5 Markov Chains

In various applications one considers collections of random variables which evolve in time in some random but prescribed manner (think, eg., about consecutive flips of a coin combined with counting the number of heads observed). Such collections are called random (or stochastic) processes. A typical random process X is a family $\{X_t : t \in T\}$ of random variables indexed by elements of some set T. When $T = \{0, 1, 2, \dots\}$ one speaks about a 'discrete-time' process, alternatively, for $T = \mathbb{R}$ or $T = [0, \infty)$ one has a 'continuous-time' process. In what follows we shall only consider discrete-time processes.

5.1 Markov property

Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space and let $\{X_0, X_1, \dots\}$ be a sequence of random variables[39] which take values in some countable set S, called the state space. We assume that each X_n is a discrete[40] random variable which takes one of N possible values, where $N = |S|$ (N may equal $+\infty$).

✍ **Definition 5.1.** The process X is a *Markov chain* if it satisfies the *Markov property*:

$$\begin{aligned}\mathsf{P}(X_{n+1} = x_{n+1} \,|\, X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \\ = \mathsf{P}(X_{n+1} = x_{n+1} \,|\, X_n = x_n)\end{aligned} \tag{5.1}$$

for all $n \geq 1$ and all x_0, x_1, ..., $x_{n+1} \in S$.

Interpreting n as the 'present' and $n+1$ as a 'future' moment of time, we can re-phrase the Markov property (5.1) as "given the present value of a Markov chain, its future behaviour does not depend on the past".

Remark 5.1.1. It is straightforward to check that the Markov property (5.1) is equivalent to the following statement:
for each $s \in S$ and every sequence $\{x_k : k \geq 0\}$ in S,

$$\mathsf{P}(X_{n+m} = s \,|\, X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = \mathsf{P}(X_{n+m} = s \,|\, X_n = x_n)$$

for any $m, n \geq 0$.

The evolution of a chain is described by its 'initial distribution'

$$\mu_k^0 \stackrel{\mathsf{def}}{=} \mathsf{P}(X_0 = k)$$

and its 'transition probabilities'

$$\mathsf{P}(X_{n+1} = j \,|\, X_n = i)\,;$$

it can be quite complicated in general since these probabilities depend upon the three quantities n, i, and j.

[39] ie, each X_n is a $\mathcal{F}$-measurable mapping from Ω into S.
[40] without loss of generality, we can and *shall* assume that S is a subset of integers.

Definition 5.2. A Markov chain X is called *homogeneous* if

$$\mathsf{P}(X_{n+1} = j \mid X_n = i) \equiv \mathsf{P}(X_1 = j \mid X_0 = i)$$

for all n, i, j. The *transition matrix* $\mathbf{P} = (p_{ij})$ is the $|S| \times |S|$ matrix of *transition probabilities*

$$p_{ij} = \mathsf{P}(X_{n+1} = j \mid X_n = i)\,.$$

In what follows we shall only consider homogeneous Markov chains.

The next claim characterizes transition matrices.

✍ **Theorem 5.3.** $\mathbf{P}$ *is a* stochastic matrix, *which is to say that*

a) *every entry of* $\mathbf{P}$ *is non-negative,* $p_{ij} \geq 0$*;*

b) *each row sum of* $\mathbf{P}$ *equals one, ie., for every* $i \in S$ *we have* $\sum_j p_{ij} = 1$.

Example 5.4. [Bernoulli process] Let $S = \{0, 1, 2, \dots\}$ and define the Markov chain Y by $Y_0 = 0$ and

$$\mathsf{P}(Y_{n+1} = s + 1 \mid Y_n = s) = p\,, \qquad \mathsf{P}(Y_{n+1} = s \mid Y_n = s) = 1 - p\,,$$

for all $n \geq 0$, where $0 < p < 1$. You may think of Y_n as the number of heads thrown in n tosses of a coin.

Example 5.5. [Simple random walk] Let $S = \{0, \pm 1, \pm 2, \dots\}$ and define the Markov chain X by $X_0 = 0$ and

$$p_{ij} = \begin{cases} p, & \text{if } j = i + 1, \\ q = 1 - p, & \text{if } j = i - 1, \\ 0, & \text{otherwise.} \end{cases}$$

Example 5.6. [Ehrenfest chain] Let $S = \{0, 1, \dots, r\}$ and put

$$p_{k,k+1} = \frac{r - k}{r}\,, \qquad p_{k,k-1} = \frac{k}{r}\,, \qquad p_{ij} = 0 \quad \text{otherwise}\,.$$

In words, there is a total of r balls in two urns; k in the first and $r - k$ in the second. We pick one of the r balls at random and move it to the other urn. Ehrenfest used this to model the division of air molecules between two chambers (of equal size and shape) which are connected by a small hole.

Example 5.7. [Birth and death chains] Let $S = \{0, 1, 2, \dots, \}$. These chains are defined by the restriction $p_{ij} = 0$ when $|i - j| > 1$ and, say,

$$p_{k,k+1} = p_k\,, \qquad p_{k,k-1} = q_k\,, \qquad p_{kk} = r_k$$

with $q_0 = 0$. The fact that these processes cannot jump over any integer makes the computations particularly simple.

Definition 5.8. The *n-step transition matrix* $\mathbf{P}_n = \bigl(p_{ij}(n)\bigr)$ is the matrix of *n-step transition probabilities*

$$p_{ij}(n) \equiv p_{ij}^{(n)} \stackrel{\mathsf{def}}{=} \mathsf{P}(X_{m+n} = j \mid X_m = i)\,.$$

Of course, $\mathbf{P}_1 = \mathbf{P}$.

✍ **Theorem 5.9** (Chapman-Kolmogorov equations)**.** *We have*

$$p_{ij}(m+n) = \sum_k p_{ik}(m)\, p_{kj}(n)\,. \tag{5.2}$$

Hence $\mathbf{P}_{m+n} = \mathbf{P}_m\, \mathbf{P}_n$, and so $\mathbf{P}_n = \mathbf{P}^n \equiv (\mathbf{P})^n$, the n-th power of $\mathbf{P}$.

Proof. In view of the tower property $\mathsf{P}(A \cap B \mid C) = \mathsf{P}(A \mid B \cap C)\,\mathsf{P}(B \mid C)$ and the Markov property, we get

$$\begin{aligned} p_{ij}(m+n) &= \mathsf{P}(X_{m+n} = j \mid X_0 = i) \\ &= \sum_k \mathsf{P}(X_{m+n} = j, X_m = k \mid X_0 = i) \\ &= \sum_k \mathsf{P}(X_{m+n} = j \mid X_m = k, X_0 = i)\,\mathsf{P}(X_m = k \mid X_0 = i) \\ &= \sum_k \mathsf{P}(X_{m+n} = j \mid X_m = k)\,\mathsf{P}(X_m = k \mid X_0 = i) \\ &= \sum_k p_{kj}(n)\, p_{ik}(m) = \sum_k p_{ik}(m)\, p_{kj}(n)\,. \end{aligned}$$

□

Let $\mu_i^{(n)} \stackrel{\mathsf{def}}{=} \mathsf{P}(X_n = i)$, $i \in S$, be the mass function of X_n; we write $\boldsymbol{\mu}^{(n)}$ for the row vector with entries $(\mu_i^{(n)} : i \in S)$.

Lemma 5.10. *We have $\boldsymbol{\mu}^{(m+n)} = \boldsymbol{\mu}^{(m)}\,\mathbf{P}_n$, and hence $\boldsymbol{\mu}^{(n)} = \boldsymbol{\mu}^{(0)}\,\mathbf{P}^n$.*

Proof. We have

$$\begin{aligned} \mu_j^{(m+n)} &= \mathsf{P}(X_{m+n} = j) = \sum_i \mathsf{P}(X_{m+n} = j \mid X_m = i)\,\mathsf{P}(X_m = i) \\ &= \sum_i p_{ij}(n)\,\mu_i^{(m)} = \sum_i \mu_i^{(m)}\, p_{ij}(n) = \bigl(\boldsymbol{\mu}^{(m)}\,\mathbf{P}_n\bigr)_j \end{aligned}$$

and the result follows from Theorem 5.9. □

Example 5.11. *Consider a three-state Markov chain with the transition matrix*

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 2/3 & 1/3 \\ 1/16 & 15/16 & 0 \end{pmatrix}.$$

☞ *Find a general formula for $p_{11}^{(n)}$.*

Solution. First we compute the eigenvalues of $\mathbf{P}$ by writing down the characteristic equation

$$0 = \det\bigl(x - \mathbf{P}\bigr) = \frac{1}{3}(x-1)(3x^2 + x + 1/16) = \frac{1}{3}(x-1)(3x+1/4)(x+1/4)$$

The eigenvalues are 1, $-1/4$, $-1/12$ and from this we deduce[41] that

$$p_{11}^{(n)} = a + b\Bigl(-\frac{1}{4}\Bigr)^n + c\Bigl(-\frac{1}{12}\Bigr)^n$$

with some constants a, b, c. The first few values of $p_{11}^{(n)}$ are easy to write down, so we get equations to solve for a, b, and c:

$$\begin{aligned} 1 &= p_{11}^{(0)} = a + b + c \\ 0 &= p_{11}^{(1)} = a - b/4 - c/12 \\ 0 &= p_{11}^{(2)} = a + b/16 + c/144\,. \end{aligned}$$

From this we get $a = 1/65$, $b = -26/65$, $c = 90/65$ so that

$$p_{11}^{(n)} = \frac{1}{65} - \frac{2}{5}\Bigl(-\frac{1}{4}\Bigr)^n + \frac{18}{13}\Bigl(-\frac{1}{12}\Bigr)^n\,.$$ □

5.2 Class structure

It is sometimes possible to break a Markov chain into smaller pieces, each of which is relatively easy to understand, and which together give an understanding of the whole. This is done by identifying the communicating classes of the chain.

✍ **Definition 5.12.** We say that state i *leads to* state j and write $i \to j$ if

$$\mathsf{P}_i(X_n = j \text{ for some } n \ge 0) \equiv \mathsf{P}\bigl(X_n = j \text{ for some } n \ge 0 \,|\, X_0 = i\bigr) > 0\,.$$

State i *communicates with* state j (write $i \leftrightarrow j$) if both $i \to j$ and $j \to i$.

Theorem 5.13. *For distinct states i and j the following are equivalent:*

a) $i \to j$;

b) $p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{n-1} i_n} > 0$ *for some states* $i_0 \equiv i$, i_1, i_2, …, i_{n-1}, $i_n \equiv j$;

c) $p_{ij}^{(n)} > 0$ *for some* $n \ge 0$.

[41] The justification comes from linear algebra: having distinct eigenvalues, $\mathbf{P}$ is diagonalizable, that is for some invertible matrix U we have

$$\mathbf{P} = U \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1/4 & 0 \\ 0 & 0 & -1/12 \end{pmatrix} U^{-1}$$

and hence

$$\mathbf{P}^n = U \begin{pmatrix} 1 & 0 & 0 \\ 0 & (-1/4)^n & 0 \\ 0 & 0 & (-1/12)^n \end{pmatrix} U^{-1}$$

which forces $p_{11}^{(n)} = a + b(-1/4)^n + c(-1/12)^n$ with some constants a, b, c.

☞ **Remark 5.13.1.** It is clear from b) that $i \leftrightarrow j$ and $j \leftrightarrow k$ imply $i \leftrightarrow k$. Also, $i \leftrightarrow i$ for any state i. So $\leftrightarrow$ satisfies the conditions for an equivalence relation on S and thus partitions S into communicating classes.

✍ **Definition 5.14.** We say that a class C is *closed* if

$$i \in C\,, \quad i \to j \qquad \Longrightarrow \qquad j \in C\,.$$

In other words, a closed class is one from which there is no escape. A state i is *absorbing* if $\{i\}$ is a closed class.

Exercise 5.15. Show that every transition matrix on a **finite** state space has at least one closed communicating class. Find an example of a transition matrix with no closed communicating classes.

✍ **Definition 5.16.** A Markov chain or its transition matrix P is called *irreducible* if its state space S forms a single communicating class.

Example 5.17. *Find the communicating classes associated with the stochastic matrix*

$$\mathsf{P} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 1/3 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Solution. The classes are $\{1,2,3\}$, $\{4\}$, $\{5,6\}$, with only $\{5,6\}$ being closed. (Draw the diagram!) □

Proof of Theorem 5.13. In view of the inequality

$$\max p_{ij}^{(n)} \le \mathsf{P}_i(X_n = j \text{ for some } n \ge 0) \le \sum_{n=0}^{\infty} p_{ij}^{(n)}$$

one immediately deduces a) $\Longleftrightarrow$ c). On the other hand,

$$p_{ij}^{(n)} = \sum_{i_1, i_2, \dots, i_{n-1}} p_{ii_1} p_{i_1 i_2} \dots p_{i_{n-1} j}$$

so that b) $\Longleftrightarrow$ c). □

✍ **Definition 5.18.** The *period* $d(i)$ of a state i is defined by

$$d(i) = \gcd\bigl\{n > 0 : p_{ii}^{(n)} > 0\bigr\}\,,$$

the greatest common divisor of the epochs at which return is possible (ie., $p_{ii}^{(n)} = 0$ unless n is a multiple of $d(i)$). We call i *periodic* if $d(i) > 1$ and *aperiodic* if $d(i) = 1$.

Lemma 5.19. *If states i and j are communicating, then i and j have the same period.*

Example 5.20. It is easy to see that both the simple random walk (Example 5.5) and the Ehrenfest chain (Example 5.6) have period 2. On the other hand, the birth and death process (Example 5.7) with all $p_k \equiv p_{k,k+1} > 0$, all $q_k \equiv p_{k,k-1} > 0$ and at least one $r_k \equiv p_{kk}$ positive is aperiodic (however, if all r_{kk} vanish, the birth and death chain has period 2).

Proof of Lemma 5.19. Since $i \leftrightarrow j$, there exist m, $n > 0$ such that $p_{ij}^{(m)} p_{ji}^{(n)} > 0$. By the Chapman-Kolmogorov equations,

$$p_{ii}^{(m+r+n)} \geq p_{ij}^{(m)} p_{jj}^{(r)} p_{ji}^{(n)} > 0$$

so that $d(i)$ divides $d(j)$. In a similar way one deduces that $d(j)$ divides $d(i)$. □

5.3 Hitting times and absorption probabilities

Consider the following problem.

Example 5.21. A man is saving up to buy a new car at a cost of N units of money. He starts with k ($0 < k < N$) units and tries to win the remainder by the following gamble with his bank manager. He tosses a coin repeatedly; if the coin comes up heads then the manager pays him one unit, but if it comes up tails then he pays the manager one unit. The man plays this game repeatedly until one of two events occurs: either he runs out of money and is bankrupted or he wins enough to buy the car. What is the probability that he is ultimately bankrupted?

The example above motivates the following definition.

✍ **Definition 5.22.** Let $(X_n)_{n\geq 0}$ be a Markov chain with transition matrix $\mathbf{P}$. The *hitting time* of a set $A \subset S$ is the random variable $H^A : \Omega \to \{0, 1, 2, \dots\} \cup \{\infty\}$ given by

$$H^A(\omega) \stackrel{\mathsf{def}}{=} \inf\Bigl\{n \geq 0 : X_n(\omega) \in A\Bigr\} \tag{5.3}$$

where we agree that the infimum over the empty set $\varnothing$ is ∞. The probability starting from i that $(X_n)_{n\geq 0}$ ever hits A is

$$h_i^A \stackrel{\mathsf{def}}{=} \mathsf{P}_i(H^A < \infty) \equiv \mathsf{P}\bigl(H^A < \infty \,|\, X_0 = i\bigr) . \tag{5.4}$$

When A is a closed class, h_i^A is called the *absorption probability*.

Remarkably, the hitting probabilities h_i^A can be calculated explicitly through certain linear equations associated with the transition matrix $\mathbf{P}$.

Example 5.23. *Consider the chain on $\{1, 2, 3, 4\}$ with the following transition matrix:*

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

Starting from 2, what is the probability of absorption in 4?

Solution. Put $h_i = \mathsf{P}_i(H^{\{4\}} < \infty)$. Clearly, $h_1 = 0$, $h_4 = 1$, and thanks to the Markov property, we have

$$h_2 = \frac{1}{2}\big(h_1 + h_3\big)\,, \qquad h_3 = \frac{1}{2}\big(h_2 + h_4\big)\,.$$

As a result, $h_2 = \frac{1}{2}h_3 = \frac{1}{2}(\frac{1}{2}h_2 + \frac{1}{2})$, that is $h_2 = 1/3$. □

✍ **Theorem 5.24.** *Fix $A \subset S$. The vector of hitting probabilities $h^A \equiv (h_i^A)_{i\in S}$ solves the following system of linear equations:*

$$\begin{cases} h_i^A = 1\,, & \text{for } i \in A, \\ h_i^A = \sum_{j\in S} p_{ij}\, h_j^A\,, & \text{for } i \in A^{\mathsf{c}}. \end{cases} \tag{5.5}$$

Example 5.23 (continued) Answer the same question by solving Eqns. (5.5)

Solution. Strictly speaking, the system (5.5) reads

$$h_4 = 1\,, \quad h_3 = \frac{1}{2}(h_4 + h_2)\,, \quad h_2 = \frac{1}{2}(h_3 + h_1)\,, \quad h_1 = h_1$$

so that

$$h_2 = \frac{1}{3} + \frac{2}{3}\,h_1\,, \qquad h_3 = \frac{2}{3} + \frac{1}{3}\,h_1\,.$$

The value of h_1 is not determined by the system, but the minimality condition requires $h_1 = 0$, so we recover $h_2 = 1/3$ as before. Of course, the extra boundary condition $h_1 = 0$ was obvious from the beginning so we built it into our system of equations and did not have to worry about minimal non-negative solutions. □

Proof of Theorem 5.24. We consider two cases separately. If $X_0 = i \in A$, then $H^A = 0$ so that $h_i^A = 1$.
On the other hand, if $X_0 = i \notin A$, then $H^A \geq 1$, so by the Markov property

$$\mathsf{P}_i(H^A < \infty \,|\, X_1 = j) = \mathsf{P}_j(H^A < \infty) = h_j^A\,;$$

the formula of total probability now implies

$$h_i^A = \mathsf{P}_i(H^A < \infty) = \sum_{j\in S} \mathsf{P}_i(H^A < \infty \,|\, X_1 = j)\, \mathsf{P}_i(X_1 = j) = \sum_{j\in S} p_{ij}\, h_j^A\,. \qquad \square$$

Remark 5.24.1. Actually, one can show[42] that $h^A = (h_i^A : i \in S)$ is the smallest non-negative solution to (5.5) in that if $x = (x_i : i \in S)$ is another solution to (5.5) with $x_i \geq 0$ for all $i \in S$, then $x_i \geq h_i^A$ for all $i \in S$. This property is especially useful if the state space S is infinite.

[42]Indeed, if $x = (x_i : i \in S) \geq 0$ solves (5.5), then $x_i = h_i^A = 1$ for all $i \in A$. On the other hand, if $i \in A^{\mathsf{c}}$, then the second line of (5.5) gives

$$x_i = \sum_{j\in A} p_{ij} + \sum_{j\notin A} p_{ij}x_j \equiv \mathsf{P}_i(H^A = 1) + \sum_{j\notin A} p_{ij}x_j\,.$$

By iterating this identity, we get

$$x_i = \mathsf{P}_i(H^A \leq n) + \sum_{j_1\notin A} \cdots \sum_{j_n\notin A} p_{ij_1}\, p_{j_1j_2} \ldots p_{j_{n-1}j_n} x_{j_n} \geq \mathsf{P}_i(H^A \leq n)$$

and thus $x_i \geq \lim_{n\to\infty} \mathsf{P}_i(H^A \leq n) = \mathsf{P}_i(H^A < \infty) = h_i^A$.

Example 5.25. (Gamblers' ruin) *Imagine the you enter a casino with a fortune of £i and gamble, £1 at a time, with probability p of doubling your stake and probability q of losing it. The resources of the casino are regarded as infinite, so there is no upper limit to your fortune. What is the probability that you leave broke?*
In other words, consider a Markov chain on $\{0,1,2,\dots\}$ with transition probabilities

$$p_{00} = 1\,, \qquad p_{k,k+1} = p\,, \qquad p_{k,k-1} = q \qquad (k \ge 1)$$

where $0 < p = 1 - q < 1$. Find $h_i = \mathsf{P}_i(H^{\{0\}} < \infty)$.

Solution. The vector $(h_i : i \ge 0)$ is the minimal non-negative solution to the system

$$h_0 = 1\,, \qquad h_i = p\,h_{i+1} + q\,h_{i-1}\,, \qquad (i \ge 1)\,.$$

As we shall see below, a general solution to this system is given by (check that the following expressions solve this system!)

$$h_i = \begin{cases} A + B\bigl(q/p\bigr)^i\,, & p \ne q \\ A + B\,i\,, & p = q\,, \end{cases}$$

where A and B are some real numbers. If $p \le q$, the restriction $0 \le h_i \le 1$ forces $B = 0$ so that $h_i \equiv A = 1$ for all $i \ge 0$. If $p > q$, we get (recall that $h_0 = A + B = 1$)

$$h_i = A + (1-A)\Bigl(\frac{q}{p}\Bigr)^i = \Bigl(\frac{q}{p}\Bigr)^i + A\Bigl(1 - \Bigl(\frac{q}{p}\Bigr)^i\Bigr)\,.$$

As for the non-negative solution we must have $A \ge 0$, the minimal solution now reads $h_i = (q/p)^i$. □

✍ It is often useful to know the expected time before absorption,

$$\begin{aligned} k_i^A &\stackrel{\mathsf{def}}{=} \mathsf{E}_i\bigl(H^A\bigr) \equiv \mathsf{E}\bigl(H^A \,|\, X_0 = i\bigr) \\ &= \sum_{n<\infty} n\,\mathsf{P}_i(H^A = n) + \infty \cdot \mathsf{P}_i(H^A = \infty)\,. \end{aligned} \tag{5.6}$$

Example 5.23 (continued) *Assuming that $X_0 = 2$, find the mean time until the chain is absorbed in states 1 or 4.*

Solution. Put $k_i = \mathsf{E}_i(H^{\{1,4\}})$. Clearly, $k_1 = k_4 = 0$, and thanks to the Markov property, we have

$$k_2 = 1 + \frac{1}{2}\bigl(k_1 + k_3\bigr)\,, \qquad k_3 = 1 + \frac{1}{2}\bigl(k_2 + k_4\bigr)$$

(The 1 appears in the formulae above because we count the time for the first step). As a result, $k_2 = 1 + \frac{1}{2}k_3 = 1 + \frac{1}{2}(1 + \frac{1}{2}k_2)$, that is $k_2 = 2$. □

✍ **Theorem 5.26.** *Fix $A \subset S$. The vector of mean hitting times $k^A \equiv (k_i^A, i \in S)$ is the minimal non-negative solution to the following system of linear equations:*

$$\begin{cases} k_i^A = 0\,, & \text{for } i \in A, \\ k_i^A = 1 + \sum_{j\in S} p_{ij}\,k_j^A\,, & \text{for } i \in A^{\mathsf{c}}. \end{cases} \tag{5.7}$$

Proof. First we check that k^A satisfies (5.7). If $i \in A$, then $H^A = 0$, so $k_i^A = 0$. On the other hand, if $X_0 = i \notin A$, then $H^A \geq 1$ and by the Markov property,

$$\mathsf{E}_i\big(H^A \,|\, X_1 = j\big) = 1 + \mathsf{E}_j(H^A)\,;$$

consequently, by the partition theorem for the expectations,

$$k_i^A = \mathsf{E}_i(H^A) = \sum_{j\in S} \mathsf{E}_i\big(H^A \,|\, X_1 = j\big)\, \mathsf{P}_i(X_1 = j) = 1 + \sum_{j\in S} p_{ij}\, k_j^A\,.$$

Let $y = (y_i : i \in S)$ be a non-negative solution to (5.7). We then have $k_i^A = y_i = 0$ for $i \in A$, and for $i \notin A$,

$$\begin{aligned} y_i &= 1 + \sum_{j\notin A} p_{ij}\, y_j = 1 + \sum_{j\notin A} p_{ij} \Big(1 + \sum_{k\notin A} p_{jk}\, y_k\Big) \\ &= \mathsf{P}_i(H^A \geq 1) + \mathsf{P}_i(H^A \geq 2) + \sum_{j\notin A}\sum_{k\notin A} p_{ij}\, p_{jk}\, y_k\,. \end{aligned}$$

By induction, $y_i \geq \mathsf{P}_i(H^A \geq 1) + \ldots + \mathsf{P}_i(H^A \geq n)$ for all $n \geq 1$ so that

$$y_i \geq \sum_{n=1}^{\infty} \mathsf{P}_i(H^A \geq n) \equiv \mathsf{E}_i\big(H^A\big) = k_i^A\,.$$ □

5.4 Recurrence and transience

Let X_n, $n \geq 0$, be a Markov chain with a discrete state space S.

✍ **Definition 5.27.** State i is called *recurrent* if, starting from i the chain eventually returns to i with probability 1, ie.,

$$\mathsf{P}\big(X_n = i \text{ for some } n \geq 1 \;\,|\, X_0 = i\big) = 1\,.$$

State i is called *transient* if this probability is smaller than 1.

It is convenient to introduce the so-called first passage times of the Markov chain X_n: if $j \in S$ is a fixed state, then the first passage time T_j to state j is

$$T_j = \inf\Big\{n \geq 1 : X_n = j\Big\}\,, \tag{5.8}$$

ie., the moment of the first future visit to j; of course, $T_j = +\infty$ if X_n never visits state j. Let

$$f_{ij}^{(n)} \stackrel{\text{def}}{=} \mathsf{P}_i\big(T_j = n\big) \equiv \mathsf{P}\big(X_1 \neq j, X_2 \neq j, \ldots, X_{n-1} \neq j, X_n = j \,|\, X_0 = i\big)$$

be the probability of the event "the first future visit to state j, starting from i, takes place at nth step". Then

$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)} \equiv \mathsf{P}_i\big(T_j < \infty\big)$$

is the probability that the chain ever visits j, starting from i. Of course, state j is recurrent iff

$$f_{jj} = \sum_{n=1}^{\infty} f_{jj}^{(n)} = \mathsf{P}_j\big(T_j < \infty\big) = 1\,. \tag{5.9}$$

In this case

$$\mathsf{P}_j\big(X_n = j \textit{ for some } n \geq 1\big) \equiv \mathsf{P}_j\big(X_n \textit{ returns to state } j \textit{ at least once}\big) = 1\,,$$

so that for every $m \geq 1$ we have[43]

$$\mathsf{P}_j\big(X_n \textit{ returns to state } j \textit{ at least } m \textit{ times}\big) = 1\,.$$

Consequently, with probability one the Markov chain X_n returns to recurrent state j infinitely many times.

On the other hand, if $f_{jj} < 1$, then the number of returns to state j is a geometric random variable with parameter $1 - f_{jj} > 0$, and thus is finite (and has a finite expectation) with probability one. As a result, for every state j,

$$\mathsf{P}_j\big(X_n \textit{ returns to state } j \textit{ infinitely many times}\big) \in \big\{0, 1\big\}\,,$$

depending on whether state j is transient or recurrent.

☞ **Remark 5.27.1.** Clearly, for $i \neq j$ we have

$$f_{ij} = \mathsf{P}_i\big(T_j < \infty\big) = \mathsf{P}_i(H^{\{j\}} < \infty) = h_i^{\{j\}}\,,$$

the probability that starting from i the Markov chain ever hits j.

Remark 5.27.2. By (5.9), state j is recurrent if and only if T_j is a proper random variable w.r.t. the probability distribution $\mathsf{P}_j(\,\cdot\,)$, ie., $\mathsf{P}(T_j < \infty) = 1$. A recurrent state j is positive recurrent if the expectation

$$\mathsf{E}_j\big(T_j\big) \equiv \mathsf{E}\big(T_j \mid X_0 = j\big) = \sum_{n=1}^{\infty} n f_{jj}^{(n)}$$

is finite; otherwise state j is null recurrent.

We now express recurrent/transient properties of states in terms of $p_{jj}^{(n)}$.

Lemma 5.28. *Let i, j be two states. Then for all $n \geq 1$,*

$$p_{ij}^{(n)} = \sum_{k=1}^{n} f_{ij}^{(k)}\, p_{jj}^{(n-k)}\,. \tag{5.10}$$

---

[43] indeed, if random variables X and Y are such that $\mathsf{P}(X < \infty) = 1$ and $\mathsf{P}(X < \infty) = 1$, then the sum $X + Y$ has the same property, $\mathsf{P}(X + Y < \infty) = 1$; if X is the moment of the first return, and Y is the time between the first and the second return to a fixed state, the condition $\mathsf{P}(Y < \infty \mid X = k) = 1$ implies that

$$\mathsf{P}(X + Y < \infty) = \sum_{k \geq 1} \mathsf{P}(k + Y < \infty \mid X = k)\mathsf{P}(X = k) = 1.$$

Of course, (5.10) is nothing else than the *first passage decomposition*: every trajectory leading from i to j in n steps, visits j for the first time in k steps $(1 \le k \le n)$ and then comes back to j in the remaining $n-k$ steps.

✍ **Corollary 5.29.** *The following dichotomy holds:*

a) *if* $\sum_n p_{jj}^{(n)} = \infty$, *then the state* j *is recurrent; in this case* $\sum_n p_{ij}^{(n)} = \infty$ *for all* i *such that* $f_{ij} > 0$.

b) *if* $\sum_n p_{jj}^{(n)} < \infty$, *then the state* j *is transient; in this case* $\sum_n p_{ij}^{(n)} < \infty$ *for all* i.

Remark 5.29.1. Notice that $\sum_n p_{jj}^{(n)}$ is just the expected number of visits to state j starting from j.

Example 5.30. *Let* $\big(X_n\big)_{n\ge0}$ *be a Markov chain in* S *with* $X_0 = k \in S$. *Consider the event* $A_n = \big\{\omega : X_n(\omega) = k\big\}$ *that the chain returns to state* k *after* n *steps. Clearly,* $\mathsf{P}(A_n) = p_{kk}^{(n)}$, *the* n*-step transition probability. By Corollary 5.29, state* k *is transient iff* $\sum_n p_{kk}^{(n)} < \infty$. *On the other hand, by the first Borel-Cantelli lemma, Lemma 1.6*a), *finiteness of the sum* $\sum_n \mathsf{P}(A_n) \equiv \sum_n p_{kk}^{(n)} < \infty$ *implies* $\mathsf{P}\big(A_n \text{ i.o.}\big) = 0$, *so that, with probability one, the chain visits state* k *only a finite number of times.*

☞ **Corollary 5.31.** *If* $j \in S$ *is transient, then* $p_{ij}^{(n)} \to 0$ *as* $n \to \infty$ *for all* $i \in S$.

Example 5.32. *Determine recurrent and transient states for a Markov chain on* $\{1,2,3\}$ *with the following transition matrix:*

$$\mathbf{P} = \begin{pmatrix} 1/3 & 2/3 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Solution. We obviously have $p_{11}^{(n)} = (1/3)^n$ with $\sum_n p_{11}^{(n)} < \infty$, so state 1 is transient. On the other hand, for $i \in \{2,3\}$, we have $p_{ii}^{(n)} = 1$ for even n and $p_{ii}^{(n)} = 0$ otherwise. As $\sum_n p_{ii}^{(n)} = \infty$ for $i \in \{2,3\}$, both states 2 and 3 are recurrent.
Alternatively, $f_{11} \equiv f_{11}^{(1)} = 1/3$, $f_{22} \equiv f_{22}^{(2)} = 1$, and $f_{33} \equiv f_{33}^{(2)} = 1$, so that state 1 is transient and states 2, 3 are recurrent. □

Proof of Lemma 5.28. The event $A_n = \{X_n = j\}$ can be decomposed into a disjoint union of "the first visit to j events"

$$B_k = \{X_l \ne j \text{ for } 1 \le l < k, X_k = j\}, \qquad 1 \le k \le n,$$

ie., $A_n = \cup_{k=1}^n B_k$, so that

$$\mathsf{P}_i(A_n) = \sum_{k=1}^n \mathsf{P}_i(A_n \cap B_k).$$

On the other hand, by the "tower property" and the Markov property

$$\begin{aligned}\mathsf{P}_i(A_n \cap B_k) &\equiv \mathsf{P}(A_n \cap B_k \mid X_0 = i) = \mathsf{P}(A_n \mid B_k, X_0 = i)\,\mathsf{P}(B_k \mid X_0 = i)\\ &= \mathsf{P}(A_n \mid X_k = j)\,\mathsf{P}(B_k \mid X_0 = i) \equiv p_{jj}^{(n-k)}\, f_{ij}^{(k)}\,.\end{aligned}$$

The result follows. □

Let $\mathcal{P}_{ij}(s)$ and $\mathcal{F}_{ij}(s)$ be the generating functions of $p_{ij}^{(n)}$ and $f_{ij}^{(n)}$,

$$\mathcal{P}_{ij}(s) \stackrel{\mathsf{def}}{=} \sum_n p_{ij}^{(n)}\, s^n\,, \qquad \mathcal{F}_{ij}(s) \stackrel{\mathsf{def}}{=} \sum_n f_{ij}^{(n)}\, s^n$$

with the convention that $p_{ij}^{(0)} = \delta_{ij}$, the Kronecker delta, and $f_{ij}^{(0)} = 0$ for all i and j. By Lemma 5.28, we obviously have

$$\mathcal{P}_{ij}(s) = \delta_{ij} + \mathcal{F}_{ij}(s)\,\mathcal{P}_{jj}(s)\,. \tag{5.11}$$

Notice that this equation allow us to derive $\mathcal{F}_{ij}(s)$, the generating function of the first passage time from i to j.

Proof of Corollary 5.29. To show that j is recurrent iff $\sum_n p_{jj}^{(n)} = \infty$, we first observe that (5.11) with $i = j$ is equivalent to $\mathcal{P}_{jj}(s) = \bigl(1 - \mathcal{F}_{jj}(s)\bigr)^{-1}$ for all $|s| < 1$ and thus, as $s \nearrow 1$,

$$\mathcal{P}_{jj}(s) \to \infty \qquad \Longleftrightarrow \qquad \mathcal{F}_{jj}(s) \to \mathcal{F}_{jj}(1) \equiv f_{jj} = 1\,.$$

It remains to observe that the Abel theorem implies $\mathcal{P}_{jj}(s) \to \sum_n p_{jj}^{(n)}$ as $s \nearrow 1$. The rest follows directly from (5.11) with $i \neq j$. □

Using Corollary 5.29, we can easily classify states of finite state Markov chains into transient and recurrent. Notice that transience and recurrence are class properties:

✍ **Lemma 5.33.** *Let C be a communicating class. Then* either *all states in C are transient* or *all are recurrent.*

Thus it is natural to speak of recurrent and transient classes.

Proof. Fix i, $j \in C$ and suppose that state i is transient, so that $\sum_{l=0}^{\infty} p_{ii}^{(l)} < \infty$. There exist k, $m \geq 0$ with $p_{ij}^{(k)} > 0$ and $p_{ji}^{(m)} > 0$, so that for all $l \geq 0$, by Chapman-Kolmogorov equations,

$$p_{ii}^{(k+l+m)} \geq p_{ij}^{(k)}\, p_{jj}^{(l)}\, p_{ji}^{(m)}$$

and thus

$$\sum_{l=0}^{\infty} p_{jj}^{(l)} \leq \frac{1}{p_{ij}^{(k)}\, p_{ji}^{(m)}} \sum_{l=0}^{\infty} p_{ii}^{(k+l+m)} \equiv \frac{1}{p_{ij}^{(k)}\, p_{ji}^{(m)}} \sum_{l=k+m}^{\infty} p_{ii}^{(l)} < \infty\,,$$

so that j is also transient by Corollary 5.29. □

☞ **Lemma 5.34.** *Every recurrent class is closed.*

Proof. Let C be a recurrent class which is not closed. Then there exist $i \in C$, $j \notin C$ and $m \geq 1$ such that the event $A_m \stackrel{\text{def}}{=} \{X_m = j\}$ satisfies $\mathsf{P}_i(A_m) > 0$. Denote

$$B \stackrel{\text{def}}{=} \big\{X_n = i \text{ for infinitely many } n\ \big\}\,.$$

By assumption, events A_m and B are incompatible, so that $\mathsf{P}_i(A_m \cap B) = 0$, and therefore $\mathsf{P}_i(B) \leq \mathsf{P}_i(A_m^{\mathsf{c}}) < 1$, ie., state i is not recurrent, and so neither is C. □

Remark 5.34.1. *In fact, one can prove that* Every finite closed class is recurrent.

Sketch of the proof. Let class C be closed and finite. Start X_n in C and consider the event $B_i = \{X_n = i \text{ for infinitely many } n\}$. By finiteness of C, for some $i \in C$ and all $j \in C$, we have $0 < \mathsf{P}_j(B_i) \equiv f_{ji}\mathsf{P}_i(B_i)$, and therefore $\mathsf{P}_i(B_i) > 0$. Since $\mathsf{P}_i(B_i) \in \{0, 1\}$, we deduce $\mathsf{P}_i(B_i) = 1$, i.e., state i (and the whole class C) is recurrent.

5.5 The strong Markov property

A convenient way of reformulating the Markov property (5.1) is that, conditional on the event $\{X_n = x\}$, the Markov chain $(X_m)_{m \geq n}$ has the same distribution as $(X_m)_{m \geq 0}$ with initial state $X_0 = x$; in other words, the *Markov chain* $(X_m)_{m \geq n}$ *starts afresh from the state* $X_n = x$.

For practical purposes it is often desirable to extend the validity of this property from *deterministic* times n to *random*[44] times T. In particular, the fact that the Markov chain starts afresh upon first hitting state j at a *random moment* T_j resulted in a very useful convolution property (5.10) in Lemma 5.28.
☞ Notice however that the *random* time $T^* = T_j - 1$ is not a good choice, as given $X_{T^*} = i \neq j$ the chain is *forced* to jump to $X_{T^*+1} = X_{T_j} = j$, thus discarding any other possible transition from state i.

✍ A random time T is called a *stopping time* for the Markov chain $\big(X_n\big)_{n \geq 0}$, if for any $n \geq 0$ the event $\big\{T \leq n\big\}$ is only determined by $\big(X_k\big)_{k \leq n}$ (and thus *does not depend on the future* evolution of the chain). Typical examples of stopping times include the hitting times H^A from (5.3) and the first passage times T_j from (5.8). ☞ Notice that the example $T^* = T_j - 1$ above, predetermines the first *future jump* of the Markov chain and thus is not a *stopping time*.

☞ **Lemma 5.35** (Strong Markov property)**.** *Let T be a stopping time for a Markov chain $(X_n)_{n \geq 0}$ with state space S. Then, given $\{T < \infty\}$ and $X_T = i \in S$, the process $(Y_n)_{n \geq 0}$ defined via $Y_n = X_{T+n}$ has the same distribution as $(X_n)_{n \geq 0}$ started from $X_0 = i$.*

One can verify[45] the strong Markov property by following the ideas in the proof of Lemma 5.28; namely, by partitioning Ω with the events $B_k = \{T = k\}$ and applying the formula of total probability. Since the conditions $T = k$ and $X_T = X_k = i$ are completely determined by the values $(X_m)_{m \leq k}$, the usual Markov property (5.1) can now be applied with the condition $\{X_k = i\}$ at the *deterministic* time $T = k$.

[44] recall the argument in Lemma 5.28 or Problem GF17

[45] an optional but instructive exercise!

5.6 Stationary distributions and the Ergodic theorem

For various practical reasons it is important to know how does a Markov chain behave after a long time n has elapsed. Eg., consider a Markov chain X_n on $S = \{1, 2\}$ with transition matrix

$$\mathbf{P} = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}, \quad 0 < a < 1, \quad 0 < b < 1,$$

and the initial distribution $\mu^{(0)} = (\mu_1^{(0)}, \mu_2^{(0)})$. Since for any $n \ge 0$,

$$\mathbf{P}^n = \frac{1}{a+b}\begin{pmatrix} b & a \\ b & a \end{pmatrix} + \frac{(1-a-b)^n}{a+b}\begin{pmatrix} a & -a \\ -b & b \end{pmatrix},$$

the distribution $\mu^{(n)}$ of X_n satisfies

$$\mu^{(n)} = \mu^{(0)}\,\mathbf{P}^n \to \boldsymbol{\pi}, \qquad \text{where} \qquad \boldsymbol{\pi} \stackrel{\mathsf{def}}{=} \Bigl(\frac{b}{a+b}, \frac{a}{a+b}\Bigr);$$

Notice that $\pi\,\mathbf{P} = \pi$, ie., the distribution π is 'invariant' for $\mathbf{P}$.

In general, existence of a limiting distribution for X_n as $n \to \infty$ is closely related to existence of the so-called 'stationary distributions'.

✍ **Definition 5.36.** The vector $\pi = (\pi_j : j \in S)$ is called a *stationary distribution* of a Markov chain on S with the transition matrix $\mathbf{P}$, if:

a) π is a distribution, ie., $\pi_j \ge 0$ for all $j \in S$, and $\sum_j \pi_j = 1$;

b) π is stationary, ie., $\pi = \pi\,\mathbf{P}$, which is to say that $\pi_j = \sum_i \pi_i p_{ij}$ for all $j \in S$.

Remark 5.36.1. Property b) implies that $\pi\,\mathbf{P}^n = \pi$ for all $n \ge 0$, that is if X_0 has distribution π then so does X_n, ie., the distribution of X_n does not change.

Example 5.37. *Find a stationary distribution for the Markov chain with the transition matrix*

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 2/3 & 1/3 \\ p & 1-p & 0 \end{pmatrix}.$$

Solution. Let $\pi = (\pi_1, \pi_2, \pi_3)$ be a probability vector. By a direct computation,

$$\pi\,\mathbf{P} = \Bigl(p\pi_3, \pi_1 + \frac{2}{3}\pi_2 + (1-p)\pi_3, \frac{1}{3}\pi_2\Bigr),$$

so that for π to be a stationary distribution, we need $\pi_3 = \frac{1}{3}\pi_2$, $\pi_1 = p\pi_3 = \frac{p}{3}\pi_2$, thus implying $1 = \sum_j \pi_j = \frac{4+p}{3}\pi_2$. As a result, $\pi = \bigl(p/(4+p), 3/(4+p), 1/(4+p)\bigr)$. Recall that in Example 5.11 we had $p = \frac{1}{16}$ so that $\pi_1 = \frac{1}{65}$, consistent with the limiting value of $p_{11}^{(n)}$ as $n \to \infty$. □

✍ **Lemma 5.38.** *Let S be finite. Assume for some $i \in S$ that $p_{ij}^{(n)} \to \pi_j$ as $n \to \infty$ for all $j \in S$. Then the vector $\pi = (\pi_j : j \in S)$ is a stationary distribution.*

Proof. Since S is finite, we obviously have

$$\sum_j \pi_j = \sum_j \lim_{n\to\infty} p_{ij}^{(n)} = \lim_{n\to\infty} \sum_j p_{ij}^{(n)} = 1\,,$$

$$\pi_j = \lim_{n\to\infty} p_{ij}^{(n+1)} = \lim_{n\to\infty} \sum_k p_{ik}^{(n)} p_{kj} = \sum_k \lim_{n\to\infty} p_{ik}^{(n)} p_{kj} = \sum_k \pi_k p_{kj}\,. \qquad \square$$

✍ **Theorem 5.39** (Convergence to equilibrium). *Let X_n, $n \ge 0$, be an irreducible aperiodic Markov chain with transition matrix $\mathbf{P}$ on a finite state space S. Then there exists a unique probability distribution π such that for all i, $j \in S$*

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

In particular, π is stationary for $\mathbf{P}$, and for every initial distribution

$$\mathsf{P}(X_n = j) \to \pi_j \qquad \text{as } n \to \infty.$$

The following two examples illustrate importance of our assumptions.

Example 5.40. *For a periodic Markov chain on $S = \{1,2\}$ with transition probabilities $p_{12} = p_{21} = 1$ and $p_{11} = p_{22} = 0$ we obviously have*

$$\mathbf{P}^{2k+1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\,, \qquad \mathbf{P}^{2k} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

for all $k \ge 0$, so that there is no convergence.

Example 5.41. *For a reducible Markov chain on $S = \{1,2,3\}$ with transition matrix (with strictly positive a, b, c such that $a+b+c=1$)*

$$\mathbf{P} = \begin{pmatrix} a & b & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

every stationary distribution π satisfies the equations

$$\pi_1 = a\,\pi_1 \qquad \pi_2 = b\,\pi_1 + \pi_2\,, \qquad \pi_3 = c\,\pi_1 + \pi_3\,,$$

ie., $\pi_1 = 0$, $\pi_2 = \pi_2$, $\pi_3 = \pi_3$, so that there is a whole family of solutions: every vector $\mu_\rho \stackrel{\mathsf{def}}{=} (0, \rho, 1-\rho)$ with $0 \le \rho \le 1$ is a stationary distribution for $\mathbf{P}$.

For a Markov chain X_n, let T_j be its first passage time to state j.

☞ **Lemma 5.42.** *If the transition matrix $\mathbf{P}$ of a Markov chain X_n is γ-positive, $\min_{j,k} q_{jk} \ge \gamma > 0$, then there exist positive constants C and α such that*

$$\mathsf{P}(T_j > n) \le C \exp\{-\alpha n\} \qquad \textit{for all } n \ge 0\,.$$

Proof. By the positivity assumption,

$$\mathsf{P}(T_j > n) \le \mathsf{P}\big(X_1 \ne j, X_2 \ne j, \dots, X_n \ne j\big) \le (1-\gamma)^n\,. \qquad \square$$

By Lemma 5.42, the exponential moments $\mathsf{E}\big(e^{cT_j}\big)$ exist for all $c>0$ small enough, in particular, all polynomial moments $\mathsf{E}\big((T_j)^p\big)$ with $p>0$ are finite.

✍ **Lemma 5.43.** *Let X_n be an irreducible recurrent Markov chain with stationary distribution π. Then for all states j, the expected return time $\mathsf{E}_j(T_j)$ satisfies*

$$\pi_j\,\mathsf{E}_j(T_j)=1\,.$$

If $V_k(n)=\sum_{l=1}^{n-1}\mathbb{1}_{\{X_l=k\}}$ denotes the number of visits[46] to k before time n, then $V_k(n)/n$ is the proportion of time before n spent in state k.

The following consequence of Theorem 5.39 and Lemma 5.43 gives the long-run proportion of time spent by a Markov chain in each state. It is often referred to as the Ergodic theorem.

✍ **Theorem 5.44.** *If $(X_n)_{n\ge0}$ is an irreducible Markov chain, then*

$$\mathsf{P}\Big(\frac{V_k(n)}{n}\to\frac{1}{\mathsf{E}_k(T_k)}\text{ as }n\to\infty\Big)=1\,,$$

where $\mathsf{E}_k(T_k)$ is the expected return time to state k.

Moreover, if $\mathsf{E}_k(T_k)<\infty$, then for any bounded function $f:S\to\mathbb{R}$ we have

$$\mathsf{P}\Big(\frac{1}{n}\sum_{k=0}^{n-1}f(X_k)\to\bar f\text{ as }n\to\infty\Big)=1\,,$$

where $\bar f\stackrel{\mathsf{def}}{=}\sum_{k\in S}\pi_k\,f(k)$ and $\boldsymbol{\pi}=(\pi_k:k\in S)$ is the unique stationary distribution.

In other words, the second claim implies that if a real function f on S is bounded, then the average of f along every typical trajectory of a "nice" Markov chain $(X_k)_{k\ge0}$ is close to the space average of f w.r.t. the stationary distribution of this Markov chain.

Proof of Lemma 5.43. Let X_0 have the distribution $\boldsymbol{\pi}$ so that $\mathsf{P}(X_0=j)=\pi_j$. Then,

$$\pi_j\,\mathsf{E}_j(T_j)\equiv\sum_{n=1}^{\infty}\mathsf{P}\big(T_j\ge n\mid X_0=j\big)\,\mathsf{P}(X_0=j)=\sum_{n=1}^{\infty}\mathsf{P}\big(T_j\ge n,X_0=j\big)$$

Denote $a_n\stackrel{\mathsf{def}}{=}\mathsf{P}(X_m\ne j\text{ for }0\le m\le n)$. We then have

$$\mathsf{P}\big(T_j\ge1,X_0=j\big)\equiv\mathsf{P}(X_0=j)=1-a_0$$

and for $n\ge2$

$$\begin{aligned}\mathsf{P}\big(T_j\ge n,X_0=j\big)&=\mathsf{P}\big(X_0=j,X_m\ne j\text{ for }1\le m\le n-1\big)\\&=\mathsf{P}\big(X_m\ne j\text{ for }1\le m\le n-1\big)\\&\quad-\mathsf{P}\big(X_m\ne j\text{ for }0\le m\le n-1\big)=a_{n-2}-a_{n-1}\,.\end{aligned}$$

[46] Notice that $\mathsf{E}_j\big(V_k(n)\big)\equiv\mathsf{E}\big(V_k(n)\mid X_0=j\big)=\sum_{l=1}^{n-1}p_{jk}^{(l)}$.

Consequently,

$$\pi_j \,\mathsf{E}_j(T_j) = \mathsf{P}(X_0 = j) + \sum_{n=2}^{\infty}(a_{n-2} - a_{n-1})$$
$$= \mathsf{P}(X_0 = j) + \mathsf{P}(X_0 \neq j) - \lim_{n\to\infty} a_n = 1 - \lim_{n\to\infty} a_n = 1\,,$$

since

$$\lim_n a_n = \mathsf{P}(X_m \neq j \text{ for all } m) = \mathsf{P}(T_j = \infty) = 0\,,$$

by the recurrence assumption. □

Remark 5.43.1. Observe that the argument above shows that even in the countable state space the stationary distribution is unique, provided it exists.

5.6.1 Proof of Theorem 5.39

The material of this section is optional and thus for fun, not for exam!

Our argument shall rely upon the following important fact: for $m \geq 1$, let

$$d(\mathbf{x}, \mathbf{y}) \equiv d_m(\mathbf{x}, \mathbf{y}) \stackrel{\mathsf{def}}{=} \frac{1}{2}\sum_{j=1}^{m} |x_j - y_j| \tag{5.12}$$

denote the distance between $\mathbf{x} = (x_1, \dots, x_m) \in \mathbb{R}^m$ and $\mathbf{y} = (y_1, \dots, y_m) \in \mathbb{R}^m$.
☞ Recall that a sequence $\mathbf{x}^{(n)} \in \mathbb{R}^m$ is called Cauchy if for every $\varepsilon > 0$ there is $N > 0$ such that the inequality

$$d\big(\mathbf{x}^{(n)}, \mathbf{x}^{(n+k)}\big) \leq \varepsilon \tag{5.13}$$

holds for all $n > N$ and $k \geq 0$. The fundamental property of Cauchy sequences in $\mathbb{R}^m$ reads:

☞ **Property 5.45.** *Every Cauchy sequence in $\mathbb{R}^m$ has a unique limit.*

Recall that a matrix $\mathbf{Q} = (q_{jk})_{j,k=1}^m$ is called γ-positive, if

$$\min_{j,k} q_{jk} \geq \gamma > 0\,. \tag{5.14}$$

In what follows we shall rely upon the following property of γ-positive matrices.

☞ **Lemma 5.46.** *a) If $\mathbf{Q} = (q_{jk})_{j,k=1}^m$ is a stochastic matrix and μ is a probability vector in $\mathbb{R}^m$, then $\mu\,\mathbf{Q}$ is a probability distribution in $\mathbb{R}^m$.*

b) Let $\mathbf{Q} = (q_{jk})_{j,k=1}^m$ be a γ-positive $m \times m$ stochastic matrix. Then for all probability vectors μ', μ'' in $\mathbb{R}^m$ we have

$$d\big(\mu'\mathbf{Q}, \mu''\mathbf{Q}\big) \leq (1-\gamma)\, d(\mu', \mu'')\,. \tag{5.15}$$

The estimate (5.15) means that the mapping $\mu \mapsto \mu\mathbf{Q}$ from the "probability simplex" $\big\{\boldsymbol{x} \in \mathbb{R}^m : x_j \geq 0, \sum_j x_j = 1\big\}$ to itself is a *contraction*; therefore, it has a unique fixed point $\boldsymbol{x} = \boldsymbol{x}\mathbf{Q}$ in this simplex. Results of this type are central to various areas of mathematics.

We postpone the proof of Lemma 5.46 and verify a particular case Theorem 5.39 first.

Proof of Theorem 5.39 for γ-positive matrices. Let $\mu^{(0)}$ be a fixed initial distribution, $\mathbf{P}$ be a γ-positive stochastic matrix, and let $\mu^{(n)} = \mu^{(0)}\,\mathbf{P}$ denote the distribution at time n. By (5.15),

$$d\big(\mu^{(n)},\mu^{(n+k)}\big) = d\big(\mu^{(0)}\,\mathbf{P}^n,\mu^{(0)}\,\mathbf{P}^{n+k}\big)$$
$$\le (1-\gamma)^n\,d\big(\mu^{(0)},\mu^{(0)}\,\mathbf{P}^k\big) \le (1-\gamma)^n\,,$$

so that $\mu^{(n)}$ is a Cauchy sequence in $\mathbb{R}^m$. As a result, the limit

$$\pi \stackrel{\mathsf{def}}{=} \lim_{n\to\infty}\mu^{(n)} = \lim_{n\to\infty}\mu^{(0)}\,\mathbf{P}^n$$

exists, and (by Lemma 5.46a)) is a stationary probability distribution for $\mathbf{P}$:

$$\pi\,\mathbf{P} = \big(\lim_{n\to\infty}\mu^{(n)}\big)\mathbf{P} = \lim_{n\to\infty}\big(\mu^{(0)}\,\mathbf{P}^n\big)\mathbf{P} = \lim_{n\to\infty}\mu^{(0)}\,\mathbf{P}^{n+1} = \pi\,.$$

Moreover, this limit is unique: indeed, assuming the contrary, let π' and π'' be two distinct stationary probability distributions for $\mathbf{P}$; then, by (5.15),

$$0 < d(\pi',\pi'') \equiv d(\pi'\mathbf{P},\pi''\mathbf{P}) \le (1-\gamma)\,d(\pi',\pi'') < d(\pi',\pi'')\,;$$

this contradiction implies that $d(\pi',\pi'')=0$, ie., $\pi'=\pi''$.

We finally observe that the limit $\pi = \lim_{n\to\infty}\mu^{(0)}\mathbf{P}^n$ does not depend on the initial distribution $\mu^{(0)}$; therefore, for the initial distribution

$$\mu^{(0)}_k = \begin{cases} 1, & k=i\,,\\ 0, & k\ne i\,,\end{cases}$$

we have $p^{(n)}_{ij} \equiv \big(\mu^{(0)}\mathbf{P}\big)_j \to \pi_j$ for all $j=1,2,\dots,m$. The proof is finished. □

☞ **Corollary 5.47.** *If $\mathbf{P}$ is a γ-positive transition matrix with stationary distribution π, then for any initial distribution $\mu^{(0)}$,*

$$d\big(\mu^{(0)}\mathbf{P}^n,\pi\big) \le (1-\gamma)^n\,,$$

ie., the convergence in Theorem 5.39 is exponentially fast.

Proof of Lemma 5.46. a). If μ is a probability vector and $\mathbf{Q}$ is a stochastic matrix, then $\mu\mathbf{Q}\ge 0$ and

$$\sum_j(\mu\mathbf{Q})_j = \sum_j\Big(\sum_i\mu_i q_{ij}\Big) = \sum_i\mu_i\Big(\sum_j q_{ij}\Big) = \sum_i\mu_i = 1\,.$$

b). For two probability vectors μ' and μ'' in $\mathbb{R}^m$, denote

$$A_0 \stackrel{\mathsf{def}}{=} \big\{j:\mu'-\mu''>0\big\}\,,\qquad A_1 \stackrel{\mathsf{def}}{=} \big\{j:\mu'\mathbf{Q}-\mu''\mathbf{Q}>0\big\}\,;$$

clearly, both A_0 and A_1 are proper subsets of $\{1,2,\dots,m\}$ so that $\sum_{j\in A_1}q_{ij}\le 1-\gamma$ for all $i=1,\dots,m$. Moreover, the definition (5.12) can be rewritten as

$$d(\mu',\mu'') \equiv \frac{1}{2}\sum_{j=1}^m|x_j-y_j| = \sum_{j\in A_0}(\mu'_j-\mu''_j)$$

and thus we get

$$d(\mu'\mathbf{Q},\mu''\mathbf{Q}) = \sum_{j\in A_1}\sum_i(\mu_i'-\mu_i'')q_{ij} \le \sum_{j\in A_1}\sum_{i\in A_0}(\mu_i'-\mu_i'')q_{ij}$$
$$\le (1-\gamma)\sum_{i\in A_0}(\mu_i'-\mu_i'') \equiv (1-\gamma)\,d(\mu',\mu'')\,.$$

This finishes the proof. □

☞ **Exercise 5.48.** Let $\mathbf{P} = \bigl(p_{ij}\bigr)_{i,j=1}^m$ be an irreducible aperiodic stochastic matrix. Show that there exists an integer $l \ge 1$ such that $\mathbf{P}^l$ is γ-positive with some $\gamma \in (0,1)$.
Hint: show first that if state k is aperiodic, then the set $A_k \stackrel{\mathsf{def}}{=} \{n > 0 : p_{kk}^{(n)} > 0\}$ contains all large enough natural numbers, ie., for some natural $n_0 \ge 0$, A_k contains $\{n_0, n_0+1, \dots\}$.

5.7 Reversibility and detailed balance condition

The symmetric version of the Markov property says that given the present value of the process, its past and future are independent, and thus suggests looking at Markov chains with time running backwards. It is easy to see, that if we want complete time-symmetry, the Markov chain must begin in equilibrium. In fact, Exercise 5.53 below shows that a Markov chain in equilibrium, run backwards, is again a Markov chain; the transition matrix may however be different.

✍ **Definition 5.49.** A stochastic matrix $\mathbf{P}$ and a measure λ are said to be in *detailed balance* if

$$\lambda_i p_{ij} = \lambda_j p_{ji} \qquad \text{for all } i,\ j. \tag{5.16}$$

Equations (5.16) are called the *detailed balance equations*.

If a Markov chain has initial distribution $\boldsymbol{\lambda}$ and transition matrix $\mathbf{P}$, which satisfy DBE, then (5.16) implies that the amount of mass $\lambda_i p_{ij}$ flowing from state i to state j concides with that of $\lambda_j p_{ji}$ flowing in the opposite direction. In particular, the mass distribution $\boldsymbol{\lambda}$ does not change with time:

☞ **Lemma 5.50.** *If $\mathbf{P}$ and $\boldsymbol{\lambda}$ are in detailed balance, then $\boldsymbol{\lambda}$ is invariant for $\mathbf{P}$, ie., $\boldsymbol{\lambda} = \boldsymbol{\lambda}\mathbf{P}$.*

Proof. We have $(\boldsymbol{\lambda}\mathbf{P})_j = \sum_k \lambda_k p_{kj} = \sum_k \lambda_j p_{jk} = \lambda_j$. □

✍ **Definition 5.51.** Let $(X_n)_{n\ge0}$ be an irreducible Markov chain with state space S and transition matrix $\mathbf{P}$. A probability measure $\boldsymbol{\pi}$ on S is said to be *reversible* for the chain (or for the matrix $\mathbf{P}$) if $\boldsymbol{\pi}$ and $\mathbf{P}$ are in detailed balance, ie.,

$$\pi_i p_{ij} = \pi_j p_{ji} \qquad \text{for all } i,\ j \in S.$$

An irreducible Markov chain is said to be *reversible* if it has a reversible distribution.

Example 5.52. A Markov chain with transition matrix

$$\begin{pmatrix} 0 & p & q \\ q & 0 & p \\ p & q & 0 \end{pmatrix}, \qquad 0 < p = 1 - q < 1,$$

has stationary distribution $\boldsymbol{\pi} = (1/3, 1/3, 1/3)$. For the latter to be reversible for this Markov chain, we need $\pi_1 p_{12} = \pi_2 p_{21}$, i.e., $p = q$. If $p = q = \frac{1}{2}$, then DBE (5.16) hold for all pairs of states, ie., the chain is reversible. Otherwise the chain is not reversible.

Exercise 5.53. Let $\mathbf{P}$ be irreducible and have an invariant distribution $\boldsymbol{\pi}$. Suppose that $(X_n)_{0 \le n \le N}$ is a Markov chain with transition matrix $\mathbf{P}$ and the initial distribution $\boldsymbol{\pi}$, and set $Y_n \stackrel{\text{def}}{=} X_{N-n}$. Show that $(Y_n)_{0 \le n \le N}$ is a Markov chain with the same initial distribution $\boldsymbol{\pi}$ and with transition matrix $\widehat{\mathbf{P}} = (\hat{p}_{ij})$ given by $\pi_j \hat{p}_{ji} = \pi_i p_{ij}$ for all i, j. The chain $(Y_n)_{0 \le n \le N}$ is called the time-reversal of $(X_n)_{0 \le n \le N}$.

Exercise 5.54. Find the transition matrix $\widehat{\mathbf{P}}$ of the time-reversal for the Markov chain from Example 5.52.

☞ **Example 5.55.** [Random walk on a graph] A graph G is a countable collection of states, usually called vertices, some of which are joined by edges. The valency v_j of vertex j is the number of edges at j, and we assume that every vertex in G has finite valency. The random walk on G is a Markov chain with transition probabilities

$$p_{jk} = \begin{cases} 1/v_j\,, & \text{if } (j,k) \text{ is an edge} \\ 0\,, & \text{otherwise.} \end{cases}$$

We assume that G is connected, so that $\mathbf{P}$ is irreducible. It is easy to see that $\mathbf{P}$ is in detailed balance with $\boldsymbol{v} = (v_j : j \in G)$. As a result, if the total valency $V = \sum_j v_j$ is finite, then $\boldsymbol{\pi} \stackrel{\text{def}}{=} \boldsymbol{v}/V$ is a stationary distribution and $\mathbf{P}$ is reversible.

5.8 Some applications of Markov chains

The material of this section is optional and thus for fun, not for exam!

The ideas behind Example 5.55 help to analyse many interesting Markov chains on graphs. Indeed, this approach is central for many algorithmic applications, including Markov chain Monte Carlo (MCMC).

Example 5.56. (Independent sets) Let $G = (V, E)$ be a graph with vertices in $V = \{v_1, \dots, v_k\}$ and edges in $E = \{e_1, \dots, e_\ell\}$. A collection of vertices $A \subset V$ is called an *independent set* in G, if no two vertices in A are adjacent in G. The following Markov chain randomly selects independent sets in $S_G \stackrel{\text{def}}{=} \{$ *all independent sets in* G $\}$:

Let $A \in S_G$ be given (eg., $A = \varnothing$).

1. Pick $v \in V$ uniformly at random;

2. Flip a fair coin;
3. If the result is 'heads' and no neighbour of v is in A, add v to A: $A \mapsto A \cup \{v\}$; otherwise, remove v from A: $A \mapsto A \setminus \{v\}$.

This irreducible and aperiodic Markov chain jumps between independent sets which differ in at most one vertex. Moreover, if $A \neq A'$ are two such independent sets, then $p_{AA'} = 1/(2k) = p_{A'A}$ and thus the detailed balance equations imply that the unique stationary distribution for this chain is uniform in S_G. By Theorem 5.39, after sufficiently many steps the distribution of this chain will be almost uniform in S_G.

Notice that the cardinality $|S_G|$ of S_G is enormous: indeed, if G is an $m \times m$ subset in $\mathbb{Z}^2$, then $k = m^2$ and a simple chessboard estimate gives $|G_S| \geq 2^{m^2/2} \geq 10^{3m^2/20}$; eg., for $m = 10$ this lower bound gives 10^{15}.

Example 5.57. (Graph colouring) Let $G = (V, E)$ be a graph and $q \geq 2$ be an integer. A *q-colouring* of G is an assignment

$$V \ni v \mapsto \xi_v \in S \stackrel{\text{def}}{=} \{1, 2, \dots, q\}$$

such that *if v_1 and v_2 are adjacent in G, then $\xi_{v_1} \neq \xi_{v_2}$*. The following Markov chain in S^V randomly selects q-colourings of G:

Let a colouring $C \in S^V$ be given.

1. Pick $v \in V$ uniformly at random;
2. Re-colour v in any admissible colour (ie., not in use by any of the neighbours of v) uniformly at random.

It is easy to see that for q large enough, this Markov chain is irreducible, aperiodic, and thus has a unique stationary distribution. Consequently, after sufficiently many steps, the distribution of this Markov chain will be almost uniform in the collection of admissible colourings of G. The smallest possible value of q for which the theory works depends on the structure of G; in many cases this optimal value is unknown.

☞ **Example 5.58. (Gibbs sampler)** More generally, if one needs to sample a random element from a set S according to a probability distribution $\boldsymbol{\pi}$, one attempts to construct an irreducible aperiodic Markov chain in S, which is reversible w.r.t. $\boldsymbol{\pi}$. Many such examples can be found in the literature.

☞ For various applications, it is essential to construct Markov chains with fast convergence to equilibrium. This is an area of active research with a lot of interesting open problems.