Chapter 2

Markov chains

The main purpose of this chapter is to introduce the concept of a Markov chain. The next chapter will then treat the filtering theory for Markov chains. The first section here focuses on particular Markov chains called random walks. The second section treats Markov chains in general. The concepts introduced in the second section are further illustrated with a variety of examples in the third section of this chapter.

1. Random walks

In the following definition, $\mathbb{Z}^d$ denotes the set consisting of those points in d-dimensional Euclidean space $\mathbb{R}^d$ whose coordinates are integers—positive, negative, or zero.

1. Definition (Random walks on $\mathbb{Z}^d$). Let $(X_n, n = 0, 1, 2, ...)$ be a sequence of $\mathbb{Z}^d$-valued random variables. We say that $(X_n, n = 0, 1, 2, ...)$ is *a random walk* on $\mathbb{Z}^d$ if the random variables

$$X_1 - X_0, X_2 - X_1, ...$$

are identically distributed and the random variables

$$X_0, X_1 - X_0, X_2 - X_1, ...$$

are independent. For $n \geq 1$ the random variable $X_n - X_{n-1}$ is the *step* or *jump* of the random walk at time n, the value X_n is called *the state of the random walk at time n,* and, in particular, X_0 is the (random) *starting state.*

Notice that since the distributions of $X_n - X_{n-1}$ coincide, the probability of a step of size z is the same for all $n \geq 1$ and $z \in \mathbb{Z}^d$:

$$P(X_n = X_{n-1} + z) = P(X_n - X_{n-1} = z) = P(X_1 - X_0 = z)\,.$$

Usually we think of a random walk $(X_n, n = 0, 1, 2, \dots)$ as being obtained in the following way. First we have a random experiment which yields a $\mathbb{Z}^d$-valued random variable X_0. Then we have a random experiment, which we repeat infinitely many times, each time independently of the results of all previous experiments, and obtain independent $\mathbb{Z}^d$-valued random variables $U_1, U_2, ...$ having the same distribution. After that we set

$$X_1 = X_0 + U_1\,, \quad X_2 = X_1 + U_2\,, \quad \dots\,.$$

In this situation $X_n - X_{n-1} = U_n$ has the same distribution as U_1 for all $n \geq 1$. Also the random variables $X_0, X_1 - X_0, X_2 - X_1, \dots$ are independent.

2. Definition (Initial distribution, step distribution). Let

$$(X_n, n = 0, 1, 2, ...) \tag{1}$$

be a random walk on $\mathbb{Z}^d$. The distribution of X_0 is called the *initial distribution* of the random walk. The distribution of $X_1 - X_0$—that is, the function

$$P(X_1 - X_0 = z), \quad z \in \mathbb{Z}^d$$

—is called its *step distribution*. This distribution is common to all steps $X_1 - X_0, X_2 - X_1, ...$

3. Definition (Symmetric random walks). A random walk (1) on $\mathbb{Z}^d$ is called *symmetric* if $X_1 - X_0$ and $X_0 - X_1$ have the same distribution.

4. Definition (Simple random walks). A random walk (1) on $\mathbb{Z}^d$ is *simple* if the values of $X_1 - X_0$ belong to the set $\{\pm e_i, i = 1, ..., d\}$, where e_i are the standard basis vectors in $\mathbb{R}^d$. A random walk (1) on $\mathbb{Z}^d$ is *simple symmetric* if

$$P(X_1 - X_0 = \pm e_i) = \frac{1}{2d}, \quad i = 1, 2, ..., d.$$

Recall that e_i is the vector whose i^{th} coordinate is 1 and all remaining ones equal 0.

Observe that for simple random walks the next state is always a nearest neighbor of the current one. The steps of symmetric random walks have the same distribution as their negatives. Also notice that a random walk which is simple and symmetric need not necessarily be a simple symmetric random walk.

5. Example. Let $d = 1$, $X_0 = 0$, and

$$P(X_1 - X_0 = \pm 1) = 1/2.$$

Then we get a simple random walk on $\mathbb{Z}$. One can obtain this random walk by considering an infinite sequence of Bernoulli trials with probability of success $1/2$, setting $X_0 = 0$, and letting

$$X_n - X_{n-1} = \pm 1$$

according as a success or a failure is obtained on the n^{th} trial. Remember that if we have a series of Bernoulli trials with probability of success $p \in [0,1]$ and M_n denotes the number of successes in n trials, then, for any $k = 0, 1, 2, \dots$, we have

$$P(M_n = k) = \binom{n}{k} p^k q^{n-k}, \quad \text{where } \ q = 1 - p\,. \tag{2}$$

In this notation, obviously,

$$X_n = M_n - (n - M_n) = 2M_n - n.$$

It follows that, for any $k \in \mathbb{Z}$

$$P(X_n = k) = P(M_n = (n+k)/2) = 2^{-n} \binom{n}{(n+k)/2} \tag{3}$$

if $n + k$ is even and $|k| \le n$ and $P(X_n = k) = 0$ otherwise.

6. Example. Let $(X_n, n = 0, 1, 2, \dots)$ be a simple symmetric random walk on $\mathbb{Z}$ starting at zero and let $(Y_n, n = 0, 1, 2 \dots)$ be an independent copy of $(X_n, n \ge 0)$, that is, a simple symmetric random walk on $\mathbb{Z}$ starting at zero and independent of $(X_n, n \ge 0)$. Then the sequence of ordered pairs $(Z_n = (X_n, Y_n)\,,\, n = 0, 1, 2, \dots)$ is a symmetric (but not simple) random walk on $\mathbb{Z}^2$ since

$$(X_1 - X_0, Y_1 - Y_0)\,, \quad (X_2 - X_1, Y_2 - Y_1)\,, \quad \cdots$$

are independent and identically distributed with

$$P(Z_n - Z_{n-1} = \pm e_1 \pm e_2) = P(X_n - X_{n-1} = \pm 1)\, P(Y_n - Y_{n-1} = \pm 1) = 1/4\,.$$

7. Problem. Prove that if $(Z_n = (X_n, Y_n)\,,\, n \geq 0)$ is a random walk on $\mathbb{Z}^2$ for which $Z_0 = (0,0)$,

$$P(Z_1 - Z_0 = 0) + \sum_{i=1}^{2} P(Z_1 - Z_0 = e_i) + \sum_{i=1}^{2} P(Z_1 - Z_0 = -e_i) = 1\,,$$

and if X_1 and Y_1 are independent, then $P(X_n = 0) = 1$ for all n or $P(Y_n = 0) = 1$ for all n.

8. Problem. If $(X_n, n = 0, 1, 2, \dots)$ is a random walk on $\mathbb{Z}^d$ and b is any point in $\mathbb{Z}^d$, prove that $(X_n + bn, n = 0, 1, 2, \dots)$ is a random walk on $\mathbb{Z}^d$. Also prove that if A is a linear transformation sending $\mathbb{Z}^d$ into itself, then $(AX_n, n = 0, 1, 2, \dots)$ is a random walk, which is symmetric if $(X_n, n = 0, 1, 2, \dots)$ is symmetric.

9. Problem. Let $(Z_n = (X_n, Y_n)\,,\, n \geq 0)$, be the random walk from Example 6. Set $Z'_n = ((X'_n, Y'_n)\,,\, n \geq 0)$ where

$$X'_n = (X_n + Y_n)/2 \quad \text{and} \quad Y'_n = (X_n - Y_n)/2\,.$$

Prove that $(Z'_n, n \geq 0)$, is a simple symmetric random walk starting at zero. Check that $Z'_n = 0$ if and only if $Z_n = 0$. Also, show that

$$P(Z_n = 0) = 2^{-2n} \binom{n}{n/2}^2$$

if n is even and $= 0$ otherwise. Conclude that for *any* simple symmetric random walk $(Z'_n, n \geq 0)$ on $\mathbb{Z}^2$ starting at the origin

$$P(Z'_n = 0) = 2^{-2n} \binom{n}{n/2}^2$$

if n is even and $= 0$ otherwise. Of course, emphasizing “any” makes little sense because all simple symmetric random walks behave in the same way.

What follows below in this section is not used in our presentation of filtering and prediction issues. We just wanted to give the reader a slightly deeper insight into simple random walks.

10. Theorem. *Let* $(X_n, n = 0, 1, 2...)$ *be a random walk on* $\mathbb{Z}^d$ *with* $X_0 = 0$*. Denote*

$$T = \inf\{n \geq 1 : X_n = 0\} \qquad (\inf \emptyset := \infty)$$

the first time when X_n *returns to the origin. Then the total number* N *of times when* X_n*,* $n = 0, 1, 2, ...$*, visits the origin, has the geometric distribution on* $\{1, 2, 3...\}$ *with mean*

$$\frac{1}{P(T = \infty)} \quad \textit{if} \quad P(T = \infty) > 0, \quad \textit{and} \quad \infty \quad \textit{otherwise}$$

(*recall that geometric distributions on* $\{1, 2, 3...\}$ *is introduced in Problem* 1.2.18). *In particular,* N *is infinite with probability one if and only if* $P(T = \infty) = 0$*. Also* $P(N = 1) = 1$ *if and only if* $P(T = \infty) = 1$*.*

Proof. Define the following sequence of experiments. We split the infinite trajectory of X_n, $n = 0, 1, 2, ...$ into pieces between consecutive returns of X_n to the origin and each piece we treat as the result of an experiment. We say that the first experiment yields a success if X_n returns to the origin, that is, $X_n = 0$ for some $n \geq 1$. After the first return to the origin we start the second experiment. We say that the second experiment yields a success if X_n returns to the origin a second time and so on. Since after coming to zero the random walk continues its evolution independently of the past and in the same way as initially, our sequence of experiments can be viewed as a sequence of independent Bernoulli trials. Then N equals the number of trials needed to get the first failure in our Bernoulli trials.

We infer that, if the probability of success in the first experiment is p, then

$$P(N = k) = p^{k-1}q,$$

where $q = 1 - p$ and $k = 1, 2, ...$ Thus indeed N has the geometric distribution on $\{1, 2, ...\}$ with parameter q. One knows that $1/q$ is the mean of the geometric (q) distribution, the number of successes is infinite with probability one if and only if $p = 1$ and the number of successes is zero if and only if $p = 0$. This proves the theorem since the probability of success $p = P(T < \infty)$.

Recall that the notation I_A, where A is a random event, is used for the random variable taking value 1 if A occurs and 0 if A does not occur.

11. Corollary. *We have* $P(T=\infty)=0$ *if and only if*

$$\sum_{n=0}^{\infty} P(X_n=0)=\infty.$$

Indeed, observe that N is the number of times when the events $\{X_n=0\}$ occur, that is, the number of 1's in the sequence of $I_{X_n=0}$. Also use the fact from probability theory that the expectation of a sum of nonnegative random variables is the sum of expectations. Then we find

$$N=\sum_{n=0}^{\infty} I_{X_n=0}, \quad EN=\sum_{n=0}^{\infty} P(X_n=0).$$

Thus, the series of probabilities on the right is infinite if and only if $EN=\infty$ and the latter happens (by the lemma) if and only if $P(T=\infty)=0$.

12. Definition (Recurrent random walks). We say that a random walk starting at zero, that is, satisfying $P(X_0=0)=1$, is *recurrent* if $P(T<\infty)=1$.

13. Problem. Prove that the random walks from Example 5 and Problem 9 are recurrent (Hint: By Stirling's formula $n! \sim (\frac{n}{e})^n\sqrt{2\pi n}$.)

14. Problem. Let $(X_n, n=0,1,2,...)$ be a simple random walk on $\mathbb{Z}$ with $X_0=0$ and $P(X_1=1)=p\in(0,1)$. Use the Law of Large Numbers to prove that $X_n/n\to p-q$ as $n\to\infty$. The exact content of your conclusion will depend on which version of the law you use.

Problem 14 shows that for $p\neq 1/2$ the simple random walk is not recurrent and goes to ∞ as $n\to\infty$ if $p>1/2$. In this case the probability that the random walk ever hits a specified point to the left of the origin is less than one. We are going to find this probability. First we find the probability that a simple random walk in one dimension starting at zero hits a point $a<0$ before hitting a point $b>0$. We need the following algebraic lemma.

15. Lemma. *Let $a, b \in \mathbb{Z}$, $a < 0, b > 0$, $p \in (0,1)$, $q = 1-p$, f be an arbitrary real-valued function defined on $[a,b] \cap \mathbb{Z}$. Then the system of linear algebraic equations*

$$\begin{cases} u(a) = f(a), \\ u(n) = pu(n+1) + qu(n-1) + f(n) \quad \text{for} \quad a < n < b, \\ u(b) = f(b) \end{cases} \tag{4}$$

has a unique solution.

Proof. The system (4) consists of $b - a + 1$ linear equations in the same number of unknowns. From linear algebra we know that the system has a unique solution for any f if and only if it has only one solution for $f \equiv 0$. So assume that $f \equiv 0$ and let u be a solution of (4) for that f. Among the values

$$u(a), u(a+1), ..., u(b-1), u(b)$$

at least one is maximal. Let $u(m)$ be this value. We want to show that $u(m) = 0$. If $m = a$ or $m = b$ we are done. If $a < m < b$, then

$$u(m) = pu(m+1) + qu(m-1) \le pu(m) + qu(m) = u(m).$$

Since $u(m)$ appears at both ends of this chain, we have equality throughout and, hence, both $u(m-1)$ and $u(m+1)$ equal $u(m)$. After this we can go to the right of $m+1$ unless $m+1 = b$ and conclude that $u(m+2) = u(m)$. After finitely many steps we get $u(m) = u(b) = 0$. Thus $\max_n u(n) = 0$. In the same way $\min_n u(n) = 0$, which proves the lemma.

16. Theorem. *Let $(X_n, n = 0, 1, 2, ...)$ be a simple random walk on $\mathbb{Z}$ with $X_0 = 0$ and $P(X_1 = 1) = p \in (0,1)$. Let $A_{a,b}$ be the event that X_n reaches a before reaching b, where $a, b \in \mathbb{Z}$, $a < 0, b > 0$. Then*

$$P(A_{a,b}) = \begin{cases} \frac{p^b q^{|a|} - q^{b+|a|}}{p^{b+|a|} - q^{b+|a|}} & \text{if} \quad p \ne 1/2, \\ \\ \frac{b}{b+|a|} & \text{if} \quad p = 1/2. \end{cases}$$

Proof. For $x \in [a,b] \cap \mathbb{Z}$ define $u(x)$ as the probability of the event

$$A(x) = \{x + X_n \quad \text{reaches } a \text{ before reaching } \ b\}.$$

Of course $u(a) = 1$ and $u(b) = 0$. For $a < x < b$ we have

$$u(x) = P(A(x)) = P(A(x)|X_1 = 1)p + P(A(x)|X_1 = -1)q.$$

The conditional probability $P(A(x)|X_1 = 1)$ is the probability that X_n reaches a before reaching b for the walk starting at x, given that at the first step the walk jumped to $x+1$. From the way our random walk behaves it is clear that this probability is just the probability to reach a before reaching b for the walk starting at $x+1$. Similar argument holds for $P(A(x)|X_1 = -1)$, so that

$$P(A(x)|X_1 = 1) = u(x+1), \quad P(A(x)|X_1 = -1) = u(x-1).$$

Hence $u(x)$ satisfies (4) with $f(x) = 0$ for $a < x < b$ and $f(a) = 1$, $f(b) = 0$. Because of the uniqueness aspect of Lemma 15, to find $P(A) = u(0)$ it suffices to find a solution of (4) for this f.

First we find some solutions of the middle equation in (4). We try $u(x) = \lambda^x$, where λ is a parameter. Then, we get

$$\lambda^x = p\lambda^{x+1} + q\lambda^{x-1}, \quad 1 = p\lambda + q\lambda^{-1},$$

which is a quadratic equation in λ. For $p \neq 1/2$ this equation has two distinct solutions

$$1 \quad \text{and} \quad q/p.$$

The middle equation in (4) is homogeneous. Thus, for any constants c and d, it is satisfied by the function

$$v(x) = c1^x + d(q/p)^x.$$

We want to find c and d, so that v also satisfies the boundary conditions in (4):

$$c + d(q/p)^a = 1, \quad c + d(q/p)^b = 0.$$

The solution is

$$c = -\frac{(q/p)^b}{(q/p)^a - (q/p)^b}, \quad d = \frac{1}{(q/p)^a - (q/p)^b}.$$

Then $P(A_{a,b}) = c + d$, which for $p \neq 1/2$ is seen to be in agreement with our assertion.

If $p = 1/2$, then 1 is a double root of the above quadratic equation. What helps is that there is one more solution of the middle equation in (4): $u(x) = x$. Now, again the function

$$v(x) = c + dx$$

satisfies the middle equation in (4) for any constants c and d. This function satisfies the boundary conditions if

$$c + da = 1, \quad c + db = 0.$$

We find

$$c = b/(b-a), \quad d = -1/(b-a).$$

Thus,

$$u(x) = (b-x)/(b-a) \quad \text{and} \quad P(A_{a,b}) = u(0) = c = b/(b-a).$$

The theorem is proved.

17. Theorem. *Let $(X_n, n = 0, 1, 2, ...)$ be a simple random walk on $\mathbb{Z}$ with $X_0 = 0$ and $P(X_1 = 1) = p \in (0,1)$. Let A_a be the event that eventually X_n reaches a, where $a \in \mathbb{Z}$, $a < 0$. Then*

$$P(A_a) = \begin{cases} (q/p)^{|a|} & \text{if} \quad p > 1/2, \\ 1 & \text{if} \quad p \leq 1/2. \end{cases}$$

Proof. Obviously

$$A_a = \bigcup_{b \geq 1} A_{a,b} \quad \text{and} \quad A_{a,b} \subset A_{a,c} \quad \text{if} \quad c \geq b.$$

Therefore,

$$P(A_a) = \lim_{b \to \infty} P(A_{a,b})$$

and we easily get the result from Theorem 16. The theorem is proved.

There is an interesting issue related to Theorem 17. If $p = 1/2$, then with probability 1 X_n reaches any point on $(-\infty, 0)$ and, by symmetry, on $(0, \infty)$ as well. However, it turns out that the expected time for X_n to reach even -1 is infinite. This is derived from the following theorem by letting $b \to \infty$ and observing that the time needed to reach $a < 0$ is larger than the time needed to exit from (a, b) for every $b > 0$.

18. Theorem. *Let $(X_n, n = 0, 1, 2, ...)$ be a simple symmetric random walk on $\mathbb{Z}$ with $X_0 = 0$. Let*

$$\tau = \inf\{n \geq 0 : X_n = a \quad \text{or} \quad X_n = b\},$$

where $a, b \in \mathbb{Z}$, $a < 0, b > 0$. *Then,* $E\tau = b|a|$.

Proof. For $x \in [a, b] \cap \mathbb{Z}$ define

$$\tau(x) = \inf\{n \geq 0 : x + X_n = a \quad \text{or} \quad x + X_n = b\}, \quad u(x) = E\tau(x).$$

Obviously $\tau(a) = \tau(b) = 0$, so that $u(a) = u(b) = 0$. For $a < x < b$,

$$u(x) = (1/2)E(\tau(x)|X_1 = 1) + (1/2)E(\tau(x)|X_1 = -1).$$

Here the conditional expectation $E(\tau(x)|X_1 = 1)$ is the expected exit time of the walk staring at x from (a, b), given that at the first step the walk arrived at $x+1$. Again the way our walk evolves shows that this expected exit time equals 1 (= the time to go from x to $x + 1$) plus the expected exit time of the walk starting at $x+1$. We conclude

$$E(\tau(x)|X_1 = 1) = 1 + u(x + 1), \quad E(\tau(x)|X_1 = -1) = 1 + u(x - 1).$$

Therefore,

$$u(x) = 1 + (u(x + 1) + u(x - 1))/2$$

if $a < x < b$ and, as have been noticed above, $u(a) = u(b) = 0$. By a straightforward computation, one proves that $(x - a)(b - x)$ is a solution of this system. By uniqueness (Lemma 15), u is indeed equal to $(x - a)(b - x)$. It only remains to use $E\tau = u(0)$. The theorem is proved.

2. Discrete time and space Markov chains

The purpose of this section is to introduce a type of random sequences of which random walks are special examples. Let $\mathbb{S}$ be a countable set. For instance, $\mathbb{S}$ can be any finite set, such as a finite set of points on a circle, or the set of all positive integers, or the set $\mathbb{S}^2$ of all points in $\mathbb{R}^2$ having integer coordinates, but not the set $\mathbb{R}$ of all real numbers.

Some might want to study the examples in Section 3 in conjunction with their reading of this section.

1. Definition (Markov chain). Let $(Z_0, Z_1, Z_2, \dots)$ be a sequence of $\mathbb{S}$-valued random variables. We say that $(Z_n, n = 0, 1, 2, \dots)$ is a *Markov chain in discrete time* if, for arbitrary $n \geq 0$ and $z_0, z_1, \dots, z_{n+1} \in$

$\mathbb{S}$, the *Markov property* holds:

$$P(Z_{n+1} = z_{n+1} \mid Z_n = z_n, \ldots, Z_0 = z_0) = P(Z_{n+1} = z_{n+1} \mid Z_n = z_n)\,, \tag{1}$$

whenever

$$P(Z_n = z_n, \ldots, Z_0 = z_0) > 0.$$

The set $\mathbb{S}$ is called the *state space* of the Markov chain $(Z_n, n = 0, 1, 2, ...)$ and the members of $\mathbb{S}$ are the *states*. The distribution of Z_0 is called the *initial distribution*. If $P(Z_0 = z) = 1$ for some z, we speak of a Markov chain *starting at* z and call z the *starting state*. In order to explicitly mention the state space, a phrase such as "a Markov chain $(Z_k, k \geq 0)$ on $\mathbb{S}$" is used.

Formula (1) expresses the "memoryless" property of the chain: To determine the conditional distribution of Z_{n+1} we only need to know where Z_n is and may forget which path led the chain into the state z_n.

Notice that in (1) the points $z_0, z_1, \ldots, z_{n+1}$ are just arbitrary $n + 2$ points in $\mathbb{S}$, so one can restate (1) in the following way: For every $n = 0, 1, 2, \ldots,\ z_0, z_1, \ldots, z_n, c \in \mathbb{S}$ we have

$$P(Z_{n+1} = c \mid Z_n = z_0, Z_{n-1} = z_1, \ldots, Z_0 = z_n)$$
$$= P(Z_{n+1} = c \mid Z_n = z_0),$$

whenever $P(Z_n = z_0, Z_{n-1} = z_1, \ldots, Z_0 = z_n) > 0$.

To make some writing shorter it is convenient to introduce the following notation. For $n = 0, 1, 2, \ldots$ and $k = 0, 1, \ldots, n$ and points $z_i \in \mathbb{S}$ define

$$z_{[k,n]} = (z_k, \ldots, z_n)\,.$$

In particular, $z_{[n,n]}$ is another name for z_n. Also set

$$Z_{[k,n]} = (Z_k, \ldots, Z_n).$$

Then the Markov property becomes

$$P(Z_{n+1} = z_{n+1} \mid Z_{[0,n]} = z_{[0,n]}) = P(Z_{n+1} = z_{n+1} \mid Z_n = z_n)$$

whenever $P(Z_{[0,n]} = z_{[0,n]}) > 0$. For sets $C_k, ..., C_n \subset \mathbb{S}$ we also write

$$C_{[k,n]} = C_k \times ... \times C_n = \{z_{[k,n]} : z_k \in C_k, ..., z_n \in C_n\}.$$

The following problem has a continuous state space version stated as Problem 5.1.3.

2. Problem*. Let $Z_0, Z_1, \ldots$ be a sequence of $\mathbb{S}$-valued random variables and let $p_1(x,y),\ p_2(x,y), \ldots$ be a sequence of real-valued functions on $\mathbb{S}\times\mathbb{S}$. Assume that, for each $n \geq 0$ and $c, z_0, \ldots, z_n \in \mathbb{S}$ we have

$$P(Z_{n+1} = c \mid Z_{[0,n]} = z_{[0,n]}) = p_{n+1}(z_n, c)$$

whenever $P(Z_{[0,n]} = z_{[0,n]}) > 0$. Then prove that $(Z_0, Z_1, \ldots)$ is a Markov chain. (Hint: Use Problem 1.3.5.)

Throughout the rest of this section $(Z_n, n \geq 0)$ *is a discrete-time Markov chain on* $\mathbb{S}$.

For $n = 0, 1, 2, \ldots,$ let π^n denote the probability distribution of Z_n; thus,

$$\pi^n(z) = P(Z_n = z)$$

and π^0 is the initial distribution. Also introduce functions $p_{n+1}(z \mid c)$ on $\mathbb{S}\times\mathbb{S}$ such that

$$P(Z_{n+1} = z, Z_n = c) = p_{n+1}(z \mid c)P(Z_n = c), \quad n = 0, 1, \ldots, c, z \in \mathbb{S}. \tag{2}$$

Such functions $p_{n+1}(z \mid c)$ exist and one example of them is given by

$$P(Z_{n+1} = z \mid Z_n = c).$$

By the way, observe that if $P(Z_n = c) > 0$, then necessarily

$$p_{n+1}(z \mid c) = P(Z_{n+1} = z \mid Z_n = c).$$

3. Problem*. Prove that for all $n = 1, 2, \ldots$ and $z_0, z_1, \ldots \in \mathbb{S}$

$$P(Z_n = z_n, \ldots, Z_0 = z_0) = p_n(z_n \mid z_{n-1}) \cdot \ldots \cdot p_1(z_1 \mid z_0)\pi^0(z_0).$$

4. Problem*. On the basis of Problem 3 prove that for integers $m > n \geq 0$ and $z_0, z_1, \ldots \in \mathbb{S}$

$$\begin{aligned} &P(Z_{[n+1,m]} = z_{[n+1,m]} \mid Z_{[0,n]} = z_{[0,n]}) \\ &= p_m(z_m \mid z_{m-1}) \cdot \ldots \cdot p_{n+1}(z_{n+1} \mid z_n) \end{aligned}$$

if $P(Z_{[0,n]} = z_{[0,n]}) > 0$. Then by using Problem 1.3.5 conclude that

$$P(Z_{[n+1,m]} = z_{[n+1,m]} \mid Z_n = z_n) = p_m(z_m \mid z_{m-1}) \cdot \ldots \cdot p_{n+1}(z_{n+1} \mid z_n)$$

if $P(Z_n = z_n) > 0$.

5. Problem*. In the setting of Problem 4 with $m > n > 0$, prove that

$$P(Z_{[n+1,m]} = z_{[n+1,m]}, Z_n = z_n, Z_{[0,n-1]} = z_{[0,n-1]}) \\ = P(Z_{[n+1,m]} = z_{[n+1,m]} \mid Z_n = z_n) \times \\ P(Z_{[0,n-1]} = z_{[0,n-1]} \mid Z_n = z_n)\pi^n(z_n). \quad (3)$$

(Hint: The first term on the right is found in Problem 4 and the product of remaining terms in Problem 3 due to

$$P(Z_{[0,n-1]} = z_{[0,n-1]} \mid Z_n = z_n)\pi^n(z_n) \\ = P(Z_{[0,n-1]} = z_{[0,n-1]}, Z_n = z_n). \quad)$$

6. Theorem. *For all integers $m > n > 0$, points $z_n \in \mathbb{S}$, and sets $C_0, C_1, ... \subset \mathbb{S}$ we have*

$$P(Z_{[n+1,m]} \in C_{[n+1,m]}, Z_n = z_n, Z_{[0,n-1]} \in C_{[0,n-1]}) \\ = P(Z_{[n+1,m]} \in C_{[n+1,m]} \mid Z_n = z_n) \times \\ P(Z_{[0,n-1]} \in C_{[0,n-1]} \mid Z_n = z_n)\pi^n(z_n). \quad (4)$$

This theorem follows immediately from Problem 5 and the observation that the left-hand side of (4) equals the sum over

$$z_{[0,n-1]} \in C_{[0,n-1]} \quad \text{and} \quad z_{[n+1,m]} \in C_{[n+1,m]}$$

of the left-hand sides of (3).

By combining the last two terms on the right in (4) we arrive at the following.

7. Corollary. *Under the conditions of Theorem* 6

$$P(Z_{[n+1,m]} \in C_{[n+1,m]} \mid Z_n = z_n, Z_{[0,n-1]} \in C_{[0,n-1]}) \\ = P(Z_{[n+1,m]} \in C_{[n+1,m]} \mid Z_n = z_n)$$

provided that $P(Z_n = z_n, Z_{[0,n-1]} \in C_{[0,n-1]}) > 0$. In particular, when some of C_i's coincide with $\mathbb{S}$ and the others are singletons, for any $z_0, z_1, ... \in \mathbb{S}$ and $k = 0, ..., n$

$$P(Z_m \in C_m \mid Z_{[k,n]} = z_{[k,n]}) = P(Z_m \in C_m \mid Z_n = z_n)$$

if $P(Z_{[k,n]} = z_{[k,n]}) > 0$. In any case

$$P(Z_{[n+1,m]} \in C_{[n+1,m]}, Z_n = z_n, Z_{[0,n-1]} \in C_{[0,n-1]})$$

$$= P(Z_{[n+1,m]} \in C_{[n+1,m]} \mid Z_n = z_n) P(Z_n = z_n, Z_{[0,n-1]} \in C_{[0,n-1]}),$$

$$P(Z_m \in C_m, Z_{[k,n]} = z_{[k,n]})$$
$$= P(Z_m \in C_m \mid Z_n = z_n) P(Z_{[k,n]} = z_{[k,n]}).$$

By combining the last two and the first terms on the right in (4) and using the definition of p_{n+1} we obtain another important formula.

8. Corollary. *For $n \geq 1$, $C_0, C_1, ... \subset \mathbb{S}$ and $z_n, z_{n+1} \in \mathbb{S}$*

$$P(Z_{n+1} = z_{n+1}, Z_n = z_n, Z_{[0,n-1]} \in C_{[0,n-1]})$$
$$= P(Z_{n+1} = z_{n+1}, Z_n = z_n) P(Z_{[0,n-1]} \in C_{[0,n-1]} \mid Z_n = z_n)$$
$$= p_{n+1}(z_{n+1} \mid z_n) P(Z_{[0,n-1]} \in C_{[0,n-1]} \mid Z_n = z_n) \pi^n(z_n)$$
$$= p_{n+1}(z_{n+1} \mid z_n) P(Z_{[0,n-1]} \in C_{[0,n-1]}, Z_n = z_n).$$

9. Problem. Prove that for all integers $m > n > 0$ the random variables $Z_{[0,n-1]}$ and $Z_{[n+1,m]}$ are conditionally independent given Z_n in the sense that for any sets $C_0, C_1, ... \subset \mathbb{S}$ and $z_n \in \mathbb{S}$

$$P(Z_{[n+1,m]} \in C_{[n+1,m]}, Z_{[0,n-1]} \in C_{[0,n-1]} \mid Z_n = z_n)$$
$$= P(Z_{[n+1,m]} \in C_{[n+1,m]} \mid Z_n = z_n) P(Z_{[0,n-1]} \in C_{[0,n-1]} \mid Z_n = z_n).$$

An instructive way of paraphrasing this result is to say: If a random sequence is Markov then the past and future are conditionally independent given the present.

The result of the following problem is used in interpolation of Markov chains.

10. Problem*. In the setting of Theorem 6 prove that

$$P(Z_{[0,n-1]} \in C_{[0,n-1]} \mid Z_n = z_n, Z_{[n+1,m]} \in C_{[n+1,m]})$$
$$= P(Z_{[0,n-1]} \in C_{[0,n-1]} \mid Z_n = z_n)$$

if $P(Z_n = z_n, Z_{[n+1,m]} \in C_{[n+1,m]}) > 0$. Also take sets $D_0, D_1, ... \subset \mathbb{S}$. Assume that

$$D_n = \{z_n\}$$

and prove that

$$P(Z_{[0,n-1]} \in C_{[0,n-1]} \mid Z_{[0,m]} \in D_{[0,m]})$$
$$= P(Z_{[0,n-1]} \in C_{[0,n-1]} \mid Z_{[0,n]} \in D_{[0,n]})$$

if $P(Z_{[0,m]} \in D_{[0,m]}) > 0$, which means that conditioning on the future is irrelevant if the present state z_n is given.

Observe that any function p_{n+1} satisfying (2) is nonnegative and

$$\sum_{z \in \mathbb{S}} p_{n+1}(z \mid c) = 1$$

if $P(Z_n = c) > 0$.

11. Definition (**Transition function**). A function $p : \mathbb{S} \times \mathbb{S} \to [0,1]$ is called a *transition function* for $\mathbb{S}$ if

$$\sum_{z \in \mathbb{S}} p(c, z) = 1 \tag{5}$$

for every $c \in \mathbb{S}$.

12. Theorem. *For every $n = 0, 1, 2, \ldots$ there exist transition functions $p_{n+1}(c, z)$ such that the function*

$$p_{n+1}(z \mid c) := p_{n+1}(c, z)$$

satisfies (2).

Proof. For those c for which $P(Z_n = c) > 0$, the function

$$p_{n+1}(c, z) = P(Z_{n+1} = z \mid Z_n = c)$$

satisfies (2) and (5). If c is such that $P(Z_n = c) = 0$, there are infinitely many choices for $p_{n+1}(c, z)$ and each one will satisfy (2). To get (5) satisfied one can set $p_{n+1}(c, z) = \pi_0(z)$. The theorem is proved.

13. Definition. Any function $p_{n+1}(c, z)$ from Theorem 12 is called an $n+1$*st transition function* of $(Z_k, k = 0, 1, 2, \ldots)$.

14. Definition (**Time-homogeneity**). The Markov chain

$$(Z_n, n = 0, 1, 2, \ldots)$$

is *time-homogeneous* if there exists a transition function p (independent of n) such that

$$P(Z_{n+1} = z_{n+1} \mid Z_n = z_n) = p(z_n, z_{n+1})$$

whenever $P(Z_n = z_n) > 0$. In this case p is called a *transition function* of the time-homogeneous Markov chain $(Z_n, n = 0, 1, 2, \ldots)$.

In terms of Definition 14 the result of Problem 4 implies the following.

15. Theorem. *Let* $(Z_k, k \geq 0)$ *be a time-homogeneous Markov chain having transition function* p*. Then for any integers* $m > n \geq 0$ *and* $z_0, z_1, ... \in \mathbb{S}$

$$P(Z_{[n,m]} = z_{[n,m]}) = \pi^n(z_n) \prod_{i=n}^{m-1} p(z_i, z_{i+1}) .$$

In particular, if $P(Z_n = z_n) > 0$*, then*

$$P(Z_{[n,m]} = z_{[n,m]} \mid Z_n = z_n) = \prod_{i=n}^{m-1} p(z_i, z_{i+1}) .$$

The following example shows that some care is needed when events of probability 0 are involved.

16. Example. Let $\mathbb{S} = \{-1, 1\}$ and define $\hat{p}$ and $\tilde{p}$ on $\mathbb{S} \times \mathbb{S}$ by

$$\hat{p}(c, z) = \begin{cases} 1 & \text{if } z = -1 \\ 0 & \text{if } z = 1 \end{cases}$$

and

$$\tilde{p}(c, z) = \begin{cases} 1 & \text{if } z = c \\ 0 & \text{if } z \neq c \, . \end{cases}$$

It is trivial to check that both $\hat{p}$ and $\tilde{p}$ are transition functions. Let $(Z_0, Z_1, Z_2 \dots)$ be the sequence $(-1, -1, -1, \dots)$. Then it is obvious that $(Z_0, Z_1, Z_2, \dots)$ is a Markov chain with state space $\mathbb{S}$; moreover, it is time-homogeneous, as can be checked by using either of the transition functions $\hat{p}$ or $\tilde{p}$.

As in the above example, generally, it may happen that the Markov chain $(Z_n, n = 0, 1, 2, ...)$ actually lives on a set smaller than $\mathbb{S}$.

17. Problem. For an $\mathbb{S}$-valued Markov chain $(Z_n, n = 0, 1, 2, ...)$, denote

$$\tilde{\mathbb{S}} = \{z \in \mathbb{S} : P(Z_n = z) > 0 \quad \text{for some} \quad n \geq 0\}$$

and call $\tilde{\mathbb{S}}$ *the effective phase space of* $(Z_n, n = 0, 1, 2, ...)$. Prove that $(Z_n, n = 0, 1, 2, ...)$ is an $\tilde{\mathbb{S}}$-valued Markov chain having a unique transition function.

18. Problem. Prove that random walks are time-homogeneous Markov chains with transition function satisfying $p(a, z) = p(0, z - a)$.

In the following problem we describe a simple case when the observations Y_n of a Markov chain $(X_n, n \geq 0)$ are "corrupted" by noises W_n.

19. Problem. Let $(X_n, n = 0, 1, 2, \dots)$ be a time-homogeneous Markov chain with state space $\mathbb{S}$ and transition function $p(a, x)$. Let $\gamma(x, w)$ be an $\mathbb{S}$-valued function defined on $\mathbb{S} \times \mathbb{S}$. Finally, let W_0, $W_1, \dots$ be a sequence of independent identically distributed $\mathbb{S}$-valued discrete random variables. Set

$$q(a, w) = P(\gamma(a, W_0) = w)$$

and assume that

$$(X_0, \dots, X_n) \quad \text{and} \quad (W_0, \dots, W_n)$$

are independent for each n. Set

$$Y_n = \gamma(X_n, W_n)$$

and show that $\big((X_n, Y_n), n \geq 0\big)$ is a time-homogeneous Markov chain with state space $\mathbb{S} \times \mathbb{S}$ and express its transition function in terms of the transition function p and the function q. (Hint: Use the formula

$$P(X_{n+1} = x_{n+1}, Y_{n+1} = y_{n+1} \mid X_{[0,n]} = x_{[0,n]}, Y_{[0,n]} = y_{[0,n]})$$

$$= P(X_{n+1} = x_{n+1}, \gamma(x_{n+1}, W_{n+1}) = y_{n+1} \mid A),$$

where A is the event that

$$X_{[0,n]} = x_{[0,n]}, \gamma(x_n, W_n) = y_n, ..., \gamma(x_0, W_0) = y_0,$$

and use Problem 1.3.4.)

Generally, if (X_n, Y_n) is constructed as in Problem 19, the sequence $(Y_n, n \geq 0)$ alone need not have the Markov property (see Problem 3.9).

In the following problem we discuss a way to define Markov chains by means of "stochastic equations".

20. Problem. Let $(X_n, n = 0, 1, 2, \dots)$ be a simple random walk on $\mathbb{Z}$ starting at zero, and let $V_0, V_1, \dots$ be independent identically distributed random variables with values in $\{1, -1\}$. Moreover, assume that the $(n+1)$-tuples

$$(X_0, \dots, X_n) \quad \text{and} \quad (V_0, \dots, V_n)$$

are independent for each n. Let $Y_0 = 0$ and for $n \geq 1$, set

$$Y_n = Y_{n-1} + V_n(X_n - X_{n-1}).$$

Prove that $\big((X_n, Y_n), n = 0, 1, 2, \dots\big)$ is a time-homogeneous Markov chain on $\mathbb{Z}^2$ and express its transition function in terms of

$$\alpha = P(V_0 = 1) \quad \text{and} \quad \beta = P(X_1 - X_0 = 1).$$

Also prove that $(Y_n, n = 0, 1, 2, ...)$ is a simple random walk on $\mathbb{Z}$. (Hint: Follow the hint to Problem 19).

Markov chains have initial distributions and transition functions. A natural question arises: How arbitrary can these objects be?

21. Theorem. *Let functions $\pi^0(z)$ and $p(c, z)$ given for $c, z \in \mathbb{S}$ satisfy*

$$\pi^0(z) \geq 0, \quad p(c, z) \geq 0, \quad \sum_{b \in \mathbb{S}} \pi^0(b) = 1, \quad \sum_{b \in \mathbb{S}} p(c, b) = 1$$

for all $c, z \in \mathbb{S}$. Then there exists a time-homogeneous Markov chain with transition function p and initial distribution π^0. In particular, for any $z_0 \in \mathbb{S}$ there exists a Markov chain with transition p function starting at z_0.

To explain why Theorem 21 is true, we generate $\mathbb{S}$-valued random variables Y_0 and $Y_n(c)$ for all $c \in \mathbb{S}$ and $n = 1, 2, 3, ...$ in such a way that they are mutually independent and satisfy

$$P(Y_0 = z) = \pi^0(z), \quad P(Y_n(c) = z) = p(c, z)$$

for all $c, z \in \mathbb{S}$ and $n = 1, 2, 3....$ Then we define recursively

$$Z_0 = Y_0, \quad Z_n = Y_n(Z_{n-1}), \quad n = 1, 2, 3....$$

If we believe that such $Y_0, Y_n(c)$ exist, then $P(Z_0 = z) = \pi^0(z)$ and for $P(Z_{[0,k]} = z_{[0,k]}) > 0$ we have

$$P(Z_{k+1} = z | Z_{[0,k]} = z_{[0,k]}) = P(Y_{k+1}(Z_k) = z | Z_{[0,k]} = z_{[0,k]})$$

$$= \frac{P(Y_{k+1}(Z_k) = z, Z_{[0,k]} = z_{[0,k]})}{P(Z_{[0,k]} = z_{[0,k]})} = \frac{P(Y_{k+1}(z_k) = z, Z_{[0,k]} = z_{[0,k]})}{P(Z_{[0,k]} = z_{[0,k]})}$$

$$= \frac{P(Y_{k+1}(z_k) = z, A)}{P(A)},$$

where A is an event (that $Z_{[0,k]} = z_{[0,k]}$) which is expressed in terms of the variables $Y_0, Y_1(u), ..., Y_k(u)$ for $u \in \mathbb{S}$. By independence, the last fraction is

$$P(Y_{k+1}(z_k) = z) = p(z_k, z),$$

which is exactly what we needed as follows from Problem 2.

The above construction gives an idea of how Markov chains behave. First, one generates the initial state Z_0, then one uses a random mechanism independent of Z_0 to generate the next step in accordance with the transition function, and then again one uses a random mechanism independent of the previous ones to generate the next step, again in accordance with the transition function, and so on.

22. Definition. For a transition function p, define recursively p^{*k} for $k = 1, 2, \ldots$ by $p^{*1} = p$ and

$$p^{*k}(c, z) = \sum_w p^{*(k-1)}(c, w)\, p(w, z)\,, \quad k \geq 2\,.$$

The function p^{*k} is called the *k-step transition function* of the Markov chain $(Z_n, n \geq 0)$ if the chain is time homogeneous and p is its transition function.

23. Problem*. Prove that for $k \geq 2$

$$p^{*k}(z_0, z_k) = \sum_{z_1, \ldots, z_{k-1}} p(z_0, z_1) p(z_1, z_2) \cdot \ldots \cdot p(z_{k-1}, z_k).$$

24. Problem. Prove that

$$p^{*k}(c, z) = \sum_w p(c, w)\, p^{*(k-1)}(w, z)\,, \quad k \geq 2\,.$$

25. Problem. Let $(Z_n, n \geq 0)$ be a time-homogeneous Markov chain with initial distribution π^0 and transition function p. For each positive integer k, prove that $(Z_0, Z_k, Z_{2k}, Z_{3k}, \ldots)$ is a time-homogeneous Markov chain with initial distribution π^0 and (one-step) transition function p^{*k}.

The following theorem allows one to find the distribution of Z_n from the initial distribution and transition function. Recall that

$$\pi^n(z) = P(Z_n = z).$$

26. Theorem. *Let $(Z_n, n = 0, 1, 2, ...)$ be a time-homogeneous Markov chain with initial distribution π^0 and transition function p. Then, for any $n \geq 1$ and $m \geq 0$, we have*

$$\pi^n(y) = \sum_{z \in \mathbb{S}} \pi^{n-1}(z) p(z, y) = \sum_{z \in \mathbb{S}} \pi^0(z) p^{*n}(z, y), \tag{6}$$

$$P(Z_{m+n} = y \mid Z_m = x) = p^{*n}(x, y) \tag{7}$$

if $P(Z_m = x) > 0$. In particular, if, for a fixed $z \in \mathbb{S}$, we have $P(Z_0 = z) = 1$, then $\pi^n(y) = p^{*n}(z, y)$.

Proof. Equation (7) follows immediately from Theorem 15 and Problem 23. Next, we have

$$\begin{aligned} P(Z_n = y) &= \sum_{z \in \mathbb{S}} P(Z_n = y \mid Z_{n-1} = z) P(Z_{n-1} = z) \\ &= \sum_{z \in \mathbb{S}} p(z, y) \pi^{n-1}(z), \end{aligned}$$

which proves the first equality in (6). Finally,

$$P(Z_n = y) = \sum_{z \in \mathbb{S}} P(Z_n = y \mid Z_0 = z) P(Z_0 = z)$$

and since $P(Z_0 = z) = \pi^0(z)$ the second equality in (6) follows from (7). The theorem is proved.

27. Remark. If $\mathbb{S}$ is a finite set, it is sometimes convenient to identify it with $\{1, 2, ..., d\}$, set

$$\pi_i^n := \pi^n(i), \quad p_{ij} := p(i, j), \quad p_{ij}^k := p^{*k}(i, j),$$

and interpret $\pi^n = (\pi_1^n, ..., \pi_d^n)$ as a d-dimensional *row vector* (or *row matrix*) and both $p = (p_{ij})$ and $p^k = (p_{ij}^k)$ as d-by-d matrices. In view of this discussion one sometimes uses the terms *transition matrix, initial probability vector,* and *probability vector at time n.*

The notation p^k for the matrix (p_{ij}^k) would not be legitimate because the powers of matrices are well defined already, if the following were not true.

28. Problem. Using the above notation explain how the following equalities follow from earlier results and definitions in this chapter:

$$p^k = p^{k-1}\,p = p\,p^{k-1}\,, \quad k \geq 2\,;$$

$$\pi^n = \pi^{n-1}\,p = \pi^0 p^n\,, \quad n \geq 1\,. \tag{8}$$

29. Example. Let $d = 5$, $\pi^0 = (1/5, 1/5, 1/5, 1/5, 1/5)$, and

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

Then

$$\pi^1 = (1/5, 1/5, 1/5, 1/5, 1/5) \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}$$

$$= (1/5, 1/5, 1/5, 1/5, 1/5)\,.$$

By induction we get that

$$\pi^n = (1/5, 1/5, 1/5, 1/5, 1/5)$$

for all $n \geq 0$.

30. Problem. Calculate p^2 and p^3 for Example 29.

Quite often one is interested in knowing the distribution of Z_n for large n. According to (8), to find π^n we need to multiply π^0 by the n-th power of p, computing which may present considerable difficulties. Sometimes it is much more practical to approximate π^n for large n by its limit π as $n \to \infty$, provided that the limit exists. It turns out that this limit π exists in very many situations and can be computed relatively easily. Of course since $\pi^n = \pi^{n-1}p$, π should satisfy $\pi = \pi p$.

31. Definition (**Invariant distribution**)**.** Let π be a probability distribution on $\mathbb{S}$. We call it *an invariant* or *equilibrium* distribution for Z_n if

$$\pi(z) = \sum_{c \in \mathbb{S}} \pi(c) p(c, z). \tag{9}$$

In matrix notation (cf. Remark 27) equation (9) means that $\pi = \pi p$.

Observe that if the initial distribution π^0 is invariant, then for all $n = 0, 1, 2, ...$ the distribution π^n of Z_n coincides with π^0, so that all $Z_0, Z_1, Z_2, ...$ have the same distribution. In this case Z_n, $n = 0, 1, 2, ...$, is also a stationary process (see Problem 7.2.19), predicting which is one of our purposes.

In Example 29 the initial distribution π^0 is invariant and, as we will see later, this is the only invariant distribution. Generally, invariant distributions are not unique.

32. Example. For $\mathbb{S} = (1, 2, ..., d)$, the function δ_{ij} is a transition function. For any Markov chain with this transition function, any probability distribution is invariant.

33. Theorem. *In the situation of Remark* 27 *assume that there exists* $m \geq 1$ *such that all entries of* p^m *are strictly positive. Then there is a unique invariant distribution* $\pi = (\pi_1, ..., \pi_d)$ *and*

$$\lim_{n \to \infty} \pi_j^n = \pi_j, \quad \lim_{n \to \infty} p_{ij}^n = \pi_j \quad \forall i, j = 1, ..., d. \tag{10}$$

Proof. First we prove that the invariant distribution is unique. In our situation equation (9) means that $\pi = \pi p$. By multiplying both sides by p^{m-1} from the right, we get

$$\pi = \pi q, \tag{11}$$

where $q = p^m$. Let $\varepsilon > 0$ be such that $q_{ij} \geq \varepsilon$. By Problem 25 the function q is a transition function, so that

$$\sum_{i=1}^{d} q_{ji} = 1 \quad \forall j.$$

Now assume that we have two invariant distributions π and μ. Then $\sum_j \pi_j = \sum_j \mu_j = 1$ and

$$\sum_{i=1}^{d} |\pi_i - \mu_i| = \sum_{i=1}^{d} \big|\sum_{j=1}^{d} (\pi_j - \mu_j) q_{ji}\big| = \sum_{i=1}^{d} \big|\sum_{j=1}^{d} (\pi_j - \mu_j)(q_{ji} - \varepsilon)\big|$$

$$\leq \sum_{j=1}^{d} \sum_{i=1}^{d} |\pi_j - \mu_j| (q_{ji} - \varepsilon) = \alpha \sum_{j=1}^{d} |\pi_j - \mu_j|,$$

where $\alpha := 1 - \varepsilon d < 1$. By comparing the extreme terms, we get $\pi = \mu$.

Similarly, for any initial distribution π^0, by denoting $\pi^n = \pi^0 p^n$ and observing that $\pi^{n+1} = \pi^{n+1-k} q$ we obtain

$$\sum_{i=1}^{d} |\pi_i^{n+1} - \pi_i^n| \leq \alpha \sum_{i=1}^{d} |\pi_i^{n+1-k} - \pi_i^{n-k}|$$

$$\leq \alpha^2 \sum_{i=1}^{d} |\pi_i^{n+1-2k} - \pi_i^{n-2k}| \leq ... \leq \alpha^{[n/k]} \sum_{i=1}^{d} |\pi_i^{n+1-[n/k]k} - \pi_i^{n-[n/k]k}|$$

$$\leq \alpha^{[n/k]} 2 \leq \beta^n 2 \alpha^{-1},$$

where $[n/k]$ is the integer part of n/k and $\beta := \alpha^{1/k} < 1$. By Cauchy's criterion it follows that the sequence of vectors π^n converges as $n \to \infty$. By denoting π its limit and using $\pi^{n+1} = \pi^n p$, we see that $\pi = \pi p$, so that π is a (unique) invariant distribution.

This proves the first assertion in (10). By the above uniqueness of solutions to (11) the limit of π^n is independent of the initial distribution and to prove the second assertion in (10) it suffices to consider a Markov chain starting at i, for which $\pi_j^n = p_{ij}^n$. The theorem is proved.

34. Problem. Prove that the conditions of Theorem 33 are satisfied if and only if for any i, j and Markov chain with transition matrix p starting at i, $P(X_m = j) > 0$.

In Problems 35 and 36 random walks with absorption and reflection are, actually, not random walks. They are Markov chains.

35. Problem (Random walks with absorption)**.** Sometimes even if the conditions of Theorem 33 are not satisfied, the chain has a unique invariant distribution. Let $\mathbb{S} = \mathbb{Z}$ and for $i \neq 1$ let $p(i, i \pm 1) = 1/2$ and let $p(1,1) = 1$. Prove that there exists a unique invariant distribution and find it.

36. Problem (Simple random walk with reflection)**.** Fix $d \geq 2$ and $\mathbb{Z} = \{1, ..., d\}$. For $1 < i < d$, let $p(i, i \pm 1) = 1/2$ and let $p(1,1) = p(1,2) = p(d, d-1) = p(d,d) = 1/2$. Prove that all entries of the matrix p^{d-1} are strictly positive. In case $d \geq 3$, find $p^{d-2}_{1,d}$.

37. Problem. Find the invariant distribution in Problem 36.

38. Problem. Show that random walks on $\mathbb{Z}^d$, which are nontrivial in the sense that $P(Z_1 \neq Z_0) > 0$, do not have invariant probability distributions. (Hint: If we start off from an invariant distribution, then Z_0 and Z_1 have the same distribution and $Z_1 = Z_0 + (Z_1 - Z_0)$, where the summands are independent. It follows that

$$E \exp(i\lambda \cdot Z_0) = E \exp(i\lambda \cdot Z_0) E \exp(i\lambda \cdot (Z_1 - Z_0)).\ )$$

In contrast with Problem 38, finite-state Markov chains always have invariant distributions.

39. Theorem. *Let $\mathbb{S}$ be a finite set and $(Z_n, n = 0, 1, 2, ...)$ be an $\mathbb{S}$-valued time-homogeneous Markov chain. Then it has an invariant distribution.*

Proof. Let $\mathbb{S} = (1, 2, ..., d)$. For the matrix $a = (a_{ij}) = (p_{ij} - \delta_{ij})$ it holds that the sum of all its columns is zero. Therefore, the columns are linearly dependent. By linear algebra, so are the rows. This means that there exists a row vector $\lambda \neq 0$ such that $\lambda a = 0$. In coordinate form we have

$$\sum_{i=1}^{d} \lambda_i a_{ij} = 0, \quad \text{which means} \quad \lambda_j = \sum_{i=1}^{d} \lambda_i p_{ij} \quad \forall j.$$

Normalizing if necessary, we can assume that, for $\pi_i := |\lambda_i|$, we have

$$\sum_i \pi_i = 1.$$

Obviously

$$\pi_j \le \sum_{i=1}^{d} \pi_i p_{ij} \tag{12}$$

for all j. Summing up over all j leads to

$$1 = \sum_{j=1}^{d} \pi_j \le \sum_{j=1}^{d} \sum_{i=1}^{d} \pi_i p_{ij} = \sum_{i=1}^{d} \pi_i = 1.$$

It follows that, actually, the sides in (12) are equal for all j. This proves the theorem.

40. Problem (Random walks with absorption)**.** Fix $d \ge 2$ and $\mathbb{S} = \{1, ..., d\}$. For $1 < i < d$, let $p(i, i\pm 1) = 1/2$ and let $p(1,1) = p(d,d) = 1$. Find all invariant distributions. (Hint: By using linear algebra, first find all solutions of $\mu = \mu p$.)

41. Problem. Let μ and ν be invariant distributions for the same Markov chain. Define

$$\pi(x) = c \max(\mu(x), \nu(x)) \quad \text{with} \quad c^{-1} := \sum_{x \in \mathbb{S}} \max(\mu(x), \nu(x)).$$

Prove that π is invariant as well.

3. Further problems on Markov chains

The matrix viewpoint of Remark 2.27 is not confined to settings with a finite state space. For countably infinite state spaces, one can label the states with either the positive integers or the nonnegative integers and use matrices that have a top row and a left-hand column but which go on with out terminating in the downward and rightward directions.

1. Problem. Check that the assertions in Problem 2.28 continue to hold even for countably infinite state spaces.

Also, for countably infinite state spaces one can label the states with all the integers and use matrices that have no top or bottom row and no left-hand or right-hand column. This point of view is best for the next problem.

2. Problem. For the transition matrix p of the simple symmetric random walk in $\mathbb{Z}$ find p^2 and p^3. Also find the probability vector at time 1 if the initial probability vector consists of twenty-one entries of $1/21$ in positions $-10, -9, \ldots, 9, 10$ and zeroes elsewhere.

For random walks in $\mathbb{Z}^2$, one can also use a transition matrix, but it is not very useful since, in creating a one-to-one correspondence between $\mathbb{Z}^2$ and $\{1, 2, ...\}$ or $\mathbb{Z}$, one loses the geometrical intuition associated with $\mathbb{Z}^2$.

3. Problem. Consider the Markov chain starting at 0 with a transition matrix p whose top row and left-hand column are called the zeroth row and column, whose entry in each position $(i, i+1)$ is $3/4$, whose entry in each position $(i, 0)$ is $1/4$, and whose other entries all equal 0. Find p^2. Also find the distribution of the first return time to 0. Describe the distribution of N, the number of returns to 0.

4. Problem. Modify the preceding problem by making the starting point 1 rather than 0 with a focus on returns to 1 rather than to 0. Then do the modified problem.

5. Definition. Let $(U_n, n \geq 1)$ be a sequence of independent identically distributed random variables taking values in $\mathbb{Z}^+ \cup \{\infty\}$. Set $T_0 = 0$, and then recursively define T_n for $n > 0$ by

$$T_n = T_{n-1} + U_n ,$$

where the sum of any member of $\mathbb{Z}^+ \cup \{\infty\}$ and ∞ is defined to equal ∞. The random sequence $(T_n, n \geq 0)$, or any random sequence having the same joint distributions as $(T_n, n \geq 0)$, is called a *random walk on* $\mathbb{Z}^+ \cup \{\infty\}$ starting at 0.

The reason that the preceding definition is not encompassed in the definition of random walk given earlier is that the random variables U_n cannot necessarily be recovered from the random variables T_n. For instance, if all the values of T_n are known and $T_6 = +\infty$, then none of the values of U_n for $n > 6$ can be obtained as $T_n - T_{n-1}$.

6. Problem. Modify and adapt Problem 2.18 to random walks on $\mathbb{Z}^+ \cup \{\infty\}$, as introduced in Definition 5.

7. Problem. Let $(T_n, n \geq 0)$ be a random walk in $\mathbb{Z}^+ \cup \{\infty\}$, starting at 0, as defined in Definition 5. Assume that $P(T_1 > 0) = 1$. For $k \geq 0$, set

$$Y_k = k - \max\{T_n \colon T_n \leq k\}\,.$$

Prove that $(Y_k, k \geq 0)$ is a time-homogeneous Markov chain on $\mathbb{Z}^+$ starting at 0. Find its transition function in terms of the distribution of T_1.

8. Problem. Let $(X_k, k \geq 0)$ be a time-homogeneous Markov chain starting at x_0, and let $(T_n, n \geq 0)$ be the sequence of return times to x_0. Define Y_k in terms of the T_n's as in the preceding problem. Prove that $\big((X_k, Y_k), k \geq 0\big)$ is a time-homogeneous Markov chain starting at $(x_0, 0)$.

9. Problem. Let $(X_n, n \geq 0)$ be a simple symmetric random walk on $\mathbb{Z}$ starting at the origin and set $\gamma(a, w) = a(a^2 - 1)$ so that, in the notation of Problem 2.19, $Y_n = X_n(X_n^2 - 1)$. Prove that $(Y_n, n \geq 0)$ is *not* a Markov chain. (Hint: Compare $P(Y_6 = 6 \mid Y_5 = 0)$ with $P(Y_6 = 6 \mid Y_5 = 0, Y_3 = -24)$.)

10. Problem. Fix a positive integer u and a number $p \in (0, 1)$. Let $(X_n, n \geq 0)$ be the Markov chain on $\mathbb{Z}^+$ starting at 0 and having transition function p defined by $p(0, u) = 1$ and

$$p(a, a+1) = p = 1 - p(a, a-1)\,, \quad a > 0\,.$$

Is the expected time to return to 0 finite? Calculate the probability that the first return time is finite.

11. Problem. Fix positive integers $u \leq d$ and a number $p \in (0, 1)$. Let $(X_n, n \geq 0)$ be the Markov chain on $\{0, 1, \ldots, d\}$ starting at 0 and having transition function defined by $p(0, u) = 1$, $p(d, d) = 1$, and

$$p(a, a+1) = p = 1 - p(a, a-1)\,, \quad 0 < a < d\,.$$

Calculate the expected time to return to 0 and the probability that the first return time is finite.