Markov Chain

This is an introductory but, in my opinion, very dense tutorial for Markov Chain

1 Markov Process

Markov process is closely related to the state space modeling method which has many applications in engineering. The idea of state space modeling is that the system at any time t is modeled by a state which captures all the information about the system up to the time t that is related to the future of the system. If the state captures all the information that is related to the future of the system, no other information about the past is needed to determine the future of the system. This leads us to the notion of $Conditional\ Independence$.

Conditional Independence: Let $X, Y\ and\ Z$ be three random variables. We say that X and Z are conditional independent given Y if

$$Pr(X = i, Z = j \mid Y = k) = Pr(X = i \mid Y = k) \cdot Pr(Z = j \mid Y = k) \tag{1}$$

for discrete $X, Yand\ Z$ and all i, j, k such that $Pr(Y = k) > 0$. We denote this relationship between $X, Y\ and\ Z$ as $X—Y—Z$. Equivalently, if $X—Y—Z$, we have

$$Pr(X = i \mid Y = k, Z = j) = Pr(X = i \mid Y = k) \tag{2}$$

Given Y, X and Z are independent.

Markov Process: A random process X_t, $t \in \mathbb{T}$ is said to be a Markov process if for any $t_1, ..., t_{n+1} \in \mathbb{T}$, $t_1 < t_2 < ... < t_n$ and for any n

$$(X_{t_0}, X_{t_1}, ..., X_{t_n}) — X_{t_n} — (X_{t_{n+1}}, ..., X_{t_{n+m}})$$

Equivalently

$$(X_{t_0}, X_{t_1}, ..., X_{t_{n-1}}) — X_{t_n} — X_{t_{n+1}}$$

More precisely, we have

$$Pr(X_{t_n} = i \mid X_{t_{n-1}} = i_{n-1}, X_{t_{n-2}} = i_{n-2}, ..., X_{t_0} = i_0) = Pr(X_{t_n} = i \mid X_{t_{n-1}} = i_{n-1}) \tag{3}$$

This means that the past and the future of the system are conditionally independent given the present state.

2 Discrete Time Markov Chain

In this section, we introduce the Discrete Time Markov Chain(DTMC). Let $\mathbf{S} = \{1, 2, ...\}$ be a finite or countably infinite set (thought as the set of "states" of a chain). Let $\mathbf{T} = \{1, 2, ...\}$ represent time. $(X_n, n \in \mathbf{T})$ a collection of random variables each supported on $\mathbf{S}$ is called Markov Chain if $\exists P_{ij}, i, j \in \mathbf{S}$ and $\sum_{j=1}^{|\mathbf{S}|} P_{ij} = 1$ such that for all n

$$P_{ij} = P(X_n = j \mid X_{n-1} = i) = P(X_n = j \mid X_{n-1} = i, X_{n-2} = i_{n-2}, ..., X_0 = i_0) \tag{4}$$

The expression $P_{ij} = P(X_n = j \mid X_{n-1} = i)$ is called *Transition Probabilities*. We often write the transition probabilities in a matrix called *Transition Probability Matrix*. The transition probability matrix is often denoted as $\mathbf{P} = (P_{ij})_{i,j=1}^{|\mathbf{S}|}$. Note that $P_{ij} \geq 0$ for all i, j and

$$\sum_{i=1}^{|\mathbf{S}|} P_{ki} = 1 \tag{5}$$

Next, we will show how to use the transition probability matrix. Consider a DTMC $\{X_n, n = 0, 1, 2, ...\}$ where $X_n \in \mathbf{S}$. Suppose that we know the probability distribution at time 0 that is X_0. If we denote such distribution as a row vector $\pi^{(0)} = [P(X_0 = 1), P(X_0 = 2), ...]$. Now we aim to get the probability distribution of $X_1, X_2, ...$ Take $X_1 = j$ for example

$$P(X_1 = j) = \sum_{k=1}^{|\mathbf{S}|} P(X_0 = k) P(X_1 = j \mid X_0 = k) = \sum_{k=1}^{|\mathbf{S}|} P(X_0 = k) P_{ki} \tag{6}$$

Then in a matrix notation, we have $\pi^{(1)} = \pi^{(0)}\mathbf{P}$. More generally, we have

$$\pi^{(n)} = \pi^{(0)}\mathbf{P}^n \ for\ n \in \{1, 2, 3, ...\} \tag{7}$$

If we denote the probability of transitioning from state i to state j with n steps as $P_{ij}^{(n)}$

$$P_{ij}^{(n)} = P(X_{n+m} = j \mid X_m = i) \tag{8}$$

Chapman – Kolmogorov Equation: The Chapman-Kolmogorov equation can be written as

$$P_{ij}^{(m+n)} = P(X_{n+m} = j \mid X_0 = i) = \sum_{k=1}^{|\mathbf{S}|} P_{ik}^{(m)} P_{kj}^{(n)} \tag{9}$$

Proof. $P_{ij}^{(m+n)} = \sum_{k=1}^{|\mathbf{s}|} P(X_{m+n} = j \mid X_n = k, X_0 = i) P(X_n = k \mid X_0 = i) = \sum_{k=1}^{|\mathbf{S}|} P_{ik}^{(m)} P_{kj}^{(n)}$ □

Using the Chapman-Kolmogorov equation, we can prove that the n-step transition probability matrix $\mathbf{P}^{(n)}$ is simply nth power of one-step transition probability matrix $\mathbf{P}^n$. To conclude, if $\mathbf{a}$ is the initial distribution, then $P(X_n = j) = (P^n \cdot \mathbf{a})_j$

2.1 Classification of States

For a better understanding of the characterization of DTMC, we first introduce some definitions.

Definition 2.1. *State j is* **accessible** *from i, which can be written as $i \to j$, if $\exists n \geq 1$ such that $P_{ij}^{(n)} > 0$ for some n. We assume that all the states are accessible from itself ($P_{ii}^{(0)} = 1$).*

Definition 2.2. *Two state i and j are* **communicate***, which can be written as $i \leftrightarrow j$, if they are accessible form each other. Again, all states are communicate with itself.*

We can partition the states of a Markov chain into communicating classes such that only the members within the same class can communicate with each other. These classes are disjoint.

Definition 2.3. *$A \subset S$ is called* **closed** *if $Pr(leaving\ A) = 0$.*

Definition 2.4. *A Markov chain is called* **irreducible** *if the only closed class is* **S** *itself ($\Leftrightarrow$ all the states communicate with each other).*

The **irreducible** is equivalent to $\exists\ n > 0\ s.t.\ P_{ij}^{(n)} > 0$ for all i, j.

Let $f_{ii} = P(\bigcup_{n=1}^{+\infty} X_n = i \mid X_0 = i)$ which is the probability of returning to i at some point and N_i be the total number of visit to state i. We have$P(N_i = n) = f_{ii}^{(n-1)}(1 - f_{ii})$. We can see that $N_i \mid X_0 \sim Geometric(1 - f_{ii})$. Thus $E\{N_i\} = \frac{1}{1-f_{ii}}$. $E\{N_i\} = +\infty$ if and only if $f_{ii} = 1$.

Definition 2.5. *A state is called* **recurrent** *if $f_{ii} = 1$ and* **transient** *if $f_{ii} < 1$.*

Let $m_i = \sum_{n=1}^{\infty} nP_{i,i}^{(n)}$ be the expected time of first return to state i. Here $P_{i,i}^{(n)} = P(X_n = i, X_{n-1} \neq i, ..., X_0 \neq i)$

Definition 2.6. *A state is called* **positive recurrent** *if $m_i < +\infty$ and* **null recurrent** *if $m_i = +\infty$*

Definition 2.7. *A* **period** *of a state i is the largest integer d satisfying $P_{ii}^{(n)} = 0$ whenever n is not dividable by d. Therefore, the period of state i is d denoted as $d(i)$.*

If $d(i) > 1$, then we say that the state i is periodic. If $d(i) = 1$, then we say that state i is aperiodic. All states in the same communicating class have the same period. A class is said to be periodic if its states are periodic. Similarly, a class is said to be aperiodic if its states are aperiodic. Finally, a Markov chain is said to be aperiodic if all of its states are aperiodic.

2.2 Limiting Probabilities and Stationary Distribution

Here, we discuss the long-term behavior of the Markov Chain that is the limiting probability $\lim_{n\to\infty} P_{ij}^{(n)}$. The probability distribution $\pi = [\pi_0, \pi_1, ...]$ is called the stationary distribution of Markov chain if

$$\pi_i = \lim_{n\to+\infty} P(X_n = i \mid X_0 = j) \tag{10}$$

for $i, j \in \mathbf{S}$ and we have

$$\sum_i \pi_i = 1 \tag{11}$$

Suppose the Markov chain does not have periodic states, null recurrent states or transient states, then

- π_i exists and is independent of i
- $\pi_i = \frac{1}{m_i} > 0$ for all i where $m_i = \sum_{n=1}^{+\infty} nP_{i,i}^{(n)}$ which is the expected number of transitions need to return to i.

Theorem 2.8. *For the limiting distribution π, we have $\pi = \pi\mathbf{P}$ where* **P** *is the transition probability matrix*

Proof.

$$\begin{aligned}\pi = \lim_{n\to+\infty} \pi^{(n+1)} &= \lim_{n\to+\infty} \left[\pi^{(0)\mathbf{P}^{(n+1)}}\right] \\ &= \lim_{n\to+\infty} \left[\pi^{(0)\mathbf{P}^{(n)}}\right] \mathbf{P} \\ &= \pi\mathbf{P}\end{aligned} \tag{12}$$

□

2.2.1 Markov Chain with Finite State Space

Assume that the chain is irreducible and aperiodic. Then the stationary distribution is unique. The stationary distribution is

$$\pi_i = \lim_{n\to+\infty} P(X_n = i) \tag{13}$$

We also define that $r_i = \frac{1}{\pi_i}$ which is the expected return time of state i.

2.2.2 Markov Chain with Countably Infinite State Space

When the state space is countably infinite, we need to consider which type of the states are. Assume that the chain is irreducible and aperiodic, we have

- If all the states are positive recurrent, the stationary distribution is unique
- If all the states are null recurrent or transient, the stationary distribution does not exit.

$$\lim_{n\to+\infty} P(X_n = i \mid X_0 = j) = 0 \tag{14}$$

We still define that $r_i = \frac{1}{\pi_i}$ which is the expected return time of state i.

3 Reference

[1] `http://www.ifp.illinois.edu/~hajek/Papers/randomprocJuly14.pdf`

[2] `https://www.probabilitycourse.com/`