Introduction to Markov processes

Justin Salez

December 4, 2025

Contents

Chapter 1

Discrete-time Markov chains

1.1 Definition and basic terminology

In full generality, a stochastic process is a collection $\mathbf{X} = (X_t)_{t\in\mathbb{T}}$ of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, taking values in a measurable space $(\mathbb{X}, \mathcal{A})$ and indexed by a set $\mathbb{T}$. We will here think of $\mathbb{X}$ as *space* and $\mathbb{T}$ as *time*, so that $X = (X_t)_{t\in\mathbb{T}}$ describes the space-time evolution of a random system. With this interpretation in mind, the choices $\mathbb{X} = \mathbb{R}^d$ and $\mathbb{T} = [0, \infty)$ seem of course particularly natural. However, in the present chapter, we will restrict ourselves to the conceptually simpler discrete setup, where

- the time set is the set of natural integers $\mathbb{T} = \mathbb{N} := \{0, 1, 2, \ldots\}$;
- the state space $\mathbb{X}$ is countable and equipped with its complete sigma-algebra $\mathcal{A} = 2^{\mathbb{X}}$.

On the countable state space $\mathbb{X}$, we henceforth assume to be given a matrix $T \in \mathbb{R}^{\mathbb{X}\times\mathbb{X}}$ which is stochastic, meaning that $T(x, y) \geq 0$ for all $x, y \in \mathbb{X}$ and

$$\forall x \in \mathbb{X}, \qquad \sum_{y\in\mathbb{X}} T(x, y) \quad = \quad 1.$$

We call T the transition matrix and think of it as describing the evolution of a random system which jumps from its current location $x \in \mathbb{X}$ to a new state $y \in \mathbb{X}$ with probability $T(x, y)$, independently of the past. If we start from a random state X_0 and iterate this simple stochastic rule, we obtain a discrete-time stochastic process $(X_n)_{n\in\mathbb{N}}$ whose finite-dimensional marginals admit a particularly simple product form.

Definition 1.1 (Markov Chain). *A discrete-time Markov chain with transition matrix T is a $\mathbb{X}-$valued stochastic process $(X_n)_{n\in\mathbb{N}}$ such that for each $n \in \mathbb{N}$ and each $(x_0, \ldots, x_n) \in \mathbb{X}^{n+1}$,*

$$\mathbb{P}(X_0 = x_0, \ldots, X_n = x_n) \quad = \quad \mathbb{P}(X_0 = x_0)T(x_0, x_1) \cdots T(x_{n-1}, x_n). \tag{1.1}$$

Remark 1.1 (Law). *A general stochastic process $\mathbf{X} = (X_t)_{t\in\mathbb{T}}$ can always be viewed as a* single *random variable taking values in the configuration space $\mathbb{X}^{\mathbb{T}}$, equipped with the product $\sigma-$field. By the monotone class lemma, the law of $\mathbf{X}$ is then completely determined by its finite-dimensional marginals*

$$\mathbb{P}\left(X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n\right), \quad n \in \mathbb{N}, \quad (t_1, \ldots, t_n) \in \mathbb{T}^n, \quad (A_1, \ldots, A_n) \in \mathcal{A}^n. \tag{1.2}$$

In particular, the law of a discrete-time Markov chain $\mathbf{X} = (X_n)_{n\in\mathbb{N}}$ is completely determined by its initial law $\mu_0 := law(X_0)$ and its transition matrix T, and we will simply write $\mathbf{X} \sim \mathrm{DTMC}(\mu_0, T)$.

The product form in (1.1) reflects the *lack of memory* inherent to Markov chains: only their current state can influence where they go next. This fundamental principle is known as the Markov property, and it admits several alternative formulations. For example, Definition 1.1 is clearly equivalent to the requirement that

$$\mathbb{P}\left(X_{n+1} = x_{n+1} \,|\, X_0 = x_0, \dots, X_n = x_n\right) \quad = \quad T(x_n, x_{n+1}),$$

for each $n \in \mathbb{N}$ and each $(x_0, \dots, x_{n+1}) \in \mathbb{X}^{n+2}$ such that $\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) > 0$. By linearity of expectations, this is in turns equivalent to the condition

$$\mathbb{E}\left[f(X_{n+1}) \,|\, X_0 = x_0, \dots, X_n = x_n\right] \quad = \quad Tf(x_n), \tag{1.3}$$

for each bounded observable $f\colon \mathbb{X} \to \mathbb{R}$, each $n \in \mathbb{N}$ and each $(x_0, \dots, x_n) \in \mathbb{X}^{n+1}$ such that $\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) > 0$, where we have used the standard notation

$$\forall x \in \mathbb{X}, \qquad Tf(x) \quad := \quad \sum_{y \in \mathbb{X}} T(x,y) f(y). \tag{1.4}$$

Finally, because it holds for all possible realization $(x_0, \dots, x_n)$ of $(X_0, \dots, X_n) > 0$, the identity (1.3) can be conveniently re-written as an almost-sure equality between random variables:

$$\mathbb{E}\left[f(X_{n+1}) \,|\, X_0, \dots, X_n\right] \quad = \quad \mathbb{E}\left[f(X_{n+1}) \,|\, X_n\right] \quad = \quad Tf(X_n). \tag{1.5}$$

We will come back to this identity later. When studying a particular Markov chain, it is useful to draw (or at least have in mind) its diagram, i.e. the directed graph $G = (\mathbb{X}, E)$ whose vertex set is the state space $\mathbb{X}$, and whose edges represent the possible transitions:

$$E \quad := \quad \left\{(x,y) \in \mathbb{X}^2 \colon T(x,y) > 0\right\}. \tag{1.6}$$

As any directed graph, G can be partitioned into strongly connected components called communication classes, which are the equivalence classes induced by the equivalence relation

$$x \leftrightarrow y \quad \Longleftrightarrow \quad \begin{cases} \exists k \in \mathbb{N} & T^k(x,y) > 0 \\ \exists \ell \in \mathbb{N} & T^\ell(y,x) > 0. \end{cases}$$

In this formula, $T^k(x,y)$ represents the (x,y) entry of the $k-$th power of the matrix T, i.e.

$$T^k(x,y) \quad = \quad \sum_{x_1 \in \mathbb{X}} \dots \sum_{x_{k-1} \in \mathbb{X}} T(x,x_1) \cdots T(x_{k-1},y), \tag{1.7}$$

which is non-zero if and only if the diagram G contains a path of length exactly k from x to y. A communication class C is called closed if there is no edge exiting C, i.e. $(C \times C^c) \cap E = \emptyset$, and open otherwise. The restriction of T to a closed class is obviously a stochastic matrix on C, which can be studied on its own. In the important case where G is strongly connected (i.e., when there is only one communication class), we say that T is irreducible.

1.2 Classical examples

Example 1.1 (Stochastic recursions). *Let $(\xi_n)_{n\geq 1}$ be i.i.d. random variables taking values in a measurable space $(E, \mathcal{E})$, and let $F\colon \mathbb{X} \times E \to \mathbb{X}$ be a measurable function. Finally, let X_0 be a $\mathbb{X}-$valued random variable independent of $(\xi_n)_{n\geq 1}$ and define $\mathbf{X} = (X_n)_{n\in\mathbb{N}}$ inductively via the recursion*

$$\forall n \in \mathbb{N}, \qquad X_{n+1} \quad := \quad F(X_n, \xi_n). \tag{1.8}$$

Prove that $\mathbf{X}$ is a Markov chain, and explicitate its transition matrix.

Example 1.2 (Rank-one chains). *Consider a sequence* $\mathbf{X} := (X_n)_{n\in\mathbb{N}}$ *of i.i.d.* $\mathbb{X}-$*valued random variables with law* $\mu \in \mathcal{P}(\mathbb{X})$. *Show that* $\mathbf{X}$ *is a Markov chain.*

Example 1.3 (Random walks on graphs). *Let* $G = (\mathbb{X}, E)$ *be a directed graph in which the out-degree*

$$\deg(x) \quad := \quad \sum_{y\in\mathbb{X}} \mathbf{1}_{(x,y)\in E}, \tag{1.9}$$

of every vertex $x \in \mathbb{X}$ *is positive and finite. A random walk on* G *is a Markov chain on* $\mathbb{X}$ *which moves by traversing an edge chosen uniformly at random among those that start from its current location:*

$$T(x,y) \quad := \quad \begin{cases} \frac{1}{\deg(x)} & \text{if } (x,y) \in E \\ 0 & \text{else.} \end{cases}$$

Two emblematic examples are the random walk on the $d-$*dimensional lattice* ($\mathbb{X} = \mathbb{Z}^d$, $E = \{(x,y) \in \mathbb{X}^2 \colon \|x-y\|_2 = 1\}$), *and the Page Rank Algorithm* ($\mathbb{X}$ *is the set of webpages, and* E *the set of hyperlinks).*

Example 1.4 (Random walks on groups). *Let* $\mu \in \mathcal{P}(\mathbb{X})$ *be a probability measure on a countable group* $\mathbb{X}$, *and let* $(Z_n)_{n\geq 1}$ *be i.i.d. samples from* μ. *Consider the process* $\mathbf{X} = (X_n)_{n\in\mathbb{N}}$ *defined by*

$$\forall n \in \mathbb{N}, \qquad X_n \quad := \quad Z_n \cdots Z_2 Z_1, \tag{1.10}$$

with the usual convention that an empty product is the identity element. Show that $\mathbf{X}$ *is a Markov chain and specify its transition matrix. When is it irreducible?*

Example 1.5 (Branching processes). *Imagine a population in which each individual gives birth to a random number of children distributed according to some prescribed offspring law* $\mu \in \mathcal{P}(\mathbb{N})$, *independently of everything else. For each* $n \in \mathbb{N}$, *let* X_n *denote the number of individuals at generation* n. *Prove that* $\mathbf{X} = (X_n)_{n\in\mathbb{N}}$ *is a Markov chain, and specify its transition matrix.*

1.3 The strong Markov property

We henceforth assume that our probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is equipped with a filtration $(\mathcal{F}_n)_{n\geq 0}$, i.e., a nested sequence of $\sigma-$fields

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots \subseteq \mathcal{F}, \tag{1.11}$$

with the idea that $\mathcal{F}_n$ represents the *information* that is available to us at time n. We can then refine our definition of a Markov chain by taking the filtration into consideration.

Definition 1.2 (Markov chain on a filtered space). *A Markov chain with transition matrix* T *is a stochastic process* $\mathbf{X} := (X_n)_{n\in\mathbb{N}}$ *satisfying the following two properties:*

1. $\mathbf{X}$ *is adapted to the filtration, i.e. for each* $n \in \mathbb{N}$, X_n *is* $\mathcal{F}_n-$*measurable;*
2. *for each* $n \in \mathbb{N}$ *and each bounded function* $f \colon \mathbb{X} \to \mathbb{R}$, $\mathbb{E}[f(X_{n+1}) \,|\, \mathcal{F}_n] = Tf(X_n)$.

We will speak of a $(\mathcal{F}_n)_{n\geq 0}-$Markov chain when we want to explicitate the underlying filtration. It is important to realize that this definition refines the previous one. More precisely, any Markov chain in the sense of Definition 1.1 is by (1.5) a $(\mathcal{F}_n)_{n\geq 0}-$Markov chain with respect to the natural filtration

$$\forall n \in \mathbb{N}, \qquad \mathcal{F}_n \quad := \quad \sigma(X_0, \ldots, X_n). \tag{1.12}$$

The latter is clearly the smallest filtration satisfying the first condition in Definition 1.2, and focusing on this concrete and intuitive choice will help the unfamiliar reader without too much loss in terms of generality. We emphasize, however, that working with a richer filtration is always stronger, because the second property implies (1.5) by the *tower property* of conditional expectation. Here is an important generic example where the filtration can be much richer.

Example 1.6 (Stochastic recursion). *Check that the stochastic recursion defined in Example 1.1 is in fact a $(\mathcal{F}_n)_{n\in\mathbb{N}}-$Markov chain for the choice $\mathcal{F}_n := \sigma(X_0, \xi_1, \dots, \xi_n)$.*

In order to warm up and get used to the formalism of filtrations, let us establish the following result, which can be seen as yet another version of the Markov property.

Lemma 1.1 (Markov property). *Consider a Markov chain $(X_n)_{n\in\mathbb{N}}$ with transition matrix T, and a time $m \in \mathbb{N}$. Then, the conditional law of the shifted process $(X_{n+m})_{n\geq 0}$ given $\mathcal{F}_m$ is* $\mathrm{DTMC}(\delta_{X_m}, T)$.

Proof. Fix an event $A \in \mathcal{F}_m$, an integer $n \in \mathbb{N}$, and a sequence $(y_0, \dots, y_n) \in \mathbb{X}^{n+1}$. Using the short-hand $A_n := A \cap \{Y_0 = y_0, \dots, Y_n = y_n\}$, we have

$$\mathbb{P}(A_n) \quad = \quad \mathbb{P}(A_{n-1} \cap \{Y_n = y_n\}) \quad = \quad \mathbb{P}(A_{n-1})T(y_{n-1}, y_n),$$

thanks to the second identity in Definition 1.2 with $f = \mathbf{1}_{\{y_n\}}$. Iterating this, we arrive at

$$\mathbb{P}(A_n) \quad = \quad \mathbb{P}(A_0)T(y_0, y_1)\cdots T(y_{n-1}, y_n).$$

Dividing through by $\mathbb{P}(A \cap \{Y_0 = y_0\})$, we deduce that

$$\mathbb{P}(Y_1 = y_1, \dots, Y_n = y_n \,|\, A \cap \{Y_0 = y_0\}) \quad = \quad T(y_0, y_1)\cdots T(y_{n-1}, y_n),$$

provided $\mathbb{P}(A \cap \{Y_0 = y_0\}) > 0$. The fact that this is true for all $n \in \mathbb{N}$ and $(y_0, \dots, y_n) \in \mathbb{X}^{n+1}$ shows that, under the probability measure $\mathbb{P}(\cdot|A \cap \{Y_0 = y_0\})$, the process $\mathbf{Y}$ has law $\mathrm{DTMC}(\delta_{y_0}, T)$. The claim is just a compact (but sophisticated) way of saying the same thing. □

Our goal now is to establish a powerful refinement of the above result, in which the deterministic time $m \in \mathbb{N}$ is replaced with an *appropriate* random time, in the following sense.

Definition 1.3 (Stopping time). *A stopping time is a function $\tau\colon \Omega \to \mathbb{N} \cup \{+\infty\}$ such that*

$$\forall n \in \mathbb{N}, \qquad \{\tau \leq n\} \quad \in \quad \mathcal{F}_n.$$

Intuitively, one should think of a stopping time τ as a random time with the property that for each $n \in \mathbb{N}$, the occurrence or non-occurrence of the event $\{\tau \leq n\}$ is completely determined by the information available at time n. For example, the first time that a random walk on $\mathbb{Z}^3$ lies at distance 10 from its initial position is a stopping time with respect to its natural filtration, but the last time that it does so is not: its occurrence can not be determined without *looking into the future*. The most important examples of stopping times are hitting times.

Example 1.7 (Hitting time). *Consider an adapted process $\mathbf{Z} := (Z_n)_{n\in\mathbb{N}}$ taking values in a measurable space $(E, \mathcal{E})$, and fix a set $A \in \mathcal{E}$. Prove that the hitting time of A by $\mathbf{Z}$, defined as*

$$\tau_A(\mathbf{Z}) \quad := \quad \inf\{n \geq 0\colon Z_n \in A\},$$

is a stopping time, with the usual convention that $\inf\emptyset = +\infty$.

We will need the following last, small – but quite subtle – definition, which turns out to be the right way to make sense of *the information available at time n*, when the time n is random.

Definition 1.4 (Past $\sigma-$field). *The Past $\sigma-$field of a stopping time τ is defined as*

$$\mathcal{F}_\tau \;:=\; \{A \in \mathcal{F} : \forall m \in \mathbb{N},\, A \cap \{\tau = m\} \in \mathcal{F}_m\}.$$

Exercise 1.1 (Basic properties). *Prove that $\mathcal{F}_\tau$ is a $\sigma-$field with the following properties:*

1. *τ is $\mathcal{F}_\tau-$measurable;*
2. *If $(X_n)_{n\in\mathbb{N}}$ is adapted and $\mathbb{P}(\tau < \infty) = 1$, then $X_\tau \colon \omega \mapsto X_{\tau(\omega)}(\omega)$ is $\mathcal{F}_\tau-$measurable;*
3. *If τ' is another stopping time satisfying $\tau \le \tau'$, then $\mathcal{F}_\tau \subseteq \mathcal{F}_{\tau'}$.*

We are now ready to state and prove our promised generalization of Lemma 1.1.

Theorem 1.1 (Strong Markov property). *Consider a Markov chain $(X_n)_{n\in\mathbb{N}}$ with transition matrix T and a stopping time τ. Then, on $\{\tau < \infty\}$, the conditional law of $(X_{\tau+n})_{n\in\mathbb{N}}$ given $\mathcal{F}_\tau$ is* DTMC(δ_{X_τ}, T).

Proof. Fix $A \in \mathcal{F}_\tau$, $m, n \in \mathbb{N}$, and $(x_0, \dots, x_n) \in \mathbb{X}^{n+1}$. Writing $A_m := A \cap \{\tau = m\}$, we have

$$\begin{aligned}
\mathbb{P}\left(A_m \cap \{X_\tau = x_0, \dots, X_{\tau+n} = x_n\}\right) &= \mathbb{P}\left(A_m \cap \{X_m = x_0, \dots, X_{m+n} = x_n\}\right) \\
&= \mathbb{P}\left(A_m \cap \{X_m = x_0\}\right) T(x_0, x_1) \cdots T(x_{n-1}, x_n) \\
&= \mathbb{P}\left(A_m \cap \{X_\tau = x_0\}\right) T(x_0, x_1) \cdots T(x_{n-1}, x_n),
\end{aligned}$$

where the second line crucially uses the weak Markov property at time m (Lemma 1.1) and the fact that $A_m \in \mathcal{F}_m$. Summing over all $m \in \mathbb{N}$, and writing $A' := A \cap \{\tau < \infty\}$, we obtain

$$\mathbb{P}\left(A' \cap \{X_\tau = x_0, \dots, X_{\tau+n} = x_n\}\right) \;=\; \mathbb{P}\left(A' \cap \{X_\tau = x_0\}\right) T(x_0, x_1) \cdots T(x_{n-1}, x_n).$$

Since this is true for all $n \in \mathbb{N}$ and all $(x_1, \dots, x_n) \in \mathbb{X}^n$, we conclude that the process $(X_{\tau+n})_{n\in\mathbb{N}}$ has law DTMC(δ_{x_0}, T) under the probability measure $\mathbb{P}(\cdot \mid A \cap \{\tau < \infty, X_\tau = x_0\})$, whenever the latter makes sense. The claim is just a sophisticated way of saying the same thing. □

Remark 1.2 (Deterministic stopping state). *In the special case where the stopping time τ satisfies $\tau < \infty \implies X_\tau = x$ for some fixed state $x \in \mathbb{X}$, the above result asserts that, under the measure $\mathbb{P}(\cdot \mid \tau < \infty)$, the process $(X_{n+\tau})_{n\in\mathbb{N}}$ has law* DTMC(δ_x, T) *and is independent of $\mathcal{F}_\tau$.*

1.4 Recurrence and transience

In this section, we consider a Markov chain $\mathbf{X} = (X_n)_{n\ge 0}$ with transition matrix T. Assuming without loss of generality that X_0 is fully supported (meaning that $\mathbb{P}(X_0 = x) > 0$ for every $x \in \mathbb{X}$), we can introduce, for every initial distribution $\mu \in \mathcal{P}(\mathbb{X})$, the convenient notation

$$\mathbb{P}_\mu(\cdot) \;:=\; \sum_{x\in\mathbb{X}} \mu(x)\mathbb{P}(\cdot \mid X_0 = x). \tag{1.13}$$

Note that, under this probability measure, we have by construction $\mathbf{X} \sim$ DTMC(μ, T). In the extremal case where μ is a Dirac mass at some state $x \in \mathbb{X}$, we use the simpler notation

$$\mathbb{P}_x(\cdot) \;:=\; \mathbb{P}(\cdot \mid X_0 = x). \tag{1.14}$$

Under the latter, we are interested in understanding the statistics of a basic but fundamental observable, namely, the (possibly infinite) number of visits to x, known as the local time at x:

$$L_x \;:=\; \sum_{n=0}^{\infty} \mathbf{1}_{\{X_n = x\}}. \tag{1.15}$$

An important related quantity is the (possibly infinite as well) return time to x, defined as

$$\tau_x^+ \quad := \quad \inf\{n \geq 1 \colon X_n = x\}. \tag{1.16}$$

We sometimes write $L_x(X), \tau_x^+(X)$ to explicitate the process X under consideration.

Lemma 1.2 (Recurrence/transience dichotomy). *Fix $x \in \mathbb{X}$, and set $p(x) := \mathbb{P}_x(\tau_x^+ < \infty)$.*

- *If $p(x) = 1$, then the state x is recurrent, i.e. it is almost-surely visited infinitely often:*

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

- *If $p(x) < 1$, then the state x is transient, i.e. it is visited finitely many times almost-surely. In that case, the total number of visits to x has a Geometric distribution with parameter $p(x)$:*

$$\forall n \in \mathbb{N}, \qquad \mathbb{P}_x(L_x = n+1) \quad = \quad (1-p(x))\,(p(x))^n\,.$$

Proof. We assume that $p(x) > 0$, otherwise the claim is trivial. By definition, we have

$$L_x \quad = \quad \mathbf{1}_{(X_0 = x)} + \mathbf{1}_{(\tau_x^+ < \infty)} L_x(\widetilde{X}), \tag{1.17}$$

where $\widetilde{X} = (X_{n+\tau_x^+})_{n \geq 0}$. In particular, for any $k \geq 1$, we can write

$$\begin{aligned} \mathbb{P}_x\left(L_x = k+1\right) \quad &= \quad \mathbb{P}_x\left(\tau_x^+ < \infty, L_x(\widetilde{X}) = k\right) \\ = \; p(x)\mathbb{P}_x\left(L_x = k\right), \end{aligned}$$

by the strong Markov property and Remark 1.2. If $p(x) = 1$, then the sequence $k \mapsto \mathbb{P}_x\left(L_x = k\right)$ is constant on $\{1, 2, \ldots\}$, and since its total sum is $\mathbb{P}(L_x < \infty)$, we must have $\mathbb{P}_x(L_x < \infty) = 0$. On the other hand, if $p(x) = 1$, then the above recursion shows that $\mathbb{P}_x\left(L_x = k\right) = p(x)^{k-1}\mathbb{P}_x(L_x = 1) = p(x)^{k-1}(1-p(x))$, as desired. □

The following convenient representation of $p(x)$ in terms of the transition matrix T provides a useful criterion to determine whether a state $x \in \mathbb{X}$ is recurrent or transient in practice.

Lemma 1.3 (Practical criterion). *For any state $x \in \mathbb{X}$, we have*

$$\frac{1}{1-p(x)} \quad = \quad \sum_{n=0}^{\infty} T^n(x,x). \tag{1.18}$$

in particular, x is transient if and only if the series on the right-hand side is convergent.

Proof. The very definition of $\mathrm{DTMC}(\delta_x, T)$ implies that for all $n \in \mathbb{N}$,

$$\mathbb{P}_x(X_n = x) \quad = \quad \sum_{x_1, \ldots, x_{n-1} \in \mathbb{X}^{n-1}} T(x,x_1)\cdots T(x_{n-1},x) \quad = \quad T^n(x,x).$$

Summing over all $n \in \mathbb{N}$ and invoking Fubini's theorem, we deduce that

$$\mathbb{E}_x[L_x] \quad = \quad \sum_{n=0}^{\infty} T^n(x,x).$$

The claim now readily follows from Lemma 1.2. □

Corollary 1.1 (Class property)**.** *Recurrence or transience is a class property: if a state $x \in \mathbb{X}$ is recurrent (resp. transient) then so is every other state in the same communication class as x.*

Proof. Let $x, y \in \mathbb{X}$ be in the same communication class. By definition, there exists $k, \ell \in \mathbb{N}$ such that $T^k(x,y) > 0$ and $T^\ell(y,x) > 0$. Now, observe that for every $n \in \mathbb{N}$, we have

$$T^{n+k+\ell}(y,y) \quad \geq \quad T^\ell(y,x) T^n(x,x) T^k(x,y). \tag{1.19}$$

Summing over all $n \in \mathbb{N}$, we readily deduce that

$$\sum_{n=0}^{\infty} T^n(y,y) \quad \geq \quad T^k(x,y) T^\ell(y,x) \sum_{n=0}^{\infty} T^n(x,x),$$

Now, if x is recurrent, then the right-hand side is infinite, hence so must the left-hand side be, which means that y is also recurrent. Exchanging the roles of x and y yields the converse. □

In particular, when T is irreducible, either all states are recurrent or all are transient. This dichotomy turns out to take the following surprisingly strong form.

Theorem 1.2 (The strong recurrence/transience dichotomy)**.** *If T is irreducible, then*

- *either every state is visited finitely many times almost-surely, from any initial condition, i.e.*

$$\forall \mu \in \mathcal{P}(\mathbb{X}), \qquad \mathbb{P}_\mu(\forall y \in \mathbb{X},\ L_y < \infty) \quad = \quad 1;$$

- *or every state is visited infinitely often almost-surely, from every initial condition, i.e.*

$$\forall \mu \in \mathcal{P}(\mathbb{X}), \qquad \mathbb{P}_\mu(\forall y \in \mathbb{X},\ L_y = \infty) \quad = \quad 1.$$

Proof. For any $x, y \in \mathbb{X}$, (1.17) and the strong Markov property at time τ_y^+ gives

$$\mathbb{P}_x(L_y = \infty) \quad = \quad \mathbb{P}_x(\tau_y^+ < \infty) \mathbb{P}_y(L_y = \infty). \tag{1.20}$$

In the transient case, the right-hand side is zero, so the left-hand side also is. Since a countable union of negligible event is negligible, we deduce that for all $x \in \mathbb{X}$,

$$\mathbb{P}_x(\exists y \in \mathbb{X},\ L_y = \infty) \quad = \quad 0,$$

and the first claim follows by averaging against $\mu \in \mathcal{P}(\mathbb{X})$. In the recurrent case, we have $\mathbb{P}_y(L_y = \infty) = 1$ instead of 0, and since a countable intersection of almost-sure events is almost-sure, a similar reasoning allows us to conclude provided we can prove that

$$\mathbb{P}_x(\tau_y^+ < \infty) \quad = \quad 1, \tag{1.21}$$

which we now do. By irreducibility, the diagram of the chain contains a path $(z_0, \dots, z_k)$ from $z_0 = y$ to $z_k = x$, and choosing a shortest one ensures that y is visited only once. In particular,

$$\mathbb{P}_y(X_k = x, \tau_y^+ > k) \quad \geq \quad T(z_0, z_1) \cdots T(z_{k-1}, z_k) \ > \ 0.$$

On the other hand, by the weak Markov property at time k,

$$\mathbb{P}_y(\tau_y^+ = \infty, X_k = x) \quad = \quad \mathbb{P}_y(X_k = x, \tau_y^+ > k) \mathbb{P}_x(\tau_y^+ = \infty).$$

Since the left-hand side is zero (y is recurrent), the claim (1.21) now readily follows. □

Lemma 1.4 (Two easy cases)**.** *Let C be a communication class of a Markov chain.*

1. *If C is open, then C is transient.*
2. *If C is closed and finite, then C is recurrent.*

Proof. If C is open, then there is $(x,y) \in C \times (\mathbb{X} \setminus C)$ such that $T(x,y) > 0$. This forces $\mathbb{P}_y(X_k = x) = T^k(y,x) = 0$ for all $k \in \mathbb{N}$, i.e. $\mathbb{P}_y(\tau_x^+ = \infty) = 1$. By the weak Markov property at time 1, we then have

$$\mathbb{P}_x(\tau_x^+ = \infty) \;\geq\; \mathbb{P}_x(X_1 = y, \tau_x^+ = \infty) \;=\; T(x,y)\mathbb{P}_y(\tau_x^+ = \infty) \;=\; T(x,y) \;>\; 0,$$

showing that x (hence C) is transient, as desired. If on the other hand C is finite and closed, then upon restricting T to C, we can assume that T is irreducible and $\mathbb{X}$ is finite. But

$$\sum_{y\in\mathbb{X}} L_y \;=\; \sum_{y\in\mathbb{X}}\sum_{n\in\mathbb{N}} \mathbf{1}_{(X_n=y)} \;=\; \sum_{n\in\mathbb{N}}\sum_{y\in\mathbb{X}} \mathbf{1}_{(X_n=y)} \;=\; +\infty, \tag{1.22}$$

so the random variables $(L_y)_{y\in\mathbb{X}}$ can not all be almost-surely finite. □

Exercise 1.2 (Polya's theorem)**.** *Consider simple random walk on $\mathbb{Z}^d$. Prove that for all $n \in \mathbb{N}$,*

$$T^{2n}(0,0) \;=\; \binom{2n}{n}\frac{1}{(2d)^n}\sum_{k_1+\ldots+k_d=n}\binom{n}{k_1,\ldots,k_d}^2,$$

and $T^{2n+1}(0,0) = 0$. Deduce that the chain is recurrent for $d \in 1,2$ and transient for $d \geq 3$.

Exercise 1.3 (Trees)**.** *Prove that simple random walk on an infinite $d-$regular tree ($d \geq 2$) is transient.*

Exercise 1.4 (Birth-death chains)**.** *On $\mathbb{X} = \mathbb{N}$, consider the transition matrix*

$$T \;:=\; \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & \cdots \\ q_1 & 0 & p_1 & 0 & 0 & \cdots \\ 0 & q_2 & 0 & p_2 & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix},$$

where $p_n \in (0,1)$ and $q_n = 1 - p_n$ for each $n \geq 1$. Prove that the chain is transient if and only if

$$\sum_{n=1}^{\infty} \frac{q_1\cdots q_n}{p_1\cdots p_n} \;<\; +\infty. \tag{1.23}$$

Hint: prove that $f : n \mapsto \mathbb{P}_n[\tau_0 < \infty]$ is the minimal non-negative solution to a certain equation.

1.5 Invariant measures

Throughout these notes, we will view the law $\mu := \mathrm{law}(X)$ of a $\mathbb{X}-$valued random variable X as a row vector with entries indexed by $\mathbb{X}$, as follows:

$$\forall x \in \mathbb{X}, \qquad \mu(x) \;:=\; \mathbb{P}(X = x). \tag{1.24}$$

In particular, such a probability vector can be right-multiplied by a stochastic matrix Q to produce a new probability vector μQ as follows:

$$\forall x \in \mathbb{X}, \qquad (\mu Q)(x) \;:=\; \sum_{y\in\mathbb{X}} \mu(x)Q(x,y). \tag{1.25}$$

With this notation at hand, we have the following elementary but important observation.

Lemma 1.5 (Time-shifts). *If* $(X_n)_{n\geq 0} \sim \mathrm{DTMC}(\mu, T)$, *then* $(X_{n+1})_{n\geq 0} \sim \mathrm{DTMC}(\mu T, T)$.

Proof. For each $n \in \mathbb{N}$ and each $(x_0, \ldots, x_n) \in \mathbb{X}^{n+1}$, we have by definition

$$\mathbb{P}(X_0 = x_0, \ldots, X_n = x_n) \quad = \quad \mu(x_0)T(x_0, x_1) \cdots T(x_{n-1}, x_n). \tag{1.26}$$

Summing over all $x_0 \in \mathbb{X}$, we obtain

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) \quad = \quad (\mu T)(x_1)T(x_1, x_2) \cdots T(x_{n-1}, x_n). \tag{1.27}$$

Being true for all $n \in \mathbb{N}$ and $(x_1, \ldots, x_n) \in \mathbb{X}^n$, this shows that $(X_{n+1})_{n\geq 0} \sim \mathrm{DTMC}(\mu T, T)$. □

It follows from the above lemma the law $\mathrm{DTMC}(\mu, T)$ is invariant under time-shifts if and only if the initial law μ satisfies the fixed point equation:

$$\mu T \quad = \quad \mu. \tag{1.28}$$

A probability measure $\mu \in \mathcal{P}(\mathbb{X})$ satisfying this property is called an invariant (or stationary) law. Our goal in this chapter is to obtain a complete description of all invariants laws of a Markov chain. Here is a simple observation that can be taken as a motivation for doing so.

Lemma 1.6 (Limits are invariant). *Let* $(X_n)_{n\in\mathbb{N}}$ *be a Markov chain with transition matrix* T *such that*

$$\mu \quad := \quad \lim_{n\to\infty} \mathrm{law}(X_n), \tag{1.29}$$

exists in the usual weak sense in $\mathcal{P}(\mathbb{X})$. *Then,* μ *is an invariant law.*

This is perhaps a good time for a quick reminder about weak convergence on discrete spaces.

Exercise 1.5 (Weak convergence of probability measures on a discrete space). *Define the total-variation distance between two probability measures* $\mu, \nu \in \mathcal{P}(\mathbb{X})$ *as follows:*

$$\mathrm{d_{TV}}(\mu, \nu) \quad := \quad \max_{A\subseteq\mathbb{X}} |\mu(A) - \nu(A)|.$$

1. *Prove that* $\mathrm{d_{TV}}(\mu, \nu)$ *admits the following equivalent expressions:*

$$\begin{aligned}
\mathrm{d_{TV}}(\mu, \nu) &= \max_{A\subseteq\mathbb{X}} (\mu(A) - \nu(A)) \\
&= \sum_{x\in\mathbb{X}} (\mu(x) - \nu(x))_+ \\
&= \frac{1}{2} \sum_{x\in\mathbb{X}} |\mu(x) - \nu(x)| \\
&= 1 - \sum_{x\in\mathbb{X}} \mu(x) \wedge \nu(x) \\
&= \inf_{X\sim\mu, Y\sim\nu} \mathbb{P}(X \neq Y) \\
&= \sup_{f\colon \mathbb{X}\to[0,1]} |\mu f - \nu f|,
\end{aligned}$$

2. *Prove that* $\mathrm{d_{TV}}$ *is a distance on* $\mathcal{P}(\mathbb{X})$ *that metrizes weak convergence w.r.t the discrete topology.*

3. *Prove that* $\mathrm{d_{TV}}(\mu_n, \mu) \to 0$ *if and only if* $\mu_n(x) \to \mu(x)$ *for each* $x \in \mathbb{X}$.

4. *Prove that* $\mathrm{d_{TV}}(\mu T, \nu T) \leq \mathrm{d_{TV}}(\mu, \nu)$ *for any stochastic matrix* T *on* $\mathbb{X}$.

Proof of Lemma 1.6. The law $\mu_n = \mathrm{law}(X_n)$ satisfies the equation $\mu_{n+1} = \mu_n T$ for all $n \in \mathbb{N}$, and we may safely pass to the $n \to \infty$ limit to obtain (1.28) because $\mu \mapsto \mu T$ is continuous. □

The following observation will considerably simplify our analysis of invariant laws.

Lemma 1.7 (Zero for transient states)**.** *If $\mu \in \mathcal{P}(\mathbb{X})$ is invariant and x is transient, then $\mu(x) = 0$.*

Proof. Let $\mu \in \mathcal{P}(\mathbb{X})$ be invariant and let $x \in \mathbb{X}$ be transient. Then, for all $n \in \mathbb{N}$, we have

$$\mu(x) \;=\; \mathbb{P}_\mu(X_n = x) \;\le\; \mathbb{P}_\mu\left(\bigcup_{k=n}^{\infty}\{X_k = x\}\right) \;\xrightarrow[n\to\infty]{}\; \mathbb{P}_\mu\left(\bigcap_{n=0}^{\infty}\bigcup_{k=n}^{\infty}\{X_k = x\}\right) \;=\; \mathbb{P}_\mu(N_x = \infty).$$

But the right-hand side is zero because x is transient. □

Consequently, our study of invariant laws reduces to recurrent classes, and since the latter are closed, the stationarity equation $\mu = \mu T$ can in fact be solved separately on each recurrent class. Thus, we may now assume, without loss of generality, that T is irreducible and recurrent. It will actually be useful to consider a natural generalization of the equation (1.28), in which the normalization constraint $\sum_{x\in\mathbb{X}} \mu(x) = 1$ is dropped: we then speak of an invariant measure.

Theorem 1.3 (Existence and uniqueness)**.** *Assume that T is irreducible and recurrent, and fix an arbitrary origin $o \in \mathbb{X}$. Then, the quantity*

$$\mu_o(x) \;:=\; \mathbb{E}_o\left[\sum_{n=0}^{\tau_o^+-1} \mathbf{1}_{\{X_n = x\}}\right],$$

is positive and finite for every $x \in \mathbb{X}$, and the resulting vector μ_o is an invariant measure. Moreover, any other invariant measure μ is proportional to μ_o, i.e. $\mu(\cdot) = \mu(o)\mu_o(\cdot)$.

Proof. For $x \in \mathbb{X} \setminus \{o\}$, we have by Fubini's Theorem,

$$\begin{aligned}
\mu_o(x) &= \sum_{n=0}^{\infty} \mathbb{P}_o(\tau_o^+ > n, X_n = x) \\
&= \sum_{n=1}^{\infty} \mathbb{P}_o(\tau_o^+ > n-1, X_n = x) \\
&= \sum_{n=1}^{\infty} \sum_{y\in\mathbb{X}} \mathbb{P}_o(X_{n-1} = y, \tau_o^+ > n-1, X_n = x) \\
&= \sum_{n=1}^{\infty} \sum_{y\in\mathbb{X}} \mathbb{P}_o(X_{n-1} = y, \tau_o^+ > n-1) T(y,x) \\
&= \sum_{y\in\mathbb{X}} T(y,x)\mu_o(y) \\
&= (\mu_o T)(x),
\end{aligned}$$

where we have crucially used the fact that $x \neq o$ and the weak Markov property at time $n-1$. On the other hand, at the state o, we clearly have $\mu_o(o) = 1$, while

$$\begin{aligned}
(\mu_o T)(o) &= \sum_{x\in\mathbb{X}} \mu_o(x) T(x,o) \\
&= \sum_{x\in\mathbb{X}} \sum_{n\in\mathbb{N}} \mathbb{P}_o(X_n = x, \tau_o^+ > n) T(x,o) \\
&= \sum_{x\in\mathbb{X}} \sum_{n\in\mathbb{N}} \mathbb{P}_o(X_n = x, \tau_o^+ = n+1) \\
&= 1,
\end{aligned}$$

where we have used the Markov property at time n. Thus, the vector μ_o satisfies the stationarity equation $\mu_o T = \mu_o$. Iterating the latter shows that $\mu_o(x)$ is finite and positive everywhere, because it is so at $x = o$ and the chain is irreducible. Now, let μ be any invariant measure, and observe that for any $z_0 \in \mathbb{X} \setminus \{o\}$, we have

$$\begin{aligned}
\mu(z_0) &= \mu(o)T(o,z_0) + \sum_{z_1 \in \mathbb{X}\setminus\{o\}} \mu(z_1)T(z_1,z_0) \\
&= \mu(o)T(o,z_0) + \sum_{z_1 \in \mathbb{X}\setminus\{o\}} \mu(o)T(o,z_1)T(z_1,z_0) + \sum_{z_1,z_2 \in \mathbb{X}\setminus\{o\}} \mu(z_2)T(z_2,z_1)T(z_1,z_0) \\
&= \cdots \\
&= \sum_{z_1,\dots,z_n \in \mathbb{X}\setminus\{o\}} \mu(z_n) \prod_{i=1}^{n} T(z_i,z_{i-1}) + \sum_{k=0}^{n-1} \sum_{z_1,\dots,z_k \in \mathbb{X}\setminus\{o\}} \mu(o)T(o,z_k) \prod_{i=1}^{k} T(z_i,z_{i-1}) \\
&\geq \sum_{k=0}^{n-1} \sum_{z_1,\dots,z_k \in \mathbb{X}\setminus\{o\}} \mu(o)T(o,z_k) \prod_{i=1}^{k} T(z_i,z_{i-1}) \\
&= \sum_{k=0}^{n-1} \mu(o)\mathbb{P}_o(\tau_o^+ > k, X_k = z_0) \\
&= \mu(o)\mathbb{E}_o\left[\sum_{k=0}^{(\tau_o^+ \wedge n)-1} \mathbf{1}_{(X_k = z_0)}\right],
\end{aligned}$$

for every $n \in \mathbb{N}$. Sending $n \to \infty$, we conclude that

$$\mu(z_0) \;\geq\; \mu(o)\mu_o(z_0).$$

Now, this inequality was established for all $z_0 \in \mathbb{X} \setminus \{o\}$, but it trivially also holds (in fact with equality) when $z_0 = o$. Consequently, the difference $\nu(\cdot) := \mu(\cdot) - \mu(o)\mu_o(\cdot)$ is a measure on $\mathbb{X}$, and it solves the stationary equation $\nu T = \nu$ because both μ and μ_o do. In particular, $\nu T^n(o) = \nu(o) = 0$ for every $n \in \mathbb{N}$, which forces $\nu = 0$ by irreducibility. □

Corollary 1.2 (Invariant laws)**.** *If T is irreducible, then either $\mathbb{E}_x[\tau_x^+] = \infty$ for all $x \in \mathbb{X}$ and there is no invariant law, or $\mathbb{E}_x[\tau_x^+] < \infty$ for all $x \in \mathbb{X}$ and there is a unique invariant law, given by*

$$\forall x \in \mathbb{X}, \quad \pi(x) \;:=\; \frac{1}{\mathbb{E}_x[\tau_x^+]}. \tag{1.30}$$

Proof. We may assume that the chain is recurrent, otherwise the claim is trivial. The previous result ensures that the invariant measures $(\mu_o)_{o\in\mathbb{X}}$ are proportional to one another. In particular, their total masses are either all finite, or all infinite. But by construction, we have

$$\mu_o(\mathbb{X}) \;=\; \sum_{x\in\mathbb{X}} \mu_o(x) \;=\; \sum_{x\in\mathbb{X}} \mathbb{E}_o\left[\sum_{n=0}^{\tau_o^+ - 1} \mathbf{1}_{(X_n = x)}\right] \;=\; \mathbb{E}_o[\tau_o^+],$$

for all $o \in \mathbb{X}$. Thus, either $\mathbb{E}_o[\tau_o^+] = \infty$ for all $o \in \mathbb{X}$, or $\mathbb{E}_o[\tau_o^+] < \infty$ for all $o \in \mathbb{X}$. In the first case, all invariant measures have infinite mass, hence there is no invariant law. In the second case, we can divide μ_o by $\mathbb{E}[\tau_o^+]$ to produce an invariant law, and any invariant law π satisfies

$$\pi(o) \;=\; \frac{\pi(o)}{\pi(\mathbb{X})} \;=\; \frac{\mu_o(o)}{\mu_o(\mathbb{X})} \;=\; \frac{1}{\mathbb{E}_o[\tau_o^+]}, \tag{1.31}$$

for any $o \in \mathbb{X}$, by the uniqueness in the previous theorem. □

A recurrent state $x \in \mathbb{X}$ is called positive recurrent if $\mathbb{E}_x[\tau_x^+] < \infty$, and null recurrent otherwise. The above result ensures that x is positive recurrent if and only if it is in the support of some invariant law, and that this is a class property. Of course, any finite closed class is positive recurrent, because measures thereon trivially have finite total mass.

Remark 1.3 (Dropping the irreducibility assumption)**.** *For a general (not necessarily irreducible) Markov chain, the following complete characterization of invariant laws is easily deduced from the above results: if there is no positive recurrent class, then there is no invariant law. If there are positive recurrent classes, say, $(C_i)_{i\in I}$, then the invariant laws are precisely those of the form*

$$\mu(x) \quad = \quad \begin{cases} \frac{\alpha_i}{\mathbb{E}_x[\tau_x^+]} & \text{if } x \in C_i \text{ for some } i \in I \\ 0 & \text{if } x \notin \bigcup_{i\in I} C_i \end{cases}$$

where $(\alpha_i)_{i\in I}$ are arbitrary non-negative coefficients whose total sum is 1*.*

Exercise 1.6. *Prove that the simple random walks on $\mathbb{Z}$ and $\mathbb{Z}^2$ are null recurrent.*

1.6 Time reversal

Our transition matrix T can be viewed as a linear operator acting on the Banach space $L^\infty(\mathbb{X})$ of bounded functions $f : \mathbb{X} \to \mathbb{R}$ as follows:

$$\forall x \in \mathbb{X}, \qquad Tf(x) \quad := \quad \sum_{y\in\mathbb{X}} T(x,y) f(y). \tag{1.32}$$

Note that T is a contraction, in the sense that $\|Tf\|_\infty \leq \|f\|_\infty$ for all $f \in L^\infty(\mathbb{X})$. Assuming from now on that $\mathbb{X}$ is equipped with a fully-supported invariant measure μ, we can extend the above definition to $f \in L^p(\mathbb{X},\mu)$, $1 \leq p < \infty$, and the resulting operator $T : L^p \to L^p$ remains a contraction. Indeed, by Jensen's inequality,

$$\begin{aligned} \|Tf\|_p^p \quad &= \quad \sum_{x\in\mathbb{X}} \mu(x) \left| \sum_{y\in\mathbb{X}} T(x,y) f(y) \right|^p \\ &\leq \quad \sum_{x,y\in\mathbb{X}} \mu(x) T(x,y) \, |f(y)|^p \\ &= \quad \sum_{y\in\mathbb{X}} (\mu T)(y) \, |f(y)|^p \\ &= \quad \|f\|_p^p, \end{aligned}$$

where the last line crucially uses the stationarity equation $\mu T = \mu$. In particular, T is a bounded linear operator on the Hilbert space $L^2(\mathbb{X},\mu)$. As such, it admits an adjoint $T^\star : L^2 \to L^2$, characterized by the duality relation

$$\forall f, g \in L^2(\mathbb{X},\mu), \qquad \langle T^\star f, g \rangle \quad = \quad \langle f, Tg \rangle. \tag{1.33}$$

Choosing $f = \mathbf{1}_{\{y\}}$ and $g = \mathbf{1}_{\{x\}}$ yields the more concrete expression

$$T^\star(x,y) \quad = \quad \frac{\mu(y) T(y,x)}{\mu(x)}, \tag{1.34}$$

from which it readily follows that the matrix $T^\star$ is stochastic, that it admits μ as an invariant measure, and that $T^{\star\star} = T$. When μ is a probability measure, the duality $T \mapsto T^\star$ can be interpreted as time reversal, as explained in the following classical exercise.

Exercise 1.7 (Time reversal). *Assuming that $\mu(\mathbb{X}) = 1$, prove the distributional identity*

$$\forall n \in \mathbb{N}, \qquad (X_0^\star, \dots, X_n^\star) \stackrel{d}{=} (X_n, \dots, X_0), \tag{1.35}$$

in which we have used the notation $(X_n)_{n\geq 0} \sim \mathrm{DTMC}(\mu, T)$ and $(X_n^\star)_{n\geq 0} \sim \mathrm{DTMC}(\mu, T^\star)$.

Because self-adjoint operators are much better behaved than non self-adjoint ones, Markov chains that satisfy the self-duality relation $T^\star = T$ play a distinguished role in the theory.

Definition 1.5 (Reversibility). *T is called reversible when $T^\star = T$, or equivalently,*

$$\forall (x, y) \in \mathbb{X}^2, \qquad \mu(x)T(x, y) \;=\; \mu(y)T(y, x). \tag{1.36}$$

Note that summing the identity (1.36) over all $x \in \mathbb{X}$ yields exactly the stationarity property $\mu T = \mu$, sometimes called the global balance equation. For this reason, (1.36) is often referred to as the detailed balance equation. This powerful local symmetry happens to hold in many fundamental models, and admits a number of interesting consequences, as we will see.

Example 1.8 (Simple random walk). *Consider simple random walk on a locally finite graph $G = (V, E)$, as defined earlier. Prove that there exists a reversible measure if and only if E is symmetric.*

Example 1.9 (Random walks on groups). *Consider the random walk generated by a probability measure ν on a finite group $\mathbb{X}$. Prove that the counting measure ($\mu(x) = 1$ for all $x \in \mathbb{X}$) is always invariant, and find a necessary and sufficient condition for it to be reversible.*

Example 1.10 (Glauber dynamics). *Consider a probability measure π on a product space of the form $\mathbb{S}^n$, where $\mathbb{S}$ is a countable set. On the support $\mathbb{X} := \mathrm{supp}(\pi)$, consider the transition matrix*

$$T(x, y) \;:=\; \frac{1}{n}\sum_{i=1}^{n} \frac{\pi(y)}{\pi(L_{i,x})} \mathbf{1}_{L_{i,x}}(y), \tag{1.37}$$

where the 'line' $L_{i,x} = \{x_1\} \times \cdots \times \{x_{i-1}\} \times \mathbb{S} \times \{x_{i+1}\} \times \cdots \times \{x_n\}$ consists of all points that agree with x except maybe for the ith coordinate. In words, a transition consists in selecting one of the n coordinates uniformly at random and resampling its content according to the measure π, conditioned on the current values of all other coordinates. Check that this dynamics is reversible with respect to π.

Example 1.11 (Chess). *Imagine a king starting at one corner of a chessboard and performing simple random walk according to the rules of chess, until it returns to its initial position. On average, how much time will it spend in the four central squares? What about a rook? A knight? A bishop? A queen?*

1.7 The ergodic theorem

For a measure μ on $\mathbb{X}$ and a function $f : \mathbb{X} \to \mathbb{R}$ which is non-negative or in $L^1(\mathbb{X}, \mu)$, we let

$$\mu f \;:=\; \sum_{x\in\mathbb{X}} \mu(x) f(x), \tag{1.38}$$

denote the integral of f w.r.t. μ. The following fundamental result considerably generalizes the strong law of large numbers, and can be seen as the starting point of the celebrated Monte Carlo Markov Chain revolution in computational statistics. Roughly speaking, the idea there is to efficiently compute integrals of the form (1.38) with respect to a complicated high-dimensional measure μ, by running an appropriate Markov chain that admits μ as an invariant law. A simple and popular example is the Glauber dynamics, introduced at (1.37).

Theorem 1.4 (Ergodic Theorem). *Let $(X_n)_{n\in\mathbb{N}}$ be an irreducible and recurrent Markov chain, with invariant measure μ. Then, for any $f \in L^1(\mathbb{X},\mu)$, we have*

$$\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \quad \xrightarrow[n\to\infty]{a.-s.} \quad \frac{\mu f}{\mu(\mathbb{X})} \;=\; \begin{cases} \pi f & \text{in the positive-recurrent case} \\ 0 & \text{in the null-recurrent case.} \end{cases} \tag{1.39}$$

Proof. Upon replacing f with f_+ and f_- if necessary, we can assume without loss of generality that f is non-negative. Also, upon decomposing $\mathbb{P}(\cdot)$ as $\sum_{o\in\mathbb{X}} \mathbb{P}(X_0 = o)\mathbb{P}_o(\cdot)$ if necessary, we can assume that $\mathbb{P}(X_0 = o) = 1$ for some fixed state $o \in \mathbb{X}$. Now, define a sequence $(\tau_k)_{k\geq 0}$ of $\mathbb{N}\cup\{+\infty\}-$valued random variables inductively by setting $\tau_0 = 0$ and

$$\tau_{k+1} \;:=\; \inf\{n \geq \tau_k + 1 \colon X_n = o\}. \tag{1.40}$$

In words, $\tau_0, \tau_1, \ldots$ are the successive times at which o is visited. Note that those are stopping times, and that they are almost-surely finite thanks to our recurrence assumption. Moreover, the strong Markov property ensures that the random variables $(Z_k)_{k\in\mathbb{N}}$ defined by

$$Z_k \;:=\; \sum_{\ell=\tau_k}^{\tau_{k+1}-1} f(X_\ell), \tag{1.41}$$

are i.i.d.. Since they are non-negative, the strong law of large numbers guarantees that

$$\frac{1}{n}\sum_{k=0}^{\tau_n-1} f(X_k) \;=\; \frac{1}{n}\sum_{k=0}^{n-1} Z_k \quad \xrightarrow[n\to\infty]{a.-s.} \quad \mathbb{E}\left[Z_0\right] \;=\; \sum_{x\in\mathbb{X}} f(x)\mathbb{E}\left[\sum_{k=0}^{\tau_o^+-1} \mathbf{1}_{(X_k=x)}\right] \;=\; \mu_o f,$$

where μ_o is the invariant measure appearing in Theorem 1.3. Now, for each $n \in \mathbb{N}$, we have

$$\tau_{N_n} \;\leq\; n \;<\; \tau_{N_n+1}, \tag{1.42}$$

almost-surely, where $N_n := \sum_{k=1}^{n} \mathbf{1}_{(X_k=o)}$ counts the visits to o. Thus,

$$\frac{1}{N_n}\sum_{k=0}^{\tau_{N_n}-1} f(X_k) \;\leq\; \frac{1}{N_n}\sum_{k=0}^{n-1} f(X_k) \;\leq \left(1+\frac{1}{N_n}\right)\frac{1}{N_n+1}\sum_{k=0}^{\tau_{N_n+1}-1} f(X_k),$$

and since $N_n \to \infty$ almost-surely as $n \to \infty$ (o is recurrent), we deduce that

$$\frac{1}{N_n}\sum_{k=0}^{n-1} f(X_k) \quad \xrightarrow[n\to\infty]{a.-s.} \quad \mu_o f. \tag{1.43}$$

Finally, since the above argument applied to any $f\colon \mathbb{X} \to \mathbb{R}_+$, we may choose $f = 1$ to obtain

$$\frac{n}{N_n} \quad \xrightarrow[n\to\infty]{a.-s.} \quad \mu_o(\mathbb{X}), \tag{1.44}$$

and the last two displays together readily imply the claim, because μ_o is proportional to μ. □

Example 1.12 (Google's Page Rank Algorithm). *Given a finite directed graph $G = (V, E)$ such as the World Wide Web, a natural problem with numerous applications consists in attributing to every node $x \in V$ a "score" $\mu(x) \geq 0$ that measures its "importance" or "influence" in the network. Intuitively, the score $\mu(x)$ should be higher if x is the end-point of many arrows $(y, x) \in E$. However, we naturally want the contribution of an arrow $(y, x) \in E$ to be larger if y itself has a high score, and smaller if y points to many other nodes rather than just to x. A moment's thought then leads to the consistency equation*

$$\mu(x) \;=\; \sum_{y\in\mathbb{X}} \frac{\mu(y)}{\deg(y)}\mathbf{1}_{(y,x)\in E}, \tag{1.45}$$

which is exactly the stationary equation $\mu = \mu T$ associated with the simple random walk on G. Consequently, a reasonable way to attribute a score to a node $x \in \mathbb{X}$ consists in running a simple random walk on the network for a long time, and counting the number of visits to x. This simple and beautiful idea is the key behind the celebrated Page Rank algorithm developed by Larry Page and Sergey Brin, the founders of Google, while they were Ph. D. students at Stanford University in the late 1990s.

1.8 Convergence to equilibrium

Let $(X_n)_{n\in\mathbb{N}}$ be an irreducible and positive-recurrent Markov chain. By taking expectations in the ergodic theorem (Theorem 1.4), we have the convergence

$$\frac{1}{n}\sum_{k=0}^{n-1}\mathbb{P}(X_k \in \cdot) \quad \xrightarrow[n\to\infty]{} \quad \pi(\cdot), \tag{1.46}$$

in the usual sense of weak convergence on $\mathbb{X}$ (see Exercise 1.5), regardless of the initial condition. In words, the large-time marginals converge, in the Cesàro sense, to the invariant law. Our aim here is to identify the conditions that guarantee the stronger (and more satisfactory) conclusion

$$\mathbb{P}(X_n \in \cdot) \quad \xrightarrow[n\to\infty]{} \quad \pi(\cdot). \tag{1.47}$$

The fact that additional assumptions are required is made clear in the following simple example.

Example 1.13 (Alternating chain). *On $\mathbb{X} = \{0,1\}$, consider the transition matrix*

$$T \quad = \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Verify that T is irreducible and positive-recurrent, explicitate the invariant law π, and compute T^n for all $n \in \mathbb{N}$. Deduce that the laws $\mathbb{P}_0(X_n \in \cdot)$ and $\mathbb{P}_1(X_n \in \cdot)$ do not converge as $n \to \infty$.

As is clear from the above example, an obvious obstruction to the convergence (1.47) is the possibility that the Markov chain deterministically cycles through distinct parts of the state space. The purpose of the following definition is precisely to preclude such pathologies.

Definition 1.6 (Aperiodicity). *An irreducible transition matrix is called aperiodic if for every pair of states $(x,y) \in \mathbb{X}^2$, there exists a time $n_{x,y} \in \mathbb{N}$ such that*

$$\forall n \geq n_{x,y}, \qquad T^n(x,y) \quad > \quad 0. \tag{1.48}$$

Remark 1.4 (Simplification). *Fix an arbitrary origin $o \in \mathbb{X}$. Then, for any $(x,y) \in \mathbb{X}^2$, we can find $k, \ell \in \mathbb{N}$ such that $T^k(x,o) > 0$ and $T^\ell(o,y) > 0$, and the crude inequality*

$$T^{n+k+\ell}(x,y) \quad \geq \quad T^k(x,o)T^n(o,o)T^\ell(o,y),$$

shows that the ultimate positivity condition (1.48) needs only be checked at the diagonal entry (o,o). Moreover, since the set $I := \{n \in \mathbb{N} \colon T^n(o,o) > 0\}$ is stable under addition, this question further reduces to the existence of coprime elements in I, making aperiodicity rather easy to decide in practice.

Exercise 1.8 (Semi-groups of $\mathbb{N}$). *Prove that any $I \subseteq \mathbb{N}$ stable under addition is ultimately periodic:*

$$d\mathbb{N} \cap [n,\infty) \quad \subseteq \quad I \quad \subseteq \quad d\mathbb{N},$$

for some $(n,d) \in \mathbb{N}^2$, with in fact $d = \gcd(I)$.

Of course, aperiodicity is necessary for the convergence (1.47) to occur, because $\pi(\cdot)$ has full support. What is remarkable is that this condition is also sufficient, as stated in the next theorem.

Theorem 1.5 (Mixing). *Any irreducible, aperiodic and positive-recurrent Markov chain $(X_n)_{n\geq 0}$ converges in distribution to its unique invariant law, regardless of the initial condition:*

$$\mathrm{d}_{\mathrm{TV}}\left(\mathrm{law}(X_n), \pi\right) \quad \xrightarrow[n\to\infty]{} \quad 0.$$

Proof. We use an instructive argument based on coupling. Let $Y \sim \mathrm{MC}(\pi, T)$ be independent of X, and starting from the invariant law π. Then, the bi-variate process $(X_n, Y_n)_{n\geq 0}$ is easily seen to be a Markov chain on $\mathbb{X}^2$, with transition matrix

$$K\left((x,y),(x',y')\right) \quad := \quad T(x,x')T(y,y'). \tag{1.49}$$

By an immediate induction, we have for all $x, y, x', y' \in \mathbb{X}$ and all $n \in \mathbb{N}$,

$$K^n\left((x,y),(x',y')\right) \quad = \quad T^n(x,x')T^n(y,y').$$

Thus, our assumption on T ensures that K is irreducible and recurrent. In particular, $(X_n, Y_n)_{n\geq 0}$ will eventually almost-surely hit the diagonal set

$$\Delta \quad := \quad \{(x,x) \colon x \in \mathbb{X}\}. \tag{1.50}$$

Now, the associated hitting time τ is a stopping time w.r.t. the filtration $(\sigma(X_0, Y_0, \ldots, X_n, Y_n))_{n\in\mathbb{N}}$, and the strong Markov property ensures that the two processes $(X_\tau, X_{\tau+1}, \ldots)$ and $(Y_\tau, Y_{\tau+1}, \ldots)$ have the same conditional distribution given $\mathcal{F}_\tau$. In particular, the random sequence

$$Z \quad := \quad (X_0, X_1, \ldots, X_\tau, Y_{\tau+1}, Y_{\tau+2}, \ldots), \tag{1.51}$$

has the same distribution as X. Consequently, for any $A \subseteq \mathbb{X}$ and every $n \in \mathbb{N}$, we can write

$$\begin{aligned} \mathbb{P}(X_n \in A) - \pi(A) \quad &= \quad \mathbb{P}(Z_n \in A) - \mathbb{P}(Y_n \in A) \\ &\leq \quad \mathbb{P}(Z_n \in A, Y_n \notin A) \\ &\leq \quad \mathbb{P}(\tau > n). \end{aligned}$$

Since replacing A with A^c changes $\mathbb{P}(X_n \in A) - \pi(A)$ to its negative, we conclude that

$$\sup_{A\subseteq\mathbb{X}} \left|\mathbb{P}(X_n \in A) - \pi(A)\right| \quad \leq \quad \mathbb{P}(\tau > n),$$

and the claim now readily follows by sending $n \to \infty$ and recalling that $\mathbb{P}(\tau < \infty) = 1$. □

Chapter 2

Continuous-time Markov chains

2.1 Constructive definition

In this chapter, we investigate the more realistic (but more challenging) situation where time is continuous, while space is discrete. Thus, a stochastic process will here consist of a collection of random variables defined on some underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$, taking values in some countable set $\mathbb{X}$, and indexed by the continuous half-line $\mathbb{R}_+ = [0, \infty)$. A rather generic way to produce such a process $X = (X_t)_{t \geq 0}$ is to start from a discrete-time stochastic process $Y = (Y_0, Y_1, \ldots)$ and assign positive random durations $E = (E_0, E_1, \ldots)$ to the steps, as follows:

$$\forall \omega \in \Omega, \quad \forall n \in \mathbb{N}, \quad \forall t \in \left[\sum_{k=0}^{n-1} E_k(\omega), \sum_{k=0}^{n} E_k(\omega) \right), \quad X_t(\omega) \quad := \quad Y_n(\omega). \tag{2.1}$$

In order to ensure that this definition actually specifies $X = (X_t)_{t \geq 0}$ completely, we need to require that the sequence of durations E is non-explosive, meaning that

$$\sum_{n=0}^{\infty} E_n \quad = \quad +\infty. \tag{2.2}$$

Under this condition, the procedure (2.1) completely determines X, and we write $X = \Psi(Y, E)$. Note that the function $\Psi : \mathbb{X}^{\mathbb{N}} \times (0, \infty)^{\mathbb{N}} \to \mathbb{X}^{[0,\infty)}$ thus defined is measurable with respect to the corresponding product topologies. Indeed, for every $t \geq 0$ and $x \in \mathbb{X}$, we can write

$$\{X_t = x\} \quad = \quad \bigcup_{n=0}^{\infty} \{E_0 + \cdots + E_{n-1} \leq t < E_0 + \cdots + E_n\} \cap \{Y_n = x\}. \tag{2.3}$$

Thus, X is indeed a stochastic process. By construction, the latter is obviously càdlàg: each sample-path trajectory $t \mapsto X_t(\omega)$, $\omega \in \Omega$ is right-continuous and admits a well-defined left limit everywhere. Conversely, any càdlàg stochastic process must in fact be of the form (2.1)-(2.2), because our state space $\mathbb{X}$ is discrete. We emphasize, however, that the representation (2.1) is not unique, because the occurrence of consecutive repetitions in Y allows for an arbitrary repartition of the corresponding durations in E without affecting the resulting process $X = \Psi(Y, E)$.

Restricting attention to càdlàg trajectories will be very convenient from a measure-theoretical viewpoint. Indeed, such trajectories are entirely determined by their values on a dense countable set (say $\mathbb{Q}_+$), and this can be used to express any reasonable property of càdlàg trajectories as a certain event A that belongs to the product $\sigma-$field of $\mathbb{X}^{[0,\infty)}$. Consequently, the associated probability $\mathbb{P}(X \in A)$ is well defined and determined by the finite-dimensional marginals of X, and hence by those of the pair (Y, E). This fact will be used repeatedly in the sequel.

As the title of the chapter indicates, our aim here is to develop a suitable continuous-time analogue of discrete-time Markov chains, based on the same fundamental principle: *only their current state should influence what they do next*. As we will see, the right and only way to enforce this lack of memory in a càdlàg stochastic process $X = \Psi(Y, E)$ is to impose the following structural assumptions on the pair (Y, E). Throughout this chapter, $\mu \in [0,1]^{\mathbb{X}}$ is a probability vector, $T \in [0,1]^{\mathbb{X}\times\mathbb{X}}$ a stochastic matrix, and $q \in (0,\infty)^{\mathbb{X}}$ a positive vector.

Definition 2.1 (Constructive description)**.** *A continuous-time Markov chain with parameters (μ, T, q) is a stochastic process of the form $X = \Psi(Y, E)$, where*

1. *Y is a discrete-time Markov chain with initial law μ and transition matrix T;*
2. *Conditionally on Y, the variables $E_0, E_1, \ldots$ are independent Exponential of rates $q(Y_0), q(Y_1), \ldots$*
3. *E is non-explosive.*

Remark 2.1 (Recursive formulation)**.** *The law of the pair (Y, E) specified by conditions 1 and 2 in Definition 2.1 can be equivalently expressed in the following inductive way:*

$$\forall x \in \mathbb{X},\ \forall t \in \mathbb{R}_+,\ \forall n \in \mathbb{N},\ \mathbb{P}\left(Y_{n+1} = x, E_{n+1} \geq t \,\middle|\, Y_0, E_0, \ldots, Y_n, E_n\right) \quad = \quad T(Y_n, x)e^{-q(Y_n)t}.$$

In particular, choosing $t = 0$ shows that the process Y is a Markov chain w.r.t. the richer filtration

$$\forall n \in \mathbb{N}, \qquad \mathcal{F}_n \quad := \quad \sigma(Y_0, E_0, \ldots, Y_n, E_n).$$

Note that there is an existential issue in the above definition. Specifically, any triple (μ, T, q) does (uniquely) specify the law of some random pair (Y, E) satisfying the first two conditions, but the third one may fail. We will call (μ, T, q) non-explosive when

$$\mathbb{P}\left(\sum_{n=0}^{\infty} E_n = \infty\right) \quad = \quad 1.$$

Upon modifying E on a null set, we may then safely define a continuous-time Markov chain $X = \Psi(Y, E)$, and its law is then entirely determined by (μ, T, q), allowing us to write $X \sim \mathrm{CTMC}(\mu, T, q)$. A characterization of non-explosive triples will be given below, after we recall the definition and main properties of exponential random variables.

2.2 Exponential random variables

We start by recalling that a $(0,\infty)-$valued random variable E (or more accurately, its law) is Exponential with rate $\lambda \in (0,\infty)$, written $E \sim \mathrm{Exp}(\lambda)$, if its tail distribution function is

$$\forall t \geq 0, \qquad \mathbb{P}(E > t) \quad = \quad e^{-\lambda t}. \tag{2.4}$$

It readily follows from this definition that E stochastically decreases with λ, that it has mean $1/\lambda$ and density $f_\lambda(t) = \lambda e^{-\lambda t}$, and that it enjoys the following elementary scaling property:

$$E \sim \mathrm{Exp}(\lambda) \quad \Longleftrightarrow \quad \lambda E \sim \mathrm{Exp}\,(1)\,. \tag{2.5}$$

It is convenient to extend the above definition to the degenerate cases $\lambda = 0$ and $\lambda = \infty$ as follows: $\mathrm{Exp}(0) := \delta_\infty$ and $\mathrm{Exp}(\infty) := \delta_0$.

Lemma 2.1 (Summability of Exponential random variables). *Let $(E_n)_{n\in\mathbb{N}}$ be independent Exponential random variables with rates $(\lambda_n)_{n\in\mathbb{N}}$. Then,*

$$\sum_{n=0}^{\infty}\frac{1}{\lambda_n} < \infty \implies \mathbb{P}\left(\sum_{n=0}^{\infty} E_n < +\infty\right) = 1;$$
$$\sum_{n=0}^{\infty}\frac{1}{\lambda_n} = \infty \implies \mathbb{P}\left(\sum_{n=0}^{\infty} E_n = +\infty\right) = 1.$$

Proof. The first claim readily follows from the observation that

$$\mathbb{E}\left[\sum_{n=0}^{\infty} E_n\right] = \sum_{n=0}^{\infty}\frac{1}{\lambda_n}.$$

As for the second, we use instead the relation

$$\mathbb{E}\left[\exp\left(-\sum_{n=0}^{\infty} E_n\right)\right] = \prod_{n=0}^{\infty}\frac{\lambda_n}{1+\lambda_n} = \exp\left(-\sum_{n=0}^{\infty}\log\left(1+\frac{1}{\lambda_n}\right)\right),$$

along with the classical fact that $\sum_{n=0}^{\infty}\log\left(1+\frac{1}{\lambda_n}\right) < \infty$ if and only if $\sum_{n=0}^{\infty}\frac{1}{\lambda_n} < \infty$. □

This result readily implies the following criterion for non-explosive triples.

Corollary 2.1 (Non-explosion). *A triple (μ, T, q) is non-explosive if and only if*

$$\mathbb{P}\left(\sum_{n=0}^{\infty}\frac{1}{q(Y_n)} = +\infty\right) = 1,$$

where $Y \sim \mathrm{DTMC}(\mu, T)$. In particular, this holds when $\sup_{x\in\mathbb{X}} q(x) < \infty$, or when μ is supported on recurrent states. More generally, it suffices that $\sup_{x\in A} q(x) < \infty$, where $A \subseteq \mathbb{X}$ consists of those transient states that have a positive chance to be visited by Y.

The following celebrated property is the essential reason behind the prominent role played by Exponential random variabes in the theory of continuous-time Markov chains.

Lemma 2.2 (Memoryless property). *A $[0,\infty]-$valued random variable E is Exponential if and only if it satisfies the following memoryless property:*

$$\forall s, t \geq 0, \qquad \mathbb{P}(E > t+s) = \mathbb{P}(E > t)\mathbb{P}(E > s). \tag{2.6}$$

Proof. Let E be a $[0,\infty]-$valued random variable with the memoryless property. This means that its tail distribution function $F\colon t \mapsto \mathbb{P}(E > t)$ satisfies

$$\forall t, s \geq 0, \qquad F(t+s) = F(t)F(s). \tag{2.7}$$

Now, write $F(1) = e^{-\lambda}$ for some $\lambda \in [0,\infty]$ and observe that by an immediate induction, the equation (2.7) evaluated at integer times gives for all $n \in \{1, 2, \ldots\}$,

$$F(n) = F\left(\underbrace{1+\cdots+1}_{n \text{ times}}\right) = \underbrace{F(1)\times\cdots\times F(1)}_{n \text{ times}} = e^{-\lambda n}.$$

In a similar spirit, for all $m, n \in \{1, 2, \ldots\}$, we can write

$$F(n) \quad = \quad F\left(\underbrace{\frac{n}{m} + \cdots + \frac{n}{m}}_{m \text{ times}}\right) \quad = \quad \underbrace{F\left(\frac{n}{m}\right) \times \cdots \times F\left(\frac{n}{m}\right)}_{m \text{ times}}.$$

Those two lines together force $F(n/m) = e^{-\lambda n/m}$. In other words, we have shown that

$$\mathbb{P}(E > t) \quad = \quad e^{-\lambda t},$$

for all $t \in (0, \infty) \cap \mathbb{Q}$. Finally, the conclusion extends to all $t \in (0, \infty)$ because tail distribution functions are right-continuous (or alternatively, non-increasing). Conversely, it is clear that the function $F(t) = e^{-\lambda t}$ satisfies the memory-less property (2.7). □

Lemma 2.3 (Minimum of Exponential random variables)**.** *Let $(E_n)_{n\in\mathbb{N}}$ be independent Exponential random variables with rates $(\lambda_n)_{n\in\mathbb{N}}$. Then, the random variable*

$$E \quad := \quad \inf_{n\in\mathbb{N}} E_n, \tag{2.8}$$

is Exponential with rate $\lambda := \sum_{n\in\mathbb{N}} \lambda_n$. If moreover $\lambda \in (0, \infty)$, then the above infimum is almost-surely attained at a unique (random) index N, and the latter is independent of E with law

$$\forall n \in \mathbb{N}, \qquad \mathbb{P}(N = n) \quad = \quad \frac{\lambda_n}{\lambda}. \tag{2.9}$$

Proof. For all $t > 0$, we can write

$$\mathbb{P}(E \geq t) \quad = \quad \mathbb{P}\left(\bigcap_{n=0}^{\infty} \{E_n \geq t\}\right) \quad = \quad \prod_{n=0}^{\infty} \mathbb{P}(E_n \geq t) \quad = \quad \prod_{n=0}^{\infty} e^{-\lambda_n t} \quad = \quad e^{-\lambda t},$$

establishing the first claim. We henceforth assume that $\lambda \in (0, \infty)$. For $t > 0, k \in \mathbb{N}$, we have

$$\begin{aligned}
\mathbb{P}\left(E \geq t, \inf_{n\in\mathbb{N}} E_n \text{ is uniquely attained at } k\right) &= \mathbb{P}\left(\{E_k \geq t\} \cap \bigcap_{n\in\mathbb{N}\setminus\{k\}} \{E_n > E_k\}\right) \\
&= \mathbb{E}\left[\mathbf{1}_{(E_k\geq t)} \mathbb{P}\left(\bigcap_{n\in\mathbb{N}\setminus\{k\}} \{E_n > E_k\} \middle| E_i\right)\right] \\
&= \mathbb{E}\left[\mathbf{1}_{(E_k\geq t)} \prod_{n\in\mathbb{N}\setminus\{k\}} e^{-\lambda_n E_k}\right] \\
&= \mathbb{E}\left[\mathbf{1}_{(E_k\geq t)} e^{(\lambda_k - \lambda) E_k}\right] \\
&= \frac{\lambda_k}{\lambda} e^{-\lambda t}.
\end{aligned}$$

Sending $t \to 0$ and summing over $k \in \mathbb{N}$ shows that $\inf_{n\in\mathbb{N}} E_n$ is indeed almost-surely uniquely attained, so that N is well-defined. The above computation then rewrites as follows:

$$\forall t \in (0, \infty), \quad \forall k \in \mathbb{N}, \qquad \mathbb{P}(E \geq t, N = k) \quad = \quad \frac{\lambda_k}{\lambda} e^{-\lambda t},$$

showing that E and N are independent, with the appropriate marginals. □

2.3 Markov property, semi-group and generator

We are now ready to establish the following fundamental property of continuous-time Markov chains, which formalizes the idea that *only their current state can influence what they do next*.

Theorem 2.1 (Markov property). *If $X \sim \mathrm{CTMC}(\mu, T, q)$, then for any fixed time $s \geq 0$,*

$$\mathrm{Law}\left((X_{t+s})_{t\geq 0} \,\middle|\, (X_t)_{t\in[0,s]}\right) \quad = \quad \mathrm{CTMC}(\delta_{X_s}, T, q).$$

As anticipated, the memory-less property of Exponential random variables will play a crucial role in the proof. Specifically, we use it to establish the following shift-invariance property.

Lemma 2.4 (Random shifts of Exponential sequences). *Let $E_0, E_1, \ldots$ be independent Exponential random variables whose rates $\lambda_0, \lambda_1, \ldots$ are positive and satisfy $\sum_{n=0}^{\infty} \frac{1}{\lambda_n} = +\infty$. Fix $s \in [0, \infty)$ and set*

$$N \quad := \quad \min\{n \geq 0 \colon E_0 + \cdots + E_n > s\}.$$

Then, the process $\widetilde{E} := (E_1 + \cdots + E_N - s, E_{N+1}, E_{N+2}, \ldots)$ satisfies

$$\mathrm{Law}\left(\widetilde{E} \,\middle|\, N, E_0, \ldots, E_{N-1}\right) \quad = \quad \bigotimes_{n=0}^{\infty} \mathrm{Exp}\left(\lambda_{N+n}\right).$$

Proof. The condition $\sum_{n=0}^{\infty} \frac{1}{\lambda_n} = +\infty$ guarantees that the random variable N is well-defined, by Lemma 2.1. Now fix $n, m \in \mathbb{N}$ and $t_0, \ldots, t_m \geq 0$, and put $S := s - (E_0 + \cdots + E_{n-1})$. Observing that $\{N = n\} = \{0 \leq S < E_n\}$, we can write

$$\begin{aligned}
&\mathbb{P}\left(N = n, \widetilde{E}_0 \geq t_0, \ldots, \widetilde{E}_m \geq t_m \,\middle|\, E_0, \ldots, E_{n-1}\right) \\
&= \mathbf{1}_{(S\geq 0)} \mathbb{P}\left(E_n > t_0 + S \,\middle|\, E_0 \ldots, E_{n-1}\right) e^{-\sum_{i=1}^{m} t_i \lambda_{n+i}} \\
&= \mathbf{1}_{(S\geq 0)} \mathbb{P}\left(E_n > S \,\middle|\, E_0 \ldots, E_{n-1}\right) \mathbb{P}\left(E_n > t_0 \,\middle|\, E_0 \ldots, E_{n-1}\right) e^{-\sum_{i=1}^{m} t_i \lambda_{n+i}} \\
&= \mathbb{P}\left(N = n \,\middle|\, E_0 \ldots, E_{n-1}\right) e^{-\sum_{i=0}^{m} t_i \lambda_{n+i}},
\end{aligned}$$

where at the third line we have crucially used the memoryless property of E_n under the conditional law $\mathbb{P}(\cdot \,|\, E_0, \ldots, E_{n-1})$. This shows that, conditionally on the event $\{N = n\}$ and on the random variables $(E_0, \ldots, E_{n-1})$, the variables $(\widetilde{E}_0, \ldots, \widetilde{E}_m)$ are independent Exponential with rates $(\lambda_n, \ldots, \lambda_{n+m})$. Since $n, m \in \mathbb{N}$ are arbitrary, the claim is proved. □

Proof of Theorem 2.1. By assumption, we have $X = \Psi(Y, E)$ for some random pair (Y, E) satisfying the three conditions in Definition 2.1. In particular, the random variable

$$N \quad := \quad \min\{n \geq 0 \colon E_0 + \cdots + E_n > s\},$$

is well-defined, and since $E_1 + \cdots + E_{N-1} \leq s < E_1 + \cdots + E_N$, the shifted process $\widetilde{X} := (X_{t+s})_{t\geq 0}$ can be written as $\widetilde{X} = \Psi(\widetilde{Y}, \widetilde{E})$ with $(\widetilde{Y}, \widetilde{E})$ obtained from (Y, E) as follows:

$$\widetilde{Y} \;:=\; (Y_N, Y_{N+1}, Y_{N+2}, \ldots), \quad \text{and} \quad \widetilde{E} \;:=\; (E_0 + \cdots + E_N - s, E_{N+1}, E_{N+2}, \ldots).$$

Note that the non-explosion of $\widetilde{E}$ trivially follows from that of E. Also, $(X_t)_{t\in[0,s]}$ is a measurable function of $(N, Y_0, \ldots, Y_N, E_0, \ldots, E_{N-1})$, since for all $t \in [0, s]$ and all $x \in \mathbb{X}$ we can write

$$\{X_t = x\} \quad = \quad \bigcup_{n=0}^{\infty} \{N = n\} \bigcap \left[\left\{ Y_n = x, \sum_{i=0}^{n-1} E_i \leq t \right\} \cup \bigcup_{k=0}^{n-1} \left\{ Y_k = x, \sum_{i=0}^{k-1} E_i \leq t < \sum_{i=0}^{k} E_i \right\} \right].$$

Thus, our task boils down to verifying the following two properties:

$$\begin{aligned} \mathrm{Law}\left(\widetilde{Y} \,\middle|\, N, E_0, \dots, E_{N-1}, Y_0, \dots, Y_N\right) &= \mathrm{DTMC}(\delta_{X_s}, T); \\ \mathrm{Law}\left(\widetilde{E} \,\middle|\, N, E_0, \dots, E_{N-1}, Y\right) &= \bigotimes_{n=0}^{\infty} \mathrm{Exp}\left(q(\widetilde{Y}_n)\right). \end{aligned}$$

Now, the first claim is a consequence of the strong Markov property (w.r.t. the filtration appearing in Remark 2.1) for the discrete-time Markov chain Y at the stopping time N. Note that the latter is finite, with $Y_N = X_s$. On the other hand, the second claim is simply obtained by applying Lemma 2.4 conditionally on Y. □

Exercise 2.1 (Poisson process)**.** *Let $(E_n)_{n\geq 0}$ be independent Exponential random variables with rate $\lambda \in (0,\infty)$. The Poisson process of rate $\lambda \in (0,\infty)$ is the stochastic process $X = (X_t)_{t\geq 0}$ defined by*

$$\forall t \geq 0, \qquad X_t := \sum_{n=0}^{\infty} \mathbf{1}_{(E_0+\cdots+E_n \leq t)}.$$

1. *Check that X is a continuous-time Markov chain, and explicitate its parameters.*
2. *Prove that for each $n \in \mathbb{N}$, the random variable $E_0 + \cdots + E_n$ has the $\Gamma(n+1,\lambda)$ density:*

$$\forall t \in [0,\infty), \qquad f_\lambda^{\otimes(n+1)}(t) = \frac{\lambda^{n+1} t^n e^{-\lambda t}}{n!}.$$

3. *Deduce that for every $t \in \mathbb{R}_+$, the random variable X_t has the $\mathrm{Poisson}(\lambda t)$ distribution:*

$$\forall n \in \mathbb{N}, \qquad \mathbb{P}(X_t = n) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}.$$

4. *Fix $s \geq 0$. Prove that $(X_{t+s} - X_s)_{t\geq 0}$ is a Poisson process of rate λ, independent of $(X_t)_{t\in[0,s]}$.*

Given a continuous-time Markov chains X with parameters (μ, T, q), our aim here is to provide a more analytical understanding of its law. We will assume that μ has full support, so that we can define a collection of probability measures $(\mathbb{P}_x)_{x\in\mathbb{X}}$ on $(\Omega, \mathcal{F})$ as follows:

$$\forall x \in \mathbb{X}, \qquad \mathbb{P}_x(\cdot) := \mathbb{P}(\cdot \,|\, X_0 = x).$$

Of course, the law of X under $\mathbb{P}_x$ is $\mathrm{CTMC}(\delta_x, T, q)$, regardless of our particular choice for μ. Let us now introduce the following important family of stochastic matrices $(P_t)_{t\geq 0}$:

$$\forall t \geq 0, \qquad \forall (x,y) \in \mathbb{X}^2, \qquad P_t(x,y) := \mathbb{P}_x(X_t = y).$$

By left or right multiplication, the matrix P_t provides a direct access to the marginal law $\mu_t := \mathrm{law}(X_t)$ or the expectation of (bounded or non-negative) functions $f : \mathbb{X} \to \mathbb{R}$ as follows:

$$\mu_t(y) = \mathbb{P}(X_t = y) = \sum_{x\in\mathbb{X}} \mu(x) P_t(x,y) = \mu P_t(y) \tag{2.10}$$

$$\mathbb{E}_x[f(X_t)] = \sum_{y\in\mathbb{X}} P_t(x,y) f(y) = (P_t f)(x). \tag{2.11}$$

In fact, much more is true: for $s,t \geq 0$ and $f : \mathbb{X} \to [0,\infty)$, we have by the Markov property,

$$\mathbb{E}\left[f(X_{s+t}) \,\middle|\, (X_u)_{u\in[0,s]}\right] = P_t f(X_s), \tag{2.12}$$

which, by successive conditioning, provides an access to the finite-dimensional marginals of X: specifically, for every $n \in \mathbb{N}$, $(x_1, \ldots, x_n) \in \mathbb{X}^n$, and $0 = t_0 \leq t_1 \leq t_2 \leq \ldots \leq t_n$, we have

$$\mathbb{P}(X_{t_0} = x_0, X_{t_1} = x_1, \ldots, X_{t_n} = x_n) \quad = \quad \mu(x_0) P_{t_1 - t_0}(x_0, x_1) \cdots P_{t_n - t_{n-1}}(x_{n-1}, x_n). \quad (2.13)$$

Thus, the law of X is determined by μ and $(P_t)_{t \geq 0}$, allowing us to use the alternative notation

$$X \sim \mathrm{CTMC}(\mu, (P_t)_{t \geq 0}).$$

Remark 2.2 (Time-discretization). *If we fix $h > 0$ and choose $t_i = ih$ in (2.13), we obtain that*

$$(X_t)_{t \geq 0} \sim \mathrm{CTMC}(\mu, (P_t)_{t \geq 0}) \quad \Longrightarrow \quad (X_{nh})_{n \in \mathbb{N}} \sim \mathrm{DTMC}(\mu, P_h).$$

In other words, time-discretizations of continuous-time Markov chains are discrete-time Markov chains !

The stochastic matrices $(P_t)_{t \geq 0}$ form what is called a semi-group.

Lemma 2.5 (Semi-group property). *The stochastic matrices $(P_t)_{t \geq 0}$ satisfy the semi-group property*

$$\forall s, t \in \mathbb{R}_+, \qquad P_{s+t} \quad = \quad P_s P_t, \qquad \text{and} \qquad P_0 \quad = \quad \mathrm{Id}. \quad (2.14)$$

Proof. Using (2.12), we have for any $f : \mathbb{X} \to [0, \infty)$, $x \in \mathbb{X}$, and $s, t \geq 0$,

$$\begin{aligned} P_{s+t} f(x) &= \mathbb{E}_x[f(X_{s+t})] \\ &= \mathbb{E}_x\left[\mathbb{E}_x\left[f(X_{s+t}) | (X_u)_{u \in [0,s]}\right]\right] \\ &= \mathbb{E}_x\left[P_t f(X_s)\right] \\ &= P_s P_t f(x), \end{aligned}$$

as desired. On the other hand, $P_0 f(x) = \mathbb{E}_x[f(X_0)] = f(x)$, by definition of $\mathbb{P}_x$. □

Note that the semi-group is entry-wise right-continuous: for each $x, y \in \mathbb{X}$ and $t, h \geq 0$, we can use the right-continuity of X and dominated convergence to write

$$P_{t+h}(x, y) \quad = \quad \mathbb{P}_x(X_{t+h} = y) \quad \xrightarrow[h \to 0+]{} \quad \mathbb{P}_x(X_t = y) \quad = \quad P_t(x, y).$$

In the scalar case, the argument outlined in the proof of Lemma 2.2 shows that any right-continuous function $F : [0, \infty) \to (0, \infty)$ satisfying the "semi-group property" $F(t + s) = F(t)F(s)$ and $F(0) = 1$ must in fact be an exponential of the form $F(t) = e^{qt}$, with a rate which is then necessarily given by the formula $q := \lim_{t \to 0+} \frac{F(t) - F(0)}{t}$. In light of this, it is natural to hope for a similar conclusion in our infinite-dimensional setting, namely

$$\forall t \geq 0, \qquad P_t \quad = \quad e^{tQ}, \qquad \text{where} \qquad Q \quad = \quad \lim_{t \to 0+} \frac{P_t - P_0}{t}. \quad (2.15)$$

Of course, giving an honest mathematical content to this dream is far from trivial:

- Does the limit defining Q exist? in what sense?
- Assuming that Q exists, what is then meaning of e^{tQ}?
- Assuming that e^{tQ} makes sense, why should it equal P_t?

Note, however, than an explicit expression for the cancidate matrix Q appearing in (2.15) can easily be guessed. Indeed, for any bounded function $f : \mathbb{X} \to \mathbb{R}$ and any $x \in \mathbb{X}$, neglecting the unlikely event $\{E_0 + E_1 \leq t\}$ when t is small yields the (non-rigorous) approximation

$$\begin{aligned}(P_t f)(x) - (P_0 f)(x) &= \mathbb{E}_x\left[f(X_t) - f(X_0)\right] \\ &\approx \mathbb{E}_x\left[\mathbf{1}_{(E_0 \leq t)}(f(Y_1) - f(x))\right] \\ &= \left(1 - e^{-tq(x)}\right)(Tf(x) - f(x)).\end{aligned}$$

Dividing by t and letting $t \to 0$, we see that the action of the matrix Q in (2.15) should be given by $Qf(x) = q(x)\left((Tf)(x) - f(x)\right)$ or equivalently,

$$\forall (x,y) \in \mathbb{X}^2, \qquad Q(x,y) := q(x)T(x,y) - q(x)\mathbf{1}_{(y=x)}. \tag{2.16}$$

This explicit matrix $Q \in \mathbb{R}^{\mathbb{X}\times\mathbb{X}}$ is called the rate matrix, or generator of the chain X. Note that it is non-negative off the diagonal, non-positive on the diagonal, and with zero-sum rows. Conversely, any such matrix is of the form (2.16), for example by choosing (q, T) as follows:

$$q(x) := \begin{cases} -Q(x,x) & \text{if } Q(x,x) \neq 0 \\ 1 & \text{if } Q(x,x) = 0 \end{cases} \tag{2.17}$$

$$T(x,y) := \begin{cases} -\frac{Q(x,y)}{Q(x,x)}\mathbf{1}_{(y\neq x)} & \text{if } Q(x,x) \neq 0 \\ \mathbf{1}_{(y=x)} & \text{if } Q(x,x) = 0. \end{cases} \tag{2.18}$$

We will now try to address the issues raised above, and give a rigorous meaning to (2.15).

2.4 Kolmogorov equations

To make our life simpler, we will assume in this section that the jump rates are bounded:

$$q_{\max} := \sup_{x\in\mathbb{X}} q(x) < \infty. \tag{2.19}$$

This fairly reasonable assumption will allow us to give a very satisfactory meaning to (2.15), without much effort. Moreover, a suitable limit procedure will then enable us to address the general case. Let us remark that (2.19) is more than enough to guarantee the non-explosion of (μ, T, q), as requested at the beginning of the section. We start with an easy and classical exercise about bounded linear operators on $L^\infty(\mathbb{X})$ and their operator norm.

Exercise 2.2 (Strong convergence). *Given a matrix $A \in \mathbb{R}^{\mathbb{X}\times\mathbb{X}}$ with summable rows and a bounded function $f : \mathbb{X} \to \mathbb{R}$, we define a function $Af : \mathbb{X} \to \mathbb{R}$ by matrix-vector multiplication as follows:*

$$\forall x \in \mathbb{X}, \qquad (Af)(x) := \sum_{y\in\mathbb{X}} A(x,y)f(y).$$

1. *Show that Af is well-defined, but not necessarily bounded.*
2. *Show that $Af \in L^\infty(\mathbb{X})$ for all $f \in L^\infty(\mathbb{X})$ if and only if $\|A\| < \infty$, where*

$$\|A\| := \sup_{x\in\mathbb{X}} \left\{ \sum_{y\in\mathbb{X}} |A(x,y)| \right\}.$$

3. *Prove that, whenever finite, $\|A\|$ is in fact the operator norm of $A\colon L^\infty \to L^\infty$, i.e.*

$$\|A\| \quad = \quad \sup\left\{\frac{\|Af\|_\infty}{\|f\|_\infty}\colon f \in L^\infty(\mathbb{X}) \setminus \{0\}\right\}.$$

4. *Establish the sub-multiplicativity property $\|AB\| \leq \|A\|\,\|B\|$ for all $A, B \in \mathbb{R}^{\mathbb{X}\times\mathbb{X}}$.*

5. *Prove that $\{A \in \mathbb{R}^{\mathbb{X}\times\mathbb{X}}\colon \|A\| < \infty\}$ equipped with $\|\cdot\|$ is a Banach space, denoted $\mathcal{B}(L^\infty(\mathbb{X}))$.*

6. *Check that the transition and rate matrices satisfy $\|T\| = 1$ and $\|Q\| \leq 2q_{\max}$.*

We will use the term strong convergence for the convergence of matrices in $\mathcal{B}(L^\infty(\mathbb{X}))$, i.e.

$$\|A_n - A\| \quad \xrightarrow[n\to\infty]{} \quad 0, \tag{2.20}$$

in order to distinguish it from the much weaker entrywise convergence

$$\forall (x,y) \in \mathbb{X}^2, \quad A_n(x,y) \quad \xrightarrow[n\to\infty]{} \quad A(x,y). \tag{2.21}$$

This *strong* versus *entrywise* distinction naturally extends to all matrix-theoretical notions that involve convergence, e.g. summability of a matrix-valued sequence $(A_n)_{n\geq 0}$, continuity and differentiability of a matrix-valued function $t \mapsto A_t$, etc. The following result is a (strong) confirmation of our earlier guess that the matrix Q defined at (2.16) is the limit of $\frac{P_t - P_0}{t}$ as $t \to 0$.

Theorem 2.2 (Kolmogorov equations)**.** *Under the bounded-rate assumption (2.19), the semi-group $(P_t)_{t\geq 0}$ is strongly continuously differentiable, with derivative*

$$\forall t \in \mathbb{R}_+, \qquad \frac{\mathrm{d}P_t}{\mathrm{d}t} \quad = \quad QP_t \quad = \quad P_tQ. \tag{2.22}$$

Proof. Fix $f \in L^\infty(\mathbb{X})$, $x \in \mathbb{X}$ and $h > 0$. By definition of $X = \Psi(Y, E)$, we have

$$f(X_h) \quad = \quad \sum_{n=0}^{\infty} f(Y_n)\mathbf{1}_{(E_0+\cdots+E_{n-1}\leq h<E_0+\cdots+E_n)}.$$

In particular, $f(X_h) - f(X_0) = 0$ on $\{E_0 > h\}$, so

$$|P_hf(x) - P_0f(x)| \quad = \quad |\mathbb{E}_x[f(X_h) - f(X_0)]| \quad \leq \quad 2\|f\|_\infty(1 - e^{hq(x)}) \quad \leq \quad 2\|f\|_\infty hq_{\max}.$$

Since this is true for all $x \in \mathbb{X}$ and $f \in L^\infty(\mathbb{X})$, we already see that

$$\|P_h - P_0\| \quad \leq \quad 2hq_{\max},$$

which already shows that $(P_t)_{t\geq 0}$ is strongly continuous at $0+$. To go one step further, observe that $f(X_h) - f(X_0) = (f(Y_1) - f(Y_0))\mathbf{1}_{(E_0\leq h)}$ on the event $\{E_0 + E_1 > h\}$, hence

$$\left|\mathbb{E}_x\left[f(X_h) - f(X_0) - (f(Y_1) - f(Y_0))\mathbf{1}_{(E_0\leq h)}\right]\right| \quad \leq \quad 4\|f\|_\infty \mathbb{P}_x\left(E_0 + E_1 \leq h\right).$$

Now, conditionally on Y, the random variables $E_0, E_1, \ldots$ are independent Exponential with rates $q(Y_0), q(Y_1), \ldots$, and those rates are almost-surely at most $q_{\max}$. Thus,

$$\mathbb{P}_x\left(E_0 + E_1 \leq h\right) \quad \leq \quad \mathbb{P}_x\left(E_0 \leq h, E_1 \leq h\right) \quad \leq \quad (1 - e^{-hq_{\max}})^2 \quad \leq \quad (hq_{\max})^2.$$

On the other hand, recalling that $Qf(x) = q(x)(Tf - f)(x)$, we have

$$\left|\mathbb{E}_x\left[(f(Y_1) - f(Y_0))\mathbf{1}_{(E_0 \le h)}\right] - hQf(x)\right| = \left|\left(1 - e^{-hq(x)} - hq(x)\right)(Tf - f)(x)\right| \le \|f\|_\infty (hq_{\max})^2,$$

thanks to the classical estimate $|1 - e^{-u} - u| \le u^2/2$ for $u \ge 0$. Recalling that $(P_h f)(x) - (P_0 f)(x) = \mathbb{E}_x[f(X_h) - f(X_0)]$ and putting things together, we obtain

$$|(P_h f)(x) - (P_0 f)(x) - hQf(x)| \le 5(hq_{\max})^2 \|f\|_\infty.$$

Since this is true for all $x \in \mathbb{X}$ and $f \in L^\infty(\mathbb{X})$, we conclude that

$$\|P_h - P_0 - hQ\| \le 5(hq_{\max})^2.$$

Thus, $t \mapsto P_t$ is strongly differentiable at $0+$, with derivative Q. Now using the semi-group property, the sub-multiplicativity of $\|\cdot\|$ and $\|P_t\| = 1$, we can write

$$\|P_{t+h} - P_t - hQP_t\| = \|(P_h - P_0 - hQ)P_t\| \le 5(hq_{\max})^2.$$

Assuming that $h \le t$, we may further replace t with $t - h$ above to also obtain

$$\|P_{t-h} - P_t + hQP_{t-h}\| \le 5(hq_{\max})^2,$$

and the term hQP_{t-h} can further be replaced with hQP_t at the extra cost

$$\|hQP_{t-h} - hQP_t\| \le h\|Q\|\|P_{t-h}\|\|P_h - P_0\| \le 4(hq_{\max})^2. \tag{2.23}$$

In conclusion, we have proved that for each $t \ge 0$ and all $h \in [-t, \infty)$,

$$\|P_{t+h} - P_t + hQP_t\| \le 9(hq_{\max})^2.$$

Thus, $t \mapsto P_t$ is strongly differentiable everywhere, with

$$\forall t \in \mathbb{R}_+, \qquad \frac{\mathrm{d}P_t}{\mathrm{d}t} = QP_t.$$

Finally, the same argument with $P_{t+h} - P_t - hP_tQ = P_t(P_h - P_0 - hQ)$ yields

$$\forall t \in \mathbb{R}_+, \qquad \frac{\mathrm{d}P_t}{\mathrm{d}t} = P_tQ,$$

as desired. □

Remark 2.3 (Practical interest). *In view of (2.10), the forward Kolomogorov equation $\frac{\mathrm{d}P_t}{\mathrm{d}t} = P_tQ$ implies that the marginal law $\mu_t := \mathrm{law}(X_t)$ evolves as follows:*

$$\forall t \ge 0, \qquad \frac{\mathrm{d}\mu_t}{\mathrm{d}t} = \mu_tQ. \tag{2.24}$$

Dually, in view of (2.11), the backward Kolomogorov equation $\frac{\mathrm{d}P_t}{\mathrm{d}t} = QP_t$ implies that expectations of the form $f_t(x) := \mathbb{E}_x[f(X_t)]$ ($f \in L^\infty(\mathbb{X}), x \in \mathbb{X}$) evolve as follows:

$$\forall t \ge 0, \qquad \frac{\mathrm{d}f_t}{\mathrm{d}t} = Qf_t. \tag{2.25}$$

Exercise 2.3 (Poisson process). *Let X be a Poisson process of rate $\lambda \in (0, \infty)$. Write down the Kolmogorov forward equation, and solve it to conclude that*

$$\forall t \geq 0, \qquad X_t \quad \sim \quad \text{Poisson}(\lambda t). \tag{2.26}$$

Give an alternative proof of this using the Kolmogorov backward equation, with $f(x) = e^{-\theta x}$, $\theta \geq 0$.

In the scalar case, the differential equation $\frac{\mathrm{d}p_t}{\mathrm{d}t} = qp_t$ is classically uniquely solved by the exponential function $p(t) = p(0)e^{qt}$. This turns out to admit a matrix-valued extension, as we now show. The exponential of a matrix $A \in \mathcal{B}(L^\infty(\mathbb{X}))$ is defined as follows:

$$e^A \quad := \quad \sum_{n=0}^{\infty} \frac{A^n}{n!}. \tag{2.27}$$

The sub-multiplicativity of $\|\cdot\|$ ensures that this series is strongly convergent, and that

$$\left\| e^A \right\| \quad \leq \quad \sum_{n=0}^{\infty} \frac{\|A\|^n}{n!} \quad = \quad e^{\|A\|}. \tag{2.28}$$

Many familiar properties of the scalar exponential extend to the matrix case. Here is an example.

Lemma 2.6 (Matrix exponential). *If $A, B \in \mathcal{B}(L^\infty(\mathbb{X}))$ commute, then*

$$e^{A+B} \quad = \quad e^A e^B. \tag{2.29}$$

In particular, e^A is always invertible, with inverse e^{-A}.

Proof. The commutation property $AB = BA$ ensures the validity of Newton's formula:

$$\forall n \in \mathbb{N}, \qquad (A+B)^n \quad = \quad \sum_{k=0}^{n} \binom{n}{k} A^k B^{n-k}.$$

Dividing through by $n!$ and summing over all $n \in \mathbb{N}$, we obtain

$$\begin{aligned} e^{A+B} \quad &= \quad \sum_{n=0}^{\infty} \frac{(A+B)^n}{n!} \\ &= \quad \sum_{n=0}^{\infty} \sum_{k=0}^{n} \frac{A^k B^{n-k}}{k!(n-k)!} \\ &= \quad \sum_{k=0}^{\infty} \frac{A^k}{k!} \sum_{n=k}^{\infty} \frac{B^{n-k}}{(n-k)!} \\ &= \quad e^A e^B. \end{aligned}$$

In particular, choosing $B = -A$ yields $e^A e^{-A} = \mathrm{Id} = e^{-A} e^A$, hence the second point. □

Theorem 2.3 (Resolution of Kolmogorov equations). *Fix A and $(H_t)_{t \geq 0}$ in $\mathcal{B}(L^\infty(\mathbb{X}))$. Then,*

1. *$(H_t)_{t \geq 0}$ is strongly differentiable with $\frac{\mathrm{d}H_t}{\mathrm{d}t} = H_t A$ if and only if $H_t = H_0 e^{tA}$ for all $t \geq 0$.*
2. *$(H_t)_{t \geq 0}$ is strongly differentiable with $\frac{\mathrm{d}H_t}{\mathrm{d}t} = A H_t$ if and only if $H_t = e^{tA} H_0$ for all $t \geq 0$.*

Proof. We focus on the first assertion, as the second is proved in a similar way. Suppose that

$$\forall t \geq 0, \qquad H_t \;=\; H_0 e^{tA},$$

and let us extend the definition of H_t to $t < 0$ by the same formula. Then, for any $t, h \in \mathbb{R}$, the properties of matrix exponentials allow us to write

$$\|H_{t+h} - H_t - hH_tA\| \;=\; \left\| H_t \left(\sum_{n=2}^{\infty} \frac{(hA)^n}{n!} \right) \right\| \;\leq\; \|H_t\| \left(e^{|h|\|A\|} - 1 - |h|\|A\| \right).$$

For each $t \in \mathbb{R}$, the right-hand side is $o(|h|)$ as $h \to 0$, so $(H_t)_{t\in\mathbb{R}}$ is strongly differentiable and

$$\forall t \in \mathbb{R}, \quad \frac{\mathrm{d}H_t}{\mathrm{d}t} \;=\; H_tA,$$

which is more than needed. Conversely, suppose that $(H_t)_{t\geq 0}$ is strongly differentiable and satisfies $\frac{\mathrm{d}H_t}{\mathrm{d}t} = H_tA$, and consider the matrices $(M_t)_{t\geq 0}$ defined as follows:

$$\forall t \geq 0, \qquad M_t \;:=\; H_t e^{-tA}.$$

Being a product of strongly differentiable functions, $(M_t)_{t\geq 0}$ is strongly differentiable and

$$\frac{\mathrm{d}M_t}{\mathrm{d}t} \;=\; \left(\frac{\mathrm{d}H_t}{\mathrm{d}t}\right) e^{-tA} + H_t \left(\frac{\mathrm{d}e^{-tA}}{\mathrm{d}t}\right) \;=\; H_tAe^{-tA} - H_tAe^{-tA} \;=\; 0.$$

Thus, $M_t = M_0$ or equivalently $H_t = H_0 e^{tA}$, for all $t \geq 0$. □

Corollary 2.2 (Determination of the semi-group)**.** *Under the assumption $q_{\max} < \infty$, we have*

$$\forall t \geq 0, \qquad P_t \;=\; e^{tQ}.$$

Exercise 2.4 (Poisson semi-group)**.** *Explicitate $(P_t)_{t\geq 0}$ when X is a Poisson process of rate $\lambda \in (0, \infty)$.*

2.5 Extension to unbounded rates

We now address the more delicate case where q is unbounded. We will see that the conclusions obtained in the previous section remain essentially valid, albeit in a weaker, entry-wise sense.

Theorem 2.4 (Backward Kolmogorov equation)**.** *The semi-group $(P_t)_{t\geq 0}$ is the only entry-wise differentiable family of stochastic matrices solving the backward Kolmogorov equation*

$$P_0 \;=\; \mathrm{Id}, \qquad \textit{and} \qquad \forall t \geq 0, \qquad \frac{\mathrm{d}P_t}{\mathrm{d}t} \;=\; QP_t.$$

We will use the following simple observation, which will be considerably generalized later.

Lemma 2.7 (Strong Markov property at the first jump time)**.** *Let $X = \Psi(Y, E)$ be a continuous-time Markov chains with parameters (μ, T, q). Then,*

$$\mathrm{law}\left((X_{t+E_0})_{t\geq 0} \,|\, E_0, Y_0, Y_1\right) \;=\; \mathrm{CTMC}(\delta_{Y_1}, T, q).$$

Proof. The properties of (Y, E) listed in Definition 2.1 readily imply that the shifted sequences $\widetilde{Y} := (Y_1, Y_2, \dots)$ and $\widetilde{E} := (E_1, E_2, \dots)$ satisfy

$$\begin{aligned} \operatorname{law}(\widetilde{Y} \,|\, E_0, Y_0, Y_1) &= \operatorname{DTMC}(\delta_{Y_1}, T) \\ \operatorname{law}(\widetilde{E} \,|\, E_0, Y) &= \bigotimes_{i \in \mathbb{N}} \operatorname{Exp}\left(q(\widetilde{Y}_i)\right) \\ \sum_{i=0}^{\infty} \widetilde{E}_i &= \sum_{i=1}^{\infty} E_i \;<\; \infty. \end{aligned}$$

This establises the claim, since $\Psi(\widetilde{Y}, \widetilde{E}) = (X_{t+E_0})_{t \geq 0}$ by construction. □

Proof of Theorem 2.4. Fix $t \geq 0$ and $(x, y) \in \mathbb{X}^2$. Using the above lemma, we can compute $P_t(x, y) = \mathbb{P}_x(X_t = y)$ by conditionning on (E_0, Y_1) as follows:

$$\begin{aligned} P_t(x,y) &= e^{-tq(x)} \mathbf{1}_{(x=y)} + \int_0^t \sum_{z \in \mathbb{X}} T(x,z) \mathbb{P}_z(X_{t-u} = y) q(x) e^{-uq(x)} \, \mathrm{d}u \\ &= e^{-tq(x)} \mathbf{1}_{(x=y)} + \int_0^t (TP_{t-u})(x,y) q(x) e^{-uq(x)} \, \mathrm{d}u. \end{aligned}$$

Multiplying through by $e^{tq(x)}$ and performing the change of variables $u \mapsto t - u$, we obtain

$$e^{tq(x)} P_t(x,y) = \mathbf{1}_{(x=y)} + \int_0^t (TP_u)(x,y) q(x) e^{uq(x)} \, \mathrm{d}u.$$

This shows that $t \mapsto P_t(x, y)$ is continuously differentiable, and differentiating both sides gives

$$\frac{\mathrm{d}P_t(x,y)}{\mathrm{d}t} + q(x) P_t(x,y) = q(x) T P_t(x,y),$$

i.e. $\frac{\mathrm{d}P_t}{\mathrm{d}t} = QP_t$. Conversely, if $(\widehat{P}_t)_{t \geq 0}$ is an entry-wise differentiable family of stochastic matrices solving the backward equation, then reversing the above computation gives

$$\widehat{P}_t(x,y) = e^{-tq(x)} \mathbf{1}_{(x=y)} + \int_0^t (T\widehat{P}_{t-u})(x,y) q(x) e^{-uq(x)} \, \mathrm{d}u.$$

On the other hand, writing $P_t^{(n)}(x, y) := \mathbb{P}_x(X_t = y, \sum_{i=0}^{n-1} E_i > t)$ and conditionning on (E_0, Y_1) as above yields for $n \geq 1$,

$$P_t^{(n)}(x,y) = e^{-tq(x)} \mathbf{1}_{(x=y)} + \int_0^t (TP_u^{(n-1)})(x,y) q(x) e^{-uq(x)} \, \mathrm{d}u.$$

Comparing the last two displays, we conclude by induction that for all $n \in \mathbb{N}$,

$$\widehat{P}_t(x,y) \geq P_t^{(n)}(x,y).$$

Sending $n \to \infty$ shows that $\widehat{P}_t \geq P_t$, and equality must hold because both sides are stochastic. □

The forward Kolmogorov equation requires a bit more effort, as we need to condition on the last step instead of the first. To this end, the following time-reversal identity will be useful.

Lemma 2.8 (Time-reversal identity). *Writing $S_n =: E_0 + \cdots + E_{n-1}$ (with $S_0 := 0$), we have*

$$q(y_n) \mathbb{P}(S_n \leq t < S_{n+1} \,|\, Y_0 = y_0, \dots, Y_n = y_n) = q(y_0) \mathbb{P}(S_n \leq t < S_{n+1} \,|\, Y_n = y_0, \dots, Y_0 = y_n),$$

for every $t \in \mathbb{R}_+$, every $n \in \mathbb{N}$, and every $(y_0, \dots, y_n) \in \mathbb{X}^{n+1}$.

Proof. Writing $A = \{(u_0, \dots, u_{n-1}) \in \mathbb{R}_+^n : u_0 + \cdots + u_{n-1} \le t\}$, we have by construction

$$q(Y_n)\mathbb{P}(S_n \le t < S_{n+1} \,|\, Y_0, \dots, Y_n) \;=\; \left(\prod_{i=0}^{n} q(Y_i)\right) \int_A e^{-q(Y_n)(t-\sum_{i=0}^{n-1} u_i)} \prod_{i=0}^{n-1} e^{-q(Y_i)u_i}\, \mathrm{d}u_i,$$

and the involutive change of variables $(u_0, \dots, u_{n-1}) \mapsto (t - \sum_{i=0}^{n-1} u_i, u_{n-1}, \dots, u_1)$ shows that the right-hand side is invariant under replacing $(Y_0, \dots, Y_n)$ with $(Y_n, \cdots, Y_0)$. □

Theorem 2.5 (Forward Kolmogorov equation)**.** *The semi-group $(P_t)_{t\ge 0}$ is the only entry-wise differentiable family of stochastic matrices solving the forward Kolmogorov equation*

$$P_0 \;=\; \mathrm{Id}, \qquad \textit{and} \qquad \forall t \ge 0, \qquad \frac{\mathrm{d}P_t}{\mathrm{d}t} \;=\; P_t Q.$$

Proof. Write $X = \Psi(Y, E)$ with (Y, E) as in Definition 2.1, and set $S_n := E_0 + \cdots + E_{n-1}$, with $S_0 = 0$. Then, for each $t \ge 0$ and $(x, y) \in \mathbb{X}^2$, we have by construction

$$\begin{aligned} P_t(x,y) \;&=\; \sum_{n=0}^{\infty} \mathbb{P}_x(S_n \le t < S_{n+1}, Y_n = y) \\ &=\; e^{-q(x)t}\mathbf{1}_{(x=y)} + \sum_{z\in\mathbb{X}} \sum_{n=1}^{\infty} \mathbb{P}_x(S_n \le t < S_{n+1}, Y_{n-1} = z, Y_n = y). \end{aligned}$$

Now, using the two previous lemmas, we have for each $n \ge 1$ and $z \in \mathbb{X}$,

$$\begin{aligned} &\mathbb{P}_x(S_n \le t < S_{n+1} \,|\, Y_{n-1} = z, Y_n = y) \\ &= \frac{q(x)}{q(y)} \mathbb{P}_y(S_n \le t < S_{n+1} \,|\, Y_1 = z, Y_n = x) \\ &= q(x) \int_0^t e^{-uq(y)} \mathbb{P}_z(S_{n-1} \le t - u < S_n \,|\, Y_{n-1} = x)\, \mathrm{d}u \\ &= q(z) \int_0^t e^{-uq(y)} \mathbb{P}_x(S_{n-1} \le t - u < S_n \,|\, Y_{n-1} = z)\, \mathrm{d}u. \end{aligned}$$

Reinserting this back into the previous computation, we obtain

$$\begin{aligned} P_t(x,y) \;&=\; e^{-q(x)t}\mathbf{1}_{(x=y)} + \sum_{z\in\mathbb{X}} q(z)T(z,y) \int_0^t e^{-uq(y)} \sum_{n=1}^{\infty} \mathbb{P}_x(S_{n-1} \le t - u < S_n, Y_{n-1} = z)\, \mathrm{d}u \\ &=\; e^{-q(y)t}\mathbf{1}_{(x=y)} + \sum_{z\in\mathbb{X}} q(z)T(z,y) \int_0^t e^{-uq(y)} P_{t-u}(x,z)\, \mathrm{d}u \\ &=\; e^{-q(y)t}\mathbf{1}_{(x=y)} + e^{-tq(y)} \sum_{z\in\mathbb{X}} q(z)T(z,y) \int_0^t e^{uq(y)} P_u(x,z)\, \mathrm{d}u. \end{aligned}$$

Multiplying through by $e^{tq(y)}$, we arrive at

$$e^{tq(y)} P_t(x,y) \;=\; \mathbf{1}_{(x=y)} + \sum_{z\in\mathbb{X}} q(z)T(z,y) \int_0^t e^{uq(y)} P_u(x,z)\, \mathrm{d}u.$$

This shows that $(P_t)_{t\ge 0}$ is entry-wise differentiable (which we already knew), and that

$$\frac{\mathrm{d}P_t(x,y)}{\mathrm{d}t} + P_t(x,y)q(y) \;=\; \sum_{z\in\mathbb{X}} P_t(x,z)q(z)T(z,y).$$

This is the forward equation. Conversely, if $(\widehat{P}_t)_{t\geq 0}$ is an entry-wise differentiable family of stochastic matrices solving the forward equation, then reversing the above computation gives

$$\widehat{P}_t(x,y) \quad = \quad e^{-tq(x)}\mathbf{1}_{(x=y)} + \sum_{z\in\mathbb{X}} q(z)T(z,y)\int_0^t e^{-uq(y)}\widehat{P}_{t-u}(x,z)\,\mathrm{d}u.$$

On the other hand, by a straightforward variation of the above computation, the quantity $P_t^{(n)}(x,y) := \mathbb{P}_x(X_t = y, S_n > t)$ satisfies the recursion

$$P_t^{(n)}(x,y) \quad = \quad e^{-tq(x)}\mathbf{1}_{(x=y)} + \sum_{z\in\mathbb{X}} q(z)T(z,y)\int_0^t e^{-uq(y)}P_{t-u}^{(n-1)}(x,z)\,\mathrm{d}u.$$

Comparing the last two displays, we conclude by induction that for all $n \in \mathbb{N}$,

$$\widehat{P}_t(x,y) \quad \geq \quad P_t^{(n)}(x,y).$$

Sending $n \to \infty$ shows that $\widehat{P}_t \geq P_t$, and equality must hold because both sides are stochastic. □

Remark 2.4 (Uniqueness). *The forward and backward Kolmogorov equations show that the semi-group $(P_t)_{t\geq 0}$ is determined (or generated) by the rate matrix Q. In view of (2.13), we deduce that the law of X is determined by the pair (μ, Q), allowing us to write $X \sim \mathrm{CTMC}(\mu, Q)$. Conversely, starting from a matrix $Q \in \mathbb{R}^{\mathbb{X}\times\mathbb{X}}$ with non-negative off-diagonal entries and zero-sum rows, we may set (q, T) as in (2.17-2.18) and then consider a discrete-time Markov chain $Y = (Y_0, Y_1, \ldots)$ with transition matrix T and, conditionally on Y, a sequence $E = (E_0, E_1, \ldots)$ of independent Exponential random variables with rates $q(Y_0), q(Y_1), \ldots$. A careful inspection of the above proofs shows that the matrices $(P_t)_{t\geq 0}$ defined by*

$$P_t(x,y) \quad := \quad \sum_{n=0}^{\infty} \mathbb{P}_x\left(Y_n = y, \sum_{i=0}^{n-1} E_i \leq t < \sum_{i=0}^{n} E_i\right),$$

form then the minimal non-negative solution to both the backward and the forward Kolmogorov equation generated by Q. Clearly, the matrices $(P_t)_{t\geq 0}$ thus defined are stochastic if and only if (δ_x, T, q) is non-explosive for every $x \in \mathbb{X}$. When this is the case, we say that Q is non-explosive. We can then safely set $X := \Psi(Y, E)$, and we have $P_t(x,y) = \mathbb{P}_x(X_t = y)$ by construction.

2.6 Filtrations, stopping times and martingales

In this section, we assume that our probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is equipped with a continuous-time filtration $(\mathcal{F}_t)_{t\geq 0}$, i.e. a non-decreasing family of sub-$\sigma-$algebras of $\mathcal{F}$. A continuous-time stochastic process $X = (X_t)_{t\geq 0}$ is called adapted if for each $t \geq 0$, X_t is $\mathcal{F}_t-$measurable. We can then refine our earlier definition of a continuous-time Markov chain as follows. As usual, $Q \in \mathbb{R}^{\mathbb{X}\times\mathbb{X}}$ denotes a non-explosive rate matrix, and $(P_t)_{t\geq 0}$ the associated semi-group.

Definition 2.2 (Continuous-time Markov chain, abstract definition). *A stochastic process $X = (X_t)_{t\geq 0}$ is called a continuous-time Markov chain with generator Q if it is càdlàg, adapted, and*

$$\mathbb{E}\left[f(X_{t+s}) \,|\, \mathcal{F}_s\right] \quad = \quad P_t f(X_s), \tag{2.30}$$

for all times $s, t \geq 0$ and all functions $f \in L^\infty(\mathbb{X})$, where $(P_t)_{t\geq 0}$ is the semi-group generated by Q.

Note that by (2.12), any Markov chain in the sense of Definition 2.1 is a Markov chain in the sense of Definition 2.2 w.r.t. its natural filtration, defined as follows:

$$\mathcal{F}_s \;:=\; \sigma\left(X_u \colon u \in [0,s]\right).$$

Conversely, if X satisfies Definition 2.2, then by successive conditioning, we have

$$\mathbb{P}(X_0 = x_0, X_{t_1} = x_1, \ldots, X_{t_n} = x_n) \;=\; \mu(x_0)P_{t_1}(x_0,x_1)P_{t_2-t_1}(x_1,x_2)\cdots P_{t_n-t_{n-1}}(x_{n-1},x_n),$$

for every $n \in \mathbb{N}$, $(x_0,\ldots,x_n) \in \mathbb{X}^{n+1}$ and $0 \le t_1 \le \ldots \le t_n$, where $\mu := \mathrm{law}(X_0)$. Thus, $X \sim \mathrm{CTMC}(\mu, Q)$. Furthermore, we can upgrade this distributional correspondance into an almost-sure one under the mild assumption that $Q(x,x) \neq 0$ for all $x \in \mathbb{X}$. Indeed, our càdlàg process X then almost-surely makes infinitely many jumps, allowing us to write $X = \Psi(Y,E)$ where $Y = (Y_0, Y_1, \ldots)$ and $E = (E_0, E_1, \ldots)$ are inductively defined as follows:

$$\begin{aligned} E_n &:= \inf\left\{t \ge 0 \colon X_{t+E_0+\cdots+E_{n-1}} \neq X_{E_0+\cdots+E_{n-1}}\right\} \\ Y_n &:= X_{E_0+\cdots+E_{n-1}}, \end{aligned}$$

with the usual understanding that an empty sum is zero. Being a measurable function of X, the random pair (Y,E) thus defined must then have the same law as if X had really been constructed according to Definition 2.1, with $T(x,y) := -\frac{Q(x,y)}{Q(x,x)}\mathbf{1}_{(x\neq y)}$ and $q(x) := -Q(x,x)$. In other words,

$$\mathrm{law}(Y) \;=\; \mathrm{DTMC}(\mu,T), \quad \text{and} \quad \mathrm{law}(E \,|\, Y) \;=\; \bigotimes_{i=0}^{\infty} q(Y_i),$$

showing that X is really a continuous-time Markov chain in the sense of our original definition. Thus, the two viewpoints are equivalent. Let us now describe a closely related, third approach, which has two advantages. First, it only involves the rate-matrix Q, without any mention of the associated semi-group $(P_t)_{t\ge0}$. Second, it provides us with a concrete family of martingales that constitute a powerful toolbox for the concrete analysis of the considered Markov chain.

Definition 2.3 (Martingale). *A martingale is an adapted, integrable stochastic process $(M_t)_{t\ge0}$ such that $\mathbb{E}[M_{t+s} \,|\, \mathcal{F}_s] = M_s$ almost-surely, for every $t,s \ge 0$.*

In addition to the trivial observation that their mean is constant, martingales enjoy a variety of nice properties, including maximal inequalities, concentration inequalities, optional stopping theorems, representation theorems, and convergence theorems. We do not have enough time for an exhaustive account, unfortunately. The connection between Markov chains and martingales is summarized in the following fundamental result, which can be taken as an alternative definition.

Theorem 2.6 (Martingale characterization of Markov chains). *A càdlàg stochastic process $(X_t)_{t\ge0}$ is a Markov chain with generator Q if and only if the process $(M^f_t)_{t\ge0}$ defined by*

$$\forall t \ge 0, \qquad M^f_t \;:=\; f(X_t) - f(X_0) - \int_0^t (Qf)(X_u)\,\mathrm{d}u, \tag{2.31}$$

is a martingale for each finitely supported function $f \colon \mathbb{X} \to \mathbb{R}$.

Proof. The martingale property of M^f is equivalent to the identity

$$\forall t \ge 0, \qquad \mathbb{E}\left[\left(f(X_{t+s}) - f(X_s) - \int_0^t (Qf)(X_{u+s})\,\mathrm{d}u\right) \,\middle|\, A \cap \{X_s = x\}\right] \;=\; 0,$$

for every choice of $s \geq 0$, $x \in \mathbb{X}$ and $A \in \mathcal{F}_s$ such that $\mathbb{P}(A \cap \{X_s = x\}) > 0$. In other words, for any such triple (s, x, A), we want the collection of probability vectors $(\mu_t)_{t\geq 0}$ defined by $\mu_t := \text{law}(X_{t+s} \,|\, A \cap \{X_s = x\})$ to solve the forward Kolmogorov equation

$$\forall t \geq 0, \qquad \mu_t \;=\; \delta_x + \int_0^t \mu_u Q \,\mathrm{d}u.$$

In view of Theorem 2.5, this equation is uniquely solved by $\mu_t = P_t(x, \cdot)$, i.e.

$$\forall t \geq 0, \qquad \text{law}\,(X_{t+s} \,|\, A \cap \{X_s = x\}) \;=\; P_t(x, \cdot).$$

The fact that this must hold for all $x \in \mathbb{X}$, $A \in \mathcal{F}_s$ such that $\mathbb{P}(A \cap \{X_s = x\}) > 0$ means that

$$\forall t \geq 0, \qquad \text{law}\,(X_{t+s} \,|\, \mathcal{F}_s) \;=\; P_t(X_s, \cdot),$$

which exactly says that X is a continuous-time Markov chain with generator Q. □

Remark 2.5 (Extension). *Note that the formula (2.31) actually defines a (càdlàg and adapted) stochastic process M^f for every function $f \in L^\infty(\mathbb{X})$, regardless of its support. When Q has bounded entries, this process is integrable and it inherits the martingale property from the finitely-supported case, thanks to the uniform bound $|M_t^f| \leq 2\|f\|_\infty + t\|Q\|\|f\|_\infty$. In the general case, M^f needs not be integrable, but it is always a local martingale, as witnessed by the localizing sequence $(\tau_n)_{n\in\mathbb{N}}$ defined as follows:*

$$\forall n \in \mathbb{N}, \qquad \tau_n \;:=\; \inf\{t \geq 0 \colon q(X_t) > n\}. \tag{2.32}$$

We leave the details to the interested reader.

Definition 2.4 (Stopping time). *A stopping time is a function $\tau \colon \Omega \to [0, \infty]$ such that*

$$\forall t \geq 0, \qquad \{\tau \leq t\} \in \mathcal{F}_t,$$

and its past $\sigma-$field is $\mathcal{F}_\tau := \{A \in \mathcal{F} \colon \forall t \geq 0,\ A \cap \{\tau \leq t\} \in \mathcal{F}_t\}$.

This is the natural continuous-time analogue of the definition given in the previous chapter. It carries the same intuition and enjoys similar properties, summarized in the following exercise.

Exercise 2.5 (Properties of stopping times). *Establish the following properties.*

1. *The past $\sigma-$field of a stopping time is a $\sigma-$field.*
2. *The minimum, maximum and sum of two stopping times are stopping times.*
3. *If τ, τ' are two stopping times with $\tau \leq \tau'$, then $\mathcal{F}_\tau \subseteq \mathcal{F}_{\tau'}$.*
4. *For any fixed $t \geq 0$, the deterministic function $\tau \colon \omega \mapsto t$ is a stopping time, and $\mathcal{F}_\tau = \mathcal{F}_t$.*
5. *If τ is a stopping time, then so is $\tau_n := \lceil n\tau \rceil / n$, for every integer $n \geq 1$.*
6. *If τ is a stopping time, then τ is $\mathcal{F}_\tau-$measurable. Moreover, so is X_τ provided $\tau < \infty$, and X is a $\mathbb{X}-$valued, right-continuous, adapted process.*
7. *The hitting time of any set by a $\mathbb{X}-$valued, right-continuous adapted process is a stopping time.*

We are now ready to state and prove the following substantial generalization of Theorem 2.1.

Theorem 2.7 (Strong Markov property)**.** *Let X be a continuous-time Markov chain with generator Q and let τ be a stopping time. Then, almost-surely on the event $\{\tau < \infty\}$, we have*

$$\text{law}\left((X_{t+\tau})_{t\geq 0} \,|\, \mathcal{F}_\tau\right) \quad = \quad \text{CTMC}(\delta_{X_\tau}, Q).$$

Proof. By successive conditioning, it is enough to prove that on $\{\tau < \infty\}$,

$$\mathbb{P}\left(X_{t+\tau} = y \,|\, \mathcal{F}_\tau\right) \quad = \quad P_t(X_\tau, y),$$

for all $y \in \mathbb{X}$ and $t \geq 0$. In other words, given $x, y \in \mathbb{X}$, $t \geq 0$ and $A \in \mathcal{F}_\tau$, we want to show that

$$\mathbb{P}\left(A \cap \{\tau < \infty, X_\tau = x, X_{t+\tau} = y\}\right) \quad = \quad \mathbb{P}\left(A \cap \{\tau < \infty, X_\tau = x\}\right) P_t(x, y).$$

This is easily done if τ takes values in $\mathbb{Q}_+ \cup \{\infty\}$: indeed, we can then write

$$\begin{aligned}
\mathbb{P}\left(A \cap \{\tau < \infty, X_\tau = x, X_{t+\tau} = y\}\right) &= \sum_{s\in\mathbb{Q}_+} \mathbb{P}\left(A \cap \{\tau = s, X_s = x, X_{t+s} = y\}\right) \\
&= \sum_{s\in\mathbb{Q}_+} \mathbb{P}\left(A \cap \{\tau = s, X_s = x\}\right) P_t(x, y) \\
&= \mathbb{P}\left(A \cap \{\tau < \infty, X_\tau = x\}\right) P_t(x, y),
\end{aligned}$$

where in the second line, we have used Definition 2.2 with $f = \mathbf{1}_{\{y\}}$, and the observation that $A \cap \{\tau = s, X_s = x\}$ is in $\mathcal{F}_s$. Now, in the general case, we fix $n \geq 1$ and define

$$\tau_n \quad := \quad \frac{\lceil n\tau \rceil}{n},$$

with $\lceil \infty \rceil = \infty$. Note that τ_n is a stopping time, since for every $s \geq 0$ we have by definition

$$\{\tau_n \leq s\} \quad = \quad \left\{\tau \leq \frac{\lfloor ns \rfloor}{n}\right\} \in \mathcal{F}_{\frac{\lfloor ns \rfloor}{n}} \subseteq \mathcal{F}_s.$$

Since τ_n is $\mathbb{Q}_+ \cup \{\infty\}-$valued, and since $A \in \mathcal{F}_\tau \subseteq \mathcal{F}_{\tau_n}$, the first part of the proof ensures that

$$\mathbb{P}\left(A \cap \{\tau_n < \infty, X_{\tau_n} = x, X_{t+\tau_n} = y\}\right) \quad = \quad \mathbb{P}\left(A \cap \{\tau_n < \infty, X_{\tau_n} = x\}\right) P_t(x, y).$$

The conclusion now follows by sending $n \to \infty$, because $\tau_n \to \tau+$ and X is right-continuous. □

2.7 Recurrence, transience and invariant laws

As usual, we let X denote a continuous-time Markov chain with parameters (μ, T, q). In this section, we use the representation $X = \Psi(Y, E)$ to effortlessly transfer the various trajectorial properties of discrete-time chains established in Chapter 1 to the continuous-time setting. Let

$$\tau_x(X) := \quad \inf\{t \geq 0 \colon X_t = x\}, \tag{2.33}$$

denote the hitting time of the state $x \in \mathbb{X}$ by X. The following lemma shows that the communication relations have the same interpretation in discrete and continuous time.

Lemma 2.9 (Communication relations)**.** *For any $x, y \in \mathbb{X}$, the following conditions are equivalent:*

1. $\exists n \in \mathbb{N}, T^n(x, y) > 0.$
2. $\exists t > 0, \mathbb{P}_x(X_t = y) > 0.$

3. $\forall t > 0, \mathbb{P}_x(X_t = y) > 0$.
4. $\mathbb{P}_x(\tau_y(X) < \infty) > 0$.

Proof. By construction, the process $X = \Psi(Y, E)$ satisfies, for every $t > 0$ and $x, y \in \mathbb{X}$,

$$\mathbb{P}_x(X_t = y) \quad = \quad \sum_{n=0}^{\infty} \mathbb{P}_x \left(Y_n = y, \sum_{i=0}^{n-1} E_i \leq t < \sum_{i=0}^{n} E_i \right).$$

Moreover, $\mathbb{P}_x \left(\sum_{i=0}^{n-1} E_i \leq t < \sum_{i=0}^{n} E_i \,|\, Y \right) > 0$ almost-surely for each $n \in \mathbb{N}$ and $t > 0$, by property of independent exponentials. It follows that $\mathbb{P}_x(X_t = y) > 0$ if and only if $\mathbb{P}_x(Y_n = y) > 0$ for some $n \in \mathbb{N}$. Recalling that $\mathbb{P}_x(Y_n = y) = T^n(x, y)$, the equivalence between the first three conditions is proved. For the last one, observe that by construction,

$$\{\tau_y(X) < \infty\} \quad = \quad \bigcup_{n=0}^{\infty} \{Y_n = y\},$$

and recall that a countable union of events has zero probability if and only if each event has. □

We next turn to recurrence and transience. The return time to $x \in \mathbb{X}$ by $X = \Psi(Y, E)$ is

$$\tau_x^+(X) := \quad \inf\{t \geq E_0 \colon X_t = x\},$$

and the return probability is $p(x) := \mathbb{P}_x(\tau_x^+(X) < \infty)$. We then naturally declare x recurrent if $p(x) = 1$, and transient otherwise. Since $\{\tau_x^+(X) < \infty\} = \{\tau_x^+(Y) < \infty\}$, this classification is exactly the same as for the discrete-time chain Y. In particular, recurrence and transience are class properties. To go further, let us introduce the local time spent by X at x:

$$L_x(X) \quad := \quad \int_0^{\infty} \mathbf{1}_{(X_t = x)} \, \mathrm{d}t.$$

The next result shows that this quantity satisfies the same dichotomy as in the discrete-time case.

Theorem 2.8 (Recurrence/transience dichotomy)**.** *The following dichotomy holds:*

1. *We have $\mathbb{P}_x(L_x(X) = \infty) = 1$ for every recurrent state $x \in \mathbb{X}$.*
2. *We have $\mathbb{E}_x[L_x(X)] = \frac{1}{(1-p(x))q(x)} < \infty$ for every transient state $x \in \mathbb{X}$.*
3. *If the chain is irreducible and recurrent, then*

$$\forall \mu \in \mathcal{P}(\mathbb{X}), \qquad \mathbb{P}_\mu \left(\forall x \in \mathbb{X}, L_x(X) = \infty \right) \quad = \quad 1.$$

4. *If the chain in irreducible and transient, then*

$$\forall \mu \in \mathcal{P}(\mathbb{X}), \qquad \mathbb{P}_\mu \left(\forall x \in \mathbb{X}, L_x(X) < \infty \right) \quad = \quad 1.$$

Proof. The relation $X = \Psi(Y, E)$, together with Fubini's theorem, allows one to write

$$L_x(X) \quad = \quad \int_0^{\infty} \sum_{n=0}^{\infty} \mathbf{1}_{(Y_n = x, E_0 + \cdots + E_{n-1} \leq t < E_0 + \cdots + E_n)} \, \mathrm{d}t \quad = \quad \sum_{n=0}^{\infty} E_n \mathbf{1}_{(Y_n = x)}.$$

Conditionally on Y, the local time $L_x(X)$ is therefore the sum of $L_x(Y)$ i.i.d. Exponential variables with rate $q(x)$, where $L_x(Y) := \sum_{n=0}^{\infty} \mathbf{1}_{(Y_n = x)}$ is the local time spent at x by the discrete-time chain Y. In particular, $L_x(X)$ is finite if and only if $L_x(Y)$ is, and $\mathbb{E}[L_x(X)] = \mathbb{E}[L_x(Y)]/q(x)$. In light of this correspondence, the four claims readily follow from their discrete-time counterparts. □

We now turn to stationary laws. First observe that by the Markov property, the shifted process $(X_{s+t})_{s\geq 0}$ has law $\mathrm{CTMC}(\mu_t, Q)$ for every $t \geq 0$, where $\mu_t := \mathrm{law}(X_t)$. Recall also that $\mu_t = \mu P_t$, as noted at (2.10). In particular, $(\mu_t)_{t\geq 0}$ solves the forward Kolmogorov equation

$$\forall t \geq 0, \qquad \frac{\mu_t}{\mathrm{d}t} \;=\; \mu_t Q,$$

by Theorem 2.5. Those observations legitimate the following definition.

Definition 2.5 (Invariant law)**.** *A law $\mu \in \mathcal{P}(\mathbb{X})$ is called invariant (or stationary) when any (hence every) of the following equivalent conditions hold:*

1. *If $X \sim \mathrm{CTMC}(\mu, Q)$, then so does $(X_{s+t})_{s\geq 0}$, for each $t \geq 0$.*
2. *$\mu P_t = \mu$ for all $t \geq 0$;*
3. *$\mu Q = 0$.*

As in the discrete-time case, invariant laws do not charge transient states.

Lemma 2.10 (No mass at transient states)**.** *If $\mu \in \mathcal{P}(\mathbb{X})$ is invariant and x is transient then $\mu(x) = 0$.*

Proof. If $\mu \in \mathcal{P}(\mathbb{X})$ is invariant, then for any state x in the support of μ, we have

$$\mathbb{E}_\mu\left[L_x(X)\right] \;=\; \mathbb{E}_\mu\left[\int_0^\infty \mathbf{1}_{(X_t=x)}\,\mathrm{d}t\right] \;=\; \int_0^\infty \mathbb{P}_\mu\left(X_t = x\right)\,\mathrm{d}t \;=\; \int_0^\infty \mu\left(x\right)\,\mathrm{d}t \;=\; \infty.$$

On the other hand, by the strong Markov property at time $\tau_x(X)$, we have

$$\mathbb{E}_\mu\left[L_x(X)\right] \;=\; \mathbb{P}_\mu(\tau_x(X) < \infty)\mathbb{E}_x[L_x(X)].$$

Those two lines together show that $\mathbb{E}_x[L_x(X)] = \infty$, which means that x is recurrent. □

Thus, the stationarity equation can be restricted to recurrent classes, and since the latter are closed, it can be solved on each of them separately. This reduces our study to irreducible recurrent chains. As in the discrete case, we first seek invariant measures, i.e. non-zero vectors $\mu \in \mathbb{R}_+^{\mathbb{X}}$ such that $\mu Q = 0$. The following observation reduces this to a task we already solved.

Lemma 2.11 (Reduction to discrete time)**.** *A vector $\mu \in \mathbb{R}_+^{\mathbb{X}}$ satisfies $\mu Q = 0$ if and only if the vector $\nu \in \mathbb{R}_+^{\mathbb{X}}$ defined by $\nu(x) := \mu(x)q(x)$ satisfies $\nu T = \nu$.*

Proof. Since $Q(x,y) = q(x)(T(x,y) - \mathbf{1}_{(y=x)})$ for all $(x,y) \in \mathbb{X}^2$, we have $\mu Q = \nu T - \nu$. □

Corollary 2.3 (Invariant measures)**.** *Assume that the chain is irreducible and recurrent, and fix a reference state $o \in \mathbb{X}$. Then, the formula*

$$\forall x \in \mathbb{X}, \qquad \mu_o(x) \;:=\; \mathbb{E}_o\left[\int_0^{\tau_o^+(X)} \mathbf{1}_{(X_t=x)}\right], \tag{2.34}$$

defines an invariant measure $\mu_o \in (0,\infty)^{\mathbb{X}}$, and every other invariant measure is proportional to it.

Proof. Using the correspondence $X = \Psi(Y, E)$, we have for every $x \in \mathbb{X}$,

$$\int_0^{\tau_o^+(X)} \mathbf{1}_{(X_t=x)} \;=\; \sum_{n=0}^{\tau_o^+(Y)-1} E_n \mathbf{1}_{(Y_n=x)}.$$

Taking expectations gives $\mu_o(x) = \nu_o(x)/q(x)$, where $\nu_o(x) = \mathbb{E}_o[\sum_{n=0}^{\tau_o^+(Y)-1} \mathbf{1}_{(Y_n=x)}]$ is precisely the invariant measure of the discrete-time chain Y appearing in Theorem 1.3. Thus, the claim readily follows from the above lemma and Theorem 1.3. □

Corollary 2.4 (Invariant laws). *For an irreducible and recurrent chain, either $\mathbb{E}_o[\tau_o^+(X)] = \infty$ for all $o \in \mathbb{X}$ and there is no invariant law, or $\mathbb{E}_o[\tau_o^+(X)] < \infty$ for all $o \in \mathbb{X}$ and the unique invariant law is:*

$$\forall o \in \mathbb{X}, \qquad \pi(o) \quad = \quad \frac{1}{q(o)\mathbb{E}_o[\tau_o^+(X)]}.$$

Proof. This follows from the previous result, since $\mu_o(o) = 1/q(o)$ and $\mu_o(\mathbb{X}) = \mathbb{E}_o[\tau_o^+(X)]$. □

As in the discrete-time case, a recurrent state $x \in \mathbb{X}$ is called positive-recurrent if $\mathbb{E}_x[\tau_x^+(X)] < \infty$ and null-recurrent otherwise. Note that this actually depends on the entire pair (T, q) and not just on T itself, unlike the transience/recurrence classification. In particular, it may happen that the discrete-time chain Y admits a stationary law, while the continuous-time chain X does not.

Remark 2.6 (Reversible measures). *A measure μ on $\mathbb{X}$ is called reversible if*

$$\forall (x, y) \in \mathbb{X}^2, \qquad \mu(x)Q(x, y) \quad = \quad \mu(y)Q(y, x). \tag{2.35}$$

This powerful symmetry condition is easy to verify in practice, and it implies that μ is stationary by summing over all $x \in \mathbb{X}$. As in the discrete-time case, many concrete models admit reversible measures.

Exercise 2.6 (Birth-death chains). *On $\mathbb{X} = \mathbb{N}$, consider the rate-matrix Q defined as follows:*

$$\forall (x, y) \in \mathbb{X}^2, \qquad Q(x, y) \quad = \quad \begin{cases} b(x) & \text{if } y = x + 1 \\ d(x) & \text{if } y = x - 1 \\ -b(x) - d(x) & \text{if } y = x \\ 0 & \text{else,} \end{cases}$$

for some vectors $b, d \in \mathbb{R}_+^{\mathbb{X}}$ whose entries are all positive, except for the obvious constraint $d(0) = 0$.

1. *Prove that the discrete-time chain $T(x, y) = -Q(x, y)/Q(x, x)\mathbf{1}_{(y \neq x)}$ is recurrent if and only if*

$$\sum_{n=0}^{\infty} \frac{d(1) \cdots d(n)}{b(1) \cdots b(n)} \quad = \quad +\infty. \tag{2.36}$$

2. *Deduce that Q is non-explosive if and only if either (2.36) holds, or*

$$\sum_{n=0}^{\infty} \frac{1}{b(n) + d(n)} \quad = \quad +\infty.$$

3. *Find the unique invariant law in discrete time and in continuous time, when they exist.*

2.8 Ergodicity and convergence to equilibrium

Theorem 2.9 (Ergodic theorem). *If $(X_t)_{t \geq 0}$ is an irreducible chain, transient or null-recurrent, then*

$$\frac{1}{t} \int_0^t f(X_s)\, \mathrm{d}s \quad \xrightarrow[t \to \infty]{a.-s.} \quad 0,$$

for any finitely supported function $f : \mathbb{X} \to \mathbb{R}$, while if $(X_t)_{t \geq 0}$ is positive-recurrent, then

$$\frac{1}{t} \int_0^t f(X_s)\, \mathrm{d}s \quad \xrightarrow[t \to \infty]{a.-s.} \quad \pi(f),$$

for any non-negative or bounded function f, where π denotes the unique invariant law.

Proof. The transient case is easy: for any $x \in \mathbb{X}$, we have $\mathbb{P}(L_x(X) < \infty) = 1$, hence

$$\frac{1}{t}\int_0^t \mathbf{1}_{(X_s=x)}\,\mathrm{d}s \quad \leq \quad \frac{L_x(X)}{t} \quad \xrightarrow[t\to\infty]{a.-s.} \quad 0,$$

from which the claim follows by linearity. We now turn to the recurrent case. Upon modifying the pair (T, q) if needed, we may assume that $Y_n \neq Y_{n+1}$ for all $n \in \mathbb{N}$ (note that by irreducibility, there is no absorbing state unless $|\mathbb{X}| = 1$, in which case the claim is trivial). Also, in view of the decomposition $\mathbb{P}(\cdot) = \sum_{o\in\mathbb{X}} \mathbb{P}_o(\cdot)$, we may assume that $\mathbb{P}(X_0 = o) = 1$ for some fixed state $o \in \mathbb{X}$. Finally, upon replacing f with $f^{\pm}$, we may assume without loss of generality that $f \geq 0$. Now, let $\tau_0, \tau_1, \ldots$ denote the successive entrance times to o by the process X. More precisely, we have $\tau_0 := 0$ and then inductively, for $n \geq 1$,

$$\tau_n \quad := \quad \inf\{t > \tau_{n-1} \colon X_{t-} \neq o, X_t = o\}.$$

It is easy to check that for each $n \in \mathbb{N}$, τ_n is a well-defined, almost-surely finite stopping time such that $X_{\tau_n} = o$. By the strong Markov property, the variables $Z_0, Z_1, \ldots$ defined by

$$\forall n \in \mathbb{N}, \qquad Z_n \quad := \quad \int_{\tau_n}^{\tau_{n+1}} f(X_s)\,\mathrm{d}s,$$

are i.i.d. and non-negative. By the strong law of large numbers, we have

$$\frac{1}{n}\int_0^{\tau_n} f(X_s)\,\mathrm{d}s \;=\; \frac{1}{n}\sum_{k=0}^{n-1} Z_k \quad \xrightarrow[n\to\infty]{a.-s.} \quad \mathbb{E}[Z_0] \;=\; \mu_o(f),$$

where μ_o is the measure appearing in Theorem 2.3. Now, for each $t > 0$, set $N_t := \sup\{n \in \mathbb{N} \colon \tau_n \leq t\}$, so that $N_t \leq t < 1 + N_t$, and use the non-negativity of f to write

$$\frac{1}{N_t}\int_0^{\tau_{N_t}} f(X_s)\,\mathrm{d}s \quad \leq \quad \frac{1}{N_t}\int_0^{t} f(X_s)\,\mathrm{d}s \quad \leq \quad \left(1 + \frac{1}{N_t}\right)\frac{1}{N_t+1}\int_0^{\tau_{N_t+1}} f(X_s)\,\mathrm{d}s.$$

By non-explosion, we have $N_t \to \infty$ almost-surely as $t \to \infty$, so the last two displays imply

$$\frac{1}{N_t}\int_0^t f(X_s)\,\mathrm{d}s \quad \xrightarrow[t\to\infty]{a.-s.} \quad \mu_o(f).$$

Finally, this was established for any $f \colon \mathbb{X} \to \mathbb{R}_+$, so we may choose $f = 1$ to obtain $t/N_t \to \mu_o(\mathbb{X})$ almost-surely as $t \to \infty$. Thus, we deduce that

$$\frac{1}{t}\int_0^t f(X_s)\,\mathrm{d}s \quad \xrightarrow[n\to\infty]{a.-s.} \quad \frac{\mu_o(f)}{\mu_o(\mathbb{X})},$$

as long as $\mu_o(f) < \infty$ or $\mu_o(\mathbb{X}) < \infty$, which suffices to conclude. □

Theorem 2.10 (Convergence to equilibrium)**.** *If X is irreducible and positive recurrent, then*

$$\mathrm{law}(X_t) \quad \xrightarrow[t\to\infty]{} \quad \pi,$$

in the usual sense of weak convergence on $\mathbb{X}$ (see Exercise 1.5), where π denotes the unique invariant law.

Proof. Write $\mu_t := \mathrm{law}(X_t)$. Recall that by Remark 2.2, the time-discretization $(X_n)_{n\in\mathbb{N}}$ is a Markov chain with transition matrix P_1. The latter is irreducible and aperiodic because all entries are positive, and positive-recurrent because $\pi P_1 = \pi$. Thus, Theorem 1.5 ensures that

$$\mathrm{d_{TV}}(\mu_n, \pi) \quad \xrightarrow[n\to\infty]{} \quad 0,$$

establishing the desired convergence along some sub-sequence. To fill the gaps, we simply note that the function $t \mapsto \mathrm{d_{TV}}(\mu_t, \pi)$ is non-increasing, by (2.10) and Exercise 1.5. □

Chapter 3

Markov processes on Polish spaces

In this chapter, we considerably extend the theory developed so far by allowing our state space $\mathbb{X}$ to be an arbitrary Polish space, equipped with its Borel $\sigma-$field $\mathcal{B}$. Let us perhaps recall at this point that a Polish space is a separable topological space which can be metrized so as to be complete. Simple concrete examples are discretes spaces, the Euclidean space $\mathbb{R}^d$ for any $d \geq 1$, and any closed or open subset thereof. We use the standard notation $\mathcal{P}(\mathbb{X})$ for the set of probability measures on $(\mathbb{X}, \mathcal{B})$, and $L^\infty(\mathbb{X})$ for the Banach space of bounded measurable functions $f\colon \mathbb{X} \to \mathbb{R}$, equipped with the supremum norm $\|f\|_\infty := \sup_{x\in\mathbb{X}} f(x)$. We also write $\mathcal{C}_b(\mathbb{X}) \subseteq L^\infty(\mathbb{X})$ for the subspace of bounded continuous functions, and $\mathcal{C}_0(\mathbb{X}) \subseteq \mathcal{C}_b(\mathbb{X})$ for those that vanish at infinity, meaning that $\{x \in \mathbb{X}\colon |f(x)| \geq \varepsilon\}$ is compact for each $\varepsilon > 0$.

3.1 Discrete-time Markov processes

Transition matrices played a central role in the theory of Markov chains on discrete spaces, and our first task consists in defining an appropriate generalization thereof.

Definition 3.1 (Transition kernel). *A transition kernel is a function $T\colon \mathbb{X} \times \mathcal{B} \to [0,1]$ such that*

1. *For each $x \in \mathbb{X}$, the function $B \mapsto T(x,B)$ is a probability measure on $(\mathbb{X}, \mathcal{B})$.*
2. *For each $B \in \mathcal{B}$, the function $x \mapsto T(x,B)$ is Borel-measurable.*

A transition kernel T can be viewed as a linear operator acting on $L^\infty(\mathbb{X})$ as follows:

$$\forall x \in \mathbb{X}, \qquad Tf(x) \;:=\; \int_{\mathbb{X}} f(y)T(x,\mathrm{d}y). \tag{3.1}$$

Note that the transition kernel can then be recovered from this operator via $T(x,B) = (T\mathbf{1}_B)(x)$ for all $(x,B) \in \mathbb{X} \times \mathcal{B}$. Dually, the transition kernel T acts on $\mathcal{P}(\mathbb{X})$ as follows:

$$\forall B \in \mathcal{B}, \qquad \mu T(B) \;:=\; \int_{\mathbb{X}} T(x,B)\mu(\mathrm{d}x).$$

Again, the transition kernel can then be recovered via $T(x,B) = \delta_x T(B)$, for all $(x,B) \in \mathbb{X} \times \mathcal{B}$. Finally, two transition kernels T_1 and T_2 can be composed in the following way:

$$\forall x \in \mathbb{X}, \quad \forall B \in \mathcal{B}, \qquad (T_1T_2)(x,B) \;:=\; \int_{\mathbb{X}} T_2(y,B)T_1(x,\mathrm{d}y).$$

We invite the reader to check that those definitions make sense, that Tf is a bounded measurable function, μT a probability measure, and T_1T_2 a transition kernel, and that $(T_1T_2)f = T_1(T_2f)$.

Remark 3.1 (Conditional laws). *One can think of transition kernels as conditional laws. More precisely, if (X, Y) is a $\mathbb{X}^2-$valued random variable, then there exists a transition kernel T such that*

$$\forall B \in \mathcal{B}, \qquad \mathbb{P}\left[Y \in B \,|\, X\right] \;=\; T(X, B), \tag{3.2}$$

almost-surely. We call T (a version of) the conditional law of Y given X, and write $\mathrm{law}(Y \,|\, X) = T(X, \cdot)$. *Conversely, given $\mu \in \mathcal{P}(\mathbb{X})$ and a transition kernel T on $\mathbb{X}$, the formula*

$$\forall (A, B) \in \mathcal{B}^2, \qquad \nu(A \times B) \;:=\; \int_{\mathbb{X}} \mathbf{1}_A(\mathrm{d}x) T(x, B) \mu(\mathrm{d}x),$$

specifies a unique probability measure ν on the product space $(\mathbb{X}^2, \mathcal{B} \otimes \mathcal{B})$. Moreover, for any $f \in L^\infty(\mathbb{X}^2)$, the function $x \mapsto \int_{\mathbb{X}} f(x, y) T(x, \mathrm{d}y)$ is measurable, and we have

$$\int_{\mathbb{X}^2} f \,\mathrm{d}\nu \;=\; \int_{\mathbb{X}} \left(\int_{\mathbb{X}} f(x, y) T(x, \mathrm{d}y) \right) \mu(\mathrm{d}x),$$

justifying the notation $\nu(\mathrm{d}x\,\mathrm{d}y) = T(x, \mathrm{d}y)\mu(\mathrm{d}x)$. Note that if $(X, Y) \sim \nu$, then (3.2) holds.

Example 3.1 (Idle kernel). *The idle kernel is the following particular transition kernel:*

$$\forall x \in \mathbb{X}, \forall B \in \mathcal{B}, \qquad T(x, B) \;=\; \delta_x(B) \;=\; \begin{cases} 1 & \text{if } x \in B \\ 0 & \text{else.} \end{cases}$$

Clearly, we have $Tf = f$ for all $f \in L^\infty(\mathbb{X})$, justifying the notation $T = \mathrm{Id}$.

Definition 3.2 (Discrete-time Markov process). *A discrete-time Markov process with transition kernel T is a $\mathbb{X}-$valued stochastic process $X = (X_n)_{n \in \mathbb{N}}$ such that*

$$\mathbb{E}\left[f(X_{n+1}) \,|\, X_0, \ldots, X_n\right] \;=\; Tf(X_n), \tag{3.3}$$

almost-surely, for every $n \in \mathbb{N}$ and every $f \in L^\infty(\mathbb{X})$.

Note that by successive conditioning, the identity (3.3) allows one to reduce the computation of expectations of the form $\mathbb{E}\left[f_0(X_0) \cdots f_n(X_n)\right]$ with $n \in \mathbb{N}$ and $f_0, \ldots, f_n \in L^\infty(\mathbb{X})$, to the base case where $n = 0$. Specifically, writing $\mu = \mathrm{law}(X_0)$, we have

$$\begin{aligned}
\mathbb{E}\left[f_0(X_0)\right] &= \int f_0(x_0) \mu(\mathrm{d}x_0) \\
\mathbb{E}\left[f_0(X_0) f_1(X_1)\right] &= \int f_0(x_0) \left[\int f_1(x_1) T(x_0, \mathrm{d}x_1) \right] \mu(\mathrm{d}x_0) \\
\mathbb{E}\left[f_0(X_0) f_1(X_1) f_2(X_2)\right] &= \int f_0(x_0) \left[\int f_1(x_1) \left[\int f_2(x_2) T(x_1, \mathrm{d}x_2) \right] T(x_0, \mathrm{d}x_1) \right] \mu(\mathrm{d}x_0), \\
&\vdots
\end{aligned}$$

and so on. Using the short-hand $F(x_0, \ldots, x_n) = f_0(x_0) \cdots f_n(x_n)$ and dropping the brackets for better visibility, we arrive at the general expression:

$$\mathbb{E}\left[F(X_0, \ldots, X_n)\right] \;=\; \int \cdots \int F(x_0, \ldots, x_n) T(x_{n-1}, \mathrm{d}x_n) \cdots T(x_0, \mathrm{d}x_1) \mu(\mathrm{d}x_0). \tag{3.4}$$

By the monotone class lemma, this formula automatically extends to any bounded measurable function $F \colon \mathbb{X}^{n+1} \to \mathbb{R}$, and it determines the law of $(X_0, \ldots, X_n)$. Thus, the law of the process

$X = (X_n)_{n\in\mathbb{N}}$, viewed as a single random variable taking values in the product space $(\mathbb{X}^{\mathbb{N}}, \mathcal{B}^{\otimes\mathbb{N}})$, is completely determined by the transition kernel T and the initial law μ. We will write

$$X \quad \sim \quad \mathrm{DTMP}(\mu, T).$$

Conversely, we can always produce a discrete-time Markov process X with a given transition kernel T and initial law $\mu \in \mathcal{P}(\mathbb{X})$ using the following canonical construction. First, the formula

$$\forall B \in \mathcal{B}^{\otimes n+1}, \qquad \nu_n(B) \quad := \quad \int \cdots \int \mathbf{1}_B(x_0, \ldots, x_n) T(x_{n-1}, \mathrm{d}x_n) \cdots T(x_0, \mathrm{d}x_1) \mu(\mathrm{d}x_0),$$

defines a probability measure ν_n on $\mathbb{X}^{n+1}$ for each $n \in \mathbb{N}$, and those measures are consistent:

$$\forall n \in \mathbb{N}, \qquad \forall B \in \mathcal{B}^{\otimes n+1}, \qquad \nu_{n+1}(B \times \mathbb{X}) \quad = \quad \nu_n(B). \tag{3.5}$$

By Kolmogorov's extension theorem, there is a (unique) probability measure ν_∞ on $(\mathbb{X}^{\mathbb{N}}, \mathcal{B}^{\otimes\mathbb{N}})$, whose image under the projection $x \mapsto (x_0, \ldots, x_n)$ is ν_n, for each $n \in \mathbb{N}$. This means that, on the probability space $(\Omega, \mathcal{F}, \mathbb{P}) := (\mathbb{X}^{\mathbb{N}}, \mathcal{B}^{\otimes\mathbb{N}}, \nu_\infty)$, the identity map X is a $\mathbb{X}-$valued stochastic process satisfying (3.4). In particular, we have $\mathrm{law}(X_0) = \mu$, and

$$\mathbb{E}\left[F(X_0, \ldots, X_n) f(X_{n+1})\right] \quad = \quad \mathbb{E}\left[F(X_0, \ldots, X_n) T f(X_n)\right],$$

for every $n \in \mathbb{N}$, every $F \in L^\infty(\mathbb{X}^{n+1})$ and every $f \in L^\infty(\mathbb{X})$, establishing (3.3).

Remark 3.2 (Filtration). *When the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is equipped with a filtration $(\mathcal{F}_n)_{n\geq 0}$, we may strengthen the above definition by requiring that X be adapted, and that*

$$\mathbb{E}\left[f(X_{n+1}) \,|\, \mathcal{F}_n\right] \quad = \quad Tf(X_n), \tag{3.6}$$

almost-surely, for each $n \in \mathbb{N}$ and $f \in L^\infty(\mathbb{X})$. We then naturally speak of a $(\mathcal{F}_n)_{n\geq 0}-$Markov process. By the tower property of conditional expectations, this always implies our earlier requirement (3.3), which corresponds to the minimal choice $\mathcal{F}_n := \sigma(X_0, \ldots, X_n)$ for all $n \in \mathbb{N}$.

As in the case of discrete state-spaces, the property (3.6) guarantees that discrete-time Markov processes enjoy the following fundamental memoryless property.

Theorem 3.1 (Weak Markov property). *Let $(X_n)_{n\in\mathbb{N}}$ be a discrete-time $(\mathcal{F}_n)_{n\geq 0}-$Markov process with transition kernel T, and fix $m \in \mathbb{N}$. Then,*

$$\mathrm{law}\left((X_{m+n})_{n\in\mathbb{N}} \,|\, \mathcal{F}_m\right) \quad = \quad \mathrm{DTMP}(\delta_{X_m}, T).$$

Proof. Let us introduce the convenient notation $\mathbf{E}_x[\cdot]$ for the expectation w.r.t. the probability measure $\mathrm{DTMP}(\delta_x, T)$ on the canonical space $(\mathbb{X}^{\mathbb{N}}, \mathcal{B}^{\otimes\mathbb{N}})$, and Z for the identity map thereon. With that in hands, the desired identity rewrites as follows: for each $F \in L^\infty(\mathbb{X}^{\mathbb{N}})$, almost-surely,

$$\mathbb{E}\left[F(X_m, X_{m+1}, \ldots) \,\middle|\, \mathcal{F}_m\right] \quad = \quad \mathbf{E}_{X_m}\left[F(Z_0, Z_1, \ldots)\right].$$

By the monotone class lemma, it is enough to consider functions of the form $F(x_0, x_1, \ldots) = f_0(x_0) \cdots f_n(x_n)$, with $n \in \mathbb{N}$ and $f_0, \ldots, f_n \in L^\infty(\mathbb{X})$. We naturally proceed by induction over n. The base case $n = 0$ is trivial, and if the desired property holds for some $n \in \mathbb{N}$, then for any $f_1, \ldots, f_{n+1} \in L^\infty(\mathbb{X})$, we can write

$$\begin{aligned}
\mathbb{E}\left[f_0(X_m) \cdots f_{n+1}(X_{m+n+1}) \,\middle|\, \mathcal{F}_m\right] \quad &= \quad \mathbb{E}\left[f_0(X_m) \cdots f_n(X_{m+n})(T f_{n+1})(X_{m+n}) \,\middle|\, \mathcal{F}_m\right] \\
&= \quad \mathbf{E}_{X_m}\left[f_0(Z_0) \cdots f_n(Z_n)(T f_{n+1})(Z_n)\right] \\
&= \quad \mathbf{E}_{X_m}\left[f_0(Z_0) \cdots f_{n+1}(Z_{n+1})\right],
\end{aligned}$$

where we have successively used the property (3.6), the induction hypthesis, and finally the fact that Z is a discrete-time Markov process with transition matrix T. □

Theorem 3.2 (Strong Markov property). *Let $(X_n)_{n\in\mathbb{N}}$ be a discrete-time $(\mathcal{F}_n)_{n\geq 0}-$Markov process with transition kernel T, and let τ be a $(\mathcal{F}_n)_{n\geq 0}-$stopping time. Then, on the event $\{\tau<\infty\}$, we have*

$$\text{law}\left((X_{n+\tau})_{n\in\mathbb{N}}\,|\,\mathcal{F}_\tau\right) \;=\; \text{DTMP}(\delta_{X_\tau}, T).$$

Proof. For any $F\in L^\infty(\mathbb{X})$ and $A\in\mathcal{F}_\tau$, we can use the above result (and notation) to write

$$\begin{aligned}\mathbb{E}\left[\mathbf{1}_A\mathbf{1}_{\{\tau<\infty\}}F(X_\tau,X_{\tau+1},\ldots)\right] &= \sum_{m=0}^{\infty}\mathbb{E}\left[\mathbf{1}_{A\cap\{\tau=m\}}F(X_m,X_{m+1},\ldots)\right]\\ &= \sum_{m=0}^{\infty}\mathbb{E}\left[\mathbf{1}_{A\cap\{\tau=m\}}\mathbf{E}_{X_m}[F(Z_0,Z_1,\ldots)]\right]\\ &= \mathbb{E}\left[\mathbf{1}_A\mathbf{1}_{\{\tau<\infty\}}\mathbf{E}_{X_\tau}[F(Z_0,Z_1,\ldots)]\right].\end{aligned}$$

Since is true for all $A\in\mathcal{F}_\tau$, we have shown that almost-surely,

$$\mathbb{E}\left[\mathbf{1}_{\{\tau<\infty\}}F(X_\tau,X_{\tau+1},\ldots)\,|\,\mathcal{F}_\tau\right] \;=\; \mathbf{1}_{\{\tau<\infty\}}\mathbf{E}_{X_\tau}[F(Z_0,Z_1,\ldots)],$$

and the indicator on the left-hand side can be pulled out of the expectation because it is $\mathcal{F}_\tau-$measurable. Since this holds for any $F\in L^\infty(\mathbb{X}^{\mathbb{N}})$, the result is proved. □

Lemma 3.1 (Marginal evolution). *If X is a discrete-time Markov process with transition kernel T, then for every $n\in\mathbb{N}$, the marginal law $\mu_n:=\text{law}(X_n)$ is given by $\mu_n=\mu_0T^n$.*

Proof. For each $n\in\mathbb{N}$ and each $B\in\mathcal{B}$, we have by definition

$$\mu_{n+1}(B) \;=\; \mathbb{E}\left[\mathbf{1}_B(X_{n+1})\right] \;=\; \mathbb{E}\left[(T\mathbf{1}_B)(X_n)\right] \;=\; \int_{\mathbb{X}}T(x,B)\mu_n(\,\mathrm{d}x) \;=\; (\mu_nT)(B),$$

and the result now readily follows by induction. □

Together, the Markov property and this lemma imply that if $(X_0,X_1,\ldots)\sim\text{DTMP}(\mu,T)$, then $(X_1,X_2,\ldots)\sim\text{DTMP}(\mu T,T)$. It follows that the distribution $\text{DTMP}(\mu,T)$ is preserved under time-shifts if and only if the initial law μ is invariant, in the sense that $\mu T=\mu$. The question of existence, uniqueness and global attractiveness of invariant laws is of course of fundamental importance, but an exhaustive treatment would take us too far.

3.2 Classical examples

As in the case of discrete spaces, stochastic recursions form a natural and important source of examples of discrete-time Markov processes. It is actually generic, in the sense that any transition kernel on $(\mathbb{X},\mathcal{B})$ has a representation of the form (3.7), with ξ_1 being even just a uniform random variable on $[0,1]$.

Example 3.2 (Stochastic recursions). *Consider a sequence $\xi=(\xi_n)_{n\geq 1}$ of i.i.d. random variables taking values in an arbitrary measurable space $(E,\mathcal{E})$, and a measurable function $F\colon\mathbb{X}\times E\to\mathbb{X}$. Let also X_0 be a $\mathbb{X}-$valued random variable, independent of ξ. Define $X_1,X_2,\ldots$ inductively as follows:*

$$\forall n\in\mathbb{N},\qquad X_{n+1} \;:=\; F(X_n,\xi_{n+1}).$$

Then, it is immediate to check that $(X_n)_{n\in\mathbb{N}}$ is a discrete-time Markov process with respect to the filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ defined by $\mathcal{F}_n:=\sigma(X_0,\xi_1,\cdots,\xi_n)$, and that its transition kernel is as follows:

$$\forall x\in\mathbb{X},\quad \forall B\in\mathcal{B},\quad T(x,B) \;:=\; \mathbb{P}\left(F(x,\xi_1)\in B\right). \tag{3.7}$$

A simple instance of a stochastic recursion is a random walk on a topological group.

Example 3.3 (Random walks on groups). *Suppose that $\mathbb{X}$ is equipped with a group structure, and let $(\xi_n)_{n\geq 1}$ be i.i.d. elements drawn according to some distribution $\nu \in \mathcal{P}(\mathbb{X})$. Define*

$$\forall n \in \mathbb{N}, \qquad X_n \;:=\; \xi_n \cdots \xi_1,$$

in the sense of multiplication in the group, and with the usual convention that an empty product is the identity element. Then, it follows from the previous example that the process $X := (X_n)_{n\in\mathbb{N}}$ is a discrete-time Markov process with respect to the filtration $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$, with transition kernel

$$T(x,\, B) \;:=\; \nu(Bx^{-1}).$$

Such a process is called a random walk on the group $\mathbb{X}$, generated by the measure ν. Some natural and widely-studied examples are sums of i.i.d. random variables in $\mathbb{R}^d$, product of i.i.d. $d \times d$ complex random matrices, or compositions of i.i.d. rotations of the $d-$dimensional unit sphere.

Example 3.4 (Gibbs sampler). *A very common strategy to produce approximate samples from a given $d-$dimensional continuous probability density $f\colon \mathbb{R}^d \to \mathbb{R}_+$ consists in running the following discrete-time Markov process on $\mathbb{X} = \{x \in \mathbb{R}^d \colon f(x) > 0\}$: start from an arbitrary initial condition and, at each step, modify the current state $x = (x_0, \ldots, x_n)$ by choosing a coordinate $i \in \{1, \ldots, d\}$ uniformly at random and replacing the corresponding value x_i by a fresh sample from the conditional density*

$$f_{i,x}(z) \;:=\; \frac{f(x_1, \ldots, x_{i-1}, z, x_{i+1}, \ldots, x_n)}{\int_{-\infty}^{\infty} f(x_1, \ldots, x_{i-1}, u, x_{i+1}, \ldots, x_n)\,\mathrm{d}u}.$$

This is known as the Gibbs sampler with target f. For concreteness, we invite the reader to explicitate it in the important cases where f is the Gaussian density with mean 0 and covariance matrix $\Sigma \succ 0$.

Exercise 3.1 (Averaging process). *Fix a finite undirected graph $G = (V, E)$, and equip each vertex $i \in V$ with a real value x_i. Now, consider the following simple stochastic dynamics: at each step, an edge $\{i, j\} \in E$ is chosen uniformly at random, and the values x_i, x_j at its end-points are both replaced with their average $(x_i + x_j)/2$. Describe this as a Markov process, and find its invariant laws.*

3.3 Continuous-time Markov processes

We now turn to continuous-time Markov processes taking values in our Polish space $(\mathbb{X}, \mathcal{B})$. The central ingredient needed to describe its evolution is a transition semi-group.

Definition 3.3 (Transition semi-group). *A transition semi-group (also known as a Markov semi-group) on $\mathbb{X}$ is a family of transition kernels $(P_t)_{t\geq 0}$ on $\mathbb{X}$ satisfying*

$$\forall x \in \mathbb{X}, \quad P_0 = \mathrm{Id}, \quad \textit{and} \quad \forall s, t \in \mathbb{R}_+, \qquad P_s P_t \;=\; P_{t+s}. \tag{3.8}$$

As in the case of discrete spaces, the number $P_t(x, B)$ will represent the probability that our process evolves from x to B within a given time-interval of duration t.

Definition 3.4 (Continuous-time Markov process). *A continuous-time Markov process with transition semi-group $(P_t)_{t\geq 0}$ is a $\mathbb{X}-$valued stochastic process $X = (X_t)_{t\geq 0}$ such that*

$$\mathbb{E}\left[f(X_{t+s}) \,|\, (X_u)_{u\in[0,s]}\right] \;=\; P_t f(X_s), \tag{3.9}$$

almost-surely, for every $s, t \geq 0$ and every $f \in L^\infty(\mathbb{X})$.

Remark 3.3 (Refinement). *When the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is equipped with a filtration $(\mathcal{F}_t)_{t\geq 0}$, we may strengthen the above definition by requiring that X be adapted to $(\mathcal{F}_t)_{t\geq 0}$ and that*

$$\mathbb{E}\left[f(X_{t+s}) \,|\, \mathcal{F}_s\right] \;=\; P_t f(X_s), \tag{3.10}$$

almost-surely, for every $s, t \geq 0$ and every $f \in L^\infty(\mathbb{X})$. We then naturally speak of a $(\mathcal{F}_t)_{t\geq 0}$−Markov process. This definition implies the previous one, which is the minimal choice $\mathcal{F}_t = \sigma(X_u \colon u \in [0,t])$.

Remark 3.4 (Consistency). *By the tower property of conditional expectation, we have*

$$P_{t+s}f(X_0) \;=\; \mathbb{E}\left[f(X_{t+s}) \,|\, \mathcal{F}_0\right] \;=\; \mathbb{E}\left[\mathbb{E}\left[f(X_{t+s})|\mathcal{F}_s\right] \,|\, \mathcal{F}_0\right] \;=\; \mathbb{E}\left[P_t f(X_s) \,|\, \mathcal{F}_0\right] \;=\; P_s P_t f(X_0),$$

almost-surely, for every $s, t \geq 0$ and every $f \in L^\infty(\mathbb{X})$. In addition,

$$(P_0 f)(X_0) \;=\; \mathbb{E}[f(X_0) \,|\, \mathcal{F}_0] \;=\; f(X_0),$$

almost-surely, in agreement with the semi-group structure required by Definition 3.3.

Remark 3.5 (Finite-dimensional marginals). *By successive conditioning, (3.9) provides access to the finite-dimensional marginals of X. Specifically, writing $\mu := \mathrm{law}(X_0)$, we have by induction*

$$\mathbb{E}\left[F(X_{t_1}, \ldots, X_{t_n})\right] \;=\; \int \cdots \int F(x_1, \ldots, x_n) P_{t_n - t_{n-1}}(x_{n-1}, \mathrm{d}x_n) \cdots P_{t_1 - t_0}(x_0, \mathrm{d}x_1) \mu(\mathrm{d}x_0),$$

for any $n \in \mathbb{N}$, any $0 = t_0 \leq t_1 \leq \ldots \leq t_n$, and any function of the form $F(x_1, \ldots, x_n) = f_1(x_1) \cdots f_n(x_n)$ with $f_1, \ldots, f_n \in L^\infty(\mathbb{X})$. By the monotone class lemma, this extends to all $F \in L^\infty(\mathbb{X}^n)$, showing that the law of X is determined by its initial law and transition semi-group. We write

$$X \;\sim\; \mathrm{CTMP}\left(\mu, (P_t)_{t\geq 0}\right).$$

Remark 3.6 (Canonical construction). *Given an initial law $\mu \in \mathcal{P}(\mathbb{X})$ and a transition semi-group $(P_t)_{t\geq 0}$, there always exists a continuous-time Markov process with those parameters. Indeed, the formula*

$$\nu_{\{t_1,\ldots,t_n\}}(B) \;:=\; \int \cdots \int \mathbf{1}_B(x_1, \ldots, x_n) P_{t_n - t_{n-1}}(x_{n-1}, \mathrm{d}x_n) \cdots P_{t_1 - t_0}(x_0, \mathrm{d}x_1) \mu(\mathrm{d}x_0),$$

defines a probability measure on $\mathbb{X}^{\{t_1,\ldots,t_n\}}$ for each $n \in \mathbb{N}$ and $0 = t_0 \leq \ldots \leq t_n$, and the semi-group properties ensure that those measures are consistent, in the sense that ν_I is the image of ν_J under the canonical projection $\mathbb{X}^J \to \mathbb{X}^I$, for any finite $I \subseteq J \subseteq [0,\infty)$. Thus, Kolmogorov's extension theorem guarantees the existence of a (unique) probability measure ν on $\mathbb{X}^{[0,\infty)}$ with finite-dimensional marginals $(\nu_I)_{I\subseteq[0,\infty)}$. By construction, the identity map X on the canonical space $(\Omega, \mathcal{F}, \mathbb{P}) := \left(\mathbb{X}^{[0,\infty)}, \mathcal{B}^{\otimes[0,\infty)}, \nu\right)$ is then a Markov process with initial law μ and transition semi-group $(P_t)_{t\geq 0}$.

While very general, our definition of continuous-time Markov processes does not allow to say anything really interesting about their trajectorial properties, due to the intrinsically limited expressive power of the product topology. A simple and natural way to overcome this is to restrict attention to continuous-time Markov processes with càdlàg trajectories. The following difficult result provides a simple sufficient condition for such processes to exist. Our Polish space $\mathbb{X}$ is called locally compact if every point lies in an open set whose closure is compact. This is clearly the case for discrete spaces, the Euclidean space $\mathbb{R}^d$, or any open subset of it.

Theorem 3.3 (Existence of càdlàg Markov processes). *Assume that our Polish space $\mathbb{X}$ is locally compact, and that the transition semi-group $(P_t)_{t\geq 0}$ has the Feller property:*

1. $P_t\mathcal{C}_0(\mathbb{X}) \subseteq \mathcal{C}_0(\mathbb{X})$, *for every* $t \geq 0$.
2. $P_t f \to f$ *pointwise as* $t \to 0$, *for every* $f \in \mathcal{C}_0(\mathbb{X})$.

Then for each $\mu \in \mathcal{P}(\mathbb{X})$, *there exists a continuous-time Markov process with initial law* μ, *transition semi-group* $(P_t)_{t\geq 0}$, *and càdlàg trajectories.*

A càdlàg Markov process whose transition semi-group has the Feller property is called a Feller process. Here is a celebrated example.

Exercise 3.2 (Brownian motion). *Take* $\mathbb{X} = \mathbb{R}$, *set* $P_0 := \mathrm{Id}$ *and for* $t > 0$,

$$\forall x \in \mathbb{X}, \qquad \forall B \in \mathcal{B}, \qquad P_t(x, B) \;:=\; \frac{1}{\sqrt{2\pi t}} \int_B e^{-\frac{|y-x|^2}{2t}}\, \mathrm{d}y.$$

1. *Prove that* $(P_t)_{t\geq 0}$ *is a transition semi-group, and that it has the Feller property.*
2. *Prove that any càdlàg realization of* $\mathrm{CTMP}(\delta_0, (P_t)_{t\geq 0})$ *is a Brownian motion.*

3.4 Convolution semi-groups and Levy processes

In this section, $(\mathbb{X}, \mathcal{B})$ is the $d-$dimensional Euclidean space $\mathbb{R}^d$, equipped with its Borel $\sigma-$field. Any probability measure $\mu \in \mathcal{P}(\mathbb{X})$ gives rise to a transition kernel on $\mathbb{X}$ by convolution:

$$\forall (x, B) \in \mathbb{X} \times \mathcal{B}, \qquad T_\mu(x, B) \;:=\; \mu(B - x) \;=\; \int_{\mathbb{X}} \mathbf{1}_B(x + y)\mu(\mathrm{d}y).$$

In particular, T_{δ_0} is the idle kernel. Also, by linearity and monotone convergence, we have

$$\forall x \in \mathbb{X}, \qquad \forall f \in L^\infty(\mathbb{X}), \qquad (T_\mu f)(x) \;=\; \int_{\mathbb{X}} f(x + y)\mu(\mathrm{d}y).$$

Now, recall that the convolution $\mu \star \nu$ of two probability measures $\mu, \nu \in \mathcal{P}(\mathbb{X})$ is the image of the product measure $\mu \otimes \nu$ under addition $(x, y) \mapsto x + y$. Thus, for all $x \in \mathbb{X}, f \in L^\infty(\mathbb{X})$,

$$\begin{aligned}
(T_{\mu\star\nu} f)(x) &= \int_{\mathbb{X}} f(x + y)(\mu \star \nu)(\mathrm{d}y) \\
&= \int_{\mathbb{X}^2} f(x + y + z)\mu(\mathrm{d}y)\nu(\mathrm{d}z) \\
&= \int_{\mathbb{X}} T_\nu f(x + y)\mu(\mathrm{d}y) \\
&= (T_\mu T_\nu f)(x),
\end{aligned}$$

showing that $T_{\mu\star\nu} = T_\mu T_\nu$. In particular, if $(\mu_t)_{t\geq 0}$ are probability measures on $\mathbb{X}$ satisfying

$$\mu_0 = \delta_0, \qquad \text{and} \qquad \forall s, t \in \mathbb{R}_+, \qquad \mu_{t+s} \;=\; \mu_t \star \mu_s, \tag{3.11}$$

then, the transition kernels $(P_t)_{t\geq 0}$ defined by $P_t := T_{\mu_t}$ form a transition semi-group, called a convolution semi-group. Note that the underlying family $(\mu_t)_{t\geq 0}$ can then be recovered via $\mu_t = P_t(0, \cdot)$. Markov processes driven by convolution semi-groups are particularly nice:

Theorem 3.4 (Processes with stationary and independent increments). *Let* X *be a continuous-time,* $(\mathcal{F}_t)_{t\geq 0}-$*adapted stochastic process taking values in* $\mathbb{X} = \mathbb{R}^d$. *Then, the following are equivalent:*

(i) For each $t, s \geq 0$, *the increment* $X_{t+s} - X_s$ *is independent of* $\mathcal{F}_s$ *and distributed as* X_t.

(ii) X *is a* $(\mathcal{F}_t)_{t\geq 0} -$ CTMP *with parameters* $(\delta_0, (P_t)_{t\geq 0})$*, for some convolution semi-group* $(P_t)_{t\geq 0}$*.*

Proof. Suppose that (i) holds, and set $\mu_t := \text{law}(X_t)$. Note that $\mu_0 = \delta_0$, by (i) with $t = s$. Also, the decomposition $X_{t+s} = (X_{t+s} - X_s) + X_s$ and the property (i) implies $\mu_{t+s} = \mu_t \star \mu_s$. Thus, $(\mu_t)_{t\geq 0}$ satisfies (3.11), and we write $(P_t)_{t\geq 0}$ for the associated convolution semi-group. Now, for any $t, s \geq 0$ and $f \in L^\infty(\mathbb{X})$, we can invoke the freezing lemma to obtain

$$\mathbb{E}\left[f(X_{t+s}) \,|\, \mathcal{F}_s\right] \;=\; \mathbb{E}\left[f(X_s + (X_{t+s} - X_s)) \,|\, \mathcal{F}_s\right] \;=\; F(X_s),$$

where the function $F\colon \mathbb{X} \to \mathbb{R}$ is defines as follows: for any $x \in \mathbb{X}$,

$$\begin{aligned} F(x) &= \mathbb{E}\left[f(x + (X_{t+s} - X_s))\right] \\ &= \mathbb{E}\left[f(x + X_t)\right] \\ &= \int_{\mathbb{X}} f(x+y)\mu_t(\mathrm{d}y) \\ &= P_t f(x). \end{aligned}$$

Thus, X is a $(\mathcal{F}_t)_{t\geq 0} -$ CTMP with parameters $(\delta_0, (P_t)_{t\geq 0})$ and (ii) is established. Conversely, assume that X is a $(\mathcal{F}_t)_{t\geq 0} -$ CTMP with parameters $(\delta_0, (P_t)_{t\geq 0})$, where $(P_t)_{t\geq 0}$ is the convolution semi-group induced by some family $(\mu_t)_{t\geq 0}$ satisfying (3.11). Then, for any $t, s \geq 0$, we have

$$\text{law}\left(X_{t+s} \,|\, \mathcal{F}_s\right) \;=\; P_t(X_s, \cdot) \;=\; \delta_{X_s} \star \mu_t,$$

because $T_\mu(x, \cdot) = \delta_x \star \mu$ for any $x \in \mathbb{X}$ and $\mu \in \mathcal{P}(\mathbb{X})$, by definition. Thus,

$$\text{law}\left(X_{t+s} - X_s \,|\, \mathcal{F}_s\right) \;=\; \delta_{-X_s} \star \delta_{X_s} \star \mu_t \;=\; \mu_t,$$

which shows that $X_{t+s} - X_s$ is independent of $\mathcal{F}_s$, and with law μ_t. Choosing $s = 0$ and recalling that $X_0 = 0$, we obtain $\mu_t = \text{law}(X_t)$ and (i) follows. □

Processes of this form are called processes with stationary and independent increments. The above result does not say anything about the existence of càdlàg versions of such processes. The following result shows that the trivial necessary condition $\mu_t \to \delta_0$ is in fact also sufficient.

Lemma 3.2 (Feller property for convolution semi-groups)**.** *A convolution semi-group* $(P_t)_{t\geq 0}$ *has the Feller property if and only if the underlying measure* $\mu_t := P_t(0, \cdot)$ *satisfies the weak convergence*

$$\mu_t \;\xrightarrow[t\to\infty]{}\; \delta_0. \tag{3.12}$$

Proof. By definition, we have for any $f \in \mathcal{C}_0(\mathbb{X})$, any $t \geq 0$, and any $x \in \mathbb{X}$,

$$(P_t f)(x) \;=\; \int_{\mathbb{X}} f(x+y)\mu_t(\mathrm{d}y).$$

Thus, the first requirement in Theorem 3.3 is always guaranteed, by dominated convergence. Moreover, since weak convergence of probability measures on a locally finite Polish space can be tested against $\mathcal{C}_0(\mathbb{X})$, the condition (3.12) is equivalent to the second requirement in Theorem 3.3, and the latter is trivially necessary for the existence of a càdlàg representation. □

Definition 3.5 (Levy processes)**.** *A continuous-time Markov process with càdlàg trajectories and independent, stationary increments is called a Levy process.*

Famous examples include linear processes $(\mu_t = \delta_t)$, Poisson processes $(\mu_t = \text{Poisson}(t))$, Brownian motion $(\mu_t = \mathcal{N}(0, t))$, Gamma processes $(\mu_t = \Gamma(t, 1))$, and Cauchy processes $(\mu_t = \text{Cauchy}(t))$. Of course, one could re-scale time by any positive constant, and combine those examples to obtain many others, since any finite linear combination of independent Levy processes is a Levy process (exercise !).

3.5 Strong Markov property

Throughout this section, we let $X = (X_t)_{t\geq 0}$ denote a continuous-time $(\mathcal{F}_t)_{t\geq 0}-$Markov process with a given transition semi-group $(P_t)_{t\geq 0}$ on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. As in the discrete case, we seek to establish the weak Markov property

$$\text{law}\left[(X_{t+s})_{t\geq 0} \,|\, \mathcal{F}_s\right] \quad = \quad \text{CTMP}\left(\delta_{X_s}, (P_t)_{t\geq 0}\right),$$

for any fixed time $s \geq 0$. A convenient way to formalize this is to introduce the identity map Z on the canonical space $\left(\mathbb{X}^{[0,\infty)}, \mathcal{B}^{\otimes[0,\infty)}\right)$, which we can equip with the law $\mathbf{P}_x := \text{CTMC}(\delta_x, (P_t)_{t\geq 0})$ for any initial state $x \in \mathbb{X}$. The above statement then takes the following more concrete form.

Theorem 3.5 (Weak Markov property). *For every $s \geq 0$ and $F \in L^\infty(\mathbb{X}^{[0,\infty)})$, we have almost-surely*

$$\mathbb{E}\left[F\left((X_{t+s})_{t\geq 0}\right) \,|\, \mathcal{F}_s\right] \quad = \quad \mathbf{E}_{X_s}\left[F(Z)\right]. \tag{3.13}$$

Proof. By the monotone class lemma, it is enough to consider functions F of the form

$$F((x_t)_{t\geq 0}) \quad := \quad f_1(x_{t_1}) \cdots f_n(x_{t_n}),$$

with $n \in \mathbb{N}$, $0 \leq t_1 \leq \ldots \leq t_n$, and $f_1, \ldots, f_n \in L^\infty(\mathbb{X})$. We naturally proceed by induction over n. The case $n = 0$ is trivial, since F is then identically 1. Now, assume that the claim holds for some $n \in \mathbb{N}$. Then, for any $0 \leq t_1 \leq \ldots \leq t_{n+1}$ and $f_1, \ldots, f_{n+1} \in L^\infty(\mathbb{X})$, we can write

$$\begin{aligned}
\mathbb{E}\left[f_1(X_{s+t_1}) \cdots f_{n+1}(X_{s+t_{n+1}}) \,|\, \mathcal{F}_s\right] &= \mathbb{E}\left[f_1(X_{s+t_1}) \cdots f_n(X_{s+t_n}) \mathbb{E}\left[f_{n+1}(X_{s+t_{n+1}}) \,|\, \mathcal{F}_{s+t_n}\right] \,|\, \mathcal{F}_s\right] \\
&= \mathbb{E}\left[f_1(X_{s+t_1}) \cdots f_n(X_{s+t_n}) P_{t_{n+1}-t_n} f_{n+1}(X_{s+t_n}) \,|\, \mathcal{F}_s\right] \\
&= \mathbf{E}_{X_s}\left[f_1(Z_{t_1}) \cdots f_n(Z_{t_n}) P_{t_{n+1}-t_n} f_{n+1}(Z_{t_n})\right] \\
&= \mathbf{E}_{X_s}\left[f_1(Z_{t_1}) \cdots f_{n+1}(Z_{t_{n+1}})\right],
\end{aligned}$$

where we have successively used the tower property of conditional expectations, the fact that X is a $(\mathcal{F}_t)_{t\geq 0}-$Markov process with transition semi-group $(P_t)_{t\geq 0}$, the induction hypothesis, and the fact that under $\mathbf{P}_x$, we have $Z \sim \text{CTMP}(\delta_x, (P_t)_{t\geq 0})$, for any $x \in \mathbb{X}$. □

Unlike in the discrete case, the strong Markov property requires additional assumptions.

Exercise 3.3 (A pathology). *Let us modify the Brownian semi-group on $\mathbb{X} = \mathbb{R}$ as follows:*

$$\forall x \in \mathbb{X}, \forall t \geq 0, \qquad P_t(x, \cdot) \quad := \quad \begin{cases} \mathcal{N}(x, t) & \text{if } x \neq 0 \\ \delta_0 & \text{if } x = 0. \end{cases}$$

1. *Check that this defines a transition semi-group. Does it have the Feller property?*
2. *Fix $x \neq 0$ and a Brownian motion B. Show that $X := x + B$ has law $\text{CTMP}(\delta_x, (P_t)_{t\geq 0})$.*
3. *Set $\tau = \inf\{t \geq 0 \colon X_\tau = 0\}$. Prove that τ is a stopping time w.r.t. $\mathcal{F}^X$, and that it is a.-s. finite.*
4. *Do we have $\mathbf{E}[f(X_{t+\tau})|\mathcal{F}_\tau] = P_t f(X_\tau)$, for all $t \geq 0$ and $f \in L^\infty(\mathbb{X})$?*

The latter are easily seen to be satisfied, in particular, when X is a Feller process.

Theorem 3.6 (Strong Markov property). *Assume that X has right-continuous trajectories and that*

$$\forall t \geq 0, \quad P_t \mathcal{C}_b(\mathbb{X}) \quad \subseteq \quad \mathcal{C}_b(\mathbb{X}). \tag{3.14}$$

Then, X satisfies the strong Markov property: for any stopping time τ and any function $F \in L^\infty(\mathbb{X}^{[0,\infty)})$,

$$\mathbb{E}\left[F((X_{\tau+t})_{t\geq 0}) \,|\, \mathcal{F}_\tau\right] \quad = \quad \mathbb{E}_{X_\tau}\left[F(Z)\right],$$

almost-surely on the event $\{\tau < \infty\}$.

Proof. Using successive conditioning and the monotone class lemma as in the above argument, the claim reduces to the following special case: for every $t \geq 0$ and $f \in \mathcal{C}_b(\mathbb{X})$,

$$\mathbb{E}\left[f\left(X_{\tau+t}\right) \mid \mathcal{F}_\tau\right] \quad = \quad (P_t f)(X_\tau), \tag{3.15}$$

almost-surely on the event $\{\tau < \infty\}$. As in the case of discrete state spaces, we first assume that τ takes values in $\mathbb{Q}_+ \cup \{+\infty\}$. For any $A \in \mathcal{F}_\tau$ and $f \in \mathcal{C}_b(\mathbb{X})$, we can write

$$\begin{aligned}
\mathbb{E}\left[f\left(X_{\tau+t}\right)\mathbf{1}_{(A\cap\{\tau<\infty\})}\right] &= \sum_{s\in\mathbb{Q}_+} \mathbb{E}\left[f\left(X_{s+t}\right)\mathbf{1}_{(A\cap\{\tau=s\})}\right] \\
&= \sum_{s\in\mathbb{Q}_+} \mathbb{E}\left[\mathbb{E}\left[f\left(X_{s+t}\right)\mathbf{1}_{(A\cap\{\tau=s\})} \mid \mathcal{F}_s\right]\right] \\
&= \sum_{s\in\mathbb{Q}_+} \mathbb{E}\left[\mathbf{1}_{(A\cap\{\tau=s\})}(P_t f)(X_s)\right] \\
&= \mathbb{E}\left[\mathbf{1}_{(A\cap\{\tau<\infty\})}(P_t f)(X_\tau)\right],
\end{aligned}$$

which proves (3.15). Note that we have here crucially used the fact that $A \cap \{\tau = s\} \in \mathcal{F}_s$, which follows from the decomposition $A \cap \{\tau = s\} = A \cap \{\tau \leq s\} \cap \bigcap_{n=1}^{\infty} \{\tau \leq s - \frac{1}{n}\}^c$ and the fact that τ is a stopping time. Now, in the general case, we fix $n \geq 1$ and consider the stopping time

$$\tau_n \quad := \quad \frac{\lceil n\tau \rceil}{n},$$

with $\lceil \infty \rceil := \infty$. Since the latter is $\mathbb{Q}_+ \cup \{+\infty\}-$valued, the first part of the proof allows us to write, for any $f \in \mathcal{C}_b(\mathbb{X})$ and any $A \in \mathcal{F}_\tau \subseteq \mathcal{F}_{\tau_n}$,

$$\mathbb{E}\left[\mathbf{1}_{(A\cap\{\tau<\infty\})} f\left(X_{\tau_n}\right)\right] \quad = \quad \mathbb{E}\left[\mathbf{1}_{(A\cap\{\tau<\infty\})} P_t f(X_{\tau_n})\right].$$

We finally send $n \to \infty$, and obtain (3.15) by dominated convergence. Indeed, on the event $\tau < \infty\}$, we have $\tau_n \to \tau+$ by construction, so that $X_{\tau_n} \to X_\tau$ by right-continuity of X, implying in turn that $f(X_{\tau_n}) \to f(X_\tau)$ and $P_t f(X_{\tau_n}) \to P_t f(X_\tau)$, by continuity of f and $P_t f$. □

Note that when $\mathbb{X}$ is locally finite, we can replace $\mathcal{C}_b(\mathbb{X})$ with $\mathcal{C}_0(\mathbb{X})$ in (3.14). The strong Markov property is, of course, a very useful tool for the analysis of Markov processes. Here is an emblematic example.

Exercise 3.4 (Reflection principle). *Let $B = (B_t)_{t\geq 0}$ be a one-dimensional Brownian motion.*

1. *Fix a stopping time τ. Prove that the process $\widetilde{B} = (\widetilde{B}_t)_{t\geq 0}$ defined as follows is a Brownian motion:*

$$\forall t \geq 0, \qquad \widetilde{B}_t \quad := \quad B_t \mathbf{1}_{(\tau>t)} + (2B_\tau - B_t)\mathbf{1}_{\tau\leq t}.$$

2. *Fix $a > 0$ and set $\tau := \{\inf t \geq 0 \colon B_t \geq a\}$. Show that for any $t \geq 0$,*

$$\{\tau \leq t\} \quad = \quad \{B_t \geq a\} \dot{\cup} \{\widetilde{B}_t > a\}.$$

3. *Deduce that for each $t \geq 0$, $\sup_{s\in[0,t]} B_s$ has the same law as $|B_t|$.*

3.6 The Hille-Yosida theorem

As in the case of discrete state spaces, we would like to have a representation of the form

$$\forall t \geq 0, \qquad P_t \;=\; e^{tQ}, \tag{3.16}$$

for our transition semi-group, where Q is a suitably defined infinitesimal generator. The beautiful Hille-Yosida theorem allows one to do so, in a much broader generality. We only give a partial description of the theory, restricted to our needs. Consider a Banach space $(E, \|\cdot\|)$, and a family of linear operators $(P_t)_{t\geq 0}$ from E to E, with the following properties:

(i) Semi-group: $P_0 = \mathrm{Id}$, and $P_{t+s} = P_t P_s$ for every $t, s \geq 0$.

(ii) Contraction: $\|P_t v\| \leq \|v\|$ for each $t \geq 0$ and $v \in E$.

(iii) Strong continuity: $P_t v \to v$ as $t \to 0$, for each $v \in E$.

The family $(P_t)_{t\geq 0}$ is naturally called a strongly continuous contraction semi-group on E. Note that for each $v \in E$, the function $t \mapsto P_t v$ is uniformly continuous on $\mathbb{R}_+$, because

$$\|P_t v - P_s v\| \;=\; \|P_{t\wedge s}(P_{|t-s|}v - v)\| \;\leq\; \|P_{|t-s|}v - v\| \xrightarrow[|t-s|\to 0]{} 0.$$

Note also that for each $t \geq 0$, the action of P_t on E is continuous:

$$\forall u, v \in E, \qquad \|P_t u - P_t v\| \;\leq\; \|u - v\|.$$

With (3.16) in mind, we define the infinitesimal generator $Q\colon \mathrm{Dom}(Q) \to E$ by the formula

$$Qv \;:=\; \lim_{t\to 0} \left\{ \frac{P_t v - v}{t} \right\}, \tag{3.17}$$

for any $v \in E$ such that the limit exists. The set $\mathrm{Dom}(Q)$ of vectors v with this property is clearly a linear subspace of E, and the resulting map $Q\colon \mathrm{Dom}(Q) \to E$ is linear.

Remark 3.7 (Banach-valued Riemann integrals)**.** *Let us recall that the Riemann integral of a continuous function $f\colon [a,b] \to E$ is obtained by the Riemann-sum approximation*

$$\int_a^b f(t)\,\mathrm{d}t \;=\; \lim_{n\to\infty} \left[\sum_{k=1}^{n} f(t_k^n)(t_k^n - t_{k-1}^n) \right],$$

along any subdivision $a = t_0^n \leq t_1^n \leq \ldots \leq t_n^n = b$ of $[a,b]$ whose mesh $\max_{1\leq k\leq n} t_k^n - t_{k-1}^n$ tends to 0 as $n \to \infty$. The existence of the limit and the irrelevance of the chosen subdivision are classical, and left to the reader. The following familiar properties will be used without notice in the sequel:

1. *$f \mapsto \int_a^b f(t)\,\mathrm{d}t$ is linear from $\mathcal{C}^0([a,b], \mathbb{X})$ to $\mathbb{X}$.*
2. *$T\left(\int_a^b f(t)\,\mathrm{d}t\right) = \int_a^b (Tf)(t)\,\mathrm{d}t$ for any bounded linear operator $T\colon E \to F$, with F Banach.*
3. *$\|\int_a^b f(t)\,\mathrm{d}t\| \leq \int_a^b \|f(t)\|\,\mathrm{d}t$;*
4. *$\int_a^c f(t)\,\mathrm{d}t = \int_a^b f(t)\,\mathrm{d}t + \int_c^b f(t)\,\mathrm{d}t$ for any $c \in [a,b]$;*
5. *$\frac{1}{\varepsilon}\int_a^{a+\varepsilon} f(t)\,\mathrm{d}t \to f(a)$ as $\varepsilon \to 0$;*

6. *If f is $\mathcal{C}^1$, then $f(t) = f(a) + \int_a^t f(s)\,\mathrm{d}s$ for all $t \in [a,b]$.*

Of course, one can also define $f \mapsto \int_a^\infty f(t)\,\mathrm{d}t$ provided $\int_a^\infty \|f(t)\|\,\mathrm{d}t < \infty$...

Theorem 3.7 (Hille-Yosida)**.** *Let $(P_t)_{t\geq 0}$ be a strongly continuous contraction semi-group on a Banach space E, and let Q be its infinitesimal generator.*

1. *If $v \in \mathrm{Dom}(Q)$, then $P_t v \in \mathrm{Dom}(Q)$ for all $t \geq 0$. Moreover, $t \mapsto P_t v$ is in $\mathcal{C}^1(\mathbb{R}_+, E)$, and*
$$\forall t \geq 0, \qquad \frac{\mathrm{d}P_t v}{\mathrm{d}t} \;=\; QP_t v \;=\; P_t Q v. \tag{3.18}$$
2. *$\mathrm{Dom}(Q)$ is dense in E, and the graph $\Gamma_Q := \{(v, Qv) : v \in \mathrm{Dom}(Q)\}$ is closed in $E \times E$.*
3. *For each $\lambda > 0$, the operator $\lambda \mathrm{Id} - Q : \mathrm{Dom}(Q) \to E$ is invertible, and its inverse R_λ is given by*
$$\forall v \in E, \qquad R_\lambda v \;:=\; \int_0^\infty e^{-\lambda t} P_t v\,\mathrm{d}t. \tag{3.19}$$
4. *The generator Q characterizes the semi-group $(P_t)_{t\geq 0}$: if $(P'_t)_{t\geq 0}$ is a strongly continuous contraction semi-group on E whose generator Q' coincides with Q, then $P'_t = P_t$ for all $t \geq 0$.*

Proof of item 1. Fix $v \in \mathrm{Dom}(Q)$, $t \geq 0, s > 0$. Using the properties of $(P_t)_{t\geq 0}$, we have
$$\left\| \frac{P_s(P_t v) - P_t v}{s} - P_t Q v \right\| \;=\; \left\| P_t \left(\frac{P_s v - v}{s} - Qv \right) \right\| \;\leq\; \left\| \frac{P_s v - v}{s} - Qv \right\|,$$
which tends to 0 as $s \to 0$ by definition of Q. Thus, $P_t v \in \mathrm{Dom}(Q)$ and $QP_t v = P_t Q v$. This computation shows that the function $t \mapsto P_t v$ is right-differentiable on $\mathbb{R}_+$, with right-derivative $t \mapsto QP_t v$. Similarly, for any $s \in (0, t]$, we can write
$$\begin{aligned} \left\| \frac{P_t v - P_{t-s} v}{s} - P_t Q v \right\| &= \left\| P_{t-s} \left(\frac{P_s v - v}{s} - P_s Q v \right) \right\| \\ &\leq \left\| \frac{P_s v - v}{s} - P_s Q v \right\| \\ &\leq \left\| \frac{P_s v - v}{s} - Qv \right\| + \|P_s Q v - Q v\| \;\xrightarrow[s\to 0]{}\; 0, \end{aligned}$$
thanks to the definition of Q and the continuity property of the semi-group. Thus, the function $t \mapsto P_t v$ is in fact differentiable on $\mathbb{R}_+$, with derivative $t \mapsto QP_t v = P_t Q v$. Since the latter is continuous, the first claim is established. In integral form, this reads:
$$\forall t \geq 0, \qquad P_t v \;=\; v + \int_0^t QP_s v\,\mathrm{d}s \;=\; v + \int_0^t P_s Q v\,\mathrm{d}s, \tag{3.20}$$
an identity which will be used several times in the sequel. □

Proof of item 2. Fix $v \in E$ and $\varepsilon > 0$, and consider the approximation
$$v_\varepsilon \;:=\; \frac{1}{\varepsilon} \int_0^\varepsilon P_t v\,\mathrm{d}t,$$
which tends to v as $\varepsilon \to 0$ by the continuity property of $(P_t)_{t\geq 0}$. Now, for $s \in (0, \varepsilon)$, we have
$$\begin{aligned} \frac{P_s v_\varepsilon - v_\varepsilon}{s} &= \frac{1}{\varepsilon s} \left(\int_s^{\varepsilon+s} P_t v\,\mathrm{d}t - \int_0^\varepsilon P_t v\,\mathrm{d}t \right) \\ &= \frac{1}{\varepsilon s} \int_\varepsilon^{\varepsilon+s} P_t v\,\mathrm{d}t - \int_0^s P_t v\,\mathrm{d}t \\ &= \frac{1}{\varepsilon s} \int_0^s (P_{t+\varepsilon} v - P_t v)\,\mathrm{d}t, \end{aligned}$$

which tends to $\frac{P_\varepsilon v - v}{\varepsilon}$ as $s \to 0$ because the integrand is continuous. Thus, $v_\varepsilon \in \mathrm{Dom}(Q)$, showing that $\mathrm{Dom}(Q)$ is dense. Now, consider a sequence $(v_n, w_n)_{n\geq 1}$ of points in the graph Γ_Q, and suppose that it converges to some point $(v, w) \in E \times E$. Recalling (3.20), we have

$$\forall t \geq 0, \qquad P_t v_n \;=\; v_n + \int_0^t P_s w_n \,\mathrm{d}s,$$

for each $n \geq 1$. Sending $n \to \infty$, we deduce that

$$\forall t \geq 0, \qquad P_t v \;=\; v + \int_0^t P_s w \,\mathrm{d}s.$$

In particular, this implies that $(P_t v - v)/t \to w$ as $t \to 0$, showing that $(v, w) \in \Gamma_Q$. □

Proof of item 3. Fix $\lambda > 0$ and $v \in E$. We must show that $u := R_\lambda v$ is one and the only solution to

$$u \in \mathrm{Dom}(Q) \quad \text{and} \quad (\lambda \mathrm{Id} - Q)u \;=\; v. \tag{3.21}$$

First note that $R_\lambda v$ is well defined, because the function $t \mapsto e^{-\lambda t} P_t v$ is continuous on $\mathbb{R}_+$ and bounded by $t \mapsto e^{-\lambda t}\|v\|$, which is integrable. Now, for any $s > 0$, we have

$$P_s R_\lambda v \;=\; \int_0^\infty e^{-\lambda t} P_{t+s} v \,\mathrm{d}t \;=\; e^{\lambda s}\left(R_\lambda v - \int_0^s e^{-\lambda t} P_t v \,\mathrm{d}t\right).$$

Consequently,

$$\frac{P_s R_\lambda v - R_\lambda v}{s} \;=\; \frac{e^{\lambda s} - 1}{s} R_\lambda v - \frac{1}{s}\int_0^s e^{-\lambda t} P_t v \,\mathrm{d}t \;\underset{s\to 0}{\longrightarrow}\; \lambda R_\lambda v - v,$$

showing that $u := R_\lambda v$ solves (3.21). Now, consider another solution u', and set $w := u - u'$. By linearity, we then have $w \in \mathrm{Dom}(Q)$ and $Qw = \lambda w$. But the second item in the theorem guarantees that $t \mapsto e^{-\lambda t} P_t w$ is continuously differentiable, with derivative

$$\forall t \geq 0, \qquad \frac{\mathrm{d}}{\mathrm{d}t}\left(e^{-\lambda t} P_t w\right) \;=\; e^{-\lambda t} P_t Q w - \lambda e^{-\lambda t} P_t w \;=\; 0.$$

Thus, $P_t w = e^{\lambda t} w$ for all $t \geq 0$, which contradicts the contraction property unless $w = 0$. □

Proof of item 4. Let $(P'_t)_{t\geq 0}$ be a strongly continuous contraction semi-group on E whose generator Q' coincides with Q. Then $\lambda \mathrm{Id} - Q = \lambda \mathrm{Id} - Q'$ for all $\lambda > 0$, so by the previous item,

$$\forall \lambda > 0, \qquad \int_0^\infty e^{-\lambda t} P_t v \,\mathrm{d}t \;=\; \int_0^\infty e^{-\lambda t} P'_t v \,\mathrm{d}t.$$

Thus, the continuous bounded function $f\colon \mathbb{R}_+ \to \mathbb{X}$ defined by $f(t) = P_t v - P'_t v$ satisfies

$$\forall \lambda > 0, \qquad \int_0^\infty e^{-\lambda t} f(t) \,\mathrm{d}t \;=\; 0.$$

In particular, for any continuous linear form $\phi\colon E \to \mathbb{R}$, we have

$$\forall \lambda > 0, \qquad \int_0^\infty e^{-\lambda t} \phi\left(f(t)\right) \,\mathrm{d}t \;=\; \phi\left(\int_0^\infty e^{-\lambda t} f(t) \,\mathrm{d}t\right) \;=\; 0,$$

which forces $\phi(f(t)) = 0$ for every $t \geq 0$, by the injectivity of the Laplace transform on $\mathcal{C}^0(\mathbb{R}_+, \mathbb{R})$. This is true for any continuous linear form $\phi\colon E \to \mathbb{R}$, so the Hahn-Banach theorem guarantees that for all $t \geq 0$, $f(t) = 0$, i.e. $P_t v = P_t v'$. Since $v \in E$ is arbitrary, the claim is proved. □

Remark 3.8 (The bounded case)**.** *The situation is particuarly nice when Q is bounded, meaning that*

$$\|Q\| \quad := \quad \sup_{v \in \mathrm{Dom}(Q) \setminus \{0\}} \frac{\|Qv\|}{\|v\|} \quad < \quad \infty. \tag{3.22}$$

Indeed, we know that any $v \in E$ is the limit of some $(v_n)_{n \geq 0}$ in $\mathrm{Dom}(Q)$, and (3.22) implies that $(Qv_n)_{n \geq 0}$ is Cauchy, hence convergent. Since Γ_Q is closed, we must have $v \in \mathrm{Dom}(Q)$, showing that

$$\mathrm{Dom}(Q) \quad = \quad E.$$

Thus, Q^n is well defined for all $n \in \mathbb{N}$, and (3.22) ensures that for each $t \in \mathbb{R}$, the series

$$e^{tQ} \quad := \quad \sum_{n=0}^{\infty} \frac{t^n Q^n}{n!}$$

is convergent in the Banach algebra of bounded linear operators from E to E. From this definition, it easily follows that $t \mapsto e^{tQ}$ is continuously differentiable, with derivative

$$\forall t \in \mathbb{R}, \qquad \frac{\mathrm{d}e^{tQ}}{\mathrm{d}t} \quad = \quad Qe^{tQ} \quad = \quad e^{tQ}Q.$$

Recalling the Kolmogorov equation (3.18), we see that for any $v \in E$, the function $t \mapsto e^{-tQ}P_t v$ is differentiable with derivative 0. Thus, it must be constant equal to v, i.e. $P_t v = e^{tQ}v$. Thus,

$$\forall t \geq 0, \qquad P_t \quad = \quad e^{tQ},$$

which gives a perfectly rigorous meaning to (3.16).

3.7 Generator of a Markov process

In this section, we consider a Markov process X in a Polish space $(\mathbb{X}, \mathcal{B})$, with transition semi-group $(P_t)_{t \geq 0}$. Recall that the latter acts on the the Banach space $E = L^\infty(\mathbb{X})$ via the formula

$$\forall x \in \mathbb{X}, \quad \forall t \geq 0, \quad P_t f(x) \quad := \quad \int_{\mathbb{X}} f(y) P_t(x, \mathrm{d}y),$$

and the semi-group and contraction properties from the previous section are clearly satisfied. If in addition, the strong continuity property

$$\forall f \in L^\infty(\mathbb{X}), \qquad \|P_t f - f\|_\infty \quad \xrightarrow[t \to 0^+]{} \quad 0, \tag{3.23}$$

holds, then we can apply the previous result to define the generator Q of our process X and give a rigorous meaning to (3.16). The above assumption is satisfied, for example, when the state space $\mathbb{X}$ is discrete and (2.19) holds, in which case $\|Q\| < \infty$ and we recover exactly the results from Section 2.4. However, the assumption (3.23) is far too restrictive: it even fails for Brownian motion. A simple workaround consists in restricting our ambient Banach space so as to enforce the strongly continuity of the semi-group. Specifically, we define

$$E \quad := \quad \left\{ f \in L^\infty(\mathbb{X}) \colon \|P_t f - f\|_\infty \xrightarrow[t \to 0+]{} 0 \right\}.$$

The following lemma shows that $(E, \|\cdot\|_\infty)$ is indeed a Banach space, and that $(P_t)_{t \geq 0}$ is a strongly continuous contraction semi-group once restricted to E.

Lemma 3.3 (Restricted semi-group). *The set E is a closed linear subspace of $L^\infty(\mathbb{X})$, hence a Banach space once equipped with the induced norm $\|\cdot\|_\infty$. Moreover, $P_t E \subseteq E$ for each $t \geq 0$, and the restriction of $(P_t)_{t\geq 0}$ to E is a strongly continuous contraction semi-group on E.*

Proof. Let $(f_n)_{n\geq 1}$ be elements of E that converge in L^∞ to some f. Then, for $n \in \mathbb{N}, t \geq 0$,

$$\begin{aligned} \|P_t f - f\|_\infty &\leq \|P_t f - P_t f_n\|_\infty + \|P_t f_n - f_n\|_\infty + \|f_n - f\|_\infty \\ &\leq 2\|f_n - f\|_\infty + \|P_t f_n - f_n\|_\infty, \end{aligned}$$

because P_t is a contraction. Sending first $t \to 0+$, and then $n \to \infty$, we deduce that $f \in E$, showing that E is closed. The fact that it is a linear space is clear. Now, for $f \in E$ and $t, s \geq 0$, the semi-group property and the fact that $(P_t)_{t\geq 0}$ are contractions on L^∞ allow us to write

$$\|P_t f - P_s f\|_\infty \leq \big\|P_{|t-s|} f - f\big\|_\infty \xrightarrow[|t-s|\to 0]{} 0,$$

showing that $t \mapsto P_t f$ is continuous. In particular, $P_t f \in E$ for each $t \geq 0$, as desired. □

We can then use the Hille-Yosida theorem to define the infinitesimal generator $Q\colon \mathrm{Dom}(Q) \subseteq E \to E$ of our Markov process, and obtain the following result.

Corollary 3.1 (Generator of a Markov process). *The generator Q enjoys the following properties:*

1. *$\mathrm{Dom}(Q)$ is dense in E, Γ_Q is closed in E^2, and the action of $(P_t)_{t\geq 0}$ on E is determined by Q.*
2. *$P_t \mathrm{Dom}(Q) \subseteq \mathrm{Dom}(Q)$ for each $t \geq 0$.*
3. *For every $f \in \mathrm{Dom}(Q)$, the function $t \mapsto P_t f$ is $\mathcal{C}^1$ and satisfies the Kolmogorov equations*

$$\forall t \geq 0, \qquad \frac{\mathrm{d}P_t f}{\mathrm{d}t} = QP_t f = P_t Q f. \tag{3.24}$$

Of course, the larger the continuity set E is, the stronger this result will be. The situation becomes particularly nice when E is rich enough to be measure-determining, meaning that

$$\forall \mu, \nu \in \mathcal{P}(\mathbb{X}), \qquad \mu = \nu \iff \left(\forall f \in E, \qquad \int f \,\mathrm{d}\mu = \int f \,\mathrm{d}\nu\right). \tag{3.25}$$

Here is a list of classical examples of measure-determining classes.

- $\mathcal{C}_b(\mathbb{X})$, when $\mathbb{X}$ is Polish;
- $\mathcal{C}_c(\mathbb{X})$, when $\mathbb{X}$ is a locally compact Polish space;
- $\mathcal{C}_c^\infty(\mathbb{X})$ or the Schwartz space $\mathcal{S}(\mathbb{X})$, when $\mathbb{X} = \mathbb{R}^d$;
- $\{x \mapsto \cos(\theta x), x \mapsto \sin(\theta x)\colon \theta \in [-\pi, \pi)\}$, when $\mathbb{X} = \mathbb{R}$;
- $\{x \mapsto x^n\colon n \in \mathbb{N}\}$, when $\mathbb{X}$ is a compact interval of $\mathbb{R}$;
- $\{x \mapsto e^{-\lambda x}\colon \lambda > 0\}$, when $\mathbb{X} = [0, \infty)$.

Corollary 3.2 (Law of X). *Under the condition (3.25), the generator Q determines the semi-group $(P_t)_{t\geq 0}$. In particular, the law of X is determined by the pair (μ, Q), where $\mu := \mathrm{law}(X_0)$. We write*

$$X \sim \mathrm{CTMP}(\mu, Q).$$

Proof. Suppose that $(P'_t)_{t\geq 0}$ is another transition semi-group whose infinitesimal generator Q' coincides with Q. This implies $\mathrm{Dom}(Q') = \mathrm{Dom}(Q)$ hence $E' = E$, by taking the closure in $L^\infty(\mathbb{X})$. Now fix $t \geq 0$ and $x \in \mathbb{X}$, and use the last item in the previous corollary to write

$$\forall f \in E, \qquad \int f(y) P_t(x, \mathrm{d}y) \;=\; P_t f(x) \;=\; P'_t f(x) \;=\; \int f(y) P'_t(x, \mathrm{d}y).$$

In view of the measure-determining assumption (3.25), this forces

$$\forall B \in \mathcal{B}, \qquad P_t(x, B) \;=\; P'_t(x, B),$$

showing that $(P'_t)_{t\geq 0} = (P_t)_{t\geq 0}$. The second claim readily follows from this and Remark 3.5. □

Here is another nice consequence of the measure-determining assumption (3.25).

Corollary 3.3 (Stationarity). *Assume (3.25) and fix $\mu \in \mathcal{P}(\mathbb{X})$. Then, the following are equivalent:*

1. *The law $\mathrm{CTMC}(\mu, Q)$ is stationary (i.e. invariant under deterministic time shifts);*
2. *$\mu P_t = \mu$ for all $t \geq 0$;*
3. *$\mu Q = 0$, in the sense that $\int_{\mathbb{X}} (Qf)\, \mathrm{d}\mu = 0$ for all $f \in \mathrm{Dom}(Q)$.*

Proof. Fix $t \geq 0$. By the weak Markov property, we have $(X_{t+s})_{s\geq 0} \sim \mathrm{CTMP}(\mu_t, Q)$, where $\mu_t := \mathrm{law}(X_t)$ denotes the marginal law at time t. Moreover, for any $B \in \mathcal{B}$ we have

$$\mathbb{P}(X_t \in B) \;=\; \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_B(X_t) \,|\, X_0\right]\right] \;=\; \mathbb{E}[P_t(X_0, B)] \;=\; \int_{\mathbb{X}} \mu_0(\mathrm{d}x) P_t(x, B) \;=\; (\mu_0 P_t)(B),$$

showing that $\mu_t = \mu_0 P_t$. Thus, we have

$$\begin{aligned}
\mathrm{CTMC}(\mu, Q) \text{ is stationary} &\iff t \mapsto \mathrm{CTMP}(\mu P_t, Q) \text{ is constant} \\
&\iff t \mapsto \mu P_t \text{ is constant} \\
&\iff t \mapsto \mu P_t f \text{ is constant for each } f \in E \\
&\iff t \mapsto \mu P_t f \text{ is constant for each } f \in \mathrm{Dom}(Q),
\end{aligned}$$

where we have used the previous corollary, (3.25), and the density of $\mathrm{Dom}(Q)$ in E. Now, in view of the Kolmogorov equations (3.24), we have for each $t \geq 0$ and $f \in \mathrm{Dom}(Q)$,

$$\begin{aligned}
\mu P_t f - \mu f &= \int_{\mathbb{X}} \left(\int_0^t Q P_s f(x)\, \mathrm{d}s \right) \mu(\mathrm{d}x) \\
&= \int_0^t \left(\int_{\mathbb{X}} Q P_s f(x)\, \mathrm{d}\mu(\mathrm{d}x) \right) \mathrm{d}s \\
&= \int_0^t (\mu Q P_s f)\, \mathrm{d}s.
\end{aligned}$$

Since the integrand is continuous, we deduce that $t \mapsto \mu P_t f$ is constant if and only if $\mu Q P_t f = 0$ for all $t \geq 0$. This trivially implies that $\mu Q f = 0$ (take $t = 0$), and conversely, if $\mu Q f = 0$ for all $f \in \mathrm{Dom}(Q)$, then $\mu Q P_t f = 0$ for all $f \in \mathrm{Dom}(Q)$, because $P_t \mathrm{Dom}(Q) \subseteq \mathrm{Dom}(Q)$. □

Here is a last, very useful property of the infinitesimal generator of a Markov process.

Corollary 3.4 (Dynkin's lemma). *If X is a right-continuous $(\mathcal{F}_t)_{t\geq 0}$−Markov process with generator Q, then for any $f \in \mathrm{Dom}(Q)$, the process M^f defined as follows is a $(\mathcal{F}_t)_{t\geq 0}$-martingale.*

$$\forall t \geq 0, \qquad M^f_t \;:=\; f(X_t) - f(X_0) - \int_0^t Q f(X_s)\, \mathrm{d}s,$$

Proof. Being right-continuous and adapted, X is progressively measurable. Also, $Qf \in E \subseteq L^\infty(\mathbb{X})$. Thus, the process M^f is well-defined, adapted and integrable. Now, for $t, s \geq 0$, we have

$$\mathbb{E}\left[M^f_{t+s} - M^f_t \mid \mathcal{F}_s\right] \quad = \quad \mathbb{E}\left[F_t\left((X_{u+s})_{u\geq 0}\right) \mid \mathcal{F}_s\right] \quad = \quad \mathbf{E}_{X_s}\left[F_t(Z)\right],$$

by the Markov property, where Z denotes the identity map on the canonical space $\left(\mathbb{X}^{[0,\infty)}, \mathcal{E}^{\otimes[0,\infty)}\right)$ and $\mathbf{P}_x := \mathrm{CTMP}(\delta_x, (P_t)_{t\geq 0})$, and where we have introduced the function

$$F_t(Z) \quad := \quad f(Z_t) - f(Z_0) - \int_0^t Qf(Z_u)\,\mathrm{d}u.$$

Finally, for any $x \in \mathbb{X}$, we have by construction

$$\mathbb{E}_x[F_t(Z)] \quad = \quad P_t f(x) - f(x) - \int_0^t P_u Q f(x)\,\mathrm{d}u,$$

which is zero thanks to the Kolmogorov equation (3.24). □

In fact, Dynkins' lemma admits a converse, which yields a martingale characterization of Q.

Corollary 3.5 (Martingale characterization). *For each $x \in \mathbb{X}$, consider a right-continuous process $X^x \sim \mathrm{CTMP}(\delta_x, (P_t)_{t\geq 0})$. Then, a pair $(f, g) \in E^2$ is in Γ_Q if and only if the process M^x defined by*

$$\forall t \geq 0, \qquad M^x_t \quad := \quad f(X^x_t) - f(X^x_0) - \int_0^t g(X^x_s)\,\mathrm{d}s,$$

is a martingale for each $x \in \mathbb{X}$.

Proof. The 'only if' part is (a special case of) the previous corollary. Conversely, suppose that M^x is a martingale for all $x \in \mathbb{X}$, and fix $t > 0$. Then, we have $\mathbb{E}[M^x_t] = \mathbb{E}[M^x_0] = 0$ for all $x \in \mathbb{X}$, i.e.

$$P_t f - f - \int_0^t P_s g\,\mathrm{d}s \quad = \quad 0.$$

Now, $s \mapsto P_s g$ is continuous because $g \in E$, so we may divide through by $t > 0$ and send $t \to 0$ to deduce that $f \in \mathrm{Dom}(Q)$ and that $Qf = g$, which means that $(f, g) \in \Gamma_Q$. □

For concreteness, let us end this section with an explicit computation of the action of the generator on a measure-determining class of test functions, in two important special cases.

Exercise 3.5 (Deterministic shift). *Take $\mathbb{X} = \mathbb{R}^d$, fix a direction $v \in \mathbb{R}^d$ and let Q be the generator of the deterministic process $X_t = X_0 + tv$. Prove that $\mathrm{Dom}(Q)$ contains the space $\mathcal{C}^1_c(\mathbb{X})$ of compactly supported continuously differentiable functions, and that for any $f \in \mathcal{C}^1_c(\mathbb{X})$,*

$$Qf \quad = \quad \langle v, \nabla f \rangle.$$

Exercise 3.6 (Brownian motion). *Take $X = \mathbb{R}^d$, and let Q be the generator of Brownian motion. Prove that $\mathrm{Dom}(Q)$ contains $\mathcal{C}^2_c(\mathbb{X})$, and that for any $f \in \mathcal{C}^2_c(\mathbb{X})$,*

$$Qf \quad = \quad \frac{1}{2}\Delta f.$$

3.8 Markov diffusions

This final section is devoted to the study of Markov diffusions, which form a rich and important class of Markov processes on the Euclidean space $\mathbb{X} = \mathbb{R}^d$, equipped with its borel $\sigma-$field $\mathcal{B} = \mathcal{B}(\mathbb{R}^d)$. We assume that the reader is familiar with stochastic calculus and, in particular, the theory of stochastic differential equations (SDE) with time-homogenous coefficients.

Throughout this section, we consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F})_{t\geq 0}, \mathbb{P})$ on which is defined a $d-$dimensional $(\mathcal{F}_t)_{t\geq 0}-$Brownian motion $B = (B_t)_{t\geq 0}$. Also, we fix two Lipschitz function $b\colon \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$. Given an initial condition $X_0 \in L^2(\Omega, \mathcal{F}_0, \mathbb{P})$, we let $X = (X_t)_{t\geq 0}$ denote the unique solution, in the space $\mathbb{M}^2$, of the stochastic differential equation

$$\mathrm{d}X_t \;=\; b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}B_t, \tag{3.26}$$

with initial condition X_0. In integral form, this reads

$$\forall t \geq 0, \qquad X_t \;=\; X_0 + \int_0^t b(X_u)\,\mathrm{d}u + \int_0^t \sigma(X_u)\,\mathrm{d}B_u. \tag{3.27}$$

Note that the process X is continuous. Recall also that it is in fact adapted to the filtration generated by X_0 and the Brownian motion B, allowing us to write, for each $t \geq 0$,

$$X_t \;=\; \Psi_t\left(X_0, (B_s)_{s\in[0,t]}\right), \tag{3.28}$$

for some measurable function $\Psi_t\colon \mathbb{X}\times\mathbb{X}^{[0,t]} \to \mathbb{X}$, which only depends on t, b and σ. For any state $x \in \mathbb{R}^d$, any time $t \geq 0$, and any Borel set $A \subseteq \mathbb{R}^d$, we introduce the notation

$$P_t(x, A) \;:=\; \mathbb{P}\left(X_t^x \in A\right), \quad \text{where} \quad X_t^x \;:=\; \Psi_t\left(x, (B_s)_{s\in[0,t]}\right).$$

Theorem 3.8 (Markov property). *X is a $(\mathcal{F}_t)_{t\geq 0}-$Markov process with transition semi-group $(P_t)_{t\geq 0}$.*

Proof. We will naturally exploit the Markov property for Brownian motion, which we have established earlier, and which guarantees that the process $\widetilde{B} := (B_{t+s} - B_s)_{t\geq 0}$ is a Brownian motion, independent of $\mathcal{F}_s$. First, the stochastic integral satisfies the change-of-variable formula

$$\int_s^{t+s} \phi_u\,\mathrm{d}B_u \;=\; \int_0^t \phi_{u+s}\,\mathrm{d}\widetilde{B}_u,$$

for any $\phi \in \mathbb{M}^2_{\mathrm{LOC}}$. This is clear in the elementary case where $\phi_t(\omega) = X(\omega)\mathbf{1}_{]u,v]}(t)$ with $X \in L^2(\Omega, \mathcal{F}_u, \mathbb{P})$, and the general case follows by linearity and density. In particular, we have

$$\begin{aligned} X_{t+s} &= X_s + \int_s^{t+s} b(X_u)\,\mathrm{d}u + \int_s^{t+s} \sigma(X_u)\,\mathrm{d}B_u \\ &= X_s + \int_0^t b(X_{u+s})\,\mathrm{d}u + \int_0^t \sigma(X_{u+s})\,\mathrm{d}\widetilde{B}_u. \end{aligned}$$

In other words, the shifted process $\widetilde{X} := (X_{t+s})_{t\geq 0}$ solves the well-posed SDE

$$\mathrm{d}\widetilde{X}_t \;=\; b(\widetilde{X}_t)\,\mathrm{d}t + \sigma(\widetilde{X}_t)\,\mathrm{d}\widetilde{B}_t,$$

initialized at $\widetilde{X}_0 = X_s$ and driven by the Brownian motion $\widetilde{B}$ on the filtered space $(\Omega, \mathcal{F}, (\widetilde{\mathcal{F}}_t)_{t\geq 0}, \mathbb{P})$, with $\widetilde{\mathcal{F}}_t := \mathcal{F}_{t+s}$. Thus, for any $t \geq 0$, we have $\widetilde{X}_t = \Psi_t(\widetilde{X}_s, (\widetilde{B}_u)_{u\in[0,t]})$, i.e.

$$X_{t+s} \;=\; \Psi_t(X_s, (\widetilde{B}_u)_{u\in[0,t]}).$$

Recalling that $\widetilde{B}$ is independent of $\mathcal{F}_s$, we can now invoke the freezing lemma to write $\mathbb{E}[f(X_{t+s})|\mathcal{F}_s] = F(X_s)$, for any $f \in L^\infty(\mathbb{X})$, where $F\colon \mathbb{R}^d \to \mathbb{R}$ is given by

$$F(x) \;=\; \mathbb{E}\left[f\left(\Psi_t(x,(\widetilde{B}_u)_{u\in[0,t]})\right)\right] \;=\; \mathbb{E}[f(X_t^x)] \;=\; P_t f(x).$$

Thus, $\mathbb{E}[f(X_{t+s})|\mathcal{F}_s] = P_t f(X_s)$, as desired. Note that the semi-group property automatically follows from the tower property of conditional expectation (see Remark 3.4). □

The following result shows that the strong continuity space E of the semi-group $(P_t)_{t\geq 0}$ is measure-determining, and provides an explicit access to the generator Q on a rich class.

Theorem 3.9 (Generator). *We have $\mathcal{C}_0(\mathbb{R}^d) \subseteq E$, $\mathcal{C}_c^2(\mathbb{R}^d) \subseteq \mathrm{Dom}(Q)$, and*

$$\forall f \in \mathcal{C}_c^2(\mathbb{R}^d), \qquad Qf(x) \;=\; \nabla f(x)\cdot b(x) + \frac{1}{2}\mathrm{Tr}\left[\sigma^T(x)\,(\mathrm{Hess} f(x))\,\sigma(x)\right].$$

The proof uses the following lemma, which shows that our Markov process X can not travel too fast. We write $|\cdot|$ for both the Euclidean norm in $\mathbb{R}^d$ and the Hilbert-Schmidt norm on $\mathbb{R}^{d\times d}$.

Lemma 3.4 (Speed control). *There is $c \in (0,\infty)$ such that for any $x \in \mathbb{R}^d$ and $t \geq 0$, we have*

$$\mathbb{E}\left[|X_t^x - x|^2\right] \;\leq\; ct(1+|x|^2)e^{ct^2}.$$

Proof. In view of (3.27), we have

$$\begin{aligned} X_t^x - x &= \int_0^t b(X_u^x)\,\mathrm{d}u + \int_0^t \sigma(X_u^x)\,\mathrm{d}B_u \\ &= tb(x) + \sigma(x)B_t + \int_0^t [b(X_u^x) - b(x)]\,\mathrm{d}u + \int_0^t [\sigma(X_u^x) - \sigma(x)]\,\mathrm{d}B_u. \end{aligned}$$

We now take expectations, use the Cauchy-Schwarz inequality, the isometry property of Itô's integral, and the Lipshitz assumption on b and σ to obtain

$$\mathbb{E}\left[|X_t^x - x|^2\right] \;\leq\; ct(1+t)\left(1+|x|^2\right) + c(1+t)\int_0^t \mathbb{E}\left[|X_u^x - x|^2\right]\mathrm{d}u,$$

where c depends only on the coefficient b, σ. The claim now follows from Gronwall's lemma. □

Proof of Theorem 3.9. Fix $f \in \mathcal{C}_0(\mathbb{R}^d)$, $\delta > 0$, and $r > 0$, and let us introduce the quantities

$$m(\delta) \;:=\; \sup_{|x-y|\leq\delta} |f(x) - f(y)| \quad \text{and} \quad v(r) \;:=\; \sup_{|x|>r} |f(x)|. \tag{3.29}$$

Then for all $x, y \in \mathbb{X}$, we have

$$|f(y) - f(x)| \;\leq\; 2v(r) + m(\delta) + 2\|f\|_\infty \mathbf{1}_{|x|\leq 2r, |y-x|>\delta} + 2\|f\|_\infty \mathbf{1}_{|x|>2r, |y-x|\geq\frac{|x|}{2}},$$

where we have used the fact that $|y| \leq r \leq |x|/2 \Rightarrow |y-x| \geq |x|/2$. Thus, for any $t \geq 0$,

$$\begin{aligned} \|P_t f - f\|_\infty &= \sup_{x\in\mathbb{R}^d} |\mathbb{E}\left[f(X_t^x) - f(x)\right]| \\ &\leq 2v(r) + m(\delta) + 2\|f\|_\infty \left[\sup_{|x|\leq 2r} \mathbb{P}(|X_t^x - x| > \delta) + \sup_{|x|>2r} \mathbb{P}(|X_t^x - x| > |x|/2)\right]. \end{aligned}$$

Now, our assumption of f ensures that the first two terms can be made arbitrarily small by choosing δ small enough and r large enough, and for any such choice, the last two terms tend

to 0 as $t \to 0$, by the previous lemma. Thus, $f \in E$, establishing that $E \supseteq \mathcal{C}_0(\mathbb{R}^d)$. Now, fix $f \in \mathcal{C}_c^2(\mathbb{R}^d)$. By Itô's formula any solution X to the SDE (3.26) satisfies

$$\begin{aligned} \mathrm{d}f(X_t) &= \nabla f(X_t) \cdot \mathrm{d}X_t + \frac{1}{2}\mathrm{Hess} f(X_t)\, \mathrm{d}X_t \cdot \mathrm{d}X_t \\ &= g(X_t)\, \mathrm{d}t + \nabla f(X_t) \cdot \sigma(X_t)\, \mathrm{d}B_t, \end{aligned}$$

where $g \colon \mathbb{R}^d \to \mathbb{R}$ is defined as follows:

$$\forall x \in \mathbb{R}^d, \qquad g(x) := \nabla f(x) \cdot b(x) + \frac{1}{2}\mathrm{Tr}\left[\sigma^T(x)\, (\mathrm{Hess} f(x))\, \sigma(x)\right].$$

In integral form, this reads

$$\forall t \geq 0, \qquad f(X_t) = f(X_0) + \int_0^t g(X_u)\, \mathrm{d}u + \int_0^t \nabla f(X_t) \cdot \sigma(X_t)\, \mathrm{d}B_t,$$

for any choice of the initial condition X_0. Since the integrand in the stochastic integral is bounded, its integral forms a martingale. Moreover, the functions f and g are in $\mathcal{C}_c(\mathbb{R}^d)$ hence in E. By the martingale characterization of the generator, we deduce that $f \in \mathrm{Dom}(Q)$ and $Qf = g$. □

Exercise 3.7 (Feller property). *Fix two initial conditions $x, y \in \mathbb{R}^d$. Show that*

$$\forall t \geq 0, \qquad \mathbb{E}\left[|X_t^x - X_t^y|^2\right] \leq c|x-y|^2 e^{ct^2},$$

where $c \in (0, \infty)$ depends only on the coefficients b, σ, and deduce that $(P_t)_{t \geq 0}$ has the Feller property.

We now focus on a special case which is particularly important for MCMC applications. We fix a probability measure $\pi \in \mathcal{P}(\mathbb{R}^d)$ of the form

$$\pi(\mathrm{d}x) := e^{-V(x)}\, \mathrm{d}x, \tag{3.30}$$

for some function $V \in \mathcal{C}^2(\mathbb{R}^d)$, which is assumed to satisfy

$$\alpha \mathrm{Id} \preceq \mathrm{Hess} V \preceq \beta \mathrm{Id}, \tag{3.31}$$

for some constants $0 < \alpha \leq \beta$. A popular and widely-used strategy for producing approximate samples from π consists in running the Langevin dynamics with potential V. More precisely, one starts from an arbitrary initial condition $X_0 \in L^2(\Omega, \mathcal{F}_0, \mathbb{P})$ and simulates the stochastic process $X = (X_t)_{t \geq 0}$ whose evolution is given by the SDE:

$$\mathrm{d}X_t = \sqrt{2}\, \mathrm{d}B_t - \nabla V(X_t)\, \mathrm{d}t. \tag{3.32}$$

Note that ∇V is $\beta-$Lipschitz, by (3.31), so that (3.32) is a special case of our general SDE (3.26). The following result shows that X_t converges in distribution to π as $t \to \infty$, and does so at an exponential rate, when measured in the $2-$Wassertein distance:

$$W_2(\mu, \nu) := \inf_{X \sim \mu, Y \sim \nu} \sqrt{\mathbb{E}[|X-Y|^2]}. \tag{3.33}$$

Theorem 3.10 (Exponential mixing). *The Markov process X defined by (3.32) has a unique stationary law π with finite second moment, given by (3.30). Moreover, the marginal law $\mu_t := \mathrm{law}(X_t)$ satisfies*

$$\forall t \geq 0, \qquad W_2(\mu_t, \pi) \leq e^{-\alpha t} W_2(\mu_0, \pi).$$

Proof. Fix two measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ with finite second moment, and let (X_0, Y_0) realize the infimum in the definition of $W_2(\mu, \nu)$. Now, consider a Brownian motion $(B_t)_{t\geq 0}$, independent of (X_0, Y_0), and let $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ denote the solutions to the SDE (3.32) with initial conditions X_0 and Y_0. Then, their difference $Z := Y - X$ satisfies

$$\mathrm{d}|Z_t|^2 \;=\; 2Z_t \cdot \mathrm{d}Z_t \;=\; -2(Y_t - X_t) \cdot (\nabla V(Y_t) - \nabla V(X_t))\, \mathrm{d}t.$$

On the other hand, using the assumption (3.31), we have for all $x, y \in \mathbb{R}^d$,

$$\begin{aligned}(y-x) \cdot (\nabla V(y) - \nabla V(x)) &= \int_0^1 (y-x) \cdot (\mathrm{Hess} V)\,(x + t(y-x))\,(y-x)\,\mathrm{d}t \\ &\geq \alpha |y-x|^2.\end{aligned}$$

Combining those two facts, we obtain the differential inequality $\mathrm{d}|Z_t|^2 \leq -2\alpha |Z_t|^2\,\mathrm{d}t$, yielding

$$\forall t \geq 0, \qquad |Z_t|^2 \;\leq\; |Z_0|^2 e^{-2\alpha t}$$

Taking expectations, we conclude that the marginal laws $\mu_t = \mathrm{law}(X_t), \nu_t = \mathrm{law}(Y_t)$ satisfy

$$W_2(\mu_t, \nu_t) \;\leq\; e^{-\alpha t} W_2(\mu, \nu).$$

In particular, if μ and ν are invariant, then $W_2(\mu, \nu) \leq e^{-\alpha t} W_2(\mu, \nu)$, forcing $\mu = \nu$. This shows the uniqueness part of the result, and the exponential convergence to the stationary law, if there is one with finite second moment. Thus, it only remains to show that π has finite second moment and is stationary. The first assertion is clear because $\pi(x) = e^{-V(x)}\,\mathrm{d}x$ and by (3.31),

$$\forall x \in \mathbb{R}^2, \qquad V(x) \;\geq\; V(0) + \nabla V(0) \cdot x + \frac{\alpha |x|^2}{2}.$$

As for the second, we note that by Theorem 3.9, we have $Qf = \Delta f - \nabla V \cdot \nabla f$, for all $f \in \mathcal{C}^2_c(\mathbb{R}^d)$. Thus, writing $\langle \cdot, \cdot \rangle$ for the scalar product in $L^2(\mathbb{R}^d, \mathrm{d}x)$, we have for any $g \in \mathcal{C}^2(\mathbb{R}^d)$,

$$\begin{aligned}\langle Qf, g \rangle &= \int_{\mathbb{R}^d} (\Delta f - \nabla V \cdot \nabla f) g\,\mathrm{d}x \\ &= \int_{\mathbb{R}^d} (\Delta g + \nabla g \cdot \nabla V + g \Delta V) f\,\mathrm{d}x \\ &= \langle f, \Delta g + \nabla g \cdot \nabla V + g \Delta V \rangle,\end{aligned}$$

thanks to the integration-by-parts formula. In particular, choosing $g := e^{-V}$ gives

$$\pi Q f \;=\; \langle Qf, g \rangle \;=\; \langle f, \Delta g + \nabla g \cdot \nabla V + g \Delta V \rangle \;=\; 0.$$

Now, this was established for $f \in \mathcal{C}^2_c(\mathbb{R}^d)$, but it extends to functions in the Schwartz space $\mathcal{S}(\mathbb{R}^d)$. Indeed, the graph of Q is closed and for each $f \in \mathcal{S}(\mathbb{R}^d)$, there is $(f_n)_{n\geq 1}$ in $\mathcal{C}^2_c(\mathbb{R}^d)$ such that f_n, ∇f_n and $\mathrm{Hess}(f_n)$ converge uniformly to f, ∇f and $\mathrm{Hess} f$ respectively, implying that Qf_n converges to $\Delta f - \nabla V \cdot \nabla f$ uniformly. Finally, for each $t \geq 0$, it can be shown that $\mathcal{S}(\mathbb{R}^d)$ is preserved under P_t, so that $\pi Q P_t f = 0$. By (3.24), this implies that $\pi P_t f = \pi f$, and hence $\pi P_t = \pi$, because $\mathcal{S}(\mathbb{R}^d)$ is measure-determining. □