Markov Chains Tutorial

Adapted from my lecture notes for ST4231: Computer Intensive Statistical Methods

Department of Statistics and Applied Probability

National University of Singapore

LI Cheng

stalic@nus.edu.sg

Outline

Introduction

Discrete Markov Chains

Motivation

- In this tutorial, we review some fundamental results for discrete Markov chains.
- We only discuss the basic discrete-time Markov chains in this tutorial.
- The theory of Markov chains with uncountable state spaces is not discussed in this tutorial, although we will use such chains in the applications of MCMC.

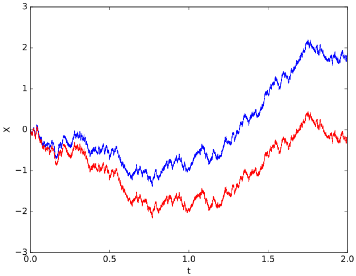
Outline

Introduction

Discrete Markov Chains

Stochastic processes

- Notation: $X = \{X_t\}_{t \geq 0}$.
- Sequence of random variables indexed by a (time) index $t \geq 0$.
- Discrete stochastic processes: t is discrete, such as $t = 0, 1, 2, \ldots$.
- Continuous stochastic processes: t is continuous, such as $t \in [0, +\infty)$.
- Example: temperature history in Singapore, stock prices, Brownian motion



Markov Property

- Roughly speaking, the Markov property is

$$P[(\text{future})|(\text{past})] \underbrace{=}_{\text{Markov}} P[(\text{future})|(\text{today})]$$

- A Markov chain X is a discrete time stochastic process $\{X_0, X_1, \cdots\}$ with the property that the distribution of X_t given all previous values of the process, X_0, X_1, $\cdots$, X_{t-1}, only depends upon X_{t-1}, i.e.,

$$P[X_t \in A | X_0, \cdots, X_{t-1}] = P[X_t \in A | X_{t-1}]$$

for any set A.

Transition

- Usually, the transition of a Markov chain determines its property.
- Let

$$p_{ij} = P(X_{t+1} = j | X_t = i)$$

  denote the transition probability from state i to state j at time $t+1$.
- Since the p_{ij} are conditional probabilities, all transition probabilities must satisfy two conditions:

$$\begin{aligned} p_{ij} \geq 0 \quad & \text{for all } (i,j), \\ \sum_j p_{ij} = 1 \quad & \text{for all } i. \end{aligned}$$

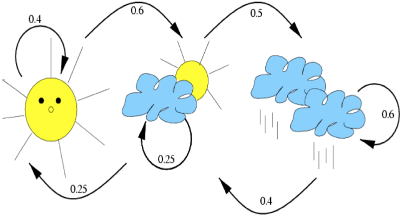
- We assume that $P(X_{t+1} = j | X_t = i)$ does not change with the time t. In such a case the Markov chain X is said to be homogeneous in time.

Weather in Singapore

- If X can only take a finite number (K) of possible values (called states), then the transition probabilities constitute a K by K transition matrix, denoted by P.

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1K} \\ p_{21} & p_{22} & \cdots & P_{2K} \\ \vdots & \vdots & & \vdots \\ p_{K1} & p_{K2} & \cdots & p_{KK} \end{pmatrix}.$$

- Exercise: Write down the transition matrix.



Multi-Step Transition

- The definition of one-step transition probability p_{ij} may be generalized to cases where the transition from one state to another takes place in some fixed number of steps.
- Let $p_{ij}(m) =$ be the m-step transition probability from state i to state j:

$$p_{ij}(m) = P(X_{t+m} = j | X_t = i), \quad m = 1, 2, \ldots$$

- We may view $p_{ij}(m)$ as the sum over all possible "paths" with length m that connects i to j.
- They can be calculated by the recursion formula

$$p_{ij}(m+1) = \sum_k p_{ik}(m) p_{kj}, \quad m = 1, 2, \ldots$$

Multi-Step Transition: Matrix Form

- We can define

$$P^{(m)} = \begin{pmatrix} p_{11}(m) & p_{12}(m) & \cdots & p_{1K}(m) \\ p_{21}(m) & p_{22}(m) & \cdots & P_{2K}(m) \\ \vdots & \vdots & & \vdots \\ p_{K1}(m) & p_{K2}(m) & \cdots & p_{KK}(m) \end{pmatrix},$$

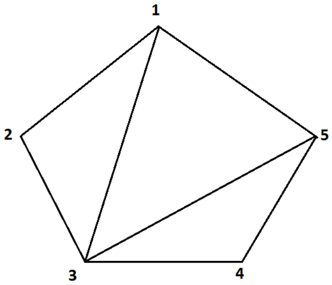
  so $P^{(1)} = P$.

- Then it can be shown (from the recursion formula) that

$$P^{(m)} = P^m = \underbrace{P \cdot P \cdots P}_{m}.$$

Example 1: Simple Random Walk on a Set (Graph)

- In a simple random walk on a graph, the state space is the set of vertices of a graph.
- At each time t, the Markov chain looks at its current position, X_t, and chooses one of its neighbors uniformly at random.
- It then jumps to the chosen neighbor, which becomes X_{t+1}.



Example 1: Simple Random Walk on a Set (Graph)

- More formally, the transition probabilities are given by

$$p_{ij} = \begin{cases} 1/\deg(i) & \text{if } i \text{ and } j \text{ are neighbors} \\ 0 & \text{otherwise (including } i = j) \end{cases}$$

  where $\deg(i)$ is the "degree of i", i.e. the number of neighbors of the vertex i.

- The transition matrix is

$$\begin{pmatrix} 0 & 1/3 & 1/3 & 0 & 1/3 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 1/4 & 1/4 & 0 & 1/4 & 1/4 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 1/3 & 0 & 1/3 & 1/3 & 0 \end{pmatrix}$$

How State Distribution Evolves with Markov Chain

- Suppose that there is an initial distribution (at time $t = 0$) over all possible states, denoted by the state distribution vector $\pi^{(0)}$.
- Note that $\pi^{(0)}$ is a **row vector**.
- In the previous example of random walk on a set (graph), such $\pi^{(0)}$ is a discrete distribution on $i = 1, 2, 3, 4, 5$, for example $\pi^{(0)} = (1/5, 1/5, 1/5, 1/5, 1/5)$.
- It can be shown that the state distribution at time t, denoted by $\pi^{(t)}$, can be found recursively by

$$\pi^{(t)} = \pi^{(t-1)} P.$$

- By iteration, we can see that

$$\pi^{(t)} = \pi^{(t-1)} P = \pi^{(t-2)} P^2 = \ldots = \pi^{(0)} P^t.$$

Stationary / Invariant Distribution

- Suppose that as time t approaches infinity, $p_{ij}^{(t)}$ tends to a number π_j independent of i, where π_j is the "steady-state probability" of state j.
- Correspondingly, for large t, the matrix P^t approaches the limiting form of a square matrix with identical rows:

$$\lim_{t\to\infty} P^t = \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_K \\ \pi_1 & \pi_2 & \cdots & \pi_K \\ \vdots & \vdots & & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_K \end{pmatrix} = \begin{pmatrix} \boldsymbol{\pi} \\ \boldsymbol{\pi} \\ \vdots \\ \boldsymbol{\pi} \end{pmatrix} = \mathbf{1}\boldsymbol{\pi},$$

where $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)$ and $\mathbf{1} = (\underbrace{1, 1, \ldots, 1}_{K})^\top$.

Stationary / Invariant Distribution

- If such a limit $\boldsymbol{\pi}$ exists, then if we start with an arbitrary initial distribution, represented by the state distribution vector $\pi^{(0)}$, we have (why?)

$$\lim_{t \to \infty} \pi^{(0)} P^t = \pi^{(0)} \mathbf{1} \boldsymbol{\pi} = \boldsymbol{\pi}.$$

- This implies that no matter which state distribution $\pi^{(0)}$ we start with, the Markov chain will eventually converge to the distribution $\boldsymbol{\pi}$. (This also implies that $\boldsymbol{\pi}$, as a limit of P^t, must represent a distribution.)
- The probability distribution $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)$ is called an **invariant** or **stationary distribution**. It is so called because it persists forever once it is established.

Example 1: Random Walk on a Set/Graph (continued)

- Exercise: Verify that the equilibrium (stationary) distribution is given by

$$\pi_i = \deg(i)/Z,$$

  where $Z = \sum_{k \in S} \deg(k)$ is the normalizing constant, and S is the state space ($|S| = K$).

- Exercise: Verify that for the example given before, the stationary distribution is $\boldsymbol{\pi} = (3/14, 2/14, 4/14, 2/14, 3/14)$.

Overview of Properties about the Stationary Distribution

- **Questions**:
  1. When does the stationary distribution $\boldsymbol{\pi}$ exist? Is it unique?
  2. If $\boldsymbol{\pi}$ exists and is unique, when does $\lim_{t\to\infty} P^t = \mathbf{1}\boldsymbol{\pi}$ hold true?
  3. How to find the stationary distribution $\boldsymbol{\pi}$ given the transition matrix P?

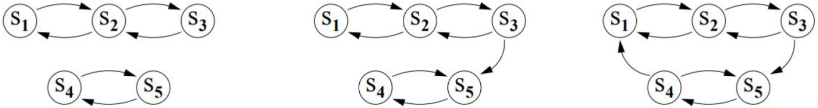
- **Short Answers**:
  1. The stationary distribution $\boldsymbol{\pi}$ exists and is unique, if the Markov chain is irreducible and positive recurrent.
  2. $\lim_{t\to\infty} P^t = \mathbf{1}\boldsymbol{\pi}$ holds true, if the Markov chain is irreducible, positive recurrent, and aperiodic.
  3. Solve the equation system $\boldsymbol{\pi}P = \boldsymbol{\pi}$, or use detailed balance condition.

Irreducible

- The state j of a Markov chain is said to be accessible from state i if there is a finite sequence of transitions from i to j with a positive probability.
- If the states i and j are accessible to each other, they are said to communicate with each other, and the communication is denoted by $i \leftrightarrow j$. Clearly, if $i \leftrightarrow j$ and $j \leftrightarrow k$, then $i \leftrightarrow k$.
- If two states of a Markov chain communicate with each other, they are said to belong to the same class.
- In general, the states of a Markov chain consist of one or more disjoint classes.
- If, however, all the states consist of a single class, then the Markov chain is said to be **irreducible**.
- Reducible chains are of little practical interest in most areas of application.

Irreducible

- A Markov chain X is called **irreducible** if for all pair of states i, j, there exists a $t > 0$ such that $p_{ij}(t) > 0$.
- In the picture below, only the third chain is irreducible.



Recurrent and Transient

- The state i is said to be a **recurrent state** if the Markov chain returns to state i with probability 1, i.e.

$$f_i = P(\text{ever returning to state } i) = 1.$$

- If the probability $f_i < 1$, state i is said to be a **transient state**.
- If a Markov chain starts in a recurrent state, the state reoccurs for an infinite number of times.
- If a Markov chain starts in a transient state, that state reoccurs only for a finite number of times.

Recurrent and Transient

- Let τ_{ii} be the time of the first return to state i:
  $\tau_{ii} = \min\{t > 0 : X_t = i | X_0 = i\}$.
- An irreducible Markov chain X is **recurrent** if $P[\tau_{ii} < \infty] = 1$ for some (and hence for all) states i. If this is not true, then X is called **transient**.
- Another equivalent condition for being a recurrent chain is

$$\sum_t p_{ii}(t) = \infty, \qquad \text{for all } i.$$

Example 2: A Seven-state Chain

- Consider the following transition matrix (bold numbers are the indexes of states):

	1	**2**	**3**	**4**	**5**	**6**	**7**
1	.3	0	0	0	.7	0	0
2	.1	.2	.3	.4	0	0	0
3	0	0	.5	.5	0	0	0
4	0	0	0	.5	0	.5	0
5	.6	0	0	0	.4	0	0
6	0	0	0	.1	0	.1	.8
7	0	0	0	1	0	0	0

- Which states are recurrent? Which states are transient? How many irreducible closed sets are there?

Positive Recurrent

- An irreducible recurrent Markov chain X is called **positive recurrent** if $E[\tau_{ii}] < \infty$ for all i. Otherwise, it is called **null recurrent**.

- All states in a communication class C are all together either positive recurrent, null recurrent, or transient. In particular, for an irreducible Markov chain, all states together must be positive recurrent, null recurrent or transient.

- If a Markov chain only has a finite number of states, and if it is irreducible, then it must be positive recurrent.

Positive Recurrent

- Another equivalent condition for **positive recurrence** is the existence of a stationary pmf $\boldsymbol{\pi}(\cdot)$ on the state space of X, such that

$$\sum_i \pi_i p_{ij}(t) = \pi_j, \text{ for all } j \text{ and } t \geq 0. \tag{1}$$

- If at time t, the distribution $\pi^{(t)} = \boldsymbol{\pi}$, i.e. it satisfies $\pi_i^{(t)} \equiv P(X_t = i) = \pi_i$, then for all state j,

$$\begin{aligned} P(X_{t+1} = j) &= \sum_i P(X_{t+1} = j | X_t = i) \cdot P(X_t = i) \\ &= \sum_i \pi_i p_{ij} = \pi_j. \end{aligned}$$

  In other words, $\pi^{(t+1)} = \boldsymbol{\pi}$.

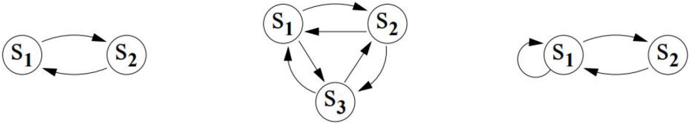
- Every irreducible Markov chain with a finite number of states has a unique stationary distribution.

Aperiodic

- An irreducible chain X is called **aperiodic** if for some (and hence for all) i,

  The greatest common divider of $\{t : p_{ii}(t) > 0\} = 1$.

- In the picture below, the 2nd and the 3rd chains are aperiodic.



Convergence Theorem

- If $X = \{X_1, X_2, \ldots\}$ is a positive recurrent and aperiodic Markov chain, then its stationary distribution $\boldsymbol{\pi}(\cdot)$ is the unique probability distribution satisfying (1). We then say that X is **ergodic**, and the following consequences hold:

  (i) $p_{ij}(t) \to \pi_j$ as $t \to \infty$ for all i, j.

  (ii) (Ergodic Theorem) For a function $h(x)$, if $E_{\boldsymbol{\pi}}[|h(X)|] < \infty$, then

  $$N^{-1} \sum_{k=1}^{N} h(X_k) \to E_{\boldsymbol{\pi}}[h(X)], \quad \text{as } N \to \infty, \text{ with probability 1,}$$

  where $E_{\boldsymbol{\pi}}[h(X)] = \sum_i h(i)\pi_i$, the expectation of $h(x)$ with respect to $\boldsymbol{\pi}(\cdot)$.

Finding the Stationary Probabilities

- In general, finding the stationary distribution $\boldsymbol{\pi}$ requires solving the system of equations (1).
- However, sometimes, we can avoid solving the set of equations (1).
- Suppose that we can find positive numbers x_j, $j = 1, 2, \cdots, K$ (for a finite state space), such that

$$x_i p_{ij} = x_j p_{ji}, \quad \textit{for } i \neq j, \quad \sum\nolimits_{j=1}^{K} x_j = 1.$$

- Then summing the preceding equations over all states i yields

$$\sum\nolimits_{i=1}^{K} x_i p_{ij} = x_j \sum\nolimits_{i=1}^{K} p_{ji} = x_j,$$

  which implies that the stationary distribution $\boldsymbol{\pi}$ satisfies $\pi_j \propto x_j$, $j = 1, 2, \cdots, K$, because $\{\pi_j, j = 1, 2, \cdots, K\}$ are the unique solution of (1) according to the previous theorem.

Example 1: Random Walk on a Set/Graph (continued)

- We claim that the stationary distribution has the probabilities $\pi_i = \deg(i)/Z$, where $Z = \sum_{k \in S} \deg(k)$.
- Exercise: Verify that: If i and j are not neighbors, then $\pi_i p_{ij} = 0 = \pi_j p_{ji}$; If i and j are neighbors, then $\pi_i p_{ij} = 1/Z = \pi_j p_{ji}$.

Example 3

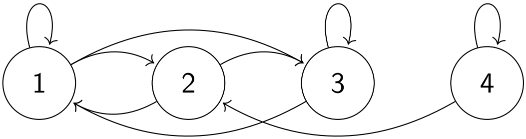
- Consider the Markov chain with state space $\mathcal{S} = \{1, 2, 3, 4\}$ and the transition matrix

$$P = \begin{pmatrix} 1/3 & 1/3 & 1/3 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \end{pmatrix}$$

- Q1. Find the stationary distribution $\boldsymbol{\pi}$.
- Q2. What is the value of $\lim_{t\to\infty} p_{23}(t)$?
- Q3. What is the value of $\lim_{N\to\infty} N^{-1} \sum_{k=1}^{N} X_k$?

Example 3

- Q1. First draw the graph.



We can see that "4" is a transient state, while "1", "2", and "3" form an irreducible closed set. This means that in the stationary distribution $\boldsymbol{\pi}$, $\pi_4 = 0$.

Example 3

- Q1 (cont'd). Solve the equation $\boldsymbol{\pi} P = \boldsymbol{\pi}$.

$$\begin{pmatrix} \pi_1 & \pi_2 & \pi_3 \end{pmatrix} \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix} = \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 \end{pmatrix}$$

$$\Longrightarrow \begin{cases} \frac{1}{3}\pi_1 + \frac{1}{2}\pi_2 + \frac{1}{2}\pi_3 = \pi_1 \\ \frac{1}{3}\pi_1 + \frac{1}{2}\pi_2 = \pi_2 \\ \frac{1}{3}\pi_1 + \frac{1}{2}\pi_3 = \pi_3 \end{cases}$$

$$\Longrightarrow \pi_1 = \frac{3}{7}, \pi_2 = \frac{2}{7}, \pi_3 = \frac{2}{7}.$$

So the stationary distribution $\boldsymbol{\pi} = (3/7, 2/7, 2/7, 0)$.

Example 3

- Q2 The set of "1", "2", and "3" has finitely many states and irreducible, hence it is positive recurrent. The Markov chain on $\{1, 2, 3\}$ is also aperiodic. This implies that the Markov chain is ergodic, and $p_{ij}(t) \to \pi_j$ as $t \to \infty$ for all i, j. Therefore,

$$\lim_{t\to\infty} p_{23}(t) = \pi_2 = \frac{2}{7}$$

- Q2 Using the ergodic property, we have

$$\begin{aligned}\lim_{N\to\infty} N^{-1} \sum_{k=1}^{N} X_k &= E_{\boldsymbol{\pi}}[X] \\ &= 1 \cdot \frac{3}{7} + 2 \cdot \frac{2}{7} + 3 \cdot \frac{2}{7} \\ &= \frac{13}{7}.\end{aligned}$$