3. Markov Chain

We consider a countable state space S with σ-algebra $\mathcal{S} := \sigma(S)$.

Definition 3.1. Given a filtration $\mathbb{F} = (\mathcal{F}_t)$ on a probability space $(\Omega, \mathcal{F}, P)$, we call a process X with values in S a *Markov chain*, if

(i) X is adapted,

(ii) $P[X_{t+1} \in B|\mathcal{F}_t] = P[X_{t+1} \in B|X_t]$ for all t and $B \in \mathcal{S}$.

The property *(ii)* is called the Markov property.
A Markov chain is called *time-homogeneous*, if

$$P\left[X_{t+1} \in B|X_t = x\right] = P\left[X_1 \in B|X_0 = x\right]$$

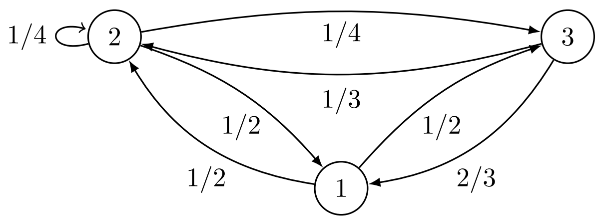
holds for all $x \in S$, $B \in \mathcal{S}$ and t.

For a time-homogeneous Markov chain we define

$$\mu_x := P[X_0 = x],$$
$$p_{xy} := P\left[X_{t+1} = y|X_t = x\right]$$

for $x, y \in S$. The initial distribution $\mu := (\mu_x)_{x\in S}$ is a random vector, that is, it holds $\sum_{x\in S} \mu_x = 1$. We call p_{ij} the transition probability from x to y. The transition matrix $p = (p_{xy})_{x,y\in S}$ is a stochastic matrix, that is $\sum_{y\in S} p_{xy} = 1$ for all $x \in S$.

Example 3.2. Let $S = \{1, 2, 3\}$ and $\mu := \delta_1$.



We obtain

$$p = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 1/4 & 1/4 \\ 2/3 & 1/3 & 0 \end{pmatrix}. \qquad \Diamond$$

Example 3.3 (random walk). Let Y be a stochastic process of independent random variables with values in $\mathbb{Z}^d$. Define $X_t := \sum_{s=0}^{t} Y_t$ and $\mathcal{F}_t := \sigma(X_s \colon s \le t)$. For $y \in \mathbb{Z}^d$, it holds

$$\begin{aligned} P[X_{t+1} = y|\mathcal{F}_t] = P[Y_{t+1} = y - X_t|\mathcal{F}_t] &= \sum_{x\in\mathbb{Z}^d} P[Y_{t+1} = y - x|\mathcal{F}_t]1_{\{X_t=x\}} \\ = \sum_{x\in\mathbb{Z}^d} P[Y_{t+1} = y - x]1_{\{X_t=x\}} &= \sum_{x\in\mathbb{Z}^d} P[Y_{t+1} = y - x|X_t]1_{\{X_t=x\}} \\ &= P[Y_{t+1} = y - X_t|X_t] = P[X_{t+1} = y|X_t]. \end{aligned}$$

Hence, for $B \subseteq \mathbb{Z}^d$, by monotone convergence we get

$$P\left[X_{t+1} \in B | \mathcal{F}_t\right] = \sum_{y \in \mathbb{Z}^d} P\left[X_{t+1} = y | \mathcal{F}_t\right] = \sum_{y \in \mathbb{Z}^d} P\left[X_{t+1} = y | X_t\right] = P[X_{t+1} \in B | X_t]$$

Therefore, X is a Markov chain. Suppose furthermore, that the process Y is identically distributed, then it holds

$$P\left[X_{t+1} = y | X_t = x\right] = P\left[Y_t = y - x | X_t = x\right] = P[Y_1 = y - x] = P[X_1 = y | X_0 = x]$$

Therefore, it is in that case time-homogeneous. $\Diamond$

Given a stochastic vector μ and a stochastic matrix p, the question is whether there exists a probability space $(\Omega, \mathcal{F}, P_\mu)$, a filtration $\mathbb{F}$ and a stochastic process X such that X is a time-homogeneous Markov chain with start distribution μ and transition probability p. To do so, we define

- $\Omega = S^{\mathbb{N}_0} = \{\omega = (\omega_t) \colon \omega_t \in S\}$;
- $\mathcal{F} = \otimes_{t \in \mathbb{N}_0} \mathcal{S}$;
- X as being the canonical process, that is

$$X_t(\omega) = \omega_t, \quad \omega = (\omega_t) \in \Omega$$

- $\mathbb{F}$ being the filtration generated by X, that is

$$\mathcal{F}_t = \sigma(X_s \colon s \leq t)$$

The product σ-algebra $\mathcal{F}$ is generated by the semi-ring of finite product cylinders

$$A = A_0 \times \cdots \times A_t \times S \times S \times \cdots = \{X_0 \in A_0, X_1 \in A_1, \cdots, X_t \in A_t\}, \quad A_s \in \mathcal{S}.$$

We define the function $P_\mu : \mathcal{R} \to [0,1]$ as follows

$$P_\mu[A] := \sum_{x_0 \in A_0, \cdots, x_t \in A_t} \mu_{x_0} p_{x_0 x_1} \cdots p_{x_{t-1} x_t}$$

This function is well defined. Indeed, let

$$A = A_0 \times \ldots \times A_t \times S \times S \times \ldots \quad \text{and} \quad B = B_0 \times \ldots \times B_s \times S \times S \ldots.$$

be such that $A = B$ and without loss of generality, suppose that $s \leq t$. It follows that $A_u = B_u$ for every $u \leq s$ and $A_u = \mathcal{S}$ for $u = s+1, \ldots, t$. Hence, since p is a stochastic matrix, it follows that

$$\begin{aligned} P_\mu[A] = & \sum_{x_0 \in A_0, \cdots, x_t \in A_t} \mu_{x_0} p_{x_0 x_1} \cdots p_{x_{t-1} x_t} \\ & = \sum_{x_0 \in A_0, \cdots, x_s \in A_s} \mu_{x_0} p_{x_0 x_1} \cdots p_{x_{s-1} x_s} \left(\sum_{x_{s+1} \in S, \cdots, x_t \in S} p_{x_s x_{s+1}} \cdots p_{x_{t-1} x_t} \right) \\ & = \sum_{x_0 \in A_0, \cdots, x_s \in A_s} \mu_{x_0} p_{x_0 x_1} \cdots p_{x_{s-1} x_s} = P_\mu[B] \end{aligned}$$

Clearly, $P[\emptyset] = 0$ and since μ is a stochastic vector, it holds for $A_0 = S$

$$P_\mu[\Omega] = P[A_0 \times S \times S \times \cdots] = \sum_{x_0 \in S} \mu_{x_0} = 1$$

Let us show that P_μ is additive by taking a pairwise disjoint finite family $(A^k)_{k\leq n}$ of elements in $\mathcal{R}$ such that $A = \cup_{k\leq n} A^k$ is also in $\mathcal{R}$. Denote by t the maximal dimension of the $(A^k)_{k\leq n}$ and A. By definition, it follows from the disjointness of the (A^k) that

$$\begin{aligned} P_\mu[A] = &\sum_{x_0\in A_0,\cdots,x_t\in A_t} \mu_{x_0} p_{x_0 x_1}\cdots p_{x_{t-1}x_t} = \sum_{\omega\in A=\cup_{k\leq n}A^k} \mu_{\omega_0} p_{\omega_0,\omega_1}\cdots p_{\omega_{t-1}\omega_t} \\ &= \sum_{k\leq n}\sum_{\omega\in A^k} \mu_{\omega_0} p_{\omega_0,\omega_1}\cdots p_{\omega_{t-1}\omega_t} = \sum_{k\leq n}\sum_{x_0\in A_0^k,\dots,x_t\in A_t^k} \mu_{x_0} p_{x_0x_1}\cdots p_{x_{t-1}x_t} = \sum_{k\leq n} P_\mu[A^k] \end{aligned}$$

Hence P_μ is finitely additive. We can therefore extend this measure to the ring $\mathcal{C}$ generated by the semi-ring $\mathcal{R}$. Indeed, as mentioned after the Definition 1.33, $\mathcal{C}$ is given by

$$\mathcal{C} = \left\{\cup_{k\leq n} A^k \colon A^1,\dots,A^n \in \mathcal{R} \text{ pairwise disjoint}\right\}.$$

Therefore as mentioned after Definition 1.34, we can extend the function P_μ to $\mathcal{C}$ as follows

$$P_\mu[A] := \sum_{k=1}^n P\left[A^k\right], \quad A = \cup_{k\leq n} A^k, A^1,\dots,A^n \text{ disjoints elements in } \mathcal{R}$$

You can also check that $P_\mu : \mathcal{C} \to [0,1]$ is well defined and inherits the properties $P_\mu[\emptyset] = 0$, $P_\mu[\Omega] = 1$ and additivity. Since $\mathcal{C}$ is in particular a semi-ring, we just have to show that P_μ is σ-additive. However, $\mathcal{C}$ being a ring, we can apply Lemma 1.36, that tells that σ-additivity is is equivalent to continuity at $\emptyset$. In other terms we have to show that if (A^n) is a decreasing sequence of sets in $\mathcal{C}$ such that $\lim A^n = \cap A^n = \emptyset$, then it follows that $P_\mu[A^n] \to 0$. Suppose by contradiction that $P_\mu[A^n] > \varepsilon > 0$ for every n. Since each A^n is a finite disjoint union of elements in $\mathcal{R}$, there exists t^n for every n such that after the coordinate t^n, A^n is the infinite product of S. Without loss of generality, up to re-indexing or adding some new sets, we can assume also that $t_n = n$. For ease of notations, given a set $A \subseteq \Omega$, we denote by A_k the projection of the first $k+1$ coordinates, so that it holds for instance[38]

$$A^n = A^n_n \times S \times S \times \cdot$$

And reciprocally, for a set $A_n \subseteq \prod_{k=0}^n S$, we denote by $A := A_n \times S \times S \cdots$.
Since the state space $\prod_{k=0}^n S$ is countable and $P_\mu[A^n] = P_\mu[A^n_n \times S \times S \times \cdots] \geq \varepsilon$, it follows that we can choose a non-empty finite set $K^n_n \subseteq A^n_n$ such that

$$P_\mu[A^n \setminus K^n] = P_\mu[(A^n_n \setminus K^n_n) \times S \times S \times \cdots] \leq \varepsilon/2^{n+1}$$

by the definition of P_μ. It follows that

$$P_\mu[K^n] = P_\mu\left[K^n_n \times S \times S \times \dots\right] \geq \varepsilon - \varepsilon/2^{n+1}$$

Now we define the sequence $\tilde{K}^n = \cap_{k=0}^n K^k$ which is a decreasing sequence per definition. Furthermore it holds

$$P_\mu[A^n \setminus \tilde{K}^n] = P_\mu\left[(\cap_{k=0}^n A^k) \setminus (\cap_{k=0}^n K^k)\right] \leq \sum_{k=0}^n P_\mu\left[A^k \setminus K^k\right] \leq \varepsilon \sum_{k=0}^n 2^{-(k+1)} \leq \varepsilon/2$$

[38] Indeed, after n, A^n is the infinite product of S.

Hence

$$P\left[\tilde{K}^n\right] = P_\mu\left[A^n\right] - P_\mu\left[A^n \setminus \tilde{K}^n\right] \geq \varepsilon - \varepsilon/2 = \varepsilon/2$$

for every n. However, since $\tilde{K}_0^n$ is finite, decreasing and $P_\mu[\tilde{K}_0^n \times S \times S \times \cdots] \geq P_\mu[\tilde{K}^n] \geq \varepsilon > 2$, it follows that there must exists $\omega_0 \in \tilde{K}_n^0$ for every n. The same argumentation for $\tilde{K}_1^n$, shows that there must exists ω_1 such that $(\omega_0, \omega_1) \in \tilde{K}_1^n$ for every n. Doing so, construct a sequence $\omega = (\omega_0, \omega_1, \ldots) \in \Omega$ such that $(\omega_0, \omega_1, \ldots, \omega_n) \in \tilde{K}_n^n$ for every n. Since $\tilde{K}^n = \tilde{K}_n^n \times S \times S \times \ldots$, it follows that $\omega \in \tilde{K}^n$ for every n, which contradicts however the emptyness of $\cap \tilde{K}^n$. Hence, P_μ is σ-additive.
This argumentation allows to show the following proposition.

Proposition 3.4. *Let μ be a probability vector and p a probability matrix with values in $\mathbb{R}^S$. Then there exists a probability measure P_μ on $(\Omega, \mathcal{F})$ where $\Omega = S^{\mathbb{N}_0}$, $\mathcal{F} = \otimes_{\mathbb{N}_0}\mathcal{S}$ such that the canonical process X given by*

$$X_t(\omega) = \omega_t, \quad \omega = (\omega_t) \in \Omega,$$

is a time-homegeneous Markov chain under P_μ with initial distribution μ and transition probability p in its own filtration $\mathbb{F}$ given by

$$\mathcal{F}_t = \sigma\left(X_s \colon s \leq t\right)$$

Proof. We already constructed the probability measure P_μ, we are left to show that X is a time-homogeneous Markov chain under P_μ with initial distribution μ and transition probability p in its own filtration $\mathbb{F}$. Adaptiveness of X follows immediately. As for the Markov and time homogenity property, on the one hand it holds

$$\begin{aligned} &P_\mu\left[X_{t+1} = x_{t+1} | X_t = x_t, \ldots, X_0 = x_0\right] \\ &\qquad = \frac{P_\mu\left[X_{t+1} = x_{t+1}, X_t = x_t, \ldots, X_0 = x_0\right]}{P_\mu\left[X_t = x_t, \ldots, X_0 = x_0\right]} = \frac{\mu_{x_0} p_{x_0 x_1} \cdots p_{x_{t-1} x_t} p_{x_t, x_{t+1}}}{\mu_{x_0} p_{x_0 x_1} \cdots p_{x_{t-1} x_t}} = p_{x_t x_{t+1}} \end{aligned}$$

whereas

$$\begin{aligned} &P_\mu\left[X_{t+1} = x_{t+1} | X_t = x_t\right] \\ &\qquad = \frac{P_\mu\left[X_{t+1} = x_{t+1}, X_t = x_t\right]}{P_\mu\left[X_t = x_t\right]} = \frac{P_\mu\left[X_{t+1} = x_{t+1}, X_t = x_t, X_{t-1} \in S, \ldots, X_0 \in S\right]}{P_\mu\left[X_t = x_t, X_{t-1} \in S \ldots, X_0 \in S\right]} \\ &\qquad = \frac{\left(\sum_{x_0 \in S, \ldots, x_{t-1} \in S} \mu_{x_0} p_{x_0 x_1} \cdots p_{x_{t-2} x_{t-1}}\right) p_{x_{t-1} x_t} p_{x_t, x_{t+1}}}{\left(\sum_{x_0 \in S, \ldots, x_{t-1} \in S} \mu_{x_0} p_{x_0 x_1} \cdots p_{x_{t-2} x_{t-1}}\right) p_{x_{t-1} x_t}} = p_{x_t x_{t+1}} \end{aligned}$$

showing that

$$P_\mu\left[X_{t+1} = x_{t+1} | X_t = x_t, \ldots, X_0 = x_0\right] = P_\mu\left[X_{t+1} = x_{t+1} | X_t = x_t\right] = P_\mu\left[X_1 = x_{t+1} | X_0 = x_t\right]$$

and therefore the time-homogeneity property. As for the Markov property, by monotone convergence, it holds

$$
\begin{aligned}
P_\mu\left[X_{t+1} \in A|\mathcal{F}_t\right] = \sum_{x_0,\dots x_t \in S} P_\mu\left[X_{t+1} \in B|X_t = x_t, \dots, X_0 = x_0\right] 1_{\{X_t = x_t, \dots, X_0 = x_0\}} \\
= \sum_{x_{t+1} \in B} \sum_{x_0,\dots x_t \in S} P_\mu\left[X_{t+1} = x_{t+1}|X_t = x_t, \dots, X_0 = x_0\right] 1_{\{X_t = x_t, \dots, X_0 = x_0\}} \\
= \sum_{x_{t+1} \in B} \sum_{x_0,\dots x_t \in S} P_\mu\left[X_{t+1} = x_{t+1}|X_t = x_t\right] 1_{\{X_t = x_t, \dots, X_0 = x_0\}} \\
= \sum_{x_0,\dots x_t \in S} P_\mu\left[X_{t+1} \in B|X_t = x_t\right] 1_{\{X_t = x_t, \dots, X_0 = x_0\}} \\
= \sum_{x_t \in S} P_\mu\left[X_{t+1} \in B|X_t = x_t\right] 1_{\{X_t = x_t\}} \left(\sum_{x_0,\dots x_{t-1} \in S} 1_{\{X_{t-1} = x_{t-1}, \dots, X_0 = x_0\}} \right) \\
= \sum_{x_t \in S} P_\mu\left[X_{t+1} \in B|X_t = x_t\right] 1_{\{X_t = x_t\}} = P_\mu\left[X_{t+1} \in B|X_t\right]
\end{aligned}
$$

which ends the proof. □

Remark 3.5. For a time-homogeneous Markov chain X on some arbitrary filtrated probability space $(\Omega, \mathcal{F}, \mathbb{F}, P)$ with values in S, start distribution $\mu_x = P[X_0 = x]$ and transition probability $p_{xy} = P[X_{t+1} = y|X_t = x]$ let P_μ denote the respective probability measure on the canonical space $(S^{\mathbb{N}_0}, \otimes_{\mathbb{N}_0}\mathcal{S})$. Then it holds

$$
P_\mu[A] = P_X[A] = P[X \in A], \quad A \in \mathcal{F}.
$$

So, without loss of generality, it is therefore enough to consider time-homogeneous Markov chains on the canonical space. ♦

Throughout this Chapter we always consider a time homogeneous Markov chain on the canonical space! Hence, from now on, X denotes the canonical process on $(\Omega, \mathcal{F}) = (S^{\mathbb{N}_0}, \otimes_{\mathbb{N}_0}\mathcal{S})$.
We define the shift-operator

$$
\begin{aligned}
\theta_s \colon \Omega &\longrightarrow \Omega \\
\omega = (\omega_t) &\longmapsto \theta_s(\omega) = (\omega_{t+s})
\end{aligned}
$$

Theorem 3.6 (Markov property). *Let H be a bounded random variable. Then*

$$
E_\mu\left[H \cdot \theta_t|\mathcal{F}_t\right] = E_{X_t}\left[H\right]
$$

where

$$
E_{X_t}[H] = \sum_{x \in S} E_x\left[H\right] 1_{\{X_t = x\}}
$$

and E_x is the expectation under the measure $P_x = P_{\delta_x}$ given by the Markov chain starting at time 0 from x with probability 1.[39]

Proof. Since $E_{X_t}\left[H\right] = \sum_{y \in S} E_y\left[H\right] 1_{\{X_t = y\}}$, it follows that $E_{X_t}[H]$ is $\mathcal{F}_t$-measurable. So we just have to show that for every $A \in \mathcal{F}_t$, it holds

$$
E_\mu\left[1_A H \circ \theta_t\right] = E_\mu\left[1_A E_{X_t}[H]\right].
$$

[39]That is $P[X_0 = x] = 1$ and $P[X_0 = y] = 0$ for every $y \neq x$.

Step 1: We consider in this step bounded random variable H of the form $H = 1_B$ where $B = \{X_0 = y_0, \dots, X_s = y_s\} \in \mathcal{F}$ for $y_0, \dots, y_s \in S$ whereby s is an arbitrary time.

Starting with $A = \{X_0 = x_0, X_1 = x_1, \dots, X_t = x_t\} \in \mathcal{F}_t$ for $x_0, \dots, x_t \in S$, we have on the one hand

$$\begin{aligned} E_\mu\left[1_A H \cdot \theta_t\right] &= P_\mu\left[A \cap \{X_t = y_0, \dots, X_{t+s} = x_s\}\right] \\ &= P_\mu\left[X_0 = x_0, \dots, X_t = x_t, X_t = y_0, \dots, X_{t+s} = y_s\right] \\ &= \delta_{x_t}(y_0)\mu_{x_0} \cdots p_{x_0 x_1} \cdots p_{x_{t-1}x_t} p_{y_0 y_1} \cdots p_{y_{s-1}y_s} \end{aligned}$$

where $\delta_{x_t}(y_0) = 1$ if $x_t = y_0$ and 0 otherwise. On the other hand, it holds

$$\begin{aligned} E_\mu\left[1_A E_{X_t}\left[H\right]\right] &= E_\mu\left[1_{\{X_0=x_0,\dots,X_t=x_t\}} E_{X_t}[H]\right] = E_\mu\left[1_{\{X_0=x_0,\dots,X_t=x_t\}} E_{x_t}\left[H\right]\right] \\ &= E_\mu\left[1_{\{X_0=x_0,\dots,X_t=x_t\}} P_{x_t}\left[X_0 = y_0, \dots, X_t = y_t\right]\right] \\ &= P_\mu\left[X_0 = x_0, \dots, X_t = x_t\right] P_{x_t}\left[X_0 = y_0, \dots, X_t = y_t\right] \\ &= \mu_{x_0} \cdots p_{x_0 x_1} \cdots p_{x_{t-1}x_t} \delta_{x_t}(y_0) p_{y_0 y_1} \cdots p_{y_{s-1}y_s} \end{aligned}$$

Showing that $E_\mu\left[1_A H \cdot \theta_t\right]$ for every A of the form $A = \{X_0 = x_0, X_1 = x_1, \dots, X_t = x_t\}$.

However, every set $A \in \mathcal{F}_t$ is a countable disjoint union of sets (A^n) of the form $\{X_0 = x_0, X_1 = x_1, \dots, X_t = x_t\}$. Hence, by dominated convergence, it follows that

$$E_\mu\left[1_A H \circ \theta_t\right] = \sum E_\mu\left[1_{A_n} H \circ \theta_t\right] = \sum E_\mu\left[1_{A_n} E_{X_t}[H]\right] = E_\mu\left[1_A E_{X_t}[H]\right]$$

for every $A \in \mathcal{F}_t$ showing the assertion in the case where $H = 1_B$.

Step 2: Every positive bounded random variable H is the increasing limit of simple functions H^n of the form $H^n = \sum_{k \leq m_n} \alpha_k^n 1_{B_k^n}$ where $\alpha_k^n \in \mathbb{R}$, and each B_k^n is of the form $\{X_0 = y_0, \dots, X_s = y_s\}$. Hence, by monotone convergence, for every $A \in \mathcal{F}_t$ it holds from the previous point that

$$E_\mu\left[1_A H \circ \theta_t\right] = \lim_n \sum_{k=1}^{n} \alpha_k^n E_\mu\left[1_A 1_{B_k^n} \circ \theta_t\right] = \lim_n \sum_{k=1}^{n} \alpha_k^n E_\mu\left[1_A E_{X_t}[1_{B_n^k}]\right] = E_\mu\left[1_A E_{X_t}[H]\right]$$

showing that for every bounded positive random variable, it holds

$$E_\mu\left[H \circ \theta_t | \mathcal{F}_t\right] = E_{X_t}\left[H\right].$$

The general case of bounded random variables follows from applying the positive case to H^+ and H^- and taking the difference. □

Theorem 3.7 (Chapman-Kolmogorov). *For every two times s and t as well as every two states x and z in S, it holds*

$$P_x\left[X_{t+s} = z\right] = \sum_{y \in S} P_x\left[X_t = y\right] P_y\left[X_s = z\right]$$

Proof. By Theorem 3.6 we get

$$\begin{aligned} P_x[X_{t+s} = z] &= E_x\left[1_{\{X_{t+s}=z\}}\right] = E_x\left[E_x\left[1_{\{X_{t+s}=z\}} | \mathcal{F}_t\right]\right] \\ &= E_x\left[E_x\left[1_{\{X_{s}=z\}} \circ \theta_t | \mathcal{F}_t\right]\right] = E_x\left[E_{X_t}\left[1_{\{X_s=z\}}\right]\right] = \sum_{y \in S} E_x\left[E_y\left[1_{\{X_s=z\}}\right] 1_{\{X_t=y\}}\right] \\ &= \sum_{y \in S} E_x\left[1_{\{X_t=y\}}\right] E_y\left[1_{\{X_s=z\}}\right] = \sum_{y \in S} P_x\left[X_t = y\right] P_y\left[X_s = z\right]. \end{aligned}$$ □

Remark 3.8 (a repetition on Markov chains). Given a time homogeneous Markov Chain, we define per induction[40]

$$p^1_{xy} := p_{xy}; \quad p^2_{xy} := \sum_{z\in S} p_{xz}p_{zy}, \quad \dots, \quad p^k_{xy} := \sum_{z\in S} p^{k-1}_{xz}p_{zy}, \quad \dots$$

so that

$$P_\mu[X_t = y|X_0 = x] = p^t_{xy}$$

and by Chapman-Kolmogorov, it holds

$$p^{t+s}_{ij} = \sum_{z\in S} p^s_{xz}p^t_{zy}. \tag{3.1}$$

♦

Theorem 3.9 (strong Markov property). *Let H be a bounded random variable and τ be a stopping time. Then it holds*

$$1_{\{\tau<\infty\}}E_\mu\left[H \circ_\tau |\mathcal{F}_\tau\right] = 1_{\{\tau<\infty\}}E_{X_\tau}\left[H\right].$$

Proof. For every $A \in \mathcal{F}_\tau$ we get by Theorem 3.6 and the fact that $A \cap \{\tau = t\} \in \mathcal{F}_t$ for every t,

$$\begin{aligned} E_\mu\left[1_A 1_{\{\tau<\infty\}} H\circ\theta_\tau\right] &= \sum_t E_\mu\left[1_A 1_{\{\tau=t\}} H\circ\theta_t\right] \\ &= \sum_t E_\mu\left[1_A 1_{\{\tau=t\}} E_{X_t}\left[H\right]\right] = E_\mu\left[1_A 1_{\{\tau<\infty\}} E_{X_\tau}\left[H\right]\right], \end{aligned}$$

The claim follows. □

3.1. Recurrence and transience

We are still considering a time homogeneous Markov chain with start distribution μ and transition probability p on the canonical space.

Definition 3.10. Given $y \in S$, define recursively

$$\tau^0_y := 0, \quad \tau_y = \tau^1_y := \inf\left\{t > \tau^0_y \colon X_t = y\right\}, \quad \dots, \quad \tau^k_y := \inf\left\{t > \tau^{k-1}_y \colon X_t = y\right\}, \quad \cdots$$

Here τ^k_y denotes the k-th return time to the state y.
We further denote by

$$\rho_{xy} := P_x\left[\tau_y < \infty\right]$$

the probability that starting from x, the Markov chain visit the state y at least once. We say that a time homogeneous Markov chain is

- *recurrent*: if $\rho_{xx} = 1$;
- *transient*: if $\rho_{xx} < 1$.

Theorem 3.11. *For $x, y \in S$, it holds*

$$P_x[\tau^k_y < \infty] = \rho_{xy}\rho^{k-1}_{yy}.$$

[40] Be aware that p^k_{xy} is not p_{xy} to the power k.

Proof. We show it per induction. Per definition, it holds $P_x[\tau_y^1 < \infty] = \rho_{xy}$. Suppose therefore that the claim holds for every $l = 1, \dots, k-1$ and we show it for k. Define $\tau = \tau_y^{k-1}$ and $H := 1_{\{\tau_y < \infty\}}$. Since $\{\tau_y^k < \infty\} = \{\tau_y \circ \theta_\tau < \infty\}$, we get that $1_{\{\tau<\infty\}}$ is bounded and $\mathcal{F}_\tau$-measurable. Using the strong Markov property, Theorem 3.9, it follows that

$$\begin{aligned} P_x[\tau_y^k < \infty] &= P_x[\tau_y \circ \theta_\tau < \infty] = E_x\left[1_{\{\tau<\infty\}}(H \circ \theta_\tau)\right] \\ &= E_x\left[1_{\{\tau<\infty\}} E_x\left[H \circ \theta_\tau | \mathcal{F}_\tau\right]\right] = E_x\left[1_{\{\tau<\infty\}} E_{X_\tau}\left[H\right]\right] = E_x\left[1_{\{\tau<\infty\}} P_y[\tau_y < \infty]\right] \\ &= \rho_{yy} P_x[\tau_y^{k-1} < \infty] = \rho_{yy}\rho_{xy}\rho_{yy}^{k-2} = \rho_{xy}\rho_{yy}^{k-1}. \end{aligned}$$ □

Remark 3.12. It holds

$$\cap_{k\in\mathbb{N}}\{\tau_x^k < \infty\} = \{X_t = x \text{ for infinitely many time } t\} = \limsup\{X_t = x\}.$$

If y is recurrent it holds $P_x[\tau_x^k < \infty] = \rho_{xx}^t = 1$ so that $P_x[\limsup\{X_t = x\}] = 1$. ♦

For $y \in S$ we define

$$N_y := \sum_t 1_{\{X_t = y\}},$$

counting how often X visits in the state y.

Theorem 3.13. *For $x \in S$, it holds*

(i) x is recurrent implies $E_[N_x] = \infty$.

(ii) $y \in S$ is transient implies $E_x[N_y] = \rho_{xy}/(1-\rho_{yy}) < \infty$.

In particular, x is recurrent if and only if $E_x[N_x] = \infty$.

Proof. *(i)* If x is recurrent, it holds

$$E_x\left[N_x\right] = \sum_{k\in\mathbb{N}} P_x[N_x \geq k] = \sum_{k\in\mathbb{N}} P_x[\tau_x^k < \infty] = \sum_{k\in\mathbb{N}} \rho_{xx}^k = \sum_{k\in\mathbb{N}} 1 = \infty$$

(ii) If y is transient, then by Theorem 3.11 we obtain

$$E_x[N_y] = \sum_{k\in\mathbb{N}} P_x[N_y \geq k] = \sum_{k\in\mathbb{N}} P_x[\tau_y^k < \infty] = \sum_{k\in\mathbb{N}} \rho_{xy}\rho_{yy}^{k-1} = \rho_{xy}\frac{1}{1-\rho_{yy}}.$$ □

Theorem 3.14. *Let x and y be two states in S, where x is recurrent and $\rho_{xy} > 0$. Then y is recurrent and $\rho_{yx} = 1$.*

Proof. Let us first show that $\rho_{yx} = 1$. Since x is recurrent, it follows that $\tau_x(\omega) < \infty$ for almost all $\omega \in \Omega$. Hence, for almost all $\omega \in \Omega$ such that $\tau_y(\omega) < \infty$ we get $\tau_x \circ \theta_{\tau_y}(\omega) < \infty$. Thus with $H := 1_{\{\tau_x = \infty\}}$, Theorem 3.9, and the fact that $X_{\tau_y} = y$, we get

$$\begin{aligned} 0 = P_x[\tau_y < \infty, \tau_x \circ \theta_{\tau_y} = \infty] &= E_x\left[1_{\{\tau_y<\infty\}} 1_{\{\tau_x\circ\theta_{\tau_y} = \infty\}}\right] = E_x\left[1_{\{\tau_y<\infty\}} E_x\left[H \circ \theta_{\tau_y} | \mathcal{F}_{\tau_y}\right]\right] \\ &= E_x\left[1_{\{\tau_y<\infty\}} E_{X_{\tau_y}}\left[H\right]\right] = E_x\left[1_{\{\tau_y<\infty\}} P_y[\tau_x = \infty]\right] = \rho_{xy}(1-\rho_{yx}). \end{aligned}$$

Since $\rho_{xy} > 0$, it must be that $\rho_{yx} = 1$.

Let us finally show that y is recurrent. Let i and j be two states in S and $k \in \mathbb{N}$. Then by Theorem 3.7 and an induction we get $P_i[X_k = j] = p^k_{ij}$, see Remark 3.8. Since $\rho_{xy} > 0$ and $\rho_{yx} > 0$, there exist $k_1, k_2 \in \mathbb{N}$ with $p^{k_1}_{xy} > 0$ and $p^{k_2}_{yx} > 0$. By Theorem 3.7 we get

$$p^{k_1+t+k_2}_{yy} \geq p^{k_2}_{yx} p^t_{xx} p^{k_1}_{xy}$$

so that

$$E_y[N_y] = \sum_t P_y[X_t = y] = \sum_t p^t_{yy} \geq \sum_t p^{k_2}_{yx} p^t_{xx} p^{k_1}_{xy}$$
$$= p^{k_2}_{yx} \left(\sum_{t \in \mathbb{N}} p^t_{xx} \right) p^{k_1}_{xy} = p^{k_2}_{yx} E_x[N_x] p^{k_1}_{xy} = \infty.$$

Theorem 3.13 implies that y is recurrent. □