

## Markov Chains with a Finite Number of States

### 5.1 Stochastic Matrices

The theory of Markov chains makes use of stochastic matrices. We therefore begin with a small digression of an algebraic nature.

**Definition 5.1.** An  $r \times r$  matrix  $Q = (q_{ij})$  is said to be stochastic if

1.  $q_{ij} \geq 0$ .
2.  $\sum_{j=1}^r q_{ij} = 1$  for any  $1 \leq i \leq r$ .

A column vector  $f = (f_1, \dots, f_r)$  is said to be non-negative if  $f_i \geq 0$  for  $1 \leq i \leq r$ . In this case we write  $f \geq 0$ .

**Lemma 5.2.** The following statements are equivalent.

- (a) The matrix  $Q$  is stochastic.
- (b1) For any  $f \geq 0$  we have  $Qf \geq 0$ , and
- (b2) If  $\mathbf{1} = (1, \dots, 1)$  is a column vector, then  $Q\mathbf{1} = \mathbf{1}$ , that is the vector  $\mathbf{1}$  is an eigenvector of the matrix  $Q$  corresponding to the eigenvalue 1.
- (c) If  $\mu = (\mu_1, \dots, \mu_r)$  is a probability distribution, that is  $\mu_i \geq 0$  and  $\sum_{i=1}^r \mu_i = 1$ , then  $\mu Q$  is also a probability distribution.

*Proof.* If  $Q$  is a stochastic matrix, then (b1) and (b2) hold, and therefore (a) implies (b). We now show that (b) implies (a). Consider the column vector  $\delta_j$  all of whose entries are equal to zero, except the  $j$ -th entry which is equal to one. Then  $(Q\delta_j)_i = q_{ij} \geq 0$ . Furthermore,  $(Q\mathbf{1})_i = \sum_{j=1}^r q_{ij}$ , and it follows from the equality  $Q\mathbf{1} = \mathbf{1}$  that  $\sum_{j=1}^r q_{ij} = 1$  for all  $i$ , and therefore (b) implies (a).

We now show that (a) implies (c). If  $\mu' = \mu Q$ , then  $\mu'_j = \sum_{i=1}^r \mu_i q_{ij}$ . Since  $Q$  is stochastic, we have  $\mu'_j \geq 0$  and

$$\sum_{j=1}^r \mu'_j = \sum_{j=1}^r \sum_{i=1}^r \mu_i q_{ij} = \sum_{i=1}^r \sum_{j=1}^r \mu_i q_{ij} = \sum_{i=1}^r \mu_i = 1.$$

Therefore,  $\mu'$  is also a probability distribution.

Now assume that (c) holds. Consider the row vector  $\delta_i$  all of whose entries are equal to zero, except the  $i$ -th entry which is equal to one. It corresponds to the probability distribution on the set  $\{1, \dots, r\}$  which is concentrated at the point  $i$ . Then  $\delta_i Q$  is also a probability distribution. It follows that  $q_{ij} \geq 0$  and  $\sum_{j=1}^r q_{ij} = 1$ , that is (c) implies (a).  $\square$

**Lemma 5.3.** *Let  $Q' = (q'_{ij})$  and  $Q'' = (q''_{ij})$  be stochastic matrices and  $Q = Q'Q'' = (q_{ij})$ . Then  $Q$  is also a stochastic matrix. If  $q''_{ij} > 0$  for all  $i, j$ , then  $q_{ij} > 0$  for all  $i, j$ .*

*Proof.* We have

$$q_{ij} = \sum_{k=1}^r q'_{ik} q''_{kj}.$$

Therefore,  $q_{ij} \geq 0$ . If all  $q''_{kj} > 0$ , then  $q_{ij} > 0$  since  $q'_{ik} \geq 0$  and  $\sum_{k=1}^r q'_{ik} = 1$ . Furthermore,

$$\sum_{j=1}^r q_{ij} = \sum_{j=1}^r \sum_{k=1}^r q'_{ik} q''_{kj} = \sum_{k=1}^r q'_{ik} \sum_{j=1}^r q''_{kj} = \sum_{k=1}^r q'_{ik} = 1.$$

$\square$

*Remark 5.4.* We can also consider infinite matrices  $Q = (q_{ij})$ ,  $1 \leq i, j < \infty$ . An infinite matrix is said to be stochastic if

1.  $q_{ij} \geq 0$ , and
2.  $\sum_{j=1}^{\infty} q_{ij} = 1$  for any  $1 \leq i < \infty$ .

It is not difficult to show that Lemmas 5.2 and 5.3 remain valid for infinite matrices.

## 5.2 Markov Chains

We now return to the concepts of probability theory. Let  $\Omega$  be the space of sequences  $(\omega_0, \dots, \omega_n)$ , where  $\omega_k \in X = \{x^1, \dots, x^r\}$ ,  $0 \leq k \leq n$ . Without loss of generality we may identify  $X$  with the set of the first  $r$  integers,  $X = \{1, \dots, r\}$ .

Let  $P$  be a probability measure on  $\Omega$ . Sometimes we shall denote by  $\omega_k$  the random variable which assigns the value of the  $k$ -th element to the sequence  $\omega = (\omega_0, \dots, \omega_n)$ . It is usually clear from the context whether  $\omega_k$  stands for such a random variable or simply the  $k$ -th element of a particular sequence. We shall denote the probability of the sequence  $(\omega_0, \dots, \omega_n)$  by  $p(\omega_0, \dots, \omega_n)$ . Thus,

$$p(i_0, \dots, i_n) = P(\omega_0 = i_0, \dots, \omega_n = i_n).$$

Assume that we are given a probability distribution  $\mu = (\mu_1, \dots, \mu_r)$  on  $X$  and  $n$  stochastic matrices  $P(1), \dots, P(n)$  with  $P(k) = (p_{ij}(k))$ .

**Definition 5.5.** *The Markov chain with the state space  $X$  generated by the initial distribution  $\mu$  on  $X$  and the stochastic matrices  $P(1), \dots, P(n)$  is the probability measure  $P$  on  $\Omega$  such that*

$$P(\omega_0 = i_0, \dots, \omega_n = i_n) = \mu_{i_0} \cdot p_{i_0 i_1}(1) \cdot \dots \cdot p_{i_{n-1} i_n}(n) \quad (5.1)$$

for each  $i_0, \dots, i_n \in X$ .

The elements of  $X$  are called the states of the Markov chain. Let us check that (5.1) defines a probability measure on  $\Omega$ . The inequality  $P(\omega_0 = i_0, \dots, \omega_n = i_n) \geq 0$  is clear. It remains to show that

$$\sum_{i_0=1}^r \dots \sum_{i_n=1}^r P(\omega_0 = i_0, \dots, \omega_n = i_n) = 1.$$

We have

$$\begin{aligned} & \sum_{i_0=1}^r \dots \sum_{i_n=1}^r P(\omega_0 = i_0, \dots, \omega_n = i_n) \\ &= \sum_{i_0=1}^r \dots \sum_{i_n=1}^r \mu_{i_0} \cdot p_{i_0, i_1}(1) \cdot \dots \cdot p_{i_{n-1} i_n}(n). \end{aligned}$$

We now perform the summation over all the values of  $i_n$ . Note that  $i_n$  is only present in the last factor in each term of the sum, and the sum  $\sum_{i_n=1}^r p_{i_{n-1} i_n}(n)$  is equal to one, since the matrix  $P(n)$  is stochastic. We then fix  $i_0, \dots, i_{n-2}$ , and sum over all the values of  $i_{n-1}$ , and so on. In the end we obtain  $\sum_{i_0=1}^r \mu_{i_0}$ , which is equal to one, since  $\mu$  is a probability distribution.

In the same way one can prove the following statement:

$$P(\omega_0 = i_0, \dots, \omega_k = i_k) = \mu_{i_0} \cdot p_{i_0 i_1}(1) \cdot \dots \cdot p_{i_{k-1} i_k}(k)$$

for any  $1 \leq i_0, \dots, i_k \leq r$ ,  $k \leq n$ . This equality shows that the induced probability distribution on the space of sequences of the form  $(\omega_0, \dots, \omega_k)$  is also a Markov chain generated by the initial distribution  $\mu$  and the stochastic matrices  $P(1), \dots, P(k)$ .

The matrices  $P(k)$  are called the transition probability matrices, and the matrix entry  $p_{ij}(k)$  is called the transition probability from the state  $i$  to the state  $j$  at time  $k$ . The use of these terms is justified by the following calculation.

Assume that  $P(\omega_0 = i_0, \dots, \omega_{k-2} = i_{k-2}, \omega_{k-1} = i) > 0$ . We consider the conditional probability  $P(\omega_k = j | \omega_0 = i_0, \dots, \omega_{k-2} = i_{k-2}, \omega_{k-1} = i)$ . By the definition of the measure  $P$ ,

$$\begin{aligned}
& P(\omega_k = j | \omega_0 = i_0, \dots, \omega_{k-2} = i_{k-2}, \omega_{k-1} = i) \\
&= \frac{P(\omega_0 = i_0, \dots, \omega_{k-2} = i_{k-2}, \omega_{k-1} = i, \omega_k = j)}{P(\omega_0 = i_0, \dots, \omega_{k-2} = i_{k-2}, \omega_{k-1} = i)} \\
&= \frac{\mu_{i_0} \cdot p_{i_0 i_1}(1) \cdot \dots \cdot p_{i_{k-2} i}(k-1) \cdot p_{ij}(k)}{\mu_{i_0} \cdot p_{i_0 i_1}(1) \cdot \dots \cdot p_{i_{k-2} i}(k-1)} = p_{ij}(k).
\end{aligned}$$

The right-hand side here does not depend on  $i_0, \dots, i_{k-2}$ . This property is sometimes used as a definition of a Markov chain. It is also easy to see that  $P(\omega_k = j | \omega_{k-1} = i) = p_{ij}(k)$ . (This is proved below for the case of a homogeneous Markov chain.)

**Definition 5.6.** *A Markov chain is said to be homogeneous if  $P(k) = P$  for a matrix  $P$  which does not depend on  $k$ ,  $1 \leq k \leq n$ .*

The notion of a homogeneous Markov chain can be understood as a generalization of the notion of a sequence of independent identical trials. Indeed, if all the rows of the stochastic matrix  $P = (p_{ij})$  are equal to  $(p_1, \dots, p_r)$ , where  $(p_1, \dots, p_r)$  is a probability distribution on  $X$ , then the Markov chain with such a matrix  $P$  and the initial distribution  $(p_1, \dots, p_r)$  is a sequence of independent identical trials.

In what follows we consider only homogeneous Markov chains. Such chains can be represented with the help of graphs. The vertices of the graph are the elements of  $X$ . The vertices  $i$  and  $j$  are connected by an oriented edge if  $p_{ij} > 0$ . A sequence of states  $(i_0, i_1, \dots, i_n)$  which has a positive probability can be represented as a path of length  $n$  on the graph starting at the point  $i_0$ , then going to the point  $i_1$ , and so on. Therefore, a homogeneous Markov chain can be represented as a probability distribution on the space of paths of length  $n$  on the graph.

Let us consider the conditional probabilities  $P(\omega_{s+l} = j | \omega_l = i)$ . It is assumed here that  $P(\omega_l = i) > 0$ . We claim that

$$P(\omega_{s+l} = j | \omega_l = i) = p_{ij}^{(s)},$$

where  $p_{ij}^{(s)}$  are elements of the matrix  $P^s$ . Indeed,

$$\begin{aligned}
& P(\omega_{s+l} = j | \omega_l = i) = \frac{P(\omega_{s+l} = j, \omega_l = i)}{P(\omega_l = i)} \\
&= \frac{\sum_{i_0=1}^r \dots \sum_{i_{l-1}=1}^r \sum_{i_{l+1}=1}^r \dots \sum_{i_{s+l-1}=1}^r P(\omega_0 = i_0, \dots, \omega_l = i, \dots, \omega_{s+l} = j)}{\sum_{i_0=1}^r \dots \sum_{i_{l-1}=1}^r P(\omega_0 = i_0, \dots, \omega_l = i)} \\
&= \frac{\sum_{i_0=1}^r \dots \sum_{i_{l-1}=1}^r \sum_{i_{l+1}=1}^r \dots \sum_{i_{s+l-1}=1}^r \mu_{i_0} p_{i_0 i_1} \dots p_{i_{l-1} i} p_{i i_{l+1}} \dots p_{i_{s+l-1} j}}{\sum_{i_0=1}^r \dots \sum_{i_{l-1}=1}^r \mu_{i_0} p_{i_0 i_1} \dots p_{i_{l-1} i}}
\end{aligned}$$

$$\begin{aligned}
 &= \frac{\sum_{i_0=1}^r \cdots \sum_{i_{l-1}=1}^r \mu_{i_0} p_{i_0 i_1} \cdots p_{i_{l-1} i} \sum_{i_{l+1}=1}^r \cdots \sum_{i_{s+l-1}=1}^r p_{i i_{l+1}} \cdots p_{i_{s+l-1} j}}{\sum_{i_0=1}^r \cdots \sum_{i_{l-1}=1}^r \mu_{i_0} p_{i_0 i_1} \cdots p_{i_{l-1} i}} \\
 &= \sum_{i_{l+1}=1}^r \cdots \sum_{i_{s+l-1}=1}^r p_{i i_{l+1}} \cdots p_{i_{s+l-1} j} = p_{ij}^{(s)}.
 \end{aligned}$$

Thus the conditional probabilities  $p_{ij}^{(s)} = P(\omega_{s+l} = j | \omega_l = i)$  do not depend on  $l$ . They are called  $s$ -step transition probabilities. A similar calculation shows that for a homogeneous Markov chain with initial distribution  $\mu$ ,

$$P(\omega_s = j) = (\mu P^s)_j = \sum_{i=1}^r \mu_i p_{ij}^{(s)}. \tag{5.2}$$

Note that by considering infinite stochastic matrices, Definition 5.5 and the argument leading to (5.2) can be generalized to the case of Markov chains with a countable number of states.

### 5.3 Ergodic and Non-ergodic Markov Chains

**Definition 5.7.** A stochastic matrix  $P$  is said to be ergodic if there exists  $s$  such that the  $s$ -step transition probabilities  $p_{ij}^{(s)}$  are positive for all  $i$  and  $j$ . A homogeneous Markov chain is said to be ergodic if it can be generated by some initial distribution and an ergodic stochastic matrix.

By (5.2), ergodicity implies that in  $s$  steps one can, with positive probability, proceed from any initial state  $i$  to any final state  $j$ .

It is easy to provide examples of non-ergodic Markov Chains. One could consider a collection of non-intersecting sets  $X_1, \dots, X_n$ , and take  $X = \bigcup_{k=1}^n X_k$ . Suppose the transition probabilities  $p_{ij}$  are such that  $p_{ij} = 0$ , unless  $i$  and  $j$  belong to consecutive sets, that is  $i \in X_k, j \in X_{k+1}$  or  $i \in X_n, j \in X_1$ . Then the matrix  $P$  is block diagonal, and any power of  $P$  will contain zeros, thus  $P$  will not be ergodic.

Another example of a non-ergodic Markov chain arises when a state  $j$  cannot be reached from any other state, that is  $p_{ij} = 0$  for all  $i \neq j$ . Then the same will be true for the  $s$ -step transition probabilities.

Finally, there may be non-intersecting sets  $X_1, \dots, X_n$  such that  $X = \bigcup_{k=1}^n X_k$ , and the transition probabilities  $p_{ij}$  are such that  $p_{ij} = 0$ , unless  $i$  and  $j$  belong to the same set  $X_k$ . Then the matrix is not ergodic.

The general classification of Markov chains will be discussed in Sect. 5.6.

**Definition 5.8.** A probability distribution  $\pi$  on  $X$  is said to be stationary (or invariant) for a matrix of transition probabilities  $P$  if  $\pi P = \pi$ .

Formula (5.2) means that if the initial distribution  $\pi$  is a stationary distribution, then the probability distribution of any  $\omega_k$  is given by the same vector  $\pi$  and does not depend on  $k$ . Hence the term ‘‘stationary’’.

**Theorem 5.9 (Ergodic Theorem for Markov chains).** *Given a Markov chain with an ergodic matrix of transition probabilities  $P$ , there exists a unique stationary probability distribution  $\pi = (\pi_1, \dots, \pi_r)$ . The  $n$ -step transition probabilities converge to the distribution  $\pi$ , that is*

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \pi_j.$$

The stationary distribution satisfies  $\pi_j > 0$  for  $1 \leq j \leq r$ .

*Proof.* Let  $\mu' = (\mu'_1, \dots, \mu'_r)$ ,  $\mu'' = (\mu''_1, \dots, \mu''_r)$  be two probability distributions on the space  $X$ . We set  $d(\mu', \mu'') = \frac{1}{2} \sum_{i=1}^r |\mu'_i - \mu''_i|$ . Then  $d$  can be viewed as a distance on the space of probability distributions on  $X$ , and the space of distributions with this distance is a complete metric space. We note that

$$0 = \sum_{i=1}^r \mu'_i - \sum_{i=1}^r \mu''_i = \sum_{i=1}^r (\mu'_i - \mu''_i) = \sum^+ (\mu'_i - \mu''_i) - \sum^+ (\mu''_i - \mu'_i),$$

where  $\sum^+$  denotes the summation with respect to those indices  $i$  for which the terms are positive. Therefore,

$$d(\mu', \mu'') = \frac{1}{2} \sum_{i=1}^r |\mu'_i - \mu''_i| = \frac{1}{2} \sum^+ (\mu'_i - \mu''_i) + \frac{1}{2} \sum^+ (\mu''_i - \mu'_i) = \sum^+ (\mu'_i - \mu''_i).$$

It is also clear that  $d(\mu', \mu'') \leq 1$ .

Let  $\mu'$  and  $\mu''$  be two probability distributions on  $X$  and  $Q = (q_{ij})$  a stochastic matrix. By Lemma 5.2,  $\mu'Q$  and  $\mu''Q$  are also probability distributions. Let us demonstrate that

$$d(\mu'Q, \mu''Q) \leq d(\mu', \mu''), \quad (5.3)$$

and if all  $q_{ij} \geq \alpha$ , then

$$d(\mu'Q, \mu''Q) \leq (1 - \alpha)d(\mu', \mu''). \quad (5.4)$$

Let  $J$  be the set of indices  $j$  for which  $(\mu'Q)_j - (\mu''Q)_j > 0$ . Then

$$\begin{aligned} d(\mu'Q, \mu''Q) &= \sum_{j \in J} (\mu'Q - \mu''Q)_j = \sum_{j \in J} \sum_{i=1}^r (\mu'_i - \mu''_i) q_{ij} \\ &\leq \sum_i^+ (\mu'_i - \mu''_i) \sum_{j \in J} q_{ij} \leq \sum_i^+ (\mu'_i - \mu''_i) = d(\mu', \mu''), \end{aligned}$$

which proves (5.3). We now note that  $J$  can not contain all the indices  $j$  since both  $\mu'Q$  and  $\mu''Q$  are probability distributions. Therefore, at least one index

$j$  is missing in the sum  $\sum_{j \in J} q_{ij}$ . Thus, if all  $q_{ij} > \alpha$ , then  $\sum_{j \in J} q_{ij} < 1 - \alpha$  for all  $i$ , and

$$d(\mu'Q, \mu''Q) \leq (1 - \alpha) \sum_i^+ (\mu'_i - \mu''_i) = (1 - \alpha)d(\mu', \mu''),$$

which implies (5.4).

Let  $\mu_0$  be an arbitrary probability distribution on  $X$  and  $\mu_n = \mu_0 P^n$ . We shall show that the sequence of probability distributions  $\mu_n$  is a Cauchy sequence, that is for any  $\varepsilon > 0$  there exists  $n_0(\varepsilon)$  such that for any  $k \geq 0$  we have  $d(\mu_n, \mu_{n+k}) < \varepsilon$  for  $n \geq n_0(\varepsilon)$ . By (5.4),

$$\begin{aligned} d(\mu_n, \mu_{n+k}) &= d(\mu_0 P^n, \mu_0 P^{n+k}) \leq (1 - \alpha)d(\mu_0 P^{n-s}, \mu_0 P^{n+k-s}) \leq \dots \\ &\leq (1 - \alpha)^m d(\mu_0 P^{n-ms}, \mu_0 P^{n+k-ms}) \leq (1 - \alpha)^m, \end{aligned}$$

where  $m$  is such that  $0 \leq n - ms < s$ . For sufficiently large  $n$  we have  $(1 - \alpha)^m < \varepsilon$ , which implies that  $\mu_n$  is a Cauchy sequence.

Let  $\pi = \lim_{n \rightarrow \infty} \mu_n$ . Then

$$\pi P = \lim_{n \rightarrow \infty} \mu_n P = \lim_{n \rightarrow \infty} (\mu_0 P^n) P = \lim_{n \rightarrow \infty} (\mu_0 P^{n+1}) = \pi.$$

Let us show that the distribution  $\pi$ , such that  $\pi P = \pi$ , is unique. Let  $\pi_1$  and  $\pi_2$  be two distributions with  $\pi_1 = \pi_1 P$  and  $\pi_2 = \pi_2 P$ . Then  $\pi_1 = \pi_1 P^s$  and  $\pi_2 = \pi_2 P^s$ . Therefore,  $d(\pi_1, \pi_2) = d(\pi_1 P^s, \pi_2 P^s) \leq (1 - \alpha)d(\pi_1, \pi_2)$  by (5.4). It follows that  $d(\pi_1, \pi_2) = 0$ , that is  $\pi_1 = \pi_2$ .

We have proved that for any initial distribution  $\mu_0$  the limit

$$\lim_{n \rightarrow \infty} \mu_0 P^n = \pi$$

exists and does not depend on the choice of  $\mu_0$ . Let us take  $\mu_0$  to be the probability distribution which is concentrated at the point  $i$ . Then, for  $i$  fixed,  $\mu_0 P^n$  is the probability distribution  $(p_{ij}^{(n)})$ . Therefore,  $\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \pi_j$ .

The proof of the fact that  $\pi_j > 0$  for  $1 \leq j \leq r$  is left as an easy exercise for the reader.  $\square$

*Remark 5.10.* Let  $\mu_0$  be concentrated at the point  $i$ . Then

$$d(\mu_0 P^n, \pi) = d(\mu_0 P^n, \pi P^n) \leq \dots \leq (1 - \alpha)^m d(\mu_0 P^{n-ms}, \pi P^{n-ms}) \leq (1 - \alpha)^m,$$

where  $m$  is such that  $0 \leq n - ms < s$ . Therefore,

$$d(\mu_0 P^n, \pi) \leq (1 - \alpha)^{\frac{n}{s} - 1} \leq (1 - \alpha)^{-1} \beta^n,$$

where  $\beta = (1 - \alpha)^{\frac{1}{s}} < 1$ . In other words, the rate of convergence of  $p_{ij}^{(n)}$  to the limit  $\pi_j$  is exponential.

*Remark 5.11.* The term ergodicity comes from statistical mechanics. In our case the ergodicity of a Markov chain implies that a certain loss of memory regarding initial conditions occurs, as the probability distribution at time  $n$  becomes nearly independent of the initial distribution as  $n \rightarrow \infty$ . We shall discuss further the meaning of this notion in Chap. 16.

## 5.4 Law of Large Numbers and the Entropy of a Markov Chain

As in the case of a homogeneous sequence of independent trials, we introduce the random variable  $\nu_i^n(\omega)$  equal to the number of occurrences of the state  $i$  in the sequence  $\omega = (\omega_0, \dots, \omega_n)$ , that is the number of those  $0 \leq k \leq n$  for which  $\omega_k = i$ . We also introduce the random variables  $\nu_{ij}^n(\omega)$  equal to the number of those  $1 \leq k \leq n$  for which  $\omega_{k-1} = i, \omega_k = j$ .

**Theorem 5.12.** *Let  $\pi$  be the stationary distribution of an ergodic Markov chain. Then for any  $\varepsilon > 0$*

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{\nu_i^n}{n} - \pi_i\right| \geq \varepsilon\right) = 0, \quad \text{for } 1 \leq i \leq r,$$

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{\nu_{ij}^n}{n} - \pi_i p_{ij}\right| \geq \varepsilon\right) = 0, \quad \text{for } 1 \leq i, j \leq r.$$

*Proof.* Let

$$\chi_i^k(\omega) = \begin{cases} 1 & \text{if } \omega_k = i, \\ 0 & \text{if } \omega_k \neq i, \end{cases}$$

$$\chi_{ij}^k(\omega) = \begin{cases} 1 & \text{if } \omega_{k-1} = i, \omega_k = j, \\ 0 & \text{otherwise,} \end{cases}$$

so that

$$\nu_i^n = \sum_{k=0}^n \chi_i^k, \quad \nu_{ij}^n = \sum_{k=1}^n \chi_{ij}^k.$$

For an initial distribution  $\mu$

$$\mathbb{E}\chi_i^k = \sum_{m=1}^r \mu_m p_{mi}^{(k)}, \quad \mathbb{E}\chi_{ij}^k = \sum_{m=1}^r \mu_m p_{mi}^{(k)} p_{ij}.$$

As  $k \rightarrow \infty$  we have  $p_{mi}^{(k)} \rightarrow \pi_i$  exponentially fast. Therefore, as  $k \rightarrow \infty$ ,

$$\mathbb{E}\chi_i^k \rightarrow \pi_i, \quad \mathbb{E}\chi_{ij}^k \rightarrow \pi_i p_{ij}$$

exponentially fast. Consequently

$$\mathbb{E}\frac{\nu_i^n}{n} = \mathbb{E}\frac{\sum_{k=0}^n \chi_i^k}{n} \rightarrow \pi_i, \quad \mathbb{E}\frac{\nu_{ij}^n}{n} = \mathbb{E}\frac{\sum_{k=1}^n \chi_{ij}^k}{n} \rightarrow \pi_i p_{ij}.$$

For sufficiently large  $n$

$$\{\omega : \left|\frac{\nu_i^n(\omega)}{n} - \pi_i\right| \geq \varepsilon\} \subseteq \left\{\omega : \left|\frac{\nu_i^n(\omega)}{n} - \frac{1}{n}\mathbb{E}\nu_i^n\right| \geq \frac{\varepsilon}{2}\right\},$$

$$\{\omega : \left|\frac{\nu_{ij}^n(\omega)}{n} - \pi_i p_{ij}\right| \geq \varepsilon\} \subseteq \left\{\omega : \left|\frac{\nu_{ij}^n(\omega)}{n} - \frac{1}{n}\mathbb{E}\nu_{ij}^n\right| \geq \frac{\varepsilon}{2}\right\}.$$



The probabilities of the events on the right-hand side can be estimated using the Chebyshev Inequality:

$$\begin{aligned} \mathbb{P}\left(\left|\frac{\nu_i^n}{n} - \frac{1}{n}\mathbb{E}\nu_i^n\right| \geq \frac{\varepsilon}{2}\right) &= \mathbb{P}\left(|\nu_i^n - \mathbb{E}\nu_i^n| \geq \frac{\varepsilon n}{2}\right) \leq \frac{4\text{Var}(\nu_i^n)}{\varepsilon^2 n^2}, \\ \mathbb{P}\left(\left|\frac{\nu_{ij}^n}{n} - \frac{1}{n}\mathbb{E}\nu_{ij}^n\right| \geq \frac{\varepsilon}{2}\right) &= \mathbb{P}\left(|\nu_{ij}^n - \mathbb{E}\nu_{ij}^n| \geq \frac{\varepsilon n}{2}\right) \leq \frac{4\text{Var}(\nu_{ij}^n)}{\varepsilon^2 n^2}. \end{aligned}$$

Thus the matter is reduced to estimating  $\text{Var}(\nu_i^n)$  and  $\text{Var}(\nu_{ij}^n)$ . If we set  $m_i^k = \mathbb{E}\chi_i^k = \sum_{s=1}^r \mu_s p_{si}^{(k)}$ , then

$$\begin{aligned} \text{Var}(\nu_i^n) &= \mathbb{E}\left(\sum_{k=0}^n (\chi_i^k - m_i^k)\right)^2 = \\ &= \mathbb{E}\sum_{k=0}^n (\chi_i^k - m_i^k)^2 + 2 \sum_{k_1 < k_2} \mathbb{E}(\chi_i^{k_1} - m_i^{k_1})(\chi_i^{k_2} - m_i^{k_2}). \end{aligned}$$

Since  $0 \leq \chi_i^k \leq 1$ , we have  $-1 \leq \chi_i^k - m_i^k \leq 1$ ,  $(\chi_i^k - m_i^k)^2 \leq 1$  and  $\sum_{k=0}^n \mathbb{E}(\chi_i^k - m_i^k)^2 \leq n + 1$ . Furthermore,

$$\begin{aligned} \mathbb{E}(\chi_i^{k_1} - m_i^{k_1})(\chi_i^{k_2} - m_i^{k_2}) &= \mathbb{E}\chi_i^{k_1}\chi_i^{k_2} - m_i^{k_1}m_i^{k_2} = \\ &= \sum_{s=1}^r \mu_s p_{si}^{(k_1)} p_{si}^{(k_2-k_1)} - m_i^{k_1}m_i^{k_2} = R_{k_1, k_2}. \end{aligned}$$

By the Ergodic Theorem (see Remark 5.10),

$$\begin{aligned} m_i^k &= \pi_i + d_i^k, \quad |d_i^k| \leq c\lambda^k, \\ p_{si}^{(k)} &= \pi_i + \beta_{s,i}^k, \quad |\beta_{s,i}^k| \leq c\lambda^k, \end{aligned}$$

for some constants  $c < \infty$  and  $\lambda < 1$ . This gives

$$\begin{aligned} |R_{k_1, k_2}| &= \left| \sum_{s=1}^r \mu_s (\pi_i + \beta_{s,i}^{k_1})(\pi_i + \beta_{s,i}^{k_2-k_1}) - (\pi_i + d_i^{k_1})(\pi_i + d_i^{k_2}) \right| \leq \\ &= c_1(\lambda^{k_1} + \lambda^{k_2} + \lambda^{k_2-k_1}) \end{aligned}$$

for some constant  $c_1 < \infty$ . Therefore,  $\sum_{k_1 < k_2} R_{k_1, k_2} \leq c_2 n$ , and consequently  $\text{Var}(\nu_i^n) \leq c_3 n$  for some constants  $c_2$  and  $c_3$ . The variance  $\text{Var}(\nu_{ij}^n)$  can be estimated in the same way.  $\square$

We now draw a conclusion from this theorem about the entropy of a Markov chain. In the case of a homogeneous sequence of independent trials, for large  $n$  the entropy is approximately equal to  $-\frac{1}{n} \ln p(\omega)$  for typical  $\omega$ , that is for  $\omega$  which constitute a set whose probability is arbitrarily close to one.

In order to use this property to derive a general definition of entropy, we need to study the behavior of  $\ln p(\omega)$  for typical  $\omega$  in the case of a Markov chain. For  $\omega = (\omega_0, \dots, \omega_n)$  we have

$$p(\omega) = \mu_{\omega_0} \prod_{i,j} p_{ij}^{\nu_{ij}^n(\omega)} = \exp(\ln \mu_{\omega_0} + \sum_{i,j} \nu_{ij}^n(\omega) \ln p_{ij}),$$

$$\ln p(\omega) = \ln \mu_{\omega_0} + \sum_{i,j} \nu_{ij}^n(\omega) \ln p_{ij}.$$

From the Law of Large Numbers, for typical  $\omega$

$$\frac{\nu_{ij}^n(\omega)}{n} \sim \pi_i p_{ij}.$$

Therefore, for such  $\omega$

$$-\frac{1}{n} \ln p(\omega) = -\frac{1}{n} \ln \mu_{\omega_0} - \sum_{i,j} \nu_{ij}^n(\omega) \ln p_{ij} \sim -\sum_{i,j} \pi_i p_{ij} \ln p_{ij}.$$

Thus it is natural to define the entropy of a Markov chain to be

$$h = -\sum_i \pi_i \sum_j p_{ij} \ln p_{ij}.$$

It is not difficult to show that with such a definition of  $h$ , the MacMillan Theorem remains true.

## 5.5 Products of Positive Matrices

Let  $A = (a_{ij})$  be a matrix with positive entries,  $1 \leq i, j \leq r$ . Let  $A^* = (a_{ij}^*)$  be the transposed matrix, that is  $a_{ij}^* = a_{ji}$ . Let us denote the entries of  $A^n$  by  $a_{ij}^{(n)}$ . We shall use the Ergodic Theorem for Markov chains in order to study the asymptotic behavior of  $a_{ij}^{(n)}$  as  $n \rightarrow \infty$ . First, we prove the following:

**Theorem 5.13 (Perron-Frobenius Theorem).** *There exist a positive number  $\lambda$  (eigenvalue) and vectors  $e = (e_1, \dots, e_r)$  and  $f = (f_1, \dots, f_r)$  (right and left eigenvectors) such that*

1.  $e_j > 0, f_j > 0, 1 \leq j \leq r$ .
2.  $Ae = \lambda e$  and  $A^*f = \lambda f$ .

*If  $Ae' = \lambda'e'$  and  $e'_j > 0$  for  $1 \leq j \leq r$ , then  $\lambda' = \lambda$  and  $e' = c_1 e$  for some positive constant  $c_1$ . If  $A^*f' = \lambda'f'$  and  $f'_j > 0$  for  $1 \leq j \leq r$ , then  $\lambda' = \lambda$  and  $f' = c_2 f$  for some positive constant  $c_2$ .*

*Proof.* Let us show that there exist  $\lambda > 0$  and a positive vector  $e$  such that  $Ae = \lambda e$ , that is

$$\sum_{j=1}^r a_{ij}e_j = \lambda e_j, \quad 1 \leq i \leq r.$$

Consider the convex set  $\mathcal{H}$  of vectors  $h = (h_1, \dots, h_r)$  such that  $h_i \geq 0$ ,  $1 \leq i \leq r$ , and  $\sum_{i=1}^r h_i = 1$ . The matrix  $A$  determines a continuous transformation  $\mathcal{A}$  of  $\mathcal{H}$  into itself through the formula

$$(\mathcal{A}h)_i = \frac{\sum_{j=1}^r a_{ij}h_j}{\sum_{i=1}^r \sum_{j=1}^r a_{ij}h_j}.$$

The Brouwer Theorem states that any continuous mapping of a convex closed set in  $\mathbb{R}^n$  to itself has a fixed point. Thus we can find  $e \in \mathcal{H}$  such that  $\mathcal{A}e = e$ , that is,

$$e_i = \frac{\sum_{j=1}^r a_{ij}e_j}{\sum_{i=1}^r \sum_{j=1}^r a_{ij}e_j}.$$

Note that  $e_i > 0$  for all  $1 \leq i \leq r$ . By setting  $\lambda = \sum_{i=1}^r \sum_{j=1}^r a_{ij}e_j$ , we obtain  $\sum_{j=1}^r a_{ij}e_j = \lambda e_i$ ,  $1 \leq i \leq r$ .

In the same way we can show that there is  $\bar{\lambda} > 0$  and a vector  $f$  with positive entries such that  $A^*f = \bar{\lambda}f$ . The equalities

$$\lambda(e, f) = (Ae, f) = (e, A^*f) = (e, \bar{\lambda}f) = \bar{\lambda}(e, f)$$

show that  $\lambda = \bar{\lambda}$ .

We leave the uniqueness part as an exercise for the reader. □

Let  $e$  and  $f$  be positive right and left eigenvectors, respectively, which satisfy

$$\sum_{i=1}^r e_i = 1 \quad \text{and} \quad \sum_{i=1}^r e_i f_i = 1.$$

Note that these conditions determine  $e$  and  $f$  uniquely. Let  $\lambda > 0$  be the corresponding eigenvalue. Set

$$p_{ij} = \frac{a_{ij}e_j}{\lambda e_i}.$$

It is easy to see that the matrix  $P = (p_{ij})$  is a stochastic matrix with strictly positive entries. The stationary distribution of this matrix is  $\pi_i = e_i f_i$ . Indeed,

$$\sum_{i=1}^r \pi_i p_{ij} = \sum_{i=1}^r e_i f_i \frac{a_{ij}e_j}{\lambda e_i} = \frac{1}{\lambda} e_j \sum_{i=1}^r f_i a_{ij} = e_j f_j = \pi_j.$$

We can rewrite  $a_{ij}^{(n)}$  as follows:

$$\begin{aligned} a_{ij}^{(n)} &= \sum_{1 \leq i_1, \dots, i_{n-1} \leq r} a_{ii_1} \cdot a_{i_1 i_2} \cdot \dots \cdot a_{i_{n-2} i_{n-1}} \cdot a_{i_{n-1} j} \\ &= \lambda^n \sum_{1 \leq i_1, \dots, i_{n-1} \leq r} p_{ii_1} \cdot p_{i_1 i_2} \cdot \dots \cdot p_{i_{n-2} i_{n-1}} \cdot p_{i_{n-1} j} \cdot e_i \cdot e_j^{-1} = \lambda^n e_i p_{ij}^{(n)} e_j^{-1}. \end{aligned}$$

The Ergodic Theorem for Markov chains gives  $p_{ij}^{(n)} \rightarrow \pi_j = e_j f_j$  as  $n \rightarrow \infty$ . Therefore,

$$\frac{a_{ij}^{(n)}}{\lambda^n} \rightarrow e_i \pi_j e_j^{-1} = e_i f_j$$

and the convergence is exponentially fast. Thus

$$a_{ij}^{(n)} \sim \lambda^n e_i f_j \text{ as } n \rightarrow \infty.$$

*Remark 5.14.* One can easily extend these arguments to the case where the matrix  $A^s$  has positive matrix elements for some integer  $s > 0$ .

## 5.6 General Markov Chains and the Doeblin Condition

Markov chains often appear as random perturbations of deterministic dynamics. Let  $(X, \mathcal{G})$  be a measurable space and  $f: X \rightarrow X$  a measurable mapping of  $X$  into itself. We may wish to consider the trajectory of a point  $x \in X$  under the iterations of  $f$ , that is the sequence  $x, f(x), f^2(x), \dots$ . However, if random noise is present, then  $x$  is mapped not to  $f(x)$  but to a nearby random point. This means that for each  $C \in \mathcal{G}$  we must consider the transition probability from the point  $x$  to the set  $C$ . Let us give the corresponding definition.

**Definition 5.15.** *Let  $(X, \mathcal{G})$  be a measurable space. A function  $P(x, C)$ ,  $x \in X, C \in \mathcal{G}$ , is called a Markov transition function if for each fixed  $x \in X$  the function  $P(x, C)$ , as a function of  $C \in \mathcal{G}$ , is a probability measure defined on  $\mathcal{G}$ , and for each fixed  $C \in \mathcal{G}$  the function  $P(x, C)$  is measurable as a function of  $x \in X$ .*

For  $x$  and  $C$  fixed,  $P(x, C)$  is called the transition probability from the initial point  $x$  to the set  $C$ . Given a Markov transition function  $P(x, C)$  and an integer  $n \in \mathbb{N}$ , we can define the  $n$ -step transition function

$$P^n(x, C) = \int_X \dots \int_X \int_X P(x, dy_1) \dots P(y_{n-2}, dy_{n-1}) P(y_{n-1}, C).$$

It is easy to see that  $P^n$  satisfies the definition of a Markov transition function.

A Markov transition function  $P(x, C)$  defines two operators:

1. The operator  $P$  which acts on bounded measurable functions

$$(Pf)(x) = \int_X f(y)P(x, dy); \quad (5.5)$$

2. The operator  $P^*$  which acts on the probability measures

$$(P^*\mu)(C) = \int_X P(x, C)d\mu(x). \quad (5.6)$$

It is easy to show (see Problem 15) that the image of a bounded measurable function under the action of  $P$  is again a bounded measurable function, while the image of a probability measure  $\mu$  under  $P^*$  is again a probability measure.

*Remark 5.16.* Note that we use the same letter  $P$  for the Markov transition function and the corresponding operator. This is partially justified by the fact that the  $n$ -th power of the operator corresponds to the  $n$ -step transition function, that is

$$(P^n f)(x) = \int_X f(y)P^n(x, dy).$$

**Definition 5.17.** A probability measure  $\pi$  is called a stationary (or invariant) measure for the Markov transition function  $P$  if  $\pi = P^*\pi$ , that is

$$\pi(C) = \int_X P(x, C)d\pi(x)$$

for all  $C \in \mathcal{G}$ .

Given a Markov transition function  $P$  and a probability measure  $\mu_0$  on  $(X, \mathcal{G})$ , we can define the corresponding homogeneous Markov chain, that is the measure on the space of sequences  $\omega = (\omega_0, \dots, \omega_n)$ ,  $\omega_k \in X$ ,  $k = 0, \dots, n$ . Namely, denote by  $\mathcal{F}$  the  $\sigma$ -algebra generated by the elementary cylinders, that is by the sets of the form  $A = \{\omega : \omega_0 \in A_0, \omega_1 \in A_1, \dots, \omega_n \in A_n\}$  where  $A_k \in \mathcal{G}$ ,  $k = 0, \dots, n$ . By Theorem 3.19, if we define

$$P(A) = \int_{A_0 \times \dots \times A_{n-1}} d\mu_0(x_0)P(x_0, dx_1) \dots P(x_{n-2}, dx_{n-1})P(x_{n-1}, A_n),$$

there exists a measure on  $\mathcal{F}$  which coincides with  $P(A)$  on the elementary cylinders. Moreover, such a measure on  $\mathcal{F}$  is unique.

*Remark 5.18.* We could also consider a measure on the space of infinite sequences  $\omega = (\omega_0, \omega_1, \dots)$  with  $\mathcal{F}$  still being the  $\sigma$ -algebra generated by the elementary cylinders. In this case, there is still a unique measure on  $\mathcal{F}$  which coincides with  $P(A)$  on the elementary cylinder sets. Its existence is guaranteed by the Kolmogorov Consistency Theorem which is discussed in Chap. 12.

We have already seen that in the case of Markov chains with a finite state space the stationary measure determines the statistics of typical  $\omega$  (the Law of Large Numbers). This is also true in the more general setting which we are considering now. Therefore it is important to find sufficient conditions which guarantee the existence and uniqueness of the stationary measure.

**Definition 5.19.** A Markov transition function  $P$  is said to satisfy the strong Doeblin condition if there exist a probability measure  $\nu$  on  $(X, \mathcal{G})$  and a function  $p(x, y)$  (the density of  $P(x, dy)$  with respect to the measure  $\nu$ ) such that

1.  $p(x, y)$  is measurable on  $(X \times X, \mathcal{G} \times \mathcal{G})$ .
2.  $P(x, C) = \int_C p(x, y) d\nu(y)$  for all  $x \in X$  and  $C \in \mathcal{G}$ .
3. For some constant  $a > 0$  we have

$$p(x, y) \geq a \quad \text{for all } x, y \in X.$$

**Theorem 5.20.** If a Markov transition function satisfies the strong Doeblin condition, then there exists a unique stationary measure.

*Proof.* By the Fubini Theorem, for any measure  $\mu$  the measure  $P^*\mu$  is given by the density  $\int_X d\mu(x)p(x, y)$  with respect to the measure  $\nu$ . Therefore, if a stationary measure exists, it is absolutely continuous with respect to  $\nu$ . Let  $M$  be the space of measures which are absolutely continuous with respect to  $\nu$ . For  $\mu_1, \mu_2 \in M$ , the distance between them is defined via  $d(\mu^1, \mu^2) = \frac{1}{2} \int |m^1(y) - m^2(y)| d\nu(y)$ , where  $m^1$  and  $m^2$  are the densities of  $\mu^1$  and  $\mu^2$  respectively. We claim that  $M$  is a complete metric space with respect to the metric  $d$ . Indeed,  $M$  is a closed subspace of  $L^1(X, \mathcal{G}, \nu)$ , which is a complete metric space. Let us show that the operator  $P^*$  acting on this space is a contraction.

Consider two measures  $\mu^1$  and  $\mu^2$  with the densities  $m^1$  and  $m^2$ . Let  $A^+ = \{y : m^1(y) - m^2(y) \geq 0\}$  and  $A^- = X \setminus A^+$ . Similarly let  $B^+ = \{y : \int_X p(x, y)(m^1(x) - m^2(x)) d\nu(x) \geq 0\}$  and  $B^- = X \setminus B^+$ . Without loss of generality we can assume that  $\nu(B^-) \geq \frac{1}{2}$  (if the contrary is true and  $\nu(B^+) > \frac{1}{2}$ , we can replace  $A^+$  by  $A^-$ ,  $B^+$  by  $B^-$  and reverse the signs in some of the integrals below).

As in the discrete case,  $d(\mu^1, \mu^2) = \int_{A^+} (m^1(y) - m^2(y)) d\nu(y)$ . Therefore,

$$\begin{aligned} d(P^*\mu^1, P^*\mu^2) &= \int_{B^+} \left[ \int_X p(x, y)(m^1(x) - m^2(x)) d\nu(x) \right] d\nu(y) \\ &\leq \int_{B^+} \left[ \int_{A^+} p(x, y)(m^1(x) - m^2(x)) d\nu(x) \right] d\nu(y) \\ &= \int_{A^+} \left[ \int_{B^+} p(x, y) d\nu(y) \right] (m^1(x) - m^2(x)) d\nu(x). \end{aligned}$$

The last expression contains the integral  $\int_{B^+} p(x, y) d\nu(y)$  which we estimate as follows

$$\int_{B^+} p(x, y) d\nu(y) = 1 - \int_{B^-} p(x, y) d\nu(y) \leq 1 - a\nu(B^-) \leq 1 - \frac{a}{2}.$$

This shows that

$$d(P^* \mu^1, P^* \mu^2) \leq (1 - \frac{a}{2})d(\mu^1, \mu^2).$$

Therefore  $P^*$  is a contraction and has a unique fixed point, which completes the proof of the theorem. □

The strong Doeblin condition can be considerably relaxed, yet we may still be able to say something about the stationary measures. We conclude this section with a discussion of the structure of a Markov chain under the Doeblin condition. We shall restrict ourselves to formulation of results.

**Definition 5.21.** *We say that  $P$  satisfies the Doeblin condition if there is a finite measure  $\mu$  with  $\mu(X) > 0$ , an integer  $n$ , and a positive  $\varepsilon$  such that for any  $x \in X$*

$$P^n(x, A) \leq 1 - \varepsilon \quad \text{if} \quad \mu(A) \leq \varepsilon.$$

**Theorem 5.22.** *If a Markov transition function satisfies the Doeblin condition, then the space  $X$  can be represented as the union of non-intersecting sets:*

$$X = \bigcup_{i=1}^k E_i \cup T,$$

where the sets  $E_i$  (ergodic components) have the property  $P(x, E_i) = 1$  for  $x \in E_i$ , and for the set  $T$  (the transient set) we have  $\lim_{n \rightarrow \infty} P^n(x, T) = 0$  for all  $x \in X$ . The sets  $E_i$  can in turn be represented as unions of non-intersecting subsets:

$$E_i = \bigcup_{j=0}^{m_i-1} C_i^j,$$

where  $C_i^j$  (cyclically moving subsets) have the property

$$P(x, C_i^{j+1(\text{mod } m_i)}) = 1 \quad \text{for} \quad x \in C_i^j.$$

Note that if  $P$  is a Markov transition function on the state space  $X$ , then  $P(x, A)$ ,  $x \in E_i$ ,  $A \subseteq E_i$  is a Markov transition function on  $E_i$ . We have the following theorem describing the stationary measures of Markov transition functions satisfying the Doeblin condition (see “Stochastic Processes” by J.L. Doob).

**Theorem 5.23.** *If a Markov transition function satisfies the Doeblin condition, and  $X = \bigcup_{i=1}^k E_i \cup T$  is a decomposition of the state space into ergodic components and the transient set, then*

1. The restriction of the transition function to each ergodic component has a unique stationary measure  $\pi_i$ .
2. Any stationary measure  $\pi$  on the space  $X$  is equal to a linear combination of the stationary measures on the ergodic components:

$$\pi = \sum_{i=1}^k \alpha_i \pi_i$$

with  $\alpha_i \geq 0$ ,  $\alpha_1 + \dots + \alpha_k = 1$ .

Finally, we formulate the Strong Law of Large Numbers for Markov chains (see “Stochastic Processes” by J.L. Doob).

**Theorem 5.24.** *Consider a Markov transition function which satisfies the Doeblin condition and has only one ergodic component. Let  $\pi$  be the unique stationary measure. Consider the corresponding Markov chain (measure on the space of sequences  $\omega = (\omega_0, \omega_1, \dots)$ ) with some initial distribution. Then for any function  $f \in L^1(X, \mathcal{G}, \pi)$  the following limit exists almost surely:*

$$\lim_{n \rightarrow \infty} \frac{\sum_{k=0}^n f(\omega_k)}{n+1} = \int_X f(x) d\pi(x).$$

## 5.7 Problems

1. Let  $P$  be a stochastic matrix. Prove that there is at least one non-negative vector  $\pi$  such that  $\pi P = \pi$ .
2. Consider a homogeneous Markov chain on a finite state space with the transition matrix  $P$  and the initial distribution  $\mu$ . Prove that for any  $0 < k < n$  the induced probability distribution on the space of sequences  $(\omega_k, \omega_{k+1}, \dots, \omega_n)$  is also a homogeneous Markov chain. Find its initial distribution and the matrix of transition probabilities.
3. Consider a homogeneous Markov chain on a finite state space  $X$  with transition matrix  $P$  and the initial distribution  $\delta_x$ ,  $x \in X$ , that is  $P(\omega_0 = x) = 1$ . Let  $\tau$  be the first  $k$  such that  $\omega_k \neq x$ . Find the probability distribution of  $\tau$ .
4. Consider the one-dimensional simple symmetric random walk (Markov chain on the state space  $\mathbb{Z}$  with transition probabilities  $p_{i,i+1} = p_{i,i-1} = 1/2$ ). Prove that it does not have a stationary distribution.
5. For a homogeneous Markov chain on a finite state space  $X$  with transition matrix  $P$  and initial distribution  $\mu$ , find  $P(\omega_n = x^1 | \omega_0 = x^2, \omega_{2n} = x^3)$ , where  $x^1, x^2, x^3 \in X$ .
6. Consider a homogeneous ergodic Markov chain on the finite state space  $X = \{1, \dots, r\}$  with the transition matrix  $P$  and the stationary distribution  $\pi$ . Assuming that  $\pi$  is also the initial distribution, find the following limit

$$\lim_{n \rightarrow \infty} \frac{\ln P(\omega_i \neq 1 \text{ for } 0 \leq i \leq n)}{n}.$$



7. Consider a homogeneous ergodic Markov chain on the finite state space  $X = \{1, \dots, r\}$ . Define the random variables  $\tau_n, n \geq 1$ , as the consecutive times when the Markov chain is in the state 1, that is

$$\tau_1 = \inf(i \geq 0 : \omega_i = 1),$$

$$\tau_n = \inf(i > \tau_{n-1} : \omega_i = 1), \quad n > 1.$$

Prove that  $\tau_1, \tau_2 - \tau_1, \tau_3 - \tau_2, \dots$  is a sequence of independent random variables.

8. Consider a homogeneous ergodic Markov chain on a finite state space with the transition matrix  $P$  and the stationary distribution  $\pi$ . Assuming that  $\pi$  is also the initial distribution, prove that the distribution of the inverse process  $(\omega_n, \omega_{n-1}, \dots, \omega_1, \omega_0)$  is also a homogeneous Markov chain. Find its matrix of transition probabilities and stationary distribution.
9. Find the stationary distribution of the Markov chain with the countable state space  $\{0, 1, 2, \dots, n, \dots\}$ , where each point, including 0, can either return to 0 with probability 1/2 or move to the right  $n \mapsto n + 1$  with probability 1/2.
10. Let  $P$  be a matrix of transition probabilities of a homogeneous ergodic Markov chain on a finite state space such that  $p_{ij} = p_{ji}$ . Find its stationary distribution.
11. Consider a homogeneous Markov chain on the finite state space  $X = \{1, \dots, r\}$ . Assume that all the elements of the transition matrix are positive. Prove that for any  $k \geq 0$  and any  $x^0, x^1, \dots, x^k \in X$ ,

$$P(\text{there is } n \text{ such that } \omega_n = x^0, \omega_{n+1} = x^1, \dots, \omega_{n+k} = x^k) = 1.$$

12. Consider a Markov chain on a finite state space. Let  $k_1, k_2, l_1$  and  $l_2$  be integers such that  $0 \leq k_1 < l_1 \leq l_2 < k_2$ . Consider the conditional probabilities

$$f(i_{k_1}, \dots, i_{l_1-1}, i_{l_2+1}, \dots, i_{k_2}) =$$

$$P(\omega_{l_1} = i_{l_1}, \dots, \omega_{l_2} = i_{l_2} | \omega_{k_1} = i_{k_1}, \dots, \omega_{l_1-1} = i_{l_1-1}, \omega_{l_2+1}$$

$$= i_{l_2+1}, \dots, \omega_{k_2} = i_{k_2})$$

with  $i_{l_1}, \dots, i_{l_2}$  fixed. Prove that whenever  $f$  is defined, it depends only on  $i_{l_1-1}$  and  $i_{l_2+1}$ .

13. Consider a Markov chain whose state space is  $\mathbb{R}$ . Let  $P(x, A), x \in \mathbb{R}, A \in \mathcal{B}(\mathbb{R})$ , be the following Markov transition function,

$$P(x, A) = \lambda([x - 1/2, x + 1/2] \cap A),$$

where  $\lambda$  is the Lebesgue measure. Assuming that the initial distribution is concentrated at the origin, find  $P(|\omega_2| \leq 1/4)$ .

- 14.** Let  $p_{ij}$ ,  $i, j \in \mathbb{Z}$ , be the transition probabilities of a Markov chain on the state space  $\mathbb{Z}$ . Suppose that

$$p_{i,i-1} = 1 - p_{i,i+1} = r(i)$$

for all  $i \in \mathbb{Z}$ , where  $r(i) = r_- < 1/2$  if  $i < 0$ ,  $r(0) = 1/2$ , and  $r(i) = r_+ > 1/2$  if  $i > 0$ . Find the stationary distribution for this Markov chain. Does this Markov chain satisfy the Doeblin condition?

- 15.** For a given Markov transition function, let  $P$  and  $P^*$  be the operators defined by (5.5) and (5.6), respectively. Prove that the image of a bounded measurable function under the action of  $P$  is again a bounded measurable function, while the image of a probability measure  $\mu$  under  $P^*$  is again a probability measure.
- 16.** Consider a Markov chain whose state space is the unit circle. Let the density of the transition function  $P(x, dy)$  be given by

$$p(x, y) = \begin{cases} 1/(2\varepsilon) & \text{if } \text{angle}(y, x) < \varepsilon, \\ 0 & \text{otherwise,} \end{cases}$$

where  $\varepsilon > 0$ . Find the stationary measure for this Markov chain.