Advanced Topics in Markov chains

J.M. Swart

October 9, 2023

Abstract This is a short advanced course in Markov chains, i.e., Markov processes with discrete space and time. The first chapter recalls, without proof, some of the basic topics such as the (strong) Markov property, transience, recurrence, periodicity, and invariant laws, as well as some necessary background material on martingales. The main aim of the lecture is to show how topics such as harmonic functions, coupling, Perron-Frobenius theory, Doob transformations and intertwining are all related and can be used to study the properties of concrete chains, both qualitatively and quantitatively. In particular, the theory is applied to the study of first exit problems and branching processes.

Notation

$\mathbb{N}$	natural numbers $\{0, 1, \ldots\}$
$\mathbb{N}_+$	positive natural numbers $\{1, 2, \ldots\}$
$\overline{\mathbb{N}}$	$\mathbb{N} \cup \{\infty\}$
$\mathbb{Z}$	integers
$\overline{\mathbb{Z}}$	$\mathbb{Z} \cup \{-\infty, \infty\}$
$\mathbb{Q}$	rational numbers
$\mathbb{R}$	real numbers
$\overline{\mathbb{R}}$	extended real numbers $[-\infty, \infty]$
$\mathbb{C}$	complex numbers
$\mathcal{B}(E)$	Borel-σ-algebra on a topological space E
1_A	indicator function of the set A
$A \subset B$	A is a subset of B, which may be equal to B
$A^{\rm c}$	complement of A
$A \backslash B$	set difference
$\overline{A}$	closure of A
$\mathrm{int}(A)$	interior of A
$(\Omega, \mathcal{F}, \mathbb{P})$	underlying probability space
ω	typical element of Ω
$\mathbb{E}$	expectation with respect to $\mathbb{P}$
$\sigma(\ldots)$	σ-field generated by sets or random variables
$\\|f\\|_\infty$	supremumnorm $\\|f\\|_\infty := \sup_x \lvert f(x)\rvert$
$\mu \ll \nu$	μ is absolutely continuous w.r.t. ν
$f_k \ll g_k$	$\lim f_k/g_k = 0$
$f_k \sim g_k$	$\lim f_k/g_k = 1$
$o(n)$	any function such that $o(n)/n \to 0$
$O(n)$	any function such that $\sup_n o(n)/n \leq \infty$
δ_x	delta measure in x
$\mu \otimes \nu$	product measure of μ and ν
$\Rightarrow$	weak convergence of probability laws

Contents

Chapter 0

Preliminaries

0.1 Stochastic processes

Let I be a (possibly infinite) interval in $\mathbb{Z}$. By definition, a *stochastic process* with *discrete time* is a collection of random variables $X = (X_k)_{k \in I}$, defined on some underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in some measurable space $(E, \mathcal{E})$. We call the random function

$$I \ni k \mapsto X_k(\omega) \in E$$

the *sample path* of the process X. The sample path of a discrete-time stochastic process is in fact itself a random variable $X = (X_k)_{k \in I}$, taking values in the product space $(E^I, \mathcal{E}^I)$, where

$$E^I := \{x = (x_k)_{k \in I} : x_k \in E \ \forall k \in I\}$$

is the space of all functions $x : I \to E$ and $\mathcal{E}^I$ denotes the product-σ-field. It is well-known that a probability law on $(E^I, \mathcal{E}^I)$ is uniquely characterized by its finite-dimensional marginals, i.e., even if I is infinite, the law of the sample path X is uniquely determined by the finite dimensional distributions

$$\mathbb{P}\big[(X_k, \ldots, X_{k+n}) \in \cdot\,\big] \qquad (\{k, \ldots, k+n\} \subset I).$$

of the process. Conversely, if $(E, \mathcal{E})$ is a Polish space equipped with its Borel-σ-field, then by the Daniell-Kolmogorov extension theorem, any consistent collection of probability measures on the finite-dimensional product spaces $(E^J, \mathcal{E}^J)$, with $J \subset I$ a finite interval, uniquely defines a probability measure on $(E^I, \mathcal{E}^I)$. Polish spaces include many of the most commonly used spaces, such as countable spaces equipped with the discrete topology, $\mathbb{R}^d$, separable Banach spaces, and much more. Moreover, open or closed subsets of Polish spaces are Polish, as are countable carthesian products of Polish spaces, equipped with the product topology.

0.2 Filtrations and stopping times

As before, let I be an interval in $\mathbb{Z}$. A discrete *filtration* is a collection of σ-fields $(\mathcal{F}_k)_{k\in I}$ such that $\mathcal{F}_k \subset \mathcal{F}_{k+1}$ for all $k, k+1 \in I$. If $X = (X_k)_{k\in I}$ is a stochastic process, then

$$\mathcal{F}^X_k := \sigma\big(\{X_j : j \in I,\ j \leq k\}\big) \qquad (k \in I)$$

is a filtration, called the filtration *generated* by X. For any filtration $(\mathcal{F}_k)_{k\in I}$, we set

$$\mathcal{F}_\infty := \sigma\big(\bigcup_{k\in I} \mathcal{F}_k\big).$$

In particular, $\mathcal{F}^X_\infty = \sigma((X_k)_{k\in I})$.

A stochastic process $X = (X_k)_{k\in I}$ is *adapted* to a filtration $(\mathcal{F}_k)_{k\in I}$ if X_k is $\mathcal{F}_k$-measurable for each $k \in I$. Then $(\mathcal{F}^X_k)_{k\in I}$ is the smallest filtration that X is adapted to, and X is adapted to a filtration $(\mathcal{F}_k)_{k\in I}$ if and only if $\mathcal{F}^X_k \subset \mathcal{F}_k$ for all $k \in I$.

Let $(\mathcal{F}_k)_{k\in I}$ be a filtration. An $\mathcal{F}_k$- *stopping time* is a function $\tau : \Omega \to I \cup \{\infty\}$ such that the $\{0,1\}$-valued process $k \mapsto 1_{\{\tau\leq k\}}$ is $\mathcal{F}_k$-adapted. Obviously, this is equivalent to the statement that

$$\{\tau \leq k\} \in \mathcal{F}_k \qquad (k \in I).$$

If $(X_k)_{k\in I}$ is an E-valued stochastic process and $A \subset E$ is measurable, then the *first entrance time* of X into A

$$\tau_A := \inf\{k \in I : X_k \in A\}$$

with $\inf \emptyset := \infty$ is an $\mathcal{F}^X_k$-stopping time. More generally, the same is true for the *first entrance time* of X into A *after* σ

$$\tau_{\sigma,A} := \inf\{k \in I : k > \sigma,\ X_k \in A\},$$

where σ is an $\mathcal{F}_k$-stopping time. Deterministic times are stopping times (w.r.t. any filtration). Moreover, if σ, τ are $\mathcal{F}_k$-stopping times, then also

$$\sigma \vee \tau, \quad \sigma \wedge \tau$$

are $\mathcal{F}_k$-stopping times. If $f : I \cup \{\infty\} \to I \cup \{\infty\}$ is measurable and $f(k) \geq k$ for all $k \in I$, and τ is an $\mathcal{F}_k$-stopping time, then also $f(\tau)$ is an $\mathcal{F}_k$-stopping time.

If $X = (X_k)_{k\in I}$ is an $\mathcal{F}_k$-adapted stochastic process and τ is an $\mathcal{F}_k$-stopping time, then the *stopped process*

$$\omega \mapsto X_{k\wedge\tau(\omega)}(\omega) \qquad (k \in I)$$

is also an $\mathcal{F}_k$-adapted stochastic process. If $\tau < \infty$ a.s., then moreover $\omega \mapsto X_{\tau(\omega)}(\omega)$ is a random variable. If τ is an $\mathcal{F}_k$-stopping time defined on some *filtered probability space* $(\Omega, \mathcal{F}, (\mathcal{F}_k)_{k\in I}, \mathbb{P})$ (with $\mathcal{F}_k \subset \mathcal{F}$ for all $k \in I$), then the *σ-field of events observable before* τ is defined as

$$\mathcal{F}_\tau := \big\{A \in \mathcal{F}_\infty : A \cap \{\tau \leq k\} \in \mathcal{F}_k \ \forall k \in I\big\}.$$

Exercise 0.1 If $(\mathcal{F}_k)_{k\in I}$ is a filtration and σ, τ are $\mathcal{F}_k$-stopping times, then show that $\mathcal{F}_{\sigma\wedge\tau} = \mathcal{F}_\sigma \wedge \mathcal{F}_\tau$.

Exercise 0.2 Let $(\mathcal{F}_k)_{k\in I}$ be a filtration, let $X = (X_k)_{k\in I}$ be an $\mathcal{F}_k$-adapted stochastic process and let τ be an $\mathcal{F}^X_k$-stopping time. Let $Y_k := X_{k\wedge\tau}$ denote the stopped process Show that the filtration generated by Y is given by

$$\mathcal{F}^Y_k = \mathcal{F}^X_{k\wedge\tau} \qquad \big(k \in I \cup \{\infty\}\big).$$

In particular, since this formula holds also for $k = \infty$, one has

$$\mathcal{F}^X_\tau = \sigma\big((X_{k\wedge\tau})_{k\in I}\big),$$

i.e., $\mathcal{F}^X_\tau$ is the σ-algebra generated by the stopped process.

0.3 Martingales

By definition, a real stochastic process $M = (M_k)_{k\in I}$, where $I \subset \mathbb{Z}$ is an interval, is an $\mathcal{F}_k$-*submartingale* with respect to some filtration $(\mathcal{F}_k)_{k\in I}$ if M is $\mathcal{F}_k$-adapted, $\mathbb{E}[|M_k|] < \infty$ for all $k \in I$, and

$$\mathbb{E}[M_{k+1}|\mathcal{F}_k] \geq M_k \qquad (\{k, k+1\} \subset I). \tag{0.1}$$

We say that M is a *supermartingale* if the reverse inequality holds, i.e., if $-M$ is a submartingale, and a *martingale* if equality holds in (0.1), i.e., M is both a submartingale and a supermartingale. By induction, it is easy to show that (0.1) holds more generally when $k, k+1$ are replaced by more general times $k, m \in I$ with $k \leq m$.

If M is an $\mathcal{F}_k$-submartingale and $(\mathcal{F}'_k)_{k\geq 0}$ is a *smaller* filtration (i.e., $\mathcal{F}'_k \subset \mathcal{F}_k$ for all $k \in I$) that M is also adapted to, then

$$\mathbb{E}[M_{k+1}|\mathcal{F}'_k] = \mathbb{E}\big[\mathbb{E}[M_{k+1}|\mathcal{F}_k]|\mathcal{F}'_k\big] \geq E[M_k|\mathcal{F}'_k] = M_k \qquad (\{k, k+1\} \subset I),$$

which shows that M is also an $\mathcal{F}_k$-submartingale. In particular, a stochastic process M is a submartingale with respect to *some* filtration if and only if it is a submartingale with respect to its own filtration $(\mathcal{F}^M_k)_{k\in I}$. In this case, we simply say that M is a *submartingale* (resp. supermartingale, martingale).

Let $(\mathcal{F}_k)_{k\in I}$ be a filtration and let $(\mathcal{F}_{k-1})_{k\in I}$ be the filtration shifted one step to left, where we set $\mathcal{F}_{k-1} := \{\emptyset, \Omega\}$ if $k-1 \notin I$. Let $X = (X_k)_{k\in I}$ be a real $\mathcal{F}_k$-adapted stochastic process such that $\mathbb{E}[|X_k|] < \infty$ for all $k \in I$. By definition, a *compensator* of X w.r.t. the filtration $(\mathcal{F}_k)_{k\in I}$ is an $\mathcal{F}_{k-1}$-adapted real process $K = (K_k)_{k\in I}$ such that $\mathbb{E}[|K_k|] < \infty$ for all $k \in I$ and $(X_k - K_k)_{k\in I}$ is an $\mathcal{F}_k$-martingale. It is not hard to show that K is a compensator if and only if K is $\mathcal{F}_{k-1}$-adapted, $\mathbb{E}[|K_k|] < \infty$ for all $k \in I$ and

$$K_{k+1} - K_k = \mathbb{E}\big[X_{k+1} \,\big|\, \mathcal{F}_k\big] - X_k \qquad (\{k, k+1\} \subset I).$$

It follows that any two compensators must be equal up to an additive $\bigcap_{k\in I} \mathcal{F}_{k-1}$-measurable random constant. In particular, if $I = \mathbb{N}$, then because of the way we have defined $\mathcal{F}_{-1}$, such a constant must be deterministic. In this case, it is customary to put $K_0 := 0$, i.e., we call

$$K_n := \sum_{k=1}^{n} \big(\mathbb{E}\big[X_k \,\big|\, \mathcal{F}_{k-1}\big] - X_{k-1}\big) \qquad (n \geq 0)$$

the (unique) compensator of X with respect to the filtration $(\mathcal{F}_k)_{k\in\mathbb{N}}$. We note that X is a submartingale if and only if its compensator is a.s. nondecreasing.

The proof of the following basic fact can be found in, e.g., [Lach12, Thm 2.4].

Proposition 0.3 (Optional stopping) *Let $I \subset \mathbb{Z}$ be an interval, $(\mathcal{F}_k)_{k\in I}$ a filtration, let τ be an $\mathcal{F}_k$-stopping time and let $(M_k)_{k\in I}$ be an $\mathcal{F}_k$-submartingale. Then the stopped process $(M_{k\wedge\tau})_{k\in I}$ is an $\mathcal{F}_k$-submartingale.*

The following proposition is a special case of [Lach12, Prop. 2.1].

Proposition 0.4 (Conditioning on events up to a stopping time) *Let $I \subset \mathbb{Z}$ be an interval, $(\mathcal{F}_k)_{k \in I}$ a filtration, let τ be an $\mathcal{F}_k$-stopping time and let $(M_k)_{k \in I}$ be an $\mathcal{F}_k$-submartingale. Then*

$$\mathbb{E}\big[M_k \,\big|\, \mathcal{F}_{k \wedge \tau}\big] \geq M_{k \wedge \tau} \qquad (k \in I).$$

0.4 Martingale convergence

If $\mathcal{F}, \mathcal{F}_k$ $(k \geq 0)$ are σ-fields, then we say that $\mathcal{F}_k \uparrow \mathcal{F}$ if $\mathcal{F}_k \subset \mathcal{F}_{k+1}$ $(k \geq 0)$ and $\mathcal{F} = \sigma(\bigcup_{k \geq 0} \mathcal{F}_k)$. Note that this is the same as saying that $(\mathcal{F}_k)_{k \geq 0}$ is a filtration and $\mathcal{F} = \mathcal{F}_\infty$, as we have defined it above. Similarly, if $\mathcal{F}, \mathcal{F}_k$ $(k \geq 0)$ are σ-fields, then we say that $\mathcal{F}_k \downarrow \mathcal{F}$ if $\mathcal{F}_k \supset \mathcal{F}_{k+1}$ $(k \geq 0)$ and $\mathcal{F} = \bigcap_{k \geq 0} \mathcal{F}_k$.

Exercise 0.5 Let $(\mathcal{F}_k)_{k \in \mathbb{N}}$ be a filtration and let τ be an $\mathcal{F}_k$-stopping time. Show that

$$\mathcal{F}_{k \wedge \tau} \uparrow \mathcal{F}_\tau \quad \text{as } k \uparrow \infty.$$

The following proposition says that conditional expectations are continuous w.r.t. convergence of σ-fields. A proof can be found in, e.g., [Lach12, Prop. 4.12], [Chu74, Thm 9.4.8] or [Bil86, Thms 3.5.5 and 3.5.7].

Proposition 0.6 (Continuity in σ-field) *Let X be a real random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{F}_\infty, \mathcal{F}_k \subset \mathcal{F}$ $(k \geq 0)$ be σ-fields. Assume that $\mathbb{E}[|X|] < \infty$ and $\mathcal{F}_k \uparrow \mathcal{F}_\infty$ or $\mathcal{F}_k \downarrow \mathcal{F}_\infty$. Then*

$$\mathbb{E}[X \,|\, \mathcal{F}_k] \underset{k \to \infty}{\longrightarrow} \mathbb{E}[X \,|\, \mathcal{F}_\infty] \qquad \textit{a.s. and in } L^1\textit{-norm.}$$

Note that if $\mathcal{F}_k \uparrow \mathcal{F}$ and $\mathbb{E}[|X|] < \infty$, then $M_k := \mathbb{E}[X \,|\, \mathcal{F}_k]$ defines a martingale. Proposition 0.6 says that such a martingale always converges. Conversely, we would like to know for which martingales $(M_k)_{k \geq 0}$ there exists a final element X such that $M_k = \mathbb{E}[X \,|\, \mathcal{F}_k]$. This leads to the problem of martingale convergence. Since each submartingale is the sum of a martingale and a nondecreasing compensator and since nondecreasing functions always converge, we may more or less equivalently ask the same question for submartingales. For a proof of the following fact we refer to, e.g., [Lach12, Thm 4.1].

Proposition 0.7 (Submartingale convergence) *Let $(M_k)_{k \in \mathbb{N}}$ be a submartingale such that $\sup_{n \geq 0} \mathbb{E}[M_k \vee 0] < \infty$. Then there exists a random variable M_∞ with $\mathbb{E}[|M_\infty|] < \infty$ such that*

$$M_k \underset{k \to \infty}{\longrightarrow} M_\infty \quad \text{a.s.}$$

In particular, this implies that nonnegative supermartingales converge almost surely. The same is not true for nonnegative submartingales: a counterexample is one-dimensional random walk reflected at the origin.

In general, even if M is a martingale, it need not be true that $\mathbb{E}[M_\infty] \geq \mathbb{E}[M_0]$ (a counterexample is random walk stopped at the origin). We recall that a collection of random variables $(X_k)_{k\in I}$ is *uniformly integrable* if

$$\lim_{n\to\infty} \sup_{k\in I} \mathbb{E}\big[|X_k| 1_{\{|X_k|\geq n\}}\big] = 0.$$

Sufficient[1] for this is that $\sup_{k\in I} E[\psi(|X_k|)] < \infty$, where $\psi : [0,\infty) \to [0,\infty)$ is nonnegative, increasing, convex, and satisfies $\lim_{r\to\infty} \psi(r)/r = \infty$. Possible choices are for example $\psi(r) = r^2$ or $\psi(r) = (1+r)\log(1+r) - r$. It is well-known that uniform integrability and a.s. convergence of a sequence of real random variables imply convergence in L_1-norm. For submartingales, the following result is known [Lach12, Thm 4.8].

Proposition 0.8 (Final element) *In addition to the assumptions of Proposition 0.7, assume that $(M_k)_{k\in\mathbb{N}}$ is uniformly integrable. Then*

$$\mathbb{E}\big[|M_k - M_\infty|\big] \underset{k\to\infty}{\longrightarrow} 0 \quad \text{a.s.}$$

and $\mathbb{E}[M_\infty|\mathcal{F}_k] \geq M_k$ for all $k \geq 0$. If M is a martingale, then $M_k = \mathbb{E}[M_\infty|\mathcal{F}_k]$ for all $k \geq 0$.

Note that if M is a martingale, then the relation $M_k = \mathbb{E}[M_\infty|\mathcal{F}_k]$ shows that all information about the process M is hidden in its final element M_∞.

Combining Propositions 0.8 and 0.3, we see that if τ is an $\mathcal{F}_k$-stopping time such that $\tau < \infty$ a.s., $(M_k)_{k\in\mathbb{N}}$ is an $\mathcal{F}_k$-submartingale, and $(M_{k\wedge\tau})_{k\in\mathbb{N}}$ is uniformly integrable, then $\mathbb{E}[M_\tau] = \lim_{k\to\infty} \mathbb{E}[M_{k\wedge\tau}] \geq M_0$.

There also exist convergence results for 'backward' martingales $(M_k)_{k\in\{-\infty,\ldots,0\}}$.

0.5 Markov chains

Proposition 0.9 (Markov property) *Let $(E,\mathcal{E})$ be a measurable space, let $I \subset \mathbb{Z}$ be an interval and let $(X_k)_{k\in I}$ be an E-valued stochastic process. For each $n \in I$,*

[1] By the De la Valle-Poussin theorem, this condition is in fact also necessary.

set $I_n^- := \{k \in I : k \leq n\}$ *and* $I_n^+ := \{k \in I : k \geq n\}$*, and let* $\mathcal{F}_n^X := \sigma((X_k)_{k \in I_n^-})$ *be the filtration generated by* X*. Then the following conditions are equivalent.*

(i) $\mathbb{P}\big[(X_k)_{k\in I_n^-} \in A,\ (X_k)_{k\in I_n^+} \in B \,\big|\, X_n\big]$
$$= \mathbb{P}\big[(X_k)_{k\in I_n^-} \in A \,\big|\, X_n\big]\mathbb{P}\big[(X_k)_{k\in I_n^+} \in B \,\big|\, X_n\big] \text{ a.s.}$$
for all $A \in \mathcal{E}^{I_n^-}$, $B \in \mathcal{E}^{I_n^+}$, $n \in I$.

(ii) $\mathbb{P}\big[(X_k)_{k\in I_n^+} \in B \,\big|\, \mathcal{F}_n^X\big] = \mathbb{P}\big[(X_k)_{k\in I_n^+} \in B \,\big|\, X_n\big]$ *a.s. for all* $B \in \mathcal{E}^{I_n^+}$, $n \in I$.

(iii) $\mathbb{P}\big[X_{n+1} \in C \,\big|\, \mathcal{F}_n^X\big] = \mathbb{P}\big[X_{n+1} \in C \,\big|\, X_n\big]$ *a.s. for all* $C \in \mathcal{E}$, $\{n, n+1\} \subset I$.

Remarks Property (i) says that the past and future are conditionally independent given the present. Property (ii) says that the future depends on the past only through the present, i.e., after we condition on the present, conditioning on the whole past does not give any extra information. Property (iii) says that it suffices to check (ii) for single time steps.

Proof of Proposition 0.9 Set $\mathcal{G}_n^X := \sigma((X_k)_{k\in I_n^+})$. If (i) holds, then, for any $A \in \mathcal{F}_n^X$ and $B \in \mathcal{G}_n^X$, we have

$$\begin{aligned}
&\mathbb{E}\big[1_A \mathbb{P}[B \,|X_n]\big] = \mathbb{E}\big[\mathbb{E}\big[1_A\mathbb{P}[B \,|\, X_n] \,\big|\, X_n\big]\big] \\
&\quad = \mathbb{E}\big[\mathbb{E}[1_A \,|\, X_n]\mathbb{P}[B \,|\, X_n]\big] = \mathbb{E}\big[\mathbb{P}[A \,|\, X_n]\mathbb{P}[B \,|\, X_n]\big] \\
&\quad \overset{\text{(i)}}{=} \mathbb{E}\big[\mathbb{P}[A\cap B \,|\, X_n]\big] = \mathbb{P}[A \cap B],
\end{aligned}$$

where in the second equality we have pulled the $\sigma(X_n)$-measurable random variable $\mathbb{P}[B \,|\, X_n]$ out of the conditional expectation. Since this holds for arbitrary $A \in \mathcal{F}_n^X$ and since $\mathbb{P}[B \,|X_n]$ is $\mathcal{F}_n^X$-measurable, it follows that $\mathbb{P}[B \,|X_n]$ satisfies all properties of the definition of $\mathbb{P}[B \,|\, \mathcal{F}_n^X]$ and hence

$$\mathbb{P}[B \,|X_n] = \mathbb{P}[B \,|\, \mathcal{F}_n^X] \quad \text{a.s.},$$

which is just another way of writing (ii). Conversely, if (ii) holds, then for any $C \in \sigma(X_n)$,

$$\begin{aligned}
&\mathbb{E}\big[\mathbb{P}[A \,|\, X_n]\mathbb{P}[B \,|\, X_n]1_C\big] = \mathbb{E}\big[\mathbb{E}[\mathbb{P}[B \,|\, X_n]1_C 1_A \,|\, X_n]\big] \\
&\quad = \mathbb{E}\big[1_{A\cap C}\mathbb{P}[B \,|\, X_n]\big] \overset{\text{(ii)}}{=} \mathbb{E}\big[1_{A\cap C}\mathbb{P}[B \,|\, \mathcal{F}_n^X]\big] = \mathbb{P}[A\cap B\cap C],
\end{aligned}$$

where in the first equality we have pulled the $\sigma(X_n)$-measurable random variable $\mathbb{P}[B \,|\, X_n]1_C$ into the conditional expectation $\mathbb{E}[1_A \,|\, X_n]$ and in the final equality we have used the definition of $\mathbb{P}[B \,|\, \mathcal{F}_n^X]$ and the fact that $A \cap C \in \mathcal{F}_n^X$. Since this holds for any $C \in \sigma(X_n)$, it follows that $\mathbb{P}[A \,|\, X_n]\mathbb{P}[B \,|\, X_n]$ satisfies all properties of the definition of $\mathbb{P}[A \cap B \,|\, X_n]$ and hence

$$\mathbb{P}[A \,|\, X_n]\mathbb{P}[B \,|\, X_n] = \mathbb{P}[A \cap B \,|\, X_n] \quad \text{a.s.}$$

To see that (iii) is sufficient for (ii), one first proves by induction that

$$\mathbb{P}\big[X_{n+1} \in C_1, \ldots, X_{n+m} \in C_m \,\big|\, \mathcal{F}_n^X\big] = \mathbb{P}\big[X_{n+1} \in C_1, \ldots, X_{n+m} \in C_m \,\big|\, X_n\big],$$

and then uses that these sort events uniquely determine conditional probabilities of events in $\mathcal{G}_n^X$. ∎

If a process $X = (X_k)_{k \in I}$ satisfies the equivalent conditions of Proposition 0.9, then we say that X has the *Markov property*. For processes with countable state spaces, there is an easier formulation.

Proposition 0.10 (Markov chains) *Let $I \subset \mathbb{Z}$ be an interval and let $X = (X_k)_{k \in I}$ be a stochastic process taking values in a countable space S. Then X has the Markov property if and only if for each $\{k, k+1\} \subset I$ there exists a probability kernel $P_{k,k+1}(x, y)$ on S such that*

$$\begin{aligned} &\mathbb{P}[X_k = x_k, \ldots, X_{k+n} = x_{k+n}] \\ &\quad = \mathbb{P}[X_k = x_k] P_{k,k+1}(x_k, x_{k+1}) \cdots P_{k+n-1,k+n}(x_{k+n-1}, x_{k+n}) \end{aligned} \tag{0.2}$$

for all $\{k, \ldots, k+n\} \subset I$, $x_k, \ldots, x_{k+n} \in S$.

Proof See, e.g., [LP11, Thm 2.1]. ∎

If $I = \mathbb{N}$, then Proposition 0.10 shows that the finite dimensional distributions, and hence the whole law of a Markov chain X are defined by its *initial law* $\mathbb{P}[X_0 \in \cdot\,]$ and its *transition probabilities* $P_{k,k+1}(x, y)$. If the initial law and transition probabilities are given, then it is easy to see that the probability laws defined by (0.2) are consistent, hence by the Daniell-Kolmogorov extension theorem, there exists a Markov chain X, unique in distribution, with this initial law and transition probabilies.

We note that conversely, a Markov chain X determines its transition probabilities $P_{k,k+1}(x, y)$ only for a.e. $x \in S$ w.r.t. the law of X_k. If it is possible to choose the *transition kernels* $P_{k,k+1}$'s in such a way that they do not depend on k, then we say that the Markov chain is *homogeneous*. We are usually not interested in the problem to find $P_{k,k+1}$ given X, but typically we start with a given probability kernel P on S and are interested in all Markov chains that have P as their transition probability in each time step, and arbitrary initial law. Often, the word *Markov chain* is used in this more general sense. Thus, we often speak of 'the Markov chain with state space S that jumps from x to y with probability...' without specifying the initial law. From now on, we use the convention that *all Markov chains are homogeneous, unless explicitly stated otherwise.* Moreover, if we don't mention the initial law, then we mean the process started in an arbitrary initial law.

We note from Proposition 0.9 (i) that the Markov property is symmetric under time reversal, i.e., if $(X_k)_{k\in I}$ has the Markov property and $I' := \{-k : k \in I\}$, then the process $X' = (X'_k)_{k\in I'}$ defined by $X'_k := X_{-k}$ $(k \in I')$ also has the Markov property. It is usually not true, however, that X' is homogeneous if X is. An exception are stationary processes, which leads to the useful concept of reversible laws.

Exercise 0.11 (Stopped Markov chain) Let $X = (X_k)_{k\geq 0}$ be a Markov chain with countable state space S and transition kernel P, let $A \subset S$ and let $\tau_A := \inf\{k \geq 0 : X_k \in A\}$ be the first entrance time of B. Let Y be the stopped process $Y_k := X_{k\wedge\tau_A}$ $(k \geq 0)$. Show that Y is a Markov chain and determine its transition kernel. If we replace τ_A by the second entrance time of A, is Y then still Markov?

By definition, a *random mapping representation* of a probability kernel P on a countable state space S is a random variable M taking values in the space of all functions $m : S \to S$ such that

$$P(x, y) = \mathbb{P}\big[M(x) = y\big] \qquad (x, y \in S).$$

Note that if M is such a random map, then $(M(x))_{x\in S}$ are S-valued random variables such that $M(x)$ has law $P(x,\,\cdot\,)$. Thus, the kernel P determines the individual laws of the random variables $(M(x))_{x\in S}$, but says nothing about their joint law. In view of this, there are usually many different ways to make a random mapping representation of a given probability kernel. Often, the key to a good proof is choosing the right one.

If $(M_k)_{k\geq 1}$ are i.i.d. random variables with the same law as M, and X_0 is an independent S-valued random variable, then setting inductively

$$X_k := M_k(X_{k-1}) \qquad (k \geq 1)$$

defines a Markov chain $(X_k)_{k\geq 0}$ with transition kernel P and initial state X_0. Random mapping representations are essential for simulating Markov chains on a computer. In addition, they have plenty of theoretical applications, for example for coupling Markov chains with different initial states. (See Section 1.3 for an introduction to coupling.)

Example 1 Let $(Z_k)_{k\geq 1}$ be i.i.d. $\mathbb{Z}$-valued random variables and set $M_k(x) := x + Z_k$. Then the inductive formula $X_k := M_k(X_{k-1})$ with $X_0 = 0$ defines the random walk $X_n = \sum_{k=1}^n Z_k$.

Example 2 Let S_n be the group of all permutations $\pi : \{1,\ldots,n\} \to \{1,\ldots,n\}$. For each $j = 1,\ldots,n$, define $\sigma_j \in S_n$ by

$$\sigma_j(1) := k \quad \text{and} \quad \sigma_j(i) := \begin{cases} i-1 & \text{if } i \leq j, \\ i & \text{if } i > k, \end{cases} \qquad (i > 1).$$

Let $(J_k)_{k\geq 1}$ be i.i.d. uniformly distributed on $\{1,\ldots,n\}$ and set $M_k(\pi) := \sigma_{J_k} \circ \pi$. Then the inductive formula $X_k := M_k(X_{k-1})$ with X_0 the identity function defines a Markov chain with state space S_n that describes a deck of cards where in each step, we take the card that is on top and place it on an arbitrary level in the deck.

We note that it is in general not true that a function $f(X_k)$ of a Markov chain X_k are themselves Markov chains. An exception is the case when

$$\mathbb{P}\big[f(X_{k+1}) \in \cdot \,|\, X_k\big]$$

depends only on $f(X_k)$. In this case, we say that $f(X)$ is an *autonomous* Markov chain.

Lemma 0.12 (Autonomous Markov functional) *Let $X = (X_k)_{k\in I}$ be a Markov chain with countable state space S and transition kernel P. Let S' be a countable set and let $f : S \to S'$ be a function. Assume that there exists a probability kernel P' on S' such that*

$$P'(x',y') = \sum_{y:\, f(y)=y'} P(x,y') \qquad (x \in S,\ f(x) = x').$$

Then $f(X) := (f(X_k))_{k\in I}$ is a Markov chain with state space S' and transition kernel P'.

Example Let $(X_k, Y_k)_{k\geq 0}$ be a two-dimensional random walk with values in $\mathbb{Z}^2$ and transition kernel

$$\begin{aligned} P(x,y;x+1,y) &= \tfrac{1}{4}, \quad P(x,y;x-1,y) = \tfrac{1}{4}, \\ P(x,y;x,y+1) &= \tfrac{1}{4}, \quad P(x,y;x,y-1) = \tfrac{1}{4}. \end{aligned}$$

Then $(X_k)_{k\geq 0}$ is an autonomous Markov chain with values in $\mathbb{Z}$ and transition kernel

$$P'(x,x+1) = \tfrac{1}{4}, \quad P'(x,x+1) = \tfrac{1}{4}, \quad \text{and} \quad P'(x,x) = \tfrac{1}{2}.$$

0.6 Kernels, operators and linear algebra

Let $X = (X_k)_{k\in I}$ be a stochastic process taking values in a countable space S, and let P be a probability kernel on S. Then it is not hard to see that X is a Markov process with transition kernel P (and arbitrary initial law) if and only if

$$\mathbb{P}[X_{k+1} = y \,|\, \mathcal{F}_k^X] = P(X_k, y) \quad \text{a.s.} \qquad (y \in S,\ \{k, k+1\} \subset I),$$

where $(\mathcal{F}_k^X)_{k\in I}$ is the filtration generated by X. More generally, one has

$$\mathbb{P}[X_{k+n} = y \,|\, \mathcal{F}_k^X] = P^n(X_k, y) \quad \text{a.s.} \qquad (y \in S,\ n \geq 0,\ \{k, k+n\} \subset I),$$

where P^n denotes the n-th power of the transition kernel P. Here, if K, L are probability kernels on S, then we define their product as

$$KL(x, z) := \sum_{y\in S} K(x, y)L(y, z) \qquad (x, z \in S),$$

which is again a probability kernel on S. Then K^n is defined as the product of n times K with itself, where $K^0(x, y) := 1_{\{x=y\}}$. We may associate each probability kernel on S with a linear operator, acting on bounded real functions $f : S \to \mathbb{R}$, defined as

$$Kf(x) := \sum_{y\in S} K(x, y)f(y) \qquad (x \in S).$$

Then KL is just the composition of the operators K and L, and for each bounded $f : S \to \mathbb{R}$, one has

$$\mathbb{E}[f(X_{k+n}) \,|\, \mathcal{F}_k^X] = P^n f(X_k) \quad \text{a.s.} \qquad (n \geq 0,\ \{k, k+n\} \subset I), \tag{0.3}$$

and this formula holds more generally provided the expectations are well-defined (e.g., if $\mathbb{E}[|f(X_{k+n})|] < \infty$ or $f \geq 0$).

If μ is a probability measure on S and K is a probability kernel on S, then we may define a new probability measure μK on S by

$$\mu K(y) := \sum_{x\in S} \mu(x)K(x, y) \qquad (y \in S).$$

In this notation, if X is a Markov process with transition kernel P and initial law $\mathbb{P}[X_0 \in \cdot\,] = \mu$, then $\mathbb{P}[X_n \in \cdot\,] = \mu P^n$ is its law at time n.

We may view transition kernels as (possibly infinite) matrices that act on row vectors μ or column vectors f by left and right multiplication, respectively.

0.7 Strong Markov property

We assume that the reader is familiar with some basic facts about Markov chains, as taught in the lecture [LP11]. In particular, this concerns the strong Markov property, which we formulate now.

Let $X = (X_k)_{k\geq 0}$ be a Markov chain with countable state space S and transition kernel P. As usual, it goes without saying that X is homogeneous (i.e., we use the same P in each time step) and when we don't mention the initial law, we mean the process started in an arbitrary initial law. Often, it is notationally convenient to assume that our process X is always the same, while the dependence on the initial law is expressed in the choice of the probability measure on our underlying probability space.

More precisely, we assume that we have a measurable space $(\Omega, \mathcal{F})$ and a collection $X = (X_k)_{k\geq 0}$ of measurable maps $X_k : \Omega \to S$, as well as a collection $(\mathbb{P}^x)_{x\in S}$ of probability measures on $(\Omega, \mathcal{F})$, such that under the measure $\mathbb{P}^x$, the process X is a Markov chain with initial law $\mathbb{P}^x[X_0 = x] = 1$ and transition kernel P. In this set-up, if μ is any probability measure on S, then under the law $\mathbb{P} := \sum_{x\in S} \mu(x)\mathbb{P}^x$, the process X is distributed as a Markov chain with initial law μ and transition kernel P.

If $X, \mathbb{P}, \mathbb{P}^x$ are as just described and $(\mathcal{F}^X_k)_{k\geq 0}$ is the filtration generated by X, then it follows from Proposition 0.9 (ii) and homogeneity that

$$\mathbb{P}\big[(X_{n+k})_{k\geq 0} \in \cdot \,\big|\, \mathcal{F}^X_n\big] = \mathbb{P}^{X_n}\big[(X_k)_{k\geq 0} \in \cdot\,\big] \quad \text{a.s.} \tag{0.4}$$

Here, for fixed $n \geq 0$, we consider $(X_{n+k})_{k\geq 0}$ as a random variable taking values in $S^{\mathbb{N}}$ (i.e., this is the process Y defined by $Y_k := X_{n+k}$ $(k \geq 0)$). Since $S^{\mathbb{N}}$ is a nice (in particular Polish) space, we can choose a regular version of the conditional probability on the left-hand side of (0.4), i.e., this is a random probability measure on $S^{\mathbb{N}}$. Since X_n is random, the same is true for the right-hand side. In words, formula (0.4) says that given the process up to time n, the process after time n is distributed as the process started in X_n. The *strong Markov property* extends this to stopping times.

Proposition 0.13 (Strong Markov property) *Let $X, \mathbb{P}, \mathbb{P}^x$ be as defined above. Then, for any $\mathcal{F}^X_k$-stopping time τ such that $\tau < \infty$ a.s., one has*

$$\mathbb{P}\big[(X_{\tau+k})_{k\geq 0} \in \cdot \,\big|\, \mathcal{F}^X_\tau\big] = \mathbb{P}^{X_\tau}\big[(X_k)_{k\geq 0} \in \cdot\,\big] \quad \text{a.s.} \tag{0.5}$$

Proof This follows from [LP11, Thm 2.3]. ∎

Remark 1 Even if $\mathbb{P}[\tau = \infty] > 0$, formula (0.5) still holds a.s. on the event $\{\tau < \infty\}$.

Remark 2 Homogeneity is essential for the strong Markov property, at least in the (useful) formulation of (0.5).

Since this is closely related to formula (0.4), we also mention the following useful principle here.

Proposition 0.14 (What can happen must eventually happen) *Let $X = (X_k)_{k\geq 0}$ be a Markov chain with countable state space S. Let $B \subset S^{\mathbb{N}}$ be measurable and set $\rho(x) := \mathbb{P}^x[(X_k)_{k\geq 0} \in B]$. Then the event*

$$\big\{(X_{n+k})_{k\geq 0} \in B \text{ for infinitely many } n \geq 0\big\} \cup \big\{\rho(X_n) \underset{n\to\infty}{\longrightarrow} 0\big\}$$

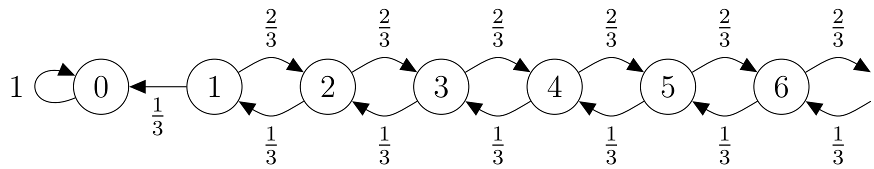
has probability one.

Proof Let $\mathcal{A}$ denote the event that $(X_{n+k})_{k\geq 0} \in B$ for some $n \geq 0$. Then by Proposition 0.6,

$$\begin{aligned}\rho(X_n) &= \mathbb{P}^{X_n}\big[(X_k)_{k\geq 0} \in B\big] = \mathbb{P}\big[(X_{n+k})_{k\geq 0} \in B \,\big|\, \mathcal{F}^X_n\big] \\ &\leq \mathbb{P}\big[\mathcal{A} \,\big|\, \mathcal{F}^X_n\big] \underset{n\to\infty}{\longrightarrow} \mathbb{P}\big[\mathcal{A} \,\big|\, \mathcal{F}^X_\infty\big] = 1_{\mathcal{A}} \quad \text{a.s.}\end{aligned}$$

In particular, this shows that $\rho(X_n) \to 0$ a.s. on the event $\mathcal{A}$. Applying the same argument to $\mathcal{A}_m := \{(X_{n+k})_{k\geq 0} \in B \text{ for some } n \geq m\}$, we see that the event $\mathcal{A}_m \cup \{\rho(X_n) \to 0\}$ has probability one for each m. Letting $m \uparrow \infty$ and observing that $\mathcal{A}_m \downarrow \{(X_{n+k})_{k\geq 0} \in B \text{ for infinitely many } n \geq 0\}$, the claim follows. ∎

Example 1 Consider a Markov chain $(X_k)_{k\geq 0}$ with state space $\mathbb{N}$ and transition kernel $P(0,0) = 1$, $P(x,x+1) = \frac{2}{3}$, $P(x,x-1) = \frac{1}{3}$ $(x \geq 1)$.



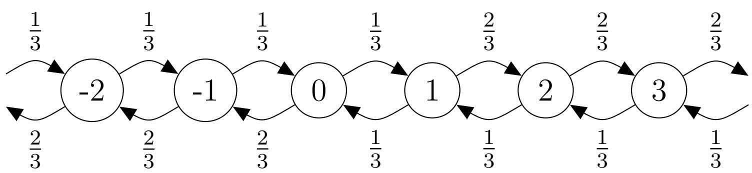
Let $B := \{(x_k)_{k\geq 0} : x_k = 0 \text{ for some } k \geq 0\}$. Then $\rho(x) = \mathbb{P}^x[(X_k)_{k\geq 0} \in B]$ is the probability that the chain ends up in the trap (absorbing state) 0. One can prove that $\rho(x) = 2^{-x}$ and hence $\rho(X_k) \to 0$ if and only if $X_k \to \infty$. Since 0 is a trap, $\{(X_{n+k})_{k\geq 0} \in B$ for infinitely many $n \geq 0\}$ is a.s. the same as $\{X_k = 0$ for some $k \geq 0\}$. Now Proposition 0.14 tells us that

$$\big\{X_k = 0 \text{ for some } k \geq 0\big\} \cup \big\{X_k \underset{k\to\infty}{\longrightarrow} \infty\big\}$$

has probability one.

Example 2 Consider a Markov chain $(X_k)_{k\geq 0}$ with state space $\mathbb{Z}$ and transition kernel

$$\begin{aligned} P(x,x+1) &:= \tfrac{2}{3}, \quad P(x,x-1) = \tfrac{1}{3} \quad (x > 0), \\ P(x,x+1) &:= \tfrac{1}{3}, \quad P(x,x-1) = \tfrac{2}{3} \quad (x \leq 0). \end{aligned}$$



Let $B := \{(x_k)_{k\geq 0} : \lim_{k\to\infty} x_k = -\infty\}$. One can check that $\rho(x) = \mathbb{P}^x[\lim_{k\to\infty} = -\infty]$ is bounded away from zero on intervals on the form $\{\ldots, k-1, k\}$ and hence $\rho(X_k) \to 0$ if and only if $X_k \to \infty$. Moreover, the event $\{(X_{n+k})_{k\geq 0} \in B$ for infinitely many $n \geq 0\}$ is the same as $\{\lim_{k\to\infty} X_k = -\infty\}$. Now Proposition 0.14 tells us that

$$\big\{X_k \underset{k\to\infty}{\longrightarrow} -\infty\big\} \cup \big\{X_k \underset{k\to\infty}{\longrightarrow} \infty\big\}$$

has probability one.

0.8 Classification of states

Let X be a Markov chain with countable state space S and transition kernel P. For each $x, y \in S$, we write $x \rightsquigarrow y$ if $P^n(x,y) > 0$ for some $n \geq 0$. Note that $x \rightsquigarrow y \rightsquigarrow z$ implies $x \rightsquigarrow z$. We say that two states x, y *communicate* if $x \rightsquigarrow y$ and $y \rightsquigarrow x$. It is easy to see that this defines an equivalence relation on S. The corresponding equivalence classes are called *communicating classes*. A Markov chain / transition kernel is called *irreducible* if all states communicate with each other, i.e., there is a single communicating class.

A state x is called *recurrent* if

$$\mathbb{P}^x[X_k = x \text{ for some } k \geq 1] = 1,$$

otherwise it is called *transient*. If two states communicate and one of them is recurrent (resp. transient), then so is the other. Fix $x \in S$, let $\tau_0 := 0$ and let

$$\tau_k := \inf\{n > \tau_{k-1} : X_n = x\} \qquad (k \geq 1)$$

be the times of the k-th visit to x after time zero. Consider the process started in $X_0 = x$. If x is recurrent, then $\tau_1 < \infty$ a.s. It follows from the strong Markov property that $\tau_2 - \tau_1$ is equally distributed with and independent of τ_1. By induction, $(\tau_k - \tau_{k-1})_{k\geq 1}$ are i.i.d. In particular, $\tau_k < \infty$ for all $k \geq 1$, i.e., the process returns to x infinitely often.

On the other hand, if x is transient, then by the same sort of argument we see that the number $N_x = \sum_{k\geq 1} 1_{\{X_k = x\}}$ of returns to x is geometrically distributed

$$\mathbb{P}^x[N_x = n] = \theta^n(1-\theta) \quad \text{where} \quad \theta := \mathbb{P}^n[X_k = x \text{ for some } k \geq 1].$$

In particular, $\mathbb{E}^x[N_x] < \infty$ if and only if x is transient.

Lemma 0.15 (Recurrent classes are closed) *Let X be a Markov chain with countable state space S and transition kernel P. Assume that $S' \subset S$ is an communicating class of recurrent states. Then $P(x,y) = 0$ for all $x \in S',\ y \in S\backslash S'$.*

Proof Imagine that $P(x,y) > 0$ for some $x \in S',\ y \in S\backslash S'$. Then, since S' is an communicating class, $y \not\rightsquigarrow x$, i.e., the process cannot return from y to x. Since $P(x,y) > 0$, this shows that the process started in x has a positive probability never to return to x, a contradiction. ∎

A state x is called *positively recurrent* if

$$\mathbb{E}^x[\inf\{n \geq 1 : X_n = x\}] < \infty.$$

Recurrent states that are not positively recurrent are called *null recurrent*. If two states are equivalent and one of them is positively recurrent (resp. null recurrent), then so is the other. From this, it is easy to see that a finite communicating class of states can never be null recurrent.

The following lemma is an easy application of the principle 'what can happen must happen' (Proposition 0.14).

Lemma 0.16 (Finite state space) *Let $X = (X_k)_{k\geq 0}$ be a Markov chain with finite state space S and transition kernel P. Let $S_{\rm pos}$ denote the set of all positively recurrent states. Then $\mathbb{P}[X_k \in S_{\rm pos}$ for some $k \geq 0] = 1$.*

By definition, the *period* of a state x is the greatest common divisor of $\{n \geq 1 : P(x,x) > 0\}$. Equivalent states have the same period. States with period one are called *aperiodic*. Irreducible Markov chains are called aperiodic if one, and hence all states have period one. If $X = (X_k)_{k\geq 0}$ is an irreducible Markov chain with period n, then $X'_k := X_{kn}$ $(k \geq 0)$ defines an aperiodic Markov chain $X' = (X'_k)_{k\geq 0}$.

The following example is of special importance.

Lemma 0.17 (Recurrence of one-dimensional random walk) *The Markov chain X with state space $\mathbb{Z}$ and transition kernel $P(k,k-1) = P(k,k+1) = \frac{1}{2}$ is null recurrent.*

Proof Note that this Markov chain is irreducible and has period two, as it takes value alternatively in the even and odd integers. Using Stirling's formula, it is not hard to show that (see [LP11, Example 2.9])

$$P^{2k}(0,0) \sim \frac{1}{\sqrt{\pi k}} \qquad \text{as } k \to \infty.$$

In particular, this shows that the expected number of returns to the origin $\mathbb{E}^0[N_0] = \sum_{k=1}^{\infty} P^{2k}(0,0)$ is infinite, hence X is recurrent. On the other hand, it is not hard to check that any invariant measure for X must be infinite, hence X has no invariant law, so it cannot be positively recurrent. ∎

We will later see that, more generally, random walks on $\mathbb{Z}^d$ are recurrent in dimensions $d = 1,2$ and transient in dimensions $d \geq 3$.

0.9 Invariant laws

By definition, an *invariant law* for a Markov process with transition kernel P and countable state space S is a probability measure μ on S that is invariant under left-multiplication with P, i.e., $\mu P = \mu$, or, written out per coordinate,

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

More generally, a (possibly infinite) measure μ on S satisfying this equation is called an *invariant measure.* A probability measure μ on S is an invariant law if and only if the process $(X_k)_{k\geq 0}$ started in the initial law $\mathbb{P}[X_0 \in \cdot\,] = \mu$ is (strictly) stationary. If μ is an invariant law, then there also exists a stationary process $X = (X_k)_{k\in\mathbb{Z}}$, unique in distribution, such that X is a Markov process with transition kernel P and $\mathbb{P}[X_k \in \cdot\,] = \mu$ for all $k \in \mathbb{Z}$ (including negative times).

The following lemma (the proof of which we skip) implies that an irreducible Markov chain that is transient or null-recurrent does not have an invariant law (even though there may exist one or more invariant measures).

Lemma 0.18 (Transient and null-recurrent states) *Let X be a Markov chain with countable state space S. Assume that $x \in S$ is transient or null recurrent. Then the Markov chain X started in an arbitrary initial law satisfies*

$$\lim_{n\to\infty} \mathbb{P}[X_n = x] = 0,$$

A detailed proof of the following theorem can be found in [LP11, Thms 2.10 and 2.26].

Theorem 0.19 (Invariant laws) *Let X be a Markov chain with countable state space S and transition kernel P. Then*

(a) *If μ is an invariant law and x is not positively recurrent, then $\mu(x) = 0$.*

(b) *If $S' \subset S$ is an communicating class of positively recurrent states, then there exists a unique invariant law μ of X such that $\mu(x) > 0$ for all $x \in S'$ and $\mu(x) = 0$ for all $x \in S\backslash S'$.*

(c) *The invariant law μ from part (b) is given by*

$$\mu(x) = \mathbb{E}^x\big[\inf\{k \geq 1 : X_k = x\}\big]^{-1}. \tag{0.6}$$

Sketch of proof For any $x \in S$, define $\mu(x)$ as in (0.6), with $1/\infty := 0$. Part (a) follows from Lemma 0.18. Assume that x is positively recurrent and let $\sigma_x := \inf\{n \geq 1 : X_n = x\}$ denote the first return time to x. One can check that

$$\mu(y) := \mathbb{E}^x[\sigma_x]^{-1}\mathbb{E}\big[\sum_{k=1}^{\sigma_x} 1_{\{X_k = y\}}\big] \qquad (y \in S) \tag{0.7}$$

defines an invariant law such that $\mu(y) > 0$ for all y in the communnicating class of x. Formula (0.6) follows from the fact that in the stationary process with invariant law μ, consecutive return times to x are i.i.d. and the process spends a $\mu(x)$-fraction of its time in x. Since (0.6) holds for any invariant law μ on the communnicating class of x, there can be only one such invariant law. ∎

Remark Using Lemma 0.15, it is not hard to prove that a general invariant law of the process is a convex combination of invariant laws that are concentrated on one communicating class of positively recurrent states.

Theorem 0.20 (Convergence to invariant law) *Let X be an irreducible, positively recurrent, aperiodic Markov chain with invariant law μ. Then the process started in any initial law satisfies*

$$\mathbb{P}[X_k = x] \underset{k\to\infty}{\longrightarrow} \mu(x) \qquad (x \in S).$$

If all states of X are transient or null recurrent, then the process started in any initial law satisfies

$$\mathbb{P}[X_k = x] \underset{k\to\infty}{\longrightarrow} 0 \qquad (x \in S).$$

Proof See [LP11, Thm 2.26] or Theorem 1.30 below. ∎

The following theorem generalizes Theorem 0.19 (b) to null recurrent Markov chains.

Theorem 0.21 (Invariant measures) *Let X be an irreducible recurrent Markov chain with countable state space S and transition kernel P. Then X has an invariant measure μ that is unique up to scalar multiples. This invariant measure satisfies $\mu(x) > 0$ for all $x \in S$.*

Sketch of proof Let $S' \subset S$ be finite and let $\sigma_{S'} := \inf\{k \geq 1 : X_k \in S'\}$ denote the first return time to S'. Since P is recurrent, setting

$$Q(x, y) := \mathbb{P}^x[X_{\sigma_{S'}} = y] \qquad (x, y \in S') \tag{0.8}$$

defines a probability kernel on S'. Since P is irreducible, the same is true for Q. If μ is an invariant measure of P, then one can check that the restriction of μ to S' is an invariant measure for Q. In particular, normalizing the restriction of μ to S' yields the unique invariant law of the probability kernel Q, which is positive recurrent in view of the finiteness of S'. Since for each $x, y \in S$ we can choose some finite $S' \supset \{x, y\}$ this argument shows that $\mu(x)/\mu(y)$ is uniquely determined and hence μ is uniquely determined up to scalar multiples. To prove existence of μ, we fix some reference point $z \in S$ and finite $z \in S_n \uparrow S$. Define positive recurrent Q_n on S_n as in (0.8) and let π_n denote their invariant laws. Let $\mu_n := \pi_n/\pi_n(z)$ be π_n normalized so that $\mu_n(z) = 1$. By our previous arguments, the μ_n are consistent in the sense that the restriction of μ_n to S_{n-1} is μ_{n-1}. In view of this, there exists a function $\mu : S \to (0, \infty)$ whose restriction to S_n is μ_n for each n. One can check that μ is an invariant measure for P. ∎

If μ is an invariant law and $X = (X_k)_{k\in\mathbb{Z}}$ is a stationary process such that $\mathbb{P}[X_k \in \cdot\,] = \mu$ for all $k \in \mathbb{Z}$, then by the symmetry of the Markov property w.r.t. time reversal, the process $X' = (X'_k)_{k\in\mathbb{Z}}$ defined by $X'_k := X_{-k}$ $(k \in \mathbb{Z})$ is also a Markov process. By stationarity, X' is moreover homogeneous, i.e., there exists a transition kernel P' such that the transition probabilities $P'_{k,k+1}$ of X' satisfy $P'(x, y) = P'_{k,k+1}(x, y)$ for a.e. x w.r.t. μ. In general, it will not be true that $P' = P$. We say that μ is a *reversible law* if μ is invariant and in addition, the stationary processes X and X' are equal in law. One can check that this is equivalent to the *detailed balance* condition

$$\mu(x)P(x, y) = P(x, y)\mu(y) \qquad (x, y \in S),$$

which says that the process X started in $\mathbb{P}[X_0 \in \cdot\,] = \mu$ satisfies $\mathbb{P}[X_0 = x,\ X_1 = y] = \mathbb{P}[X_0 = y,\ X_1 = x]$. More generally, a (possibly infinite) measure μ on S satisfying detaied balance is called an *reversible measure*. If μ is reversible measure and we define a (semi-) inner product of real functions $f : S \to \mathbb{R}$ by

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

then P is self-adjoint w.r.t. this inner product:

$$\langle f, Pg\rangle_\mu = \langle Pf, g\rangle_\mu.$$

Chapter 1

Harmonic functions

1.1 (Sub-) harmonicity

Let X be a Markov chain with countable state space S and transition kernel P. As we have seen, an invariant law of X is a vector that is invariant under left-multiplication with P. *Harmonic functions* [1] are functions that are invariant under right-multiplication with P. More precisely, we will say that a function $h : S \to \mathbb{R}$ is *subharmonic* for X if

$$\sum_y P(x,y)|h(y)| < \infty \qquad (x \in S),$$

and

$$h(x) \leq \sum_y P(x,y)h(y) \qquad (x \in S).$$

We say that h is *superharmonic* if $-h$ is subharmonic, and *harmonic* if it is both subharmonic and superharmonic.

Lemma 1.1 (Harmonic functions and martingales) *Assume that h is subharmonic for the Markov chain $X = (X_k)_{k\geq 0}$ and that $\mathbb{E}[|h(X_k)|] < \infty$ $(k \geq 0)$. Then $M_k := h(X_k)$ $(k \geq 0)$ defines a submartingale $M = (M(X_k))_{k\geq 0}$ w.r.t. to the filtration $(\mathcal{F}^X_k)_{k\geq 0}$ generated by X.*

[1] Historically, the term *harmonic function* was first used, and is still commonly used, for a smooth function $f : U \to \mathbb{R}$, defined on some open domain $U \subset \mathbb{R}^d$, that solves the *Laplace equation* $\sum_{i=1}^d \frac{\partial^2}{\partial x_i{}^2} f(x) = 0$. This is basically the same as our definition, but with our Markov chain X replaced by Brownian motion $B = (B_t)_{t\geq 0}$. Indeed, a smooth function f solves the Laplace equation if and only if $(f(B_t))_{t\geq 0}$ is a local martingale.

Proof This follows by writing (using (0.3)),

$$\mathbb{E}\big[h(X_{k+1})\,\big|\,\mathcal{F}^X_k\big]=\sum_y P(X_k,y)h(y)\geq h(X_k)\qquad(k\geq 0).$$

∎

We will say that a state x is an *absorbing state* or *trap* for a Markov chain X if $P(x,x)=1$.

Lemma 1.2 (Trapping probability) *Let X be a Markov chain with countable state space S and transition kernel P, and let $z\in S$ be a trap. Then the trapping probability*

$$h(x):=\mathbb{P}^x\big[X_k=z\text{ for some }k\geq 0\big]$$

is a harmonic function for X.

Proof Since $0\leq h\leq 1$, integrability is not an issue. Now

$$\begin{aligned}h(x)&=\mathbb{P}^x\big[X_k=z\text{ for some }k\geq 0\big]\\&=\sum_y\mathbb{P}^x\big[X_k=z\text{ for some }k\geq 0\,\big|\,X_1=y\big]\mathbb{P}^x[X_1=y]\\&=\sum_y P(x,y)\mathbb{P}^y\big[X_k=z\text{ for some }k\geq 0\big]=\sum_y P(x,y)h(y).\end{aligned}$$

∎

Lemma 1.3 (Trapping estimates) *Let X be a Markov chain with countable state space S and transition kernel P, and let $T:=\{z\in S: z\text{ is a trap}\}$. Assume that the chain gets trapped a.s., i.e., $\mathbb{P}[\exists n\geq 0\text{ s.t. }X_n\in T]=1$ (regardless of the initial law). Let $z\in T$ and let $h:S\to[0,1]$ be a subharmonic function such that $h(z)=1$ and $h\equiv 0$ on $T\backslash\{z\}$. Then*

$$h(x)\leq\mathbb{P}^x\big[X_k=z\text{ for some }k\geq 0\big]$$

If h is superharmonic, then the same holds with the inequality sign reversed.

Proof Since h is subharmonic, $M_k:=h(X_k)$ is a submartingale. Since h is bounded, M is uniformly integrable. Therefore, by Propositions 0.7 and 0.8, $M_k\to M_\infty$ a.s. and in L_1-norm, where M_∞ is some random variable such that $\mathbb{E}^x[M_\infty]\geq M_0=h(x)$. Since the chain gets trapped a.s., we have $M_\infty=h(X_\tau)$, where $\tau := \inf\{k \geq 0 : X_k \in T\}$ is the trapping time. Since $h(z) = 1$ and $h \equiv 0$ on $T\backslash\{z\}$, we have $M_\infty = 1_{\{X_\tau = z\}}$ and therefore $\mathbb{P}^x[X_\tau = z] = \mathbb{E}^x[M_\infty] \geq h(x)$. If h is superharmonic, the same holds with the inequality sign reversed. ∎

Remark 1 If $S' \subset S$ is a 'closed' set in the sense that $\mathbb{P}(x, y) = 0$ for all $x \in S'$, $y \in S\backslash S'$, then define $\phi : S \to (S\backslash S') \cup \{*\}$ by $\phi(x) := *$ if $x \in S'$ and $\phi(x) := x$ if $x \in S\backslash S'$. Now $(\phi(X_k))_{k\geq 0}$ is an autonomous (in the sense of Lemma 0.12) Markov chain that gets trapped in $*$ if and only if the original chain enters the closed set S'. In this way, Lemma 1.3 can easily be generalized to Markov chains that eventually get 'trapped' in one of finitely many equivalence classes of recurrent states.

Remark 2 Lemma 0.16 tells us that a Markov chain with finite state space eventually ends up in the set $S_{\rm pos}$ of positively recurrent states. In particular, if all states in $S_{\rm pos}$ are traps, then the chain gets trapped a.s. In general, we can partition $S_{\rm pos}$ into classes of equivalent states and use Remark 1 to calculate the probability of ending up in a given equivalence class.

Remark 3 Lemma 1.3 tells us in particular that, provided that the chain gets trapped a.s., the function h from Lemma 1.2 is the *unique* harmonic function satisfying $h(z) = 1$ and $h \equiv 0$ on $T\backslash\{z\}$. For a more general statement of this type, see Exercise 1.8 below.

Remark 4 Even in situations where it is not feasable to calculate trapping probabilities exactly, Lemma 1.3 can sometimes be used to derive lower and upper bounds for these trapping probabilities.

The following excercise shows that in the superharmonic case, the assumptions of Lemma 1.3 can be weakened considerably.

Exercise 1.4 (Weaker conditions in the superharmonic case) Let X be a Markov chain with countable state space S and transition kernel P, and let $T := \{z \in S : z \text{ is a trap}\}$. Let $z \in T$ and let $h : S \to [0, \infty)$ be a superharmonic function such that $h(z) = 1$. Show that

$$h(x) \geq \mathbb{P}^x\big[X_k = z \text{ for some } k \geq 0\big].$$

The following transformation is usually called an *h-transform* or *Doob's h-transform*. Following [LPW09], we will simply call it a *Doob transform*.[2]

[2] The term h-transform is somewhat inconvenient for several reasons. First of all, having mathematical symbols in names of chapters or articles causes all kinds of problems for referencing. Secondly, if one performs an h-transform with a function g, then should one speak of a g-transform or an h-transform? The situation becomes even more confusing when there are several functions around, one of which may be called h.

Lemma 1.5 (Doob transform) *Let X be a Markov chain with countable state space S and transition kernel P, and let $h : S \to [0,\infty)$ be a nonnegative harmonic function. Then setting $S' := \{x \in S : h(x) > 0\}$ and*

$$P^{(h)}(x,y) := \frac{P(x,y)h(y)}{h(x)} \qquad (x,y \in S')$$

defines a transition kernel $P^{(h)}$ on S'.

Proof Obviously $P^{(h)}(x,y) \geq 0$ for all $x,y \in S'$. Since

$$\sum_{y\in S'} P^{(h)}(x,y) = h(x)^{-1} \sum_{y\in S'} P(x,y)h(y) = h(x)^{-1}Ph(x) = 1 \quad (x \in S'),$$

$P^{(h)}$ is a transition kernel. ∎

Proposition 1.6 (Conditioning on the future) *Let $X = (X_k)_{k\geq 0}$ be a Markov chain with countable state space S and transition kernel P, and let $z \in S$ be a trap. Set $S' := \{y \in S : y \rightsquigarrow z\}$ and assume that $\mathbb{P}[X_0 \in S'] > 0$. Then, under the conditional law*

$$\mathbb{P}\big[(X_k)_{k\geq 0} \in \cdot \,\big|\, X_m = z \text{ for some } m \geq 0\big],$$

the process X is a Markov process in S' with Doob-transformed transition kernel $P^{(h)}$, where

$$h(x) := \mathbb{P}^x\big[X_m = z \text{ for some } m \geq 0\big]$$

satisfies $h(x) > 0$ if and only if $x \in S'$.

Proof Using the Markov property (in its strong form (0.4)), we observe that

$$\begin{aligned}
&\mathbb{P}\big[X_{n+1} = y \,\big|\, (X_k)_{0\leq k\leq n} = (x_k)_{0\leq k\leq n},\ X_m = z \text{ for some } m \geq 0\big]\\
&\quad= \mathbb{P}\big[X_{n+1} = y \,\big|\, (X_k)_{0\leq k\leq n} = (x_k)_{0\leq k\leq n},\ X_m = z \text{ for some } m \geq n+1\big]\\
&\quad= \frac{\mathbb{P}\big[X_{n+1} = y,\ X_m = z \text{ for some } m \geq n+1 \,\big|\, (X_k)_{0\leq k\leq n} = (x_k)_{0\leq k\leq n}\big]}{\mathbb{P}\big[X_m = z \text{ for some } m \geq n+1 \,\big|\, (X_k)_{0\leq k\leq n} = (x_k)_{0\leq k\leq n}\big]}\\
&\quad= \frac{P(x_n,y)\mathbb{P}^y[X_m = z \text{ for some } m \geq 0]}{\mathbb{P}^{x_n}[X_m = z \text{ for some } m \geq 1]} = P^{(h)}(x_n,y)
\end{aligned}$$

for each $(x_k)_{0\leq k\leq n}$ and y such that $\mathbb{P}[(X_k)_{0\leq k\leq n} = (x_k)_{0\leq k\leq n}] > 0$ and $x_n, y \in S'$. ∎

Remark At first sight, it is surprising that conditioning on the future may preserve the Markov property. What is essential here is that being trapped in x is a tail event, i.e., an event measurable w.r.t. the tail-σ-algebra

$$\mathcal{T} := \bigcap_{k\geq 0} \sigma(X_k, X_{k+1}, \ldots).$$

Similarly, if we condition a Markov chain $(X_k)_{0\leq k\leq n}$ that is defined on finite time interval on its final state X_n, then under the conditional law, $(X_k)_{0\leq k\leq n}$ is still Markov, although no longer homogeneous.

Exercise 1.7 (Sufficient conditions for integrability) Let $h : S \to \mathbb{R}$ be any function. Assume that $\mathbb{E}[|h(X_0)|] < \infty$ and there exists a constant $K < \infty$ such that $\sum_y P(x, y)|h(y)| \leq K|h(x)|$. Show that $\mathbb{E}[|h(X_k)|] < \infty$ $(k \geq 0)$.

Exercise 1.8 (Boundary conditions) Let X be a Markov chain with countable state space S and transition kernel P, and let $T := \{z \in S : z \text{ is a trap}\}$. Assume that the chain gets trapped a.s., i.e., $\mathbb{P}[\exists n \geq 0 \text{ s.t. } X_n \in T] = 1$ (regardless of the initial law). Show that for each bounded real function $\phi : T \to \mathbb{R}$ there exists a unique bounded harmonic function $h : S \to \mathbb{R}$ such that $h = \phi$ on T. Hint: take $h(x) := \mathbb{E}^x[\phi(X_\tau)]$, where $\tau := \inf\{k \geq 0 : X_k \in T\}$ is the trapping time.

Exercise 1.9 (Conditions for getting trapped) If we do not know a priori that a Markov chain eventually gets trapped, then the following fact is often useful. Let X be a Markov chain with countable state space S and transition kernel P, and let $h : S \to [0, 1]$ be a sub- or superharmonic function. Assume that for all $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\mathbb{P}^x\big[|h(X_1) - h(x)| \geq \delta\big] \geq \delta \quad \text{whenever } \varepsilon \leq h(x) \leq 1 - \varepsilon.$$

Show that $\lim_{k\to\infty} h(X_k) \in \{0, 1\}$ a.s. Hint: use martingale convergence to prove that $\lim_{k\to\infty} h(X_k)$ exists. Now you can use the principle 'what can happen must happen' (Proposition 0.14) to show that the limit cannot take values in $(0, 1)$.

Exercise 1.10 (Trapping estimate) Let $X, S, , P$ and h be as in Excercise 1.9. Assume that h is subharmonic and there is a point $z \in S$ such that $h(z) = 1$ and $\sup_{x\in S\setminus\{z\}} h(x) < 1$. Show that

$$h(x) \leq \mathbb{P}^x[X_k = z \text{ for some } k \geq 0].$$

Exercise 1.11 (Compensator) Let $X = (X_k)_{k\geq 0}$ be a Markov chain with countable state space S and transition kernel P, and let $f : S \to \mathbb{R}$ be a function such that $\sum_y P(x,y)|f(y)| < \infty$ for all $x \in S$. Assume that, for some given initial law, the process X satisfies $\mathbb{E}[|f(X_k)|] < \infty$ for all $k \geq 0$. Show that the compensator of $(f(X_k))_{k\geq 0}$ is given by

$$K_n = \sum_{k=0}^{n-1} \big(Pf(X_k) - f(X_k) \big) \qquad (n \geq 0).$$

Exercise 1.12 (Expected time till absorption: part 1) Let X be a Markov chain with countable state space S and transition kernel P, and let $T := \{z \in S : z \text{ is a trap}\}$. Let $\tau := \inf\{k \geq 0 : X_k \in T\}$ and assume that $\mathbb{E}^x[\tau] < \infty$ for all $x \in S$. Show that the function

$$f(x) := \mathbb{E}^x[\tau]$$

satisfies $Pf(x) - f(x) = -1$ $(x \in S\backslash T)$ and $f \equiv 0$ on T.

Exercise 1.13 (Expected time till absorption: part 2) Let X be a Markov chain with countable state space S and transition kernel P, let $T := \{z \in S : z \text{ is a trap}\}$, and set $\tau := \inf\{k \geq 0 : X_k \in T\}$. Assume that $f : S \to [0,\infty)$ satisfies $Pf(x) - f(x) \leq -1$ $(x \in S\backslash T)$ and $f \equiv 0$ on T. Show that

$$\mathbb{E}^x[\tau] \leq f(x) \qquad (x \in S).$$

Hint: show that the compensator K of $(f(X_k))_{k\geq 0}$ satisfies $K_n \leq -(n \wedge \tau)$. Now apply Exercise 1.11. To check the conditions of Exercise 1.11, you can use Exercise 1.7.

Exercise 1.14 (Foster's theorem: part 1) Let X be an irreducible Markov chain X with countable state space S and transition kernel P. Show that of the following conditions, (i) implies (ii).

(i) There exists a finite set $S' \subset S$, a function $f : S \to [0,\infty)$, and an $\varepsilon > 0$ such that

$$Pf(x) \begin{cases} < \infty & (x \in S'), \\ \leq f(x) - \varepsilon & (x \in S\backslash S'). \end{cases}$$

(ii) X is positively recurrent.

Hint: Use Lemma 0.18 and Exercise 1.11.

Exercise 1.15 (Foster's theorem: part 2) Prove the reverse implication (ii)$\Rightarrow$(i) in Exercise 1.14. Hint: Exercise 1.12.

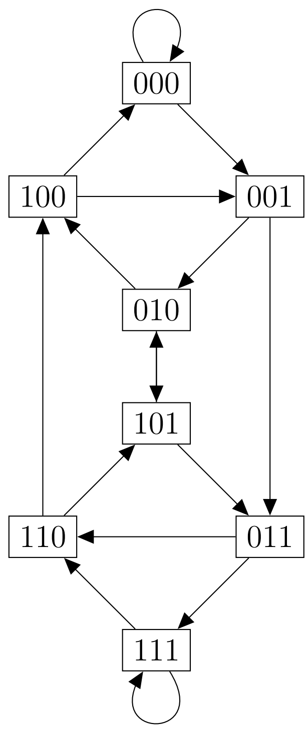
Exercise 1.16 (Absorption of random walk) Consider a random walk $W = (W_k)_{k\geq 0}$ on $\mathbb{Z}$ that jumps from x to $x+1$ with probability p and to $x-1$ with the remaining probability $q := 1-p$, where $0 < p < 1$. Fix $n \geq 1$ and set $\tau := \inf\big\{k \geq 0 : W_k \in \{0, n\}\big\}$. Calculate, for each $0 \leq x \leq n$, the probability $\mathbb{P}[W_\tau = n]$.

Exercise 1.17 (First ocurrence of a pattern: part 1)

Let $(X_k)_{k\geq 0}$ be i.i.d. Bernoulli random variables with $\mathbb{P}[X_k = 0] = \mathbb{P}[X_k = 1] = \frac{1}{2}$ $(k \geq 0)$. Set

$$\tau_{110} := \inf\big\{k \geq 0 : (X_k, X_{k+1}, X_{k+2}) = (1, 1, 0)\big\},$$

and define τ_{010} similarly. Calculate $\mathbb{P}[\tau_{010} < \tau_{110}]$. Hint: Set $\vec{X}_k := (X_k, X_{k+1}, X_{k+2})$ $(k \geq 0)$. Then $(\vec{X}_k)_{k\geq 0}$ is a Markov chain with transition probabilities as in the picture on the right. Now the problem amounts to calculating the trapping probabilities for the chain stopped at $\tau_{010} \wedge \tau_{110}$.



Exercise 1.18 (First ocurrence of a pattern: part 2) In the set-up of the previous exercise, calculate $\mathbb{E}[\tau_{110}]$ and $\mathbb{E}[\tau_{111}]$. Hints: you need to solve a system of linear equations. To find the solution, it helps to use Theorem 0.19 (c) and the fact that the uniform distribution is an invariant law. In the case of τ_{111}, it also helps to observe that $\mathbb{E}^x[\tau_{111}]$ depends only on the number of ones at the end of x.

Exercise 1.19 Let X be a Markov chain on $\mathbb{N}$ with transition kernel

$$P(0,1) = 1, \quad P(x, x+1) = p_x, \quad P(x, x-1) = 1 - p_x \quad (x \geq 1).$$

Show that the function f defined by $f(0) = 0$, $f(1) = 1$, and

$$f(x+1) := 1 + \sum_{y=1}^{x} \prod_{z=1}^{y} \frac{1-p_z}{p_z} \qquad (x \geq 1)$$

is harmonic for X. Prove that X is recurrent if and only if

$$\lim_{x\to\infty} \sum_{y=1}^{x} \prod_{z=1}^{y} \frac{1-p_z}{p_z} = \infty. \tag{1.1}$$

1.2 Random walk on a tree

In this section, we study random walk on an infinite tree in which every vertex has three neighbors. Such random walks have many interesting properties. At present they are of interest to us because they have many different bounded harmonic functions. As we will see in the next section, the situation for random walks on $\mathbb{Z}^d$ is quite different.

Let $\mathbb{T}_2$ be an infinite tree, (i.e., a connected graph without cycles) in which each vertex has degree 3 (i.e., there are three edges incident to each vertex). We will be interested in the Markov chain whose state space are the vertices of $\mathbb{T}_2$ and that jumps in each step with equal probabilities to one of the three neighboring sites.

We first need a convenient way to label vertices in such a tree. Consider a finitely generated group with generators a, b, c satisfying $a = a^{-1}$, $b = b^{-1}$ and $c = c^{-1}$. More formally, we can construct such a group as follows. Let G be the set of all finite sequences $x = x(1) \cdots x(n)$ where $n \geq 0$ (we allow for the empty sequence $\emptyset$), $x(i) \in \{a, b, c\}$ for all $1 \leq i \leq n$, and $x(i) \neq x(i+1)$ for all $1 \leq i \leq i+1 \leq n$. We define a product on V by concatenation, where we apply the rule that any two a's, b's or c's next to each other cancel each other, inductively, till we obtain an element of G. So, for example,

$$(abacb)(cab) = abacbcab, \quad (abacb)(bab) = abacbbab = abacab,$$
$$\text{and} \quad (abacb)(bcb) = abacbbcb = abaccb = abab.$$

With these rules, G is a group with unit element $\emptyset$, the empty sequence, and inverse $(x(1)\cdots x(n))^{-1} = x(n)\cdots x(1)$. Note that G is not abelian, i.e., the group product is not commutative.

We can make G into a graph by drawing an edge between two elements $x, y \in G$ if $x = ya$, $x = yb$, or $x = yc$. It is not hard to see that the resulting graph is an infinite tree in which each vertex has degree 3; see Figure 1.1.[3] We let $|x| = |x(1)\cdots x(n)| := |n|$ denote the length of an element $x \in G$. It is not hard to see that this is the same as the graph distance of x to the 'origin' $\emptyset$, i.e., the length of the shortest path connecting x to $\emptyset$.

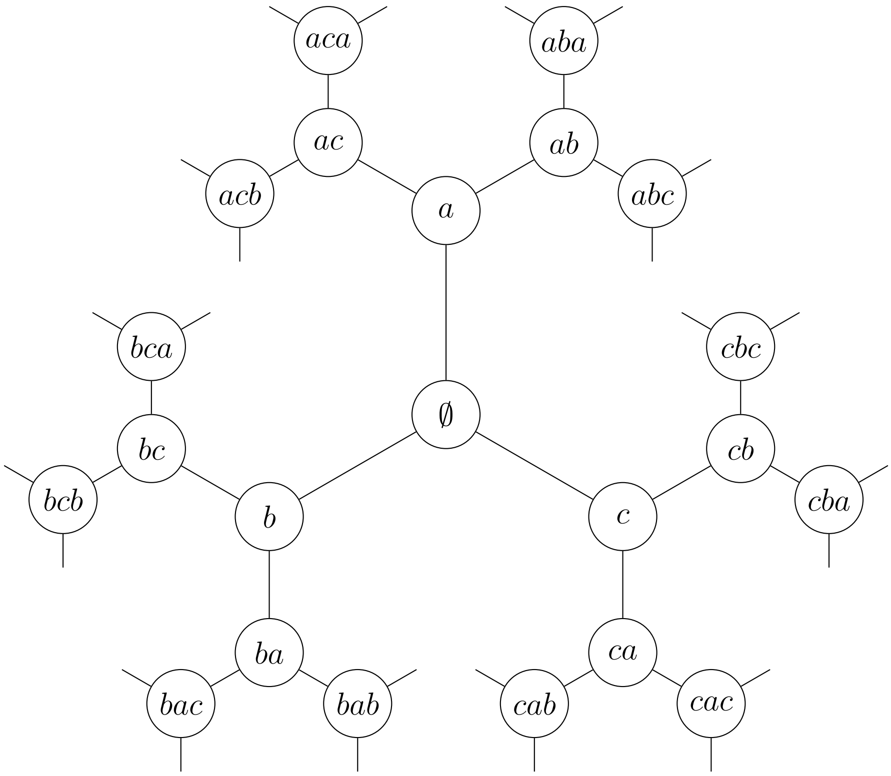


Figure 1.1: The regular tree $\mathbb{T}_2$

[3]This is a special case of a much more general construction. Let G be a finitely generated group and let $\Delta \subset G$ be a finite, symmetric (in the sense that $a \in \Delta$ implies $a^{-1} \in \Delta$) set of elements that generates G. Draw a vertex between two elements $a, b \in G$ if $a = cb$ for some $c \in \Delta$ (or equivalently, by the symmetry of Δ, if $b = c'a$ for some $c' \in \Delta$). The resulting graph is called the *left Cayley graph* associated with G and Δ. This is a very general method of making graphs with some sort of translation-invariant structure.

Let $X = (X_k)_{k\geq 0}$ be the Markov chain with state space G and transition probabilities

$$P(x, xa) = P(x, xb) = P(x, xc) = \frac{1}{3} \qquad (x \in G),$$

i.e., X jumps in each step to a uniformly chosen neighboring vertex in the graph. We call X the *nearest neighbor random walk* on G.

We observe that if X is such a random walk on G, then $|X| = (|X_k|)_{k\geq 0}$ is an autonomous (in the sense of Lemma 0.12) Markov chain with state space $\mathbb{N}$ and transition probabilities given by

$$Q(n, n-1) := \frac{1}{3} \quad \text{and} \quad Q(n, n+1) := \frac{2}{3} \qquad (n \geq 1),$$

and $Q(0, 1) := 1$.

For each $x = x(1)\cdots x(n) \in G$, let us write $x(i) := \emptyset$ if $i > n$. The following lemma shows that the random walk X is transient and walks away to infinity in a well-defined 'direction'.

Lemma 1.20 (Transience) *Let X be the random walk on G described above, started in any initial law. Then there exists a random variable $X_\infty \in \{a, b, c\}^{\mathbb{N}_+}$ such that*

$$\lim_{n\to\infty} X_n(i) = X_\infty(i) \quad a.s. \qquad (i \in \mathbb{N}_+).$$

Proof We may compare $|X|$ to a random walk $Z = (Z_k)_{k\geq 0}$ on $\mathbb{Z}$ that jumps from n to $n-1$ or $n+1$ with probabilities $1/3$ and $2/3$, respectively. Such a random walk has *independent increments*, i.e., $(Z_k - Z_{k-1})_{k\geq 1}$ are i.i.d. random variables that take the values -1 and $+1$ with probabilities $1/3$ and $2/3$. Therefore, by the strong law of large numbers, $(Z_n - Z_0)/n \to 1/3$ a.s. and therefore $Z_n \to \infty$ a.s. In particular Z visits each state only finitely often, which shows that all states are transient. It follows that the process Z started in $Z_0 = 0$ has a positive probability of not returning to 0. Since $Z_n \to \infty$ a.s. and since $|X|$ has the same dynamics as Z as long as it is in $\mathbb{N}_+$, this shows that the process started in $X_0 = a$ satisfies

$$\mathbb{P}^a\big[|X_k| \geq 1 \ \forall k \geq 1\big] = \mathbb{P}^1\big[Z_k \geq 1 \ \forall k \geq 1\big] > 0.$$

This shows that a is a transient state for X. By irreducibility, all states are transient and $|X_k| \to \infty$ a.s., which is easily seen to imply the lemma. ∎

We are now ready to prove the existence of a many bounded harmonic functions for the Markov chain X. Let

$$\partial G := \big\{x \in \{a, b, c\}^{\mathbb{N}_+} : x(i) \neq x(i+1) \ \forall i \geq 1\big\}.$$

Elements in ∂G correspond to different ways of walking to infinity. Note that ∂G is an uncountable set. In fact, if we identify elements of ∂G with points in $[0,1]$ written in base 3, then ∂G corresponds to a sort of Cantor set. We equip ∂G with the product-σ-field, which we denote by $\mathcal{B}(\partial G)$. (Indeed, one can check that this is the Borel-σ-field associated with the product topology.)

Proposition 1.21 (Bounded harmonic functions) *Let $\phi : \partial G \to \mathbb{R}$ be bounded and measurable, let X be the random walk on the tree G described above, and let X_∞ be as in Lemma 1.20. Then*

$$h(x) := \mathbb{E}^x\big[\phi(X_\infty)\big] \qquad (x \in G)$$

defines a bounded harmonic function for X. Moreover, the process started in an arbitrary initial law satisfies

$$h(X_n) \underset{n\to\infty}{\longrightarrow} \phi(X_\infty) \quad \text{a.s..}$$

Proof It follows from the Markov property (in the form (0.4)) that

$$h(x) = \mathbb{E}^x[\phi(X_\infty)] = \sum_y P(x,y)\mathbb{E}^y[\phi(X_\infty)] = \sum_y P(x,y)h(y) \qquad (x \in G),$$

which shows that h is harmonic. Since $\|h\|_\infty \leq \|\phi\|_\infty$, the function h is bounded. Moreover, by (0.4) and Proposition 0.6,

$$h(X_n) = \mathbb{E}^{X_n}[\phi(X_\infty)] = \mathbb{E}[\phi(X_\infty)\,|\,\mathcal{F}^X_n] \underset{n\to\infty}{\longrightarrow} \mathbb{E}[\phi(X_\infty)\,|\,\mathcal{F}^X_\infty] = \phi(X_\infty) \quad \text{a.s.}$$

∎

For example, in Figure 1.2, we have drawn a few values of the harmonic function

$$h(x) := \mathbb{P}^x[X_\infty(1) = a] \qquad (x \in G).$$

Although Proposition 1.21 proves that each bounded measurable function ϕ on ∂G yields a bounded harmonic function for the process X, we have not actually shown that different ϕ's yield different h's.

Lemma 1.22 (Many bounded harmonics) *Let μ be the probability measure on ∂G defined by $\mu := \mathbb{P}^\emptyset[X_\infty \in \cdot\,]$. Let $\phi, \psi : \partial G \to \mathbb{R}$ be bounded and measurable and let*

$$h(x) := \mathbb{E}^x\big[\phi(X_\infty)\big] \quad \textit{and} \quad g(x) := \mathbb{E}^x\big[\psi(X_\infty)\big] \qquad (x \in G).$$

Then $h = g$ if and only if $\phi = \psi$ a.s. w.r.t. μ.

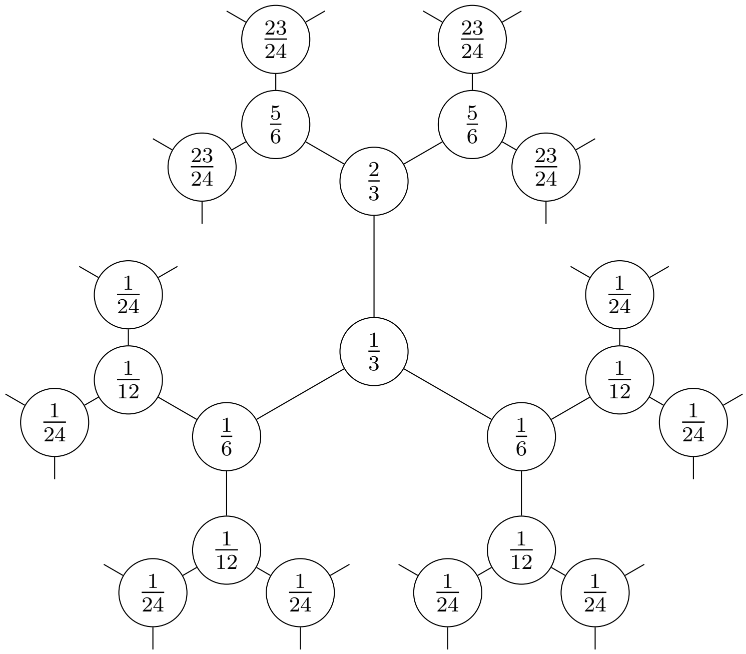


Figure 1.2: A bounded harmonic function

Proof Let us define more generally $\mu_x = \mathbb{P}^x[X_\infty \in \cdot\,]$. Since

$$\mu_x(A) = \sum_z P^n(x,z)\mathbb{P}^z[X_\infty \in \cdot\,] \le P^n(x,y)\mu_y(A)$$

$(x, y \in G,\ n \ge 0,\ A \in \mathcal{B}(\partial G))$ and P is irreducible, we see that $\mu_x \ll \mu_y$ for all $x, y \in G$, hence the measures $(\mu_x)_{x\in G}$ are all equivalent. Thus, if $\phi = \psi$ a.s. w.r.t. μ, then they are a.s. equal w.r.t. to μ_x for each $x \in G$, and therefore

$$h(x) = \int \phi \mathrm{d}\mu_x = \int \psi \mathrm{d}\mu_x = g(x) \qquad (x \in G).$$

On the other hand, if the set $\{\phi \neq \psi\}$ has positive probability under μ, then by Proposition 1.21

$$\mathbb{P}^\emptyset\big[\lim_{n\to\infty} h(X_n) \neq \lim_{n\to\infty} g(X_n)\big] > 0,$$

which shows that there must exist $x \in G$ with $h(x) \neq g(x)$. ∎

Exercise 1.23 (Escape probability) Let $Z = (Z_k)_{k\ge 0}$ be the Markov chain with state space $\mathbb{Z}$ that jumps in each step from n to $n-1$ with probability 1/3 and to $n+1$ with probability $2/3$. Calculate $\mathbb{P}^1[Z_k \geq 1 \ \forall k \geq 0]$. Hint: find a suitable harmonic function for the process stopped at zero.

Exercise 1.24 (Independent increments) Let $(Y_k)_{k\geq 1}$ be i.i.d. and uniformly distributed on $\{a, b, c\}$. Define $(X_n)_{n\geq 0}$ by the random group product (in the group G)

$$X_n := Y_1 \cdots Y_n \qquad (n \geq 1),$$

with $X_0 := \emptyset$. Show that X is the Markov chain with transition kernel P as defined above.

1.3 Coupling

For any $x = (x(1), \ldots, x(d)) \in \mathbb{Z}^d$, let $|x|_1 := \sum_{i=1}^d |x(i)|$ denote the 'L_1-norm' of x. Set $\Delta := \{x \in \mathbb{Z}^d : |x|_1 = 1\}$. Let $(Y_k)_{k\geq 1}$ be i.i.d. and uniformly distributed on Δ and let

$$X_n := \sum_{k=1}^n Y_k \qquad (n \geq 1),$$

with $X_0 := 0$. (Here we also use the symbol 0 to denote the origin $0 = (0, \ldots, 0) \in \mathbb{Z}^d$.) Then, just as in Excercise 1.24, X is a Markov chain, that jumps in each time step from its present position x to a uniformly chosen position in $x + \Delta = \{x + y : y \in \Delta\}$. We call X the *symmetric nearest neighbor random walk* on the *integer lattice* $\mathbb{Z}^d$. Sometimes X is also called *simple random walk*.

Let P denote its transition kernel. We will be interested in bounded harmonic functions for P. We will show that in contrast to the random walk on the tree, the random walk on the integer lattice has very few bounded harmonic functions. Indeed, all such functions are constant. We will prove this using *coupling*, which is a technique of much more general interest, with many applications.

Usually, when we talk about a random variable X (which may be the path of a process $X = (X_k)_{k\geq 0}$), we are not so much interested in the concrete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ that X is defined on. Rather, all that we usually care about is the law $\mathbb{P}[X \in \cdot\,]$ of X. Likewise, when we have in mind two random variables X and Y (for example, one binomially and the other normally distributed, or X and Y may be two Markov chains with possibly different initial states or transition kernels), then we usually do not *a priori* know what their joint distribution is, even if we know there individual distributions. A *coupling* of two random variables

X and Y, in the most general sense of the word, is *a way to construct X and Y together on one underlying probability space* $(\Omega, \mathcal{F}, \mathbb{P})$. More precisely, if X and Y are random variables defined on different underlying probability spaces, then a coupling of X and Y is a pair of random variables (X', Y') defined on one underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$, such that X' is equally distributed with X and Y' is equally distributed with Y. Equivalently, since the laws of X and Y are all we really care about, we may say that a *coupling* of two probability laws μ, ν defined on measurable spaces $(E, \mathcal{E})$ and $(F, \mathcal{F})$, respectively, is a probability measure ρ on the product space $(E \times F, \mathcal{E} \otimes \mathcal{F})$ such that the first marginal of ρ is μ and its second marginal is ν.

Obviously, a trivial way to couple any two random variables is to make them independent, but this is usually not what we are after. A typical coupling is designed to compare two random variables, for example by showing that they are close, or one is larger than the other. The next excercise gives a simple example.

Exercise 1.25 (Monotone coupling) Let X be uniformly distributed on $[0, \lambda]$ and let Y be exponentially distributed with mean $\lambda > 0$. Show that X and Y can be coupled such that $X \leq Y$ a.s. (Hint: note that this says that you have to construct a probability measure on $[0, \lambda] \times [0, \infty)$ that is concentrated on $\{(x, y) : x \leq y\}$ and has the 'right' marginals.) Use your coupling to prove that $\mathbb{E}[X^\alpha] \leq \mathbb{E}[Y^\alpha]$ for all $\alpha > 0$.

Now let $\Delta \subset \mathbb{Z}^d$ be as defined at the beginning of this section and let P be the transition kernel on $\mathbb{Z}^d$ defined by

$$P(x, y) := \frac{1}{2d} 1\{y - x \in \Delta\} \qquad (x, y \in \mathbb{Z}^d).$$

We are interested in bounded harmonic functions for P, i.e., bounded functions $h : \mathbb{Z}^d \to \mathbb{R}$ such that $Ph = h$. It is somewhat inconvenient that P is periodic.[4] In light of this, we define a 'lazy' modification of our transition kernel by

$$P_{\text{lazy}}(x, y) := \tfrac{1}{2} P(x, y) + \tfrac{1}{2} 1_{\{x=y\}}.$$

Obviously, $P_{\text{lazy}} f = \frac{1}{2} P f + \frac{1}{2} f$, so a function h is harmonic for P if and only if it is harmonic for P_{lazy}.

[4] Indeed, the Markov chain with transition kernel P takes values alternatively in $\mathbb{Z}^d_{\text{even}} := \{x \in \mathbb{Z}^d : \sum_{i=1}^d x(i) \text{ is even}\}$ and $\mathbb{Z}^d_{\text{odd}} := \{x \in \mathbb{Z}^d : \sum_{i=1}^d x(i) \text{ is odd}\}$.

Proposition 1.26 (Coupling of lazy walks) *Let X^x and X^y be two lazy random walks, i.e., Markov chains on $\mathbb{Z}^d$ with transition kernel $P_{\rm lazy}$, and initial states $X^x_0 = x$ and $X^y_0 = y$, $x, y \in \mathbb{Z}^d$. Then X^x and X^y can be coupled such that*

$$\exists n \geq 0 \ s.t. \ X^x_k = X^y_k \quad \forall k \geq n \quad \text{a.s.}$$

Proof We start by choosing a suitable random mapping representation. Let $(U_k)_{k\geq 1}$, $(I_k)_{k\geq 1}$, and $(W_k)_{k\geq 1}$ be collections of i.i.d. random variables, each collection independent of the others, such that for each $k \geq 1$, U_k is uniformly distributed on $\{0,1\}$, I_k is uniformly distributed on $\{1,\ldots,d\}$, and W_k is uniformly distributed on $\{-1,+1\}$. Let $e_i \in \mathbb{Z}^d$ be defined as $e_i(j) := 1_{\{i=j\}}$. Then we may construct X^x inductively by setting $X^x_0 := x$ and

$$X^x_k = X^x_{k-1} + U_k W_k e_{I_k} \qquad (k \geq 1).$$

Note that this says that U_k decides if we jump at all, I_k decides which coordinate jumps, and W_k decides whether up or down.

To construct also X^y on the same probability space, we define inductively $X^y_0 := y$ and

$$X^y_k = \begin{cases} X^y_{k-1} + (1-U_k) W_k e_{I_k} & \text{if } X^y_{k-1}(I_k) \neq X^x_{k-1}(I_k), \\ X^y_{k-1} + U_k W_k e_{I_k} & \text{if } X^y_{k-1}(I_k) = X^x_{k-1}(I_k), \end{cases} \qquad (k \geq 1).$$

Note that this says that X^x and X^y always select the same coordinate $I_k \in \{1,\ldots,d\}$ that is allowed to move. As long as X^x and X^y differ in this coordinate, they jump at different times, but after the first time they agree in this cordinate, they always increase or decrease this coordinate by the same amount at the same time. In particular, these rules ensure that

$$X^x_k(i) = X^y_k(i) \quad \text{for all } k \geq \tau_i := \inf\{n \geq 0 : X^x_n(i) = X^y_n(i)\}.$$

Since $(X^x_k, X^y_k)_{k\geq 0}$ is defined in terms of i.i.d. random variables $(U_k, I_k, W_k)_{k\geq 1}$ by a random mapping representation, the joint process (X^x, X^y) is clearly a Markov chain. We have already seen that X^x, on its own, is also a Markov chain, with the right transition kernel $P_{\rm lazy}$. It is straightforward to check that $\mathbb{P}[X^y_k = z \,|\, (X^x_k, X^y_k)] = P_{\rm lazy}(X^y_k, z)$ a.s. In particular, this transition probability depends only on X^y_k, hence by Lemma 0.12, X^y is an autonomous Markov chain with transition kernel $P_{\rm lazy}$.

In view of this, our claim will follow provided we show that $\tau_i < \infty$ a.s. for each $i = 1,\ldots,d$. Fix i and define inductively $\sigma_0 := 0$ and

$$\sigma_k := \inf\{k > \sigma_{k-1} : I_k = i\}.$$

Consider the difference process

$$D_k := X^x_{\sigma_k} - X^y_{\sigma_k} \qquad (k \geq 0).$$

Then $D = (D_k)_{k\geq 0}$ is a Markov process on $\mathbb{Z}$ that in each step jumps from z to $z+1$ or $z-1$ with equal probabilities, except when it is in zero, which is a trap. In other words, this says that D is a simple random walk stopped at the first time it hits zero. By Lemma 0.17, there a.s. exists some (random) $k \geq 0$ such that $D_k = 0$ and hence $\tau_i = \sigma_k < \infty$ a.s. ∎

As a corollary of Proposition 1.26, we obtain that all bounded harmonic functions for nearest-neighbor random walk on the d-dimensional integer lattice are constant.

Corollary 1.27 (Bounded harmonic functions are constant) *Let $P(x,y) = (2d)^{-1}1_{\{|x-y|_1=1\}}$ be the transition kernel of nearest-neighbor random walk on the d-dimensional integer lattice $\mathbb{Z}^d$. If $h : \mathbb{Z}^d \to \mathbb{R}$ is bounded and satisfies $Ph = h$, then h is constant.*

Proof Couple X^y and X^y as in Proposition 1.26. Since h is harmonic and bounded, $(h(X^x_k))_{k\geq 0}$ and $(h(X^y_k))_{k\geq 0}$ are martingales. It follows that

$$\begin{aligned} h(x) - h(y) &= \mathbb{E}[h(X^x_k)] - \mathbb{E}[h(X^y_k)] \\ &= \mathbb{E}[h(X^x_k) - h(X^y_k)] \leq 2\|h\|_\infty \mathbb{P}[X^x_k \neq X^y_k] \underset{k\to\infty}{\longrightarrow} 0 \end{aligned}$$

for each $x, y \in \mathbb{Z}^d$, proving that h is constant. ∎

Remark Actually, a much stronger statement than Corollary 1.27 is true: for nearest-neighbor random walk on $\mathbb{Z}^d$, all nonnegative harmonic functions are constant. This is called the strong Liouville property, see [Woe00, Corollary 25.5]. In general, the problem of finding all positive harmonic functions for a Markov chain leads to the (rather difficult) problem of determining the *Martin boundary* of a Markov chain.

1.4 Convergence in total variation norm

In this section we turn our attention away from harmonic functions and instead show another application of coupling. We will use coupling to give a proof of the statement in Theorem 0.20 (stated without proof in the Introduction) that any aperiodic, irreducible, positively recurrent Markov chain is ergodic, in the sense that regardless of the initial state, its law at time n converges to the invariant law as $n \to \infty$.

Recall that the *total variation distance* between two probability measures μ, ν on a countable set S is defined as

$$\|\mu - \nu\|_{\mathrm{TV}} := \max_{A \subset S} \big|\mu(A) - \nu(A)\big|$$

The following lemma gives another formula for $\|\cdot\|_{\mathrm{TV}}$.

Lemma 1.28 (Total variation distance) *For any probability measures μ, ν on a countable set S, one has*

$$\|\mu - \nu\|_{\mathrm{TV}} = \sum_{x:\,\mu(x) \geq \nu(x)} \big(\mu(x) - \nu(x)\big) = \sum_{x:\,\mu(x) < \nu(x)} \big(\nu(x) - \mu(x)\big) = \tfrac{1}{2} \sum_{x \in S} \big|\mu(x) - \nu(x)\big|.$$

Proof Set $S_> := \{x \in S : \mu(x) > \nu(x)\}$, $S_= := \{x \in S : \mu(x) = \nu(x)\}$, and $S_< := \{x \in S : \mu(x) < \nu(x)\}$. Define finite measures μ^+, ν^+ and ρ by

$$\rho(x) := \mu(x) \wedge \nu(x), \quad \mu^+(x) := \mu(x) - \rho(x), \quad \text{and} \quad \nu^+(x) := \nu(x) - \rho(x).$$

Note that μ^+ is concentrated on $S_>$ and ν^+ is concentrated on $S_<$. For any $A \subset S$,

$$\mu(A) - \nu(A) = \mu^+(A \cap S_>) - \nu^+(A \cap S_<).$$

It follows that

$$-\nu^+(S_<) \leq \mu(A) - \nu(A) \leq \mu^+(S_>)$$

where either inequality may be an equality for a suitable choice of A ($A = S_<$ or $A = S_>$, respectively). Here

$$\mu^+(S_>) = \sum_{x \in S_>} \big(\mu(x) - \rho(x)\big) = 1 - \sum_{x \in S} \rho(x) = \nu^+(S_<).$$

∎

Lemma 1.29 (Coupling and total variation distance) *Let S be a countable set, and let X and Y be S-valued random variables with laws μ and ν respectively. Then*

$$\|\mu - \nu\|_{\mathrm{TV}} \leq \mathbb{P}[X \neq Y]. \tag{1.2}$$

Moreover, given μ, ν, we can couple random variables X and Y with laws μ, ν such that equality holds in (1.2).

Proof Using notation as in the previous proof, set $p := \mu^+(S_>) = \nu^+(S_<)$. Let $X_>, X_<, X_=$ be random variables such that

$$p\mathbb{P}[X_> = x] = \mu^+(x), \quad p\mathbb{P}[X_< = x] = \nu^+(x), \quad \text{and} \quad (1-p)\mathbb{P}[X_= = x] = \rho(x)$$

$(x \in S)$, and let B be an independent Bernoulli random variable with $\mathbb{P}[B = 1] = p$. Set

$$X := \begin{cases} X_> & \text{if } B = 1, \\ X_= & \text{if } B = 0, \end{cases} \quad \text{and} \quad X := \begin{cases} X_< & \text{if } B = 1, \\ X_= & \text{if } B = 0. \end{cases}$$

Then $\mathbb{P}[X = x] = p\mathbb{P}[X_> = x] + (1-p)\mathbb{P}[X_= = x] = \mu^+(x) + \rho(x) = \mu(x)$ $(x \in S)$ and in the same way we see that Y has law ν. Since $\mathbb{P}[X \neq Y] = \mathbb{P}[B = 1] = p = \|\mu - \nu\|_{\mathrm{TV}}$ we see that equality holds in (1.2). ∎

Remark Lemmas 1.28 and 1.29 can be generalized to probability measures μ, ν on a general measurable space $(S, \mathcal{S})$. Let $S_> := \{x \in S : \mathrm{d}\mu/(\mathrm{d}\mu + \mathrm{d}\nu) > \frac{1}{2}\}$, where $\mathrm{d}\mu/(\mathrm{d}\mu + \mathrm{d}\nu)$ denotes the Radon-Nikodym derivative of μ with respect to $\mu + \nu$, define $S_<$ similarly, and for measurable A set

$$\begin{aligned} &\mu^+(A) := \mu(A \cap S_>) - \nu(A \cap S_>), \quad \nu^+(A) := \nu(A \cap S_<) - \mu(A \cap S_<), \\ &\quad \text{and} \quad \rho(A) := \mu(A) - \mu^+(A) = \nu(A) - \nu^+(A). \end{aligned}$$

Then the rest of Lemmas 1.28 and 1.29, including the proofs, are the same as four countable spaces.

Theorem 1.30 (Convergence to invariant law) *Let X be an irreducible, aperiodic, positively recurrent Markov chain with transition kernel P, state space S, and invariant law μ. Then the process started in any initial law satisfies*

$$\big\|\mathbb{P}[X_n \in \cdot\,] - \mu\big\|_{\mathrm{TV}} \underset{n\to\infty}{\longrightarrow} 0.$$

Proof We take the existence of an invariant law μ as proven. Uniqueness will follow from our proof. Let X and $\overline{X}$ be two independent Markov chains with transition kernel P, where X is started in an arbitrary initial law and $\mathbb{P}[\overline{X}_0 \in \cdot\,] = \mu$. It is easy to see that the joint process $(X, \overline{X}) = (X_k, \overline{X}_k)_{k\geq 0}$ is a Markov process with state space $S \times S$. Let us denote its transition kernel by P_2, i.e., by independence,

$$P_2\big((x,\overline{x}),(y,\overline{y})\big) = P(x,y)P(\overline{x},\overline{y}) \qquad (x,\overline{x},y,\overline{y} \in S).$$

We claim that P_2 is irreducible. Fix $x, \overline{x}, y, \overline{y} \in S$. Since P is irreducible *and* aperiodic, it is not hard to see that there exists an $m_1 \geq 1$ such that $P^n(x,y) > 0$

for all $n \geq m_1$. Likewise, there exists an $m_2 \geq 1$ such that $P^n(\overline{x}, \overline{y}) > 0$ for all $n \geq m_2$. Choosing $n \geq m_1 \vee m_2$, we see that

$$P_2^n\big((x,\overline{x}),(y,\overline{y})\big) = P^n(x,y)P^n(\overline{x},\overline{y}) > 0,$$

proving that P_2 is irreducible.

By Theorem 0.19 (a) and (b), an irreducible Markov chain is positively recurrent if and only if it has an invariant law. Obviously, the product measure $\mu \otimes \mu$ is an invariant law for P_2, so P_2 is positively recurrent. In particular, this proves that the stopping time

$$\tau := \inf\{k \geq 0 : X_k = \overline{X}_k\}$$

is a.s. finite and has, in fact, finite expectation. Let $X' = (X'_k)_{k\geq 0}$ be the process defined by

$$X'_k := \begin{cases} X_k & \text{if } k < \tau, \\ \overline{X}_k & \text{if } k \geq \tau. \end{cases}$$

It is not hard to see that X' is a Markov chain with transition kernel P and initial law $\mathbb{P}[X'_0 \in \cdot\,] = \mathbb{P}[X_0 \in \cdot\,]$, hence X' is equal in law with X. Now by Lemma 1.29

$$\big\|\mathbb{P}[X_n \in \cdot\,] - \mu\big\|_{\mathrm{TV}} \leq \mathbb{P}\big[X'_k \neq \overline{X}_k\big] = \mathbb{P}[k < \tau] \underset{k\to\infty}{\longrightarrow} 0.$$

∎

Exercise 1.31 (Periodic kernels) Show that the probability kernel P_2 in the proof of Theorem 1.30 is not irreducible if P is periodic.

Chapter 2

Positive eigenfunctions

2.1 Introduction

In the previous chapter, we have seen that positive solutions h to the equation $Ph = h$ are harmonic functions of the Markov chain with transition kernel P and that they give us information about the limit behavior of a Markov chain as time tends to infinity; e.g., they tell us in which trap a chain ends up or to which part of the (Martin) boundary a transient Markov chain converges. Moreover, through the Doob transform $P^{(h)}(x,y) := h(x)^{-1}P(x,y)h(y)$ they give us information about the process conditioned on some form of long-time behavior. Also we have seen that functions f such that $Pf = f - 1$ tell us about the expected time before a Markov chain gets trapped, which through Foster's theorem yields necessary and sufficient conditions for positive recurrence.

In the present chapter, we will look at positive eigenfunctions of P, i.e., solutions of the equation $Ph = ch$. These will tell us, e.g., that in many situations a Markov chain gets trapped exponentially fast. Through a generalized Doob transform, such functions also give information about Markov chains conditioned not to get trapped for a long time. For reasons that will become clear soon, it is useful to generalize a bit and replace P by a general nonnegative matrix $A = (A(x,y))_{x,y\in S}$ indexed by a countable set S. Here "nonnegative" means[1] that $A(x,y) \geq 0$ for all

[1]The statement that a matrix A is nonnegative should not be confused with the statement that A is *nonnegative definite*. The latter concept, which is defined only for linear spaces that are equipped with an inner product $\langle \phi, \psi \rangle$, means that $\langle \phi, A\phi \rangle$ is real and nonnegative for all ϕ. One can show that A is nonnegative definite if and only if there exists an orthonormal basis of eigenvectors of A and all eigenvalues are nonnegative. Operators whose matrix is nonnegative

x, y. We define irreducibility and aperiodicity of nonnegative matrices just as for probability kernels, i.e., A is irreducible if $\forall x, y \ \exists n$ s.t. $A^n(x,y) > 0$ and $x \in S$ is aperiodic if the greatest common divisor of $\{n \geq 1 : A^n(x,x) > 0\}$ is one.

2.2 The spectral radius

The following lemma was first proved by Kingman [Kin63]. The quantity $\rho(A)$ below is called the *spectral radius* of A. This name is motivated by the fact that if S is finite, then $\rho(A) = \sup\{|\lambda| : \lambda \text{ is a complex eigenvalue of } A\}$ (see Lemma A.1 in the appendix). This formula does not hold in general for infinite S, but we use the name spectral radius regardless.

Lemma 2.1 (Spectral radius) *Let A be a nonnegative matrix indexed by a countable set S and let $x \in S$ be aperiodic. Then the limit*

$$\rho(A) := \lim_{n\to\infty} \big(A^n(x,x)\big)^{1/n} = \sup_{n\geq 1} \big(A^n(x,x)\big)^{1/n} \in (0,\infty] \tag{2.1}$$

exists. Similarly, if x has period k, then

$$\rho(A) := \lim_{n\to\infty} \big(A^{kn}(x,x)\big)^{1/kn} = \sup_{n\geq 1} \big(A^{kn}(x,x)\big)^{1/kn} \in (0,\infty]. \tag{2.2}$$

If A is irreducible, then the limit in (2.1) does not depend on $x \in S$.

The proof of Lemma 2.1 follows from a superadditivity argument. By definition, a function $f : \mathbb{N}_+ \to \mathbb{R}$ is *subadditive* if

$$f(n+m) \leq f(n) + f(m) \qquad (n, m \geq 1),$$

and *superadditive* if the reverse inequality holds, i.e., if $-f$ is subadditive. The following simple lemma has many applications. Proofs can be found in many places, e.g. [Lig99, Thm B.22].

Lemma 2.2 (Fekete's lemma) *If $f : \mathbb{N}_+ \to [-\infty, \infty)$ is subadditive, then the limit*

$$\lim_{n\to\infty} \frac{1}{n} f(n) = \inf_{n\geq 1} \frac{1}{n} f(n)$$

exists in $[-\infty, \infty)$. The same conclusion holds if f takes values in $[-\infty, \infty]$ (with the convention $\infty - \infty = \infty$) but $\{n \geq 1 : f(n) = \infty\}$ is finite.

---

w.r.t. some basis, by contrast, need not be diagonalizable and their eigenvalues can be negative or even complex.

Proof Note that we can always extend f to a subadditive function $f : \mathbb{N} \to \mathbb{R}$ by setting $f(0) = 0$. Fix $m \geq 1$ and for each $n \geq 0$ write $n = k_m(n)m + r_m(n)$ where $k_m(n) \geq 0$ and $0 \leq r_m(n) < m$, i.e., $k_m(n)$ is n/m rounded off to below and $r_m(n)$ is the remainder. Setting $s_m := \sup_{1 \leq r < m} f(r)$, we see that

$$\frac{f(n)}{n} = \frac{f\big(k_m(n)m + r_m(n)\big)}{k_m(n)m + r_m(n)} \leq \frac{k_m(n)f(m) + s_m}{k_m(n)m} \underset{n\to\infty}{\longrightarrow} \frac{f(m)}{m},$$

which proves that

$$\limsup_{n\to\infty} \frac{f(n)}{n} \leq \frac{f(m)}{m} \qquad (m \geq 1).$$

Taking the infimum over m we conclude that

$$\limsup_{n\to\infty} \frac{f(n)}{n} \leq \inf_{m\geq 1} \frac{f(m)}{m}.$$

This shows in particular that the limit superior is less or equal than the limit inferior, hence the limit exists. Moreover, the limit (which equals the limit superior) is given by the infimum. Since f takes values in $[-\infty, \infty)$, the infimum is clearly less than $+\infty$. To see that we can allow f to take the value $+\infty$ finitely often, it suffices to note that this implies $s_m < \infty$ for m sufficiently large and we need the arguments above only for n, m sufficiently large. ∎

Proof of Lemma 2.1 It suffices to prove the statement for aperiodic x. The general statement then follows since if x has period k with respect to A, then x is aperiodic with respect to A^k. Since

$$A^{n+m}(x, x) = \sum_y A^n(x, y)A^m(y, x) \geq A^n(x, x)A^m(x, x),$$

we see that the function $n \mapsto \log A^n(x, x)$ is superadditive or equivalently $n \mapsto -\log A^n(x, x)$ is subadditive. Here we allow for the case that $A^n(x, x) = \infty$ which means $-\log A^n(x, x) = -\infty$ and $A^n(x, x) = 0$ which means $-\log A^n(x, x) = +\infty$. After taking $-\log$, the convention $0 \cdot \infty := 0$ translates into the convention $\infty - \infty = \infty$. Since we are assuming that $x \in S$ is aperiodic, $A^n(x, x) = 0$ for finitely many n which means $-\log A^n(x, x) = +\infty$ for finitely many n and hence Fekete's lemma is applicable and tells us that

$$\lim_{n\to\infty} \frac{1}{n} \log A^n(x, x) = \sup_{n\geq 1} \frac{1}{n} \log A^n(x, x) \in (-\infty, \infty].$$

Taking exponentials, we see that the limit in (2.1) exists in $(0, \infty]$ and equals the supremum.

Let $\rho_x(A) := \rho(A)$ denote the limit in (2.1), which may in general depend on $x \in S$. We observe that (2.1) says that

$$A^n(x,x) = e^{n \log \rho_x(A) - o(n)} \quad (n \geq 1) \quad \text{with} \quad 0 \leq o(n)/n \underset{n\to\infty}{\longrightarrow} 0. \tag{2.3}$$

If A is irreducible, then for every $y \in S$ we can find $k, m \geq 1$ such that $A^k(y,x) > 0$ and $A^m(x,y) > 0$. Now the estimate

$$A^{k+n+m}(y,y) \geq A^k(y,x)A^n(x,x)A^m(x,y) \tag{2.4}$$

shows that $\rho_y(A) \geq \rho_x(A)$, and reversing the roles of x and y we also obtain $\rho_x(A) \geq \rho_y(A)$, so $\rho_x(A)$ does not depend on $x \in S$. ∎

Exercise 2.3 Let A be an irreducible nonnegative matrix and let $f : S \to (0, \infty)$ be a function such that $Af \leq Kf$ for some $K < \infty$. Prove that $\rho(A) \leq K$.

2.3 R-recurrence

Let A be a nonnegative matrix indexed by a countable set S and let $h : S \to (0, \infty)$ be a positive eigenfunction of A, i.e., $Ah = ch$ for some $c > 0$. Then setting

$$P(x,y) := c^{-1}h(x)^{-1}A(x,y)h(y) \qquad (x, y \in S) \tag{2.5}$$

defines a probability kernel on S. (Indeed, $\sum_y P(x,y) = c^{-1}h(x)^{-1}Ah(x) = 1$.) Since this generalizes the Doob transform of Lemma 1.5, we call this a *generalized Doob transform*. By definition, we say that an irreducible nonnegative matrix A is *R-recurrent*[2] if there exists a positive eigenfunction h and a positive constant c such that (2.5) defines a *recurrent* probability kernel P. The following facts were proved by David Vere-Jones [Ver62, Ver67].

[2]Originally, the letter R was mathematical notation for $1/\rho(A)$. This then gave rise to a number of similar definitions such as r-recurrence (where r can be any real constant) [Ver67] and λ-recurrence (which applies in a continuous time setting) [Kin63]. For us the 'R' in the words R-transience, R-recurrence etc. will just be part of the name and not refer to any mathematical constant.

Theorem 2.4 (R-recurrence) *An irreducible nonnegative matrix A is R-recurrent if and only if $\rho(A) < \infty$ and*

$$\sum_{n=1}^{\infty} \rho(A)^{-n} A^n(x,x) = \infty \tag{2.6}$$

for some, and hence for all $x \in S$. In this case, there exists a function $h : S \to (0,\infty)$, which is unique up to scalar multiples, and a unique constant $c > 0$, such that (2.5) defines a recurrent probability kernel. Moreover, $c = \rho(A)$.

We will not give a full proof of Theorem 2.4 although we will prove most of the statements and we give a sketch of the proof of the remaining part in Section 2.7. Here, we only prove that the condition (2.6) is necessary for R-recurrence. We start with a preliminary lemma.

Lemma 2.5 (Spectral radius of probability kernels) *If P is an irreducible probability kernel, then $\rho(P) \leq 1$. If P is recurrent, then $\rho(P) = 1$.*

Proof Since $P^n(x,x) \leq 1$ for all n, it is clear from (2.1) that $\rho(P) \leq 1$. If $\rho(P) < 1$, then (2.3) shows that $P^n(x,x)$ tends to zero exponentially fast. In particular, this shows that the expected number of returns to x,

$$\sum_{n \geq 1} P^n(x,x) \tag{2.7}$$

is finite, which proves that P is transient. (Recall from Section 0.8 that the expected number of returns to x is finite if and only if P is transient.) ∎

Lemma 2.6 (Necessary conditions for R-recurrence) *Let A be an irreducible nonnegative matrix indexed by a countable set S. Assume that for some $h : S \to (0,\infty)$ and $0 < c < \infty$, formula (2.5) defines a recurrent probability kernel. Then $c = \rho(A)$ and A satisfies (2.6).*

Proof We observe that

$$\begin{aligned} P^2(x,z) &= \sum_y c^{-1} h(x)^{-1} A(x,y) h(y) c^{-1} h(y)^{-1} A(y,z) h(z) \\ &= c^{-2} h(x)^{-1} A^n(x,z) h(z), \end{aligned}$$

and more generally (with $A^0 := I$, the identity matrix)

$$P^n(x,y) = c^{-n}h(x)^{-1}A^n(x,y)h(y) \qquad (n \geq 0). \tag{2.8}$$

In particular, $P^n(x,x) = c^{-n}A^n(x,x)$ and hence by (2.1)

$$\rho(P) = c^{-1}\rho(A).$$

Since P is recurrent, $\rho(P) = 1$ by Lemma 2.5 and hence $c = \rho(A)$. The fact that P is recurrent moreover implies that the sum in (2.7) is infinite, which by the fact that $P^n(x,x) = c^{-n}A^n(x,x)$ implies (2.6). ∎

As a consequence of Theorem 2.4 we can also prove the following result.

Theorem 2.7 (Unique eigenfunction) *Let A be an irreducible, R-recurrent nonnegative matrix indexed by a countable set S. Then there exists a function $h : S \to (0,\infty)$, unique up to scalar multiples, such that $Ah = \rho(A)h$.*

Proof Since A is R-recurrent, there exists a function $h : S \to (0,\infty)$ and a constant $c > 0$ such that (2.5) defines a recurrent probability kernel P. Since P is a probability kernel, $\sum_y P(x,y) = Ah(x)/ch(x) = 1$ $(x \in S)$ which shows that $Ah = ch$. By Lemma 2.6, $c = \rho(A)$. This proves existence of h; we are left with the task of showing uniqueness.

Let $A^\dagger(x,y) := A(y,x)$ $(x,y \in S')$ denote the adjoint of A. We observe from (2.1) that

$$\rho(A^\dagger) = \rho(A).$$

Combining this with (2.6) we see that $A^\dagger$ is R-recurrent if and only if A is R-recurrent. In view of this, by our previous arguments, there exists a function $\eta : S \to (0,\infty)$ such that

$$\eta A = \rho(A)\eta. \tag{2.9}$$

Instead of showing uniqueness up to scalar multiples of positive solutions to the equation $Ah = \rho(A)h$, we can alternatively show uniqueness up to scalar multiples of positive solutions to the equation $\eta A = \rho(A)\eta$. (Applying this to $A^\dagger$ then also gives the result for h.)

We claim that (2.9) implies that $\mu(x) := \eta(x)h(x)$ $(x \in S)$ is an invariant measure for the Markov chain with transition kernel P as in (2.5). Indeed,

$$\begin{aligned}\mu P(y) &= \sum_x \mu(x)P(x,y) = \sum_x \eta(x)h(x)\rho(A)^{-1}h(x)^{-1}A(x,y)h(y) \\ &= \rho(A)^{-1}(\eta A)(y)h(y) = \eta(y)h(y).\end{aligned} \tag{2.10}$$

Since P is recurrent, by Theorem 0.21, its invariant measure is up to scalar multiples unique and hence the same is true for η. ∎

Exercise 2.8 Improve Theorem 2.7 by showing that if A is an irreducible, R-recurrent nonnegative matrix indexed by a countable set S and $f : S \to [0, \infty)$ satisfies $Af \leq \rho(A)f$, then $f = rh$ for some $r \geq 0$, where h is the function from Theorem 2.7.

Remark Exercise 2.8 implies that if A is an irreducible, R-recurrent nonnegative matrix indexed by a countable set S, and $Ah = ch$ for some function $h : S \to (0, \infty)$ and $c > 0$, then $c \geq \rho(A)$. If S is finite, then the Perron-Frobenius theorem (Theorem 2.15 below) shows that in fact $c = \rho(A)$, but for infinite matrices there may exist positive eigenfunctions with eigenvalues $c > \rho(A)$. In this case, (2.5) still defines a probability kernel, but P is not recurrent. Note that in view of this, the fact that $\rho(A)$ is called the "spectral radius" is somewhat misleading.

An irreducible nonnegative matrix A that is not R-recurrent is called *R-transient*. Warning: it is possible for a transient probability kernel to be R-recurrent (see Exercise 2.29 below). If A is R-recurrent, then Theorem 2.4 says that there exist a unique $c > 0$ and a function $h : S \to (0, \infty)$ that is unique up to scalar multiples, such that (2.5) defines a recurrent probability kernel P. Since multiplying h by a scalar has no effect on P, this means in particular that such a recurrent P is unique. We call A *R-positive* if P is positive recurrent and *R-null recurrent* if P is null recurrent.

We conclude the present section with one more definition and lemma.

Lemma 2.9 (Positive recurrence) *Let A be an irreducible, aperiodic nonnegative matrix indexed by a countable set S. Let $x \in S$ and assume that $\rho(A) < \infty$. Then the limit*

$$\lim_{n\to\infty} \rho(A)^{-n} A^n(x,x) \tag{2.11}$$

exists in $[0, \infty)$ and A is R-positive if and only if this limit is > 0. Similarly, if A has period k, then the limit $\lim_{n\to\infty} \rho(A)^{-kn} A^{kn}(x,x)$ exists and A is R-positive if and only if this limit is positive.

Proof If A is not R-recurrent, then Theorem 2.4 says that $\sum_{n=1}^{\infty} \rho(A)^{-n} A^n(x,x) < \infty$ and hence $\rho(A)^{-n} A^n(x,x) \to 0$, so we can without loss of generality assume that A is R-recurrent.

Now Lemma 2.6 and (2.8) tell us that $\rho(A)^{-n}A^n(x,x) = P^n(x,x)$. If P is null recurrent, then $\lim_{n\to\infty} P^n(x,x) = 0$ by Lemma 0.18, and if P is positive recurrent, then Theorem 0.20 tells us that $\lim_{n\to\infty} P^{kn}(x,x) = \mu(x)$, where μ is the invariant law of P^k which satisfies $\mu(x) > 0$ by Theorem 0.19. ∎

Exercise 2.10 Let A be an irreducible nonnegative matrix indexed by a countable set S. Assume that $\rho(A) < \infty$. Prove that A is R-positive if and only if there exist functions $\eta : S \to (0,\infty)$ and $h : S \to (0,\infty)$ such that $\eta A = \rho(A)\eta$, $Ah = \rho(A)h$, and

$$\sum_x \eta(x)h(x) < \infty.$$

Hint: formula (2.10).

We postpone the proof of the remaining, deeper statements of Theorem 2.4 for a while and in the next section first look at some motivating applications.

2.4 Conditioning to stay inside a set

Let X be a Markov chain with countable state space S and transition kernel P. For any $S' \subset S$, we let $P|_{S'} := (P(x,y))_{x,y\in S'}$ denote the restriction of P to S'. Note that $P|_{S'}$ is in general not a probability kernel, but it is a *subprobability kernel*, i.e., $\sum_{y\in S'} P|_{S'}(x,y) \leq 1$ for all $x \in S'$. In particular, $Q := P|_{S'}$ is a nonnegative matrix. We will be interested in the case that Q is irreducible and R-positive.

Let $Q^\dagger(x,y) := Q(y,x)$ $(x,y \in S)$ denote the adjoint of Q. As we observed in the proof of Theorem 2.7, $\rho(Q) = \rho(Q^\dagger)$, and therefore by Lemma 2.9 $Q^\dagger$ is R-positive as well. Applying Theorem 2.7, we see that there exist functions $\eta : S' \to (0,\infty)$ and $h : S' \to (0,\infty)$, unique up to scalar multiples, such that

$$\eta Q = \rho(Q)\eta \quad \text{and} \quad Qh = \rho(Q)h. \tag{2.12}$$

By the definition of R-positivity,

$$Q^{(h)}(x,y) := \rho(Q)^{-1}h(x)^{-1}Q(x,y)h(y) \qquad (x,y \in S') \tag{2.13}$$

defines a positive recurrent probability kernel on S'. Note that $Q^{(h)}$ is irreducible since Q is. By the calculation in (2.10),

$$\pi(x) := \eta(x)h(x) \qquad (x \in S') \tag{2.14}$$

is an invariant measure for $Q^{(h)}$. By Theorem 0.21 and the fact that $Q^{(h)}$ is positive recurrent, π is up to a scalar multiple equal to the unique invariant law of $Q^{(h)}$. In view of this we normalize η and h such that

$$\sum_{x\in S'} \eta(x)h(x) = 1, \tag{2.15}$$

which guarantees that π in (2.14) is the invariant law of $Q^{(h)}$.

Theorem 2.11 (Process conditioned not to leave a set - first version) *Let X be a Markov chain with countable state space S and transition kernel P, and let $S' \subset S$. Assume that $Q := P\big|_{S'}$ is irreducible, aperiodic, and R-positive. Let η, h as in (2.12) be positive left and right eigenfunctions of Q with eigenvalue $\rho(Q)$, normalized as in (2.15). Set*

$$\tau := \inf\big\{k \geq 0 : X_k \notin S'\big\}.$$

Then, for each $m \geq 1$ and $x, z \in S'$,

$$\mathbb{P}^x\big[(X_k)_{0\leq k\leq m} \in \cdot \,\big|\, n < \tau,\ X_n = z\big] \underset{n\to\infty}{\Longrightarrow} \mathbb{P}^x\big[(X^h_k)_{0\leq k\leq m} \in \cdot\,\big], \tag{2.16}$$

where X^h denotes the Markov chain with state space S' and Doob transformed transition kernel $Q^{(h)}$ defined in (2.13). If Q is periodic, then (2.16) remains true provided we restrict ourselves to those values of n for which $Q^n(x, z) > 0$.

Proof Fix $0 \leq m < n$ such that $Q^n(x, z) > 0$ and $x_0, \ldots, x_{m+1} \in S'$. Then

$$\begin{aligned}
&\mathbb{P}^{x_0}\big[X_{m+1} = x_{m+1} \,\big|\, (X_0, \ldots, X_m) = (x_0, \ldots, x_m),\ n < \tau,\ X_n = z\big] \\
&\quad= \frac{\mathbb{P}[(X_0, \ldots, X_{m+1}) = (x_0, \ldots, x_{m+1}),\ n < \tau,\ X_n = z]}{\mathbb{P}[(X_0, \ldots, X_m) = (x_0, \ldots, x_m),\ n < \tau,\ X_n = z]} \\
&\quad= \frac{\prod_{k=1}^{m+1} P(x_{k-1}, x_k) \cdot Q^{n-(m+1)}(x_{m+1}, z)}{\prod_{k=1}^{m} P(x_{k-1}, x_k) \cdot Q^{n-m}(x_m, z)} \\
&\quad= \frac{P(x_m, x_{m+1})Q^{n-(m+1)}(x_{m+1}, z)}{Q^{n-m}(x_m, z)}.
\end{aligned}$$

This shows that under the conditional law given the event $\{n < \tau,\ X_n = z\}$, the process $(X_0, \ldots, X_n)$ is a time-inhomogeneous Markov chain whose transition kernel in the $(m + 1)$-th step is given by

$$P^{(n)}_{m,m+1}(x, y) := \frac{P(x, y)Q^{n-m-1}(y, z)}{Q^{n-m}(x, z)}. \tag{2.17}$$

We observe that (recall (2.8)

$$Q^n(x,y) = \rho(Q)^n h(x)(Q^{(h)})^n(x,y)h(y)^{-1} \qquad (x,y \in S').$$

Since $Q^{(h)}$ is positively recurrent with invariant law $\pi(x) = \eta(x)h(x)$ as in (2.14), in the aperiodic case, it follows from Theorem 0.20 that

$$\rho(Q)^{-n}Q^n(x,y) \underset{n\to\infty}{\longrightarrow} h(x)\eta(y) \qquad (x,y \in S'). \tag{2.18}$$

Inserting this into (2.17) yields

$$\begin{aligned} P^{(n)}_{m,m+1}(x,y) &= \frac{P(x,y)\rho(Q)^{-(n-m-1)}Q^{n-m-1}(y,z)}{\rho(Q)^{-(n-m-1)}Q^{n-m}(x,z)} \\ &\underset{n\to\infty}{\longrightarrow} \frac{P(x,y)h(y)\eta(z)}{\rho(Q)h(x)\eta(z)} = Q^{(h)}(x,y). \end{aligned} \tag{2.19}$$

If Q has period k, then (2.18) changes in the sense that we must restrict ourselves to those values of n for which $Q^n(x,y) > 0$ and we pick up an extra factor k on the right-hand side. This yields extra factors k in the nominator and denominator of (2.19), which cancel, and the result is the same. ∎

Remark 1 If A is a nonnegative matrix indexed by a countable set S, and $0 < A^n(x,y) < \infty$ for some $x,y \in S$ and $n \geq 1$, then we can define a probability measure $\mu^{A,n}_{x,y}$ on the space of all sequences $(x_0,\ldots,x_n)$ of elements of S by setting

$$\mu^{A,n}_{x,y}(x_0,\ldots,x_n) := \frac{1}{A^n(x,y)} 1_{\{x_0=x,\ x_n=y\}} \prod_{k=1}^{n} A(x_{k-1},x_k). \tag{2.20}$$

Such a measure is called a one-dimensional *Gibbs measure* with *transfer matrix* A and *boundary conditions* x,y. The proof of Theorem 2.11 shows more generally that if A is irreducible and R-positive, then for fixed $x,y \in S$, the Gibbs measures $\mu^{A,n}_{x,y}$ converge as $n\to\infty$ to the law of the Markov chain with transition kernel P as in (2.5) and initial state x.

Remark 2 If $S'' \subset S'$ is a finite set, then since (2.16) holds for all $z \in S''$, we also obtain that

$$\mathbb{P}^x\big[(X_k)_{0\leq k\leq m} \in \cdot \,\big|\, n < \tau,\ X_n \in S''\big] \underset{n\to\infty}{\Longrightarrow} \mathbb{P}^x\big[(X^h_k)_{0\leq k\leq m} \in \cdot\,\big].$$

In particular, if S' is finite, then it suffices to condition only on the event that $\{n < \tau\}$. In the next section, we will moreover see that if S' is finite, then $P\big|_{S'}$ is automatically R-positive.

2.5 The Perron-Frobenius theorem

In this section, we prove the following fact.

Proposition 2.12 (Finite matrices) *Let A be an irreducible nonnegative matrix indexed by a finite set S. Then A is R-positive.*

Proof It suffices to prove that A is R-recurrent, because then there exists a function $h : S \to (0, \infty)$, unique up to scalar multiples, and a unique constant $c > 0$ such that formula (2.5) defines a recurrent probability kernel on S. Since S is finite, it then automatically follows that P is positive recurrent and hence A is R-positive. In view of this, it suffices to check condition (2.6) of Theorem 2.4. If A has period k, then A satisfies (2.6) if and only if A^k satisfies (2.6), so without loss of generality we may assume that A is aperiodic.

Through the formula $Af(x) := \sum_y A(x,y)f(y)$, the matrix A defines a linear operator $A : \mathbb{C}^S \to \mathbb{C}^S$. Let $\|\cdot\|$ be any norm on $\mathbb{C}^S$ and let $\|A\|$ denote the associated operator norm of A, i.e., $\|A\|$ is the smallest constant such that

$$\|Af\| \leq \|A\|\,\|f\| \qquad \forall f \in \mathbb{C}^S.$$

It follows that $\|ABf\| \leq \|A\|\,\|B\|\,\|f\|$, which in turn shows that

$$\|AB\| \leq \|A\|\,\|B\| \qquad \big(A, B \in \mathcal{L}(\mathbb{C}^S, \mathbb{C}^S)\big). \tag{2.21}$$

It follows that the map $n \mapsto \log \|A^n\|$ is subadditive, and hence Fekete's lemma (Lemma 2.2) tells us that the limit

$$\tilde\rho(A) := \lim_{n\to\infty} \|A^n\|^{1/n} = \inf_{n\geq 1} \|A^n\|^{1/n} \in [-\infty, \infty). \tag{2.22}$$

exists, or equivalently (compare (2.3)), that

$$\|A^n\| = e^{n \log \tilde\rho(A) + o(n)} \quad (n \geq 1) \quad \text{with} \quad 0 \leq o(n)/n \underset{n\to\infty}{\longrightarrow} 0. \tag{2.23}$$

Note that since $n \mapsto \log \|A^n\|$ is subadditive while $n \mapsto \log A^n(x,x)$ is superadditive, the error term here is positive while in (2.3) it is negative. We claim that $\tilde\rho(A) = \rho(A)$. It actually does not matter which operator norm we use (Exercise 2.13 below) but for concreteness let us take the operator norm associated with the supremumnorm on $\mathbb{C}^S$. Then Exercise 2.14 below tells us that

$$\|A^n\| = \sup_y \sum_z A^n(y,z).$$

This immediately implies that $A^n(x,x) \le \|A^n\|$. Since A is irreducible and aperiodic and S is finite, we can choose some $m \ge 1$ such that $A^m(y,z) > 0$ for all $y,z \in S$. Set $\varepsilon := \inf_{y,z} A^m(y,z)$. Then

$$A^{n+2m}(x,x) \ge A^m(x,y)A^n(y,z)A^m(z,x) \ge \varepsilon^2 A^n(y,z),$$

and hence

$$A^n(x,x) \le \|A^n\| \le \varepsilon^{-2} A^{n+2m}(x,x), \tag{2.24}$$

which shows that

$$\begin{aligned}\log \rho(A) &= \lim_{n\to\infty} n^{-1} \log A^n(x,x) \\ &\le \log \tilde\rho(A) \le \lim_{n\to\infty} n^{-1} \log\big(\varepsilon^{-2} A^{n+2m}(x,x)\big) = \log \rho(A).\end{aligned}$$

It now follows from (2.24) and (2.23) that

$$\sum_{n\ge1} \rho(A)^{-n} A^n(x,x) \ge \varepsilon^2 \sum_{n\ge1} \rho(A)^{-n} \|A^{n-2m}\| \ge \varepsilon^{-2} \sum_{n\ge1} \rho(A)^{-n} \rho(A)^{n-2m} = \infty,$$

which shows that condition (2.6) of Theorem 2.4 is satisfied. ∎

Exercise 2.13 (Choice of the norm) Show that the limit $\tilde\rho(A)$ from (2.22) does not depend on the choice of the norm on $\mathbb{C}^S$. Hint: you can use the fact that on a finite-dimensional space, all norms are equivalent. In particular, if $\|\cdot\|$ and $\|\cdot\|'$ are two different norms on the space of matrixes indexed by S, then there exist constants $0 < c < C < \infty$ such that $c\|A\| \le \|A\|' \le C\|A\|$ for all A.

Exercise 2.14 (Operator norm induced by supremumnorm) If $\|\cdot\|_\infty$ denotes the supremumnorm on $\mathbb{C}^S$, then show that the associated operator norm of a complex matrix $(A(x,y))_{x,y\in S}$ is given by

$$\|A\|_\infty = \sup_{x\in S} \sum_{y\in S} |A(x,y)|.$$

As a corollary of Theorem 2.4 and Proposition 2.12, we obtain the following version of the Perron-Frobenius theorem. This theorem was first proved in [Per07, Fro12], at around the same time when Markov chains were introduced [Mar06], although the link between the two subjects was discovered only much later. See [Gan00, Section 8.3] or [Sen73, Chapter 1] for a modern statement of the Perron-Frobenius theorem.

Theorem 2.15 (Perron-Frobenius) *Let A be a irreducible nonnegative matrix indexed by a finite set S. Then there exists a function $h : S \to (0, \infty)$, unique up to scalar multiples, and a unique constant $c > 0$ such that $Ah = ch$.*

Proof By Proposition 2.12, A is R-positive, so by Theorem 2.4, there exist a function $h : S \to (0, \infty)$, unique up to scalar multiples, and a unique constant $c > 0$ such that formula (2.5) defines a recurrent probability kernel. Imagine that $h' : S \to (0, \infty)$ and $c' > 0$ satisfy $Ah' = c'h$. Then formula (2.5) defines a probability kernel P. Since S is finite, P is positive recurrent and hence Theorem 2.4 tells us that $c' = c$ and $h' = rh$ for some $r > 0$. ∎

Remark In view of Proposition 2.12, we can think of Theorem 2.4 as an infinite-dimensional generalization of the Perron-Frobenius theorem. Another way to generalize the Perron-Frobenius theorem to infinite dimensions is to view A as a linear operator acting on a Banach space of functions $f : S \to \mathbb{C}$ and then use (2.22) as a starting point. This approach works even for uncountable spaces; a well-known result in this direction is the Krein-Rutman theorem [KR48].

2.6 Quasi-stationary laws

The following theorem removes the condition $X_n = z$ from the conditional probability in (2.16), at the cost of introducing the condition (2.25). Note that if S' is finite, then (2.25) is automatically satisfied while $P\big|_{S'}$ is R-positive by Proposition 2.12.

Theorem 2.16 (Process conditioned not to leave a set - second version) *Let X be a Markov chain with countable state space S and transition kernel P, and let $S' \subset S$. Assume that $Q := P\big|_{S'}$ is irreducible and R-positive. Let η, h as in (2.12) be positive left and right eigenfunctions of Q with eigenvalue $\rho(Q)$. Assume that*

$$\inf_{x \in S'} h(x) > 0. \tag{2.25}$$

Then η and h can be normalized such that

$$\sum_{x \in S'} \eta(x) = 1 \quad \textit{and} \quad \sum_{x \in S'} \eta(x) h(x) = 1. \tag{2.26}$$

Set

$$\tau := \inf \big\{ k \geq 0 : X_k \notin S' \big\}.$$

Then, for each $m \geq 1$ and $x \in S'$,

$$\mathbb{P}^x\big[(X_k)_{0\leq k\leq m} \in \cdot \,\big|\, n < \tau\big] \underset{n\to\infty}{\Longrightarrow} \mathbb{P}^x\big[(X^h_k)_{0\leq k\leq m} \in \cdot\,\big], \tag{2.27}$$

where X^h denotes the Markov chain with state space S' and Doob transformed transition kernel $Q^{(h)}$ defined in (2.13).

Proof It has already been shown in (2.15) that we can normalize η and h such that $\sum_{x\in S'} \eta(x)h(x) = 1$. Let 1 denote the function on S' that is constantly one. Arguing as in the proof of Theorem 2.11, we see that conditional on the event $\{n < \tau\}$, the process $(X_0, \ldots, X_n)$ is a time-inhomogeneous Markov chain whose transition kernel in the $(m + 1)$-th step is given by

$$P^{(n)}_{m,m+1}(x, y) := \frac{P(x, y)Q^{n-m-1}1(y)}{Q^{n-m}1(x)}. \tag{2.28}$$

Here

$$Q^n 1(x) = \sum_{y\in S'} Q^n(x, y) = \sum_{y\in S'} \rho(Q)^n h(x)(Q^{(h)})^n(x, y)h(y)^{-1}.$$

Condition (2.25) guarantees that h^{-1} is a bounded function. Combining this with Theorem 0.20, it follows that

$$\sum_{y\in S'} (Q^{(h)})^n(x, y)h(y)^{-1} \underset{n\to\infty}{\longrightarrow} \sum_{y\in S'} \pi(y)h(y)^{-1},$$

where $\pi(y) = \eta(y)h(y)$ as in (2.14) is the invariant law of $Q^{(h)}$. It follows that

$$\rho(Q)^{-n}1(x) \underset{n\to\infty}{\longrightarrow} h(x)\sum_{y\in S'} \pi(y)h(y)^{-1} = h(x)\sum_{y\in S'} \eta(y).$$

In particular, $\sum_{y\in S'} \eta(y)$ is finite so we can normalize η and h as in (2.26) and with this convention

$$\rho(Q)^{-n}1(x) \underset{n\to\infty}{\longrightarrow} h(x). \tag{2.29}$$

Inserting this into (2.28), the rest of the proof goes in the same way as the proof of Theorem 2.11. ∎

The proof of Theorem 2.16 yields the following corollary, which gives a nice interpretation to the right eigenfunction h.

Corollary 2.17 (Asymptotical probability to stay in a set) *Under the conditions of Theorem 2.16, for each $x \in S'$,*

$$\rho(Q)^{-n}\mathbb{P}^x[\tau > n] \underset{n\to\infty}{\longrightarrow} h(x) \qquad (x \in S).$$

Proof This is just another way to write (2.29). ∎

The left eigenvector η also has a nice interpretation. Let X be a Markov chain with countable state space S and transition kernel P, and let $S' \subset S$. By definition, we say that a probability measure ρ on S' is a *quasi-stationary law*[3] of P on S' if

$$\mathbb{P}[X_0 \in \cdot\,] = \rho \quad \text{implies} \quad \mathbb{P}[X_1 \in \cdot \,|\, X_1 \in S'] = \rho.$$

It seems that quasi-stationary laws were first introduced by Darroch and Seneta in [DS67].

Corollary 2.18 (Convergence to quasi-stationary law) *Under the conditions of Theorem 2.16, η is a quasi-stationary law of X on S'. Moreover, for each $x \in S'$,*

$$\mathbb{P}^x\big[X_n = y \,\big|\, \tau > n\big] \underset{n\to\infty}{\longrightarrow} \eta(y) \qquad (x, y \in S).$$

Proof Let X be the process started in some initial law μ_0 on S' and define finite measures μ_n on S' by

$$\mu_n(y) := \mathbb{P}[X_n = y,\ n < \tau] \quad (y \in S',\ n \geq 0).$$

Then

$$\mu_{n+1}(y) = \mathbb{P}[X_{n+1} = y,\ n+1 < \tau] = \sum_{z \in S'} \mathbb{P}[X_n = z,\ n < \tau]Q(z,y) = \mu_n Q(y)$$

($y \in S'$), i.e., $\mu_{n+1} = \mu_n Q$. In particular, μ_0 is a quasi-stationary law if and only if $\mu_0 Q = c\mu_0$ for some $c > 0$.

In the special case that $X_0 = x$, we see that

$$\mu_n(y) = Q^n(x,y) = \rho(Q)^n h(x)(Q^{(h)})^n(x,y)h(y)^{-1}$$

[3]Usually, the terms 'stationary law' and 'invariant law' can be used exchangeably. In this case, this may lead to confusion, however, since the term 'quasi-invariant measure' is normally used in ergodic theory for a measure that is mapped into an *equivalent* measure by some transformation.

and

$$\mathbb{P}^x\big[X_n = y \,\big|\, \tau > n\big] = \frac{\mu_n(y)}{\sum_{z\in S'}\mu_n(z)} = \frac{\rho(Q)^n h(x)(Q^{(h)})^n(x,y)h(y)^{-1}}{\rho(Q)^n h(x)\sum_z (Q^{(h)})^n(x,z)h(z)^{-1}}.$$

By Theorem 0.20, $(Q^{(h)})^n(x,y) \to \pi(y) = \eta(y)h(y)$ $(y \in S')$ as $n \to \infty$, so using the fact that h^{-1} is a bounded function because of (2.25), we see that

$$\mathbb{P}^x\big[X_n = y \,\big|\, \tau > n\big] = \frac{(Q^{(h)})^n(x,y)h(y)^{-1}}{\sum_z (Q^{(h)})^n(x,z)h(z)^{-1}} \underset{n\to\infty}{\longrightarrow} \frac{\eta(y)}{\sum_z \eta(z)} = \eta(y) \qquad (y \in S').$$

for all $y \in S'$. ∎

Remark In the proof of Corollary 2.18, we have seen that a probability measure ν on S' is a quasi-stationary law if and only if ν is a positive left eigenvector of Q. If S' is finite, then the Perron-Frobenius Theorem (Theorem 2.15) tells us that such a positive left eigenvector is unique (if we normalize it to get a probability measure). In general, we can invoke Theorem 2.7 to conclude that η is the unique quasi-stationary law with the additional property that $\eta Q = \rho(Q)\eta$. It is not so easy, however, to rule out the possibility that there could exist another quasi-stationary law ν with $\nu Q = c\nu$ for some $c > \rho(Q)$.

Exercise 2.19 (Behavior at typical times) Let $0 \leq m_n \leq n$ be such that $m_n \to \infty$ and $n - m_n \to \infty$ as $n \to \infty$. Show that for any $x \in S'$,

$$\mathbb{P}^x\big[X_{m_n} \in \cdot \,\big|\, n < \tau\big] \underset{n\to\infty}{\Longrightarrow} \pi,$$

where $\pi(y) = \eta(y)h(y)$ is the invariant law of the Doob transformed probability kernel $Q^{(h)}$.

Excercise 2.19 shows that conditional on the unlikely event that $n < \tau$ where n is large, *most* of the time up to n, the process X is approximately distributed according to the invariant law $\pi(y) = \eta(y)h(y)$. Corollary 2.18 shows that *at* the time n, the situation is different.

2.7 Sketch of the proof of the main theorem

In this section, we give the proof of Theorem 2.4. We need one preparatory lemma.

Lemma 2.20 (Speed of growth) *Let A be an irreducible aperiodic nonnegative matrix indexed by a countable set S and let $\rho(A) \in (0, \infty]$ denote its spectral radius. Then*

$$\lim_{n\to\infty} \big(A^n(x,y)\big)^{1/n} = \rho(A) \qquad (x, y \in S). \tag{2.30}$$

If A is periodic, then (2.30) remains true provided we restrict ourselves to those values of n for which $A^n(x,y) > 0$.

Proof In the special case $x = y$ this follows from Lemma 2.1. By irreducibility, for each $x, y \in S$ there exist $k, m \geq 1$ such that $A^k(x,y) > 0$ and $A^m(y,x) > 0$. Then

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n}\log A^n(x,y) &\geq \liminf_{n\to\infty} \frac{1}{n}\log\big(A^{n-k}(x,x)A^k(x,y)\big)\\ &= \lim_{n\to\infty} \frac{1}{n}\log A^{n-k}(x,x) = \rho(A)\end{aligned}$$

and

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n}\log A^n(x,y) &= \limsup_{n\to\infty} \frac{1}{n}\log\big(A^n(x,y)A^m(y,x)\big)\\ &\leq \lim_{n\to\infty} \frac{1}{n}\log A^{n+m}(x,x) = \rho(A).\end{aligned}$$

∎

The most difficult part of Theorem 2.4 is the following proposition.

Proposition 2.21 (Existence of an eigenfunction) *Let A be an irreducible nonnegative matrix such that $\rho(A) < \infty$ and (2.6) holds. Then there exists a function $h : S \to (0, \infty)$ such that $Ah = \rho(A)h$.*

We first show how this implies what we want.

Proof of Theorem 2.4 By Lemma 2.6, condition (2.6) is necessary for R-recurrence. Conversely, by Proposition 2.21, condition (2.6) implies the existence of a function $h : S \to (0, \infty)$ such that $Ah = \rho(A)h$. It follows that

$$P(x,y) := \rho(A)^{-1}h(x)^{-1}A(x,y)h(y) \qquad (x, y \in S) \tag{2.31}$$

defines a probability kernel P. In view of (2.8)

$$\sum_{n=0}^{\infty} P^n(x,x) = \sum_{n=0}^{\infty} \rho(A)^{-n}A^n(x,x)$$

which is ∞ by our assumption (2.6), showing that P is recurrent. This proves that A is R-recurrent. Imagine that for some function $f : S \to (0, \infty)$ and constant $c > 0$, setting

$$Q(x, y) := c^{-1} f(x)^{-1} A(x, y) f(y) \qquad (x, y \in S)$$

also defines a recurrent probability kernel. Then $c = \rho(A)$ by Lemma 2.6 and hence $Af = \rho(A)f$ since Q is a probability kernel. A careful inspection of the proof of Theorem 2.7 shows that although we have used Theorem 2.4 in that proof, all that we really needed was Proposition 2.21 which guarantees the existence of left and right eigenvectors η and h with eigenvalue $\rho(A) = \rho(A^\dagger)$, as well as the fact that the kernel in (2.31) is recurrent. ∎

Proof of Proposition 2.21 (sketch) For each $\lambda \geq 0$, we define a *Green's function* G_λ by

$$G_\lambda(x, y) := \sum_{n=0}^{\infty} \lambda^n A^n(x, y).$$

It follows from Lemma 2.20 that $G_\lambda(x, y) < \infty$ for all x, y as long as $\lambda\rho(A) < 1$. We observe that

$$\begin{aligned} AG_\lambda(x, z) &= \sum_y A(x, y) \sum_{n=0}^{\infty} \lambda^n A^n(y, z) = \sum_{n=0}^{\infty} \lambda^n \sum_y A(x, y) A^n(y, z) \\ &= \sum_{n=0}^{\infty} \lambda^n A^{n+1}(x, z) = \lambda^{-1} G_\lambda(x, z) - 1_{\{x=z\}}. \end{aligned} \tag{2.32}$$

Pick some reference point $z \in S$ and define $h_\lambda : S \to (0, \infty)$ by

$$h_\lambda(x) := \frac{G_\lambda(x, z)}{G_\lambda(z, z)} \qquad (x \in S).$$

Then, for each $\lambda < \rho(A)^{-1}$,

$$Ah_\lambda(x) = \lambda^{-1} h_\lambda(x) - G_\lambda(z, z)^{-1} 1_{\{x=z\}} \qquad (x \in S). \tag{2.33}$$

It follows from monotone convergence and (2.6) that

$$\lim_{\lambda \uparrow \rho(A)^{-1}} G_\lambda(z, z) = \lim_{\lambda \uparrow \rho(A)^{-1}} \sum_{n=0}^{\infty} \lambda^n A^n(x, y) = \sum_{n=0}^{\infty} \rho(A)^{-n} A^n(x, y) = \infty. \tag{2.34}$$

In view of this, we may expect that letting $\lambda \uparrow \rho(A)^{-1}$ in (2.33), the function h_λ converges to a function h such that $Ah = \rho(A)h$.

If S is finite, then this last step can be made rigorous in a straightforward manner (Exercise 2.22 below), but if S is infinite the only proofs I know are quite complicated. Proposition 2.21 follows from [Ver67, Thm 4.1], which is also treated in the books [Sen81, Woe00] (with basically the same proof). Below, I sketch a somewhat different proof (as yet unpublished, but I can give the details on request).

For each $\lambda < \rho(A)^{-1}$, set

$$P_\lambda(x,y) := \lambda h_\lambda(x)^{-1} A(x,y) h_\lambda(y) \qquad (x,y \in S)$$

By (2.32),

$$\sum_y P_\lambda(x,y) = \lambda G_\lambda(x,z)^{-1} A G_\lambda(x,z) = 1 - \lambda G_\lambda(z,z)^{-1} 1_{\{x=z\}} =: 1 - p_\lambda 1_{\{x=z\}}$$

which shows that P_λ is a subprobability kernel on S. By definition, an *excursion* away from z is a finite sequence $(x_0, \ldots, x_n)$ such that $x_0 = z = x_n$ and $x_k \neq z$ for all $0 < k < n$. For each $\lambda \leq \rho(A)^{-1}$, we define a measure ν_λ on the space Γ_z of excursions away from z by

$$\nu_\lambda(x_0, \ldots, x_n) := \lambda^n \prod_{k=1}^{n} A(x_{k-1}, x_k).$$

One can prove that $\nu_\lambda(\Gamma_z) = 1 - p_\lambda$ for $\lambda < \rho(A)^{-1}$ and the subprobability kernel P_λ corresponds to a Markov chain with killing at z, where after each visit to z either the chain makes with probability $1 - p_\lambda$ an excursion away from z according to the subprobability measure ν_λ, or is killed with the remaining probability p_λ.

It follows from (2.34) that $p_\lambda \to 0$ as $\lambda \uparrow \rho(A)^{-1}$. Therefore, $\nu := \nu_{\rho(A)^{-1}}$ is a probability measure on Γ_z. One can check that the process that makes i.i.d. excursions away from z with law ν is a recurrent Markov chain with some transition kernel P. Using notation as in (2.20), it is easy to see from the definition of ν that

$$\mu^{A,n}_{z,z} = \mu^{P,n}_{z,z} \qquad \big(z \geq 1,\ A^n(z,z) > 0\big),$$

i.e., A and P, viewed as transfer matrices, define the same Gibbs measures. It can be show that this implies that A and P must be related as in (2.5) for some $h : S \to (0, \infty)$ and $c > 0$. For strictly positive matrices this is proved in [Geo88, Remark (11.4)]; with some care the argument can be adapted to irreducible nonnegative matrices. Since P is recurrent, Lemma 2.6 implies that $c = \rho(A)$ and now (2.5) implies that $Ah = \rho(A)h$.

The proof of [Ver67, Thm 4.1] is at first sight different from the proof we have just sketched, since it is formulated entirely in terms of functions that are defined on the set S. On closer inspection, however, these functions are closely related to the excursions that played such an important role in our proof, so on a deeper level both proofs are probably similar. ∎

Exercise 2.22 (Compactness argument) Prove that if S is finite, then there exist $\lambda_n \uparrow \rho(A)^{-1}$ such that the functions h_{λ_n} from (2.33) converge to a function $h : S \to (0, \infty)$ satisfying $Ah = \rho(A)h$. Note that toghether with what we have already proved, this completes the proof of the Perron-Frobenius theorem (Theorem 2.15).

2.8 Further results

Let X be a Markov chain with countable state space S and transition kernel P. Let $\sigma_x := \inf\{k \geq 1 : X_k = x\}$ denote the first return time to a point $x \in S$. Let us say that a point $x \in S$ is *strongly positive recurrent* if

$$\mathbb{E}^x\big[e^{\varepsilon\sigma_x}\big] < \infty \quad \text{for some } \varepsilon > 0.$$

It is known [Ken59] that strong positive recurrence is a class property, i.e., if one point in a communicating class is strongly positive recurrent then the same is true for all points in this communicating class. In the aperiodic case, strong positive recurrence is equivalent to exponentially fast convergence to the invariant law; it is then said that the Markov chainn is *geometrically ergodic*. If x is positive recurrent but not strongly so, then we say that x is *weakly positive recurrent*. Recall that by Theorem 2.4, if A is an R-recurrent irreducible nonnegative matrix, then there exists a unique recurrent probability kernel P that is related to A as in (2.5). We say that A is *strongly R-positive* resp. *weakly R-positive* if P is strongly or weakly positive recurrent respectively.

In general, it is not easy to verify conditions (2.6) or (2.11) to check that a given nonnegative matrix A is R-recurrent or R-positive, respectively. In our proof that finite matrices are R-positive (Proposition 2.12) we used a subadditivity argument. Such arguments sometimes work in an infinite dimensional setting as well. The following theorem, which we cite from [Swa17] without proof, gives necessary and sufficient conditions for strong R-positivity.

Theorem 2.23 (Strong R-positivity) *Let A be an irreducible nonnegative matrice indexed by a countable set S. Let $B \leq A$ be another nonnegative matrix such* *that $B(x,y) > 0$ if and only if $A(x,y) > 0$, $B \neq A$, and $\{(x,y) : B(x,y) \neq A(x,y)\}$ is finite. Then A is strongly R-positive if and only if $\rho(B) < \rho(A)$.*

In words, Theorem 2.23 says that A is strongly R-positive if and only if lowering the value of finitely many entries of the matrix lowers its spectral radius. This reduces the problem of proving R-positivity to finding good upper and lower bounds on the spectral radii of two matrices. We note that upper bounds can be obtained with Exercise 2.3.

We conclude this chapter with two lemmas and a series of exercises. We extend our definition of the spectral radius to nonnegative matrices A that are not necessarily irreducible by putting

$$\rho(A) := \sup_{x\in S}\sup_{n\geq 1}\big(A^n(x,x)\big)^{1/n}. \tag{2.35}$$

In view of Lemma 2.1, this coincides with our previous definition if A is irreducible. (Note that this is true even in the periodic case.)

Lemma 2.24 (Lower semi-continuity of the spectral radius) *The function $A \mapsto \rho(A)$ is lower semi-continuous with respect to pointwise convergence, i.e., if S is a countable set and A_m, A are nonnegative matrices indexed by S such that $\lim_{m\to\infty} A_m(x,y) = A(x,y)$ for all $x,y \in S$, then*

$$\liminf_{m\to\infty}\rho(A_m) \geq \rho(A).$$

Proof It is well-known that the supremum of a collection of lower semi-continuous functions is again lower semi-continuous. Therefore, in view of (2.35), it suffices to prove that for fixed x and n, the map $A \mapsto \big(A^n(x,x)\big)^{1/n}$ is lower semi-continuous. Equivalently, we can show that $A \mapsto A^n(x,x)$ is lower semi-continuous. Let $\Omega_{x,x}$ denote the space of all sequences $(x_0,\ldots,x_n) \in S^{n+1}$ such that $x_0 = x = x_n$. Then we can write, using Fatou's lemma,

$$\begin{aligned}&\liminf_{m\to\infty} A^n_m(x,x)\\&\quad=\liminf_{m\to\infty}\sum_{(x_0,\ldots,x_n)\in\Omega_{x,x}}\prod_{k=1}^n A_m(x_{k-1},x_k) \geq \sum_{(x_0,\ldots,x_n)\in\Omega_{x,x}}\prod_{k=1}^n A(x_{k-1},x_k).\end{aligned}$$

∎

Lemma 2.25 (Increasing limits) *The function $A \mapsto \rho(A)$ is continuous with respect to increasing sequences, i.e., if $A_1 \leq A_2 \leq \cdots$ are nonnegative matrices indexed by a countable set S and $A(x,y) := \lim_{m\to\infty} A_m(x,y) = A(x,y)$ $(x,y \in S)$, then*

$$\lim_{m\to\infty} \rho(A_m) = \rho(A).$$

Proof Lemma 2.24 implies that $\liminf_{m\to\infty} \rho(A_m) \geq \rho(A)$. Since $A_m \leq A$ and hence $\rho(A_m) \leq \rho(A)$ for all m, the other inequality is trivial. ∎

Exercise 2.26 (R-transience) Let P be an irreducible probability kernel indexed by a countable set S that is transient and satisfies $\rho(P) = 1$. Show that P is R-transient. Now let A be an irreducible nonnegative matrix and let $h : S \to (0,\infty)$ satisfy $Ah = ch$ for some $c > 0$. Assume that the probability kernel defined by $P(x,y) := c^{-1}h(x)^{-1}A(x,y)h(y)$ $(x,y \in S)$ is transient and satisfies $\rho(P) = 1$. Show that this implies that A is R-transient.

Exercise 2.27 (Examples of R-transient matrices) Let P be the transition kernel of nearest-neighbor random walk on $\mathbb{Z}^d$, i.e., $P(x,y) := 1/(2d)$ if $|x-y| = 1$ and $P(x,y) := 0$ otherwise. The *Local Central Limit Theorem* [LL10, Thm 2.3.9] implies that

$$P^{2n}(0,0) \sim \Big(\frac{d}{2\pi(2n)}\Big)^{d/2} \quad \text{as } n \to \infty, \tag{2.36}$$

where the notation $f_n \sim g_n$ means $f_n/g_n \to 1$. Use this to prove that $\rho(P) = 1$ and that P is R-null recurrent in dimensions $d = 1,2$ and R-transient in dimensions $d \geq 3$.

Exercise 2.28 (Strong Liouville property) Let P be the transition kernel of a recurrent Markov chain with countable state space S and let $f : S \to (0,\infty)$ be a harmonic function, i.e., $Pf = f$. Prove that f is constant. Note that this strengthens Corollary 1.27 in dimensions $d = 1,2$.

Exercise 2.29 (R-null recurrence) For $p \in (0,1)$ let P_p be the probability kernel on $\mathbb{Z}$ defined as

$$P_p(x,x+1) := p, \quad P_p(x,x-1) := 1-p, \quad \text{and} \quad P(x,y) := 0 \quad (|x-y| \neq 1).$$

Show that P_p is R-null recurrent and determine $\rho(P_p)$. Hint: transform P_p into a null-recurrent probability kernel.

Exercise 2.30 (Restricted kernel) Using notation as in Exercise 2.29, let $Q_p := P_p|_{\mathbb{N}}$ denote the restriction of P_p to $\mathbb{N}$. Prove that $\rho(Q_p) = \rho(P_p)$. Hint: apply Lemma 2.25.

Exercise 2.31 (Entropic repulsion of driftless walk) Using notation as in the previous exercise, prove that $Q_{1/2}$ is R-transient. Hint: You can use that by Exercises 2.29 and 2.30, $\rho(Q_{1/2}) = \rho(P_{1/2}) = 1$. Now use the function $h(x) := x+1$ $(x \in \mathbb{Z})$ to transform $Q_{1/2}$ into a probability kernel $Q^{(h)}$ such that $\rho(Q^{(h)}) = 1$ and apply Exercise 1.19 to check that $Q^{(h)}$ is transient.

Exercise 2.32 (Entropic repulsion of drifted walk) Generalize Exercise 2.31 by showing that Q_p is R-transient for each $p \in (0,1)$.

Exercise 2.33 (Quasi-stationary law) Let X be a Markov chain with state space $\mathbb{Z}$ and transition kernel P given by

$$\begin{array}{lll} P(x,x+1) := p & P(x,x-1) = 1-p & (x \neq 0) \\ P(0,1) := \frac{1}{2} & P(0,-1) = \frac{1}{2}. & \end{array}$$

Let $\tau := \inf\{k \geq 0 : X_k < 0\}$. Assume that $p < \frac{1}{4}$. Show that there exists a probability kernel $Q^{(h)}$ on $\mathbb{N}$ such that, for each $x > 0$,

$$\mathbb{P}^x\big[(X_k)_{0 \leq k \leq m} \in \cdot \,\big|\, n < \tau\big] \underset{n\to\infty}{\Longrightarrow} \mathbb{P}^x\big[(X^h_k)_{0 \leq k \leq m} \in \cdot\,\big],$$

where X^h denotes the Markov chain with transition kernel $Q^{(h)}$, and give a formula for $Q^{(h)}$. Hint: try positive eigenfunctions of the form $h(x) = \theta^x$ with $\theta > 0$.

Chapter 3

Intertwining

3.1 Intertwining of Markov chains

In linear algebra,[1] an *intertwining relation* is a relation between linear operators of the form

$$AB = B\tilde{A}.$$

In particular, if B is invertible, then this says that $A = B\tilde{A}B^{-1}$, or, in other words, that the matrices A and $\tilde{A}$ are *similar*. In this case, we may associate B with a change of basis, and $\tilde{A}$ corresponds to the matrix A written in terms of a different basis. In general, however, B need not be invertible. It is especially in this case that the word intertwining is used.

We will be especially interested in the case that $A, \tilde{A}$ and B are (linear operators corresponding to) probability kernels. So let us assume that we have probability kernels $P, \tilde{P}$ on countable spaces $S, \tilde{S}$, respectively, and a probability kernel K from S to $\tilde{S}$ (i.e., $S \times \tilde{S} \ni (x, y) \mapsto K(x, y)$), such that

$$PK = K\tilde{P}. \tag{3.1}$$

Note that $P, \tilde{P}$ and K correspond to linear operators $P : \mathbb{C}^S \to \mathbb{C}^S$, $\tilde{P} : \mathbb{C}^{\tilde{S}} \to \mathbb{C}^{\tilde{S}}$, and $K : \mathbb{C}^{\tilde{S}} \to \mathbb{C}^S$, so both sides of the equation (3.1) correspond to a linear operator from $\mathbb{C}^{\tilde{S}}$ into $\mathbb{C}^S$. We need some examples to see this sort of relations between probability kernels can really happen.

[1] For example, in the theory of representations of Lie algebras and Lie groups

Exclusion process. Fix $n \geq 2$ and let $C_n := \mathbb{Z}/n$, i.e., $C_n = \{0, \ldots, n-1\}$ with addition modulo n. (I.e., C_n is the cyclic group with n elements.) Let $S := \{0,1\}^{C_n}$, i.e., S consists of all finite sequences $x = (x(0), \ldots, x(n-1))$ indexed by C_n. Let $(I_k)_{k\geq 1}$ be i.i.d. and uniformly distributed on C_n. For each $x \in S$, we may define a Markov chain $X = (X_k)_{k\geq 0}$, started in $X_0 = x$ and with values in S, by setting

$$X_{k+1}(i) := \begin{cases} X_k(I+1) & \text{if } i = I, \\ X_k(I) & \text{if } i = I+1, \\ X_k(i) & \text{otherwise.} \end{cases} \tag{3.2}$$

In words, this says that in each time step, we choose a uniformly distributed position $I \in C_n$ and exchange the values of X in I and $I+1$ (where we calculate modulo n). Note that we have described our Markov chain in terms of a random mapping representation. In particular, it is clear from this construction that X is a Markov chain.

Thinning Let C be any countable set and let $S := \{0,1\}^C$ be the set of all $x = (x(i))_{i\in C}$ with $x(i) \in \{0,1\}$ for all $i \in C$, i.e., S is the set of all sequences of zeros and ones indexed by C. Fix $0 \leq p \leq 1$, and let $(\chi_i)_{i\in C}$ be i.i.d. *Bernoulli* (i.e., $\{0,1\}$-valued) random variables with $\mathbb{P}[\chi_i = 1] = p$. We define a probability kernel K_p from S to S by

$$K_p(x,\,\cdot\,) = \mathbb{P}\big[\big(\chi_i x(i)\big)_{i\in C} \in \cdot\,\big] \qquad (x \in S). \tag{3.3}$$

Note that $\big(\chi_i x(i)\big)_{i\in C}$ is obtained from x by setting some coordinates of x to zero, independently for each $i \in C$, where each coordinate $x(i)$ that is one has probability p to remain one and probability $1-p$ to become a zero. We describe this procedure as *thinning* the ones with parameter p.

Exercise 3.1 (Thinning of exclusion processes) Let P be the transition kernel of the exclusion process on C_n described above, with state space $S = \{0,1\}^{C_n}$, and for $0 \leq p \leq 1$, let K_p be the kernel from S to S corresponding to thinning with parameter p. Show that

$$PK_p = K_pP \qquad (0 \leq p \leq 1).$$

Counting process Let P be a transition kernel on a countable space S and let $Y^1 = (Y^1_k)_{k\geq 0}$ and $Y^2 = (Y^2_k)_{k\geq 0}$ be two independent Markov chains with transition kernel P and possibly different deterministic initial states $Y^i_0 = y^i$ $(i = 1,2)$. Then

$(Y^1, Y^2) = (Y^1_k, Y^2_k)_{k\geq 0}$ is a Markov chain with values in the product space $S \times S$. We may view (Y^1, Y^2) as two particles, walking around in the space S. Now maybe we are not interested in which particle is where, but only in how many particles are on which place. In that case, we may look at the process

$$X_k(y) := 1_{\{Y^1_k = y\}} + 1_{\{Y^2_k = y\}} \qquad (y \in S,\ k \geq 0),$$

which takes values in the space $\tilde{S}$ consisting of all functions $x : S \to \{0, 1, 2\}$ such that $\sum_{y\in S} x(y) = 2$. Note that X just counts how many particles are present on each site $x \in S$. It is not hard to see that $X = (X_k)_{k\geq 0}$ is an autonomous Markov chain.

Exercise 3.2 (Counting process) Let (Y^1, Y^2) and X be the Markov chains with state spaces $S \times S$ and $\tilde{S}$ described above. Let P_2 be the transition kernel of (Y^1, Y^2) and let Q be the transition kernel of X. For each $x \in \tilde{S}$, let $K(x, \cdot\,)$ be the uniform distribution on the set

$$U_x := \big\{ \big(y^1, y^2\big) \in S \times S : x = 1_{y^1} + 1_{y^2} \big\},$$

where $1_y(y') := 1_{\{y=y'\}}$. Note that K is a probability kernel from $\tilde{S}$ to $S \times S$. From x, we can see that there are two particles and where these are, but not which is the first and which is the second particle. All K does is arbitrarily ordering the particles in x. Show that

$$QK = KP_2.$$

This example can easily be generalized to any number of independent Markov chains (all with the same transition kernel).

Conditioning on the future As a third example, we look at a Markov chain X with finite state space S and transition kernel Q. We assume that Q is such that all states in S are transient, except for two states z_1 and z_2, which are traps. We set

$$h_i(x) := \mathbb{P}^x\big[X_k = z_i \text{ for some } k \geq 0\big] \qquad (i = 1, 2).$$

By Lemma 1.2, we know that h_1 and h_2 are harmonic functions. Since all states except z_1 and z_2 are transient and our state space is finite, we have $h_1 + h_2 = 1$. By Proposition 1.6, the process X conditioned to be eventually trapped in z_i is itself a Markov chain, with state space $\{x \in S : h_i(x) > 0\}$ and Doob transformed transition kernel $Q^{(h_i)}$. Set

$$\tilde{S} := \big\{(x, i) : x \in S,\ i \in \{0, 1\},\ h_i(x) > 0\big\}.$$

We define a probability kernel $\tilde{P}$ on $\tilde{S}$ by

$$\tilde{P}\big((x,i),(y,j)\big) := Q^{(h_i)}(x,y)1_{\{i=j\}} \qquad (x,y \in S,\ i,j \in \{1,2\}).$$

Note that if $(\tilde{X},I) = (\tilde{X}_k,I_k)_{k\geq 0}$ is a Markov chain with transition kernel $\tilde{P}$, then I never changes its value, and depending on whether $I_k = 1$ for all $k \geq 0$ or $I_k = 2$ for all $k \geq 0$, the process $\tilde{X}$ is our original Markov chain X conditioned to be trapped in either z_1 or z_2.

Exercise 3.3 (Conditioning on the future) In the example above, define a probability kernel $\tilde{K}$ from S to $\tilde{S}$ by

$$\tilde{K}\big(x,(y,i)\big) := 1_{\{x=y\}}h_i(x) \qquad (x,y \in S,\ i \in \{1,2\}).$$

Show that

$$Q\tilde{K} = \tilde{K}\tilde{P}.$$

Returning to our general set-up, we observe that the intertwining relation (3.1) implies that for any probability measure μ on S

$$\mu P^n K = \mu K \tilde{P}^n \qquad (n \geq 0).$$

This function has the following interpretation. If we start the Markov chain with transition kernel P in the initial law μ, run it till time n, and then apply the kernel K to its law, then the result is the same as if we start the the Markov chain with transition kernel $\tilde{P}$ in the initial law μK, run it till time n, and look at its law.

More concretely, in our three examples, this says:

- If we start an exclusion process in some initial law, run it till time n, and then thin it with the parameter p, then the result is the same as when we thin the initial state with p, and then run the process till time n.
- If we start the counting process Y in any initial law, run it to time n, and then arbitrarily order the particles, then the result is the same as when we first arbitrarily order the particles, and then run the process $(X(1),X(2))$ till time n.
- If we run the two-trap Markov chain X till time n, and then assign it a value 1 or 2 according to the probability, given its present state X_n, that it will eventually get trapped in z_1 or z_2, respectively, then the result is the same as when we assign such values at time zero, and then run the appropriate Doob transformed Markov chain till time n.

None of these statement comes as a big surprise, but we will later see less trivial examples of this phenomenon.

3.2 Markov functionals

We already know that functions of Markov chains usually do not have the Markov property. An exception, as we have seen in Lemma 0.12, is the case when a function of a Markov chain is *autonomous*. Let us quickly recall what this means. Let Y be a Markov chain with countable state space T and transition kernel P, and let $\psi : T \to S$ be a function from T into some other countable set S. Then we say that $(X_k)_{k\geq 0} := (\psi(Y_k))_{k\geq 0}$ is an *autonomous Markov chain* if

$$\mathbb{P}\big[\psi(Y_{k+1}) = x \,\big|\, Y_k = y\big]$$

depends on y only through $\psi(y)$. Equivalently, this says that there exists a transition kernel Q on S such that

$$Q(x,x') = \sum_{y' :\, \psi(y')=x'} P(y,y') \qquad \forall y \in S \text{ s.t. } \psi(y) = x. \tag{3.4}$$

If (3.4) holds, then, regardless of the initial law of Y, one has that the process $(X_k)_{k\geq 0} = (\psi(Y_k))_{k\geq 0}$ is a Markov chain with transition kernel Q.

In Theorem 3.5 below we will see that sometimes, a function $(\psi(Y_k))_{k\geq 0}$ of a Markov chain $(Y_k)_{k\geq 0}$ can be a Markov chain itself, even when it is not autonomous. In this case, however, this is usually only true for certain special initial laws of X. It seems this result is due to Rogers and Pitman [RP81]. We start with a preliminary result that makes clear why in general, functions of a Markov chain do not have the Markov property.

Theorem 3.4 (Filtering equations) *Let $(Y_k)_{k\geq 0}$ be a Markov chain with countable state space T and transition kernel P, let $\psi : T \to S$ be a function from T into some other countable set S, and let $X_k := \psi(Y_k)$ $(k \geq 0)$. Set*

$$\pi(y \,|\, x_0,\ldots,x_n) := \mathbb{P}\big[Y_n = y \,\big|\, (X_0,\ldots,X_n) = (x_0,\ldots,x_n)\big]$$

for all $n \geq 0$ and $x_0,\ldots,x_n$ for which $\mathbb{P}[(X_0,\ldots,X_n) = (x_0,\ldots,x_n)] > 0$. Then

$$\pi(y \,|\, x_0,\ldots,x_n) = \frac{\sum_{y'} P(y',y) 1_{\{\psi(y)=x_n\}} \pi(y' \,|\, x_0,\ldots,x_{n-1})}{\sum_{y',y''} P(y',y'') 1_{\{\psi(y'')=x_n\}} \pi(y' \,|\, x_0,\ldots,x_{n-1})}. \tag{3.5}$$

Moreover,

$$\begin{aligned} &\mathbb{P}\big[X_n = x_n \,\big|\, (X_0, \ldots, X_{n-1}) = (x_0, \ldots, x_{n-1})\big] \\ &\quad = \sum_{y,y'} P(y', y) 1_{\{\psi(y)=x_n\}} \pi(y' \,|\, x_0, \ldots, x_{n-1}). \end{aligned} \tag{3.6}$$

Remark 1 Equation (3.5) is called the *filtering equation* , which is the basis for filtering theory. It shows how we have to update in real time our information about the present state of a Markov chain Y based on partial information in the form of an observed process X which is a function of Y.

Remark 2 Since in general, $\pi(y \,|\, x_0, \ldots, x_n)$ depends on the whole history of the process X and not just on its current state x_n, we see that the process X is in general not a Markov chain.

Proof of Theorem 3.4 We start by proving the filtering equation (3.5). Conditioning first on $X_0, \ldots, X_{n-1}$ and then using the formula for conditional probabilities to condition also on X_n, we obtain

$$\pi(y \,|\, x_0, \ldots, x_n) = \frac{\mathbb{P}[Y_n = y,\ X_n = x_n \,|\, (X_0, \ldots, X_{n-1}) = (x_0, \ldots, x_{n-1})]}{\mathbb{P}[X_n = x_n \,|\quad ,, \quad]}, \tag{3.7}$$

where ,, is shorthand for $(X_0, \ldots, X_{n-1}) = (x_0, \ldots, x_{n-1})$. Conditioning on the value of Y_{n-1}, the nominator in (3.7) can be rewritten as

$$\begin{aligned} &\mathbb{P}\big[Y_n = y,\ X_n = x_n \,\big|\quad ,, \quad\big] \\ &\quad = \sum_{y'} \mathbb{P}\big[Y_n = y,\ X_n = x_n \,\big|\, Y_{n-1} = y' \text{ and } \quad ,, \quad\big] \, \mathbb{P}\big[Y_{n-1} = y' \,\big|\quad ,, \quad\big] \\ &\quad = \sum_{y'} P(y', y) 1_{\{\psi(y)=x_n\}} \pi(y' \,|\, x_0, \ldots, x_{n-1}), \end{aligned} \tag{3.8}$$

where in the last step we have used the Markov property of Y. The denominator in (3.7) is the same as the nominator summed over all values of y, so inserting (3.8) into (3.7) we obtain (3.5). Summing (3.8) over y we also obtain (3.6). ∎

Although in general the right-hand side of (3.6) depends on the whole history of the process X and hence X is not a Markov chain, the following theorem shows that in special cases and for special initial laws, it may happen that X is Markovian after all, even when it is not autonomous. Note formula (3.9) below says that if X and Y are random variables such that $\mathbb{P}[Y \in \cdot\,|X] = K(X, \cdot\,)$, then that $\psi(Y) = X$ a.s. Thus, the probability kernel K from S to T is in a sense the 'inverse' of the function $\psi : T \to S$.

Theorem 3.5 (Markov functionals) *Let $(Y_k)_{k\geq 0}$ be a Markov chain with countable state space T and transition kernel P, let $\psi : T \to S$ be a function from T into some other countable set S, and let $X_k := \psi(Y_k)$ $(k \geq 0)$. Assume that there exists a probability kernel K from S to T such that*

$$\big\{y \in T : K(x,y) > 0\big\} \subset \big\{y \in T : \psi(y) = x\big\} \qquad (x \in S), \tag{3.9}$$

and a probability kernel Q on S such that

$$QK = KP. \tag{3.10}$$

Then

$$\mathbb{P}\big[Y_0 = y \,\big|\, X_0\big] = K(X_0, y) \quad \text{a.s.} \qquad (y \in T) \tag{3.11}$$

implies that

$$\mathbb{P}\big[Y_n = y \,\big|\, (X_0, \ldots, X_n)\big] = K(X_n, y) \quad \text{a.s.} \qquad (y \in T,\ n \geq 0). \tag{3.12}$$

Moreover, (3.11) implies that the process $(X_k)_{k\geq 0}$, on its own, is a Markov chain with transition kernel Q.

Proof Formula (3.12) says that

$$\pi(y \,|\, x_0, \ldots, x_n) = K(x_n, y).$$

We will prove this by induction. For $n = 0$ the statement is just (3.11). Assuming that the statement is true for $n-1$, the filtering equation (3.5) tells us that

$$\pi(y \,|\, x_0, \ldots, x_n) = \frac{\sum_{y'} P(y', y) 1_{\{\psi(y) = x_n\}} K(x_{n-1}, y')}{\sum_{y', y''} P(y', y'') 1_{\{\psi(y'') = x_n\}} K(x_{n-1}, y')}. \tag{3.13}$$

We can rewrite the nominator as

$$\begin{aligned}
&\sum_{y'} P(y', y) 1_{\{\psi(y) = x_n\}} K(x_{n-1}, y') = KP(x_{n-1}, y) 1_{\{\psi(y) = x_n\}} \\
&\quad = QK(x_{n-1}, y) 1_{\{\psi(y) = x_n\}} = \sum_x Q(x_{n-1}, x) K(x, y) 1_{\{\psi(y) = x_n\}} \\
&\quad = Q(x_{n-1}, x_n) K(x_n, y),
\end{aligned} \tag{3.14}$$

where we have used (3.10) and (3.9). The denominator in (3.13) is the same as the nominator summed over y, so we obtain

$$\pi(y \,|\, x_0, \ldots, x_n) = \frac{Q(x_{n-1}, x_n) K(x_n, y)}{\sum_{y''} Q(x_{n-1}, x_n) K(x_n, y'')} = \frac{Q(x_{n-1}, x_n) K(x_n, y)}{Q(x_{n-1}, x_n)} = K(x_n, y).$$

This proves (3.12).

Inserting (3.12) into (3.6), using (3.14), we obtain

$$\begin{aligned}&\mathbb{P}\big[X_n = x_n \,\big|\, (X_0,\ldots,X_{n-1}) = (x_0,\ldots,x_{n-1})\big]\\&\quad= \sum_{y,y'} P(y',y)1_{\{\psi(y)=x_n\}}K(x_{n-1},y') = \sum_y Q(x_{n-1},x_n)K(x_n,y) = Q(x_{n-1},x_n),\end{aligned}$$

which shows that $(X_k)_{k\geq 0}$ is a Markov chain with transition kernel Q. ∎

Let us see how Theorem 3.5 relates to the examples developed in the previous section.

Counting process Let (Y^1,Y^2) be two independent Markov processes, where each process takes values in S and has transition kernel P, as in Excercise 3.2. Let X and $\tilde{S}$ be as defined there and let $\psi : S\times S\to\tilde{S}$ be the function

$$\psi(y^1,y^2) := 1_{y^1} + 1_{y^2},$$

so that $X = \psi(Y^1,Y^2)$. Then the kernel K defined in Excercise 3.2 satisfies

$$\big\{(y^1,y^2) : K\big(x,(y^1,y^2)\big) > 0\big\} \subset \big\{(y^1,y^2) : \psi(y^1,y^2) = x\big\}.$$

Since moreover $QK = KP_2$, all assumptions of Theorem 3.5 are fulfilled, so we find that

$$\mathbb{P}\big[(Y_0^1,Y_0^2) = (y^1,y^2)\,\big|\, X_0\big] = K\big(X_0,(y^1,y^2)\big) \quad \text{a.s.} \tag{3.15}$$

for all $(y^1,y^2)\in S\times S$ implies that

$$\mathbb{P}\big[(Y_n^1,Y_n^2) = (y^1,y^2)\,\big|\,(X_0,\ldots,X_n)\big] = K\big(X_n,(y^1,y^2)\big) \quad \text{a.s.} \tag{3.16}$$

for all $(y^1,y^2)\in S\times S$ and $n\geq 0$, and under the same assumption, $X_k = \psi(Y^1)k, Y_k^2)$ is a Markov chain with transition kernel Q. Note that this latter conclusion holds in fact even without the assumption (3.15), since X is autonomous. Formula (3.16) tells us that if initially we do not know the order of the particles, then by observing the process X up to time n, we obtain no information about the order of the particles.

Conditioning on the future Let X and $(\tilde{X},I)$ be Markov chains with state spaces S and $\tilde{S}$ and transition kernels Q and $\tilde{P}$ as in Excercise 3.3. Thus, X is a Markov chain that eventually gets trapped in one of two traps z_1 and z_2, and $(\tilde{X},I)$ is the same process where the second coordinate I tells us from the beginning in which of the two traps we will end up.

Define $\psi : \tilde{S} \to S$ by

$$\psi(x,i) := x \qquad \big((x,i) \in \tilde{S}\big),$$

i.e., ψ is just projection on the first coordinate. Then the kernel $\tilde{K}$ from Excercise 3.3 satisfies

$$\big\{(\tilde{x},i) \in \tilde{S} : \tilde{K}\big(x,(\tilde{x},i)\big) > 0\big\} \subset \big\{(\tilde{x},i) \in \tilde{S} : \psi(\tilde{x},i) = x\big\}$$

and $P\tilde{K} = \tilde{K}\tilde{P}$, so all assumptions of Theorem 3.5 are fulfilled. We therefore conclude that

$$\mathbb{P}\big[(\tilde{X}_0, I_0) = (\tilde{x},i) \,\big|\tilde{X}_0\big] = \tilde{K}\big(X_0,(\tilde{x},i)\big) \quad \text{a.s.} \qquad \big((\tilde{x},i) \in \tilde{S}\big)$$

implies that

$$\mathbb{P}\big[(\tilde{X}_n, I_n) = (\tilde{x},i) \,\big|(\tilde{X}_0,\ldots,\tilde{X}_n)\big] = \tilde{K}\big(X_n,(\tilde{x},i)\big) \quad \text{a.s.}$$

for all $n \geq 0$ and $(\tilde{x},i) \in \tilde{S}$. More simply formulated, this says that

$$\mathbb{P}\big[I_0 = i \,\big|\, \tilde{X}_0\big] = h_i(X_0) \quad \text{a.s.} \qquad (i = 0,1)$$

implies that

$$\mathbb{P}\big[I_n = i \,\big|\, (\tilde{X}_0,\ldots,\tilde{X}_n)\big] = h_i(\tilde{X}_n) \quad \text{a.s.} \qquad (n \geq 0,\ i = 1,2). \tag{3.17}$$

Moreover, under the same assumption, the process $\tilde{X}$, on its own, is a Markov chain with transition kernel P. Note that this is more surprising than in the previous example, since in the present set-up, $\tilde{X}$ is *not* autonomous as part of the joint Markov chain $(\tilde{X}, I)$. Indeed, $\tilde{X}$ evolves according to the transition kernel $Q^{(h_1)}$, resp. $Q^{(h_2)}$, depending on whether $I = 1$ or $= 2$.

To understand how it is possible that $\tilde{X}$ has the Markov property even though it is not autonomous, we observe that by (3.17), if we observe the whole path of the process $\tilde{X}$ up to time n, then we do not gain more information about the present state of I than we would get from knowing X_n. As a result,

$$\begin{aligned}
&\mathbb{P}\big[\tilde{X}_{n+1} = x_{n+1} \,\big|\, (\tilde{X}_0,\ldots,\tilde{X}_n) = (x_0,\ldots,x_n)\big] \\
&\quad = \mathbb{P}\big[\tilde{X}_{n+1} = x_{n+1} \,\big|\, (\tilde{X}_n, I_n) = (x_n,1)\big] h_1(x_n) \\
&\qquad + \mathbb{P}\big[\tilde{X}_{n+1} = x_{n+1} \,\big|\, (\tilde{X}_n, I) = (x_n,2)\big] h_2(x_n),
\end{aligned}$$

which depends only on x_n and not on $(x_0,\ldots,x_{n-1})$.

Thinning of exclusion processes This example does not satisfy condition (3.9), hence Theorem 3.5 is not applicable. In view of this and other examples, we will prove a more general theorem in the next section.

3.3 Intertwining and coupling

In Theorem 3.5 we have seen how intertwining is related to the problem of Markov functionals, i.e., the question whether certain functions of a Markov chain themselves have the Markov property. The intertwinings in Theorem 3.5 are of a special kind, because of the condition (3.9). In this section, we will see that intertwining relations between two Markov chains X and Y in general give rise to couplings between X and Y such that (3.12) holds. The following result is due to Diaconis and Fill [DF90, Thm 2.17]; the continuous-time analogue has been proved in [Fil92, Thm. 2].

Theorem 3.6 (Intertwining coupling) *Let Q and P be probability kernels on countable state spaces S and T, respectively. Assume that K is a probability kernel from S to T such that*

$$QK = KP.$$

Let

$$Q_{y'}(x,x') := \frac{Q(x,x')K(x',y')}{QK(x,y')} \qquad \big(x,x' \in S,\ y' \in T,\ QK(x,y') > 0\big), \tag{3.18}$$

and choose for $Q_{y'}(x,\,\cdot\,)$ any probability law on S if $QK(x,y') = 0$. Then $Q_{y'}$ is a probability kernel on S for each $y' \in T$ and

$$\hat{P}(x,y;x',y') := P(y,y')Q_{y'}(x,x') \qquad \big((x,y),(x',y') \in \hat{S}\big) \tag{3.19}$$

defines a probability kernel on $\hat{T} := \{(x,y) \in S \times T : K(x,y) > 0\}$, where $\hat{P}$ does not depend on the freedom in the choice of the $Q_{y'}$. If (X,Y) is the Markov chain with transition kernel $\hat{P}$ started in an initial law such that

$$\mathbb{P}\big[Y_0 = y \,\big|\, X_0\big] = K(X_0,y) \qquad (y \in \tilde{S}),$$

then

$$\mathbb{P}\big[Y_n = y \,\big|\, (X_k)_{0\le k\le n}\big] = K(X_n,y) \qquad (y \in \tilde{S},\ n \ge 0), \tag{3.20}$$

and X, on its own, is a Markov chain with transition kernel Q.

Remark It is clear from (3.19) that in the joint Markov chain (X,Y), the second component Y is autonomous with transition kernel P, but X is in general not autonomous, unless $Q_{y'}$ can be chosen so that it does not depend on y'.

Proof of Theorem 3.6 We observe that

$$\sum_{x'\in S} Q_{y'}(x,x') = \frac{QK(x,y')}{QK(x,y')} = 1 \qquad (QK(x,y')>0),$$

which shows that $Q_{y'}$ is a probability kernel. To show that the definition of $\hat{P}$ does not depend on how we define $Q_{y'}(x,\,\cdot\,)$ when $QK(x,y')=0$, we observe that $(x,y)\in\hat{S}$ and $P(y,y')>0$ imply that $K(x,y)P(y,y')>0$ and hence $QK(x,y') = KP(x,y')>0$.

To prove the remaining statements, let $\hat{K}$ be the probability kernel from S to $\hat{T}$ defined by

$$\hat{K}\big(x,(x',y')\big) := 1_{\{x=x'\}}K(x,y') \qquad \big(x\in S,\ (x',y')\in\hat{S}\big),$$

and define $\psi:\hat{T}\to S$ by $\psi(x,y):=x$. If we show that

$$Q\hat{K} = \hat{K}\hat{P},$$

then the remaining claims follow from Theorem 3.5. We observe that

$$Q\hat{K}(x;x',y') = \sum_{x''\in S} Q(x,x'')1_{\{x''=x'\}}K(x'',y') = Q(x,x')K(x',y')$$

and

$$\begin{aligned}\hat{K}\hat{P}(x;x',y') &= \sum_{(x'',y'')\in\hat{T}} 1_{\{x=x''\}}K(x,y'')\hat{P}(x'',y'';x',y')\\ &= \sum_{y''\in\tilde{S}} K(x,y'')\hat{P}(x,y'';x',y'),\end{aligned}$$

so that $Q\hat{K}=\hat{K}\hat{P}$ can be written coordinatewise as

$$Q(x,x')K(x',y') = \sum_{y''\in\tilde{S}} K(x,y'')\hat{P}(x,y'';x',y') \qquad \big(x\in S,\ (x',y')\in\hat{S}\big).$$

To check this, we write

$$\begin{aligned}\sum_{y''\in\tilde{S}} K(x,y'')\hat{P}(x,y'';x',y') &= \sum_{y''\in\tilde{S}} K(x,y'')P(y'',y')\frac{Q(x,x')K(x',y')}{QK(x,y')}\\ &= \frac{KP(x,y')}{QK(x,y')}Q(x,x')K(x',y') = Q(x,x')K(x',y'),\end{aligned}$$

where we have used our assumption that $QK = KP$. ∎

Below, we revisit two of our previous examples of intertwining relations in the light of Theorem 3.6.

Thinning of exclusion processes In this example,

$$\hat{S} = \big\{(x, y) : x, y \in \{0,1\}^{C_n},\ y \leq x\big\},$$

and the joint evolution of (X, Y) can be described in the same way as in (3.2), using the same random pair I for both X and Y.

Condition on the future This example was already covered by Theorem 3.5, but we can also look at it a bit differently. Let K be the probability kernel from S to $T := \{0,1\}$ defined by

$$K(x, i) := h_i(x) \qquad (x \in S,\ i = 0, 1),$$

and let P be the trivial kernel on T defined by $P(0,0) := 1$ and $P(1,1) := 1$. Then one can check that

$$QK = KP$$

so Theorem 3.6 is applicable. Now Q_1 and Q_2 from (3.18) are just the Doob transformed kernels $Q^{(h_1)}$ and $Q^{(h_1)}$, while $\hat{T}$, $\hat{P}$, and the kernel $\hat{K}$ from the proof of Theorem 3.6 are what we called $\tilde{S}$, $\tilde{P}$, and $\tilde{K}$ before.

Exercise 3.7 (Intertwining based on generalized Doob transform) Let X be a Markov chain with finite state space S and transition kernel Q. Assume that $z \in S$ is a trap and set $S' := S \backslash \{z\}$. Assume that $Q' := Q|_{S'}$ is irreducible, and that $Q(x, z) > 0$ for at least one $x \in S'$. Let $c = \rho(Q')$ and let $h : S' \to (0, \infty)$ be the Perron-Frobenius eigenvector of Q', normalized such that $\max_{x \in S'} h(x) = 1$. Extend h to S by setting $h(z) := 0$ and observe that this implies that h is also an eigenvector of Q, i.e., $Qh = ch$. Define a probability kernel K from S to $\{1,2\}$ by setting

$$K(x, 1) := h(x) \quad \text{and} \quad K(x, 2) := 1 - h(x).$$

Find a probability kernel P on $T := \{1,2\}$ such that

$$QK = KP$$

and give expressions for the kernels Q_1 and Q_2 from (3.18). Check, in particular, that $Q_1 = Q^{(h_1)}$ is a generalized Doob transform of Q in the sense of (2.5).

Exercise 3.8 (Survival estimate) Let X be the Markov chain from Exercise 3.7 and let K and P be as in that exercise. Apply Theorem 3.6 to couple X to Markov chain Y with transition kernel P such that (3.20) holds. Use this coupling to derive the estimate

$$\mathbb{P}^x[X_n \in S'] \geq c^n h(x). \tag{3.21}$$

Remark 1 In view of Corollary 2.17, the estimate (3.21) is quite good, but it is not completely sharp since the function h in Theorem 2.16 is normalized in a different way than in the present section. Indeed, in Section 2.4 we chose h such that $\sum_{x \in S'} \eta(x)h(x) = 1$, where η is a probability measure, while at present we need $\sup_{x \in S'} h(x) \leq 1$. Thus (unless h is constant on S'), the h in our present section is always smaller than the one in Section 2.4.

Remark 2 Interesting examples of intertwining relations for birth-and-death processes can be found in [DM09, Swa11]. Lower estimates in the spirit of (3.21) but in the more complicated set-up of hierarchical contact processes have been derived in [AS10]. The 'evolving set process' in [LPW09, Thm 17.23] (which can be defined for quite general Markov chains) provides another nontrivial example of an intertwining relation.

Chapter 4

Branching processes

4.1 The branching property

Let S be a countable set and let $\mathcal{N}(S)$ be the set of all functions $x : S \to \mathbb{N}$ such that $\sum_{i\in S} x(i) < \infty$. It is easy to check that $\mathcal{N}(S)$ is a countable set (even though the set of *all* functions $x : S \to \mathbb{N}$ is uncountable). We interpret $x \in \mathcal{N}(S)$ as a collection of finitely many particles or individuals, where $x(i)$ is the number of individuals of *type* i, where S is the *type space*, or sometimes also the number of particles at the position i, where S represents physical space. Each $x \in \mathcal{N}(S)$ can be written as

$$x = \sum_{\beta=1}^{|x|} \delta_{i_\beta}, \tag{4.1}$$

where $|x| := \sum_i x(i)$, $i_1, \ldots, i_{|x|} \in S$, and $\delta_i \in \mathcal{N}(S)$ is defined as

$$\delta_i(j) := 1_{\{i=j\}} \qquad (i, j \in S).$$

This way of writing x is of course not unique but depends on the way we order the individuals. We will be interested in Markov processes with state space $\mathcal{N}(S)$, where in each time step, each individual, independently of the others, is replaced by a finite number of new individuals (its *offspring*). Let Q be a probability kernel from S to $\mathcal{N}(S)$. For a given $x \in \mathcal{N}(S)$ of the form (4.1), we can construct independent $\mathcal{N}(S)$-valued random variables $V^1, \ldots, V^{|x|}$ such that

$$\mathbb{P}[V^\beta \in \cdot\,] = Q(i_\beta, \,\cdot\,) \qquad (\beta = 1, \ldots, |x|).$$

Then

$$P(x,\,\cdot\,) := \mathbb{P}\big[\sum_{\beta=1}^{|x|} V^\beta \in \cdot\,\big]$$

defines a probability law on $\mathcal{N}(S)$. Doing this for each $x \in \mathcal{N}(S)$ defines a probability kernel P on $\mathcal{N}(S)$. By definition, the Markov chain X with state space $\mathcal{N}(S)$ and transition kernel P is called the *multitype branching process* with *offspring distribution* Q.

We define the *convolution* of two probability laws μ, ν on $\mathcal{N}(S)$ by

$$\mu * \nu(z) := \sum_{x' \leq z} \mu(x')\nu(z - x').$$

Note that this says that if X and Y are independent $\mathcal{N}(S)$-valued random variables with laws μ and ν, respectively, then $X + Y$ has law $\mu * \nu$. This is similar to the convolution of two probability laws on $\mathbb{R}^d$, except that in our case the space is $\mathcal{N}(S)$.

Lemma 4.1 (Branching property - first version) *The transition kernel P of a branching process has the property that*

$$P(x+y,\,\cdot\,) = P(x,\,\cdot\,) * P(y,\,\cdot\,) \qquad \big(x, y \in \mathcal{N}(S)\big). \tag{4.2}$$

Proof We need to check that

$$P(x+y, z) = \sum_{x' \leq z} P(x, x')P(y, z - x') \qquad \big(x, y, z \in \mathcal{N}(S)\big). \tag{4.3}$$

Let us write

$$x = \sum_{\beta=1}^{|x|} \delta_{i_\beta} \quad \text{and} \quad y = \sum_{\gamma=1}^{|y|} \delta_{j_\gamma},$$

and let $V^1, \ldots, V^{|x|}$ and $W^1, \ldots, W^{|y|}$ be all independent of each other such that

$$\begin{aligned} \mathbb{P}[V^\beta \in \cdot\,] &= Q(i_\beta,\,\cdot\,) \qquad (\beta = 1, \ldots, |x|), \\ \mathbb{P}[W^\gamma \in \cdot\,] &= Q(j_\gamma,\,\cdot\,) \qquad (\gamma = 1, \ldots, |y|). \end{aligned}$$

Then

$$\begin{aligned} P(x+y, z) &= \mathbb{P}\big[\sum_{\beta=1}^{|x|} V^\beta + \sum_{\gamma=1}^{|y|} W^\gamma = z\,\big] \\ &= \sum_{x' \leq z} \mathbb{P}\big[\sum_{\beta=1}^{|x|} V^\beta = x',\ \sum_{\gamma=1}^{|y|} W^\gamma = z - x'\,\big] = \sum_{x' \leq z} P(x, x')P(y, z - x') \end{aligned}$$

as required. ∎

In general, any probability kernel on $\mathcal{N}(S)$ for which (4.2) holds is said to have the *branching property*. In particular (4.2) implies that if

$$x = \sum_{\beta=1}^{|x|} \delta_{i_\beta},$$

then

$$P(x,\,\cdot\,) = P(\delta_{i_1},\,\cdot\,) * \cdots * P(\delta_{i_{|x|}},\,\cdot\,),$$

where $P(\delta_i,\,\cdot\,) = Q(i,\,\cdot\,)$. Thus, each Markov process that has the branching property is a branching process, and its transition probabilities are uniquely characterized by the offspring distribution.

Lemma 4.2 (Branching property - stronger version) *Let $X = (X_k)_{k\geq 0}$ and $Y = (Y_k)_{k\geq 0}$ be independent branching processes with the same type space $\mathcal{N}(S)$ and offspring distribution Q. Then $Z = (Z_k)_{k\geq 0}$, defined by*

$$Z_k(i) := X_k(i) + Y_k(i) \qquad (k \geq 0,\ i \in S) \tag{4.4}$$

is distributed as a branching processes with type space $\mathcal{N}(S)$ and offspring distribution Q.

Proof We need to check that

$$\mathbb{P}\big[Z_{k+1} = z \,\big|\, \mathcal{F}_k^Z\big] = P(Z_k, z) \quad \text{a.s.} \qquad \big(k \geq 0,\ z \in \mathcal{N}(S)\big),$$

where $(\mathcal{F}_k^Z)_{k\geq 0}$ is the filtration generated by Z. We will show that actually

$$\mathbb{P}\big[Z_{k+1} = z \,\big|\, \mathcal{F}_k^{(X,Y)}\big] = P(Z_k, z) \quad \text{a.s.} \qquad \big(k \geq 0,\ z \in \mathcal{N}(S)\big),$$

where $(\mathcal{F}_k^{(X,Y)})_{k\geq 0}$ is the filtration generated by $(X,Y) = (X_k, Y_k)_{k\geq 0}$. Since $\mathcal{F}_k^Z \subset \mathcal{F}_k^{(X,Y)}$ and $P(Z_k, z)$ is $\mathcal{F}_k^Z$-measurable, this then implies that

$$\mathbb{P}\big[Z_{k+1} = z \,\big|\, \mathcal{F}_k^Z\big] = \mathbb{E}\big[\mathbb{P}\big[Z_{k+1} = z \,\big|\, \mathcal{F}_k^{(X,Y)}\big] \,\big|\, \mathcal{F}_k^Z\big] = \mathbb{E}\big[P(Z_k, z) \,\big|\, \mathcal{F}_k^Z\big] = P(Z_k, z).$$

Thus, by the Markov property of (X,Y), it suffices to show that

$$\mathbb{P}\big[Z_{k+1} = z \,\big|\, X_k = x,\ Y_k = y\big] = P(x + y, z) \qquad \big(k \geq 0,\ x, y, z \in \mathcal{N}(S)\big).$$

Here

$$\begin{aligned}
&\mathbb{P}\big[Z_{k+1}=z\,\big|\,X_k=x,\ Y_k=y\big]\\
&\quad=\sum_{x'\leq z}\mathbb{P}\big[X_{k+1}=x',\ Y_{k+1}=z-x'\,\big|\,X_k=x,\ Y_k=y\big]\\
&\quad=\sum_{x'\leq z}P(x,x')P(y,z-x')=P(x+y,z),
\end{aligned}$$

where we have used that by the independence of X and Y, the process (X,Y) is a Markov process with transition kernel $P_2(x,y;x',y'):=P(x,x')P(y,y')$, and in the last equality we have used Lemma 4.1. ∎

It is not hard to see that a Markov process with state space $\mathcal{N}(S)$ has the branching property (4.4) if and only if its transition kernel has the branching property (4.2).

4.2 Generating functions

For each function $\phi:S\to\mathbb{R}$ and $x\in\mathcal{N}(S)$, let us write

$$\phi^x:=\prod_{i\in S}\phi(i)^{x(i)}=\prod_{\beta=1}^{|x|}\phi(i_\beta)\quad\text{where}\quad x=\sum_{\beta=1}^{|x|}\delta_{i_\beta},$$

where $\phi^0:=1$. It is easy to see that

$$\phi^{x+y}=\phi^x\phi^y\qquad\big(x,y\in\mathcal{N}(S),\ \phi:S\to\mathbb{R}\big).$$

Because of all the independence coming from the branching property, the linear operator P associated with the transition kernel of a branching process maps such 'multiplicative functions' into multiplicative functions. We will especially be interested in the case that ϕ takes values in $[0,1]$. We let $[0,1]^S$ denote the space of all functions $\phi:S\to[0,1]$.

Lemma 4.3 (Generating operator) *Let P denote the transition kernel of a multitype branching process with type space S and offspring distribution Q. Let U be the nonlinear operator defined by*

$$1-U\phi(i):=\sum_{x\in\mathcal{N}(S)}Q(i,x)(1-\phi)^x\qquad\big(i\in S,\ \phi\in[0,1]^S\big).$$

Then

$$Pf_\phi=f_{U\phi}\qquad\big(\phi\in[0,1]^S\big),$$

where for any $\phi\in[0,1]^S$, we define $f_\phi:\mathcal{N}(S)\to[0,1]$ by $f_\phi(x):=(1-\phi)^x$.

Proof If X is started in $X_0 = x$ with $x = \sum_{\beta=1}^{|x|} \delta_{i_\beta}$, then X_1 is equal in distribution to $\sum_{\beta=1}^{|x|} V^\beta$ where the V^β's are independent with distribution $Q(i_\beta,\,\cdot\,)$. It follows that

$$\begin{aligned} Pf_\phi(x) &= \mathbb{E}\big[(1-\phi)^{\sum_{\beta=1}^{|x|} V^\beta}\big] = \mathbb{E}\big[\prod_{\beta=1}^{|x|}(1-\phi)^{V^\beta}\big] \\ &= \prod_{\beta=1}^{|x|}\mathbb{E}\big[(1-\phi)^{V^\beta}\big] = \prod_{\beta=1}^{|x|}(1-U\phi)(i_\beta) = f_{U\phi}(x). \end{aligned}$$

∎

Remark 1 It would seem that the formulation of the lemma is simpler if we replace ϕ by $1-\phi$ everywhere, but as we will see later there are good reasons to formulate things in terms of $1-\phi$.

Remark 2 Our assumption that $0 \leq \phi \leq 1$ guarantees that the sum $\sum_x Q(i,x)\phi^x$ in the definition of $U\phi(i)$ is finite. Under more restrictive assumptions on Q, we can define $U\phi$ also for more general real-valued ϕ.

We call the nonlinear operator U from Lemma 4.3 the *generating operator* of the branching process with offspring distribution Q. By induction, Lemma 4.3 shows that $P^n f_\phi = f_{U^n\phi}$, or, in other words

$$\mathbb{E}^x\big[(1-\phi)^{X_n}\big] = (1-U^n\phi)^x \qquad \big(n \geq 0,\ \phi \in [0,1]^S\big). \tag{4.5}$$

The advantage of the operator U is that it acts on functions 'living' on the space S, while P acts on functions on the much larger space $\mathcal{N}(S)$. The price we pay for this is that U, unlike P, is not linear.

The next lemma shows that U contains, in a sense 'all information we need'.

Lemma 4.4 (Generating functions are distribution determining) *Let μ, ν be probability measures on $\mathcal{N}(S)$ such that*

$$\sum_{x\in\mathcal{N}(S)} \mu(x)(1-\phi)^x = \sum_{x\in\mathcal{N}(S)} \nu(x)(1-\phi)^x \qquad \big(\phi \in [0,1]^S\big).$$

Then $\mu = \nu$.

Proof We will prove the statement first under the assumtion that S is finite. Let $\mathcal{N}(S) \cup \{\infty\}$ be the one-point compactification of $\mathcal{N}(S)$. For each $\psi \in [0,1)^S$, define $g_\psi(x) := \psi^x$ and $g_\psi(\infty) := 0$. Then the g_ψ's are continuous functions on the compact space $\mathcal{N}(S)\cup\{\infty\}$. It is not hard to see that they *separate points*, i.e., for each $x \neq x'$ there exists a $\psi \in [0,1)^S$ such that $g_\psi(x) \neq g_\psi(x')$. Since $g_\psi g_{\psi'} = g_{\psi\psi'}$, the class $\{g_\psi : \psi \in [0,1)^S\}$ is closed under multiplication. Let $\mathcal{H}$ be the space of all linear combinations of functions from this class and the identity function. Then $\mathcal{H}$ is an algebra that separates points, hence by the Stone-Weierstras theorem $\mathcal{H}$ is dense in the space of continuous functions on $\mathcal{N}(S)\cup\{\infty\}$, equipped with the supremumnorm. By linearity and because μ and ν are probability measures, $\sum_x \mu(x)f(x) = \sum_x \nu(x)f(x)$ for all $f \in \mathcal{H}$. Since $\mathcal{H}$ is dense, it follows that $\mu = \nu$.

If S is not finite, then by applying our argument to functions ψ that are zero outside a finite set, we see that the finite-dimensional marginals of μ and ν agree, which shows that $\mu = \nu$ in general. ∎

There is a nice suggestive way of writing the relation (4.5). Generalizing (3.3), for any $\phi \in [0,1]^S$, we define a probability kernel K_ϕ from $\mathcal{N}(S)$ to $\mathcal{N}(S)$ by

$$K_\phi(x,\,\cdot\,) := P\big[\sum_{\beta=1}^{|x|}\chi_\beta\delta_{i_\beta} \in \cdot\,\big], \tag{4.6}$$

where $x = \sum_{\beta=1}^{|x|}\delta_{i_\beta}$ and the $\chi_1,\ldots,\chi_{|x|}$ are independent Bernoulli random variables with $\mathbb{P}[\chi_\beta = 1] = \phi(i_\beta)$. Thus, if Z is distributed according to the law $K_\phi(x,\,\cdot\,)$, then Z is obtained from x by independent *thinning*, where a particle of type i is kept with probability $\phi(i)$ and thrown away with the remaining probability. Note that K_ϕ has the branching property (4.2) and corresponds in fact to the offspring distribution

$$Q_\phi(i,y) := K_\phi(\delta_i,y) = \rho(i)1_{\{y=\delta_i\}} + (1-\rho(i))1_{\{y=0\}} \qquad \big(i \in S,\ y \in \mathcal{N}(S)\big).$$

Let $\mathrm{Thin}_\phi(x)$ denote a random variable with law $K_\phi(x,\,\cdot\,)$. Then

$$\mathbb{P}\big[\mathrm{Thin}_\phi(x) = 0\big] = (1-\phi)^x \qquad \big(x \in \mathcal{N}(S),\ \phi \in [0,1]^S\big),$$

where 0 denotes the configuration in $\mathcal{N}(S)$ with no particles. In view of this,

$$1 - U\phi(i) = \sum_{x\in\mathcal{N}(S)} Q(i,x)(1-\phi)^x = \delta_i PK_\phi(0)$$

and hence

$$U\phi(i) = \delta_i QK_\phi(\mathcal{N}(S)\backslash\{0\}). \tag{4.7}$$

Note that this says that if we start with one particle of type i, let it produce offspring, and then thin with ϕ, then $U\phi(i)$ is the probability that we are left with at least one individual. Likewise, we may rewrite (4.5) in the form

$$\delta_x P^n K_\phi(0) = \delta_x K_{U^n\phi}(0), \tag{4.8}$$

where $\mu P^n K_\phi$ is the law obtained by starting the branching process in x, running it till time t, and then applying the kernel K_ϕ, while $\delta_x K_{U^n\phi}$ is the law obtained by thinning x with $U^n\phi$.

Exercise 4.5 (Repeated thinning) Show that $K_\phi K_\psi = K_{\phi\psi}$ $(\phi, \psi \in [0,1]^S)$.

Exercise 4.6 (Thinning characterization) Let μ, ν be probability laws on $\mathcal{N}(S)$ such that $\mu K_\phi(0) = \nu K_\phi(0)$ for all $\phi \in [0,1]^S$. Show that $\mu = \nu$.

4.3 The survival probability

Let X be a branching process with type space S and generating operator U. We observe that

$$\begin{aligned}
&\mathbb{P}^{\delta_i}\big[X_n \neq 0\big] = 1 - \mathbb{P}^{\delta_i}\big[\mathrm{Thin}_1(X_n) = 0\big] \\
&\quad = 1 - \mathbb{P}^{\delta_i}\big[\mathrm{Thin}_{U^n 1}(\delta_i) = 0\big] = 1 - \big(1 - U^n 1(i)\big) = U^n 1(i),
\end{aligned}$$

where we use the symbol 1 also to denote the function that is constantly one. Since 0 is a trap for any branching process,

$$\mathbb{P}^{\delta_i}\big[X_n \neq 0\big] = \mathbb{P}^{\delta_i}\big[X_k \neq 0\ \forall 0 \leq k \leq n\big] \underset{n\to\infty}{\longrightarrow} \mathbb{P}^{\delta_i}\big[X_k \neq 0\ \forall k \geq 0\big],$$

where we have used the continuity of our probability measure with respect to decreasing sequences of events. In view of this, let us write

$$\rho(i) := \lim_{n\to\infty} U^n 1(i) = \mathbb{P}^{\delta_i}\big[X_k \neq 0\ \forall k \geq 0\big] \tag{4.9}$$

for the probability to survive starting from a single particle of type i.

Lemma 4.7 (Survival probability) *The function ρ in (4.9) is the largest solution (in $[0,1]^S$) of the equation*

$$U\rho = \rho,$$

i.e., ρ solves this equation and any other solutions ρ' of this equation, if they exist, satisfy $\rho' \leq \rho$.

Proof Since

$$U\phi(i) = \sum_x Q(i,x)\big(1-(1-\phi)^x\big),$$

and since $\phi \mapsto 1-(1-\phi)^x$ is a nondecreasing function, we see that

$$\phi \leq \psi \quad \text{implies} \quad U\phi \leq U\psi. \tag{4.10}$$

By monotone convergence, we see moreover that

$$\phi_n \downarrow \phi \quad \text{implies} \quad U\phi_n \downarrow U\phi. \tag{4.11}$$

Since $1 \geq U1$ we see by (4.10) and induction that $U^n 1 \geq U^{n+1}1$ and $U^n 1 \downarrow \rho$, which was in fact also clear from our probabilistic interpretation. By (4.11), it follows that

$$U\rho = U \lim_{n\to\infty} U^n 1 = \lim_{n\to\infty} U^{n+1} = \rho.$$

Now if $\rho' \in [0,1]^S$ is any other fixed point of U, then by (4.10) and (4.11)

$$\rho' \leq 1 \quad \text{implies} \quad \rho' = U^n \rho' \leq U^n 1 \underset{\rho\to\infty}{\longrightarrow},$$

which shows that $\rho' \leq \rho$. ∎

Exercise 4.8 (Galton-Watson process) Let X be a branching process whose type space $S = \{1\}$ consists of a single point, and let Q be its offspring distribution. We identify $\mathcal{N}(\{1\}) \cong \mathbb{N}$. Since there is only one type of individual, we only need to know with which probability a single individual produces n offspring ($n \geq 0$). Thus, we simply write

$$Q(n) = Q(1, n\delta_1)$$

which is a probability law on $\mathbb{N}$. Assume that Q has a finite second moment and let

$$a := \sum_{n=0}^{\infty} nQ(n)$$

denote its mean. We identify $[0,1]^S \cong [0,1]$ and let $U : [0,1] \to [0,1]$ be the generating operator of X, which is now just a (nonlinear) function from $[0,1]$ to $[0,1]$.

(a) Show that U is a concave function and that $U'(0) = a$.

(b) Assume that $Q(1) < 1$. Show that

$$P^1[X_k \neq 0 \ \forall k \geq 0] > 0$$

if and only if $a > 1$.

Remark Single-type branching processes as in Excercise 4.8 are called *Galton-Watson processes* after the seminal paper (from 1875) [WG75], where they proved that the survival probability ρ solves $U\rho = \rho$ but incorrectly concluded from this that the process dies out for *all* $a \geq 0$, since 'obviously the solution of this equation is $\rho = 0$'.

Exercise 4.9 (Spatial branching) Let $(I_k)_{k\geq 0}$ be i.i.d. with $\mathbb{P}[I_k = -1] = 1/2 = \mathbb{P}[I_k = 1]$, and let N be a Poisson distributed random variable with mean a, independent of $(I_k)_{k\geq 0}$. Let X be the branching process with type space $\mathbb{Z}$ and offspring distribution Q given by

$$Q(i,\,\cdot\,) := \mathbb{P}\big[\sum_{k=1}^{N} \delta_{i+I_k} \in \cdot\,\big] \qquad (i \in \mathbb{Z}),$$

which says that a particle at i produces $\mathrm{Pois}(a)$ offspring which are independently placed on either $i-1$ or $i+1$, with equal probabilities. Show that

$$\mathbb{P}^{\delta_0}\big[X_k \neq 0 \ \forall k \geq 0\big] > 0$$

if and only if $a > 1$.

Exercise 4.10 (Two-type process) Let X be a branching process with type space $S = \{1,2\}$ and the following offspring distribution. Individuals of both types produce a Poisson number of offspring, with mean a. If the parent is of type 1, then its offspring are, independently of each other, of type 1 or 2 with probability 1/2 each. All offspring of individuals of type 2 are again of type 2. Starting with a single individual of type 1, for what values of a is there a positive probability that there will be individuals of type 1 at all times?

Exercise 4.11 (Poisson offspring) Let X be a Galton-Watson process where each individual produces a Poisson number of offspring with mean a, and let

$$\rho_n := \mathbb{P}^1[X_n = 0] \qquad (n \geq 0)$$

be the probability that the process started with a single individual is extinct after n steps. Prove that $\rho_{n+1} = e^{a(\rho_n - 1)}$.

4.4 First moment formula

Proposition 4.12 (First moment formula) *Let X be a branching process with type space S and offspring distribution Q. Assume that the matrix*

$$A(i,j) := \sum_{x\in\mathcal{N}(S)} Q(i,x)x(j) \qquad (i,j\in S) \tag{4.12}$$

satisfies

$$\sup_{i\in S}\sum_{j\in S} A(i,j) < \infty. \tag{4.13}$$

Then

$$\mathbb{E}^x[X_n(j)] = \sum_i x(i)A^n(i,j) \qquad \big(x\in\mathcal{N}(S),\ j\in S,\ n\geq 0\big).$$

Proof We first prove the statement for $n = 1$. If $x = \sum_{\beta=1}^{|x|}\delta_{i_\beta}$, then X_1 is equal in distribution to $\sum_{\beta=1}^{|x|} V^\beta$ where the V^β's are independent with distribution $Q(i_\beta,\,\cdot\,)$. Therefore,

$$\begin{aligned}\mathbb{E}^x[X_1(j)] &= \mathbb{E}\big[\sum_{\beta=1}^{|x|} V^\beta(j)\big] = \sum_{\beta=1}^{|x|}\mathbb{E}\big[V^\beta(j)\big]\\ &= \sum_{\beta=1}^{|x|}\sum_y Q(i_\beta,y)y(j) = \sum_{\beta=1}^{|x|} A(i_\beta,j) = \sum_i x(i)A(i,j).\end{aligned}$$

By induction, it follows that

$$\begin{aligned}\mathbb{E}^x[X_{n+1}(j)] &= \sum_{x'}\mathbb{P}^x[X_n = x']\mathbb{E}^x[X_{n+1}(j)\,|\,X_n = x']\\ &= \sum_{x'}\mathbb{P}^x[X_n = x']\sum_i x'(i)A(i,j) = \sum_i A(i,j)\sum_{x'}\mathbb{P}^x[X_n = x']x'(i)\\ &= \sum_i A(i,j)\sum_{x'}\mathbb{E}^x[X_n(i)] = \sum_i A(i,j)\sum_k x(k)A^n(k,i) = A^{n+1}(i,k),\end{aligned}$$

where all expressions are finite by (4.13). ∎

Lemma 4.13 (Subcritical processes) *Let X be a branching process with finite type space S and offspring distribution Q. Assume that its first moment matrix A defined in (4.12) is irreducible and satisfies (4.13), and let α be its Perron-Frobenius eigenvalue. If $\alpha < 1$, then*

$$\mathbb{P}^x\big[X_k \neq 0\ \forall k\geq 0\big] = 0 \qquad \big(x\in\mathcal{N}(S)\big).$$

Proof For any $f : S \to \mathbb{R}$, let $l_f : \mathcal{N}(S) \to \mathbb{R}$ denote the 'linear' function

$$l_f(x) := \sum_{i \in S} x(i) f(i) \qquad \big(x \in \mathcal{N}(S),\ f : S \to \mathbb{R}\big).$$

Then by Lemma 4.12,

$$\begin{aligned} P^n l_f(x) &= \mathbb{E}^x\big[l_f(X_n)\big] = \sum_i f(i) \mathbb{E}^x\big[X_n(i)\big] \\ &= \sum_i f(i) \sum_j x(j) A^n(j,i) = \sum_j x(j) A^n f(j) = l_{A^n f}(x). \end{aligned}$$

In particular, if h is the (strictly positive) right eigenvector of A with eigenvalue α, then

$$P^n l_h = l_{A^n h} = l_{\alpha^n h},$$

which says that

$$\mathbb{E}^x\big[h(X_n)\big] = \alpha^n h(x),$$

which tends to zero by our assumption that $\alpha < 1$. Since h is strictly positive, it follows that $\mathbb{P}^x[X_n \neq 0] \to 0$. ∎

Proposition 4.14 (Critical processes) *Let X be a branching process with finite type space S and offspring distribution Q. Assume that its first moment matrix A defined in (4.12) is irreducible and satisfies (4.13), and let α be its Perron-Frobenius eigenvalue. If $\alpha = 1$ and there exists some $i \in S$ such that $Q(i,0) > 0$, then*

$$\mathbb{P}^x\big[X_k \neq 0\ \forall k \geq 0\big] = 0 \qquad \big(x \in \mathcal{N}(S)\big).$$

Proof Let $\mathcal{A} := \{X_n \neq 0\ \forall n \geq 0\}$ and let

$$\rho(i) := \mathbb{P}^{\delta_i}(\mathcal{A}).$$

By the branching property,

$$\mathbb{P}^x(\mathcal{A}^{\mathrm{c}}) = (1-\rho)^x \qquad \big(x \in \mathcal{N}(S)\big).$$

By the principle 'what can happen must happen' (Proposition 0.14), we have

$$(1-\rho)^{X_n} \underset{0 \to \infty}{\longrightarrow} \quad \text{a.s. on the event } \mathcal{A}.$$

By irreducibility and the fact that $Q(j,0)>0$ for some j, it is not hard to see that $\rho(i)<1$ for all $i\in S$. By the finiteness of S, it follows that $\sup_{i\in S}\rho(i)<1$ and hence

$$\inf_{x\in\mathcal{N}(S),\ |x|\leq N}(1-\rho)^x>0\qquad(N\geq 0).$$

It follows that

$$|X_n|\underset{n\to\infty}{\longrightarrow}\infty\quad\text{a.s. on the event }\mathcal{A},\tag{4.14}$$

i.e., the only way for the process to survive is to let the number of particles tend to infinity.

Let h be the (strictly positive) right eigenvector of A with eigenvalue α. Then

$$\mathbb{E}\big[l_h(X_{n+1})\,\big|\,\mathcal{F}^X_n\big]=(Pl_h)(X_n)=l_{\alpha h}(X_n)=\alpha l_h(X_n),$$

which shows that (provided that $\mathbb{E}[h(X_0)]<\infty$, which is satisfied for processes started in deterministic initial states) the process

$$M_n:=\alpha^{-n}\sum_{i\in S}h(i)X_n(i)\qquad(n\geq 0)\tag{4.15}$$

is a nonnegative martingale. By martingale convergence, it follows that there exists a random variable M_∞ such that

$$M_n\underset{n\to\infty}{\longrightarrow}M_\infty\quad\text{a.s.}\tag{4.16}$$

In particular, if $\alpha=1$, this proves that

$$|X_n|\underset{n\to\infty}{\not\longrightarrow}\infty\quad\text{a.s.},$$

which by (4.14) implies that $\mathbb{P}(\mathcal{A})=0$. ∎

4.5 Second moment formula

We start with some general Markov chain theory. For any probability law μ on a countable set S and functions $f,g:S\to\mathbb{R}$, let us write

$$\mathrm{Cov}_\mu(f,g):=\mu(fg)-(\mu f)(\mu g)$$

for the covariace of f and g, whenever this is well-defined. Note that if X is a random variable with law μ, then $\mathrm{Cov}_\mu(f,g)=\mathbb{E}[f(X)g(X)]-\mathbb{E}[f(X)]\mathbb{E}[g(X)]$, in accordance with the usual formula for the covariance of two functions.

Lemma 4.15 (Covariance formula) *Let X be a Markov chain with countable state space S and transition kernel P. Let μ be a probability law on S and let $\mathcal{C}$ be a class of functions $f : S \to \mathbb{R}$ such that*

(i) *$f \in \mathcal{C}$ implies $Pf \in \mathcal{C}$.*

(ii) *$\sum_{x,y} \mu(x)P(x,y)|f(y)g(y)| < \infty$ for all $x \in S$ and $f, g \in \mathcal{C}$.*

Then, for any $f, g \in \mathcal{C}$,

$$\mathrm{Cov}_{\mu P^n}(f,g) = \mathrm{Cov}_\mu(P^n f, P^n g) + \sum_{k=1}^{n} \mu P^{n-k} \Gamma(P^{k-1} f, P^{k-1} g), \tag{4.17}$$

where

$$\Gamma(f,g) := P(fg) - (Pf)(Pg) \qquad (f, g \in \mathcal{C}).$$

Remark 1 If $X = (X_k)_{k \geq 0}$ is the Markov chain with initial law μ and transition kernel P, then

$$\begin{aligned}
\mathrm{Cov}_{\mu P^n}(f,g) &= \mathbb{E}\big[f(X_n)g(X_n)\big] - \mathbb{E}\big[f(X_n)\big]\mathbb{E}\big[g(X_n)\big], \\
&=: \mathrm{Cov}\big(f(X_n), g(X_n)\big),
\end{aligned}$$

and similarly

$$\mathrm{Cov}_\mu(P^n f, P^n g) = \mathrm{Cov}\big((P^n f)(X_0), (P^n g)(X_0)\big).$$

Remark 2 The assumptions of the lemma are trivially fulfilled if we take for $\mathcal{C}$ the class of all bounded real functions on S. Often, we also need the lemma for certain unbounded functions, but in this case we need to find a class $\mathcal{C}$ satisfying the assumptions of the lemma to ensure that all second moments are finite.

Proof of Lemma 4.15 The statement is trivial for $n = 0$. Fix $n \geq 1$ and define a function $H : \{0, \ldots, n\} \to \mathbb{R}$ by

$$H(k) := P^k\big((P^{n-k} f)(P^{n-k} g)\big) \qquad (0 \leq k \leq n).$$

Then

$$\begin{aligned}
&\mu\big(H(n) - H(0)\big) = \mu P^n(fg) - \mu\big((P^n f)(P^n g)\big) \\
&\quad = \big[\mu P^n(fg) - (\mu P^h f)(\mu P^n g)\big] - \big[\mu\big((P^n f)(P^n g)\big) - (\mu P^h f)(\mu P^n g)\big] \\
&\quad = \mathrm{Cov}_{\mu P^n}(f,g) - \mathrm{Cov}_\mu(P^n f, P^n g).
\end{aligned}$$

It follows that

$$\begin{aligned}
&\mathrm{Cov}_{\mu P^n}(f,g) - \mathrm{Cov}_\mu(P^n f, P^n g) = \sum_{k=1}^n \mu\big[H(k) - H(k-1)\big] \\
&= \sum_{k=1}^n \mu\big[P^k\big((P^{n-k}f)(P^{n-k}g)\big) - P^{k-1}\big((P^{n-k+1}f)(P^{n-k+1}g)\big)\big] \\
&= \sum_{k=1}^n \mu P^{k-1}\Gamma(P^{n-k}f, P^{n-k}g).
\end{aligned}$$

Changing the summation order (setting $k' := n-k+1$), we arrive at (4.17). ∎

We now apply this general formula to branching processes. To simplify matters, we will only look at finite type spaces.

Proposition 4.16 (Second moment formula) *Let X be a branching process with finite type space S and offspring distribution Q. Let V^i denote a random variable with law $Q(i,\cdot\,)$ and assume that*

$$\begin{aligned}
A(i,j) &:= \mathbb{E}[V^i(j)], \\
C(i;j,k) &:= \mathbb{E}[V^i(j)V^i(k)]
\end{aligned} \tag{4.18}$$

are finite for all $i,j,k \in S$. Let A be the linear operator with matrix $A(i,j)$ and for functions $f,g : S \to \mathbb{R}$, let $C(f,g) : S \to \mathbb{R}$ be defined by

$$C(f,g)(i) := \sum_{j,k\in S} C(i;j,k) f(j) g(k).$$

Fox $x \in \mathcal{N}(S)$ and $f : S \to \mathbb{R}$, let $xf := \sum_i x(i)f(i)$. Then, for functions $f,g : S \to \mathbb{R}$, one has

$$\begin{aligned}
\mathbb{E}^x\big[X_n f\big] &= xA^n f, \\
\mathrm{Cov}^x\big(X_n f, X_n g\big) &= \sum_{k=1}^n xA^{n-k} C(A^{k-1}f, A^{k-1}g),
\end{aligned} \tag{4.19}$$

where Cov^x *denotes covariance w.r.t. to the law* $\mathbb{P}^x$.

Proof The first formula in (4.19) has already been proved in Lemma 4.12. As in the proof of Lemma 4.13, for any real function f on S, let $l_f : \mathcal{N}(S) \to \mathbb{R}$ denote the 'linear' function $l_f(x) := \sum_i f(i)x(i)$. Then the first formula in (4.19) says that

$$P^n l_f = l_{A^n f} \qquad (f \in \mathbb{R}^S),$$

which motivates us to take for the class $\mathcal{C}$ in Lemma 4.15 the class of 'linear' functions l_f with $f : S \to \mathbb{R}$ any function. Using the fact that $C(i; j, k) < \infty$ for all $i, j, k \in S$, it is not hard to prove that $\mathcal{C}$ satisfies the assumptions of Lemma 4.15. Let $x = \sum_{\beta=1}^{|x|} \delta_{i_\beta}$ and let $V^1, \ldots, V^{|x|}$ be independent such that V^β is distributed according to $Q(i_\beta, \cdot\,)$. We calculate

$$\begin{aligned}
\Gamma(l_f, l_g)(x) &= \big(P(l_f l_g) - (Pl_f)(Pl_g)\big)(x) \\
&= \mathrm{Cov}\Big(\sum_{j,\beta} f(j)V^\beta(j)\,,\,\sum_{k,\gamma} g(k)V^\gamma(k)\Big) = \sum_{jk} f(j)g(k)\sum_\beta \mathrm{Cov}\big(V^\beta(j), V^\beta(k)\big) \\
&= \sum_{ijk} x(i)f(j)g(k)C(i; j, k) = l_{C(f,g)}(x),
\end{aligned}$$

where we have used that $\mathrm{Cov}\big(V^\beta(i), V^\gamma(j)\big) = 0$ for $\beta \neq \gamma$ by independence. Then Lemma 4.15 tells us that

$$\begin{aligned}
\mathrm{Cov}_{\delta_x P^n}(l_f, l_g) &= \sum_{k=1}^n \delta_x P^{n-k}\Gamma(P^{k-1}l_f, P^{k-1}l_g) = \sum_{k=1}^n \delta_x P^{n-k}\Gamma(l_{A^{k-1}f}, l_{A^{k-1}g}) \\
&= \sum_{k=1}^n \delta_x P^{n-k} l_{C(A^{k-1}f,\, A^{k-1}g)} = \sum_{k=1}^n l_{A^{n-k}C(A^{k-1}f,\, A^{k-1}g)}(x),
\end{aligned}$$

which proves the second formula in (4.19). ∎

4.6 Supercritical processes

The aim of this section is to prove that supercritical branching processes survive.

Proposition 4.17 (Supercritical process) *Let X be a branching process with finite type space S and offspring distribution Q. Assume that the first and second moments $A(i, j)$ and $C(i; j, k)$ of Q, defined in (4.18), are all finite and that A is irreducible. Assume that the Perron-Frobenius eigenvalue α of A satisfies $\alpha > 1$. Then*

$$\mathbb{P}^x\big[X_k \neq 0\ \forall k \geq 0\big] > 0 \qquad \big(x \in \mathcal{N}(S),\ x \neq 0\big).$$

Proof Let h be the (strictly positive) right eigenvector of A with eigenvalue α. We have already seen in formulas (4.15)–(4.16) in the proof of Proposition 4.14 that

$M_n = \alpha^{-n} X_n h$ is a nonnegative martingale that converges to an a.s. limit M_∞. By Proposition 0.8, if we can show that M is uniformly integrable, then

$$\mathbb{E}^x[M_\infty] = xh > 0$$

for all $x \neq 0$, which shows that the process survives with positive probability. Thus, it suffices to show that

$$\sup_{n\geq 0} \mathbb{E}^x\big[M_n^2\big] < \infty.$$

We write

$$\mathbb{E}^x\big[M_n^2\big] = \big(\mathbb{E}^x[\alpha^{-n}X_n h]\big)^2 + \mathrm{Var}^x(\alpha^{-n}X_n h),$$

where by the first formula in (4.19)

$$\big(\mathbb{E}^x[\alpha^{-n}X_n h]\big)^2 = (\alpha^{-n} x A^n h)^2 = (xh)^2$$

is clearly bounded uniformly in n. We observe that since h is strictly positive and S is finite, we can find a constant $K < \infty$ such that

$$C(h,h) \leq Kh.$$

Therefore, applying the second formula in (4.19), we see that

$$\begin{aligned}
&\mathrm{Var}^x(\alpha^{-n}X_n h) = \alpha^{-2n}\sum_{k=1}^n xA^{n-k}C(A^{k-1}h, A^{k-1}h)\\
&= \alpha^{-2n}\sum_{k=1}^n xA^{n-k}\alpha^{2(k-1)}C(h,h) \leq \alpha^{-2n}K\sum_{k=1}^n xA^{n-k}\alpha^{2(k-1)}h\\
&= \alpha^{-2n}K\sum_{k=1}^n \alpha^{n-k}\alpha^{2(k-1)}xh = Kxh\sum_{k=1}^n \alpha^{k-n-2} \leq Kxh\alpha^{-2}\sum_{k=0}^\infty \alpha^{-k} < \infty,
\end{aligned}$$

uniformly in $n \geq 1$, where we have used that $\alpha > 1$. ∎

4.7 Trimmed processes

Let X be a multitype branching process with countable type space S and offspring distribution Q. Recall from Lemma 4.7 that

$$\rho(i) := \mathbb{P}^{\delta_i}\big[X_k \neq 0 \ \forall k \geq 0\big] \qquad (i \in S)$$

is the largest solution of the equation $U\rho = \rho$, where U is the generating operator of X. Let $\rho \in [0,1]^S$ be any solution of $U\rho = \rho$. The aim of the present section is to prove a result similar to conditioning on the future (as in Proposition 1.6) or intertwining of processes with one trap (Exercise 3.7), but now on the level of the individuals in a branching process. More precisely, we will divide the population into two 'sorts' of individuals, with probabilities (depending on the type) $\rho(i)$ and $1-\rho(i)$. In particular, if ρ is the survival probability, then we divide the population at each time k into those individuals which we know are going to survive (or, more precisely, that have living descendants at all times), and those whose descendants are going to die out completely.

To this aim, let $x = \sum_{\beta=1}^{|x|} \delta_{i_\beta} \in \mathcal{N}(S)$ and let $\chi_1, \ldots, \chi_{|x|}$ be independent Bernoulli random variables with $\mathbb{P}[\chi_\beta = 1] = \rho(i_\beta)$. Doing this for any $x \in \mathcal{N}(S)$, we define a probability kernel L_ρ from $\mathcal{N}(S)$ to $\mathcal{N}(S \times \{0,1\})$ by

$$L_\rho(x,\,\cdot\,) := \mathbb{P}\big[\sum_{\beta=1}^{|x|} \delta_{(i_\beta,\chi_\beta)} \in \cdot\,\big].$$

Another way to describe L_ρ is to note that L_ρ has the branching property (4.2) and

$$L_\rho(\delta_i, y) = \rho(i)1_{\{y=\delta_{(i,1)}\}} + (1-\rho(i))1_{\{y=\delta_{(i,0)}\}} \qquad \big(i \in S,\ y \in \mathcal{N}(\hat{S})\big).$$

Note that this is very similar to the thinning kernel K_ρ defined in (4.6).

We set

$$\hat{S} := \big\{(i,\sigma) : i \in S,\ \sigma \in \{0,1\},\ K(\delta_i, \delta_{(i,\sigma)}) > 0\big\},$$

i.e.,

$$\begin{aligned} &\hat{S} := S_0 \cup S_1, \quad \text{where} \\ &S_0 := \{(i,0) : 1-\rho(i) > 0\} \quad \text{and} \quad S_1 := \{(i,0) : 1-\rho(i) > 0\}. \end{aligned}$$

Then L_ρ is in effect a probability kernel from $\mathcal{N}(S)$ to $\mathcal{N}(\hat{S})$.

For any $x \in \mathcal{N}(S)$ and $S' \subset S$, let us write

$$x\big|_{S'} := \big(x(i)\big)_{i\in S'} \in \mathcal{N}(S')$$

for the restriction of x to S' (and similarly for subsets of $\hat{S}$). With this notation, if Y is a random variable with law $L_\rho(x,\,\cdot\,)$, then $Y|_{S_1}$ (resp. $Y|_{S_0}$) is a thinning of x with the function ρ (resp. $1-\rho$).

Recall that the offspring distribution Q of X is a probability kernel from S to $\mathcal{N}(S)$. Let

$$\mathcal{E} := \{y \in \mathcal{N}(\hat{S}) : y\big|_{S_1} = 0\}.$$

Observe that QL_ρ is a probability kernel from S to $\mathcal{N}(\hat{S})$. For given $i \in S$, the probability of $\mathcal{E}$ under $QL_\rho(i,\,\cdot\,)$ is given by

$$QL_\rho(i,\mathcal{E}) = \sum_{x\in\mathcal{N}(S)} Q(i,x)(1-\rho)^x = 1 - U\rho(i) = 1 - \rho(i), \qquad (4.20)$$

where we have used that $U\rho = \rho$. We define a probability kernel $\hat{Q}$ from $\hat{S}$ to $\mathcal{N}(\hat{S})$ by

$$\hat{Q}(i,\sigma;\,\cdot\,) := \begin{cases} QL_\rho\big(i,\,\cdot\,\big|\,\mathcal{E}\big) & \text{if } \sigma = 0, \\ QL_\rho\big(i,\,\cdot\,\big|\,\mathcal{E}^{\mathrm{c}}\big) & \text{if } \sigma = 1, \end{cases}$$

where we are conditioning the probability law $QL_\rho\big(i,\,\cdot\,\big)$ on the event $\mathcal{E}$ and its complement, respectively. By (4.20), these conditional probabilities are well-defined, i.e., $QL_\rho(i,\mathcal{E}) > 0$ for all $(i,0) \in S_0$ and $QL_\rho(i,\mathcal{E}^{\mathrm{c}}) > 0$ for all $(i,1) \in S_1$.

In words, our definition of $\hat{Q}$ says that an individual of type $(i,0) \in S_0$ produces offspring in the following manner. First, we produce offspring according to the law $Q(i,\,\cdot\,)$ of our original branching process. Next, we assign to these individuals independent 'signs' $0,1$ with probabilities depending on the type of the individual through the function ρ. Finally, we condition on producing only offspring with sign 0. Individuals of type $(i,1) \in S_1$ reproduce similarly, but in this case we condition on producing at least one offspring with sign 1. Note that individuals of a type in S_1 always produce at least one offspring in S_1, and possibly also offspring in S_0. Individuals in S_0 produce only offspring in S_0, and possibly no offspring at all.

Theorem 4.18 (Distinguishing surviving particles) *Let $X = (X_k)_{k\geq 0}$ be a branching process with finite type space S and irreducible offspring distribution Q. Assume that X survives (with positive probability). Let ρ, L_ρ, $\hat{S}$, and $\hat{Q}$ be as defined above. Then X can be coupled to a branching process Y with type space $\hat{S}$ and offspring distribution $\hat{Q}$, in such a way that*

$$\mathbb{P}\big[Y_n = y\,\big|\,(X_k)_{0\leq k\leq n}\big] = L_\rho(X_n,\,\cdot\,) \qquad (y \in \hat{S},\ n \geq 0). \qquad (4.21)$$

Proof We apply Theorem 3.6. In fact, for our present purpose Theorem 3.5 is sufficient, where the function ψ occurring there is given by

$$\psi(y)(i) := y(i,0) + y(i,1) \qquad \big(y \in \mathcal{N}(\hat{S}),\ i \in S\big).$$

Let P and $\hat{P}$ denote the transition kernels of X and Y, respectively. We need to check that

$$PL_\rho = L_\rho \hat{P}.$$

Since P, L_ρ, and $\hat{P}$ have the branching property (4.2), it suffices to check that

$$\delta_i PL_\rho = \delta_i L_\rho \hat{P} \qquad (i \in S).$$

Indeed, by our definition of $\hat{Q}$,

$$\begin{aligned}\delta_i L_\rho \hat{P} &= (1-\rho(i))Q(i,0;\,\cdot\,) + \rho(i)Q(i,1;\,\cdot\,)\\ &= (1-\rho(i))QL_\rho(i,\,\cdot\,|\,\mathcal{E}) + \rho(i)QL_\rho(i,\,\cdot\,|\,\mathcal{E}^{\mathrm{c}}) = QL_\rho(i,\,\cdot\,) = \delta_i PL_\rho,\end{aligned}$$

where we have used (4.20). ∎

Proposition 4.19 (Trimmed process) *Let X and Y be the branching processes with type spaces S and $\hat{S}$ in Theorem 4.18, let ρ be as in that theorem and let U be the generating operator of X. Then*

$$\big(Y_k\big|_{S_1}\big)_{k\geq 0}$$

is a branching process with type space S_1 and generating operator U^ρ given by

$$U^\rho\phi(i) = \rho^{-1}(i)U(\rho\phi)(i) \qquad \big(i \in S_1,\ \phi \in [0,1]^{S_1}\big),$$

where $\rho\phi$ denotes the pointwise product of ρ and ϕ, which is extended to a function on S by setting $\rho\phi(j) = 0$ for all $j \in S\backslash S_1$.

Proof Since individuals in S_0 never produce offpring in S_1, it is clear that the restriction of Y to S_1 is a branching process. Let Q' be the offspring distribution of the restricted process. Then $Q'(i,\,\cdot\,)$ is just $QK_\rho(i,\,\cdot\,)$ conditioned on producing at least one offspring, where K_ρ is the thinning kernel defined in (4.6). Then, setting $\mathcal{G} := \mathcal{N}(S_1)\backslash\{0\}$, we have by (4.7)

$$U^\rho\phi(i) = \delta_i Q' K_\phi(\mathcal{G}) = \frac{\delta_i QK_\rho K_\phi(\mathcal{G})}{\delta_i QK_\rho(\mathcal{G})} = \frac{\delta_i QK_{\rho\phi}(\mathcal{G})}{\delta_i QK_\rho(\mathcal{G})} = \frac{U(\rho\phi)(i)}{U\rho(i)} = \rho(i)^{-1}U(\rho\phi)(i),$$

where we have used that $U\rho = \rho$. ∎

Remark In particular, if ρ is the survival probability, then the branching process Y from Proposition 4.19 has been called the *trimmed tree* of X in [FS04], which deals with continuous type spaces and continuous time. Similar constructions have been used in branching theory long before this paper and go under different names. It seems *skeletal process* is the term that is most used nowadays. An overview of the literature on skeletal processes can be found in [EKW15].

Exercise 4.20 (Nonbranching process) Let S be a countable set, let P' be a probability kernel on S, and let Q be the offspring distribution defined by

$$Q(i,\delta_j) := P'(i,j) \qquad (i,j\in S),$$

and $Q(i,x) := 0$ if $|x| \neq 1$. In other words: a particle at the position i produces exactly one offspring on a position that is distributed according to $P'(i,\,\cdot\,)$. Let X be the branching process with type space S and ofspring distribution Q and let U be its generating operator. Show that $U = P'$. In particular, this says that a function $\rho\in[0,1]^S$ solves $U\rho=\rho$ if and only if ρ is harmonic for P' and $U^\rho = (P')^\rho$ is just the classical Doob transform of P'.

If X is a multitype branching process with type space S and offspring distribution Q, then let us write $i\to j$ if $Q(i,\{x : x(j)>0\})>0$ and $i \rightsquigarrow j$ if there exist $i=i_0\to\cdots\to i_n=j$. We say that Q is *irreducible* if $i\rightsquigarrow j$ for all $i,j\in S$. In particular, if Q has finite first moments, then this is equivalent to irreducibility of the matrix $A(i,j)$ in (4.18). We also define aperiodicity of Q in the obvious way, i.e., an irreducible Q is aperiodic if the greatest common divisior of $\{n\geq 1 : P^n(\delta_i,\{x : x\geq\delta_i\})>0\}$ is one for some, and hence for all $i\in S$. If Q is irreducible and

$$\rho(i) = \mathbb{P}^{\delta_i}\big[X_k\neq 0\ \forall k\geq 0\big] \qquad (i\in S), \tag{4.22}$$

then it is not hard to see that either $\rho(i)>0$ for all $i\in S$, or $\rho(i)=0$ for all $i\in S$. In the first case, we say that X *survives*, while in the second case we say that X *dies out*.

Exercise 4.21 (Immortal process) Let X be a branching process with finite type space S, offspring distribution Q, and generating operator U. Assume that Q is irreducible and aperiodic and that $Q(i,0)=0$ for each $i\in S$, i.e., each individual always produces at least one offspring. Assume also that $Q(i,\{x : |x|\geq 2\})>0$ for at least one $i\in S$. Then it is not hard to show that

$$\mathbb{P}^{\delta_i}[X_n(j)\geq N] \underset{n\to\infty}{\longrightarrow} 1 \qquad (i\in S,\ N<\infty).$$

Use this to show that for any $\phi\in[0,1]^S$ that is not identically zero,

$$U^n\phi(i) \underset{n\to\infty}{\longrightarrow} 1 \qquad (i\in S).$$

In particular, this shows that the equation $U\rho=\rho$ has only two solutions: $\rho=0$ and $\rho=1$.

Exercise 4.22 (Fixed points of generating operator) Let X be a branching process with finite type space S, offspring distribution Q, and generating operator U. Assume that Q is irreducible and aperiodic and that $Q(i,\{x : |x| \geq 2\}) > 0$ for at least one $i \in S$. Assume that the survival probability ρ in (4.22) is positive for some, and hence for all $i \in S$. Show that for any $\phi \in [0,1]^S$ that is not identically zero,

$$U^n\phi(i) \underset{n\to\infty}{\longrightarrow} \rho(i) \qquad (i \in S).$$

In particular, this shows that ρ is the only nonzero solution of the equation $U\rho = \rho$. Hint: use Proposition 4.19 to reduce the problem to the set-up of Excercise 4.21.

Exercise 4.23 (Exponential growth) In the set-up of Excercise 4.22, assume that Q has finite first moments. Let α be the Perron-Frobenis eigenvalue of the first moment matrix A defined in (4.12) and let $h > 0$ be the associated right eigenvector. Assume that $\alpha > 0$, let $M = (M_n)_{n\geq 0}$ be the martingale defined in (4.15), and let $M_\infty := \lim_{n\to\infty} M_n$. Set

$$\rho'(i) := \mathbb{P}^{\delta_i}\big[M_\infty > 0\big].$$

Prove that $\rho' = \rho$, where ρ is the survival probability defined in (4.22). Hint: show that $U\rho' = \rho'$.

Appendix A

Supplementary material

A.1 The spectral radius

For any complex matrix A indexed by a finite set S, we define

$$\rho(A) := \lim_{n\to\infty} \|A^n\|^{1/n},$$

where $\|\cdot\|$ denote the operator norm associated with some norm on $\mathbb{C}^S$. The arguments around (2.22) show that the limit exists and does not depend on the choice of the norm on $\mathbb{C}^S$. It is also shown there that in the special case that A is a nonnegative matrix, this definition coincides with the definition in (2.1). Let $\mathrm{spec}(A)$ denote the spectrum of A, i.e., the collection of all eigenvalues. The following lemma links $\rho(A)$ to $\mathrm{spec}(A)$. In particular, for nonnegative matrices, the Perron-Frobenius theorem tells us that $\rho(A) \in \mathrm{spec}(A)$ and the following lemma identifies $\rho(A)$ as the largest eigenvalue (in absolute value).

Lemma A.1 (Gelfand's formula) *One has*

$$\rho(A) = \sup\{|\lambda| : \lambda \in \mathrm{spec}(A)\}.$$

Proof of Lemma Suppose that $\lambda \in \mathrm{spec}(A)$ and let f be an associated eigenvector. Then $\|A^n f\| = |\lambda|^n \|f\|$ which shows that $\|A^n\|^{1/n} \geq |\lambda|$ and hence $\rho(A) \geq |\lambda|$.

To complete the proof, it suffices to show that $\rho(A) \leq \lambda_+$, where $\lambda_+ := \sup\{|\lambda| : \lambda \in \mathrm{spec}(A)\}$. We start with the case that A can be diagonalized, i.e., there exists a basis $\{e_1, \ldots, e_d\}$ of eigenvectors with associated eigenvalues $\lambda_1, \ldots, \lambda_d$.

By Excercise 2.13 the choice of our norm on V is irrelevant. We choose the ℓ_1-norm with respect to the basis $\{e_1, \ldots, e_d\}$, i.e.,

$$\|\phi\| := \sum_{i=1} |\phi(i)|,$$

where $\phi(1), \ldots, \phi(d)$ are the coordinates of ϕ w.r.t. this basis. Then

$$\|A^n\phi\| = \| \sum_{i=1}^{d} \phi(i)\lambda_i^n e_i\| = \sum_{i=1}^{d} |\phi(i)|\, |\lambda_i|^n \leq \lambda_+^n \|\phi\|,$$

which proves that $\|A^n\| \leq \lambda_+^n$ for each $n \geq 1$ and hence $\rho(A) \leq \lambda_+$.

In general, A need not be diagonalizable, but we can choose a basis such that the matrix of A w.r.t. this basis has a Jordan normal form. Then we may write $A = D + E$ where D is the diagonal part of A and E has ones only on some places just above the diagonal and zeroes elsewhere. One can check that E is *nilpotent*, i.e., $E^m = 0$ for some $m \geq 1$. Moreover E commutes with D and $\|E\| \leq 1$ if we choose the ℓ_1-norm with respect to the basis $\{e_1, \ldots, e_d\}$. Now

$$A^n = (D + E)^n = \sum_{k=0}^{m-1} \binom{n}{k} D^{n-k} E^k$$

and therefore

$$\|A^n\| \leq \sum_{k=0}^{m-1} \binom{n}{k} \|D^{n-k}\| \leq \sum_{k=0}^{m-1} \binom{n}{k} \lambda_+^{n-k} = \lambda_+^n \sum_{k=0}^{m-1} \binom{n}{k} \lambda_+^{-k} =: q(n)\lambda_+^n,$$

where

$$q(n) = 1 + n\lambda_+^{-1} + \tfrac{1}{2}n(n-1)\lambda_+^{-2} + \cdots$$

is a polynomial in n of degree m. In particular, this shows that $\|A^n\| \ll (\lambda_+ + \varepsilon)^n$ as $n \to \infty$ for all $\varepsilon > 0$, which again yields that $\rho(A) \leq \lambda_+$. ∎

English-Czech glossary

stochastic process	nahodný/stochastický proces
state space	stavový prostor
Borel sigma-field	borelovská sigma-algebra
trajectory	trajektorie
distribution function	distribuční funkce
mean	střední hodnota
standard deviation	rozptyl
continuous in probability	stochasticky spojitý
	/ spojitý podle pravděpodobnosti
complement	doplněk
white noise	bílý šum
random walk on the real line	náhodná procházka na přímce
Galton-Watson branching process	Galton-Watsonův proces větvení
Poisson process	Poisonův proces
(Wiener process) Brownian motion	Wienerův proces/Brownův pohyb
homogenoeus Markov process	homogenní Markovový proces
continuous-time Markov chain	Markovový řetězec se spojitým časem
Markov chain with discrete time	Markovový řetězec s diskrétním časem
transition probabilities	pravděpodobnosti přechodů
(n-step)	(n-tého řádu)
probability kernel	pravděpodobnostní jádro
stochastic matrix	stochastická matice
square matrix	čtverecová matice
initial distribution	počáteční rozdělení
(conditional) probability	podmíněná/absolutní pravděpodobnost
integer valued random variable	celočíselná náhodná veličina
event	jev
stopping time	Markovský čas, zastavovací čas
strong markov property	silná markovská vlastnost
first hitting time	čas prvního nástupu, vstupu do stavu j
recurrent / transient state	tvrvalý/přechodný stav
	rekurentní/transientní
null/positive recurrent state	nulový/nenulový trvalý stav
(a)periodic state	(ne)periodický stav
reachable (from state i)	dosažitelný stav (ze stavu i)
closed set	uzavřená množina
closure	uzávěr
irreducible	(ne)rozložitelná/irreducibilní
proper subset	vlastní podmnožina

trap	absorpční stav
singleton	jednoprvková množina
first exit time	čas výstupu
assume values	nabývat hodnot
invariant law	stacionární rozdělení
invariant measure	invariantní míra
reversible	vratný
transition rates	intenzity přechodů
Q-matrix, generator	matice intenzit (přechodů)
process with independent increments	proces s nezávislými přírůstky
birth-and-death process	proces množení a zániku
binary branching process	lineární proces množení a zániku
step function	schodovitá funkce
explosion time	čas exploze
nonexplosive chain	regulární řetězec
embedded Markov chain	vnořený diskrétní řetězec skoků
Kolmogorov differential equation	Kolmogorovovy diferenciální rovnice
backwards/forwards	retrospektivní a prospektivní
differenticiable	diferencovatelný
queueing systems	systémy hromadné obsluhy

Bibliography

[AS10] S.R. Athreya and J.M. Swart. Survival of contact processes on the hierarchical group. *Probab. Theory Relat. Fields* 147(3), 529–563, 2010.

[Bil86] P. Billingsley. *Probability and Measure.* Wiley, New York, 1986.

[Chu74] K.L. Chung. *A Course in Probability Theory, 2nd ed.* Academic Press, Orlando, 1974.

[DF90] Strong stationary times via a new form of duality. P. Diaconis and J. Fill. *Ann. Probab.* 18(4), 1483–1522, 1990.

[DM09] P. Diaconis and L. Miclo. On times to quasi-stationarity for birth and death processes. *J. Theor. Probab.* 22, 558–586, 2009.

[DS67] J.N. Darroch and E. Seneta. On quasi-stationary distributions in absorbing continuous-time finite Markov chains. *J. Appl. Probab.* 4, 192–196, 1967.

[DZ98] A. Dembo and O. Zeitouni. *Large deviations techniques and applications 2nd edition.* Applications of Mathematics 38. Springer, New York, 1998.

[EKW15] M. Eckhoff, A.E. Kyprianou, and M. Winkel. Spines, skeletons and the Strong Law of Large Numbers for superdiffusions. *Ann. Probab.* 43(5) (2015), 2545–2610.

[Fil92] J.A. Fill. Strong stationary duality for continuous-time Markov chains. I. Theory. *J. Theor. Probab.* 5(1), 45–70, 1992.

[Fro12] G. Frobenius. Über Matrizen aus nicht negativen Elementen. *Sitzungsber. Königl. Preuss. Akad. Wiss.* (1912), 456–477.

[FS04] K. Fleischmann and J.M. Swart. Trimmed trees and embedded particle systems. *Ann. Probab.* 32(3A), 2179-2221, 2004.

[Gan00] F.R. Gantmacher. *The Theory of Matrices, Vol. 2.* AMS, Providence RI, 2000.

[Geo88] H.-O. Georgii. *Gibbs Measures and Phase Transitions.* De Gruyter, Berlin, 1988.

[Ken59] D.G. Kendall. Unitary dilations of Markov transition operators, and the corresponding integral representations for transition-probability matrices. Pages 139–161 in: U. Grenander (ed), *Probability and Statistics. The Harald Cramér volume.* Almqvist & Wiksell, Stockholm, 1959.

[Kin63] J.F.C. Kingman. The exponential decay of Markov transition probabilities. *Proc. London Math. Soc. (3)* 13 (1963), 337–358.

[KR48] M.G. Kreĭn and M.A. Rutman. Linear operators leaving invariant a cone in a Banach space [in Russian]. *Uspehi Matem. Nauk* 3 (1948), No. 1(23), 3–95.

[Lach12] P. Lachout. Disktrétní martingály. Skripta MFF UK, 2012.

[Lig10] T.M. Liggett. *Continuous Time Markov Processes.* AMS, Providence, 2010.

[Lig99] T.M. Liggett. *Stochastic interacting systems: contact, voter and exclusion processes.* Springer-Verlag, Berlin, 1999.

[LL10] G.F. Lawler and V. Limic. *Random Walk: A modern introduction.* Cambridge Studies in Advances Mathematics 123. Cambridge University Press, Cambridge, 2010.

[LP11] P. Lachout and Z. Prášková. Základy náhodných procesů I. Skripta MFF UK, 2011.

[LPW09] D.A. Levin, Y. Peres and E.L. Wilmer. *Markov Chains and Mixing Times.* AMS, Providence, 2009.

[Mar06] A.A. Markov. Extension of the law of large numbers to dependent quantities [in Russian]. *Izv. Fiz.-Matem. Obsch. Kazan Univ.* (2nd Ser.) 15 (1906), 135–156.

[Per07] O. Perron. Zur Theorie der Matrices. *Mathematische Annalen* 64(2) (1907), 248–263.

[RP81] L.C.G. Rogers and J.W. Pitman. Markov functions. *Ann. Probab.* 9(4), 573–582, 1981.

[Sen73] E. Seneta. *Non-Negative Matrices: An Introduction to Theory and Applications.* George Allen & Unwin, London, 1973.

[Sen81] E. Seneta. *Non-negative Matrices and Markov Chains. 2nd ed.* Springer, New York, 1981.

[Swa11] J.M. Swart. Intertwining of birth-and-death processes. *Kybernetika* 47(1), 1–14, 2011.

[Swa17] J.M. Swart. Necessary and sufficient conditions for a nonnegative matrix to be strongly R-positive. preprint (2017), ArXiv:1709.09459.

[Ver62] D. Vere-Jones. Geometric ergodicity in denumerable Markov chains. *Quart. J. Math. Oxford Ser. 2* 13 (1962), 7–28.

[Ver67] D. Vere-Jones. Ergodic properties of non-negative matrices - I. *Pacific J. Math.* 22(2) (1967), 361–386.

[WG75] H.W. Watson and F. Galton. On the Probability of the Extinction of Families. *Journal of the Anthropological Institute of Great Britain* 4, 138–144, 1875.

[Woe00] W. Woess. *Random walks on infinite graphs and groups.* Cambridge Tracts in Mathematics 138. Cambridge University Press, Cambridge, 2000.

[Woe09] W. Woess. *Denumerable Markov chains. Generating functions, boundary theory, random walks on trees.* EMS Textbooks in Mathematics. EMS, Zürich, 2009.

Index