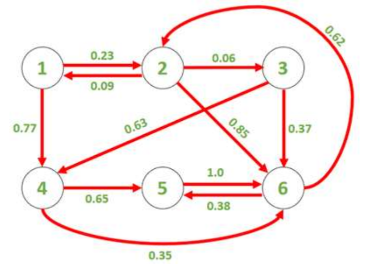
Chapter 5.
Computations with Transition Functions



Let X_n, $n \geq 0$, be a Markov chain on $\mathscr{S}$ having transition function P. In this section we will show how various conditional probabilities can be expressed in terms of P. We will also define the n-step transition function of the Markov chain.

We begin with the formula

$$\boxed{\begin{aligned} P(X_{n+1} = x_{n+1}, \ldots, X_{n+m} = x_{n+m} \mid X_0 = x_0, \ldots, X_n = x_n) \\ = P(x_n, x_{n+1}) \cdots P(x_{n+m-1}, x_{n+m}). \end{aligned}} \tag{16}$$

To prove (16) we write the left side of this equation as

$$= \frac{\pi_0(x_0)P(x_0, x_1) \cdots P(x_{n+m-1}, x_{n+m})}{\pi_0(x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n)},$$

which reduces to the right side of (16).

It is convenient to rewrite (16) as

$$(17)\quad P(X_{n+1} = y_1, \dots, X_{n+m} = y_m \mid X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x) = P(x, y_1)P(y_1, y_2) \cdots P(y_{m-1}, y_m).$$

Let $A_0, \dots, A_{n-1}$ be subsets of $\mathscr{S}$. It follows from (17) that

$$(18)\quad P(X_{n+1} = y_1, \dots, X_{n+m} = y_m \mid X_0 \in A_0, \dots, X_{n-1} \in A_{n-1}, X_n = x) = \boxed{\phantom{P(x, y_1)P(y_1, y_2) \cdots P(y_{m-1}, y_m)}}$$

Let $B_1, \dots, B_m$ be subsets of $\mathscr{S}$. It follows from (18) that

$$(19)\quad \boxed{P(X_{n+1} \in B_1, \dots, X_{n+m} \in B_m \mid X_0 \in A_0, \dots, X_{n-1} \in A_{n-1}, X_n = x) = \sum_{y_1 \in B_1} \cdots \sum_{y_m \in B_m} P(x, y_1)P(y_1, y_2) \cdots P(y_{m-1}, y_m).}$$

The *m-step transition function* $P^m(x, y)$, which gives the probability of going from x to y in m steps, is defined by

$$(20)\quad P^m(x, y) = \sum_{y_1} \cdots \sum_{y_{m-1}} P(x, y_1)P(y_1, y_2) \cdots P(y_{m-2}, y_{m-1})P(y_{m-1}, y)$$

for $m \geq 2$, by $P^1(x, y) = P(x, y)$, and by

$$P^0(x, y) = \begin{cases} 1, & x = y, \\ 0, & \text{elsewhere.} \end{cases}$$

We see by setting $B_1 = \cdots = B_{m-1} = \mathscr{S}$ and $B_m = \{y\}$ in (19) that

$$(21)\quad \boxed{\phantom{P(X_{n+m} = y \mid X_0 \in A_0, \dots, X_{n-1} \in A_{n-1}, X_n = x) = P^m(x, y)}}$$

In particular, by setting $A_0 = \cdots = A_{n-1} = \mathscr{S}$, we see that

$$(22)\quad P(X_{n+m} = y \mid X_n = x) = P^m(x, y).$$

It also follows from (21) that

(23) $$P(X_{n+m} = y \mid X_0 = x, X_n = z) = P^m(z, y).$$

Since

$$P^{n+m}(x, y) = P(X_{n+m} = y \mid X_0 = x)$$
$$= \sum_z P(X_n = z \mid X_0 = x) P(X_{n+m} = y \mid X_0 = x, X_n = z)$$
$$= \sum_z P^n(x, z) P(X_{n+m} = y \mid X_0 = x, X_n = z),$$

we conclude from (23) that

(24)

Chapman–Kolmogorov equation

For Markov chains having a finite number of states, (24) allows us to think of P^n as the nth power of the matrix P.

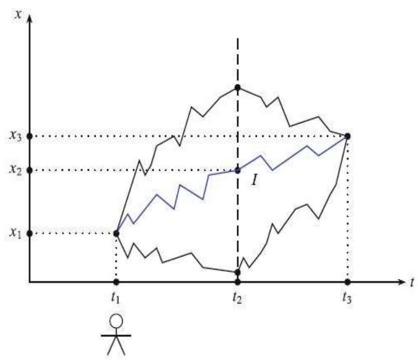
$$\mathbf{P}^{(n+m)} = \mathbf{P}^{(n)}\mathbf{P}^{(m)}$$

or equivalently:

$$p_{ij}^{(n+m)} = \sum_{k \in S} p_{ik}^{(n)} p_{kj}^{(m)}$$

Chapman-Kolmogorov equation

Proof :



Let π_0 be an initial distribution for the Markov chain. Since

$$P(X_n = y) = \sum_x P(X_0 = x, X_n = y)$$

$$= \sum_x P(X_0 = x)P(X_n = y \mid X_0 = x),$$

we see that

(25)

This formula allows us to compute the distribution of X_n in terms of the initial distribution π_0 and the n-step transition function P^n.

For an alternative method of computing the distribution of X_n, observe that

$$P(X_{n+1} = y) = \sum_x P(X_n = x, X_{n+1} = y)$$

$$= \sum_x P(X_n = x)P(X_{n+1} = y \mid X_n = x),$$

so that

(26) $$\boxed{P(X_{n+1} = y) = \sum_x P(X_n = x)P(x, y).}$$

If we know the distribution of X_0, we can use (26) to find the distribution of X_1. Then, knowing the distribution of X_1, we can use (26) to find the distribution of X_2. Similarly, we can find the distribution of X_n by applying (26) n times.

We will use the notation $\boxed{P_x(\ )}$ to denote probabilities of various events defined in terms of a Markov chain starting at x. Thus

$$P_x(X_1 \neq a, X_2 \neq a, X_3 = a)$$

denotes the probability that a Markov chain starting at x is in a state a at time 3 but not at time 1 or at time 2. In terms of this notation, (19) can be rewritten as

(27)

(19) $$P(X_{n+1} \in B_1, \ldots, X_{n+m} \in B_m \mid X_0 \in A_0, \ldots, X_{n-1} \in A_{n-1}, X_n = x) = \sum_{y_1 \in B_1} \cdots \sum_{y_m \in B_m} P(x, y_1)P(y_1, y_2) \cdots P(y_{m-1}, y_m).$$

Hitting times

Let A be a subset of $\mathscr{S}$. The *hitting time* T_A of A is defined by

$$T_A = \min\{n > 0 : X_n \in A\}$$

if $X_n \in A$ for some $n > 0$, and by $T_A = \infty$ if $X_n \notin A$ for all $n > 0$. In other words, T_A is the first positive time the Markov chain is in (*hits*) A.

Hitting times play an important role in the theory of Markov chains. In this book we will be interested mainly in hitting times of sets consisting of a single point. We denote the hitting time of a point $a \in \mathscr{S}$ by T_a rather than by the more cumbersome notation $T_{\{a\}}$.

An important equation involving hitting times is given by

(28)

In order to verify (28) we note that the events $\{T_y = m, X_n = y\}$, $1 \le m \le n$, are disjoint and that

$$\{X_n = y\} = \bigcup_{m=1}^{n} \{T_y = m, X_n = y\}.$$

We have in effect decomposed the event $\{X_n = y\}$ according to the hitting time of y. We see from this decomposition that

$$\begin{aligned} P^n(x, y) = P_x(X_n = y) &= \sum_{m=1}^{n} P_x(T_y = m, X_n = y) \\ &= \sum_{m=1}^{n} P_x(T_y = m)P(X_n = y \mid X_0 = x, T_y = m) \\ &= \sum_{m=1}^{n} P_x(T_y = m)P(X_n = y \mid X_0 = x, X_1 \ne y, \ldots, \\ &\qquad\qquad X_{m-1} \ne y, X_m = y) \\ &= \sum_{m=1}^{n} P_x(T_y = m)P^{n-m}(y, y), \end{aligned}$$

and hence that (28) holds.

Example. Show that if a is an absorbing state, then

$$P^n(x, a) = P_x(T_a \le n),\ n \ge 1.$$

If a is an absorbing state, then $P^{n-m}(a, a) = 1$ for $1 \le m \le n$, and hence (28) implies that

$$\begin{aligned} P^n(x, a) &= \sum_{m=1}^{n} P_x(T_a = m)P^{n-m}(a, a) \\ &= \sum_{m=1}^{n} P_x(T_a = m) = P_x(T_a \le n). \qquad \blacksquare \end{aligned}$$

Observe that

$$P_x(T_y = 1) = P_x(X_1 = y) = P(x, y)$$

and that

$$P_x(T_y = 2) = \sum_{z \ne y} P_x(X_1 = z, X_2 = y) = \sum_{z \ne y} P(x, z)P(z, y).$$

For higher values of n the probabilities $P_x(T_y = n)$ can be found by using the formula

(29)

This formula is a consequence of (27), but it should also be directly obvious. For in order to go from x to y for the first time at time $n + 1$, it is necessary to go to some state $z \neq y$ at the first step and then go from z to y for the first time at the end of n additional steps.

Transition matrix

Suppose now that the state space $\mathscr{S}$ is finite, say $\mathscr{S} = \{0, 1, \ldots, d\}$. In this case we can think of P as the *transition matrix* having $d + 1$ rows and columns given by

$$\begin{array}{c} \\ 0 \\ \vdots \\ d \end{array} \begin{array}{c} \begin{array}{ccc} 0 & \cdots & d \end{array} \\ \begin{bmatrix} P(0,0) & \cdots & P(0,d) \\ \vdots & & \vdots \\ P(d,0) & \cdots & P(d,d) \end{bmatrix} \end{array}.$$

For example, the transition matrix of the gambler's ruin chain on $\{0, 1, 2, 3\}$ is

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ q & 0 & p & 0 \\ 0 & q & 0 & p \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}.$$

Similarly, we can regard P^n as an *n-step transition matrix*. Formula (24) with $m = n = 1$ becomes

$$P^2(x, y) = \sum_z P(x, z)P(z, y).$$

Recalling the definition of ordinary matrix multiplication, we observe that the two-step transition matrix P^2 is the product of the matrix P with itself. More generally, by setting $m = 1$ in (24) we see that

(30)

It follows from (30) by induction that the n-step transition matrix P^n is the nth power of P.

An initial distribution π_0 can be thought of as a $(d + 1)$-dimensional row vector

$$\pi_0 = (\pi_0(0), \ldots, \pi_0(d)).$$

If we let π_n denote the $(d + 1)$-dimensional row vector

$$\pi_n = (P(X_n = 0), \ldots, P(X_n = d)),$$

then (25) and (26) can be written respectively as

$$\pi_n = \pi_0 P^n$$

and

$$\pi_{n+1} = \pi_n P.$$

Example. Consider the two-state Markov chain having one-step transition matrix

$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}, \quad \text{where } p + q > 0.$$

Find P^n.

In order to find $P^n(0, 0) = P_0(X_n = 0)$, we set $\pi_0(0) = 1$ in (3) and obtain

$P(X_n = 0 \mid X_0 = 0)$

$$P^n(0, 0) = \frac{q}{p+q} + (1-p-q)^n \frac{p}{p+q}.$$

In order to find $P^n(0, 1) = P_0(X_n = 1)$, we set $\pi_0(1) = 0$ in (4) and obtain

$P(X_n = 1 \mid X_0 = 0)$

$$P^n(0, 1) = \frac{p}{p+q} - (1-p-q)^n \frac{p}{p+q}.$$

Similarly, we conclude that

$P(X_n = 0 \mid X_0 = 1)$

$$P^n(1, 0) = \frac{q}{p+q} - (1-p-q)^n \frac{q}{p+q}$$

and

$P(X_n = 1 \mid X_0 = 1)$

$$P^n(1, 1) = \frac{p}{p+q} + (1-p-q)^n \frac{q}{p+q}.$$

It follows that