Markov Chain

Azmain Biswas

April 2023

Contents

Chapter 1

Introduction

Chapter 2

Markov Chain

2.1 Definition

Definition 2.1.1 (Markov Chain)**.** A discrete time stochastic process $\{X_n, n = 1, 2, 3, \ldots\}$ is defined to be *Discrete Time Markov Chain* or simply *Markov Chain* if it takes value the state space $\mathbf{S}$, and for every $n \geq 0$ it satisfy the property

$$\mathbf{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \mathbf{P}(X_{n+1} = j | X_n = i) \tag{2.1}$$

Unless otherwise mentioned we take the state space $\mathbf{S}$ to be $\{0, 1, 2, 3, \ldots\}$. If $X_n = i$ we say that the process is in ith state at time n. In the definition eq. (2.1) may be interpreted as for Markov Chain, the conditional distribution of any future state X_{n+1}, given the past states $X_0, X_1, \ldots, X_{n-1}$ and the present state X_n, is independent of the past and only depend on the present state. This property is called *Markovian Property.* In other word for markov chain predicting the future we only need information about the present state.

Note: Notice that Assumption we are making for markov chain to forget the past as long as present in known is very strong assumption.

2.2 Homogeneous Markov Chain

Definition 2.2.1 (Homogeneous Markov Chain)**.** We say a markov chain $\{X_n, n \geq 0\}$ is homogeneous if $\mathbf{P}(X_{n+1} = j | X_n = i) = \mathbf{P}(X_2 = j | X_1 = i)$ $\forall n > 0$.

The quantity $\mathbf{P}(X_{n+1} = j | X_n = i)$ is called the *transition probability* from state i to state j. For homogeneous Markov Chain we can specify the transition probabilities $\mathbf{P}(X_{n+1} = j | X_n = i)$ by a sequence of value $p_{xy} = \mathbf{P}(X_{n+1} = y | X_n = x)$.

For the case of finite state Markov chain, say the state space is $\{1, 2, 3, \ldots\}$. Then the transition probabilities are p_{ij}, $1 \leq i, j \leq N$ for transition from state i to state j. The $N \times$

N matrix $P = (p_{ij})_{N\times N}$ is called The *Transition Matrix* of chain. Since probabilities are nonnegative and since the process must make a transition into some state, we have

$$P_{ij} \geq 0, \quad i,j \geq 0, \text{ and } \sum_{j=0}^{N} P_{ij} = 1, \quad \forall\, i = 0,1,2,\ldots.$$

For, infinite state markov chain the probability transition matrix will be infinite order. Then,

$$P = \begin{bmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{11} & P_{12} & p_{12} & \cdots \\ \vdots & \vdots & \vdots & \ldots \\ p_{i0} & \cdots & p_{ij} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Example 2.2.1 (Rain and sunny)**.** Suppose that whether it rains today depends on previous weather conditions only from the last two days. Specifically, suppose that if it has rained for the past two days, then it will rain tomorrow with probability 0.7; if it rained today but not yesterday, then it will rain tomorrow with probability 0.5; if it rained yesterday but not today, then it will rain tomorrow with probability 0.4; if it has not rained in the past two days, then it will rain tomorrow with probability 0.2.

We can transform it into a Markov chain by letting the state on any day be determined by the weather conditions during both that day and the preceding one. For instance, we can say that the process is in

State 0: if it rained both today and yesterday
State 1: if it rained today but not yesterday
State 2: if it rained yesterday but not today
State 3: if it rained neither today nor yesterday

The it will be 4-state Markov chain whose transition matrix will be,

$$P = \begin{bmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{bmatrix}$$

2.3 Chapman-Kolmogorov Equations

We have already defined the one step transition probabilities p_{ij}. We now define the n-step transition probabilities p_{ij}^n to be the probability that a process in state i will be in state j after

n additional transitions. i.e.

$$p_{ij}^n = \mathbf{P}(X_{n+m} = j | X_m = i), \; n \geq 0, \; i,j \geq 0.$$

By definition of markov chain we get,

$$\begin{aligned}
p_{ij}^{n+m} &= \mathbf{P}(X_{n+m} = j | X_0 = i) \\
&= \sum_{k=0}^{\infty} \mathbf{P}(X_{n+m} = j, X_n = k | X_0 = i) \text{ (By theorem of total probability)} \\
&= \sum_{k=0}^{\infty} \mathbf{P}(X_{n+m} = j | X_n = k, X_0 = i) \mathbf{P}(X_n = k | X_0 = i) \\
&= \sum_{k=0}^{\infty} \mathbf{P}(X_{n+m} = j | X_n = k) \mathbf{P}(X_n = k | X_0 = i) \\
&= \sum_{k=0}^{\infty} p_{kj}^m p_{ik}^n. \qquad (2.2)
\end{aligned}$$

If we take $n = m = 1$. Then

$$p_{ij}^2 = \sum_{k=0}^{\infty} p_{kj} p_{ik} \qquad (2.3)$$

the above expression is (i, j) element of P^2 matrix then we see eq. (2.3) in matrix form,

$$P^2 = P * P$$

Hence, eq. (2.2) can also written in matrix form,

$$P^{n+m} = P^n * P^m$$

Where P^n & P^m are the n-step and m-step transition matrix respectively.

Example 2.3.1 (Transition matrix of 4-state Markov chain)**.** Consider the 4-state Markov chain depicted in fig. 2.1 When no probabilities are written over the arrows, as in this case, it means all arrows originating from a given state are equally likely. For example, there are 3 arrows originating from state 1, so the transitions $1 \to 3$, $1 \to 2$, and $1 \to 1$ all have probability 1/3. Therefore the transition matrix of the chain is.

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

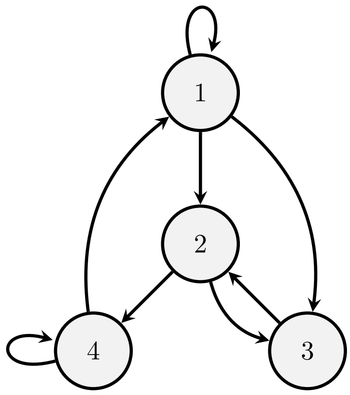

Figure 2.1: Example of A 4 state markov chain

To compute the probability that the chain is in state 3 after 5 steps, starting at state 1, we would look at the (1,3) entry of P^5.

$$P^5 = \begin{bmatrix} \frac{853}{3888} & \frac{509}{1944} & \frac{52}{243} & \frac{395}{1296} \\ \frac{173}{864} & \frac{85}{432} & \frac{31}{108} & \frac{91}{288} \\ \frac{37}{144} & \frac{29}{72} & \frac{1}{9} & \frac{11}{48} \\ \frac{499}{2592} & \frac{395}{1296} & \frac{71}{324} & \frac{245}{864} \end{bmatrix}$$

so, $p_{13}^5 = \mathbf{P}(X_5 = 3 | X_0 = 1) = \frac{52}{243}$.

To get the marginal distributions of $X_0, X_1, ...$, we need to specify not just the transition matrix, but also the initial conditions of the chain. This can be done by setting the initial state X_0 to be a particular state, or by randomly choosing X_0 according to some distribution. Let $(t_1, t_2, ..., t_N)$ be the PMF of X_0 displayed as a vector, that is, $t_i = \mathbf{P}(X_0 = i)$. Then the marginal distribution of the chain at any time can be computed from the transition matrix.

Proposition 2.3.1 (Marginal Distribution of X_n)**.** *Define* $\mathbf{t} = (t_1, t_2, \ldots, t_n)$ *by* $t_i = \mathbf{P}(X_0 = i)$, *and view* $\mathbf{t}$ *as a row vector. Then the marginal distribution of* X_n *is given by the vector* $\mathbf{t}P^n$*. That is the* j*-th component of* $\mathbf{t}P^n$ *is* $\mathbf{P}(X_n = j)$*.*

Proof. By the law of total probability we get,

$$
\begin{aligned}
\mathbf{P}(X_n = j) &= \sum_{i=0}^{N} \mathbf{P}(X_n, X_0 = i) \\
&= \sum_{i=0}^{N} \mathbf{P}(X_0 = i)\mathbf{P}(X_n = j | X_0 = i) \\
&= \sum_{i=0}^{N} t_i p_{ij}^n
\end{aligned}
$$

Which is the $j0$th component of $\mathbf{t}P^n$. $\square$

Example 2.3.2 (Marginal distribution of 4-state Markov chain)**.** Again consider the 4-state Markov chain in fig. 2.1. Suppose the initial conditions are $\mathbf{t} = (1/4, 1/4, 1/4, 1/4)$, Let X_n be the position of the chain at time n. Then distribution of X_1 is

$$
\begin{aligned}
\mathbf{t}P &= \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{bmatrix} \\
&= \begin{bmatrix} \frac{5}{24} & \frac{1}{3} & \frac{5}{24} & \frac{1}{4} \end{bmatrix}
\end{aligned}
$$

The marginal distribution of X_5 is

$$
\begin{aligned}
\mathbf{t}P^5 &= \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} \frac{853}{3888} & \frac{509}{1944} & \frac{52}{243} & \frac{395}{1296} \\ \frac{173}{864} & \frac{85}{432} & \frac{31}{108} & \frac{91}{288} \\ \frac{37}{144} & \frac{29}{72} & \frac{1}{9} & \frac{11}{48} \\ \frac{499}{2592} & \frac{395}{1296} & \frac{71}{324} & \frac{245}{864} \end{bmatrix} \\
&= \begin{bmatrix} \frac{707}{3254} & \frac{472}{1619} & \frac{101}{486} & \frac{1469}{5184} \end{bmatrix}
\end{aligned}
$$

2.4 Classification of states

Definition 2.4.1. We say that state j is *accessible* from state i, written as $i \to j$, If $p_{ij}^n > 0$ for some $n \in \mathbb{N}$.
We assume every state is accessible from itself since,

$$p_{ii}^0 = \mathbf{P}(X_0 = i | X_0 = i) = 1.$$

Definition 2.4.2. Two states i and j are said to *communicate* , written as $i \longleftrightarrow j$, if they are accessible from each other.
i.e.

$$i \longleftrightarrow j \implies i \to j \quad \& \; j \to i$$

Communication is an equivalence relation. That means that,

1. $i \longleftrightarrow i$.
2. if $i \longleftrightarrow j$ then $j \longleftrightarrow i$.
3. if $i \longleftrightarrow j$ and $j \longleftrightarrow k$ then $i \longleftrightarrow k$.

First two property is obvious for last one, let for some n, m in $\mathbb{N}$ then, $p_{ij}^n, p_{jk}^m > 0$ by assumption. By Chapman-Kolmogorov equatation.

$$p_{ik}^{n+m} = \sum_{r=0}^{\infty} p_{ir}^n p_{rk}^m \geq p_{ij}^n p_{jk}^m > 0.$$

Hence state k is accessible from state i. By same we can see tha converse.

Therefore, the states of a Markov chain can be partitioned into communicating classes such that only members of the same class communicate with each other. i.e. two states i & j belong to same class if and only if $i \longleftrightarrow j$.

Example 2.4.1. Consider the markov chain define in the picture fig. 2.2.

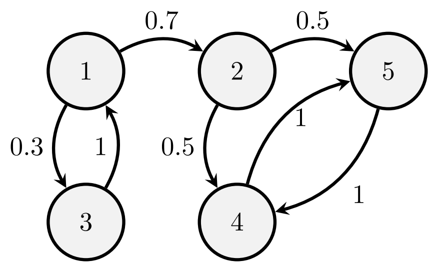

Figure 2.2:

Here the classes are $\{1, 3\}, \{2\}, \{4, 5\}$

Definition 2.4.3 (Irreducible Markov chain)**.** A Markov chain is said to be irreducible if it has only one communicating class. That is, every state communicate with each other.
That is, for any states i, j there is some positive integer n such that the (i, j) entry of P^n is positive.

A Markov chain that is not irreducible called reducible.

For any state i and j define f_{ij}^n to be the probability that, starting from i, the first transition into j occurs at n time.
i.e.

$$f_{ij}^n = \mathbf{P}(X_n, X_k \neq j, k = 1, 2, \ldots n-1 | X_0 = i).$$

Let,

$$f_{ij} = \sum_{n=0}^{\infty} f_{ij}^n$$

Then, f_{ii} denote the probability of ever making a transition into step j when start from state i. If j is not accessible from i f_{ij} will be zero.

Definition 2.4.4 (Recurrent and Transient state)**.** A state j of a Markov chain is said to be *recurrent* $f_{ii} = 1$ and *transient* if $f_{ii} < 0$.

In other word, if a markov chain start in a recurrent state, there is a guarantee that it will visit that state again in the future (eventually return to that state with probability 1).

In contrast, a transient state in a Markov chain is a state where, once you reach it, there is a positive probability that you will never return to that state. i.e. if you begin in a transient state, there's a chance you won't return there.

Corollary 2.4.0.1. *state i is recurrent if and only if* $\sum_{n=0}^{\infty} p_{ii}^n = \infty$

Proof. The state i is recurrent if and only if, starting in state i, the expected number of time periods that the process is in state i is infinite.
Let,

$$I_n = \begin{cases} 1, & \text{if } X_n = i \\ 0, & \text{if } X_n \neq i \end{cases}$$

we then have $\sum_{n=0}^{\infty} I_n$ represent the number of periods that the process is in state i Then,

$$\begin{aligned} E\left[\sum_{n=0}^{\infty} I_n | X_0 = i\right] &= \sum_{n=0}^{\infty} E[I_n | X_0 = i] \\ &= \sum_{n=0}^{\infty} \mathbf{P}(X_n = i | X_0 = i) \\ &= \sum_{n=0}^{\infty} p_{ii}^n \end{aligned}$$

Hence the result. □

Theorem 2.4.1. *If i is recurrent and $i \longrightarrow j$, then j is also recurrent.*

Proof. Let m and k be such that $p_{ij}^{k} > 0, p_{ji}^{m} > 0$. Now for any $n \geq 0$

$$p_{jj}^{m+n+k} \geq p_{ji}^{m} p_{ii}^{k} p_{ij}^{m}$$

This follows because the left-hand side of the preceding equation is the probability of going from $j \rightarrow j$ in $m + n + k$ steps, whereas the right-hand side is the probability of going from $j \rightarrow j$ in $m + n + k$ steps via a path that goes from $j \rightarrow i$ in m steps, then from $i \rightarrow i$ in n additional steps, then from $i \rightarrow$j in k additional steps. By summing the preceding over n. we obtain

$$\sum_{n=0}^{\infty} p_{jj}^{m+n+k} \geq p_{ji}^{m} p_{ii}^{k} \sum_{n=0}^{\infty} p_{ij}^{n} = \infty$$

where $\sum_{n=0}^{\infty} p_{ij}^{n} = \infty$ because state i is recurrent. Thus, j is also recurrent. □

Proposition 2.4.2. *In an irreducible Markov chain with a finite state space, all states are recurrent.*

Chapter 3

Stationary Probability

Consider a two-state Markov chain with transition probability matrix

$$P = \begin{bmatrix} 0.35 & 0.65 \\ 0.89 & 0.11 \end{bmatrix}$$

Then 2-step transition matrix will be

$$P^2 = \begin{bmatrix} 0.7010 & 0.2990 \\ 0.4094 & 0.5906 \end{bmatrix}$$

Then 8-step transition matrix will be

$$P^8 = \begin{bmatrix} 0.5810 & 0.4190 \\ 0.5737 & 0.4263 \end{bmatrix}$$

Once again 12-state transition matrix is,

$$P^{12} = \begin{bmatrix} 0.5782 & 0.4218 \\ 0.5776 & 0.4224 \end{bmatrix}$$

Note that The 8-step transition matrix is almost identical to 12-step transition matrix. So it seems that p_{ij}^n is converging to some value as $n \to \infty$ that is same for all i. In other words, In other word it seems to exist a limiting probability that the process will be in state j after a large number of transitions, and this value is independent of initial state. This limiting probability is known as stationary probability and the distribution of X_j^n as $n \to \infty$ is known as stationary distribution.

State i is said to have *period* d if $p_{ii}^n = 0$ whenever n is not divisible by d, and d is the largest integer with this property. A state with period 1 is said to be *aperiodic*. We denote period of i by $d(i)$.

If state i is recurrent, then it is said to be positive recurrent if, starting in state the expected time until the process returns to state i is finite. While there exist recurrent states that an not positive recurrent (such states are called null recurrent). Postie recurrent, aperiodic states are called *ergodic*.

Definition 3.0.1 (Stationary Distribution)**.** A probability distribution $\{p_j, j \geq 0\}$ is called stationary for the Markov chain if

$$p_j = \sum_{i=0}^{\infty} p_i p_{ij},\ j \geq 0 \tag{3.1}$$

i.e. If $\mathbf{t} = (p_1, p_2, \ldots, p_j, \ldots)$ is a probability distribution vector and P is transition matrix, Then

$$\mathbf{t} = \mathbf{t}P. \tag{3.2}$$

From eq. (3.2) we see that 1 is a eigenvalue of transition matrix P and $\mathbf{t}$ is eigenvector corresponding to 1. Since in transition matrix such that $\sum_{j=0}^{\infty} p_{ij} = 1\ \forall i$ i.e. sum of all elements of row is 1, 1 must be a eigenvalue.

Now the question aeries is eigenvector corresponding to 1 is a probability vector(sum of all component in 1) or not.

Theorem 3.0.1 (Existence and Uniqueness of Stationary Distribution)**.** *An irreducible aperiodic Markov chain belongs to one of the following two classes:*

1. *Either the state are all transient or all null recurrent; in this case $P_{ij}^n \to 0$ as $n \to \infty\ \forall i, j$ and there exist no stationary distribution.*
2. *Or else, all states are positive recurrent, i.e.*

$$\pi_j = \lim_{n\to\infty} p_{ij}^n > 0.$$

In this case, $\{\pi_j, j = 0, 1, \ldots\}$ is stationary distribution and there exist no other stationary distribution.

Proof. we will first proof (2), we have,

$$\sum_{j=0}^{M} p_{ij}^n \leq \sum_{j=0}^{\infty} p_{ij}^n = 1\ \forall M.$$

Letting $n \to \infty$ we get,

$$\sum_{j=0}^{M} \pi_j \leq 1 \; \forall M.$$
$$\sum_{j=0}^{\infty} \leq 1.$$

Now,

$$p_{ij}^{n+1} = \sum_{k=0}^{\infty} p_{ik}^{n} p_{kj} \geq \sum_{k=0}^{M} p_{ik}^{n} p_{kj} \; \forall M.$$

Letting $n \to \infty$ gives us,

$$\pi_j \geq \sum_{k=0}^{M} \pi_k p_{kj} \; \forall M,$$

i.e.

$$\pi_j \geq \sum_{k=0}^{\infty} \pi_k p_{kj}, \;\; \forall j \geq 0.$$

To show that the above is actually an equality, suppose the inequality is strict for some j. Then we obtain,

$$\sum_{j=0}^{\infty} \pi_j > \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \pi_k p_{jk} = \sum_{k=0}^{\infty} \pi_k \sum_{j=0}^{\infty} p_{kj} = \sum_{k=0}^{\infty} \pi_k.$$

which is contradiction. Therefore,

$$\pi_j = \sum_{k=0}^{\infty} \pi_k p_{kj}, \quad \forall j, \ldots$$

Potting $p_j = \pi_j / \sum_0^\infty \pi_k$, we see that $\{p_j, j = 0, 1, 2, \ldots\}$ is stationary distribution, and hence at least one stationary distribution exist. Now let $\{p_j, j = 0, 1, 2, \ldots\}$ be any stationary distribution. Then if $\{p_j, j = 0, 1, 2, \ldots\}$ is the probabolity distribution of X_0 then

$$\begin{aligned} p_j &= \mathbf{P}(X_n = j) \\ &= \sum_{i=0}^{\infty} \mathbf{P}(X_n = j | X_0 = i) \mathbf{P}(X_0 = i) \\ &= \sum_{i=0}^{\infty} p_{ij}^{n} p_i \end{aligned} \tag{3.3}$$

Then, from eq. (3.3) we see,

$$p_j \geq \sum_{i=0}^{M} p_{ij}^{n} p_i \;\; \forall M.$$

Letting n and then M approach ∞ we get,

$$p_j \geq \sum_{i=0}^{\infty} \pi_j p_i = \pi_j.$$

Now we show that $p_j \leq \pi_j$ use eq. (3.3) and that fact that $p_{ij}^n \leq 1$ to obtain

$$p_j \leq \sum_{i=0}^{M} p_{ij}^n p_i + \sum_{i=M+1}^{\infty} p_i \ \forall M.$$

and letting $n \to \infty$ gives,

$$p_j \leq \sum_{i=0}^{M} \pi_j p_i + \sum_{i=M+1}^{\infty} p_i \ \forall M.$$

Since $\sum_0^\infty = 1$, we obtain upon letting $M \to \infty$ that

$$p_j \leq \sum_{i=0}^{\infty} \pi_j p_i = \pi_j. \tag{3.4}$$

If the state are transient or null recurrent and $\{p_j, j = 0, 1, 2, \ldots\}$ is a stationary distribution, then eq. (3.3) hold and $p_{ij}^n \to 0$, which is clearly impossible. Thus, for case (1), no stationary distribution exists and the proof is complete. $\square$