ 0.15 & 0.85 \end{bmatrix}^{10}$ is the limiting matrix.

Reference

Notes, equations, and figures in the lecture are based on or taken from materials in the course textbook:

“Introduction to Probability, Statistics, and Random Processes”, by H. Pishro-Nik, ISBN 987-0-9906372-0-2 (Kappa Research, LLC.)