

## 16.3: Introduction to Discrete-Time Chains

In this and the next several sections, we consider a Markov process with the discrete time space  $\mathbb{N}$  and with a discrete (countable) state space. Recall that a Markov process with a discrete state space is called a *Markov chain*, so we are studying *discrete-time Markov chains*.

### Review

We will review the basic definitions and concepts in the general introduction. With both time and space discrete, many of these definitions and concepts simplify considerably. As usual, our starting point is a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , so  $\Omega$  is the sample space,  $\mathcal{F}$  the  $\sigma$ -algebra of events, and  $\mathbb{P}$  the probability measure on  $(\Omega, \mathcal{F})$ . Let  $\mathbf{X} = (X_0, X_1, X_2, \dots)$  be a stochastic process defined on the probability space, with time space  $\mathbb{N}$  and with countable state space  $S$ . In the context of the general introduction,  $S$  is given the power set  $\mathcal{P}(S)$  as the  $\sigma$ -algebra, so all subsets of  $S$  are measurable, as are all functions from  $S$  into another measurable space. Counting measure  $\#$  is the natural measure on  $S$ , so integrals over  $S$  are simply sums. The same comments apply to the time space  $\mathbb{N}$ : all subsets of  $\mathbb{N}$  are measurable and counting measure  $\#$  is the natural measure on  $\mathbb{N}$ .

The vector space  $\mathcal{B}$  consisting of bounded functions  $f : S \rightarrow \mathbb{R}$  will play an important role. The norm that we use is the *supremum norm* defined by

$$\|f\| = \sup\{|f(x)| : x \in S\}, \quad f \in \mathcal{B} \quad (16.3.1)$$

For  $n \in \mathbb{N}$ , let  $\mathcal{F}_n = \sigma\{X_0, X_1, \dots, X_n\}$ , the  $\sigma$ -algebra generated by the process up to time  $n$ . Thus  $\mathcal{F} = \{\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \dots\}$  is the natural filtration associated with  $\mathbf{X}$ . We also let  $\mathcal{G}_n = \sigma\{X_n, X_{n+1}, \dots\}$ , the  $\sigma$ -algebra generated by the process from time  $n$  on. So if  $n \in \mathbb{N}$  represents the present time, then  $\mathcal{F}_n$  contains the events in the past and  $\mathcal{G}_n$  the events in the future.

### Definitions

We start with the basic definition of the Markov property: the past and future are conditionally independent, given the present.

$\mathbf{X} = (X_0, X_1, X_2, \dots)$  is a Markov chain if  $\mathbb{P}(A \cap B \mid \mathcal{F}_n) = \mathbb{P}(A \mid \mathcal{F}_n)\mathbb{P}(B \mid \mathcal{G}_n)$  for every  $n \in \mathbb{N}$ ,  $A \in \mathcal{F}_n$  and  $B \in \mathcal{G}_n$ .

There are a number of equivalent formulations of the Markov property for a discrete-time Markov chain. We give a few of these.

$\mathbf{X} = (X_0, X_1, X_2, \dots)$  is a Markov chain if either of the following equivalent conditions is satisfied:

1.  $\mathbb{P}(X_{n+1} = x \mid \mathcal{F}_n) = \mathbb{P}(X_{n+1} = x \mid X_n)$  for every  $n \in \mathbb{N}$  and  $x \in S$ .
2.  $\mathbb{E}[f(X_{n+1}) \mid \mathcal{F}_n] = \mathbb{E}[f(X_{n+1}) \mid X_n]$  for every  $n \in \mathbb{N}$  and  $f \in \mathcal{B}$ .

Part (a) states that for  $n \in \mathbb{N}$ , the conditional probability density function of  $X_{n+1}$  given  $\mathcal{F}_n$  is the same as the conditional probability density function of  $X_{n+1}$  given  $X_n$ . Part (b) also states, in terms of expected value, that the conditional distribution of  $X_{n+1}$  given  $\mathcal{F}_n$  is the same as the conditional distribution of  $X_{n+1}$  given  $X_n$ . Both parts are the Markov property looking just one time step in the future. But with discrete time, this is equivalent to the Markov property at general future times.

$\mathbf{X} = (X_0, X_1, X_2, \dots)$  is a Markov chain if either of the following equivalent conditions is satisfied:

1.  $\mathbb{P}(X_{n+k} = x \mid \mathcal{F}_n) = \mathbb{P}(X_{n+k} = x \mid X_n)$  for every  $n, k \in \mathbb{N}$  and  $x \in S$ .
2.  $\mathbb{E}[f(X_{n+k}) \mid \mathcal{F}_n] = \mathbb{E}[f(X_{n+k}) \mid X_n]$  for every  $n, k \in \mathbb{N}$  and  $f \in \mathcal{B}$ .

Part (a) states that for  $n, k \in \mathbb{N}$ , the conditional probability density function of  $X_{n+k}$  given  $\mathcal{F}_n$  is the same as the conditional probability density function of  $X_{n+k}$  given  $X_n$ . Part (b) also states, in terms of expected value, that the conditional distribution of  $X_{n+k}$  given  $\mathcal{F}_n$  is the same as the conditional distribution of  $X_{n+k}$  given  $X_n$ . In discrete time and space, the Markov property can also be stated without explicit reference to  $\sigma$ -algebras. If you are not familiar with measure theory, you can take this as the starting definition.

$\mathbf{X} = (X_0, X_1, X_2, \dots)$  is a Markov chain if for every  $n \in \mathbb{N}$  and every sequence of states  $(x_0, x_1, \dots, x_{n-1}, x, y)$

$$\mathbb{P}(X_{n+1} = y \mid X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}, X_n = x) = \mathbb{P}(X_{n+1} = y \mid X_n = x) \quad (16.3.2)$$

The theory of discrete-time Markov chains is simplified considerably if we add an additional assumption.

A Markov chain  $\mathbf{X} = (X_0, X_1, X_2, \dots)$  is *time homogeneous* if

$$\mathbb{P}(X_{n+k} = y \mid X_k = x) = \mathbb{P}(X_n = y \mid X_0 = x) \quad (16.3.3)$$

for every  $k, n \in \mathbb{N}$  and every  $x, y \in S$ .

That is, the conditional distribution of  $X_{n+k}$  given  $X_k = x$  depends only on  $n$ . So if  $\mathbf{X}$  is homogeneous (we usually don't bother with the *time* adjective), then the chain  $\{X_{k+n} : n \in \mathbb{N}\}$  given  $X_k = x$  is equivalent (in distribution) to the chain  $\{X_n : n \in \mathbb{N}\}$  given  $X_0 = x$ . For this reason, the initial distribution is often unspecified in the study of Markov chains—if the chain is in state  $x \in S$  at a particular time  $k \in \mathbb{N}$ , then it doesn't really matter how the chain got to state  $x$ ; the process essentially “starts over”, independently of the past. The term *stationary* is sometimes used instead of homogeneous.

From now on, we will usually assume that our Markov chains are homogeneous. This is not as big of a loss of generality as you might think. A non-homogeneous Markov chain can be turned into a homogeneous Markov process by enlarging the state space, as shown in the introduction to general Markov processes, but at the cost of creating an uncountable state space. For a homogeneous Markov chain, if  $k, n \in \mathbb{N}$ ,  $x \in S$ , and  $f \in \mathcal{B}$ , then

$$\mathbb{E}[f(X_{k+n}) \mid X_k = x] = \mathbb{E}[f(X_n) \mid X_0 = x] \quad (16.3.4)$$

### Stopping Times and the Strong Markov Property

Consider again a stochastic process  $\mathbf{X} = (X_0, X_1, X_2, \dots)$  with countable state space  $S$ , and with the natural filtration  $\mathfrak{F} = (\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \dots)$  as given above. Recall that a random variable  $\tau$  taking values in  $\mathbb{N} \cup \{\infty\}$  is a *stopping time* or a *Markov time* for  $\mathbf{X}$  if  $\{\tau = n\} \in \mathcal{F}_n$  for each  $n \in \mathbb{N}$ . Intuitively, we can tell whether or not  $\tau = n$  by observing the chain up to time  $n$ . In a sense, a stopping time is a random time that does not require that we see into the future. The following result gives the quintessential examples of stopping times.

Suppose again  $\mathbf{X} = \{X_n : n \in \mathbb{N}\}$  is a discrete-time Markov chain with state space  $S$  as defined above. For  $A \subseteq S$ , the following random times are stopping times:

1.  $\rho_A = \inf\{n \in \mathbb{N} : X_n \in A\}$ , the *entrance time* to  $A$ .
2.  $\tau_A = \inf\{n \in \mathbb{N}_+ : X_n \in A\}$ , the *hitting time* to  $A$ .

Proof

For  $n \in \mathbb{N}$

1.  $\{\rho_A = n\} = \{X_0 \notin A, X_1 \notin A, \dots, X_{n-1} \notin A, X_n \in A\} \in \mathcal{F}_n$
2.  $\{\tau_A = n\} = \{X_1 \notin A, X_2 \notin A, \dots, X_{n-1} \notin A, X_n \in A\} \in \mathcal{F}_n$

An example of a random time that is generally *not* a stopping time is the *last* time that the process is in  $A$ :

$$\zeta_A = \max\{n \in \mathbb{N}_+ : X_n \in A\} \quad (16.3.5)$$

We cannot tell if  $\zeta_A = n$  without looking into the future:  $\{\zeta_A = n\} = \{X_n \in A, X_{n+1} \notin A, X_{n+2} \notin A, \dots\}$  for  $n \in \mathbb{N}$ .

If  $\tau$  is a stopping time for  $\mathbf{X}$ , the  $\sigma$ -algebra associated with  $\tau$  is

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau = n\} \in \mathcal{F}_n \text{ for all } n \in \mathbb{N}\} \quad (16.3.6)$$

Intuitively,  $\mathcal{F}_\tau$  contains the events that can be described by the process up to the random time  $\tau$ , in the same way that  $\mathcal{F}_n$  contains the events that can be described by the process up to the deterministic time  $n \in \mathbb{N}$ . For more information see the section on filtrations and stopping times.

The *strong Markov property* states that the future is independent of the past, given the present, when the present time is a stopping time. For a discrete-time Markov chain, the ordinary Markov property implies the strong Markov property.

If  $\mathbf{X} = (X_0, X_1, X_2, \dots)$  is a discrete-time Markov chain then  $\mathbf{X}$  has the strong Markov property. That is, if  $\tau$  is a finite stopping time for  $\mathbf{X}$  then

1.  $\mathbb{P}(X_{\tau+k} = x \mid \mathcal{F}_\tau) = \mathbb{P}(X_{\tau+k} = x \mid X_\tau)$  for every  $k \in \mathbb{N}$  and  $x \in S$ .
2.  $\mathbb{E}[f(X_{\tau+k}) \mid \mathcal{F}_\tau] = \mathbb{E}[f(X_{\tau+k}) \mid X_\tau]$  for every  $k \in \mathbb{N}$  and  $f \in \mathcal{B}$ .

Part (a) states that the conditional probability density function of  $X_{\tau+k}$  given  $\mathcal{F}_\tau$  is the same as the conditional probability density function of  $X_{\tau+k}$  given just  $X_\tau$ . Part (b) also states, in terms of expected value, that the conditional distribution of  $X_{\tau+k}$  given  $\mathcal{F}_\tau$  is the same as the conditional distribution of  $X_{\tau+k}$  given just  $X_\tau$ . Assuming homogeneity as usual, the Markov chain  $\{X_{\tau+n} : n \in \mathbb{N}\}$  given  $X_\tau = x$  is equivalent in distribution to the chain  $\{X_n : n \in \mathbb{N}\}$  given  $X_0 = x$ .

### Transition Matrices

Suppose again that  $\mathbf{X} = (X_0, X_1, X_2, \dots)$  is a homogeneous, discrete-time Markov chain with state space  $S$ . With a discrete state space, the transition kernels studied in the general introduction become transition *matrices*, with rows and columns indexed by  $S$  (and so perhaps of infinite size). The kernel operations become familiar matrix operations. The results in this section are special cases of the general results, but we sometimes give independent proofs for completeness, and because the proofs are simpler. You may want to review the section on kernels in the chapter on expected value.

For  $n \in \mathbb{N}$  let

$$P_n(x, y) = \mathbb{P}(X_n = y \mid X_0 = x), \quad (x, y) \in S \times S \quad (16.3.7)$$

The matrix  $P_n$  is the  $n$ -step *transition probability matrix* for  $\mathbf{X}$ .

Thus,  $y \mapsto P_n(x, y)$  is the probability density function of  $X_n$  given  $X_0 = x$ . In particular,  $P_n$  is a *probability matrix* (or *stochastic matrix*) since  $P_n(x, y) \geq 0$  for  $(x, y) \in S^2$  and  $\sum_{y \in S} P(x, y) = 1$  for  $x \in S$ . As with any nonnegative matrix on  $S$ ,  $P_n$  defines a kernel on  $S$  for  $n \in \mathbb{N}$ :

$$P_n(x, A) = \sum_{y \in A} P_n(x, y) = \mathbb{P}(X_n \in A \mid X_0 = x), \quad x \in S, A \subseteq S \quad (16.3.8)$$

So  $A \mapsto P_n(x, A)$  is the probability distribution of  $X_n$  given  $X_0 = x$ . The next result is the *Chapman-Kolmogorov equation*, named for Sydney Chapman and Andrei Kolmogorov. It gives the basic relationship between the transition matrices.

If  $m, n \in \mathbb{N}$  then  $P_m P_n = P_{m+n}$

Proof

This follows from the Markov and time-homogeneous properties and a basic conditioning argument. If  $x, z \in S$  then

$$P_{m+n}(x, z) = \mathbb{P}(X_{m+n} = z \mid X_0 = x) = \sum_{y \in S} \mathbb{P}(X_{m+n} = z \mid X_0 = x, X_m = y) \mathbb{P}(X_m = y \mid X_0 = x) \quad (16.3.9)$$

But by the Markov property and time-homogeneous properties

$$\mathbb{P}(X_{m+n} = z \mid X_0 = x, X_m = y) = \mathbb{P}(X_n = z \mid X_0 = y) = P_n(y, z) \quad (16.3.10)$$

Of course also  $\mathbb{P}(X_m = y \mid X_0 = x) = P_m(x, y)$ . Hence we have

$$P_{m+n}(x, z) = \sum_{y \in S} P_m(x, y) P_n(y, z) \quad (16.3.11)$$

The right side, by definition, is  $P_m P_n(x, z)$ .

It follows immediately that the transition matrices are just the matrix powers of the one-step transition matrix. That is, letting  $P = P_1$  we have  $P_n = P^n$  for all  $n \in \mathbb{N}$ . Note that  $P^0 = I$ , the identity matrix on  $S$  given by  $I(x, y) = 1$  if  $x = y$  and 0 otherwise. The right operator corresponding to  $P^n$  yields an expected value.

Suppose that  $n \in \mathbb{N}$  and that  $f : S \rightarrow \mathbb{R}$ . Then, assuming that the expected value exists,

$$P^n f(x) = \sum_{y \in S} P^n(x, y) f(y) = \mathbb{E}[f(X_n) \mid X_0 = x], \quad x \in S \quad (16.3.12)$$

Proof

This follows easily from the definitions:

$$P^n f(x) = \sum_{y \in S} P^n(x, y) f(y) = \sum_{y \in S} \mathbb{P}(X_n = y \mid X_0 = x) f(y) = \mathbb{E}[f(X_n) \mid X_0 = x], \quad x \in S \quad (16.3.13)$$

The existence of the expected value is only an issue if  $S$  is infinite. In particular, the result holds if  $f$  is nonnegative or if  $f \in \mathcal{B}$  (which in turn would always be the case if  $S$  is finite). In fact,  $P^n$  is a linear *contraction operator* on the space  $\mathcal{B}$  for  $n \in \mathbb{N}$ . That is, if  $f \in \mathcal{B}$  then  $P^n f \in \mathcal{B}$  and  $\|P^n f\| \leq \|f\|$ . The left operator corresponding to  $P^n$  is defined similarly. For  $f : S \rightarrow \mathbb{R}$

$$f P^n(y) = \sum_{x \in S} f(x) P^n(x, y), \quad y \in S \quad (16.3.14)$$

assuming again that the sum makes sense (as before, only an issue when  $S$  is infinite). The left operator is often restricted to nonnegative functions, and we often think of such a function as the density function (with respect to  $\#$ ) of a positive measure on  $S$ . In this sense, the left operator maps a density function to another density function.

A function  $f : S \rightarrow \mathbb{R}$  is *invariant* for  $P$  (or for the chain  $\mathbf{X}$ ) if  $fP = f$ .

Clearly if  $f$  is invariant, so that  $fP = f$  then  $fP^n = f$  for all  $n \in \mathbb{N}$ . If  $f$  is a probability density function, then so is  $fP$ .

If  $X_0$  has probability density function  $f$ , then  $X_n$  has probability density function  $fP^n$  for  $n \in \mathbb{N}$ .

Proof

Again, this follows easily from the definitions and a conditioning argument.

$$\mathbb{P}(X_n = y) = \sum_{x \in S} \mathbb{P}(X_0 = x) \mathbb{P}(X_n = y \mid X_0 = x) = \sum_{x \in S} f(x) P^n(x, y) = f P^n(y), \quad y \in S \quad (16.3.15)$$

In particular, if  $X_0$  has probability density function  $f$ , and  $f$  is invariant for  $\mathbf{X}$ , then  $X_n$  has probability density function  $f$  for all  $n \in \mathbb{N}$ , so the sequence of variables  $\mathbf{X} = (X_0, X_1, X_2, \dots)$  is identically distributed. Combining two results above, suppose that  $X_0$  has probability density function  $f$  and that  $g : S \rightarrow \mathbb{R}$ . Assuming the expected value exists,  $\mathbb{E}[g(X_n)] = f P^n g$ . Explicitly,

$$\mathbb{E}[g(X_n)] = \sum_{x \in S} \sum_{y \in S} f(x) P^n(x, y) g(y) \quad (16.3.16)$$

It also follows from the last theorem that the distribution of  $X_0$  (the *initial distribution*) and the one-step transition matrix determine the distribution of  $X_n$  for each  $n \in \mathbb{N}$ . Actually, these basic quantities determine the *finite dimensional distributions* of the process, a stronger result.

Suppose that  $X_0$  has probability density function  $f_0$ . For any sequence of states  $(x_0, x_1, \dots, x_n) \in S^n$ ,

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = f_0(x_0) P(x_0, x_1) P(x_1, x_2) \cdots P(x_{n-1}, x_n) \quad (16.3.17)$$

Proof

This follows directly from the Markov property and the multiplication rule of conditional probability:

$$\begin{aligned} \mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) &= \mathbb{P}(X_0 = x_0) \mathbb{P}(X_1 = x_1 \mid X_0 = x_0) \mathbb{P}(X_2 = x_2 \mid X_0 = x_0, X_1 = x_1) \\ &\quad \cdots \mathbb{P}(X_n = x_n \mid X_0 = x_0, \dots, X_{n-1} = x_{n-1}) \end{aligned} \quad (16.3.18)$$

But by the Markov property, this reduces to

$$\begin{aligned}\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) &= \mathbb{P}(X_0 = x_0) \mathbb{P}(X_1 = x_1 | X_0 = x_0) \mathbb{P}(X_2 = x_2 | X_1 = x_1) \cdots \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}) \\ &= f_0(x_0) P(x_0, x_1) P(x_1, x_2) \cdots P(x_{n-1}, x_n)\end{aligned}$$

Computations of this sort are the reason for the term *chain* in the name *Markov chain*. From this result, it follows that given a probability matrix  $P$  on  $S$  and a probability density function  $f$  on  $S$ , we can construct a Markov chain  $\mathbf{X} = (X_0, X_1, X_2, \dots)$  such that  $X_0$  has probability density function  $f$  and the chain has one-step transition matrix  $P$ . In applied problems, we often know the one-step transition matrix  $P$  from modeling considerations, and again, the initial distribution is often unspecified.

There is a natural graph (in the combinatorial sense) associated with a homogeneous, discrete-time Markov chain.

Suppose again that  $\mathbf{X} = (X_0, X_1, X_2, \dots)$  is a Markov chain with state space  $S$  and transition probability matrix  $P$ . The *state graph* of  $\mathbf{X}$  is the directed graph with vertex set  $S$  and edge set  $E = \{(x, y) \in S^2 : P(x, y) > 0\}$ .

That is, there is a directed edge from  $x$  to  $y$  if and only if state  $x$  leads to state  $y$  in one step. Note that the graph may well have loops, since a state can certainly lead back to itself in one step. More generally, we have the following result:

Suppose again that  $\mathbf{X} = (X_0, X_1, X_2, \dots)$  is a Markov chain with state space  $S$  and transition probability matrix  $P$ . For  $x, y \in S$  and  $n \in \mathbb{N}_+$ , there is a directed path of length  $n$  in the state graph from  $x$  to  $y$  if and only if  $P^n(x, y) > 0$ .

*Proof*

This follows since  $P^n(x, y) > 0$  if and only if there exists a sequence of states  $(x_1, x_2, \dots, x_{n-1})$  with  $P(x, x_1) > 0, P(x_1, x_2) > 0, \dots, P(x_{n-1}, y) > 0$ . This is also precisely the condition for the existence of a directed path  $(x, x_1, \dots, x_{n-1}, y)$  of length  $n$  from  $x$  to  $y$  in the state graph.

## Potential Matrices

For  $\alpha \in (0, 1]$ , the  $\alpha$ -potential matrix  $R_\alpha$  of  $\mathbf{X}$  is

$$R_\alpha = \sum_{n=0}^{\infty} \alpha^n P^n, \quad (x, y) \in S^2 \quad (16.3.19)$$

1.  $R = R_1$  is simply the *potential matrix* of  $\mathbf{X}$ .
2.  $R(x, y)$  is the expected number of visits by  $\mathbf{X}$  to  $y \in S$ , starting at  $x \in S$ .

*Proof*

First the definition of  $R_\alpha$  as an infinite series of matrices makes sense since  $P^n$  is a nonnegative matrix for each  $n$ . The interpretation of  $R(x, y)$  for  $(x, y) \in S^2$  comes from interchanging sum and expected value, again justified since the terms are nonnegative.

$$R(x, y) = \sum_{n=0}^{\infty} P^n(x, y) = \sum_{n=0}^{\infty} \mathbb{E}[\mathbf{1}(X_n = y) | X_0 = x] = \mathbb{E}\left(\sum_{n=0}^{\infty} \mathbf{1}(X_n = y) \mid X_0 = x\right) = \mathbb{E}[\#\{n \in \mathbb{N} : X_n = y\} | X_0 = x] \quad (16.3.20)$$

Note that it's quite possible that  $R(x, y) = \infty$  for some  $(x, y) \in S^2$ . In fact, knowing when this is the case is of considerable importance in recurrence and transience, which we study in the next section. As with any nonnegative matrix, the  $\alpha$ -potential matrix defines a kernel and defines left and right operators. For the kernel,

$$R_\alpha(x, A) = \sum_{y \in A} R_\alpha(x, y) = \sum_{n=0}^{\infty} \alpha^n P^n(x, A), \quad x \in S, A \subseteq S \quad (16.3.21)$$

In particular,  $R(x, A)$  is the expected number of visits by the chain to  $A$  starting in  $x$ :

$$R(x, A) = \sum_{y \in A} R(x, y) = \sum_{n=0}^{\infty} P^n(x, A) = \mathbb{E}\left[\sum_{n=0}^{\infty} \mathbf{1}(X_n \in A)\right], \quad x \in S, A \subseteq S \quad (16.3.22)$$

If  $\alpha \in (0, 1)$ , then  $R_\alpha(x, S) = \frac{1}{1-\alpha}$  for all  $x \in S$ .

*Proof*

Using geometric series,

$$R_\alpha(x, S) = \sum_{n=0}^{\infty} \alpha^n P^n(x, S) = \sum_{n=0}^{\infty} \alpha^n = \frac{1}{1-\alpha} \quad (16.3.23)$$

Hence  $R_\alpha$  is a bounded matrix for  $\alpha \in (0, 1)$  and  $(1-\alpha)R_\alpha$  is a probability matrix. There is a simple interpretation of this matrix.

If  $\alpha \in (0, 1)$  then  $(1-\alpha)R_\alpha(x, y) = \mathbb{P}(X_N = y | X_0 = x)$  for  $(x, y) \in S^2$ , where  $N$  is independent of  $\mathbf{X}$  and has the geometric distribution on  $\mathbb{N}$  with parameter  $1-\alpha$ .

Proof

Let  $(x, y) \in S^2$ . Conditioning on  $N$  gives

$$\mathbb{P}(X_N = y \mid X_0 = x) = \sum_{n=0}^{\infty} \mathbb{P}(N = n) \mathbb{P}(X_N = y \mid X_0 = x, N = n) \quad (16.3.24)$$

But by the substitution rule and the assumption of independence,

$$\mathbb{P}(X_N = y \mid N = n, X_0 = x) = \mathbb{E}(X_n = y \mid N = n, X_0 = x) = \mathbb{P}(X_n = y \mid X_0 = x) = P^n(x, y) \quad (16.3.25)$$

Since  $N$  has the geometric distribution on  $N$  with parameter  $1 - \alpha$  we have  $\mathbb{P}(N = n) = (1 - \alpha)\alpha^n$ . Hence

$$\mathbb{P}(X_N = y \mid X_0 = x) = \sum_{n=0}^{\infty} (1 - \alpha)\alpha^n P^n(x, y) = (1 - \alpha)R_\alpha(x, y) \quad (16.3.26)$$

So  $(1 - \alpha)R_\alpha$  can be thought of as a *transition matrix* just as  $P^n$  is a transition matrix, but corresponding to the *random time*  $N$  (with  $\alpha$  as a parameter) rather than the *deterministic time*  $n$ . An interpretation of the potential matrix  $R_\alpha$  for  $\alpha \in (0, 1)$  can also be given in economic terms. Suppose that we receive one monetary unit each time the chain visits a fixed state  $y \in S$ . Then  $R(x, y)$  is the expected total reward, starting in state  $x \in S$ . However, typically money that we will receive at times distant in the future have less value to us *now* than money that we will receive soon. Specifically suppose that a monetary unit at time  $n \in \mathbb{N}$  has a *present value* of  $\alpha^n$ , so that  $\alpha$  is an *inflation factor* (sometimes also called a *discount factor*). Then  $R_\alpha(x, y)$  gives the expected total discounted reward, starting at  $x \in S$ .

The potential kernels  $\mathbf{R} = \{R_\alpha : \alpha \in (0, 1)\}$  completely determine the transition kernels  $\mathbf{P} = \{P_n : n \in \mathbb{N}\}$ .

Proof

Note that for  $(x, y) \in S^2$ , the function  $\alpha \mapsto R_\alpha(x, y)$  is a power series in  $\alpha$  with coefficients  $n \mapsto P^n(x, y)$ . In the language of combinatorics,  $\alpha \mapsto R_\alpha(x, y)$  is the *ordinary generating function* of the sequence  $n \mapsto P^n(x, y)$ . As noted above, this power series has radius of convergence at least 1, so we can extend the domain to  $\alpha \in (-1, 1)$ . Thus, given the potential matrices, we can recover the transition matrices by taking derivatives and evaluating at 0:

$$P^n(x, y) = \frac{1}{n!} \left[ \frac{d^n}{d\alpha^n} R_\alpha(x, y) \right]_{\alpha=0} \quad (16.3.27)$$

Of course, it's really only necessary to determine  $P$ , the one step transition kernel, since the other transition kernels are powers of  $P$ . In any event, it follows that the matrices  $\mathbf{R} = \{R_\alpha : \alpha \in (0, 1)\}$ , along with the initial distribution, completely determine the finite dimensional distributions of the Markov chain  $\mathbf{X}$ . The potential matrices commute with each other and with the transition matrices.

If  $\alpha, \beta \in (0, 1]$  and  $k \in \mathbb{N}$ , then

1.  $P^k R_\alpha = R_\alpha P^k = \sum_{n=0}^{\infty} \alpha^n P^{n+k}$
2.  $R_\alpha R_\beta = R_\beta R_\alpha = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \alpha^m \beta^n P^{m+n}$

Proof

Distributing matrix products through matrix sums is allowed since the matrices are nonnegative.

1. Directly

$$R_\alpha P^k = \sum_{n=0}^{\infty} \alpha^n P^n P^k = \sum_{n=0}^{\infty} \alpha^n P^{n+k} \quad (16.3.28)$$

The other direction requires an interchange.

$$P^k R_\alpha = P^k \sum_{n=0}^{\infty} \alpha^n P^n = \sum_{n=0}^{\infty} \alpha^n P^k P^n = \sum_{n=0}^{\infty} \alpha^n P^{n+k} \quad (16.3.29)$$

2. First,

$$R_\alpha R_\beta = \sum_{m=0}^{\infty} \alpha^m P^m R_\beta = \sum_{m=0}^{\infty} \alpha^m P^m \left( \sum_{n=0}^{\infty} \beta^n P^n \right) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \alpha^m \beta^n P^m P^n = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \alpha^m \beta^n P^{m+n} \quad (16.3.30)$$

The other direction is similar.

The fundamental equation that relates the potential matrices is given next.

If  $\alpha, \beta \in (0, 1]$  with  $\alpha \geq \beta$  then

$$\alpha R_\alpha = \beta R_\beta + (\alpha - \beta) R_\alpha R_\beta \quad (16.3.31)$$

Proof

If  $\alpha = \beta$  the equation is trivial, so assume  $\alpha > \beta$ . From the previous result,

$$R_\alpha R_\beta = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \alpha^j \beta^k P^{j+k} \quad (16.3.32)$$

Changing variables to sum over  $n = j + k$  and  $k$  gives

$$R_\alpha R_\beta = \sum_{n=0}^{\infty} \sum_{k=0}^n \alpha^{n-k} \beta^k P^n = \sum_{n=0}^{\infty} \sum_{k=0}^n \left(\frac{\beta}{\alpha}\right)^k \alpha^n P^n = \sum_{n=0}^{\infty} \frac{1 - \left(\frac{\beta}{\alpha}\right)^{n+1}}{1 - \frac{\beta}{\alpha}} \alpha^n P^n \quad (16.3.33)$$

Simplifying gives

$$R_\alpha R_\beta = \frac{1}{\alpha - \beta} [\alpha R_\alpha - \beta R_\beta] \quad (16.3.34)$$

Note that since  $\beta < 1$ , the matrix  $R_\beta$  has finite values, so we don't have to worry about the dreaded indeterminate form  $\infty - \infty$ .

If  $\alpha \in (0, 1]$  then  $I + \alpha R_\alpha P = I + \alpha P R_\alpha = R_\alpha$ .

Proof

From the result above,

$$I + \alpha R_\alpha P = I + \alpha P R_\alpha = I + \sum_{n=0}^{\infty} \alpha^{n+1} P^{n+1} = \sum_{n=0}^{\infty} \alpha^n P^n = R_\alpha \quad (16.3.35)$$

This leads to an important result: when  $\alpha \in (0, 1)$ , there is an inverse relationship between  $P$  and  $R_\alpha$ .

If  $\alpha \in (0, 1)$ , then

1.  $R_\alpha = (I - \alpha P)^{-1}$
2.  $P = \frac{1}{\alpha} (I - R_\alpha^{-1})$

Proof

The matrices have finite values, so we can subtract. The identity  $I + \alpha R_\alpha P = R_\alpha$  leads to  $R_\alpha (I - \alpha P) = I$  and the identity  $I + \alpha P R_\alpha = R_\alpha$  leads to  $(I - \alpha P) R_\alpha = I$ . Hence (a) holds. Part (b) follows from (a).

This result shows again that the potential matrix  $R_\alpha$  determines the transition operator  $P$ .

### Sampling in Time

If we sample a Markov chain at multiples of a fixed time  $k$ , we get another (homogeneous) chain.

Suppose that  $\mathbf{X} = (X_0, X_1, X_2, \dots)$  is a Markov chain with state space  $S$  and transition probability matrix  $P$ . For fixed  $k \in \mathbb{N}_+$ , the sequence  $\mathbf{X}_k = (X_0, X_k, X_{2k}, \dots)$  is a Markov chain on  $S$  with transition probability matrix  $P^k$ .

If we sample a Markov chain at a general increasing sequence of time points  $0 < n_1 < n_2 < \dots$  in  $\mathbb{N}$ , then the resulting stochastic process  $\mathbf{Y} = (Y_0, Y_1, Y_2, \dots)$ , where  $Y_k = X_{n_k}$  for  $k \in \mathbb{N}$ , is still a Markov chain, but is not time homogeneous in general.

Recall that if  $A$  is a nonempty subset of  $S$ , then  $P_A$  is the matrix  $P$  restricted to  $A \times A$ . So  $P_A$  is a sub-stochastic matrix, since the row sums may be less than 1. Recall also that  $P_A^n$  means  $(P_A)^n$ , not  $(P^n)_A$ ; in general these matrices are different.

If  $A$  is a nonempty subset of  $S$  then for  $n \in \mathbb{N}$ ,

$$P_A^n(x, y) = \mathbb{P}(X_1 \in A, X_2 \in A, \dots, X_{n-1} \in A, X_n = y \mid X_0 = x), \quad (x, y) \in A \times A \quad (16.3.36)$$

That is,  $P_A^n(x, y)$  is the probability of going from state  $x$  to  $y$  in  $n$  steps, remaining in  $A$  all the while. In terms of the state graph of  $\mathbf{X}$ , it is the sum of products of probabilities along paths of length  $n$  from  $x$  to  $y$  that stay inside  $A$ .

### Examples and Applications

#### Computational Exercises

Let  $\mathbf{X} = (X_0, X_1, \dots)$  be the Markov chain on  $S = \{a, b, c\}$  with transition matrix

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} \\ 1 & 0 & 0 \end{bmatrix} \quad (16.3.37)$$

For the Markov chain  $\mathbf{X}$ ,

1. Draw the state graph.
2. Find  $\mathbb{P}(X_1 = a, X_2 = b, X_3 = c \mid X_0 = a)$

- Find  $P^2$
- Suppose that  $g: S \rightarrow \mathbb{R}$  is given by  $g(a) = 1$ ,  $g(b) = 2$ ,  $g(c) = 3$ . Find  $\mathbb{E}[g(X_2) | X_0 = x]$  for  $x \in S$ .
- Suppose that  $X_0$  has the uniform distribution on  $S$ . Find the probability density function of  $X_2$ .

Answer

- The edge set is  $E = \{(a, a), (a, b), (b, a), (b, c), (c, a)\}$
- $P(a, a)P(a, b)P(b, c) = \frac{3}{16}$
- By standard matrix multiplication,

$$P^2 = \begin{bmatrix} \frac{3}{8} & \frac{1}{4} & \frac{3}{8} \\ \frac{7}{8} & \frac{1}{8} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \quad (16.3.38)$$

- In matrix form,

$$g = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad P^2 g = \begin{bmatrix} 2 \\ \frac{9}{8} \\ \frac{3}{2} \end{bmatrix} \quad (16.3.39)$$

- In matrix form,  $X_0$  has PDF  $f = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$ , and  $X_2$  has PDF  $fP^2 = \begin{bmatrix} \frac{7}{12} & \frac{7}{24} & \frac{1}{8} \end{bmatrix}$ .

Let  $A = \{a, b\}$ . Find each of the following:

- $P_A$
- $P_A^2$
- $(P^2)_A$

Proof

$$\begin{aligned} 1. P_A &= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & 0 \\ \frac{3}{8} & \frac{1}{4} \end{bmatrix} \\ 2. P_A^2 &= \begin{bmatrix} \frac{1}{8} & \frac{1}{8} \\ \frac{1}{8} & \frac{1}{8} \end{bmatrix} \\ 3. (P^2)_A &= \begin{bmatrix} \frac{3}{8} & \frac{1}{4} \\ \frac{7}{8} & \frac{1}{8} \end{bmatrix} \end{aligned}$$

Find the invariant probability density function of  $\mathbf{X}$

Answer

Solving  $fP = f$  subject to the condition that  $f$  is a PDF gives  $f = \begin{bmatrix} \frac{8}{15} & \frac{4}{15} & \frac{3}{15} \end{bmatrix}$

Compute the  $\alpha$ -potential matrix  $R_\alpha$  for  $\alpha \in (0, 1)$ .

Answer

Computing  $R_\alpha = (I - \alpha P)^{-1}$  gives

$$R_\alpha = \frac{1}{(1 - \alpha)(8 + 4\alpha + 3\alpha^2)} \begin{bmatrix} 8 & 4\alpha & 3\alpha^2 \\ 2\alpha + 6\alpha^2 & 8 - 4\alpha & 6\alpha - 3\alpha^2 \\ 8\alpha & 4\alpha^2 & 8 - 4\alpha - \alpha^2 \end{bmatrix} \quad (16.3.40)$$

As a check on our work, note that the row sums are  $\frac{1}{1 - \alpha}$ .

### The Two-State Chain

Perhaps the simplest, non-trivial Markov chain has two states, say  $S = \{0, 1\}$  and the transition probability matrix given below, where  $p \in (0, 1)$  and  $q \in (0, 1)$  are parameters.

$$P = \begin{bmatrix} 1 - p & p \\ q & 1 - q \end{bmatrix} \quad (16.3.41)$$

For  $n \in \mathbb{N}$ ,

$$P^n = \frac{1}{p + q} \begin{bmatrix} q + p(1 - p - q)^n & p - p(1 - p - q)^n \\ q - q(1 - p - q)^n & p + q(1 - p - q)^n \end{bmatrix} \quad (16.3.42)$$

Proof

The eigenvalues of  $P$  are 1 and  $1 - p - q$ . Next,  $B^{-1}PB = D$  where

$$B = \begin{bmatrix} 1 & -p \\ 1 & q \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 \\ 0 & 1 - p - q \end{bmatrix} \quad (16.3.43)$$

Hence  $P^n = BD^nB^{-1}$ , which gives the expression above.

As  $n \rightarrow \infty$ ,

$$P^n \rightarrow \frac{1}{p+q} \begin{bmatrix} q & p \\ q & p \end{bmatrix} \quad (16.3.44)$$

Proof

Note that  $0 < p + q < 2$  and so  $-1 < 1 - (p + q) < 1$ . Hence  $(1 - p - q)^n \rightarrow 0$  as  $n \rightarrow \infty$ .

Open the simulation of the two-state, discrete-time Markov chain. For various values of  $p$  and  $q$ , and different initial states, run the simulation 1000 times. Compare the relative frequency distribution to the limiting distribution, and in particular, note the rate of convergence. Be sure to try the case  $p = q = 0.01$

The only invariant probability density function for the chain is

$$f = \left[ \frac{q}{p+q} \quad \frac{p}{p+q} \right] \quad (16.3.45)$$

Proof

Let  $f = [a \quad b]$ . The matrix equation  $fP = f$  leads to  $-pa + qb = 0$  so  $b = a\frac{p}{q}$ . The condition  $a + b = 1$  for  $f$  to be a PDF then gives  $a = \frac{q}{p+q}$ ,  $b = \frac{p}{p+q}$

For  $\alpha \in (0, 1)$ , the  $\alpha$ -potential matrix is

$$R_\alpha = \frac{1}{(p+q)(1-\alpha)} \begin{bmatrix} q & p \\ q & p \end{bmatrix} + \frac{1}{(p+q)^2(1-\alpha)} \begin{bmatrix} p & -p \\ -q & q \end{bmatrix} \quad (16.3.46)$$

Proof

In this case,  $R_\alpha$  can be computed directly as  $\sum_{n=0}^{\infty} \alpha^n P^n$  using geometric series.

In spite of its simplicity, the two state chain illustrates some of the basic limiting behavior and the connection with invariant distributions that we will study in general in a later section.

### Independent Variables and Random Walks

Suppose that  $\mathbf{X} = (X_0, X_1, X_2, \dots)$  is a sequence of independent random variables taking values in a countable set  $S$ , and that  $(X_1, X_2, \dots)$  are identically distributed with (discrete) probability density function  $f$ .

$\mathbf{X}$  is a Markov chain on  $S$  with transition probability matrix  $P$  given by  $P(x, y) = f(y)$  for  $(x, y) \in S \times S$ . Also,  $f$  is invariant for  $P$ .

Proof

As usual, let  $\mathcal{F}_n = \sigma\{X_0, X_1, \dots, X_n\}$  for  $n \in \mathbb{N}$ . Since the sequence  $\mathbf{X}$  is independent,

$$\mathbb{P}(X_{n+1} = y \mid \mathcal{F}_n) = \mathbb{P}(X_{n+1} = y) = f(y), \quad y \in S \quad (16.3.47)$$

Also,

$$fP(y) = \sum_{x \in S} f(x)P(x, y) = \sum_{x \in S} f(x)f(y) = f(y), \quad y \in S \quad (16.3.48)$$

As a Markov chain, the process  $\mathbf{X}$  is not very interesting, although of course it is very interesting in other ways. Suppose now that  $S = \mathbb{Z}$ , the set of integers, and consider the *partial sum process* (or *random walk*)  $\mathbf{Y}$  associated with  $\mathbf{X}$ :

$$Y_n = \sum_{i=0}^n X_i, \quad n \in \mathbb{N} \quad (16.3.49)$$

$\mathbf{Y}$  is a Markov chain on  $\mathbb{Z}$  with transition probability matrix  $Q$  given by  $Q(x, y) = f(y - x)$  for  $(x, y) \in \mathbb{Z} \times \mathbb{Z}$ .

Proof

Again, let  $\mathcal{F}_n = \sigma\{X_0, X_1, \dots, X_n\}$  for  $n \in \mathbb{N}$ . Then also,  $\mathcal{F}_n = \sigma\{Y_0, Y_1, \dots, Y_n\}$  for  $n \in \mathbb{N}$ . Hence

$$\mathbb{P}(Y_{n+1} = y \mid \mathcal{F}_n) = \mathbb{P}(Y_n + X_{n+1} = y \mid \mathcal{F}_n) = \mathbb{P}(Y_n + X_{n+1} = y \mid Y_n), \quad y \in \mathbb{Z} \quad (16.3.50)$$

since the sequence  $\mathbf{X}$  is independent. In particular,

$$\mathbb{P}(Y_{n+1} = y \mid Y_n = x) = \mathbb{P}(x + X_{n+1} = y \mid Y_n = x) = \mathbb{P}(X_{n+1} = y - x) = f(y - x), \quad (x, y) \in \mathbb{Z}^2 \quad (16.3.51)$$

Thus the probability density function  $f$  governs the distribution of a step size of the random walker on  $\mathbb{Z}$ .

Consider the special case of the random walk on  $\mathbb{Z}$  with  $f(1) = p$  and  $f(-1) = 1 - p$ , where  $p \in (0, 1)$ .

1. Give the transition matrix  $Q$  explicitly.
2. Give  $Q^n$  explicitly for  $n \in \mathbb{N}$ .

Answer

1.  $Q(x, x - 1) = 1 - p$ ,  $Q(x, x + 1) = p$  for  $x \in \mathbb{Z}$ .
2. For  $k \in \{0, 1, \dots, n\}$

$$Q^n(x, x + 2k - n) = \binom{n}{k} p^k (1 - p)^{n-k} \quad (16.3.52)$$

This corresponds to  $k$  steps to the right and  $n - k$  steps to the left.

This special case is the *simple random walk* on  $\mathbb{Z}$ . When  $p = \frac{1}{2}$  we have the *simple, symmetric random walk*. The simple random walk on  $\mathbb{Z}$  is studied in more detail in the section on random walks on graphs. The simple symmetric random walk is studied in more detail in the chapter on Bernoulli Trials.

### Doubly Stochastic Matrices

A matrix  $P$  on  $S$  is *doubly stochastic* if it is nonnegative and if the row and columns sums are 1:

$$\sum_{u \in S} P(x, u) = 1, \quad \sum_{u \in S} P(u, y) = 1, \quad (x, y) \in S \times S \quad (16.3.53)$$

Suppose that  $\mathbf{X}$  is a Markov chain on a finite state space  $S$  with doubly stochastic transition matrix  $P$ . Then the uniform distribution on  $S$  is invariant.

Proof

Constant functions are left invariant. Suppose that  $f(x) = c$  for  $x \in S$ . Then

$$fP(y) = \sum_{x \in S} f(x)P(x, y) = c \sum_{x \in S} P(x, y) = c, \quad y \in S \quad (16.3.54)$$

Hence if  $S$  is finite, the uniform PDF  $f$  given by  $f(x) = 1/\#(S)$  for  $x \in S$  is invariant.

If  $P$  and  $Q$  are doubly stochastic matrices on  $S$ , then so is  $PQ$ .

Proof

For  $y \in S$ ,

$$\sum_{x \in S} PQ(x, y) = \sum_{x \in S} \sum_{z \in S} P(x, z)Q(z, y) = \sum_{z \in S} Q(z, y) \sum_{x \in S} P(x, z) = \sum_{z \in S} Q(z, y) = 1 \quad (16.3.55)$$

The interchange of sums is valid since the terms are nonnegative.

It follows that if  $P$  is doubly stochastic then so is  $P^n$  for  $n \in \mathbb{N}$ .

Suppose that  $\mathbf{X} = (X_0, X_1, \dots)$  is the Markov chain with state space  $S = \{-1, 0, 1\}$  and with transition matrix

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} \quad (16.3.56)$$

1. Draw the state graph.
2. Show that  $P$  is doubly stochastic
3. Find  $P^2$ .
4. Show that the uniform distribution on  $S$  is the only invariant distribution for  $\mathbf{X}$ .
5. Suppose that  $X_0$  has the uniform distribution on  $S$ . For  $n \in \mathbb{N}$ , find  $\mathbb{E}(X_n)$  and  $\text{var}(X_n)$ .
6. Find the  $\alpha$ -potential matrix  $R_\alpha$  for  $\alpha \in (0, 1)$ .

Proof

1. The edge set is  $E = \{(-1, -1), (-1, 0), (0, 0), (0, 1), (1, -1), (1, 1)\}$
2. Just note that the row sums and the column sums are 1.
3. By matrix multiplication,

$$P^2 = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} \quad (16.3.57)$$

4. Let  $f = [p \quad q \quad r]$ . Solving the equation  $fP = f$  gives  $p = q = r$ . The requirement that  $f$  be a PDF then forces the common value to be  $\frac{1}{3}$ .
5. If  $X_0$  has the uniform distribution on  $S$ , then so does  $X_n$  for every  $n \in \mathbb{N}$ , so  $\mathbb{E}(X_n) = 0$  and  $\text{var}(X_n) = \mathbb{E}(X_0^2) = \frac{2}{3}$ .

6.

$$R_\alpha = (I - \alpha P)^{-1} = \frac{1}{(1 - \alpha)(4 - 2\alpha + \alpha^2)} \begin{bmatrix} 4 - 4\alpha + \alpha^2 & 2\alpha - \alpha^2 & \alpha^2 \\ \alpha^2 & 4 - 4\alpha + \alpha^2 & 2\alpha - \alpha^2 \\ 2\alpha - \alpha^2 & \alpha^2 & 4 - 4\alpha + \alpha^2 \end{bmatrix} \quad (16.3.58)$$

Recall that a matrix  $M$  indexed by a countable set  $S$  is *symmetric* if  $M(x, y) = M(y, x)$  for all  $x, y \in S$ .

If  $P$  is a symmetric, stochastic matrix then  $P$  is doubly stochastic.

Proof

This is trivial since

$$\sum_{x \in S} P(x, y) = \sum_{x \in S} P(y, x) = 1, \quad y \in S \quad (16.3.59)$$

The converse is not true. The doubly stochastic matrix in the [exercise above](#) is not symmetric. But since a symmetric, stochastic matrix on a finite state space is doubly stochastic, the uniform distribution is invariant.

Suppose that  $\mathbf{X} = (X_0, X_1, \dots)$  is the Markov chain with state space  $S = \{-1, 0, 1\}$  and with transition matrix

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{3}{4} \\ 0 & \frac{3}{4} & \frac{1}{4} \end{bmatrix} \quad (16.3.60)$$

1. Draw the state graph.
2. Show that  $P$  is symmetric
3. Find  $P^2$ .
4. Find all invariant probability density functions for  $\mathbf{X}$ .
5. Find the  $\alpha$ -potential matrix  $R_\alpha$  for  $\alpha \in (0, 1)$ .

Proof

1. The edge set is  $E = \{(-1, -1), (0, 0), (0, 1), (1, 0), (1, 1)\}$
2. Just note that  $P$  is symmetric with respect to the main diagonal.
3. By matrix multiplication,

$$P^2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{5}{8} & \frac{3}{8} \\ 0 & \frac{3}{8} & \frac{5}{8} \end{bmatrix} \quad (16.3.61)$$

4. Let  $f = [p \quad q \quad r]$ . Solving the equation  $fP = f$  gives simply  $r = q$ . The requirement that  $f$  be a PDF forces  $p = 1 - 2q$ . Thus the invariant PDFs are  $f = [1 - 2q \quad q \quad q]$  where  $q \in [0, \frac{1}{2}]$ . The special case  $q = \frac{1}{3}$  gives the uniform distribution on  $S$ .

5.

$$R_\alpha = (I - \alpha P)^{-1} = \frac{1}{2(1 - \alpha)^2(2 + \alpha)} \begin{bmatrix} 4 - 2\alpha - 2\alpha^2 & 0 & 0 \\ 0 & 4 - 5\alpha + \alpha^2 & 3\alpha - 3\alpha^2 \\ 0 & 3\alpha - 3\alpha^2 & 4 - 5\alpha + \alpha^2 \end{bmatrix} \quad (16.3.62)$$

### Special Models

The Markov chains in the following exercises model interesting processes that are studied in separate sections.

Read the introduction to the Ehrenfest chains.

Read the introduction to the Bernoulli-Laplace chain.

Read the introduction to the reliability chains.

Read the introduction to the branching chain.

Read the introduction to the queuing chains.

Read the introduction to random walks on graphs.

Read the introduction to birth-death chains.

---

This page titled [16.3: Introduction to Discrete-Time Chains](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist](#) ([Random Services](#)) via [source content](#) that was edited to the style and standards of the LibreTexts platform.