Applied Stochastic Processes

Jochen Geiger

(last update: July 18, 2007)

Contents

1 Discrete Markov chains

Markov processes form an important class of random processes with many applications in areas like physics, biology, computer science or finance. The characteristic property of a Markov process is its lack of memory, that is, the decision where to go next may (and typically does) depend on the current state of the process but not on how it got there. If the process can take only countably many different values then it is referred to as a *Markov chain*.

1.1 Basic properties and examples

A stochastic process $X = (X_t)_{t\in T}$ is a random variable which takes values in some *path space* $\mathcal{S}^T := \{x = (x_t)_{t\in T} : T \to \mathcal{S}\}$. Here, the space of possible outcomes $\mathcal{S}$ is some discrete (i.e., finite or countably infinite) *state space* and X_t is the state at time t (with values in $\mathcal{S}$). In this chapter, we will assume that time is discrete, i.e., we take the index set T to be the non-negative integers $\mathbb{N}_0 := \{0, 1, 2, \ldots\}$.

The distribution of X can be specified by describing the dynamics of the process, i.e., how to start and how to proceed. By the multiplication rule, we have

$$\begin{aligned}
&\mathbb{P}\{(X_0, X_1, \ldots, X_n) = (x_0, x_1, \ldots, x_n)\} \\
&\quad= \mathbb{P}\{X_0 = x_0\}\, \mathbb{P}\{X_1 = x_1 \,|\, X_0 = x_0\}\, \mathbb{P}\{X_2 = x_2 \,|X_0 = x_0, X_1 = x_1\} \cdots \\
&\qquad\qquad\qquad \cdots\ \mathbb{P}\{X_n = x_n \,|\, X_0 = x_0, \ldots, X_{n-1} = x_{n-1}\} \\
&\quad=: p_0(x_0)\, p_1(x_0, x_1) \cdots p_n(x_0, \ldots, x_n\}. \qquad (1.1)
\end{aligned}$$

The left-hand side of equation (1.1) can be written as:

$$\mathbb{P}\{X \in B_{x_0,\ldots,x_n}\}, \quad \text{where } B_{x_0,\ldots,x_n} = \{x_0\} \times \{x_1\} \times \cdots \times \{x_n\} \times \mathcal{S}^{\{n+1,n+2,\ldots\}}. \qquad (1.2)$$

Note that the functions p_j, $j \geq 0$ have to satisfy the following conditions.

1. $p_j(x_0, \ldots, x_j) \geq 0$ for all $j \geq 0$, $x_0, \ldots, x_j \in \mathcal{S}$;
2. $\sum\limits_{x_j\in\mathcal{S}} p_j(x_0, \ldots, x_j) = 1$ for all $j \geq 0$, $x_0, \ldots, x_{j-1} \in \mathcal{S}$.

Remark. The measures in (1.1) uniquely extend to a probability measure on $(\mathcal{S}^{\mathbb{N}_0}, \mathcal{B})$, where $\mathcal{B}$ is the σ-algebra generated by all sets of the form (1.2) (see Theorem 3.1 in [1]).

In general, the p_j may depend on the entire collection $x_0, \ldots, x_j$. However, little can be said about interesting properties of a stochastic process in this generality. This is quite different for so-called Markovian stochastic dynamics, where the p_j depend on x_{j-1} and x_j only.

Definition 1.1 *Let $\mathcal{S}$ be a countable space and $P = (P_{xy})_{x,y\in\mathcal{S}}$ a stochastic matrix (i.e., $P_{xy} \geq 0$ and $\sum_{y\in\mathcal{S}} P_{xy} = 1$, $\forall\, x \in \mathcal{S}$). A sequence of $\mathcal{S}$-valued random variables (r.v.'s) $X_0, X_1, \ldots$ is called a* Markov chain *(MC) with* state space $\mathcal{S}$ *and* transition matrix P*, if*

$$\mathbb{P}\{X_{n+1} = y \,|\, X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = x\} = P_{xy} \qquad (1.3)$$

holds for every $n \in \mathbb{N}_0$ *and* $x_0, \ldots, x_{n-1}, x, y \in \mathcal{S}$ *(provided that the conditional probability is well defined).*

Remarks.

- To be precise, the process $(X_n)_{n\geq 0}$ defined above is a *time-homogeneous* Markov chain (in general, the matrix P in (1.3) may depend on n).
- The dynamics (described by P) will be considered as fixed, but we will vary the initial distribution $\mu := \mathbb{P}\{X_0 \in \cdot\,\}$. It is standard notation to add the initial distribution as a subscript:

$$\begin{aligned} \mathbb{P}_\mu\{X_0 = x_0, \ldots, X_n = x_n\} &:= \mu(x_0)\,\mathbb{P}\{X_1 = x_1, \ldots, X_n = x_n | X_0 = x_0\} \\ &= \mu(x_0)\,P_{x_0x_1}\cdots P_{x_{n-1}x_n}. \end{aligned} \tag{1.4}$$

  If $\mu = \delta_z$, then we write $\mathbb{P}_z := \mathbb{P}_{\delta_z}$ for short.

The following proposition formalizes the Markov property, which says that the future and the past are conditionally independent given the present state of the chain.

Proposition 1.2 (Markov property) *Let* $(X_n)_{n\geq 0}$ *be a (time-homogeneous) Markov chain with state space* $\mathcal{S}$ *and initial distribution* μ*, then*

$$\mathbb{P}_\mu\{(X_m, \ldots, X_{m+n}) \in B \,|\, X_m = x,\, (X_0, \ldots, X_{m-1}) \in B'\} \;=\; \mathbb{P}_x\{(X_0, \ldots, X_n) \in B\}$$

holds for every $m, n \in \mathbb{N}_0$*,* $x \in \mathcal{S}$*,* $B \subset \mathcal{S}^{n+1}$ *and* $B' \subset \mathcal{S}^m$*.*

In other words, conditional on the event $\{X_m = x\}$ the process $(X_{m+n})_{n\geq 0}$ is a Markov chain started at x, independent of $(X_0, \ldots, X_{m-1})$.

Proof. By (1.4), we have for any $x_0, \ldots, x_{m-1}, x, y_0, \ldots, y_n$

$$\begin{aligned} &\mathbb{P}_\mu\{X_0 = x_0, \ldots, X_{m-1} = x_{m-1}, X_m = x, X_m = y_0, X_{m+1} = y_1, \ldots, X_{m+n} = y_n\} \\ &\quad= \delta_{xy_0}\,\mu(x_0)P_{x_0x_1}\cdots P_{x_{m-1}x}\,P_{y_0y_1}\cdots P_{y_{n-1}y_n} \\ &\quad= \mathbb{P}_\mu\{X_0 = x_0, \ldots, X_m = x\}\,\mathbb{P}_x\{X_0 = y_0, \ldots, X_n = y_n\}. \end{aligned} \tag{1.5}$$

Summation over all $(x_0, \ldots, x_{m-1}) \in B'$ and $(y_0, \ldots, y_n) \in B$ gives

$$\begin{aligned} &\mathbb{P}_\mu\{(X_0, \ldots, X_{m-1}) \in B',\, X_m = x, (X_m, \ldots, X_{m+n}) \in B\} \\ &\quad= \mathbb{P}_\mu\{(X_0, \ldots, X_{m-1}) \in B',\, X_m = x\}\,\mathbb{P}_x\{(X_0, \ldots, X_n) \in B\}. \end{aligned} \tag{1.6}$$

Dividing either side by $\mathbb{P}_\mu\{(X_0, \ldots, X_{m-1}) \in B',\, X_m = x\}$ gives the assertion of the proposition.

A simple consequence of the Markov property is the following formula for the n-step transition probabilities of the Markov chain $(X_n)_{n\geq 0}$.

Lemma 1.3 (n-step transition probabilities) *Let* $(X_n)_{n\geq 0}$ *be a Markov chain with state space* $\mathcal{S}$ *and transition matrix* P*. Then, for every* $x, y \in \mathcal{S}$ *and every* $n \in \mathbb{N}_0$

$$\mathbb{P}_x\{X_n = y\} = P^n_{xy}$$

holds, where

$$P^n = (P^n_{xy})_{x,y\in\mathcal{S}} = \underbrace{P\cdots P}_{n\ times},\ \ n \geq 1$$

and $P^0 := Id$*.*

Proof. By induction. For $n = 0$ we have

$$\mathbb{P}_x\{X_0 = y\} = \delta_{xy} = Id_{xy} = P^0_{xy}.$$

To get from n to $n+1$ note that by means of the law of total probability we have

$$\begin{aligned}\mathbb{P}_x\{X_{n+1} = y\} &= \sum_{z\in\mathcal{S}} \mathbb{P}_x\{X_n = z\}\,\mathbb{P}_x\{X_{n+1} = y \,|\, X_n = z\} \\ &= \sum_{z\in\mathcal{S}} P^n_{xz}\,P_{zy} \ = \ P^{n+1}_{xy},\end{aligned}$$

where for the second equality we have used the induction hypothesis and the Markov property (Proposition 1.2).

Remark. If the chain is started with initial distribution μ then the law of total probability and Lemma 1.3 give

$$\begin{aligned}\mathbb{P}_\mu\{X_n = y\} &= \sum_{x\in\mathcal{S}} \mu(x)\,\mathbb{P}_x\{X_n = y\} \\ &= \sum_{x\in\mathcal{S}} \mu(x)\,P^n_{xy} \ =: \ (\mu P^n)(y),\ \ y \in \mathcal{S}.\end{aligned}$$

Examples.

- **Random products**. Let $Y_0, Y_1, \ldots$ be independent and identically distributed (i.i.d.) random variables with values in a discrete set $\mathcal{S} \subset \mathbb{R}$. Set

  $$X_n := Y_0 \cdots Y_n,\ \ n \geq 0.$$

  and

  $$X'_n := Y_{n-1}Y_n,\ \ n \geq 1 \ \text{ and } \ X'_0 := 1.$$

Then $(X_n)_{n\geq 0}$ is a Markov chain with transition probabilities

$$P_{xy} = \mathbb{P}\Big\{Y_0 = \frac{y}{x}\Big\}, \quad x, y \in \bigcup_{k=0}^{\infty} \{y_0 \cdots y_k \colon y_i \in \mathcal{S}\}.$$

The process $(X'_n)_{n\geq 0}$ is not Markovian since the past gives precise information on the factor Y_{n-1}, which is contained in X'_n but not in X'_{n+1}.

- **Simple random walk (SRW) on $\mathbb{Z}$.** Imagine a random walker on the integers, who at each time independently takes a step to the right (with probability $p \in (0,1)$) or to the left (with probability $q = 1-p$). Let X_n be the state of the random walker at time n. Then $(X_n)_{n\geq 0}$ is a Markov chain on the graph $\mathcal{S} = \mathbb{Z}$ with transition probabilities

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

  The simple random walk is called *symmetric* (or *driftless*), if $p = \frac{1}{2}$.

- **Renewal chains.** Some device (e.g., a light bulb) is replaced whenever it expires. The lifetimes of the successive items are assumed i.i.d. with distribution $(q_y)_{y\in\mathbb{N}}$. Define X_n to be the residual lifetime of the item currently in use at time n. The Markov chain $(X_n)_{n\geq 0}$ moves down at unit speed and whenever the chain hits 0 it performs an independent jump according to the lifetime distribution (shifted by 1). Its transition probabilities are $(x, y \in \mathbb{N}_0)$

  $$P_{xy} = \begin{cases} 1, & \text{if } y = x-1 \geq 0, \\ q_{y+1}, & \text{if } x = 0, \\ 0, & \text{else.} \end{cases}$$

1.2 Hitting times and hitting probabilities

Let $B \subset \mathcal{S}$, $B \neq \emptyset$ and define τ_B to be the *first hitting time* of the set B by the Markov chain $(X_n)_{n\geq 0}$,

$$\tau_B := \min\{n \geq 0 : X_n \in B\}.$$

We use the convention $\min\emptyset := \infty$. For $z \in B$ and $x \in \mathcal{S}$ let $h_z(x)$ be the probability that the chain started at x is at z when it first hits B,

$$h_z(x) := \mathbb{P}_x\{X_{\tau_B} = z,\, \tau_B < \infty\} = \sum_{n=0}^{\infty} \mathbb{P}_x\{X_n = z,\, \tau_B = n\}.$$

We note that it might well be that

$$h_B(x) := \mathbb{P}_x\{\tau_B < \infty\} = \sum_{z\in B} h_z(x)$$

is strictly less than 1.

For $f\colon \mathcal{S} \to \mathbb{R}_0^+$ we define the function $Pf\colon \mathcal{S} \to \mathbb{R}_0^+$ through

$$(Pf)(x) := \sum_{y\in\mathcal{S}} P_{xy} f(y) \ \Big(= \mathbb{E}_x f(X_1)\Big), \ x \in \mathcal{S}.$$

Proposition 1.4 *For every $z \in B$ the function h_z satisfies*

$$h_z(x) = (Ph_z)(x), \quad \text{if } x \notin B \tag{1.7}$$

and

$$h_z(x) = \delta_{xz}, \quad \text{if } x \in B. \tag{1.8}$$

A function which satisfies equation (1.7) is called *harmonic* (with respect to P) on $B^c = \mathcal{S}\backslash B$. Equation (1.8) is the so-called boundary condition.

Proof. Fix $z \in B$. By the law of total probability, we have

$$h_z(x) = \sum_{y\in\mathcal{S}} \mathbb{P}_x\{X_{\tau_B} = z,\, \tau_B < \infty \,|\, X_1 = y\} \underbrace{\mathbb{P}_x\{X_1 = y\}}_{= P_{xy}} \tag{1.9}$$

for every $x \in \mathcal{S}$. If $x \notin B$, then $\tau_B \geq 1$ and the Markov property implies

$$\begin{aligned}
\mathbb{P}_x\{X_{\tau_B} = z,\, \tau_B < \infty \,|\, X_1 = y\} &= \sum_{n=1}^{\infty} \mathbb{P}_x\{X_n = z,\, \tau_B = n \mid X_1 = y\} \\
&= \sum_{n=0}^{\infty} \mathbb{P}_x\{X_1 \notin B, \ldots, X_n \notin B, X_{n+1} = z \,|\, X_1 = y\} \\
&= \sum_{n=0}^{\infty} \mathbb{P}_y\{X_0 \notin B, \ldots, X_{n-1} \notin B,\, X_n = z\} \\
&= \mathbb{P}_y\{X_{\tau_B} = z,\, \tau_B < \infty\} \ = \ h_z(y).
\end{aligned} \tag{1.10}$$

Combining (1.9) and (1.10) implies (1.7). The boundary condition (1.8) follows from the fact that $\tau_B = 0$, if $X_0 \in B$.

Corollary 1.5

$$h_B(x) = \begin{cases} (Ph_B)(x), & \text{if } x \notin B, \\ 1, & \text{if } x \in B. \end{cases} \tag{1.11}$$

Proof. Summing (1.7) and (1.8) over $z \in \mathcal{S}$ gives (1.11).

Example (Gambler's ruin). What is the probability of building up a fortune of N Euro before going bankrupt when playing a fair game (loose or win 1 Euro with probability $\frac{1}{2}$) and starting with an initial capital of x Euro. Note that the gambler's fortune performs a symmetric simple random walk, i.e.,

$$P_{zy} = \begin{cases} \frac{1}{2}, & \text{if } z = y \pm 1, \\ 0, & \text{else.} \end{cases}$$

The probability we are interested in is

$$\mathbb{P}_x\{(X_n) \text{ hits } N \text{ before it hits } 0\} = \mathbb{P}_x\{X_{\tau_{\{0,N\}}} = N,\, \tau_{\{0,N\}} < \infty\} = h_N(x)$$

with $B = \{0, N\}$. By Proposition 1.4, we have

$$h_N(x) = \frac{1}{2}h_N(x+1) + \frac{1}{2}h_N(x-1), \quad x \notin \{0, N\} \tag{1.12}$$

and

$$h_N(0) = 0, \quad h_N(N) = 1.$$

Equations (1.12) show that h_N is linear outside of $\{0, N\}$. Consequently,

$$h_N(x) = \begin{cases} 0, & \text{if } x < 0, \\ \frac{x}{N}, & \text{if } 0 \le x \le N, \\ 1, & \text{if } x > N. \end{cases} \tag{1.13}$$

In particular, (1.13) shows that $\mathbb{P}_x\{\tau_N < \infty\}$ for every $x, N \in \mathbb{Z}$.

Example (SRW with drift). In general, the solution to the system of equations (1.7) and (1.8) is not unique: Let

$$P_{xy} = \begin{cases} p, & \text{if } y = x+1, \\ q = 1-p, & \text{if } y = x-1, \\ 0, & \text{else.} \end{cases}$$

If we take $B = \{0\}$, then

$$f \equiv 1 \text{ and } g(x) = \left(\frac{q}{p}\right)^x, \quad x \in \mathbb{Z}$$

are both solutions to (1.7) and (1.8). If $p \neq \frac{1}{2}$ then f and g differ.

However, the function h_z is distinguished being the minimal (non-negative) solution to (1.7) which satisfies boundary condition (1.8)

Proposition 1.6 *Let $z \in B$ and suppose that $f = f_z : \mathcal{S} \to \mathbb{R}_0^+$ satisfies $f = 1_{\{z\}}$ on B and $Pf \le f$ on B^c. Then*

$$h_z(x) \le f(x) \text{ for all } x \in \mathcal{S}.$$

Proof. We will show

$$\mathbb{P}_x\{X_{\tau_B} = z,\, \tau_B \le n\} \le f(x) \text{ for all } x \in \mathcal{S}, n \in \mathbb{N}_0. \tag{1.14}$$

If $x \in B$, then $\mathbb{P}_x\{X_{\tau_B} = z,\, \tau_B \le n\} = \delta_{xz} = f(x)$. For $x \notin B$ we proceed by induction. First, note that

$$\mathbb{P}_x\{X_{\tau_B} = z,\, \tau_B \le 0\} = 0 \le f(x)$$

by the assumed non-negativity of f. For the induction step use $\tau_B \geq 1$ $\mathbb{P}_x$-a.s. to deduce that

$$\begin{aligned}\mathbb{P}_x\{X_{\tau_B} = z,\, \tau_B \leq n+1\} &= \sum_{y\in\mathcal{S}} \mathbb{P}_x\{X_{\tau_B} = z,\, \tau_B \leq n+1 \,|\, X_1 = y\}\, P_{xy} \\ &= \sum_{y\in\mathcal{S}} \mathbb{P}_y\{X_{\tau_B} = z,\, \tau_B \leq n\}\, P_{xy} \\ &\leq \sum_{y\in\mathcal{S}} f(y) P_{xy} \;=\; (Pf)(x) \;\leq\; f(x).\end{aligned}$$

Passing to the limit $n \to \infty$ in (1.14) proves the assertion of Proposition 1.6.

Now let $e_B(x)$ be the expected time to hit B when starting at x, i.e.,

$$e_B(x) := \mathbb{E}_x \tau_B \;=\; \sum_{n=0}^{\infty} n\, \mathbb{P}_x\{\tau_B = n\} + \infty\, \mathbb{P}_x\{\tau_B = \infty\},\ x \in \mathcal{S}.$$

Proposition 1.7 *For every $B \subset \mathcal{S}$ the function e_B is the minimal non-negative solution to*

$$e(x) = \begin{cases} 1 + (Pe)(x), & \text{if } x \notin B \\ 0, & \text{if } x \in B. \end{cases} \tag{1.15}$$

Proof. We only prove the minimality of e_B (the proof that e_B solves (1.15) is similar to the proof of Proposition 1.4 and is left as an excercise). So let e be a non-negative solution of (1.15). We will show

$$e(x) \;\geq\; \sum_{k=1}^{n} \mathbb{P}_x\{\tau_B \geq k\} \text{ for all } n \in \mathbb{N}_0 \text{ and } x \in \mathcal{S}. \tag{1.16}$$

To see that (1.16) implies that e_B is minimal recall that if T is a random variable with values in $\mathbb{N}_0 \cup \{\infty\}$, then

$$\begin{aligned}\mathbb{E}T &= \sum_{j=1}^{\infty} j\, \mathbb{P}\{T = j\} \,+\, \infty\, \mathbb{P}\{T = \infty\} \\ &= \sum_{j=1}^{\infty}\sum_{k=1}^{j} \mathbb{P}\{T = j\} \,+\, \infty\, \mathbb{P}\{T = \infty\} \\ &= \sum_{k=1}^{\infty}\sum_{j=k}^{\infty} \mathbb{P}\{T = j\} \,+\, \infty\, \mathbb{P}\{T = \infty\} \\ &= \sum_{k=1}^{\infty} \mathbb{P}\{T \geq k\} \,+\, \infty\, \mathbb{P}\{T = \infty\} \;=\; \sum_{k=1}^{\infty} \mathbb{P}\{T \geq k\}. \end{aligned} \tag{1.17}$$

Hence, the minimality of e_B follows from (1.16) by passing to the limit $n \to \infty$.

Inequality (1.16) is obvious for $x \in B$ or if $n = 0$. To establish the general case we proceed by induction. Note that that for $x \notin B$

$$\begin{aligned}\sum_{k=1}^{n+1} \mathbb{P}_x\{\tau_B \geq k\} &= 1 + \sum_{k=2}^{n+1} \sum_{y \in \mathcal{S}} P_{xy}\, \mathbb{P}_y\{\tau_B \geq k-1\} \\ &= 1 + \sum_{y \in \mathcal{S}} P_{xy} \sum_{k=1}^{n} \mathbb{P}_y\{\tau_B \geq k\} \\ &\leq 1 + \sum_{y \in \mathcal{S}} P_{xy} e(y) \\ &= 1 + (Pe)(x) \;=\; e(x). \end{aligned} \tag{1.18}$$

Example. Let (X_n) be symmetric SRW on $\mathbb{Z}$ and take $B = \{0, N\}$. Then (1.15) implies

$$e_B(x) = 1 + \frac{1}{2} e_B(x-1) + \frac{1}{2} e_B(x+1) \tag{1.19}$$

for $x \notin \{0, N\}$. Equation (1.19) shows that on $\mathbb{Z} \setminus \{0, N\}$ the function e_B is quadratic with $e_B''(x) = -2$ (if $e_B''(x)$ exists). Since $e_B(0) = e_B(N) = 0$ we obtain

$$e_B(x) = \begin{cases} x(N-x), & \text{if } 0 \leq x \leq N, \\ \infty, & \text{else.} \end{cases} \tag{1.20}$$

Example (An algorithm for finding the maximum). Let $Y_1, Y_2, \ldots, Y_r$ be i.i.d. random variables with values in $\mathbb{R}$ and a density f (this assumption is merely to ensure that the Y_i are pairwise different).

1. Set $M := Y_1$
2. For $i = 2, \ldots, r$:
   If $Y_i > M$, then set $M := Y_i$
3. Output: M

Let I_{j+1} be the time of the jth exchange of the current record, i.e., we set $I_1 := 1$ and

$$I_{j+1} := \begin{cases} \min\{i > I_j : Y_i > Y_{I_j}\}, & \text{if } \{i : Y_i > Y_{I_j}\} \neq \emptyset, \\ I_j, & \text{else.} \end{cases}$$

We write J for the total number of record holders during the execution of the algorithm,

$$J := \min\{j \geq 1 : I_{j+1} = I_j\}.$$

Note that $J - 1$ is the total number of exchanges which we take as a measure for the costs of the algorithm. We will do an average case analysis and calculate $\mathbb{E}J$. For this purpose we introduce the process $(X_n)_{n \geq 0}$ with $X_0 := 0$ and

$$X_j := \text{rank}(Y_{I_j}),\ j \geq 1.$$

Claim: $(X_j)_{j\geq 0}$ is a Markov chain on $\{0,\ldots,r\}$ with transition matrix

$$P_{xy} = \begin{cases} (r-x)^{-1}, & \text{if } x<y, \\ 1, & \text{if } x=y=r, \\ 0, & \text{else.} \end{cases}$$

Indeed, note that given the event $\{X_0=x_0, X_1=x_1,\ldots,X_n=x_n\}$ we know that

- Comparison with the element of rank x_n has taken place.
- Comparison with the elements of ranks $x_n+1,\ldots,r$ are still to come.

By our assumption the succession of those latter elements is purely random. Hence, any of these $r-x_n$ elements is equally likely to be the next to be compared with.

Note that $J=\min\{j\geq 1: X_j=r\}$ and, consequently,

$$\mathbb{E}\,J = \mathbb{E}_0\tau_{\{r\}} = e_{\{0\}}(0).$$

By Proposition 1.7, we have for $x<r$

$$e_{\{0\}}(x) \;=\; 1+(Pe_{\{0\}})(x) \;=\; 1+\frac{1}{r-x}\Big(e_{\{0\}}(x+1)+\ldots+e_{\{0\}}(r)\Big) \tag{1.21}$$

and $e_{\{0\}}(r)=0$. Solving the recursive equation (1.21) yields

$$e_{\{0\}}(x)=\sum_{j=1}^{r-x}\frac{1}{j}$$

and, consequently,

$$\mathbb{E}J=\sum_{j=1}^{r}\frac{1}{j} \;\sim\; r\log r \quad \text{as } r\to\infty. \tag{1.22}$$

The intention of this example was to illustrate the technique suggested by Propositions 1.4 and 1.7. We note that there is a much simpler way to derive (1.22) using linearity of expectation (Exercise !).

1.3 Recurrence and transience

For $x\in\mathcal{S}$ let

$$\sigma_x := \min\{n\geq 1:\, X_n=x\}$$

be the *first passage time* of $(X_n)_{n\geq 0}$ to x (or *return time*, if $X_0=x$). Clearly, $\sigma_x\geq\tau_x$ and $\sigma_x=\tau_x$ on $\{X_0\neq x\}$.

Definition 1.8 *A state $x\in\mathcal{S}$ is called* recurrent, *if $\mathbb{P}_x\{\sigma_x<\infty\}=1$. Otherwise, it is called* transient. *A Markov chain is called* recurrent/transient, *if all states are recurrent/transient.*

Let L_x be the number of visits in x by $(X_n)_{n\geq 0}$, i.e.,

$$L_x := \sum_{n=0}^{\infty} I_{\{X_n=x\}}.$$

The random variable L_x is called the *occupation time* (or *local time*) at x. The following theorem gives a handy criterion for recurrence of a state x in terms of L_x.

Theorem 1.9 *For any state $x \in \mathcal{S}$ the following are equivalent:*

i) *x is recurrent.*

ii) $\mathbb{P}_x\{L_x = \infty\} = 1$.

iii) $\mathbb{E}_x L_x = \infty$.

Note that, by definition, x is recurrent if and only if $\mathbb{P}_x\{L_x \geq 2\} = 1$. The typical application of Theorem 1.9 is to conclude recurrence through verification of *iii*).

The result is an immediate consequence of the fact that, under $\mathbb{P}_x$, the occupation time at x is a geometric random variable with parameter $p_x := \mathbb{P}_x\{\sigma_x = \infty\}$ (= escape probability):

Proposition 1.10

$$\mathbb{P}_x\{L_x = k\} = p_x(1-p_x)^{k-1},\ k \in \mathbb{N}. \tag{1.23}$$

Proof of proposition. Let $\sigma_x^{(j)}$ be the time of the jth passage to x, i.e.,

$$\sigma_x^{(0)} := 0 \ \text{ and } \ \sigma_x^{(j)} := \min\{n > \sigma_x^{(j-1)} :\ X_n = x\},\ j \geq 1.$$

Note that

$$L_x \;=\; I_{\{X_0=x\}} + \sum_{j=1}^{\infty} I_{\{\,\sigma_x^{(j)}<\infty\}} \;=\; I_{\{X_0=x\}} + \max\{j \geq 0\colon\ \sigma_x^{(j)} < \infty\}.$$

By the strong Markov property (Exercise 1.2), conditional on $\{\sigma_x^{(j)} < \infty\}$ the process $(X_{\sigma_x^{(j)}+n})_{n\geq 0}$ is a Markov chain with transition matrix P started at x, independent of $(X_m)_{0\leq m<\sigma_x^{(j)}}$. Hence, for every $k \geq 1$,

$$\begin{aligned}\mathbb{P}_x\{L_x \geq k+1 \,|\, L_x \geq k\} \;&=\; \mathbb{P}_x\{\sigma_x^{(k)} < \infty \,|\, \sigma_x^{(k-1)} < \infty\} \\ &=\; \mathbb{P}_x\{\sigma_x < \infty\} \;=\; 1 - p_x. \end{aligned} \tag{1.24}$$

Iteration of (1.24) gives (1.23).

Proof of Theorem 1.9. By definition, x is recurrent if and only if $p_x = 0$. Now use Proposition 1.10 and the fact that a geometric distribution on $\mathbb{N}$ with parameter p has mean $\frac{1}{p}$.

Corollary 1.11 *A state $x \in \mathcal{S}$ is recurrent if and only if* $\sum_{n=0}^{\infty} P_{xx}^n = \infty$.

Proof. In view of Lemma 1.3 we have

$$\mathbb{E}_x L_x \;=\; \mathbb{E}_x\Big(\sum_{n=0}^{\infty} I_{\{X_n=x\}}\Big) \;=\; \sum_{n=0}^{\infty} \mathbb{P}_x\{X_n = x\} \;=\; \sum_{n=0}^{\infty} P_{xx}^n.$$

The claim now follows by Theorem 1.9.

Example (Symmetric SRW on $\mathbb{Z}^2$).

$$P_{xy} = \begin{cases} \frac{1}{4}, & \text{if } ||x-y|| = 1, \\ 0, & \text{else.} \end{cases}$$

Note that in order to return to his starting point the random walker must take as many steps to the North as to the South and as many steps to the East as to the West. Hence, $P_{xx}^{2n+1} = 0$ and

$$\begin{aligned} P_{xx}^{2n} &= \sum_{i=0}^{n} \binom{2n}{i,i,n-i,n-i} \left(\frac{1}{4}\right)^{2n}, \\ &= \left(\frac{1}{4}\right)^{2n} \binom{2n}{n} \sum_{i=0}^{n} \binom{n}{i}\binom{n}{n-i} \\ &= \left(\frac{1}{4}\right)^{2n} \binom{2n}{n}^2, \end{aligned}$$

where for the last equality notice that $\binom{n}{i}\binom{n}{n-i}/\binom{2n}{n}$ is the weight at i of a hypergeometric distribution with parameter $(n, n, 2n)$. By Stirling's formula ($n! \sim (\frac{n}{e})^n \sqrt{2\pi n}\,$), we thus have

$$P_{xx}^{2n} \sim \frac{1}{\pi n} \quad \text{as } n \to \infty.$$

Now use Corollary 1.11 to conclude that symmetric SRW on $\mathbb{Z}^2$ is recurrent.

In much the same way it can be shown that symmetric SRW is transient in dimensions 3 and higher.

We now turn to the question whether it can be that one state is transient while another is recurrent. In principle, the answer is yes. For instance, take symmetric simple random walk on $\mathbb{Z}$ with absorption at the origin (i.e., $P_{00} = 1$). For this chain the origin is recurrent while all other states are transient. However, note that $\mathbb{Z} \setminus \{0\}$ "cannot be reached" from the origin.

Definition 1.12 *We say that a state y* can be reached *from state x (and write "$x \to y$"), if*

$$P_{xy}^{n} > 0 \textit{ for some } n \in \mathbb{N}_0.$$

The following equivalent description of reachability is often more convenient:

$$x \to y \iff \mathbb{P}_x\{\tau_y < \infty\} > 0. \tag{1.25}$$

Indeed, note that for every $n \in \mathbb{N}_0$

$$P_{xy}^{n} = \mathbb{P}_x\{X_n = y\} \le \mathbb{P}_x\{\tau_y \le n\} \le \mathbb{P}_x\{\tau_y < \infty\}.$$

On the other hand,

$$\mathbb{P}_x\{\tau_y < \infty\} = \mathbb{P}_x\{X_n = y \text{ for some } n \ge 0\} \le \sum_{n=0}^{\infty} P_{xy}^{n}.$$

Lemma 1.13 *If x is recurrent and $x \to y$, then*

i) $\mathbb{P}_x\{L_y = \infty\} = 1$.

ii) $y \to x$.

iii) *y is recurrent.*

Proof.

i) Clearly, we can bound

$$L_y \;=\; \sum_{n=0}^{\infty} I_{\{X_n = y\}} \;\geq\; \sum_{j=0}^{\infty} I\{X_n = y \text{ for some } \sigma_x^{(j)} \leq n < \sigma_x^{(j+1)}\}.$$

By the strong Markov property and recurrence of x, under $\mathbb{P}_x$ the indicator variables on the right-hand side above are i.i.d. Bernoulli random variables with success probability

$$p := \mathbb{P}_x\{X_n = y \text{ for some } 0 \leq n < \sigma_x\} \;=\; \mathbb{P}_x\{\tau_y < \sigma_x\} > 0.$$

For positivity of p notice that

$$0 < \mathbb{P}_x\{\tau_y < \infty\} \leq \sum_{j=0}^{\infty} \mathbb{P}_x\{X_n = y \text{ for some } n \in \{\sigma_x^{(j)}, \ldots, \sigma_x^{(j+1)} - 1\}\} = \sum_{n=0}^{\infty} p\,.$$

ii) By Theorem 1.9 and the strong Markov property, we have

$$\begin{aligned} 0 \;=\; \mathbb{P}_x\{L_x < \infty\} \;&\geq\; \mathbb{P}_x\{\tau_y < \infty,\, X_n \neq x \text{ for all } n > \tau_y\} \\ &=\; \mathbb{P}_x\{\tau_y < \infty\}\,\mathbb{P}_y\{\sigma_x = \infty\}. \end{aligned}$$

By the assumed reachability of y from x the first factor is positive. Hence, the second factor must be zero.

iii) The strong Markov property and *i*), *ii*) imply

$$\begin{aligned} \mathbb{P}_y\{L_y = \infty\} \;&\geq\; \mathbb{P}_y\{L_y = \infty \,|\, \tau_x < \infty\}\,\mathbb{P}_y\{\tau_x < \infty\} \\ &=\; \mathbb{P}_x\{L_y = \infty\}\,\mathbb{P}_y\{\tau_x < \infty\} \;>\; 0. \end{aligned}$$

Recurrence of y follows by Theorem 1.9.

Remarks. If x and y can be reached from each other we say that x and y *communicate* (and write "$x \leftrightarrow y$"). Mutual reachability defines an equivalence relation on the state space $\mathcal{S}$. The equivalence classes are called *irreducible components*. Lemma 1.13 shows that communicating states are of the same type (either all recurrent or all transient).

Definition 1.14 *A Markov chain $(X_n)_{n \geq 0}$ with transition matrix P is called* irreducible*, if any two states communicate. It is called* irreducible recurrent*, if it consists of exactly one irreducible recurrent component.*

1.4 Stationary distributions

Definition 1.15 *A probability measure* π *on* $\mathcal{S}$ *is called a* stationary distribution *(or* equilibrium distribution*) for* P*, if*

$$\pi(x) = \sum_{y \in \mathcal{S}} \pi(y)\, P_{yx}, \qquad x \in \mathcal{S}. \tag{1.26}$$

Remarks.

- In matrix notation (1.26) reads $\pi = \pi P$. By induction, we see that

  $$\pi(x) = \sum_{y \in \mathcal{S}} \pi(y) P^n_{yx}, \quad n \in \mathbb{N}_0,\, x \in \mathcal{S}. \tag{1.27}$$

  Note that the right-hand side of (1.27) equals $\mathbb{P}_\pi\{X_n = x\}$. In fact, a Markov chain with transition matrix P and initial distribution μ is a stationary sequence (i.e., $\mathbb{P}_\mu\{(X_n, X_{n+1}, \dots, X_{n+k}) \in \cdot\,\} = \mathbb{P}_\mu\{(X_0, \dots, X_k) \in \cdot\,\}$ for all $n, k \in \mathbb{N}_0$) if and only if μ is a stationary distribution for P (verify!).

- (1.26) can be viewed as a *balance equation*: Imagine masses placed at each of the states in $\mathcal{S}$ and suppose that proportion P_{xy} of the mass at x is passed to y. If the mass distribution is stationary, then at each state the mass sent off is the same as the total mass received.

The chain is called *reversible* with respect to π, if P satisfies the *detailed balance equations*

$$\pi(x) P_{xy} = \pi(y) P_{yx}, \quad x, y \in \mathcal{S}. \tag{1.28}$$

Clearly, (1.28) implies (1.27).

Example. If $\mathcal{S}$ is finite and the transition matrix P is symmetric or doubly stochastic, then the uniform distribution on $\mathcal{S}$ is a stationary distribution for P. If P is doubly stochastic, then the chain is reversible if and only if it is symmetric.

In the sequel we will derive conditions for existence and uniqueness of stationary distributions. Later we will see that subject to some quite natural conditions a Markov chain converges towards its equilibrium distribution. Often it is instructive to think of $\pi(x)$ as the asymptotic proportion of time that the chain spends at x.

First we have a look at an important example. Recall the definition of a renewal chain with lifetime distribution $(q_y)_{y \in \mathbb{N}}$,

$$P_{xy} = \begin{cases} 1, & \text{if } y = x - 1 \geq 0, \\ q_{y+1}, & \text{if } x = 0, \\ 0, & \text{else.} \end{cases}$$

If π is a stationary distribution for P, then π must satisfy

$$\pi(x) = \pi(x+1) + \pi(0)q_{x+1} \text{ for all } x \in \mathbb{N}_0. \tag{1.29}$$

Summing (1.29) from $x = 0$ to $z - 1$ gives

$$\pi(z) = \pi(0)\Big(1 - \sum_{x=0}^{z-1} q_{x+1}\Big) = \pi(0) \sum_{x=z+1}^{\infty} q_x = \pi(0)\,\mathbb{P}_0\{\sigma_0 \geq z+1\}. \tag{1.30}$$

To determine the weight $\pi(0)$ we sum (1.30) over $z \in \mathbb{N}_0$ and use (1.17) to deduce

$$1 = \pi(0) \sum_{z=1}^{\infty} \mathbb{P}_0\{\sigma_0 \geq z\} = \pi(0)\,\mathbb{E}_0\sigma_0. \tag{1.31}$$

Hence, the equilibrium weight at 0 is the inverse of the expected duration of an excursion from 0,

$$\pi(0) = \frac{1}{\mathbb{E}_0\sigma_0}. \tag{1.32}$$

This identity will turn out to hold in quite some generality. Note that the renewal chain has a stationary distribution only if $\mathbb{E}_0\sigma_0 < \infty$.

Definition 1.16 *A state $x \in \mathcal{S}$ is said to be* positive recurrent*, if $\mathbb{E}_x\sigma_x < \infty$. A Markov chain is called positive recurrent if all states are positive recurrent.*

Example. Symmetric SRW on $\mathbb{Z}$ is recurrent, but not positive recurrent. This is since (recall (1.20))

$$\mathbb{E}_0\sigma_0 = 1 + \mathbb{E}_1\tau_0 = 1 + \lim_{N\to\infty} \mathbb{E}_1 \min(\tau_0, \tau_N) = \infty.$$

Our next result says that a stationary distribution is supported by recurrent states.

Proposition 1.17 *If π is a stationary distribution for P and $\pi(x) > 0$, then x is recurrent.*

Proof. Assume x is transient. By the strong Markov property and Theorem 1.9, we have

$$\begin{aligned}\sum_{n=0}^{\infty} P^n_{yx} = \sum_{n=0}^{\infty} \mathbb{E}_y I_{\{X_n=x\}} = \mathbb{E}_y\Big(\sum_{n=0}^{\infty} I_{\{X_n=x\}}\Big) &= \mathbb{E}_y L_x \\ &= \mathbb{E}_y\{L_x \,|\, \tau_x < \infty\}\,\mathbb{P}_y\{\tau_x < \infty\} \leq \mathbb{E}_x L_x < \infty \end{aligned}\tag{1.33}$$

for every $y \in \mathcal{S}$. Consequently,

$$P^n_{yx} \to 0 \text{ as } n \to \infty \text{ for every } y \in \mathcal{S}.$$

Hence (recall (1.27)),

$$\pi(x) = \sum_{y\in\mathcal{S}} \pi(y)P^n_{yx} \to 0 \text{ as } n \to \infty$$

by the dominated convergence theorem. Since $\pi(x)$ does not depend on n, we see that $\pi(x) = 0$.

Definition 1.18 *A stationary distribution is called* ergodic*, if it is supported by a single irreducible component.*

Remarks.

- Note that this component is equal to supp $\pi := \{x : \pi(x) > 0\}$ (because all states of the component communicate with each other).
- Suppose that π is a stationary distribution for P with irreducible components $\mathcal{S}_1, \mathcal{S}_2, \ldots$. For $i \in I = \{j : \pi(\mathcal{S}_j) > 0\}$ define the probability measure π_i as

  $$\pi_i(x) := \begin{cases} \dfrac{\pi(x)}{\pi(\mathcal{S}_i)}, & \text{if } x \in \mathcal{S}_i, \\ 0, & \text{else.} \end{cases}$$

  Then, π_i is an ergodic stationary distribution for P (verify!) and

  $$\pi = \sum_{i \in I} \pi(\mathcal{S}_i)\, \pi_i$$

  is the so-called *ergodic decomposition* of π.

Theorem 1.19 *If π is an ergodic stationary distribution for P, then*

$$\pi(x) = \frac{1}{\mathbb{E}_x \sigma_x} \quad \textit{for every } x \in \operatorname{supp} \pi.$$

Proof. We reduce the general case to that of a renewal chain. Fix $x \in \operatorname{supp} \pi$. Then, by the assumed ergodicity, Lemma 1.13 and Proposition 1.17,

$$\begin{aligned} \mathbb{P}_\pi\{\sigma_x < \infty\} \ &\geq\ \mathbb{P}_\pi\{L_x = \infty\} \ =\ \sum_{y \in \mathcal{S}} \pi(y)\, \mathbb{P}_y\{L_x = \infty\} \\ &=\ \sum_{y: y \leftrightarrow x} \pi(y)\, \mathbb{P}_y\{L_x = \infty\} \ =\ \sum_{y: y \leftrightarrow x} \pi(y) \ =\ \sum_{y \in \mathcal{S}} \pi(y) \ =\ 1. \end{aligned} \tag{1.34}$$

Now let Y_n be the residual duration of the excursion from state x at time n,

$$Y_n := \min\{j \geq 0 : X_{n+j} = x\}, \quad n \in \mathbb{N}_0.$$

Clearly,

i) $Y_0 = \tau_x$.

ii) $Y_n = 0 \iff n = \sigma_x^{(j)}$ for some $j \geq 1$ or $n = \tau_x = 0$.

Now observe that under $\mathbb{P}_\pi$ the chain $(Y_n)_{n \geq 0}$ is a stationary renewal chain with lifetime distribution

$$(q_j)_{j \geq 1} = (\mathbb{P}_x\{\sigma_x = j\})_{j \geq 1}$$

and initial distribution

$$(\mu(k))_{k \geq 0} = (\mathbb{P}_\pi\{\tau_x = k\})_{k \geq 0}.$$

Note that (1.34) and *i)* imply that μ is a probabilty measure on $\mathbb{N}_0$. Hence, letting $\tilde{\sigma}_0$ denote the first passage time of the renewal chain $(Y_n)_{n \geq 0}$ to 0, relation (1.32) gives

$$\pi(x) = \mathbb{P}_\pi\{X_0 = x\} = \mathbb{P}_\mu\{Y_0 = 0\} = \mu(0) = \frac{1}{\mathbb{E}_0 \tilde{\sigma}_0} = \frac{1}{\mathbb{E}_x \sigma_x}.$$

Corollary 1.20 (Uniqueness) *An irreducible Markov chain has at most one stationary distribution. If it has one, then all states are positive recurrent.*

How about existence of a stationary distribution? This question will be answered in the affirmative, i.e., π is indeed an equilibrium distribution. To prove the result we need a more convenient representation of $\pi(x)$ than the one given in Theorem 1.19.

Fix $z \in \mathcal{S}$. Let $m_z(x)$ be the expected number of visits to x during an excursion from z,

$$m_z(x) := \mathbb{E}_z\Big(\sum_{n=1}^{\sigma_z} I_{\{X_n=x\}}\Big), \quad x \in \mathcal{S}.$$

Note that m_z induces a measure on $\mathcal{S}$ with total mass

$$m_z(\mathcal{S}) = \sum_{x\in\mathcal{S}} \mathbb{E}_z\Big(\sum_{n=1}^{\sigma_z} I_{\{X_n=x\}}\Big) = \mathbb{E}_z\Big(\sum_{n=1}^{\sigma_z}\sum_{x\in\mathcal{S}} I_{\{X_n=x\}}\Big) = \mathbb{E}_z\sigma_z.$$

Theorem 1.21 (Existence) *If $z \in \mathcal{S}$ is recurrent, then m_z is P-invariant, i.e.,*

$$m_z(x) = \sum_{y\in\mathcal{S}} m_z(y)P_{yx} \quad \textit{for all } x \in \mathcal{S}. \tag{1.35}$$

In particular, if $(X_n)_{n\geq 0}$ is an irreducible Markov chain and z is positive recurrent, then

$$\pi = \frac{m_z}{m_z(\mathcal{S})}$$

is the unique stationary distribution.

Moreover,

$$\sum_{y\in\mathcal{S}} f(y)\pi(y) = \mathbb{E}_z\Big(\sum_{n=1}^{\sigma_z} f(X_n)\Big)\Big/\,\mathbb{E}_z\sigma_z \tag{1.36}$$

for every π-integrable $f\colon \mathcal{S} \to \mathbb{R}$.

Note that Theorem 1.21 implies

$$\frac{1}{\mathbb{E}_x\sigma_x} = \frac{m_z(x)}{m_z(\mathcal{S})} \text{ for all } x, z \in \mathcal{S}$$

(which is what we expect in view of our interpretation of $\pi(x)$ as the asymptotic proportion of time spent in x). If P is symmetric or doubly stochastic then $m \equiv 1$ is P-invariant.

Proof. To understand identity (1.35) note that P_{yx} is the probability to move to x when presently at y and that $m_z(y)$ is the expected number of trials to go from y to x during an excursion from z. Hence, ignoring dependencies between the number of visits to different states during an excursion from z the law of total probability should give the P-invariance of m_z. To give a formal proof we use a so-called "last exit decomposition". For $x \in \mathcal{S}$ we have

$$\begin{aligned} m_z(x) &= \mathbb{E}_z\Big(\sum_{n=1}^{\sigma_z} I_{\{X_n=x\}}\Big) \\ &= \mathbb{E}_z\Big(\sum_{n=1}^{\infty}\sum_{y\in\mathcal{S}} I_{\{X_{n-1}=y,\, X_n=x,\, n\le\sigma_z\}}\Big) \\ &= \sum_{y\in\mathcal{S}}\sum_{n=1}^{\infty} \mathbb{P}_z\{X_{n-1}=y, X_n=x,\, n\le\sigma_z\}. \end{aligned}$$

Since $\{\sigma_z \ge n\} = \{X_1 \ne z, X_2 \ne z, \ldots, X_{n-1} \ne z\}$, the Markov property implies

$$\mathbb{P}_z\{X_{n-1}=y, X_n=x,\, \sigma_z\ge n\} = \mathbb{P}_z\{X_{n-1}=y,\, \sigma_z\ge n\}\, P_{yx}.$$

To complete the proof of the P-invariance of m_z it remains to show that

$$\sum_{n=1}^{\infty}\mathbb{P}_z\{X_{n-1}=y,\, \sigma_z\ge n\} = m_z(y).$$

Since $\sigma_z < \infty$ (by recurrence of z) and since $X_0 = X_{\sigma_z} = z$ under $\mathbb{P}_z$, we have

$$\begin{aligned} \sum_{n=1}^{\infty}\mathbb{P}_z\{X_{n-1}=y, \sigma_z\ge n\} &= \mathbb{E}_z\Big(\sum_{n=1}^{\sigma_z} I_{\{X_{n-1}=y\}}\Big) \\ = \mathbb{E}_z\Big(\sum_{n=0}^{\sigma_z-1} I_{\{X_n=y\}}\Big) &= \mathbb{E}_z\Big(\sum_{n=1}^{\sigma_z} I_{\{X_n=y\}}\Big) = m_z(y). \end{aligned}$$

If z is positive recurrent (i.e., $m_z(\mathcal{S}) = \mathbb{E}_z\sigma_z < \infty$), then $\pi = m_z/m_z(\mathcal{S})$ is a stationary distribution. If $(X_n)_{n\ge0}$ is irreducible, then π is unique (by Corollary 1.20).

For functions f of the form $f = 1_{\{y\}}$ for some $y \in \mathcal{S}$ assertion (1.36) is the same as (1.35). Since any function f can be represented as

$$f = \sum_{y\in\mathcal{S}} f(y)\, 1_{\{y\}},$$

the case of general π-integrable f follows by linearity of expectation.

1.5 Limit theorems

Ergodic theorems are limit theorems which deal with the equality of space and time averages. Our first result is of that kind. It states that the time average of a Markov chain sample converges towards the mean w.r.t. its stationary distribution.

Theorem 1.22 (Law of large numbers) *Let $(X_n)_{n\ge0}$ be an irreducible and positive recurrent Markov chain with stationary distribution π. Then*

$$\lim_{n\to\infty}\frac{1}{n+1}\sum_{k=0}^{n} f(X_k) = \sum_{y\in\mathcal{S}} f(y)\pi(y) \quad \mathbb{P}_z\text{-}a.s.$$

for every $z \in \mathcal{S}$ and every π-integrable $f\colon\, \mathcal{S} \to \mathbb{R}$.

Remark. Taking $f = 1_{\{x\}}$ gives

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} I_{\{X_k = x\}} = \pi(x) \quad \mathbb{P}_z\text{-}a.s.,$$

i.e., $\pi(x)$ is indeed the asymptotic proportion of time spent at x.

Proof. With no loss of generality we may assume $f \geq 0$ (else decompose $f = f^+ - f^-$). Fix $z \in \mathcal{S}$ and let J_n be the number of passages to z until time n,

$$J_n := \max\{j \geq 0 : \sigma_z^{(j)} \leq n\}.$$

Note that

i) $J_n \overset{\mathbb{P}_z\text{-}a.s.}{\to} \infty$ as $n \to \infty$ (by recurrence of z).

ii) $\sigma_z^{(J_n)} \leq n < \sigma_z^{(J_n+1)}$ (by definition of J_n).

Since $f \geq 0$ we have

$$\frac{1}{\sigma_z^{(J_n+1)}} \sum_{k=0}^{\sigma_z^{(J_n)}-1} f(X_k) \;\leq\; \frac{1}{n+1} \sum_{k=0}^{n} f(X_k) \;\leq\; \frac{1}{\sigma_z^{(J_n)}} \sum_{k=0}^{\sigma_z^{(J_n+1)}-1} f(X_k) \quad \mathbb{P}_z\text{-}a.s. \tag{1.37}$$

Under $\mathbb{P}_z$,

$$Y_i := \sum_{k=\sigma_z^{(i-1)}}^{\sigma_z^{(i)}-1} f(X_k), \quad i \geq 1$$

are i.i.d. random variables with mean (compare (1.36))

$$\mathbb{E}_z Y_1 = \mathbb{E}_z\left(\sum_{k=0}^{\sigma_z - 1} f(X_k)\right) = \mathbb{E}_z\left(\sum_{k=1}^{\sigma_z} f(X_k)\right) = (\mathbb{E}_z \sigma_z) \sum_{y \in \mathcal{S}} f(y)\pi(y).$$

Hence, the standard law of large numbers implies

$$\frac{1}{j} \sum_{k=0}^{\sigma_z^{(j)}-1} f(X_k) = \frac{1}{j} \sum_{i=1}^{j} Y_i \overset{\mathbb{P}_z\text{-}a.s.}{\longrightarrow} (\mathbb{E}_z\sigma_z) \sum_{y\in\mathcal{S}} f(y)\pi(y) \;\text{ as }\; j \to \infty. \tag{1.38}$$

Also, under $\mathbb{P}_z$,

$$\sigma_z^{(j)} = \sum_{i=1}^{j} \left(\sigma_z^{(i)} - \sigma_z^{(i-1)}\right)$$

is the sum of j i.i.d. random variables with mean $\mathbb{E}_z\sigma_z$. Hence,

$$\frac{1}{j}\,\sigma_z^{(j)} \overset{\mathbb{P}_z\text{-}a.s.}{\longrightarrow} \mathbb{E}_z\sigma_z \;\text{ as }\; j \to \infty. \tag{1.39}$$

Combining (1.38) and(1.39) with *ii)* shows that both sides of (1.37) converge to $\sum_{y\in\mathcal{S}} f(y)\pi(y)$ $\mathbb{P}_z$ - a.s.

We now turn to the question, under which circumstances we have the stronger convergence $\mathbb{P}\{X_n = y\} \to \pi(y)$ as $n \to \infty$. We need some additional conditions as the following example shows.

Consider symmetric SRW on $\mathbb{Z}_{2k}$ for some $k \in \mathbb{N}$. By symmetry, the uniform distribution $\pi(x) = (2k)^{-1}$, $x \in \mathbb{Z}_{2k}$ is the (unique) stationary distribution. However, $\mathbb{P}_x\{X_{2n+1} = x\} = 0$ for every $n \in \mathbb{N}_0$. We will see, however, that this kind of periodicity is all that can go wrong.

Definition 1.23 *A state x of a Markov chain with transition matrix P is called* aperiodic *if*

$$gcd\{n \geq 1 : P_{xx}^n > 0\} = 1.$$

A Markov chain is called aperiodic*, if all states are aperiodic.*

Lemma 1.24

i) *x is aperiodic $\Longleftrightarrow$ $P_{xx}^n > 0$ for all n sufficiently large.*

ii) *If $(X_n)_{n\geq 0}$ is irreducible and some state is aperiodic, then the chain is aperiodic.*

Proof.

i) A fundamental (and elementary) theorem from number theory states that, if $A \subset \mathbb{N}$ is closed under addition and *gcd* $A = 1$, then $\mathbb{N} \setminus A$ is finite (for a proof see, e.g., the appendix of [2]). Clearly, $P_{xx}^{n+k} \geq P_{xx}^n P_{xx}^k$. Hence, $\{n \geq 1 : P_{xx}^n > 0\}$ is such a set.

ii) Let $z \neq x$ and suppose x is aperiodic. By the assumed irreducibility, states x and z communicate, i.e., $P_{zx}^{\ell}, P_{xz}^m > 0$ for some $\ell, m \in \mathbb{N}$. Consequently,

$$P_{zz}^{\ell+m+n} \geq P_{zx}^{\ell}\, P_{xx}^n\, P_{xz}^m > 0$$

for all n sufficiently large. Part *i)* implies aperiodicity of z.

Theorem 1.25 (Convergence theorem) *Let $(X_n)_{n\geq 0}$ be an irreducible and aperiodic Markov chain with stationary distribution π. Then*

$$\lim_{n\to\infty} \mathbb{P}_x\{X_n = y\} = \pi(y) \quad \text{for all } x, y \in \mathcal{S}. \tag{1.40}$$

Remark. By the dominated convergence theorem, assertion (1.40) is equivalent to

$$\lim_{n\to\infty} d_{TV}[\mathbb{P}_\mu\{X_n \in \ \cdot\ \}, \pi] = 0$$

for every initial distribution μ. (Recall that the total variation distance between probability measures ρ_1 and ρ_2 on a discrete space $\mathcal{S}$ equals $\frac{1}{2}\sum_{x\in\mathcal{S}} |\rho_1(x) - \rho_2(x)|$.)

Proof. The idea is to *couple* the chain $(X_n)_{n\geq 0}$ with an independent chain $(Y_n)_{n\geq 0}$ which has the same transition matrix P but initial distribution π, i.e., we will follow the path of $(X_n)_{n\geq 0}$ until the two chains first meet, then follow the path of $(Y_n)_{n\geq 0}$. The exact construction is as follows.

Let $(Y_n)_{n\geq 0}$ be a stationary version of $(X_n)_{n\geq 0}$, i.e., a Markov chain with transition matrix P and initial distribution π, independent of $(X_n)_{n\geq 0}$ started at x. Set

$$X'_n := \begin{cases} X_n, & \text{if } n < T, \\ Y_n, & \text{if } n \geq T, \end{cases}$$

where

$$T := \min\{n \geq 0 : X_n = Y_n\}$$

is the so-called *coupling time*. We claim that

i) $(X_n)_{n\geq 0}$ and $(X'_n)_{n\geq 0}$ have the same distribution.

ii) The *coupling inequality* holds:

$$|\,\mathbb{P}_x\{X'_n \in B\} - \mathbb{P}_\pi\{Y_n \in B\}| \leq \mathbb{P}\{T > n\} \text{ for all } n \in \mathbb{N}_0 \text{ and } B \subset \mathcal{S}.$$

iii) $\mathbb{P}\{T < \infty\} = 1$.

Once we have proved $i), -, iii)$ we are done since then

$$|\,\mathbb{P}_x\{X_n = y\} - \pi(y)| \overset{i)}{=} |\,\mathbb{P}_x\{X'_n = y\} - \mathbb{P}_\pi\{Y_n = y\}| \overset{ii)}{\leq} \mathbb{P}\{T > n\} \xrightarrow{iii)} 0 \text{ as } n \to \infty.$$

Claim $i)$ is intuitively clear, here comes a formal proof. For $n < k \leq \infty$ we have

$$\mathbb{P}\{X'_0 = x_0, \ldots, X'_n = x_n,\, T = k\} = \mathbb{P}_x\{X_0 = x_0, \ldots, X_n = x_n,\, T = k\} \tag{1.41}$$

and for $k \leq n$ (using independence of (X_n) and (Y_n) and the Markov property)

$$\begin{aligned} &\mathbb{P}_x\{X'_0 = x_0, \ldots, X'_n = x_n,\, T = k\} \\ &\quad= \mathbb{P}\{X_0 = x_0, \ldots, X_k = x_k, Y_0 \neq x_0, \ldots, Y_{k-1} \neq x_{k-1}, Y_k = x_k, \ldots, Y_n = x_n\} \\ &\quad= \mathbb{P}_x\{X_0 = x_0, \ldots, X_k = x_k\}\,\mathbb{P}\{Y_{k+1} = x_{k+1}, \ldots, Y_n = x_n \,|\, Y_k = x_k\} \\ &\qquad\qquad \cdot \mathbb{P}_\pi\{Y_0 \neq x_0, \ldots, Y_{k-1} \neq x_{k-1}, Y_k = x_k\} \\ &\quad= \mathbb{P}\{X_0 = x_0, \ldots, X_n = x_n, Y_0 \neq x_0, \ldots, Y_{k-1} \neq x_{k-1}, Y_k = x_k\} \\ &\quad= \mathbb{P}_x\{X_0 = x_0, \ldots, X_n = x_n,\, T = k\}. \end{aligned} \tag{1.42}$$

Summing (1.41) and (1.42) over k gives the desired distributional identity.
For $ii)$ observe that

$$\begin{aligned} |\,\mathbb{P}_x\{X'_n \in B\} - \mathbb{P}_\pi\{Y_n \in B\}| &= |\,\mathbb{P}_x\{X'_n \in B,\, T \leq n\} + \mathbb{P}_x\{X'_n \in B,\, T > n\} \\ &\qquad - \mathbb{P}\{Y_n \in B,\, T \leq n\} - \mathbb{P}\{Y_n \in B,\, T > n\}| \\ &= |\,\mathbb{P}_x\{X'_n \in B,\, T > n\} - \mathbb{P}\{Y_n \in B,\, T > n\}| \\ &\leq \mathbb{P}\{T > n\}. \end{aligned}$$

For $iii)$ observe that the pair (X_n, Y_n) is a Markov chain on $\mathcal{S} \times \mathcal{S}$ with transition matrix

$$\tilde{P}_{(x,y)(x',y')} = P_{xx'}P_{yy'}, \quad x, x', y, y' \in \mathcal{S}$$

and stationary distribution (verify !)

$$\tilde{\pi}(x,y) = \pi(x)\pi(y), \quad x,y \in \mathcal{S}.$$

We claim that $\tilde{P}$ is irreducible. Indeed, by the assumed irreducibility of P we have $P^{\ell}_{xx'} > 0$ and $P^{m}_{yy'} > 0$ for some $\ell, m \geq 0$. By aperiodicity of P, we have $P^{n}_{x'x'}, P^{n}_{y'y'} > 0$ for all n sufficiently large. Hence, for all k sufficiently large

$$\tilde{P}^{k}_{(x,y)(x'y')} = P^{k}_{xx'}P^{k}_{yy'} \geq P^{\ell}_{xx'}P^{k-\ell}_{x'x'}P^{m}_{yy'}P^{k-m}_{y'y'} > 0.$$

Since $\tilde{P}$ is irreducible and has a stationary distribution, it is (positive) recurrent (by Corollary 1.20). Hence (recall Lemma 1.13),

$$T = \min\{n \geq 0 : X_n = Y_n\} \leq \min\{n \geq 0 : (X_n, Y_n) = (z,z)\} \overset{a.s.}{<} \infty \text{ for all } z \in \mathcal{S}.$$

1.6 Optimal stopping

Let $(X_n)_{n\geq 0}$ be a Markov chain on $\mathcal{S}$ with transition matrix P and let $f : \mathcal{S} \to \mathbb{R}^{+}_{0}$. We think of $f(x)$ as the payoff when the chain is being stopped at x. Our aim is to find an optimal strategy. To this end we have to agree on a notion of optimality and on the set of allowed strategies. The function

$$v(x) := \sup_{T} \mathbb{E}_x f(X_T), \quad x \in \mathcal{S},$$

where the supremum extends over all finite (!) stopping times T, is called the *value*. Note that $\mathbb{E}_x f(X_T)$ is the expected payoff when starting the chain at x and using "strategy" T. Natural questions that come to mind are

- What is v in terms of f and P ?
- What is an optimal strategy ?

By the Markov property and since no costs are involved to run the chain, it seems intuitively clear, that if an optimal strategy T exists, then it should be of the form $T = \tau_B$ for some $B \subset \mathcal{S}$. However, this will not be assumed in the sequel.

Theorem 1.26

i) The value v is the smallest superharmonic majorant of the payoff function f.

ii) $v = \max(f, Pv)$.

iii) $v = \lim_{n\to\infty} v_n$, where $v_0 := f$ and $v_{n+1} := \max(v_n, Pv_n)$, $n \geq 0$.

Proof. *i)* Since $T \equiv 0$ is a stopping time, we have $v \geq f$. To show that the value v is a superharmonic function let T_k, $k \geq 1$ be a sequence of finite stopping times such that for every $x \in \mathcal{S}$

$$\mathbb{E}_x f(X_{T_k}) \uparrow v(x) \ \text{ as } k \to \infty.$$

Define the stopping time $\tilde{T}_k$ through

$$\tilde{T}_k(X_0, X_1, X_2, \ldots) := 1 + T_k(X_1, X_2, \ldots).$$

Intuitively, the strategy $\tilde{T}_k$ is to make one step and then stop according to rule T_k. Decomposing the expected payoff according to the first step of the chain and using the strong Markov property and the monotone convergence theorem we get

$$\begin{aligned} v(x) &\geq \mathbb{E}_x f(X_{\tilde{T}_k}) \\ &= \sum_{y \in \mathcal{S}} P_{xy}\, \mathbb{E}_x(f(X_{\tilde{T}_k}) \,|\, X_1 = y) \\ &= \sum_{y \in \mathcal{S}} P_{xy}\, \mathbb{E}_y f(X_{T_k}) \\ &\to \sum_{y \in \mathcal{S}} P_{xy}\, v(y) \;=\; Pv(x). \end{aligned} \tag{1.43}$$

We now show that v is the smallest superharmonic function: Suppose $g \geq f$ and $g \geq Pg$. Then, for every $x \in \mathcal{S}$ and every finite stopping time T,

$$g(x) = \mathbb{E}_x g(X_0) \geq \mathbb{E}_x g(X_T) \geq \mathbb{E}_x f(X_T),$$

where for the first inequality we have used the stopping theorem for supermartingales (to come in Chapter 5). Taking the supremum over all finite stopping times T gives $g \geq v$.

iii) Define

$$\tilde{v} := \lim_{n\to\infty} v_n.$$

To prove that $\tilde{v} = v$ we will show that $\tilde{v}$ is the smallest superharmonic majorant of v. Assertion *iii*) then follows by part *i*).

By the monotone convergence theorem and definition of $\tilde{v}$ we have

$$P\tilde{v} = \lim_{n\to\infty} Pv_n \leq \lim_{n\to\infty} v_{n+1} = \tilde{v}.$$

Now suppose that $g \geq f$ and $g \geq Pg$. We claim that

$$g \geq v_n, \quad n \geq 0. \tag{1.44}$$

Inequality (1.44) is obvious for $n = 0$. Suppose that it holds for n, then

$$Pv_n \leq Pg \leq g,$$

which implies

$$v_{n+1} = \max(v_n, Pv_n) \leq g.$$

ii) We will show

$$v_{n+1} = \max(f, Pv_n), \; n \geq 0. \tag{1.45}$$

The assertion then follows by letting n tend to ∞. We prove (1.45) by induction. First note that, by definition,

$$v_1 = \max(v_0, Pv_0) = \max(f, Pv_0).$$

Now suppose (1.45) holds for n, then

$$\begin{aligned} v_{n+1} &= \max(v_n, Pv_n) \\ &= \max(f, Pv_{n-1}, Pv_n) \;=\; \max(f, Pv_n). \end{aligned}$$

Remark. It can be shown (using the stopping theorem) that $\tau_{\{v=f\}}$ is the *smallest* optimal stopping time provided that one of the following (equivalent) conditions hold:

i) $\mathbb{P}\{\tau_{\{v=f\}} < \infty\} = 1$.

ii) $v(x) = \mathbb{E}_x f(X_T)$ for some finite stopping time T.

1.7 Exercises

Exercise 1.1 *Let Y_i, $i \geq 1$ be i.i.d. random variables with values in $\mathbb{R}$. Define*

$$\bar{X}_0 = 0,\ \bar{X}_n := Y_1 + \cdots + Y_n,\ n \geq 1$$

and

$$\tilde{X}_0 = 0,\ \tilde{X}_1 = Y_1 \text{ and } \tilde{X}_n := Y_{n-1} + Y_n,\ n \geq 2.$$

Which of the processes $(\bar{X}_n)_{n\geq 0}$ and $(\tilde{X}_n)_{n\geq 0}$ is a Markov chain?

Exercise 1.2 *A random variable T taking values in $\mathbb{N}_0 \cup \{\infty\}$ is called a* stopping time *for the Markov chain $(X_n)_{n\geq 0}$, if*

$$\{X_0 = x_0, \ldots, X_n = x_n\} \subset \{T = n\} \quad \text{or} \quad \{X_0 = x_0, \ldots, X_n = x_n\} \subset \{T \neq n\}$$

for every $n \in \mathbb{N}_0$ and $x_0, \ldots, x_n \in \mathcal{S}$.

a) Show that the first hitting time of the set $B \subset \mathcal{S}$,

$$\tau_B \;:=\; \min\{\, n \geq 0 : X_n \in B \,\}$$

is a stopping time.

b) ***Strong Markov property.*** *Show that*

$$\begin{aligned} &\mathbb{P}_\mu\{(X_T, \ldots, X_{T+n}) \in B \,|\, T < \infty,\, X_T = x,\, (X_0, \ldots, X_{T-1}) \in B'\} \\ &\quad = \;\mathbb{P}_x\{(X_0, \ldots, X_n) \in B\} \end{aligned}$$

for every $n \in \mathbb{N}_0$, $x \in \mathcal{S}$, $B \subset \mathcal{S}^{n+1}$ and $B' \subset \bigcup_{k=0}^{\infty} \mathcal{S}^k$.

Exercise 1.3 *Let $(X_n)_{n\geq 0}$ be a Markov chain with state space $\mathcal{S}$ and transition matrix P. Set $G_n(x,y) := \sum_{k=0}^{n} P_{xy}^k$. Show that*

$$G_n(x,y) \;\leq\; G_n(y,y) \quad \text{for all } n \in \mathbb{N}_0 \text{ and } x,y \in \mathcal{S}.$$

Hint: Observe that $G_n(x,y)$ is the expected number of visits of the chain in y until time n when started at x.

Exercise 1.4 *Let $(X_n)_{n\geq 0}$ be simple random walk on $\mathbb{Z}$ with drift (i.e., $p := P_{x,x+1} = 1 - P_{x,x-1} \neq \frac{1}{2}$.) Calculate*

$$\mathbb{P}_x\{\tau_y < \tau_0\} \quad \text{and} \quad \mathbb{E}_x \min(\tau_0, \tau_y) \quad \text{for } 0 \leq x \leq y.$$

Exercise 1.5 *Show that simple symmetric random walk on $\mathbb{Z}^2$ is recurrent.*

Exercise 1.6 *Let T be the infinite rooted d-ary tree (i.e., T is the infinite graph where all vertices have degree $d+1$, except the root which has degree d). At each time a random walker on the vertex set of T independently jumps from the present vertex to any of the adjacent vertices equally likely. Show that for $d \geq 2$ the random walk on T is transient.*

Hint: Use the first part of Exercise 1.4.

Exercise 1.7 ***Downward excursions on the way up.*** *Let $(X_n)_{n\geq 0}$ be simple symmetric random walk on $\mathbb{Z}$. The maximum of the random walk until time n is defined as $M_n := \max_{0\leq k\leq n} X_k$ and the downward excursion depth at n as $D_n := M_n - X_n$. Show that, for $z \in \mathbb{Z}$,*

$$\mathbb{P}_0\Big\{ \max_{0\leq k\leq \sigma_z} Y_k < y \Big\} \;=\; \left(\frac{y}{1+y}\right)^z, \; y = 1, 2, \ldots$$

where τ_z denotes the first hitting time of z.

Exercise 1.8 ***Time reversal.*** *Let $(X_n)_{n\geq 0}$ be a Markov chain with state space $\mathcal{S}$, transition matrix P, and stationary distribution π. Define the matrix Q by*

$$Q_{xy} \;:=\; \frac{\pi(y)P_{yx}}{\pi(x)}.$$

a) Show that Q is the transition matrix of a Markov chain with stationary distribution π.
b) Show that Q is the transition matrix, corresponding to the time-reversed chain, i.e.

$$P_\pi\{X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n\} \;=\; \pi(x_n) Q_{x_n x_{n-1}} \cdots Q_{x_1 x_0}.$$

Exercise 1.9 *The **Galton-Watson process** is a basic stochastic model for the growth of a population. It is assumed that each individual independently has k children with probability p_k ($p_k \geq 0$ and $\sum_{k=0}^{\infty} p_k = 1$). The generation sizes Z_n can be recursively defined as*

$$Z_n := \sum_{i=1}^{Z_{n-1}} X_{n,i}, \; n \geq 1,$$

where the $X_{n,i}$, $n, i \geq 1$ are independent random variables with the offspring distribution $(p_k)_{k\geq 0}$, independent of the number of founding ancestors Z_0.

a) Show that $(Z_n)_{n\geq 0}$ is a Markov chain on $\mathbb{N}_0$ with transition probabilities

$$P_{xy} = \sum_{\substack{y_1,\ldots,y_x \in \mathbb{N}_0:\\ y_1+\cdots+y_x=y}} p_{y_1}\cdots p_{y_x}.$$

b) Show that, if $p_1 \neq 1$ then all states $z \neq 0$ are transient.

c) Prove that, if $p_1 \neq 1$ then

$$\mathbb{P}_z\{Z_n \to 0\} + \mathbb{P}_z\{Z_n \to \infty\} = 1 \text{ for all } z \in \mathbb{N}.$$

Exercise 1.10 *Let $y \neq z$ be states from the same recurrent component of the state space $\mathcal{S}$. Recall that m_z was defined as*

$$m_z(y) \;:=\; \mathbb{E}_z\Big(\sum_{n=1}^{\sigma_z} I_{\{X_n=y\}}\Big) \quad \text{for all } y \in \mathcal{S},$$

where σ_z is the time of the first return of the chain $(X_n)_{n\geq 0}$ to z. Show that

$$m_z(y) \;=\; \frac{\mathbb{P}_z\{\tau_y < \sigma_z\}}{\mathbb{P}_y\{\tau_z < \sigma_y\}}.$$

Exercise 1.11 *The **Metropolis algorithm** is a recipe for the construction of a Markov chain that has a given probability distribution $\pi > 0$ on $\mathcal{S}$ as its stationary distribution: Let Q be an irreducible stochastic matrix on $\mathcal{S}$ and define*

$$\begin{aligned} P_{xy} &= \min\Big(1, \frac{\pi(y)\,Q_{yx}}{\pi(x)\,Q_{xy}}\Big)\,Q_{xy}, \quad x \neq y \in \mathcal{S};\\ P_{xx} &= 1 - \sum_{y\neq x} P_{xy}, \quad x \in \mathcal{S}. \end{aligned}$$

a) Show that P is a stochastic matrix and that a Markov chain with transition matrix P is reversible with respect to π. Can you think of a two-step random mechanism that generates transitions according to P?

b) Recall that a Markov chain converges to its unique stationary distribution, if it is aperiodic and irreducible. Show that if Q is not irreducible, then P is not irreducible either. Is the converse true?

c) For $m, n \in \mathbb{N}$ let $\mathcal{S} = \{(x_1,\ldots,x_m) \in \{0,\ldots,20\}^m \mid \sum_{i=1}^m x_i = n\}$. We want to generate a uniform sample from $\mathcal{S}$, however, we do not know the exact size of $\mathcal{S}$. How can we use the Metropolis algorithm here? Propose a stochastic matrix Q and check aperiodicity and irreducibility of the corresponding matrix P.

d) Write a computer program to implement the method from part c).

Exercise 1.12 *Let $(X_n)_{n\in\mathbb{N}}$ be the symmetric random walk on $\{0,1,\ldots,N\}$ with reflecting boundaries, i.e. transition probabilities*

$$P_{0,1} = P_{N,N-1} = 1$$

and

$$P_{k,k+1} = P_{k,k-1} = \frac{1}{2} \quad \text{for every } k \in \{1,\dots,N-1\}.$$

a) Is the chain irreducible?
b) Is the chain aperiodic?
c) Find a stationary distribution for $(X_n)_{n\geq 0}$.

Exercise 1.13 ***The Propp-Wilson algorithm.*** *Let* P *be a stochastic matrix on the finite set* $\mathcal{S}$ *with stationary distribution* π *and let* F *be a random mapping from* $\mathcal{S}$ *to* $\mathcal{S}$ *with*

$$\mathbb{P}\{F(x) = y\} = P_{xy} \quad \text{for all } x, y \in \mathcal{S}.$$

a) Show that if X *is a random variable with values in* $\mathcal{S}$ *and distribution* π*, independent of* F*, then* X *and* $F(X)$ *have the same distribution.*

b) Define

$$G_n := F_1 \circ F_2 \circ \cdots \circ F_n,\ n \geq 1,$$

where the F_k, $k \geq 1$ *are independent copies of the random mapping* F*. Let*

$$T := \min\{n \geq 1 \colon |G_n(\mathcal{S})| = 1\}.$$

be the so-called backwards coupling time.

Show that if $\mathbb{P}\{T < \infty\} = 1$*, then* π *is unique and*

$$\mathbb{P}\{G_T(\mathcal{S}) = \{y\}\} = \pi(y),\ \forall\ y \in \mathcal{S}.$$

(Hint: Consider the sequence $G_n(X)$ *and observe that* $G_{T+j} = G_T$ *for all* $j \geq 0$*.)*

c) Suppose that P *is irreducible and aperiodic. Show that if the random variables* $F(x)$, $x \in \mathcal{S}$ *are independent, then* $\mathbb{P}\{T < \infty\} = 1$*.*

Exercise 1.14 *Suppose that* P *is a stochastic matrix on the state space* $\mathcal{S}$ *with stationary distribution* π*. Show that the Markov chain with transition matrix*

$$\alpha P + (1-\alpha)\mathrm{Id}_{\mathcal{S}},\ 0 < \alpha < 1$$

is an aperiodic chain with stationary distribution π*.*

Exercise 1.15 *Show that any Markov chain on a finite state space* $\mathcal{S}$ *has at least one stationary distribution.*

Exercise 1.16 *Let* $(X_n)_{n\geq 0}$ *be a random walk on* $\mathbb{Z}$ *with increments*

$$X_n - X_{n-1} = Y_n,\ n \geq 1,$$

where the Y_j, $j \geq 1$ *are i.i.d. random variables with values in* $\mathbb{Z}$*. For* n *fixed the* ***dual random walk*** $(X'_k)_{0\leq k\leq n}$ *is defined as*

$$X'_k := X_n - X_{n-k},\ 0 \leq k \leq n.$$

a) *Show that*

$$(X'_1, \ldots, X'_n) \overset{d}{=} (X_1 - X_0, X_2 - X_0, \ldots, X_n - X_0).$$

b) *Show that for* $x \in \mathbb{N}_0$, $y \in \mathbb{N}$ *and* $n \geq 2$

$$\begin{aligned} &\mathbb{P}_0\{\max_{1 \leq j \leq n-2} X_j < 0,\, X_{n-1} = -y,\, X_n = x\} \\ &= \mathbb{P}_0\{X_1 = x + y,\, \min_{2 \leq j \leq n-1} X_j > x,\, X_n = x\}. \end{aligned}$$

c) *Suppose* $\mathbb{P}\{Y_1 \geq -1\} = 1$ *and let* $\sigma := \min\{n \geq 1 : X_n \geq 0\}$. *Use part b) to conclude that*

$$\mathbb{P}_0\{\sigma < \infty,\, X_{\sigma-1} = -y,\, X_\sigma = x\} = \mathbb{P}\{Y_1 = x + y\}\, \mathbb{P}_y\{\tau_0 < \infty\}$$

for $x \in \mathbb{N}_0$, $y \in \mathbb{N}$.

Exercise 1.17 *Let* $\mathcal{S} = \{0,1\}^3$ *and for* $x = (x^{(1)}, x^{(2)}, x^{(3)}) \in \mathcal{S}$ *let* $f(x) = \sum_{i=1}^3 x^{(i)}$ *be the sum of its components. Consider the following stochastic dynamics on* $\mathcal{S}$: *Given* x *choose* I *uniformly at random from* $\{1,2,3\}$ *and flip the* I*th component of* x.

a) *Let* $(X_n)_{n \geq 0}$ *be a Markov chain following this dynamics. Give an argument that the process* $Y_n = f(X_n)$, $n \geq 0$ *is a Markov chain on* $\mathcal{S}_0 = \{0,1,2,3\}$.

b) *Compute the transition matrix* P *of* $(Y_n)_{n \geq 0}$ *and its stationary distribution* π.

c) *Show that the chain with transition matrix* $Q = \frac{1}{2}(P + I)$ *satisfies the conditions of the convergence theorem.*

Exercise 1.18 *Let* ρ, ρ', ρ'' *and* $\rho_n, n \geq 1$ *be probability measures on a discrete space* $\mathcal{S}$. *Recall that the total variation distance between* ρ *and* ρ' *is defined as*

$$d_{TV}(\rho, \rho') := \sup_{B \subset \mathcal{S}} |\rho(B) - \rho'(B)|.$$

a) *Prove the* Δ*-inequality:* $d_{TV}(\rho, \rho'') \leq d_{TV}(\rho, \rho') + d_{TV}(\rho', \rho'')$.

b) *Show that the supremum in the definition of* d_{TV} *is attained and that*

$$d_{TV}(\rho, \rho') = \frac{1}{2} \sum_{x \in \mathcal{S}} |\rho(x) - \rho'(x)|.$$

c) *Show that*

$$\lim_{n \to \infty} \rho_n(x) = \rho(x) \text{ for all } x \in \mathcal{S} \implies \lim_{n \to \infty} d_{TV}(\rho_n, \rho) = 0.$$

2 Renewal processes

2.1 Limit theorems

Let T_j, $j \geq 1$ be non-negative i.i.d. random variables with values in $\overline{\mathbb{R}}_0^+$ and let T_0 be a $\mathbb{R}_0^+$-valued random variable, independent of $(T_j)_{j\geq 1}$. Set $S_0 := 0$ and

$$S_{n+1} := T_0 + \sum_{j=1}^{n} T_j, \quad n \geq 0. \tag{2.1}$$

The times S_n, $n \geq 1$ are called *renewal points*. The number of renewals until time t is denoted by

$$N_t := \max\{n \geq 0 : S_n \leq t\},\ t \geq 0.$$

Note the relation

$$\{N_t \geq n\} = \{S_n \leq t\} \text{ for all } n \in \mathbb{N} \text{ and } t \geq 0.$$

Definition 2.1 *The process* $(Y_t)_{t\geq 0}$ *with*

$$Y_t := S_{N_t+1} - t, \quad t \geq 0,$$

is called a renewal process *with lifetime distribution* $\nu := \mathcal{L}(T_1)$ *and delay* T_0.

$(Y_t)_{t\geq 0}$ is the continuous-time analogue to the renewal chain studied in Chapter 1. Traditionally, the process $(N_t)_{t\geq 0}$ is called renewal process. Note that in general the latter process is not Markovian.

Proposition 2.2 (Law of large numbers) *Suppose* $\mu := \mathbb{E}\, T_1 \in (0, \infty]$, *then*

$$\lim_{t\to\infty} \frac{N_t}{t} = \frac{1}{\mu} \ \text{a.s.} \tag{2.2}$$

Proof. If $\mathbb{P}\{T_1 = \infty\} = \theta > 0$ then $N_t \uparrow N_\infty < \infty$ a.s. as $t \to \infty$, where

$$\mathbb{P}\{N_\infty = k\} = \theta(1-\theta)^{k-1}, \quad k \in \mathbb{N}.$$

Hence,

$$\lim_{t\to\infty} \frac{N_t}{t} = 0 = \frac{1}{\mu} \ \text{a.s.}$$

Now assume $\mathbb{P}\{T_1 = \infty\} = 0$. By definition of N_t, we have $S_{N_t} \leq t < S_{N_t+1}$ for all $t \geq 0$. Division by N_t gives

$$\frac{N_t - 1}{N_t} \frac{S_{N_t}}{N_t - 1} \leq \frac{t}{N_t} < \frac{S_{N_t+1}}{N_t}. \tag{2.3}$$

Now observe that

i) $\dfrac{S_{n+1}}{n} = \dfrac{T_0}{n} + \dfrac{\sum_{j=1}^{n} T_j}{n} \overset{a.s.}{\to} \mu$ as $n \to \infty$ (by the standard LLN);

ii) $N_t \overset{a.s.}{\to} \infty$ (since $S_n \overset{a.s.}{<} \infty$ for all n).

Asymptotics *i)* and *ii)* show that when passing to the limit $t \to \infty$ in (2.3), both the lower and upper bound converge to μ a.s.

The function

$$u(t) := \mathbb{E}N_t = \sum_{n=1}^{\infty} \mathbb{P}\{N_t \geq n\} = \sum_{n=1}^{\infty} \mathbb{P}\{S_n \leq t\}$$

is the so-called *renewal function*. The following theorem shows that we may interchange expectation and limiting procedures in (2.2).

Theorem 2.3 (Elementary renewal theorem) *Suppose that* $\mu \in (0, \infty]$, *then*

$$\lim_{t\to\infty} \frac{u(t)}{t} = \frac{1}{\mu}.$$

To prove Theorem 2.3 we need the following result which is interesting in its own.

Lemma 2.4 (Wald's Identity) *Let* $Z_1, Z_2, Z_3, \ldots$ *be i.i.d. random variables with* $\mathbb{E}|Z_1| < \infty$. *Let* τ *be a stopping time for* $(Z_n)_{n\geq 0}$ *with* $\mathbb{E}\,\tau < \infty$. *Then*

$$\mathbb{E}\left(\sum_{i=1}^{\tau} Z_i\right) = \mathbb{E}\,\tau\,\mathbb{E}Z_1. \tag{2.4}$$

Relation (2.4) states that the mean of the random sum $\sum_{i=1}^{\tau} Z_i$ is the same as if the Z_i and τ were independent.

Proof. Clearly,

$$\sum_{i=1}^{\tau} Z_i = \sum_{i=1}^{\infty} Z_i\, I\{i \leq \tau\}.$$

Note that

$$I\{i \leq \tau\} \;=\; 1 - I\{\tau \leq i-1\} \;=\; 1 - \sum_{j=0}^{i-1} I\{\tau = j\}. \tag{2.5}$$

Identity (2.5) and the fact that τ is a stopping time for (Z_n) show that the random variables Z_i and $I\{i \leq \tau\}$ are independent. Thus (compare (1.17))

$$\begin{aligned} \mathbb{E}\left(\sum_{i=1}^{\tau} Z_i\right) &= \sum_{i=1}^{\infty} \mathbb{E}(Z_i\, I\{i \leq \tau\}) \\ &= \sum_{i=1}^{\infty} \mathbb{E}Z_i\, \mathbb{P}\{\tau \geq i\} \\ &= \mathbb{E}Z_1 \sum_{i=1}^{\infty} \mathbb{P}\{\tau \geq i\} \;=\; \mathbb{E}Z_1\, \mathbb{E}\,\tau. \end{aligned}$$

Proof of Thm. 2.3. We may assume $\mu < \infty$ (else $u(t) \uparrow u(\infty) = \frac{1}{\theta} < \infty$). We first prove the lower bound

$$\liminf_{t\to\infty} \frac{u(t)}{t} \geq \frac{1}{\mu}. \tag{2.6}$$

Since

$$\{N_t = n\} = \{S_n \leq t,\, S_{n+1} > t\} = \left\{ \sum_{j=0}^{n-1} T_j \leq t < \sum_{j=0}^{n} T_j \right\}$$

depends on $T_0, \ldots, T_n$ only, for each $t \geq 0$ the random variable N_t is a stopping time for $(T_k)_{k \geq 0}$. Using Wald's identity we get

$$t \;\leq\; \mathbb{E} S_{N_t+1} = \mathbb{E}\bigg(T_0 + \sum_{j=1}^{N_t} T_j\bigg) = \mathbb{E}\, T_0 + u(t)\mu. \tag{2.7}$$

Consequently,

$$\frac{u(t)}{t} \geq \frac{1}{\mu} - \frac{\mathbb{E}\, T_0}{\mu t}$$

which gives (2.6) provided that $\mathbb{E}\, T_0 < \infty$.

In general (i.e., possibly $\mathbb{E}\, T_0 = \infty$), we have

$$u(t) \geq \mathbb{E}(N_t\, I\{T_0 \leq x\}) = \mathbb{E}(N_t \,|\, T_0 \leq x)\, \mathbb{P}\{T_0 \leq x\}.$$

Given $\{T_0 \leq x\}$ the process $(Y_t)_{t \geq 0}$ is a renewal process with the same lifetime distribution and a delay with mean $\mathbb{E}(T_0 \,|\, T_0 \leq x) \leq x$. Hence, the first part gives

$$\liminf_{t \to \infty} \frac{u(t)}{t} \geq \mathbb{P}\{T_0 \leq x\}\, \frac{1}{\mu}. \tag{2.8}$$

Passing to the limit $x \to \infty$ in (2.8) gives (2.6).

For the upper bound Wald's identity is not directly applicable since $N_t - 1$ is not a stopping time for $(T_k)_{k \geq 0}$. Instead, we will use the inequality

$$t + T_{N_t} \geq S_{N_t+1}.$$

In general, the mean $\mathbb{E} T_{N_t}$ cannot be bounded in terms of μ. To circumvent this difficulty we use a truncation procedure. For $x \geq 0$ let

$$T_j^x := T_j \wedge x,\ \ j \geq 0$$

and define S_j^x, N_t^x, $u_x(t)$ and μ_x in the obvious way. Since the maximal overshoot is bounded by x, we have

$$\mathbb{E} S^x_{N_t^x+1} \leq t + x,$$

Hence (recall (2.7)),

$$t + x \geq \mathbb{E} T_0^x + u_x(t)\, \mathbb{E} T_1^x \geq u_x(t)\, \mu_x$$

so that

$$\limsup_{t \to \infty} \frac{u_x(t)}{t} \leq \frac{1}{\mu_x}.$$

Since $S_n^x \leq S_n$ a.s. we have $N_t \leq N_t^x$ a.s., which implies

$$\limsup_{t \to \infty} \frac{u(t)}{t} \leq \limsup_{x \to \infty} \limsup_{t \to \infty} \frac{u_x(t)}{t} \leq \limsup_{x \to \infty} \frac{1}{\mu_x} = \frac{1}{\mu}. \tag{2.9}$$

2.2 Stationary renewal processes

This section deals with the question when the renewal process $(Y_t)_{t\geq 0}$ is a stationary process (i.e., for what $\mathcal{L}(T_0)$). By the Markov property, for $(Y_t)_{t\geq 0}$ to be stationary it is sufficient that $\mathcal{L}(Y_t)$ does not depend on t. Now recall from Section 1.5 that if $(X_n)_{n\geq 0}$ is an aperiodic, irreducible and positive recurrent Markov chain then

$$\mathbb{P}_x\{X_n = y\} \to \pi(y) \text{ as } n \to \infty \text{ for all } x, y \in \mathcal{S},$$

where π is the unique stationary distribution of $(X_n)_{n\geq 0}$. Hence, our opening question boils down to finding the law of $Y_t = S_{N_t+1} - t$ as $t \to \infty$. Let

$$A_t := t - S_{N_t}$$

be the age of the item in use at time t. The first guess (?!) that the total lifetime $T_{N_t} = A_t + Y_t$ of the item in use at t has asymptotic distribution $\nu = \mathcal{L}(T_1)$ is wrong. This is because of the so-called *size-biasing effect*: It is more likely that t falls in a large renewal interval than that it is covered by a small one.

To explain things in a simple setting we first have a look at stationary renewal chains (i.e., we take time to be discrete). Consider the augmented process $(Y_n, Z_n)_{n\geq 0}$, where Z_n is the total lifetime of the item in use at n. Observe that $(Y_n, Z_n)_{n\geq 0}$ is a Markov chain on the state space $\mathcal{S} = \{(y,z) \in \mathbb{N}_0^2 : 0 \leq y < z < \infty,\ q_z > 0\}$ with transition matrix

$$P_{(y,z)(y',z')} = \begin{cases} 1, & \text{if } y' = y-1,\ z' = z, \\ q_{z'}, & \text{if } y = 0,\ y' = z'-1, \\ 0, & \text{else.} \end{cases}$$

Note that the chain $(Y_n, Z_n)_{n\geq 0}$ is irreducible, since $(y,z) \to (0,z) \to (z'-1, z') \to (y', z')$.

Proposition 2.5 *Suppose* $0 < \mu := \mathbb{E}T_1 = \sum_{k=1}^{\infty} k q_k < \infty$. *Then the stationary distribution of the Markov chain* $(Y_n, Z_n)_{n\geq 0}$ *is the distribution of the pair* $(\lfloor U\widehat{T}_1\rfloor, \widehat{T}_1)$, *where* $\widehat{T}_1$ *has the size-biased distribution of* T_1,

$$\mathbb{P}\{\widehat{T}_1 = y\} = \frac{y\,\mathbb{P}\{T_1 = y\}}{\mu},\ \ y \in \mathbb{N}_0,$$

and U *is uniformly distributed on the interval* $[0,1]$, *independent of* $\widehat{T}_1$.

Remark. Note that $\mathbb{P}\{\widehat{T}_1 \geq 1\} = 1$ and that $\mathbb{E}\widehat{T}_1 = \mathbb{E}T_1^2/\mathbb{E}T_1 \geq \mathbb{E}T_1$ by Jensen's inequality. In fact, $\widehat{T}_1$ is stochastically larger than T_1.

Proof. We first compute the weights of the distribution $(\lfloor U\widehat{T}_1\rfloor, \widehat{T}_1)$ and then verify stationarity. For $0 \leq y < z < \infty$ with $q_z > 0$ we have

$$\mathbb{P}\{\lfloor U\widehat{T}_1\rfloor = y \,|\, \widehat{T}_1 = z\} \ = \ \mathbb{P}\{\lfloor Uz\rfloor = y\} \ = \ \mathbb{P}\{y \leq Uz < y+1\} \ = \ \frac{1}{z}. \tag{2.10}$$

In other words, given $\widehat{T}_1 = z$ the random variable $\lfloor U\widehat{T}_1 \rfloor$ is uniformly distributed on the set $\{0, \ldots, z-1\}$. Hence,

$$\begin{aligned}\pi(y,z) &:= \mathbb{P}\{(\lfloor U\widehat{T}_1\rfloor, \widehat{T}_1) = (y,z)\} \\ &= \mathbb{P}\{(\lfloor U\widehat{T}_1\rfloor, \widehat{T}_1) = (y,z) \,|\, \widehat{T}_1 = z\}\, \mathbb{P}\{\widehat{T}_1 = z\}.\end{aligned}$$

For π being a stationary measure it must satisfy the balance equations

$$\pi(z, z+1) = \sum_{x=1}^{\infty} \pi(0,x)\, q_{z+1}, \quad z \in \mathbb{N}_0 \tag{2.11}$$

and

$$\pi(y,z) = \pi(y+1, z), \quad 0 \le y < z-1. \tag{2.12}$$

The validity of (2.12) is obvious. For (2.11) observe that

$$\pi(z, z+1) = \frac{q_{z+1}}{\mu} = q_{z+1} \sum_{x=1}^{\infty} \frac{q_x}{\mu} = \sum_{x=1}^{\infty} \pi(0,x) q_{z+1}.$$

Remark. Note that the projection onto the first coordinate (the residual lifetime) gives

$$\pi^*(y) := \sum_{z=y+1}^{\infty} \pi(y,z) = \frac{1}{\mu} \sum_{z=y+1}^{\infty} q_z = \frac{\mathbb{P}\{T_1 > y\}}{\mathbb{E} T_1},$$

which is consistent with our considerations in Chapter 1.

We now return to the continuous-time setting.

Lemma 2.6 *Suppose* $0 < \mu < \infty$ *and let* $\mathcal{L}(T_0)$ *have density*

$$g(t) = \frac{\mathbb{P}\{T_1 > t\}}{\mu}, \quad t \ge 0. \tag{2.13}$$

Then

i) For every $n \ge 2$, $\mathcal{L}(S_n)$ *has density*

$$g_n(t) = \frac{1}{\mu}\Big(\mathbb{P}\{S_n - S_1 \le t\} - \mathbb{P}\{S_{n+1} - S_1 \le t\}\Big), \quad t \ge 0.$$

ii)

$$u(t) = \frac{t}{\mu}, \quad t \ge 0.$$

Remarks. The density g in (2.13) is the analogue to π^* (verify that it integrates to 1 !). Note that, by *ii*),

$$\mathbb{E}(N_{t+h} - N_t) = u(t+h) - u(t) = \frac{h}{\mu}$$

does not depend on t. Obviously, this property is crucial for stationarity.

Proof. Since S_n is the sum of independent random variables at least one of which has a density, $\mathcal{L}(S_n)$ has a density, too. For $k \geq 1$ let

$$F_k(s) := \mathbb{P}\{S_{k+1} - S_1 \leq s\}, \quad s \geq 0.$$

Decomposing S_n w.r.t. T_0 $(= S_1)$ gives

$$\begin{aligned} \frac{d\,\mathbb{P}\{S_n \leq t\}}{dt} &= \int_0^t g(t-s)\, dF_{n-1}(s) \\ &= \frac{1}{\mu}\int_0^t (1 - F_1(t-s))\, dF_{n-1}(s) \\ &= \frac{1}{\mu}(F_{n-1}(t) - F_n(t)). \end{aligned}$$

For *ii*) observe that

$$\begin{aligned} u(t) &= \sum_{n=1}^{\infty} \mathbb{P}\{S_n \leq t\} \\ &= \int_0^t g(s)ds + \sum_{n=2}^{\infty}\int_0^t \frac{1}{\mu}(F_{n-1}(s) - F_n(s))ds \\ &= \frac{1}{\mu}\int_0^t \Big(1 - F_1(s) + \sum_{n=2}^{\infty}(F_{n-1}(s) - F_n(s))\Big)ds \\ &= \frac{1}{\mu}\int_0^t ds \;=\; \frac{t}{\mu}. \end{aligned}$$

Proposition 2.7 *If $\mathcal{L}(T_0)$ has density g, then $(Y_t)_{t\geq 0}$ is a stationary process.*

Proof. It suffices to show that $\mathcal{L}(Y_t)$ has density g for all $t \geq 0$. For every $y \geq 0$ we have

$$\begin{aligned} \mathbb{P}\{Y_t \geq y\} &= \sum_{n=0}^{\infty} \mathbb{P}\{Y_t \geq y,\, N_t = n\} \\ &= \sum_{n=0}^{\infty} \mathbb{P}\{S_n \leq t,\, S_{n+1} \geq t+y\} \\ &= \mathbb{P}\{T_0 \geq t+y\} + \sum_{n=1}^{\infty}\int_0^t \mathbb{P}\{S_n \in ds\}\,\mathbb{P}\{T_n \geq t-s+y\} \\ &= \int_{t+y}^{\infty} g(s)ds + \mu\int_0^t g(t-s+y)\, du(s) \\ &= \int_y^{\infty} g(s)ds \;=\; \mathbb{P}\{Y_0 \geq y\}. \end{aligned}$$

In analogy with (2.5) the density g should arise from $\mathcal{L}(T_1)$ by some form of size-biasing. Following we give a formal definition of this notion.

Definition 2.8 *Let X be a non-negative random variable with $0 < \mathbb{E}X < \infty$. The random variable $\widehat{X}$ is said to have the* size-biased distribution *of X, if*

$$\mathbb{E}h(\widehat{X}) = \frac{\mathbb{E}(Xh(X))}{\mathbb{E}X}$$

for every non-negative measurable function h.

Remark. If X takes values in $\mathbb{N}_0$, then (take $h = 1_{\{k\}}$)

$$\mathbb{P}\{\widehat{X} = k\} = \frac{k\,\mathbb{P}\{X = k\}}{\mathbb{E}X}, \quad k \in \mathbb{N}_0.$$

If $\mathcal{L}(X)$ has density f, then $\mathcal{L}(\widehat{X})$ has density

$$\hat{f}(x) = \frac{xf(x)}{\mathbb{E}X}, \quad x \geq 0.$$

Lemma 2.9 *Let $\widehat{T}_1$ and U be independent random variables, where $\widehat{T}_1$ has the size-biased distribution of T_1 and U is uniformly distributed on the interval $(0,1)$. Then $\mathcal{L}(U\widehat{T}_1)$ has density g.*

Proof.

$$\begin{aligned}
\mathbb{P}\{U\widehat{T}_1 \leq t\} &= \int_0^1 \mathbb{P}\{U\widehat{T}_1 \leq t \,|\, U = u\}\, du \\
&= \int_0^1 \mathbb{P}\{\widehat{T}_1 \leq \tfrac{t}{u}\}\, du \\
&= \frac{1}{\mu}\int_0^1 \mathbb{E}(T_1\, I\{T_1 \leq \tfrac{t}{u}\})\, du \\
&= \frac{1}{\mu}\int_0^1 \Big(\int_0^{\frac{t}{u}} s\, dF_1(s)\Big) du \\
&= \int_0^t \Big(\frac{1}{\mu}\int_r^\infty dF_1(s)\Big) dr \;=\; \int_0^t g(r)dr.
\end{aligned}$$

As in the discrete case one can consider the pair process $(Y_t, Z_t)_{t\geq 0}$, where $Z_t = S_{N_t+1} - S_{N_t}$ is the total lifetime of the item in use at t. The following two convergence theorems for renewal processes are not surprising in light of the results in the discrete time setting. Their proofs, however, are much more involved and omitted here.

Theorem 2.10 *Suppose $0 < \mu < \infty$ and that $\mathcal{L}(T_1)$ is non-lattice (i.e., $\mathbb{P}\{T_1 \in d\mathbb{N}\} < 1$ for all $d \geq 2$). Then*

$$\mathcal{L}_{(y,z)}(Y_t, Z_t) \to \mathcal{L}(U\widehat{T}_1, \widehat{T}_1) \text{ weakly as } t \to \infty$$

for every $0 \leq y \leq z < \infty$, $q_z > 0$.

Theorem 2.11 (Blackwell's renewal theorem) *Suppose* $0 < \mu < \infty$ *and that* $\mathcal{L}(T_1)$ *is non-lattice. Then*

$$u(t+h) - u(t) \to \frac{h}{\mu} \quad \text{as } t \to \infty$$

for every $h \geq 0$.

2.3 The homogeneous Poisson process on the half line

We now consider the case where the lifetimes T_j, $j \geq 1$ have exponential distribution.

Definition 2.12 *A non-negative random variable* T *is said to be exponentially distributed with parameter* $\lambda > 0$ *("Exp(λ)-distributed" for short), if*

$$\mathbb{P}\{T \geq t\} = e^{-\lambda t}, \quad t \geq 0.$$

We list some elementary properties of the exponential distribution.

1. $\mathcal{L}(T)$ has density $f(t) = \lambda e^{-\lambda t}$, $t \geq 0$.
2. T has mean $\mathbb{E}\, T = 1/\lambda$ and variance $\operatorname{Var} T = 1/\lambda^2$.
3. The Laplace transform of T is
$$\varphi(\theta) := \mathbb{E} \exp(-\theta T) = \frac{\lambda}{\lambda + \theta}, \quad \theta \geq 0.$$
4. The "lack of memory":
$$\mathbb{P}\{T \geq t + s \,|\, T \geq t\} = \mathbb{P}\{T \geq s\} \text{ for all } s, t \geq 0. \tag{2.14}$$

Property (2.14) says that knowledge of the age gives no information on the residual lifetime. This explains why in the case of exponential lifetimes the stationary density of T_0 and the lifetime distribution agree,

$$g(t) = \frac{\mathbb{P}\{T_1 > t\}}{\mu} = \lambda e^{-\lambda t} = f(t), \quad t \geq 0.$$

This is commonly referred to as the bus stop paradox: If the interarrival times are exponentially distributed then arriving at the bus stop at time t one has to wait as long as if arriving immediately after a bus has left.

Definition 2.13 *Let* T_j, $j \geq 0$ *be independent Exp(λ)-distributed random variables. Then the random set* $\Phi := \{S_n : n \geq 1\}$ *is called a* homogeneous Poisson process *on* $\mathbb{R}_0^+$ *with intensity* λ.

The "homogeneous" refers to the process $(N_t)_{t \geq 0}$ having stationary increments. The name Poisson process is due to the following.

Proposition 2.14 *Let* $N_1 := |\, \Phi \cap [0,1] \,|$ *be the number of points of* Φ *in the unit interval. Then*

i) N_1 *is Poisson distributed with mean* λ.

ii) Given $N_1 = n$, *the random variable* $(S_1, \ldots, S_n)$ *is distributed like the order statistics* $(U_{(1)}, \ldots, U_{(n)})$ *of* n *independent uniform random variables* $U_1, \ldots, U_n$.

Recall that the order statistics is the U_i in increasing order. The respective assertions hold for any interval $[a, b]$.

Definition 2.15 *A random variable X with values in the non-negative integers $\mathbb{N}_0$ is called* Poisson-distributed *with parameter $\lambda > 0$, if*

$$\mathbb{P}\{X = k\} = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \mathbb{N}_0.$$

Sometimes it is convenient to have the case $\lambda = \infty$ included: The Poisson distribution with parameter ∞ is the delta measure at ∞. We list some elementary properties of the Poisson distribution.

1. X has mean $\mathbb{E}X = \lambda$ and variance $\operatorname{Var} X = \lambda$.
2. The generating function of X is
$$\mathbb{E}s^X = e^{-\lambda(1-s)}, \ \ 0 \le s \le 1.$$
3. The Poisson approximation of the binomial distribution $B(n, p_n)$:
If $np_n \to \lambda$ as $n \to \infty$, then
$$B(n, p_n, k) = \binom{n}{k} p_n^k (1 - p_n)^{n-k} \longrightarrow e^{-\lambda} \frac{\lambda^k}{k!} \ \text{ as } n \to \infty.$$

Proof of Proposition 2.14. Let

$$B \subset \{(s_1, s_2, \ldots, s_n) : 0 \le s_1 \le s_2 \le \ldots \le s_n \le 1\}$$

be a measurable set. Write $\tilde{B}$ for the corresponding inter arrival times,

$$\tilde{B} := \{(s_1, s_2 - s_1, \ldots, s_n - s_{n-1}) : (s_1, \ldots, s_n) \in B\}.$$

Then,

$$\begin{aligned}
&\mathbb{P}\{N_1 = n, (S_1, \ldots, S_n) \in B\} \\
&\quad = \mathbb{P}((S_1, \ldots, S_n) \in B,\, S_{n+1} > 1) \\
&\quad = \mathbb{P}\Big\{(T_0, \ldots, T_{n-1}) \in \tilde{B},\, \sum_{j=0}^{n} T_j > 1\Big\} \\
&\quad = \int_{\substack{(t_0,\ldots,t_{n-1}) \in \tilde{B}, \\ t_0 + \cdots + t_n > 1}} \lambda e^{-\lambda t_0}\, \lambda e^{-\lambda t_1} \cdots \lambda e^{-\lambda t_n} dt_0 \cdots dt_n \\
&\quad = \int_{(s_1,\ldots,s_n) \in B,\, s_{n+1} > 1} \lambda^n\, \lambda e^{-\lambda s_{n+1}} ds_1 \cdots ds_{n+1} \\
&\quad = \frac{\lambda^n e^{-\lambda}}{n!}\, n! \int_B ds_1 \cdots ds_n. \qquad (2.15)
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\mathbb{P}\{(U_{(1)},\dots,U_{(n)}) \in B\} \\
&= \sum_{\rho \text{ permutation}} \mathbb{P}\{(U_{(1)},\dots,U_{(n)}) \in B,\, U_{\rho(1)} < U_{\rho(2)} < \cdots < U_{\rho(n)}\} \\
&= \sum_{\rho \text{ permutation}} \mathbb{P}\{(U_{\rho(1)},\dots,U_{\rho(n)}) \in B\} \\
&= n!\,\mathbb{P}\{(U_1,\dots,U_n) \in B\} \\
&= n! \int_B ds_1 \cdots ds_n. \qquad (2.16)
\end{aligned}$$

Combining (2.15) and (2.16) establishes the claim of the proposition.

Observe that Proposition 2.14 suggests the following two-step construction of a homogeneous Poisson point process.

1. Choose a $Poisson(\lambda)$-distributed random number N, say.
2. Generate N independent uniform random variables on $[0,1]$.

Note that this construction has an obvious extension to more general state spaces (other than the construction via lifetimes which uses the linear ordering of $\mathbb{R}$). We will get back to this construction in the next chapter.

2.4 Exercises

Exercise 2.1 (Stationary age processes) *Consider a renewal chain* $(Y_n)_{n\geq 0}$ *with lifetime distribution* $(q_k)_{k\geq 1}$ *and finite expected lifetime* $\mathbb{E}\,T_1$.

a) Compute the transition matrix of the time-reversed chain (compare Exercise 1.8 and recall that $\pi(y) = \mathbb{P}\{T_1 > y\}/\mathbb{E}T_1,\ y \in \mathbb{N}_0$*).*

b) Compute the so-called ***hazard function****: the conditional probability that the lifetime* T_1 *equals* y *given that it exceeds* y*.*

Exercise 2.2 *Let* $(Y_t)_{t\geq 0}$ *be a renewal process with renewal points* $S_{n+1} = \sum_{j=0}^{n} T_j,\ n \geq 0$ *and let* $X_1, X_2, \dots$ *be a sequence of i.i.d. random variables with finite mean* $m := \mathbb{E}X_1$. *The renewal reward process is defined as*

$$R_t := \sum_{j=1}^{N_t} X_j, \quad t \geq 0.$$

a) Let $\mu = \mathbb{E}T_1$ *be the expected lifetime. Show that* $R_t/t \overset{a.s.}{\to} m/\mu$.

b) Define $A_t := t - S_{N_t}$ *to the time at* t *since the last renewal (= age of the item in use). Show that*

$$t^{-1}\int_0^t A_s ds \ \xrightarrow{a.s.}\ \frac{\mathbb{E}T_1^2}{2\mu}.$$

Exercise 2.3 ***Patterns of discrete random variables.*** *Let* $X_1, X_2, \dots$ *be i.i.d. random variables with* $\mathbb{P}\{X_1 = j\} = p_j$, $j \in \mathbb{N}$. *Let* τ *be the first time the pattern* $j_1, \dots, j_k$ *occurs,*

$$\tau := \min\{n \geq k : X_{n-k+1} = j_1, \dots, X_n = j_k\}.$$

Suppose that $j_1 \neq j_k$. *Use the elementary renewal theorem to compute* $\mathbb{E}\,\tau$.

Exercise 2.4 *Consider a single server bank in which customers arrive at the times of a Poisson point process on* $\mathbb{R}^+$ *with rate* λ, *but only enter when the server is free. Let the service times of the customers entering the bank be i.i.d. with finite mean* m.

- *a) Argue that the times at which customers enter the bank form a renewal process.*
- *b) Compute the expected number of customers arriving while one customer is served.*
- *c) What fraction of potential customers enter the bank in the long run?*
- *d) What fraction of time is the server busy?*

Exercise 2.5 *Let* X *be a non-negative random variable with positive finite mean. Recall that a random variable* $\widehat{X}$ *is said to have the* ***size-biased*** *distribution of* X, *if*

$$\mathbb{E}h(\widehat{X}) = \frac{\mathbb{E}(Xh(X))}{\mathbb{E}X}$$

for every non-negative measurable function h.

- *a) Show that* $\mathbb{P}\{\widehat{X} \in \ \cdot\ \}$ *is a probability measure on* $(\mathbb{R}_0^+, \mathcal{B})$.
- *b) Show that* $\widehat{X}$ *is stochastically larger than* X,

$$\mathbb{P}\{\widehat{X} \geq x\} \ \geq \ \mathbb{P}\{X \geq x\} \quad \textit{for all } x \geq 0.$$

- *c) Show that if* $X_1, \dots, X_n$ *are i.i.d. random variables then*

$$\widehat{\sum_{i=1}^{n} X_i} \ \stackrel{d}{=} \ \widehat{X}_1 + X_2 + \cdots + X_n,$$

  where $\widehat{X}_1$ *is independent of* $X_2, \dots, X_n$.

- *d) Suppose that* X *takes values in the non-negative integers* $\mathbb{N}_0$. *Show that* $\widehat{Y}$ *has probability generating function (p.g.f.)*

$$\hat{f}(s) = \frac{sf'(s)}{f'(1)}, \ 0 \leq s \leq 1,$$

  where

$$f(s) := \sum_{k=0}^{\infty} \mathbb{P}(X = k)s^k, \ 0 \leq s \leq 1,$$

  is the p.g.f. of X.

e) Let X have (shifted) geometric distribution,

$$\mathbb{P}\{X = k\} = p(1-p)^k, \ k \in \mathbb{N}_0.$$

Show that

$$\widehat{X} \overset{d}{=} 1 + X_1 + X_2 \quad \text{and} \quad \lfloor U\widehat{X} \rfloor \overset{d}{=} X,$$

where X_1 and X_2 are independent copies of X, and U is uniformly distributed on the interval $[0,1]$, independent of $\widehat{X}$.

3 Poisson point processes

The Poisson point process is a fundamental model in the theory of stochastic processes. It is used to model spatial phenomena (like the spread of colonies of bacteria on a culture medium or the positions of visible stars in a patch of the sky) or a random series of events occurring in time (like the emission of radioactive particles or the arrival times of phone calls).

Formally, a (Poisson) point process is a countable random subset of some state space $\mathcal{S}$. The distribution of Φ can be described by specifying the law of the number of points of Φ falling into "test sets" B_i (where the B_i are from some σ-field $\mathcal{B}$ on $\mathcal{S}$). To this end one needs to ensure that there are enough sets in $\mathcal{B}$. This can be done by the weak assumption that the diagonal $D = \{(x,y) : x = y\}$ is measurable in the product space $(\mathcal{S} \times \mathcal{S}, \mathcal{B} \otimes \mathcal{B})$, which implies that singletons $\{x\}$ are in $\mathcal{B}$ for every $x \in \mathcal{S}$.

We will not address the measure theoretic niceties here (but be aware that there is some demand!) and only list a few important properties (for a thorough account we refer to the monographs [3, 6]).

- The set of countable subsets of a measurable set $\mathcal{S}$ is a Polish space when equipped with the vague topology (identify $\Phi = \{x_i : i \in I\}$ with the counting measure $\sum_{i \in I} \delta_{x_i}$).
- The distribution of Φ is uniquely determined by the "finite-dimensional distributions" of $(N(B_1), \ldots, N(B_n))$, where $N(B_i) := |\Phi \cap B_i|$ is the number of points that fall into the set B_i.

3.1 Construction and basic properties

Definition 3.1 *Let λ be a σ-finite (and non-atomic) measure on the measurable space $(\mathcal{S}, \mathcal{B})$. A random countable subset Φ of $\mathcal{S}$ is called a* (simple) Poisson point process *on $\mathcal{S}$ with* intensity (*or* mean) measure λ, *if*

i) The random variables $N(B_1), \ldots, N(B_n)$ are independent for any finite collection of disjoint *measurable subsets $B_1, \ldots, B_n \in \mathcal{B}$.*

ii) For every $B \in \mathcal{B}$ the random variable $N(B)$ has Poisson distribution with mean $\lambda(B)$.

Recall that a measure λ is called non-atomic if $\lambda(\{x\}) = 0$ for all $x \in \mathcal{S}$. This property is not essential for the definition of a Poisson point process, but guarantees that Φ has no multiple points. A measure λ is called σ-finite if there exist $B_1 \subset B_2 \subset \cdots \subset \mathcal{S}$ such that $\bigcup_{i \geq 1} B_i = \mathcal{S}$ and $\lambda(B_i) < \infty$ for all i, i.e., if $\mathcal{S}$ can be exhausted by sets of finite measure.

The inevitability of the Poisson distribution. We give a (somewhat heuristic) argument that the Poisson distribution is inevitable if one requires the maximal independence in condition i) of Definition 3.1. Suppose that we independently throw points into the infinitesimal volume elements dy with probability

$$\mathbb{P}\{N(dy) = 1\} = \mathbb{E}N(dy) = \lambda(dy).$$

Then, the so-called avoidance function or taboo probability of Φ is

$$\begin{aligned}\mathbb{P}\{N(B)=0\} = \prod_{y\in B}(1-\lambda(dy)) &= \prod_{y\in B}\exp(-\lambda(dy)) \\ = \exp\Big(-\int_B \lambda(dy)\Big) &= \exp(-\lambda(B))\end{aligned}$$

and

$$\begin{aligned}\mathbb{P}\{N(B)=k\} &= \frac{1}{k!}\sum_{y_1,\dots,y_k\in B}\lambda(dy_1)\cdots\lambda(dy_k)\prod_{y\in B\setminus\{y_1,\dots,y_k\}}(1-\lambda(dy)) \\ &= \frac{\lambda(k)}{k!}\exp(-\lambda(B)),\quad k\in\mathbb{N}.\end{aligned}$$

This kind of argument is instructive and almost always leads to the right answer/guess. It can be justified by a limiting procedure (decompose the set B into partitions $B_i^{(n)}$, $1\le i\le n$ with $\max_{1\le i\le n}\lambda(B_i^{(n)})\to 0$ as $n\to\infty$).

As a first step towards showing the existence of Poisson point processes we turn things around and verify that condition *ii)* does not contradict the independence from condition *i)*. Note that for disjoint sets B_1 and B_2

$$N(B_1\cup B_2)=N(B_1)+N(B_2).$$

Consequently, the sum of independent Poisson distributed random variables has to have a Poisson distribution again. This result is the content of the following lemma.

Lemma 3.2 *Let X_j, $j\ge 1$ be independent Poisson(λ_j)- distributed random variables. Then $\sum_{j=1}^\infty X_j$ has Poisson distribution with mean $\sum_{j=1}^\infty \lambda_j$.*

Proof. Recall that the probabilty generating function of a Poisson(λ)-distributed random variable X is $\mathbb{E}s^X=\exp(-\lambda(1-s))$. Hence,

$$\mathbb{E}s^{\sum_{j=1}^n X_j} = \prod_{j=1}^n \mathbb{E}s_j^X = \prod_{j=1}^n \exp(-\lambda_j(1-s)) = \exp\Big(-\sum_{j=1}^n \lambda_j(1-s)\Big).$$

For $n=\infty$ the monotone convergence theorem implies

$$\mathbb{E}s^{\sum_{j=1}^\infty X_j} = \lim_{n\to\infty}\mathbb{E}s^{\sum_{j=1}^n X_j} = \lim_{n\to\infty}\exp\Big(-\sum_{j=1}^n\lambda_j(1-s)\Big) = \exp\Big(-\sum_{j=1}^\infty \lambda_j(1-s)\Big).$$

Theorem 3.3 (Construction) *Let λ be a finite (!) and non-atomic measure on the measurable space $(\mathcal{S},\mathcal{B})$. Let N be a Poisson distributed random variable with mean $\lambda(\mathcal{S})$ and let $Y_1,Y_2,\dots$ be i.i.d. random variables with distribution $\lambda/\lambda(\mathcal{S})$, independent of N. Then the random set*

$$\Phi := \{Y_j : 1\le j\le N\}$$

is a Poisson point process on $\mathcal{S}$ with intensity measure λ.

Proof. Let $B_1, \ldots, B_n \in \mathcal{B}$ be disjoint subsets of $\mathcal{S}$ and set

$$B_{n+1} := \mathcal{S} \setminus \bigcup_{i=1}^{n} B_i .$$

Note that given $N = \ell$ the counts $(N(B_1), \ldots, N(B_{n+1}))$ have multinomial distribution with parameter $(\ell, p_1, \ldots, p_{n+1})$, where $p_j = \lambda(B_j)/\lambda(\mathcal{S})$, $1 \le j \le n+1$. Hence, for $k_1, \ldots, k_{n+1} \in \mathbb{N}_0$ and $\ell = k_1 + \cdots + k_{n+1}$ we have

$$\begin{aligned}
&\mathbb{P}\{N(B_1) = k_1, \ldots, N(B_{n+1}) = k_{n+1}\} \\
&= \mathbb{P}\{N = \ell\}\, \mathbb{P}\{N(B_1) = k_1, \ldots, N(B_{n+1}) = k_{n+1} \,|\, N = \ell\} \\
&= \exp(-\lambda(\mathcal{S}))\, \frac{\lambda(\mathcal{S})^{\ell}}{\ell!} \binom{\ell}{k_1, \ldots, k_{n+1}} \left(\frac{\lambda(B_1)}{\lambda(\mathcal{S})}\right)^{k_1} \cdots \left(\frac{\lambda(B_{n+1})}{\lambda(\mathcal{S})}\right)^{k_{n+1}} \\
&= \exp\Big(-\sum_{j=1}^{n+1} \lambda(B_j)\Big)\, \frac{\lambda(B_1)^{k_1}}{k_1!} \cdots \frac{\lambda(B_{n+1})^{k_{n+1}}}{k_{n+1}!} \\
&= \prod_{j=1}^{n+1} \exp(-\lambda(B_j))\, \frac{\lambda(B_j)^{k_j}}{k_j!} . \qquad (3.1)
\end{aligned}$$

Properties *i)* and *ii)* follow from (3.1) by summation over k_{n+1}.

Theorem 3.4 (Restriction theorem) *Let Φ be a Poisson point process on $\mathcal{S}$ with mean measure λ and let B be a measurable subset of $\mathcal{S}$. Then the random countable set $\Phi_B := \Phi \cap B$ is a Poisson point process on $\mathcal{S}$ with mean measure $\lambda_B = \lambda(B \cap \cdot)$.*

Proof. Conditions *i)* and *ii)* are easily verified for Φ_B.

The following result will be useful to extend the construction of a Poisson point process given in Theorem 3.3 to the case $\lambda(\mathcal{S}) = \infty$.

Theorem 3.5 (Superposition theorem) *Let Φ_j, $j \ge 1$ be independent Poisson point processes on $\mathcal{S}$ with non-atomic intensity measures λ_j. If $\lambda := \sum_{j=1}^{\infty} \lambda_j$ is σ-finite, then the union*

$$\Phi := \bigcup_{j=1}^{\infty} \Phi_j$$

is a Poisson point process on $\mathcal{S}$ with intensity measure λ.

Proof. If X and Y are independent random variables with values in $\mathcal{S}$ and non-atomic distributions ν_X and ν_Y, then, by means of Fubini's theorem,

$$\mathbb{P}\{X = Y\} = \int_{\mathcal{S}} \mathbb{P}\{X = Y \,|\, Y = y\}\, \nu_Y(dy) = \int_{\mathcal{S}} \nu_X(\{x\})\, \nu_Y(dy) = 0.$$

Hence, letting X_i and X_j have distribution $\lambda_i(\,\cdot\cap B)/\lambda_i(B)$ and $\lambda_j(\,\cdot\cap B)/\lambda_j(B)$, respectively, Theorem 3.3 shows that

$$\begin{aligned}
\mathbb{P}\{\Phi_i \cap \Phi_j \cap B \neq \emptyset\} &\leq \mathbb{E}\,|\Phi_i \cap \Phi_j \cap B| \\
&= \mathbb{E}\Big(\mathbb{E}\Big(\,|\Phi_i \cap \Phi_j \cap B|\,\Big|\,N_i(B), N_j(B)\Big)\Big) \\
&\leq \mathbb{P}\{X_i = X_j\}\,\mathbb{E}(N_i(B)N_j(B)) \\
&= \mathbb{P}\{X_i = X_j\}\,\lambda_i(B)\,\lambda_j(B) \;=\; 0
\end{aligned} \tag{3.2}$$

provided that $\lambda_i(B), \lambda_j(B) < \infty$. Taking $B_1 \subset B_2 \subset \cdots \subset \mathcal{S}$ such that $\bigcup_{n\geq 1} B_n = \mathcal{S}$ and $\lambda(B_n) < \infty$ for all n, we see that (3.2) holds for all B. Hence,

$$N(B) := |\Phi \cap B| \overset{a.s.}{=} \sum_{j=1}^{\infty} N_j(B).$$

By Lemma 3.2, $N(B)$ is Poisson($\lambda(B)$)-distributed. For the independence note that

$$(N(B_1), \ldots, N(B_n)) \overset{a.s.}{=} \Big(\sum_{j=1}^{\infty} N_j(B_1), \ldots, \sum_{j=1}^{\infty} N_j(B_n)\Big),$$

with the $N_j(B_k)$, $1 \leq k \leq n$ and $j \geq 1$ being mutually independent if the B_k are disjoint.

Corollary 3.6 (Existence theorem) *Let λ be a σ-finite (non-atomic) measure on $(\mathcal{S}, \mathcal{B})$. Then there exists a Poisson point process on $\mathcal{S}$ with mean measure λ.*

Proof. Let $B_1 \subset B_2 \subset \cdots$ with $\lambda(B_j) < \infty$ and $\bigcup_{j=1}^{\infty} B_j = \mathcal{S}$. Set

$$B_j^* := B_j \setminus \bigcup_{i=1}^{j-1} B_i,\ \ j \geq 1.$$

Note that the B_j^*, $j \geq 1$ are disjoint and that $\bigcup_{i=1}^{j} B_i^* = \bigcup_{i=1}^{j} B_i$ for all $j \geq 1$. Following Theorem 3.3 we can construct independent Poisson point processes Φ_j^* on $\mathcal{S}$ with intensity measures $\lambda_j^* = \lambda(\,\cdot\,\cap B_j^*)$. By Theorem 3.5, the superposition

$$\Phi^* := \bigcup_{j=1}^{\infty} \Phi_j^*$$

is a Poisson point process on $\mathcal{S}$ with intensity measure

$$\lambda^* \;=\; \sum_{j=1}^{\infty} \lambda_j^* \;= \sum_{j=1}^{\infty} \lambda(\,\cdot\,\cap B_j^*) \;=\; \lambda\Big(\bigcup_{j=1}^{\infty} \cdot\,\cap B_j^*\Big) \;=\; \lambda.$$

Another nice property of Poisson point processes is that, if the state space is mapped into some other space, then the transformed points again form a Poisson point process. The only thing one has to pay attention to is that the function might pile distinct points. This possibility, however, can be detected by merely looking at the transformed intensity measure.

Theorem 3.7 (Mapping theorem) *Let Φ be a Poisson point process on $\mathcal{S}$ with mean measure λ and let $f : \mathcal{S} \to \mathcal{S}'$ be a measurable function such that the image (or induced) measure*

$$\lambda_f = \lambda \circ f^{-1}$$

on $\mathcal{S}'$ is σ-finite and non-atomic. Then $f(\Phi) = \{f(x) : x \in \Phi\}$ is a Poisson point process on $\mathcal{S}'$ with mean measure λ_f.

Proof. Let $B_n \uparrow \mathcal{S}$ with $\lambda(B_n) < \infty$ and let $Y_{n,1}$ and $Y_{n,2}$ be i.i.d. with distribution λ_{B_n}. Similar to (3.2) we deduce

$$\begin{aligned}\mathbb{P}\{f(x) = f(y) \text{ for some } x \neq y \in \Phi\} &= \lim_{n\to\infty} \mathbb{P}\{f(x) = f(y) \text{ for some } x \neq y \in \Phi \cap B_n\} \\ &\leq \lim_{n\to\infty} \mathbb{P}\{f(Y_{n,1}) = f(Y_{n,2})\}\, \mathbb{E}N(B_n)^2 = 0.\end{aligned}$$

Hence,

$$N_f(B') := |f(\Phi) \cap B'| \overset{a.s.}{=} |\Phi \cap f^{-1}(B')| = N(f^{-1}(B')), \tag{3.3}$$

which shoes that $N_f(B')$ has Poisson distribution with mean $\lambda(f^{-1}(B')) = \lambda_f(B')$. For the independence properties note that $f^{-1}(B'_i) \cap f^{-1}(B'_j) = f^{-1}(B'_i \cap B'_j) = \emptyset$, if $B'_i \cap B'_j = \emptyset$.

3.2 Sums over Poisson point processes

In this section we study random quantities $\sum_{x\in\Phi} f(x)$, where Φ is a Poisson point process on $\mathcal{S}$ and $f : \mathcal{S} \to \mathbb{R}$ is some measurable function. We will address questions about absolute convergence, the expectation and variance of such sums and their Laplace functionals. The general strategy will be to first look at simple functions $f = \sum_{i=1}^n \beta_i 1_{B_i}$ with non-negative β_i and disjoint measurable sets B_i, $1 \leq i \leq n$, and then apply standard techniques from integration theory to extend the results to general f.

Theorem 3.8 *Let Φ be a Poisson point process on $\mathcal{S}$ with mean measure λ and let $f : \mathcal{S} \to \mathbb{R}$ be a measurable function.*

1. *If $f \geq 0$, then*

$$\mathbb{E}\exp\Big(-\sum_{x\in\Phi} f(x)\Big) = \exp\Big(-\int_{\mathcal{S}} \Big(1 - e^{-f(z)}\Big)\lambda(dz)\Big) \tag{3.4}$$

and

$$\mathbb{P}\Big\{\sum_{x\in\Phi} f(x) < \infty\Big\} = \begin{cases} 1, & \text{if } \int_{\mathcal{S}} \min(1, f(z))\,\lambda(dz) < \infty; \\ 0, & \text{else.}\end{cases} \tag{3.5}$$

2. *If* $f \geq 0$ *or* $\int_{\mathcal{S}} |f(z)|\,\lambda(dz) < \infty$, *then*

$$\mathbb{E}\Big(\sum_{x\in\Phi} f(x)\Big) = \int_{\mathcal{S}} f(z)\,\lambda(dz).$$

3. *If* $\int_{\mathcal{S}} |f(z)|\,\lambda(dz) < \infty$, *then*

$$\mathrm{Var}\Big(\sum_{x\in\Phi} f(x)\Big) = \int_{\mathcal{S}} f^2(z)\,\lambda(dz).$$

The *Laplace functional* $\varphi(f)$ in (3.4) uniquely determines the distribution of Φ. For fixed f the value is the same as the Laplace transform of $\sum_{x\in\Phi} f(x)$ evaluated at 1.

Proof. Let

$$f = \sum_{i=1}^{n} \beta_i 1_{B_i}$$

where $\beta_i \geq 0$ and $B_i \cap B_j = \emptyset$ for $i \neq j$. Then

$$\begin{aligned}\sum_{x\in\Phi} f(x) &= \sum_{x\in\Phi}\sum_{i=1}^{n} \beta_i\, 1_{B_i}(x) \\ &= \sum_{i=1}^{n} \beta_i \sum_{x\in\Phi} 1_{B_i}(x) \;=\; \sum_{i=1}^{n} \beta_i N(B_i). \end{aligned} \tag{3.6}$$

Now recall that if N is Poisson(ν)-distributed, then $\mathbb{E}\exp(-\theta N) = \exp\big(-\nu(1-e^{-\theta})\big)$. Hence, using the independence of the $N(B_i)$ and identity (3.6) we get

$$\begin{aligned}\mathbb{E}\exp\Big(-\sum_{x\in\Phi} f(x)\Big) &= \mathbb{E}\exp\Big(-\sum_{i=1}^{n} \beta_i N(B_i)\Big) \\ &= \prod_{i=1}^{n} \exp\Big(-\lambda(B_i)\Big(1-e^{-\beta_i}\Big)\Big) \\ &= \exp\Big(-\sum_{i=1}^{n} \lambda(B_i)\Big(1-e^{\beta_i}\Big)\Big) \\ &= \exp\Big(-\int_{\mathcal{S}} \Big(1-e^{-\sum_{i=1}^{n} \beta_i\, 1_{B_i}(z)}\Big)\lambda(dz)\Big) \\ &= \exp\Big(-\int_{\mathcal{S}} \Big(1-e^{-f(z)}\Big)\lambda(dz)\Big). \end{aligned}$$

Since any non-negative measurable function f is the increasing limit of simple functions f_n, say, the monotone convergence theorem gives

$$\mathbb{E}\exp\Big(-\sum_{x\in\Phi} f(x)\Big) = \mathbb{E}\exp\Big(-\sum_{x\in\Phi} \lim_{n\to\infty} f_n(x)\Big)$$

$$\begin{aligned}
&= \lim_{n\to\infty} \mathbb{E}\exp\Big(-\sum_{x\in\Phi} f_n(x)\Big)\\
&= \lim_{n\to\infty} \exp\Big(-\int_{\mathcal{S}}\Big(1-e^{-f_n(z)}\Big)\lambda(dz)\Big)\\
&= \exp\Big(-\int_{\mathcal{S}}\Big(1-e^{-f(z)}\Big)\lambda(dz)\Big). \qquad (3.7)
\end{aligned}$$

For the convergence criterion of the random series $\sum_{x\in\Phi} f(x)$ recall that for any non-negative random variable X one has $\mathbb{P}\{X<\infty\} = \lim_{\theta\downarrow 0}\mathbb{E}\exp(-\theta X)$. Using the continuity of the exponential function we get

$$\begin{aligned}
\mathbb{P}\Big\{\sum_{x\in\Phi} f(x)<\infty\Big\} &= \lim_{\theta\downarrow 0}\mathbb{E}\exp\Big(-\theta\sum_{x\in\Phi} f(x)\Big)\\
&= \exp\Big(-\lim_{\theta\downarrow 0}\int_{\mathcal{S}}\Big(1-e^{-\theta f(z)}\Big)\lambda(dz)\Big). \qquad (3.8)
\end{aligned}$$

We write I_θ for the integral on the right-hand side of (3.8). By the monotone convergence theorem, to prove (3.5) it suffices to show

$$\int_{\mathcal{S}}\min(1,f(z))\,\lambda(dz)\ \begin{matrix}<\infty\\[1em] =\infty\end{matrix}\quad\Longrightarrow\quad\begin{matrix}I_\theta<\infty \text{ for some } \theta\\[1em] I_\theta=\infty \text{ for all } \theta\ .\end{matrix}$$

For the first implication note that

$$I_\theta \ \le \int_{\mathcal{S}}\min(\theta f(z),1)\,\lambda(dz) \le\ \max(\theta,1)\int_{\mathcal{S}}\min(f(z),1)\,\lambda(dz).$$

Similarly,

$$I_\theta \ \ge \int_{\mathcal{S}}\min(\tfrac{\theta}{2}f(z),1)\,\lambda(dz) \ge\ \min(\tfrac{\theta}{2},c_\theta)\int_{\mathcal{S}}\min(f(z),1)\,\lambda(dz)$$

for some $c_\theta>0$. This completes the proof of the first part of the Theorem.

For 2. we again first take f of the form $f=\sum_{i=1}^n \beta_i 1_{B_i}$. Then (recall (3.6))

$$\begin{aligned}
\mathbb{E}\Big(\sum_{x\in\Phi} f(x)\Big) &= \mathbb{E}\Big(\sum_{i=1}^n \beta_i N(B_i)\Big) = \sum_{i=1}^n \beta_i\lambda(B_i)\\
&= \int_{\mathcal{S}}\sum_{i=1}^n \beta_i 1_{B_i}(z)\,\lambda(dz) = \int_{\mathcal{S}} f(z)\lambda(dz).
\end{aligned}$$

For general $f\ge 0$ take $f_n\uparrow f$ and use monotone convergence. For f with $\int|f|\,d\lambda<\infty$ decompose $f=f^+-f^-$ and use linearity of expectation.

Finally, for $f=\sum_{i=1}^n \beta_i 1_{B_i}$ we have

$$\begin{aligned}
\mathrm{Var}\Big(\sum_{x\in\Phi} f(x)\Big) &= \mathrm{Var}\Big(\sum_{i=1}^n \beta_i\, 1_{B_i}(x)\Big)\\
&= \sum_{i=1}^n \mathrm{Var}(\beta_i N(B_i)) = \sum_{i=1}^n \beta_i^2\lambda(B_i) = \int_{\mathcal{S}} f^2(z)\,\lambda(dz).
\end{aligned}$$

If $\int |f|\, d\lambda < \infty$, then

$$\begin{aligned}
\mathrm{Var}\Big(\sum_{x\in\Phi} f(x)\Big) &= \mathrm{Var}\Big(\sum_{x\in\Phi\cap\{f\ge 0\}} f(x) + \sum_{x\in\Phi\cap\{f<0\}} f(x)\Big) \\
&= \mathrm{Var}\Big(\sum_{x\in\Phi} f^+(x)\Big) + \mathrm{Var}\Big(\sum_{x\in\Phi} f^-(x)\Big) \\
&= \int_{\mathcal{S}} \Big(f^+(z)^2 + f^-(z)^2\Big)\, \lambda(dz) \\
&= \int_{\mathcal{S}} f(z)^2 \lambda\,(dz).
\end{aligned}$$

3.3 Labelled Poisson point processes

Recall the following characteristic property of the Poisson distribution in relation with the binomial distribution.

Lemma 3.9 *Let $M = \sum_{i=1}^N X_i$, where N is a Poisson(λ)-distributed random variable and the X_i, $i \ge 1$ are i.i.d. Bernoulli(p)-distributed random variables, independent of N. Then M and $N-M$ are independent Poisson-distributed random variables with mean λp and $\lambda(1-p)$, respectively.*

Proof. Given $\{N = n\}$ the random variable M is binomially distributed with parameter (n, p). Hence, for every $k, \ell \ge 0$,

$$\begin{aligned}
\mathbb{P}\{M = k, N - M = \ell\} &= \mathbb{P}\{N = k+\ell\}\, \mathbb{P}\{M = k \,|\, N = k+\ell\} \\
&= e^{-\lambda} \frac{\lambda^{k+\ell}}{(k+\ell)!} \binom{k+\ell}{k} p^k (1-p)^\ell \\
&= e^{-\lambda p} \frac{(\lambda p)^k}{k!}\, e^{-\lambda(1-p)} \frac{(\lambda(1-p))^\ell}{\ell!}.
\end{aligned}$$

Corollary 3.10 (Coloring Theorem) *Let Φ be a Poisson point process on $\mathcal{S}$ with mean measure λ. Color the points of Φ independently (of each other and of their positions) with k different colors. For each j let Φ_j be the set of points of color j. Then Φ_j, $1 \le j \le k$ are independent Poisson point processes on $\mathcal{S}$ with intensity measure $\lambda_j = p_j \lambda$, where p_j is the probability that a point is colored j.*

Proof. Let $k = 2$ (the case $k \ge 3$ follows by induction). If $B_1, \ldots, B_n$ are disjoint measurable subsets of $\mathcal{S}$, then, by Lemma 3.9, the counts

$$N_j(B_1) = |\Phi_j(B_i)|, \quad j = 1,2 \text{ and } 1 \le i \le n$$

are independent random variables with distribution Poisson($p_j \lambda(B_i)$).

The assumptions of Theorem 3.10 are far too restrictive. The demand to relax these conditions leads to the concept of labelled (or marked) Poisson point processes.

Let $\Phi = \{X_i : i \in I\}$ be a Poisson point process. Attach to every point X_i a label L_i whose distribution may depend on X_i, but is independent of (X_j, L_j), $j \neq i$ (you may think of spatially inhomogeneous color probabilities).

For a formal definition recall that P is called a *transition kernel* from $(\mathcal{S}, \mathcal{B})$ to $(\mathcal{S}', \mathcal{B}')$, if

1. $P(x, \cdot\,)$ is a probability measure on $(\mathcal{S}', \mathcal{B}')$ for each $x \in \mathcal{S}$.
2. $P(\,\cdot\,, B')$ is a measurable function on $\mathcal{S}$ for each $B' \in \mathcal{B}'$.

Theorem 3.11 *Let Φ be a Poisson point process on $\mathcal{S}$ with mean measure λ. Given $\Phi = \{x_i : i \in I\}$, let L_i, $i \in I$ be independent with distribution $P(x_i, \cdot\,)$. Then*

$$\Psi := \{(X_i, L_i) : i \in I\}$$

is a Poisson point process on $\mathcal{S} \times S'$ with mean measure

$$\lambda P(dx\, d\ell) := \lambda(dx) P(x, d\ell).$$

Proof. By the superposition theorem, it suffices to consider the case $\lambda(\mathcal{S}) < \infty$. We use the construction of Φ from Theorem 3.3:

$$\Phi = \{X_i : 1 \leq i \leq N\},$$

where N is a Poisson($\lambda(\mathcal{S})$)-distributed random variable, and the X_i are i.i.d. with distribution $\lambda/\lambda(\mathcal{S})$, independent of N.

Note that if X has distribution $\lambda/\lambda(\mathcal{S})$ and the conditional distribution of L given $\{X = x\}$ is $P(x, \cdot\,)$, then the pair (X, L) has distribution $\lambda P/\lambda(\mathcal{S})$. Hence,

$$\Psi \overset{d}{=} \{(X_i, L_i) : 1 \leq i \leq N\},$$

where N is Poisson-distributed with mean

$$\begin{aligned} \lambda(\mathcal{S}) &= \int_{\mathcal{S}} \lambda(dx)\, P(x, \mathcal{S}') \\ &= \int_{\mathcal{S}} \lambda(dx) \int_{\mathcal{S}'} P(x, d\ell) \;=\; \lambda P(\mathcal{S} \times \mathcal{S}'), \end{aligned}$$

and the (X_i, L_i), $i \geq 1$ are i.i.d. with distribution

$$\frac{\lambda}{\lambda(\mathcal{S})} P \;=\; \frac{\lambda P}{\lambda P(\mathcal{S} \times \mathcal{S}')}.$$

The assertion of Theorem 3.11 now follows by yet another application of Theorem 3.3.

Corollary 3.12

1. *If $\lambda P(\mathcal{S} \times \cdot\,)$ is a σ-finite and non-atomic measure on $\mathcal{S}'$, then the set of labels $\{L_i : i \in I\}$ form a Poisson point process on $\mathcal{S}'$ with intensity measure $\lambda P(\mathcal{S} \times \cdot\,)$.*

2. *For measurable disjoint subsets $B'_j \subset \mathcal{S}'$, the sets of points with a label in B'_j,*

$$\Psi_{B'_j} = \Psi \cap \mathcal{S} \times B'_j,\ j \geq 1$$

are independent Poisson point processes with intensity measures $\lambda P(\,\cdot\, \times B'_j)$.

Note that the coloring theorem is the particular case $P(x, dy) = \mu(dy)$, i.e., $\lambda P = \lambda \otimes \mu$.

Proof. The first part follows by Theorem 3.11 and the mapping theorem, the second part by Theorem 3.11 and the restriction theorem.

3.4 Exercises

Exercise 3.1 *Let Φ be a homogeneous Poisson point process on $\mathbb{R}^2$ (i.e., Φ has Lebesgue measure as its mean measure λ). Set $D_0 := 0$ and let D_j, $j \in \mathbb{N}$ denote the distance of the j-th closest point in Φ to the origin. Show that*

a) $\mathbb{P}\{D_1 \geq y\} = \exp(-\pi y^2)$ *for all* $y \geq 0$.

b) $D_{j+1}^2 - D_j^2$, $j \in \mathbb{N}_0$ *are independent and exponentially distributed random variables with parameter* π.

Exercise 3.2 *Let ν be a probability measure on $\mathbb{R}\backslash\{0\}$ with expectation m and finite variance σ^2. Let $(Y_t)_{t\geq 0}$ be the process defined through*

$$Y_t := \sum_{0 \leq S_i \leq t} H_i, \quad t \geq 0,$$

where (S_i, H_i) are the points of a Poisson point process Φ on $\mathbb{R}_0^+ \times \mathbb{R}$ with mean measure $\lambda(ds\,dh) = ds\,\nu(dh)$.

a) *Show that (Y_t) is a renewal reward process with Exp(1)-distributed lifetimes.*

b) *Compute the mean and the variance of Y_t.*

c) *Suppose $\nu(\mathbb{R}^+) = 1$ and let $\varphi(\theta) = \int_0^\infty \exp(-\theta x)\nu(dx)$, $\theta \geq 0$. Compute the Laplace transform $\varphi_t(\theta) := \mathbb{E}\exp(-\theta Y_t)$.*

Exercise 3.3 *The* Gamma(α, β)-distribution *with shape and scale parameters $\alpha, \beta > 0$ has density*

$$g_{\alpha,\beta}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1} e^{-\beta x},\ x \in \mathbb{R}_0^+,$$

where $\Gamma(\alpha) = \int\limits_0^\infty y^{\alpha-1} e^{-y} dy$.

a) *Compute the Laplace transform of a Gamma(α, β)-distributed random variable X.*

b) *Show that the sum of n independent Gamma$(\frac{\alpha}{n}, \beta)$-distributed random variables has distribution Gamma(α, β).*

c) Let Φ be a Poisson point process on $\mathbb{R}^+$ with intensity measure

$$\lambda_{\alpha,\beta}(dz) = \frac{\alpha\, e^{-\beta z}}{z}\, dz.$$

Show that

$$Y := \sum_{x \in \Phi} x$$

has distribution Gamma(α, β).

Exercise 3.4 ***PPP representation of the Gamma process.*** *Let Φ be a Poisson point process on $\mathbb{R}_0^+ \times \mathbb{R}^+$ with intensity measure $\lambda(ds\, dz) = ds\, \lambda_{\alpha,\beta}(dz)$. Set*

$$Y_t := \sum_{(u,y) \in \Phi \cap [0,t]} y, \quad t \geq 0.$$

Show that:

a) $Y_t = \lim_{h \downarrow 0} Y_{t+h}$ and $Y_{t-} := \lim_{h \downarrow 0} Y_{t-h}$ exists a.s. for every $t \geq 0$.

b) In every non-empty bounded time interval $[t_1, t_2]$ the process $(Y_t)_{t \geq 0}$ takes infinitely many jumps, only finitely many of which exceed $\varepsilon\,(> 0)$.

c) For every $0 \leq t_1 < t_2 < \cdots < t_{n+1}$ the increments $Y_{t_{j+1}} - Y_{t_j}$, $1 \leq j \leq n$ are independent random variables with distribution Gamma$(\alpha(t_{j+1} - t_j), \beta)$.

4 Markov chains in continuous time

In this chapter time is continuous while the state space $\mathcal{S}$ is still discrete.

4.1 Definition and basic properties

Definition 4.1 *Let P^t, $t \geq 0$ be a sequence of stochastic matrices on a discrete space $\mathcal{S}$. A stochastic process $(X_t)_{t\geq 0}$ is called a (time-homogeneous) continuous-time Markov chain with state space $\mathcal{S}$ and transition probabilities P^s_{yz}, if*

$$\mathbb{P}_x\{X_{t+s} = z \,|\, X_{t_1} = y_1, \ldots, X_{t_n} = y_n, X_t = y\} = P^s_{yz} \tag{4.1}$$

for every $n \in \mathbb{N}_0$, $0 \leq t_1 \leq t_2 \leq \ldots \leq t_n \leq t$, $s \geq 0$ and $x, y, y_1, \ldots, y_n, z \in \mathcal{S}$ (provided that the conditional probability is well defined). Here, $\mathbb{P}_x$ is the law of $(X_t)_{t\geq 0}$ when started at x.

Note that in the continuous-time setting we need a whole sequence of stochastic matrices to describe the dynamics of the chain rather than a single transition matrix as in the discrete-time case. The analogue of Lemma 1.3 are the so-called *Chapman-Kolmogorov-equations*: The law of total probability and (4.1) imply that the P^t satisfy

$$P^{t+s}_{xy} = \sum_{z\in\mathcal{S}} P^t_{xz}\, P^s_{zy} = (P^t P^s)_{xy} \quad \text{for all } s, t \geq 0 \text{ and } x, y \in \mathcal{S}.$$

In other words, the sequence P^t, $t \geq 0$ is a *semigroup* of stochastic matrices.

The distribution of the path $(X_t)_{t\geq 0}$ is not uniquely determined by the finite-dimensional distributions (4.1) as can be seen by the following example. Let

$$(X^{(1)}_t)_{t>0} \equiv 0$$

and

$$X^{(2)}_t = \begin{cases} 1, & \text{if } \; t = U, \\ 0, & \text{else,} \end{cases}$$

where U is uniformly distributed on the unit interval $[0, 1]$, say. Both processes have transition probabilities

$$P^t_{xy} = \begin{cases} 1, & \text{if } y = 0, \\ 0, & \text{else.} \end{cases}$$

In order to have uniqueness of $\mathcal{L}((X_t)_{t\geq 0})$ one has to require some smoothness of the paths (e.g., that the paths are a.s. continuous or càdlàg).

4.2 Jump processes

Jump processes are an important class of Markov chains in continuous time. They can be thought of as disrete-time Markov chains with randomized holding times.

Let $T_0 := \inf\{t > 0 : X_t \neq X_0\}$ and assume $\mathbb{P}_x\{T_0 > 0\} = 1$, i.e., the chain does not instantaneously move away from its initial point (later on we will learn of processes where this is not the case). The Markov property (4.1) then implies

$$\mathbb{P}_x\{T_0 \geq t+s \,|\, T_0 \geq t\} = \mathbb{P}_x\{T_0 \geq t+s \,|\, X_{u^-} = x \text{ for all } 0 \leq u \leq t\} = \mathbb{P}_x\{T_0 \geq s\}. \tag{4.2}$$

The *lack of memory* property (4.2) uniquely characterizes the exponential distribution, i.e., $\mathcal{L}(T_0) = \mathrm{Exp}(q_x)$ for some $0 \leq q_x < \infty$. This suggests the following construction of a continuous-time Markov chain. As ingredients we need

- state dependent jump rates q_x, $x \in \mathcal{S}$;
- a (stochastic) jump matrix $(J_{xy})_{x,y\in\mathcal{S}}$.

For simplicity we will assume

$$J_{xx} = 0 \text{ or } 1 \text{ for all } x \in \mathcal{S}. \tag{4.3}$$

$$\text{If } J_{xx} = 1, \text{ then } q_x = 0. \tag{4.4}$$

Relations (4.3) and (4.3) say that "any jump is a true jump" (see Exercise 4.1.a) for the fact that these assumptions mean no loss of generality). A verbal description of a (q, J)-*jump process* is as follows: When started at x the process

- stays at x for an $\mathrm{Exp}(q_x)$-distributed time T_0
- jumps to y with probability J_{xy}
- stays at y for an $\mathrm{Exp}(q_y)$-distributed time T_1 (independent of T_0)
- jumps to z with probability J_{yz}
- ...

Construction 4.2 *Let $(Y_n)_{n\geq 0}$ be a discrete-time Markov chain with transition matrix J. Given $(Y_0, Y_1, \ldots) = (y_0, y_1, \ldots)$ let T_n, $n \geq 0$ be independent Exp(y_n)-distributed random variables. Set*

$$X_t = y_n, \quad \textit{if } S_n \leq t < S_{n+1},$$

where $S_0 := 0$ and $S_{n+1} := S_n + T_n$, $n \geq 0$. Then $(X_t)_{t\geq 0}$ is said to be a (q,J)- jump process *and $(Y_n)_{n\geq 0} = (X_{S_n})_{n\geq 0}$ is called the* embedded *Markov chain.*

4.3 Explosion

Our construction of the (q, J)-jump process defines the path $(X_t)_{t\geq 0}$ for all $t \geq 0$ only if the sum of holding times exhausts the time axis, i.e., if

$$\zeta := \sum_{n=0}^{\infty} T_n = \infty.$$

However, if the transition rates q_x grow to fast, then ζ might be finite.

Definition 4.3 *If $\zeta < \infty$, then the jump process is said to* explode *and ζ is called the* explosion time. *If $\mathbb{P}_x\{\zeta = \infty\} = 1$ for all $x \in \mathcal{S}$, then the process $(X_t)_{t\geq 0}$ is called* regular.

The following result gives a handy characterization of regularity.

Lemma 4.4 *A (q,J)-jump process is regular if and only if*

$$\mathbb{P}_x\Big\{\sum_{n=0}^{\infty}\frac{1}{q_{Y_n}}=\infty\Big\}=1 \text{ for all } x\in\mathcal{S}.$$

Proof. We first consider the case where the embedded Markov chain $(Y_n)_{n\geq 0}$ is deterministic, i.e., we assume that for each $x\in\mathcal{S}$ we have

$$\mathbb{P}_x\{\,(Y_n)_{n\geq 0}=(y_n^x)_{n\geq 0}\}=1$$

for some $(y_n^x)_{n\geq 0}\in\mathcal{S}^{\mathbb{N}_0}$. In this case we need to show that

$$\mathbb{P}_x\{\zeta=\infty\}=1 \iff \sum_{n=0}^{\infty}\frac{1}{q_{y_n^x}}=\infty. \tag{4.5}$$

For the necessity part of (4.5) observe that

$$\mathbb{E}_x\zeta=\mathbb{E}_x\Big(\sum_{n=0}^{\infty}T_n\Big)=\sum_{n=0}^{\infty}\mathbb{E}_xT_n=\sum_{n=0}^{\infty}\frac{1}{q_{y_n^x}}.$$

Hence,

$$\sum_{n=0}^{\infty}\frac{1}{q_{y_n^x}}<\infty \implies \mathbb{E}_x\zeta<\infty \implies \mathbb{P}_x\{\zeta<\infty\}=1.$$

For the other direction first observe that

$$\mathbb{P}_x\{\zeta=\infty\}=1 \iff \mathbb{E}_x\exp(-\zeta)=0.$$

Clearly, we may assume $q_{y_n^x}>0$ for all n. By the assumed independence of the T_n, we have

$$\begin{aligned}\mathbb{E}_x\exp(-\zeta) &= \mathbb{E}_x\exp\Big(-\sum_{n=0}^{\infty}T_n\Big) = \prod_{n=0}^{\infty}\mathbb{E}_x\exp(-T_n)\\ &= \prod_{n=0}^{\infty}\frac{q_{y_n^x}}{1+q_{y_n^x}} = \frac{1}{\prod_{n=0}^{\infty}\Big(1+\frac{1}{q_{y_n^x}}\Big)} \leq \frac{1}{\sum_{n=0}^{\infty}\frac{1}{q_{y_n^x}}}.\end{aligned}$$

Hence,

$$\sum_{n=0}^{\infty}\frac{1}{q_{y_n^x}}=\infty \implies \mathbb{E}_x\exp(-\zeta)=0 \implies \mathbb{P}_x\{\zeta<\infty\}=1.$$

Now let $(X_t)_{t\geq 0}$ have a general jump matrix J. By the law of total probability,

$$\mathbb{P}_x\{\zeta=\infty\}=\mathbb{P}_x\Big\{\sum_{n=0}^{\infty}T_n=\infty\Big\}=\mathbb{E}_xg(Y_0,Y_1,\ldots), \tag{4.6}$$

where $g:\mathcal{S}^{\mathbb{N}_0}\to[0,1]$ is defined as

$$g(y_0,y_1,\ldots):=\mathbb{P}\Big\{\sum_{n=0}^{\infty}T_n=\infty\,\Big|\,Y_n=y_n,\,n\geq 0\Big\}.$$

Since $0 \le g \le 1$, we have

$$\mathbb{P}_x\{\zeta = \infty\} = 1 \Longleftrightarrow \mathbb{P}_x\{(Y_n)_{n\ge0} \in g^{-1}(1)\} = 1.$$

From the first part of our proof, we know that

$$g^{-1}(1) = \Big\{(y_0, y_1, \ldots) : \sum_{n=0}^{\infty} \frac{1}{q_{y_n}} = \infty\Big\}.$$

Hence,

$$\mathbb{P}_x\{\zeta = \infty\} = 1 \quad \Longleftrightarrow \quad \mathbb{P}_x\Big\{\sum_{n=0}^{\infty} \frac{1}{q_{Y_n}} = \infty\Big\} = 1.$$

Corollary 4.5 *A (q, J)-jump process is regular if one of the following conditions holds:*

i) $\sup_{x\in\mathcal{S}} q_x < \infty$*;*

ii) $|\mathcal{S}| < \infty$*;*

iii) $(Y_n)_{n\ge0}$ *is a recurrent Markov chain.*

Proof. Clearly, condition *ii*) implies *i*). If *i*) holds, then

$$\sum_{n=0}^{\infty} \frac{1}{q_{Y_n}} \;\ge\; \sum_{n=0}^{\infty} \frac{1}{\sup_{x\in\mathcal{S}} q_x} \;=\; \infty.$$

For *iii*) note that

$$\sum_{n=0}^{\infty} \frac{1}{q_{Y_n}} \;\ge\; \sum_{n=0}^{\infty} \frac{1}{q_{Y_n}} I\{Y_n = x\} \;=\; \frac{L_x}{q_x},$$

where L_x is the number of visits at x by $(Y_n)_{n\ge0}$. By Theorem 1.9, $\mathbb{P}_x\{L_x = \infty\} = 1$, if x is a recurrent state of $(Y_n)_{n\ge0}$.

Examples.

- **Pure birth process.** Let $J_{x,x+1} = 1$ for all $x \in \mathcal{S} = \mathbb{N}$ and $q_x = x^\alpha$. Then

  $$\sum_{n=0}^{\infty} \frac{1}{q_{Y_n}} \;=\; \sum_{n=Y_0}^{\infty} \frac{1}{n^\alpha}$$

  Hence, the chain explodes, if $\alpha > 1$, while it is regular, if $0 < \alpha \le 1$.

- **Binary Galton-Watson process.** Let $q_x = \lambda x$, $x \in \mathbb{N}_0$ and

  $$P_{xy} = \begin{cases} p_2, & \text{if } y = x+1, \\ p_0 = 1 - p_2, & \text{if } y = x-1\ . \end{cases}$$

  Then

  $$\sum_{n=0}^{\infty} \frac{1}{q_{Y_n}} \;\ge\; \sum_{n=0}^{\infty} \frac{1}{\lambda(Y_0 + n)} \;=\; \frac{1}{\lambda} \sum_{n=Y_0}^{\infty} \frac{1}{n} \;=\; \infty.$$

  This shows that the binary continuous-time Galton-Watson process is regular.

Is there a Markovian continuation of a non-regular (q, J)-jump process after its time of explosion? In what follows we briefly discuss two possible extensions.

The minimal process. Let Δ be some element not in $\mathcal{S}$ and set $\mathcal{S}^{\Delta} := \mathcal{S} \cup \{\Delta\}$. The *minimal process* $(X_t^{\Delta})_{t\geq 0}$ following the dynamics (q, J) is constructed as above for all times before ζ and is set equal to Δ for times after the explosion (the absorbing external state Δ is sometimes called *cemetery*). The minimal process is thus a continuous-time Markov chain on $\mathcal{S}^{\Delta}$ with transition probabilities

$$P_{xy}^{\Delta\, t} = \begin{cases} P_{xy}^t, & x, y \in \mathcal{S}; \\ \mathbb{P}\{\zeta \leq t\}, & x \in \mathcal{S},\, y = \Delta; \\ \delta_{xy}, & x = \Delta. \end{cases}$$

Revival after explosion. Instead of letting the process being absorbed in the cemetery Δ we can modify our (q, J)-jump process such that $\mathbb{P}_x\{X_t^{\mathrm{mod}} \in \mathcal{S}\} = 1$ for all $t \geq 0$: At time of explosion let the process immediately jump back into $\mathcal{S}$ landing at z with probability $\nu(z)$ where ν is some probability measure on $\mathcal{S}$. At the next explosion we independently repeat this procedure (same ν !). The resulting process $(X_t^{\mathrm{mod}})_{t\geq 0}$ is a continuous-time Markov chain on $\mathcal{S}$ whose distribution depends on ν.

4.4 Backward and forward equations

The backward and forward equations are two systems of differential equations for $P^t : \mathbb{R}_0^+ \to [0,1]$. The idea behind is that, in general, computation of P_{xy}^t is difficult whereas computation of $\frac{dP_{xy}^t}{dt}$ is easy. By the Markov property, those infinitesimal characteristics should contain (almost) all information on the transition semi-group $(P^t)_{t\geq 0}$.

Heuristics. We analyze P^t near $t = 0$. Neglecting effects of multiple jumps (compare Exercise 4.3) we have (recall our assumption $J_{xx} \in \{0, 1\}$)

$$1 - P_{xx}^{dt} \;=\; \mathbb{P}_x\{T_0 \leq dt\} \;=\; 1 - e^{-q_x dt} \;=\; q_x dt \tag{4.7}$$

and

$$P_{xy}^{dt} \;=\; \mathbb{P}_x\{T_0 \leq dt\}\, J_{xy} \;=\; q_x J_{xy}\, dt, \quad x \neq y. \tag{4.8}$$

The *Q-matrix* associated with q and J (or *infinitesimal generator* of the jump process $(X_t)_{t\geq 0}$) is defined as

$$Q_{xy} := \begin{cases} q_x J_{xy}, & \text{if } x \neq y, \\ -q_x, & \text{if } x = y. \end{cases}$$

Note that q and J can be recovered from Q. Also, since $P_{xy}^0 = \delta_{xy}$, we can rewrite (4.7) and (4.8) in matrix notation as

$$\frac{d}{dt} P^t \Big|_{t=0} = \; Q\,. \tag{4.9}$$

For the **backward equations** we decompose the evolution until time $t + dt$ with respect to X_{dt}, i.e., we write

$$P^{t+dt} - P^t \;=\; P^{dt} P^t - P^t \;=\; (P^{dt} - Id_{\mathcal{S}}) P^t.$$

Identity (4.9) suggests that

$$\frac{d}{dt}P^t = QP^t.$$

For the **forward equations** we decompose the evolution up to time $t+dt$ with respect to X_t instead. This yields

$$P^{t+dt} - P^t \; = \; P^tP^{dt} - P^t \; = \; P^t(P^{dt} - Id_{\mathcal{S}}),$$

so that we expect

$$\frac{d}{dt}P^t = P^tQ.$$

Theorem 4.6 (Backward equations) *Let $(P^t)_{t\geq 0}$ be the transition semi-group of a (q,J)-jump process. Then P^t_{xy} is continuously differentiable with*

$$\frac{d}{dt}P^t_{xy} = \sum_{z\in\mathcal{S}} Q_{xz}P^t_{zy} \tag{4.10}$$

for all $x, y \in \mathcal{S}$ and $t \geq 0$. In particular,

$$\frac{d}{dt}P^t_{xy}\Big|_{t=0} = Q_{xy}.$$

Note that for each fixed $y \in \mathcal{S}$, (4.10) is a system of differential equations for $u_t := P^t_{\cdot y}$.

Proof. We will derive an integral equation for P^t_{xy}. Using the Markov property of $(X_t)_{t\geq 0}$ we get

$$\begin{aligned} P^t_{xy} &= \mathbb{P}_x\{X_t = y,\, t < \zeta\} \\ &= \mathbb{P}_x\{X_t = y,\, T_0 > t\} \; + \; \sum_{z\neq x}\mathbb{P}_x\{X_t = y, T_0 \leq t, X_{T_0} = z, \zeta > t\} \\ &= e^{-q_xt}\delta_{xy} + \sum_{z\neq x}\int_0^t q_xe^{-q_xs}J_{xz}\,\mathbb{P}_z\{X_{t-s} = y, \zeta > t-s\}\,ds \\ &= e^{-q_xt}\Big(\delta_{xy} + q_x\int_0^t e^{q_xu}\Big(\sum_{z\neq x}J_{xz}P^u_{zy}\Big)\,du\Big). \end{aligned} \tag{4.11}$$

Observe that

$$h(u) := e^{q_xu}\Big(\sum_{z\neq x}J_{xz}P^u_{zy}\Big),\; u \geq 0$$

is a continuous function, since

$$\begin{aligned} \Big|\sum_{z\neq x}J_{xz}P^{u+h}_{zy} - \sum_{z\neq x}J_{xz}P^u_{zy}\Big| &\leq \sum_{z\neq x}J_{xz}|P^{u+h}_{zy} - P^u_{zy}| \\ &\leq \sum_{z\neq x,\, z\in K}|P^{u+h}_{zy} - P^u_{zy}| \; + \; 2\sum_{z\neq x,\, z\in K^c}J_{xz}. \end{aligned} \tag{4.12}$$

By (4.11), the first sum goes to zero as $h \downarrow 0$ for each finite K, while the second sum can be made arbitrarily small by choosing the set K large enough.

Hence, the product formula for differentiation yields

$$\frac{d}{dt} P_{xy}^t \;=\; -q_x P_{xy}^t + e^{-q_x t} q_x e^{q_x t} \sum_{z \neq x} e J_{xz} P_{zy}^t \;=\; \sum_{z \in S} Q_{xz} P_{zy}^t. \tag{4.13}$$

For continuity of the derivative on the right-hand side of (4.13) recall (4.12).

Theorem 4.7 (Forward equations) *Let* $(P^t)_{t \geq 0}$ *be the transition semi-group of a (q,J)-jump process and assume*

$$\sum_{y \in \mathcal{S}} P_{xy}^t q_y < \infty \tag{4.14}$$

for all $x, y \in \mathcal{S}$ *and* $t \geq 0$. *Then* P_{xy}^t *is continuously differentiable with*

$$\frac{d}{dt} P_{xy}^t = \sum_{z \in S} P_{xz}^t Q_{zy}. \tag{4.15}$$

Remarks.

- The assertion of Theorem 4.7 holds without assumption (4.12) (see [10], pp.100–103 for a proof).
- Condition (4.14) can be interpreted as the chain having finite expected speed at time t.
- For each fixed $x \in \mathcal{S}$, (4.15) is a system of differential equations for $v_t := P_{x\cdot}^t$.

Proof. For $x \neq y$ we have

$$\frac{P_{xy}^h}{h} \;\leq\; \frac{\mathbb{P}_x\{T_0 \leq h\}}{h} \;=\; \frac{1 - e^{-q_x h}}{h} \;\leq\; q_x.$$

Also,

$$\frac{P_{xy}^{t+h} - P_{xy}^t}{h} \;=\; \sum_{z \neq y} P_{xz}^t \frac{P_{zy}^h}{h} + P_{xy}^t \, \frac{P_{yy}^h - 1}{h}.$$

Hence, by assumption (4.14) and the dominated convergence theorem,

$$\lim_{h \downarrow 0} \frac{P_{xy}^{t+h} - P_{xy}^t}{h} \;=\; \lim_{h \downarrow 0} \sum_{z \neq y} P_{xz}^t \frac{P_{zy}^h}{h} \;+\; P_{xy}^t \lim_{h \downarrow 0} \frac{P_{yy}^h - 1}{h} \;=\; \sum_{z \in S} P_{xz}^t Q_{zy}.$$

To stress the different role of the backward and forward equations consider a function $f : \mathcal{S} \to \mathbb{R}$ and a measure μ on $\mathcal{S}$. Note that

$$\mathbb{E}_x f(X_t) = \sum_{y \in S} P^t_{xy} f(y) =: (P^t f)(x)$$

is the expected payoff at t when started at x and that

$$\mathbb{P}_\mu\{X_t = y\} = \sum_{x \in \mathcal{S}} \mu(x) P^t_{xy} =: (\mu P^t)(y)$$

is the distribution of X_t when started with initial distribution μ. The function $P_t f$ satisfies the backward equations while the measure $\mu_t P$ satisfies the forward equations. Indeed,

$$\begin{aligned}\frac{d}{dt}(P^t f)(x) &= \sum_{y \in \mathcal{S}} \frac{d}{dt} P^t_{xy} f(y) \;=\; \sum_{y \in \mathcal{S}} f(y) \sum_{z \in \mathcal{S}} Q_{xz} P^t_{zy} \\ &= \sum_{z \in \mathcal{S}} Q_{xz} \sum_{y \in \mathcal{S}} P^t_{zy} f(y) \;=\; \sum_{z \in \mathcal{S}} Q_{xz} (P^t f)(z) =: (QP^t f)(x).\end{aligned}$$

Similarly,

$$\frac{d}{dt}(\mu P^t)(y) = (\mu P^t Q)(y).$$

The interchange of summation and differentiation has to be justified from case to case.

4.5 Stationary distributions

The notions of stationarity and ergodicity introduced in Chapter 1 have a natural extension to the continuous-time setting. Also, the main results extend to continuous time with only minor modifications.

Definition 4.8 *A probability measure π on $\mathcal{S}$ is called a* stationary *distribution for $(P^t)_{t \geq 0}$, if*

$$\pi P^t = \pi \ \text{ for all } t \geq 0.$$

Theorem 4.9 *Suppose that the embedded chain $(Y_n)_{n \geq 0}$ is irreducible and positive recurrent with stationary distribution ν. If $(P^t)_{t \geq 0}$ has a stationary distribution π, then*

$$\pi(x) = c\, \frac{\nu(x)}{q_x}, \ x \in \mathcal{S},$$

where $c^{-1} = \sum_{y \in S} \frac{\nu(y)}{q_y}$.

The result (see [2], pp. 358–359 for a proof) is not surprising in view of our interpretation of $\pi(x)$ as the asymptotic proportion of time that the chain spends in x. Note that $\nu(x)$ is the asymptotic proportion of time*s* at x while $1/q_x$ is the expected time time spent at x per visit.

The following theorem is the continuous-time analogue of Theorem 1.25 (see Exercise 4.6 for a proof).

Theorem 4.10 (Convergence theorem) *Let $(Y_n)_{n \geq 0}$ be an irreducible Markov chain and suppose that $(P^t)_{t \geq 0}$ has stationary distribution π. Then,*

$$\lim_{t \to \infty} P^t_{xy} = \pi(y) \ \text{ for all } x, y \in \mathcal{S}.$$

4.6 Standard transition semi-groups

In Section 4.2 we obtained a semi-group of stochastic matrices starting from the description of the dynamics of a continuous-time Markov chain in terms of the transition rates q and the jump matrix J. Now we turn things around and start from the sequence of stochastic matrices.

Definition 4.11 *A sequence of stochastic matrices* P^t, $t \geq 0$ *is called a* standard transition semi-group *on* $\mathcal{S}$, *if*

i) $P^0 = Id_{\mathcal{S}}$.

ii) $P^{s+t} = P^s P^t$ *for all* $s, t \geq 0$.

iii) $\lim_{h \downarrow 0} P^h_{xx} = 1$ *for all* $x \in \mathcal{S}$.

Properties *i)*,–, *iii)* imply a great deal more than one might expect:

Theorem 4.12 *Let* $(P^t)_{t \geq 0}$ *be a* standard transition semi-group *on* $\mathcal{S}$. *Then*

$$\lim_{h \downarrow 0} \frac{P^h - Id_{\mathcal{S}}}{h} = Q \tag{4.16}$$

exists with

$$\begin{aligned} -\infty \leq Q_{xx} \leq 0 &\quad \textit{for all } x \in \mathcal{S}; \\ 0 \leq Q_{xy} < \infty &\quad \textit{for all } x \neq y \in \mathcal{S}; \\ \textstyle\sum_{y \neq x} Q_{xy} \leq -Q_{xx} &\quad \textit{for all } x \in \mathcal{S}. \end{aligned} \tag{4.17}$$

For a proof of Theorem 4.12, see [8], pp. 138–142.

Definition 4.13 *A state* $x \in \mathcal{S}$ *is called*

$$\begin{aligned} \text{instantaneous,} &\quad \textit{if } Q_{xx} = -\infty, \\ \text{stable ,} &\quad \textit{if } Q_{xx} > -\infty. \end{aligned}$$

A stable state $x \in \mathcal{S}$ *is called*

$$\text{conservative,} \quad \textit{if } \sum_{y \in \mathcal{S}} Q_{xy} = 0$$

and non-conservative, *else.*

We will now discuss the probabilistic meaning of instantaneous and non-conservative states. Note that the process has to jump away from an instantaneous state x immediately (since $Q_{xx} = -\infty$). However, by the assumed continuity *iii)*, the chain is at x for small times h with probability close to 1. For a stable but non-conservative state x note that the process should exit from $\mathcal{S}$ at the time it jumps away from x. However, since the P^t are stochastic matrices, it should immediately return.

The following two examples due to Kolmogorov illustrate these effects. They have become known as K1 and K2. Starting from a Q-matrix which satisfies (4.17) (allowing $Q_{xx} = -\infty$ or $\sum_y Q_{xy} < 0$) we will construct a Markov chain $(X_t)_{t \geq 0}$. The associated semi-group $(P^t)_{t \geq 0}$ will then be shown to satisfy (4.16).

Example K2. This is an example of a continous-time Markov chain with state space $\mathcal{S} = \mathbb{N}_0$ that has a stable but non-conservative state.

$$Q = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & \cdots & & & & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & & & & \cdots \\ 0 & 4 & -4 & 0 & 0 & \cdots & & & & \cdots \\ 0 & 0 & 9 & -9 & 0 & \cdots & & & & \cdots \\ & & & & & \ddots & & & & \\ \vdots & & & & & \cdots & 0 & i^2 & -i^2 & 0 & \cdots \\ & & & & & & & & \ddots & & \\ & & \cdots & & & \cdots & & & & \cdots \end{pmatrix} \tag{4.18}$$

At a state $i \geq 2$ thechain stays an $\mathrm{Exp}(i^2)$-distributed time before it jumps to $i-1$. When started at 0 the process waits an $\mathrm{Exp}(1)$-distributed time before it jumps to ∞, then immediately "implodes" and is finally absorbed in 1. Such a process can be constructed as follows. Let T_i, $i \geq 1$ be independent $\mathrm{Exp}(i^2)$-distributed random variables and note that

$$\mathbb{E}\Big(\sum_{i=1}^{\infty} T_i\Big) = \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty \implies \sum_{i=1}^{\infty} T_i \overset{a.s.}{<} \infty.$$

Set

$$X_t := \begin{cases} 0, & \text{if } \ 0 \leq t \leq T_1; \\ k, & \text{if } \ T_1 + \sum\limits_{i=k+1}^{\infty} T_i \leq t < T_1 + \sum\limits_{i=k}^{\infty} T_i; \\ 1, & \text{if } \ t \geq \sum\limits_{i=1}^{\infty} T_i. \end{cases}$$

It is easily verified that the transition probabilities of the Markov chain $(X_t)_{t\geq 0}$ satisfy (4.16).

Example K1. The second process is an example of a continous-time Markov chain with an instantaneous state. The state space is again the set of non-negative integers $\mathbb{N}_0$.

$$Q = \begin{pmatrix} -\infty & 1 & 1 & 1 & 1 & \cdots & & & & \cdots \\ 1 & -1 & 0 & 0 & 0 & \cdots & & & & \cdots \\ 4 & 0 & -4 & 0 & 0 & \cdots & & & & \cdots \\ 9 & 0 & 0 & -9 & 0 & \cdots & & & & \cdots \\ \vdots & & & & & \ddots & & & & \\ i^2 & 0 & & \cdots & & & 0 & -i^2 & 0 & \cdots \\ & & & & & & & & \ddots & \\ \vdots & & & \cdots & & & & & & \cdots \end{pmatrix} \tag{4.19}$$

At $i \geq 1$ the process stays an Exp(i^2)-distributed time, then jumps to 0. To better understand what happens at state 0, we first have a look at an approximation of the chain with a finite state space $\mathcal{S}_M = \{0, 1, \ldots, M\}$. Let

$$Q_M = \begin{pmatrix} -M & 1 & 1 & 1 & 1 & & \cdots & 1 \\ 1 & -1 & 0 & 0 & 0 & & \cdots & 0 \\ 4 & 0 & -4 & 0 & 0 & & \cdots & 0 \\ 9 & 0 & 0 & -9 & 0 & & \cdots & 0 \\ \vdots & & & & & \ddots & & \vdots \\ M^2 & 0 & \cdots & & & & \cdots & -M^2 \end{pmatrix} \tag{4.20}$$

Now we are back again in the setting of Section 4.2 and may use Construction 4.2 to build a jump process $(X_t^M)_{t \geq 0}$ with infinitesimal generator Q_M. At 0 this process stays an Exp(M)-distributed time before it takes a jump to a uniformly distributed element of $\mathcal{S}_M$. Note that the expected height of the jump is $(M+1)/2$ whereas the expected return time is $\approx c/M$ (with $c = \sum_{k \geq 1} k^{-2}$). In order to pass to the limit $M \to \infty$ we need an alternative construction of $(X_t^M)_{t \geq 0}$. Let

$$L_t^M := \int_0^t I\{X_s^M = 0\} ds, \ t \geq 0,$$

be the local (or occupation) time of the jump process at 0 before t. We describe an excursion of the jump process away from 0 starting at (real) time t by

- the local time L_t^M,
- its height H^M,

– its duration D^M.

To do so let

$$\Phi^M = \{(L_i^M, H_i^M, D_i^M),\ i \in I\}$$

be a Poisson point process on $\mathbb{R}^+ \times \{1, \ldots, M\} \times \mathbb{R}^+$ with intensity measure

$$\lambda(ds \times \{k\} \times du) = ds\, k^2 \exp(-k^2 u) du. \tag{4.21}$$

(Note that the intensity measure depends on M only through the underlying space.) The i-th excursion starts at real time

$$\bar{S}_i^M := L_i^M + \sum_{L_j^M < L_i^M} D_j^M$$

and ends at $\bar{S}_i^M + D_i^M$. The jump process $(X_t^M)_{t \geq 0}$ can be recovered from the point process Φ^M through

$$X_t^M = \sum_{i \in I} H_i^M I\{t \in [\bar{S}_i^M, \bar{S}_i^M + D_i^M)\}. \tag{4.22}$$

Other than Construction 4.2 the Poisson point process construction above also works for $M = \infty$:

Let $\Phi = \{(L_i, H_i, D_i),\ i \in I\}$ be a Poisson point process on $\mathbb{R}^+ \times \mathbb{N} \times \mathbb{R}^+$ with intensity λ as in (4.21). Define A_t to be the time spent away from 0 until the moment when (first) spent time t at 0,

$$A_t := \int_0^{L_t^{-1}} I\{X_s \neq 0\}\, ds,$$

where

$$L_t^{-1} := \inf\{s \geq 0 \,|\, L_s = t\}.$$

Note that

$$\begin{aligned} \mathbb{E}A_t \ &= \ \mathbb{E}\Big(\sum_{L_i < t} D_i\Big) \ = \ \mathbb{E}\Big(\sum_{k=1}^{\infty} \sum_{L_i < t,\, H_i = k} D_i\Big) \\ &= \ \sum_{k=1}^{\infty} \mathbb{E}\Big(\sum_{L_i < t,\, H_i = k} D_i\Big) \ = \ \sum_{k=1}^{\infty} \int_{(0,t) \times \mathbb{R}^+} u\, k^2 \exp(-k^2 u) du\, ds \ = \ t \sum_{k=1}^{\infty} \frac{1}{k^2} \ < \ \infty. \end{aligned}$$

Hence, $A_t \overset{a.s.}{<} \infty$ which implies

$$\bar{S}_i := L_i + \sum_{L_j < L_i} D_j \overset{a.s.}{<} \infty$$

for every $i \in I$. In analogy with (4.23) we may thus define

$$X_t := \sum_{i \in I} H_i\, I\{t \in [\bar{S}_i, \bar{S}_i + D_i)\},\ \ t \geq 0. \tag{4.23}$$

We are left with the question whether the process $(X_t)_{t \geq 0}$ has a standard transition matrix that satisfies (4.16). The answer is yes. Propeties *i)* and *ii)* as well as (4.16) are easy to verify whereas the continuity *iii)* is a little involved. We will not give a complete proof here. However, the next result gives some strong evidence.

Lemma 4.14 *For all* $\varepsilon > 0$

$$\mathbb{P}_0\{A_t \le \varepsilon t\} \ge 1 - \varepsilon \ \textit{ for all } t \textit{ sufficiently small.} \tag{4.24}$$

Assertion (4.24) states that the proportion of time before t that the process spends away from state 0 tends to 0 as $t \downarrow 0$. Indeed, note that (4.24) is equivalent to $\mathbb{P}_0\{L_t \ge (1-\varepsilon)\,t\} \ge 1-\varepsilon$ for all t sufficiently small.

Proof. For $k \in \mathbb{N}$ set

$$A_{k,t} := \int_0^{L_t^{-1}} I_{\{X_s = k\}} ds,$$

so that $A_t = \sum_{k=1}^{\infty} A_{k,t}$. Observe that

$$\begin{aligned}
\mathbb{P}_0\{A_t > \varepsilon t\} &\le \mathbb{P}_0\Big\{\sum_{k=1}^{M} A_{k,t} > 0 \ \text{ or } \sum_{k=M+1}^{\infty} A_{k,t} \ge \varepsilon t\Big\} \\
&\le \mathbb{P}_0\Big\{\sum_{k=1}^{M} A_{k,t} > 0\Big\} + \mathbb{P}_0\Big\{\sum_{k=M+1}^{\infty} A_{k,t} \ge \varepsilon t\Big\}. \qquad (4.25)
\end{aligned}$$

Now

$$\begin{aligned}
\mathbb{P}_0\Big\{\sum_{k=1}^{M} A_{k,t} > 0\Big\} &= \mathbb{P}\Big\{\Phi \cap (0,t) \times \{1,\ldots,M\} \times \mathbb{R}^+ \ne \emptyset\Big\} \\
&= 1 - \exp\Big(-\lambda((0,t) \times \{1,\ldots,M\} \times \mathbb{R}^+)\Big) \\
&= 1 - \exp(-Mt) \ \to \ 0 \ \text{ as } t \downarrow 0. \qquad (4.26)
\end{aligned}$$

For the other term on the right-hand side of (4.25) the Markov inequality gives

$$\begin{aligned}
\mathbb{P}_0\Big\{\sum_{k=M+1}^{\infty} A_{k,t} \ge \varepsilon t\Big\} &\le (\varepsilon t)^{-1} \mathbb{E}_0\Big(\sum_{k=M+1}^{\infty} A_{k,t}\Big) \\
&= (\varepsilon t)^{-1} \sum_{k=M+1}^{\infty} \mathbb{E}_0 A_{k,t} \\
&= \varepsilon^{-1} \sum_{k=M+1}^{\infty} \frac{1}{k^2} \ \to \ 0 \ \text{ as } M \to \infty. \qquad (4.27)
\end{aligned}$$

Combining (4.26) and (4.27) with (4.25) gives

$$\limsup_{t\downarrow 0} \mathbb{P}_0\{A_t > \varepsilon t\} \ \le \ \limsup_{M\to\infty} \limsup_{t\downarrow 0} \Big(\mathbb{P}_0\Big\{\sum_{k=1}^{M} A_{k,t} > 0\Big\} + \mathbb{P}_0\Big\{\sum_{k=M+1}^{\infty} A_{k,t} > \varepsilon t\Big\}\Big) = 0\,.$$

4.7 Exercises

Exercise 4.1 *Let $J = (J_{xy})$ be a stochastic matrix on $\mathcal{S}$ and let q_x, $x \in \mathcal{S}$ be a sequence of non-negative real numbers.*

a) Find a stochastic matrix $\bar{J} = (\bar{J}_{xy})$ and real numbers $\bar{q}_x$, $x \in \mathcal{S}$ with

$$\bar{J}_{xx} \in \{0,1\},\ \forall\, x \in \mathcal{S} \quad \text{and} \quad \bar{J}_{xx} = 1 \Longrightarrow \bar{q}_x = 0,$$

such that the (q, J)-jump process and the $(\bar{q}, \bar{J})$-jump process are identical in law.

b) Suppose that $\sup_{x\in\mathcal{S}} q_x =: v < \infty$. Find a stochastic matrix $\tilde{J} = (\tilde{J}_{xy})$ such that

$$P^t_{xy} \;=\; \sum_{n=0}^{\infty} e^{-vt}\,\frac{(vt)^n}{n!}\,\tilde{J}^n_{xy}$$

for every $x, y \in \mathcal{S}$ and $t \geq 0$, where P^t_{xy} are the transition probabilities of the (q, J)-jump process.

Hint: Consider first the case where $q_x = v$ for all $x \in \mathcal{S}$.

Exercise 4.2 *The **Yule process** $(X_t)_{t\geq 0}$ is a pure birth process in which each individual of the population independently gives birth in accordance with a Poisson process having rate $\lambda > 0$. Suppose that the population starts from a single founding ancestor at time 0. Show that:*

a) The time S_n of the nth jump has the same distribution as $R_1 + \cdots + R_n$, where R_j, $j \geq 1$ are independent Exp$(j\lambda)$-distributed random variables.

b) S_n has the same distribution as $\max_{1\leq i\leq n} V_i$, where the V_i, $i \geq 1$ are independent Exp(λ)-distributed random variables.

c) X_t has geometric distribution with mean $e^{\lambda t}$.

d) $\mathbb{P}\{X_{s+t} = j \,|\, X_s = i\} = \binom{j-1}{i-1} e^{-\lambda t i}(1 - e^{-\lambda t})^{j-i}$ *for all $s, t \geq 0$ and $j \geq i$.*

Exercise 4.3 *Let S_n, $n \geq 1$ be the transition times of a (q, J)-jump process $(X_t)_{t\geq 0}$, i.e.,*

$$S_{n+1} := \inf\{t > S_n : X_t \neq X_{S_n}\},\ n \geq 0 \ \text{ and } \ S_0 := 0.$$

Show that or every non-absorbing state $x \in S$

$$\lim_{h\downarrow 0} \frac{\mathbb{P}_x\{S_2 \leq h\}}{\mathbb{P}_x\{S_1 \leq h\}} = 0.$$

Exercise 4.4 *Let $(X_t)_{t\geq 0}$ be a continuous-time (rate 1) Galton-Watson process with offspring distribution $(p_j)_{j\geq 0}$. Assuming $p_1 = 0$ the Q-matrix associated with $(X_t)_{t\geq 0}$ has entries*

$$Q_{xy} := \begin{cases} -x, & \text{if } x = y,\\ x p_{y-x+1}, & \text{if } y = x-1 \text{ or } y \geq x+1,\\ 0, & \text{else.} \end{cases}$$

a) *Recursively define*

$$Y'_{n+1} \;:=\; Y'_n + (W_{n+1} - 1)\, I\{Y'_n > 0\}, \quad n \geq 0,$$

where the random variables W_n, $n \geq 1$ *are i.i.d. with distribution* $(p_j)_{j\geq 0}$. *Show that* $(Y'_n)_{n\geq 0}$ *has the same transition probabilities as the embedded Markov chain* $(Y_{n\wedge T})_{n\geq 0}$ *stopped at* $T := \min\{k \geq 0 : Y_k = 0\}$.

b) *Suppose that the mean offspring number* $m := \sum_{j=0}^{\infty} jp_j$ *is finite. Use part a) and the law of large numbers to show that* $(X_t)_{t\geq 0}$ *is regular.*

c) *Suppose that* $p_0 = p_2 = \frac{1}{2}$. *Show that*

$$\mathbb{P}_1\{X_t = 0\} = \frac{t}{2+t}, \; t \geq 0.$$

Hint: Establish a differential equation for $\phi(t) := \mathbb{P}_1\{X_t = 0\}$ *and observe that* $\mathbb{P}_x\{X_t = 0\} = (\mathbb{P}_1\{X_t = 0\})^x$.

Exercise 4.5 *Recall Pólya's urn model: Initially, there are* r *red and* s *black balls in the urn* $(r, s \in \mathbb{N})$. *At each draw the chosen ball is replaced and one extra ball of the drawn color is added to the urn. Let* $(X_{1,t})_{t\geq 0}$ *and* $(X_{2,t})_{t\geq 0}$ *be two independent Yule processes started at* $X_{1,0} = r$ *and* $X_{2,0} = s$, *respectively and write* $Z_t = X_{1,t} + X_{2,t}$, $t \geq 0$.

a) *Show that* $(Z_t)_{t\geq 0}$ *is a pure birth process with linear jump rates started at* $r + s$.

b) *Let* S_n *denote the time of the nth jump of* $(Z_t)_{t\geq 0}$. *Show that* X_{1,S_n}, $n \geq 1$ *has the same distribution as the number of red balls in a Pólya urn after* n *drawings, when initially filled with* r *red and* s *black balls.*

c) *Argue that, as* $n \to \infty$, *the asymptotic proportion of red balls in the Pólya urn is distributed like*

$$\frac{W_1}{W_1 + W_2},$$

where W_1 *and* W_2 *are independent with distribution Gamma*$(r, 1)$ *and Gamma*$(s, 1)$, *respectively.*

Hint: Use part b) of this and Exercises 3.3 and 4.2.

Exercise 4.6 *Let* $(X_t)_{t\geq 0}$ *be a* (q, J)*-jump process with* $q_x > 0$ *for all* $x \in \mathcal{S}$. *Show that, if* π *is a stationary distribution for* $(X_t)_{t\geq 0}$ *and the embedded Markov chain* $(Y_n)_{n\geq 0}$ *is irreducible, then*

$$\lim_{t\to\infty} \mathbb{P}_x\{X_t = y\} = \pi(y) \;\; \textit{for all } x, y \in \mathcal{S}.$$

Hint: Consider first the skeleton $(X_n)_{n\geq 0}$.

5 Martingales

6 Brownian motion

Bibliography

[1] Billingsley, P. (1995) *Probability and measure*, 3rd ed., Wiley, Chichester.

[2] Brémaud, P. (1999) *Markov chains. Gibbs fields, Monte Carlo simulation, and queues.* Springer, New York.

[3] Daley, D.J. and Vere-Jones, D. (1988). *An introduction to the theory of point processes.* Springer, New York.

[4] Durrett, R. (1999). *Essentials of stochastic processes.* Springer, New York.

[5] Häggström, O. (2002). *Finite Markov chains and algorithmic applications*, Cambridge University Press, Cambridge.

[6] Kallenberg, O. *Foundations of modern probability.* 2nd ed. Springer, New York.

[7] Karlin, S. and Taylor, H. M. (1975). *A first course in stochastic processes*, 2nd ed. Academic Press, New York.

[8] Karlin, S. and Taylor, H. M. (1981). *A second course in stochastic processes.* Academic Press, New York.

[9] Kingman, J.F.C. (1993). *Poisson processes.* Clarendon Press, Oxford.

[10] Norris, J. R.(1997). *Markov chains.* Cambridge University Press, Cambridge.