CHAPTER 8

Markov Chains

This chapter is inspired partly by [**Nev70**] and [**R.88**].

8.1. Definitions and Basic Properties

Let S be a finite or countable set, which we call state space, endowed with the σ-algebra $\mathcal{P}(S) = \{A : A \subset S\}$.

Definition 8.1.1. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. A sequence of S-valued random variables $(X_n)_{n\geq 0}$ is a Markov chain (with state space S) if for all $n \geq 0$, $X_n : \Omega \to S$ is measurable and*

$$P(X_{n+1} = x_{n+1}|X_n = x_n, \dots, X_0 = x_0) = P(X_{n+1} = x_{n+1}|X_n = x_n) \qquad (8.1.1)$$

for all $x_0, \dots, x_{n+1} \in S$.

We avoided to mention that (8.1.1) holds P-almost surely, and will usually continue doing so in the sequel. We don't yet worry about the structure of the underlying probability space $(\Omega, \mathcal{F}, P)$, although a canonical choice will be made in Section 8.1.1.

We will mostly consider the case where the probability $P(X_{n+1} = x_{n+1}|X_n = x_n)$ does not depend on n, that is where

$$P(X_{n+1} = y|X_n = x) = P(X_1 = y|X_0 = x)$$

for all $n \geq 1$. In such case, the chain is called homogeneous, and the dependence among the random variables is determined by the numbers $P(X_1 = y|X_0 = x)$, called transition probabilities. Observe that these satisfy $\sum_{y\in S} P(X_1 = y|X_0 = x) = 1$ (P-a.s.) for all $x \in S$. We are interested in the study of Markov chains for which the transition probabilities are specified a priori.

Definition 8.1.2. *A collection $Q(x, y)$, $x, y \in S$, is called a transition probability matrix if $Q(x, y) \in [0, 1]$ and if $\sum_{y\in S} Q(x, y) = 1$ for all $x \in S$. A homogeneous Markov chain $(X_n)_{n\geq 0}$ has transition probability matrix Q if*

$$P(X_{n+1} = y|X_n = x) = Q(x, y) \quad P\text{-a.s.}$$

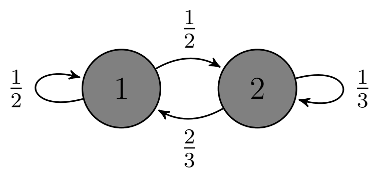
for all $n \geq 0$, $x, y \in S$.

Although it might seem trivial at this point, observe that for all $y \in S$, $x \mapsto Q(x, y)$ is measurable. The existence of a Markov chain associated to a transition probability matrix will be shown in Section 8.1.1. Before going further we give a serie of examples.

Example 8.1.1. Independent variables furnish a trivial example of Markov chain. Let $(X_n)_{n\geq 0}$ be a sequence of i.i.d random variables with distribution μ over $(S, \mathcal{P}(S))$. If we define $Q(x,y) := \mu(y)$, then by independence,

$$P(X_{n+1} = x_{n+1} | X_n = x_n, \ldots, X_0 = x_0) = \mu(x_{n+1}) \equiv Q(x_n, x_{n+1}) \,.$$

Example 8.1.2. The two state Markov chain is defined for $S = \{1, 2\}$. An example of a transition matrix is given in the following graphical representation:



Example 8.1.3. The random walk on $S = \mathbb{Z}^d$. We considered the simplest case of random walk in Section 6. Consider a sequence $(Y_n)_{n\geq 1}$ of $\mathbb{Z}^d$-valued independent identically distributed random variables, and denote their common distribution by p. Define $S_0 := 0$, and for all $n \geq 1$, $S_n := \sum_{k=1}^{n} Y_k$. The sequence $(S_n)_{n\geq 0}$ is called a random walk on $\mathbb{Z}^d$. Observe that, since Y_{n+1} is independent of $S_1, \ldots, S_n$, we have

$$\begin{aligned} P(S_{n+1} &= x_{n+1} | S_n = x_n, \ldots, S_0 = x_0) \\ &= P(Y_{n+1} = x_{n+1} - x_n | S_n = x_n, \ldots, S_0 = x_0) \\ &= P(Y_{n+1} = x_{n+1} - x_n) \\ &= P(Y_{n+1} = x_{n+1} - x_n | S_n = x_n) \\ &= P(S_{n+1} = x_{n+1} | S_n = x_n) \,. \end{aligned}$$

Therefore, since $P(Y_{n+1} = x_{n+1} - x_n) = p(x_{n+1} - x_n)$, $(S_n)_{n\geq 0}$ is a Markov chain with state space $S = \mathbb{Z}^d$ and transition matrix $Q(x,y) = p(y - x)$. When

$$p(x) = \begin{cases} \frac{1}{2d} & \text{if } \|x\|_1 = 1 \,. \\ 0 & \text{otherwise,} \end{cases} \tag{8.1.2}$$

that is when $p(\pm e_i) = \frac{1}{2d}$ where the $e_1, \ldots, e_d$ are the canonical unit vectors of $\mathbb{R}^d$, the random walk is called simple, symmetric. More will be said on random walks in Section 8.3.2.

Example 8.1.4. Uniform Random Walk on a Graph. Let $G = (V, E)$ be a simple graph without loops. For each $x \in V$, we assume that $A_x := \{y \in V : \{x, y\} \in E\}$ is finite: $|A_x| < \infty$. Setting $S \equiv V$, one can define a transition matrix by

$$Q(x,y) := \begin{cases} \frac{1}{|A_x|} & \text{if } \{x, y\} \in E \,, \\ 0 & \text{otherwise} \,. \end{cases} \tag{8.1.3}$$

The simple random walk of the previous example is a particular case.

Example 8.1.5. The Ehrenfest chain. Consider two urns with a total of r balls. Each urn can be considered as a box with a certain number of molecules, the total number of molecules being r. At each time step, a ball is chosen at random (in either box) and its position switched to the other box. Let X_n be the number of balls in the first box at time n. Then $(X_n)_{n\geq 0}$ is a Markov Chain with state space $S = \{0, 1, 2, \dots, r\}$ and transition matrix Q given by

$$Q(k, k+1) = \frac{r-k}{r}, \quad Q(k, k-1) = \frac{k}{r},$$

and zero otherwise.

EXAMPLE 8.1.6. Birth and death chains. Consider $S = \{0, 1, 2, \dots\}$, in which $X_n = x$ means that population at time n is x, and $Q(x, y) > 0$ only if $|x - y| \leq 1$. Therefore, the chain is determined by the transition probabilities $r_x = Q(x, x)$, $q_x = Q(x, x-1)$ (clearly, $q_0 = 0$), $p_x = Q(x, x+1)$. See Figure 1.

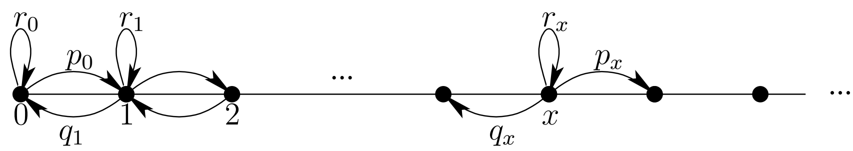


FIGURE 1. The birth and death chain.

EXAMPLE 8.1.7. Renewal chains. Consider $S = \{0, 1, 2, \dots\}$ and a sequence $(p_k)_{k\geq 1}$ with $\sum_k p_k = 1$. Then $Q(0, k) = p_k$, and $Q(k, k-1) = 1$ for all $k \geq 2$.

EXAMPLE 8.1.8. The Branching Process was introduced by Galton and Watson to understand extinction or survival of family names. Let $(Y_k^{(n)})_{n\geq 0, k\geq 1}$ be an array of i.i.d. $\mathbb{N}$-valued random variables, with distribution ρ: $P(Y_k^{(n)} = j) = \rho(j)$ for all $j \geq 1$. $Y_k^{(n)}$ is the number of children of the kth individual of the nth generation. Let $X_0 := 1$, and define the total number of individuals of the $n+1$th generation:

$$X_{n+1} := \sum_{k=1}^{X_n} Y_k^{(n)}. \tag{8.1.4}$$

Let us show that $(X_n)_{n\geq 0}$ is a Markov chain with state space $S = \{0, 1, 2, \dots\}$.

$$\begin{aligned} P(X_{n+1} = y | X_n = x_n, \dots, X_0 = x_0) &= P\Big(\sum_{k=1}^{X_n} Y_k^{(n)} = y \Big| X_n = x_n, \dots, X_0 = x_0\Big) \\ &= P\Big(\sum_{k=1}^{x_n} Y_k^{(n)} = y\Big). \end{aligned}$$

Since the variables $Y_k^{(n)}$ are independent, the distribution of the sum $\sum_{k=1}^{x_n} Y_k^{(n)}$ is given by the convolution $\rho * \rho * \cdots * \rho$ (x_n times), which we denote by ρ^{*x_n}. This shows that $(X_n)_{n\geq 0}$ is a Markov chain with transition probability matrix given by

$$Q(x, y) = \rho^{*x}(y) \quad \forall x, y \in S \tag{8.1.5}$$

It is well known that in the subcritical case, i.e. when $\lambda := E[Y_1^{(0)}] \leq 1$, the population dies out P-almost surely. In the supercritical case, i.e. for $\lambda > 1$, then the population explodes with positive probability.

We define the iterates of a transition matrix as follows: $Q^{(1)} := Q$, and for $n \geq 2$,

$$Q^{(n)}(x,z) := \sum_{y \in S} Q^{(n-1)}(x,y)Q(y,z)\,. \tag{8.1.6}$$

Clearly, each $Q^{(n)}$ is well defined and is again a transition matrix. Let us give an important equivalent characterization of Markov chains.

LEMMA 8.1.1. *Let Q be a transition probability matrix. A sequence $(X_n)_{n\geq 0}$ is a Markov chain with transition matrix Q if and only if for all $n \geq 1$ and all $x_0, \dots, x_n \in S$,*

$$P(X_0 = x_0, \dots, X_n = x_n) = P(X_0 = x_0)Q(x_0,x_1)\dots Q(x_{n-1},x_n)\,. \tag{8.1.7}$$

In particular, if $P(X_0 = x_0) > 0$, then

$$P(X_n = y | X_0 = x_0) = Q^{(n)}(x_0, y)\,. \tag{8.1.8}$$

PROOF. Assume $(X_n)_{n\geq 0}$ is a Markov chain with transition matrix Q. If $n = 1$, (8.1.7) is trivial. Indeed, if $P(X_0 = x_0) = 0$ then $P(X_0 = x_0, X_1 = x_1) = 0$ and so $P(X_0 = x_0, X_1 = x_1) = P(X_0 = x_0)Q(x_0,x_1)$. If $P(X_0 = x_0) = 0 > 0$ the same holds. So assume that (8.1.7) holds for n. Again, if $P(X_0 = x_0, \dots, X_n = x_n) = 0$ then $P(X_0 = x_0, \dots, X_{n+1} = x_{n+1}) = 0$ and the result follows. If $P(X_0 = x_0, \dots, X_n = x_n) > 0$ then

$$\begin{aligned}
&P(X_0 = x_0, \dots, X_{n+1} = x_{n+1}) \\
&\quad = P(X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_0 = x_0)P(X_0 = x_0, \dots, X_n = x_n) \\
&\quad = P(X_{n+1} = x_{n+1} | X_n = x_n)P(X_0 = x_0)Q(x_0,x_1)\dots Q(x_{n-1},x_n) \\
&\quad = P(X_0 = x_0)Q(x_0,x_1)\dots Q(x_{n-1},x_n)Q(x_n,x_{n+1})\,,
\end{aligned}$$

which shows the validity of (8.1.7) for $n+1$. For (8.1.8), use (8.1.7) as follows:

$$\begin{aligned}
P(X_n = x_n, X_0 = x_0) &= \sum_{x_1,\dots,x_{n-1}} P(X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) \\
&= P(X_0 = x_0) \sum_{x_1,\dots,x_{n-1}} Q(x_0,x_1)\dots Q(x_{n-1},x_n) \\
&\equiv P(X_0 = x_0)Q^{(n)}(x_0,x_n)\,, \qquad (8.1.9)
\end{aligned}$$

which gives (8.1.8) if $P(X_0 = x_0) > 0$. □

Observe that by (8.1.7), the transition matrix Q completely specifies the evolution of the chain, once the distribution of X_0 is known. Let therefore μ be a distribution on $(S, \mathcal{P}(S))$. When the distribution of X_0 is given by μ, we will denote the law of $(X_n)_{n\geq 0}$ by P_μ. That is, by (8.1.7),

$$P_\mu(X_0 = x_0, \dots, X_n = x_n) = \mu(x_0)Q(x_0,x_1)\dots Q(x_{n-1},x_n)\,. \tag{8.1.10}$$

When μ is a Dirac mass, i.e. $\mu(x) = 1$ for some $x \in S$, we will write P_x rather than P_μ, and interpret x as being a deterministic initial condition. For example, (8.1.8) gives

$$P_x(X_n = y) = Q^{(n)}(x, y)\,. \tag{8.1.11}$$

As can be easily verified, the measure P_μ can be reconstructed by convex combination of the measures $\{P_x\}_{x\in S}$:

$$P_\mu = \sum_{x\in S} \mu(x) P_x\,.$$

We denote expectations with respect to P_x by E_x. We have

$$\begin{aligned} E_x(f(X_n)) &= \sum_{y\in S} f(y) P_x(X_n = y) \\ &= \sum_{y\in S} f(y) Q^{(n)}(x, y) \\ &\equiv Q^{(n)} f(x)\,, \end{aligned} \tag{8.1.12}$$

where for each $n \geq 1$, the function $Q^{(n)} f : S \to \mathbb{R}$ is defined by

$$Q^{(n)} f(x) := \sum_{y\in S} Q^{(n)}(x, y) f(y)\,. \tag{8.1.13}$$

(8.1.8) says that the distribution of X_n, conditionned on X_0, is given by the nth iterate of Q. This distribution can be written as $P(X_n = y|X_0 = x_0) = E[1_{\{X_n=y\}}|X_0 = x_0]$. Since we will also be interested in functions depending on the process, of the form $f : S \to \mathbb{R}$, we might therefore be interested in studying more general conditional expectations of the form $E[f(X_n)|X_0 = x_0]$.

Lemma 8.1.2. *Let $(X_n)_{n\geq 1}$ be a Markov chain with transition matrix Q. If $f : S \to \mathbb{R}$, then for all $n \geq 0$,*

$$E[f(X_{n+1})|X_n = x_n, \dots, X_0 = x_0] = Qf(x_n)\,. \tag{8.1.14}$$

More generally, for any set $\{i_1, \dots, i_k\} \subset \{1, 2, \dots, n-1\}$,

$$E[f(X_{n+1})|X_n = x_n, X_{i_k} = x_{i_k} \dots, X_{i_1} = x_{i_1}] = Qf(x_n)\,. \tag{8.1.15}$$

8.1.1. The Canonical Chain. Up to now the underlying probability space on which the chain is defined hasn't had an important role, but one should of course verify that at least one such space exists.

Theorem 8.1.1. *Let μ be a probability distribution on $(S, \mathcal{P}(S))$ and Q a transition probability matrix. Then there exists a probability space $(\Omega', \mathcal{F}', P'_\mu)$ and a sequence of S-valued random variables $(X_n)_{n\geq 0}$ on $(\Omega', \mathcal{F}', P'_\mu)$ which form a Markov Chain with transition probability matrix Q:*

$$P'_\mu(X_0 = x_0, \dots, X_n = x_n) = \mu(x_0) Q(x_0, x_1) \dots Q(x_{n-1}, x_n)\,. \tag{8.1.16}$$

PROOF. By Theorem 7.1.1, one can construct simultaneously a family of i.i.d. random variables $(Y_n)_{n\geq 1}$ on the product space $\Omega' = [0,1]^{\mathbb{N}}$, with uniform distribution on $[0,1]$ with respect to the Lebesgue measure. Ω' is endowed with the product σ-algebra $\mathcal{F}'$ and P' is the product of Lebesgue measures. Let us enumerate S in an arbitrary way: $S = \{y_1, y_2, \dots\}$. Fix some initial condition $x \in S$. We define a process $(X_n^x)_{n\geq 0}$ on $(\Omega', \mathcal{F}', P')$ as follows. First, $X_0^x := x$. Then, we need to define X_1^x in such a way that $P'(X_1^x = y_k | X_0^x = x) = Q(x, y_k)$ for all $k \geq 1$. Define, for all $z \in S$,

$$\alpha_k(z) := \sum_{1\leq i\leq k} Q(z, y_i)\,.$$

Observe that $0 \leq \alpha_1(z) \leq \alpha_2(z) \leq \cdots \leq 1$, and $\alpha_k(z) \to 1$ when $k \to \infty$. Then, set

$$X_1^x = y_k \quad \text{if and only if} \quad \alpha_{k-1}(x) < Y_1 \leq \alpha_k(x)\,.$$

Clearly, $P'(X_1^x = y_k | X_0^x = x) = P'(\alpha_{k-1}(x) < Y_1 \leq \alpha_k(x)) \equiv Q(x, y_k)$. For $n \geq 2$, X_n^x is defined by

$$X_n^x = y_k \quad \text{if and only if} \quad \alpha_{k-1}(X_{n-1}^x) < Y_n \leq \alpha_k(X_{n-1}^x)\,.$$

One then gets, by the independence and uniformity of the Y_ns,

$$\begin{aligned} P'(X_{n+1}^x = y_k | X_n^x = x_n, \dots, X_0^x = x_0) &= \\ &= P'(\alpha_{k-1}(X_n^x) < Y_n \leq \alpha_k(X_n^x) | X_n^x = x_n, \dots, X_0^x = x_0) \\ &= P'(\alpha_{k-1}(x_n) < Y_n \leq \alpha_k(x_n)) \\ &= Q(x_n, y_k)\,, \end{aligned} \tag{8.1.17}$$

which shows that $(X_n^x)_{n\geq 0}$ is a Markov chain with transition probability matrix Q and initial condition x. One can obtain a chain with initial distribution μ by taking convex combinations. Write the process constructed above $(X_n)_{n\geq 0}$, and denote its law by P'_x, in order to have $P'_x(X_0 = x) = 1$. Now define

$$P'_\mu := \sum_{x\in S} \mu(x) P'_x\,.$$

Then, using Lemma 8.1.1,

$$\begin{aligned} P'_\mu(X_0 = x_0, \dots, X_n = x_n) &= \sum_{x\in S} \mu(x) P'_x(X_0 = x_0, \dots, X_n = x_n) \\ &= \sum_{x\in S} \mu(x) 1_{\{x=x_0\}} Q(x_0, x_1) \dots Q(x_{n-1}, x_n) \\ &= \mu(x_0) Q(x_0, x_1) \dots Q(x_{n-1}, x_n)\,, \end{aligned}$$

which is (8.1.16). □

As will become clearer in the sequel, the study of homogeneous Markov chains is greatly facilitated by the introduction of a certain time translation operator on the process and of its random version, which will lead to the proofs of all recurrence results of Section 8.3. In the present section we construct a canonical space on which this operator will be naturally defined.

Each realization $\omega \in \Omega$ yields a sequence $X_1(\omega), X_2(\omega), \dots,$ which we call a **trajectory** of the chain. A natural candidate for the simplest probability space describing an S-valued Markov Chain $(X_n)_{n\geq 0}$ is therefore the space in which each element ω *is itself* a trajectory, that is, the elements of which are the sequences $\omega = (\omega_0, \omega_1, \dots)$ where each $\omega_k \in S$:

$$\Omega := S^{\{0,1,2,\dots\}}\,. \tag{8.1.18}$$

For each $k \geq 0$, consider the **coordinate map** $X_k : \Omega \to \mathbb{R}$ defined by $X_k(\omega) := \omega_k$. The σ-algebra $\mathcal{F}$ is defined as the smallest collection of subsets of Ω for which each X_k is measurable, that is $\mathcal{F} := \sigma(X_k,\, k \geq 0)$. The σ-algebra $\mathcal{F}$ can also be obtained by considering the σ-algebra generated by **thin cylinders**, i.e. subsets of Ω of the form

$$[x_0, x_1, \dots, x_n] = \{\omega \in \Omega : \omega_0 = x_0, \omega_1 = x_1, \dots, \omega_n = x_n\}\,, \tag{8.1.19}$$

where $x_0, \dots, x_n \in S$. The intersection of two thin cylinders is either empty or is again a thin cylinder. The **algebra of cylinders** is obtained by taking finite unions of thin cylinders, and is denoted $\mathcal{C}$. Then clearly, $\mathcal{F} = \sigma(\mathcal{C})$.

THEOREM 8.1.2. *Let μ be a probability distribution on $(S, \mathcal{P}(S))$ and Q be a transition probability matrix. Then there exists a unique probability measure P_μ on $(\Omega, \mathcal{F})$ such that on $(\Omega, \mathcal{F}, P_\mu)$, the coordinate maps $(X_n)_{n\geq 1}$ form a Markov Chain with state space S, transition probability matrix Q, and initial distribution μ:*

$$P_\mu(X_0 = x_0, \dots, X_n = x_n) = \mu(x_0)Q(x_0, x_1)\dots Q(x_{n-1}, x_n)\,. \tag{8.1.20}$$

PROOF. Consider the probability space $(\Omega', \mathcal{F}', P'_\mu)$ constructed in Theorem 8.1.1, together with the process constructed therein, which we temporarily denote by $(X'_n)_{n\geq 0}$ in order to distinguish it from the coordinate maps on Ω. Consider the map $\varphi : \Omega' \to \Omega$ defined by $\varphi(\omega')_n := X'_n(\omega')$ for all $n \geq 0$.

LEMMA 8.1.3. *φ is measurable: $\varphi^{-1}(A) \in \mathcal{F}'$ for all $A \in \mathcal{F}$.*

PROOF. Let $\mathcal{A} := \{A \in \mathcal{F} : \varphi^{-1}(A) \in \mathcal{F}'\}$. Then $\mathcal{A}$ is a σ-algebra. Moreover, it contains all sets of the form $X_n^{-1}(\{x\})$, $x \in S$, $n \geq 0$, since $\varphi^{-1}(X_n^{-1}(\{x\})) = (X_n \circ \varphi)^{-1}(\{x\}) = {X'}_n^{-1}(\{x\}) \in \mathcal{F}'$ by definition (the ${X'}_n$s are random variables). Therefore, $\mathcal{A} \equiv \mathcal{F}$. □

Since φ is measurable, we can define the image measure $P_\mu := P'_\mu \circ \varphi^{-1}$. We have

$$\begin{aligned} P_\mu(X_0 = x_0, \dots, X_n = x_n) &= P'_\mu(X'_0 = x_0, \dots, X'_n = x_n) \\ &= \mu(x_0)Q(x_0, x_1)\dots Q(x_{n-1}, x_n)\,, \end{aligned}$$

which shows that $(X_n)_{n\geq 0}$ has the wanted properties. Regarding uniqueness, assume $\tilde{P}_\mu$ is another measure also satisfying (8.1.20). But (8.1.20) implies that P_μ and $\tilde{P}_\mu$ coincide on thin cylinders, and since these generate $\mathcal{F}$, they are equal. □

ALTERNATE PROOF OF THEOREM 8.1.2: Let $[x_0, x_1, \dots, x_n]$ be any thin cylinder and define

$$P([x_0, x_1, \dots, x_n]) := \mu(x_0)Q(x_0, x_1)\dots Q(x_{n-1}, x_n)\,. \tag{8.1.21}$$

We need to show that P extends uniquely to a probability on $(\Omega, \mathcal{F})$ and that under P. But this follows immediately from Kolmogorov's Extension Theorem 7.1.2. By (8.1.21), the coordinate maps $(X_n)_{n\geq 0}$ clearly define a Markov chain with transition probability matrix Q and initial distribution μ. □

The following proposition shows that the canonical representation is sufficient for the study of Markov chains, in the sense that one cannot distinguish the distribution of the canonical chain from any other.

PROPOSITION 8.1.1. *Let $(Y_n)_{n\geq 0}$ be a Markov chain with initial distribution μ and transition matrix Q, constructed on some probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P})$. Let $(X_n)_{n\geq 0}$ be the Canonical Markov chain with initial distribution μ and transition matrix Q, constructed on the product space $(\Omega, \mathcal{F}, P)$ as above. Then P is the image of $\widetilde{P}$ under the measurable map $\varphi : \widetilde{\Omega} \to \Omega$ defined by $\varphi(\tilde{\omega}) := (Y_n(\tilde{\omega}))_{n\geq 0}$. That is, $P = \widetilde{P} \circ \varphi^{-1}$.*

PROOF. We already saw in Lemma 8.1.3 that φ is measurable. Now for any thin cylinder $[x_0, \ldots, x_n]$,

$$\begin{aligned}\widetilde{P} \circ \varphi^{-1}([x_0, \ldots, x_n]) &= \widetilde{P}(Y_0 = x_0, \ldots, Y_n = x_n) \\ &= \mu(x_0)Q(x_0, x_1) \cdots Q(x_{n-1}, x_n) \\ &\equiv P(X_0 = x_0, \ldots, X_n = x_n)\,. \end{aligned} \tag{8.1.22}$$

Since thin cylinders generate $\mathcal{F}$, this proves the proposition. □

8.2. The Markov Property

The basic relation defining a Markov chain, (8.1.1), says that conditionnally on a given past up to time n, $X_0, X_1, \ldots, X_n$, the distribution of X_{n+1} depends only on X_n. Since this and time homogeneity suggest a certain translation in time, the canonical space constructed in the previous section appears well adapted to the precise formulation of a more general version of this property: conditionnally on a given past up to time n, $X_0, X_1, \ldots, X_n$, the distribution of the *entire future* $X_{n+1}, X_{n+2}, \ldots$ depends only on X_n. So from now on, the Markov chain under consideration will always be considered as built on the canonical product space Ω defined in (8.1.18). Define the transformation $\theta : \Omega \to \Omega$, called the shift, by

$$\theta(\omega)_n := \omega_{n+1} \quad \forall n \geq 0\,.$$

Since $\theta^{-1}(X_n^{-1}(\{x\})) = X_{n+1}^{-1}(\{x\}) \in \mathcal{F}$, θ is measurable. One can of course iterate the shift: $\theta_1 := \theta$, and $\theta_{n+1} := \theta_n \circ \theta$.

To use the language of conditional expectation, we encode the information contained in the past of n, $X_0, X_1, \ldots, X_n$, in the σ-algebra $\mathcal{F}_n := \sigma(X_0, X_1, \ldots, X_n)$.

THEOREM 8.2.1 (Simple Markov Property). *Let $x \in S$, and $n \geq 1$. Let $\varphi : \Omega \to \mathbb{R}$ be bounded, positive and $\mathcal{F}_n$-measurable. Then for all bounded, positive, measurable $\psi : \Omega \to \mathbb{R}$,*

$$E_x[\varphi \cdot \psi \circ \theta_n] = E_x[\varphi \cdot E_{X_n}(\psi)]\,. \tag{8.2.1}$$

In the right-hand side of (8.2.1) appears the random variable $E_{X_n}(\psi)$, which is just $E_x(\psi)$ evaluated at X_n [1]. Observe that by taking $\varphi = 1_A$ for each $A \in \mathcal{F}_n$, (8.2.1) is equivalent to the P_x-almost sure statement

$$E_x[\psi \circ \theta_n | \mathcal{F}_n] = E_{X_n}[\psi]\,. \tag{8.2.2}$$

Proof. We first consider the case where φ and ψ are indicators of thin cylinders: $\varphi = 1_C$, with $C = [x_0, \dots, x_n]$, $\psi = 1_D$ with $D = [y_0, \dots, y_p]$. We have

$$\begin{aligned} E_{X_n}[\psi] &= \sum_{x'_0,\dots,x'_p} \psi(x'_0, \dots, x'_p) P_{X_n}(X_0 = x'_0, \dots, X_p = x'_p) \\ &= 1_{\{X_n = y_0\}} Q(y_0, y_1) \dots Q(y_{p-1}, y_p)\,, \end{aligned}$$

which leads to

$$E_x[\varphi \cdot E_{X_n}(\psi)] = 1_{\{x_0 = x\}} Q(x_0, x_1) \dots Q(x_{n-1}, x_n) 1_{\{x_n = y_0\}} Q(y_0, y_1) \dots Q(y_{p-1}, y_p)\,.$$

On the other hand,

$$\begin{aligned} E_x[\varphi \cdot \psi \circ \theta_n] &= E_x[1_{\{X_0 = x_0\}} \dots 1_{\{X_n = x_n\}} 1_{\{X_n = y_0\}} 1_{\{X_{n+1} = y_1\}} \dots 1_{\{X_{n+p} = y_p\}}] \\ &= P_x(X_0 = 0, \dots X_n = x_n, X_n = y_0, X_{n+1} = y_1, \dots, X_{n+p} = y_p) \\ &= 1_{\{x_0 = x\}} Q(x_0, x_1) \dots Q(x_{n-1}, x_n) 1_{\{x_n = y_0\}} Q(y_0, y_1) \dots Q(y_{p-1}, y_p)\,, \end{aligned}$$

which shows (8.2.1) in the particular case. We then show that for the same φ, (8.2.1) holds also in the case where $\psi = 1_A$, where $A \in \mathcal{F}$. Consider the class $\mathcal{A} = \{A \in \mathcal{F} : E_x[\varphi \cdot 1_A \circ \theta_n] = E_x[\varphi \cdot E_{X_n}(1_A)]\}$. We know that $\mathcal{A}$ contains all thin cylinders, and therefore all cylinders by summation. It is then easy to verify that $\mathcal{A}$ is a Dynkin system, and so $\mathcal{A} = \mathcal{F}$ by Theorem 3.0.1. Now, the extension to arbitrary bounded positive functions follows by uniform approximation by simple functions. □

A simple application of the Markov Property is the following identity, known as the Chapman-Kolmogorov Equation:

$$P_x(X_{m+n} = y) = \sum_{z \in S} P_x(X_m = z) P_z(X_n = y)\,. \tag{8.2.3}$$

Namely, one can write $P_x(X_{m+n} = y) = E_x[E_x(1_{X_{m+n} = y} | \mathcal{F}_m)]$, and then

$$\begin{aligned} E_x[E_x(1_{X_{m+n} = y} | \mathcal{F}_m)] &= E_x[E_x(1_{X_n = y} \circ \theta_m | \mathcal{F}_m)] \\ &= E_x[E_{X_m}(1_{X_n} = y)] \\ &= E_x[P_{X_m}(X_n = y)] \\ &= \sum_{z \in S} P_x(X_m = z) P_z(X_n = y)\,. \end{aligned} \tag{8.2.4}$$

The reader should convince himself that any other proof of (8.2.3) will necessary end up requiring one or another form of the Simple Markov Property.

[1] Observe here that $x \mapsto E_x(\psi)$ is $\mathcal{P}(S)$-measurable, and that $\omega \mapsto E_{X_{n(\omega)}}(\psi)$ is a random variable.

The Markov Property deserves an extension to the case there the time n is replaced by a random time T. The reason is the following. Suppose we are interested in the following question: if the chain returns back to its starting point with probability one, is it true that it will do so an infinite number of times? This seems clear since at the time of first return, the chain is back at its original position and therefore by the Markov Property the probability of coming back a second time is again one, and so on. Nevertheless, the times at which the chain returns to its starting point are *random*, and the simple Markov Property can't be used in its actual form.

Random times are usually called *stopping times*. We will define them here in the framework of Markov chains; in Section 9 these will be used extensively in the chapter on martingales. A stopping time satisfies a list of properties which we first illustrate on a simple example. Let $(X_n)_{n\geq 0}$ be the random walk on the integers with initial condition $X_0 = 0$. Considering n as a parameter describing time, an example of a random time is the first return of the walk to the origin, which we already encountered in Section 6:

$$T_0 := \inf\{n \geq 1 : X_n = 0\} . \tag{8.2.5}$$

If the walk never returns to the origin, i.e. $\{n \geq 1 : X_n = 0\} = \emptyset$, we set $T_0 = \infty$. So T_0 is a random variable taking values in $\{1, 2, \dots\} \cup \{\infty\}$. Moreover, the event $\{T_0 = n\}$ is insensitive to the change of any of the variables X_k for $k > n$. This is made clear by noting that $\{T_0 = n\} = \{X_1 \neq 0, \dots, X_{n-1} \neq 0, X_n = 0\}$. In other words, $\{\tau_0 = n\}$ is $\mathcal{F}_n$-measurable, where $\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n)$. We call the sequence $(\mathcal{F}_n)_{n\geq 0}$ the **natural filtration** associated to the chain $(X_n)_{n\geq 0}$. Clearly, $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ for all $n \geq 0$. The natural filtration can be defined for any random process.

Definition 8.2.1. *Consider the natural filtration $(\mathcal{F}_n)_{n\geq 0}$ associated to a Markov chain $(X_n)_{n\geq 0}$. A* **stopping time** *is a $\{1, 2, \dots\} \cup \{\infty\}$-valued random variable T such that for all $n \geq 0$, $\{T = n\} \in \mathcal{F}_n$.*

In the Simple Markov Property, we considered a Markov chain at a fixed time n, conditionned with respect to $\mathcal{F}_n$. We now want to consider the same chain at a random time T, and condition with respect to the σ-algebra which contains events that depend only on what happened *before* T. Since it doesn't make sense to write "$\sigma(X_0, X_1, \dots, X_T)$", we say that $A \in \mathcal{F}_T$ if each time $T \leq n$ then $A \in \mathcal{F}_n$. So define the **stopped σ-algebra generated by T**:

$$\mathcal{F}_T := \{A \in \mathcal{F} : A \cap \{T \leq n\} \in \mathcal{F}_n \, \forall n \geq 0\} . \tag{8.2.6}$$

Fot example, $\{T < \infty\} \in \mathcal{F}_T$. It can be easily verified that $\mathcal{F}_T$ is a σ-algebra (see Exercise 8.12). The position of a Markov chain at time T is naturally defined by the random variable

$$X_T(\omega) := \begin{cases} X_n(\omega) & \text{if } T(\omega) = n \, , \\ \text{“0”} & \text{if } T(\omega) = \infty \, , \end{cases}$$

where "0" is any fixed point of S. We also define $\theta_\infty := \mathrm{id}$. We are now ready to move on to the study of the Markov Property when the conditionning is done with respect to the past of a random stopping time.

THEOREM 8.2.2 (Strong Markov Property). *Let $x \in S$. Let T be a stopping time adapted to the natural filtration $(\mathcal{F}_n)_{n\geq 0}$. Let $\varphi : \Omega \to \mathbb{R}$ be bounded, positive and $\mathcal{F}_T$-measurable. Then for all bounded, positive, measurable $\psi : \Omega \to \mathbb{R}$,*

$$E_x[1_{\{T<\infty\}} \cdot \varphi \cdot \psi \circ \theta_T] = E_x[1_{\{T<\infty\}} \cdot \varphi \cdot E_{X_T}(\psi)] \,. \tag{8.2.7}$$

In particular, if $P_x(T < \infty) = 1$, then

$$E_x[\varphi \cdot \psi \circ \theta_T] = E_x[\varphi \cdot E_{X_T}(\psi)] \,. \tag{8.2.8}$$

The analogue of (8.2.2) for random times reads

$$E_x[1_{\{T<\infty\}} \cdot \psi \circ \theta_T | \mathcal{F}_T] = 1_{\{T<\infty\}} E_{X_T}[\psi] \quad P_x\text{-a.s.} \tag{8.2.9}$$

PROOF. The proof follows by writing $\{T < \infty\} = \bigcup_{n\geq 0}\{T = n\}$. Then, since $1_{\{T=n\}} \cdot \varphi$ is $\mathcal{F}_n$-measurable, by the Simple Markov Property,

$$\begin{aligned} E_x[1_{\{T=n\}} \cdot \varphi \cdot \psi \circ \theta_T] &= E_x[1_{\{T=n\}} \cdot \varphi \cdot \psi \circ \theta_n] \\ &= E_x[1_{\{T=n\}} \cdot \varphi \cdot E_{X_n}(\psi)] \\ &= E_x[1_{\{T=n\}} \cdot \varphi \cdot E_{X_T}(\psi)] \,. \end{aligned}$$

Summing over n gives (8.2.7). □

8.3. Recurrence and Classification

We now consider the *recurrence* problem mentionned before in the case of the random walk: when does a Markov chain come back to its starting point? As before, we will always consider the canonical chain constructed on the product space $\Omega = S^{\{0,1,2,\dots\}}$.

Two random variables are relevant in the study of recurrence. For each $x \in S$, the first visit at x is defined by

$$T_x := \inf\{n \geq 1 : X_n = x\} \,, \tag{8.3.1}$$

where we set $T_x := \infty$ if $\{n \geq 1 : X_n = x\} = \emptyset$. Observe that T_x is a stopping time since $\{T_x > n\} = \{X_1 \neq x, \dots, X_n \neq x\} \in \mathcal{F}_n$. On the other hand, the number of visits at site x is defined by

$$N_x := \sum_{n\geq 1} 1_{\{X_n = x\}} \,. \tag{8.3.2}$$

Clearly, $N_x \geq 1$ if and only if $T_x < \infty$, and so $P_x(N_x \geq 1) = P_x(T_x < \infty)$. A cornerstone in the study of recurrence for Markov chains is a generalization to the situation where $N_x \geq k$.

LEMMA 8.3.1. *Let $x, y \in S$, $k \geq 1$. Then*

$$P_x(N_y \geq k) = P_x(T_y < \infty) P_y(N_y \geq k - 1) \,. \tag{8.3.3}$$

In particular, $P_x(N_x \geq k) = P_x(T_x < \infty)^{k-1}$.

PROOF. Observe that $N_y = N_y \circ \theta_{T_y} + 1$. Therefore, $N_y \geq k+1$ if and only if $T_y < \infty$ and $N_y \circ \theta_{T_y} \geq k$:

$$\begin{aligned} P_x(N_y \geq k+1) &= P_x(T_y < \infty, N_y \circ \theta_{T_y} \geq k) \\ &= E_x[1_{\{T_y<\infty\}} \cdot 1_{\{N_y \geq k\}} \circ \theta_{T_y}] \,. \end{aligned}$$

By the Strong Markov Property (with $\varphi = 1$) and since $X_{T_y} = y$,

$$\begin{aligned} E_x[1_{\{T_y<\infty\}} \cdot 1_{\{N_y \geq k\}} \circ \theta_{T_y}] &= E_x[1_{\{T_y<\infty\}} \cdot E_{X_{T_y}}[1_{\{N_y \geq k\}}]] \\ &= E_x[1_{\{T_y<\infty\}}]E_y[1_{\{N_y \geq k\}}] \\ &\equiv P_x(T_y < \infty)P_y(N_y \geq k) \,. \end{aligned} \tag{8.3.4}$$

The second affirmation follows from the first by induction. □

As an immediate corollary, we obtain the following formula:

$$E_x[N_x] = \sum_{k \geq 1} E_x[N_x \geq k] = \frac{1}{1 - P_x(T_x < \infty)} = \frac{1}{P_x(T_x = \infty)} \,. \tag{8.3.5}$$

This formula makes sense also when $P_x(T_x = \infty) = 0$, in which case $E_x[N_x] = \infty$.

DEFINITION 8.3.1. *A point $x \in S$ is called*

- *recurrent if $P_x(T_x < \infty) = 1$,*
- *transient if $P_x(T_x < \infty) < 1$.*

PROPOSITION 8.3.1. *Let $x \in S$. Then*

(1) *x is recurrent if and only if $P_x(N_x = \infty) = 1$,*
(2) *x is transient if and only if $P_x(N_x = \infty) = 0$.*

PROOF. Assume x is recurrent. Then since $\{N_x \geq k\} \searrow \{N_x = +\infty\}$, and since $P_x(N_x \geq k) = 1$ by the previous lemma, we have $P_x(N_x = \infty) = 1$. Conversely, if $P_x(N_x = \infty) = 1$ then $P_x(N_x \geq k) = 1$ for all $k \geq 1$, which implies $P_x(T_x < \infty) = 1$ by the previous lemma: x is recurrent. If x is transient then $P_x(T_x < \infty) < 1$, and by Lemma 8.3.1, $P_x(N_x = \infty) = \lim_{k\to\infty} P_x(N_x \geq k) = 0$. □

Observe that $\{N_x = \infty\}$ is a tail event, and we have proved that with respect to P_x, its probability is either 0 (when x is transient) or 1 (when x is recurrent). Nevertheless, we have not yet proved a 0-1 Law for Markov chains.

The goal of the rest of this section is to study the partition of S into recurrent and transient states. Comparison of recurrence properties of different points x, y, will be done by studying the expected number of visits at y when started from x:

$$u(x, y) := E_x[N_y] \,. \tag{8.3.6}$$

LEMMA 8.3.2. *For any $x, y \in S$,*

(1) *$u(x, y) = \sum_{n \geq 0} Q^{(n)}(x, y)$.*
(2) *x is recurrent if and only if $u(x, x) = \infty$.*

(3) *If $x \neq y$, then*

$$u(x,y) = P_x(T_y < \infty)u(y,y)\,. \tag{8.3.7}$$

Each of these properties is intuitive: together, (1) and (2) give a computable criterium for verifying whether a point is recurrent, which will be used repeatedly in the sequel (in particular for random walks, in Section 8.3.2). The identity (8.3.7) gives a simple way of comparing recurrence properties of different points.

Proof of Lemma 8.3.2. (1) follows by the definition of N_y and (8.1.8), (2) was shown in (8.3.5). For (3), we use the Strong Markov Property. Since $N_y = 0$ on $\{T_y = \infty\}$,

$$\begin{aligned} E_x[N_y] = E_x[N_y, T_y < \infty] &= E_x[1_{\{T_y<\infty\}} \cdot N_y \circ \theta_{T_y}] \\ &= E_x[1_{\{T_y<\infty\}} \cdot E_y[N_y]] \\ &= P_x(T_y < \infty)E_y[N_y]\,, \end{aligned}$$

which is (8.3.7). □

"Recurrence is contagious", as seen hereafter.

Lemma 8.3.3. *Let x be recurrent, and $y \neq x$. If $u(x,y) > 0$, then $P_y(T_x < \infty) = 1$, $u(y,x) > 0$, y is recurrent and $P_x(T_y < \infty) = 1$. If y is transient, then $u(x,y) = 0$.*

Proof. Since x is recurrent, $P_x(N_x = \infty) = 1$ (Proposition 8.3.1), and so

$$\begin{aligned} 0 = P_x(N_x < \infty) &\geq P_x(T_y < \infty, T_x \circ \theta_{T_y} = \infty) \\ &= E_x[1_{\{T_y<\infty\}} \cdot 1_{\{T_x=\infty\}} \circ \theta_{T_y}] \\ &= E_x[1_{\{T_y<\infty\}} \cdot E_y[1_{\{T_x=\infty\}}]] \\ &= P_x(T_y < \infty)P_y(T_x = \infty)\,. \end{aligned} \tag{8.3.8}$$

Since $u(x,y) > 0$, there exists $n \geq 1$ such that $Q^{(n)}(x,y) > 0$, which implies $P_x(T_y < \infty) \geq Q^{(n)}(x,y) > 0$. (8.3.8) thus gives $P_y(T_x = \infty) = 0$, i.e. $P_y(T_x < \infty) = 1$. By (8.3.7), we obtain $u(y,x) = P_y(T_x < \infty)u(x,x) = \infty > 0$. Since $u(y,x) > 0$, there exists $m \geq 1$ such that $Q^{(m)}(y,x) > 0$. Then, for all $p \geq 0$, by the Chapman-Kolmogorov Equation,

$$Q^{(n+m+p)}(y,y) \geq Q^{(m)}(y,x)Q^{(p)}(x,x)Q^{(n)}(x,y)\,,$$

and so

$$u(y,y) \geq \sum_{p\geq 0} Q^{(n+m+p)}(y,y) \geq Q^{(m)}(y,x)\Big[\sum_{p\geq 0} Q^{(p)}(x,x)\Big]Q^{(n)}(x,y) = \infty\,,$$

which implies that y is recurrent. Proceeding as above from y to x gives $P_x(T_y < \infty) = 1$. The last claim is then obvious. □

As an application, consider the simple random walk on $\mathbb{Z}$ with $0 < p < 1$. Let $x, y \in S$, $x < y$. Then $u(x,y) \geq Q^{(y-x)}(x,y) \geq p^{y-x} > 0$. Similarly, $u(y,x) \geq q^{y-x} > 0$. Therefore, all points are either recurrent, or transient. It is thus sufficient to consider the recurrence properties of the origin. By Theorem

6.1.1, we have that all points are recurrent if $p = \frac{1}{2}$, transient otherwise.

Going back to the general case, let us write S as a disjoint union $\mathcal{R} \cup \mathcal{T}$, where $\mathcal{R}$ are the recurrent points and $\mathcal{T}$ are the transient points. Define the following relation on $\mathcal{R}$: $x \sim y$ if and only if $u(x,y) > 0$. Then obviously $x \sim x$, and Lemma 8.3.3 shows that $\sim$ is reflexive: $x \sim y$ implies $y \sim x$. On the other hand, Then, if $x \sim y$ then there exists $n \geq 1$ with $Q^{(n)}(x,y) > 0$, if $y \sim z$ then there exists $m \geq 1$ with $Q^{(n)}(y,z) > 0$, and so $Q^{(n+m)}(x,z) > 0$, i.e. $x \sim z$. That is, $\sim$ is an equivalence relation, and we can consider the partition of $\mathcal{R}$ into equivalence classes. Since $\mathcal{R}$ is countable, this partition also is, and we denote it by $\mathcal{R} = \bigcup_{j \geq 1} \mathcal{R}_j$. Each $\mathcal{R}_j$ is called a recurrence class.

The Classification Theorem hereafter proves the following intuitive properties: the chain started at $x \in \mathcal{R}_j$ stays in $\mathcal{R}_j$ forever and visits any other $y \in \mathcal{R}_j$ an infinite number of times. The chain started at $x \in \mathcal{T}$ either never visits $\mathcal{R}$ and visits any transient point a finite number of times, or eventually enters a recurrence class $\mathcal{R}_j$ and stays there forever.

THEOREM 8.3.1. *The decomposition $S = \mathcal{T} \cup \bigcup_{j \geq 1} \mathcal{R}_j$ has the following properties:*

(1) *If $x \in \mathcal{R}_j$ then, P_x-almost surely, $N_y = \infty$ for all $y \in \mathcal{R}_j$ and $N_y = 0$ for all $y \in S \setminus \mathcal{R}_j$.*
(2) *If $x \in \mathcal{T}$ and $T_{\mathcal{R}} := \inf\{n \geq 1 : X_n \in \mathcal{R}\}$ then, P_x-almost surely,*
   (a) *either $T_{\mathcal{R}} = \infty$ and then $N_y < \infty$ for all $y \in S$,*
   (b) *or $T_{\mathcal{R}} < \infty$ and there exists a random $j \geq 1$ such that $X_n \in \mathcal{R}_j$ for all $n \geq T_{\mathcal{R}}$.*

PROOF. (1) Let $x \in \mathcal{R}_j$. Then $E_x(N_y) = u(x,y) = 0$ for all $y \in \mathcal{T}$ by Lemma 8.3.3, and for all $y \in \mathcal{R}_i$ $(i \neq j)$ by definition. Therefore, $N_y = 0$ P_x-a.s. for all $y \in S \setminus \mathcal{R}_j$. If $y \in \mathcal{R}_j$, then by taking $k \to \infty$ in Lemma 8.3.1, we get

$$P_x(N_y = \infty) = P_x(T_y < \infty) P_y(N_y = \infty). \tag{8.3.9}$$

But $P_x(T_y < \infty) = 1$ by Lemma 8.3.3, and $P_y(N_y = \infty) = 1$ by Proposition 8.3.1. Therefore, $N_y = \infty$ P_x-a.s.
(2) Let $x \in \mathcal{T}$. We first show (2a), which means

$$P_x(T_{\mathcal{R}} = \infty) = P_x(T_{\mathcal{R}} = \infty, N_y < \infty \, \forall y \in \mathcal{T}). \tag{8.3.10}$$

Since

$$P_x(T_{\mathcal{R}} = \infty, N_y < \infty \, \forall y \in \mathcal{T}) = P_x(T_{\mathcal{R}} = \infty) - P_x\Big(\{T_{\mathcal{R}} = \infty\} \cap \bigcup_{y \in \mathcal{T}} \{N_y = \infty\}\Big),$$

it suffices to notice that for each $y \in \mathcal{T}$,

$$P_x(T_{\mathcal{R}} = \infty, N_y = \infty) \leq P_x(N_y = \infty),$$

which is zero since y is transient (use (8.3.9) and Proposition 8.3.1). This proves (8.3.10). Then we show (2b), which means

$$P_x(T_{\mathcal{R}} < \infty) = P_x(T_{\mathcal{R}} < \infty, \exists j \geq 1 \text{ s.t. } X_n \in \mathcal{R}_j \forall n \geq T_{\mathcal{R}}). \tag{8.3.11}$$

Since the recurrence classes $\mathcal{R}_j$ are disjoint, we can compute

$$\begin{aligned} P_x(T_{\mathcal{R}} < \infty, X_n \in \mathcal{R}_j \forall n \geq T_{\mathcal{R}}) &= E_x[1_{\{T_{\mathcal{R}}<\infty\}} \cdot 1_{\{X_n \in \mathcal{R}_j \forall n \geq 0\}} \circ \theta_{T_{\mathcal{R}}}] \\ &= E_x[1_{\{T_{\mathcal{R}}<\infty\}} \cdot P_{X_{T_{\mathcal{R}}}}(X_n \in \mathcal{R}_j \forall n \geq 0)] \end{aligned}$$

But clearly, $P_{X_{T_{\mathcal{R}}}}(X_n \in \mathcal{R}_j \forall n \geq 0) = 1$ if $X_{T_{\mathcal{R}}} \in \mathcal{R}_j$, 0 if $X_{T_{\mathcal{R}}} \notin \mathcal{R}_j$. Therefore, the right hand side of (8.3.11) equals

$$\begin{aligned} \sum_{j\geq 1} E_x[1_{\{T_{\mathcal{R}}<\infty\}} \cdot P_{X_{T_{\mathcal{R}}}}(X_n \in \mathcal{R}_j \forall n \geq 0)] &= E_x\Big[1_{\{T_{\mathcal{R}}<\infty\}} \sum_{j\geq 1} 1_{\{X_{T_{\mathcal{R}}} \in \mathcal{R}_j\}}\Big] \\ &\equiv E_x[1_{\{T_{\mathcal{R}}<\infty\}}] \\ &= P_x(T_{\mathcal{R}} < \infty)\,. \end{aligned}$$

We have used the fact that $\sum_{j\geq 1} 1_{\{X_{T_{\mathcal{R}}} \in \mathcal{R}_j\}} = 1_{\{X_{T_{\mathcal{R}}} \in \mathcal{R}\}} = 1$ on $\{T_{\mathcal{R}} < \infty\}$. This finishes the proof of the theorem. □

8.3.1. Irreducibility. The Classification Theorem shows that the long time evolution of a Markov chain depends on how the state space S splits into equivalence classes, via the use of the function u. It is natural to consider the case in which the chain has a single class.

DEFINITION 8.3.2. *A chain is called irreducible if $u(x, y) > 0$ for all $x, y \in S$.*

An equivalent definition of irreducibility is: for all $x, y \in S$, there exists an $n \geq 1$ such that $Q^{(n)}(x, y) > 0$. As seen hereafter, in an irreducible chain, all the points are of the same type.

THEOREM 8.3.2. *Let the chain be irreducible. Then*

(1) *either all the points are recurrent, there exists a single recurrence class $S \equiv \mathcal{R}_1$, and $P_x(N_y = \infty\, \forall y \in S) = 1$ for all $x \in S$,*
(2) *or all states are transient, $S = \mathcal{T}$, and $P_x(N_y < \infty\, \forall y \in S) = 1$ for all $x \in S$.*

When S is finite, only the first case can happen.

PROOF. (1) If there is a recurrent point, then by the irreducibility hypothesis and Lemma 8.3.3, all points are recurrent, and clearly there can exist only one recurrence class. The statement, as well as (2), follow from Theorem 8.3.1. For the last statement, assume $|S| < \infty$. If some $x \in S$ were transient, then by (2), we would have, P_x-a.s., $N_y < \infty$ for all $y \in S$. In particular, $\sum_{y\in S} N_y < \infty$. But this is absurd since

$$\sum_{y\in S} N_y = \sum_{y\in S} \sum_{n\geq 0} 1_{\{X_n=y\}} = \sum_{n\geq 0} \sum_{y\in S} 1_{\{X_n=y\}} = \infty\,.$$

(Indeed, for each $n \geq 0$, $\sum_{y\in S} 1_{\{X_n=y\}} = 1$.) □

Before going further and introduce invariant measures, we apply these results to the study of recurrence of random walks on $\mathbb{Z}^d$.

8.3.2. The Simple Symmetric Random Walk on $\mathbb{Z}^d$. The simple random walk on $\mathbb{Z}^d$ was introduced in Example 8.1.3: $S_n = \sum_{k=1}^n X_k$, where $S_0 = 0$ and the variables X_k are $\mathbb{Z}^d$-valued, i.i.d., with distribution p defined in (8.1.2). We denote the probability describing the walk by P (rather than P_0). Clearly, the chain is irreducible. By Theorem 8.3.2, the points are either all recurrent, or all transient. It is thus enough to consider the origin, whose time of first return is denoted T_0. The random walk is recurrent if $P(T_0 < \infty) = 1$, and transient otherwise (i.e. if $P(T_0 < \infty) < 1$). The main result for the simple random walk is the following.

THEOREM 8.3.3. *The simple symmetric random walk is recurrent for $d = 1, 2$, and transient for $d \geq 3$.*

Since $Q^{(n)}(0,0) = P(S_n = 0)$, which is zero when n is odd, Lemma 8.3.2 gives the following criterium for recurrence.

$$\text{The walk is recurrent } \Leftrightarrow \sum_{n\geq 1} P(S_{2n} = 0) = \infty\,. \tag{8.3.12}$$

Recurrence for $d = 1, 2$ will be obtained with the following property of symmetric random walks.

LEMMA 8.3.4. *If the walk is symmetric, then*

$$P(S_{2n} = 0) = \sup_{z\in\mathbb{Z}^d} P(S_{2n} = z)\,. \tag{8.3.13}$$

PROOF. We sum over the position at the nth step and use independence:

$$\begin{aligned} P(S_{2n} = z) &= \sum_{y\in\mathbb{Z}^d} P(S_n = y, S_{2n} = z) \\ &= \sum_{y\in\mathbb{Z}^d} P(S_n = y, S_{2n} - S_n = z - y) \\ &= \sum_{y\in\mathbb{Z}^d} P(S_n = y) P(S_n = z - y)\,. \end{aligned} \tag{8.3.14}$$

By the Cauchy-Schwartz Inequality and a change of variable,

$$\begin{aligned} \sum_{y\in\mathbb{Z}^d} P(S_n = y)P(S_n = z-y) &\leq \Big[\sum_{y\in\mathbb{Z}^d} P(S_n = y)^2\Big]^{\frac12} \Big[\sum_{y\in\mathbb{Z}^d} P(S_n = z-y)^2\Big]^{\frac12} \\ &= \sum_{y\in\mathbb{Z}^d} P(S_n = y)^2\,. \end{aligned}$$

Now if the walk is symmetric then $P(S_n = y) = P(S_n = -y)$, and so using again (8.3.14) with $z = 0$, we get

$$\sum_{y\in\mathbb{Z}^d} P(S_n = y)^2 = \sum_{y\in\mathbb{Z}^d} P(S_n = y)P(S_n = -y) = P(S_{2n} = 0)\,,$$

which proves the claim. □

Below, $\|\cdot\|$ denotes Euclidian distance in $\mathbb{Z}^d$.

PROOF OF THEOREM 8.3.3: First consider $d = 1$: by Lemma 8.3.4,

$$1 = \sum_{y \in \mathbb{Z}: \|y\| \leq 2n} P(S_{2n} = y) \leq (4n+1) P(S_{2n} = 0)\,,$$

which gives $P(S_{2n} = 0) \geq (4n+1)^{-1}$. By (8.3.12), the walk is recurrent. For $d = 2$, we proceed in the same way. A straightforward computation using independence of the $X_k s$ yields $E[\|S_{2n}\|^2] = 2n$. By the Chebychev Inequality,

$$P(\|S_{2n}\| > 2\sqrt{n}) \leq \frac{E[\|S_{2n}\|^2]}{4n} = \frac{1}{2}\,.$$

One can thus proceed as before and obtain

$$\frac{1}{2} \leq P(\|S_{2n}\| \leq 2\sqrt{n}) = \sum_{y \in \mathbb{Z}^2: \|y\| \leq 2\sqrt{n}} P(S_{2n} = y) \leq (8\sqrt{n}+1)^2 P(S_{2n} = 0)\,.$$

By (8.3.12), the walk is recurrent. For $d = 3$, we need an upper bound. Let $n_i \geq 0$, $i \in \{1, 2, 3\}$, be the number of positive steps done along the direction e_i. To be back at the origin after $2n$ steps, we must choose a triple (n_1, n_2, n_3) satisfying $n_1 + n_2 + n_3 = n$, and then choose a path which contains, for each $i = 1, 2, 3$, n_i steps along $+e_i$, and n_i steps along $-e_i$. There are

$$\binom{2n}{n_1\ n_1\ n_2\ n_2\ n_3\ n_3} = \frac{(2n)!}{(n_1! n_2! n_3!)^2}$$

ways of doing so. Since each path has probability $(\frac{1}{6})^{2n}$,

$$\begin{aligned}
P(S_{2n} = 0) &= \sum_{\substack{(n_1,n_2,n_3):\\ n_1+n_2+n_3=n}} \frac{(2n)!}{(n_1! n_2! n_3!)^2} \frac{1}{6^{2n}} \\
&= \frac{1}{2^{2n}} \binom{2n}{n} \sum_{\substack{(n_1,n_2):\\ 0 \leq n_1+n_2 \leq n}} \left[\frac{n!}{n_1! n_2! (n-n_1-n_2)!} \frac{1}{3^n} \right]^2 \\
&\leq \frac{1}{2^{2n}} \binom{2n}{n} \max_{\substack{(n_1,n_2):\\ 0 \leq n_1+n_2 \leq n}} \frac{n!}{n_1! n_2! (n-n_1-n_2)!} \frac{1}{3^n}\,. \qquad (8.3.15)
\end{aligned}$$

We have used the fact that the numbers in the brackets add up to one.

LEMMA 8.3.5. *There exists $C > 0$ such that*

$$\max_{\substack{(n_1,n_2):\\ 0 \leq n_1+n_2 \leq n}} \frac{n!}{n_1! n_2! (n-n_1-n_2)!} \frac{1}{3^n} \leq \frac{C}{n}\,. \qquad (8.3.16)$$

PROOF. As can be easily verified, the denominator in (8.3.16) decreases when the difference between the three numbers $n_1, n_2, n - n_1 - n_2$ is reduced. One can therefore bound the maximum over all triples in which each term lies within distance at most one from $\frac{n}{3}$. This implies that for large n, the Stirling Formula can be used for each of the terms apearing in the ratio, which proves the lemma. □

Using the lemma and again the Stirling Formula for the first term in (8.3.15),

$$P(S_{2n}=0) \leq \frac{D}{n^{\frac{3}{2}}} .$$

With (8.3.12), we conclude that the simple random walk on $\mathbb{Z}^3$ is transient. The proof that the walk is transient in higher dimensions is left as an exercise. □

Observe that all the estimates we have obtained above for $P(S_{2n}=0)$ follow from a more general Local Limit Theorem, valid in all dimension (see Exercise 8.19):

$$P(S_{2n}=0) \sim \frac{1}{\sqrt{(2\pi n)^d}} .$$

8.4. Equilibrium: Stationary Distributions

Theorems (8.3.1) and 8.3.2 give a first general picture of what the asymptotic behaviour of a Markov chain looks like: starting from an arbitrary point x, it either falls into one of the recurrence classes $\mathcal{R}_j$, or remains transient forever. Our next objective is to take a closer look at what can happen in each of these cases. More precisely, we will look at things such at the average time spent by the chain at each point $x \in S$, leading to the natural notion of *invariant* measure. Before this we to introduce some notations for probability distributions on $(S, \mathcal{P}(S))$.

8.4.1. Invariant Measures. Let μ be a measure on $(S, \mathcal{P}(S))$, i.e. a collection of non-negative numbers $(\mu(x))_{x\in S}$. To avoid misleading it with E_μ, which acts on random variables living in another space, we denote the expectation, with respect to μ, of a positive bounded measurable function $f : S \to \mathbb{R}$ by either of the symbols

$$\int f d\mu = \mu(f) := \sum_{x\in S} \mu(x) f(x) .$$

It is sometimes useful to think of functions $f : S \to \mathbb{R}$ as *column vectors* and of measures μ on S as *row vectors*. The expectation $\mu(f)$ can then be naturally written as an inner product:

$$\langle f, \mu \rangle := \sum_{x\in S} \mu(x) f(x) .$$

If Q is a transition probability matrix, we define a new measure μQ by

$$\mu Q(x) := \sum_{y\in S} \mu(y) Q(y,x) . \tag{8.4.1}$$

Remembering (8.1.13):

$$Q^{(n)} f(x) := \sum_{y\in S} Q^{(n)}(x,y) f(y) , \tag{8.4.2}$$

we have the following identity:

$$\langle f, \mu Q \rangle = \langle Qf, \mu \rangle .$$

It does then make sense to say that Q act *from the left* on functions and *from the right* on measures. If μ is a probability (i.e. $\sum_x \mu(x) = 1$), then μQ is again a probability. Going back to Markov chains: if μ is the probability distribution of X_0 for the Markov chain $(X_n)_{n\geq 0}$ whose transition matrix is Q, i.e. $P_\mu(X_0 = x) = \mu(x)$, then μQ is the distribution of X_1. Indeed, by Lemma 8.1.1,

$$P_\mu(X_1 = x) = \sum_{y\in S} P_\mu(X_1 = x, X_0 = y) = \sum_{y\in S} \mu(y)Q(y,x) \equiv \mu Q(x)\,.$$

Similarly, the distribution of X_n is given by $\mu Q^{(n)}$:

$$P_\mu(X_n = x) = \mu Q^{(n)}(x)\,.$$

We see that understanding the large-n-behaviour of the chain goes through the study of the limits

$$\pi(x) := \lim_{n\to\infty} \mu Q^{(n)}(x)\,. \tag{8.4.3}$$

Giving a meaning to (8.4.3), conditions under which this limit exists, and its possible independence of μ, will be done in details later.

There is also a formula for the expectation of $f(X_n)$ with respect to E_μ:

$$E_\mu(f(X_n)) = \sum_{x\in S} P_\mu(X_n = x) f(x) = \mu Q^{(n)}(f)\,. \tag{8.4.4}$$

(8.4.4) says that the expectation of an observable made on the evolution can be obtained by an expectation of this observable over S with respect to the measure $\mu Q^{(n)}$. Observe that $\mu Q^{(n)}(f) = \langle f, \mu Q^{(n)}\rangle = \langle Q^{(n)} f, \mu\rangle = \mu(Q^{(n)} f)$. To motivate the following definition, assume for a while that the limit defining π in (8.4.3) exists for all $x \in S$. Then for all bounded f,

$$\langle f, \pi Q\rangle = \langle Qf, \pi\rangle = \lim_{n\to\infty} \langle Qf, \mu Q^{(n)}\rangle = \lim_{n\to\infty} \langle f, \mu Q^{(n+1)}\rangle = \langle f, \pi\rangle\,,$$

which implies that $\pi Q = \pi$. This motivates the following definition.

DEFINITION 8.4.1. *Let Q be a transition matrix, μ a measure on $(S, \mathcal{P}(S))$. If*

$$\mu Q = \mu\,, \tag{8.4.5}$$

then μ is called invariant with respect to *Q.*

(8.4.5) is sometimes called the balance relation. Consider the random walk of Example 8.1.3, with $Q(x,y) = p(y-x)$. Then the counting measure ($\mu(x) = 1$ for all x) is invariant:

$$\mu Q(x) = \sum_{y\in S} Q(y,x) = \sum_{y\in S} p(y-x) = 1 = \mu(x)\,.$$

By induction we see that if μ is invariant, then $\mu Q^{(n)} = \mu$ for all $n \geq 1$. Moreover, when the initial distribution μ of a Markov chain $(X_n)_{n\geq 0}$ with transition matrix Q is invariant under Q, then X_n has the same distribution as X_0. Namely, by (8.4.4),

$$E_\mu(f(X_n)) = \mu Q^{(n)}(f) = \mu(f) \equiv E_\mu(f(X_0))\,.$$

In such a case, i.e. when the distribution of the chain is insensitive to the evolution under the transition matrix Q, we say that X_n is at equilibirum for all $n \geq 1$. Invariant measures will play an important role in the study of the asymptotics of the chain.

We will first be interested in the existence of invariant measures, then of invariant probability measures, and then we shall move on to the study of the existence of the limits (8.4.3).

8.4.2. Existence of Invariant Measures. Finding an invariant measure means, for the time being, solving a system of equations for $(\mu(x))_{x\in S}$:

$$\mu(x) = \sum_{y\in S} \mu(y)Q(y,x) \quad \forall x \in S\,.$$

Definition 8.4.2. *A measure μ is reversible (with respect to Q) if*

$$\mu(x)Q(x,y) = \mu(y)Q(y,x) \quad \forall x,y \in S\,. \tag{8.4.6}$$

The set relations (8.4.6) are sometimes called the relation of detailed balance, since it is stronger than (8.4.5). Observe that if μ is reversible, then for all $x \in S$,

$$\mu Q(x) = \sum_{y\in S} \mu(y)Q(y,x) = \sum_{y\in S} \mu(x)Q(x,y) = \mu(x)\,.$$

We have thus shown

Lemma 8.4.1. *If μ is reversible, then it is invariant.*

This result gives an easy way of finding invariant measures. For example, consider the uniform random walk on the graph, introduced in Example 8.1.4. Then the measure $\mu(x) := |A_x|$ is invariant. Namely, if $\{x,y\} \in E$,

$$\mu(x)Q(x,y) = |A_x|\frac{1}{|A_x|} = 1 = |A_y|\frac{1}{|A_y|} = \mu(y)Q(y,x)\,.$$

Another example is the simple random walk on $\mathbb{Z}$ with $Q(x,x+1) = p < 1$. It easy to verify, using the above criterium, that the measure

$$\mu(x) = \Big(\frac{p}{1-p}\Big)^x\,, \quad \forall x \in \mathbb{Z}$$

is invariant. Observe that $\mu(x)$ is bounded if and only if $p = \frac{1}{2}$. When $p > \frac{1}{2}$ (resp. $p < \frac{1}{2}$), then μ gives unbounded weight to points far to the right (resp. left), which reflects the transience of the walk. As an exercise, the reader can also compute an invariant measure for the Ehrenfest Model of Example 8.1.5 (Exercise 8.24).

The following result shows that the existence of at least one recurrent point x guarantees the existence of an invariant measure.

THEOREM 8.4.1. *Let $x \in S$ be recurrent. For all $y \in S$, define*

$$\nu_x(y) := E_x\Big[\sum_{k=0}^{T_x-1} 1_{\{X_k=y\}}\Big] \equiv E_x[N_y, T_y < T_x]\,. \tag{8.4.7}$$

Then ν_x is an invariant measure [2]. *Moreover, $\nu_x(y) > 0$ if and only if y belongs to the recurrence class of x. Finally, $\nu_x(y) < \infty$ for all $y \in S$.*

PROOF. Observe that $\nu_x(x) = E_x(1) = 1$. We compute, for all $z \in S$,

$$\begin{aligned}\sum_{y\in S}\nu_x(y)Q(y,z) &= \sum_{y\in S}\sum_{k\geq 0} E_x[1_{\{k<T_x\}}1_{\{X_k=y\}}]Q(y,z)\\ &= \sum_{y\in S}\sum_{k\geq 0} E_x[1_{\{k<T_x\}}1_{\{X_k=y\}}1_{\{X_{k+1}=z\}}] \qquad (8.4.8)\\ &= \sum_{k\geq 0} E_x[1_{\{k<T_x\}}1_{\{X_{k+1}=z\}}]\,.\end{aligned}$$

This identity (8.4.8) is justified by observing that, since $1_{\{k<T_x\}}1_{\{X_k=y\}}$ is $\mathcal{F}_k$-measurable, the Markov Property at time k gives

$$\begin{aligned}E_x[1_{\{k<T_x\}}1_{\{X_k=y\}}1_{\{X_{k+1}=z\}}] &= E_x[1_{\{k<T_x\}}1_{\{X_k=y\}}1_{\{X_1=z\}}\circ\theta_k]\\ &= E_x[1_{\{k<T_x\}}1_{\{X_k=y\}}E_{X_k}[1_{\{X_1=z\}}]]\\ &= E_x[1_{\{k<T_x\}}1_{\{X_k=y\}}E_y[1_{\{X_1=z\}}]]\\ &= E_x[1_{\{k<T_x\}}1_{\{X_k=y\}}]Q(y,z)\,.\end{aligned}$$

Now, if $z \neq x$, then clearly $1_{\{k<T_x\}}1_{\{X_{k+1}=z\}} = 1_{\{k+1<T_x\}}1_{\{X_{k+1}=z\}}$, and so

$$\sum_{y\in S}\nu_x(y)Q(y,z) = \sum_{k\geq 0}E_x[1_{\{k+1<T_x\}}1_{\{X_{k+1}=z\}}] = E_x\Big[\sum_{k=0}^{T_x-2}1_{\{X_{k+1}=z\}}\Big] = \nu_x(z)\,.$$

On the other hand, when $z = x$, then $E_x[1_{\{k<T_x\}}1_{\{X_{k+1}=x\}}] = P_x(T_x = k+1)$, and so, since x is recurrent,

$$\sum_{y\in S}\nu_x(y)Q(y,x) = \sum_{k\geq 0}P_x(T_x = k+1) = P_x(T_x<\infty) = 1 = \nu_x(x)\,.$$

This proves that ν_x is invariant. Then, if y belongs to the recurrence class of x, there exists some $m \geq 1$ such that $Q^{(m)}(x,y) > 0$, and so

$$\nu_x(y) = \sum_{z\in S}\nu_x(z)Q^{(m)}(z,y) \geq \nu_x(x)Q^{(m)}(x,y) > 0\,.$$

On the other hand, if y is not in the recurrence class of x, then $N_y = 0$ i.e. $1_{\{X_k=y\}} = 0$ for all $k \geq 0$ P_x-a.s. by Theorem 8.3.1, and so $\nu_x(y) = 0$. To show that $\nu_x(y)$ is finite, observe that invariance of ν_x implies that $\nu_x = \nu_x Q^{(n)}$ for all $n \geq 1$. In particular,

$$1 = \nu_x(x) = \nu_x Q^{(n)}(x) \geq \nu_x(y)Q^{(n)}(y,x) \quad \forall y \in S\,,$$

[2]To see that ν_x is not completely trivial, i.e. that $\nu_x(y) < \infty$ for all $y \in S$, see [**R.88**] p. 301. of Neveu p. 50.

which implies $\nu_x(y) < \infty$ if $n \geq 1$ is such that $Q^{(n)}(y, x) > 0$. But this is true for at least one n when y belongs to the recurrence class of x. If y is not in the recurrence class of x, we have $\nu_x(y) = 0 < \infty$, as seen above. □

Observe that if there is more than one recurrence class, then the theorem above allows to construct invariant measures with disjoint supports.

THEOREM 8.4.2. *Let the chain be irreducible and all points be recurrent. Then the invariant measure (which exists by Theorem 8.4.1) is unique, up to a multiplicative constant.*

PROOF. Let $x \in S$ and consider the invariant measure ν_x of Theorem 8.4.1. We will show that for any other invariant measure μ,

$$\mu(y) \geq \mu(x)\nu_x(y) \quad \forall y \in S \,. \tag{8.4.9}$$

Assume for a while that this is true. We have, for all $n \geq 1$,

$$\mu(x) = \sum_{z \in S} \mu(z) Q^{(n)}(z, x) \geq \sum_{z \in S} \mu(x)\nu_x(z) Q^{(n)}(z, x) = \mu(x)\,,$$

which gives

$$\sum_{z \in S} [\mu(z) - \mu(x)\nu_x(z)] Q^{(n)}(z, x) = 0\,.$$

Therefore, $\mu(z) = \mu(x)\nu_x(z)$ each time $Q^{(n)}(z, x) > 0$ for some $n \geq 1$. But this is guaranteed by the irreducibility of the chain. Therefore, $\mu = c\nu_x$, with $c = \mu(x)$, proving the theorem. To obtain (8.4.9), we will show, by induction on $p \geq 0$, that ($a \wedge b := \min\{a, b\}$)

$$\mu(y) \geq \mu(x) E_x\Big[\sum_{k=0}^{p \wedge (T_x - 1)} 1_{\{X_k = y\}}\Big]\,. \tag{8.4.10}$$

From this, (8.4.9) follows by taking $p \to \infty$. The inequality (8.4.10) is an equality when $y = x$, so we may always consider $y \neq x$. For $p = 0$, the inequality is trivial. Assuming (8.4.10) holds for p,

$$\begin{aligned}
\mu(y) = \sum_{z \in S} \mu(z) Q(z, y) &\geq \mu(x) \sum_{z \in S} E_x\Big[\sum_{k=0}^{p \wedge (T_x - 1)} 1_{\{X_k = z\}}\Big] Q(z, y) \\
&= \mu(x) \sum_{z \in S} \sum_{k=0}^{p} E_x[1_{\{k < T_x\}} 1_{\{X_k = z\}}] Q(z, y)\,. \\
&= \mu(x) \sum_{z \in S} \sum_{k=0}^{p} E_x[1_{\{k < T_x\}} 1_{\{X_k = z\}} 1_{\{X_{k+1} = y\}}] \qquad (8.4.11) \\
&= \mu(x) \sum_{k=0}^{p} E_x[1_{\{k < T_x\}} 1_{\{X_{k+1} = y\}}]
\end{aligned}$$

In (8.4.11) we used the Markov Property, as in the proof of Theorem 8.4.1. Now, since $y \neq x$,

$$\begin{aligned}\sum_{k=0}^{p} E_x[1_{\{k<T_x\}}1_{\{X_{k+1}=y\}}] &= \sum_{k=0}^{p} E_x[1_{\{k+1<T_x\}}1_{\{X_{k+1}=y\}}] \\ &= \sum_{l=1}^{p+1} E_x[1_{\{l<T_x\}}1_{\{X_l=y\}}] = E_x\Big[\sum_{l=0}^{(p+1)\wedge(T_x-1)} 1_{\{X_l=y\}}\Big] .\end{aligned}$$

This proves (8.4.10) for $p+1$. □

Now that the existence and uniqueness of invariant measures is settled, we turn to the problem of determining whether there exist *finite* measures, i.e. for which $\mu(S) < \infty$, or, which is equivalent, to finding probability distributions on S invariant under Q. This will require a further distinction among recurrent points. Before this, we give a simple result showing that invariant probability measures concentrate on recurrent points. From now on, invariant probability measures will be denoted by π.

LEMMA 8.4.2. *Assume there exists an invariant probability π, then each point $x \in S$ with $\pi(x) > 0$ is recurrent.*

PROOF. Since π is invariant we have $\pi Q^{(n)} = \pi$ for all $n \geq 1$. Assume $\pi(x) > 0$. Then, using Fubini's Theorem and recalling the definition (8.3.6),

$$\infty = \sum_{n\geq 1} \pi(x) = \sum_{n\geq 1}\sum_{y\in S} \pi(y)Q^{(n)}(y,x) \leq \sum_{y\in S} \pi(y)u(y,x) \leq u(x,x) .$$

We used (3) of Lemma 8.3.2 and the fact that π is a probability. By (2) of the same lemma, we conclude that x is recurrent. □

PROPOSITION 8.4.1. *If the chain is irreducible and if there exists an invariant probability π, then it has the form*

$$\pi(x) = \frac{1}{E_x(T_x)} \quad \forall x \in S . \tag{8.4.12}$$

PROOF. If there exists an invariant probability, then all points are recurrent. Indeed, if there existed a transient point then all points would be transient (since the chain is irreducible), and so $\pi(x) = 0$ for all x by Lemma 8.4.2, a contradiction. We choose any $x \in S$ and show that $\pi(x)$ has the form (8.4.12). By Theorem 8.4.1 there exists an invariant measure ν_x, given in (8.4.7). By Theorem 8.4.2 the invariant measure is unique up to a multiplicative constant. Therefore, if there exists an invariant probability π, then the total mass of ν_x must be finite, $\nu_x(S) < \infty$, and π have the form $\pi = \frac{\nu_x}{\nu_x(S)}$. But

$$\nu_x(S) = \sum_{y\in S} \nu_x(y) = E_x\Big[\sum_{k=0}^{T_x-1}\sum_{y\in S} 1_{\{X_k=y\}}\Big] \equiv E_x(T_x) .$$

In particular, $\pi(x) = \frac{\nu_x(x)}{E_x(T_x)} = \frac{1}{E_x(T_x)}$. This shows the theorem. □

The previous result shows that for an invariant measure to exist, one must have $E_x(T_x) < \infty$ for all recurrent point x. This leads to the following distinction among recurrent points.

DEFINITION 8.4.3. *A recurrent point $x \in S$ is called*

- *positive-recurrent if $E_x(T_x) < \infty$,*
- *null-recurrent if $E_x(T_x) = \infty$.*

For example, the simple symmetric random walk on $\mathbb{Z}$ is recurrent, but null-recurrent, as we saw in Theorem 6.1.1. Positive recurrence is a class property: points belonging to the same recurrence class are either all positive-recurrent, or all null-recurrent.

LEMMA 8.4.3. *Let the chain be irreducible. Then the following are equivalent.*

(1) *There exists one positive-recurrent point $x \in S$.*
(2) *There exists an invariant probability π.*
(3) *All points $x \in S$ are positive-recurrent.*

PROOF. (1) implies (2): Assume $x \in S$ is positive-recurrent. Consider the invariant measure ν_x. Then $\nu_x(S) = E_x[T_x] < \infty$, and so $\pi := \nu_x(S)^{-1}\nu_x$ is an invariant probability. (2) implies (3): As we saw in Lemma 8.4.2, the existence of an invariant probability implies that $\pi(x) > 0$ for all x. But $\pi(x) = E_x[T_x]^{-1}$ by Proposition 8.4.1, and so $E_x[T_x] < \infty$. (3) implies (1) trivially. □

We gather the results about invariant for irreducible chains in a theorem.

THEOREM 8.4.3. *Let the chain be irreducible and all points be recurrent. Then*

(1) *either each point is positive-recurrent, and there exists a unique invariant probability measure π, $\pi(S) = 1$, given by*

$$\pi(x) = \frac{1}{E_x(T_x)} \quad \forall x \in S\,, \tag{8.4.13}$$

(2) *or each point is null-recurrent, and any invariant measure μ has infinite mass ($\mu(S) = \infty$).*

8.5. Approach to Equilibrium

We now turn to the study of how equilibrium is approached along the time evolution of a Markov chain. Our main purpose is to show that the distribution of the chain converges, in the limit $n \to \infty$, to the invariant measure constructed in Theorem 8.4.3. We will therefore study the limits which appeared in (8.4.3). A detailed study of the convergence to equilibrium can be found in [**Str05**].

The convergence of distribution will be in the sense of the total variation norm, defined, for each $\rho : S \to \mathbb{R}$, by

$$\|\rho\|_{\mathrm{TV}} := \sum_{x \in S} |\rho(x)|\,. \tag{8.5.1}$$

We will say that a sequence of measures $(\mu_n)_{n\geq 1}$ on $(S, \mathcal{P}(S))$ converges to μ if $\|\mu_n - \mu\|_{\mathrm{TV}} \to 0$. As a short hand, we write $\mu_n \Rightarrow \mu$. Our aim is to find under which conditions can one obtain, for a recurrent chain with unique invariant probability π,

$$\mu Q^{(n)} \Rightarrow \pi .$$

We will present a standard proof based on a *coupling* argument.

Consider a Markov chain with state space S and transition matrix Q. A *coupling* consists in building two copies of this chain on the cartesian product $\mathbb{S} := S \times S$. We endow $\mathbb{S}$ with the σ-field $\mathcal{P}(\mathbb{S})$. If μ, ν are probability distributions on S, $\mu \otimes \nu$ denotes the probability distribution on $\mathbb{S}$ defined by $(\mu \otimes \nu)(x, y) := \mu(x)\nu(y)$. We then define the following transition matrix on $\mathbb{S}$:

$$\mathbb{Q}((x, y), (x'y')) := Q(x, x')Q(y, y') . \tag{8.5.2}$$

By Theorem 8.1.2, we can construct a canonical version of a Markov chain $(X_n, Y_n)_{n\geq 0}$ with state space $\mathbb{S}$, initial distribution $\mu \otimes \nu$ and transition matrix $\mathbb{Q}$. We denote the associated measure by $\mathbb{P}_{\mu\otimes\nu}$. It is clear that under $\mathbb{P}_{\mu\otimes\nu}$, the coupled chain $(X_n, Y_n)_{n\geq 0}$ describes two independent copies of the original markov chain. Its marginals are given by

$$\begin{aligned}
&\mathbb{P}_{\mu\otimes\nu}(X_{n+1} = x' | X_n = x) = Q(x, x') , \quad \mathbb{P}_{\mu\otimes\nu}(X_n = x) = \mu Q^{(n)}(x) , \\
&\mathbb{P}_{\mu\otimes\nu}(Y_{n+1} = y' | Y_n = y) = Q(y, y') , \quad \mathbb{P}_{\mu\otimes\nu}(Y_n = y) = \nu Q^{(n)}(y) .
\end{aligned}$$

A key idea is then to choose $\nu := \pi$, where π is the invariant measure of Q. This implies that under $\mathbb{P}_{\mu\otimes\nu}$, $(Y_n)_{n\geq 0}$ is at equilibrium for all $n \geq 0$:

$$\mathbb{P}_{\mu\otimes\pi}(Y_n = y) = \pi Q^{(n)}(y) = \pi(y) = \mathbb{P}_{\mu\otimes\pi}(Y_0 = y) .$$

Therefore,

$$\begin{aligned}
P_\mu(X_n = y) - \pi(y) &= \mathbb{P}_{\mu\otimes\pi}(X_n = y) - \mathbb{P}_{\mu\otimes\pi}(Y_n = y) \\
&= \mathbb{E}_{\mu\otimes\pi}\big[1_{\{X_n=y\}} - 1_{\{Y_n=y\}}\big] .
\end{aligned}$$

Let now $\mathbb{T}$ define the stopping time at which X_n and Y_n meet for the first time:

$$\mathbb{T} := \inf\{n \geq 1 : X_n = Y_n\} .$$

In other words, $\mathbb{T}$ is the first time the chain $(X_n, Y_n)_{n\geq 0}$ hits the diagonal $\{(x, x) : x \in S\}$. The point is that if the two chains meet at some time N, then the Markov Property implies that they become probabilistically undistinguishable for times $> N$. We therefore decompose the last expectation with respect to the stopping time $\mathbb{T}$ and to the position of the chain at time $\mathbb{T}$:

$$\begin{aligned}
\mathbb{E}_{\mu\otimes\pi}\big[1_{\{X_n=y\}} - 1_{\{Y_n=y\}}\big] &= \mathbb{E}_{\mu\otimes\pi}\big[1_{\{\mathbb{T}>n\}}(1_{\{X_n=y\}} - 1_{\{Y_n=y\}})\big] \\
&\quad + \sum_{k=1}^{n} \sum_{z\in S} \mathbb{E}_{\mu\otimes\pi}\big[1_{\{\mathbb{T}=k, X_k=Y_k=z\}}(1_{\{X_n=y\}} - 1_{\{Y_n=y\}})\big] .
\end{aligned}$$

This last sum is zero. Indeed, using twice the Markov Property at time k,

$$\begin{aligned}\mathbb{E}_{\mu\otimes\pi}[1_{\{\mathbb{T}=k,X_k=Y_k=z\}}1_{\{X_n=y\}}] &= \mathbb{E}_{\mu\otimes\pi}[1_{\{\mathbb{T}=k,X_k=Y_k=z\}}1_{\{X_{n-k}=y\}}\circ\theta^k]\\ &= \mathbb{E}_{\mu\otimes\pi}[1_{\{\mathbb{T}=k,X_k=Y_k=z\}}]Q^{(n-k)}(z,y)\\ &= \mathbb{E}_{\mu\otimes\pi}[1_{\{\mathbb{T}=k,X_k=Y_k=z\}}1_{\{Y_{n-k}=y\}}\circ\theta^k]\\ &= \mathbb{E}_{\mu\otimes\pi}[1_{\{\mathbb{T}=k,X_k=Y_k=z\}}1_{\{Y_n=y\}}]\,.\end{aligned}$$

Therefore,

$$\begin{aligned}\sum_{y\in S}|P_\mu(X_n=y)-\pi(y)| &= \sum_{y\in S}\left|\mathbb{E}_{\mu\otimes\pi}\left[1_{\{\mathbb{T}>n\}}(1_{\{X_n=y\}}-1_{\{Y_n=y\}})\right]\right|\\ &\leq 2\sum_{y\in S}\mathbb{E}_{\mu\otimes\pi}[1_{\{\mathbb{T}>n\}}1_{\{Y_n=y\}}] = 2\mathbb{P}_{\mu\otimes\pi}(\mathbb{T}>n)\,,\end{aligned}$$

We are left with

$$\|\mu Q^{(n)}-\pi\|_{\mathrm{TV}} \leq 2\mathbb{P}_{\mu\otimes\pi}(\mathbb{T}>n)\,, \tag{8.5.3}$$

which is the standard coupling inequality. We will thus obtain $\mu Q^{(n)}\Rightarrow\pi$ if we can show that the chain $(X_n,Y_n)_{n\geq 0}$ is recurrent. The most general way of obtaining this recurrence is under a condition on the chain S called *aperiodicity*, to which we shall turn in a while. Before this we consider a more restrictive condition, but which gives a rate of convergence for the speed at which $\|\mu Q^{(n)}-\pi\|_{\mathrm{TV}}\to 0$.

LEMMA 8.5.1. *Assume the chain S satisfies the following condition: there exists $\ell\geq 1$ such that*

$$\inf_{x,y\in S} Q^{(\ell)}(x,y)\geq\delta>0\,. \tag{8.5.4}$$

Then, for all probability distributions μ, ν, we have

$$\mathbb{P}_{\mu\otimes\nu}(\mathbb{T}>k\ell)\leq(1-\delta)^k\,,\quad\forall k\geq 1\,. \tag{8.5.5}$$

PROOF. We will prove the lemma for Dirac masses $\mu=\delta_x$, $\nu=\delta_y$, in which case the measure is denoted $\mathbb{P}_{(x,y)}$. That is, we will show that for all $k\geq 1$,

$$\mathbb{P}_{(x,y)}(\mathbb{T}>k\ell)\leq(1-\delta)^k\,,\quad\forall(x,y)\in\mathbb{S}\,. \tag{8.5.6}$$

The general case (8.5.5) then follows by summation over $(x,y)\in\mathbb{S}$ [3]. We show (8.5.6) by induction on k. Consider first the case $k=1$. For any pair (x,y), we have, by (8.5.4),

$$\begin{aligned}\mathbb{P}_{(x,y)}(\mathbb{T}\leq\ell)\geq\mathbb{P}_{(x,y)}(X_\ell=Y_\ell) &= \sum_{z\in S}\mathbb{P}_{(x,y)}(X_\ell=Y_\ell=z)\\ &= \sum_{z\in S}Q^{(\ell)}(x,z)Q^{(\ell)}(y,z)\geq\delta\sum_{z\in S}Q^{(\ell)}(y,z)=\delta\,,\end{aligned}$$

which shows (8.5.6) for $k=1$. Assume then that (8.5.6) holds for k and for all pair (x,y). Then,

$$\mathbb{P}_{(x,y)}(\mathbb{T}>(k+1)\ell)=\sum_{(s,t)\in\mathbb{S}}\mathbb{P}_{(x,y)}(\mathbb{T}>(k+1)\ell,X_{k\ell}=s,Y_{k\ell}=t)\,.$$

[3] A mettre en exercice!

Using the Markov Property at time $k\ell$,

$$\mathbb{P}_{(x,y)}(\mathbb{T} > (k{+}1)\ell, X_{k\ell} = s, Y_{k\ell} = t) = \mathbb{P}_{(x,y)}(\mathbb{T} > k\ell, X_{k\ell} = s, Y_{k\ell} = t)\mathbb{P}_{(s,t)}(\mathbb{T} > \ell)\,.$$

Using $\mathbb{P}_{(s,t)}(\mathbb{T} > \ell) \leq 1 - \delta$, resumming over $(s,t) \in \mathbb{S}$ and using the induction hypothesis yields (8.5.6) for $k+1$. □

A direct corollary is then

Theorem 8.5.1. *Assume the chain S is irreducible and positive recurrent and satisfies* (8.5.4) *for some $\delta > 0$, $\ell \geq 1$. Let π denote the unique invariant probability measure. Then*

$$\|\mu Q^{(n)} - \pi\|_{\mathrm{TV}} \leq 2(1-\delta)^{\lfloor \frac{n}{\ell} \rfloor} \tag{8.5.7}$$

uniformly in all initial distribution μ. In particular, $\mu Q^{(n)} \Rightarrow \pi$.

To emphasize the fact that a chain as above *forgets about its initial condition*, consider two distinct initial distributions μ, μ'. By the triangle inequality,

$$\|\mu Q^{(n)} - \mu' Q^{(n)}\|_{\mathrm{TV}} \leq \|\mu Q^{(n)} - \pi\|_{\mathrm{TV}} + \|\pi - \mu Q^{(n)}\|_{\mathrm{TV}} \to 0\,,$$

and so the distribution of X_n with initial distribution μ becomes, asymptotically, indistinguishable from the one started with μ'.

Assumption (8.5.4) is a strong mixing condition. It forces trajectories to meet, in the sense that *any* pair of points (x,y) can be joined during a time interval of length ℓ with positive probability. This implies that two trajectories meet at *some* of the times $\ell, 2\ell, 3\ell, \dots$, and so the coupled chain $\mathbb{S}$ is recurrent. Clearly, (8.5.4) is not realistic when S is infinite, and one can cook up simple examples in which it is not satisfied even in case where S is finite. Consider for example the case where S are the vertices of a square, where the particle can jump to either of its two nearest neighbours with probability $\frac{1}{2}$.

Definition 8.5.1. *Let $x \in S$ be recurrent. Let $I(x) := \{n \geq 1 : Q^{(n)}(x,x) > 0\}$ be the set of times at which a return to x is possible when starting from x. The greatest common divisor of $I(x)$, denoted $d(x)$, is called the* **period** *of x.*

Since x is recurrent, $u(x,x) = \infty > 0$, and so $Q^{(n)}(x,x) > 0$ for infinitely many ns. Therefore, $I(x)$ contains an infinite number of numbers. Moreover, observe that $I(x)$ is stable under addition: if $n, m \in I(x)$ then by the Chapman-Kolmogorov Equation (8.2.3),

$$Q^{(n+m)}(x,x) \geq Q^{(n)}(x,x) Q^{(m)}(x,x) > 0\,,$$

and so $n + m \in I(x)$.

Lemma 8.5.2. *If $x, y \in S$ belong to the same recurrence class, then $d(x) = d(y)$.*

Proof. Since x, y are in the same class, there exists $K \geq 1$ such that $Q^{(K)}(x,y) > 0$ and $L \geq 1$ such that $Q^{(L)}(y,x) > 0$. Therefore,

$$Q^{(K+L)}(y,y) \geq Q^{(L)}(y,x) Q^{(K)}(x,y) > 0\,,$$

which means that $K+L \in I(y)$, and therefore, $d(y)$ divides $K+L$. Then, consider any $n \in I(x)$. We have

$$Q^{(K+n+L)}(y,y) \geq Q^{(L)}(y,x)Q^{(n)}(x,x)Q^{(K)}(x,y) > 0\,,$$

which means that $K+n+L \in I(y)$ and therefore, $d(y)$ divides $K+n+L$. Therefore, $d(y)$ divides n. Since this holds for all $n \in I(x)$, $d(y)$ is a divisor of $I(x)$. As a consequence, $d(y)$ divides $d(x)$. Changing the roles of y and x shows that $d(x)$ divides $d(y)$, and so $d(x) = d(y)$. □

Lemma 8.5.3. *If $d(x) = 1$, then there exists m_0 such that $Q^{(n)}(x,x) > 0$ for all $n \geq m_0$.*

Proof. We first show that when $d(x) = 1$, $I(x)$ must contain two consecutive integers. So let $n_0, n_0 + k \in I(x)$. If $k = 1$ then there is nothing to do. If $k > 1$ then (since $k > d(x)$) there must exist some $l \in I(x)$ which k does not divide. Write $l = km+r$, $0 < r < k$. Since $I(x)$ is stable under addition, the two numbers $(m+1)(n_0+k)$, $(m+1)n_0 + n_1$, are both in $I(x)$. But these numbers differ by less than k:

$$(m+1)(n_0+k) - (m+1)n_0 + n_1 = (m+1)k - (km+r) = k - r < k\,.$$

Proceeding by induction we finally obtain a number N such that $\{N, N+1\} \in I(x)$. Let $m_0 := N^2$. Then each $n \geq N^2$ can be written as $n = N^2 + kN + r$ for some $k \geq 0$, $0 \leq r < N$. We can therefore write n as $n = (N+1)r + N(N-r+k)$, which shows that $n \in I(x)$. □

This lemma says that any point with $d(x) = 1$ can come back to its original position in an arbitray number n of steps, as long as n is sufficiently large. This clearly means that if two independent walks are started at points x, x' with $d(x) = d(x') = 1$, they can meet at any $y \in S$ at time n with positive probability, as soon as n is taken sufficiently large. Of course, depending on x, y, n might have to be taken larger. This shows that imposing $d(x) = 1$ for all $x \in S$ leads to the same recurrence property as (8.5.4), without uniformity in x, y.

Definition 8.5.2. *If $d(x) = 1$ for all $x \in S$, the chain is called* **aperiodic.**

Aperiodicity is an algebraic property that turns all initial conditions equivalent; it does entail that two trajectories started at two different points have a positive probability of meeting along the evolution, but only just (with no uniformity on the time or points). This is enough to guarantee convergence to equilibrium.

Theorem 8.5.2. *Let the chain be irreducible and aperiodic. Assume π is an invariant probability. Then for all initial distribution μ,*

$$\|\mu Q^{(n)} - \pi\|_{\mathrm{TV}} \to 0\,.$$

Proof. We first show that $\mathbb{S}$ is irreducible. So let $(x,y), (x',y')$ be points in $\mathbb{S}$. Since the original chain is irreducible there exist $K \geq 1$ such that $Q^{(K)}(x,x') > 0$ and $L \geq 1$ such that $Q^{(L)}(y,y') > 0$. By Lemma 8.5.3 there exists $n_0 \geq 1$ such that $Q^{(n)}(x,x) > 0$ for all $n \geq n_0$, and $m_0 \geq 1$ such that $Q^{(m)}(y,y) > 0$ for all $m \geq m_0$. For all $n \geq \max\{n_0, m_0\}$, we have

$$\begin{aligned}\mathbb{Q}^{(K+L+n)}((x,y),(x',y')) &= Q^{(K+L+n)}(x,x')Q^{(K+L+n)}(y,y') \\ &\geq Q^{(L+n)}(x,x)Q^{(K)}(x,x')Q^{(K+n)}(y,y)Q^{(L)}(y,y') > 0\,.\end{aligned}$$

Then, since $\pi\otimes\pi$ is an invariant probability for $\mathbb{Q}$, the chain $\mathbb{S}$ is recurrent (Lemma 8.4.2). By the coupling inequality (8.5.3), this shows that $\|\mu Q^{(n)} - \pi\|_{\mathrm{TV}} \to 0$. □

Observe that there exists a chain which is irreducible, aperiodic, recurrent, but in which two copies don't necessarily meet (see [**R.88**] p. 313).

8.6. The Ergodic Theorem

The notion of *invariant measure*, together with the convergence properties described in Theorems 8.5.1 and 8.5.2, gives a fairly satisfactory description of the asymptotic behaviour of an irreducible Markov chain. What still needs to be done is to see how the *empirical quantities* relate to this asymptotic behaviour. For example: what is, up to time n, the time spent by a chain at a site $y \in S$?

THEOREM 8.6.1 (Ergodic Theorem). *Assume the chain is irreducible and positive recurrent. Let π denote the unique invariant probability measure, and consider a non-negative function $f : S \to \mathbb{R}$, integrable with respect to π: $\int |f| d\pi < \infty$. Then for all $x \in S$,*

$$\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \longrightarrow \int f d\pi\,, \quad P_x\text{-a.s.} \tag{8.6.1}$$

This results answers the previous question (in the case of an irreducible, positive recurrent chain). Namely, take $f = \delta_y$. Then $\int f d\pi = \pi(y)$ and by (8.6.1), the fraction of time spent by the chain at y is

$$\frac{1}{n}\sharp\{0 \leq k \leq n-1 : X_k = y\} = \frac{1}{n}\sum_{k=0}^{n-1} 1_{\{X_k=y\}} \longrightarrow \pi(y)\,, \quad P_x\text{-a.s.} \tag{8.6.2}$$

This is very different from the convergence obtained in the previous section. Namely, in the aperiodic case for example, we had obtained $\|\mu Q^{(n)} - \pi\|_{\mathrm{TV}} \to 0$, which implies $P_\mu(X_n = y) \to \pi(y)$ for all $y \in S$, which is a probability of what happens *at time n*. On the other hand, (8.6.2) gives an *almost sure* convergence of the time the *trajectory* spends at y *up to* time n.

PROOF OF THEOREM 8.6.1: Let $x \in S$. We consider the partition of the trajectory into the successive returns of the chain at x:

$$0 =: T_x^0 < T_x^1 < T_x^2 < \dots\,,$$

where $T_x^1 := T_x$, and for $k \geq 2$,

$$T_x^k := \inf\{n > T_x^{k-1} : X_n = x\}\,.$$

Since the chain is irreducuble and since there exists an invariant probability, the chain is recurrent (Lemma 8.4.2), and each T_x^k is P_x-almost surely finite. The result will follow from the fact that the events happening during the time intervals $[T_x^k, T_x^{k+1})$ are independent, and from the Law of Large Numbers. Fix $f : S \to \mathbb{R}$ and define, for all $k \geq 0$,

$$Z_k := \sum_{j=T_x^{(k)}}^{T_x^{(k+1)}-1} f(X_j)\,.$$

Clearly, $Z_k = Z_0 \circ \theta_{T_x^{(k)}}$.

LEMMA 8.6.1. *The sequence $(Z_n)_{n\geq 0}$ is i.i.d.*

PROOF. First observe that for all positive measurable bounded $g : \mathbb{R} \to \mathbb{R}$, the Markov Property at time $T_x^{(k)}$ gives

$$E_x[g(Z_k)] = E_x[(g \circ Z_0) \circ \theta_{T_x^{(k)}}] = E_x[g(Z_0)]\,,$$

and so the Z_ks are identically distributed. For the independence, it is sufficient to show that for all $k \geq 0$,

$$E_x[g_0(Z_0)\dots g_k(Z_k)] = E_x[g_0(Z_0)]\dots E_x[g_k(Z_0)]\,, \tag{8.6.3}$$

where $g_j : \mathbb{R} \to \mathbb{R}$, $j = 0, 1, \dots, k$ are arbitrary bounded functions. This is trivially true when $k = 0$, so assume (8.6.3) holds for $k - 1$. Using again the Markov property at time $T_x^{(k)}$ and the induction hypothesis,

$$\begin{aligned} E_x[g_0(Z_0)\dots g_k(Z_k)] &= E_x[g_0(Z_0)\dots g_{k-1}(Z_{k-1})]E_x[g_k(Z_0)] \\ &= E_x[g_0(Z_0)]\dots E_x[g_k(Z_0)]\,. \end{aligned}$$

This shows (8.6.3) for k. □

Now, observe that since π is invariant it must have the form $\pi = \pi(x)\nu_x$ for all x, where ν_x is the invariant measure of Theorem 8.4.1. This implies that

$$\begin{aligned} E_x[|Z_0|] = E_x[Z_0] &= E_x\Big[\sum_{j=0}^{T_x-1} f(X_j)\Big] \\ &= \sum_{y\in S} f(y) E_x\Big[\sum_{j=0}^{T_x-1} 1_{\{X_j=y\}}\Big] \\ &= \int f d\nu_x = \frac{1}{\pi(x)} \int f d\pi < \infty\,. \end{aligned}$$

Therefore, by the Strong Law of Large Numbers,

$$\frac{1}{n}\sum_{k=0}^{n-1} Z_k \longrightarrow \frac{1}{\pi(x)}\int f d\pi \quad P_x\text{-a.s.} \tag{8.6.4}$$

Let $N_x(n)$ be the number of visits of the chain at x up to time n. Then $T_x^{N_x(n)} \leq n < T_x^{N_x(n)+1}$, and since f is non-negative,

$$\frac{1}{N_x(n)} \sum_{k=0}^{T_x^{N_x(n)}-1} f(X_k) \leq \frac{1}{N_x(n)} \sum_{k=0}^{n} f(X_k) \leq \frac{1}{N_x(n)} \sum_{k=0}^{T_x^{N_x(n)}+1} f(X_k)\,,$$

which is the same as

$$\frac{1}{N_x(n)}\sum_{j=0}^{N_x(n)-1} Z_j \leq \frac{1}{N_x(n)}\sum_{k=0}^{n} f(X_k) \leq \frac{1}{N_x(n)}\sum_{j=0}^{N_x(n)} Z_j\,,$$

By (8.6.4) and since $N_x(n) \to \infty$ P_x-a.s. when $n \to \infty$ (Proposition 8.3.1),

$$\frac{1}{N_x(n)}\sum_{k=0}^{n} f(X_k) \longrightarrow \int f d\nu_x\,.$$

The same expression with $f = 1$ gives $\frac{n}{N_x(n)} \to \nu_x(S) = \frac{1}{\pi(x)}$. This finishes the proof. □

8.7. Exercises

Generalities.

EXERCISE 8.1. [**GS05**] p. 219. A Die is rolled repeatedly. Which of the following are Markov chains? For those that are, supply the transition matrix.

- The largest number X_n shown up to time n.
- The number N_n of sixes in n rolls.
- At time r, the time C_r since the most recent six.
- At time r, the time B_r until the next six.

EXERCISE 8.2. [**GS05**] p. 219. Let $(X_n)_{n\geq 0}$ be the simple random walk starting at the origin. Are $(|X_n|)_{n\geq 0}$ and $(M_n)_{n\geq 0}$ Markov chains? (We defined $M_n := \max\{X_k : 0 \leq k \leq n\}$.) When this is the case, compute the transition matrix. Show that $Y_n := M_n - X_n$ defines a Markov chain. What happens if $X_0 \neq 0$.

EXERCISE 8.3. [**GS05**] p. 220. Let X_n, Y_n be Markov chains on $S = \mathbb{Z}$. is $X_n + Y_n$ necessarily a Markov chain?

EXERCISE 8.4. [**GS05**] p. 220. Let X_n be a Markov chain. Show that for all $1 < r < n$,

$$\begin{aligned} P(X_r = x | X_i = x_i,\, i = 1, 2, \ldots, r-1, r+1, \ldots, n) \\ = P(X_r = x | X_{r-1} = x_{r-1}, X_{r+1} = x_{r+1})\,. \end{aligned}$$

EXERCISE 8.5. Consider "Markov's Other chain" ([**GS05**] p. 218): let $Y_1, Y_3, Y_5, \ldots$ be a sequence of independent identically distributed random variables such that

$$P(Y_{2k+1} = -1) = P(Y_{2k+1} = +1) = \frac{1}{2}.$$

Define then $Y_{2k} := Y_{2k-1}Y_{2k+1}$. Check that $Y_2, Y_4, Y_6, \ldots$ are identically distributed, with the same distribution as above. Is $(Y_k)_{k\geq 1}$ a Markov chain? Enlarge the state space to $\{\pm 1\}^2$ and define $Z_n := (Y_n, Y_{n+1})$.

EXERCISE 8.6. [**R.88**] p.281. Two state Markov chain. Let $S = \{0, 1\}$ with transition matrix

$$Q = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}$$