Markov chains

J.M. Swart

February 20, 2012

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^{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
$\mu \ll \nu$	μ is absolutely continuous w.r.t. ν
$\|f\|_\infty$	supremumnorm $\|f\|_\infty := \sup_x \|f(x)\|$
$f_k \sim g_k$	$\lim f_k / g_k = 1$

Preface

To be filled in.

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

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

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

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 $(\mathcal{F}_k^X)_{k\in I}$ 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 $\mathcal{F}_k^X \subset \mathcal{F}_k$ 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 $\mathcal{F}_k^X$-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

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

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 $k \geq 0$.

Note that in particular, if M is a martingale, then Proposition 0.8 says that $M_k = \mathbb{E}[M_\infty|\mathcal{F}_k]$, which shows that all information about the martingale 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 $n \in I$, 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.*

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

(i) $\mathbb{P}\big[(X_k)_{k\in I_n^-} \in A,\ (X_k)_{k\in I_n^+} \in B \,\big|\, X_n\big]$
$= \mathbb{P}\big[(X_k)_{k\in I_n^-} \in A \,\big|\, X_n\big]\mathbb{P}\big[(X_k)_{k\in I_n^+} \in B \,\big|\, X_n\big]$ *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

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

where 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 from the definition of conditional expectations that

$$\mathbb{P}[B \,|\, \mathcal{F}_n^X] = \mathbb{P}[B \,|X_n],$$

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

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

Since this holds for any $C \in \sigma(X_n)$, it follows from the definition of conditional expectations that

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

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*

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

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

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 probability space $(E, \mathcal{E}, \mu)$ together with a measurable function $f : S \times E \to S$ such that if Z_1 is a random variable taking values in $(E, \mathcal{E})$ with law μ, then

$$P(x, y) = \mathbb{P}[f(x, Z_1) = y] \qquad (x, y \in S).$$

If $(Z_k)_{k\geq 1}$ are i.i.d. random variables with values in $(E, \mathcal{E})$ and common law μ and X_0 is a random variable taking values in S, independent of the $(Z_k)_{k\geq 1}$, then setting, inductively,

$$X_k := f(X_{k-1}, Z_k) \qquad (k \geq 1)$$

defines a Markov chain 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.) One can show that each probability kernel has a random mapping representation, but such a representation is far from unique. Often, the key to a good proof is choosing the right one.

We note that it is in general not true that functions of Markov chains are themselves Markov chains. An exception is the case when

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

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'.

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 te 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 $\mathcal{A}_m := \{(X_{n+k})_{k\geq 0} \in B$ 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 $\mathcal{A}_m \downarrow \{(X_{n+k})_{k\geq 0} \in B$ for infinitely many $n \geq 0\}$, the claim follows. ∎

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 $x \rightsquigarrow z$. Two states x, y are called *equivalent* if $x \rightsquigarrow y$ and $y \rightsquigarrow x$. It is easy to see that this defines an equivalence relation on S. A Markov chain / transition kernel is called *irreducible* if all states are equivalent.

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 are equivalent 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 equivalence 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 equivalence 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 equivalence 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])

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

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).

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

Theorem 0.18 (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 equivalence 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 $1/\infty := 0$. Since consecutive return times are i.i.d., it is not hard to prove that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{P}^x[X_k = x] = \mu(x), \tag{0.7}$$

i.e., the process started in x spends a $\mu(x)$-fraction of its time in x. As a result, it is not hard to show that if x is transient or null-recurrent, then the process started in any initial law satisfies $\mathbb{P}[X_n = x] \to 0$ for $n \to \infty$, hence no invariant law can give positive probability to such a state.

On the other hand, if $S' \subset S$ is an equivalence class of positively recurrent states, then one can check that (0.7) holds more generally for the process started in any initial law on S'. It follows that for any such process, the Césaro-limit

$$\mu = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{P}[X_k \in \cdot\,]$$

exists and does not depend on the initial law. In particular, if we start in an invariant law, then this limit must be μ, which proves uniqueness. It is not hard to check that any such Césaro-limit must be an invariant law, from which we obtain existence. ∎

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 equivalence class of positively recurrent states.

Theorem 0.19 (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]. ∎

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) meaure 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

Eigenfunctions

1.1 Harmonic functions

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)),

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

∎

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

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

∎

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 a 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. In particular, this applies when S is finite.

Remark 2 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 1.7 below.

Remark 3 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 transformation is usually called an *h-transform* or *Doob's h-transform*. Following [LPW09], we will simply call it a *Doob transform*.[2]

Lemma 1.4 (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. ∎

[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.

Proposition 1.5 (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 $x \in S$ be a trap. Set $S' := \{y \in S : y \rightsquigarrow x\}$ and assume that $\mathbb{P}[X_0 \in S'] > 0$. Then, under the conditional law*

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

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

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

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

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

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 $\mathcal{T} := \bigcap_{k\geq 0} \mathcal{F}_k^X$. 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 1.6 (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 1.7 (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 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 $h(x) := \mathbb{E}[\phi(X_\tau)]$, where $\tau := \inf\{k \geq 0 : X_k \in T\}$ is the trapping time.

Exercise 1.8 (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 and use the principle 'what can happen must happen' (Proposition 0.14) to show that the limit cannot take values in $(0,1)$.

Exercise 1.9 (Trapping estimate) Let $X, S, , P$ and h be as in Excercise 1.8. Assume that h is a submartingale and there is a point $z \in S$ such that $h(z) = 1$ and $\sup_{x \in S\backslash\{z\}} h(x) < 1$. Show that

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

Exercise 1.10 (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 1.11 (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 $\mathbb{E}^x[\tau < \infty] < \infty$ for all $x \in S$. Show that the function

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

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

Exercise 1.12 (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

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

Hint: show that the compensator K of $(f(X_k))_{k \geq 0}$ satisfies $K_n \leq -(n \wedge \tau)$.

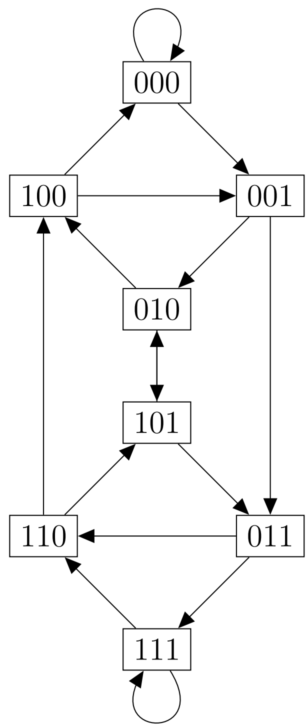
Exercise 1.13 (Absorption of random walk) Consider a random walk $W = (W_k)_{k \geq 0}$ on $\mathbb{Z}$ that jumps from x to $x + 1$ with probability p and to $x - 1$ with the remaining probability $q := 1 - p$, where $0 < p < 1$. Fix $n \geq 1$ and set $\tau := \inf\big\{k \geq 0 : W_k \in \{0, n\}\big\}$. Calculate, for each $0 \leq x \leq n$, the probability $\mathbb{P}[W_\tau = n]$.

Exercise 1.14 (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 1.15 (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 0.18 (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.

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,

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

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. 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 $\emptyset$.

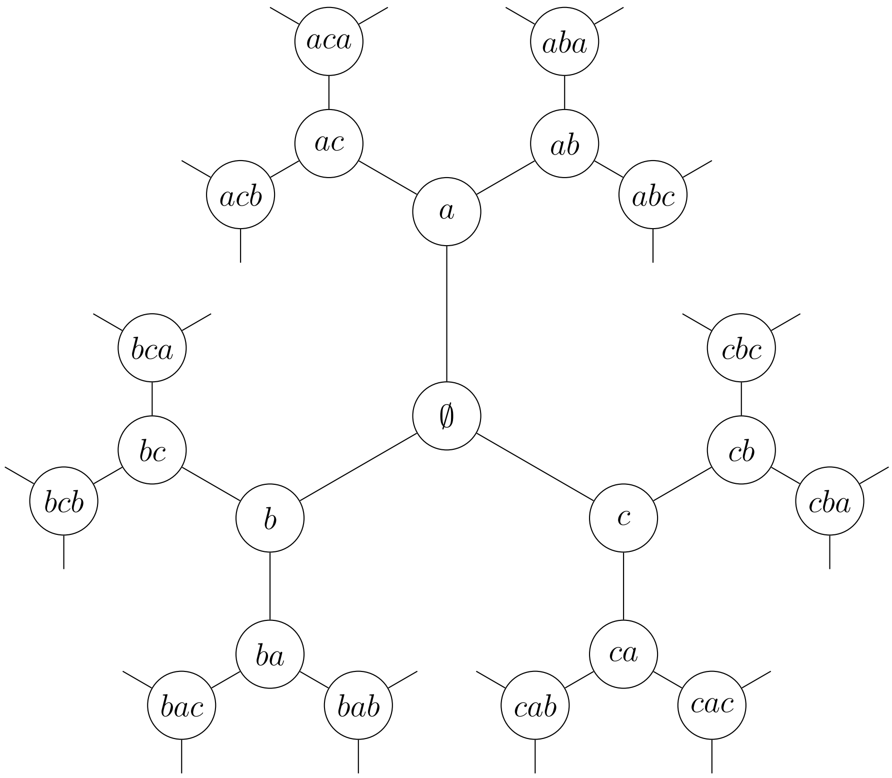


Figure 1.1: The regular tree $\mathbb{T}_2$

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 a 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 1.16 (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 1.17 (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 1.16. 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,

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

∎

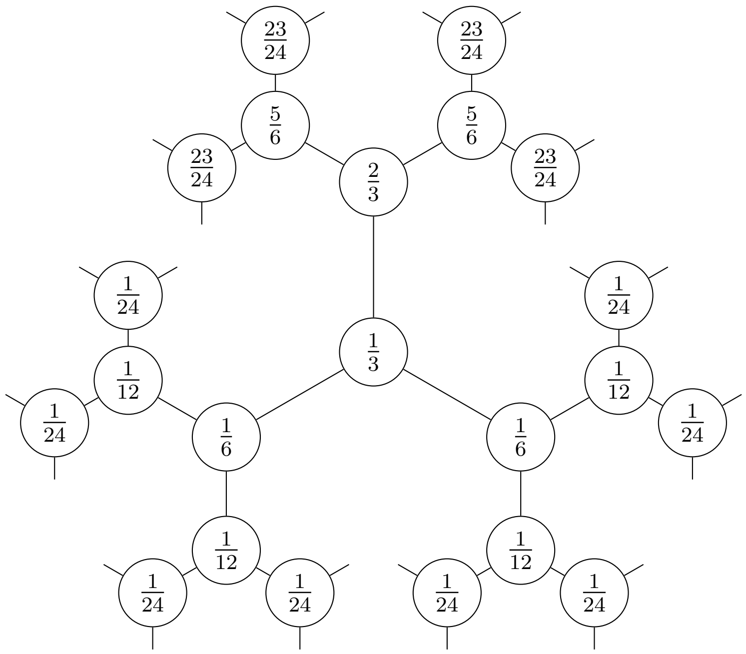


Figure 1.2: A bounded harmonic function

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 1.17 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 1.18 (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.

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

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

$(x, y \in G,\ n \geq 0,\ A \in \mathcal{B}(\partial G))$ 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 1.17

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

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

Exercise 1.19 (Escape probability) Let $Z = (Z_k)_{k\geq 0}$ 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 1.20 (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 $|x|_1 := \max_{i=1}^d |x(i)|$ 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 1.20, 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 *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 1.21 (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 aperiodic.[3] 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, $P_{\text{lazy}} f = \frac{1}{2} Pf + \frac{1}{2} f$, so a function h is harmonic for P if and only if it is harmonic for P_{lazy}.

Proposition 1.22 (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_{lazy}, and initial states $X_0^x = x$ and $X_0^y = y$, $x, y \in \mathbb{Z}^d$. Then X^x and X^y can be coupled such that*

$$\exists n \geq 0 \text{ s.t. } X_k^x = X_k^y \quad \forall k \geq n \quad \text{a.s.}$$

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 $X_0^x := x$ and

$$X_k^x = X_{k-1}^x + U_k W_k e_{I_k} \qquad (k \geq 1).$$

[3] 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 $\mathbb{Z}^d_{\text{odd}} := \{x \in \mathbb{Z}^d : \sum_{i=1}^d x(i) \text{ is odd}\}$.

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_{lazy}. It is straightforward to check that $\mathbb{P}[X^y_k = z \,|\, (X^x_k, X^y_k)] = P_{\text{lazy}}(X^y_k, z)$ 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_{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 1.22, we obtain that all bounded harmonic functions for nearest-neighbor random walk on the d-dimensional integer lattice are constant.

Corollary 1.23 (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 1.22. Since h is harmonic and bounded, $(h(X_k^x))_{k\geq 0}$ and $(h(X_k^y))_{k\geq 0}$ are martingales. It follows that

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

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

Bibliography

[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.

[Lach12] P. Lachout. Disktrétní martingály. Skripta MFF UK, 2012.

[Lig10] T.M. Liggett. *Continuous Time Markov Processes.* AMS, Providence, 2010.

[LP11] P. Lachout and Z. Prášková. Základy náhodných procesù I. Skripta MFF UK, 2011.

[LPW09] D.A. Levin, Y. Peres and E.L. Wilmer. *Markov Chains and Mixing Times.* AMS, Providence, 2009.

Index