Chapter 17
Markov Chains

In spite of their simplicity, Markov processes with countable state space (and discrete time) are interesting mathematical objects with which a variety of real-world phenomena can be modeled. We give an introduction to the basic concepts and then study certain examples in more detail. The connection with discrete potential theory will be investigated later, in Chapter 19. Some readers might prefer to skip the somewhat technical construction of general Markov processes in Section 17.1.

There is a vast literature on Markov chains. For further reading, see, e.g., [21, 27, 64, 66, 91, 116, 123, 124, 143, 152].

17.1 Definitions and Construction

In the following, E is always a Polish space with Borel σ-algebra $\mathcal{B}(E)$, $I \subset \mathbb{R}$ and $(X_t)_{t\in I}$ is an E-valued stochastic process. We assume that $(\mathcal{F}_t)_{t\in I} = \mathbb{F} = \sigma(X)$ is the filtration generated by X.

Definition 17.1 We say that X has the *Markov property* (MP) if, for every $A \in \mathcal{B}(E)$ and all $s, t \in I$ with $s \le t$,

$$\mathbf{P}[X_t \in A \mid \mathcal{F}_s] = \mathbf{P}[X_t \in A \mid X_s].$$

Remark 17.2 If E is a countable space, then X has the Markov property if and only if, for all $n \in \mathbb{N}$, all $s_1 < \ldots < s_n < t$ and all $i_1, \ldots, i_n, i \in E$ with $\mathbf{P}[X_{s_1} = i_1, \ldots, X_{s_n} = i_n] > 0$, we have

$$\mathbf{P}[X_t = i \mid X_{s_1} = i_1, \ldots, X_{s_n} = i_n] = \mathbf{P}[X_t = i \mid X_{s_n} = i_n]. \tag{17.1}$$

In fact, (17.1) clearly implies the Markov property. On the other hand, if X has the Markov property, then (see (8.6)) $\mathbf{P}[X_t = i \mid X_{s_n}](\omega) = \mathbf{P}[X_t = i \mid X_{s_n} = i_n]$ for almost all $\omega \in \{X_{s_n} = i_n\}$. Hence, for $A := \{X_{s_1} = i_1, \ldots, X_{s_n} = i_n\}$ (using the

A. Klenke, *Probability Theory*, Universitext, DOI 10.1007/978-1-4471-5361-0_17,

Markov property in the second equation),

$$\begin{aligned}\mathbf{P}[X_t = i, X_{s_1} = i_1, \ldots, X_{s_n} = i_n] \\ = \mathbf{E}\big[\mathbf{E}[\mathbb{1}_{\{X_t=i\}} \mid \mathcal{F}_{s_n}]\mathbb{1}_A\big] = \mathbf{E}\big[\mathbf{E}[\mathbb{1}_{\{X_t=i\}} \mid X_{s_n}]\mathbb{1}_A\big] \\ = \mathbf{E}\big[\mathbf{P}[X_t = i \mid X_{s_n} = i_n]\mathbb{1}_A\big] = \mathbf{P}[X_t = i \mid X_{s_n} = i_n]\mathbf{P}[A].\end{aligned}$$

Dividing both sides by $\mathbf{P}[A]$ yields (17.1). ◊

Definition 17.3 Let $I \subset [0,\infty)$ be closed under addition and assume $0 \in I$. A stochastic process $X = (X_t)_{t\in I}$ is called a time-homogeneous *Markov process* with distributions $(\mathbf{P}_x)_{x\in E}$ on the space $(\Omega, \mathcal{A})$ if:

(i) For every $x \in E$, X is a stochastic process on the probability space $(\Omega, \mathcal{A}, \mathbf{P}_x)$ with $\mathbf{P}_x[X_0 = x] = 1$.
(ii) The map $\kappa : E \times \mathcal{B}(E)^{\otimes I} \to [0, 1]$, $(x, B) \mapsto \mathbf{P}_x[X \in B]$ is a stochastic kernel.
(iii) X has the time-homogeneous *Markov property* (MP): For every $A \in \mathcal{B}(E)$, every $x \in E$ and all $s, t \in I$, we have

$$\mathbf{P}_x[X_{t+s} \in A \mid \mathcal{F}_s] = \kappa_t(X_s, A) \quad \mathbf{P}_x\text{-a.s.}$$

Here, for every $t \in I$, the *transition kernel* $\kappa_t : E \times \mathcal{B}(E) \to [0, 1]$ is the stochastic kernel defined for $x \in E$ and $A \in \mathcal{B}(E)$ by

$$\kappa_t(x, A) := \kappa\big(x, \big\{y \in E^I : y(t) \in A\big\}\big) = \mathbf{P}_x[X_t \in A].$$

The family $(\kappa_t(x, A), t \in I, x \in E, A \in \mathcal{B}(E))$ is also called the family of *transition probabilities* of X.

We write $\mathbf{E}_x$ for expectation with respect to $\mathbf{P}_x$, $\mathcal{L}_x[X] = \mathbf{P}_x$ and $\mathcal{L}_x[X \mid \mathcal{F}] = \mathbf{P}_x[X \in \cdot \mid \mathcal{F}]$ (for a regular conditional distribution of X given $\mathcal{F}$).

If E is countable, then X is called a *discrete Markov process*.

In the special case $I = \mathbb{N}_0$, X is called a *Markov chain*. In this case, κ_n is called the family of n-step transition probabilities.

Remark 17.4 We will see that the existence of the transition kernels (κ_t) implies the existence of the kernel κ. Thus, a time-homogeneous Markov process is simply a stochastic process with the Markov property and for which the transition probabilities are time-homogeneous. Although it is sometimes convenient to allow also time-inhomogeneous Markov processes, for a wide range of applications it is sufficient to consider time-homogeneous Markov processes. We will not go into the details but will henceforth assume that all Markov processes are time-homogeneous. ◊

In the following, we will use the somewhat sloppy notation $\mathbf{P}_{X_s}[X \in \cdot] := \kappa(X_s, \cdot)$. That is, we understand X_s as the initial value of a *second* Markov process with the same distributions $(\mathbf{P}_x)_{x\in E}$.

Example 17.5 Let $Y_1, Y_2, \ldots$ be i.i.d. $\mathbb{R}^d$-valued random variables and let

$$S_n^x = x + \sum_{i=1}^{n} Y_i \quad \text{for } x \in \mathbb{R}^d \text{ and } n \in \mathbb{N}_0.$$

Define probability measures $\mathbf{P}_x$ on $((\mathbb{R}^d)^{\mathbb{N}_0}, (\mathcal{B}(\mathbb{R}^d))^{\otimes \mathbb{N}_0})$ by $\mathbf{P}_x = \mathbf{P} \circ (S^x)^{-1}$. Then the canonical process $X_n : (\mathbb{R}^d)^{\mathbb{N}_0} \to \mathbb{R}^d$ is a Markov chain with distributions $(\mathbf{P}_x)_{x \in \mathbb{R}^d}$. The process X is called a random walk on $\mathbb{R}^d$ with initial value x. ◊

Example 17.6 In the previous example, it is simple to pass to continuous time; that is, $I = [0, \infty)$. To this end, let $(\nu_t)_{t \geq 0}$ be a convolution semigroup on $\mathbb{R}^d$ and let $\kappa_t(x, dy) = \delta_x * \nu_t(dy)$. In Theorem 14.47, for every $x \in \mathbb{R}^d$, we constructed a measure $\mathbf{P}_x$ on $((\mathbb{R}^d)^{[0,\infty)}, \mathcal{B}(\mathbb{R}^d)^{\otimes [0,\infty)})$ with

$$\mathbf{P}_x \circ (X_0, X_{t_1}, \ldots, X_{t_n})^{-1} = \delta_x \otimes \bigotimes_{i=0}^{n-1} \kappa_{t_{n+1} - t_n}$$

for any choice of finitely many points $0 = t_0 < t_1 < \ldots < t_n$. It is easy to check that the map $\kappa : \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d)^{\otimes [0,\infty)}$, $(x, A) \mapsto \mathbf{P}_x[A]$ is a stochastic kernel. The time-homogeneous Markov property is immediate from the fact that the increments are independent and stationary. ◊

Example 17.7 (See Example 9.5 and Theorem 5.36.) Let $\theta > 0$ and $\nu_t^\theta(\{k\}) = e^{-\theta t} \frac{t^k \theta^k}{k!}$, $k \in \mathbb{N}_0$, the convolution semigroup of the Poisson distribution. The Markov process X on $\mathbb{N}_0$ with this semigroup is called a *Poisson process* with (jump) rate θ. ◊

As in Example 17.6, we will construct a Markov process for a more general Markov semigroup of stochastic kernels.

Theorem 17.8 *Let $I \subset [0, \infty)$ be closed under addition and let $(\kappa_t)_{t \in I}$ be a Markov semigroup of stochastic kernels from E to E. Then there is a measurable space $(\Omega, \mathcal{A})$ and a Markov process $((X_t)_{t \in I}, (\mathbf{P}_x)_{x \in E})$ on $(\Omega, \mathcal{A})$ with transition probabilities*

$$\mathbf{P}_x[X_t \in A] = \kappa_t(x, A) \quad \text{for all } x \in E,\ A \in \mathcal{B}(E),\ t \in I. \tag{17.2}$$

Conversely, for every Markov process X, Eq. (17.2) defines a semigroup of stochastic kernels. By (17.2), the finite-dimensional distributions of X are uniquely determined.

Proof "$\Longrightarrow$" We construct X as a canonical process. Let $\Omega = E^{[0,\infty)}$ and $\mathcal{A} = \mathcal{B}(E)^{\otimes [0,\infty)}$. Further, let X_t be the projection on the tth coordinate. For $x \in E$, define (see Corollary 14.43) on $(\Omega, \mathcal{A})$ the probability measure $\mathbf{P}_x$ such that, for finitely many time points $0 = t_0 < t_1 < \ldots < t_n$, we have

$$\mathbf{P}_x \circ (X_{t_0}, \ldots, X_{t_n})^{-1} = \delta_x \otimes \bigotimes_{i=0}^{n-1} \kappa_{t_{i+1}-t_i}.$$

Then

$$\begin{aligned}\mathbf{P}_x[X_{t_0} \in A_0, \ldots, X_{t_n} \in A_n] \\ &= \int_{A_{n-1}} \mathbf{P}_x[X_{t_0} \in A_0, \ldots, X_{t_{n-2}} \in A_{n-2}, X_{t_{n-1}} \in dx_{n-1}] \\ &\quad \times \kappa_{t_n - t_{n-1}}(x_{n-1}, A_n);\end{aligned}$$

hence $\mathbf{P}_x[X_{t_n} \in A_n \mid \mathcal{F}_{t_{n-1}}] = \kappa_{t_n - t_{n-1}}(X_{t_{n-1}}, A_n)$. Thus X is recognized as a Markov process. Furthermore, we have $\mathbf{P}_x[X_t \in A] = (\delta_x \cdot \kappa_t)(A) = \kappa_t(x, A)$.

"$\Longleftarrow$" Now let $(X, (\mathbf{P}_x)_{x \in E})$ be a Markov process. Then a stochastic kernel κ_t is defined by

$$\kappa_t(x, A) := \mathbf{P}_x[X_t \in A] \quad \text{for all } x \in E,\, A \in \mathcal{B}(E),\, t \in I.$$

By the Markov property, we have

$$\begin{aligned}\kappa_{t+s}(x, A) = \mathbf{P}_x[X_{t+s} \in A] &= \mathbf{E}_x\big[\mathbf{P}_{X_s}[X_t \in A]\big] \\ &= \int \mathbf{P}_x[X_s \in dy]\mathbf{P}_y[X_t \in A] \\ &= \int \kappa_s(x, dy)\kappa_t(y, A) = (\kappa_s \cdot \kappa_t)(x, A).\end{aligned}$$

Hence $(\kappa_t)_{t \in I}$ is a Markov semigroup. □

Theorem 17.9 *A stochastic process $X = (X_t)_{t \in I}$ is a Markov process if and only if there exists a stochastic kernel $\kappa : E \times \mathcal{B}(E)^{\otimes I} \to [0, 1]$ such that, for every bounded $\mathcal{B}(E)^{\otimes I} - \mathcal{B}(\mathbb{R})$-measurable function $f : E^I \to \mathbb{R}$ and for every $s \geq 0$ and $x \in E$, we have*

$$\mathbf{E}_x\big[f\big((X_{t+s})_{t \in I}\big) \,\big|\, \mathcal{F}_s\big] = \mathbf{E}_{X_s}\big[f(X)\big] := \int_{E^I} \kappa(X_s, dy) f(y). \tag{17.3}$$

Proof "$\Longleftarrow$" The time-homogeneous Markov property follows by (17.3) with the function $f(y) = \mathbb{1}_A(y(t))$ since $\mathbf{P}_{X_s}[X_t \in A] = \mathbf{P}_x[X_{t+s} \in A \mid \mathcal{F}_s] = \kappa_t(X_s, A)$.

"$\Longrightarrow$" By the usual approximation arguments, it is enough to consider functions f that depend only on finitely many coordinates $0 \leq t_1 \leq t_2 \leq \ldots \leq t_n$. We perform the proof by induction on n.

For $n = 1$ and f an indicator function, this is the (time-homogeneous) Markov property. For general measurable f, the statement follows by the usual approximation arguments.

Now assume the claim is proved for $n \in \mathbb{N}$. Again it suffices to assume that f is an indicator function of the type $f(x) = \mathbb{1}_{B_1 \times \ldots \times B_{n+1}} l(x_{t_1}, \ldots, x_{t_{n+1}})$ (with $B_1, \ldots, B_{n+1} \in \mathcal{B}(E)$). Using the Markov property (third and fifth equalities in the following equation) and the induction hypothesis (fourth equality), we get

$$\begin{aligned}
&\mathbf{E}_x\big[f\big((X_{t+s})_{t\geq 0}\big) \,\big|\, \mathcal{F}_s\big] \\
&\quad = \mathbf{E}_x\big[\mathbf{E}_x\big[f\big((X_{t+s})_{t\geq 0}\big) \,\big|\, \mathcal{F}_{t_n+s}\big] \,\big|\, \mathcal{F}_s\big] \\
&\quad = \mathbf{E}_x\big[\mathbf{E}_x[\mathbb{1}_{\{X_{t_{n+1}+s} \in B_{n+1}\}} \mid \mathcal{F}_{t_n+s}] \mathbb{1}_{B_1}(X_{t_1+s}) \ldots \mathbb{1}_{B_n}(X_{t_n+s}) \,\big|\, \mathcal{F}_s\big] \\
&\quad = \mathbf{E}_x\big[\mathbf{P}_{X_{t_n+s}}[X_{t_{n+1}-t_n} \in B_{n+1}] \mathbb{1}_{B_1}(X_{t_1+s}) \ldots \mathbb{1}_{B_n}(X_{t_n+s}) \,\big|\, \mathcal{F}_s\big] \\
&\quad = \mathbf{E}_{X_s}\big[\mathbf{P}_{X_{t_n}}[X_{t_{n+1}-t_n} \in B_{n+1}] \mathbb{1}_{B_1}(X_{t_1}) \ldots \mathbb{1}_{B_n}(X_{t_n})\big] \\
&\quad = \mathbf{E}_{X_s}\big[\mathbf{P}_{X_0}[X_{t_{n+1}} \in B_{n+1} \mid \mathcal{F}_{t_n}] \mathbb{1}_{B_1}(X_{t_1}) \ldots \mathbb{1}_{B_n}(X_{t_n})\big] \\
&\quad = \mathbf{E}_{X_s}\big[\mathbf{P}_{X_0}[X_{t_1} \in B_1, \ldots, X_{t_{n+1}} \in B_{n+1} \mid \mathcal{F}_{t_n}]\big] \\
&\quad = \mathbf{E}_{X_s}\big[f(X)\big].
\end{aligned}$$

□

Corollary 17.10 *A stochastic process $(X_n)_{n\in\mathbb{N}_0}$ is a Markov chain if and only if*

$$\mathcal{L}_x\big[(X_{n+k})_{n\in\mathbb{N}_0} \,\big|\, \mathcal{F}_k\big] = \mathcal{L}_{X_k}\big[(X_n)_{n\in\mathbb{N}_0}\big] \quad \textit{for every } k \in \mathbb{N}_0. \tag{17.4}$$

Proof If the conditional distributions exist, then, by Theorem 17.9, the equation (17.4) is equivalent to X being a Markov chain. Hence we only have to show that the conditional distributions exist.

Since E is Polish, $E^{\mathbb{N}_0}$ is also Polish and we have $\mathcal{B}(E^{\mathbb{N}_0}) = \mathcal{B}(E)^{\otimes \mathbb{N}_0}$ (see Theorem 14.8). Hence, by Theorem 8.37, there exists a regular conditional distribution of $(X_{n+k})_{n\in\mathbb{N}_0}$ given $\mathcal{F}_k$. □

Theorem 17.11 *Let $I = \mathbb{N}_0$. If $(X_n)_{n\in\mathbb{N}_0}$ is a stochastic process with distributions $(\mathbf{P}_x, x \in E)$, then the Markov property in Definition 17.3*(iii) *is implied by the existence of a stochastic kernel $\kappa_1 : E \times \mathcal{B}(E) \to [0,1]$ with the property that for every $A \in \mathcal{B}(E)$, every $x \in E$ and every $s \in I$, we have*

$$\mathbf{P}_x[X_{s+1} \in A \mid \mathcal{F}_s] = \kappa_1(X_s, A). \tag{17.5}$$

In this case, the n-step transition kernels κ_n can be computed inductively by

$$\kappa_n = \kappa_{n-1} \cdot \kappa_1 = \int_E \kappa_{n-1}(\,\boldsymbol{\cdot}\,, dx)\kappa_1(x, \,\boldsymbol{\cdot}\,).$$

In particular, the family $(\kappa_n)_{n\in\mathbb{N}}$ is a Markov semigroup and the distribution X is uniquely determined by κ_1.

Proof In Theorem 17.9, let $t_i = i$ for every $i \in \mathbb{N}_0$. For the proof of that theorem, only (17.5) was needed. □

The (time-homogeneous) Markov property of a process means that, for fixed time t, the future (after t) depends on the past (before t) only via the present (that is, via the value X_t). We can generalize this concept by allowing random times τ instead of fixed times t.

Definition 17.12 Let $I \subset [0,\infty)$ be closed under addition. A Markov process $(X_t)_{t\in I}$ with distributions $(\mathbf{P}_x, x \in E)$ has the *strong Markov property* if, for every a.s. finite stopping time τ, every bounded $\mathcal{B}(E)^{\otimes I} - \mathcal{B}(\mathbb{R})$ measurable function $f : E^I \to \mathbb{R}$ and every $x \in E$, we have

$$\mathbf{E}_x\big[f\big((X_{\tau+t})_{t\in I}\big)\,\big|\,\mathcal{F}_\tau\big] = \mathbf{E}_{X_\tau}\big[f(X)\big] := \int_{E^I} \kappa(X_\tau, dy) f(y). \tag{17.6}$$

Remark 17.13 If I is countable, then the strong Markov property holds if and only if, for every almost surely finite stopping time τ, we have

$$\mathcal{L}_x\big[(X_{\tau+t})_{t\in\mathbb{N}_0}\,\big|\,\mathcal{F}_\tau\big] = \mathcal{L}_{X_\tau}\big[(X_t)_{t\in\mathbb{N}_0}\big] := \kappa(X_\tau, \cdot\,). \tag{17.7}$$

This follows just as in Corollary 17.10. ◊

Most Markov processes one encounters have the strong Markov property. In particular, for countable time sets, the strong Markov property follows from the Markov property. For continuous time, however, in general, some work has to be done to establish the strong Markov property.

Theorem 17.14 *If $I \subset [0,\infty)$ is countable and closed under addition, then every Markov process $(X_n)_{n\in I}$ with distributions $(\mathbf{P}_x)_{x\in E}$ has the strong Markov property.*

Proof Let $f : E^I \to \mathbb{R}$ be measurable and bounded. Then, for every $s \in I$, the random variable $\mathbb{1}_{\{\tau=s\}}\mathbf{E}_x[f((X_{s+t})_{t\in I}) \mid \mathcal{F}_\tau]$ is measurable with respect to $\mathcal{F}_s$. Using the tower property of the conditional expectation and Theorem 17.9 in the third equality, we thus get

$$\begin{aligned}
\mathbf{E}_x\big[f\big((X_{\tau+t})_{t\in I}\big)\,\big|\,\mathcal{F}_\tau\big] &= \sum_{s\in I} \mathbb{1}_{\{\tau=s\}}\mathbf{E}_x\big[f\big((X_{s+t})_{t\in I}\big)\,\big|\,\mathcal{F}_\tau\big] \\
&= \sum_{s\in I} \mathbf{E}_x\big[\mathbb{1}_{\{\tau=s\}}\mathbf{E}_x\big[f\big((X_{s+t})_{t\in I}\big)\,\big|\,\mathcal{F}_s\big]\,\big|\,\mathcal{F}_\tau\big] \\
&= \sum_{s\in I} \mathbf{E}_x\big[\mathbb{1}_{\{\tau=s\}}\mathbf{E}_{X_s}\big[f\big((X_t)_{t\in I}\big)\big]\,\big|\,\mathcal{F}_\tau\big] \\
&= \mathbf{E}_{X_\tau}\big[f\big((X_t)_{t\in I}\big)\big].
\end{aligned}$$
□

As a simple application of the strong Markov property, we show the reflection principle for random walks.

Theorem 17.15 (Reflection principle) *Let $Y_1, Y_2, \dots$ be i.i.d. real random variables with symmetric distribution $\mathcal{L}[Y_1] = \mathcal{L}[-Y_1]$. Define $X_0 = 0$ and $X_n := Y_1 + \ldots + Y_n$ for $n \in \mathbb{N}$. Then, for every $n \in \mathbb{N}_0$ and $a > 0$,*

$$\mathbf{P}\Big[\sup_{m \le n} X_m \ge a\Big] \le 2\mathbf{P}[X_n \ge a] - \mathbf{P}[X_n = a]. \tag{17.8}$$

If we have $\mathbf{P}[Y_1 \in \{-1, 0, 1\}] = 1$, then for $a \in \mathbb{N}$ equality holds in (17.8).

Proof Let $a > 0$ and $n \in \mathbb{N}$. Define the time of first excess of a (truncated at $(n+1)$),

$$\tau := \inf\{m \ge 0 : X_m \ge a\} \wedge (n+1).$$

Then τ is a bounded stopping time and

$$\sup_{m \le n} X_m \ge a \quad \Longleftrightarrow \quad \tau \le n.$$

Let $f(m, X) = \mathbb{1}_{\{m \le n\}}(\mathbb{1}_{\{X_{n-m} > a\}} + \frac{1}{2}\mathbb{1}_{\{X_{n-m} = a\}})$. Then

$$f\big(\tau, (X_{\tau+m})_{m \in \mathbb{N}_0}\big) = \mathbb{1}_{\{\tau \le n\}}\bigg(\mathbb{1}_{\{X_n > a\}} + \frac{1}{2}\mathbb{1}_{\{X_n = a\}}\bigg).$$

The strong Markov property of X yields

$$\mathbf{E}_0\big[f\big(\tau, (X_{\tau+m})_{m \ge 0}\big) \,\big|\, \mathcal{F}_\tau\big] = \varphi(\tau, X_\tau),$$

where $\varphi(m, x) = \mathbf{E}_x[f(m, X)]$. (Recall that $\mathbf{E}_x$ denotes the expectation for X if $X_0 = x$.)

Due to the symmetry of Y_i, we have

$$\varphi(m, x) \begin{cases} \ge \frac{1}{2}, & \text{if } m \le n \text{ and } x \ge a, \\ = \frac{1}{2}, & \text{if } m \le n \text{ and } x = a, \\ = 0, & m > n. \end{cases}$$

Hence

$$\begin{aligned} \{\tau \le n\} = \{\tau \le n\} \cap \{X_\tau \ge a\} &\subset \bigg\{\varphi(\tau, X_\tau) \ge \frac{1}{2}\bigg\} \cap \{\tau \le n\} \\ &= \big\{\varphi(\tau, X_\tau) > 0\big\} \cap \{\tau \le n\}. \end{aligned}$$

Now (17.8) is implied by

$$\begin{aligned} \mathbf{P}[X_n > a] + \frac{1}{2}\mathbf{P}[X_n = a] &= \mathbf{E}\big[f\big(\tau, (X_{\tau+m})_{m \ge 0}\big)\big] \\ &= \mathbf{E}_0\big[\varphi(\tau, X_\tau)\mathbb{1}_{\{\tau \le n\}}\big] \ge \frac{1}{2}\mathbf{P}_0[\tau \le n]. \end{aligned} \tag{17.9}$$

Now assume $\mathbf{P}[Y_1 \in \{-1, 0, 1\}] = 1$ and $a \in \mathbb{N}$. Then $X_\tau = a$ if $\tau \le n$. Hence

$$\big\{\varphi(\tau, X_\tau) > 0\big\} \cap \{\tau \le n\} = \left\{\varphi(\tau, X_\tau) = \frac{1}{2}\right\} \cap \{\tau \le n\}.$$

Thus, in the last step of (17.9), equality holds and hence also in (17.8). □

Exercise 17.1.1 Let $I \subset \mathbb{R}$ and let $X = (X_t)_{t\in I}$ be a stochastic process. For $t \in I$, define the σ-algebras that code the past before t and the future beginning with t by

$$\mathcal{F}_{\le t} := \sigma(X_s : s \in I, s \le t) \quad \text{and} \quad \mathcal{F}_{\ge t} := \sigma(X_s : s \in I, s \ge t).$$

Show that X has the Markov property if and only if, for every $t \in I$, the σ-algebras $\mathcal{F}_{\le t}$ and $\mathcal{F}_{\ge t}$ are independent given $\sigma(X_t)$ (compare Definition 12.20).

In other words, a process has the (possibly time-inhomogeneous) Markov property if and only if past and future are independent given the present.

17.2 Discrete Markov Chains: Examples

Let E be countable and $I = \mathbb{N}_0$. By Definition 17.3, a Markov process $X = (X_n)_{n\in\mathbb{N}_0}$ on E is a discrete Markov chain (or Markov chain with discrete state space).

If X is a discrete Markov chain, then $(\mathbf{P}_x)_{x\in E}$ is determined by the *transition matrix*

$$p = \big(p(x, y)\big)_{x,y\in E} := \big(\mathbf{P}_x[X_1 = y]\big)_{x,y\in E}.$$

The n-step transition probabilities

$$p^{(n)}(x, y) := \mathbf{P}_x[X_n = y]$$

can be computed as the n-fold matrix product

$$p^{(n)}(x, y) = p^n(x, y),$$

where

$$p^n(x, y) = \sum_{z\in E} p^{n-1}(x, z)\,p(z, y)$$

and where $p^0 = I$ is the unit matrix.

By induction, we get the *Chapman–Kolmogorov equation* (see (14.14)) for all $m, n \in \mathbb{N}_0$ and $x, y \in E$,

$$p^{(m+n)}(x, y) = \sum_{z\in E} p^{(m)}(x, z)\,p^{(n)}(z, y). \tag{17.10}$$

Definition 17.16 A matrix $(p(x,y))_{x,y\in E}$ with nonnegative entries and with

$$\sum_{y\in E} p(x,y) = 1 \quad \text{for all } x \in E$$

is called a *stochastic matrix* on E.

A stochastic matrix is essentially a stochastic kernel from E to E. In Theorem 17.8 we saw that, for the semigroup of kernels $(p^n)_{n\in\mathbb{N}}$, there exists a unique discrete Markov chain whose transition probabilities are given by p. The arguments we gave there were rather abstract. Here we give a construction for X that could actually be used to implement a computer simulation of X.

Let $(R_n)_{n\in\mathbb{N}_0}$ be an independent family of random variables with values in E^E and with the property

$$\mathbf{P}\big[R_n(x) = y\big] = p(x,y) \quad \text{for all } x, y \in E. \tag{17.11}$$

For example, choose $(R_n(x), x \in E, n \in \mathbb{N})$ as an independent family of random variables with values in E and distributions

$$\mathbf{P}\big[R_n(x) = y\big] = p(x,y) \quad \text{for all } x, y \in E \text{ and } n \in \mathbb{N}_0.$$

Note, however, that in (17.11) we have *required* neither independence of the random variables $(R_n(x), x \in E)$ nor that all R_n had the same distribution. Only the one-dimensional marginal distributions are determined. In fact, in many applications it is useful to have subtle dependence structures in order to *couple* Markov chains with different initial chains. We pick up this thread again in Section 18.2.

For $x \in E$, define

$$X_0^x = x \quad \text{and} \quad X_n^x = R_n\big(X_{n-1}^x\big) \quad \text{for } n \in \mathbb{N}.$$

Finally, let $\mathbf{P}_x := \mathcal{L}[X^x]$ be the distribution of X^x. Recall that this is a probability measure on the space of sequences $(E^{\mathbb{N}_0}, \mathcal{B}(E)^{\otimes\mathbb{N}_0})$.

Theorem 17.17

(i) *With respect to the distribution $(\mathbf{P}_x)_{x\in E}$, the canonical process X on $(E^{\mathbb{N}_0}, \mathcal{B}(E)^{\otimes\mathbb{N}_0})$ is a Markov chain with transition matrix p.*

(ii) *In particular, to any stochastic matrix p, there corresponds a unique discrete Markov chain X with transition probabilities p.*

Proof "(ii)" This follows from (i) since Theorem 17.11 yields uniqueness of X.

"(i)" For $n \in \mathbb{N}_0$ and $x, y, z \in E$, by construction,

$$\begin{aligned}\mathbf{P}_x[X_{n+1} = z \mid \mathcal{F}_n, X_n = y] &= \mathbf{P}\big[X^x_{n+1} = z \,\big|\, \sigma(R_m, m \le n), X^x_n = y\big] \\ &= \mathbf{P}\big[R_{n+1}\big(X^x_n\big) = z \,\big|\, \sigma(R_m, m \le n), X^x_n = y\big] \\ &= \mathbf{P}\big[R_{n+1}(y) = z\big] \\ &= p(y, z).\end{aligned}$$

Hence, by Theorem 17.11, X is a Markov chain with transition matrix p. □

Example 17.18 (Random walk on $\mathbb{Z}$) Let $E = \mathbb{Z}$, and assume

$$p(x, y) = p(0, y - x) \quad \text{for all } x, y \in \mathbb{Z}.$$

In this case, we say that p is *translation invariant*. A discrete Markov chain X with transition matrix p is a random walk on $\mathbb{Z}$. Indeed, $X_n \stackrel{\mathcal{D}}{=} X_0 + Z_1 + \ldots + Z_n$, where $(Z_n)_{n\in\mathbb{N}}$ are i.i.d. with $\mathbf{P}[Z_n = x] = p(0, x)$.

The R_n that we introduced in the explicit construction are given by $R_n(x) := x + Z_n$. ◊

Example 17.19 (Computer simulation) Consider the situation where the state space $E = \{1, \ldots, k\}$ is finite. The aim is to simulate a Markov chain X with transition matrix p on a computer. Assume that the computer provides a random number generator that generates an i.i.d. sequence $(U_n)_{n\in\mathbb{N}}$ of random variables that are uniformly distributed on $[0, 1]$. (Of course, this is wishful thinking. But modern random number generators produce sequences that for many purposes are close enough to really random sequences.)

Define $r(i, 0) = 0$, $r(i, j) = p(i, 1) + \ldots + p(i, j)$ for $i, j \in E$, and define Y_n by

$$R_n(i) = j \quad \Longleftrightarrow \quad U_n \in \big[r(i, j-1), r(i, j)\big).$$

Then, by construction, $\mathbf{P}[R_n(i) = j] = r(i, j) - r(i, j-1) = p(i, j)$. ◊

Example 17.20 (Branching process as a Markov chain) We want to understand the Galton–Watson branching process (see Definition 3.9) as a Markov chain on $E = \mathbb{N}_0$.

To this end, let $(q_k)_{k\in\mathbb{N}_0}$ be a probability vector, the offspring distribution of one individual. Define $q_k^{*0} = \mathbb{1}_{\{0\}}(k)$ and

$$q_k^{*n} = \sum_{l=0}^{k} q_{k-l}^{*(n-1)} q_l \quad \text{for } n \in \mathbb{N}$$

as the n-fold convolutions of q. Hence, for n individuals, q_k^{*n} is the probability to have exactly k offspring. Finally, define the matrix p by $p(x, y) = q_y^{*x}$ for $x, y \in \mathbb{N}_0$.

Now let $(Y_{n,i}, n \in \mathbb{N}_0, i \in \mathbb{N}_0)$ be i.i.d. with $\mathbf{P}[Y_{n,i} = k] = q_k$. For $x \in \mathbb{N}_0$, define the branching process X with x ancestors and offspring distribution q by $X_0 = x$ and $X_n := \sum_{i=1}^{X_{n-1}} Y_{n-1,i}$. In order to show that X is a Markov chain, we compute

$$\begin{aligned}
&\mathbf{P}[X_n = x_n \mid X_0 = x, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}] \\
&\quad = \mathbf{P}[Y_{n-1,1} + \ldots + Y_{n-1,x_{n-1}} = x_n] \\
&\quad = \mathbf{P}_{Y_{1,1}}^{*x_{n-1}}\big(\{x_n\}\big) = q_{x_n}^{*x_{n-1}} = p(x_{n-1}, x_n).
\end{aligned}$$

Hence X is a Markov chain on $\mathbb{N}_0$ with transition matrix p. ◊

Example 17.21 (Wright's evolution model) In population genetics, Wright's evolution model [171] describes the hereditary transmission of a genetic trait with two possible specifications (say A and B); for example, resistance/no resistance to a specific antibiotic. It is assumed that the population has a constant size of $N \in \mathbb{N}$ individuals and the generations change at discrete times and do not overlap. Furthermore, for simplicity, the individuals are assumed to be *haploid*; that is, cells bear only one copy of each chromosome (like certain protozoans do) and not two copies (as in mammals).

Here we consider the case where none of the traits is favored by selection. Hence, it is assumed that each individual of the new generation chooses independently and uniformly at random one individual of the preceding generation as ancestor and becomes a perfect clone of that. Thus, if the number of individuals of type A in the current generation is $k \in \{0, \ldots, N\}$, then in the new generation it will be random and binomially distributed with parameters N and k/N.

The gene frequencies k/N in this model can be described by a Markov chain X on $E = \{0, 1/N, \ldots, (N-1)/N, 1\}$ with transition matrix $p(x, y) = b_{N,x}(\{Ny\})$. Note that X is a (bounded) martingale. Hence, by the martingale convergence theorem (Theorem 11.7), X converges $\mathbf{P}_x$-almost surely to a random variable X_∞ with $\mathbf{E}_x[X_\infty] = \mathbf{E}_x[X_0] = x$. As with the voter model (see Example 11.16) that is closely related to Wright's model, we can argue that the limit X_∞ can take only the stable values 0 and 1. That is, $\mathbf{P}_x[\lim_{n\to\infty} X_n = 1] = x = 1 - \mathbf{P}_x[\lim_{n\to\infty} X_n = 0]$. ◊

Example 17.22 (Discrete Moran model) In contrast to Wright's model, the Moran model also allows overlapping generations. The situation is similar to that of Wright's model; however, now in each time step, only (exactly) one individual gets replaced by a new one, whose type is chosen at random from the whole population.

As the new and the old types of the replaced individual are independent, as a model for the gene frequencies, we obtain a Markov chain X on $E = \{0, \frac{1}{N}, \ldots, 1\}$ with transition matrix

$$p(x, y) = \begin{cases} x(1-x), & \text{if } y = x + 1/N, \\ x^2 + (1-x)^2, & \text{if } y = x, \\ x(1-x), & \text{if } y = x - 1/N, \\ 0, & \text{else.} \end{cases}$$

Here also, X is a bounded martingale and we can compute the square variation process,

$$\langle X\rangle_n = \sum_{i=1}^{n} \mathbf{E}\big[(X_i - X_{i-1})^2 \,\big|\, X_{i-1}\big] = \frac{2}{N^2}\sum_{i=0}^{n-1} X_i(1-X_i). \tag{17.12}$$

$\diamondsuit$

Exercise 17.2.1 (Discrete martingale problem) Let $E \subset \mathbb{R}$ be countable and let X be a Markov chain on E with transition matrix p and with the property that, for any x, there are at most three choices for the next step; that is, there exists a set $A_x \subset E$ of cardinality 3 with $p(x, y) = 0$ for all $y \in E \setminus A_x$. Let $d(x) := \sum_{y\in E}(y - x)p(x, y)$ for $x \in E$.

(i) Show that $M_n := X_n - \sum_{k=0}^{n-1} d(X_k)$ defines a martingale M with square variation process $\langle M\rangle_n = \sum_{i=0}^{n-1} f(X_i)$ for a unique function $f : E \to [0, \infty)$.
(ii) Show that the transition matrix p is uniquely determined by f and d.
(iii) For the Moran model (Example 17.22), use the explicit form (17.12) of the square variation process to compute the transition matrix.

17.3 Discrete Markov Processes in Continuous Time

Let E be countable and let $(X_t)_{t\in[0,\infty)}$ be a Markov process on E with transition probabilities $p_t(x, y) = \mathbf{P}_x[X_t = y]$ (for $x, y \in E$). (Some authors call such a process a Markov chain in continuous time.)

Let $x, y \in E$ with $x \neq y$. We say that X jumps *with rate* $q(x, y)$ from x to y if the following limit exists:

$$q(x, y) := \lim_{t\downarrow 0} \frac{1}{t}\mathbf{P}_x[X_t = y].$$

Henceforth we assume that the limit $q(x, y)$ exists for all $y \neq x$ and that

$$\sum_{y\neq x} q(x, y) < \infty \quad \text{for all } x \in E. \tag{17.13}$$

Then we define

$$q(x, x) = -\sum_{y\neq x} q(x, y). \tag{17.14}$$

Finally we assume that (which is equivalent to exchangeability of the limit and the sum over $y \neq x$ in the display preceding (17.13))

$$\lim_{t\downarrow 0} \frac{1}{t}\big(\mathbf{P}_x[X_t = y] - \mathbb{1}_{\{x=y\}}\big) = q(x, y) \quad \text{for all } x, y \in E. \tag{17.15}$$

Definition 17.23 If (17.13), (17.14) and (17.15) hold, then q is called the *Q-matrix* of X. Sometimes q is also called the *generator* of the semigroup $(p_t)_{t\geq 0}$.

Example 17.24 (Poisson process) The Poisson process with rate $\alpha > 0$ (compare Section 5.5) has the Q-matrix $q(x, y) = \alpha(\mathbb{1}_{\{y=x+1\}} - \mathbb{1}_{\{y=x\}})$. ◊

Theorem 17.25 *Let q be an $E \times E$ matrix such that $q(x, y) \geq 0$ for all $x, y \in E$ with $x \neq y$. Assume that (17.13) and (17.14) hold and that*

$$\lambda := \sup_{x\in E} \bigl|q(x, x)\bigr| < \infty. \tag{17.16}$$

Then q is the Q-matrix of a unique Markov process X.

Intuitively, (17.15) suggests that we define $p_t = e^{tq}$ in a suitable sense. Then, formally, $q = \frac{d}{dt} p_t\big|_{t=0}$. The following proof shows that this formal argument can be made rigorous.

Proof Let I be the unit matrix on E. Define

$$p(x, y) = \frac{1}{\lambda} q(x, y) + I(x, y) \quad \text{for } x, y \in E,$$

if $\lambda > 0$ and $p = I$ otherwise. Then p is a stochastic matrix and $q = \lambda(p - I)$. Let $((Y_n)_{n\in\mathbb{N}_0}, (\mathbf{P}_x^Y)_{x\in E})$ be a discrete Markov chain with transition matrix p and let $((T_t)_{t\geq 0}, (\mathbf{P}_n^T)_{n\in\mathbb{N}_0})$ be a Poisson process with rate λ. Let $X_t := Y_{T_t}$ and $\mathbf{P}_x = \mathbf{P}_x^Y \otimes \mathbf{P}_0^T$. Then $\mathfrak{X} := ((X_t)_{t\geq 0}, (\mathbf{P}_x)_{x\in E})$ is a Markov process and

$$\begin{aligned} p_t(x, y) := \mathbf{P}_x[X_t = y] &= \sum_{n=0}^{\infty} \mathbf{P}_0^T[T_t = n] \mathbf{P}_x^Y[Y_n = y] \\ &= e^{-\lambda t} \sum_{n=0}^{\infty} \frac{\lambda^n t^n}{n!} p^n(x, y). \end{aligned}$$

This power series (in t) converges everywhere (note that as a linear operator, p has finite norm $\|p\|_2 \leq 1$) to the matrix exponential function $e^{\lambda t p}(x, y)$. Furthermore,

$$p_t(x, y) = e^{-\lambda t} e^{\lambda t p}(x, y) = e^{\lambda t(p-I)}(x, y) = e^{tq}(x, y).$$

Differentiating the power series termwise yields $\frac{d}{dt} p_t(x, y)\big|_{t=0} = q(x, y)$. Hence $\mathfrak{X}$ is the required Markov process.

Now assume that $(\widetilde{p}_t)_{t\geq 0}$ are the transition probabilities of another Markov process $\widetilde{\mathfrak{X}}$ with the same generator q; that is, with

$$\lim_{s\downarrow 0} \frac{1}{s} \bigl(\widetilde{p}_s(x, y) - I(x, y)\bigr) = q(x, y).$$

It is easy to check that

$$\lim_{s\downarrow 0} \frac{1}{s} \bigl(p_{t+s}(x, y) - p_t(x, y)\bigr) = (q \cdot p_t)(x, y).$$

That is, we have $(d/dt)p_t(x,y) = qp_t(x,y)$. Similarly, we get $(d/dt)\widetilde{p}_t = q\widetilde{p}_t(x,y)$. Hence also,

$$p_t(x,y) - \widetilde{p}_t(x,y) = \int_0^t \big(q(p_s - \widetilde{p}_s)\big)(x,y)\,ds.$$

If we let $r_s = p_s - \widetilde{p}_s$, then $\|r_s\|_2 \le 2$ and $\|q\|_2 \le 2\lambda$; hence

$$\sup_{s\le t}\|r_s\|_2 \le \sup_{s\le t}\int_0^s \|qr_u\|_2\,du \le \|q\|_2 \sup_{s\le t}\int_0^s \|r_u\|_2\,du \le 2\lambda t \sup_{s\le t}\|r_s\|_2.$$

For $t < 1/2\lambda$, this implies $r_t = 0$; hence $\widetilde{p}_t = p_t$. For general $t > 0$, choose $n \in \mathbb{N}$ such that $t/n < 1/2\lambda$ to obtain $\widetilde{p}_t = (\widetilde{p}_{t/n})^n = (p_{t/n})^n = p_t$. □

Remark 17.26 The condition (17.16) cannot be dropped easily, as the following example shows. Let $E = \mathbb{N}$ and

$$q(x,y) = \begin{cases} x^2, & \text{if } y = x+1,\\ -x^2, & \text{if } y = x,\\ 0, & \text{else.}\end{cases}$$

We construct explicitly a candidate X for a Markov process with Q-matrix q. Let $T_1, T_2, \ldots$ be independent, exponentially distributed random variables with $\mathbf{P}_{T_n} = \exp_{n^2}$. Define $S_n = T_1 + \ldots + T_{n-1}$ and $X_t = \sup\{n \in \mathbb{N}_0 : S_n \le t\}$. Then, at any time, X makes at most one step to the right. Furthermore, due to the lack of memory of the exponential distribution (see Exercise 8.1.1),

$$\begin{aligned}
&\mathbf{P}[X_{t+s} \ge n+1 \mid X_t = n]\\
&\quad = \mathbf{P}[S_{n+1} \le t+s \mid S_n \le t,\, S_{n+1} > t]\\
&\quad = \mathbf{P}[T_n \le s+t-S_n \mid S_n \le t,\, T_n > t - S_n] = \mathbf{P}[T_n \le s]\\
&\quad = 1 - \exp\big(-n^2 s\big).
\end{aligned}$$

Therefore,

$$\lim_{s\downarrow 0} s^{-1}\mathbf{P}[X_{t+s} = n+1 \mid X_t = n] = n^2$$

and

$$\lim_{s\downarrow 0} s^{-1}\big(\mathbf{P}[X_{t+s} = n \mid X_t = n] - 1\big) = -n^2;$$

hence

$$\lim_{s\downarrow 0} s^{-1}\big(\mathbf{P}[X_{t+s} = m \mid X_t = n] - I(m,n)\big) = q(m,n) \quad \text{for all } m,n \in \mathbb{N}.$$

Let

$$\tau^n = \inf\{t \ge 0 : X_t = n\} = S_n \quad \text{for } n \in \mathbb{N}.$$

Then $\mathbf{E}_1[\tau^n] = \sum_{k=1}^{n-1} \frac{1}{k^2}$. By monotone convergence, $\mathbf{E}_1[\sup_{n\in\mathbb{N}} \tau^n] < \infty$. That is, in finite time, X exceeds all levels. We say that X *explodes*. ◊

Example 17.27 (A variant of Pólya's urn model) Consider a variant of Pólya's urn model with black and red balls (compare Example 12.29). In contrast to the original model, we do not simply add *one* ball of the same color as the ball that we return. Rather, the number of balls that we add varies from time to time. More precisely, the kth ball of a given color will be returned together with r_k more balls of the same color. The numbers $r_1, r_2, \ldots \in \mathbb{N}$ are parameters of the model. In particular, the case $1 = r_1 = r_2 = \ldots$ is the classical Pólya's urn model. Let

$$X_n := \begin{cases} 1, & \text{if the } n\text{th ball is black,} \\ 0, & \text{else.} \end{cases}$$

For the classical model, we saw (Example 12.29) that the fraction of black balls in the urn converges a.s. to a Beta-distributed random variable Z. Furthermore, given Z, the sequence $X_1, X_2, \ldots$ is independent and Ber_Z-distributed. A similar statement holds for the case where $r = r_1 = r_2 = \ldots$ for some $r \in \mathbb{N}$. Indeed, here only the parameters of the Beta distribution change. In particular (as the Beta distribution is continuous and, in particular, does not have atoms at 0 or 1), almost surely we draw infinitely many balls of each color. Formally, $\mathbf{P}[B] = 0$ where B is the event where there is one color of which only finitely many balls are drawn.

The situation changes when the numbers r_k grow quickly as $k \to \infty$. Assume that in the beginning there is one black and one red ball in the urn. Denote by $w_n = 1 + \sum_{k=1}^{n} r_k$ the total number of balls of a given color after n balls of that color have been drawn already ($n \in \mathbb{N}_0$).

For illustration, first consider the extreme situation where w_n grows very quickly; for example, $w_n = 2^n$ for every $n \in \mathbb{N}$. Denote by

$$S_n = 2(X_1 + \ldots + X_n) - n$$

the number of black balls drawn in the first n steps minus the number of red balls drawn in these steps. Then, for every $n \in \mathbb{N}_0$,

$$\mathbf{P}[X_{n+1} = 1 \mid S_n] = \frac{2^{S_n}}{1 + 2^{S_n}} \quad \text{and} \quad \mathbf{P}[X_{n+1} = 0 \mid S_n] = \frac{2^{-S_n}}{1 + 2^{-S_n}}.$$

We conclude that $(Z_n)_{n\in\mathbb{N}_0} := (|S_n|)_{n\in\mathbb{N}_0}$ is a Markov chain on $\mathbb{N}_0$ with transition matrix

$$p(z, z') = \begin{cases} 2^z/(1 + 2^z), & \text{if } z' = z + 1 > 1, \\ 1, & \text{if } z' = z + 1 = 1, \\ 1/(1 + 2^z), & \text{if } z' = z - 1, \\ 0, & \text{else.} \end{cases}$$

The event B from above can be written as

$$B = \{Z_{n+1} < Z_n \text{ only finitely often}\}.$$

Let $A = \{Z_{n+1} > Z_n \text{ for all } n \in \mathbb{N}_0\}$ denote the event where Z *flees directly to* ∞ and let $\tau_z = \inf\{n \in \mathbb{N}_0 : Z_n \geq z\}$. Evidently,

$$\mathbf{P}_z[A] = \prod_{z'=z}^{\infty} p\big(z', z'+1\big) \geq 1 - \sum_{z'=z}^{\infty} \frac{1}{1+2^{z'}} \geq 1 - 2^{1-z}.$$

It is easy to check that $\mathbf{P}_0[\tau_z < \infty] = 1$ for all $z \in \mathbb{N}_0$. Using the strong Markov property, we get that, for all $z \in \mathbb{N}_0$,

$$\mathbf{P}_0[B] \geq \mathbf{P}_0[Z_{n+1} > Z_n \text{ for all } n \geq \tau_z] = \mathbf{P}_z[A] \geq 1 - 2^{1-z},$$

and thus $\mathbf{P}_0[B] = 1$. In prose, almost surely eventually only balls of one color will be drawn.

This example was a bit extreme. In order to find a necessary and sufficient condition on the growth of (w_n), we need more subtle methods that appeal to the above example of the explosion of a Markov process.

We will show that $\mathbf{P}[B] = 1$ if and only if $\sum_{n=0}^{\infty} \frac{1}{w_n} < \infty$. To this end, consider independent random variables $T_1^s, T_1^r, T_2^s, T_2^r, \ldots$ with $\mathbf{P}_{T_n^r} = \mathbf{P}_{T_n^s} = \exp_{w_{n-1}}$. Let $T_\infty^r = \sum_{n=1}^{\infty} T_n^r$ and $T_\infty^s = \sum_{n=1}^{\infty} T_n^s$. Clearly, $\mathbf{E}[T_\infty^r] = \sum_{n=0}^{\infty} 1/w_n < \infty$; hence, in particular, $\mathbf{P}[T_\infty^r < \infty] = 1$. The corresponding statement holds for T_∞^s. Note that T_∞^r and T_∞^s are independent and have densities (since T_1^r and T_1^s have densities); hence we have $\mathbf{P}[T_\infty^r = T_\infty^s] = 0$.

Now let

$$R_t := \sup\big\{n \in \mathbb{N} : T_1^r + \ldots + T_{n-1}^r \leq t\big\}$$

and

$$S_t := \sup\big\{n \in \mathbb{N} : T_1^s + \ldots + T_{n-1}^s \leq t\big\}.$$

Let $R := \{T_1^r + \ldots + T_n^r, n \in \mathbb{N}\}$ and let $S := \{T_1^s + \ldots + T_n^s, n \in \mathbb{N}\}$ be the jump times of (R_t) and (S_t). Define $U := R \cup S = \{u_1, u_2, \ldots\}$, where $u_1 < u_2 < \ldots$. Let

$$X_n = \begin{cases} 1, & \text{if } u_n \in S, \\ 0, & \text{else.} \end{cases}$$

Let $L_n = x_1 + \ldots + x_n$. Then

$$\begin{aligned}
&\mathbf{P}[X_{n+1} = 1 \mid X_1 = x_1, \ldots, X_n = x_n] \\
&= \mathbf{P}\big[u_{n+1} \in S \,\big|\, (u_k \in S \iff x_k = 1) \text{ for every } k \le n\big] \\
&= \mathbf{P}\big[T_1^s + \ldots + T_{L_n+1}^s < T_1^r + \ldots + T_{n-L_n+1}^r \,\big| \\
&\qquad T_1^s + \ldots + T_{L_n+1}^s > T_1^r + \ldots + T_{n-L_n}^r\big] \\
&= \mathbf{P}\big[T_{L_n+1}^s < T_{n-L_n+1}^r\big] = \frac{w_{L_n}}{w_{L_n} + w_{n-L_n}}.
\end{aligned}$$

Hence $(X_n)_{n\in\mathbb{N}_0}$ is our generalized urn model with weights $(w_n)_{n\in\mathbb{N}_0}$. Consider now the event B^c where infinitely many balls of each color are drawn. Evidently, $\{X_n = 1 \text{ infinitely often}\} = \{\sup S = \sup U\}$ and $\{X_n = 0 \text{ infinitely often}\} = \{\sup R = \sup U\}$. Since $\sup S = T_\infty^s$ and $\sup R = T_\infty^r$, we thus have $\mathbf{P}[B^c] = \mathbf{P}[T_\infty^r = T_\infty^s] = 0$. ◊

Exercise 17.3.1 Let $r, s, R, S \in \mathbb{N}$. Consider the generalized version of Pólya's urn model $(X_n)_{n\in\mathbb{N}_0}$ with $r_k = r$ and $s_k = s$ for all $k \in \mathbb{N}$. Assume that in the beginning there are R red balls and S black balls in the urn. Show that the fraction of black balls converges almost surely to a random variable Z with a Beta distribution and determine the parameters. Show that $(X_n)_{n\in\mathbb{N}_0}$ is i.i.d. given Z and $X_i \sim \mathrm{Ber}_Z$ for all $i \in \mathbb{N}_0$.

Exercise 17.3.2 Show that, almost surely, infinitely many balls of each color are drawn if

$$\sum_{n=0}^{\infty} \frac{1}{w_n} = \infty.$$

17.4 Discrete Markov Chains: Recurrence and Transience

In the following, let $X = (X_n)_{n\in\mathbb{N}_0}$ be a Markov chain on the countable space E with transition matrix p.

Definition 17.28 For any $x \in E$, let $\tau_x := \tau_x^1 := \inf\{n > 0 : X_n = x\}$ and

$$\tau_x^k = \inf\{n > \tau_x^{k-1} : X_n = x\} \quad \text{for } k \in \mathbb{N}, k \ge 2.$$

τ_x^k is the kth *entrance time* of X for x. For $x, y \in E$, let

$$F(x, y) := \mathbf{P}_x\big[\tau_y^1 < \infty\big] = \mathbf{P}_x[\text{there is an } n \ge 1 \text{ with } X_n = y]$$

be the probability of ever going from x to y. In particular, $F(x, x)$ is the return probability (after the first jump) from x to x.

Note that $\tau_x^1 > 0$ even if we start the chain at $X_0 = x$.

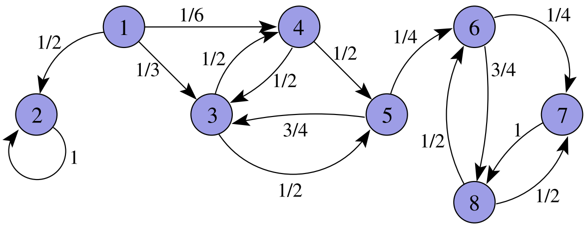

Fig. 17.1 Markov chain with eight states. The *numbers* are the transition probabilities for the corresponding *arrows*. State 2 is absorbing, the states 1, 3, 4 and 5 are transient and the states 6, 7 and 8 are (positive) recurrent

Theorem 17.29 *For all $x, y \in E$ and $k \in \mathbb{N}$, we have*

$$\mathbf{P}_x\left[\tau_y^k < \infty\right] = F(x, y)F(y, y)^{k-1}.$$

Proof We carry out the proof by induction on k. For $k = 1$, the claim is true by definition. Now let $k \geq 2$. Using the strong Markov property of X (see Theorem 17.14), we get

$$\begin{aligned}\mathbf{P}_x\left[\tau_y^k < \infty\right] &= \mathbf{E}_x\left[\mathbf{P}_x\left[\tau_y^k < \infty \,\middle|\, \mathcal{F}_{\tau_y^{k-1}}\right]\mathbb{1}_{\{\tau_y^{k-1}<\infty\}}\right] \\ &= \mathbf{E}_x\left[F(y, y)\cdot \mathbb{1}_{\{\tau_y^{k-1}<\infty\}}\right] \\ &= F(y, y)\cdot F(x, y)F(y, y)^{k-2} = F(x, y)F(y, y)^{k-1}.\end{aligned}$$

□

Definition 17.30 A state $x \in E$ is called

- *recurrent* if $F(x, x) = 1$,
- *positive recurrent* if $\mathbf{E}_x[\tau_x^1] < \infty$,
- *null recurrent* if x is recurrent but not positive recurrent,
- *transient* if $F(x, x) < 1$, and
- *absorbing* if $p(x, x) = 1$.

The Markov chain X is called (positive/null) recurrent if every state $x \in E$ is (positive/null) recurrent and is called transient if every recurrent state is absorbing.

Remark 17.31 Clearly, we have:

$$\text{"absorbing"} \quad \Longrightarrow \quad \text{"positive recurrent"} \quad \Longrightarrow \quad \text{"recurrent"}.$$

◊

Example 17.32

(i) In Fig. 17.1, the state 2 is absorbing. If it does not get trapped in 2, the chain will eventually jump from 5 to 6 and will not return after that. Hence 1, 3, 4 and 5 are transient. The states 6, 7 and 8 are positive recurrent. One can show (see Exercise 17.6.1) that $\mathbf{E}_6[\tau_6] = \frac{17}{4}$, $\mathbf{E}_7[\tau_7] = \frac{17}{5}$ and $\mathbf{E}_8[\tau_8] = \frac{17}{8}$.

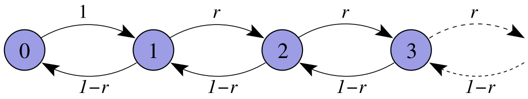

Fig. 17.2 Markov chain on $\mathbb{N}_0$ with parameter $r \in (0, 1)$. The chain is positive recurrent if $r \in (0, 1/2)$, null recurrent if $r = 1/2$ and transient if $r \in (1/2, 1)$

(ii) The chain in Fig. 17.2 has a drift to the right if $r > \frac{1}{2}$. Hence, in this case, every state is transient. On the other hand, if $r \in (0, \frac{1}{2})$, then the chain has a drift to the left (except at the point 0) and hence visits every state again and again. Thus the chain is recurrent. With a little thought, one can show (see Exercise 17.6.4) that in this case, the chain is actually positive recurrent and in the remaining case $r = \frac{1}{2}$ it is null recurrent. ◊

Definition 17.33 Denote by $N(y) = \sum_{n=0}^{\infty} \mathbb{1}_{\{X_n = y\}}$ the total number of visits of X to y and by

$$G(x, y) = \mathbf{E}_x\big[N(y)\big] = \sum_{n=0}^{\infty} p^n(x, y)$$

the *Green function* of X.

Theorem 17.34

(i) *For all $x, y \in E$, we have (with the convention $1/0 := \infty$, $0/0 := 0$ and $0 \cdot \infty := 0$)*

$$G(x, y) = \left\{ \begin{array}{ll} \frac{F(x,y)}{1-F(y,y)}, & \text{if } x \neq y \\ \frac{1}{1-F(y,y)}, & \text{if } x = y \end{array} \right\} = F(x, y)G(y, y) + \mathbb{1}_{\{x=y\}}. \tag{17.17}$$

(ii) *A non-absorbing state $x \in E$ is recurrent if and only if $G(x, x) = \infty$.*

Proof (ii) follows by (i). Hence, it remains to show (17.17). By Theorem 17.29, we have

$$\begin{aligned} G(x, y) &= \mathbf{E}_x\big[N(y)\big] = \sum_{k=1}^{\infty} \mathbf{P}_x\big[N(y) \geq k\big] \\ &= \mathbb{1}_{\{x=y\}} + \sum_{k=1}^{\infty} \mathbf{P}_x\big[\tau_y^k < \infty\big] = \mathbb{1}_{\{x=y\}} + \sum_{k=1}^{\infty} F(x, y)F(y, y)^{k-1} \\ &= \begin{cases} \frac{F(x,y)}{1-F(y,y)}, & \text{if } x \neq y, \\ \frac{1}{1-F(x,x)}, & \text{if } x = y. \end{cases} \end{aligned}$$

The second equality in (17.17) is an immediate consequence. □

Theorem 17.35 *If x is recurrent and $F(x, y) > 0$, then y is also recurrent, and $F(x, y) = F(y, x) = 1$.*

Proof Let $x, y \in E$, $x \neq y$, be such that $F(x, y) > 0$. Then there is a $k \in \mathbb{N}$ and states $x_1, \ldots, x_k \in E$ with $x_k = y$ and $x_i \neq x$ for all $i = 1, \ldots, k$ and such that

$$\mathbf{P}_x[X_i = x_i \text{ for all } i = 1, \ldots, k] > 0.$$

In particular, $p^k(x, y) > 0$. By the Markov property, we have

$$\begin{aligned} 1 - F(x, x) = \mathbf{P}_x\big[\tau_x^1 = \infty\big] &\geq \mathbf{P}_x\big[X_1 = x_1, \ldots, X_k = x_k, \tau_x^1 = \infty\big] \\ &= \mathbf{P}_x[X_1 = x_1, \ldots, X_k = x_k] \cdot \mathbf{P}_y\big[\tau_x^1 = \infty\big] \\ &= \mathbf{P}_x[X_1 = x_1, \ldots, X_k = x_k]\big(1 - F(y, x)\big). \end{aligned}$$

If now $F(x, x) = 1$, then also $F(y, x) = 1$. Since $F(y, x) > 0$, there exists an $l \in \mathbb{N}$ with $p^l(y, x) > 0$. Hence, for $n \in \mathbb{N}_0$,

$$p^{l+n+k}(y, y) \geq p^l(y, x) p^n(x, x) p^k(x, y).$$

If x is recurrent, then we conclude that

$$G(y, y) \geq \sum_{n=0}^{\infty} p^{l+n+k}(y, y) \geq p^l(y, x) p^k(x, y) G(x, x) = \infty$$

and hence also that y is recurrent. Changing the roles of x and y in the above argument, we get $F(x, y) = 1$. □

Definition 17.36 A discrete Markov chain is called

- *irreducible* if $F(x, y) > 0$ for all $x, y \in E$, or equivalently $G(x, y) > 0$, and
- *weakly irreducible* if $F(x, y) + F(y, x) > 0$ for all $x, y \in E$.

Theorem 17.37 *An irreducible discrete Markov chain is either recurrent or transient. If $|E| \geq 2$, then there is no absorbing state.*

Proof This follows directly from Theorem 17.35. □

Theorem 17.38 *If E is finite and X is irreducible, then X is recurrent.*

Proof Evidently, for all $x \in E$,

$$\sum_{y \in E} G(x, y) = \sum_{n=0}^{\infty} \sum_{y \in E} p^n(x, y) = \sum_{n=0}^{\infty} 1 = \infty.$$

As E is finite, there is a $y \in E$ with $G(x, y) = \infty$. Since $F(y, x) > 0$, there exists a $k \in \mathbb{N}$ with $p^k(y, x) > 0$. Therefore, since $p^{n+k}(x, x) \geq p^n(x, y) p^k(y, x)$, we have

$$G(x,x) \geq \sum_{n=0}^{\infty} p^n(x,y)p^k(y,x) = p^k(y,x)G(x,y) = \infty. \qquad \square$$

Exercise 17.4.1 Let x be positive recurrent and let $F(x,y) > 0$. Show that y is also positive recurrent.

17.5 Application: Recurrence and Transience of Random Walks

In this section, we study recurrence and transience of random walks on the D-dimensional integer lattice $\mathbb{Z}^D$, $D = 1, 2, \ldots$. A more exhaustive investigation can be found in Spitzer's book [158].

Consider first the simplest situation of symmetric simple random walk X on $\mathbb{Z}^D$. That is, at each step, X jumps to any of its $2D$ neighbors with the same probability $1/2D$. Hence, in terms of the Markov chain notation, we have $E = \mathbb{Z}^D$ and

$$p(x,y) = \begin{cases} \frac{1}{2D}, & \text{if } |x-y| = 1, \\ 0, & \text{else.} \end{cases}$$

Is this random walk recurrent or transient?

The central limit theorem suggests that

$$p^n(0,0) \approx C_D n^{-D/2} \quad \text{as } n \to \infty$$

for some constant C_D that depends on the dimension D. However, first we have to exclude the case where n is odd since here clearly $p^n(0,0) = 0$. Thus let $Y_1, Y_2, \ldots$ be independent $\mathbb{Z}^D$-valued random variables with $\mathbf{P}[Y_i = x] = p^2(0,x)$. Then $X_{2n} \stackrel{\mathcal{D}}{=} S_n := Y_1 + \ldots + Y_n$ for $n \in \mathbb{N}_0$; hence $G(0,0) = \sum_{n=0}^{\infty} \mathbf{P}[S_n = 0]$. Clearly, $Y_1 = (Y_1^1, \ldots, Y_1^D)$ has covariance matrix $C_{i,j} := \mathbf{E}[Y_1^i \cdot Y_1^j] = \frac{2}{D}\mathbb{1}_{\{i=j\}}$. By the local central limit theorem (see, e.g., [20, pp. 224ff] for a one-dimensional version of that theorem or Exercise 17.5.1 for an analytic derivation), we have

$$n^{D/2} p^{2n}(0,0) = n^{D/2}\mathbf{P}[S_n = 0] \stackrel{n\to\infty}{\longrightarrow} 2(4\pi/D)^{-D/2}. \tag{17.18}$$

Now $\sum_{n=1}^{\infty} n^{-\alpha} < \infty$ if and only if $\alpha > 1$. Hence $G(0,0) < \infty$ if and only if $D > 2$. We have thus shown the following theorem of Pólya [134].

Theorem 17.39 (Pólya (1921) [134]) *Symmetric simple random walk on $\mathbb{Z}^D$ is recurrent if and only if $D \leq 2$.*

The procedure we used here to derive Pólya's theorem has the disadvantage that it relies on the local central limit theorem, which we have not proved (and will not). Hence we will consider different methods of proof that yield further insight into the problem.

Consider first the one-dimensional simple random walk that with probability p jumps one step to the right and with probability $1-p$ jumps one step to the left.

Then

$$G(0,0) = \sum_{n=0}^{\infty} \binom{2n}{n} \big(p(1-p)\big)^n = \sum_{n=0}^{\infty} \binom{-1/2}{n} \big(-4p(1-p)\big)^n.$$

Using the generalized binomial theorem (see Lemma 3.5), we get (since we have $(1-4p(1-p))^{1/2} = |2p-1|$)

$$G(0,0) = \begin{cases} \frac{1}{|2p-1|}, & \text{if } p \neq \frac{1}{2}, \\ \infty, & \text{if } p = \frac{1}{2}. \end{cases} \tag{17.19}$$

Thus, simple random walk on $\mathbb{Z}$ is recurrent if and only if it is symmetric; that is, if $p = \frac{1}{2}$.

Of course, transience in the case $p \neq \frac{1}{2}$ could also be deduced directly from the strong law of large numbers since $\lim_{n\to\infty} \frac{1}{n} X_n = \mathbf{E}_0[X_1] = 2p - 1$ almost surely. In fact, this argument is even more robust since it uses only that the single steps of X have an expectation that is not zero.

Consider now the situation where X does not necessarily jump to one of its nearest neighbors but where we still have $\mathbf{E}_0[|X_1|] < \infty$ and $\mathbf{E}_0[X_1] = 0$. The strong law of large numbers does not yield recurrence immediately and we have to do some work:

By the Markov property, for every $N \in \mathbb{N}$ and every $y \neq x$,

$$G_N(x,y) := \sum_{k=0}^{N} \mathbf{P}_x[X_k = y] = \sum_{k=0}^{N} \mathbf{P}_x\big[\tau_y^1 = k\big] \sum_{l=0}^{N-k} \mathbf{P}_y[X_l = y] \leq G_N(y,y).$$

This implies for all $L \in \mathbb{N}$

$$\begin{aligned} G_N(0,0) &\geq \frac{1}{2L+1} \sum_{|y| \leq L} G_N(0,y) \\ &= \frac{1}{2L+1} \sum_{k=0}^{N} \sum_{|y|\leq L} p^k(0,y) \\ &\geq \frac{1}{2L+1} \sum_{k=1}^{N} \sum_{y:|y/k| \leq L/N} p^k(0,y). \end{aligned}$$

By the weak law of large numbers, we have $\liminf_{k\to\infty} \sum_{|y|\leq \varepsilon k} p^k(0,y) = 1$ for every $\varepsilon > 0$. Hence, letting $L = \varepsilon N$, we get

$$\liminf_{N\to\infty} G_N(0,0) \geq \frac{1}{2\varepsilon} \quad \text{for every } \varepsilon > 0.$$

Thus $G(0,0) = \infty$, which shows that X is recurrent.

We summarize the discussion in a theorem.

Theorem 17.40 *A random walk on $\mathbb{Z}$ with $\sum_{x=-\infty}^{\infty} |x| p(0,x) < \infty$ is recurrent if and only if $\sum_{x=-\infty}^{\infty} x p(0,x) = 0$.*

Now what about symmetric simple random walk in dimension $D = 2$ or in higher dimensions? In order that the random walk be at the origin after $2n$ steps, it must perform k_i steps in the ith direction and k_i steps in the opposite direction for some numbers $k_1, \ldots, k_D \in \mathbb{N}_0$ with $k_1 + \ldots + k_D = n$. We thus get

$$p^{2n}(0,0) = (2D)^{-2n} \sum_{k_1+\ldots+k_D=n} \binom{2n}{k_1, k_1, \ldots, k_D, k_D}, \tag{17.20}$$

where $\binom{N}{l_1,\ldots,l_r} = \frac{N!}{l_1!\ldots l_r!}$ is the multinomial coefficient. In particular, for $D = 2$,

$$\begin{aligned} p^{2n}(0,0) &= 4^{-2n} \sum_{k=0}^{n} \frac{(2n)!}{(k!)^2((n-k)!)^2} \\ &= 4^{-2n} \binom{2n}{n} \sum_{k=0}^{n} \binom{n}{k}\binom{n}{n-k} = \left(2^{-2n}\binom{2n}{n}\right)^2. \end{aligned}$$

Note that in the last step, we used a simple combinatorial identity that follows, e.g., by the convolution formula $(b_{n,p} * b_{n,p})(\{n\}) = b_{2n,p}(\{n\})$. Now, by Stirling's formula,

$$\lim_{n\to\infty} \sqrt{n} 2^{-2n} \binom{2n}{n} = \frac{1}{\sqrt{\pi}},$$

hence $\lim_{n\to\infty} n p^{2n}(0,0) = \frac{1}{\pi}$. In particular, we have $\sum_{n=1}^{\infty} p^{2n}(0,0) = \infty$. That is, two-dimensional symmetric simple random walk is recurrent.

For $D \geq 3$, the sum over the multinomial coefficients cannot be computed in a satisfactory way. However, it is not too hard to give an estimate that shows that there exists a $c = c_D$ such that $p^{2n}(0,0) \leq c n^{-D/2}$, which implies $G(0,0) \leq c \sum_{n=1}^{\infty} n^{-D/2} < \infty$ (see, e.g., [53, p. 361] or [59, Example 6.31]). Here, however, we follow a different route.

Things would be easy if the individual coordinates of the chain were *independent* one-dimensional random walks. In this case, the probability that at time $2n$ all coordinates are zero would be the Dth power of the probability that the first coordinate is zero. For one coordinate, however, which moves only with probability $1/D$ and thus has variance $1/D$, the probability of being back at the origin at time $2n$ is approximately $(n\pi/D)^{-1/2}$. Up to a factor, we would thus get (17.18) without using the multidimensional local central limit theorem.

An elegant way to decouple the coordinates is to pass from discrete time to continuous time in such a way that the individual coordinates become independent but such that the Green function remains unchanged.

We give the details. Let $(T_t^i)_{t\geq 0}$, $i = 1, \ldots, D$ be independent Poisson processes with rate $1/D$. Let $Z^1, \ldots, Z^D$ be independent (and independent of the Poisson processes) symmetric simple random walks on $\mathbb{Z}$. Define $T := T^1 + \ldots + T^D$, $Y_t^i := Z^i_{T_t^i}$ for $i = 1, \ldots, D$ and let $Y_t = (Y_t^1, \ldots, Y_t^D)$. Then Y is a Markov chain in continuous time with Q-matrix $q(x, y) = p(x, y) - \mathbb{1}_{\{x=y\}}$. As T is a Poisson process with rate 1, $(X_{T_t})_{t\geq 0}$ is also a Markov process with Q-matrix q. It follows that $(X_{T_t})_{t\geq 0} \stackrel{\mathcal{D}}{=} (Y_t)_{t\geq 0}$. We now compute

$$G_Y := \int_0^\infty \mathbf{P}_0[Y_t = 0]\,dt = \int_0^\infty \sum_{n=0}^\infty \mathbf{P}_0[X_{2n} = 0, T_t = 2n]\,dt$$
$$= \sum_{n=0}^\infty p^{2n}(0,0) \int_0^\infty e^{-t} \frac{t^{2n}}{(2n)!}\,dt = G(0,0).$$

The two processes $(X_n)_{n\in\mathbb{N}_0}$ and $(Y_t)_{t\in[0,\infty)}$ thus have the same Green function. As the coordinates of Y are independent, we have

$$G_Y = \int_0^\infty \mathbf{P}_0\big[Y_t^1 = 0\big]^D\,dt.$$

Hence we only have to compute the asymptotics of $\mathbf{P}_0[Y_t^1 = 0]$ for large t. We can argue as follows. By the law of large numbers, we have $T_t^1 \approx t/D$ for large t. Furthermore, $\mathbf{P}_0[Y_t^1 \text{ is even}] \approx \frac{1}{2}$. Hence we have, with $n_t = \lfloor t/2D \rfloor$ for $t \to \infty$ (compare Exercise 17.5.2),

$$\mathbf{P}_0\big[Y_t^1 = 0\big] \sim \frac{1}{2}\mathbf{P}\big[Z^1_{2n_t} = 0\big] = \frac{1}{2}\binom{2n_t}{n_t}4^{-n_t} \sim (2\pi/D)^{-1/2}t^{-1/2}. \qquad (17.21)$$

Since $\int_1^\infty t^{-\alpha}\,dt < \infty$ if and only if $\alpha > 1$, we also have $G_Y < \infty$ if and only if $D > 2$. However, this is the statement of Pólya's theorem.

Finally, we present a third method of studying recurrence and transience of random walks that does not rely on the Euclidean properties of the integer lattice but rather on the Fourier inversion formula.

First consider a general (discrete time) irreducible random walk with transition matrix p on $\mathbb{Z}^D$. By $\phi(t) = \sum_{x\in\mathbb{Z}^D} e^{i\langle t,x\rangle} p(0,x)$ denote the characteristic function of a single transition. The convolution of the transition probabilities translates into powers of the characteristic function; hence

$$\phi^n(t) = \sum_{x\in\mathbb{Z}^D} e^{i\langle t,x\rangle} p^n(0,x).$$

By the Fourier inversion formula (Theorem 15.10), we recover the n-step transition probabilities from ϕ^n by

$$p^n(0,x) = (2\pi)^{-D} \int_{[-\pi,\pi)^D} e^{-i\langle t,x\rangle} \phi^n(t)\,dt.$$

In particular, for $\lambda \in (0,1)$,

$$\begin{aligned}
R_\lambda &:= \sum_{n=0}^{\infty} \lambda^n p^n(0,0) \\
&= (2\pi)^{-D} \sum_{n=0}^{\infty} \int_{[-\pi,\pi)^D} \lambda^n \phi^n(t)\, dt \\
&= (2\pi)^{-D} \int_{[-\pi,\pi)^D} \frac{1}{1-\lambda\phi(t)}\, dt. \\
&= (2\pi)^{-D} \int_{[-\pi,\pi)^D} \operatorname{Re}\left(\frac{1}{1-\lambda\phi(t)}\right) dt.
\end{aligned}$$

Now $G(0,0) = \lim_{\lambda \uparrow 1} R_\lambda$ and hence

$$X \text{ is recurrent} \quad \iff \quad \lim_{\lambda\uparrow 1} \int_{[-\pi,\pi)^D} \operatorname{Re}\left(\frac{1}{1-\lambda\phi(t)}\right) dt = \infty. \tag{17.22}$$

If we had $\phi(t) = 1$ for some $t \in (-2\pi, 2\pi)^D \setminus \{0\}$, then we would have $\phi^n(t) = 1$ for every $n \in \mathbb{N}$ and hence, by Exercise 15.2.1, $\mathbf{P}_0[\langle X_n, t/(2\pi)\rangle \in \mathbb{Z}] = 1$. Thus X would not be irreducible contradicting the assumption. Due to the continuity of ϕ for all $\varepsilon > 0$, we thus have

$$\inf\left\{|\phi(t) - 1| : t \in [-\pi,\pi)^D \Big\backslash (-\varepsilon,\varepsilon)^D\right\} > 0.$$

We summarize the discussion in a theorem due to Chung and Fuchs [26].

Theorem 17.41 (Chung–Fuchs (1951) [26]) *An irreducible random walk on $\mathbb{Z}^D$ with characteristic function ϕ is recurrent if and only if, for every $\varepsilon > 0$,*

$$\lim_{\lambda\uparrow 1} \int_{(-\varepsilon,\varepsilon)^D} \operatorname{Re}\left(\frac{1}{1-\lambda\phi(t)}\right) dt = \infty. \tag{17.23}$$

Now consider symmetric simple random walk. Here $\phi(t) = \frac{1}{D}\sum_{i=1}^D \cos(t_i)$. Expanding the cosine function in a Taylor series, we get $\cos(t_i) = 1 - \frac{1}{2}t_i^2 + O(t_i^4)$; hence $1 - \phi(t) = \frac{1}{2D}\|t\|_2^2 + O(\|t\|_2^4)$. We infer that X is recurrent if and only if $\int_{\|t\|_2<\varepsilon} \|t\|_2^{-2}\, dt = \infty$. We compute this integral in polar coordinates (with C_D the surface of the unit sphere in $\mathbb{R}^D$):

$$\int_{\|t\|_2<\varepsilon} \|t\|_2^{-2}\, dt = C_D \int_0^\varepsilon r^{D-1} r^{-2}\, dr = \infty \quad \iff \quad D \le 2.$$

Hence, X is recurrent if and only if $D \le 2$.

In Section 19.3, we will encounter a further method of proving Pólya's theorem that has a completely different structure and that is based on the connection between Markov chains and electrical networks.

In fact, the Chung–Fuchs theorem can be used to compute the numerical values of the Green function $G_D(0,0)$ of symmetric simple random walk on $\mathbb{Z}^D$ if we compute numerically the so-called *Watson integral*

$$G_D(0,0) = (2\pi)^{-D} \int_{[-\pi,\pi)^D} \frac{D}{D - (\cos(x_1) + \ldots + \cos(x_D))}\, dx. \tag{17.24}$$

For this purpose, we follow [80] (where there are further refinements of the method) to transform the D-fold integral into a double integral. Denote by

$$I_0(t) := \frac{1}{\pi} \int_0^{\pi} e^{t\cos(\theta)}\, d\theta$$

the so-called modified Bessel function of the first kind. Using the identity $\frac{1}{\lambda} = \int_0^\infty e^{-\lambda t}\, dt$ for the integrand and applying Fubini's theorem, we get

$$G_D(0,0) = \frac{D}{(2\pi)^D} \int_0^\infty e^{-Dt} \left(\int_{[-\pi,\pi)^D} e^{t(\cos(x_1)+\ldots+\cos(x_D))}\, dx \right) dt$$

and thus

$$G_D(0,0) = D \int_0^\infty e^{-Dt} I_0(t)^D\, dt. \tag{17.25}$$

The right-hand side of (17.25) can quickly be computed numerically with great accuracy (see Table 17.1).

For the case $D = 3$, Watson [168] found the expression

$$G_3(0,0) = 12\frac{18 + 12\sqrt{2} - 10\sqrt{3} - 7\sqrt{6}}{\pi^2} K\big((2-\sqrt{3})(\sqrt{3}-\sqrt{2})\big)^2,$$

where $K(m) = \int_0^1 ((1-t^2)(1-mt^2))^{-1/2}\, dt$ is the complete elliptic integral of the first kind with modulus $m \in (-1,1)$. This in turn can be expressed as a (quickly convergent) series

$$K(m) = \frac{\pi}{2}\left(1 + \sum_{n=1}^{\infty} \left(\frac{(2n)!}{4^n (n!)^2}\right)^2 m^2\right).$$

Glasser and Zucker [61] found an expression as a product of four Gamma functions,

$$G_3(0,0) = \frac{\sqrt{6}}{32\pi^3} \Gamma\left(\frac{1}{24}\right)\Gamma\left(\frac{5}{24}\right)\Gamma\left(\frac{7}{24}\right)\Gamma\left(\frac{11}{24}\right) = 1.516386059151978018\,1\ldots$$

Exercise 17.5.1 For $n \in \mathbb{N}_0$, let p^n be the matrix of n-step transition probabilities of simple symmetric random walk on $\mathbb{Z}^D$. For $n \in \mathbb{N}$, derive the formula (see Theorem 15.10)

$$p^{2n}(0,0) = (2\pi)^{-D} \int_{[-\pi,\pi)^D} D^{-2n} \big(\cos(t_1) + \ldots + \cos(t_D)\big)^{2n}\, dt.$$

Table 17.1 Green function $G_D(0,0)$ and return probability $F_D(0,0)$ of simple symmetric random walk on $\mathbb{Z}^D$. The numerical computations are based on (17.25)

D	$G_D(0,0)$	$F_D(0,0)$
2	∞	1
3	1.51638605915	0.34053732955
4	1.23946712185	0.19320167322
5	1.15630812484	0.13517860982
6	1.11696337322	0.10471549562
7	1.09390631559	0.08584493411
8	1.07864701202	0.07291264996
9	1.06774608638	0.06344774965
10	1.05954374789	0.05619753597
11	1.05313615291	0.05045515982
12	1.04798637482	0.04578912090
13	1.04375406289	0.04191989708
14	1.04021240323	0.03865787709
15	1.03720412092	0.03586962312
16	1.03461657857	0.03345836447
17	1.03236691238	0.03135214040
18	1.03039276285	0.02949628913
19	1.02864627888	0.02784852234
20	1.02709011674	0.02637559869

By a suitable bound for the integral, conclude the convergence $n^{D/2} p^{2n}(0,0) \overset{n\to\infty}{\longrightarrow} 2(4\pi/D)^{-D/2}$ (see (17.18)).

Exercise 17.5.2 Show (17.21) formally.

Exercise 17.5.3 Use Theorem 17.41 to show that a random walk on $\mathbb{Z}^2$ with $\sum_{x\in\mathbb{Z}^2} x p(0,x) = 0$ is recurrent if $\sum_{x\in\mathbb{Z}^2} \|x\|_2^2 p(0,x) < \infty$.

Exercise 17.5.4 Use Theorem 17.41 to show that, for $D \geq 3$ every irreducible random walk on $\mathbb{Z}^D$ is transient.

Exercise 17.5.5 Show (17.25) for $G_D(0,0)$ directly with the $p^{2n}(0,0)$ from (17.20) and using the representation of $I_0(t)$ as the series $I_0(t) = \sum_{k=0}^{\infty} (k!)^{-2}(t/2)^k$.

17.6 Invariant Distributions

In the following, let p be a stochastic matrix on the discrete space E and let $(X_n)_{n\in\mathbb{N}_0}$ be a corresponding Markov chain.

This section is devoted to the question: Which distributions are preserved under the dynamics of the Markov chain? Of course, often the chain will not stay put in a specific state but the *distribution* of the random state of the chain might nevertheless be the same for all times. If such an invariant distribution exists, we will see in Chapter 18 that under rather weak conditions, the distribution of a Markov chain (started in an arbitrary state) converges in the large time limit to such an invariant distribution.

Definition 17.42 If μ is a measure on E and $f : E \to \mathbb{R}$ is a map, then we write $\mu p(\{x\}) = \sum_{y \in E} \mu(\{y\}) p(y, x)$ and $pf(x) = \sum_{y \in E} p(x, y) f(y)$ if the sums converge.

Definition 17.43

(i) A σ-finite measure μ on E is called an *invariant measure* if

$$\mu p = \mu.$$

A probability measure that is an invariant measure is called an *invariant distribution*. Denote by $\mathcal{I}$ the set of invariant distributions.

(ii) A function $f : E \to \mathbb{R}$ is called *subharmonic* if pf exists and if $f \le pf$. f is called superharmonic if $f \ge pf$ and harmonic if $f = pf$.

Remark 17.44 In the terminology of linear algebra, an invariant measure is a left eigenvector of p corresponding to the eigenvalue 1. A harmonic function is a right eigenvector corresponding to the eigenvalue 1. ◇

Lemma 17.45 *If f is bounded and (sub-, super-) harmonic, then $(f(X_n))_{n \in \mathbb{N}_0}$ is a (sub-, super-) martingale with respect to the filtration $\mathbb{F} = \sigma(X)$ generated by X.*

Proof Let f be bounded and subharmonic. Then

$$\begin{aligned}\mathbf{E}_x\big[f(X_n)\,\big|\,\mathcal{F}_{n-1}\big] = \mathbf{E}_{X_{n-1}}\big[f(X_1)\big] &= \sum_{y \in E} p(X_{n-1}, y) f(y) \\ &= pf(X_{n-1}) \ge f(X_{n-1}).\end{aligned}$$

□

Theorem 17.46 *If any point is transient, then an invariant distribution does not exist.*

Proof By assumption, $G(x, y) = \sum_{n=0}^{\infty} p^n(x, y) < \infty$ for all $x, y \in E$; hence $p^n(x, y) \overset{n \to \infty}{\longrightarrow} 0$. For every probability measure μ on E, we thus have that $\mu p^n(\{x\}) \overset{n \to \infty}{\longrightarrow} 0$. If μ was invariant, however, then we would have $\mu p^n(\{x\}) = \mu(\{x\})$ for all $n \in \mathbb{N}$. □

Theorem 17.47 *Let x be a recurrent state and let $\tau_x^1 = \inf\{n \ge 1 : X_n = x\}$. Then one invariant measure μ_x is defined by*

$$\mu_x\big(\{y\}\big) = \mathbf{E}_x\left[\sum_{n=0}^{\tau_x^1-1} \mathbb{1}_{\{X_n=y\}}\right] = \sum_{n=0}^{\infty} \mathbf{P}_x\big[X_n = y, \tau_x^1 > n\big].$$

Proof First we have to show that $\mu_x(\{y\}) < \infty$ for all $y \in E$. For $y = x$, clearly $\mu_x(\{x\}) = 1$. For $y \neq x$ and $F(x, y) = 0$, we have $\mu_x(\{y\}) = 0$. Now let $y \neq x$ and $F(x, y) > 0$. As x is recurrent, we have $F(x, y) = F(y, x) = 1$ and y is recurrent (Theorem 17.35). Let

$$\widehat{F}(x, y) = \mathbf{P}_x\big[\tau_x^1 > \tau_y^1\big].$$

Then $\widehat{F}(x, y) > 0$ (otherwise y would not be visited). Changing the roles of x and y, we also get $\widehat{F}(y, x) > 0$.

By the strong Markov property (Theorem 17.14), we have

$$\begin{aligned}\mathbf{E}_y\left[\sum_{n=0}^{\tau_x^1-1} \mathbb{1}_{\{X_n=y\}}\right] &= 1 + \mathbf{E}_y\left[\sum_{n=\tau_y^1}^{\tau_x^1-1} \mathbb{1}_{\{X_n=y\}};\, \tau_x^1 > \tau_y^1\right] \\ &= 1 + \big(1 - \widehat{F}(y, x)\big)\mathbf{E}_y\left[\sum_{n=0}^{\tau_x^1-1} \mathbb{1}_{\{X_n=y\}}\right].\end{aligned}$$

Hence,

$$\mathbf{E}_y\left[\sum_{n=0}^{\tau_x^1-1} \mathbb{1}_{\{X_n=y\}}\right] = \frac{1}{\widehat{F}(y, x)}.$$

Therefore,

$$\mu_x\big(\{y\}\big) = \mathbf{E}_x\left[\sum_{n=0}^{\tau_x^1-1} \mathbb{1}_{\{X_n=y\}}\right] = \mathbf{E}_x\left[\sum_{n=\tau_y^1}^{\tau_x^1-1} \mathbb{1}_{\{X_n=y\}};\, \tau_x^1 > \tau_y^1\right] = \frac{\widehat{F}(x, y)}{\widehat{F}(y, x)} < \infty.$$

Define $\overline{p}_n(x, y) = \mathbf{P}_x[X_n = y; \tau_x^1 > n]$. Then, for every $z \in E$,

$$\mu_x p\big(\{z\}\big) = \sum_{y\in E} \mu_x\big(\{y\}\big) p(y, z) = \sum_{n=0}^{\infty} \sum_{y\in E} \overline{p}_n(x, y) p(y, z).$$

Case 1: $x \neq z$. In this case,

$$\begin{aligned}\sum_{y\in E} \overline{p}_n(x, y) p(y, z) &= \sum_{y\in E} \mathbf{P}_x\big[X_n = y, \tau_x^1 > n, X_{n+1} = z\big] \\ &= \mathbf{P}_x\big[\tau_x^1 > n + 1; X_{n+1} = z\big] = \overline{p}_{n+1}(x, z).\end{aligned}$$

Hence (since $\overline{p}_0(x,z)=0$)

$$\mu_x p(\{z\}) = \sum_{n=0}^{\infty} \overline{p}_{n+1}(x,z) = \sum_{n=1}^{\infty} \overline{p}_n(x,z) = \sum_{n=0}^{\infty} \overline{p}_n(x,z) = \mu_x(\{z\}).$$

Case 2: $x = z$. In this case, we have

$$\sum_{y\in E} \overline{p}_n(x,y)p(y,x) = \sum_{y\in E} \mathbf{P}_x\big[X_n = y;\, \tau_x^1 > n;\, X_{n+1} = x\big] = \mathbf{P}_x\big[\tau_x^1 = n+1\big].$$

Thus (since $\mathbf{P}_x[\tau_x^1 = 0] = 0$)

$$\mu_x p(\{x\}) = \sum_{n=0}^{\infty} \mathbf{P}_x\big[\tau_x^1 = n+1\big] = 1 = \mu_x(\{x\}).$$

□

Corollary 17.48 *If X is positive recurrent, then $\pi := \frac{\mu_x}{\mathbf{E}_x[\tau_x^1]}$ is an invariant distribution for any $x \in E$.*

Theorem 17.49 *If X is irreducible, then X has at most one invariant distribution.*

Remark 17.50

(i) One could in fact show that if X is irreducible and recurrent, then an invariant measure of X is unique up to a multiplicative factor. However, the proof is a little more involved. Since we will not need the statement here, we leave its proof as an exercise (compare Exercise 17.6.6; see also [39, Section 6.5]).

(ii) For transient X, there can be more than one invariant measure. For example, consider the asymmetric random walk on $\mathbb{Z}$ that jumps one step to the right with probability r and one step to the left with probability $1-r$ (for some $r \in (0,1)$). The invariant measures are the nonnegative linear combinations of the measures μ_1 and μ_2 given by $\mu_1(\{x\}) \equiv 1$ and $\mu_2(\{x\}) = (r/(1-r))^x$, $x \in \mathbb{Z}$. X is transient if and only if $r \neq 1/2$, in which case we have $\mu_1 \neq \mu_2$. ◊

Proof Let π and ν be invariant distributions. Choose an arbitrary probability vector $(g_n)_{n\in\mathbb{N}}$ with $g_n > 0$ for all $n \in \mathbb{N}$. Define the stochastic matrix $\widetilde{p}(x,y) = \sum_{n=1}^{\infty} g_n p^n(x,y)$. Then $\widetilde{p}(x,y) > 0$ for all $x, y \in E$ and $\pi\widetilde{p} = \pi$ as well as $\nu\widetilde{p} = \nu$.

Consider now the signed measure $\mu = \pi - \nu$. We have $\mu\widetilde{p} = \mu$. If we had $\mu \neq 0$, then there would exist (since $\mu(E) = 0$) points $x_1, x_2 \in E$ with $\mu(\{x_1\}) > 0$ and $\mu(\{x_2\}) < 0$. Clearly, for every $y \in E$, this would imply $|\mu(\{x_1\})\widetilde{p}(x_1,y) + \mu(\{x_2\})\widetilde{p}(x_2,y)| < |\mu(\{x_1\})\widetilde{p}(x_1,y)| + |\mu(\{x_2\})\widetilde{p}(x_2,y)|$; hence

$$\begin{aligned}\|\mu\widetilde{p}\|_{TV} &= \sum_{y\in E}\Big|\sum_{x\in E}\mu(\{x\})\widetilde{p}(x,y)\Big| \\ &< \sum_{y\in E}\sum_{x\in E}\big|\mu(\{x\})\big|\widetilde{p}(x,y) = \sum_{x\in E}\big|\mu(\{x\})\big| = \|\mu\|_{TV}.\end{aligned}$$

Since this is a contradiction, we conclude that $\mu = 0$. □

Recall that $\mathcal{I}$ is the set of invariant distributions of X.

Theorem 17.51 *Let X be irreducible. X is positive recurrent if and only if $\mathcal{I} \neq \emptyset$. In this case, $\mathcal{I} = \{\pi\}$ with*

$$\pi\big(\{x\}\big) = \frac{1}{\mathbf{E}_x[\tau_x^1]} > 0 \quad \textit{for all } x \in E.$$

Proof If X is positive recurrent, then $\mathcal{I} \neq \emptyset$ by Corollary 17.48. Now let $\mathcal{I} \neq \emptyset$ and $\pi \in \mathcal{I}$. As X is irreducible, we have $\pi(\{x\}) > 0$ for all $x \in E$. Let $\mathbf{P}_\pi = \sum_{x \in E} \pi(\{x\})\mathbf{P}_x$. Fix an $x \in E$ and for $n \in \mathbb{N}_0$, let

$$\sigma_x^n = \sup\{m \le n : X_m = x\} \in \mathbb{N}_0 \cup \{-\infty\}$$

be the time of last entrance in x before time n. (Note that this is not a stopping time.) By the Markov property, for all $k \le n$,

$$\begin{aligned}
\mathbf{P}_\pi\big[\sigma_x^n = k\big] &= \mathbf{P}_\pi[X_k = x, X_{k+1} \neq x, \ldots, X_n \neq x] \\
&= \mathbf{P}_\pi[X_{k+1} \neq x, \ldots, X_n \neq x \mid X_k = x]\mathbf{P}_\pi[X_k = x] \\
&= \pi\big(\{x\}\big)\mathbf{P}_x[X_1, \ldots, X_{n-k} \neq x] \\
&= \pi\big(\{x\}\big)\mathbf{P}_x\big[\tau_x^1 \ge n - k + 1\big].
\end{aligned}$$

Hence, for every $n \in \mathbb{N}_0$ (since $\mathbf{P}_y[\tau_x^1 < \infty] = 1$ for all $y \in E$),

$$\begin{aligned}
1 &= \sum_{k=0}^{n} \mathbf{P}_\pi\big[\sigma_x^n = k\big] + \mathbf{P}_\pi\big[\sigma_x^n = -\infty\big] \\
&= \pi\big(\{x\}\big) \sum_{k=0}^{n} \mathbf{P}_x\big[\tau_x^1 \ge n - k + 1\big] + \mathbf{P}_\pi\big[\tau_x^1 \ge n + 1\big] \\
&\xrightarrow{n\to\infty} \pi\big(\{x\}\big) \sum_{k=1}^{\infty} \mathbf{P}_x\big[\tau_x^1 \ge k\big] = \pi\big(\{x\}\big)\mathbf{E}_x\big[\tau_x^1\big].
\end{aligned}$$

Therefore, $\mathbf{E}_x[\tau_x^1] = \frac{1}{\pi(\{x\})} < \infty$, and thus X is positive recurrent. □

Example 17.52 Let $(p_x)_{x \in \mathbb{N}_0}$ be numbers in $(0, 1]$ and let X be an irreducible Markov chain on $\mathbb{N}_0$ with transition matrix

$$p(x, y) = \begin{cases} p_x, & \text{if } y = x + 1, \\ 1 - p_x, & \text{if } y = 0, \\ 0, & \text{else.} \end{cases}$$

If μ is an invariant measure, then the equations for $\mu p = \mu$ read

$$\mu(\{n\}) = p_{n-1}\mu(\{n-1\}) \quad \text{for } n \in \mathbb{N},$$
$$\mu(\{0\}) = \sum_{n=0}^{\infty} \mu(\{n\})(1 - p_n).$$

Hence we get

$$\mu(\{n\}) = \mu(\{0\}) \prod_{k=0}^{n-1} p_k$$

and (note that the sum is a telescope sum)

$$\mu(\{0\}) = \mu(\{0\}) \sum_{n=0}^{\infty} (1 - p_n) \prod_{k=0}^{n-1} p_k = \mu(\{0\}) \left(1 - \prod_{n=0}^{\infty} p_n\right).$$

Hence there exists a nontrivial invariant measure μ (that is, $\mu(\{0\})$ can be chosen strictly positive) if and only if $\prod_{n=0}^{\infty} p_n = 0$. This, however, is true if and only if $\sum_{n=0}^{\infty}(1 - p_n) = \infty$. Using a Borel–Cantelli argument, it is not hard to show that this is exactly the condition for recurrence of X.

If $\mu \neq 0$, then μ is a finite measure if and only if

$$M := \sum_{n=0}^{\infty} \prod_{k=0}^{n-1} p_k < \infty.$$

Hence X is positive recurrent if and only if $M < \infty$. In fact, it is not hard to show that M is the expected time to return to 0; hence the criterion for positive recurrence could also be deduced by Theorem 17.51.

A necessary condition for $M < \infty$ is of course that the series $\sum_{n=0}^{\infty}(1 - p_n)$ diverge; that is, that X is recurrent. One sufficient condition for $M < \infty$ is

$$\sum_{n=0}^{\infty} \exp\left(-\sum_{k=0}^{n-1}(1 - p_k)\right) < \infty.$$

◊

Exercise 17.6.1 Consider the Markov chain from Fig. 17.1 (p. 368). Determine the set of all invariant distributions. Show that the states 6, 7 and 8 are positive recurrent and compute the expected first entrance times

$$\mathbf{E}_6[\tau_6] = \frac{17}{4}, \qquad \mathbf{E}_7[\tau_7] = \frac{17}{5} \quad \text{and} \quad \mathbf{E}_8[\tau_8] = \frac{17}{8}.$$

Exercise 17.6.2 Let $X = (X_t)_{t \geq 0}$ be a Markov chain on E in continuous time with Q-matrix q. Show that a probability measure π on E is an invariant distribution for X if and only if $\sum_{x \in E} \pi(\{x\}) q(x, y) = 0$ for all $y \in E$.

Exercise 17.6.3 Let G be a countable Abelian group and let p be the transition matrix of an irreducible random walk X on G. That is, we have $p(hg, hf) = p(g, f)$ for all $h, g, f \in G$. (This generalizes the notion of a random walk on $\mathbb{Z}^D$.) Use Theorem 17.51 to show that X is positive recurrent if and only if G is finite.

Exercise 17.6.4 Let $r \in [0, 1]$ and let X be the Markov chain on $\mathbb{N}_0$ with transition matrix (see Fig. 17.2 on p. 369)

$$p(x, y) = \begin{cases} 1, & \text{if } x = 0 \text{ and } y = 1, \\ r, & \text{if } y = x + 1 \geq 2, \\ 1 - r, & \text{if } y = x - 1, \\ 0, & \text{else.} \end{cases}$$

Compute the invariant measure and show the following using Theorem 17.51:

(i) If $r \in (0, \frac{1}{2})$, then X is positive recurrent.
(ii) If $r = \frac{1}{2}$, then X is null recurrent.
(iii) If $r \in \{0\} \cup (\frac{1}{2}, 1]$, then X is transient.

Exercise 17.6.5

(i) Use a direct argument to show that the Markov chain in Example 17.52 is recurrent if and only if $\sum_{n=0}^{\infty}(1 - p_n) = \infty$.
(ii) Show that the expected time to return to 0 is M and infer that the chain is positive recurrent if and only if $M < \infty$.
(iii) Give examples of sequences $(p_x)_{x\in\mathbb{N}_0}$ such that the chain is (a) transient, (b) null recurrent, (c) positive recurrent, and (d) positive recurrent but

$$\sum_{n=0}^{\infty} \exp\left(-\sum_{k=0}^{n-1}(1 - p_k)\right) = \infty.$$

Exercise 17.6.6 Let X be irreducible and recurrent. Show that, as claimed in Remark 17.50, the invariant measure is unique up to constant multiples.

Hint: Let $\pi \neq 0$ be an invariant measure for X and abbreviate

$$\mathbf{P}_\pi = \sum_{x\in E} \pi\big(\{x\}\big)\mathbf{P}_x$$

(note that, in general, this need not be a finite measure). Let $x, y \in E$ with $x \neq y$ and deduce by induction that

$$\pi\big(\{y\}\big) = \mathbf{P}_\pi\big[\tau_x^1 \geq n, X_0 \neq x, X_n = y\big] + \sum_{k=1}^{n} \mathbf{P}_\pi\big[\tau_x^1 \geq k, X_0 = x, X_k = y\big].$$

Infer that

$$\pi\big(\{y\}\big) \ge \sum_{k=1}^{\infty} \mathbf{P}_\pi\big[\tau_x^1 \ge k,\, X_0 = x,\, X_k = y\big] = \pi\big(\{x\}\big)\mu_x\big(\{y\}\big),$$

where μ_x is the invariant measure defined in Theorem 17.47. Now use the fact that $\pi p^n = \pi$ and $\mu_x p^n = \mu_x$ for all $n \in \mathbb{N}$ to conclude that even $\pi(\{y\}) = \pi(\{x\})\mu_x(\{y\})$ holds.

17.7 Stochastic Ordering and Coupling

In many situations, for the comparison of two distributions, it is helpful to construct a product space such that the two distributions are the marginal distributions but are not necessarily independent. We first introduce the abstract principle of such *couplings* and then give some examples.

There are many concepts to order probability measures on $\mathbb{R}$ or $\mathbb{R}^d$ such that the "larger" one has a greater preference for large values than the "smaller" one. As one of the most prominent orders we present here the so-called stochastic order and illustrate its connection with couplings. As an excuse for presenting this section in a chapter on Markov chains, we fill finally use a simple Markov chain in order to prove a theorem on the stochastic order of binomial distributions.

Definition 17.53 Let $(E_1, \mathcal{E}_1, \mu_1)$ and $(E_2, \mathcal{E}_2, \mu_2)$ be probability spaces. A probability measure μ on $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ with $\mu(\,\cdot \times E_2) = \mu_1$ and $\mu(E_1 \times \cdot\,) = \mu_2$ is called a *coupling* of μ_1 and μ_2.

Clearly, the product measure $\mu = \mu_1 \otimes \mu_2$ is a coupling, but in many situations there are more interesting ones.

Example 17.54 Let X be a real random variable and let $f, g : \mathbb{R} \to \mathbb{R}$ be monotone increasing functions with $\mathbf{E}[f(X)^2] < \infty$ and $\mathbf{E}[g(X)^2] < \infty$. We want to show that the random variables $f(X)$ and $g(X)$ are nonnegatively correlated.

To this end, let Y be an *independent copy* of X; that is, a random variable with $\mathbf{P}_Y = \mathbf{P}_X$ that is independent of X. Note that $\mathbf{E}[f(X)] = \mathbf{E}[f(Y)]$ and $\mathbf{E}[g(X)] = \mathbf{E}[g(Y)]$. For all numbers $x, y \in \mathbb{R}$, we have $(f(x) - f(y))(g(x) - g(y)) \ge 0$. Hence

$$\begin{aligned} 0 &\le \mathbf{E}\big[\big(f(X) - f(Y)\big)\big(g(X) - g(Y)\big)\big] \\ &= \mathbf{E}\big[f(X)g(X)\big] - \mathbf{E}\big[f(X)\big]\mathbf{E}\big[g(Y)\big] + \mathbf{E}\big[f(Y)g(Y)\big] - \mathbf{E}\big[f(Y)\big]\mathbf{E}\big[g(X)\big] \\ &= 2\mathbf{Cov}\big[f(X), g(X)\big]. \end{aligned}$$

◊

Example 17.55 Let (E, ϱ) be a Polish space. For two probability measures P and Q on $(E, \mathcal{B}(E))$, denote by $K(P, Q) \subset \mathcal{M}_1(E \times E)$ the set of all couplings of P

and Q. The so-called *Wasserstein metric* on $\mathcal{M}_1(E)$ is defined by

$$d_W(P, Q) := \inf\left\{ \int \varrho(x, y)\varphi\big(d(x, y)\big) : \varphi \in K(P, Q) \right\}. \tag{17.26}$$

It can be shown that (this is the Kantorovich–Rubinstein theorem [84]; see also [37, pp. 420ff])

$$d_W(P, Q) = \sup\left\{ \int f\, d(P - Q) : f \in \mathrm{Lip}_1(E; \mathbb{R}) \right\}. \tag{17.27}$$

Compare this representation of the Wasserstein metric with that of the total variation norm,

$$\|P - Q\|_{TV} = \sup\left\{ \int f\, d(P - Q) : f \in \mathcal{L}^\infty(E) \text{ with } \|f\|_\infty \le 1 \right\}. \tag{17.28}$$

In fact, we can also give a definition for the total variation in terms of a coupling: Let $D := \{(x, x) : x \in E\}$ be the diagonal in $E \times E$. Then

$$\|P - Q\|_{TV} = \inf\big\{ \varphi\big((E \times E) \setminus D\big) : \varphi \in K(P, Q) \big\}. \tag{17.29}$$

See [60] for a comparison of different metrics on $\mathcal{M}_1(E)$. ◊

As an example of a more involved coupling, we quote the following theorem that is due to Skorohod.

Theorem 17.56 (Skorohod coupling) *Let $\mu, \mu_1, \mu_2, \ldots$ be probability measures on a Polish space E with $\mu_n \overset{n\to\infty}{\longrightarrow} \mu$. Then there exists a probability space $(\Omega, \mathcal{A}, \mathbf{P})$ with random variables $X, X_1, X_2, \ldots$ with $\mathbf{P}_X = \mu$ and $\mathbf{P}_{X_n} = \mu_n$ for every $n \in \mathbb{N}$ such that $X_n \overset{n\to\infty}{\longrightarrow} X$ almost surely.*

Proof See, e.g., [83, p. 79]. □

We now come to the concept of stochastic order.

Definition 17.57 Let $\mu_1, \mu_2 \in \mathcal{M}_1(\mathbb{R}^d)$. We write $\mu_1 \le_{\mathrm{st}} \mu_2$ if

$$\int f\, d\mu_1 \le \int f\, d\mu_2$$

for every monotone increasing bounded function $f : \mathbb{R}^d \to \mathbb{R}$. In this case, we say that μ_2 is *stochastically larger* than μ_1.

Evidently, $\le_{\mathrm{st}}$ is a partial order on $\mathcal{M}_1(\mathbb{R}^d)$. The stochastic order belongs to the class of so-called integral orders that are defined by the requirement that the integrals with respect to a certain class of functions (here: monotone increasing and bounded) are ordered. Other classes of functions that are often considered are convex functions or indicator functions on lower or upper orthants.

Let F_1 and F_2 be the distribution functions of μ_1 and μ_2. Clearly, $\mu_1 \le_{\text{st}} \mu_2$ implies $F_1(x) \ge F_2(x)$ for all $x \in \mathbb{R}^d$. If $d = 1$, then both statements are equivalent. However, for $d \ge 2$, the condition $F_1 \ge F_2$ is weaker than $\mu_1 \le_{\text{st}} \mu_2$. For example, consider $d = 2$ and

$$\mu_1 = \frac{1}{2}\delta_{(0,0)} + \frac{1}{2}\delta_{(1,1)} \quad \text{and} \quad \mu_2 = \frac{1}{2}\delta_{(1,0)} + \frac{1}{2}\delta_{(0,1)}.$$

The partial order defined by the comparison of the distribution functions is called (lower) orthant order.

For a survey on different orders of probability measures, see, e.g., [120].

The following theorem was shown by Strassen [160] in larger generality for integral orders.

Theorem 17.58 (Strassen's theorem) *Let*

$$L := \big\{(x_1, x_2) \in \mathbb{R}^d \times \mathbb{R}^d : x_1 \le x_2\big\}.$$

Then $\mu_1 \le_{\text{st}} \mu_2$ if and only if there is a coupling φ of μ_1 and μ_2 with $\varphi(L) = 1$.

Proof Let φ be such a coupling. For monotone increasing bounded $f : \mathbb{R}^d \to \mathbb{R}$, we have $f(x_1) - f(x_2) \le 0$ for every $x = (x_1, x_2) \in L$; hence $\int f\, d\mu_1 - \int f\, d\mu_2 = \int_L (f(x_1) - f(x_2))\varphi(dx) \le 0$ and thus $\mu_1 \le_{\text{st}} \mu_2$.

Now assume $\mu_1 \le_{\text{st}} \mu_2$. We only consider the case $d = 1$ (see [120, Theorem 3.3.5] for $d \ge 2$). Here $F((x_1, x_2)) := \min(F_1(x_1), F_2(x_2))$ defines a distribution function on $\mathbb{R} \times \mathbb{R}$ (see Exercise 1.5.5) that corresponds to a coupling φ with $\varphi(L) = 1$. A somewhat more explicit representation can be obtained using random variables. Let U be a random variable that is uniformly distributed on $(0, 1)$. Then

$$X_i := F_i^{-1}(U) := \inf\big\{x \in \mathbb{R} : F_i(x) \ge U\big\}$$

is a real random variable with distribution μ_i (see proof of Theorem 1.104). Clearly, we have $X_1 \le X_2$ almost surely; that is, $\mathbf{P}[(X_1, X_2) \in L] = 1$. Evidently, the distribution function of (X_1, X_2) is F. □

While Strassen's theorem yields the existence of an abstract coupling, in many examples a natural coupling can be established and used as a tool for proving, e.g., stochastic orders.

Example 17.59 Let $n \in \mathbb{N}$ and $0 \le p_1 \le p_2 \le 1$. Let $Y_1, \ldots, Y_n$ be independent random variables that are uniformly distributed on $[0, 1]$. Define $X_i = \#\{k \le n : Y_k \le p_i\}$, $i = 1, 2$. Then $X_i \sim b_{n,p_i}$ and $X_1 \le X_2$ almost surely. This coupling shows that $b_{n,p_1} \le_{\text{st}} b_{n,p_2}$.

An even simpler coupling can be used to show that $b_{m,p} \le_{\text{st}} b_{n,p}$ for $m \le n$ and $p \in [0, 1]$. ◊

Theorem 17.60 *Let $n_1, n_2 \in \mathbb{N}$ and $p_1, p_2 \in (0,1)$. We have $b_{n_1,p_1} \le_{\text{st}} b_{n_2,p_2}$ if and only if*

$$(1-p_1)^{n_1} \ge (1-p_2)^{n_2} \tag{17.30}$$

and

$$n_1 \le n_2. \tag{17.31}$$

Proof (The proof follows the exposition in [100, Section 3].)

Since $b_{n_i,p_i}(\{0\}) = (1-p_i)^{n_i}$, conditions (17.30) and (17.31) are clearly necessary for $b_{n_1,p_1} \le_{\text{st}} b_{n_2,p_2}$. Hence we only have to show sufficiency of the two conditions.

Assume that (17.30) and (17.31) hold. By Example 17.59, it is enough to consider the smallest p_2 that fulfills (17.30). Hence we assume $(1-p_1)^{n_1} = (1-p_2)^{n_2}$. Define $\lambda := -n_1 \log(1-p_1) = -n_2 \log(1-p_2)$. We will construct a binomially distributed random variable by throwing a Poi_λ-distributed number T of balls in n_i boxes and count the number of nonempty boxes. More precisely, let $T \sim \text{Poi}_\lambda$ and let $X_1, X_2, \ldots$ be independent and uniformly distributed on $[0,1]$ and independent of T. For $n \in \mathbb{N}$, $t \in \mathbb{N}_0$ and $l = 1, \ldots, n$, define

$$M_{n,t,l} = \#\big\{s \le t : X_s \in \big((l-1)/n, l/n\big]\big\}$$

and the number of nonempty boxes after t balls are thrown:

$$N_{n,t} := \sum_{l=1}^{n} \mathbb{1}_{\{M_{n,t,l} > 0\}}.$$

By Theorem 5.35, the random variables $M_{n,T,1}, \ldots, M_{n,T,n}$ are independent and $\text{Poi}_{\lambda/n}$-distributed. In particular, we have

$$\mathbf{P}[M_{n_i,T,l} > 0] = 1 - e^{-\lambda/n_i} = p_i$$

and thus $N_{n_i,T} \sim b_{n_i,p_i}$, $i = 1, 2$. Hence it suffices to show that $N_{n_1,T} \le_{\text{st}} N_{n_2,T}$. For this in turn it is enough to show

$$N_{n_1,t} \le_{\text{st}} N_{n_2,t} \quad \text{for all } t \in \mathbb{N}_0. \tag{17.32}$$

In fact, let $f : \{0, \ldots, n\} \to \mathbb{R}$ be monotone increasing. Then

$$\begin{aligned}\mathbf{E}\big[f(N_{n_1,T})\big] &= \sum_{t=0}^{\infty} \mathbf{E}\big[f(N_{n_1,t})\big]\mathbf{P}[T=t] \\ &\le \sum_{t=0}^{\infty} \mathbf{E}\big[f(N_{n_2,t})\big]\mathbf{P}[T=t] = \mathbf{E}\big[f(N_{n_2,T})\big].\end{aligned}$$

We use an induction argument to show (17.32). For $t = 0$, the claim holds trivially. Now assume that (17.32) holds for some given $t \in \mathbb{N}_0$. We are now at the point to use a Markov chain. Note that (for fixed n), $(N_{n,t})_{t=0,1,\ldots}$ is a Markov chain with state space $\{0,\ldots,n\}$ and transition matrix

$$p_n(k,l)=\begin{cases}k/n, & \text{if } l=k,\\ 1-k/n, & \text{if } l=k+1,\\ 0, & \text{otherwise.}\end{cases}$$

We define for $k,l=0,\ldots,n$

$$h_{n,l}(k)=\sum_{j=l}^{n}p_n(k,j)=\begin{cases}0, & \text{if } k<l-1,\\ 1-k/n, & \text{if } k=l-1,\\ 1, & \text{if } k>l-1.\end{cases}$$

Then $\mathbf{P}[N_{n,t+1}\geq l]=\mathbf{E}[h_{n,l}(N_{n,t})]$ and $h_{n,l}(k)$ is monotone increasing both in k and in n. Hence by the induction hypothesis, we have

$$\begin{aligned}\mathbf{P}[N_{n_1,t+1}\geq l]=\mathbf{E}\big[h_{n_1,l}(N_{n_1,t})\big]&\leq\mathbf{E}\big[h_{n_1,l}(N_{n_2,t})\big]\\ &\leq\mathbf{E}\big[h_{n_2,l}(N_{n_2,t})\big]=\mathbf{P}[N_{n_2,t+1}\geq l].\end{aligned}$$

We conclude that $N_{n_1,t+1}\leq_{\mathrm{st}}N_{n_2,t+1}$ which completes the induction and the proof of the theorem. □

Exercise 17.7.1 Use an elementary direct coupling argument to show the claim of Theorem 17.60 for the case $n_2/n_1\in\mathbb{N}$.

Exercise 17.7.2 For the Poisson distribution, show that

$$\mathrm{Poi}_{\lambda_1}\leq_{\mathrm{st}}\mathrm{Poi}_{\lambda_2}\quad\Longleftrightarrow\quad\lambda_1\leq\lambda_2.$$

Exercise 17.7.3 Let $n\in\mathbb{N}$, $p\in(0,1)$ and $\lambda>0$. Show that

$$b_{n,p}\leq_{\mathrm{st}}\mathrm{Poi}_\lambda\quad\Longleftrightarrow\quad(1-p)^n\geq e^{-\lambda}.$$