Chapter 7 Stochastic Simulation

7.1 Markov Chains

7.1.1 Introduction

Suppose that $\mathcal{M} = \{X_n\}_{n=0}^{\infty}$, (or perhaps $\mathcal{M} = \{X_n\}_{n=-\infty}^{\infty}$) is a sequence of correlated random variables, where each X_n comes from some set Ω, .*called* the **state space.** We assume that states in Ω can be labeled by the integers, i.e., Ω is discrete. The process is a **Markov chain** if it satisfies the **Markov condition**

$$\Pr\left(X_{n+1} = j | X_n = i, X_{n-1} = x_{n-1}, ..., X_0 = x_0\right) = \Pr\left(X_{n+1} = j | X_n = i\right). \tag{7.1}$$

Fixing an **initial distribution** $\Pr\left(X_0 = i\right)$ for X_0 and the **transition probability** for X_{n+1} given X_n, $\Pr\left(X_{n+1} = j | X_n = i\right)$ determines a Markov chain.

If the transition probability does not depend on n, i.e., if

$$\Pr\left(X_{n+m+1} = j | X_{n+m} = i\right) = \Pr\left(X_{n+1} = j | X_n = i\right) \text{ for all } m \in \mathbb{Z}, \tag{7.2}$$

we say that $\mathcal{M}$ is **homogeneous** and we write the transition probability as a matrix $\mathbf{P}$ where

$$P_{ij} = \Pr\left(X_{n+1} = j | X_n = i\right). \tag{7.3}$$

Note that P_{ij} denotes the conditional probability to enter state j on the next step, given that the current state is i. The transition probabilities satisfy the normalization condition

$$\sum_{j \in \Omega} P_{ij} = 1 \tag{7.4}$$

since the chain must be in **some** state on the next step. A matrix with rows which sum to one is called **stochastic**.

Example 1 *Suppose that* $\Omega = \{1, 2, 3\}$, *the transition matrix is*

$$\mathbf{P} = \left(\begin{array}{ccc} \frac{2}{5} & \frac{1}{2} & \frac{1}{10} \\ \frac{1}{5} & \frac{7}{10} & \frac{1}{10} \\ \frac{2}{5} & \frac{2}{5} & \frac{1}{5} \end{array} \right),$$

and that the initial distribution is $\Pr\left(X_0 = i\right) = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. *We may represent the transition matrix of the Markov chain as a graph with a vertex for each state and a directed edge from vertex i to vertex j when there is a nonzero transition probability P_{ij} from i to j. Thus for the Markov chain above, we would have the digraph (directed graph) shown in Figure 7.1.*

Simulation Note that if $a + b + c = 1$ and we wish to pick a state in $\{1, 2, 3\}$ with probability a for 1, b for 2 and c for 3, we need only generate a random number p distributed uniformly on $[0, 1]$ and if $p < a$, pick 1, if $a < p < a + b$, pick 2 and if $a + b < p < a + b + c = 1$ pick c.

We will use this technique to simulate the chain in Example 1. The random numbers

u_1	u_2	u_3	u_4	u_5	u_6
0.429	0.156	0.146	0.951	0.921	0.644

have been sampled from a uniform distribution on $[0, 1]$. Our simulation proceeds as follows:

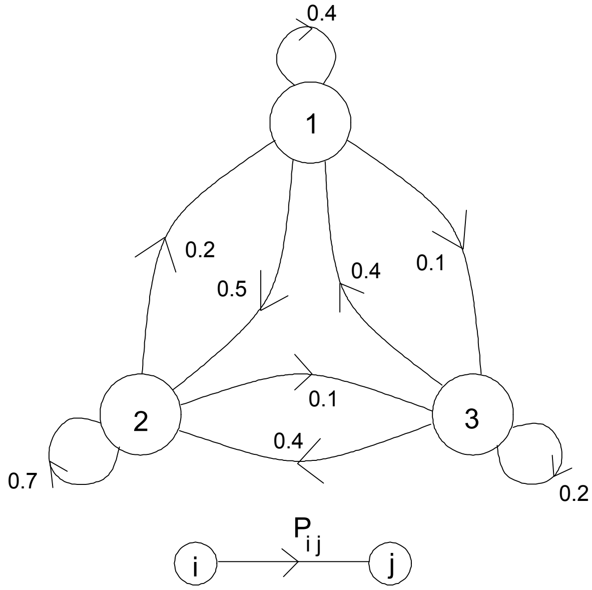

Figure 7.1 Digraph corresponding to the transition matrix of Example 1.

1. Pick X_0 using the initializing distribution $\Pr(X_0 = i) = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. Since we have $a = b = c = \frac{1}{3}$ and $u_1 = 0.429$, we select $X_0 = 2$.
2. We are now in state 2. We must choose a new state by jumping from state 2. Since $P_{2j} = \left(\frac{1}{5}, \frac{7}{10}, \frac{1}{10}\right)$, we have $a = \frac{1}{5}$, $b = \frac{7}{10}$ and $c = \frac{1}{10}$. Since $u_2 = 0.156 < a$, we select $X_1 = 1$.
3. We are now in state 1. Iterating, since $P_{1j} = \left(\frac{2}{5}, \frac{1}{2}, \frac{1}{10}\right)$, we have $a = \frac{2}{5}$, $b = \frac{1}{2}$ and $c = \frac{1}{10}$. Since $u_3 = 0.146 < a$, we select $X_2 = 1$.

We thus obtain

X_0	X_1	X_2	X_3	X_4	X_5	...
2	1	1	3	3	2	...

which we call a a **realization** of the chain. By simulating the process we obtain a realization of the process. We see from the simulation procedure that the stochastic process satisfies the Markov condition: we only used the value of the last state X_n to calculate the distribution forX_{n+1}.

7.1.2 The Distribution of X_n

Consider $\Pr(X_n = j)$, which is the probability that after a simulation of a Markov chain for n steps, the state reached is $X_n = j$. These probabilities may be arranged in a **row vector** $\pi^{(n)}$ where, by definition $\pi_j^{(n)} = \Pr(X_n = j)$. When $n = 1$, we see that

$$\begin{aligned}\Pr(X_1 = j) &= \sum_{i\in\Omega} \Pr(X_1 = j,\ X_0 = i) \\ &= \sum_{i\in\Omega} \Pr(X_1 = j | X_0 = i) \Pr(X_0 = i)\,. \qquad (7.5)\end{aligned}$$

This may be written in matrix form as

$$\pi_j^{(1)} = \sum_{i\in\Omega} \pi_i^{(0)} P_{ij} \text{ or } \pi^{(1)} = \pi^{(0)}\mathbf{P}. \qquad (7.6)$$

Similarly,

$$\pi^{(n)} = \pi^{(n-1)}\mathbf{P}. \tag{7.7}$$

Suppose that for some π, we have that

$$\pi = \pi\mathbf{P} \tag{7.8}$$

i.e., π is a left eigenvector of $\mathbf{P}$ with eigenvalue 1, normalized such that $\sum \pi_i = 1$. Then π is called a **stationary distribution** for $\mathbf{P}$, since if $\pi^{(n)} = \pi$, then $\pi^{(n+1)} = \pi^{(n)}\mathbf{P} = \pi$ also, i.e., once the chain is in distribution π, it stays in that distribution.

Example 2 $\pi = \left(\frac{5}{18}, \frac{11}{18}, \frac{1}{9}\right)$ *is the stationary distribution for the matrix in the first example, i.e.,*

$$\left(\begin{array}{ccc} \frac{5}{18} & \frac{11}{18} & \frac{1}{9} \end{array} \right) = \left(\begin{array}{ccc} \frac{5}{18} & \frac{11}{18} & \frac{1}{9} \end{array} \right) \left(\begin{array}{ccc} \frac{2}{5} & \frac{1}{2} & \frac{1}{10} \\ \frac{1}{5} & \frac{7}{10} & \frac{1}{10} \\ \frac{2}{5} & \frac{2}{5} & \frac{1}{5} \end{array} \right) \tag{7.9}$$

Definition 3 *Suppose that $\pi^{(n)} \to \pi$ as $n \to \infty$ for any $\pi^{(0)}$. Then π is the* ***equilibrium distribution*** *of the chain $\mathcal{M}$ and the chain is said to be* ***ergodic.***

For an ergodic chain, and for sufficiently large n, the states of $\mathcal{M}$ are distributed like π and the system is "in equilibrium".

Exercise 7.1.1 *Suppose that $\mathcal{M}$ is ergodic with equilibrium distribution π. Show that as $n \to \infty$,*

$$\mathbf{P}^n \to \left(\begin{array}{ccc} \cdots & \pi & \cdots \\ \cdots & \pi & \cdots \\ & \vdots & \\ \cdots & \pi & \cdots \end{array} \right). \tag{7.10}$$

where $\mathbf{P}^n = \mathbf{P}\mathbf{P}\ldots\mathbf{P}$ matrix-multiplied n times.

Solution: By iterating $\pi^{(n)} = \pi^{(n-1)}\mathbf{P}$ we see that $\pi^{(n)} = \pi^{(0)}\mathbf{P}^n$. If $\mathcal{M}$ is ergodic, then $\pi^{(n)} \to \pi$ for any $\pi^{(0)}$. In particular if we choose $\pi^{(0)} = (0, 0, ..., 1, ..., 0)$ where the 1 is in the k'th location, $\pi^{(0)}\mathbf{P}^n$ is equal to the k'th row of $\mathbf{P}^n$ and, by assumption, this tends to π as $n \to \infty$. Since this holds for every k, the matrix $\mathbf{P}^n$ has the form shown.

Exercise 7.1.2 *Verify that*

$$\left(\begin{array}{ccc} \frac{2}{5} & \frac{1}{2} & \frac{1}{10} \\ \frac{1}{5} & \frac{7}{10} & \frac{1}{10} \\ \frac{2}{5} & \frac{2}{5} & \frac{1}{5} \end{array} \right)^n \to \frac{1}{18} \left(\begin{array}{ccc} 5 & 11 & 2 \\ 5 & 11 & 2 \\ 5 & 11 & 2 \end{array} \right)$$

as $n \to \infty$ as follows from the examples above.

Exercise 7.1.3 *Determine the stationary distribution if* $\mathbf{P} = \left(\begin{array}{cc} 1-a & a \\ b & 1-b \end{array} \right)$ *where* $0 \le a, b \le 1$.

Solution: If $0 < a + b < 2$, then $\pi = \left(\begin{array}{cc} b/(a+b) & a/(a+b) \end{array} \right)$ is the stationary distribution. If $a = b = 0$, then $\Omega = \{0, 1\}$ is **reducible** under $\mathbf{P}$, i.e., depending upon the initial state, $\pi^{(n)} \to \left(\begin{array}{cc} 1 & 0 \end{array} \right)$ or $\pi^{(n)} \to \left(\begin{array}{cc} 0 & 1 \end{array} \right)$ and $\pi = c\left(\begin{array}{cc} 0 & 1 \end{array} \right) + d\left(\begin{array}{cc} 0 & 1 \end{array} \right)$ is a stationary distribution for any c and d. On the other hand, if $a = b = 1$, the Markov chain is periodic.

7.1.3 Equilibrium Distributions

Under what circumstances does a stationary distribution exist ? If one exists, is it unique ? Does $\pi^{(n)} \to \pi$ for any initial $\pi^{(0)}$? We seek sufficient conditions for ergodicity.

7.1.3.1 Irreducibility

If we can find a sequence of states

$$i \to k_1 \to k_2 \to \cdots \to k_n \to j \tag{7.11}$$

such that the transition probabilities $P_{i,k_1} \neq 0$, $P_{k_m,k_{m+1}} \neq 0$, $P_{k_n,j} \neq 0$, then there is a sequence of states from i to j with a non-zero probability of occurring in $\mathcal{M}$. We say that "i and j communicate" and write $i \to j$. If j and i also communicate, i.e., if $j \to i$, we say that i and j "intercommunicate" and write $i \leftrightarrow j$. Sets of intercommunicating states form equivalence classes, since $i \leftrightarrow m$ and $m \leftrightarrow j \Rightarrow i \leftrightarrow j$, and likewise $i \leftrightarrow m$ but $m \not\leftrightarrow j \Rightarrow i \not\leftrightarrow j$.

If all states in Ω intercommunicate, then Ω is said to be **irreducible under P**, i.e., for any two states i and j in Ω, there is a path with non-zero probability which links i to j and a path with non-zero probability which links j to i. Otherwise, Ω is **reducible under P**.

If there is more than one distinct equivalence class of intercommunicating states in Ω, the Markov chain is reducible under $\mathbf{P}$ and a stationary distribution need not be unique.

Example 4 *Suppose that* $\Omega = \{1, 2, 3, 4\}$ *and that*

$$\mathbf{P} = \begin{pmatrix} 0.4 & 0.6 & 0 & 0 \\ 0.2 & 0.8 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 \\ 0 & 0 & 0.2 & 0.8 \end{pmatrix} \tag{7.12}$$

Then it is clear (from the associated digraph) that $\{1, 2\}$ *and* $\{3, 4\}$ *are the equivalence classes of intercommunicating states. There are two left eigenvectors of* $\mathbf{P}$ *with eigenvalue* 1, *namely* $\sigma = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} & 0 & 0 \end{pmatrix}$ *and* $\rho = \begin{pmatrix} 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix}$. *If the initial state* $X_0 \in \{1, 2\}$, *the stationary distribution is* σ *and if* $X_0 \in \{3, 4\}$ *the stationary distribution is* ρ.

7.1.3.2 Reversibility

If I give you a realization of a reversible Markov process, but don't tell you in which direction the chain was simulated, you will not be able to figure out the simulation direction by looking at the sequence of states in the realization. This is a rare and special property.

Consider, for example, a more hum-drum chain on $\Omega = \{1, 2, 3\}$ with the transition matrix

$$\mathbf{P} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \tag{7.13}$$

This defines an irreducible chain. We notice that the sequence $1 \to 3 \to 2 \to 1$ is possible in this chain but the reverse sequence $1 \to 2 \to 3 \to 1$ is not possible since $P_{23} = 0$. Thus there is a sequence of states for which it is possible to tell in which direction the simulation occurred and the chain is not reversible.

Lemma 5 *Let $\mathcal{M} = \{X_n\}_{n=-\infty}^{\infty}$ be a homogeneous Markov chain with transition matrix* $\mathbf{P}$*. Then if we define the random variables $Y_n = X_{-n}$, the sequence $\mathcal{M}' = \{Y_n\}_{n=-\infty}^{\infty}$ is also a Markov chain, called the **reversed chain** of $\mathcal{M}$.*

Proof: Consider the conditional probability $\Pr(Y_{n+1}|Y_n, Y_{n-1}, ...)$. We wish to show that this is equal to $\Pr(Y_{n+1}|Y_n)$. By definition of conditional probabilities, we see that

$$\begin{aligned}\Pr(Y_{n+1}|Y_n, Y_{n-1}, ...) &= \frac{\Pr(Y_{n+1}, Y_n, Y_{n-1}, ...)}{\Pr(Y_n, Y_{n-1}, ...)} \\ &= \frac{\Pr(Y_{n-1}, Y_{n-2}, ...|Y_{n+1}, Y_n)\Pr(Y_{n+1}, Y_n)}{\Pr(Y_n, Y_{n-1}, ...)} \\ &= \frac{\Pr(X_{-n+1}, X_{-n+2}, ...|X_{-n-1}, X_{-n})\Pr(Y_{n+1}, Y_n)}{\Pr(Y_n, Y_{n-1}, ...)} \qquad (7.14)\end{aligned}$$

Using the Markov property of $\mathcal{M}$ we have that

$$\Pr(X_{-n+1}, X_{-n+2}, ...|X_{-n-1}, X_{-n}) = \Pr(X_{-n+1}, X_{-n+2}, ...|X_{-n}), \qquad (7.15)$$

hence

$$\begin{aligned}\Pr(Y_{n+1}|Y_n, Y_{n-1}, ...) &= \frac{\Pr(X_{-n+1}, X_{-n+2}, ...|X_{-n})\Pr(Y_{n+1}, Y_n)}{\Pr(Y_n, Y_{n-1}, ...)} \\ &= \frac{\Pr(Y_{n-1}, Y_{n-2}, ...|Y_n)\Pr(Y_{n+1}, Y_n)}{\Pr(Y_n, Y_{n-1}, ...)} \\ &= \frac{\Pr(Y_{n+1}, Y_n)}{\Pr(Y_n)} = \Pr(Y_{n+1}|Y_n). \qquad (7.16)\end{aligned}$$

Thus $\mathcal{M}'$ is also a Markov chain.

Definition 6 *A homogeneous Markov chain $\mathcal{M}$ is **reversible** if the transition matrix for the reversed chain $\mathcal{M}'$ coincides with that for $\mathcal{M}$, so that*

$$\Pr(X_{n+1} = j|X_n = i) = \Pr(X_n = j|X_{n+1} = i)$$

If the transition matrices are the same, the value of any statistic we care to measure on a realization will have the same distribution, whether the realization came from the forward or reversed chain. That's why we cant determine what the simulation direction was by looking at the output.

7.1.3.3 Necessary and sufficient conditions for reversibility

Theorem 7 *Suppose that $\mathcal{M} = \{X_n\}_{n=-\infty}^{\infty}$ is a Markov chain with transition matrix* $\mathbf{P}$*, unique stationary distribution π, and that for all n, X_n is distributed as π. $\mathcal{M}$ is reversible iff*

$$\pi_i P_{ij} = \pi_j P_{ji} \text{ for all } i, j \in \Omega. \qquad (7.17)$$

*This is often called the **detailed balance condition.***

Proof: Let $\mathbf{Q} = (Q_{ij})$ be the transition matrix of the reversed chain $\mathcal{M}' = \{Y_n\}_{n=-\infty}^{\infty}$ where $Y_n = X_{-n}$, i.e.,

$$Q_{ij} = \Pr(Y_{n+1} = j|Y_n = i). \qquad (7.18)$$

We must prove that $Q_{ij} = P_{ij}$ iff detailed balance holds. By definition of the process Y,

$$\begin{aligned} Q_{ij} &= \Pr\left(X_{-n-1} = j | X_{-n} = i\right) \\ &= \frac{\Pr\left(X_{-n-1} = j, X_{-n} = i\right)}{\Pr\left(X_{-n} = i\right)} \\ &= \frac{\Pr\left(X_{-n} = i | X_{-n-1} = j\right) \Pr\left(X_{-n-1} = j\right)}{\Pr\left(X_{-n} = i\right)} \end{aligned} \tag{7.19}$$

which is essentially a statement of Bayes' theorem. Since each X_n is distributed as π for all n,

$$\begin{aligned} Q_{ij} &= \frac{\Pr\left(X_{-n} = i | X_{-n-1} = j\right) \pi_j}{\pi_i} \\ &= \frac{P_{ji}\pi_j}{\pi_i}, \end{aligned} \tag{7.20}$$

by definition of the transition matrix $\mathbf{P}$ of $\mathcal{M}$. Then $Q_{ij} = P_{ij}$ and the chain is reversible if and only if detailed balance holds.

Example 8 *Ehrenfest model of Diffusion*

Given the microscopic dynamics of a reversible physical process, this example shows how the stationary distribution may be found by using the detailed balance condition.

Consider two boxes A and B connected to each other. Within the two boxes are m particles. At time $t \in \mathbb{Z}$, let X_t represent the number of particles in box A, so that there are $m - X_t$ particles in box B. Select one of the m particles at random and transfer it to the other box (whichever that is). This leads to the situation at time $t + 1$.

We shall verify that $\mathcal{M} = \{X_t\}_{t=0}^{\infty}$ is a homogeneous Markov chain. Its transition matrix $\mathbf{P}$ satisfies $P_{ij} = 0$ unless $j = i - 1$ or $j = i + 1$. Now, $P_{i,i-1}$ is the probability that $X_{t+1} = X_t - 1 = i - 1$ or the probability that the randomly selected particle is in box A which is i/m. Similarly, $P_{i,i+1} = (m - i)/m$.

Any given sequence of moves is equally likely to occur forward or in reverse (imagine putting a label on each ball), so the process is reversible. Thus the stationary distribution π must satisfy the detailed balance condition. Setting $j = i + 1$ in the condition,

$$\pi_i P_{i,i+1} = \pi_{i+1} P_{i+1,i} \tag{7.21}$$

or

$$\begin{aligned} \pi_i \left(\frac{m-i}{m}\right) &= \pi_{i+1} \left(\frac{i+1}{m}\right) \\ \pi_{i+1} &= \left(\frac{m-i}{i+1}\right) \pi_i \end{aligned} \tag{7.22}$$

Using this relationship, we see that

$$\pi_1 = m\pi_0,\ \pi_2 = \frac{m-1}{2}\pi_1,\ \pi_3 = \frac{m-2}{3}\pi_2, ... \tag{7.23}$$

so that

$$\pi_i = \frac{m(m-1)\cdots(m-i+1)}{i!}\pi_0 = \binom{m}{i}\pi_0. \tag{7.24}$$

We can find all π_i by the normalization condition since

$$1 = \sum_{i=0}^{m} \pi_i = \sum_{i=0}^{m} \binom{m}{i} \pi_0 = 2^m \pi_0. \tag{7.25}$$

Hence the desired stationary distribution is

$$\pi_i = 2^{-m} \binom{m}{i}, \tag{7.26}$$

which is a binomial distribution $B\left(m, \frac{1}{2}\right)$.

7.1.3.4 Periodic chains (a technical detail)

If the transition matrix $\mathbf{P}$ of a Markov chain has a zero diagonal, i.e., if $P_{ii} = 0$ for all i, the chain may be **periodic**. Let $\mathcal{M}$ be an irreducible Markov chain with transition matrix $\mathbf{P}$ and let i be a fixed state. Define the set

$$T = \left\{ k : \left(\mathbf{P}^k \right)_{ii} > 0,\ k > 0 \right\}. \tag{7.27}$$

These are the steps on which it is possible for a chain which starts in state i to revisit i. The greatest common divisor of the integers in T is called the **period** of the state i.

It is possible to show that the period of an irreducible chain for the state i is a constant independent of i. The chain is said to be periodic if the period of any of its states is greater than one.

Example 9 *Consider the chain defined by the state space* $\Omega = \{1, 2, 3, 4\}$, *the transition matrix*

$$\mathbf{P} = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}, \tag{7.28}$$

and the initial state $X_0 = 1$.

This is periodic with $d = 2$ *since*

$$\Pr\left(X_n = 2 | X_0 = 1\right) = \begin{cases} 0 & \text{if } n \text{ is even} \\ \frac{1}{2} & \text{if } n \text{ is odd} \end{cases} \tag{7.29}$$

Periodic chains are not ergodic, though the difficulty can be eliminated by sub-sampling.

7.1.3.5 Ergodicity theorem for reversible chains

Theorem 10 *For an irreducible, aperiodic Markov chain* $\mathcal{M}$ *on a countable state space with transition matrix* $\mathbf{P}$, *if there exists* $\pi = (\pi_i)$ *such that* $0 \leq \pi_i \leq 1$, $\sum_i \pi_i = 1$ *and*

$$\pi_i P_{ij} = \pi_j P_{ji}, \tag{7.30}$$

then $\mathcal{M}$ *is reversible and ergodic with unique equilibrium distribution* π.

Notice that if we sum the detailed balance condition with respect to j, we obtain

$$\pi_i \left(\sum_j P_{ij} \right) = \sum_j \pi_j P_{ji} \tag{7.31}$$

or

$$\pi_i = \sum_j \pi_j P_{ji} \tag{7.32}$$

which shows that π is a stationary distribution. In general, given a stochastic process with transition matrix $\mathbf{P}$, it is very difficult to find the equilibrium distribution π unless the chain is reversible. We have also assumed that Ω is countable and that $\mathcal{M}$ is homogeneous.

Exercise 7.1.4 *Suppose $\mathcal{M}$ is an irreducible Markov chain on a finite space of N states, with $P_{ji} = P_{ij}$, ie the transition matrix is symmetric. Prove $\mathcal{M}$ is reversible, and that the uniform distribution $\pi_i = 1/N$ is its unique equilibrium distribution.*

Solution: See Example 11 in the next section.

Reading

G.R. Grimmet and D.R. Stirzaker, *Probability and Random Processes.*

7.2 Markov Chain Monte Carlo

7.2.1 Some special cases

In the previous section, we considered the problem of finding the equilibrium distribution π of a chain with a given transition matrix $\mathbf{P}$. In this section, we consider the reverse problem. Given a distribution π, how do we construct the transition matrix $\mathbf{P}$ of a Markov chain so that the equilibrium distribution of this Markov chain is π?

A Markov chain can be specified by a transition matrix, or by giving the microscopic dynamics (*ie*, an algorithm which determines X_{n+1} given X_n). The algorithm implicitly fixes the transition matrix. Real problems are usually too complex for a transition matrix to be given explicitly. So the problem of specifying a MC with a given desired equilibrium distribution will boil down to the problem of providing an algorithm which we can prove generates a MC with the right equilibrium distribution. We use the idea of **reversibility.**

Example 11 *Construct an ergodic, reversible Markov chain $\mathcal{M}$ on the state space $\Omega = \{1, 2, 3, 4\}$ with equilibrium distribution $\pi_i = \frac{1}{4}$, (i.e., uniform on Ω.)*

Since $\mathcal{M}$ is to be reversible, the transition matrix must satisfy

$$\pi_i P_{ij} = \pi_j P_{ji}. \tag{7.33}$$

In order for $\pi_i = \pi_j$, we have $P_{ij} = P_{ji}$. If Ω is irreducible under $\mathbf{P}$ (i.e., the chain can get to any state in Ω) then we have satisfied the conditions of the ergodicity theorem for reversible chains. So any **symmetric, irreducible** transition matrix $\mathbf{P}$ will do the job.

If for example,

$$\mathbf{P} = \begin{pmatrix} 3/4 & 1/4 & 0 & 0 \\ 1/4 & 1/2 & 1/4 & 0 \\ 0 & 1/4 & 1/2 & 1/4 \\ 0 & 0 & 1/4 & 3/4 \end{pmatrix} \tag{7.34}$$

we satisfy $\sum_j P_{ij} = 1$ and $P_{ij} = P_{ji}$. Now $\pi = \begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}$ is a left eigenvector of $\mathbf{P}$. All the conditions for ergodicity are satisfied so we expect that $\pi^{(n)} = \pi^{(0)}\mathbf{P}^n$ tends to $\begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}$ as $n \to \infty$ from any start. Explicitly, we find that

$$\mathbf{P}^2 = \begin{pmatrix} .625 & .3125 & .0625 & 0 \\ .3125 & .375 & .25 & .0625 \\ .0625 & .25 & .375 & .3125 \\ 0 & .0625 & .3125 & .625 \end{pmatrix}, \tag{7.35}$$

$$\mathbf{P}^4 = \begin{pmatrix} .49219 & .32813 & .14063 & .03906 \\ .32813 & .30469 & .22656 & .14063 \\ .14063 & .22656 & .30469 & .32813 \\ .03906 & .14063 & .32813 & .49219 \end{pmatrix}, \tag{7.36}$$

$$\mathbf{P}^{100} = \begin{pmatrix} .25000\,00567 & .25000\,00235 & .24999\,99765 & .24999\,99433 \\ .25000\,00235 & .25000\,00097 & .24999\,99903 & .24999\,99765 \\ .24999\,99765 & .24999\,99903 & .25000\,00097 & .25000\,00235 \\ .24999\,99433 & .24999\,99765 & .25000\,00235 & .25000\,00567 \end{pmatrix}. \tag{7.37}$$

Note that this problem is trivial as we could have simply taken $P_{ij} = 1/4$ for all i, j in Ω. At each update, we sample uniformly on Ω. Then $\pi^{(n)} = \pi$ for all n.

Exercise 7.2.1 *Consider the state space* $\Omega = \{(a, b, c) : a + b + c = 0 \text{ and } a, b, c \in \{-9, -8, ..., 8, 9\}\}$. *Construct a reversible Markov chain with an equilibrium distribution which is uniform on* Ω.

Solution: In order to move from state $X_n = i$ to state $X_{n+1} = j$ where i and j are in Ω, pick a vector $\mathbf{u}$ with elements $(0, 1, -1)$ in random order, and set $j = i + \mathbf{u}$. This move respects the constraint that $a + b + c = 0$. If $i + \mathbf{u}$ is not in Ω, the move is not made and we set $X_{n+1} = X_n = i$. There are six choices for $\mathbf{u}$. We notice that

1. For each pair of states i and j, there is either one or no vector $\mathbf{u}$ of the above form which relates them. If $j = i + \mathbf{u}$, then $i = j + (-\mathbf{u})$. Since the probability to pick $\mathbf{u}$ is the same as the probability to pick $-\mathbf{u}$, we see that $P_{ij} = P_{ji} = \frac{1}{6}$ if there is a $\mathbf{u}$ such that $i = j + \mathbf{u}$. Also, $P_{ij} = P_{ji} = 0$ if no such $\mathbf{u}$ exists.

2. The rule that we do not make moves taking us outside Ω only affects the probability P_{ii} for some i and does not spoil the symmetry of the transition matrix $\mathbf{P}$.

Since $\mathbf{P}$ is symmetric, it is clear that a uniform distribution π on Ω will satisfy $\pi_i P_{ij} = \pi_j P_{ji}$. The chain is clearly irreducible as for any pair of states i and j in Ω, there is a sequence of valid vectors $\mathbf{u}$ which can take us from i to j. By the ergodicity theorem for reversible chains,

$$\Pr(X_n = i) \to \frac{1}{|\Omega|}, \text{ i.e., uniform on } \Omega \tag{7.38}$$

for any initial distribution. Note that $|\Omega|$ denotes the number of states in Ω.

7.2.2 Metropolis-Hastings Markov Chain Monte Carlo

We seem to be able to handle requests for samples from a uniform distribution. How about generating non-uniform distributions ? Metropolis-Hastings Markov Chain Monte Carlo is a certain type of algorithm which generates a MC with equilibrium distribution π. The user gets to say what kind of pie they want. The algorithm is as follows:

Let $X_n = i$. X_{n+1} is determined in the following way.

1. **Generation step:** Generate a candidate state j from i with some distribution $g(j|i)$. $g(j|i)$ is a fixed distribution that we are free to choose, so long that it satisfies the conditions
    (a) $g(j|i) = 0 \Rightarrow g(i|j) = 0$, (cant go forward implies cant go back)
    (b) $g(j|i)$ is the transition matrix of an irreducible Markov chain on Ω.
2. **Acceptance step:** With probability
$$\alpha(j|i) \equiv \min\left\{1, \frac{\pi_j}{\pi_i}\frac{g(i|j)}{g(j|i)}\right\}, \tag{7.39}$$
set $X_{n+1} = j$ (i.e., "accept" j), otherwise set $X_n = i$ (i.e., "reject" j).

Note that since the generation and acceptance steps are independent, the transition probability to go from i to j is

$$\Pr(X_{n+1} = j | X_n = i) = g(j|i)\,\alpha(j|i) \text{ provided that } i \neq j. \tag{7.40}$$

This says that the probability to land in state j from the current state i is equal to the probability to generate j from the current state i times the probability that the new state j is accepted. The probability P_{ii} can be found from the requirement that $\sum_j P_{ij} = 1$.

Assertion: (Metropolis *et al.* 1953, Hastings 1970)

Let π be a given probability distribution. The Markov chain simulated by the Metropolis-Hastings algorithm is reversible with respect to π. If it is also irreducible and aperiodic, then it defines an ergodic Markov chain with unique equilibrium distribution π.

Proof: We have to show that the transition matrix $\mathbf{P}$ determined by the MH algorithm satisfies

$$\pi_i P_{ij} = \pi_j P_{ji} \tag{7.41}$$

for all $i \neq j$ (since the case for $i = j$ is trivial). If this is the case, then the chain is reversible and the rest of the assertion follows from the ergodicity theorem of the first section.

Assume without loss of generality that

$$\pi_j g(i|j) > \pi_i g(j|i). \tag{7.42}$$

Since

$$\begin{aligned} P_{ij} &= g(j|i)\,\alpha(j|i) \\ &= g(j|i)\min\left\{1, \frac{\pi_j}{\pi_i}\frac{g(i|j)}{g(j|i)}\right\} = g(j|i), \end{aligned} \tag{7.43}$$

by assumption. On the other hand

$$\begin{aligned} P_{ji} &= g(i|j)\,\alpha(i|j) \\ &= g(i|j)\min\left\{1, \frac{\pi_i}{\pi_j}\frac{g(j|i)}{g(i|j)}\right\} = g(i|j)\,\frac{\pi_i}{\pi_j}\frac{g(j|i)}{g(i|j)} \\ &= \frac{\pi_i}{\pi_j} g(j|i) \end{aligned} \tag{7.44}$$

again by assumption. From these expressions, it is clear that $\pi_i P_{ij} = \pi_j P_{ji}$ as required.

Note that if $g(j|i)$ is not irreducible, then P_{ij} is not irreducible. However, irreducibility of $g(j|i)$ is not sufficient to guarantee irreducibility of P_{ij} since it may be that $\alpha(j|i) = 0$ (if $\pi_j = 0$), and this may prevent the chain from reaching all states in Ω.

Example 12 *Use the Metropolis-Hastings method to construct a reversible, ergodic Markov chain on* $\Omega = \{1, 2, 3\}$ *with equilibrium distribution* $\pi = \left(\frac{5}{18}, \frac{11}{18}, \frac{2}{18}\right)$.

Solution: Let us suppose that $g_{ij} \equiv g(j|i) = \frac{1}{3}$ for all i and j, i.e., the generation step picks a candidate state uniformly from Ω. By the definition of $\alpha_{ij} \equiv \alpha(j|i)$, we see that

$$\alpha_{ij} = \min\left\{1, \frac{\pi_j}{\pi_i}\right\} = \begin{pmatrix} 1 & 1 & 2/5 \\ 5/11 & 1 & 2/11 \\ 1 & 1 & 1 \end{pmatrix}. \tag{7.45}$$

Since $P_{ij} = g_{ij}\alpha_{ij}$ for $i \neq j$ and $P_{ii} = 1 - \sum_{j \neq i} P_{ij}$, gives the diagonal entries, we calculate

$$\mathbf{P} = \begin{pmatrix} 8/15 & 1/3 & 2/15 \\ 5/33 & 26/33 & 2/33 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}.$$

It is easy to check that $\pi\mathbf{P} = \pi$ and that the rows of $\mathbf{P}^n$ tend to π as n becomes large.

Example 13 *Construct the transition matrix* $\mathbf{P}$ *of a Markov chain with state space* $\{0, 1, 2, ...\}$ *and equilibrium distribution*

$$\pi_i = \frac{\mu^i \exp(-\mu)}{i!}, \tag{7.46}$$

i.e., a Poisson distribution with mean $\mu \in \mathbb{R}^+$.

As usual, there are two parts, an update scheme to generate candidate states and a scheme to accept or reject these states.

Let $X_n = i$. X_{n+1} is determined in the following way.

1. **Generation:** Any simple scheme will do, e.g.,

$$g(j|i) = \begin{cases} 1/2 & \text{if } j = i + 1 \\ 1/2 & \text{if } j = i - 1 \\ 0 & \text{otherwise} \end{cases} \tag{7.47}$$

   i.e., given $X_n = i$, the candidate state for X_{n+1} is chosen uniformly from $\{i + 1, i - 1\}$.

2. **Accept / reject:** The acceptance probability is determined by the Metropolis-Hastings formula
$$\alpha\,(j|i) = \min\left\{1, \frac{\pi_j\, g\,(i|j)}{\pi_i\, g\,(j|i)}\right\}. \tag{7.48}$$
Now $g\,(i|j) = g\,(j|i) = 1/2$ so
$$\alpha\,(i+1|i) = \min\left\{1, \frac{\pi_{i+1}}{\pi_i}\right\} = \min\left\{1, \frac{\mu}{i+1}\right\}, \tag{7.49}$$
and
$$\alpha\,(i-1|i) = \min\left\{1, \frac{\pi_{i-1}}{\pi_i}\right\} = \min\left\{1, \frac{i}{\mu}\right\}. \tag{7.50}$$
Note that if $X_n = 0$ and the candidate state $j = -1$ is chosen, we reject it and set $X_{n+1} = 0$. This affects P_{ii} when $i = 0$ but leaves P_{ij} correct for detailed balance to hold.

Exercise 7.2.2 *Implement the algorithm for the above example in you favorite computer language. Using $\mu = 1$, check the mean and variance of the samples $X_0 \ldots X_n$ are close to μ when n is large.*

Example 14 *The Binary Markov Random Field*

Consider the space $\Omega = \{(x_1, x_2, ..., x_{N^2}) : x_m \in \{-1, 1\}\}$. We may regard x_m as the colour of the m'th pixel where $x_m = -1$ denotes a black pixel and $x_m = 1$ denotes a white pixel. Then Ω is the space of all black and white images (Figure 7.2) with N^2 pixels. Suppose that the probability of $x \in \Omega$ is given by

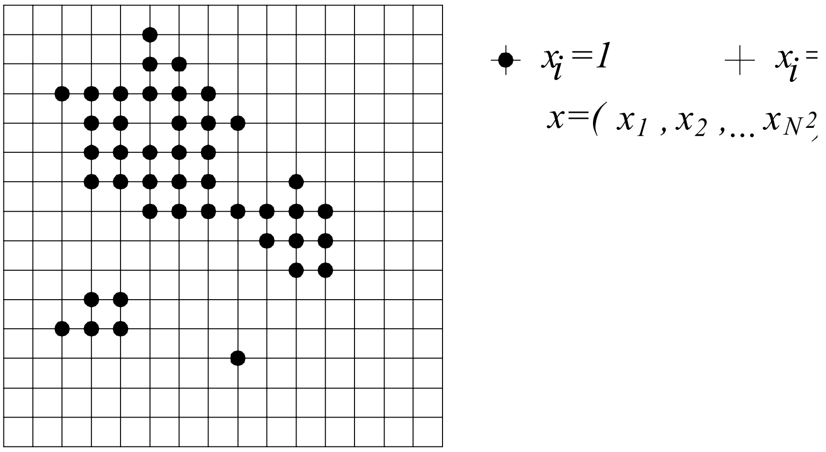

Figure 7.2 The Ising model on a square lattice with free boundaries.

$$\Pr\,(x) = \frac{1}{\mathcal{Z}'} \exp\left(J \sum_{\langle m,n\rangle} x_m x_n\right) \tag{7.51}$$
$$= \frac{1}{\mathcal{Z}} \exp\,(-2J\#x) \tag{7.52}$$

for some $J > 0$ and where $\langle m, n\rangle$ indicates a sum over pixels m and n which are adjacent on the image lattice, and $\#x$ is the number of edges connecting pixels of unequal value. The constant $\mathcal{Z}$ normalizes the probability distribution.

Exercise 7.2.3 *Show that* $\sum_{\langle m,n\rangle} x_m x_n = -2\#x + 2N^2 - 2N$. *What happened to* $\exp(2N^2 - 2N)$ *?*

The probability distribution above favours smooth images. We wish to generate samples $X \sim \Pr(X = x)$ from this distribution. Each sample X is an array of N^2 binary values $x = (x_1, x_2, ..., x_{N^2})$. The normalising constant $\mathcal{Z}$ is not in general available (though see L. Onsager, Phys Rev, vol65, pp117, 1944) and classical sampling methods will not work. The distribution (7.51) is called a binary Markov random field, or "Ising" model.

Exercise 7.2.4 *Construct a reversible Markov chain on* Ω *with equilibrium distribution* $\pi_x = \Pr(x)$.

Solution: We use the Metropolis Hastings prescription:

Let $X_n = x$. X_{n+1} is determined in the following way.

1. Generate a candidate state x' from x using some process we can choose. Here is a simple method: given $x = (x_1, x_2, ..., x_{N^2})$, pick a pixel n at random from $1, 2, ..., N^2$. Set $x' = (x_1, x_2, ..., -x_n, ..., x_{N^2})$. Notice that

   (a) we can get to any state in Ω by a sequence of such updates,

   (b) if x' and x differ by more than one pixel, $g(x'|x) = g(x|x') = 0$,

   (c) if x' and x differ by exactly one pixel, $g(x'|x) = g(x|x') = 1/N^2$,

   so our generation scheme is suitable for MH MCMC.

2. Work out the Metropolis-Hastings acceptance probability with $\pi_x = \Pr(x)$.

$$\begin{aligned}\alpha\left(x'|x\right) &= \min\left\{1, \frac{\pi_{x'}}{\pi_x}\frac{g(x|x')}{g(x'|x)}\right\} \\ &= \min\left\{1, \exp\left(-2J(\#x' - \#x)\right)\right\} \qquad (7.53)\end{aligned}$$

   Notice that both the g's and the normalization $\mathcal{Z}$ cancel. Since x and x' are the same except at x_n where $x'_n = -x_n$,, we write $\#x' - \#x = \#x'_n - \#x_n$ where $\#x_n$ is the number of disagreeing neighbours around x_n. So we set $X_{n+1} = x'$ with probability

$$\alpha\left(x'|x\right) = \min\left\{1, \exp\left(-2J\#\Delta_n\right)\right\} \qquad (7.54)$$

   (where $\#\Delta_n = \#x'_n - \#x_n$ is the change in the number of disagreeing neighbours at pixel n). Otherwise we set $X_{n+1} = x$, *ie* no change.

Notice that

1. If $\#\Delta_n < 0$, i.e., the change from x to x' gives a smoother image with fewer disagreeing neighbours, then $\alpha = 1$ and the proposed change is accepted with probability 1.
2. If $\#\Delta_n > 0$, the change leads to a more irregular image with fewer agreeing neighbours. Then $\alpha = \exp(-2J\#\Delta_n)$ and the proposed change is accepted with probability < 1.

Algorithm for generating a realization of the Markov chain

A Matlab script file for generating realizations from a random binary Markov field is

```
Xpix = 64;
Ypix = 64;
J = 1;
l = 0;
F = -ones(Ypix,Xpix);
while 1,
   for k = 1:4096
      % Select a pixel at random
      ix = ceil(Xpix*rand(1)); iy = ceil(Ypix*rand(1));
      Fc = F(iy,ix); pos = (ix-1)*Ypix + iy; % Index of pixel
      nbhrs = pos + [-1 1 -Ypix Ypix]; % Find indicies of neighbours
      nbhrs(find([iy==1 iy==Ypix ix==1 ix==Xpix])) = []; % Remove those outside picture
      nagree = sum(Fc==F(nbhrs)); ndisagree = sum(Fc~=F(nbhrs));
      change = nagree - ndisagree;
      if rand(1)<exp(-2*J*change) % if change<0, this happens with certainty
         F(iy,ix) = -Fc;
      end
      l = l + 1;
   end
   figure(1); image(63*F);colormap(gray); title(['Iteration ' num2str(l)]);
   drawnow
end
```

The only non-straightforward part is handling pixels on the edges which have fewer than four neighbours. If you run it you will see images sampled from $\Pr(x)$. Increasing the value J leads to images dominated by one color, as the average number of disagreeing edges decreases with increasing J. Some examples are given in Figure 7.3.

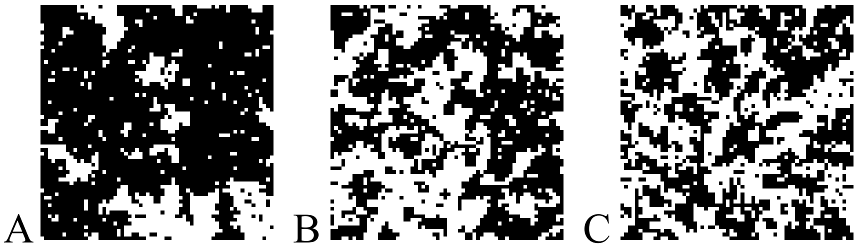

Figure 7.3 Samples $x \sim \exp(-2J\#x)/\mathcal{Z}$ with $N = 64$ and (A) $J = 0.45$ (B) $J = 0.4$ (C) $J = 0.35$.