1

Discrete-time Markov chains

This chapter is the foundation for all that follows. Discrete-time Markov chains are defined and their behaviour is investigated. For better orientation we now list the key theorems: these are Theorems 1.3.2 and 1.3.5 on hitting times, Theorem 1.4.2 on the strong Markov property, Theorem 1.5.3 characterizing recurrence and transience, Theorem 1.7.7 on invariant distributions and positive recurrence. Theorem 1.8.3 on convergence to equilibrium, Theorem 1.9.3 on reversibility, and Theorem 1.10.2 on long-run averages. Once you understand these you will understand the basic theory. Part of that understanding will come from familiarity with examples, so a large number are worked out in the text. Exercises at the end of each section are an important part of the exposition.

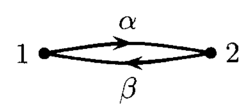
1.1 Definition and basic properties

Let I be a countable set. Each $i \in I$ is called a *state* and I is called the *state-space*. We say that $\lambda = (\lambda_i : i \in I)$ is a *measure* on I if $0 \le \lambda_i < \infty$ for all $i \in I$. If in addition the *total mass* $\sum_{i \in I} \lambda_i$ equals 1, then we call λ a *distribution*. We work throughout with a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

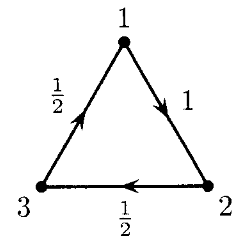
Then λ defines a distribution, the *distribution of* X. We think of X as modelling a random state which takes the value i with probability λ_i. There is a brief review of some basic facts about countable sets and probability spaces in Chapter 6.

We say that a matrix $P = (p_{ij} : i, j \in I)$ is *stochastic* if every row $(p_{ij} : j \in I)$ is a distribution. There is a one-to-one correspondence between stochastic matrices P and the sort of diagrams described in the Introduction. Here are two examples:

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$



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



We shall now formalize the rules for a Markov chain by a definition in terms of the corresponding matrix P. We say that $(X_n)_{n\geq 0}$ is a *Markov chain* with *initial distribution* λ and *transition matrix* P if

(i) X_0 has distribution λ;

(ii) for $n \geq 0$, conditional on $X_n = i$, X_{n+1} has distribution $(p_{ij} : j \in I)$ and is independent of $X_0, \dots, X_{n-1}$.

More explicitly, these conditions state that, for $n \geq 0$ and $i_0, \dots, i_{n+1} \in I$,

(i) $\mathbb{P}(X_0 = i_0) = \lambda_{i_0}$;

(ii) $\mathbb{P}(X_{n+1} = i_{n+1} \mid X_0 = i_0, \dots, X_n = i_n) = p_{i_n i_{n+1}}$.

We say that $(X_n)_{n\geq 0}$ is *Markov*(λ, P) for short. If $(X_n)_{0\leq n\leq N}$ is a finite sequence of random variables satisfying (i) and (ii) for $n = 0, \dots, N-1$, then we again say $(X_n)_{0\leq n\leq N}$ is *Markov*(λ, P).

It is in terms of properties (i) and (ii) that most real-world examples are seen to be Markov chains. But mathematically the following result appears to give a more comprehensive description, and it is the key to some later calculations.

Theorem 1.1.1. *A discrete-time random process* $(X_n)_{0\leq n\leq N}$ *is Markov*(λ, P) *if and only if for all* $i_0, \dots, i_N \in I$

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \dots, X_N = i_N) = \lambda_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{N-1} i_N}. \tag{1.1}$$

Proof. Suppose $(X_n)_{0\le n\le N}$ is Markov(λ, P), then

$$\begin{aligned}
&\mathbb{P}(X_0 = i_0, X_1 = i_1, \dots, X_N = i_N)\\
&\quad = \mathbb{P}(X_0 = i_0)\mathbb{P}(X_1 = i_1 \mid X_0 = i_0)\\
&\qquad \dots \mathbb{P}(X_N = i_N \mid X_0 = i_0, \dots, X_{N-1} = i_{N-1})\\
&\quad = \lambda_{i_0} p_{i_0 i_1} \dots p_{i_{N-1} i_N}.
\end{aligned}$$

On the other hand, if (1.1) holds for N, then by summing both sides over $i_N \in I$ and using $\sum_{j\in I} p_{ij} = 1$ we see that (1.1) holds for $N-1$ and, by induction

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \lambda_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i_n}$$

for all $n = 0, 1, \dots, N$. In particular, $\mathbb{P}(X_0 = i_0) = \lambda_{i_0}$ and, for $n = 0, 1, \dots, N-1$,

$$\begin{aligned}
&\mathbb{P}(X_{n+1} = i_{n+1} \mid X_0 = i_0, \dots, X_n = i_n)\\
&\quad = \mathbb{P}(X_0 = i_0, \dots, X_n = i_n, X_{n+1} = i_{n+1}) / \mathbb{P}(X_0 = i_0, \dots, X_n = i_n)\\
&\quad = p_{i_n i_{n+1}}.
\end{aligned}$$

So $(X_n)_{0\le n\le N}$ is Markov(λ, P). □

The next result reinforces the idea that Markov chains have no memory. We write $\delta_i = (\delta_{ij} : j \in I)$ for the *unit mass* at i, where

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 1.1.2 (Markov property). *Let $(X_n)_{n\ge 0}$ be Markov(λ, P). Then, conditional on $X_m = i$, $(X_{m+n})_{n\ge 0}$ is Markov(δ_i, P) and is independent of the random variables $X_0, \dots, X_m$.*

Proof. We have to show that for any event A determined by $X_0, \dots, X_m$ we have

$$\begin{aligned}
&\mathbb{P}(\{X_m = i_m, \dots, X_{m+n} = i_{m+n}\} \cap A \mid X_m = i)\\
&\quad = \delta_{i i_m} p_{i_m i_{m+1}} \dots p_{i_{m+n-1} i_{m+n}} \mathbb{P}(A \mid X_m = i)
\end{aligned} \tag{1.2}$$

In that case we have to show

$$\begin{aligned}&\mathbb{P}(X_0 = i_0, \dots , X_{m+n} = i_{m+n} \text{ and } i = i_m)/\mathbb{P}(X_m = i)\\ &= \delta_{ii_m} p_{i_m i_{m+1}} \cdots p_{i_{m+n-1} i_{m+n}}\\ &\quad \times \mathbb{P}(X_0 = i_0, \dots , X_m = i_m \text{ and } i = i_m)/\mathbb{P}(X_m = i)\end{aligned}$$

which is true by Theorem 1.1.1. In general, any event A determined by $X_0, \dots , X_m$ may be written as a countable disjoint union of elementary events

$$A = \bigcup_{k=1}^{\infty} A_k.$$

Then the desired identity (1.2) for A follows by summing up the corresponding identities for A_k. $\square$

The remainder of this section addresses the following problem: *what is the probability that after n steps our Markov chain is in a given state?* First we shall see how the problem reduces to calculating entries in the nth power of the transition matrix. Then we shall look at some examples where this may be done explicitly.

We regard distributions and measures λ as row vectors whose components are indexed by I, just as P is a matrix whose entries are indexed by $I \times I$. When I is finite we will often label the states $1, 2, \dots , N$; then λ will be an N-vector and P an $N \times N$-matrix. For these objects, matrix multiplication is a familiar operation. We extend matrix multiplication to the general case in the obvious way, defining a new measure λP and a new matrix P^2 by

$$(\lambda P)_j = \sum_{i \in I} \lambda_i p_{ij}, \quad (P^2)_{ik} = \sum_{j \in I} p_{ij} p_{jk}.$$

We define P^n similarly for any n. We agree that P^0 is the identity matrix I, where $(I)_{ij} = \delta_{ij}$. The context will make it clear when I refers to the state-space and when to the identity matrix. We write $p_{ij}^{(n)} = (P^n)_{ij}$ for the (i, j) entry in P^n.

In the case where $\lambda_i > 0$ we shall write $\mathbb{P}_i(A)$ for the conditional probability $\mathbb{P}(A \mid X_0 = i)$. By the Markov property at time $m = 0$, under $\mathbb{P}_i$, $(X_n)_{n \geq 0}$ is Markov(δ_i, P). So the behaviour of $(X_n)_{n \geq 0}$ under $\mathbb{P}_i$ does not depend on λ.

Theorem 1.1.3. *Let $(X_n)_{n \geq 0}$ be Markov(λ, P). Then, for all $n, m \geq 0$,*

(i) $\mathbb{P}(X_n = j) = (\lambda P^n)_j$;

(ii) $\mathbb{P}_i(X_n = j) = \mathbb{P}(X_{n+m} = j \mid X_m = i) = p_{ij}^{(n)}$.

Proof. (i) By Theorem 1.1.1

$$\begin{aligned}\mathbb{P}(X_n = j) &= \sum_{i_0 \in I} \cdots \sum_{i_{n-1} \in I} \mathbb{P}(X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = j) \\ &= \sum_{i_0 \in I} \cdots \sum_{i_{n-1} \in I} \lambda_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} j} = (\lambda P^n)_j.\end{aligned}$$

(ii) By the Markov property, conditional on $X_m = i$, $(X_{m+n})_{n \geq 0}$ is Markov (δ_i, P), so we just take $\lambda = \delta_i$ in (i). $\square$

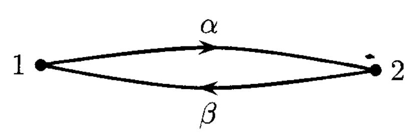
In light of this theorem we call $p_{ij}^{(n)}$ the *n-step transition probability from i to j*. The following examples give some methods for calculating $p_{ij}^{(n)}$.

Example 1.1.4

The most general two-state chain has transition matrix of the form

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

and is represented by the following diagram:



We exploit the relation $P^{n+1} = P^n P$ to write

$$p_{11}^{(n+1)} = p_{12}^{(n)} \beta + p_{11}^{(n)} (1 - \alpha).$$

We also know that $p_{11}^{(n)} + p_{12}^{(n)} = \mathbb{P}_1(X_n = 1 \text{ or } 2) = 1$, so by eliminating $p_{12}^{(n)}$ we get a recurrence relation for $p_{11}^{(n)}$:

$$p_{11}^{(n+1)} = (1 - \alpha - \beta) p_{11}^{(n)} + \beta, \quad p_{11}^{(0)} = 1.$$

This has a unique solution (see Section 1.11):

$$p_{11}^{(n)} = \begin{cases} \dfrac{\beta}{\alpha+\beta} + \dfrac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n & \text{for } \alpha + \beta > 0 \end{cases}$$

Example 1.1.5 (Virus mutation)

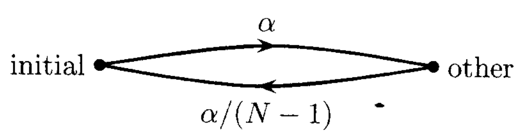
Suppose a virus can exist in N different strains and in each generation either stays the same, or with probability α mutates to another strain, which is chosen at random. What is the probability that the strain in the nth generation is the same as that in the 0th?

We could model this process as an N-state chain, with $N \times N$ transition matrix P given by

$$p_{ii} = 1 - \alpha, \quad p_{ij} = \alpha/(N-1) \quad \text{for } i \neq j.$$

Then the answer we want would be found by computing $p_{11}^{(n)}$. In fact, in this example there is a much simpler approach, which relies on exploiting the symmetry present in the mutation rules.

At any time a transition is made from the initial state to another with probability α, and a transition from another state to the initial state with probability $\alpha/(N-1)$. Thus we have a two-state chain with diagram



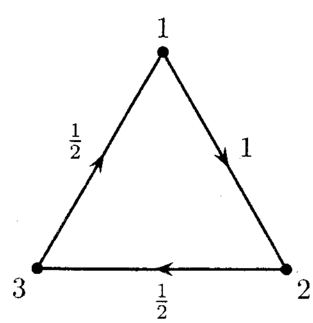
and by putting $\beta = \alpha/(N-1)$ in Example 1.1.4 we find that the desired probability is

$$\frac{1}{N} + \left(1 - \frac{1}{N}\right)\left(1 - \frac{\alpha N}{N-1}\right)^n.$$

Beware that in examples having less symmetry, this sort of lumping together of states may not produce a Markov chain.

Example 1.1.6

Consider the three-state chain with diagram



and transition matrix

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

The problem is to find a general formula for $p_{11}^{(n)}$.

First we compute the eigenvalues of P by writing down its characteristic equation

$$0 = \det(x - P) = x(x - \tfrac{1}{2})^2 - \tfrac{1}{4} = \tfrac{1}{4}(x-1)(4x^2+1).$$

The eigenvalues are $1, i/2, -i/2$ and from this we deduce that $p_{11}^{(n)}$ has the form

$$p_{11}^{(n)} = a + b\left(\frac{i}{2}\right)^n + c\left(-\frac{i}{2}\right)^n$$

for some constants a, b and c. (The justification comes from linear algebra: having distinct eigenvalues, P is diagonalizable, that is, for some invertible matrix U we have

$$P = U \begin{pmatrix} 1 & 0 & 0 \\ 0 & i/2 & 0 \\ 0 & 0 & -i/2 \end{pmatrix} U^{-1}$$

and hence

$$P^n = U \begin{pmatrix} 1 & 0 & 0 \\ 0 & (i/2)^n & 0 \\ 0 & 0 & (-i/2)^n \end{pmatrix} U^{-1}$$

which forces $p_{11}^{(n)}$ to have the form claimed.) The answer we want is real and

$$\left(\pm\frac{i}{2}\right)^n = \left(\frac{1}{2}\right)^n e^{\pm in\pi/2} = \left(\frac{1}{2}\right)^n \left(\cos\frac{n\pi}{2} \pm i\sin\frac{n\pi}{2}\right)$$

so it makes sense to rewrite $p_{11}^{(n)}$ in the form

$$p_{11}^{(n)} = \alpha + \left(\frac{1}{2}\right)^n \left\{\beta\cos\frac{n\pi}{2} + \gamma\sin\frac{n\pi}{2}\right\}$$

for constants α, β and γ. The first few values of $p_{11}^{(n)}$ are easy to write down, so we get equations to solve for α, β and γ:

so $\alpha = 1/5$, $\beta = 4/5$, $\gamma = -2/5$ and

$$p_{11}^{(n)} = \frac{1}{5} + \left(\frac{1}{2}\right)^n \left\{\frac{4}{5}\cos\frac{n\pi}{2} - \frac{2}{5}\sin\frac{n\pi}{2}\right\}.$$

More generally, the following method may in principle be used to find a formula for $p_{ij}^{(n)}$ for any M-state chain and any states i and j.

(i) Compute the eigenvalues $\lambda_1, \dots, \lambda_M$ of P by solving the characteristic equation.

(ii) If the eigenvalues are distinct then $p_{ij}^{(n)}$ has the form

$$p_{ij}^{(n)} = a_1\lambda_1^n + \dots + a_M\lambda_M^n$$

for some constants $a_1, \dots, a_M$ (depending on i and j). If an eigenvalue λ is repeated (once, say) then the general form includes the term $(an + b)\lambda^n$.

(iii) As roots of a polynomial with real coefficients, complex eigenvalues will come in conjugate pairs and these are best written using sine and cosine, as in the example.

Exercises

1.1.1 Let $B_1, B_2, \dots$ be disjoint events with $\bigcup_{n=1}^{\infty} B_n = \Omega$. Show that if A is another event and $\mathbb{P}(A|B_n) = p$ for all n then $\mathbb{P}(A) = p$.

Deduce that if X and Y are discrete random variables then the following are equivalent:

(a) X and Y are independent;

(b) the conditional distribution of X given $Y = y$ is independent of y.

1.1.2 Suppose that $(X_n)_{n\geq 0}$ is Markov (λ, P). If $Y_n = X_{kn}$, show that $(Y_n)_{n\geq 0}$ is Markov (λ, P^k).

1.1.3 Let X_0 be a random variable with values in a countable set I. Let $Y_1, Y_2, \dots$ be a sequence of independent random variables, uniformly distributed on $[0, 1]$. Suppose we are given a function

$$G : I \times [0, 1] \to I$$

and define inductively

$$X_{n+1} = G(X_n, Y_{n+1}).$$

Show that $(X_n)_{n\geq 0}$ is a Markov chain and express its transition matrix P in terms of G. Can all Markov chains be realized in this way? How would you simulate a Markov chain using a computer?