Chapter 3

Markov Chains

Markov chains are the simplest examples among stochastic processes, i.e. random variables that evolve in time. Markov chains are relatively simple because the random variable is discrete and time is discrete as well. More importantly, Markov chain (and for that matter Markov processes in general) have the basic property that their future evolution is determine by their state at present time does not depend on their past.

3.1 Discrete Time Markov Chains

Here is an elementary example of Markov chain. Take a sequence $\{\xi_j\}_{j\in\mathbb{N}}$, where ξ_j are i.i.d random variables such that $\xi_j = \pm 1$ with probability $\frac{1}{2}$, and define the sequence $\{X_n\}_{n\in\mathbb{N}}$ via the following recurrence relation:

$$X_0 = 0 \qquad X_n = X_{n-1} + \xi_n \mod N,$$

This defines a random walk on $\{0, 1, \ldots, N-1\}$. Clearly, given the present state X_n, the distribution of X_{n+1} is determined, and one does not need to know about the past. This can be restated as the *Markov property*

$$\mathbf{P}(X_{n+1} = i_{n+1} | \{X_m = i_m\}_{m=1}^n) = \mathbf{P}(X_{n+1} = i_{n+1} | X_n = i_n),$$

i.e. the probability of X_{n+1} conditional on the whole past history of the sequence, $\{X_m\}_{m=1}^n$, reduces to probability of X_{n+1} conditional on the latest value alone, X_n. This last conditional probability is called the transition probability and it is of special importance for Markov chains. In this particular example, it is given by

$$\mathbf{P}(X_{n+1} = i \pm 1 | X_n = i) = \mathbf{P}(\xi_{n+1} = \pm 1) = \tfrac{1}{2}.$$

and $\mathbf{P}(X_{n+1} = \text{anything else } | X_n = i) = 0$.

A Markov chain does not need to be given as explicitly as in the recurrence relation above. More generally, given a state space S which we will assume finite

for simplicity, one can define the *transition probabilities*, i.e. the probabilities that the process is in state i at time $n+1$ given that it was in state j at time n

$$p_n(i|j) = \mathbf{P}(X_{n+1} = i | X_n = j).$$

The Markov chain is called *stationary* if $p_n(i|j)$ is independent of n, and from now on we will discuss only stationary Markov chains and let $p(i|j) = p_n(i|j)$. The *Markov property* then requires that

$$\begin{aligned}\mathbf{P}(X_{n+1} = j | X_n = k_n, X_{n-1} = k_{n-1}, \ldots, X_1 = k_1) \\ = \mathbf{P}(X_{n+1} = j | X_n = k_n),\end{aligned}$$

i.e. the value of X_{n+1} is determined completely by X_n and does not depend on the past history of X_n. This property allow to express the probability of any finite sequence $\{X_0, X_1, \ldots X_n\}$ in terms of the the initial distribution

$$\mu(i) = \mathbf{P}(X_0 = i),$$

and $p(i|j)$. Indeed, the Markov property implies that

$$\begin{aligned}\mathbf{P}(X_n = i_n, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \\ = p(i_n|i_{n-1}) p(i_{n-1}|i_{n-2}) \ldots p(i_1|i_0) \mu(i_0).\end{aligned}$$

From this we also get

$$\begin{aligned}\mathbf{P}(X_n = i | X_0 = j) &= \sum_{i_{n-1}, \ldots, i_1 \in S} p(i|i_{n-1}) p(i_{n-1}|i_{n-2}) \ldots p(i_1|j) \\ &=: p^{(n)}(i|j)\end{aligned}$$

$p^{(n)}(i|j)$ may be viewed as the (i,j)-th entry of the matrix P^n, where $P = (p(i|j))$. P is a *stochastic matrix*, i.e. $p(i|j)$ satisfies

$$p(i|j) \geq 0, \qquad \sum_{i \in S} p(i|j) = 1. \tag{3.1}$$

Given the initial distribution of the Markov chain p_0, the distribution of X_n is then given by

$$\mu_n = P^n \mu_0.$$

Exercise 3.1.1. μ_n satisfies the recurrence relation $\mu_n = P\mu_{n-1}$. Show that this equation can always be written as

$$\mu_n(i) = \Big(1 - \sum_{\substack{j \in S \\ j \neq i}} p(j|i)\Big) \mu_{n-1}(i) + \sum_{\substack{j \in S \\ j \neq i}} p(i|j) \mu_{n-1}(j).$$

In words, this equation says that the probability that the process be in state i at time n is the sum of the probability that the process was in state i at time $n-1$ and made no jump to another state (first term at the right hand-side), and the probabilities that the process was in state j at time $n-1$ and made a jump to state i (second term at the right hand-side).

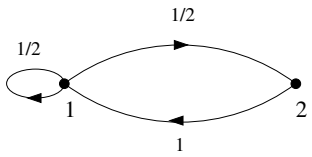

Figure 3.1: The graph of the Markov chain with $P = \begin{pmatrix} \frac{1}{2} & 1 \\ \frac{1}{2} & 0 \end{pmatrix}$

A Markov chain has a representation in terms of an oriented and weighted graph which facilitates many discussion about its properties. The nodes of the graph are the states of the chain, and an arrow with weight $p(i,j)$ connects state j to state i if $p(i,j) > 0$. For instance, the graph of the two state Markov chain with stochastic matrix

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

is shown in figure 3.1.

3.2 Ergodic chain and invariant distribution

Of special interest are the long-time properties of a Markov chain. In particular, we may ask:

1. Is there a unique *invariant distribution*? π is called an invariant distribution if

$$\pi = P\pi.$$

2. What are the convergence properties of an initial distribution μ_0 towards this invariant distribution?

When the state space S is finite, these can be phrased as standard linear algebra quuestions. Indeed the existence and uniqueness of an equilibrium distribution is equivalent to the existence of a nonnegative eigenvector of P with eigenvalue equal to 1 and multiplicity 1. Note that 1 is an eigenvalue of P because of the condition $\sum_{i \in S} p(i|j) = 1$ for all j, which can be written has $(1, 1, \ldots, 1)P = (1, 1, \ldots, 1)$.

Recall that if there exists a permutation matrix Q such that

$$QPQ^T = \begin{pmatrix} A_1 & B \\ 0 & A_2 \end{pmatrix}$$

then P is called *reducible*. Otherwise P is called *irreducible*. The following theorem essentially is a re-phrasing of Perron-Frobenius theorem which states that a irreducible positive matrix with spectral radius 1 (like P) has an eigenvalue 1 with multiplicity 1 and a strictly positive eigenvector:

Theorem 3.2.1. *If P is irreducible, then:*

1. *There exists a unique invariant distribution π. π is strictly positive.*

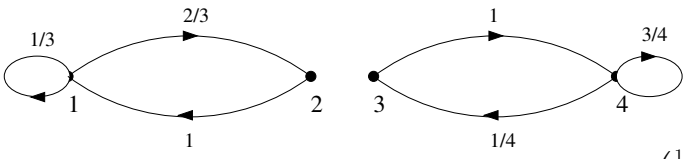

Figure 3.2: The graph of the disconnected Markov chain with $P = \begin{pmatrix} \frac{1}{3} & 1 & 0 & 0 \\ \frac{2}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} \\ 0 & 0 & 1 & \frac{3}{4} \end{pmatrix}$

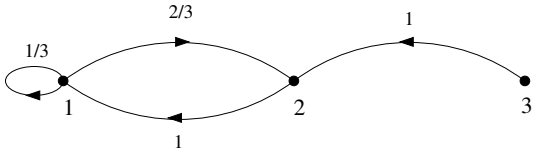

Figure 3.3: The graph of the non-ergodic Markov chain with $P = \begin{pmatrix} \frac{1}{3} & 1 & 0 \\ \frac{2}{3} & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$

2. *For any initial distribution* μ_0,

$$\mu_n = \bar{P}_n \mu_0 \to \pi$$

where

$$\bar{P}_n = \frac{1}{n} \sum_{j=1}^{n} P^n.$$

Exercise 3.2.1. Recall that the spectral radius of a matrix is the maximum of the absolute value of its eigenvalue. Show that a stochastic matrix has spectral radius 1.

A Markov chain whose stochastic matrix is irreducible is said to be *ergodic*. It implies that every state of the chain has a positive probability to be visited from every other state (can you see why this is not the case if P is reducible?). In terms of the graph of the chain, it means that from any node one can follow arrows to reach any other node. A Markov chain can fail to be ergodic for two reasons (or combinations thereof).

1. The chain may be disconnected, as in the example shown in figure 3.2. In this example, the chain can be decomposed into two sub-chains composed respectively of states 1 and 2 and states 3 and 4 which are each ergodic.

2. Some state of the chain may lead to others but may not be reachable from these, as in the example shown in figure 3.3. In this example, the chain can be made ergodic by removing state 3.

Generally, a non-ergodic Markov chain can be split into sub-chains which, after appropriate removal of non-communicating states like 3 in figure 3.3, are each ergodic.

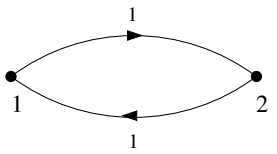

Figure 3.4: The graph of the periodic Markov chain with $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

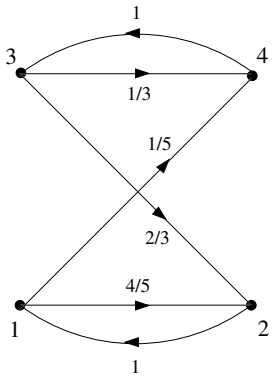

Figure 3.5: The graph of the periodic Markov chain with $P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{4}{5} & 0 & \frac{2}{3} & 0 \\ 0 & 0 & 0 & 1 \\ \frac{1}{5} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$

This also means that one need to carefully distinguish existence of a unique invaraint distribution and ergodicity of the Markov chain. For instance, the chain whose graph is shown in figure 3.3 is non-ergodic but it has a unique invariant distribution,

$$\pi = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Note that this invariant distribution is not strictly positive.

Exercise 3.2.2. Consider the two-state Markov chain with

$$P = \begin{pmatrix} 1-p & 1 \\ p & 0 \end{pmatrix}$$

Show that the chain is ergodic for $p \in (0, 1]$, and non-ergodic for $p = 0$. Find the invariant distribution π of the chain and show that every initial condition converges exponentially fast to π for $p \in [0, 1)$, and does not converge to π for $p = 1$ (in this case the time-averaged distribution converges to π).

It is also worth noting that Theorem 3.2.1 states that the time average of P^n and not P^n itself converges to π. To get a stronger result one needs to exclude periodic chains, like the one shown on figure 3.4 or the more general one shown in figure 3.5. A *periodic chain* is such that by lumping some of

its states together (like state 1 with state 3 and state 2 with state 4 in the example shown in figure 3.5), a new Markov chain is created whose dynamics is a deterministic cycling between its new states, i.e. whose stochastic matrix is of the form

$$P = \begin{pmatrix} 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}$$

If periodic Markov chains are excluded from the ergodic ones, we obtain a stronger result:

Theorem 3.2.2. *Assume that there exists an s such that $p^{(s)}(i|j) > 0$ for any pairs (i, j). Then*

1. *There exists a unique invariant distribution π. π is strictly positive.*
2. *For any initial distribution μ_0,*

$$\mu_n = P^n \mu_0 \to \pi \quad \textit{exponentially fast as } n \to \infty.$$

Proof of Theorem 3.2.2. We follow Sinai [12]. Given two distribution, μ_0 and μ_0', we define their distance has

$$d(\mu_0, \mu_0') = \tfrac{1}{2} \sum_{i \in S} |\mu_0(i) - \mu_0'(i)|.$$

Since

$$\begin{aligned} 0 &= \sum_{i \in S} (\mu_0(i) - \mu_0'(i)) \\ &= \sum_{i \in S} (\mu_0(i) - \mu_0'(i))_+ - \sum_{i \in S} (\mu_0'(i) - \mu_0(i))_+ , \end{aligned}$$

where $(a)_+ = a$ if $a > 0$ and $(a)_+ = 0$ otherwise, we also have

$$\begin{aligned} d(\mu_0, \mu_0') &= \tfrac{1}{2} \sum_{i \in S} (\mu_0(i) - \mu_0'(i))_+ + \tfrac{1}{2} \sum_{i \in S} (\mu_0'(i) - \mu_0(i))_+ \\ &= \sum_{i \in S} (\mu_0(i) - \mu_0'(i))_+ . \end{aligned}$$

This also shows that $d(\mu_0, \mu_0') \leq 1$. Let $\mu_s = \mathbf{P}^s \mu_0$, $\mu_s' = P^s \mu_0'$ and consider $d(\mu_s, \mu_s')$. We have

$$\begin{aligned} d(\mu_s, \mu_s') &= \sum_{i \in S} \Big(\sum_{j \in S} \Big(p^{(s)}(i|j) \mu_0(j) - p^{(s)}(i|j) \mu_0'(j) \Big) \Big)_+ \\ &\leq \sum_{i \in B_+} p^{(s)}(i|j) \sum_{j \in S} (\mu_0(j) - \mu_0'(j)) , \end{aligned}$$

where $B_+ \subset S$ is the subset where $\sum_{j\in S}\left(p^{(s)}(i|j)\mu_0(j) - p^{(s)}(i|j)\mu_0'(j)\right) > 0$. We now note that B_+ cannot contain all the elements of S for otherwise one must have $\sum_{j\in S} p^{(s)}(i|j)\mu_0(j) > \sum_{j\in S} p^{(s)}(i|j)\mu_0'(j)$ for all i and hence

$$\sum_{i,j\in S} p^{(s)}(i|j)\mu_0(j) > \sum_{i,j\in S} p^{(s)}(i|j)\mu_0(j),$$

which is impossible since both sides sum to 1. Therefore at least one element is missing in B_+ and, since $p^{(s)}(i|j) \geq \alpha > 0$ by assumption, we have $\sum_{i\in B_+} p^{(s)}(i|j) \leq (1-\alpha)$. Therefore

$$d(\mu_s, \mu_s') \leq d(\mu_0, \mu_0')(1-\alpha),$$

implying contraction. Similarly,

$$d(\mu_{sn}, \mu_{sn+m}) \leq d(\mu_0, \mu_m)(1-\alpha)^n \leq (1-\alpha)^n.$$

For n sufficiently large the right hand side can be made arbitrary small, implying that the sequence $\{\mu_n\}_{n\in N}$ is a Cauchy sequence. Hence it has a limit π which satisfies $\pi = \lim_{n\to\infty} P^n\mu_0 = \lim_{n\to\infty} P(P^n\mu_0) = P\pi$. The distribution π satisfying such a property is also unique for if there were two, π and π', then $d(\pi,\pi') = d(P^s\pi, P^s\pi') < d(\pi,\pi')$, implying $d(\pi,\pi') = 0$, i.e $\pi = \pi'$. □

Exercise 3.2.3. Show that necessarily the s in the theorem satisfies $s \leq N-1$, where N is the number of states of the chain.

3.3 Modification of the chain: First passage problems, etc.

Several questions about a Markov chain can be answered by modifying the chain. Here is a standard example. Suppose that $X_0 = i_\star$ and we ask about the first time that X_n will visit some state $j_\star \neq i_\star$, i.e

$$n_\star = \min_n\{n : X_n = j_\star, X_0 = i_\star\}.$$

$n_\star$ is a random variable, and we would like to derive an expression for its probability distribution. This can be done as follows. Take the transition probability of the original chain and modify the entries involving $j_\star$ as follows

$$\bar{p}(i|j_\star) = \begin{cases} 0 & \text{if } i \neq j_\star \\ 1 & \text{if } i = j_\star \end{cases}$$

If all the other entries are left as in the original chain, i.e. $\bar{p}(i|j) = p(i|j)$ if $j \neq j_\star$, then the motion in the modified chain is as in the original one, except that as

soon X_n reaches $j_\star$, it remains there indefinitely. Therefore, if $\bar{\mu}_n = \bar{P}^n \mu_0$ with $\mu_0(i) = 1$ if $i = i_\star$ and $\mu_0(i) = 0$ otherwise, then

$$\sum_{\substack{i \in S \\ i \neq j_\star}} \bar{\mu}_n(i)$$

gives the probability that X_n has not yet reached $j_\star$ at time n. But this is precisely the probability that $n_\star$ is bigger than n, i.e.

$$\mathbf{P}(n_\star > n) = \sum_{\substack{i \in S \\ i \neq j_\star}} \bar{\mu}_n(i).$$

Exercise 3.3.1. Let $S_\star \subset S$ and assume $X_0 = i_\star \in S_\star$. Derive an expression for the probability of the first exit time of X_n from $S_\star$.