University of Zurich
Institute of Mathematics

Continuous-Time Markov Processes on Discrete State Spaces

Master's Thesis in Mathematics

Author: Leonie Schäppi, 20-701-025
Supervisor: Prof. Dr. Jean Bertoin
Co-Supervisor: Dr. Alejandro Rosales Ortiz
Submission: January 2026

Abstract

Continuous-time Markov processes are a central object in probability theory, with applications ranging from statistical physics to population dynamics and queueing theory. This thesis studies continuous-time Markov processes on discrete state spaces within the framework of Markovian and sub-Markovian semigroups. The aim is to provide a self-contained and rigorous introduction to the theory, with detailed proofs, explanations intended to build intuition, and illustrative examples.

Beginning with Markovian semigroups of transition matrices, we construct continuous-time Markov processes on discrete state spaces and establish the simple Markov property. To exclude pathological behavior, we restrict attention to right-continuous processes, for which we establish the strong Markov property.

We analyze the structure of such right-continuous processes by decomposing their dynamics into a discrete-time Markov chain of visited states and a sequence of exponential holding times between jumps. This naturally leads to the definition of the infinitesimal generator, which encodes the jump rates and transition probabilities of the process and determines its law up to explosion. We derive the backward and forward Chapman-Kolmogorov equations, which link the infinitesimal generator to the transition semigroup via an infinite system of linear differential equations.

For any infinitesimal generator on a discrete state space, we construct the associated minimal right-continuous Markov process by adjoining a cemetery state and absorbing the trajectory at the explosion time. The associated transition matrix is a sub-Markovian semigroup and is characterized as the minimal nonnegative solution of the Chapman-Kolmogorov equations.

Finally, we study qualitative properties of continuous-time Markov processes on discrete state spaces, such as irreducibility, recurrence and transience, criteria for non-explosion, invariant measures and positive recurrence. We conclude with ergodic theorems describing the long-term behavior of the process.

Acknowledgements

I would like to thank my supervisor, Prof. Dr. Jean Bertoin, for giving me the opportunity to write this thesis and for overseeing it. My first introduction to the field of probability theory and stochastic processes began with his engaging lecture *Stochastik* in my fourth semester, which motivated me to take further courses in this area.

I am especially grateful to my co-supervisor, Dr. Alejandro Rosales Ortiz. His instructive lecture *Probability II* showed me how precise and theoretical, yet at the same time visual and intuitive, probability theory can be. I thank him for his continuous guidance and valuable feedback throughout the writing of this thesis, and for giving me the opportunity to work on a topic that truly suits my interests.

I also thank my partner, my parents and my friends for their support and encouragement during challenging phases of my studies and the writing of this thesis, and for providing balance outside of academic life and the world of mathematics.

Contents

1 Introduction **5**

2 Preliminaries **6**

3 Continuous-Time Markov Processes on Discrete State Spaces **7**
3.1 Fundamental Properties 7
3.1.1 State Space, Path Space and Filtrations 7
3.1.2 Semigroups 10
3.1.3 Definition of a Markov Process and the Simple Markov Property 17
3.1.4 Right-Continuous Processes and the Explosion Time 21
3.1.5 Strong Markov Property 25
3.2 Structure of Markov Processes and Infinitesimal Generators 33
3.2.1 Process Decomposition and the Infinitesimal Generator 33
3.2.2 Backward and Forward Chapman-Kolmogorov Equations 38
3.2.3 Examples of Markov Processes 47

4 The Minimal Markov Process Associated with a Generator **56**
4.1 Construction of the Minimal Markov Process 56
4.2 Simulation of Minimal Markov Processes 63

5 Qualitative Study of Markov Processes **69**
5.1 Recurrence and Transience 69
5.2 Criteria for Non-Explosion 74
5.3 Invariant Measures and Positive Recurrence 80
5.4 Ergodic Theorems 91

6 Conclusion and Outlook **98**

Bibliography **99**

1 Introduction

Continuous-time Markov processes on discrete state spaces form a fundamental class of stochastic models, with applications in many areas of probability theory and related fields, including statistical physics, population dynamics and queueing theory. They describe systems that evolve randomly through sudden transitions between countably many states, with random (typically exponentially distributed) holding times between successive transitions. From a mathematical perspective, they offer a setting in which probabilistic reasoning can be combined with analytic techniques and semigroup methods.

The aim of this thesis is to provide a self-contained and rigorous introduction to the theory of continuous-time Markov processes on discrete state spaces. Particular emphasis is placed on complete proofs, clear explanations and intuitive interpretations of the results. The content is mainly based on Chapter II of the lecture notes *Processus markoviens 1* by Thomas Duquesne [Duq23], originally written in French, and unless explicitly stated otherwise, all material is drawn from that source. This thesis presents an English exposition of the material and includes additional proofs, technical details, explanations and examples to improve clarity and understanding. The overall structure and notation follow Chapter II of [Duq23] closely, but the focus is restricted to the foundational and structural aspects of the theory developed there, without going into the later sections dedicated to applications and concrete model classes. Since this thesis continues the exposition developed in Chapter I of [Duq23], the reader is expected to be familiar with the definitions and basic properties of discrete-time Markov chains introduced there.

This thesis is organized as follows. After some preliminary material in Chapter 2, Chapter 3 develops the foundations of continuous-time Markov processes on discrete state spaces. We begin by describing the state space, the associated path space and the filtration framework, before introducing Markovian and sub-Markovian semigroups. We then define continuous-time Markov processes on discrete state spaces through their transition semigroups and establish the simple Markov property. To exclude irregular behavior, we restrict our attention to right-continuous processes, whose structure we then decompose in detail. We introduce jump times, the explosion time, the skeleton and holding times, which together provide a pathwise description of trajectories as successive plateaus separated by jumps. Within this framework, we also establish the strong Markov property.

The second part of Chapter 3 focuses on the structure of continuous-time Markov processes on discrete state spaces through their infinitesimal generators. We introduce the infinitesimal generator as the object encoding the jump rates and transition probabilities of the process, which together characterize its law up to explosion. We further develop the decomposition of the process into a discrete-time Markov chain of visited states and a sequence of conditionally independent and exponentially distributed holding times between jumps. This is followed by the derivation of the backward and forward Chapman-Kolmogorov equations, which link the infinitesimal generator to the transition semigroup through an infinite system of linear differential equations.

In Chapter 4, we reverse our point of view and study the construction of the minimal right-continuous Markov process associated with a given infinitesimal generator. We show that the resulting process is minimal in the sense that its sub-Markovian transition semigroup is the minimal nonnegative solution of the Chapman-Kolmogorov equations associated with the generator. The construction is then illustrated by simulation examples.

Chapter 5 is devoted to the qualitative study of continuous-time Markov processes on discrete state spaces. We investigate recurrence and transience and show that the classification of states coincides for the continuous-time process and its discrete-time chain of visited states. We establish criteria for non-explosion and then study invariant measures, which capture a balance in the jump dynamics of the process and describe how the process distributes its time among the states. This leads to a characterization of positive recurrence through the existence of an invariant probability measure. Finally, we establish ergodic theorems describing the long-term behavior of the process, including convergence to equilibrium.

Lastly, Chapter 6 concludes the thesis and provides an outlook on possible directions for further study.

2 Preliminaries

Throughout this thesis, the reader is assumed to be familiar with standard graduate-level material from probability theory, measure theory, real analysis and ordinary differential equations. As mentioned in the introduction, the reader is also expected to be familiar with the definitions and basic properties of discrete-time Markov chains introduced in Chapter I of [Duq23]. We use the convention $\mathbb{N} = \{0, 1, 2, \dots\}$ throughout.

The approximation of measurable functions by simple functions is used repeatedly in the proofs throughout this thesis. It is therefore stated explicitly here in the form of two lemmas, one for nonnegative measurable functions and one for bounded measurable functions.

Lemma 2.1 (Approximation by simple functions in the nonnegative case)**.** *Let $(\Omega, \mathcal{F})$ be a measurable space and let $f : \Omega \to \mathbb{R}_+$ be a nonnegative $\mathcal{F}$-measurable function. Then there exists an increasing sequence $(f_n)_{n\geq 1}$ of nonnegative $\mathcal{F}$-measurable simple functions of the form*

$$f_n = \sum_{k=1}^{m_n} a_{k,n} \mathbf{1}_{A_{k,n}}, \quad \text{where } a_{k,n} \geq 0, A_{k,n} \in \mathcal{F}$$

such that

$$f_n(\omega) \to f(\omega) \text{ for all } \omega \in \Omega \quad \text{and} \quad f_n \leq f_{n+1} \leq f \text{ for all } n \geq 1.$$

Proof. See Proposition A.1.18 in Section A.1 of [Duq23]. □

Lemma 2.2 (Approximation by simple functions in the bounded case)**.** *Let $(\Omega, \mathcal{F})$ be a measurable space and let $f : \Omega \to \mathbb{R}$ be a bounded $\mathcal{F}$-measurable function. Then there exists a sequence $(f_n)_{n\geq 1}$ of bounded $\mathcal{F}$-measurable simple functions of the form*

$$f_n = \sum_{k=1}^{m_n} a_{k,n} \mathbf{1}_{A_{k,n}}, \quad \text{where } A_{k,n} \in \mathcal{F}$$

such that

$$f_n(\omega) \to f(\omega) \text{ for all } \omega \in \Omega \quad \text{and} \quad \sup_{\omega\in\Omega} |f_n(\omega)| \leq \|f\|_\infty \text{ for all } n \geq 1.$$

Proof. Let $M := \|f\|_\infty < \infty$ and define $g := f + M$. Then g is $\mathcal{F}$-measurable and satisfies $0 \leq g \leq 2M$. Apply Lemma 2.1 to the nonnegative measurable function g and obtain an increasing sequence of nonnegative $\mathcal{F}$-measurable simple functions $(g_n)_{n\geq 1}$ such that $g_n(\omega) \to g(\omega)$ for all $\omega \in \Omega$ and $g_n \leq g_{n+1} \leq g$ for all $n \geq 1$. Now define $f_n := g_n - M$ for all $n \geq 1$. Then each f_n is an $\mathcal{F}$-measurable simple function and

$$f_n(\omega) = g_n(\omega) - M \longrightarrow g(\omega) - M = f(\omega) \quad \text{for all } \omega \in \Omega.$$

Finally, since $0 \leq g_n \leq g \leq 2M$, we have $-M \leq f_n \leq M$ and hence

$$\sup_{\omega\in\Omega} |f_n(\omega)| \leq M = \|f\|_\infty \quad \text{for all } n \geq 1.$$

□

3 Continuous-Time Markov Processes on Discrete State Spaces

3.1 Fundamental Properties

3.1.1 State Space, Path Space and Filtrations

Throughout this thesis, let E denote a countable set with at least two elements. We equip E with the discrete topology, which corresponds to the discrete metric $d : E \times E \to \mathbb{R}_+$ defined by

$$d(i,j) := \begin{cases} 0, & \text{if } i = j, \\ 1, & \text{if } i \neq j. \end{cases}$$

Indeed, for any $i \in E$, $B_d(i, \frac{1}{2}) = \{i\}$, so every singleton of E is open with respect to d. Since every subset of E is a union of singletons, all subsets of E are open with respect to d. Hence, the topology induced by d is exactly the discrete topology.

Note that since E is equipped with the discrete topology, whose associated σ-algebra is $\mathcal{P}(E)$, any function defined on E and taking values in a measurable space is measurable.

A sequence of states $(i_n)_{n\geq 0}$ in E converges to i with respect to d if and only if for all $\varepsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ we have $i_n \in B_d(i, \varepsilon)$. Now observe that:

- If $\varepsilon < 1$, then $B_d(i,\varepsilon) = \{i\}$. Hence, for convergence, there must exist n_0 such that $i_n = i$ for all $n \geq n_0$, i.e. the sequence must eventually be constant at i.
- If $\varepsilon \geq 1$, then $B_d(i,\varepsilon) = E$, so the convergence condition is automatically satisfied without further restriction.

Therefore, convergence in (E, d) is equivalent to eventually being constant at the limit.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A random variable $Z : \Omega \to E$ is said to be $\mathcal{F}$-measurable if and only if, for every state $i \in E$, the set $\{Z = i\} = Z^{-1}(\{i\})$ belongs to $\mathcal{F}$. In this thesis we will work with collections of random variables indexed by $\mathbb{R}_+$ and taking values in E, which we interpret as a random quantity evolving over time. Given a collection $(X_t)_{t\in\mathbb{R}_+}$ of $\mathcal{F}$-measurable random variables $X_t : \Omega \to E$, it will be fruitful, for technical reasons, to view this collection as a random function, i.e. a map which associates to each $\omega \in \Omega$ the function $t \in \mathbb{R}_+ \mapsto X_t(\omega) \in E$. We thus define $\mathbf{X}(\omega) := (X_t(\omega))_{t\in\mathbb{R}_+}$.

To formalize this, we first need to introduce some preliminary background. We will spend some time building the technical tools necessary to handle stochastic processes indexed by $\mathbb{R}_+$.

The set of all functions from $\mathbb{R}_+$ to E is denoted by

$$E^{\mathbb{R}_+} := \{\mathbf{x} = (x(t))_{t\in\mathbb{R}_+} : x(t) \in E,\ t \in \mathbb{R}_+\}.$$

In order to see $\mathbf{X} : \Omega \to E^{\mathbb{R}_+}$ as a random variable, it is necessary to define a σ-algebra on $E^{\mathbb{R}_+}$, which is the object of the following definition.

Definition 3.1 (Cylinders and product σ-algebra)**.** Let E be a countable set equipped with the discrete topology, and hence endowed with the σ-algebra $\mathcal{P}(E)$.

(a) (Set of elementary cylinders) A subset $C \subseteq E^{\mathbb{R}_+}$ of functions from $\mathbb{R}_+$ to E is called an *elementary cylinder* if there exist $n \in \mathbb{N}$, $0 \leq t_0 < t_1 < \cdots < t_n$ and $i_0, \ldots, i_n \in E$ such that

$$C = \{\mathbf{x} = (x(t))_{t\in\mathbb{R}_+} \in E^{\mathbb{R}_+} : x(t_0) = i_0, \ldots, x(t_n) = i_n\}.$$

In words, an elementary cylinder is simply the set of all paths that agree on a finite set of time points $t_0, \ldots, t_n$ with prescribed states $i_0, \ldots, i_n$, while being unconstrained at all other times.

We denote by $\mathcal{C}$ the collection of all elementary cylinders, together with $\emptyset$ and $E^{\mathbb{R}_+}$. Since the intersection of two elementary cylinders is again an elementary cylinder or the empty set (which is also in $\mathcal{C}$), the collection $\mathcal{C}$ forms a π-system.

(b) (Cylinder σ-algebra) We denote by $\mathcal{P}(E)^{\mathbb{R}_+}$ the σ-algebra generated by $\mathcal{C}$. This choice of σ-algebra on $E^{\mathbb{R}_+}$ ensures that all events depending on finitely many time points are measurable.

Now that we have endowed $E^{\mathbb{R}_+}$ with the cylinder σ-algebra $\mathcal{P}(E)^{\mathbb{R}_+}$, we can see a stochastic process as a random variable taking values in the path space $E^{\mathbb{R}_+}$. This is made precise in the next lemma.

Lemma 3.2. *Let $(X_t)_{t\in\mathbb{R}_+}$ be a collection of $\mathcal{F}$-measurable random variables $X_t : \Omega \to E$. Then the map*

$$\begin{aligned}\mathbf{X} : \Omega &\to E^{\mathbb{R}_+}\\ \omega &\mapsto (X_t(\omega))_{t\in\mathbb{R}_+}\end{aligned}$$

is $(\mathcal{F}, \mathcal{P}(E)^{\mathbb{R}_+})$-measurable. Such a random variable $\mathbf{X}$ taking values in $E^{\mathbb{R}_+}$ is called a process with values in E. *The space $E^{\mathbb{R}_+}$ can thus be interpreted as the path space, i.e. the collection of all possible trajectories of a process with state space E.*

Proof. Since $\mathcal{P}(E)^{\mathbb{R}_+} = \sigma(\mathcal{C})$, it suffices to check that $\mathbf{X}^{-1}(C) \in \mathcal{F}$ for all $C \in \mathcal{C}$. Let

$$C := \{\mathbf{x} = (x(t))_{t\in\mathbb{R}_+} \in E^{\mathbb{R}_+} : x(t_0) = i_0, \dots, x(t_n) = i_n\}.$$

Then

$$\begin{aligned}\mathbf{X}^{-1}(C) &= \{\omega \in \Omega : \mathbf{X}(\omega) \in C\}\\ &= \{\omega \in \Omega : (X_t(\omega))_{t\in\mathbb{R}_+} \in C\}\\ &= \{\omega \in \Omega : X_{t_0}(\omega) = i_0, \dots, X_{t_n}(\omega) = i_n\}\\ &= \bigcap_{p=0}^{n} \{X_{t_p} = i_p\}.\end{aligned}$$

Since each X_{t_p} is $\mathcal{F}$-measurable, the preimages $X_{t_p}^{-1}(\{i_p\}) = \{X_{t_p} = i_p\}$ belong to $\mathcal{F}$. By closure of $\mathcal{F}$ under finite intersections, their intersection also belongs to $\mathcal{F}$, which completes the proof. □

We now introduce the shift operator on trajectories, which shifts a path forward in time.

Definition 3.3 (Shift operator)**.** Fix $t_0 \in \mathbb{R}_+$. We define the map $\theta_{t_0} : E^{\mathbb{R}_+} \to E^{\mathbb{R}_+}$ by

$$\theta_{t_0}\mathbf{x} := (x(t_0 + t))_{t\in\mathbb{R}_+}, \quad \text{for } \mathbf{x} = (x(t))_{t\in\mathbb{R}_+} \in E^{\mathbb{R}_+}.$$

The map θ_{t_0} is called the shift operator at time t_0 on trajectories.

It is clear that $\theta_{t_0} \circ \theta_{t_1} = \theta_{t_0+t_1}$ for all $t_0, t_1 \in \mathbb{R}_+$. This means that shifting a path first by t_1 and then by t_0 is the same as shifting it once by $t_0 + t_1$.

To work with this operator on $E^{\mathbb{R}_+}$ in a probabilistic setting, we need to ensure that it is measurable with respect to the cylinder σ-algebra $\mathcal{P}(E)^{\mathbb{R}_+}$. The following lemma establishes this.

Lemma 3.4. *Fix $t \in \mathbb{R}_+$. The map $\theta_t : E^{\mathbb{R}_+} \to E^{\mathbb{R}_+}$ is $(\mathcal{P}(E)^{\mathbb{R}_+}, \mathcal{P}(E)^{\mathbb{R}_+})$-measurable.*

Proof. Since $\mathcal{P}(E)^{\mathbb{R}_+} = \sigma(\mathcal{C})$, it suffices to verify that $\theta_t^{-1}(C) \in \mathcal{P}(E)^{\mathbb{R}_+}$ for every $C \in \mathcal{C}$. Take

$$C := \{\mathbf{x} = (x(s))_{s\in\mathbb{R}_+} \in E^{\mathbb{R}_+} : x(t_0) = i_0, \dots, x(t_n) = i_n\}.$$

Then

$$\theta_t^{-1}(C) = \{\mathbf{x} = (x(s))_{s\in\mathbb{R}_+} \in E^{\mathbb{R}_+} : x(t_0 + t) = i_0, \dots, x(t_n + t) = i_n\}$$

which is an elementary cylinder in $\mathcal{C}$. Since $\mathcal{C} \subseteq \sigma(\mathcal{C}) = \mathcal{P}(E)^{\mathbb{R}_+}$, this proves the lemma. □

We now introduce the notion of filtration in continuous time, which tells us what can be observed about the process up to a given time.

Definition 3.5 (Continuous-time filtration)**.** Recall that $(\Omega, \mathcal{F})$ is a measurable space.

(a) A *filtration* on $(\Omega, \mathcal{F})$ is a collection $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ of sub-σ-algebras of $\mathcal{F}$ that is increasing with respect to inclusion, i.e. $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ for all $t \geq s$.

(b) The *right-continuous filtration*, denoted $(\mathcal{F}_{t+})_{t\in\mathbb{R}_+}$, is defined by

$$\mathcal{F}_{t+} := \bigcap_{h>0} \mathcal{F}_{t+h}, \text{ for } t \in \mathbb{R}_+.$$

Intuitively, $\mathcal{F}_{t+}$ contains all events whose occurrence can be determined by observing the process immediately after time t.

(c) Let $\mathbf{X} = (X_t)_{t\in\mathbb{R}_+}$ be a process defined on $(\Omega, \mathcal{F})$, with values in E, which is $(\mathcal{F}, \mathcal{P}(E)^{\mathbb{R}_+})$-measurable. Given a filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$, we say that $\mathbf{X}$ is *adapted* to $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ if and only if X_t is $\mathcal{F}_t$-measurable for every $t \in \mathbb{R}_+$.

(d) Let $\mathbf{X}$ be as in (c). For each $t \in \mathbb{R}_+$, we denote by $\mathcal{F}_t^{\mathbf{X}}$ the σ-algebra generated by the random variables X_s, $s \in [0, t]$. In other words,

$$\mathcal{F}_t^{\mathbf{X}} := \sigma(\{X_s = i\} : i \in E,\ s \in [0, t]).$$

We call $(\mathcal{F}_t^{\mathbf{X}})_{t\in\mathbb{R}_+}$ the *natural filtration* associated with $\mathbf{X}$.

Before proceeding, we take a closer look at the right-continuous filtration in (b). In particular, we show that this construction indeed defines a filtration, that each σ-algebra $\mathcal{F}_t$ is contained in its right-continuous counterpart $\mathcal{F}_{t+}$, and that applying the right-continuous operation twice does not change the result.

Lemma 3.6. *The right-continuous family $(\mathcal{F}_{t+})_{t\in\mathbb{R}_+}$ is a filtration on $(\Omega, \mathcal{F})$. Moreover, for $t \in \mathbb{R}_+$:*

$$\mathcal{F}_t \subseteq \mathcal{F}_{t+} \quad \textit{and} \quad \mathcal{F}_{t+} = \bigcap_{h>0} \mathcal{F}_{(t+h)+}.$$

Proof. For each $h > 0$, $\mathcal{F}_{t+h}$ is a sub-σ-algebra of $\mathcal{F}$. Since arbitrary intersections of σ-algebras are σ-algebras, $\mathcal{F}_{t+} = \bigcap_{h>0} \mathcal{F}_{t+h}$ is a sub-σ-algebra of $\mathcal{F}$.

If $s \le t$, then for every $h > 0$ we have $s + h \le t + h$ and hence $\mathcal{F}_{s+h} \subseteq \mathcal{F}_{t+h}$. Therefore

$$\mathcal{F}_{s+} = \bigcap_{h>0} \mathcal{F}_{s+h} \subseteq \bigcap_{h>0} \mathcal{F}_{t+h} = \mathcal{F}_{t+}$$

and $(\mathcal{F}_{t+})_{t\in\mathbb{R}_+}$ is a filtration.

Moreover, for any fixed $t \in \mathbb{R}_+$ and all $h > 0$, we have $\mathcal{F}_t \subseteq \mathcal{F}_{t+h}$ by monotonicity of the filtration. Taking intersections over $h > 0$ yields

$$\mathcal{F}_t \subseteq \bigcap_{h>0} \mathcal{F}_{t+h} = \mathcal{F}_{t+}.$$

Finally, by definition,

$$\mathcal{F}_{(t+h)+} = \bigcap_{k>0} \mathcal{F}_{t+h+k}.$$

Setting $u := h + k$, this becomes

$$\mathcal{F}_{(t+h)+} = \bigcap_{u>h} \mathcal{F}_{t+u}.$$

Therefore

$$\bigcap_{h>0} \mathcal{F}_{(t+h)+} = \bigcap_{h>0} \bigcap_{u>h} \mathcal{F}_{t+u} = \bigcap_{u>0} \mathcal{F}_{t+u} = \mathcal{F}_{t+}.$$

□

It is also natural to ask whether the construction in (d) really gives a filtration and how it relates to other filtrations. This is clarified in the next lemma.

Lemma 3.7. *The natural filtration $(\mathcal{F}_t^{\mathbf{X}})_{t\in\mathbb{R}_+}$ in (d) is indeed a filtration on $(\Omega, \mathcal{F})$ and it is the smallest filtration (with respect to inclusion) to which $\mathbf{X}$ is adapted.*

Proof. Fix $0 \leq s \leq t$. By definition,

$$\mathcal{F}_s^{\mathbf{X}} = \sigma(\{X_u = i\} : i \in E,\ u \in [0,s]) \subseteq \sigma(\{X_u = i\} : i \in E,\ u \in [0,t]) = \mathcal{F}_t^{\mathbf{X}}$$

since $[0,s] \subseteq [0,t]$. Hence $(\mathcal{F}_t^{\mathbf{X}})_{t\in\mathbb{R}_+}$ is a filtration. Moreover, for each $t \in \mathbb{R}_+$, the random variable X_t is $\mathcal{F}_t^{\mathbf{X}}$-measurable by construction, since its preimages $\{X_t = i\}$, $i \in E$, are among the generators of $\mathcal{F}_t^{\mathbf{X}}$. Thus $\mathbf{X}$ is adapted to $(\mathcal{F}_t^{\mathbf{X}})_{t\in\mathbb{R}_+}$.

Now let $(\mathcal{G}_t)_{t\in\mathbb{R}_+}$ be any filtration to which $\mathbf{X}$ is adapted. Then for each $s \leq t$, X_s is $\mathcal{G}_s$-measurable, and since $\mathcal{G}_s \subseteq \mathcal{G}_t$, also $\mathcal{G}_t$-measurable. Hence $\{X_s = i\} \in \mathcal{G}_t$ for every $i \in E$. Therefore

$$\mathcal{F}_t^{\mathbf{X}} = \sigma(\{X_s = i\} : i \in E,\ s \leq t) \subseteq \mathcal{G}_t \quad \text{for all } t \in \mathbb{R}_+,$$

so $(\mathcal{F}_t^{\mathbf{X}})_{t\in\mathbb{R}_+}$ is the smallest filtration (by inclusion) making $\mathbf{X}$ adapted. □

3.1.2 Semigroups

In the case of continuous-time processes, the object that plays the role of transition matrices is the semigroup, defined as follows.

Definition 3.8 (Markovian and sub-Markovian semigroups)**.** Recall that E is a countable space equipped with the discrete topology, and hence endowed with the σ-algebra $\mathcal{P}(E)$. For each $t \in \mathbb{R}_+$, we are given a matrix $P_t = (p_t(i,j))_{i,j\in E}$ indexed by E. The family $(P_t)_{t\in\mathbb{R}_+}$ of matrices forms a *sub-Markovian semigroup* if it satisfies:

(a) For all $t \in \mathbb{R}_+$ and $i, j \in E$, $p_t(i,j) \geq 0$.

(b) For all $s, t \in \mathbb{R}_+$ and $i, j \in E$,

$$p_{s+t}(i,j) = \sum_{k\in E} p_s(i,k)p_t(k,j), \quad \text{i.e. } P_sP_t = P_{s+t}$$

which is the semigroup property. We also require that $P_0 = \mathrm{Id}_E$.

(c) For all $t \in \mathbb{R}_+$ and $i \in E$, $\sum_{j\in E} p_t(i,j) \leq 1$. So each row of P_t has sum less than or equal to 1.

The family $(P_t)_{t\in\mathbb{R}_+}$ forms a *Markovian semigroup* if it satisfies (a), (b) and the following:

(c′) For all $t \in \mathbb{R}_+$ and $i \in E$, $\sum_{j\in E} p_t(i,j) = 1$. So each row of P_t sums to 1.

Remark 3.9 (Intuition). The conditions in the preceding definition can be made more intuitive as follows.

(b) The semigroup property expresses that the probability of moving from i to j in time $s+t$ can be obtained by first moving from i to some intermediate state k in time s (with probability $p_s(i,k)$), and then from k to j in time t (with probability $p_t(k,j)$).

(c) The condition $\sum_{j\in E} p_t(i,j) \leq 1$ means that some probability mass may be "lost". We will see that in that case the process might have left the state space, which is a phenomenon called "killing".

(c′) The stronger condition $\sum_{j\in E} p_t(i,j) = 1$ ensures that the process must always be in some state of E.

In the discrete-time case, we had a single transition matrix $Q = (\pi_{i,j})_{i,j\in E}$ describing the probabilities of jumping from one state to another in each discrete time step. In contrast, for continuous-time processes we have an entire family of matrices $(P_t)_{t\in\mathbb{R}_+}$, one for each time t, which gives the transition probabilities over any length of continuous time.

We now give two examples of Markovian semigroups to illustrate the definition.

Example 3.10. Take $E := \{1, 2, 3, \dots\}$ and fix $c > 0$. For all $t \in \mathbb{R}_+$, define

$$p_t(i,j) := \mathbf{1}_{\{i \le j\}} \tfrac{(ct)^{j-i} e^{-ct}}{(j-i)!}, \quad i, j \in \{1, 2, 3, \dots\}.$$

Then $(P_t)_{t \in \mathbb{R}_+}$ is a Markovian semigroup.

Proof. (a) All terms in the definition of $p_t(i,j)$ (indicator function, c, t, exponential function, factorial) are nonnegative.

(b) For $s, t \ge 0$ and $i \le j$,

$$\begin{aligned}(P_s P_t)(i,j) &= \sum_{k \in \{1,2,3,\dots\}} p_s(i,k) p_t(k,j) \\ &= \sum_{k=i}^{j} \tfrac{(cs)^{k-i} e^{-cs}}{(k-i)!} \tfrac{(ct)^{j-k} e^{-ct}}{(j-k)!} \\ &= e^{-c(s+t)} \sum_{m=0}^{n} \tfrac{(cs)^m}{m!} \tfrac{(ct)^{n-m}}{(n-m)!}\end{aligned}$$

where $n := j - i$ and $m := k - i$. Using the binomial theorem,

$$\begin{aligned}e^{-c(s+t)} \sum_{m=0}^{n} \tfrac{(cs)^m}{m!} \tfrac{(ct)^{n-m}}{(n-m)!} &= e^{-c(s+t)} \tfrac{c^n}{n!} \sum_{m=0}^{n} \binom{n}{m} s^m t^{n-m} \\ &= e^{-c(s+t)} \tfrac{(c(s+t))^n}{n!} \\ &= p_{s+t}(i,j).\end{aligned}$$

If $i > j$, both sides are 0.

For $t = 0$, we have

$$p_0(i,j) = \mathbf{1}_{\{i \le j\}} \tfrac{0^{j-i}}{(j-i)!}.$$

This equals 1 if $i = j$ and 0 otherwise, so $P_0 = \mathrm{Id}_{\{1,2,3,\dots\}}$.

(c′) For fixed i (using again $n := j - i$),

$$\begin{aligned}\sum_{j \in \{1,2,3,\dots\}} p_t(i,j) &= \sum_{j=i}^{\infty} \tfrac{(ct)^{j-i} e^{-ct}}{(j-i)!} \\ &= e^{-ct} \sum_{n=0}^{\infty} \tfrac{(ct)^n}{n!} \\ &= e^{-ct} e^{ct} \\ &= 1.\end{aligned}$$

□

Example 3.11. Take $E := \{0, 1\}$. For $t > 0$, set $p_t(i,j) := \frac{1}{2}$ for all $i, j \in \{0, 1\}$. Moreover, set

$$p_0(0,0) := p_0(1,1) := 1, \quad p_0(0,1) := p_0(1,0) := 0.$$

Then $(P_t)_{t \in \mathbb{R}_+}$ is a Markovian semigroup.

Proof. (a) In both cases, $t > 0$ and $t = 0$, the entries are nonnegative.

(b) For $s, t > 0$, we have

$$P_t = P_s = P_{s+t} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Thus

$$P_s P_t = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = P_{s+t}.$$

Also $P_0 = \mathrm{Id}_{\{0,1\}}$ and for all $t \in \mathbb{R}_+$, we have $P_0 P_t = P_t = P_{0+t}$ and $P_t P_0 = P_t = P_{t+0}$.

(c′) Each row of P_t sums to 1, both when $t = 0$ and when $t > 0$.

□

So far we have seen that in a sub-Markovian semigroup probability mass can be "lost". A way to handle this is to extend the state space with an extra absorbing state, so that the resulting semigroup is Markovian.

Remark 3.12 (Making a sub-Markovian semigroup Markovian). Consider a sub-Markovian semigroup $(P_t)_{t\in\mathbb{R}_+}$ on E. By extending the state space, it is possible to obtain a Markovian semigroup from $(P_t)_{t\in\mathbb{R}_+}$ via the following construction.

Introduce a cemetery state $\partial \notin E$ and set $E^* := E \cup \{\partial\}$. Intuitively, ∂ represents an absorbing state into which the process jumps whenever probability mass is "lost" in the original sub-Markovian process. For $t \in \mathbb{R}_+$, define the matrix $P_t^* = (p_t^*(i,j))_{i,j\in E^*}$ by

$$p_t^*(i,j) := p_t(i,j) \text{ for } i,j \in E, \quad p_t^*(\partial,\partial) := 1, \quad p_t^*(\partial,j) := 0 \text{ for } j \in E,$$

and for $i \in E$,

$$p_t^*(i,\partial) := 1 - \sum_{j\in E} p_t(i,j).$$

Then $(P_t^*)_{t\in\mathbb{R}_+}$ is Markovian on E^*.

Proof. (a) For $i,j \in E$ we have $p_t^*(i,j) = p_t(i,j) \geq 0$ since $(P_t)_{t\in\mathbb{R}_+}$ is sub-Markovian. For $i \in E$,

$$p_t^*(i,\partial) = 1 - \sum_{j\in E} p_t(i,j) \geq 0, \quad \text{since } \sum_{j\in E} p_t(i,j) \leq 1.$$

All other cases are 0 or 1. Hence nonnegativity holds.

(b) Let $s,t \in \mathbb{R}_+$. For $i,j \in E$,

$$(P_s^* P_t^*)(i,j) = \sum_{k\in E^*} p_s^*(i,k)p_t^*(k,j) = \sum_{k\in E} p_s(i,k)p_t(k,j)$$

since if $k = \partial$, then $p_t^*(\partial,j) = 0$, so the sum reduces to the case $k \in E$. By the semigroup property of $(P_t)_{t\in\mathbb{R}_+}$,

$$\sum_{k\in E} p_s(i,k)p_t(k,j) = p_{s+t}(i,j) = p_{s+t}^*(i,j).$$

For $i \in E$ and $j = \partial$,

$$\begin{aligned}(P_s^* P_t^*)(i,\partial) &= \sum_{k\in E^*} p_s^*(i,k)p_t^*(k,\partial) \\ &= \Big(\sum_{k\in E} p_s(i,k)(1 - \sum_{j\in E} p_t(k,j))\Big) + p_s^*(i,\partial)p_t^*(\partial,\partial) \\ &= \Big(\sum_{k\in E} p_s(i,k)(1 - \sum_{j\in E} p_t(k,j))\Big) + (1 - \sum_{k\in E} p_s(i,k)).\end{aligned}$$

This simplifies to

$$1 - \sum_{j\in E}\Big(\sum_{k\in E} p_s(i,k)p_t(k,j)\Big) = 1 - \sum_{j\in E} p_{s+t}(i,j) = p_{s+t}^*(i,\partial).$$

For $i = \partial$ and $j \in E$,

$$(P_s^* P_t^*)(\partial,j) = \sum_{k\in E^*} p_s^*(\partial,k)p_t^*(k,j) = p_t^*(\partial,j) = 0 = p_{s+t}^*(\partial,j).$$

For $i = j = \partial$,

$$(P_s^* P_t^*)(\partial,\partial) = \sum_{k\in E^*} p_s^*(\partial,k)p_t^*(k,\partial) = 1 = p_{s+t}^*(\partial,\partial).$$

Thus $P_s^* P_t^* = P_{s+t}^*$.

To show $P_0^* = \mathrm{Id}_{E^*}$, note that $P_0 = \mathrm{Id}_E$ and $p_0^*(\partial, \partial) = 1$. For $i \in E$,

$$p_0^*(i, \partial) = 1 - \sum_{j \in E} p_0(i, j) = 1 - 1 = 0$$

and for $j \in E$, $p_0^*(\partial, j) = 0$. Hence P_0^* is the identity on E^*.

(c′) For $i \in E$,

$$\sum_{j \in E^*} p_t^*(i, j) = \Big(\sum_{j \in E} p_t(i, j)\Big) + p_t^*(i, \partial) = \Big(\sum_{j \in E} p_t(i, j)\Big) + (1 - \sum_{j \in E} p_t(i, j)) = 1.$$

For $i = \partial$,

$$\sum_{j \in E^*} p_t^*(\partial, j) = p_t^*(\partial, \partial) = 1.$$

□

We are now going to construct a particularly simple family of semigroups using the exponential of finite matrices. It is essential to note that not all semigroups can be represented as exponentials of finite matrices. The forthcoming section focuses on those that can, emphasizing their structure and properties.

(Sub-)Markovian semigroups on a finite state space. When E is finite, to simplify the notation, we identify E with $\{1, \ldots, N\}$, where N is the number of elements of E. We denote by $\mathcal{M}_N(\mathbb{R})$ the set of $N \times N$ matrices with real entries. $\mathcal{M}_N(\mathbb{R})$ is a real vector space of dimension N^2 and also a finite-dimensional algebra under matrix multiplication. Therefore it can be endowed with an *algebra norm* $\|\cdot\|$, that is, in addition to the usual vector norm properties

- $\|A\| \geq 0$, with equality iff $A = 0$, for all $A \in \mathcal{M}_N(\mathbb{R})$
- $\|\lambda A\| = |\lambda| \|A\|$, for all $\lambda \in \mathbb{R}$ and $A \in \mathcal{M}_N(\mathbb{R})$
- $\|A + B\| \leq \|A\| + \|B\|$, for all $A, B \in \mathcal{M}_N(\mathbb{R})$

it also satisfies the submultiplicativity property

$$\|AB\| \leq \|A\| \|B\|, \text{ for all } A, B \in \mathcal{M}_N(\mathbb{R}).$$

If $A = (a_{i,j})_{1 \leq i,j \leq N} \in \mathcal{M}_N(\mathbb{R})$, we use the notation $[A](i, j) := a_{i,j}$ for the (i, j)-entry of the matrix A. It follows that for fixed (i, j), the map $A \mapsto [A](i, j)$ is continuous, since it is a linear coordinate projection, and in finite dimension every linear map is continuous.

We now derive the definition of the matrix exponential and establish its basic properties. Let $A \in \mathcal{M}_N(\mathbb{R})$. For all $n \in \mathbb{N}$, set

$$f_n(A) := \sum_{k=0}^{n} \tfrac{A^k}{k!}$$

with the convention $A^0 := \mathrm{Id}_N$. Our first goal is to check that this sequence has a well-defined limit, which will serve as the definition of the matrix exponential. By the triangle inequality and submultiplicativity,

$$\begin{aligned}
\sup_{n_1 > n_2 \geq n} \|f_{n_1}(A) - f_{n_2}(A)\| &= \sup_{n_1 > n_2 \geq n} \Big\| \sum_{k=n_2+1}^{n_1} \tfrac{A^k}{k!} \Big\| \\
&\leq \sup_{n_1 > n_2 \geq n} \sum_{k=n_2+1}^{n_1} \tfrac{\|A\|^k}{k!} \\
&\leq \sum_{k=n+1}^{\infty} \tfrac{\|A\|^k}{k!} \xrightarrow[n \to \infty]{} 0.
\end{aligned}$$

The last expression converges to zero since it is the tail of the scalar exponential series $\sum_{k=0}^{\infty} \frac{\|A\|^k}{k!} = e^{\|A\|}$, and the tails of a convergent series vanish. We can deduce that $(f_n(A))_{n\geq 0}$ is a Cauchy sequence in $(\mathcal{M}_N(\mathbb{R}), \|\cdot\|)$.

Since $(\mathcal{M}_N(\mathbb{R}), \|\cdot\|)$ is finite-dimensional, it is complete and the Cauchy sequence $(f_n(A))_{n\geq 0}$ converges to some limit in $\mathcal{M}_N(\mathbb{R})$. We denote this limit by

$$e^A := \|\cdot\|\text{-}\lim_{n\to\infty} f_n(A) = \sum_{k=0}^{\infty} \tfrac{A^k}{k!}.$$

Next, we verify that the series defining e^A converges entrywise. For each $1 \leq i, j \leq N$, continuity of the linear coordinate projection $A \mapsto [A](i,j)$ implies the existence of a constant $C_{i,j} > 0$ such that $|[A^k](i,j)| \leq C_{i,j}\|A^k\| \leq C_{i,j}\|A\|^k$ (since we are in finite dimension). It follows that

$$\sum_{k=0}^{\infty} \tfrac{|[A^k](i,j)|}{k!} \leq C_{i,j} \sum_{k=0}^{\infty} \tfrac{\|A\|^k}{k!} = C_{i,j} e^{\|A\|} < \infty$$

from which we conclude that

$$[e^A](i,j) = \sum_{k=0}^{\infty} \tfrac{[A^k](i,j)}{k!}$$

is absolutely convergent.

We now establish the key property that if two matrices commute, then the exponential of their sum equals the product of their exponentials. Assume $A, B \in \mathcal{M}_N(\mathbb{R})$ commute, i.e. $AB = BA$. Then

$$f_n(A)f_n(B) = \sum_{p=0}^{n}\sum_{r=0}^{n} \tfrac{A^p}{p!}\tfrac{B^r}{r!} = \sum_{q=0}^{2n} \tfrac{1}{q!} \sum_{p=\max(0,q-n)}^{\min(q,n)} \binom{q}{p} A^p B^{q-p}$$

where we used $q := p + r$. On the other hand, by the binomial theorem under $AB = BA$,

$$f_n(A+B) = \sum_{q=0}^{n} \tfrac{(A+B)^q}{q!} = \sum_{q=0}^{n} \tfrac{1}{q!} \sum_{p=0}^{q} \binom{q}{p} A^p B^{q-p}.$$

Subtracting gives

$$f_n(A)f_n(B) - f_n(A+B) = \sum_{q=n+1}^{2n} \tfrac{1}{q!} \sum_{p=q-n}^{n} \binom{q}{p} A^p B^{q-p}.$$

Taking norms and using submultiplicativity,

$$\begin{aligned}\|f_n(A)f_n(B) - f_n(A+B)\| &\leq \sum_{q=n+1}^{\infty} \tfrac{1}{q!} \sum_{p=0}^{q} \binom{q}{p} \|A\|^p \|B\|^{q-p} \\ &= \sum_{q=n+1}^{\infty} \tfrac{(\|A\|+\|B\|)^q}{q!} \xrightarrow[n\to\infty]{} 0.\end{aligned}$$

Again, the last expression converges to zero because it is the tail of the scalar exponential series $\sum_{q=0}^{\infty} \frac{(\|A\|+\|B\|)^q}{q!} = e^{\|A\|+\|B\|}$, and the tails of a convergent series vanish. This then yields

$$e^{A+B} = e^A e^B.$$

Having established the main properties of the matrix exponential, we now use it to define semigroups.

Example 3.13. Let $G \in \mathcal{M}_N(\mathbb{R})$ and define $P_t := e^{tG}$ for $t \in \mathbb{R}_+$. Since tG and sG commute and $(t+s)G = tG + sG$, by the additivity property of the exponential established above, we have

$$P_{t+s} = e^{(t+s)G} = e^{tG}e^{sG} = P_t P_s, \quad P_0 = e^0 = \mathrm{Id}_N.$$

Therefore $(P_t)_{t\in\mathbb{R}_+}$ forms a matrix semigroup.

We now state a fundamental theorem which shows that, under appropriate assumptions, every matrix semigroup arises from the exponential of a single matrix (the infinitesimal generator), and it gives the precise conditions under which the semigroup is Markovian or sub-Markovian.

Theorem 3.14. *For $t \in \mathbb{R}_+$, consider a matrix $P_t = (p_t(i,j))_{1\le i,j\le N} \in \mathcal{M}_N(\mathbb{R})$. Suppose the following assumptions hold:*

(a) $P_t P_s = P_{t+s}$, *for all* $s, t \in \mathbb{R}_+$

(b) $P_0 = \mathrm{Id}_N$

(c) The map $t \in \mathbb{R}_+ \mapsto P_t \in \mathcal{M}_N(\mathbb{R})$ is continuously differentiable from the right at 0. That is, there exists $\delta > 0$ such that $t \mapsto P_t$ is differentiable on $(0,\delta)$, right-differentiable at 0, and the derivative $\frac{\mathrm{d}}{\mathrm{d}t}P_t$ is right-continuous at 0.

Moreover, there exists a matrix $G = (q_{i,j})_{1\le i,j\le N} \in \mathcal{M}_N(\mathbb{R})$ such that for all $1 \le i, j \le N$,

$$\lim_{t\to 0^+} \tfrac{1}{t}(p_t(i,j) - p_0(i,j)) = \lim_{t\to 0^+} \tfrac{1}{t}(p_t(i,j) - \delta_{i,j}) = q_{i,j}.$$

Then the following assertions hold:

(i) For all $t \in \mathbb{R}_+$, $P_t = e^{tG}$. Moreover, the map $t \in \mathbb{R}_+ \mapsto P_t \in \mathcal{M}_N(\mathbb{R})$ is differentiable, and its derivative satisfies the ordinary differential equation

$$\tfrac{\mathrm{d}}{\mathrm{d}t}P_t = GP_t = P_t G, \quad P_0 = \mathrm{Id}_N$$

where $\frac{\mathrm{d}}{\mathrm{d}t}P_t := (p_t'(i,j))_{1\le i,j\le N}$.

(ii) The family $(P_t)_{t\in\mathbb{R}_+}$ is Markovian (resp. sub-Markovian) if and only if the coefficients $q_{i,j}$ of G satisfy

$$\text{for } 1 \le i,j \le N, \quad q_{i,j} \ge 0 \text{ for } i \ne j, \quad \sum_{\substack{j=1\\ j\ne i}}^{N} q_{i,j} = -q_{i,i} \text{ (resp. } \le -q_{i,i}). \tag{3.1}$$

We observe that $-q_{i,i} \ge 0$ for all $1 \le i \le N$.

Proof. We begin by proving (i) under assumptions (a), (b), and (c). By (c), the map $t \mapsto P_t$ is differentiable on $(0,\delta)$ and right-differentiable at 0, which implies that it is continuous on $(0,\delta)$ and right-continuous at 0. Since the domain is $\mathbb{R}_+ = [0,\infty)$, it follows that $t \mapsto P_t$ is continuous on $[0,\delta)$. We may therefore fix a small $\eta > 0$ with $2\eta < \delta$ such that the function $t \in [0,2\eta] \mapsto P_t$ is continuous. Shifting this function by η then yields that $t \in [-\eta,\eta] \mapsto P_{t+\eta}$ is continuous.

Let $t_0 \ge 2\eta$. By the semigroup property, $P_{t_0+t} = P_{t_0-\eta}P_{t+\eta}$. Since $P_{t_0-\eta}$ is a fixed matrix and $t \in [-\eta,\eta] \mapsto P_{t+\eta}$ is continuous, it follows that $t \in [-\eta,\eta] \mapsto P_{t_0+t}$ is continuous. Equivalently, $t \in [t_0-\eta, t_0+\eta] \mapsto P_t$ is continuous, which means that $t \mapsto P_t$ is continuous on a neighborhood of every $t_0 \ge 2\eta$. Together with the previously established continuity on $[0,2\eta]$, we conclude that $t \mapsto P_t$ is continuous on all of $\mathbb{R}_+$.

Since we are in finite dimension and convergence with respect to any norm on $\mathcal{M}_N(\mathbb{R})$ is equivalent to entrywise convergence, (c) implies

$$\lim_{h\to 0^+} \|\tfrac{1}{h}(P_h - \mathrm{Id}_N) - G\| = 0.$$

Next, observe that for all $t \in \mathbb{R}_+$,

$$\|\tfrac{1}{h}(P_{t+h} - P_t) - GP_t\| \le \|\tfrac{1}{h}(P_h - \mathrm{Id}_N) - G\|\|P_t\|$$

using the semigroup property $P_{t+h} = P_h P_t$ and submultiplicativity. Taking the limit $h \to 0^+$, we obtain that $t \mapsto P_t$ is right-differentiable on $\mathbb{R}_+$ with right-derivative GP_t. Similarly, for all $t \in \mathbb{R}_+$,

$$\|\tfrac{1}{h}(P_{t+h} - P_t) - P_t G\| \le \|P_t\|\|\tfrac{1}{h}(P_h - \mathrm{Id}_N) - G\|$$

using the semigroup property $P_{t+h} = P_t P_h$ and submultiplicativity. Again taking the limit $h \to 0^+$, we obtain that $t \mapsto P_t$ is right-differentiable on $\mathbb{R}_+$ with right-derivative $P_t G$. This shows that $GP_t = P_t G$ for all $t \in \mathbb{R}_+$.

Let us next address the left-differentiability of $(P_t)_{t\in\mathbb{R}_+}$. To this end, fix $t > 0$ and $h \in (0, t)$. Since the map $s \in \mathbb{R}_+ \mapsto P_s$ is continuous and every norm on a finite-dimensional vector space is continuous, the composition $s \in \mathbb{R}_+ \mapsto \|P_s\|$ is continuous. Therefore it is bounded on the compact interval $[0, t]$ and $c = \sup_{s\in[0,t]} \|P_s\| < \infty$. Then by triangle inequality, semigroup property and submultiplicativity,

$$\begin{aligned}
\|\tfrac{1}{h}(P_{t-h} - P_t) + GP_t\| &\le \|\tfrac{1}{h}(P_{t-h} - P_t) + GP_{t-h}\| + \|GP_t - GP_{t-h}\| \\
&\le \|P_{t-h}(\tfrac{1}{h}(\mathrm{Id}_N - P_h) + G)\| + \|G\|\|P_{t-h}\|\|P_h - \mathrm{Id}_N\| \\
&\le \|P_{t-h}\|\|(\tfrac{1}{h}(\mathrm{Id}_N - P_h) + G)\| + c\|G\|\|P_h - \mathrm{Id}_N\| \\
&\le c\|\tfrac{1}{h}(\mathrm{Id}_N - P_h) + G\| + c\|G\|\|P_h - \mathrm{Id}_N\|.
\end{aligned}$$

By assumption (c) and right-continuity of $h \mapsto P_h$ at 0, we get that

$$\lim_{h\to 0^+} \|\tfrac{1}{h}(P_{t-h} - P_t) + GP_t\| = 0.$$

Hence $t \mapsto P_t$ is also left-differentiable on $\mathbb{R}_+$ and its left-derivative coincides with its right-derivative. Therefore $t \mapsto P_t$ is differentiable on $\mathbb{R}_+$ and satisfies

$$\tfrac{\mathrm{d}}{\mathrm{d}t} P_t = GP_t, \quad P_0 = \mathrm{Id}_N.$$

This is a first-order linear ODE with a given initial condition. Since $e^{tG} = \mathrm{Id}_N + tG + \frac{t^2}{2!}G^2 + \frac{t^3}{3!}G^3 + \cdots$, we may differentiate term by term and obtain $\frac{\mathrm{d}}{\mathrm{d}t} e^{tG} = G + tG^2 + \frac{t^2}{2!}G^3 + \cdots = Ge^{tG}$. Moreover $e^{0\cdot G} = \mathrm{Id}_N$, so $t \mapsto e^{tG}$ solves this ODE with the given initial condition. By uniqueness of solutions to linear ODEs in finite dimension, we conclude that $P_t = e^{tG}$ for all $t \in \mathbb{R}_+$, which proves (i).

We now prove (ii). Suppose first that $(P_t)_{t\in\mathbb{R}_+}$ is Markovian (resp. sub-Markovian) and we will show that this implies (3.1). Since by (i) we have $P_t = e^{tG}$, we can write $P_t = \mathrm{Id}_N + tG + t^2 B_t$ where

$$B_t := \sum_{k=2}^{\infty} \tfrac{t^{k-2}}{k!} G^k \in \mathcal{M}_N(\mathbb{R}).$$

For $t \in [0, 1]$, we have $\|\frac{t^{k-2}}{k!}G^k\| \le \frac{1}{k!}\|G\|^k$ and since $\sum_{k=2}^{\infty} \frac{\|G^k\|}{k!} \le e^{\|G\|} < \infty$, the series defining B_t converges uniformly on $[0, 1]$. Using this and the fact that each term of the series defining B_t depends continuously on t, we deduce that $t \mapsto B_t$ is continuous on $[0, 1]$ (since the uniform limit of continuous functions is continuous). Moreover, as a continuous function on the compact interval $[0, 1]$, $t \mapsto B_t$ is bounded. Finally, since the coordinate projection $A \mapsto [A](i, j)$ is continuous, the entries $b_t(i, j)$ of B_t are also continuous and bounded functions of $t \in [0, 1]$.

Thus for all $t \in [0, 1]$, $1 \le i, j \le N$,

$$p_t(i, j) = tq_{i,j} + t^2 b_t(i, j) \quad \text{for } i \ne j, \quad p_t(i, i) = 1 + tq_{i,i} + t^2 b_t(i, i). \tag{3.2}$$

Since $(P_t)_{t\in\mathbb{R}_+}$ is Markovian (resp. sub-Markovian), $p_t(i, j) \ge 0$ for all $t \in \mathbb{R}_+$, $1 \le i, j \le N$. We then deduce from (3.2) that for $i \ne j$ and $t \in (0, 1]$, $q_{i,j} \ge -tb_t(i, j)$. Letting $t \to 0^+$ gives $q_{i,j} \ge 0$ since $b_t(i, j)$ is bounded on $(0, 1]$.

If the semigroup is sub-Markovian, then

$$1 \ge \sum_{j=1}^{N} p_t(i, j) = 1 + t\sum_{j=1}^{N} q_{i,j} + t^2 \sum_{j=1}^{N} b_t(i, j).$$

Therefore $\sum_{j=1}^{N} q_{i,j} + t\sum_{j=1}^{N} b_t(i, j) \le 0$ for $t \in (0, 1]$. Letting $t \to 0^+$ gives $\sum_{j=1}^{N} q_{i,j} \le 0$ since $b_t(i, j)$ is bounded on $(0, 1]$. This implies

$$\sum_{\substack{j=1 \\ j\ne i}}^{N} q_{i,j} \le -q_{i,i}.$$

The Markovian case is analogous with equality instead of inequality. This proves one implication of (ii).

Conversely, assume that $(P_t)_{t\in\mathbb{R}_+}$ satisfies (a), (b), (c) and (3.1). We want to show that $(P_t)_{t\in\mathbb{R}_+}$ is Markovian (resp. sub-Markovian).

We first show that P_t has nonnegative entries for all $t \in \mathbb{R}_+$. Assume that $q_{i,j} \neq 0$ for all $1 \leq i, j \leq N$. Then necessarily $q_{i,j} > 0$ if $i \neq j$ and $q_{i,i} < 0$. Since $P_t = e^{tG}$ by (i), we can again write $P_t = \mathrm{Id}_N + tG + t^2 B_t$ and obtain (3.2). For the case $i \neq j$, the linear term $tq_{i,j} > 0$ dominates for sufficiently small t since $b_t(i,j)$ is bounded on $[0,1]$. For the case $i = j$, the constant term 1 dominates near $t = 0$ since $b_t(i,i)$ is bounded on $[0,1]$. Thus there exists $\varepsilon > 0$ such that $p_t(i,j) > 0$ for all $t \in (0,\varepsilon]$ and all $1 \leq i, j \leq N$.

Now take any $t_0 > 0$ and choose $n \geq 1$ with $0 < \frac{t_0}{n} < \varepsilon$. From above, it follows that $P_{t_0/n}$ has positive entries, hence so does $P_{t_0} = (P_{t_0/n})^n$ by the semigroup property. Since $P_0 = \mathrm{Id}_N$ has nonnegative entries, we deduce that P_t has nonnegative entries for all $t \in \mathbb{R}_+$.

For the general case (without the assumption $q_{i,j} \neq 0$), let $\eta \in (0,1)$ and define $G^\eta = (q^\eta_{i,j})$ by $q^\eta_{i,j} := q_{i,j} + \eta > 0$ for $i \neq j$ and $q^\eta_{i,i} := q_{i,i} - (N-1)\eta < 0$. We subtract $(N-1)\eta$ on the diagonal so that each row sum in G^η is the same as in G and the condition (3.1) is preserved for G^η. Then $G^\eta = G + \eta M$ for some matrix $M \in \mathcal{M}_N(\mathbb{R})$ defined by

$$M_{i,j} := \begin{cases} 1, & i \neq j, \\ -(N-1), & i = j. \end{cases}$$

Set $P^\eta_t := e^{tG^\eta}$. By the previous argument, P^η_t has nonnegative entries for all $t \in \mathbb{R}_+$ and all $\eta \in (0,1)$. Note that the exponential map $\exp : M_N(\mathbb{R}) \to M_N(\mathbb{R})$ is continuous as it is the uniform limit of the continuous partial sums $f_n(A)$. Thus $P_t = \lim_{\eta\to 0^+} P^\eta_t$ and therefore P_t has nonnegative entries for all $t \in \mathbb{R}_+$.

Finally, let $v \in \mathbb{R}^N$ be the column vector of ones. Then by (3.1), $Gv = 0$ (resp. entries non-positive). Using (i) and associativity of matrix multiplication,

$$\tfrac{\mathrm{d}}{\mathrm{d}t}(P_t v) = (\tfrac{\mathrm{d}}{\mathrm{d}t}P_t)v = (P_t G)v = P_t(Gv).$$

Since $Gv = 0$ (resp. entries non-positive), we deduce that for each $1 \leq i \leq N$, the function $t \in \mathbb{R}_+ \mapsto P_t v(i) = \sum_{j=1}^N p_t(i,j)$ is constant (resp. non-increasing). As $P_0 v = \mathrm{Id}_N v = v$, it follows that $\sum_{j=1}^N p_t(i,j) = 1$ (resp. ≤ 1) for all $t \in \mathbb{R}_+$. This proves the converse implication of (ii), and completes the proof of the theorem. □

Remark 3.15 (Intuition). This remark provides an interpretation of the statements in the preceding theorem. Some of these interpretations will be justified in later sections.

(c) The matrix G is called the *infinitesimal generator* and we will later see that it describes the instantaneous rate of change of the system.

(i) The entire semigroup $(P_t)_{t\in\mathbb{R}_+}$ is determined by its infinitesimal generator G via the matrix exponential. We will see later that the differential equations satisfied by $(P_t)_{t\in\mathbb{R}_+}$ are known as the *Chapman-Kolmogorov equations*.

(ii) In the Markovian (resp. sub-Markovian) case, we will see that the off-diagonal entries $q_{i,j}$ $(i \neq j)$ represent the instantaneous transition rates from state i to state j. Consequently, the total rate at which probability mass leaves state i (i.e. the sum of all off-diagonal entries in the i-th row) is equal to $-q_{i,i}$ (resp. $\leq -q_{i,i}$).

3.1.3 Definition of a Markov Process and the Simple Markov Property

Recall that $(\Omega, \mathcal{F}, \mathbb{P})$ denotes a probability space on which, unless otherwise specified, all random variables under consideration are defined. Throughout this section, let E be a countable set equipped with the discrete topology, and hence endowed with the σ-algebra $\mathcal{P}(E)$. Note again that any function defined on E and taking values in a measurable space is therefore measurable. We begin with a first definition of Markov processes.

Definition 3.16 (First definition of discrete-state Markov processes)**.** A process $\mathbf{X} = (X_t)_{t\in\mathbb{R}_+}$ defined on $(\Omega, \mathcal{F})$ and taking values in E is a *Markov process under* $\mathbb{P}$ if and only if there exists a Markovian semigroup on E, denoted $(P_t)_{t\in\mathbb{R}_+}$ with $P_t := (p_t(i,j))_{i,j\in E}$, such that for all $0 = t_0 < t_1 < \cdots < t_n$ and for all $i_0, \ldots, i_n \in E$, we have

$$\mathbb{P}(X_0 = i_0, X_{t_1} = i_1, \ldots, X_{t_n} = i_n) = \mathbb{P}(X_0 = i_0)p_{t_1}(i_0, i_1)p_{t_2-t_1}(i_1, i_2)\cdots p_{t_n-t_{n-1}}(i_{n-1}, i_n).$$

The law of X_0 under $\mathbb{P}$ is called the initial law (or entrance law) of $\mathbf{X}$ and is often denoted by μ. In summary, we say that $\mathbf{X}$ is a Markov process with values in E, with semigroup $(P_t)_{t\in\mathbb{R}_+}$ and initial law μ.

Note that by taking $n = 1$ and $t_1 = t$ in the above factorization, we have

$$\mathbb{P}(X_0 = i, X_t = j) = \mathbb{P}(X_0 = i)p_t(i, j).$$

For any i such that $\mu(i) = \mathbb{P}(X_0 = i) > 0$, we may divide both sides by $\mu(i)$ to obtain

$$\mathbb{P}(X_t = j \mid X_0 = i) = p_t(i, j).$$

Thus, starting from state i, the probability of being in state j after a time t is $p_t(i, j)$, which shows that the quantities $p_t(i, j)$ can indeed be interpreted as transition probabilities of the process.

Intuitively, the factorization in the definition expresses that the probability of the process following the path $i_0 \to i_1 \to \cdots \to i_n$ at the successive times $0 = t_0 < t_1 < \cdots < t_n$ is obtained by multiplying the probability of starting in i_0, given by $\mu(i_0) = \mathbb{P}(X_0 = i_0)$, with the probabilities $p_{t_k-t_{k-1}}(i_{k-1}, i_k)$ of each individual transition from state i_{k-1} to state i_k after an additional time $t_k - t_{k-1}$. In other words, the joint probability of the whole path is the product of the initial probability and all intermediate transition probabilities.

It is useful to introduce a filtration in order to formalize the dependence structure of Markov processes. This leads to the following definition.

Definition 3.17. Let $\mathbf{X} = (X_t)_{t\in\mathbb{R}_+}$ be a Markov process with values in E defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ be a filtration on $(\Omega, \mathcal{F})$. We say that $\mathbf{X}$ is a *Markov process relative to the filtration* $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ if and only if for all $s, t \in \mathbb{R}_+$ and for every bounded function $f : E \to \mathbb{R}_+$, there exists a function $g : E \to \mathbb{R}$ such that

$$\mathbb{E}[f(X_{t+s}) \mid \mathcal{F}_t] = g(X_t) \quad \mathbb{P}\text{-a.s.}$$

Note that g may depend on f, t and s, and both f and g are $\mathcal{P}(E)$-measurable. Moreover, X_t is implicitly $\mathcal{F}_t$-measurable, i.e. $\mathbf{X}$ is adapted to the filtration.

The relation in the definition expresses that the conditional expectation of the future random variable $f(X_{t+s})$, given the information $\mathcal{F}_t$ available up to time t, only depends on the present state X_t. The following lemma reformulates this property in terms of transition probabilities.

Lemma 3.18. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space endowed with a filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$. Let $\mathbf{X} = (X_t)_{t\in\mathbb{R}_+}$ be a Markov process with values in E relative to $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$. Denote by $(P_t)_{t\in\mathbb{R}_+}$ its semigroup, with $P_t := (p_t(i,j))_{i,j\in E}$. Then, for all $s, t \in \mathbb{R}_+$ and all $j \in E$, we have*

$$\mathbb{E}[\mathbf{1}_{\{X_{t+s}=j\}} \mid \mathcal{F}_t] = p_s(X_t, j) \quad \mathbb{P}\text{-a.s.}$$

Consequently, the conditional distribution of X_{t+s} given $\mathcal{F}_t$ is $P_s(X_t, \mathrm{d}y) = \sum_{j\in E} p_s(X_t, j)\mathbf{1}_{\{j\}}(y)$.

Proof. By definition, since the indicator function is bounded and nonnegative, there exists $g : E \to \mathbb{R}$ such that $\mathbb{E}[\mathbf{1}_{\{X_{t+s}=j\}} \mid \mathcal{F}_t] = g(X_t)$ $\mathbb{P}$-a.s. For all $i, i_0 \in E$, we have

$$\begin{aligned}
\mathbb{P}(X_0 = i_0)p_t(i_0, i)p_s(i, j) &= \mathbb{P}(X_0 = i_0, X_t = i, X_{t+s} = j) \\
&= \mathbb{E}[\mathbf{1}_{\{X_0=i_0, X_t=i, X_{t+s}=j\}}] \\
&= \mathbb{E}[\mathbf{1}_{\{X_0=i_0, X_t=i\}}\mathbb{E}[\mathbf{1}_{\{X_{t+s}=j\}} \mid \mathcal{F}_t]] \\
&= \mathbb{E}[\mathbf{1}_{\{X_0=i_0, X_t=i\}}g(X_t)] \\
&= \mathbb{P}(X_0 = i_0, X_t = i)g(i) \\
&= \mathbb{P}(X_0 = i_0)p_t(i_0, i)g(i)
\end{aligned}$$

where we used Definition 3.16 and the weak tower property.

We observe that the first and last expressions in the previous display differ only by the factors $g(i)$ and $p_s(i,j)$. Our goal is to conclude that $g(i) = p_s(i,j)$ for all states i that are actually visited by the process (i.e. such that $\mathbb{P}(X_t = i) > 0$ for some $t \in \mathbb{R}_+$). From this, the result will follow directly.

Assume that $\mathbb{P}(X_t = i) > 0$. To deduce from the last display that $g(i) = p_s(i,j)$, it suffices to ensure that $\mathbb{P}(X_0 = i_0)p_t(i_0, i)$ is strictly positive for some $i_0 \in E$. Indeed,

$$0 < \mathbb{P}(X_t = i) = \sum_{i_0 \in E} \mathbb{P}(X_0 = i_0)p_t(i_0, i)$$

so there must exist $i_0 \in E$ with $\mathbb{P}(X_0 = i_0)p_t(i_0, i) > 0$. □

The lemma shows that, conditionally on $\mathcal{F}_t$ (the information available up to time t), the probability of being in state j at future time $t + s$ depends on the past only through the present state X_t and is given by $p_s(X_t, j)$.

In order to gather the preceding properties into a single framework, we introduce the notion of a global Markov process.

Definition 3.19 (Global Markov process)**.** A (global) Markov process with values in E consists of the data

$$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, \mathbf{X} = (X_t)_{t\in\mathbb{R}_+}, (P_t)_{t\in\mathbb{R}_+}, \mathbb{P}_\mu, \mu \in \mathcal{M}_1(E))$$

where:

(a) $(\Omega, \mathcal{F})$ is a measurable space and $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ a filtration on $(\Omega, \mathcal{F})$.

(b) For all $t \in \mathbb{R}_+$, the map $X_t : \Omega \to E$ is $\mathcal{F}_t$-measurable, i.e. the process $\mathbf{X}$ is adapted to $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$.

(c) $(P_t)_{t\in\mathbb{R}_+}$ is a Markovian semigroup. For all $t \in \mathbb{R}_+$, we write $P_t := (p_t(i,j))_{i,j\in E}$.

(d) For every $\mu \in \mathcal{M}_1(E)$ (the set of probability measures on E), $\mathbb{P}_\mu$ is a probability measure on $(\Omega, \mathcal{F})$ such that for all $s, t \in \mathbb{R}_+$ and all $j \in E$,

$$\mathbb{E}_\mu[\mathbf{1}_{\{X_{t+s}=j\}} \mid \mathcal{F}_t] = p_s(X_t, j) \quad \mathbb{P}_\mu\text{-a.s.}$$

Equivalently, for any bounded or nonnegative function $f : E \to \mathbb{R}$,

$$\mathbb{E}_\mu[f(X_{t+s}) \mid \mathcal{F}_t] = \sum_{j\in E} p_s(X_t, j)f(j) =: (P_s f)(X_t) \quad \mathbb{P}_\mu\text{-a.s.}$$

Proof of the equivalence in (d). The implication from the second statement to the first is immediate, since taking $f = \mathbf{1}_{\{j\}}$, which is bounded and nonnegative, gives $\mathbb{E}_\mu[\mathbf{1}_{\{X_{t+s}=j\}} \mid \mathcal{F}_t] = p_s(X_t, j)$ $\mathbb{P}_\mu$-a.s.

To show the other implication, note first that since E is countable, we may enumerate $E = \{j_1, j_2, \dots\}$. Now let $f : E \to \mathbb{R}$ bounded or nonnegative, and define a sequence of simple functions $(f_n)_{n\geq 1}$ by $f_n := \sum_{k=1}^n f(j_k)\mathbf{1}_{\{j_k\}}$. Note that $f_n \to f$ pointwise on E and $\mathbb{P}_\mu$-a.s.

$$\begin{aligned}
\mathbb{E}_\mu[f_n(X_{t+s}) \mid \mathcal{F}_t] &= \sum_{k=1}^n f(j_k)\mathbb{E}_\mu[\mathbf{1}_{\{X_{t+s}=j_k\}} \mid \mathcal{F}_t] \\
&= \sum_{k=1}^n f(j_k)p_s(X_t, j_k) \\
&= \sum_{j\in E} p_s(X_t, j) \sum_{k=1}^n f(j_k)\mathbf{1}_{\{j=j_k\}} \\
&= \sum_{j\in E} p_s(X_t, j)f_n(j).
\end{aligned}$$

If f is bounded, then $|f_n(j)| \leq \|f\|_\infty = \sup_{j\in E} |f(j)|$ for all $j \in E$, $n \geq 1$ and taking limits in the last display, the dominated convergence theorem yields

$$\mathbb{E}_\mu[f(X_{t+s}) \mid \mathcal{F}_t] = \sum_{j\in E} p_s(X_t, j)f(j) \quad \mathbb{P}_\mu\text{-a.s.}$$

If instead f is nonnegative, the sequence $(f_n)_{n\geq 1}$ is non-decreasing and $f_n \uparrow f$. Again, taking limits in the display above, but now using the monotone convergence theorem gives

$$\mathbb{E}_\mu[f(X_{t+s}) \mid \mathcal{F}_t] = \sum_{j\in E} p_s(X_t, j)f(j) \quad \mathbb{P}_\mu\text{-a.s.}$$

□

Notation 3.20. We write $\delta_{i,\cdot}$ for the Dirac measure at $i \in E$, i.e. $\delta_{i,j} := 1$ if $i = j$ and $\delta_{i,j} := 0$ otherwise. We denote $\mathbb{P}_{\delta_{i,\cdot}}$ simply by $\mathbb{P}_i$, and the corresponding expectation $\mathbb{E}_{\delta_{i,\cdot}}$ by $\mathbb{E}_i$.

The following theorem formalizes the idea that, given the past up to time t (i.e. $\mathcal{F}_t$), the future $(X_{t+s})_{s\in\mathbb{R}_+}$ behaves like the Markov process started at the random state X_t.

Theorem 3.21 (Simple Markov property)**.** *Consider a Markov process*

$$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, \mathbf{X} = (X_t)_{t\in\mathbb{R}_+}, (P_t)_{t\in\mathbb{R}_+}, \mathbb{P}_\mu, \mu \in \mathcal{M}_1(E)).$$

For any $\mu \in \mathcal{M}_1(E)$, any $t \in \mathbb{R}_+$ and any bounded measurable function $F : E^{\mathbb{R}_+} \to \mathbb{R}$, we have the following equality, called the simple Markov property at time t*:*

$$\mathbb{E}_\mu[F(\theta_t\mathbf{X}) \mid \mathcal{F}_t] = \mathbb{E}_{X_t}[F(\mathbf{X})] \quad \mathbb{P}_\mu\text{-a.s.}$$

Here $\mathbb{E}_{X_t}[F(\mathbf{X})]$ denotes the composition of $\omega \in \Omega \mapsto X_t(\omega) \in E$ with $i \in E \mapsto \mathbb{E}_i[F(\mathbf{X})] \in \mathbb{R}$.

Proof. First, note that $\omega \mapsto F(\theta_t\mathbf{X}(\omega))$ is $\mathcal{F}$-measurable, since θ_t is measurable by Lemma 3.4. Moreover, $\omega \mapsto \mathbb{E}_{X_t(\omega)}[F(\mathbf{X})]$ is $\mathcal{F}_t$-measurable, since it is the composition of the $\mathcal{F}_t$-measurable map $\omega \mapsto X_t(\omega)$ with the $\mathcal{P}(E)$-measurable map $i \mapsto \mathbb{E}_i[F(\mathbf{X})]$.

Fix $\mu \in \mathcal{M}_1(E)$, $t \in \mathbb{R}_+$ and let $C \in \mathcal{C}$. Then there exist $0 \leq t_0 < \cdots < t_n$ and $i_0, \ldots, i_n \in E$ such that

$$C = \{\mathbf{x} = (x(s))_{s\in\mathbb{R}_+} \in E^{\mathbb{R}_+} : x(t_0) = i_0, \ldots, x(t_n) = i_n\}.$$

We first prove the simple Markov property for $F := \mathbf{1}_C$. In this case, we have that $F(\theta_t\mathbf{X}) = \mathbf{1}_{\{X_{t+t_0}=i_0,\ldots,X_{t+t_n}=i_n\}}$ and the event $\{X_{t+t_0} = i_0, \ldots, X_{t+t_{n-1}} = i_{n-1}\}$ lies in $\mathcal{F}_{t+t_{n-1}}$, because the process $\mathbf{X}$ is adapted to $(\mathcal{F}_s)_{s\in\mathbb{R}_+}$. Thus $\mathbb{P}_\mu$-a.s.

$$\begin{aligned}\mathbb{E}_\mu[F(\theta_t\mathbf{X}) \mid \mathcal{F}_{t+t_{n-1}}] &= \mathbf{1}_{\{X_{t+t_0}=i_0,\ldots,X_{t+t_{n-1}}=i_{n-1}\}}\mathbb{E}_\mu[\mathbf{1}_{\{X_{t+t_n}=i_n\}} \mid \mathcal{F}_{t+t_{n-1}}] \\ &= \mathbf{1}_{\{X_{t+t_0}=i_0,\ldots,X_{t+t_{n-1}}=i_{n-1}\}}p_{t_n-t_{n-1}}(X_{t+t_{n-1}}, i_n)\end{aligned}$$

where we used Lemma 3.18. Using the tower property and again Lemma 3.18, we have $\mathbb{P}_\mu$-a.s.

$$\begin{aligned}&\mathbb{E}_\mu[F(\theta_t\mathbf{X}) \mid \mathcal{F}_{t+t_{n-2}}] \\ &= \mathbb{E}_\mu[\mathbb{E}_\mu[F(\theta_t\mathbf{X}) \mid \mathcal{F}_{t+t_{n-1}}] \mid \mathcal{F}_{t+t_{n-2}}] \\ &= \mathbf{1}_{\{X_{t+t_0}=i_0,\ldots,X_{t+t_{n-2}}=i_{n-2}\}}\mathbb{E}_\mu[\mathbf{1}_{\{X_{t+t_{n-1}}=i_{n-1}\}}p_{t_n-t_{n-1}}(X_{t+t_{n-1}}, i_n) \mid \mathcal{F}_{t+t_{n-2}}] \\ &= \mathbf{1}_{\{X_{t+t_0}=i_0,\ldots,X_{t+t_{n-2}}=i_{n-2}\}}p_{t_n-t_{n-1}}(i_{n-1}, i_n)\mathbb{E}_\mu[\mathbf{1}_{\{X_{t+t_{n-1}}=i_{n-1}\}} \mid \mathcal{F}_{t+t_{n-2}}] \\ &= \mathbf{1}_{\{X_{t+t_0}=i_0,\ldots,X_{t+t_{n-2}}=i_{n-2}\}}p_{t_{n-1}-t_{n-2}}(X_{t+t_{n-2}}, i_{n-1})p_{t_n-t_{n-1}}(i_{n-1}, i_n).\end{aligned}$$

Iterating this reasoning yields

$$\mathbb{E}_\mu[F(\theta_t\mathbf{X}) \mid \mathcal{F}_t] = p_{t_0}(X_t, i_0)p_{t_1-t_0}(i_0, i_1)\cdots p_{t_n-t_{n-1}}(i_{n-1}, i_n) \quad \mathbb{P}_\mu\text{-a.s.}$$

Now fix $i \in E$, take $\mu(\cdot) = \delta_{i,\cdot}$ and $t = 0$. Since $\mathbb{P}_i(X_0 = i) = 1$, taking $\mathbb{E}_i$ of both sides in the last equation and using the weak tower property gives

$$\mathbb{E}_i[F(\mathbf{X})] = p_{t_0}(i, i_0)p_{t_1-t_0}(i_0, i_1)\cdots p_{t_n-t_{n-1}}(i_{n-1}, i_n).$$

Combining the last two equations gives, for arbitrary μ and t,

$$\mathbb{E}_\mu[F(\theta_t\mathbf{X}) \mid \mathcal{F}_t] = \mathbb{E}_{X_t}[F(\mathbf{X})] \quad \mathbb{P}_\mu\text{-a.s.}$$

This proves the simple Markov property for $F = \mathbf{1}_C$.

Next, we show that the simple Markov property holds for all $F = \mathbf{1}_A$ where $A \in \mathcal{P}(E)^{\mathbb{R}_+}$. Let $\mathcal{D}$ denote the collection of sets $A \in \mathcal{P}(E)^{\mathbb{R}_+}$ for which the simple Markov property holds with $F = \mathbf{1}_A$. By what we showed above, $\mathcal{C} \subseteq \mathcal{D}$. We claim that $\mathcal{D}$ is a monotone class. Indeed, let $(A_n)_{n\geq 1} \subseteq \mathcal{D}$. If $A_n \uparrow A := \cup_{n\geq 1} A_n$, then $\mathbf{1}_{A_n} \uparrow \mathbf{1}_A$, and by the monotone convergence theorem we obtain

$$\mathbb{E}_\mu[\mathbf{1}_{A_n}(\theta_t\mathbf{X}) \mid \mathcal{F}_t] \uparrow \mathbb{E}_\mu[\mathbf{1}_A(\theta_t\mathbf{X}) \mid \mathcal{F}_t]$$

and

$$\mathbb{E}_{X_t}[\mathbf{1}_{A_n}(\mathbf{X})] \uparrow \mathbb{E}_{X_t}[\mathbf{1}_A(\mathbf{X})].$$

Therefore, passing to the limit in the simple Markov property for $F = \mathbf{1}_{A_n}$ yields the simple Markov property for $F = \mathbf{1}_A$, so $A \in \mathcal{D}$. If $A_n \downarrow A := \cap_{n\geq 1} A_n$, a similar reasoning applies, using the dominated convergence theorem (since $0 \leq \mathbf{1}_{A_n} \leq 1$), and again we obtain $A \in \mathcal{D}$. Hence $\mathcal{D}$ is a monotone class containing the π-system $\mathcal{C}$, which generates $\mathcal{P}(E)^{\mathbb{R}_+}$. By the monotone class theorem (see Theorem A.1.10 in Section A.1 of [Duq23]), it follows that $\mathcal{D} = \mathcal{P}(E)^{\mathbb{R}_+}$. Thus the simple Markov property holds for all $F = \mathbf{1}_A$ where $A \in \mathcal{P}(E)^{\mathbb{R}_+}$.

Now we prove the simple Markov property for simple functions $F := \sum_{k=1}^m a_k \mathbf{1}_{A_k}$ with $a_k \in \mathbb{R}$ and $A_k \in \mathcal{P}(E)^{\mathbb{R}_+}$. Using the result above for indicators, we have $\mathbb{P}_\mu$-a.s.

$$\begin{aligned}
\mathbb{E}_\mu[F(\theta_t\mathbf{X}) \mid \mathcal{F}_t] &= \sum_{k=1}^m a_k \mathbb{E}_\mu[\mathbf{1}_{A_k}(\theta_t\mathbf{X}) \mid \mathcal{F}_t] \\
&= \sum_{k=1}^m a_k \mathbb{E}_{X_t}[\mathbf{1}_{A_k}(\mathbf{X})] \\
&= \mathbb{E}_{X_t}\Big[\sum_{k=1}^m a_k \mathbf{1}_{A_k}(\mathbf{X})\Big] \\
&= \mathbb{E}_{X_t}[F(\mathbf{X})].
\end{aligned}$$

We now extend the result to general bounded measurable functions $F : E^{\mathbb{R}_+} \to \mathbb{R}$. By Lemma 2.2, there exists a sequence of simple $\mathcal{P}(E)^{\mathbb{R}_+}$-measurable functions $(F_n)_{n\geq 1}$ of the form $F_n = \sum_{k=1}^{m_n} a_{k,n} \mathbf{1}_{A_{k,n}}$ with $a_{k,n} \in \mathbb{R}$, $A_{k,n} \in \mathcal{P}(E)^{\mathbb{R}_+}$ such that $|F_n| \leq \|F\|_\infty$ and $F_n \to F$ pointwise on $E^{\mathbb{R}_+}$. Using the result above for simple functions, we get for each $n \geq 1$

$$\mathbb{E}_\mu[F_n(\theta_t\mathbf{X}) \mid \mathcal{F}_t] = \mathbb{E}_{X_t}[F_n(\mathbf{X})] \quad \mathbb{P}_\mu\text{-a.s.}$$

By dominated convergence (using $|F_n(\theta_t\mathbf{X})| \leq \|F\|_\infty$ and $|F_n(\mathbf{X})| \leq \|F\|_\infty$), we obtain

$$\mathbb{E}_\mu[F(\theta_t\mathbf{X}) \mid \mathcal{F}_t] = \mathbb{E}_{X_t}[F(\mathbf{X})] \quad \mathbb{P}_\mu\text{-a.s.}$$

which completes the proof. □

3.1.4 Right-Continuous Processes and the Explosion Time

In this section, we want to address a technical difficulty in our previous definitions. These definitions do not exclude certain pathological or irregular processes, which we illustrate with the following example.

Example 3.22. Let $E := \{0, 1\}$ and fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose that for all $t \in \mathbb{R}_+$, there exists a random variable $X_t : \Omega \to \{0, 1\}$ which is $\mathcal{F}$-measurable. We assume that $\mathbb{P}(X_0 = 0) = 1$, that the random variables $(X_t)_{t\in\mathbb{R}_+}$ are mutually independent (i.e. for every finite family $t_1, \ldots, t_n \in \mathbb{R}_+$, the variables $X_{t_1}, \ldots, X_{t_n}$ are independent) and that for all $t > 0$, we have $\mathbb{P}(X_t = 0) = \mathbb{P}(X_t = 1) = \frac{1}{2}$. Thus, $(X_t)_{t>0}$ is a family of i.i.d. Bernoulli random variables with parameter $1/2$, while X_0 is almost surely equal to 0.

Lemma 3.2 ensures that $\mathbf{X} = (X_t)_{t\in\mathbb{R}_+} : \Omega \to \{0,1\}^{\mathbb{R}_+}$ is $(\mathcal{F}, \mathcal{P}(\{0,1\})^{\mathbb{R}_+})$-measurable. It is also easy to verify that $\mathbf{X}$ is a Markov process with Markovian semigroup $P_t = (p_t(i,j))_{i,j\in\{0,1\}}$, $t \in \mathbb{R}_+$, defined by $P_0 := \mathrm{Id}_{\{0,1\}}$ and $p_t(i,j) := \frac{1}{2}$ for all $i, j \in \{0,1\}, t > 0$ (see Example 3.11). Indeed, take $0 = t_0 < t_1 < \cdots < t_n$ and $i_0, \ldots, i_n \in \{0,1\}$. By mutual independence of $(X_t)_{t\in\mathbb{R}_+}$, we have

$$\mathbb{P}(X_{t_0} = i_0, \ldots, X_{t_n} = i_n) = \mathbb{P}(X_{t_0} = i_0) \prod_{k=1}^{n} \mathbb{P}(X_{t_k} = i_k).$$

For every $k \geq 1$, we have $t_k > 0$ and hence $\mathbb{P}(X_{t_k} = i_k) = \frac{1}{2}$. At the same time, for every $k \geq 1$, we have $t_k - t_{k-1} > 0$ and thus $p_{t_k - t_{k-1}}(i_{k-1}, i_k) = \frac{1}{2}$. Therefore

$$\mathbb{P}(X_{t_0} = i_0, \ldots, X_{t_n} = i_n) = \mathbb{P}(X_{t_0} = i_0) \prod_{k=1}^{n} p_{t_k - t_{k-1}}(i_{k-1}, i_k)$$

which proves that $\mathbf{X}$ is a Markov process.

This very simple Markov process fits within our definitions, but we will now see that it is extremely irregular. Taking an elementary cylinder

$$C := \{\mathbf{x} = (x(t))_{t\in\mathbb{R}_+} \in \{0,1\}^{\mathbb{R}_+} : x(t_0) = i_0, \ldots, x(t_n) = i_n\},$$

mutual independence of $(X_t)_{t\in\mathbb{R}_+}$ yields

$$\mathbb{P}(C) = \begin{cases} \mathbf{1}_{\{i_0=0\}}(\frac{1}{2})^n, & \text{if } t_0 = 0, \\ (\frac{1}{2})^{n+1}, & \text{if } t_0 > 0. \end{cases}$$

Thus, except for the deterministic constraint $X_0 = 0$, every elementary cylinder has positive probability. So the image of $\mathbf{X} : \Omega \to \{0,1\}^{\mathbb{R}_+}$ must realize all possible assignments of values on any finite set of strictly positive times. This forces the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to have cardinality at least that of $\{0,1\}^{\mathbb{R}_+}$ (i.e. $|\Omega| \geq 2^{|\mathbb{R}_+|}$), which is extremely large.

Furthermore, the trajectory of $\mathbf{X}$ is extremely irregular. Indeed, if we consider the random set of zeros $S := \{t \in \mathbb{R}_+ : X_t = 0\}$, the independence of $(X_t)_{t\in\mathbb{R}_+}$ implies that, almost surely, every nontrivial interval of $\mathbb{R}_+$ contains infinitely many points of both S and S^c. So the trajectory of $\mathbf{X}$ oscillates endlessly and is thus extremely irregular.

To avoid such highly irregular processes, it is necessary to restrict to a class of processes whose trajectories are regular. For this purpose, we introduce the notion of right-continuity for processes with values in a countable space equipped with the discrete topology.

Definition 3.23 (Right-continuous trajectories)**.** Let E be a countable set equipped with the discrete topology, and hence endowed with the σ-algebra $\mathcal{P}(E)$. As proved previously, the discrete topology is generated by the discrete metric d defined, for all $i, j \in E$, by

$$d(i,j) := \begin{cases} 0, & i = j, \\ 1, & i \neq j. \end{cases}$$

Let $\mathbf{x} = (x(t))_{t\in\mathbb{R}_+} \in E^{\mathbb{R}_+}$ be a function from $\mathbb{R}_+$ to E. We say that $\mathbf{x}$ is *right-continuous* if and only if for all $t \in \mathbb{R}_+$,

$$\lim_{h\to 0^+} d(x(t), x(t+h)) = 0.$$

This is equivalent to requiring that for all $t \in \mathbb{R}_+$, there exists $h_t > 0$ such that $x(t) = x(t+u)$ for all $u \in [0, h_t]$. In other words, a function from $\mathbb{R}_+$ to E is right-continuous if and only if for each $t \in \mathbb{R}_+$ it remains constant after t during a strictly positive time interval.

Proof of the equivalence. We first assume that for all $t \in \mathbb{R}_+$, $\lim_{h\to 0^+} d(x(t), x(t+h)) = 0$. This means that for sufficiently small $h > 0$ we have $d(x(t), x(t+h)) = 0$. Since d only takes values in $\{0,1\}$, it follows that $x(t) = x(t+h)$. Hence there exists $h_t > 0$ such that $x(t) = x(t+u)$ for all $u \in [0, h_t]$.

Conversely, suppose that there exists $h_t > 0$ such that $x(t) = x(t+u)$ for all $u \in [0, h_t]$. Then for all $0 < h \leq h_t$, $d(x(t), x(t+h)) = d(x(t), x(t)) = 0$. As $h \to 0^+$, we eventually have $h \leq h_t$, and therefore $\lim_{h\to 0^+} d(x(t), x(t+h)) = 0$. □

The following example shows that right-continuous functions with values in a countable space (equipped with the discrete topology) can nevertheless display highly irregular behavior.

Example 3.24. Take $E := \mathbb{Z}$ with the discrete topology. Define $\mathbf{x} = (x(t))_{t\in\mathbb{R}_+}$ on the interval $[0, 1)$ as

$$x(t) := (-1)^n n \quad \text{if } t \in [1 - 2^{-n}, 1 - 2^{-(n+1)}), \quad n \in \mathbb{N}.$$

Then extend $\mathbf{x}$ to all $t \geq 0$ by 1-periodicity, i.e. by requiring $x(t + 1) = x(t)$ for all $t \geq 0$.

In words, $\mathbf{x}$ evolves on each unit interval $[p - 1, p)$, $p \in \mathbb{N} \setminus \{0\}$, by alternating between positive and negative values whose magnitude increases. On each subinterval $[1 - 2^{-n}, 1 - 2^{-(n+1)})$, the function remains constant, and these constant pieces are called plateaus. As t approaches p from the left, the plateaus become shorter and shorter and accumulate at $t = p$, while the jumps between them grow in size. Hence the trajectory of $\mathbf{x}$ exhibits increasingly extreme oscillations as $t \to p^-$.

We see that $\mathbf{x}$ is right-continuous. Indeed, on each interval $I_n = [1-2^{-n}, 1-2^{-(n+1)})$, $n \in \mathbb{N}$, the function $\mathbf{x}$ is constant. Thus, for any $t \in \mathbb{R}_+$ there exists some n such that $t \in I_n$, and a small right neighborhood of t is contained in I_n, where $\mathbf{x}$ does not change. Hence $\lim_{h\to 0^+} x(t + h) = x(t)$.

However, $\mathbf{x}$ does not admit any left limits at strictly positive integer times, since for all $p \in \mathbb{N}\setminus\{0\}$

$$\liminf_{h\to 0^+} x(p - h) = -\infty, \quad \limsup_{h\to 0^+} x(p - h) = +\infty.$$

Thus, a right-continuous function does not necessarily admit left limits, and even if it progresses by successive steps, the jump times may accumulate (as they do here at $t = p$). This phenomenon is called *explosion*.

Before studying explosions formally, we recall that we have already defined what it means for a single trajectory to be right-continuous. We now extend this notion to the level of stochastic processes, by requiring that every trajectory of the process is right-continuous.

Definition 3.25 (Right-continuous processes)**.** Let E be a countable set equipped with the discrete topology, and hence endowed with the σ-algebra $\mathcal{P}(E)$. Let $(\Omega, \mathcal{F})$ be a measurable space. For all $t \in \mathbb{R}_+$, let $X_t : \Omega \to E$ be an $\mathcal{F}$-measurable random variable. We say that the process $\mathbf{X} = (X_t)_{t\in\mathbb{R}_+}$ is a *right-continuous process* (everywhere on Ω) if and only if

$$\text{for every } \omega \in \Omega, \text{ the trajectory } t \in \mathbb{R}_+ \mapsto X_t(\omega) \in E \text{ is right-continuous.}$$

Decomposition of a right-continuous trajectory. Fix a right-continuous function $\mathbf{x} = (x(t))_{t\in\mathbb{R}_+}$ in $E^{\mathbb{R}_+}$ and consider its trajectory up to its possible explosion time (i.e. the possible accumulation of its jump times). We now decompose this trajectory and introduce some notions that describe its jump structure.

(a) (Jump times) We set $J_0(\mathbf{x}) := 0$ and define recursively the sequence $J_n(\mathbf{x}) \in [0, \infty]$, $n \in \mathbb{N}$, by

$$J_{n+1}(\mathbf{x}) := \inf\{t > J_n(\mathbf{x}) : x(t) \neq x(J_n(\mathbf{x}))\}$$

with the convention that $\inf \emptyset := \infty$. The sequence $(J_n(\mathbf{x}))_{n\geq 1}$ is the *sequence of jump times* of $\mathbf{x}$ before explosion.

If $J_n(\mathbf{x}) < \infty$ but $J_{n+1}(\mathbf{x}) = \infty$, then $x(t) = x(J_n(\mathbf{x}))$ for all $t \geq J_n(\mathbf{x})$. We then say that the trajectory is *absorbed* at $x(J_n(\mathbf{x}))$, and in this case $J_p(\mathbf{x}) = \infty$ for all $p \geq n + 1$.

Remark 3.26. Note that since the trajectory $\mathbf{x} : \mathbb{R}_+ \to E$ is right-continuous, the set of (finite) jump times of $\mathbf{x}$ is at most countable. Indeed, let $\mathcal{J} \subseteq (0, \infty)$ be the set of times at which $\mathbf{x}$ changes value. By right-continuity in the discrete metric, for each $t \in \mathcal{J}$ there exists $h_t > 0$ such that $\mathbf{x}$ is constant on $[t, t + h_t]$. In particular, there can be no other change time in $(t, t + h_t)$, and hence the intervals $\{(t, t + h_t)\}_{t\in\mathcal{J}}$ are pairwise disjoint. Now since $\mathbb{Q}$ is dense in $\mathbb{R}$, we can choose a rational number $q_t \in (t, t + h_t)$ for each $t \in \mathcal{J}$. This defines an injection $t \mapsto q_t$ from $\mathcal{J}$ into $\mathbb{Q}$, so $\mathcal{J}$ is at most countable.

(b) (Explosion time) We set

$$\zeta(\mathbf{x}) := \sup_{n\geq 0} J_n(\mathbf{x}) \in [0, \infty].$$

This is the *explosion time* of $\mathbf{x}$.

If $\zeta(\mathbf{x}) = \infty$, we say that $\mathbf{x}$ does not explode. Note that this may occur in two distinct cases:

- The trajectory is absorbed, that is, there exists n such that $J_n(\mathbf{x}) < \infty$ but $J_{n+1}(\mathbf{x}) = \infty$, in which case the trajectory becomes constant after time $J_n(\mathbf{x})$.
- The sequence of jump times diverges, i.e. $J_n(\mathbf{x}) \to \infty$, so that the trajectory keeps evolving indefinitely without accumulating jumps in finite time.

(c) (Number of jumps) We denote

$$N(\mathbf{x}) := \sup\{n \geq 0 : J_n(\mathbf{x}) < \infty\}$$

the *number of jumps* before possible absorption. Thus $N(\mathbf{x}) < \infty$ if and only if the trajectory $\mathbf{x}$ is absorbed in some state.

(d) (Skeleton) We denote the successive values of $\mathbf{x}$ before explosion by

$$Y_n(\mathbf{x}) := \begin{cases} x(J_n(\mathbf{x})), & 0 \leq n \leq N(\mathbf{x}), \\ x(J_{N(\mathbf{x})}(\mathbf{x})), & n > N(\mathbf{x}). \end{cases}$$

The sequence $\mathbf{Y}(\mathbf{x}) := (Y_n(\mathbf{x}))_{n\in\mathbb{N}}$ is called the *skeleton* of $\mathbf{x}$, and it represents the sequence of states visited by $\mathbf{x}$.

We note that if $N(\mathbf{x}) < \infty$ (i.e. if the trajectory is absorbed), then the sequence $(Y_n(\mathbf{x}))_{n\in\mathbb{N}}$ is constant from some point onward.

(e) (Holding times) We define the sequence $D_n(\mathbf{x}) \in [0, \infty]$, $n \geq 1$, of *holding times* between the jumps of $\mathbf{x}$ as

$$D_{n+1}(\mathbf{x}) := \begin{cases} J_{n+1}(\mathbf{x}) - J_n(\mathbf{x}), & 0 \leq n \leq N(\mathbf{x}), \\ \infty, & n > N(\mathbf{x}). \end{cases}$$

We note that each $D_{n+1}(\mathbf{x})$ represents the length of the plateau during which the trajectory stays in the state $Y_n(\mathbf{x})$. If $N(\mathbf{x}) < \infty$ (i.e. if the trajectory is absorbed), then the last holding time is infinite, corresponding to the trajectory remaining forever in its absorbing state.

It is important to note that the data of $(D_n(\mathbf{x}))_{n\geq 1}$ and $(Y_n(\mathbf{x}))_{n\in\mathbb{N}}$ completely determine the trajectory $\mathbf{x}$ on the time interval $[0, \zeta(\mathbf{x}))$. Since $(J_n(\mathbf{x}))_{n\in\mathbb{N}}$ is non-decreasing, we also have

$$\zeta(\mathbf{x}) = \lim_{n\to\infty} J_n(\mathbf{x}) = \lim_{n\to\infty} \sum_{k=1}^{n} D_k(\mathbf{x}) = \sum_{k\geq 1} D_k(\mathbf{x}).$$

Hence $\zeta(\mathbf{x})$ can be interpreted as the sum of all holding times. In particular, $\zeta(\mathbf{x}) = \infty$ either when the trajectory is absorbed ($D_k(\mathbf{x}) = \infty$ from some k onward) or when there is no explosion ($J_n(\mathbf{x}) \to \infty$ as $n \to \infty$).

On the other hand, the two sequences $(D_n(\mathbf{x}))_{n\geq 1}$ and $(Y_n(\mathbf{x}))_{n\in\mathbb{N}}$ say nothing about the trajectory after the explosion time $\zeta(\mathbf{x})$. From now on, we will restrict our attention to the study of Markov processes before their explosion time.

To illustrate these notions and their interpretation, we now consider a simple example of a right-continuous trajectory and its decomposition.

Example 3.27. Let $\mathbf{x} = (x(t))_{t\in\mathbb{R}_+}$ be a trajectory with values in $E := \{0, 1, 2, 3\}$ defined as

$$x(t) := \begin{cases} 0, & 0 \leq t < 1, \\ 2, & 1 \leq t < 2.5, \\ 1, & 2.5 \leq t < 3.5, \\ 3, & t \geq 3.5. \end{cases}$$

This trajectory is right-continuous and absorbed in the state 3 at time $J_3(\mathbf{x}) = 3.5$. The corresponding decomposition is

$$\begin{aligned} &J_0(\mathbf{x}) = 0, \quad J_1(\mathbf{x}) = 1, \quad J_2(\mathbf{x}) = 2.5, \quad J_3(\mathbf{x}) = 3.5, \quad J_4(\mathbf{x}) = \infty, \\ &D_1(\mathbf{x}) = 1, \quad D_2(\mathbf{x}) = 1.5, \quad D_3(\mathbf{x}) = 1, \quad D_4(\mathbf{x}) = \infty, \\ &Y_0(\mathbf{x}) = 0, \quad Y_1(\mathbf{x}) = 2, \quad Y_2(\mathbf{x}) = 1, \quad Y_3(\mathbf{x}) = 3. \end{aligned}$$

Hence the explosion time is $\zeta(\mathbf{x}) = \sum_{n\geq 1} D_n(\mathbf{x}) = \infty$. The trajectory and its decomposition are illustrated in Figure 3.1.

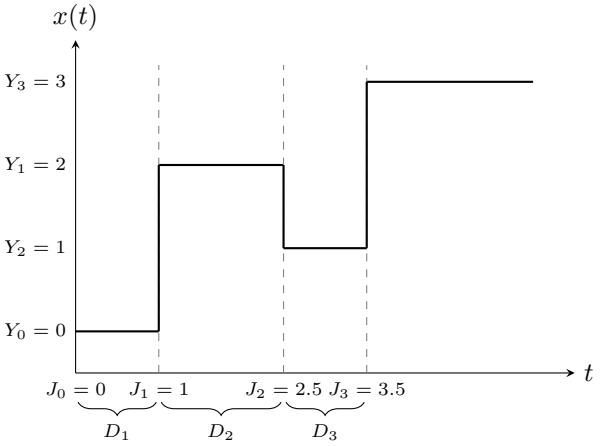

Figure 3.1: The right-continuous trajectory $\mathbf{x}$ and its decomposition into jump times, holding times and skeleton states.

3.1.5 Strong Markov Property

We restrict our study of Markov processes with values in a countable space (equipped with the discrete topology) to those that are right-continuous, and we will prove that they satisfy the strong Markov property. For this, it is necessary to define the notion of a stopping time.

Stopping times. Although very close to discrete stopping times, the notion of a stopping time taking values in $[0, \infty]$ involves some complications.

Definition 3.28. Let $(\Omega, \mathcal{F})$ be a measurable space and $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ a filtration on this space. We set $\mathcal{F}_\infty := \sigma(\mathcal{F}_t; t \in \mathbb{R}_+) \subseteq \mathcal{F}$. Recall the right-continuous filtration $(\mathcal{F}_{t+})_{t\in\mathbb{R}_+}$ defined by $\mathcal{F}_{t+} := \bigcap_{h>0} \mathcal{F}_{t+h}$, noting that $\mathcal{F}_t \subseteq \mathcal{F}_{t+}$ for all $t \in \mathbb{R}_+$.

(a) (Stopping time) A function $T : \Omega \to [0, \infty]$ is a *stopping time with respect to the filtration* $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ if and only if

$$\text{for all } t \in \mathbb{R}_+, \{\omega \in \Omega : T(\omega) \leq t\} \in \mathcal{F}_t.$$

Since $\mathcal{F}_t \subseteq \mathcal{F}$ for all $t \in \mathbb{R}_+$, this implies that T is $(\mathcal{F}, \mathcal{B}([0, \infty]))$-measurable.

(b) Let T be a stopping time with respect to $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$. We define

$$\mathcal{F}_T := \{A \in \mathcal{F}_\infty : \text{ for all } t \in \mathbb{R}_+, A \cap \{T \leq t\} \in \mathcal{F}_t\},$$

and

$$\mathcal{F}_{T+} := \{A \in \mathcal{F}_\infty : \text{ for all } t \in \mathbb{R}_+, A \cap \{T \leq t\} \in \mathcal{F}_{t+}\}.$$

Intuitively, $\mathcal{F}_T$ represents the information available up to the random time T while $\mathcal{F}_{T+}$ represents the information available immediately after time T.

Having introduced stopping times and the associated σ-algebras $\mathcal{F}_T$ and $\mathcal{F}_{T+}$, we verify that these are indeed σ-algebras, and that the "just after T" information contains the "up to T" information.

Lemma 3.29. *The collections $\mathcal{F}_T$ and $\mathcal{F}_{T+}$ are sub-σ-algebras of $\mathcal{F}_\infty$, and satisfy $\mathcal{F}_T \subseteq \mathcal{F}_{T+}$.*

Proof. We first check that $\mathcal{F}_T$ is a sub-σ-algebra of $\mathcal{F}_\infty$. Indeed, for every $t \in \mathbb{R}_+$, we have $\Omega \cap \{T \le t\} = \{T \le t\} \in \mathcal{F}_t$ (since T is an $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$-stopping time), and therefore $\Omega \in \mathcal{F}_T$. If $A \in \mathcal{F}_T$, then for each $t \in \mathbb{R}_+$ we can write $A^c \cap \{T \le t\} = \{T \le t\} \setminus (A \cap \{T \le t\})$. Both $\{T \le t\}$ and $A \cap \{T \le t\}$ belong to $\mathcal{F}_t$, so $A^c \cap \{T \le t\} \in \mathcal{F}_t$, which shows that $A^c \in \mathcal{F}_T$. Finally, if $(A_n)_{n\ge1}$ is a countable family of sets in $\mathcal{F}_T$, then for all $t \in \mathbb{R}_+$,

$$\Big(\bigcup_{n\ge1} A_n\Big) \cap \{T \le t\} = \bigcup_{n\ge1} (A_n \cap \{T \le t\})$$

and since each $A_n \cap \{T \le t\} \in \mathcal{F}_t$ and $\mathcal{F}_t$ is a σ-algebra, the right-hand side belongs to $\mathcal{F}_t$. Hence $\bigcup_{n\ge1} A_n \in \mathcal{F}_T$. These three properties show that $\mathcal{F}_T$ is a σ-algebra. Moreover, since by definition every $A \in \mathcal{F}_T$ satisfies $A \in \mathcal{F}_\infty$, it follows that $\mathcal{F}_T$ is a sub-σ-algebra of $\mathcal{F}_\infty$.

To check that $\mathcal{F}_{T+}$ is a sub-σ-algebra of $\mathcal{F}_\infty$, we can apply the same reasoning and use the fact that $\mathcal{F}_t \subseteq \mathcal{F}_{t+}$ for all $t \in \mathbb{R}_+$.

Finally, to show that $\mathcal{F}_T \subseteq \mathcal{F}_{T+}$, consider $A \in \mathcal{F}_T$. By definition, for every $t \in \mathbb{R}_+$, $A \cap \{T \le t\} \in \mathcal{F}_t$. Since $\mathcal{F}_t \subseteq \mathcal{F}_{t+}$ for all t, it follows that $A \cap \{T \le t\} \in \mathcal{F}_{t+}$. Therefore $A \in \mathcal{F}_{T+}$, which completes the proof. □

The following lemma confirms that the definition of $\mathcal{F}_T$ is consistent with its interpretation as the information available up to time T, by showing that T is $\mathcal{F}_T$-measurable.

Lemma 3.30. *If T is an $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$-stopping time, then T is $(\mathcal{F}_T, \mathcal{B}([0,\infty]))$-measurable.*

Proof. It is enough to show that $\{T \le a\} \in \mathcal{F}_T$ for every $a \in [0,\infty)$, since the family $\{(-\infty, a] : a \in [0,\infty)\}$ generates $\mathcal{B}([0,\infty])$. Fix $a \in [0,\infty)$ and $t \in \mathbb{R}_+$. Then $\{T \le a\} \cap \{T \le t\} = \{T \le \min(a,t)\}$. Since T is an $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$-stopping time, we have $\{T \le \min(a,t)\} \in \mathcal{F}_{\min(a,t)} \subseteq \mathcal{F}_t$. Thus, by the definition of $\mathcal{F}_T$, we obtain $\{T \le a\} \in \mathcal{F}_T$. □

Having established the basic properties of the σ-algebras $\mathcal{F}_T$ and $\mathcal{F}_{T+}$, we now examine several fundamental relationships between stopping times and these σ-algebras.

Proposition 3.31. *Let $(\Omega, \mathcal{F})$ be a measurable space and $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ a filtration on this space. Set $\mathcal{F}_\infty := \sigma(\mathcal{F}_t; t \in \mathbb{R}_+)$ and recall the notation $\mathcal{F}_{t+} := \bigcap_{h>0} \mathcal{F}_{t+h}$ for $t \in \mathbb{R}_+$, which defines the right-continuous filtration associated with $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$. Let $S, T : \Omega \to [0,\infty]$ be two functions. Then the following statements hold:*

(i) T is an $(\mathcal{F}_{t+})_{t\in\mathbb{R}_+}$-stopping time if and only if $\{T < t\} \in \mathcal{F}_t$ for every $t \in \mathbb{R}_+$.

(ii) Suppose T is an $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$-stopping time. Then

$$\mathcal{F}_{T+} = \{A \in \mathcal{F}_\infty : \text{ for all } t \in \mathbb{R}_+,\ A \cap \{T < t\} \in \mathcal{F}_t\}.$$

(iii) Suppose S and T are two $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$-stopping times. Then $S \wedge T = \min(S,T)$ and $S \vee T = \max(S,T)$ are two $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$-stopping times and $\mathcal{F}_{S\wedge T} = \mathcal{F}_S \cap \mathcal{F}_T$. Consequently, if $S(\omega) \le T(\omega)$ for all $\omega \in \Omega$, then $\mathcal{F}_S \subseteq \mathcal{F}_T$.

Let $(S_n)_{n\ge1}$ be a sequence of stopping times with respect to $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$.

(iv) $S^* = \sup_{n\ge1} S_n$ is an $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$-stopping time.

(v) $S_* = \inf_{n\ge1} S_n$ is an $(\mathcal{F}_{t+})_{t\in\mathbb{R}_+}$-stopping time and

$$\mathcal{F}_{S_*+} = \bigcap_{n\ge1} \mathcal{F}_{S_n+}.$$

Proof. (i) Suppose first that T is an $(\mathcal{F}_{t+})_{t\in\mathbb{R}_+}$-stopping time. Fix $t > 0$ and a strictly increasing sequence $(t_n)_{n\in\mathbb{N}}$ with $t_n < t$ and $\lim_{n\to\infty} t_n = t$.

Note that if $A \in \mathcal{F}_{t_n+}$, then by definition $A \in \mathcal{F}_{t_n+h}$ for all $h > 0$. Since $t_n < t$, we can choose $h \in (0, t - t_n]$, which gives $t_n + h \le t$ and hence $\mathcal{F}_{t_n+h} \subseteq \mathcal{F}_t$ by monotonicity of the filtration. Therefore $A \in \mathcal{F}_t$, proving $\mathcal{F}_{t_n+} \subseteq \mathcal{F}_t$.

Consequently, $\{T \le t_n\} \in \mathcal{F}_{t_n+} \subseteq \mathcal{F}_t$, so

$$\{T < t\} = \bigcup_{n\in\mathbb{N}} \{T \le t_n\} \in \mathcal{F}_t.$$

Also, $\{T < 0\} = \emptyset = \Omega^c \in \mathcal{F}_0$. Hence, if T is an $(\mathcal{F}_{t+})_{t\in\mathbb{R}_+}$-stopping time, then $\{T < t\} \in \mathcal{F}_t$ for all $t \in \mathbb{R}_+$.

Conversely, assume $\{T < t\} \in \mathcal{F}_t$ for all $t \in \mathbb{R}_+$. Fix $t \in \mathbb{R}_+$ and a strictly decreasing sequence $(t_n)_{n\in\mathbb{N}}$ with $t_n > t$ and $\lim_{n\to\infty} t_n = t$. Fix $h > 0$. For all $n \in \mathbb{N}$ such that $t_n \le t + h$, we have $\{T < t_n\} \in \mathcal{F}_{t_n} \subseteq \mathcal{F}_{t+h}$. Thus, for all $h > 0$,

$$\{T \le t\} = \bigcap_{n\in\mathbb{N}} \{T < t_n\} \in \mathcal{F}_{t+h}.$$

Therefore, $\{T \le t\} \in \bigcap_{h>0} \mathcal{F}_{t+h} = \mathcal{F}_{t+}$, which proves (i).

(ii) The proof is analogous to (i), using the same argument with the increasing and decreasing sequences $(t_n)_{n\in\mathbb{N}}$.

(iii) It is clear that for every $t \in \mathbb{R}_+$, $\{S \wedge T > t\} = \{S > t\} \cap \{T > t\} \in \mathcal{F}_t$, since the complements $\{S \le t\}$ and $\{T \le t\}$ are in $\mathcal{F}_t$. Taking complements, $\{S \wedge T \le t\} \in \mathcal{F}_t$ for all $t \in \mathbb{R}_+$, so $S \wedge T$ is a $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$-stopping time. The same reasoning applies to $S \vee T$.

Now we show that $\mathcal{F}_{S\wedge T} = \mathcal{F}_S \cap \mathcal{F}_T$. For the first inclusion, let $A \in \mathcal{F}_{S\wedge T}$. Then $A \cap \{S \wedge T \le t\} \in \mathcal{F}_t$. Since $\{T \le t\} \subseteq \{S \wedge T \le t\}$, we note that $A \cap \{T \le t\} = A \cap \{S \wedge T \le t\} \cap \{T \le t\}$ which belongs to $\mathcal{F}_t$ since T is a $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$-stopping time. This implies $A \in \mathcal{F}_T$. Similarly $A \in \mathcal{F}_S$, so $\mathcal{F}_{S\wedge T} \subseteq \mathcal{F}_S \cap \mathcal{F}_T$.

Conversely, if $A \in \mathcal{F}_S \cap \mathcal{F}_T$, then for all $t \in \mathbb{R}_+$, $A \cap \{S \le t\} \in \mathcal{F}_t$ and $A \cap \{T \le t\} \in \mathcal{F}_t$. Thus $A \cap \{S \wedge T \le t\} = (A \cap \{S \le t\}) \cup (A \cap \{T \le t\}) \in \mathcal{F}_t$. Hence $A \in \mathcal{F}_{S\wedge T}$, proving $\mathcal{F}_{S\wedge T} = \mathcal{F}_S \cap \mathcal{F}_T$.

Now assume that for every $\omega \in \Omega$, we have $S(\omega) \le T(\omega)$. If $A \in \mathcal{F}_S$, then for all $t \in \mathbb{R}_+$, $A \cap \{S \le t\} \in \mathcal{F}_t$. Since $\{T \le t\} \subseteq \{S \le t\}$, we obtain $A \cap \{T \le t\} = A \cap \{S \le t\} \cap \{T \le t\} \in \mathcal{F}_t$. Therefore $A \in \mathcal{F}_T$, which shows that $\mathcal{F}_S \subseteq \mathcal{F}_T$.

(iv) For all $t \in \mathbb{R}_+$,

$$\{S^* \le t\} = \bigcap_{n\ge1} \{S_n \le t\} \in \mathcal{F}_t$$

which shows that $S^* = \sup_{n\ge1} S_n$ is an $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$-stopping time.

(v) For all $t \in \mathbb{R}_+$ and each $n \ge 1$, since $\mathbb{Q}$ is dense in $\mathbb{R}$, $\{S_n < t\} = \bigcup_{q\in\mathbb{Q},q<t} \{S_n \le q\}$. Since each $\{S_n \le q\} \in \mathcal{F}_q \subseteq \mathcal{F}_t$, we have $\{S_n < t\} \in \mathcal{F}_t$. Therefore,

$$\{S_* < t\} = \bigcup_{n\ge1} \{S_n < t\} \in \mathcal{F}_t$$

and by (i), S_* is an $(\mathcal{F}_{t+})_{t\in\mathbb{R}_+}$-stopping time.

Finally, let $A \in \mathcal{F}_{S_*+}$. By definition, for all $t \in \mathbb{R}_+$, $A \cap \{S_* \le t\} \in \mathcal{F}_{t+}$. For each $n \ge 1$, since $\{S_n \le t\} \subseteq \{S_* \le t\}$, we have $A \cap \{S_n \le t\} = A \cap \{S_* \le t\} \cap \{S_n \le t\}$ which belongs to $\mathcal{F}_{t+}$ since $\{S_n \le t\} \in \mathcal{F}_t \subseteq \mathcal{F}_{t+}$. This implies $A \in \mathcal{F}_{S_n+}$ for all $n \ge 1$. Hence $\mathcal{F}_{S_*+} \subseteq \bigcap_{n\ge1} \mathcal{F}_{S_n+}$.

Conversely, if $A \in \bigcap_{n\ge1} \mathcal{F}_{S_n+}$, then by (ii) for all $n \ge 1$ and all $t \in \mathbb{R}_+$, $A \cap \{S_n < t\} \in \mathcal{F}_t$. Thus

$$A \cap \{S_* < t\} = \bigcup_{n\ge1} (A \cap \{S_n < t\}) \in \mathcal{F}_t \subseteq \mathcal{F}_{t+}.$$

Applying (ii) for the filtration $(\mathcal{F}_{t+})_{t\in\mathbb{R}_+}$ implies $A \in \mathcal{F}_{S_*++}$. From Lemma 3.6 we know that $\mathcal{F}_{S_*++} = \mathcal{F}_{S_*+}$, hence $A \in \mathcal{F}_{S_*+}$ and $\bigcap_{n\ge1} \mathcal{F}_{S_n+} \subseteq \mathcal{F}_{S_*+}$. Therefore $\mathcal{F}_{S_*+} = \bigcap_{n\ge1} \mathcal{F}_{S_n+}$. □

We next show that several natural random times associated with a right-continuous and adapted process, such as jump times and hitting times, are indeed stopping times.

Proposition 3.32. *Let E be a countable space equipped with the discrete topology, and hence endowed with the σ-algebra $\mathcal{P}(E)$. Let $(\Omega, \mathcal{F})$ be a measurable space and $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ a filtration on this space. For each $t \in \mathbb{R}_+$, consider a random variable $X_t : \Omega \to E$ that is $\mathcal{F}_t$-measurable. In other words, the process $\mathbf{X} = (X_t)_{t\in\mathbb{R}_+}$ is $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$-adapted. Assume moreover that $\mathbf{X}$ is right-continuous everywhere on Ω. Then the following assertions hold:*

(i) The first jump time $J_1(\mathbf{X})$ *of* $\mathbf{X}$ *is an* $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$*-stopping time. More generally, for every* $n \in \mathbb{N}$*, the* n*-th jump time* $J_n(\mathbf{X})$ *and the explosion time* $\zeta(\mathbf{X})$ *are also* $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$*-stopping times.*

(ii) Fix $i \in E$ *and define*

$$T_i(\mathbf{X}) := \inf\{t \in \mathbb{R}_+ : X_t = i\}, \quad T_i^+(\mathbf{X}) := \inf\{t \geq J_1(\mathbf{X}) : X_t = i\}$$

with the convention that $\inf\emptyset := \infty$*. Then* $T_i(\mathbf{X})$ *and* $T_i^+(\mathbf{X})$ *are both* $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$*-stopping times, where* $T_i(\mathbf{X})$ *denotes the* first hitting time of i *and* $T_i^+(\mathbf{X})$ *the* first return time to i*.*

(iii) Let $K \subseteq E$*. Define*

$$T_K(\mathbf{X}) := \inf\{t \in \mathbb{R}_+ : X_t \in K\}$$

again with $\inf\emptyset := \infty$*. Then* $T_K(\mathbf{X})$ *is an* $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$*-stopping time and represents the* first hitting time of the set K*.*

Proof. (i) For every $t \in \mathbb{R}_+$, we have $\{J_1(\mathbf{X}) > t\} = \{\text{for all } s \in [0,t], X_s = X_0\}$. Because $\mathbf{X}$ is right-continuous, checking equality for all real $s \in [0,t]$ is equivalent to checking it for all rational $r \in ([0,t] \cap \mathbb{Q}) \cup \{t\}$. Indeed, if $X_s = X_0$ for all $s \in [0,t]$, then in particular $X_r = X_0$ for all such rational r. Conversely, if $X_r = X_0$ for all rational $r \in ([0,t] \cap \mathbb{Q}) \cup \{t\}$, then for any $s \in [0,t]$ we can choose a sequence of rationals $(r_n)_{n\geq 1}$ with $r_n \downarrow s$, and by right-continuity we have $X_s = \lim_{n\to\infty} X_{r_n} = X_0$. Hence,

$$\{J_1(\mathbf{X}) > t\} = \{\text{for all } r \in ([0,t] \cap \mathbb{Q}) \cup \{t\}, X_r = X_0\} = \bigcap_{r\in([0,t]\cap\mathbb{Q})\cup\{t\}} \{X_r = X_0\}.$$

For each r in $([0,t] \cap \mathbb{Q}) \cup \{t\}$, the event $\{X_r = X_0\}$ belongs to $\mathcal{F}_r \subseteq \mathcal{F}_t$. Thus $\{J_1(\mathbf{X}) > t\} \in \mathcal{F}_t$, and so does its complement. This proves that $J_1(\mathbf{X})$ is an $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$-stopping time.

More generally, using again the right-continuity of $\mathbf{X}$, one shows that for every $n \in \mathbb{N}$ and every $t \in \mathbb{R}_+$

$$\begin{aligned}
\{J_n(\mathbf{X}) \leq t\} &= \{\text{there exist } 0 < s_1 < \cdots < s_n \leq t : X_0 \neq X_{s_1}, X_{s_1} \neq X_{s_2}, \ldots, X_{s_{n-1}} \neq X_{s_n}\} \\
&= \bigcup_{\substack{r_1,\ldots,r_n\in([0,t]\cap\mathbb{Q})\cup\{t\} \\ 0<r_1<\cdots<r_n}} \{X_0 \neq X_{r_1}, X_{r_1} \neq X_{r_2}, \ldots, X_{r_{n-1}} \neq X_{r_n}\} \\
&= \bigcup_{\substack{r_1,\ldots,r_n\in([0,t]\cap\mathbb{Q})\cup\{t\} \\ 0<r_1<\cdots<r_n}} \bigcap_{k=1}^{n} \{X_{r_{k-1}} \neq X_{r_k}\}, \quad \text{where } r_0 := 0.
\end{aligned}$$

Each event $\{X_{r_{k-1}} \neq X_{r_k}\}$ belongs to $\mathcal{F}_{r_k} \subseteq \mathcal{F}_t$, so $J_n(\mathbf{X})$ is an $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$-stopping time. Finally, by Proposition 3.31 (iv), $\zeta(\mathbf{X}) = \sup_{n\geq 0} J_n(\mathbf{X})$ is also an $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$-stopping time.

(ii) & (iii) For any $t \in \mathbb{R}_+$, we observe that

$$\{T_K(\mathbf{X}) \leq t\} = \{\text{there exists } s \in [0,t] : X_s \in K\} = \bigcup_{i\in K} \bigcup_{r\in([0,t]\cap\mathbb{Q})\cup\{t\}} \{X_r = i\}.$$

For each rational r in $([0,t] \cap \mathbb{Q}) \cup \{t\}$, the event $\{X_r = i\}$ belongs to $\mathcal{F}_r \subseteq \mathcal{F}_t$, so $T_K(\mathbf{X})$ is an $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$-stopping time. Taking $K = \{i\}$ gives the result for $T_i(\mathbf{X})$.

Next, observe that $\{T_i^+(\mathbf{X}) \leq 0\} = \emptyset \in \mathcal{F}_0$. For any $t > 0$, we have

$$\begin{aligned}
\{T_i^+(\mathbf{X}) \leq t\} &= \{\text{there exist } s' < s \in (0,t] : X_{s'} \neq X_s, X_s = i\} \\
&= \bigcup_{\substack{r,r'\in((0,t]\cap\mathbb{Q})\cup\{t\} \\ r'<r}} \{X_{r'} \neq X_r, X_r = i\}.
\end{aligned}$$

Since for each pair (r', r) the event $\{X_{r'} \neq X_r, X_r = i\}$ belongs to $\mathcal{F}_r \subseteq \mathcal{F}_t$, it follows that $T_i^+(\mathbf{X})$ is an $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$-stopping time. □

Strong Markov property. Before stating the strong Markov property, we first introduce two useful objects. The random variable X_T represents the value of the process at the (possibly random) time T, and the shifted process $\theta_T\mathbf{X}$ describes its evolution after time T.

Let $(\Omega, \mathcal{F})$ be a measurable space and $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ a filtration on it. For each $t \in \mathbb{R}_+$, let $X_t : \Omega \to E$ be $\mathcal{F}_t$-measurable. Fix $i_* \in E$ (which plays no specific role in what follows), and let $T : \Omega \to [0, \infty]$ be given. For each $\omega \in \Omega$, define

$$X_T(\omega) := \begin{cases} X_{T(\omega)}(\omega), & \text{if } T(\omega) < \infty, \\ i_*, & \text{if } T(\omega) = \infty. \end{cases}$$

Let $\mathbf{i}_* : \mathbb{R}_+ \to E$ be the constant function equal to i_*. For each $\omega \in \Omega$, define

$$\theta_T\mathbf{X}(\omega) := \begin{cases} (X_{T(\omega)+t}(\omega))_{t\in\mathbb{R}_+}, & \text{if } T(\omega) < \infty, \\ \mathbf{i}_*, & \text{if } T(\omega) = \infty. \end{cases}$$

We now state the strong Markov property, which states that the simple Markov property continues to hold when the deterministic time parameter is replaced by a stopping time.

Theorem 3.33 (Strong Markov property)**.** *Let E be a countable space equipped with the discrete topology, and hence endowed with the σ-algebra $\mathcal{P}(E)$. Consider a Markov process*

$$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, \mathbf{X} = (X_t)_{t\in\mathbb{R}_+}, (P_t)_{t\in\mathbb{R}_+}, \mathbb{P}_\mu, \mu \in \mathcal{M}_1(E)).$$

Assume $\mathbf{X}$ is right-continuous. Let T be an $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$-stopping time. Then:

(i) The random variable $X_T : \Omega \to E$ is $\mathcal{F}_T$-measurable.

(ii) The random trajectory $\theta_T\mathbf{X} : \Omega \to E^{\mathbb{R}_+}$ is $(\mathcal{F}, \mathcal{P}(E)^{\mathbb{R}_+})$-measurable.

(iii) For every initial law $\mu \in \mathcal{M}_1(E)$ and every bounded $\mathcal{P}(E)^{\mathbb{R}_+}$-measurable function $F : E^{\mathbb{R}_+} \to \mathbb{R}$, we have the following equality, called the strong Markov property *at stopping time T:*

$$\mathbb{E}_\mu[\mathbf{1}_{\{T<\infty\}} F(\theta_T\mathbf{X}) \mid \mathcal{F}_{T+}] = \mathbf{1}_{\{T<\infty\}} \mathbb{E}_{X_T}[F(\mathbf{X})] \quad \mathbb{P}_\mu\text{-a.s.}$$

Proof. (i) We use an approximation of the stopping time T. For any positive real number a, let $\lceil a \rceil$ denote the smallest integer strictly greater than a (that is, $\lceil a \rceil = n$ if $n - 1 \le a < n$), so that $\lceil a \rceil$ is the integer part of a plus one. For each $n \in \mathbb{N}$, define $T_n := 2^{-n}\lceil 2^n T \rceil$ so that $T_n = (k+1)2^{-n}$ if and only if $k2^{-n} \le T < (k+1)2^{-n}$. Also observe that $T_n = \infty$ if and only if $T = \infty$.

We now show that $(T_n)_{n\in\mathbb{N}}$ is a decreasing sequence of stopping times converging to T. Indeed, for all $n \in \mathbb{N}$,

$$T_n \ge T_{n+1} \quad \Longleftrightarrow \quad 2^{-n}\lceil 2^n T \rceil \ge 2^{-(n+1)}\lceil 2^{n+1} T \rceil \quad \Longleftrightarrow \quad 2\lceil 2^n T \rceil \ge \lceil 2^{n+1} T \rceil.$$

Let $x := 2^n T$ and $p := \lceil x \rceil$. Then $p - 1 \le x < p$, hence $2p - 2 \le 2x < 2p$. By definition of the ceiling function, $\lceil 2x \rceil$ is the smallest integer strictly greater than $2x$, so $\lceil 2x \rceil \le 2p = 2\lceil x \rceil$. Substituting back, we obtain $\lceil 2^{n+1} T \rceil \le 2\lceil 2^n T \rceil$, which proves that $T_n \ge T_{n+1}$ for all $n \in \mathbb{N}$. Therefore, $(T_n)_{n\in\mathbb{N}}$ is a decreasing sequence. To show that it converges to T, note that $\lceil 2^n T \rceil - 1 \le 2^n T < \lceil 2^n T \rceil$. Divide by 2^n to obtain $2^{-n}\lceil 2^n T \rceil - 2^{-n} \le T < 2^{-n}\lceil 2^n T \rceil$. That is, $0 < T_n - T \le 2^{-n}$, and hence $T_n \downarrow T$ as $n \to \infty$.

We still need to show that each T_n is an $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$-stopping time. Fix $n \in \mathbb{N}$ and $t \in \mathbb{R}_+$. We have

$$\begin{aligned} \{T_n \le t\} &= \{\text{there exists } 1 \le k \le \lfloor 2^n t \rfloor : \lceil 2^n T \rceil = k\} \\ &= \bigcup_{k=1}^{\lfloor 2^n t \rfloor} \{\lceil 2^n T \rceil = k\} \\ &= \bigcup_{k=1}^{\lfloor 2^n t \rfloor} \{(k-1)2^{-n} \le T < k2^{-n}\}. \end{aligned}$$

For each k, $\{(k-1)2^{-n} \le T < k2^{-n}\} = \{T < k2^{-n}\} \cap \{T \ge (k-1)2^{-n}\}$. Since $\mathbb{Q}$ is dense in $\mathbb{R}$ and T is a stopping time, we have

$$\{T < k2^{-n}\} = \bigcup_{\substack{q < k2^{-n} \\ q \in \mathbb{Q}}} \{T \le q\} \in \mathcal{F}_{k2^{-n}}.$$

Similarly, $\{T \ge (k-1)2^{-n}\} = \{T < (k-1)2^{-n}\}^c \in \mathcal{F}_{(k-1)2^{-n}} \subseteq \mathcal{F}_{k2^{-n}}$. Hence $\{(k-1)2^{-n} \le T < k2^{-n}\} \in \mathcal{F}_{k2^{-n}}$ for each k and since $k2^{-n} \le t$, it follows that $\{(k-1)2^{-n} \le T < k2^{-n}\} \in \mathcal{F}_t$. Thus $\{T_n \le t\} \in \mathcal{F}_t$ and T_n is an $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$-stopping time.

Because $\mathbf{X}$ is right-continuous and E is discrete, each sample path is locally constant from the right. That is, for every $\omega \in \Omega$, there exists $h_T(\omega) > 0$ such that $X_s(\omega) = X_T(\omega)$ for all $s \in [T(\omega), T(\omega) + h_T(\omega)]$. Since $(T_n(\omega))_{n\in\mathbb{N}}$ is a decreasing sequence converging to $T(\omega)$, there exists $n_0(\omega)$ such that $T_n(\omega) \in [T(\omega), T(\omega) + h_T(\omega)]$ for all $n \ge n_0(\omega)$. Hence, for all $n \ge n_0(\omega)$, $X_{T_n}(\omega) = X_T(\omega)$, so the sequence $(X_{T_n}(\omega))_{n\in\mathbb{N}}$ eventually becomes constant at $X_T(\omega)$.

Fix $t \in \mathbb{R}_+$. We observe that

$$\begin{aligned}\{T_n < t\} \cap \{X_{T_n} = i\} &= \{\text{there exists } 1 \le k < 2^n t : X_{k2^{-n}} = i \text{ and } (k-1)2^{-n} \le T < k2^{-n}\} \\ &= \bigcup_{1 \le k < 2^n t} \{X_{k2^{-n}} = i\} \cap \{(k-1)2^{-n} \le T < k2^{-n}\}.\end{aligned}$$

Since $\{(k-1)2^{-n} \le T < k2^{-n}\} \in \mathcal{F}_{k2^{-n}}$ for each k (previous argument) and $k < 2^n t$, it follows that $\{(k-1)2^{-n} \le T < k2^{-n}\} \in \mathcal{F}_t$. Thus $\{T_n < t\} \cap \{X_{T_n} = i\} \in \mathcal{F}_t$.

Next, since $\mathbf{X}$ is right-continuous, we have

$$\begin{aligned}\{T < t\} \cap \{X_T = i\} &= \{\text{for all } p \ge 0, \text{ there exists } n \ge p : T_n < t \text{ and } X_{T_n} = i\} \\ &= \bigcap_{p \ge 0} \bigcup_{n \ge p} \{T_n < t\} \cap \{X_{T_n} = i\}.\end{aligned}$$

From the previous step, we conclude that $\{T < t\} \cap \{X_T = i\} \in \mathcal{F}_t$. Moreover, $\{T = t\} = \{T \le t\} \cap \{T < t\}^c \in \mathcal{F}_t$, hence $\{T = t\} \cap \{X_t = i\} \in \mathcal{F}_t$. Therefore, for all $i \in E$ and $t \in \mathbb{R}_+$, $\{T \le t\} \cap \{X_T = i\} = (\{T < t\} \cap \{X_T = i\}) \cup (\{T = t\} \cap \{X_t = i\}) \in \mathcal{F}_t$. Thus $\{X_T = i\} \in \mathcal{F}_T$ for all $i \in E$, which proves that $X_T : \Omega \to E$ is $\mathcal{F}_T$-measurable.

(ii) Define

$$\Gamma : \Omega \to [0,\infty] \times E^{\mathbb{R}_+}, \quad \Gamma(\omega) := (T(\omega), (X_t(\omega))_{t\in\mathbb{R}_+}).$$

Since T is a stopping time, it is $(\mathcal{F}, \mathcal{B}([0,\infty]))$-measurable, and because X_t is $\mathcal{F}_t$-measurable with $\mathcal{F}_t \subseteq \mathcal{F}$ for each t, the map $\omega \mapsto (X_t(\omega))_{t\in\mathbb{R}_+}$ is $(\mathcal{F}, \mathcal{P}(E)^{\mathbb{R}_+})$-measurable. Hence Γ is $(\mathcal{F}, \mathcal{B}([0,\infty]) \otimes \mathcal{P}(E)^{\mathbb{R}_+})$-measurable. Next, define the (path-shift) map

$$\Phi : [0,\infty] \times E^{\mathbb{R}_+} \to E^{\mathbb{R}_+}, \quad \Phi(s, \mathbf{x}) := \theta_s \mathbf{x} = (x(s+t))_{t\in\mathbb{R}_+}$$

with the usual convention $\theta_\infty \mathbf{x} := \mathbf{i}_*$. By Lemma 3.4, the map Φ is $(\mathcal{B}([0,\infty]) \otimes \mathcal{P}(E)^{\mathbb{R}_+}, \mathcal{P}(E)^{\mathbb{R}_+})$-measurable. Finally, $\theta_T \mathbf{X} = \Phi \circ \Gamma$, so $\theta_T \mathbf{X}$ is $(\mathcal{F}, \mathcal{P}(E)^{\mathbb{R}_+})$-measurable.

(iii) Fix $\mu \in \mathcal{M}_1(E)$. We first establish the strong Markov property for a function of the form $F := \mathbf{1}_C$, where $C \in \mathcal{C}$ is a cylinder set. Then there exist $n \in \mathbb{N}$, $0 \le t_1 < \cdots < t_n$ in $\mathbb{R}_+$ and $i_1, \ldots, i_n \in E$ such that $C = \{\mathbf{x} \in E^{\mathbb{R}_+} : x(t_1) = i_1, \ldots, x(t_n) = i_n\}$.

We have shown above that for every $\omega \in \Omega$, there exists $h_T(\omega) > 0$ such that for all sufficiently large n with $T_n(\omega) \in [T(\omega), T(\omega) + h_T(\omega))$, we have $X_{T_n(\omega)} = X_T(\omega)$. Consequently, any functional that depends on the path $\mathbf{X}$ only through its values on a fixed right-neighborhood of $T(\omega)$ takes the same value when evaluated using $T_n(\omega)$ instead of $T(\omega)$ for all sufficiently large n. In particular, this implies that for all $\omega \in \Omega$

$$\lim_{n\to\infty} \mathbf{1}_{\{T_n(\omega) < \infty\}} \mathbf{1}_C(\theta_{T_n} \mathbf{X}(\omega)) = \mathbf{1}_{\{T(\omega) < \infty\}} \mathbf{1}_C(\theta_T \mathbf{X}(\omega)) \tag{3.3}$$

and

$$\lim_{n\to\infty} \mathbf{1}_{\{T_n(\omega) < \infty\}} \mathbb{P}_{X_{T_n(\omega)}}(\mathbf{X} \in C) = \mathbf{1}_{\{T(\omega) < \infty\}} \mathbb{P}_{X_T(\omega)}(\mathbf{X} \in C). \tag{3.4}$$

Fix $A \in \mathcal{F}_{T+}$. Combining (3.3) and (3.4) with the dominated convergence theorem yields

$$\lim_{n\to\infty} \mathbb{E}_\mu[\mathbf{1}_A \mathbf{1}_{\{T_n<\infty\}} \mathbf{1}_C(\theta_{T_n}\mathbf{X})] = \mathbb{E}_\mu[\mathbf{1}_A \mathbf{1}_{\{T<\infty\}} \mathbf{1}_C(\theta_T \mathbf{X})] \tag{3.5}$$

and

$$\lim_{n\to\infty} \mathbb{E}_\mu[\mathbf{1}_A \mathbf{1}_{\{T_n<\infty\}} \mathbb{P}_{X_{T_n}}(\mathbf{X} \in C)] = \mathbb{E}_\mu[\mathbf{1}_A \mathbf{1}_{\{T<\infty\}} \mathbb{P}_{X_T}(\mathbf{X} \in C)] \tag{3.6}$$

since both $\mathbf{1}_A \mathbf{1}_{\{T_n<\infty\}} \mathbf{1}_C(\theta_{T_n}\mathbf{X})$ and $\mathbf{1}_A \mathbf{1}_{\{T_n<\infty\}} \mathbb{P}_{X_{T_n}}(\mathbf{X} \in C)$ are bounded by 1. For each $n \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{E}_\mu[\mathbf{1}_A \mathbf{1}_{\{T_n<\infty\}} \mathbf{1}_C(\theta_{T_n}\mathbf{X})] &= \mathbb{E}_\mu[\mathbf{1}_A \mathbf{1}_{\{\exists!\, k\geq 1:\ (k-1)2^{-n} \leq T < k2^{-n}\}} \mathbf{1}_C(\theta_{T_n}\mathbf{X})] \\
&= \mathbb{E}_\mu[\mathbf{1}_A \mathbf{1}_{\bigsqcup_{k\geq 1}\{(k-1)2^{-n} \leq T < k2^{-n}\}} \mathbf{1}_C(\theta_{T_n}\mathbf{X})] \\
&= \mathbb{E}_\mu[\mathbf{1}_A \sum_{k\geq 1} \mathbf{1}_{\{(k-1)2^{-n} \leq T < k2^{-n}\}} \mathbf{1}_C(\theta_{k2^{-n}}\mathbf{X})] \\
&= \sum_{k\geq 1} \mathbb{E}_\mu[\mathbf{1}_{A\cap\{(k-1)2^{-n} \leq T < k2^{-n}\}} \mathbf{1}_C(\theta_{k2^{-n}}\mathbf{X})].
\end{aligned}$$

Since $A \in \mathcal{F}_{T+}$, by Proposition 3.31 (ii), the event

$$A \cap \{(k-1)2^{-n} \leq T < k2^{-n}\} = (A \cap \{T < k2^{-n}\}) \cap (A \cap \{T < (k-1)2^{-n}\})^c$$

belongs to $\mathcal{F}_{k2^{-n}}$. Therefore, using the weak tower property, we obtain

$$\mathbb{E}_\mu[\mathbf{1}_{A\cap\{(k-1)2^{-n} \leq T < k2^{-n}\}} \mathbf{1}_C(\theta_{k2^{-n}}\mathbf{X})] = \mathbb{E}_\mu[\mathbf{1}_{A\cap\{(k-1)2^{-n} \leq T < k2^{-n}\}} \mathbb{E}_\mu[\mathbf{1}_C(\theta_{k2^{-n}}\mathbf{X}) \mid \mathcal{F}_{k2^{-n}}]].$$

Because $\mathbf{1}_C$ is bounded and measurable ($C \in \mathcal{C} \subseteq \sigma(\mathcal{C}) = \mathcal{P}(E)^{\mathbb{R}_+}$), the simple Markov property applied at time $k2^{-n}$ gives

$$\mathbb{E}_\mu[\mathbf{1}_C(\theta_{k2^{-n}}\mathbf{X}) \mid \mathcal{F}_{k2^{-n}}] = \mathbb{E}_{X_{k2^{-n}}}[\mathbf{1}_C(\mathbf{X})] = \mathbb{P}_{X_{k2^{-n}}}(\mathbf{X} \in C) \quad \mathbb{P}_\mu\text{-a.s.}$$

Hence

$$\mathbb{E}_\mu[\mathbf{1}_{A\cap\{(k-1)2^{-n} \leq T < k2^{-n}\}} \mathbf{1}_C(\theta_{k2^{-n}}\mathbf{X})] = \mathbb{E}_\mu[\mathbf{1}_{A\cap\{(k-1)2^{-n} \leq T < k2^{-n}\}} \mathbb{P}_{X_{k2^{-n}}}(\mathbf{X} \in C)].$$

It follows that

$$\begin{aligned}
\mathbb{E}_\mu[\mathbf{1}_A \mathbf{1}_{\{T_n<\infty\}} \mathbf{1}_C(\theta_{T_n}\mathbf{X})] &= \sum_{k\geq 1} \mathbb{E}_\mu[\mathbf{1}_{A\cap\{(k-1)2^{-n} \leq T < k2^{-n}\}} \mathbb{P}_{X_{k2^{-n}}}(\mathbf{X} \in C)] \\
&= \mathbb{E}_\mu[\mathbf{1}_A \mathbf{1}_{\{T_n<\infty\}} \mathbb{P}_{X_{T_n}}(\mathbf{X} \in C)].
\end{aligned}$$

Then, by (3.5) and (3.6), taking limits on both sides of the last display, we obtain for all $A \in \mathcal{F}_{T+}$

$$\mathbb{E}_\mu[\mathbf{1}_A \mathbf{1}_{\{T<\infty\}} \mathbf{1}_C(\theta_T \mathbf{X})] = \mathbb{E}_\mu[\mathbf{1}_A \mathbf{1}_{\{T<\infty\}} \mathbb{P}_{X_T}(\mathbf{X} \in C)].$$

Observe that this identity is one of the two characterizing properties of the conditional expectation $\mathbb{E}_\mu[\mathbf{1}_{\{T<\infty\}} \mathbf{1}_C(\theta_T \mathbf{X}) \mid \mathcal{F}_{T+}]$. To verify the second characterizing property, which is measurability, note that $\mathbb{P}_{X_T}(\mathbf{X} \in C) = \mathbb{E}_{X_T}[\mathbf{1}_C(\mathbf{X})]$ is $\mathcal{F}_T$-measurable since X_T is $\mathcal{F}_T$-measurable and $\mathcal{F}_T \subseteq \mathcal{F}_{T+}$, while the mapping $x \mapsto \mathbb{E}_x[\mathbf{1}_C(\mathbf{X})]$ is deterministic. Hence, we deduce that

$$\mathbb{E}_\mu[\mathbf{1}_{\{T<\infty\}} \mathbf{1}_C(\theta_T \mathbf{X}) \mid \mathcal{F}_{T+}] = \mathbf{1}_{\{T<\infty\}} \mathbb{E}_{X_T}[\mathbf{1}_C(\mathbf{X})] \quad \mathbb{P}_\mu\text{-a.s.}$$

which proves the strong Markov property for all indicator functions $F = \mathbf{1}_C$ with $C \in \mathcal{C}$.

Next, we show that the strong Markov property holds for all $F = \mathbf{1}_A$ where $A \in \mathcal{P}(E)^{\mathbb{R}_+}$. Let $\mathcal{D}$ denote the collection of sets $A \in \mathcal{P}(E)^{\mathbb{R}_+}$ for which the strong Markov property holds with $F = \mathbf{1}_A$. By what we showed above, $\mathcal{C} \subseteq \mathcal{D}$. We claim that $\mathcal{D}$ is a monotone class. Indeed, let $(A_n)_{n\geq 1} \subseteq \mathcal{D}$. If $A_n \uparrow A := \cup_{n\geq 1} A_n$, then $\mathbf{1}_{A_n} \uparrow \mathbf{1}_A$, and by the monotone convergence theorem we obtain

$$\mathbb{E}_\mu[\mathbf{1}_{\{T<\infty\}} \mathbf{1}_{A_n}(\theta_T \mathbf{X}) \mid \mathcal{F}_{T+}] \uparrow \mathbb{E}_\mu[\mathbf{1}_{\{T<\infty\}} \mathbf{1}_A(\theta_T \mathbf{X}) \mid \mathcal{F}_{T+}]$$

and

$$\mathbb{E}_{X_T}[\mathbf{1}_{A_n}(\mathbf{X})] \uparrow \mathbb{E}_{X_T}[\mathbf{1}_A(\mathbf{X})].$$

Therefore, passing to the limit in the strong Markov property for $F = \mathbf{1}_{A_n}$ yields the strong Markov property for $F = \mathbf{1}_A$, so $A \in \mathcal{D}$. If $A_n \downarrow A := \cap_{n\geq 1} A_n$, a similar reasoning applies, using the dominated convergence theorem (since $0 \leq \mathbf{1}_{A_n} \leq 1$), and again we obtain $A \in \mathcal{D}$. Hence $\mathcal{D}$ is a monotone class containing the π-system $\mathcal{C}$, which generates $\mathcal{P}(E)^{\mathbb{R}_+}$. By the monotone class theorem (see Theorem A.1.10 in Section A.1 of [Duq23]), it follows that $\mathcal{D} = \mathcal{P}(E)^{\mathbb{R}_+}$. Thus the strong Markov property holds for all $F = \mathbf{1}_A$ where $A \in \mathcal{P}(E)^{\mathbb{R}_+}$.

Now we prove the strong Markov property for simple functions $F := \sum_{k=1}^m a_k \mathbf{1}_{A_k}$ with $a_k \in \mathbb{R}$ and $A_k \in \mathcal{P}(E)^{\mathbb{R}_+}$. Since we just showed that the strong Markov property holds for all indicator functions, linearity of conditional expectation gives $\mathbb{P}_\mu$-a.s.

$$\mathbb{E}_\mu[\mathbf{1}_{\{T<\infty\}} F(\theta_T \mathbf{X}) \mid \mathcal{F}_{T+}] = \sum_{k=1}^m a_k \mathbb{E}_\mu[\mathbf{1}_{\{T<\infty\}} \mathbf{1}_{A_k}(\theta_T \mathbf{X}) \mid \mathcal{F}_{T+}] = \mathbf{1}_{\{T<\infty\}} \mathbb{E}_{X_T}[F(\mathbf{X})].$$

Finally, we extend this result to general bounded $\mathcal{P}(E)^{\mathbb{R}_+}$-measurable functions $F : E^{\mathbb{R}_+} \to \mathbb{R}$. By Lemma 2.2, there exists a sequence of simple $\sigma(\mathcal{C})$-measurable functions $(F_n)_{n\geq 1}$ of the form $F_n = \sum_{k=1}^{m_n} a_{k,n} \mathbf{1}_{A_{k,n}}$ with $a_{k,n} \in \mathbb{R}$, $A_{k,n} \in \mathcal{P}(E)^{\mathbb{R}_+}$ such that $|F_n| \leq \|F\|_\infty$ and $F_n \to F$ pointwise on $E^{\mathbb{R}_+}$. Then by the previous step, we obtain

$$\mathbb{E}_\mu[\mathbf{1}_{\{T<\infty\}} F_n(\theta_T \mathbf{X}) \mid \mathcal{F}_{T+}] = \mathbf{1}_{\{T<\infty\}} \mathbb{E}_{X_T}[F_n(\mathbf{X})] \quad \mathbb{P}_\mu\text{-a.s.}$$

and by the dominated convergence theorem

$$\mathbb{E}_\mu[\mathbf{1}_{\{T<\infty\}} F(\theta_T \mathbf{X}) \mid \mathcal{F}_{T+}] = \mathbf{1}_{\{T<\infty\}} \mathbb{E}_{X_T}[F(\mathbf{X})] \quad \mathbb{P}_\mu\text{-a.s.}$$

This completes the proof of the strong Markov property. □

Lemma 3.34. *Under the assumptions of Theorem 3.33, the strong Markov property also holds when conditioning with respect to $\mathcal{F}_T$. For every initial law $\mu \in \mathcal{M}_1(E)$ and every bounded $\mathcal{P}(E)^{\mathbb{R}_+}$-measurable function $F : E^{\mathbb{R}_+} \to \mathbb{R}$,*

$$\mathbb{E}_\mu[\mathbf{1}_{\{T<\infty\}} F(\theta_T \mathbf{X}) \mid \mathcal{F}_T] = \mathbf{1}_{\{T<\infty\}} \mathbb{E}_{X_T}[F(\mathbf{X})] \quad \mathbb{P}_\mu\text{-a.s.}$$

Proof. First, the random variable $\mathbf{1}_{\{T<\infty\}} \mathbb{E}_{X_T}[F(\mathbf{X})]$ is $\mathcal{F}_T$-measurable. Indeed:

- $\mathbf{1}_{\{T<\infty\}}$ is $\mathcal{F}_T$-measurable since $\{T < \infty\} \cap \{T \leq t\} = \{T \leq t\} \in \mathcal{F}_t$;
- the map $\omega \mapsto X_T(\omega)$ is $\mathcal{F}_T$-measurable by part (i) of Theorem 3.33;
- the function $x \mapsto \mathbb{E}_x[F(\mathbf{X})]$ is deterministic and measurable from E to $\mathbb{R}$;
- hence the composition $\omega \mapsto \mathbb{E}_{X_T(\omega)}[F(\mathbf{X})]$ is $\mathcal{F}_T$-measurable.

Next, we take conditional expectations with respect to $\mathcal{F}_T$ on both sides of the equality in Theorem 3.33:

$$\mathbb{E}_\mu[\mathbb{E}_\mu[\mathbf{1}_{\{T<\infty\}} F(\theta_T \mathbf{X}) \mid \mathcal{F}_{T+}] \mid \mathcal{F}_T] = \mathbb{E}_\mu[\mathbf{1}_{\{T<\infty\}} \mathbb{E}_{X_T}[F(\mathbf{X})] \mid \mathcal{F}_T] \quad \mathbb{P}_\mu\text{-a.s.}$$

Since $\mathbf{1}_{\{T<\infty\}} \mathbb{E}_{X_T}[F(\mathbf{X})]$ is $\mathcal{F}_T$-measurable, the conditional expectation on the right-hand side reduces to itself. Using the tower property and the inclusion $\mathcal{F}_T \subseteq \mathcal{F}_{T+}$, we obtain

$$\mathbb{E}_\mu[\mathbf{1}_{\{T<\infty\}} F(\theta_T \mathbf{X}) \mid \mathcal{F}_T] = \mathbf{1}_{\{T<\infty\}} \mathbb{E}_{X_T}[F(\mathbf{X})] \quad \mathbb{P}_\mu\text{-a.s.}$$

□

3.2 Structure of Markov Processes and Infinitesimal Generators

3.2.1 Process Decomposition and the Infinitesimal Generator

We recall the notations $J_n(\mathbf{x})$, $\zeta(\mathbf{x})$, $N(\mathbf{x})$, $Y_n(\mathbf{x})$ and $D_n(\mathbf{x})$, defined for a right-continuous trajectory $\mathbf{x}$ taking values in a countable space equipped with the discrete topology. We adopt the following conventions for exponential distributions: a random variable that is almost surely zero is said to follow an exponential distribution with infinite parameter, and a random variable that is almost surely infinite is said to follow an exponential distribution with zero parameter. We then speak of an *extended exponential variable*, which means that its parameter varies in $[0, \infty]$. In contrast, the usual exponential variable admits only strictly positive and finite parameters.

Before proceeding further, we recall the memoryless property of the exponential distribution and formalize its extension to the degenerate cases $\lambda = 0$ and $\lambda = \infty$.

Lemma 3.35. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathbf{e} : \Omega \to [0, \infty]$ be an $\mathcal{F}$-measurable random variable. Then the following are equivalent:*

1. *For all $s, t \in \mathbb{R}_+$, $\mathbb{P}(\mathbf{e} > s + t) = \mathbb{P}(\mathbf{e} > s)\mathbb{P}(\mathbf{e} > t)$. Observe that this identity expresses the memoryless property $\mathbb{P}(\mathbf{e} > s + t \mid \mathbf{e} > s) = \mathbb{P}(\mathbf{e} > t)$ for all $s, t \in \mathbb{R}_+$.*

2. *The random variable $\mathbf{e}$ follows an extended exponential distribution, i.e. there exists $\lambda \in [0, \infty]$ such that for all $t \in \mathbb{R}_+$, $\mathbb{P}(\mathbf{e} > t) = e^{-\lambda t}$ with the following conventions:*
   - *when $\lambda = 0$, we have $e^{-0 \cdot t} := 1$ for all $t \in \mathbb{R}_+$, hence $\mathbb{P}(\mathbf{e} > t) = 1$ and $\mathbf{e} = \infty$ almost surely;*
   - *when $\lambda = \infty$, we have $e^{-\infty \cdot t} := 0$ for all $t > 0$, hence $\mathbb{P}(\mathbf{e} > t) = 0$ and $\mathbf{e} = 0$ almost surely.*

   In particular, for $\lambda \in (0, \infty)$, the variable $\mathbf{e}$ follows the usual exponential distribution $\mathrm{Exp}(\lambda)$.

Proof. (2) $\Rightarrow$ (1) is immediate, since $e^{-\lambda(s+t)} = e^{-\lambda s} e^{-\lambda t}$.

For (1) $\Rightarrow$ (2), write $S(t) := \mathbb{P}(\mathbf{e} > t)$. Then $S : \mathbb{R}_+ \to [0, 1]$ satisfies $S(s + t) = S(s)S(t)$ for all $s, t \in \mathbb{R}_+$ and is nonincreasing with $S(0) = 1$ (standard property of survival function).

Degenerate cases. Assume that there exists $t_0 > 0$ with $S(t_0) = 0$. By the multiplicative identity, $S(nt) = S(t)^n$ for all $n \in \mathbb{N}$ and $t \in \mathbb{R}_+$. Taking $t = t_0/n$ gives $0 = S(t_0) = S(t_0/n)^n$, hence $S(t_0/n) = 0$ for every $n \in \mathbb{N}$. Now fix any $t > 0$ and choose n large enough so that $t_0/n \le t$. Since S is nonincreasing, $S(t) \le S(t_0/n) = 0$ and thus $S(t) = 0$ for all $t > 0$. This corresponds to the case $\mathbf{e} = 0$ a.s., i.e. $\lambda = \infty$.

Now assume that there exists $t_0 > 0$ such that $S(t_0) = 1$. Then, by the multiplicative identity, $S(t_0 + t) = S(t_0)S(t) = S(t)$ for all $t \in \mathbb{R}_+$, so S is periodic with period t_0. Since S is nonincreasing, it follows that $S(t) = 1$ for all $t \in \mathbb{R}_+$, which means that $\mathbf{e} = \infty$ a.s., i.e. $\lambda = 0$.

Nondegenerate case. Assume $S(t) \in (0, 1)$ for all $t > 0$. Define $\phi(t) := -\log S(t) \in (0, \infty)$ for $t > 0$ and $\phi(0) := 0$. Then the multiplicative identity becomes the additive identity $\phi(s + t) = \phi(s) + \phi(t)$ for all $s, t \in \mathbb{R}_+$. Moreover, ϕ is nondecreasing (since S is nonincreasing). Let $\lambda := \phi(1)$. By additivity, $\phi(n) = n\lambda$ for all integers $n \ge 1$. For any integers $m, n \ge 1$, we have

$$n\phi(\tfrac{m}{n}) = \phi(\underbrace{\tfrac{m}{n} + \cdots + \tfrac{m}{n}}_{n \text{ times}}) = \phi(m) = m\lambda$$

so $\phi(\frac{m}{n}) = \frac{m}{n}\lambda$ for all rational numbers $\frac{m}{n} \ge 0$ (the case $m = 0$ gives $\phi(0) = 0$). Fix $t \in \mathbb{R}_+$ and take sequences of rationals $q_k \uparrow t$ and $r_k \downarrow t$. By monotonicity of ϕ, $\phi(q_k) \le \phi(t) \le \phi(r_k)$ for all k. Also $\phi(q_k) = \lambda q_k$ and $\phi(r_k) = \lambda r_k$ by what we just showed. Letting $k \to \infty$ yields $\phi(t) = \lambda t$ for all $t \in \mathbb{R}_+$. Consequently, $S(t) = e^{-\phi(t)} = e^{-\lambda t}$ for all $t \in \mathbb{R}_+$, which shows that $\mathbf{e} \sim \mathrm{Exp}(\lambda)$. □

Having identified the random variables with the memoryless property as (extended) exponential, we next see how this shows up in continuous-time Markov processes. The following proposition shows that, under each initial state, the holding time until the first jump of the process is exponentially distributed (possibly in the extended sense) and independent of the next visited state.

Proposition 3.36. *Let E be a countable space equipped with the discrete topology, and hence endowed with the σ-algebra $\mathcal{P}(E)$. Let*

$$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, \mathbf{X} = (X_t)_{t\in\mathbb{R}_+}, (P_t)_{t\in\mathbb{R}_+}, \mathbb{P}_\mu, \mu \in \mathcal{M}_1(E))$$

be a Markov process which is right-continuous. For each $i \in E$, under $\mathbb{P}_i$, the first jump time $J_1(\mathbf{X})$ follows an (extended) exponential law with parameter denoted $q(i) \in \mathbb{R}_+$. Moreover, under $\mathbb{P}_i$, $J_1(\mathbf{X})$ is independent of the position $Y_1(\mathbf{X})$ at the first jump.

Proof. For simplicity, write $J_1 := J_1(\mathbf{X})$. Fix $i \in E$. For all $t, s \in \mathbb{R}_+$, we have

$$\begin{aligned}\mathbb{P}_i(J_1 > s+t) &= \mathbb{P}_i(J_1 > t, X_{u+t} = X_t \text{ for all } u \in [0,s]) \\ &= \mathbb{E}_i[\mathbf{1}_{\{J_1>t\}}\mathbf{1}_{\{(\theta_t\mathbf{X})_u=(\theta_t\mathbf{X})_0 \text{ for all } u\in[0,s]\}}] \\ &= \mathbb{E}_i[\mathbf{1}_{\{J_1>t\}}\mathbf{1}_{\{J_1(\theta_t\mathbf{X})>s\}}].\end{aligned}$$

Observe that by Proposition 3.32, J_1 is a stopping time with respect to $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$, and thus $\{J_1 > t\} = \{J_1 \le t\}^c$ belongs to $\mathcal{F}_t$. Note also that the function $F(\mathbf{x}) := \mathbf{1}_{\{J_1(\mathbf{x})>t\}}$ is bounded and $\mathcal{P}(E)^{\mathbb{R}_+}$-measurable. Indeed,

$$\begin{aligned}\{J_1(\mathbf{x}) > t\} &= \bigcup_{j\in E}\{\mathbf{x} \in E^{\mathbb{R}_+} : x(s) = j \text{ for all } s \in [0,t]\} \\ &= \bigcup_{j\in E}\bigcap_{s\in[0,t]}\{\mathbf{x} \in E^{\mathbb{R}_+} : x(s) = j\}.\end{aligned}$$

Because $\mathbb{Q}$ is dense in $\mathbb{R}$ and $\mathbf{X}$ is right-continuous, we can replace the intersection over $[0,t]$ by a countable intersection over the rationals:

$$\{J_1(\mathbf{x}) > t\} = \bigcup_{j\in E}\bigcap_{r\in\mathbb{Q}\cap[0,t]}\{\mathbf{x} \in E^{\mathbb{R}_+} : x(r) = j\}.$$

Since E and $\mathbb{Q}\cap[0,t]$ are countable and each set $\{x(r) = j\}$ is an elementary cylinder belonging to $\mathcal{P}(E)^{\mathbb{R}_+}$, the event $\{J_1(\mathbf{x}) > t\}$ belongs to $\mathcal{P}(E)^{\mathbb{R}_+}$, which shows that the function $F(\mathbf{x}) = \mathbf{1}_{\{J_1(\mathbf{x})>t\}}$ is $\mathcal{P}(E)^{\mathbb{R}_+}$-measurable. Using what we just showed, by the simple Markov property at time t, we have $\mathbb{P}_i$-a.s.

$$\mathbb{E}_i[\mathbf{1}_{\{J_1>t\}}\mathbf{1}_{\{J_1(\theta_t\mathbf{X})>s\}} \mid \mathcal{F}_t] = \mathbf{1}_{\{J_1>t\}}\mathbb{E}_{X_t}[\mathbf{1}_{\{J_1>s\}}] = \mathbf{1}_{\{J_1>t\}}\mathbb{P}_{X_t}(J_1 > s).$$

Observing that $\mathbb{P}_i$-a.s. $X_t = i$ whenever $J_1 > t$, we obtain $\mathbb{P}_i$-a.s.

$$\mathbb{E}_i[\mathbf{1}_{\{J_1>t\}}\mathbf{1}_{\{J_1(\theta_t\mathbf{X})>s\}} \mid \mathcal{F}_t] = \mathbf{1}_{\{J_1>t\}}\mathbb{P}_i(J_1 > s).$$

Applying $\mathbb{E}_i$ on both sides and using the weak tower property, we get

$$\mathbb{E}_i[\mathbf{1}_{\{J_1>t\}}\mathbf{1}_{\{J_1(\theta_t\mathbf{X})>s\}}] = \mathbb{E}_i[\mathbf{1}_{\{J_1>t\}}]\mathbb{P}_i(J_1 > s)$$

which is equivalent to $\mathbb{P}_i(J_1 > t+s) = \mathbb{P}_i(J_1 > t)\mathbb{P}_i(J_1 > s)$. By Lemma 3.35, we deduce that J_1 follows an extended exponential law under $\mathbb{P}_i$. We denote its parameter by $q(i)$. Since $\mathbf{X}$ is right-continuous, for every $\omega \in \Omega$ there exists $\delta(\omega) > 0$ such that $X_t(\omega) = X_0(\omega)$ for all $t \in [0, \delta(\omega)]$. Hence $\mathbb{P}_i(J_1 > 0) = 1$. From $\mathbb{P}_i(J_1 > t) = e^{-q(i)t}$, we obtain $q(i) < \infty$. This proves the first part of the proposition.

Write $Y_1 := Y_1(\mathbf{X})$. To show independence of J_1 and Y_1 under $\mathbb{P}_i$, fix $t \in \mathbb{R}_+$ and a bounded (measurable) function $g : E \to \mathbb{R}$. Observe that $Y_1 = Y_1(\theta_t\mathbf{X})$ whenever $J_1 > t$. By the weak tower property and the simple Markov property at time t, we get

$$\begin{aligned}\mathbb{E}_i[\mathbf{1}_{\{J_1>t\}}g(Y_1)] &= \mathbb{E}_i[\mathbf{1}_{\{J_1>t\}}g(Y_1(\theta_t\mathbf{X}))] \\ &= \mathbb{E}_i[\mathbf{1}_{\{J_1>t\}}\mathbb{E}_i[g(Y_1(\theta_t\mathbf{X})) \mid \mathcal{F}_t]] \\ &= \mathbb{E}_i[\mathbf{1}_{\{J_1>t\}}\mathbb{E}_{X_t}[g(Y_1)]] \\ &= \mathbb{E}_i[\mathbf{1}_{\{J_1>t\}}]\mathbb{E}_i[g(Y_1)]\end{aligned}$$

where the last step again follows from the fact that $\mathbb{P}_i$-a.s. $X_t = i$ whenever $J_1 > t$.

Next, we show for all $A \in \mathcal{B}(\mathbb{R}_+)$ that

$$\mathbb{E}_i[\mathbf{1}_A(J_1)g(Y_1)] = \mathbb{E}_i[\mathbf{1}_A(J_1)]\mathbb{E}_i[g(Y_1)]. \tag{3.7}$$

Let $\mathcal{D}$ denote the collection of sets $A \in \mathcal{B}(\mathbb{R}_+)$ for which (3.7) holds. By what we showed above, $\{(t,\infty) : t \in \mathbb{R}_+\} \subseteq \mathcal{D}$. We claim that $\mathcal{D}$ is a monotone class. Indeed, let $(A_n)_{n\geq 1} \subseteq \mathcal{D}$. If $A_n \uparrow A := \cup_{n\geq 1} A_n$, then $\mathbf{1}_{A_n} \uparrow \mathbf{1}_A$, and by the monotone convergence theorem we obtain

$$\mathbb{E}_i[\mathbf{1}_{A_n}(J_1)] \uparrow \mathbb{E}_i[\mathbf{1}_A(J_1)].$$

Moreover, by the dominated convergence theorem (since $0 \leq \mathbf{1}_{A_n} \leq 1$ and g is bounded) we have

$$\mathbb{E}_i[\mathbf{1}_{A_n}(J_1)g(Y_1)] \longrightarrow \mathbb{E}_i[\mathbf{1}_A(J_1)g(Y_1)].$$

Therefore, passing to the limit in

$$\mathbb{E}_i[\mathbf{1}_{A_n}(J_1)g(Y_1)] = \mathbb{E}_i[\mathbf{1}_{A_n}(J_1)]\mathbb{E}_i[g(Y_1)]$$

yields (3.7) for A, so $A \in \mathcal{D}$. If $A_n \downarrow A := \cap_{n\geq 1} A_n$, a similar reasoning applies, using the dominated convergence theorem, and again we obtain $A \in \mathcal{D}$. Hence $\mathcal{D}$ is a monotone class containing the π-system $\{(t,\infty) : t \in \mathbb{R}_+\}$, which generates $\mathcal{B}(\mathbb{R}_+)$. By the monotone class theorem (see Theorem A.1.10 in Section A.1 of [Duq23]), it follows that $\mathcal{D} = \mathcal{B}(\mathbb{R}_+)$. Thus (3.7) holds for all $A \in \mathcal{B}(\mathbb{R}_+)$.

Now let $f : \mathbb{R}_+ \to \mathbb{R}$ be a simple function of the form $f := \sum_{k=1}^m a_k \mathbf{1}_{A_k}$ with $a_k \in \mathbb{R}$, $A_k \in \mathcal{B}(\mathbb{R}_+)$. Using what we just showed and linearity of expectation, we have

$$\begin{aligned}
\mathbb{E}_i[f(J_1)g(Y_1)] &= \sum_{k=1}^m a_k \mathbb{E}_i[\mathbf{1}_{A_k}(J_1)g(Y_1)] \\
&= \sum_{k=1}^m a_k \mathbb{E}_i[\mathbf{1}_{A_k}(J_1)]\mathbb{E}_i[g(Y_1)] \\
&= \mathbb{E}_i[f(J_1)]\mathbb{E}_i[g(Y_1)].
\end{aligned}$$

We now extend this to general bounded measurable functions $f : \mathbb{R}_+ \to \mathbb{R}$. By Lemma 2.2, there exists a sequence of simple $\mathcal{B}(\mathbb{R}_+)$-measurable functions $(f_n)_{n\geq 1}$ of the form $f_n = \sum_{k=1}^{m_n} a_{k,n}\mathbf{1}_{A_{k,n}}$ with $a_{k,n} \in \mathbb{R}$, $A_{k,n} \in \mathcal{B}(\mathbb{R}_+)$ such that $|f_n| \leq \|f\|_\infty$ and $f_n \to f$ pointwise on $\mathbb{R}_+$. Then by the previous step, we obtain $\mathbb{E}_i[f_n(J_1)g(Y_1)] = \mathbb{E}_i[f_n(J_1)]\mathbb{E}_i[g(Y_1)]$. Since $|f_n(J_1)g(Y_1)| \leq \|f\|_\infty\|g\|_\infty$ and $|f_n(J_1)| \leq \|f\|_\infty$, by dominated convergence we obtain $\mathbb{E}_i[f(J_1)g(Y_1)] = \mathbb{E}_i[f(J_1)]\mathbb{E}_i[g(Y_1)]$. This implies that J_1 and Y_1 are independent under $\mathbb{P}_i$. □

We now introduce the infinitesimal generator, which combines the rate of leaving each state, $q(i)$, and the probabilities of where to go next. We will see that these two quantities encode all the information characterizing the process up to explosion.

Definition 3.37 (Infinitesimal generator of a right-continuous Markov process)**.** Let

$$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, \mathbf{X} = (X_t)_{t\in\mathbb{R}_+}, (P_t)_{t\in\mathbb{R}_+}, \mathbb{P}_\mu, \mu \in \mathcal{M}_1(E))$$

be a right-continuous Markov process. Proposition 3.36 shows that under $\mathbb{P}_i$, J_1 is an (extended) exponential random variable with parameter $q(i) \in \mathbb{R}_+$. For all $i, j \in E$, define $\pi_{i,j} := \mathbb{P}_i(Y_1 = j)$. Then $Q := (\pi_{i,j})_{i,j\in E}$ is a transition matrix on E, since for each $i \in E$, $\sum_{j\in E} \pi_{i,j} = \sum_{j\in E} \mathbb{P}_i(Y_1 = j) = 1$. There are two possible cases:

- If $q(i) > 0$, then $\mathbb{P}_i(J_1 = \infty) = \lim_{t\to\infty} \mathbb{P}_i(J_1 > t) = \lim_{t\to\infty} e^{-q(i)t} = 0$, so the process cannot stay in i forever and hence $\pi_{i,i} = \mathbb{P}_i(Y_1 = i) = 0$.
- If $q(i) = 0$, then $\mathbb{P}_i(J_1 > t) = e^{-q(i)t} = 1$ for all $t \in \mathbb{R}_+$, so $\mathbb{P}_i(J_1 = \infty) = 1$. Therefore the process never leaves state i and $\pi_{i,i} = \mathbb{P}_i(Y_1 = i) = 1$. Consequently, $\pi_{i,j} = 0$ for all $j \neq i$.

In both cases, $q(i)\pi_{i,i} = 0$. We define the *infinitesimal generator* of the Markov process as the matrix $G = (q_{i,j})_{i,j\in E}$ defined by

$$\forall i, j \in E, \quad q_{i,j} := \begin{cases} q(i)\pi_{i,j} \in \mathbb{R}_+ & \text{if } i \neq j, \\ -q(i) \in \mathbb{R}_- & \text{if } i = j. \end{cases}$$

Intuitively, for $i \neq j$, $q_{i,j} = q(i)\pi_{i,j}$ is the instantaneous rate of jumping from i to j, since the process leaves i at rate $q(i)$ and then moves to j with probability $\pi_{i,j}$. Note that for all $i \in E$,

$$\sum_{j\neq i} q_{i,j} = q(i)\sum_{j\neq i}\pi_{i,j} = q(i)(1-\pi_{i,i}) = q(i) - q(i)\pi_{i,i} = -q_{i,i}.$$

So the diagonal term $q_{i,i} = -q(i)$ ensures that the total rate at which probability mass leaves state i, i.e. the sum of all off-diagonal entries in the i-th row, equals $-q_{i,i}$ (see Theorem 3.14 (ii)).

We now extend Proposition 3.36 on the first jump time to describe the full dynamics of the process. We will see that a right-continuous Markov process on a countable state space consists of two conditionally independent layers of randomness: the independent exponential holding times that determine when the jumps occur, and the discrete-time Markov chain describing the sequence of visited states.

Theorem 3.38. *Let E be a countable space equipped with the discrete topology, and hence endowed with the σ-algebra $\mathcal{P}(E)$. Let*

$$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, \mathbf{X} = (X_t)_{t\in\mathbb{R}_+}, (P_t)_{t\in\mathbb{R}_+}, \mathbb{P}_\mu, \mu \in \mathcal{M}_1(E))$$

be a right-continuous Markov process. For simplicity, write $J_n := J_n(\mathbf{X})$, $Y_n := Y_n(\mathbf{X})$ and $D_n := D_n(\mathbf{X})$ for $n \in \mathbb{N}$.

(i) The discrete-time process

$$(\Omega, \mathcal{F}, (\mathcal{F}_{J_n})_{n\in\mathbb{N}}, \mathbf{Y} = (Y_n)_{n\in\mathbb{N}}, Q = (\pi_{i,j})_{i,j\in E}, \mathbb{P}_\mu, \mu \in \mathcal{M}_1(E))$$

is a Markov chain.

(ii) For any initial distribution $\mu \in \mathcal{M}_1(E)$, under $\mathbb{P}_\mu$, conditionally on $(Y_n)_{n\in\mathbb{N}}$, the holding times $(D_n)_{n\geq 1}$ are mutually independent and each D_n follows an (extended) exponential distribution with parameter $q(Y_{n-1})$.

(iii) For every initial distribution $\mu \in \mathcal{M}_1(E)$ and every $n \geq 1$, under $\mathbb{P}_\mu$, conditionally on Y_{n-1}, the holding time D_n and the next state Y_n are independent.

Proof. (i) & (ii): To simplify notation, assume that $(\Omega, \mathcal{F})$ is sufficiently large to support a sequence of random variables $(\mathbf{e}_n)_{n\geq 1}$ taking values in $\mathbb{R}_+$ such that, for every initial distribution $\mu \in \mathcal{M}_1(E)$, under $\mathbb{P}_\mu$ the random variables $(\mathbf{e}_n)_{n\geq 1}$ are i.i.d. exponentials with parameter 1, independent of $\mathbf{X}$.

Fix $i \in E$ and suppose first that $q(i) > 0$, so that $J_1 < \infty$ $\mathbb{P}_i$-a.s. Let $F : E^{\mathbb{R}_+} \to \mathbb{R}$ be a bounded $\mathcal{P}(E)^{\mathbb{R}_+}$-measurable function. Then the strong Markov property at time J_1 (in fact, Lemma 3.34) implies $\mathbb{P}_i$-a.s. that $\mathbb{E}_i[F(\theta_{J_1}\mathbf{X}) \mid \mathcal{F}_{J_1}] = \mathbb{E}_{Y_1}[F(\mathbf{X})]$ using that $Y_1 = X_{J_1}$. Consequently, for all bounded measurable functions $f : [0,\infty] \to \mathbb{R}$ and $g : E \to \mathbb{R}$, multiplying by $f(J_1)g(Y_1)$ and applying $\mathbb{E}_i$ on both sides, we obtain

$$\mathbb{E}_i[f(J_1)g(Y_1)\mathbb{E}_i[F(\theta_{J_1}\mathbf{X}) \mid \mathcal{F}_{J_1}]] = \mathbb{E}_i[f(J_1)g(Y_1)\mathbb{E}_{Y_1}[F(\mathbf{X})]].$$

By the weak tower property and independence of J_1 and Y_1 (Proposition 3.36), this yields

$$\mathbb{E}_i[f(J_1)g(Y_1)F(\theta_{J_1}\mathbf{X})] = \mathbb{E}_i[f(J_1)]\mathbb{E}_i[g(Y_1)\mathbb{E}_{Y_1}[F(\mathbf{X})]]. \tag{3.8}$$

Note that under $\mathbb{P}_i$, $\frac{\mathbf{e}_1}{q(i)}$ has an exponential distribution with parameter $q(i)$, since

$$\mathbb{E}_i[f(\tfrac{\mathbf{e}_1}{q(i)})] = \int_0^\infty \mathrm{d}x f(\tfrac{x}{q(i)})e^{-x} = \int_0^\infty \mathrm{d}t f(t)q(i)e^{-q(i)t}$$

where we used the change of variable $t := \frac{x}{q(i)}$, so that $\mathrm{d}x = q(i)\mathrm{d}t$. Because J_1 also has an exponential distribution with parameter $q(i)$ under $\mathbb{P}_i$ (Proposition 3.36), it follows that $\mathbb{E}_i[f(J_1)] = \mathbb{E}_i[f(\frac{\mathbf{e}_1}{q(i)})]$ and (3.8) becomes

$$\mathbb{E}_i[f(J_1)g(Y_1)F(\theta_{J_1}\mathbf{X})] = \mathbb{E}_i[f(\tfrac{\mathbf{e}_1}{q(i)})]\mathbb{E}_i[g(Y_1)\mathbb{E}_{Y_1}[F(\mathbf{X})]].$$

Recall that $N := N(\mathbf{X}) = \sup\{n \in \mathbb{N} : J_n < \infty\}$ denotes the number of jumps before absorption. In the case $q(i) > 0$, we have $N \geq 1$ $\mathbb{P}_i$-a.s. and therefore the last display can be rewritten as

$$\mathbb{E}_i[f(J_1)g(Y_1)F(\theta_{J_{1\wedge N}}\mathbf{X})] = \mathbb{E}_i[f(\tfrac{\mathbf{e}_1}{q(i)})]\mathbb{E}_i[g(Y_1)\mathbb{E}_{Y_1}[F(\mathbf{X})]].$$

If $q(i) = 0$, then $\mathbb{P}_i(J_1 = \infty) = \mathbb{P}_i(Y_1 = i) = 1$, so that $N = 0$ $\mathbb{P}_i$-a.s. and $\mathbb{E}_i[f(J_1)g(Y_1)] = f(\infty)g(i)$. Therefore

$$\mathbb{E}_i[f(J_1)g(Y_1)F(\theta_{J_{1\wedge N}}\mathbf{X})] = f(\infty)g(i)\mathbb{E}_i[F(\mathbf{X})] = \mathbb{E}_i[f(\tfrac{\mathbf{e}_1}{q(i)})]\mathbb{E}_i[g(Y_1)\mathbb{E}_{Y_1}[F(\mathbf{X})]]$$

where in the last step we used the convention $\frac{\mathbf{e}_1}{q(i)} := \infty$ for $q(i) = 0$.

In summary, we have

$$\mathbb{E}_i[f(J_1)g(Y_1)F(\theta_{J_{1\wedge N}}\mathbf{X})] = \mathbb{E}_i[f(\tfrac{\mathbf{e}_1}{q(i)})]\mathbb{E}_i[g(Y_1)\mathbb{E}_{Y_1}[F(\mathbf{X})]]$$

for both cases $q(i) > 0$ and $q(i) = 0$. Taking $g = \mathbf{1}_{\{i_1\}}$ for some $i_1 \in E$ and $F \equiv 1$, the last display becomes

$$\mathbb{E}_i[f(D_1)\mathbf{1}_{\{Y_1=i_1\}}] = \mathbb{E}_i[f(\tfrac{\mathbf{e}_1}{q(i)})]\pi_{i,i_1}.$$

Applying this identity recursively, we see that for all $i_1, \ldots, i_n \in E$ and for all bounded measurable functions $f_1, \ldots, f_n : [0, \infty] \to \mathbb{R}$,

$$\mathbb{E}_i[f_1(D_1)\cdots f_n(D_n)\mathbf{1}_{\{Y_1=i_1,\ldots,Y_n=i_n\}}] = \mathbb{E}_i[f_1(\mathbf{e}_1/q(i))\cdots f_n(\mathbf{e}_n/q(i_{n-1}))]\pi_{i,i_1}\cdots\pi_{i_{n-1},i_n}. \quad (3.9)$$

By the weak tower property and the Markov property at time 0 (recalling that $Y_0 = X_{J_0} = X_0$), we have

$$\begin{aligned}\mathbb{E}_\mu[f_1(D_1)\cdots f_n(D_n)\mathbf{1}_{\{Y_0=i_0,\ldots,Y_n=i_n\}}] &= \mathbb{E}_\mu[\mathbf{1}_{\{Y_0=i_0\}}\mathbb{E}_\mu[f_1(D_1)\cdots f_n(D_n)\mathbf{1}_{\{Y_1=i_1,\ldots,Y_n=i_n\}} \mid Y_0]] \\ &= \mathbb{E}_\mu[\mathbf{1}_{\{Y_0=i_0\}}\mathbb{E}_{Y_0}[f_1(D_1)\cdots f_n(D_n)\mathbf{1}_{\{Y_1=i_1,\ldots,Y_n=i_n\}}]] \\ &= \mu(i_0)\mathbb{E}_{i_0}[f_1(D_1)\cdots f_n(D_n)\mathbf{1}_{\{Y_1=i_1,\ldots,Y_n=i_n\}}].\end{aligned}$$

Using (3.9) with initial state i_0, we obtain

$$\mathbb{E}_\mu[f_1(D_1)\cdots f_n(D_n)\mathbf{1}_{\{Y_0=i_0,\ldots,Y_n=i_n\}}] = \mathbb{E}_{i_0}[f_1(\tfrac{\mathbf{e}_1}{q(i_0)})\cdots f_n(\tfrac{\mathbf{e}_n}{q(i_{n-1})})]\mu(i_0)\pi_{i_0,i_1}\cdots\pi_{i_{n-1},i_n}.$$

Since the random variables $(\mathbf{e}_k)_{k\geq 1}$ are independent of the process $\mathbf{X}$, this is equivalently written as

$$\mathbb{E}_\mu[f_1(D_1)\cdots f_n(D_n)\mathbf{1}_{\{Y_0=i_0,\ldots,Y_n=i_n\}}] = \mathbb{E}_\mu[f_1(\tfrac{\mathbf{e}_1}{q(i_0)})\cdots f_n(\tfrac{\mathbf{e}_n}{q(i_{n-1})})]\mu(i_0)\pi_{i_0,i_1}\cdots\pi_{i_{n-1},i_n}. \quad (3.10)$$

To conclude, taking $f_k \equiv 1$ in (3.10) yields $\mathbb{P}_\mu(Y_0 = i_0, \ldots, Y_n = i_n) = \mu(i_0)\pi_{i_0,i_1}\cdots\pi_{i_{n-1},i_n}$, which shows that $(Y_n)_{n\in\mathbb{N}}$ is a discrete-time Markov chain with transition matrix $Q = (\pi_{i,j})_{i,j\in E}$, proving part (i). Moreover, dividing both sides of (3.10) by $\mathbb{P}_\mu(Y_0 = i_0, \ldots, Y_n = i_n)$, we obtain

$$\mathbb{E}_\mu[f_1(D_1)\cdots f_n(D_n) \mid Y_0 = i_0, \ldots, Y_n = i_n] = \mathbb{E}_\mu[f_1(\tfrac{\mathbf{e}_1}{q(i_0)})\cdots f_n(\tfrac{\mathbf{e}_n}{q(i_{n-1})})].$$

Since the random variables $(\mathbf{e}_k)_{k\geq 1}$ are independent, it follows that

$$\mathbb{E}_\mu[f_1(D_1)\cdots f_n(D_n) \mid Y_0 = i_0, \ldots, Y_n = i_n] = \mathbb{E}_\mu[f_1(\tfrac{\mathbf{e}_1}{q(i_0)})]\cdots\mathbb{E}_\mu[f_n(\tfrac{\mathbf{e}_n}{q(i_{n-1})})].$$

This shows that, conditionally on $(Y_n)_{n\in\mathbb{N}}$, the holding times $(D_n)_{n\geq 1}$ are mutually independent. Recall that under $\mathbb{P}_{i_{k-1}}$, each $\frac{\mathbf{e}_k}{q(i_{k-1})}$ has an (extended) exponential distribution with parameter

$q(i_{k-1})$. Hence, for every $i_{k-1} \in E$, we have $D_k \mid (Y_{k-1} = i_{k-1}) \sim \mathrm{Exp}(q(i_{k-1}))$ in the extended sense, which establishes part (ii).

(iii) Fix $n \geq 1$ and $\mu \in \mathcal{M}_1(E)$. Let $f : [0, \infty] \to \mathbb{R}$ and $g : E \to \mathbb{R}$ be bounded measurable functions. By the strong Markov property at the stopping time J_{n-1} and Proposition 3.36, we have $\mathbb{P}_\mu$-a.s.

$$\begin{aligned}\mathbb{E}_\mu[f(D_n)g(Y_n) \mid \mathcal{F}_{J_{n-1}}] &= \mathbb{E}_\mu[f(J_1(\theta_{J_{n-1}}\mathbf{X}))g(Y_1(\theta_{J_{n-1}}\mathbf{X})) \mid \mathcal{F}_{J_{n-1}}] \\ &= \mathbb{E}_{Y_{n-1}}[f(J_1(\mathbf{X}))g(Y_1(\mathbf{X}))] \\ &= \mathbb{E}_{Y_{n-1}}[f(J_1)]\mathbb{E}_{Y_{n-1}}[g(Y_1)].\end{aligned}$$

Using the fact that $\sigma(Y_{n-1}) = \sigma(X_{J_{n-1}}) \subseteq \mathcal{F}_{J_{n-1}}$ (since $\sigma(X_{J_{n-1}})$ is the smallest σ-algebra making $X_{J_{n-1}}$ measurable) and the tower property, we obtain $\mathbb{P}_\mu$-a.s.

$$\begin{aligned}\mathbb{E}_\mu[f(D_n)g(Y_n) \mid Y_{n-1}] &= \mathbb{E}_\mu[\mathbb{E}_\mu[f(D_n)g(Y_n) \mid \mathcal{F}_{J_{n-1}}] \mid Y_{n-1}] \\ &= \mathbb{E}_\mu[\mathbb{E}_{Y_{n-1}}[f(J_1)]\mathbb{E}_{Y_{n-1}}[g(Y_1)] \mid Y_{n-1}] \\ &= \mathbb{E}_{Y_{n-1}}[f(J_1)]\mathbb{E}_{Y_{n-1}}[g(Y_1)].\end{aligned}$$

Taking $g \equiv 1$ and $f \equiv 1$ in the above identity gives

$$\mathbb{E}_\mu[f(D_n) \mid Y_{n-1}] = \mathbb{E}_{Y_{n-1}}[f(J_1)], \quad \mathbb{E}_\mu[g(Y_n) \mid Y_{n-1}] = \mathbb{E}_{Y_{n-1}}[g(Y_1)].$$

Substituting these into the previous equality, we obtain

$$\mathbb{E}_\mu[f(D_n)g(Y_n) \mid Y_{n-1}] = \mathbb{E}_\mu[f(D_n) \mid Y_{n-1}]\mathbb{E}_\mu[g(Y_n) \mid Y_{n-1}].$$

Hence $D_n \perp\!\!\!\perp Y_n \mid Y_{n-1}$ under $\mathbb{P}_\mu$, which completes the proof of (iii). □

Intuitively, this theorem characterizes the evolution of the process through two interacting but conditionally independent mechanisms:

- The sequence of visited states $(Y_n)_{n\in\mathbb{N}}$ forms a discrete-time Markov chain, meaning that each state Y_n depends only on the preceding state Y_{n-1} and is conditionally independent of the entire past.
- The sequence of holding times $(D_n)_{n\geq 1}$ determines how long the process remains in each state before the next transition. Each D_n depends only on the current state Y_{n-1} through an exponential distribution with parameter $q(Y_{n-1})$ and is independent of the entire past before entering that state. Because the exponential distribution is memoryless, the time already spent in a state provides no information about the remaining duration. Conditionally on the sequence of visited states, the holding times are mutually independent, so once the path of states is fixed, they behave like independent exponential clocks that restart upon entering each new state.

Given the current state Y_{n-1}, the holding time D_n and the next state Y_n are independent random variables, meaning that the duration of the stay D_n carries no information about the destination of the next transition Y_n, and vice versa.

3.2.2 Backward and Forward Chapman-Kolmogorov Equations

When we observe a random phenomenon that evolves over time, it is often modeled by a Markov process. We can construct its infinitesimal generator $G = (q_{i,j})_{i,j\in E}$ using the rate $q(i)$ at which the process leaves state i and the probability $Q(i, j) = \pi_{i,j}$ of jumping from state i to state j (see Definition 3.37). However, knowing G alone is not sufficient if we wish to compute explicit probabilities such as $\mathbb{P}_i(X_t \in B)$. For this, we need to determine the entire semigroup $(P_t)_{t\in\mathbb{R}_+}$. Thanks to the following theorem, finding $(P_t)_{t\in\mathbb{R}_+}$ reduces to solving a system of linear differential equations known as the *Chapman-Kolmogorov equations*, which generalize the differential relations of Theorem 3.14 to infinite state spaces.

Theorem 3.39 (Chapman-Kolmogorov equations)**.** *Let E be a countable space equipped with the discrete topology, and hence endowed with the σ-algebra $\mathcal{P}(E)$. Let*

$$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, \mathbf{X} = (X_t)_{t\in\mathbb{R}_+}, (P_t)_{t\in\mathbb{R}_+}, \mathbb{P}_\mu, \mu \in \mathcal{M}_1(E))$$

be a right-continuous Markov process. We denote by $G := (q_{i,j})_{i,j\in E}$ its infinitesimal generator (see Definition 3.37). We use the notation $P_t := (p_t(i,j))_{i,j\in E}$ for $t \in \mathbb{R}_+$.

(i) For all $i, j \in E$, the maps $t \in \mathbb{R}_+ \mapsto p_t(i,j) \in [0,1]$ are continuously differentiable and are solutions of the following system of linear differential equations:

$$p_t'(i,j) = \sum_{k\in E} q_{i,k} p_t(k,j) \quad \text{for all } i,j \in E, t \in \mathbb{R}_+.$$

If we denote $\frac{\mathrm{d}}{\mathrm{d}t}P_t := (p_t'(i,j))_{i,j\in E}$ for all $t \in \mathbb{R}_+$, then we can also write this system as

$$\tfrac{\mathrm{d}}{\mathrm{d}t}P_t = GP_t \text{ for all } t \in \mathbb{R}_+ \quad \text{and} \quad P_0 = \mathrm{Id}_E$$

which is called the backward Chapman-Kolmogorov equation*. Note that $P_0 = \mathrm{Id}_E$ serves as the initial condition and follows from the fact that $p_0(i,j) = \mathbb{P}_i(X_0 = j) = \mathbf{1}_{\{i=j\}}$.*

(ii) Recall that $\zeta := \zeta(\mathbf{X})$ is the explosion time of $\mathbf{X}$. We set $p_t^{\min}(i,j) := \mathbb{P}_i(X_t = j, \zeta > t)$ for all $i, j \in E$, $t \in \mathbb{R}_+$ and denote $P_t^{\min} := (p_t^{\min}(i,j))_{i,j\in E}$. Then $(P_t^{\min})_{t\in\mathbb{R}_+}$ is a sub-Markovian semigroup which is a solution of the backward Chapman-Kolmogorov equation:

$$\tfrac{\mathrm{d}}{\mathrm{d}t}P_t^{\min} = GP_t^{\min} \text{ for all } t \in \mathbb{R}_+ \quad \text{and} \quad P_0^{\min} = \mathrm{Id}_E.$$

It is even the minimal nonnegative solution, i.e. if $\tilde{P}_t = (\tilde{p}_t(i,j))_{i,j\in E}$, $t \in \mathbb{R}_+$ is a family of matrices with nonnegative, continuously differentiable entries which satisfy $\frac{\mathrm{d}}{\mathrm{d}t}\tilde{P}_t = G\tilde{P}_t$ for all $t \in \mathbb{R}_+$ and $\tilde{P}_0 = \mathrm{Id}_E$, then $p_t^{\min}(i,j) \le \tilde{p}_t(i,j)$ for all $i, j \in E$, $t \in \mathbb{R}_+$.

(iii) $(P_t^{\min})_{t\in\mathbb{R}_+}$ is also a solution of the forward Chapman-Kolmogorov equation*:*

$$\tfrac{\mathrm{d}}{\mathrm{d}t}p_t^{\min}(i,j) = \sum_{k\in E} p_t^{\min}(i,k) q_{k,j} \quad \text{for all } i,j \in E, t \in \mathbb{R}_+$$

which we can also write as

$$\tfrac{\mathrm{d}}{\mathrm{d}t}P_t^{\min} = P_t^{\min}G \text{ for all } t \in \mathbb{R}_+ \quad \text{and} \quad P_0^{\min} = \mathrm{Id}_E.$$

It is even the minimal nonnegative solution, i.e. if $\tilde{P}_t = (\tilde{p}_t(i,j))_{i,j\in E}$, $t \in \mathbb{R}_+$ is a family of matrices with nonnegative, continuously differentiable entries which satisfy $\frac{\mathrm{d}}{\mathrm{d}t}\tilde{P}_t = \tilde{P}_t G$ for all $t \in \mathbb{R}_+$ and $\tilde{P}_0 = \mathrm{Id}_E$, then $p_t^{\min}(i,j) \le \tilde{p}_t(i,j)$ for all $i, j \in E$, $t \in \mathbb{R}_+$.

Proof. (i) Since $\mathbf{X}$ is right-continuous, for all $\omega \in \Omega$ and $t \in \mathbb{R}_+$, there exists $h_t(\omega) > 0$ such that $X_{t+u}(\omega) = X_t(\omega)$ for all $u \in [0, h_t(\omega)]$. Thus for all $j \in E$, we have $\mathbf{1}_{\{X_{t+h}(\omega)=j\}} \to \mathbf{1}_{\{X_t(\omega)=j\}}$ as $h \to 0^+$. By the dominated convergence theorem, it follows for all $i \in E$ that $\mathbb{E}_i[\mathbf{1}_{\{X_{t+h}=j\}}] \to \mathbb{E}_i[\mathbf{1}_{\{X_t=j\}}]$ as $h \to 0^+$, i.e. $\mathbb{P}_i(X_{t+h} = j) \to \mathbb{P}_i(X_t = j)$ as $h \to 0^+$. Hence for all $i, j \in E$, the map $t \mapsto p_t(i,j) = \mathbb{P}_i(X_t = j)$ is right-continuous.

Fix $i, j \in E$ and $t > 0$. Since $X_t = i$ on $\{J_1 > t\}$ under $\mathbb{P}_i$ and $X_t = (\theta_{J_1}\mathbf{X})_{t-J_1}$ on $\{J_1 \le t\}$, it follows that

$$\begin{aligned} p_t(i,j) &= \mathbb{P}_i(J_1 > t, X_t = j) + \mathbb{P}_i(J_1 \le t, X_t = j) \\ &= \mathbb{E}_i[\mathbf{1}_{\{J_1>t\}}\mathbf{1}_{\{X_t=j\}}] + \mathbb{E}_i[\mathbf{1}_{\{J_1\le t\}}\mathbf{1}_{\{X_t=j\}}] \\ &= \delta_{i,j}\mathbb{E}_i[\mathbf{1}_{\{J_1>t\}}] + \mathbb{E}_i[\mathbf{1}_{\{J_1\le t\}}\mathbf{1}_{\{(\theta_{J_1}\mathbf{X})_{t-J_1}=j\}}]. \end{aligned}$$

Using the strong Markov property at time J_1 and Proposition 3.36, we have

$$\begin{aligned}
\mathbb{E}_i[\mathbf{1}_{\{J_1\le t\}}\mathbf{1}_{\{(\theta_{J_1}\mathbf{X})_{t-J_1}=j\}}] &= \mathbb{E}_i[\mathbf{1}_{\{J_1\le t\}}\mathbb{E}_i[\mathbf{1}_{\{J_1<\infty\}}\mathbf{1}_{\{(\theta_{J_1}\mathbf{X})_{t-J_1}=j\}} \mid \mathcal{F}_{J_1}]] \\
&= \mathbb{E}_i[\mathbf{1}_{\{J_1\le t\}}\mathbf{1}_{\{J_1<\infty\}}\mathbb{E}_{Y_1}[\mathbf{1}_{\{X_{t-J_1}=j\}}]] \\
&= \mathbb{E}_i[\mathbf{1}_{\{J_1\le t\}}\mathbb{P}_{Y_1}(X_{t-J_1}=j)] \\
&= \int_0^\infty \sum_{k\in E} \mathbf{1}_{\{s\le t\}}\mathbb{P}_k(X_{t-s}=j)\mathbb{P}_i(Y_1=k)\mathbb{P}_i(J_1\in \mathrm{d}s) \\
&= \int_0^t \mathrm{d}s \sum_{k\in E} \mathbb{P}_k(X_{t-s}=j)\mathbb{P}_i(Y_1=k)q(i)e^{-q(i)s} \\
&= \int_0^t \mathrm{d}s\, \mathbb{E}_i[\mathbb{P}_{Y_1}(X_{t-s}=j)]q(i)e^{-q(i)s}.
\end{aligned}$$

Therefore, using Proposition 3.36 again and the change of variables $u = t - s$, we obtain

$$\begin{aligned}
p_t(i,j) &= \delta_{i,j}\mathbb{P}_i(J_1>t) + \int_0^t \mathrm{d}s\, \mathbb{E}_i[\mathbb{P}_{Y_1}(X_{t-s}=j)]q(i)e^{-q(i)s} \\
&= \delta_{i,j}e^{-q(i)t} + \int_0^t \mathrm{d}s\, \mathbb{E}_i[p_{t-s}(Y_1,j)]q(i)e^{-q(i)s} \\
&= \delta_{i,j}e^{-q(i)t} - e^{-q(i)t}\int_t^0 \mathrm{d}u\, \mathbb{E}_i[p_u(Y_1,j)]q(i)e^{q(i)u} \\
&= \delta_{i,j}e^{-q(i)t} + e^{-q(i)t}\int_0^t \mathrm{d}u\, \mathbb{E}_i[p_u(Y_1,j)]q(i)e^{q(i)u}.
\end{aligned}$$

Multiplying both sides by $e^{q(i)t}$ gives

$$e^{q(i)t}p_t(i,j) = \delta_{i,j} + \int_0^t \mathrm{d}u\, \mathbb{E}_i[p_u(Y_1,j)]q(i)e^{q(i)u}. \tag{3.11}$$

Since we showed that $p_{u+h}(Y_1,j) \to p_u(Y_1,j)$ as $h \to 0^+$ and $0 \le p_{u+h}(Y_1,j) \le 1$, the dominated convergence theorem yields $\mathbb{E}_i[p_{u+h}(Y_1,j)] \to \mathbb{E}_i[p_u(Y_1,j)]$ as $h \to 0^+$. Thus the map $u \mapsto \mathbb{E}_i[p_u(Y_1,j)]$ is right-continuous.

Now set $f(t) := e^{q(i)t}p_t(i,j)$ and $h(t) := \mathbb{E}_i[p_t(Y_1,j)]q(i)e^{q(i)t}$. We can then rewrite (3.11) as $f(t) - f(0) = \int_0^t \mathrm{d}u\, h(u)$. Since $0 \le p_u(Y_1,j) \le 1$, we have $0 \le h(u) \le q(i)e^{q(i)u}$, so h is bounded on compact intervals. We showed above that $u \mapsto \mathbb{E}_i[p_u(Y_1,j)]$ is right-continuous, so h is also right-continuous and hence measurable. From this, we deduce that h is locally integrable and by the fundamental theorem of calculus for Lebesgue integrals, f is absolutely continuous for $t \in \mathbb{R}_+$ and $f'(t) = h(t)$ for a.e. $t > 0$. From the identity $p_t(i,j) = e^{-q(i)t}f(t)$, we then obtain for a.e. $t > 0$,

$$\tfrac{\mathrm{d}}{\mathrm{d}t}p_t(i,j) = -q(i)p_t(i,j) + \mathbb{E}_i[p_t(Y_1,j)]q(i). \tag{3.12}$$

To extend this identity from holding a.e. to holding for all $t > 0$, recall that f is absolutely continuous and hence continuous. This implies that $t \mapsto e^{-q(i)t}f(t) = p_t(i,j)$ is continuous and by dominated convergence, $t \mapsto \mathbb{E}_i[p_t(Y_1,j)]$ is also continuous. Consequently, the right-hand side in (3.12) is continuous in t and thus holds for all $t > 0$. Using Definition 3.37, we obtain for all $t > 0$,

$$\begin{aligned}
\tfrac{\mathrm{d}}{\mathrm{d}t}p_t(i,j) &= -q(i)p_t(i,j) + \mathbb{E}_i[p_t(Y_1,j)]q(i) \\
&= -q(i)p_t(i,j) + \sum_{k\in E} p_t(k,j)\mathbb{P}_i(Y_1=k)q(i) \\
&= -q(i)p_t(i,j) + \sum_{k\in E} p_t(k,j)\pi_{i,k}q(i) \\
&= p_t(i,j)\pi_{i,i}q(i) + \sum_{k\in E} q_{i,k}p_t(k,j) \\
&= \sum_{k\in E} q_{i,k}p_t(k,j)
\end{aligned}$$

where for the last equality we used the fact that $\pi_{i,i}q(i) = 0$. This shows the backward Chapman-Kolmogorov equation for $t > 0$. Also, since we showed above that $\frac{\mathrm{d}}{\mathrm{d}t}p_t(i,j)$ is continuous in t, we conclude that $t \mapsto p_t(i,j)$ is continuously differentiable for $t > 0$.

It remains to identify the right-derivative at $t = 0$. Consider the representation

$$p_t(i,j) = \delta_{i,j}e^{-q(i)t} + \int_0^t \mathrm{d}s\, \mathbb{E}_i[p_{t-s}(Y_1,j)]q(i)e^{-q(i)s}.$$

Subtracting $\delta_{i,j}$ and dividing by t gives

$$\tfrac{p_t(i,j)-p_0(i,j)}{t} = \delta_{i,j}\tfrac{e^{-q(i)t}-1}{t} + \tfrac{1}{t}\int_0^t \mathrm{d}s\, \mathbb{E}_i[p_{t-s}(Y_1,j)]q(i)e^{-q(i)s}.$$

For the first term, by L'Hôpital's rule,

$$\lim_{t\to 0^+} \delta_{i,j}\tfrac{e^{-q(i)t}-1}{t} = -\delta_{i,j}q(i).$$

For the second term, since $t \mapsto p_t(i,j)$ is right-continuous on $\mathbb{R}_+$, we have for each fixed $s \geq 0$ that $p_{t-s}(Y_1,j) \to p_0(Y_1,j) = \mathbf{1}_{\{Y_1=j\}}$ as $t \to 0^+$. As $0 \leq p_{t-s}(Y_1,j) \leq 1$, dominated convergence yields $\mathbb{E}_i[p_{t-s}(Y_1,j)] \to \mathbb{E}_i[\mathbf{1}_{\{Y_1=j\}}] = \pi_{i,j}$ as $t \to 0^+$ for each fixed $s \geq 0$. Moreover, since $0 \leq \mathbb{E}_i[p_{t-s}(Y_1,j)]q(i)e^{-q(i)s} \leq q(i)e^{-q(i)s}$, dominated convergence in s gives

$$\begin{aligned}
\lim_{t\to 0^+} \tfrac{1}{t}\int_0^t \mathrm{d}s\, \mathbb{E}_i[p_{t-s}(Y_1,j)]q(i)e^{-q(i)s} &= \lim_{t\to 0^+} \tfrac{1}{t}\int_0^t \mathrm{d}s\, \pi_{i,j}q(i)e^{-q(i)s} \\
&= \pi_{i,j}q(i) \lim_{t\to 0^+} \tfrac{1}{t}\int_0^t \mathrm{d}s\, e^{-q(i)s} \\
&= \pi_{i,j}q(i) \lim_{t\to 0^+} \tfrac{1-e^{-q(i)t}}{q(i)t} \\
&= \pi_{i,j}q(i)
\end{aligned}$$

where the last equality again follows by L'Hôpital's rule. Combining the two limits, we obtain

$$\lim_{t\to 0^+} \tfrac{p_t(i,j)-p_0(i,j)}{t} = -\delta_{i,j}q(i) + \pi_{i,j}q(i) = q_{i,j} = \sum_{k\in E} q_{i,k}p_0(k,j).$$

Hence the right-derivative at $t = 0$ exists and the backward Chapman-Kolmogorov equation also holds at $t = 0$ in the right-derivative sense.

To show that $t \mapsto p_t(i,j)$ is continuously differentiable at $t = 0$ (from the right), recall that (3.12) holds for all $t > 0$. By right-continuity of $t \mapsto p_t(i,j)$ and dominated convergence, we get

$$\lim_{t\to 0^+} \tfrac{\mathrm{d}}{\mathrm{d}t}p_t(i,j) = -q(i)\delta_{i,j} + \pi_{i,j}q(i)$$

which is exactly the right-derivative of $p_t(i,j)$ at $t = 0$. Hence $t \mapsto p_t(i,j)$ is continuously differentiable for all $t \in \mathbb{R}_+$ (from the right at 0), completing the proof of (i).

(ii) For all $i,j \in E$ and all $t,s \in \mathbb{R}_+$, observe that

$$p^{\min}_{t+s}(i,j) = \mathbb{P}_i(X_{t+s} = j, \zeta > t+s) = \sum_{k\in E} \mathbb{P}_i(X_t = k, X_{t+s} = j, \zeta > t+s). \tag{3.13}$$

Moreover, recall that $\{\zeta > t\} = \{\zeta \leq t\}^c \in \mathcal{F}_t$ since ζ is a $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$-stopping time (see Proposition 3.32). Applying the weak tower property and the simple Markov property at time t:

$$\begin{aligned}
\mathbb{P}_i(X_t = k, X_{t+s} = j, \zeta > t+s) &= \mathbb{P}_i(X_t = k, \zeta(\mathbf{X}) > t, (\theta_t\mathbf{X})_s = j, \zeta(\theta_t\mathbf{X}) > s) \\
&= \mathbb{E}_i[\mathbf{1}_{\{X_t=k,\zeta(\mathbf{X})>t\}}\mathbb{E}_i[\mathbf{1}_{\{(\theta_t\mathbf{X})_s=j,\zeta(\theta_t\mathbf{X})>s\}} \mid \mathcal{F}_t]] \\
&= \mathbb{E}_i[\mathbf{1}_{\{X_t=k,\zeta(\mathbf{X})>t\}}\mathbb{E}_{X_t}[\mathbf{1}_{\{X_s=j,\zeta(\mathbf{X})>s\}}]] \\
&= \mathbb{P}_i(X_t = k, \zeta(\mathbf{X}) > t)\mathbb{P}_k(X_s = j, \zeta(\mathbf{X}) > s) \\
&= p^{\min}_t(i,k)p^{\min}_s(k,j).
\end{aligned}$$

Combined with (3.13), this shows the semigroup property $p^{\min}_{t+s}(i,j) = \sum_{k\in E} p^{\min}_t(i,k)p^{\min}_s(k,j)$. Also note that $p^{\min}_0(i,j) = \mathbb{P}_i(X_0 = j, \zeta > 0) = \mathbb{P}_i(X_0 = j) = \delta_{i,j}$, and that $p^{\min}_t(i,j)$ is nonnegative for all $i,j \in E$, $t \in \mathbb{R}_+$, since it is a probability. Together with the fact that, for each $i \in E$ and $t \in \mathbb{R}_+$, the sum $\sum_{j\in E} p^{\min}_t(i,j) = \mathbb{P}_i(\zeta > t)$ may be smaller than 1 (if explosion is possible), we conclude that $(P^{\min}_t)_{t\in\mathbb{R}_+}$ is a sub-Markovian semigroup.

Using the strong Markov property at time J_1 and Proposition 3.36, we obtain for all $t \in \mathbb{R}_+$, all $n \in \mathbb{N}$ and all $i,j \in E$:

$$
\begin{aligned}
&\mathbb{P}_i(X_t = j, t < J_{n+1}) \\
&= \mathbb{E}_i[\mathbf{1}_{\{X_t=j\}}\mathbf{1}_{\{t<J_{n+1}\}}\mathbf{1}_{\{J_1>t\}}] + \mathbb{E}_i[\mathbf{1}_{\{X_t=j\}}\mathbf{1}_{\{t<J_{n+1}\}}\mathbf{1}_{\{J_1\le t\}}] \\
&= \mathbb{E}_i[\mathbf{1}_{\{J_1>t\}}]\delta_{i,j} + \mathbb{E}_i[\mathbf{1}_{\{J_1\le t\}}\mathbb{E}_i[\mathbf{1}_{\{J_1<\infty\}}\mathbf{1}_{\{(\theta_{J_1}\mathbf{X})_{t-J_1}=j\}}\mathbf{1}_{\{t-J_1<J_n(\theta_{J_1}\mathbf{X})\}} \mid \mathcal{F}_{J_1}]] \\
&= e^{-q(i)t}\delta_{i,j} + \mathbb{E}_i[\mathbf{1}_{\{J_1\le t\}}\mathbf{1}_{\{J_1<\infty\}}\mathbb{E}_{Y_1}[\mathbf{1}_{\{X_{t-J_1}=j\}}\mathbf{1}_{\{t-J_1<J_n(\mathbf{X})\}}]] \\
&= e^{-q(i)t}\delta_{i,j} + \sum_{k\in E}\int_0^\infty \mathbf{1}_{\{s\le t\}}\mathbb{P}_k(X_{t-s} = j, t-s < J_n)\mathbb{P}_i(J_1 \in \mathrm{d}s)\mathbb{P}_i(Y_1 = k) \\
&= e^{-q(i)t}\delta_{i,j} + \sum_{k\in E}\int_0^t \mathrm{d}s\, \pi_{i,k}\mathbb{P}_k(X_{t-s} = j, t-s < J_n)q(i)e^{-q(i)s} \\
&= e^{-q(i)t}\delta_{i,j} + \int_0^t \mathrm{d}s\, \pi_{i,i}\mathbb{P}_i(X_{t-s} = j, t-s < J_n)q(i)e^{-q(i)s} + \\
&\quad \sum_{k\in E\setminus\{i\}}\int_0^t \mathrm{d}s\, q_{i,k}\mathbb{P}_k(X_{t-s} = j, t-s < J_n)e^{-q(i)s} \\
&= e^{-q(i)t}\delta_{i,j} + \sum_{k\in E\setminus\{i\}}\int_0^t \mathrm{d}s\, q_{i,k}\mathbb{P}_k(X_{t-s} = j, t-s < J_n)e^{-q(i)s} \qquad (3.14)
\end{aligned}
$$

where for the last equality we used the fact that $q(i)\pi_{i,i} = 0$. Since $J_n \uparrow \zeta$ as $n \to \infty$, the events $\{X_{t-s} = j, t-s < J_n\}$ increase to $\{X_{t-s} = j, t-s < \zeta\}$ as $n \to \infty$. Hence $\mathbf{1}_{\{X_{t-s}=j,t-s<J_n\}} \uparrow \mathbf{1}_{\{X_{t-s}=j,t-s<\zeta\}}$ as $n \to \infty$. By the monotone convergence theorem, we therefore obtain

$$
\begin{aligned}
\lim_{n\to\infty} \mathbb{P}_k(X_{t-s} = j, t-s < J_n) &= \lim_{n\to\infty} \mathbb{E}_k[\mathbf{1}_{\{X_{t-s}=j,t-s<J_n\}}] \\
&= \mathbb{E}_k[\mathbf{1}_{\{X_{t-s}=j,t-s<\zeta\}}] \\
&= \mathbb{P}_k(X_{t-s} = j, t-s < \zeta) \\
&= p^{\min}_{t-s}(k,j).
\end{aligned}
$$

Likewise, $\lim_{n\to\infty} \mathbb{P}_i(X_t = j, t < J_{n+1}) = \mathbb{P}_i(X_t = j, t < \zeta) = p^{\min}_t(i,j)$. Taking limits on both sides of (3.14) and again using the monotone convergence theorem, we have

$$
p^{\min}_t(i,j) = e^{-q(i)t}\delta_{i,j} + \sum_{k\in E\setminus\{i\}}\int_0^t \mathrm{d}s\, q_{i,k}p^{\min}_{t-s}(k,j)e^{-q(i)s}.
$$

Applying the change of variables $u = t - s$ to the integral, we obtain

$$
\begin{aligned}
p^{\min}_t(i,j) &= e^{-q(i)t}\delta_{i,j} - e^{-q(i)t}\sum_{k\in E\setminus\{i\}}\int_t^0 \mathrm{d}u\, q_{i,k}p^{\min}_u(k,j)e^{q(i)u} \\
&= e^{-q(i)t}\delta_{i,j} + e^{-q(i)t}\sum_{k\in E\setminus\{i\}}\int_0^t \mathrm{d}u\, q_{i,k}p^{\min}_u(k,j)e^{q(i)u}.
\end{aligned}
$$

To differentiate this, we need the integrand $u \mapsto q_{i,k}p^{\min}_u(k,j)e^{q(i)u}$ to be continuous on $[0,t]$. Indeed, $u \mapsto p^{\min}_u(k,j)$ is continuous on $\mathbb{R}_+$ by the same argument used in (i) to show the continuity of $t \mapsto p_t(i,j)$ on $\mathbb{R}_+$. Therefore, $u \mapsto q_{i,k}p^{\min}_u(k,j)e^{q(i)u}$ is continuous on $\mathbb{R}_+$ and we can apply

the fundamental theorem of calculus to obtain

$$\begin{aligned}
\tfrac{\mathrm{d}}{\mathrm{d}t}p_t^{\min}(i,j) &= -q(i)e^{-q(i)t}\delta_{i,j} - q(i)e^{-q(i)t}\sum_{k\in E\setminus\{i\}}\int_0^t \mathrm{d}u\, q_{i,k}p_u^{\min}(k,j)e^{q(i)u} + \\
&\qquad e^{-q(i)t}\sum_{k\in E\setminus\{i\}} q_{i,k}p_t^{\min}(k,j)e^{q(i)t} \\
&= -q(i)(e^{-q(i)t}\delta_{i,j} + e^{-q(i)t}\sum_{k\in E\setminus\{i\}}\int_0^t \mathrm{d}u\, q_{i,k}p_u^{\min}(k,j)e^{q(i)u}) + \sum_{k\in E\setminus\{i\}} q_{i,k}p_t^{\min}(k,j) \\
&= -q(i)p_t^{\min}(i,j) + \sum_{k\in E\setminus\{i\}} q_{i,k}p_t^{\min}(k,j) \\
&= \sum_{k\in E} q_{i,k}p_t^{\min}(k,j)
\end{aligned}$$

which implies that $(P_t^{\min})_{t\in\mathbb{R}_+}$ satisfies the backward Chapman-Kolmogorov equation.

We now want to prove minimality. Let $\tilde{P}_t = (\tilde{p}_t(i,j))_{i,j\in E}$, $t\in\mathbb{R}_+$ be a solution of the backward Chapman-Kolmogorov equation, which implies for all $t\in\mathbb{R}_+$ and all $i,j\in E$

$$\begin{aligned}
\tfrac{\mathrm{d}}{\mathrm{d}t}\tilde{p}_t(i,j) = \sum_{k\in E} q_{i,k}\tilde{p}_t(k,j) \quad &\iff \quad \tfrac{\mathrm{d}}{\mathrm{d}t}\tilde{p}_t(i,j) + q(i)\tilde{p}_t(i,j) = \sum_{k\in E\setminus\{i\}} q_{i,k}\tilde{p}_t(k,j) \\
&\iff \quad e^{q(i)t}\tfrac{\mathrm{d}}{\mathrm{d}t}\tilde{p}_t(i,j) + q(i)e^{q(i)t}\tilde{p}_t(i,j) = e^{q(i)t}\sum_{k\in E\setminus\{i\}} q_{i,k}\tilde{p}_t(k,j) \\
&\iff \quad \tfrac{\mathrm{d}}{\mathrm{d}t}(e^{q(i)t}\tilde{p}_t(i,j)) = e^{q(i)t}\sum_{k\in E\setminus\{i\}} q_{i,k}\tilde{p}_t(k,j).
\end{aligned}$$

Applying integrals to the last display and using the fundamental theorem of calculus, we obtain

$$\begin{aligned}
&\int_0^t \mathrm{d}s\, \tfrac{\mathrm{d}}{\mathrm{d}s}(e^{q(i)s}\tilde{p}_s(i,j)) = \int_0^t \mathrm{d}s\, e^{q(i)s}\sum_{k\in E\setminus\{i\}} q_{i,k}\tilde{p}_s(k,j) \\
\iff \quad & e^{q(i)t}\tilde{p}_t(i,j) - \tilde{p}_0(i,j) = \sum_{k\in E\setminus\{i\}}\int_0^t \mathrm{d}s\, e^{q(i)s}q_{i,k}\tilde{p}_s(k,j) \\
\iff \quad & \tilde{p}_t(i,j) - e^{-q(i)t}\delta_{i,j} = \sum_{k\in E\setminus\{i\}}\int_0^t \mathrm{d}s\, e^{-q(i)(t-s)}q_{i,k}\tilde{p}_s(k,j) \\
\iff \quad & \tilde{p}_t(i,j) = e^{-q(i)t}\delta_{i,j} + \sum_{k\in E\setminus\{i\}}\int_0^t \mathrm{d}u\, e^{-q(i)u}q_{i,k}\tilde{p}_{t-u}(k,j) \qquad (3.15)
\end{aligned}$$

where in the last equivalence we used the change of variables $u := t-s$. Next, we want to show by induction that

$$\mathbb{P}_i(X_t = j, t < J_n) \le \tilde{p}_t(i,j) \qquad (3.16)$$

for all $t\in\mathbb{R}_+$, all $n\in\mathbb{N}$ and all $i,j\in E$. For the base case, observe that $\mathbb{P}_i(X_t = j, t < J_0) = \mathbb{P}_i(X_t = j, t < 0) = 0 \le \tilde{p}_t(i,j)$. Suppose now that $\mathbb{P}_i(X_t = j, t < J_n) \le \tilde{p}_t(i,j)$ for some $n\in\mathbb{N}$. Using (3.14), the induction hypothesis and (3.15), we deduce that $\mathbb{P}_i(X_t = j, t < J_{n+1}) \le \tilde{p}_t(i,j)$, which concludes the induction. Passing to the limit $n\to\infty$ in (3.16) yields $\mathbb{P}_i(X_t = j, t < \zeta) = p_t^{\min}(i,j) \le \tilde{p}_t(i,j)$, for all $i,j\in E$ and all $t\in\mathbb{R}_+$, which shows minimality and completes the proof of (ii).

(iii) Recall that $J_{n+1} = J_n + D_{n+1} = J_{n-1} + D_n + D_{n+1}$. Set

$$\begin{aligned}
\alpha &:= \mathbb{E}_i[\mathbf{1}_{\{J_n \le t < J_{n+1}, Y_{n-1}=k, Y_n=j\}} \mid Y_{n-1}, J_{n-1}] \\
&= \mathbf{1}_{\{Y_{n-1}=k\}} \mathbb{E}_i[\mathbf{1}_{\{J_{n-1}+D_n \le t < J_{n-1}+D_n+D_{n+1}, Y_n=j\}} \mid Y_{n-1}, J_{n-1}] \\
&= \mathbf{1}_{\{Y_{n-1}=k\}} \sum_{\ell \in E} \mathbf{1}_{\{Y_{n-1}=\ell\}} \mathbb{E}_i[\mathbf{1}_{\{J_{n-1}+D_n \le t < J_{n-1}+D_n+D_{n+1}, Y_n=j\}} \mid Y_{n-1}=\ell, J_{n-1}] \\
&= \mathbf{1}_{\{Y_{n-1}=k\}} \mathbb{E}_i[\mathbf{1}_{\{J_{n-1}+D_n \le t < J_{n-1}+D_n+D_{n+1}, Y_n=j\}} \mid Y_{n-1}=k, J_{n-1}] \\
&= \mathbf{1}_{\{Y_{n-1}=k\}} \int_{s \in \mathbb{R}_+} \int_{u \in \mathbb{R}_+} \sum_{\ell \in E} \mathbf{1}_{\{J_{n-1}+s \le t < J_{n-1}+s+u, \ell=j\}} \\
&\qquad \mathbb{P}_i(Y_n = \ell, D_{n+1} \in \mathrm{d}u, D_n \in \mathrm{d}s \mid Y_{n-1}=k, J_{n-1}) \quad \mathbb{P}_i\text{-a.s.}
\end{aligned}$$

Note that if all integrands are nonnegative measurable functions, Tonelli's theorem justifies both the interchange of integration order and the factorization of integrals over product spaces. We know from Theorem 3.38:

- $(Y_n)_{n\in\mathbb{N}}$ is a discrete-time Markov chain with transition matrix $Q = (\pi_{i,j})_{i,j\in E}$, which implies that $Y_n \perp\!\!\!\perp (Y_0, \dots, Y_{n-2})$ given Y_{n-1}.
- Conditionally on Y_{n-1}, we have $D_n \sim \mathrm{Exp}(q(Y_{n-1}))$. This implies that, given Y_{n-1}, D_n is independent of everything before entering Y_{n-1}.
- Conditionally on Y_n, we have $D_{n+1} \sim \mathrm{Exp}(q(Y_n))$. This implies that, given Y_n, D_{n+1} is independent of everything before entering Y_n.
- $(D_n)_{n\ge1}$ are mutually independent given $(Y_n)_{n\in\mathbb{N}}$.
- Conditionally on Y_{n-1}, we have $D_n \perp\!\!\!\perp Y_n$.

Using the chain rule for conditional probabilities, we obtain

$$\begin{aligned}
&\mathbb{P}_i(Y_n = \ell, D_{n+1} \in \mathrm{d}u, D_n \in \mathrm{d}s \mid Y_{n-1}=k, J_{n-1}) \\
&= \mathbb{P}_i(D_{n+1} \in \mathrm{d}u \mid Y_n = \ell, Y_{n-1}=k, D_n = s, J_{n-1}) \mathbb{P}_i(Y_n = \ell, D_n \in \mathrm{d}s \mid Y_{n-1}=k, J_{n-1}).
\end{aligned}$$

To simplify this further, we next want to show that $J_{n-1} \perp\!\!\!\perp Y_n$ given Y_{n-1}. Indeed, note that $Y_n = Y_1(\theta_{J_{n-1}}\mathbf{X})$ and that by the strong Markov property at J_{n-1},

$$\mathbb{E}_i[F(\theta_{J_{n-1}}\mathbf{X}) \mid \mathcal{F}_{J_{n-1}}] = \mathbb{E}_{Y_{n-1}}[F(\mathbf{X})] \quad \mathbb{P}_i\text{-a.s.}$$

for all bounded measurable F and on the event $\{J_{n-1} < \infty\}$. Taking $F(\mathbf{x}) = \mathbf{1}_{\{Y_1(\mathbf{x})=j\}}$, we obtain $\mathbb{E}_i[\mathbf{1}_{\{Y_n=j\}} \mid \mathcal{F}_{J_{n-1}}] = \mathbb{E}_{Y_{n-1}}[\mathbf{1}_{\{Y_1=j\}}]$ $\mathbb{P}_i$-a.s. where the right-hand side is $\sigma(Y_{n-1})$-measurable. Now let H be $\sigma(Y_{n-1})$-measurable. By the weak tower property and the fact that $\sigma(Y_{n-1}) \subseteq \mathcal{F}_{J_{n-1}}$ (since $\sigma(Y_{n-1})$ is the smallest σ-algebra making Y_{n-1} measurable), we get that

$$\begin{aligned}
\mathbb{E}_i[\mathbf{1}_{\{Y_n=j\}}H] &= \mathbb{E}_i[\mathbb{E}_i[\mathbf{1}_{\{Y_n=j\}}H \mid \mathcal{F}_{J_{n-1}}]] \\
&= \mathbb{E}_i[\mathbb{E}_i[\mathbf{1}_{\{Y_n=j\}} \mid \mathcal{F}_{J_{n-1}}]H] \\
&= \mathbb{E}_i[\mathbb{E}_{Y_{n-1}}[\mathbf{1}_{\{Y_1=j\}}]H].
\end{aligned}$$

Therefore $\mathbb{E}_{Y_{n-1}}[\mathbf{1}_{\{Y_1=j\}}]$ satisfies the two characterizing properties of the conditional expectation $\mathbb{E}_i[\mathbf{1}_{\{Y_n=j\}} \mid Y_{n-1}]$ and we obtain $\mathbb{P}_i$-a.s.

$$\mathbb{E}_i[\mathbf{1}_{\{Y_n=j\}} \mid Y_{n-1}] = \mathbb{E}_{Y_{n-1}}[\mathbf{1}_{\{Y_1=j\}}] = \mathbb{E}_i[\mathbf{1}_{\{Y_n=j\}} \mid \mathcal{F}_{J_{n-1}}].$$

Let $B \in \mathcal{B}(\mathbb{R}_+)$. Then by the tower property

$$\begin{aligned}
\mathbb{E}_i[\mathbf{1}_{\{Y_n=j\}}\mathbf{1}_{\{J_{n-1}\in B\}} \mid Y_{n-1}] &= \mathbb{E}_i[\mathbb{E}_i[\mathbf{1}_{\{Y_n=j\}} \mid \mathcal{F}_{J_{n-1}}]\mathbf{1}_{\{J_{n-1}\in B\}} \mid Y_{n-1}] \\
&= \mathbb{E}_i[\mathbb{E}_{Y_{n-1}}[\mathbf{1}_{\{Y_1=j\}}]\mathbf{1}_{\{J_{n-1}\in B\}} \mid Y_{n-1}] \\
&= \mathbb{E}_{Y_{n-1}}[\mathbf{1}_{\{Y_1=j\}}]\mathbb{E}_i[\mathbf{1}_{\{J_{n-1}\in B\}} \mid Y_{n-1}] \\
&= \mathbb{E}_i[\mathbf{1}_{\{Y_n=j\}} \mid Y_{n-1}]\mathbb{E}_i[\mathbf{1}_{\{J_{n-1}\in B\}} \mid Y_{n-1}] \quad \mathbb{P}_i\text{-a.s.}
\end{aligned}$$

Therefore

$$\mathbb{P}_i(Y_n = j, J_{n-1} \in B \mid Y_{n-1}) = \mathbb{P}_i(Y_n = j \mid Y_{n-1})\mathbb{P}_i(J_{n-1} \in B \mid Y_{n-1})$$

and we conclude that $J_{n-1} \perp\!\!\!\perp Y_n$ given Y_{n-1}.

It follows directly from the definition of conditional independence that for any events A, B and C, if $A \perp\!\!\!\perp C \mid B$, then $\mathbb{P}(A \mid B, C) = \mathbb{P}(A \mid B)$. Using this, the conditional independence we just showed above and the fact that $J_{n-1} = J_{n-2} + D_{n-1} \perp\!\!\!\perp D_n \mid Y_{n-1}$, we get

$$\begin{aligned}\mathbb{P}_i(Y_n = \ell, D_n \in \mathrm{d}s \mid Y_{n-1} = k, J_{n-1}) &= \mathbb{P}_i(Y_n = \ell, D_n \in \mathrm{d}s \mid Y_{n-1} = k) \\ &= \mathbb{P}_i(Y_n = \ell \mid Y_{n-1} = k)\mathbb{P}_i(D_n \in \mathrm{d}s \mid Y_{n-1} = k)\end{aligned}$$

where the second equation comes from $D_n \perp\!\!\!\perp Y_n \mid Y_{n-1}$. Substituting this into the expression of α, we obtain

$$\begin{aligned}\alpha = \mathbf{1}_{\{Y_{n-1}=k\}} &\int_{s\in\mathbb{R}_+}\int_{u\in\mathbb{R}_+}\sum_{\ell\in E}\mathbf{1}_{\{J_{n-1}+s\le t<J_{n-1}+s+u,\ell=j\}} \\ &\mathbb{P}_i(D_{n+1} \in \mathrm{d}u \mid Y_n = \ell, Y_{n-1} = k, D_n = s, J_{n-1})\mathbb{P}_i(Y_n = \ell \mid Y_{n-1} = k)\mathbb{P}_i(D_n \in \mathrm{d}s \mid Y_{n-1} = k) \\ = \mathbf{1}_{\{Y_{n-1}=k\}} &\int_{s\in\mathbb{R}_+}\int_{u\in\mathbb{R}_+}\sum_{\ell\in E}\mathbf{1}_{\{J_{n-1}+s\le t<J_{n-1}+s+u,\ell=j\}}\mathbb{P}_i(D_{n+1} \in \mathrm{d}u \mid Y_n = \ell) \\ &\mathbb{P}_i(Y_n = \ell \mid Y_{n-1} = k)\mathbb{P}_i(D_n \in \mathrm{d}s \mid Y_{n-1} = k)\end{aligned}$$

where we used $D_{n+1} \perp\!\!\!\perp (Y_{n-1}, D_n, J_{n-1}) \mid Y_n$ for the second equation. Using what we know from Theorem 3.38, we get

$$\begin{aligned}\alpha &= \mathbf{1}_{\{Y_{n-1}=k\}}\int_0^\infty \mathrm{d}s\int_0^\infty \mathrm{d}u\,\mathbf{1}_{\{J_{n-1}+s\le t<J_{n-1}+s+u\}}q(j)e^{-q(j)u}\pi_{k,j}q(k)e^{-q(k)s} \\ &= \pi_{k,j}q(k)\int_0^\infty \mathrm{d}s\,e^{-q(k)s}\mathbf{1}_{\{J_{n-1}+s\le t,Y_{n-1}=k\}}\int_{t-J_{n-1}-s}^\infty \mathrm{d}u\,q(j)e^{-q(j)u} \\ &= \pi_{k,j}q(k)\int_0^{t-J_{n-1}} \mathrm{d}s\,e^{-q(k)s}\mathbf{1}_{\{J_{n-1}+s\le t,J_{n-1}\le t,Y_{n-1}=k\}}e^{-q(j)(t-J_{n-1}-s)}.\end{aligned}$$

Next, we perform the change of variables $w = t - J_{n-1} - s$ and obtain

$$\begin{aligned}\alpha &= \pi_{k,j}q(k)\int_0^{t-J_{n-1}} \mathrm{d}w\,e^{-q(k)(t-J_{n-1}-w)}e^{-q(j)w}\mathbf{1}_{\{J_{n-1}\le t,Y_{n-1}=k\}} \\ &= \pi_{k,j}q(k)\int_0^{t-J_{n-1}} \mathrm{d}w\int_{t-J_{n-1}-w}^\infty \mathrm{d}u\,q(k)e^{-q(k)u}e^{-q(j)w}\mathbf{1}_{\{J_{n-1}\le t,Y_{n-1}=k\}} \\ &= \pi_{k,j}q(k)\int_0^{t-J_{n-1}} \mathrm{d}w\,\mathbb{E}_i[\mathbf{1}_{\{D_n>t-J_{n-1}-w\}} \mid Y_{n-1}=k]e^{-q(j)w}\mathbf{1}_{\{J_{n-1}\le t,Y_{n-1}=k\}} \\ &= \pi_{k,j}q(k)\int_0^{t-J_{n-1}} \mathrm{d}w\,\mathbb{E}_i[\mathbf{1}_{\{J_{n-1}\le t,D_n>t-J_{n-1}-w,Y_{n-1}=k\}} \mid J_{n-1},Y_{n-1}]e^{-q(j)w} \\ &= \pi_{k,j}q(k)\int_0^{t-J_{n-1}} \mathrm{d}w\,\mathbb{E}_i[\mathbf{1}_{\{0\le J_{n-1}\le t-w<J_n,Y_{n-1}=k\}} \mid J_{n-1},Y_{n-1}]e^{-q(j)w} \\ &= \pi_{k,j}q(k)\int_0^{t} \mathrm{d}w\,\mathbb{E}_i[\mathbf{1}_{\{J_{n-1}\le t-w<J_n,Y_{n-1}=k\}} \mid J_{n-1},Y_{n-1}]e^{-q(j)w}\end{aligned}$$

where we again used Theorem 3.38 and the fact that for any events A, B and C, if $A \perp\!\!\!\perp C \mid B$, then $\mathbb{P}(A \mid B, C) = \mathbb{P}(A \mid B)$. Taking expectation on both sides of the previous display and using the weak tower property, we obtain

$$\begin{aligned}&\mathbb{E}_i[\mathbf{1}_{\{J_n\le t<J_{n+1},Y_{n-1}=k,Y_n=j\}}] = \pi_{k,j}q(k)\int_0^t \mathrm{d}w\,\mathbb{E}_i[\mathbf{1}_{\{J_{n-1}\le t-w<J_n,Y_{n-1}=k\}}]e^{-q(j)w} \iff \\ &\mathbb{P}_i(J_n \le t < J_{n+1}, Y_{n-1} = k, Y_n = j) = \mathbf{1}_{\{k\neq j\}}q_{k,j}\int_0^t \mathrm{d}w\,\mathbb{P}_i(J_{n-1} \le t-w < J_n, Y_{n-1} = k)e^{-q(j)w}\end{aligned}$$

since $\pi_{k,j}q(k) = 0$ if $k = j$ and $\pi_{k,j}q(k) = q_{k,j}$ if $k \neq j$. Using this equation, we then observe that

$$\begin{aligned}
&\mathbb{P}_i(X_t = j, t < J_{n+1}) \\
&= \mathbb{P}_i\Big(\bigsqcup_{m=0}^{n} \{X_t = j, J_m \leq t < J_{m+1}\}\Big) \\
&= \sum_{m=0}^{n} \mathbb{P}_i(Y_m = j, J_m \leq t < J_{m+1}) \\
&= \mathbb{P}_i(Y_0 = j, J_0 \leq t < J_1) + \sum_{m=1}^{n} \sum_{k \in E} \mathbb{P}_i(Y_{m-1} = k, Y_m = j, J_m \leq t < J_{m+1}) \\
&= \mathbb{E}_i[\mathbf{1}_{\{X_0 = j\}} \mathbf{1}_{\{J_1 > t\}}] + \sum_{m=1}^{n} \sum_{k \in E \setminus \{j\}} \mathbb{P}_i(Y_{m-1} = k, Y_m = j, J_m \leq t < J_{m+1}) \\
&= \delta_{i,j}\mathbb{P}_i(J_1 > t) + \sum_{m=1}^{n} \sum_{k \in E \setminus \{j\}} \mathbf{1}_{\{k \neq j\}} q_{k,j} \int_0^t \mathrm{d}w\, \mathbb{P}_i(J_{m-1} \leq t - w < J_m, Y_{m-1} = k) e^{-q(j)w} \\
&= \delta_{i,j} e^{-q(j)t} + \sum_{k \in E \setminus \{j\}} q_{k,j} \int_0^t \mathrm{d}w\, e^{-q(j)w} \mathbb{P}_i(X_{t-w} = k, t - w < J_n).
\end{aligned}$$

In the proof of (ii), we applied the monotone convergence theorem, using the fact that $J_n \uparrow \zeta$ as $n \to \infty$, to obtain

$$\lim_{n \to \infty} \mathbb{P}_i(X_t = j, t < J_{n+1}) = p_t^{\min}(i,j) \quad \text{and} \quad \lim_{n \to \infty} \mathbb{P}_i(X_{t-w} = k, t - w < J_n) = p_{t-w}^{\min}(i,k).$$

Letting $n \to \infty$ in the previous display, the monotone convergence theorem therefore implies that

$$p_t^{\min}(i,j) = \delta_{i,j} e^{-q(j)t} + \sum_{k \in E \setminus \{j\}} q_{k,j} \int_0^t \mathrm{d}w\, e^{-q(j)w} p_{t-w}^{\min}(i,k).$$

Multiplying by $e^{q(j)t}$ and then making the change of variables $s = t - w$ gives

$$\begin{aligned}
e^{q(j)t} p_t^{\min}(i,j) &= \delta_{i,j} + \sum_{k \in E \setminus \{j\}} q_{k,j} \int_0^t \mathrm{d}w\, e^{q(j)(t-w)} p_{t-w}^{\min}(i,k) \\
&= \delta_{i,j} + \sum_{k \in E \setminus \{j\}} q_{k,j} \int_0^t \mathrm{d}s\, e^{q(j)s} p_s^{\min}(i,k) \\
&= \delta_{i,j} + \int_0^t \mathrm{d}s \sum_{k \in E \setminus \{j\}} q_{k,j} e^{q(j)s} p_s^{\min}(i,k). \qquad (3.17)
\end{aligned}$$

Next, we show that the integrand $s \mapsto \sum_{k \in E \setminus \{j\}} q_{k,j} e^{q(j)s} p_s^{\min}(i,k)$ is continuous on $[0,t]$, which will then allow us to differentiate (3.17) with respect to t. Indeed, we have seen in (ii) that $s \mapsto p_s^{\min}(i,k)$ is continuous on $\mathbb{R}_+$, so every summand $s \mapsto q_{k,j} e^{q(j)s} p_s^{\min}(i,k)$ is continuous on $\mathbb{R}_+$. If E is finite, this directly implies that $s \mapsto \sum_{k \in E \setminus \{j\}} q_{k,j} e^{q(j)s} p_s^{\min}(i,k)$ is continuous on $\mathbb{R}_+$. If E is countably infinite, we fix an increasing sequence of finite subsets $(E_n)_{n \in \mathbb{N}}$ with $\bigcup_{n \in \mathbb{N}} E_n = E$ and define

$$S_n(s) := \sum_{k \in E_n \setminus \{j\}} q_{k,j} e^{q(j)s} p_s^{\min}(i,k), \quad S(s) := \sum_{k \in E \setminus \{j\}} q_{k,j} e^{q(j)s} p_s^{\min}(i,k).$$

Each S_n is continuous on $\mathbb{R}_+$ and $S_n(s) \to S(s)$ pointwise as $n \to \infty$ since all summands are nonnegative. Moreover, for $s \in [0,t]$ we have $|S_n(s)| \leq \sum_{k \in E \setminus \{j\}} |q_{k,j}| e^{q(j)t} < \infty$ and thus, by the Weierstrass M-test, the partial sums S_n converge uniformly to S on $[0,t]$ as $n \to \infty$. Since each

S_n is continuous, their uniform limit S is continuous on $[0,t]$. Now, by the fundamental theorem of calculus, differentiating (3.17) gives

$$e^{q(j)t}(\tfrac{\mathrm{d}}{\mathrm{d}t}p_t^{\min}(i,j) + q(j)p_t^{\min}(i,j)) = e^{q(j)t}\sum_{k\in E\setminus\{j\}} q_{k,j}p_t^{\min}(i,k)$$
$$\iff \quad e^{q(j)t}\tfrac{\mathrm{d}}{\mathrm{d}t}p_t^{\min}(i,j) = e^{q(j)t}\sum_{k\in E} p_t^{\min}(i,k)q_{k,j}.$$

Thus the semigroup $(P_t^{\min})_{t\in\mathbb{R}_+}$ satisfies the forward Chapman-Kolmogorov equation. The minimality of $(P_t^{\min})_{t\in\mathbb{R}_+}$ follows from the same argument as in (ii). □

Observe that $p_t^{\min}(i,j)$ is the probability that the process starting from i is in state j at time t before explosion, and it becomes 0 after explosion. In general, the semigroup $(P_t^{\min})_{t\in\mathbb{R}_+}$ is therefore only sub-Markovian, since probability mass may be lost when the process explodes. However, if no explosion occurs, then no probability mass is lost and $(P_t^{\min})_{t\in\mathbb{R}_+}$ is in fact Markovian.

Remark 3.40. Let E be a countable space equipped with the discrete topology, and hence endowed with the σ-algebra $\mathcal{P}(E)$. Let $(P_t)_{t\in\mathbb{R}_+}$ be a Markovian semigroup on E. For every $t\in\mathbb{R}_+$, we write $P_t := (p_t(i,j))_{i,j\in E}$. We consider a right-continuous Markov process

$$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, \mathbf{X} = (X_t)_{t\in\mathbb{R}_+}, (P_t)_{t\in\mathbb{R}_+}, \mathbb{P}_\mu, \mu\in\mathcal{M}_1(E))$$

and we denote by $G := (q_{i,j})_{i,j\in E}$ its infinitesimal generator.

If the process $\mathbf{X}$ explodes, that is, if there exists $i\in E$ such that $\mathbb{P}_i(\zeta<\infty)>0$, then the backward Chapman-Kolmogorov equation does not admit a unique solution. Indeed, explosion implies that there exist $j\in E$ and $t>0$ such that $\mathbb{P}_i(X_t=j,\zeta\le t)>0$. We observe that

$$p_t(i,j) = \mathbb{P}_i(X_t=j) = \mathbb{P}_i(X_t=j,\zeta>t) + \mathbb{P}_i(X_t=j,\zeta\le t) > \mathbb{P}_i(X_t=j,\zeta>t) = p_t^{\min}(i,j).$$

Thus $P_t\neq P_t^{\min}$, although by Theorem 3.39 both of these are solutions of the backward Chapman-Kolmogorov equation.

Observe that we do not assume here that $(P_t)_{t\in\mathbb{R}_+}$ satisfies the forward Chapman-Kolmogorov equation. This omission is intentional and the role of the forward equation, in particular its impact on uniqueness, will be explained later.

3.2.3 Examples of Markov Processes

To illustrate the results of the two previous subsections 3.2.1 and 3.2.2, we study two examples of continuous-time Markov processes on countable state spaces.

Example 3.41 ([Szn17], Example 4.16). A machine is operative for exponential times with parameter $\lambda>0$ and then inoperative for exponential times with parameter $\mu>0$. The state space is $E := \{0,1\}$, where 0 denotes the operative state and 1 the inoperative state, and $(X_t)_{t\in\mathbb{R}_+}$ describes at each time t whether the machine is operative or not. From this model, we want to determine the generator matrix G and the transition matrix Q, and then deduce $(P_t)_{t\in\mathbb{R}_+}$ using the Chapman-Kolmogorov equations.

First, we need to justify that $(X_t)_{t\in\mathbb{R}_+}$ is a right-continuous Markov process. Let $(J_n)_{n\in\mathbb{N}}$ denote the successive jump times of $(X_t)_{t\in\mathbb{R}_+}$. Since the process remains constant on each interval $[J_n, J_{n+1})$, it is right-continuous. Moreover, because the holding times are exponentially distributed and the exponential distribution is memoryless, the distribution of the next holding time, as well as the choice of the next state (according to Q), depend only on the current state and not on the past. This ensures that $(X_t)_{t\in\mathbb{R}_+}$ is a Markov process.

From the description "operative for exponential times with parameter $\lambda>0$", we know that $D_1 \mid (Y_0=0)\sim\mathrm{Exp}(\lambda)$ and therefore, by Theorem 3.38, we have $q(0)=\lambda$. Similarly, since $D_1 \mid (Y_0=1)\sim\mathrm{Exp}(\mu)$, we can deduce that $q(1)=\mu$. By Definition 3.37, we have $q_{0,0}=-q(0)=-\lambda$ and $q_{1,1}=-q(1)=-\mu$. Moreover, $\sum_{j\neq 0} q_{0,j} = -q_{0,0}$ gives $q_{0,1}=\lambda$, and $\sum_{j\neq 1} q_{1,j} = -q_{1,1}$ gives $q_{1,0}=\mu$. Thus, the generator matrix is

$$G = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}.$$

Now we want to compute the transition matrix Q of the process $(Y_n)_{n\in\mathbb{N}}$. From Definition 3.37, we deduce that $\pi_{0,0} = 0$ since $q(0) = \lambda > 0$ and $\pi_{1,1} = 0$ since $q(1) = \mu > 0$. For the off-diagonal entries of Q, we know by Definition 3.37 that $\pi_{0,1} = \frac{q_{0,1}}{q(0)} = \frac{\lambda}{\lambda} = 1$ and $\pi_{1,0} = \frac{q_{1,0}}{q(1)} = \frac{\mu}{\mu} = 1$. Thus, the transition matrix Q is given by

$$Q = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Next, using the backward Chapman-Kolmogorov equation, we obtain

$$\begin{aligned} p_t'(0,0) &= q_{0,0}p_t(0,0) + q_{0,1}p_t(1,0) = \lambda(p_t(1,0) - p_t(0,0)) \\ p_t'(0,1) &= q_{0,0}p_t(0,1) + q_{0,1}p_t(1,1) = \lambda(p_t(1,1) - p_t(0,1)) \\ p_t'(1,0) &= q_{1,0}p_t(0,0) + q_{1,1}p_t(1,0) = \mu(p_t(0,0) - p_t(1,0)) \\ p_t'(1,1) &= q_{1,0}p_t(0,1) + q_{1,1}p_t(1,1) = \mu(p_t(0,1) - p_t(1,1)). \end{aligned} \tag{3.18}$$

From this, we deduce

$$\mu p_t'(0,0) + \lambda p_t'(1,0) = \mu\lambda(p_t(1,0) - p_t(0,0)) + \lambda\mu(p_t(0,0) - p_t(1,0)) = 0$$

so that $t \mapsto \mu p_t(0,0) + \lambda p_t(1,0)$ is constant. At $t = 0$, we have

$$\mu p_0(0,0) + \lambda p_0(1,0) = \mu\mathbb{P}_0(X_0 = 0) + \lambda\mathbb{P}_1(X_0 = 0) = \mu\delta_{0,0} + \lambda\delta_{1,0} = \mu$$

which implies that $\mu p_t(0,0) + \lambda p_t(1,0) = \mu$ for all $t \in \mathbb{R}_+$. Therefore $\lambda p_t(1,0) = \mu(1 - p_t(0,0))$ and inserting this into (3.18) yields

$$p_t'(0,0) = \mu(1 - p_t(0,0)) - \lambda p_t(0,0) = \mu - (\lambda + \mu)p_t(0,0). \tag{3.19}$$

If we set $g(t) := p_t(0,0) - \frac{\mu}{\lambda+\mu}$, then $g'(t) = -(\lambda+\mu)g(t)$. This is a first-order linear homogeneous differential equation, whose solutions are $g(t) = ce^{-\int_0^t \mathrm{d}s\,(\lambda+\mu)} = ce^{-(\lambda+\mu)t}$ for $t \in \mathbb{R}_+$. It follows that $p_t(0,0) = ce^{-(\lambda+\mu)t} + \frac{\mu}{\lambda+\mu}$ and for $t = 0$, this forces $c = \frac{\lambda}{\lambda+\mu}$. We thus obtain the following entries of P_t:

$$\begin{aligned} p_t(0,0) &= \frac{\lambda}{\lambda+\mu}e^{-(\lambda+\mu)t} + \frac{\mu}{\lambda+\mu} \\ p_t(0,1) &= 1 - p_t(0,0) = -\frac{\lambda}{\lambda+\mu}e^{-(\lambda+\mu)t} + \frac{\lambda}{\lambda+\mu} \\ p_t(1,0) &= \frac{\mu}{\lambda}(1 - p_t(0,0)) = \frac{\mu}{\lambda}p_t(0,1) = -\frac{\mu}{\lambda+\mu}e^{-(\lambda+\mu)t} + \frac{\mu}{\lambda+\mu} \\ p_t(1,1) &= 1 - p_t(1,0) = \frac{\mu}{\lambda+\mu}e^{-(\lambda+\mu)t} + \frac{\lambda}{\lambda+\mu}. \end{aligned}$$

We can also use the forward Chapman-Kolmogorov equation to deduce $(P_t)_{t\in\mathbb{R}_+}$, which is often easier to solve in concrete examples:

$$\begin{aligned} p_t'(0,0) &= p_t(0,0)q_{0,0} + p_t(0,1)q_{1,0} = -\lambda p_t(0,0) + \mu p_t(0,1) \\ p_t'(1,0) &= p_t(1,0)q_{0,0} + p_t(1,1)q_{1,0} = -\lambda p_t(1,0) + \mu p_t(1,1). \end{aligned}$$

Since $p_t(0,1) = 1 - p_t(0,0)$, we find $p_t'(0,0) = -\lambda p_t(0,0) + \mu(1 - p_t(0,0)) = \mu - (\lambda + \mu)p_t(0,0)$. This equation coincides with (3.19), and we then proceed as before.

Example 3.42. [Sha22] We want to model the number of infected individuals at the initial stage of a pandemic, when the population is not yet taking preventive measures such as masking or distancing. We treat each new infection as the "birth" of a new individual, which allows us to represent the evolution of the infected population using a so-called *pure-birth process*.

Intuitively, the spread of the disease is closely related to the number of currently infected patients who have the ability to infect more people (i.e. give birth). As more and more people get infected, the time interval before the infection (or birth) of the next patient becomes shorter and shorter. Consequently, the infection (or birth) rate is not constant but increases with the number of infected individuals.

Let us now define what a *pure-birth process* is. Let $(X_t)_{t\in\mathbb{R}_+}$ denote the number of individuals (or infected patients) in the population at time t. A pure-birth process is a special case of a continuous-time Markov process that allows only one type of state transition, namely a "birth", which increases the current value of X_t by one. The birth rate at each state depends only on the current state. We say that X_t is a pure-birth process with rates λ_n, $n \in \mathbb{N}$, if for all $t \geq 0$ and for all sufficiently small $h > 0$:

- $X_{t+h} \geq X_t$,
- $\mathbb{P}(X_{t+h} - X_t = 1 \mid X_t = n) = \lambda_n h + o(h)$,
- $\mathbb{P}(X_{t+h} - X_t = 0 \mid X_t = n) = 1 - \lambda_n h + o(h)$,
- $\mathbb{P}(X_{t+h} - X_t > 1 \mid X_t = n) = o(h)$.

Intuitively, over a very small interval h, the probability of exactly one birth (or infection) is proportional to h, the probability of no birth is close to 1, and the probability of more than one birth is negligible compared to h.

To determine the generator matrix G of a pure-birth process, we first specify the state space as $E := \mathbb{N}$. It is clear that for all $i < j$ in E with $j - i > 1$, we have $q_{i,j} = q(i)\pi_{i,j} = q(i)\mathbb{P}_i(Y_1 = j) = 0$ since the state chain (skeleton) cannot jump more than one birth at a time. Moreover, by the first bullet point above, $(X_t)_{t\in\mathbb{R}_+}$ is increasing in t, so for all $i > j$ in E we also have $q_{i,j} = 0$. For the diagonal entries, we have $q_{i,i} = -q(i) = -\lambda_i$ for all $i \in E$, since by Theorem 3.38 $q(i)$ is the total rate at which the process leaves state i, which coincides with the rate of the only possible transition out of i, namely λ_i. For the remaining case $j - i = 1$, we have $q_{i,j} = q(i)\mathbb{P}_i(Y_1 = j) = q(i) = \lambda_i$ since the state chain always jumps one birth at a time. Therefore, the generator matrix G of a pure-birth process has the form

$$G = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \cdots \\ 0 & -\lambda_1 & \lambda_1 & 0 & \cdots \\ 0 & 0 & -\lambda_2 & \lambda_2 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots \end{pmatrix}.$$

We now introduce an important example of a pure-birth process, called the *Yule process*. A Yule process is a special case of a pure-birth process that assumes each individual in the population gives birth independently at a constant rate λ. Using the previous notation, the birth rate in state n is therefore $\lambda_n = \lambda n$. That is, the total birth rate is proportional to the current population size. This makes the Yule process well suited for modelling a pandemic's spread during the initial stage, when individuals have not yet changed their behavior and each infected person has roughly the same contact rate.

By the bullet-point definition above, for all $n \in \mathbb{N}$, $t \in \mathbb{R}_+$ and for all sufficiently small $h > 0$:

$$\begin{aligned}
&\mathbb{P}(X_{t+h} = n) \\
&= \mathbb{P}(X_{t+h} - X_t = 1, X_t = n-1) + \mathbb{P}(X_{t+h} - X_t = 0, X_t = n) \\
&= \mathbb{P}(X_t = n-1)\mathbb{P}(X_{t+h} - X_t = 1 \mid X_t = n-1) + \mathbb{P}(X_t = n)\mathbb{P}(X_{t+h} - X_t = 0 \mid X_t = n) \\
&= \mathbb{P}(X_t = n-1)((n-1)\lambda h + o(h)) + \mathbb{P}(X_t = n)(1 - n\lambda h + o(h)).
\end{aligned}$$

Subtracting $\mathbb{P}(X_t = n)$ from both sides, dividing by $h > 0$ and taking the limit as $h \to 0$ gives the following first-order linear inhomogeneous differential equation:

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}(X_t = n) = (n-1)\lambda\mathbb{P}(X_t = n-1) - n\lambda\mathbb{P}(X_t = n).$$

To solve this differential equation, we assume the initial condition $X_0 = n_0$. Using the integrating-factor method, we obtain

$$\mathbb{P}(X_t = n) = (C + \int_0^t \mathrm{d}u\,(n-1)\lambda\mathbb{P}(X_u = n-1)e^{n\lambda u})e^{-n\lambda t}.$$

Applying the initial condition $\mathbb{P}(X_0 = n_0) = 1$ gives $\mathbb{P}(X_t = n_0) = e^{-n_0\lambda t}$. Moreover, since $\mathbb{P}(X_0 = n) = 0$ for all $n > n_0$, we deduce that $\mathbb{P}(X_t = n_0 + 1) = n_0 e^{-n_0\lambda t}(1 - e^{-\lambda t})$. Iterating this for $n_0 + 2, n_0 + 3, \ldots$ produces the explicit solution

$$\mathbb{P}(X_t = n) = \binom{n-1}{n_0-1} e^{-n_0\lambda t}(1 - e^{-\lambda t})^{n-n_0}, \quad n = n_0, n_0 + 1, \ldots$$

That is, $X_t \sim \text{NegBin}(k = n_0, p = e^{-\lambda t})$, and in the special case $n_0 = 1$, $X_t \sim \text{Geom}(p = e^{-\lambda t})$.

Note that in a pure-birth process (such as the Yule process), the holding times $D_n = J_n - J_{n-1}$ ($n \geq 1$) represent the inter-birth intervals between the $(n-1)$-th and n-th births. By Theorem 3.38, and since the birth rate throughout the interval preceding the $(n+1)$-th birth is $\lambda_n = \lambda n$, we have $D_{n+1} \sim \text{Exp}(\lambda n)$.

Another interpretation is obtained by noting that when the process is in state n, each of the n individuals independently gives birth at rate λ. Therefore, the next birth occurs at the minimum of n exponential clocks, i.e. $D_{n+1} = \min\{Z_1, \ldots, Z_n\}$ where $Z_1, \ldots, Z_n$ are i.i.d. $\text{Exp}(\lambda)$ random variables. To verify the equivalence of these two interpretations, observe that if $Z_1, \ldots, Z_n$ are i.i.d. $\text{Exp}(\lambda)$ random variables,

$$\begin{aligned}
\mathbb{P}(\min\{Z_1, \ldots, Z_n\} \leq z) &= 1 - \mathbb{P}(\min\{Z_1, \ldots, Z_n\} > z) \\
&= 1 - \mathbb{P}(Z_1 > z, \ldots, Z_n > z) \\
&= 1 - \mathbb{P}(Z_1 > z)^n \\
&= 1 - e^{-\lambda n z} \\
&= \mathbb{P}(W \leq z)
\end{aligned}$$

where W denotes an exponential random variable with rate λn. Therefore, the two interpretations are indeed equivalent.

We will now simulate the infection (or birth) times of a Yule process in R using $D_{n+1} \sim \text{Exp}(\lambda n)$. The following `Yule_sim` function simulates the process of infections (or births) up to time T_{end}, starting from the initial state n_0 (see Listing 3.1).

Listing 3.1: Simulation of a Yule process

```r
# lambda : individual birth rate
# n_0 : initial population size
# T_end : simulate until this time
Yule_sim <- function(lambda, n_0, T_end) {

 Z <- c() # vector of inter-birth holding times: Z[i] = D_{n_0+i}
 n <- c() # vector of population size after each birth: n[i] = n_0 + i
 i <- 0 # counter for the number of births that have occurred

 # main simulation loop (terminates via break condition)
 while (TRUE) {

  # next birth index
  i <- i + 1

  # birth rate when current population size is n_0 + (i - 1)
  rate <- (n_0 + i - 1) * lambda

  # sample the inter-birth holding time from Exp(rate)
  wait_time <- rexp(1, rate = rate)

  # if adding this holding time would exceed T_end, exit loop
  if (sum(Z) + wait_time > T_end) break

  # otherwise, record this valid inter-birth holding time
  Z[i] <- wait_time
```

```
    # after this i-th birth, population increases to n_0 + i
    n[i] <- n_0 + i
  }

  # return the inter-birth times and the population sizes
  list(Z = Z, n = n)
}
```

We now use the `Yule_sim` function to repeat the Yule process 100 times (see Listing 3.2). In each simulation, the infection (or birth) rate for a patient is $\lambda = 0.2$ and the initial number of patients is $n_0 = 10$. We stop simulating after $T_{\text{end}} = 10$ and plot the trajectory (see Figure 3.2).

Listing 3.2: Multiple simulated Yule trajectories

```
set.seed(1234) # fix random seed for reproducibility
sim <- 100 # number of simulated Yule process trajectories

# first simulation (used to initialize the plot)
result <- Yule_sim(0.2, 10, 10)

# plot the first trajectory
# x-axis: birth times = cumulative sum of inter-birth intervals
# y-axis: population size after each birth
par(mar = c(5, 4, 1, 1))
plot(
  c(0, cumsum(result$Z)), # birth times
  c(10, result$n), # population sizes
  type = "l", # line plot
  col = 1, # color for first line (black)
  ylim = c(0, 150), # fix y-scale for all runs
  xlab = "Time",
  ylab = "Population"
)

# simulate and add remaining trajectories
for (i in 2:sim) {

  # generate another Yule process trajectory
  result <- Yule_sim(0.2, 10, 10)

  # add its line to the existing plot
  lines(
    c(0, cumsum(result$Z)), # birth times
    c(10, result$n), # population sizes
    type = "l", # line plot
    col = i # different color for each simulation
  )
}
```

Each colored line in Figure 3.2 shows one simulated trajectory of the Yule process, starting with 10 initially infected individuals and evolving up to time 10. The number of infected individuals only increases and the spread accelerates over time because the infection rate is proportional to the number of currently infected people. The trajectories diverge quite a bit since small random differences in early infection events lead to large differences in outbreak size later. Overall, Figure 3.2 visualizes the exponential growth of the Yule process.

Next, we explore the influence of the infection rate on X_t, the number of infected individuals at time t, caused by mutation of the virus. We compare $\lambda = 0.1, 0.15, 0.2, 0.25$ and set $T_{\text{end}} = 2$. For each λ value, we run 1000 simulations of the Yule process starting with 5 infected individuals (see Listing 3.3). In each simulation, we record the final outbreak size at time $T_{\text{end}} = 2$ and then plot histograms showing the distribution of these final outbreak sizes (see Figure 3.3).

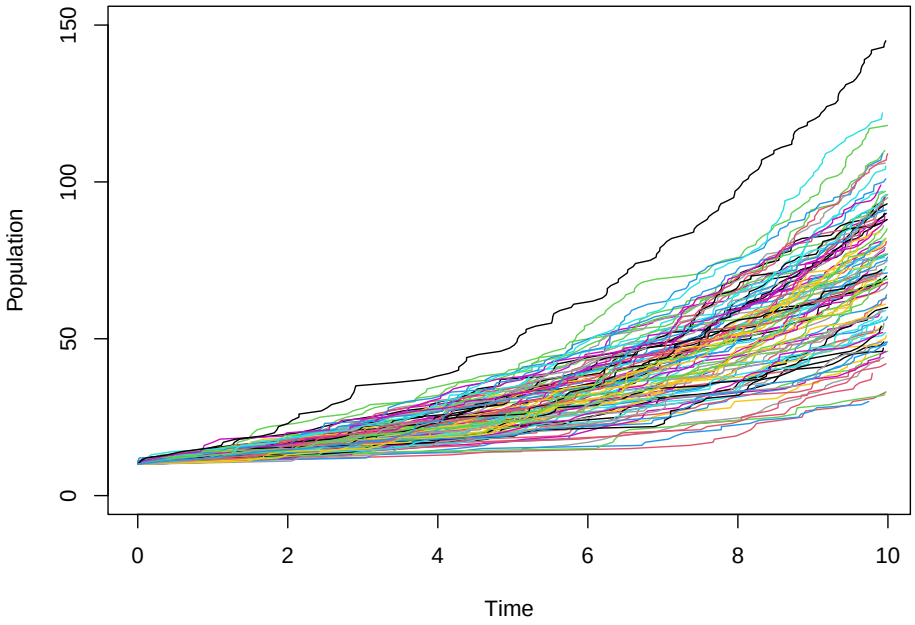

Figure 3.2: Simulated Yule process trajectories for $\lambda = 0.2$, $n_0 = 10$ and $T_{\text{end}} = 10$.

Listing 3.3: Influence of λ for $T_{\text{end}} = 2$

```r
# function that plots histograms of final outbreak sizes for a given T_end
Yule_histo <- function(T_end) {

  set.seed(123) # fix random seed for reproducibility
  lambda0 <- c(0.1, 0.15, 0.2, 0.25) # define different lambda values to test
  sim <- 1000 # number of simulations for each lambda

  par(mfrow = c(2, 2)) # set up 2x2 grid for plotting histograms

  # loop through each lambda value
  for (i in 1:length(lambda0)) {

    # initialize vector to store final population sizes
    records <- rep(0, sim)

    # run multiple simulations for current lambda
    for (j in 1:sim) {

      # simulate Yule process with current parameters
      result <- Yule_sim(lambda0[i], 5, T_end)

      # if no births occurred, population remains at initial n_0 = 5
      # otherwise, take the last population value from simulation
      records[j] <- ifelse(length(result$n) > 0, tail(result$n, 1), 5)
    }

    # plot histogram of final population sizes
    hist(records, freq = TRUE, xlab = "Population",
        main = paste0("lambda = ", lambda0[i]), breaks = 10)
  }
}

# plot histograms of final outbreak sizes for T_end = 2
Yule_histo(2)
```

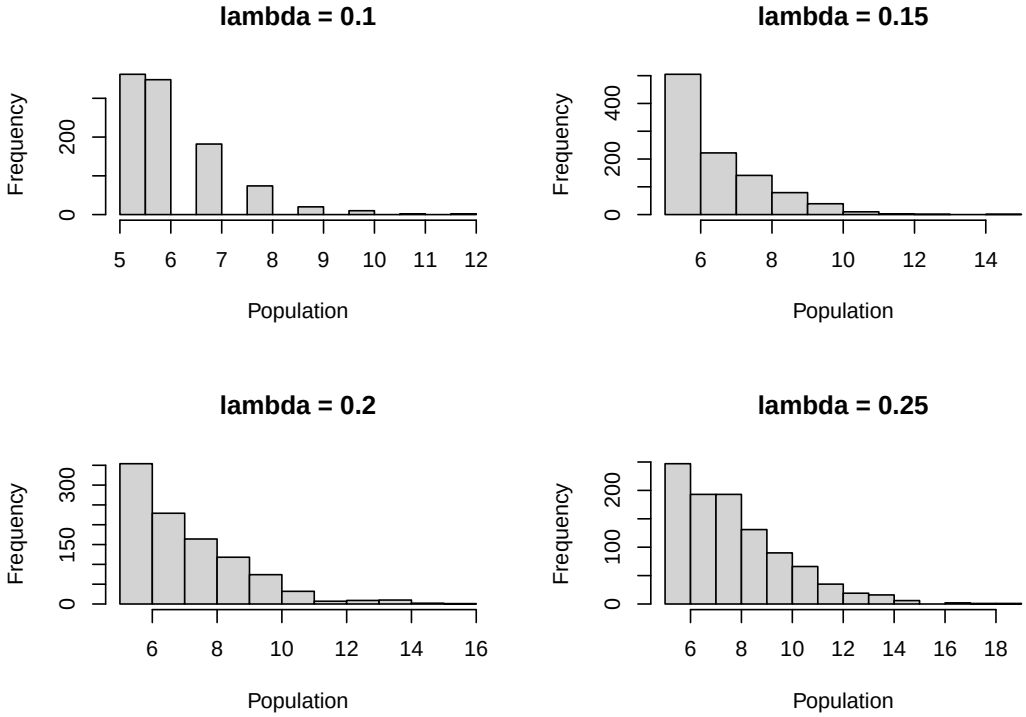

Figure 3.3: Distribution of final population sizes at $T_{\text{end}} = 2$ for different values of λ.

We may conclude from Figure 3.3 that the infection rate does not seem to have much influence on X_t. Let us now generate the histogram plots for $T_{\text{end}} = 10$ using the same `Yule_histo` function (see Listing 3.4 and Figure 3.4).

Listing 3.4: Influence of λ for $T_{\text{end}} = 10$

```
# plot histograms of final outbreak sizes for T_end = 10
Yule_histo(10)
```

In Figure 3.4, there now seems to be a clear difference among the four plots, especially with respect to the x-axis. What if we generate the histogram plots for $T_{\text{end}} = 20$ using again the `Yule_histo` function (see Listing 3.5 and Figure 3.5)?

Listing 3.5: Influence of λ for $T_{\text{end}} = 20$

```
# plot histograms of final outbreak sizes for T_end = 20
Yule_histo(20)
```

The difference becomes even larger in Figure 3.5. As time increases, even a small difference in the individual infection rate caused by mutation can result in a huge difference in the total number of infected individuals.

Finally, let us verify that the distribution of X_t follows a negative binomial distribution with parameters $k = n_0$ and $p = e^{-\lambda t}$. We set $\lambda = 0.2$, $n_0 = 5$ and $T_{\text{end}} = 20$, and we generate a plot of the empirical distribution of the simulated final population sizes (see Listing 3.6). On the same graph, we overlay the theoretical negative binomial distribution with parameters $k = n_0$ and $p = e^{-\lambda T_{\text{end}}}$ to see how well the simulated outcomes match the theoretical prediction (see Figure 3.6).

Listing 3.6: Validation of the negative binomial distribution

```
sim <- 10000 # number of simulated Yule process trajectories
lambda <- 0.2 # individual birth rate
n_0 <- 5 # initial population size
T_end <- 20 # simulate until this time
```

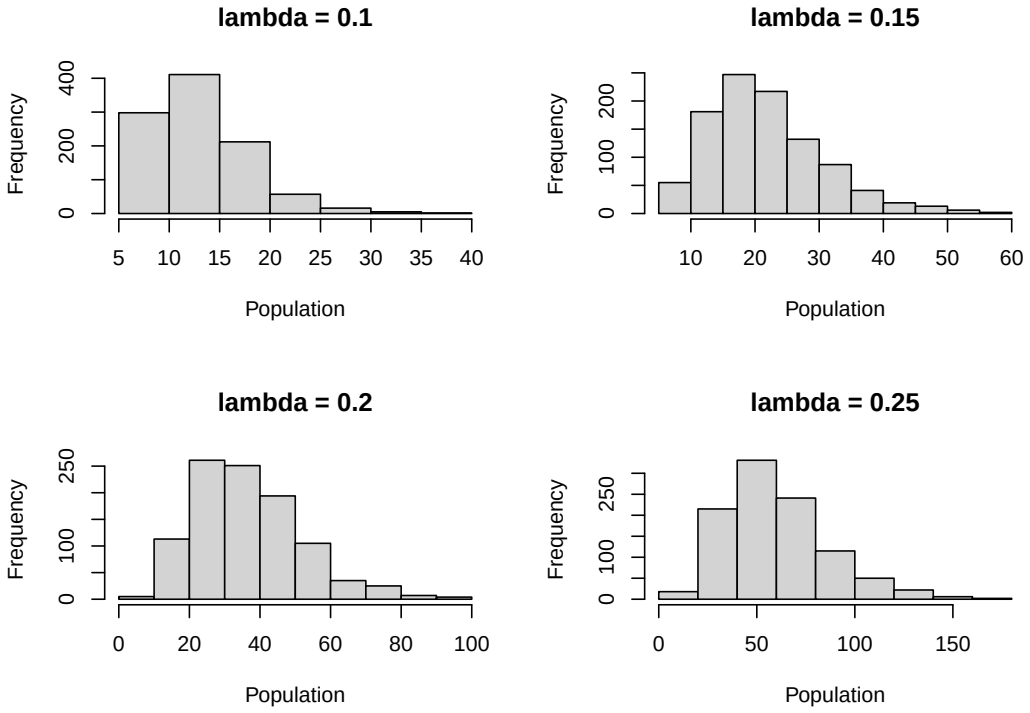

Figure 3.4: Distribution of final population sizes at $T_{\text{end}} = 10$ for different values of λ.

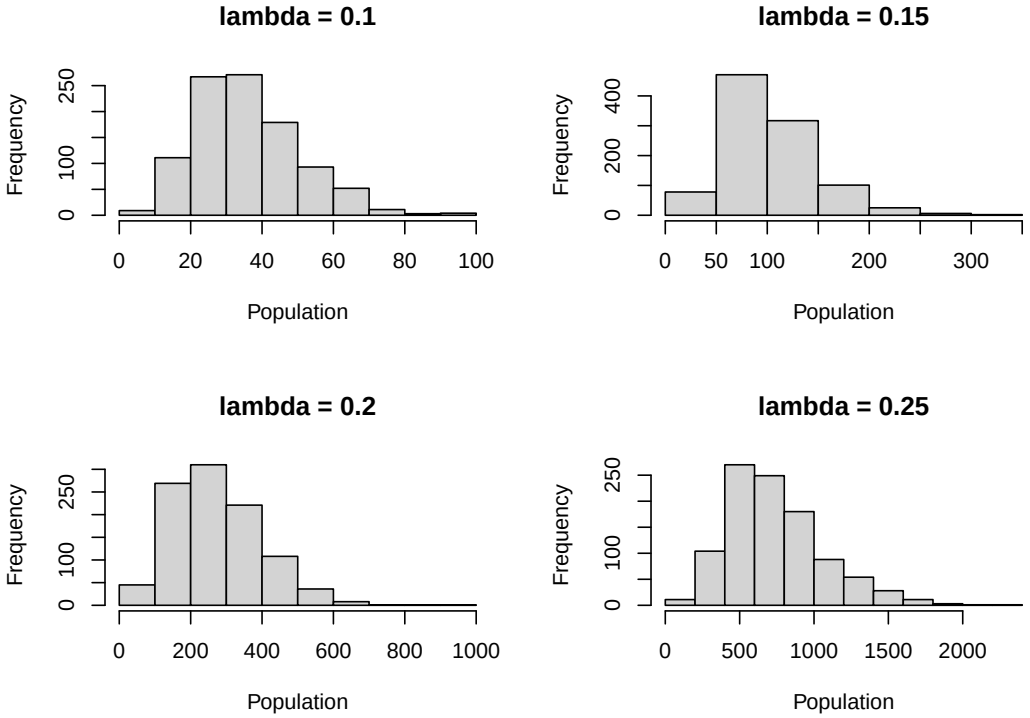

Figure 3.5: Distribution of final population sizes at $T_{\text{end}} = 20$ for different values of λ.

```r
# theoretical parameter for the negative binomial distribution
p <- exp(-T_end * lambda)

# vector to store final population sizes from each simulation
records <- rep(0, sim)

# run multiple simulations of the Yule process
for (j in 1:sim) {

 # simulate Yule process
 result <- Yule_sim(lambda, n_0, T_end)

 # if no births occurred, population remains at initial n_0 = 5
 # otherwise, take the last population value from simulation
 records[j] <- ifelse(length(result$n) > 0, tail(result$n, 1), n_0)
}

# plot histogram of final population sizes (density instead of raw counts)
par(mar = c(5, 4, 1, 1))
hist(records, probability = TRUE, xlab = "Population", breaks = 50,
    main = "")

# overlay theoretical negative binomial pmf
# (dnbinom() starts at 0 but X_t starts at n_0 -> shift)
x <- seq(n_0, max(records), 1)
lines(x, dnbinom(x - n_0, size = n_0, prob = p), col = 2)
```

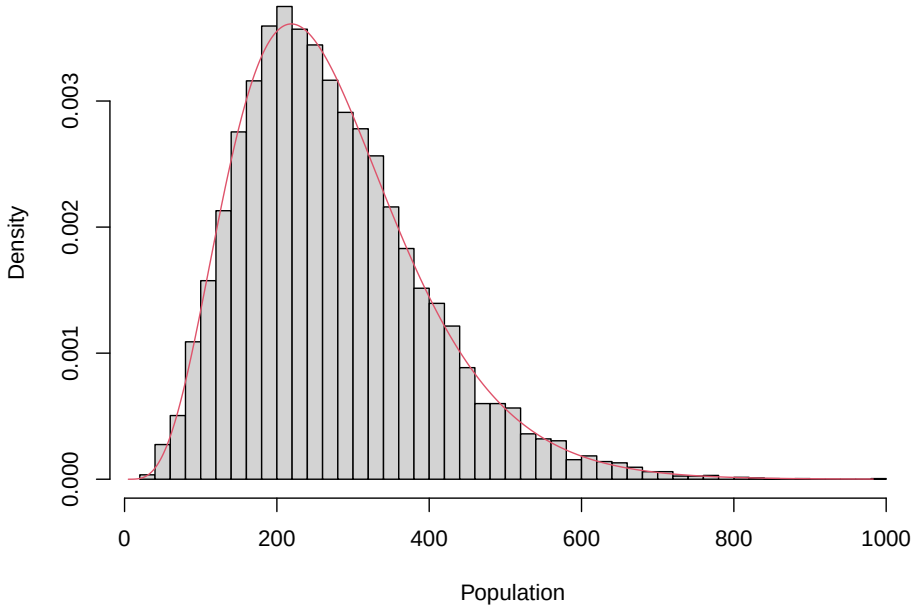

Figure 3.6: Empirical distribution of final population size at $T_{\text{end}} = 20$ with overlaid negative binomial pmf.

The histogram bars in Figure 3.6 closely follow the red curve, which shows that the simulated Yule process indeed produces a negative-binomial distribution of outbreak sizes.

4 The Minimal Markov Process Associated with a Generator

4.1 Construction of the Minimal Markov Process

As already mentioned, our study of Markov processes taking values in a countable state space (equipped with the discrete topology) concerns only what happens before a possible explosion. We wish to view the part of the trajectory prior to a possible explosion as a Markov process itself, so that we can apply the results already obtained (e.g. the simple and strong Markov property). To do this, we introduce a cemetery state and absorb the process there from the explosion time onward (if it occurs). Such processes are called minimal processes, and they form a class of Markov processes with countable state space (equipped with the discrete topology) large enough to include most real-world applications (queueing theory, biological models, statistical mechanics, etc.).

In Section 3.2.1, we began with a right-continuous Markov process and derived its infinitesimal generator. We now reverse this construction by starting from an infinitesimal generator (which must first be defined) and uniquely constructing the corresponding right-continuous minimal Markov process. Note that this construction in particular shows the existence of right-continuous Markov processes. We will also see that the infinitesimal generator completely determines the probability distribution of the minimal process.

Definition 4.1 (Infinitesimal generator)**.** Let E be a countable space equipped with the discrete topology, and hence endowed with the σ-algebra $\mathcal{P}(E)$. A matrix $G = (q_{i,j})_{i,j\in E}$ is an *infinitesimal generator* if and only if for all $i, j \in E$,

$$q_{i,j} \in \mathbb{R}_+ \text{ if } i \neq j, \quad q_{i,i} \in \mathbb{R}_- \quad \text{and} \quad -q_{i,i} = \sum_{j\in E\setminus\{i\}} q_{i,j}.$$

To an infinitesimal generator $G = (q_{i,j})_{i,j\in E}$ we associate the following two objects:

(a) A family $(q(i))_{i\in E}$ of nonnegative real numbers given by $q(i) := -q_{i,i}$ for $i \in E$, called the family of holding time parameters.

(b) A matrix $Q = (\pi_{i,j})_{i,j\in E}$ defined for all $i, j \in E$ by

$$\pi_{i,j} := \begin{cases} -\frac{q_{i,j}}{q_{i,i}}, & \text{if } i \neq j \text{ and } q_{i,i} < 0, \\ 0, & \text{if } i = j \text{ and } q_{i,i} < 0, \\ \delta_{i,j}, & \text{if } q_{i,i} = 0. \end{cases}$$

Note that Q is a transition matrix since for every $i \in E$,

$$\sum_{j\in E} \pi_{i,j} = \begin{cases} -\frac{1}{q_{i,i}} \sum_{j\in E\setminus\{i\}} q_{i,j} = 1, & \text{if } q_{i,i} < 0, \\ \sum_{j\in E} \delta_{i,j} = 1, & \text{if } q_{i,i} = 0. \end{cases}$$

Q is called the *transition matrix of the skeleton*. Observe that if $q_{i,i} = 0$, then $\pi_{i,i} = 1$ (i.e. the state i is absorbing for the Markov chain with transition matrix Q) and if $q_{i,i} < 0$, then $\pi_{i,i} = 0$.

Notice that this definition is closely related to Definition 3.37, where the holding times and transition probabilities were obtained from an existing right-continuous Markov process. Here, however, we take these quantities as starting data. We will later show that the two definitions are compatible, in the sense that every right-continuous Markov process yields such a generator, and conversely every generator gives rise to a unique minimal right-continuous Markov process. We now give the formal definition of the minimal process associated with a given infinitesimal generator.

Definition 4.2 (Minimal process)**.** Let E be a countable space equipped with the discrete topology, and hence endowed with the σ-algebra $\mathcal{P}(E)$. Let $G = (q_{i,j})_{i,j\in E}$ be an infinitesimal generator on E. Choose a cemetery point $\partial \notin E$ and set $E^* := E \cup \{\partial\}$. Then E^* is again a countable space

and we equip it with the discrete topology (and the corresponding σ-algebra $\mathcal{P}(E^*)$) by extending the discrete metric on E via $d(i,\partial)=1$ and $d(\partial,\partial)=0$ for all $i\in E$.

A Markov process taking values in E^*, denoted

$$(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\in\mathbb{R}_+},\mathbf{X}=(X_t)_{t\in\mathbb{R}_+},(P_t^*)_{t\in\mathbb{R}_+},\mathbb{P}_\mu,\mu\in\mathcal{M}_1(E^*))$$

with $P_t^*:=(p_t^*(i,j))_{i,j\in E^*}$, and right-continuous, is a *minimal process* associated with the generator G if and only if it satisfies the following conditions:

(a) $\mathbf{X}$ is right-continuous in E^* (in the sense of Definition 3.23).

(b) The cemetery point ∂ is absorbing, i.e. for all $t\in\mathbb{R}_+$, we have $p_t^*(\partial,i)=0$ for all $i\in E$ and $p_t^*(\partial,\partial)=1$.

(c) The process explodes if and only if it reaches ∂ (where it is then definitely absorbed). That is, $\zeta=\inf\{t\in\mathbb{R}_+ : X_t=\partial\}$ (with the convention $\inf\emptyset:=\infty$) and $X_t=\partial$ for all $t\geq\zeta$. This implies that $X_t\in E$ for every $t<\zeta$.

(d) Let $G^*=(q_{i,j}^*)_{i,j\in E^*}$ be the infinitesimal generator of the process, as in Definition 3.37. Then $q_{i,j}=q_{i,j}^*$ for all $i,j\in E$. In other words, when restricted to E, the infinitesimal generator G^* of the minimal process coincides with the infinitesimal generator G.

Remark 4.3. The conditions (a)-(d), combined with the results of the preceding sections (which apply to the right-continuous Markov process $\mathbf{X}$ on E^* since E^* is countable and equipped with the discrete topology), have the following immediate consequences:

(i) Since ∂ is absorbing by (b), we have $\mathbb{P}_\partial(J_1=\infty)=1$ where $J_1\sim\mathrm{Exp}(q(\partial))$ under $\mathbb{P}_\partial$ (see Proposition 3.36). By Lemma 3.35, this implies that $q(\partial)=0$. Hence, by Definition 3.37, we obtain $q_{\partial,\partial}^*=-q(\partial)=0$ and $q_{\partial,i}^*=q(\partial)\pi_{\partial,i}=0$ for all $i\in E$. Furthermore, since the explosion time is exactly the first hitting time of ∂ by (c), the process cannot reach ∂ from a state $i\in E$ without making an infinite number of jumps. In particular, the process cannot jump directly from i to ∂ and therefore $q_{i,\partial}^*=q(i)\pi_{i,\partial}=q(i)\mathbb{P}_i(Y_1=\partial)=0$ for all $i\in E$.

(ii) For all $t\in\mathbb{R}_+$ and all $i,j\in E$, set $p_t(i,j):=p_t^*(i,j)$ and write $P_t:=(p_t(i,j))_{i,j\in E}$. Since the process stays in E until the possible explosion time ζ, we see that for all $t\in\mathbb{R}_+$ and all $i,j\in E$,

$$p_t(i,j)=p_t^*(i,j)=\mathbb{P}_i(X_t=j)=\mathbb{P}_i(X_t=j,\zeta>t). \tag{4.1}$$

Define $(P_t^{\min,*})_{t\in\mathbb{R}_+}$ by $p_t^{\min,*}(i,j):=\mathbb{P}_i(X_t=j,\zeta>t)$ for all $t\in\mathbb{R}_+$ and $i,j\in E^*$. Applying Theorem 3.39 to the process $\mathbf{X}$ on E^*, we obtain that $(P_t^{\min,*})_{t\in\mathbb{R}_+}$ is a sub-Markovian semigroup which is the minimal nonnegative solution of the Chapman-Kolmogorov equations associated with the generator G^*. Restricting this semigroup to E and using (4.1) together with (d) then yields that $(P_t)_{t\in\mathbb{R}_+}$ is a sub-Markovian semigroup which is the minimal nonnegative solution of the Chapman-Kolmogorov equations associated with the generator G, i.e.

$$\tfrac{\mathrm{d}}{\mathrm{d}t}P_t=GP_t=P_tG \text{ for all } t\in\mathbb{R}_+ \quad\text{and}\quad P_0=\mathrm{Id}_E.$$

(iii) Since the Chapman-Kolmogorov equations above only depend on G and the minimal nonnegative solution is unique, we deduce that the sub-Markovian semigroup $(P_t)_{t\in\mathbb{R}_+}$ is entirely determined by G. Moreover, $(P_t)_{t\in\mathbb{R}_+}$ determines the Markovian semigroup $(P_t^*)_{t\in\mathbb{R}_+}$ via the equation

$$p_t^*(i,\partial)=1-\sum_{j\in E}p_t(i,j) \quad \text{for all } t\in\mathbb{R}_+ \text{ and } i\in E$$

and the identities $p_t^*(\partial,\partial)=1$, $p_t^*(\partial,i)=0$ and $p_t^*(i,j)=p_t(i,j)$ for all $t\in\mathbb{R}_+$ and $i,j\in E$. Therefore the Markovian semigroup $(P_t^*)_{t\in\mathbb{R}_+}$ is entirely determined by G.

Notation 4.4. Since the process

$$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, \mathbf{X} = (X_t)_{t\in\mathbb{R}_+}, (P_t^*)_{t\in\mathbb{R}_+}, \mathbb{P}_\mu, \mu \in \mathcal{M}_1(E^*))$$

is entirely determined by G, and since ∂ is absorbing, we prefer to specify the minimal process associated with G without referring explicitly to the cemetery point, writing simply

$$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, \mathbf{X} = (X_t)_{t\in\mathbb{R}_+}, G, \mathbb{P}_\mu, \mu \in \mathcal{M}_1(E)).$$

Note that the cemetery state plays no role in the initial distribution μ of the process. With this notation, we must keep in mind that the corresponding Markov process actually takes values in $E^* = E \cup \{\partial\}$.

Now that we have given the general definition of an infinitesimal generator and a minimal process, we will construct a minimal Markov process associated with an arbitrary generator. This construction is quite instructive, as it shows how an infinitesimal generator encodes the dynamics of its associated minimal process, which is also useful in applications where models are described only through their generator.

Let E be a countable space equipped with the discrete topology, and hence endowed with the σ-algebra $\mathcal{P}(E)$. Let $G = (q_{i,j})_{i,j\in E}$ be an infinitesimal generator on E, according to Definition 4.1. Denote by $(q(i))_{i\in E}$ the family of holding time parameters and by $Q := (\pi_{i,j})_{i,j\in E}$ the transition matrix of the skeleton, which are both objects associated with G (see Definition 4.1). We again introduce a cemetery point $\partial \notin E$ and set $E^* := E \cup \{\partial\}$, viewed as a countable space equipped with the discrete topology and the corresponding σ-algebra $\mathcal{P}(E^*)$. We extend Q to E^* by declaring ∂ to be an absorbing state, i.e. by setting $\pi_{\partial,i} := 0$ for all $i \in E$, $\pi_{\partial,\partial} := 1$ and $q(\partial) := 0$ (so that the holding time at ∂ is infinite). We also set $\pi_{i,\partial} := 0$ for all $i \in E$ because we want ∂ to be reached only by explosion (infinitely many accumulating jumps) and not by a single transition of the skeleton chain. We then continue to denote by Q the extended transition matrix $(\pi_{i,j})_{i,j\in E^*}$.

The construction we give below presupposes the existence of the following two objects (see Section I.1.c of [Duq23] for their construction):

(a) A Markov chain taking values in E^*,

$$(\Omega, \mathcal{F}, \mathbf{Y} = (Y_n)_{n\in\mathbb{N}}, Q, \mathbb{P}_\mu, \mu \in \mathcal{M}_1(E^*)),$$

with respect to its natural filtration $\mathcal{F}_n^{\mathbf{Y}} = \sigma(Y_0, \dots, Y_n)$ for $n \in \mathbb{N}$.

(b) A sequence of $\mathcal{F}$-measurable random variables $\mathbf{e}_n : \Omega \to (0, \infty)$, $n \geq 1$, such that for every initial distribution $\mu \in \mathcal{M}_1(E^*)$, under $\mathbb{P}_\mu$ the random variables $(\mathbf{e}_n)_{n\geq 1}$ are i.i.d. exponential with parameter 1 and independent of $\mathbf{Y}$.

We use the convention $\frac{a}{0} := \infty$ for all $a \in (0, \infty)$, so that $\frac{\mathbf{e}_n(\omega)}{0} = \infty$ for every $n \geq 1$ and $\omega \in \Omega$. This allows us to define the sequence of $\mathcal{F}$-measurable random variables $D_n : \Omega \to (0, \infty]$, $n \geq 1$, by

$$D_n(\omega) := \tfrac{\mathbf{e}_n(\omega)}{q(Y_{n-1}(\omega))} \in (0, \infty].$$

Observe that this construction already appeared in the proof of Theorem 3.38, where we showed that under $\mathbb{P}_{i_{n-1}}$, the ratio $\frac{\mathbf{e}_n}{q(i_{n-1})}$ has an (extended) exponential distribution with parameter $q(i_{n-1})$. We then set $J_0 := 0$ and for every $n \geq 1$,

$$J_n := D_1 + \cdots + D_n \in (0, \infty]$$

with the convention that $a + \infty := \infty$ for all $a \in [0, \infty]$. Finally, we define

$$N := \sup\{n \in \mathbb{N} : J_n < \infty\} \in \mathbb{N} \cup \{\infty\} \quad \text{and} \quad \zeta := \sup_{n\in\mathbb{N}} J_n \in [0, \infty].$$

Notice that D_n, J_n, N, and ζ are simply functions on Ω and at this stage they are defined independently of any process. Since $(J_n)_{n\in\mathbb{N}}$ is increasing, the interval $[0, \zeta)$ can be written as

$$[0, \zeta) = \bigsqcup_{n\in\mathbb{N}} [J_n, J_{n+1}). \tag{4.2}$$

We then construct a process $\mathbf{X} = (X_t)_{t\in\mathbb{R}_+} : \Omega \to (E^*)^{\mathbb{R}_+}$ as follows:

- If $t \in [0, \zeta)$, then by (4.2) there exists $n \in \mathbb{N}$ such that $t \in [J_n, J_{n+1})$. We then set $X_t := Y_n$.
- If $t \in [\zeta, \infty)$, we set $X_t := \partial$.

Thus, for every $t \in \mathbb{R}_+$ and every $i \in E$, we have $\{X_t = i\} = \bigsqcup_{n\in\mathbb{N}}\{J_n \le t < J_{n+1}\} \cap \{Y_n = i\} \in \mathcal{F}$, since J_n is $\mathcal{F}$-measurable and $\{Y_n = i\} \in \mathcal{F}_n^{\mathbf{Y}} \subseteq \mathcal{F}$ for every $n \in \mathbb{N}$. Also $\{X_t = \partial\} = \{\zeta \le t\} \in \mathcal{F}$ for every $t \in \mathbb{R}_+$, since ζ is $\mathcal{F}$-measurable. This shows that for every $t \in \mathbb{R}_+$ and every $i \in E^*$, the preimage $X_t^{-1}(\{i\})$ is in $\mathcal{F}$ and therefore $X_t : \Omega \to E^*$ is $(\mathcal{F}, \mathcal{P}(E^*))$-measurable for every $t \in \mathbb{R}_+$. By Lemma 3.2, we may therefore assert that $\mathbf{X} : \Omega \to (E^*)^{\mathbb{R}_+}$ is $(\mathcal{F}, \mathcal{P}(E^*)^{\mathbb{R}_+})$-measurable.

For each $\omega \in \Omega$, the trajectory $t \mapsto X_t(\omega)$ is piecewise constant on the intervals $[J_n(\omega), J_{n+1}(\omega))$ for $n \in \mathbb{N}$ and on $[\zeta(\omega), \infty)$. Hence every trajectory is right-continuous (in the sense of Definition 3.23), and $\mathbf{X}$ is a right-continuous process.

By construction, the process $\mathbf{X}$ (before possible explosion) changes its value exactly at the times $J_n < \infty$. Indeed, since $X_t = Y_n$ for every $t \in [J_n, J_{n+1})$, the process $\mathbf{X}$ (before possible explosion) can only change value at the times J_n. To see that it actually changes value at each time $J_n < \infty$, recall that $D_n = \frac{\mathbf{e}_n}{q(Y_{n-1})}$ is finite if and only if $q(Y_{n-1}) > 0$. In particular, whenever $J_n < \infty$ we must have $q(Y_{n-1}) > 0$. Moreover, for every $i \in E$ with $q(i) > 0$, we imposed $\pi_{i,i} = 0$, i.e. $\mathbb{P}_i(Y_1 = i) = 0$ (see Definition 4.1). Thus, whenever $J_n < \infty$ (and hence $q(Y_{n-1}) > 0$), we have $Y_n \neq Y_{n-1}$ and therefore the process $\mathbf{X}$ changes value at time J_n.

Combining this with the construction of $\mathbf{X}$ and the definitions in Section 3.1.4, we obtain

$$\begin{aligned}
J_{n+1}(\mathbf{X}) &= \inf\{t > J_n(\mathbf{X}) : X_t \neq X_{J_n(\mathbf{X})}\} = J_{n+1}, \quad n \in \mathbb{N},\\
J_0(\mathbf{X}) &= 0 = J_0,\\
\zeta(\mathbf{X}) &= \sup_{n\in\mathbb{N}} J_n(\mathbf{X}) = \sup_{n\in\mathbb{N}} J_n = \zeta,\\
N(\mathbf{X}) &= \sup\{n \in \mathbb{N} : J_n(\mathbf{X}) < \infty\} = \sup\{n \in \mathbb{N} : J_n < \infty\} = N,\\
Y_n(\mathbf{X}) &= X_{J_n(\mathbf{X})} = X_{J_n} = Y_n, \quad 0 \le n \le N(\mathbf{X}) = N,\\
Y_n(\mathbf{X}) &= X_{J_{N(\mathbf{X})}(\mathbf{X})} = X_{J_N} = Y_N = Y_n, \quad n > N(\mathbf{X}) = N,\\
D_{n+1}(\mathbf{X}) &= J_{n+1}(\mathbf{X}) - J_n(\mathbf{X}) = J_{n+1} - J_n = D_{n+1}, \quad 0 \le n \le N(\mathbf{X}) = N,\\
D_{n+1}(\mathbf{X}) &= \infty = D_{n+1}, \quad n > N(\mathbf{X}) = N,
\end{aligned}$$

with the convention $\inf\emptyset := \infty$. Moreover, note that the first hitting time of ∂ coincides with the explosion time, i.e. $\zeta = \inf\{t \in \mathbb{R}_+ : X_t = \partial\}$.

Next, for every $t \in \mathbb{R}_+$, define $p_t^*(i,j) := \mathbb{P}_i(X_t = j)$ for all $i, j \in E^*$ and write $P_t^* = (p_t^*(i,j))_{i,j\in E^*}$. Denote by $\mathcal{F}_t^{\mathbf{X}} := \sigma(X_s : s \in [0,t])$ the natural filtration of $\mathbf{X}$.

Proposition 4.5. *With the notations and definitions introduced above,*

$$(\Omega, \mathcal{F}, (\mathcal{F}_t^{\mathbf{X}})_{t\in\mathbb{R}_+}, \mathbf{X}, (P_t^*)_{t\in\mathbb{R}_+}, \mathbb{P}_\mu, \mu \in \mathcal{M}_1(E^*))$$

is a minimal Markov process associated with G as introduced in Definition 4.2.

Proof. By construction, $\mathbf{X}$ is a minimal process associated with G according to Definition 4.2. Indeed, $\mathbf{X}$ is right-continuous, we defined the state ∂ to be absorbing, and $\mathbf{X}$ explodes if and only if it reaches ∂ (i.e. $\zeta = \inf\{t \in \mathbb{R}_+ : X_t = \partial\}$ and $X_t = \partial$ for all $t \in [\zeta, \infty)$). It is left to verify that G^* restricted to E coincides with G.

By Definition 3.37 (applied to $\mathbf{X}$ on E^*), the infinitesimal generator $G^* = (q_{i,j}^*)_{i,j\in E^*}$ is determined by the parameters $q^*(i)$ of the (extended) exponential distribution of J_1 under $\mathbb{P}_i$ (which we verify below) and the transition probabilities $\pi_{i,j}^* = \mathbb{P}_i(Y_1 = j)$. By construction, under $\mathbb{P}_i$, we have that $J_1 = D_1 = \frac{\mathbf{e}_1}{q(Y_0)}$ follows an (extended) exponential distribution with parameter $q(i)$. Therefore $q^*(i) = q(i)$ for all $i \in E^*$. Moreover, we defined $\mathbf{Y}$ to have transition matrix $Q = (\pi_{i,j})_{i,j\in E^*}$, so $\pi_{i,j}^* = \pi_{i,j}$ for all $i, j \in E^*$.

Definition 3.37 then gives $q_{i,i}^* = -q^*(i) = -q(i)$ for all $i \in E^*$ and since by Definition 4.1 $q(i) = -q_{i,i}$ for all $i \in E$, it follows that $q_{i,i}^* = q_{i,i}$ for all $i \in E$. Therefore G^* restricted to E coincides with G on the diagonal entries.

Moreover, Definition 3.37 yields $q_{i,j}^* = q^*(i)\pi_{i,j}^* = q(i)\pi_{i,j}$ for all $i \neq j$ in E^*. If $q(i) > 0$, then by Definition 4.1 we have $\pi_{i,j} = \frac{q_{i,j}}{q(i)}$ and hence $q_{i,j}^* = q_{i,j}$ for all $i \neq j$ in E. If $q(i) = 0$, then

$q^*_{i,j} = q(i)\pi_{i,j} = 0$ for all $i \neq j$ in E^* and the row-sum condition in Definition 4.1 implies for all $i \in E$ that $\sum_{j\in E\setminus\{i\}} q_{i,j} = -q_{i,i} = q(i) = 0$, which forces $q_{i,j} = 0$ for all $i \neq j$ in E. Thus in this case we also get $q^*_{i,j} = q_{i,j}$ for all $i \neq j$ in E. This shows that G^* restricted to E coincides with G also on the off-diagonal entries. Therefore $\mathbf{X}$ is a minimal process associated with G.

It remains to show that for every $\mu \in \mathcal{M}_1(E^*)$, under $\mathbb{P}_\mu$, $\mathbf{X}$ is a Markov process. Fix $\mu \in \mathcal{M}_1(E^*)$. We first show that for every $m \in \mathbb{N}$, every $i \in E$, every $j \in E^*$, all $t, s \in \mathbb{R}_+$, all $i_0, \ldots, i_{m-1}, i_m = i \in E$ and all $0 = t_0 \leq t_1 \leq \cdots \leq t_m = t$, we have

$$\begin{aligned}&\mathbb{P}_\mu(X_0 = i_0, X_{t_1} = i_1, \ldots, X_{t_{m-1}} = i_{m-1}, X_t = i, X_{t+s} = j)\\ &= \mathbb{P}_\mu(X_0 = i_0, X_{t_1} = i_1, \ldots, X_{t_{m-1}} = i_{m-1}, X_t = i)p^*_s(i,j).\end{aligned} \tag{4.3}$$

Since the process $\mathbf{X}$ is constructed from the two sequences $(D_n)_{n\geq 1}$ and $(Y_n)_{n\in\mathbb{N}}$, we may view it as a certain function Φ of these sequences:

$$\mathbf{X} = \Phi((D_n)_{n\geq 1}, (Y_n)_{n\in\mathbb{N}}).$$

To justify this notation, we check the measurability of the map

$$\begin{aligned}\Phi : (0,\infty]^{\mathbb{N}} \times (E^*)^{\mathbb{N}} &\longrightarrow (E^*)^{\mathbb{R}_+}\\ ((d_n)_{n\geq 1}, (y_n)_{n\in\mathbb{N}}) &\longmapsto \mathbf{x} = (x(t))_{t\in\mathbb{R}_+}\end{aligned}$$

defined by

$$x(t) := \begin{cases} y_n, & \text{if } j_n \leq t < j_{n+1} \text{ for some } n \in \mathbb{N},\\ \partial, & \text{if } t \geq \zeta,\end{cases}$$

where $j_0 := 0$, $j_n := \sum_{k=1}^n d_k$ for $n \geq 1$ and $\zeta := \sup_{n\in\mathbb{N}} j_n$. We equip $(0,\infty]^{\mathbb{N}}$ with the Borel σ-algebra $\mathcal{B}((0,\infty]^{\mathbb{N}})$, $(E^*)^{\mathbb{N}}$ with the full σ-algebra $\mathcal{P}((E^*)^{\mathbb{N}})$ and $(E^*)^{\mathbb{R}_+}$ with the σ-algebra $\mathcal{P}(E^*)^{\mathbb{R}_+}$ which is generated by cylinders of the form

$$C = \{\mathbf{x} = (x(t))_{t\in\mathbb{R}_+} \in (E^*)^{\mathbb{R}_+} : x(t_0) = i_0, \ldots, x(t_m) = i_m\}$$

for $m \in \mathbb{N}$, $0 \leq t_0 < \cdots < t_m$ and $i_0, \ldots, i_m \in E^*$. To prove that Φ is measurable, it is enough to show that $\Phi^{-1}(C) \in \mathcal{B}((0,\infty]^{\mathbb{N}}) \otimes \mathcal{P}((E^*)^{\mathbb{N}})$ for every such cylinder C. Equivalently, it suffices to prove that for every finite family $0 \leq t_0 < \cdots < t_m$ and $i_0, \ldots, i_m \in E^*$, the map

$$(\mathbf{d}, \mathbf{y}) \longmapsto (\Phi(\mathbf{d}, \mathbf{y})(t_0), \ldots, \Phi(\mathbf{d}, \mathbf{y})(t_m))$$

is $\mathcal{B}((0,\infty]^{\mathbb{N}}) \otimes \mathcal{P}((E^*)^{\mathbb{N}})$-measurable. Now since a finite product of measurable maps is measurable, we only need to show that for all $t \in \mathbb{R}_+$, the map

$$(\mathbf{d}, \mathbf{y}) \longmapsto \Phi(\mathbf{d}, \mathbf{y})(t)$$

is $\mathcal{B}((0,\infty]^{\mathbb{N}}) \otimes \mathcal{P}((E^*)^{\mathbb{N}})$-measurable, i.e. $\{(\mathbf{d}, \mathbf{y}) : \Phi(\mathbf{d}, \mathbf{y})(t) = i\} \in \mathcal{B}((0,\infty]^{\mathbb{N}}) \otimes \mathcal{P}((E^*)^{\mathbb{N}})$ for all $t \in \mathbb{R}_+$ and $i \in E^*$. For this, fix $t \in \mathbb{R}_+$ and $i \in E^*$. If $i \in E$, we have

$$\begin{aligned}\{(\mathbf{d}, \mathbf{y}) : \Phi(\mathbf{d}, \mathbf{y})(t) = i\} &= \bigsqcup_{n\in\mathbb{N}} \{(\mathbf{d}, \mathbf{y}) : y_n = i, j_n(\mathbf{d}) \leq t < j_{n+1}(\mathbf{d})\}\\ &= \bigsqcup_{n\in\mathbb{N}} (\{(\mathbf{d}, \mathbf{y}) : y_n = i\} \cap \{(\mathbf{d}, \mathbf{y}) : j_n(\mathbf{d}) \leq t\} \cap \{(\mathbf{d}, \mathbf{y}) : j_{n+1}(\mathbf{d}) > t\}).\end{aligned}$$

Since the map $(\mathbf{d}, \mathbf{y}) \mapsto y_n$ is a coordinate projection for each $n \in \mathbb{N}$, it is $\mathcal{B}((0,\infty]^{\mathbb{N}}) \otimes \mathcal{P}((E^*)^{\mathbb{N}})$-measurable. The map $(\mathbf{d}, \mathbf{y}) \mapsto j_n(\mathbf{d}) = \sum_{k=1}^n d_k$ is a finite sum of coordinate projections for each $n \in \mathbb{N}$, and thus $\mathcal{B}((0,\infty]^{\mathbb{N}}) \otimes \mathcal{P}((E^*)^{\mathbb{N}})$-measurable as well. Therefore the sets $\{(\mathbf{d}, \mathbf{y}) : y_n = i\}$, $\{(\mathbf{d}, \mathbf{y}) : j_n(\mathbf{d}) \leq t\}$ and $\{(\mathbf{d}, \mathbf{y}) : j_{n+1}(\mathbf{d}) > t\}$ are in $\mathcal{B}((0,\infty]^{\mathbb{N}}) \otimes \mathcal{P}((E^*)^{\mathbb{N}})$, which implies that $\{(\mathbf{d}, \mathbf{y}) : \Phi(\mathbf{d}, \mathbf{y})(t) = i\} \in \mathcal{B}((0,\infty]^{\mathbb{N}}) \otimes \mathcal{P}((E^*)^{\mathbb{N}})$. Similarly, if $i = \partial$, we have

$$\{(\mathbf{d}, \mathbf{y}) : \Phi(\mathbf{d}, \mathbf{y})(t) = \partial\} = \{(\mathbf{d}, \mathbf{y}) : \zeta(\mathbf{d}) \leq t\} = \bigcap_{n\in\mathbb{N}} \{(\mathbf{d}, \mathbf{y}) : j_n(\mathbf{d}) \leq t\} \in \mathcal{B}((0,\infty]^{\mathbb{N}}) \otimes \mathcal{P}((E^*)^{\mathbb{N}}).$$

From this, we conclude that Φ is measurable.

Now let $F : (E^*)^{\mathbb{R}_+} \to \mathbb{R}$ be a bounded $\mathcal{P}(E^*)^{\mathbb{R}_+}$-measurable function. Fix $i \in E$ and note that

$$\begin{aligned}
&\mathbf{1}_{\{J_n \le t < J_{n+1}, X_t = i\}} F(\theta_t \mathbf{X}) \\
&= \mathbf{1}_{\{J_n \le t, Y_n = i\}} \mathbf{1}_{\{t < J_{n+1}\}} F(\theta_t \Phi((D_k)_{k \ge 1}, (Y_k)_{k \in \mathbb{N}})) \\
&= \mathbf{1}_{\{J_n \le t, Y_n = i\}} \mathbf{1}_{\{D_{n+1} > t - J_n\}} F(\Phi(D_{n+1} - (t - J_n), D_{n+2}, \dots; Y_n, Y_{n+1}, \dots))
\end{aligned}$$

since at time $t \in [J_n, J_{n+1})$, the process is in state Y_n and the remaining time until the next jump is $J_n + D_{n+1} - t = D_{n+1} - (t - J_n)$. Define $G(u) := \mathbb{E}_i[\mathbf{1}_{\{D_1 > u\}} F(\Phi(D_1 - u, D_2, \dots; Y_0, Y_1, \dots))]$ for every $u \in \mathbb{R}_+$. Note that Y_n and J_n (with J_n viewed as a function of $D_1, \dots, D_n$) are both $\sigma(D_1, \dots, D_n, Y_0, \dots, Y_n)$-measurable. Using the identity from the previous display, this measurability, and the fact that the exponential variables $(\mathbf{e}_k)_{k \in \mathbb{N}}$ are i.i.d. and independent of $\mathbf{Y}$, we obtain $\mathbb{P}_\mu$-a.s.

$$\begin{aligned}
&\mathbb{E}_\mu[\mathbf{1}_{\{J_n \le t < J_{n+1}, X_t = i\}} F(\theta_t \mathbf{X}) \mid D_1, \dots, D_n, Y_0, \dots, Y_n] \\
&= \mathbb{E}_\mu[\mathbf{1}_{\{J_n \le t, Y_n = i\}} \mathbf{1}_{\{D_{n+1} > t - J_n\}} F(\Phi(D_{n+1} - (t - J_n), D_{n+2}, \dots; Y_n, Y_{n+1}, \dots)) \mid D_1, \dots, D_n, Y_0, \dots, Y_n] \\
&= \mathbf{1}_{\{J_n \le t, Y_n = i\}} \mathbb{E}_\mu[\mathbf{1}_{\{D_{n+1} > t - J_n\}} F(\Phi(D_{n+1} - (t - J_n), D_{n+2}, \dots; Y_n, Y_{n+1}, \dots)) \mid D_1, \dots, D_n, Y_0, \dots, Y_n] \\
&= \mathbf{1}_{\{J_n \le t, Y_n = i\}} \mathbb{E}_\mu[\mathbf{1}_{\{D_{n+1} > t - J_n\}} F(\Phi(D_{n+1} - (t - J_n), D_{n+2}, \dots; Y_n, Y_{n+1}, \dots)) \mid Y_0, \dots, Y_n].
\end{aligned}$$

Applying the simple Markov property at time n for the discrete-time Markov chain $\mathbf{Y}$ (see Section I.1.e of [Duq23]) and again using that the exponential variables $(\mathbf{e}_k)_{k \in \mathbb{N}}$ are i.i.d. and independent of $\mathbf{Y}$ yields $\mathbb{P}_\mu$-a.s.

$$\begin{aligned}
&\mathbb{E}_\mu[\mathbf{1}_{\{J_n \le t < J_{n+1}, X_t = i\}} F(\theta_t \mathbf{X}) \mid D_1, \dots, D_n, Y_0, \dots, Y_n] \\
&= \mathbf{1}_{\{J_n \le t, Y_n = i\}} \mathbb{E}_{Y_n}[\mathbf{1}_{\{D_1 > t - J_n\}} F(\Phi(D_1 - (t - J_n), D_2, \dots; Y_0, Y_1, \dots))] \\
&= \mathbf{1}_{\{J_n \le t, Y_n = i\}} \mathbb{E}_i[\mathbf{1}_{\{D_1 > t - J_n\}} F(\Phi(D_1 - (t - J_n), D_2, \dots; Y_0, Y_1, \dots))] \\
&= \mathbf{1}_{\{J_n \le t, Y_n = i\}} G(t - J_n).
\end{aligned} \tag{4.4}$$

By the memoryless property of exponentials, for all $u, v \in \mathbb{R}_+$, we have $\mathbb{P}_i(D_1 - u > v \mid D_1 > u) = \mathbb{P}_i(D_1 > v + u \mid D_1 > u) = \mathbb{P}_i(D_1 > v)$. This implies $\mathbb{P}_i$-a.s. that

$$\mathbb{E}_i[F(\Phi(D_1 - u, D_2, \dots; Y_0, Y_1, \dots)) \mid D_1 > u] = \mathbb{E}_i[F(\Phi(D_1, D_2, \dots; Y_0, Y_1, \dots))] = \mathbb{E}_i[F(\mathbf{X})].$$

Hence, by the definition of conditional expectation on an event, we obtain $\mathbb{P}_i$-a.s.

$$\begin{aligned}
G(u) &= \mathbb{E}_i[\mathbf{1}_{\{D_1 > u\}} F(\Phi(D_1 - u, D_2, \dots; Y_0, Y_1, \dots))] \\
&= \mathbb{P}_i(D_1 > u) \mathbb{E}_i[F(\Phi(D_1 - u, D_2, \dots; Y_0, Y_1, \dots)) \mid D_1 > u] \\
&= \mathbb{P}_i(D_1 > u) \mathbb{E}_i[F(\mathbf{X})] \\
&= e^{-q(i)u} \mathbb{E}_i[F(\mathbf{X})].
\end{aligned}$$

Combining this with (4.4) yields $\mathbb{P}_\mu$-a.s.

$$\mathbb{E}_\mu[\mathbf{1}_{\{J_n \le t < J_{n+1}, X_t = i\}} F(\theta_t \mathbf{X}) \mid D_1, \dots, D_n, Y_0, \dots, Y_n] = \mathbf{1}_{\{J_n \le t, Y_n = i\}} e^{-q(i)(t - J_n)} \mathbb{E}_i[F(\mathbf{X})]. \tag{4.5}$$

Taking $F \equiv 1$, we obtain $\mathbb{P}_\mu$-a.s.

$$\mathbb{E}_\mu[\mathbf{1}_{\{J_n \le t < J_{n+1}, X_t = i\}} \mid D_1, \dots, D_n, Y_0, \dots, Y_n] = \mathbf{1}_{\{J_n \le t, Y_n = i\}} e^{-q(i)(t - J_n)}.$$

Reinserting this into (4.5) gives $\mathbb{P}_\mu$-a.s.

$$\begin{aligned}
&\mathbb{E}_\mu[\mathbf{1}_{\{J_n \le t < J_{n+1}, X_t = i\}} F(\theta_t \mathbf{X}) \mid D_1, \dots, D_n, Y_0, \dots, Y_n] \\
&= \mathbb{E}_\mu[\mathbf{1}_{\{J_n \le t < J_{n+1}, X_t = i\}} \mid D_1, \dots, D_n, Y_0, \dots, Y_n] \mathbb{E}_i[F(\mathbf{X})].
\end{aligned}$$

On $\{J_n \le t < J_{n+1}\}$, the variables $X_0, X_{t_1}, \dots, X_{t_{m-1}}$ depend only on $D_1, \dots, D_n, Y_0, \dots, Y_n$ and are therefore $\sigma(D_1, \dots, D_n, Y_0, \dots, Y_n)$-measurable. Thus, by this measurability and the identity

in the previous display, we obtain $\mathbb{P}_\mu$-a.s.

$$\begin{aligned}
&\mathbb{E}_\mu[\mathbf{1}_{\{X_0=i_0,X_{t_1}=i_1,\dots,X_{t_{m-1}}=i_{m-1},J_n\le t<J_{n+1},X_t=i\}}F(\theta_t\mathbf{X}) \mid D_1,\dots,D_n,Y_0,\dots,Y_n]\\
&=\mathbf{1}_{\{X_0=i_0,X_{t_1}=i_1,\dots,X_{t_{m-1}}=i_{m-1}\}}\mathbb{E}_\mu[\mathbf{1}_{\{J_n\le t<J_{n+1},X_t=i\}}F(\theta_t\mathbf{X}) \mid D_1,\dots,D_n,Y_0,\dots,Y_n]\\
&=\mathbf{1}_{\{X_0=i_0,X_{t_1}=i_1,\dots,X_{t_{m-1}}=i_{m-1}\}}\mathbb{E}_\mu[\mathbf{1}_{\{J_n\le t<J_{n+1},X_t=i\}} \mid D_1,\dots,D_n,Y_0,\dots,Y_n]\mathbb{E}_i[F(\mathbf{X})]\\
&=\mathbb{E}_\mu[\mathbf{1}_{\{X_0=i_0,X_{t_1}=i_1,\dots,X_{t_{m-1}}=i_{m-1},J_n\le t<J_{n+1},X_t=i\}} \mid D_1,\dots,D_n,Y_0,\dots,Y_n]\mathbb{E}_i[F(\mathbf{X})].
\end{aligned}$$

Taking expectations on both sides and applying the weak tower property, we obtain

$$\begin{aligned}
&\mathbb{E}_\mu[\mathbf{1}_{\{X_0=i_0,X_{t_1}=i_1,\dots,X_{t_{m-1}}=i_{m-1},J_n\le t<J_{n+1},X_t=i\}}F(\theta_t\mathbf{X})]\\
&=\mathbb{E}_\mu[\mathbf{1}_{\{X_0=i_0,X_{t_1}=i_1,\dots,X_{t_{m-1}}=i_{m-1},J_n\le t<J_{n+1},X_t=i\}}]\mathbb{E}_i[F(\mathbf{X})].
\end{aligned}$$

Summing over n yields

$$\begin{aligned}
&\mathbb{E}_\mu[\mathbf{1}_{\{X_0=i_0,X_{t_1}=i_1,\dots,X_{t_{m-1}}=i_{m-1},\zeta>t,X_t=i\}}F(\theta_t\mathbf{X})]\\
&=\mathbb{P}_\mu(X_0=i_0,X_{t_1}=i_1,\dots,X_{t_{m-1}}=i_{m-1},\zeta>t,X_t=i)\mathbb{E}_i[F(\mathbf{X})].
\end{aligned}$$

Since $i\in E$, we have $\{X_t=i\}=\{\zeta>t,X_t=i\}$. Choosing $F(\mathbf{X})=\mathbf{1}_{\{X_s=j\}}$ in the previous display directly yields (4.3). Note that (4.3) also holds when $i=\partial$. Indeed, since ∂ is absorbing, the event $\{X_t=\partial,X_{t+s}=j\}$ has probability 0 whenever $j\neq\partial$, and in this case we also have $p^*_s(\partial,j)=0$, so both sides of (4.3) vanish. If instead $j=\partial$, then $X_{t+s}=\partial$ automatically whenever $X_t=\partial$, and since $p^*_s(\partial,\partial)=1$, both sides of (4.3) reduce to $\mathbb{P}_\mu(X_0=i_0,X_{t_1}=i_1,\dots,X_{t_{m-1}}=i_{m-1},X_t=\partial)$.

Applying (4.3) recursively, we obtain for every $\mu\in\mathcal{M}_1(E^*)$, every $0=t_0<t_1<\cdots<t_m$ and every $i_0,\dots,i_m\in E^*$,

$$\mathbb{P}_\mu(X_0=i_0,X_{t_1}=i_1,\dots,X_{t_m}=i_m)=\mu(i_0)\prod_{1\le k\le m}p^*_{t_k-t_{k-1}}(i_{k-1},i_k). \tag{4.6}$$

Note that the inequalities $t_{k-1}<t_k$ avoid degenerate cases in which two observation times coincide, which would force $X_{t_{k-1}}=X_{t_k}$ and lead to trivial zero-probability events. We allow $i_0,\dots,i_m\in E^*$ because we have already verified that (4.3) remains valid when one of the states equals ∂.

Next, we verify that (4.6) implies that $(P^*_t)_{t\in\mathbb{R}_+}$ is a Markovian semigroup using Definition 3.8. Condition (a) is clearly satisfied since the entries of P^*_t are nonnegative probabilities for every $t\in\mathbb{R}_+$. Condition (c') also holds, because $\sum_{j\in E^*}p^*_t(i,j)=\sum_{j\in E^*}\mathbb{P}_i(X_t=j)=1$ for all $i\in E^*$. It remains to verify (b). First note that $P^*_0=(\mathbb{P}_i(X_0=j))_{i,j\in E^*}=(\delta_{i,j})_{i,j\in E^*}=\mathrm{Id}_{E^*}$. To establish the semigroup property, take $m=2$ and choose $t_0=0<t_1=t<t_2=t+s$ in (4.6), obtaining

$$\mathbb{P}_\mu(X_0=i_0,X_t=i_1,X_{t+s}=i_2)=\mu(i_0)\,p^*_t(i_0,i_1)\,p^*_s(i_1,i_2).$$

Summing over $i_1\in E^*$ yields

$$\mathbb{P}_\mu(X_0=i_0,X_{t+s}=i_2)=\mathbb{P}_\mu(X_0=i_0)\sum_{i_1\in E^*}p^*_t(i_0,i_1)p^*_s(i_1,i_2).$$

Taking the initial distribution $\mu=\delta_{i_0}$, we have $\mathbb{P}_{i_0}(X_0=i_0)=1$, and therefore

$$p^*_{t+s}(i_0,i_2)=\sum_{i_1\in E^*}p^*_t(i_0,i_1)p^*_s(i_1,i_2)$$

which proves the semigroup property. Hence $(P^*_t)_{t\in\mathbb{R}_+}$ is a Markovian semigroup. Moreover, by Definition 3.16, (4.6) also implies that $\mathbf{X}$ is a Markov process, completing the proof. □

This construction of the minimal Markov process $\mathbf{X}$ associated with G, together with point (ii) of Remark 4.3, yields the following result on the existence and uniqueness of nonnegative solutions to the Chapman-Kolmogorov equations.

Corollary 4.6. *For every generator G as in Definition 4.1, there exists a unique minimal nonnegative solution $(P_t)_{t\in\mathbb{R}_+}$ to the backward and forward Chapman-Kolmogorov equations*

$$\tfrac{\mathrm{d}}{\mathrm{d}t}P_t = GP_t = P_tG \text{ for all } t \in \mathbb{R}_+ \quad \text{and} \quad P_0 = \mathrm{Id}_E.$$

Moreover, this minimal solution with nonnegative coefficients is a sub-Markovian semigroup.

Note that by Proposition 4.5 and point (ii) of Remark 4.3, the sub-Markovian semigroup appearing in the corollary is actually $(P^{\min}_t)_{t\in\mathbb{R}_+}$, where $P^{\min}_t(i,j) := \mathbb{P}_i(X_t = j, \zeta > t)$ for all $t \in \mathbb{R}_+$ and $i, j \in E$, and where $\mathbf{X}$ is the constructed minimal Markov process associated with G, with explosion time ζ. The semigroup $(P^{\min}_t)_{t\in\mathbb{R}_+}$ coincides with the restriction to E of the Markovian semigroup $(P^*_t)_{t\in\mathbb{R}_+}$ of the constructed process $\mathbf{X}$ on E^*.

Remark 4.7. The construction in this section shows that the infinitesimal generator G determines the entire behavior of the associated Markov process $\mathbf{X}$ before its explosion time ζ. Indeed, from G we recover both the holding time parameters $q(i)$ and the transition matrix Q of the skeleton chain, and therefore the full dynamics of the trajectory $(X_t)_{0\le t<\zeta}$. This is particularly useful, since given G, one can quickly understand the behavior of the associated Markov process up to explosion, simply by inspecting the entries of G.

However, it is important to note that G does not specify the behavior of the process at or after the explosion time. Consequently, several right-continuous Markov processes may share the same generator G while differing in their behavior at or after the explosion time. Among all such processes, the minimal Markov process associated with G is the canonical and simplest representative.

4.2 Simulation of Minimal Markov Processes

Since the construction developed in the previous section is particularly well suited for simulation, we now present two examples of continuous-time Markov processes constructed from an infinitesimal generator and simulated in R. The first example considers a process on a finite state space, without a cemetery state or explosion. The second example builds on the code structure of the first but simulates a process on an infinite state space where explosion may occur.

Example 4.8 (adapted from [Ami], Exercise 6.7.1). We consider an infinitesimal generator $G = (q_{i,j})_{i,j\in E}$ on the finite state space $E := \{1, 2, 3, 4\}$ given by

$$G := \begin{pmatrix} -3 & 2 & 0 & 1 \\ 0 & -2 & \frac{1}{2} & \frac{3}{2} \\ 1 & 1 & -4 & 2 \\ 1 & 0 & 0 & -1 \end{pmatrix}.$$

Starting from this generator, we define the holding time parameters $(q(i))_{i\in E}$ and construct the associated skeleton transition matrix $Q = (\pi_{i,j})_{i,j\in E}$. Observe that $q(i) = -q_{i,i} > 0$ for all $i \in E$, so none of the states is absorbing. We assume that the associated process $\mathbf{X}$ starts in state $X_0 := 1$ and fix an end time $T := 3$ for the simulation.

To generate a sample path of $\mathbf{X}$, we sequentially draw independent exponential random variables $\mathbf{e}_n \sim \mathrm{Exp}(1)$ and use them to construct the holding times D_n. Based on these holding times, we recursively compute the jump times J_n and at each jump sequentially sample the corresponding skeleton states Y_n from the transition matrix Q. This procedure is continued until the next jump time would exceed T. The sample path $(X_t)_{t\in[0,T]}$ is then constructed as a piecewise constant function from the jump times $(J_n)_{n\in\mathbb{N}}$ and skeleton states $(Y_n)_{n\in\mathbb{N}}$ and evaluated on a fine time grid for plotting (see Listing 4.1 and Figure 4.1).

Since the state space E is finite, explosion cannot occur. Consequently, no cemetery state ∂ is required and no explosion time ζ needs to be considered in this example.

Listing 4.1: Simulation of a finite-state continuous-time Markov process from its generator

```
set.seed(12345)

# finite state space
```

```r
E <- 1:4

# generator matrix G = (q_{i,j})
G <- matrix(c(
  -3, 2, 0, 1,
   0, -2, 1/2, 3/2,
   1, 1, -4, 2,
   1, 0, 0, -1
), nrow = 4, byrow = TRUE)

# vector of holding time parameters q(i) = -q_{i,i}
q <- -diag(G)

# skeleton transition matrix Q = (pi_{i,j})
Q <- matrix(0, nrow = 4, ncol = 4)
for (i in E) {
  for (j in E) {
    # pi_{i,j} = q_{i,j} / q(i) for j != i (here q(i) > 0 for all i)
    if (j != i) Q[i, j] <- G[i, j] / q[i]
    # keep pi_{i,i} = 0
  }
}

# simulate path of continuous-time Markov process X given by G (up to T = 3)
T <- 3 # end time
current_state <- 1 # Y_0 = 1
current_jump_time <- 0 # J_0 = 0
Y <- c(1) # vector of visited states (skeleton)
J <- c(0) # vector of jump times

while (TRUE) {
  # sample e_{n+1} ~ Exp(1)
  exp_var <- rexp(1, rate = 1)
  # D_{n+1} = e_{n+1} / q(Y_n)
  next_holding_time <- exp_var / q[current_state]
  # J_{n+1} = J_n + D_{n+1}
  next_jump_time <- current_jump_time + next_holding_time

  # if next jump occurs after T, stop without updating J and Y
  if (next_jump_time > T) break

  # if next jump occurs before T, update J and Y
  J <- c(J, next_jump_time)
  # sample next state j from E using transition probability Q[current_state, j]
  next_state <- sample(E, size = 1, prob = Q[current_state, ])
  Y <- c(Y, next_state)

  # update current jump time and state
  current_jump_time <- next_jump_time
  current_state <- next_state
}

# build path of X from J and Y
t_grid <- seq(0, T, by = 0.001) # time grid
X <- numeric(length(t_grid)) # vector of values X_t evaluated at t_grid

for (k in seq_along(t_grid)) {
  t <- t_grid[k]
  # find index n such that J[n] <= t < J[n+1]
  n <- max(which(J <= t))
  # X_t = Y[n]
```

```
  X[k] <- Y[n]
}

# plot simulated path of X
par(mar = c(5, 4, 1, 1))
plot(
  t_grid, X,
  type = "n", # set up axes, but do not draw path yet
  lwd = 1.5,
  xlab = "t",
  ylab = expression(X[t]),
  xlim = c(0, T),
  ylim = c(1, 4),
  yaxt = "n" # suppress default y-axis tick marks
)

grid() # draw grid first (background)
lines(t_grid, X, type = "s", lwd = 1.5) # draw path (step function) on top
axis(2, at = 1:4) # add custom y-axis tick marks
```

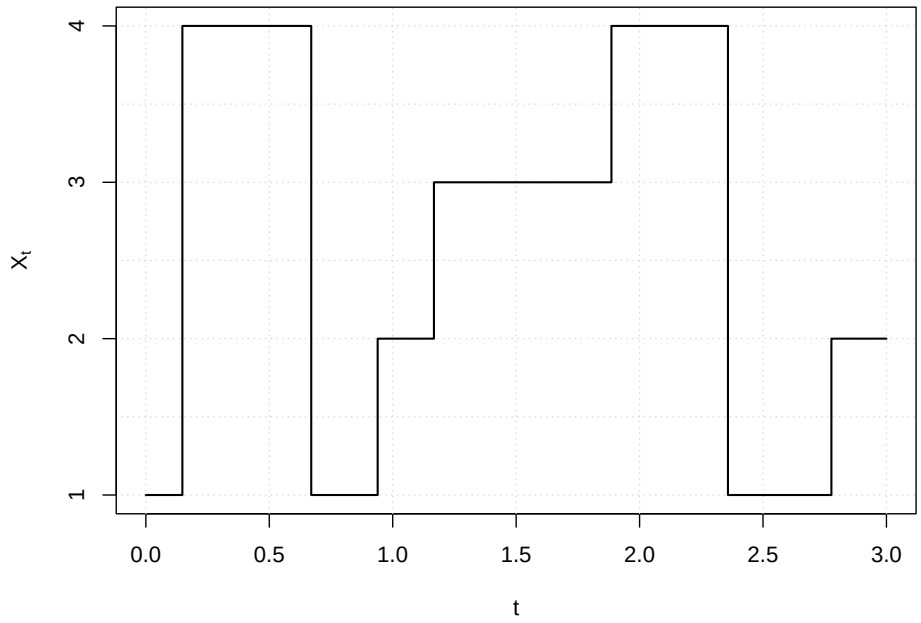

Figure 4.1: Simulated sample path of a finite-state continuous-time Markov process on $E = \{1, 2, 3, 4\}$ up to time $T = 3$.

The resulting sample path in Figure 4.1 illustrates the piecewise constant nature of a continuous-time Markov process. Both the jump times $(J_n)_{n\in\mathbb{N}}$ and the visited states $(Y_n)_{n\in\mathbb{N}}$ are determined by the generator G and the irregular spacing of the jumps reflects the exponentially distributed holding times $(D_n)_{n\geq 1}$.

Example 4.9 (adapted from [Ami], Example 6.4.11). We now consider an infinitesimal generator $G = (q_{i,j})_{i,j\in E}$ on the infinite state space $E := \{0, 1, 2, \dots\}$ and introduce a cemetery state $\partial \notin E$ which we represent numerically by $\partial := -1$. In contrast to the finite-state example, the generator G cannot be specified explicitly as an infinite matrix in R. Instead, it is defined implicitly through the holding time parameters $(q(i))_{i\in E}$ and skeleton transition matrix $Q = (\pi_{i,j})_{i,j\in E}$.

Since an infinite vector cannot be stored explicitly in R, the holding time parameters $(q(i))_{i\in E^*}$ are defined by a function rather than a vector. We set $q(i) := 2^i$ for all $i \in E$ and $q(\partial) := 0$, so that the holding time at the cemetery state is infinite. Similarly, since an infinite skeleton transition matrix Q cannot be stored explicitly, we implement the skeleton transitions by a function that samples the next state $j \in E^*$ given the current state $i \in E^*$ according to the transition probability

$\pi_{i,j}$. We choose the transition probabilities $(\pi_{i,j})_{i,j\in E^*}$ so that the associated skeleton chain $\mathbf{Y}$ moves from the initial state $Y_0 := 0$ to 1 with probability 1, and from any state $i \geq 1$ to $i+1$ with probability $a := 0.6$ and to $i-1$ with probability $b := 1-a = 0.4$. We will see later that a process $\mathbf{X}$ associated with a skeleton chain $\mathbf{Y}$ of this form is called a *birth-death process*. With this choice of transition probabilities, the cemetery state ∂ cannot be reached by a direct skeleton transition and is only reached through explosion. To make the cemetery state absorbing, we define $\pi_{\partial,\partial} := 1$.

The simulation of a sample path of $\mathbf{X}$ follows the same sequential procedure as in the finite-state example. Independent exponential random variables $\mathbf{e}_n$ are used to sequentially construct the holding times D_n and then recursively obtain the jump times J_n. The skeleton states Y_n are sampled sequentially using our defined skeleton transition function. This procedure is continued until the next jump time would exceed $T := 3$ or until explosion occurs. Since the accumulation of infinitely many jumps in finite time cannot be simulated directly, explosion is detected numerically via a tolerance threshold of 10^{-12} on the holding times D_n. Once a holding time falls below this threshold, we assume that the jump times have accumulated, approximate the explosion time ζ by the last resolved jump time, and send the process $\mathbf{X}$ to the cemetery state ∂ where it remains for all subsequent times.

As in the finite-state example, the sample path $(X_t)_{t\in[0,T]}$ is then constructed as a piecewise constant function from the jump times $(J_n)_{n\in\mathbb{N}}$ and skeleton states $(Y_n)_{n\in\mathbb{N}}$ and evaluated on a fine time grid for plotting (see Listing 4.2 and Figure 4.2).

To give some intuition for this construction, note that our choice of transition probabilities $(\pi_{i,j})_{i,j\in E^*}$ leads to a drift of the skeleton chain $\mathbf{Y}$ toward higher states. Indeed, since $a > b$, jumps to the right occur with higher probability than jumps to the left, so the states Y_n tend to increase as n grows. Moreover, our choice of holding time parameters $(q(i))_{i\in E}$ implies, by Theorem 3.38, that the holding times $(D_n)_{n\geq 1}$ satisfy $\mathbb{E}_\mu[D_n \mid Y_{n-1} = i] = \frac{1}{q(i)} = 2^{-i}$ for all $i \in E$ and $\mu \in \mathcal{M}_1(E)$, which converges to zero as $i \to \infty$. Consequently, as the process $\mathbf{X}$ moves to higher states, the expected holding times become arbitrarily small and the process makes infinitely many jumps in finite time, leading to explosion.

Listing 4.2: Simulation of an explosive continuous-time minimal Markov process

```
set.seed(12345)

# infinite state space E = {0,1,2,...}
# cemetery state partial = -1
partial <- -1

# function for holding time parameters q(i)
q <- function(i) {
  if (i == partial) {
    return(0) # q(partial) = 0 (infinite holding time)
  } else {
    return(2^i) # q(i) = 2^i for i in E
  }
}

# choose a,b > 0 with a + b = 1 and 1 < a / b < 2
a <- 0.6
b <- 1 - a

# function for skeleton transition according to Q = (pi_{i,j})
sample_from_Q <- function(i) {
  # pi_{partial, partial} = 1 (absorbing)
  if (i == partial) return(partial)
  # pi_{0, 1} = 1 (partial = -1 only reachable by explosion)
  if (i == 0) return(1)
  # pi_{i,i+1} = a and pi_{i,i-1} = b for i >= 1
  sample(x = c(i - 1, i + 1), size = 1, prob = c(b, a))
}
```

```
# simulate path of continuous-time minimal Markov process X (up to T = 3)
# (given by q and Q)
T <- 3 # end time
previous_state <- 0 # Y_0 = 0
previous_jump_time <- 0 # J_0 = 0
Y <- c(0) # visited states (skeleton)
J <- c(0) # jump times

# practical explosion detection (simulation cannot do infinitely many jumps)
max_jumps <- 200000
tol_time_increment <- 1e-12
zeta <- Inf

for (n in 1:max_jumps) {
 # cemetery state is absorbing
 if (previous_state == partial) break

 # sample e_n ~ Exp(1)
 exp_var <- rexp(1, rate = 1)
 # D_n = e_n / q(Y_{n-1}) (with convention e_n / 0 = Inf)
 current_holding_time <- if (q(previous_state) == 0) {
  Inf
 } else {
  exp_var / q(previous_state)
 }
 # J_n = J_{n-1} + D_n
 current_jump_time <- previous_jump_time + current_holding_time

 # detect explosion numerically
 # (extremely small holding times indicate accumulation of jump times)
 if (current_holding_time < tol_time_increment) {
  # approximate zeta by last resolved jump time J_{n-1} (zeta = sup_n J_n)
  zeta <- previous_jump_time
  # X_t = partial for t >= zeta (absorption after explosion)
  previous_state <- partial
  break
 }

 # if current jump occurs after T, stop without updating J and Y
 if (current_jump_time > T) break

 # if current jump occurs before T, update J and Y
 J <- c(J, current_jump_time)
 # sample current state j from E using transition proba Q[previous_state, j]
 current_state <- sample_from_Q(previous_state)
 Y <- c(Y, current_state)

 # update previous jump time and state
 previous_jump_time <- current_jump_time
 previous_state <- current_state
}

# build path of X from J and Y
t_grid <- seq(0, T, by = 0.001) # time grid
X <- numeric(length(t_grid)) # vector of values X_t evaluated at t_grid

for (k in seq_along(t_grid)) {
 t <- t_grid[k]
 # X_t = partial for t >= zeta (absorption after explosion)
 if (t >= zeta) {
  X[k] <- partial
```

```
  } else {
    # find index n such that J[n] <= t < J[n+1]
    n <- max(which(J <= t))
    # X_t = Y[n]
    X[k] <- Y[n]
  }
}

# plot simulated path of X
par(mar = c(5, 4, 1, 1))
plot(
  t_grid, X,
  type = "n", # set up axes, but do not draw path yet
  lwd = 1.5,
  xlab = "t",
  ylab = expression(X[t]),
  xlim = c(0, T),
  yaxt = "n" # suppress default y-axis tick marks
)

grid() # draw grid first (background)
lines(t_grid, X, type = "s", lwd = 1.5) # draw path (step function) on top

# add custom y-axis tick marks
y_max <- max(X[X != partial])
y_ticks <- 0:y_max
axis(2, at = y_ticks, labels = y_ticks)
axis(2, at = partial, labels = expression(partialdiff))
```

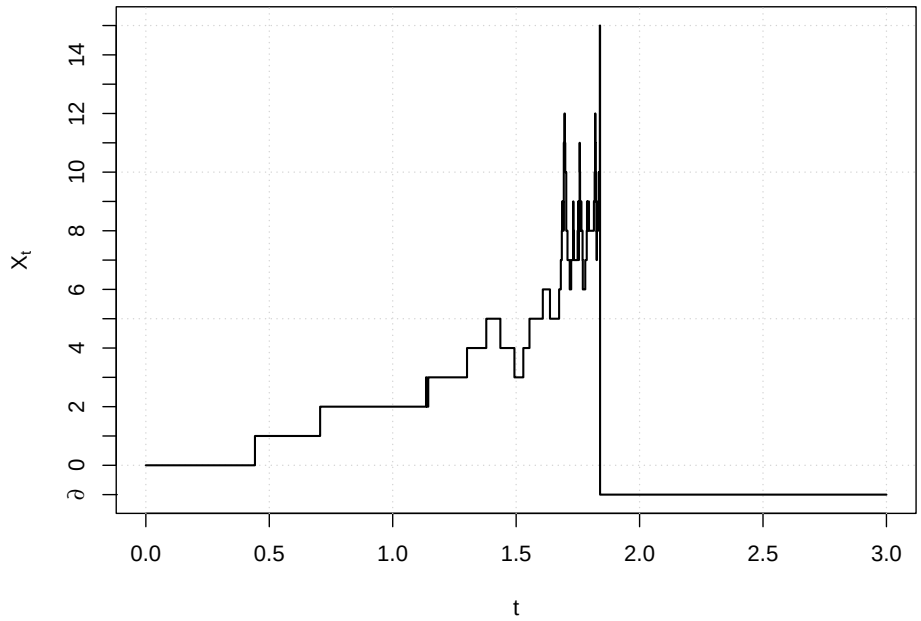

Figure 4.2: Simulated sample path of an explosive continuous-time minimal Markov process on $E = \{0, 1, 2, \dots\}$ up to time $T = 3$.

The resulting sample path in Figure 4.2 shows that as the states increase, the holding times shrink and the jumps occur more and more rapidly until explosion is detected. At the (approximate) explosion time, the path is sent to the cemetery state ∂ and remains there, which is reflected by the flat segment following the vertical drop.

5 Qualitative Study of Markov Processes

5.1 Recurrence and Transience

We now begin our qualitative study of Markov processes by introducing the classification of states, which describes how the process moves between states and, in particular, whether certain states can be reached, revisited, or are eventually left forever.

Let E be a countable space equipped with the discrete topology, and hence endowed with the σ-algebra $\mathcal{P}(E)$. Let $G = (q_{i,j})_{i,j\in E}$ be an infinitesimal generator on E (according to Definition 4.1). We introduce a cemetery point $\partial \notin E$ and denote $E^* := E \cup \{\partial\}$, which we view as a countable space equipped with the discrete topology and the corresponding σ-algebra $\mathcal{P}(E^*)$. Let

$$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, \mathbf{X} = (X_t)_{t\in\mathbb{R}_+}, G, \mathbb{P}_\mu, \mu \in \mathcal{M}_1(E))$$

be the minimal right-continuous Markov process on E^* associated with G as constructed in the last section. We denote by $(P_t)_{t\in\mathbb{R}_+}$ the minimal semigroup associated with G (see Corollary 4.6) and use the notation $P_t := (p_t(i,j))_{i,j\in E}$. Recall the notation $(q(i))_{i\in E}$ for the holding time parameters and $Q := (\pi_{i,j})_{i,j\in E}$ for the transition matrix of the skeleton chain, both of which are determined by G. To simplify notation, we set $D_n := D_n(\mathbf{X})$, $J_n := J_n(\mathbf{X})$, $\zeta := \zeta(\mathbf{X})$ and $Y_n := Y_n(\mathbf{X})$.

Note that the results in this section concern only the states of E and not those of E^*, since the process stays in E up to the explosion time and the post-explosion dynamics at ∂ are trivial.

Definition 5.1 (Irreducibility)**.** Let $i, j \in E$ be two distinct states.

(a) We say that j is *accessible* from i if and only if $\mathbb{P}_i(\exists\, t \in \mathbb{R}_+ : X_t = j) > 0$, which we write $i \to j$.

(b) We say that i and j *communicate* if and only if $i \to j$ and $j \to i$, which we write $i \leftrightarrow j$.

(c) We say that G (or the minimal Markov process $\mathbf{X}$ associated with G) is *irreducible* if and only if all its states in E communicate.

Accessibility, as defined above, can be expressed in several different but equivalent ways: through transitions of the skeleton chain $\mathbf{Y}$, through the generator G or through positive entries of the minimal semigroup $(P_t)_{t\in\mathbb{R}_+}$. The following proposition states these equivalent formulations.

Proposition 5.2. *Let $i, j \in E$ be two distinct states. The following assertions are equivalent:*

(i) $i \to j$

(ii) *j is accessible from i for the skeleton chain $\mathbf{Y}$, i.e. there exist $n \geq 1$ and states $i_1, \dots, i_{n-1} \in E$ such that $\pi_{i,i_1}\pi_{i_1,i_2}\cdots\pi_{i_{n-1},j} > 0$.*

(iii) *There exist $n \geq 1$ and states $i_1, \dots, i_{n-1} \in E$ such that $i \neq i_1$, $i_1 \neq i_2$, $\dots$, $i_{n-1} \neq j$ and $q_{i,i_1}q_{i_1,i_2}\cdots q_{i_{n-1},j} > 0$.*

(iv) *For all $t > 0$, $p_t(i,j) > 0$.*

(v) *There exists $t_0 > 0$ such that $p_{t_0}(i,j) > 0$.*

Proof. Clearly, (iv) implies (v). For the implication from (v) to (i), observe that (v) implies that there exists $t_0 > 0$ such that $\mathbb{P}_i(X_{t_0} = j) > 0$, and since $\{X_{t_0} = j\} \subseteq \{\exists\, t \in \mathbb{R}_+ : X_t = j\}$, we have $\mathbb{P}_i(\exists\, t \in \mathbb{R}_+ : X_t = j) \geq \mathbb{P}_i(X_{t_0} = j) > 0$.

For the implication from (i) to (ii), note that (i) implies that $\mathbb{P}_i(\exists\, m \geq 1 : Y_m = j) > 0$, and since

$$\mathbb{P}_i(\exists\, m \geq 1 : Y_m = j) = \mathbb{P}_i\Big(\bigsqcup_{m\geq 1}\{Y_m = j\}\Big) = \sum_{m\geq 1}\mathbb{P}_i(Y_m = j)$$

there exists $n \geq 1$ such that $\mathbb{P}_i(Y_n = j) > 0$. By Section I.2.a of [Duq23],

$$\mathbb{P}_i(Y_n = j) = [Q^n](i,j) = \sum_{k_1,\dots,k_{n-1}\in E} \pi_{i,k_1}\pi_{k_1,k_2}\cdots\pi_{k_{n-1},j}$$

so there exist $i_1, \dots, i_{n-1} \in E$ such that $\pi_{i,i_1}\pi_{i_1,i_2}\cdots\pi_{i_{n-1},j} > 0$.

Now we prove the implication from (ii) to (iii). Assume (ii). If j is absorbing and the path $i \to i_1 \to \cdots \to i_{n-1} \to j$ repeats j multiple times at the end (which is possible since $\pi_{j,j} = 1$ when $q(j) = 0$), we shorten the path so that j appears only once at the end. We may then assume that none of the states $i, i_1, \dots, i_{n-1}$ is absorbing, i.e. $q(i), q(i_1), \dots, q(i_{n-1}) > 0$, and that $i \neq i_1$, $i_1 \neq i_2$, …, $i_{n-1} \neq j$. By Definition 4.1, we then have

$$q_{i,i_1}q_{i_1,i_2}\cdots q_{i_{n-1},j} = q(i)\cdots q(i_{n-1})\pi_{i,i_1}\pi_{i_1,i_2}\cdots\pi_{i_{n-1},j} > 0.$$

Finally, to show the implication from (iii) to (iv), assume (iii). Note that if $k \neq \ell$ in E and $q_{k,\ell} > 0$, then

$$q(k) = -q_{k,k} = \sum_{m \in E\setminus\{k\}} q_{k,m} \geq q_{k,\ell} > 0.$$

By Definition 4.1, we therefore have $\pi_{k,\ell} = -\frac{q_{k,\ell}}{q_{k,k}} > 0$. Moreover, for every $s > 0$,

$$\begin{aligned}
p_s(k,\ell) &= \mathbb{P}_k(X_s = \ell) \\
&= \mathbb{P}_k(\exists\, n \geq 1 : J_n \leq s < J_{n+1}, Y_n = \ell) \\
&\geq \mathbb{P}_k(J_1 \leq s < J_2, Y_1 = \ell) \\
&= \int_0^s \mathbb{P}_k(J_1 \in \mathrm{d}u, J_2 > s, Y_1 = \ell) \\
&= \int_0^s \mathbb{P}_k(J_1 \in \mathrm{d}u, Y_1 = \ell)\mathbb{P}_k(J_2 > s \mid J_1 = u, Y_1 = \ell). \qquad (5.1)
\end{aligned}$$

By Proposition 3.36, under $\mathbb{P}_k$, J_1 is independent of Y_1 and exponentially distributed with parameter $q(k)$. Hence

$$\mathbb{P}_k(J_1 \in \mathrm{d}u, Y_1 = \ell) = \mathbb{P}_k(J_1 \in \mathrm{d}u)\mathbb{P}_k(Y_1 = \ell) = \mathrm{d}u\, q(k)e^{-q(k)u}\pi_{k,\ell}.$$

We now compute the second term in the integrand of (5.1). Because both J_1 and $Y_1 = X_{J_1}$ are $\mathcal{F}_{J_1}$-measurable (Lemma 3.30 and Theorem 3.33), the event $\{J_1 = u, Y_1 = \ell\}$ belongs to $\mathcal{F}_{J_1}$. Thus, by the tower property, we obtain $\mathbb{P}_k$-a.s.

$$\begin{aligned}
\mathbb{P}_k(J_2 > s \mid J_1 = u, Y_1 = \ell) &= \mathbb{P}_k(J_2 - J_1 > s - u \mid J_1 = u, Y_1 = \ell) \\
&= \mathbb{E}_k[\mathbf{1}_{\{J_2 - J_1 > s-u\}} \mid J_1 = u, Y_1 = \ell] \\
&= \mathbb{E}_k[\mathbb{E}_k[\mathbf{1}_{\{J_2 - J_1 > s-u\}} \mid \mathcal{F}_{J_1}] \mid J_1 = u, Y_1 = \ell].
\end{aligned}$$

Applying the strong Markov property at J_1 to the inner conditional expectation yields $\mathbb{P}_k$-a.s.

$$\begin{aligned}
\mathbb{P}_k(J_2 > s \mid J_1 = u, Y_1 = \ell) &= \mathbb{E}_k[\mathbb{E}_k[\mathbf{1}_{\{J_1<\infty\}}\mathbf{1}_{\{J_1(\theta_{J_1}\mathbf{X})>s-u\}} \mid \mathcal{F}_{J_1}] \mid J_1 = u, Y_1 = \ell] \\
&= \mathbb{E}_k[\mathbf{1}_{\{J_1<\infty\}}\mathbb{E}_{X_{J_1}}[\mathbf{1}_{\{J_1>s-u\}}] \mid J_1 = u, Y_1 = \ell] \\
&= \mathbb{E}_k[\mathbb{E}_{Y_1}[\mathbf{1}_{\{J_1>s-u\}}] \mid J_1 = u, Y_1 = \ell] \\
&= \mathbb{E}_k[\mathbb{E}_\ell[\mathbf{1}_{\{J_1>s-u\}}] \mid J_1 = u, Y_1 = \ell] \\
&= \mathbb{P}_\ell(J_1 > s - u) \\
&= e^{-q(\ell)(s-u)}
\end{aligned}$$

where we also used that $\mathbb{P}_k(J_1 < \infty) = 1$, since for $q(k) > 0$ the identity $\mathbb{P}_k(J_1 > t) = e^{-q(k)t}$ implies $\mathbb{P}_k(J_1 = \infty) = \lim_{t\to\infty} e^{-q(k)t} = 0$. Putting the two expressions back into (5.1), we get

$$p_s(k,\ell) \geq \int_0^s \mathrm{d}u\, q(k)e^{-q(k)u}\pi_{k,\ell}e^{-q(\ell)(s-u)}.$$

Note that the integrand is strictly positive since $q(k) > 0$ and $\pi_{k,\ell} > 0$. Since $s > 0$, the whole integral is therefore strictly positive, and thus $p_s(k,\ell) > 0$ for all $s > 0$ and all $k \neq \ell$ with $q_{k,\ell} > 0$.

Now let the states $i_1, \dots, i_{n-1} \in E$ be those provided by (iii). For any $t > 0$, using the semigroup property $P_t = (P_{\frac{t}{n}})^n$ together with the definition of the matrix power, we obtain

$$\begin{aligned} p_t(i,j) &= \sum_{k_1,\dots,k_{n-1} \in E} p_{\frac{t}{n}}(i,k_1) p_{\frac{t}{n}}(k_1,k_2) \cdots p_{\frac{t}{n}}(k_{n-1},j) \\ &\geq p_{\frac{t}{n}}(i,i_1) p_{\frac{t}{n}}(i_1,i_2) \cdots p_{\frac{t}{n}}(i_{n-1},j). \end{aligned}$$

Using the fact that $p_s(k,\ell) > 0$ for all $s > 0$ and all $k \neq \ell$ with $q_{k,\ell} > 0$, we then get that all the factors in the last product are strictly positive, and thus $p_t(i,j) > 0$ for all $t > 0$. This proves (iv) and completes the proof of the proposition. □

Having characterized when one state can be reached from another, we now study how often the process returns to a given state. This leads to the notions of recurrence and transience, which depend on whether the set of times at which the process visits a state is bounded or unbounded.

Note that a set $A \subseteq \mathbb{R}_+$ is said to be *bounded* if there exists $M < \infty$ such that $A \subseteq [0, M]$, or equivalently if $\sup A < \infty$. Conversely, A is said to be *unbounded* if $\sup A = \infty$.

Definition 5.3 (Recurrent and transient states)**.** Fix $i \in E$.

(a) We say that i is *recurrent* if, under $\mathbb{P}_i$, the set $\{t \in \mathbb{R}_+ : X_t = i\}$ is almost surely unbounded, i.e. $\mathbb{P}_i(\{t \in \mathbb{R}_+ : X_t = i\} \text{ is unbounded}) = 1$.

(b) We say that i is *transient* if, under $\mathbb{P}_i$, the set $\{t \in \mathbb{R}_+ : X_t = i\}$ is almost surely bounded, i.e. $\mathbb{P}_i(\{t \in \mathbb{R}_+ : X_t = i\} \text{ is bounded}) = 1$.

Because the process $\mathbf{X}$ changes state only at its jump times and follows the sequence of states of the skeleton chain $\mathbf{Y}$, recurrence and transience for $\mathbf{X}$ are closely tied to those of $\mathbf{Y}$ (see Section I.3.a of [Duq23] for the definitions of recurrence and transience for discrete-time Markov chains). Recurrence and transience also interact with the communication structure of the state space. These relationships are summarized in the next theorem.

Theorem 5.4. *Let $i, j \in E$. The following assertions hold:*

(i) If i is recurrent for the skeleton chain $\mathbf{Y}$, then i is recurrent for the Markov process $\mathbf{X}$.

(ii) If i is transient for the skeleton chain $\mathbf{Y}$, then i is transient for the Markov process $\mathbf{X}$.

(iii) Every state in E is either transient or recurrent for $\mathbf{X}$.

(iv) If $i \leftrightarrow j$, then i and j are either both transient or both recurrent for $\mathbf{X}$.

(v) If G is irreducible, then either all states are recurrent or all states are transient for $\mathbf{X}$.

Proof. (i) Suppose that i is recurrent for $\mathbf{Y}$. We define the successive return times $T_i^{(m)} := T_i^{(m)}(\mathbf{Y})$, $m \in \mathbb{N}$, of $\mathbf{Y}$ to i by

$$T_i^{(0)} := 0 \quad \text{and} \quad T_i^{(m+1)} := \inf\{n > T_i^{(m)} : Y_n = i\}.$$

Since i is recurrent for $\mathbf{Y}$, by Theorem I.3.2 of [Duq23], we have $\mathbb{P}_i$-a.s. that $T_i^{(m)} < \infty$ for all $m \in \mathbb{N}$. We then observe that

$$\{t \in [0, \zeta) : X_t = i\} = \bigsqcup_{m \in \mathbb{N}} [J_{T_i^{(m)}}, J_{T_i^{(m)}+1}) \quad \mathbb{P}_i\text{-a.s.} \tag{5.2}$$

By Theorem 3.38, under $\mathbb{P}_i$ the random variables $D_{T_i^{(m)}+1} = J_{T_i^{(m)}+1} - J_{T_i^{(m)}}$, $m \in \mathbb{N}$, are independent and follow an (extended) exponential distribution with parameter $q(i)$ (since $Y_{T_i^{(m)}} = i$ $\mathbb{P}_i$-a.s. for all $m \in \mathbb{N}$).

If $q(i) > 0$, then under $\mathbb{P}_i$ the random variables $D_{T_i^{(m)}+1}$, $m \in \mathbb{N}$, are i.i.d. with finite mean $\mathbb{E}[D_{T_i^{(m)}+1}] = \frac{1}{q(i)}$. By the law of large numbers, we obtain

$$\tfrac{1}{n} \sum_{m=0}^{n-1} D_{T_i^{(m)}+1} \xrightarrow[n\to\infty]{} \tfrac{1}{q(i)} > 0 \quad \mathbb{P}_i\text{-a.s.}$$

Thus the average of the holding times stays strictly positive for large n, which implies that

$$\sum_{m=0}^{n-1} D_{T_i^{(m)}+1} \xrightarrow[n\to\infty]{} \infty \quad \mathbb{P}_i\text{-a.s.}$$

and therefore

$$\sum_{m\in\mathbb{N}} D_{T_i^{(m)}+1} = \infty \quad \mathbb{P}_i\text{-a.s.} \tag{5.3}$$

If $q(i) = 0$, then $\mathbb{P}_i(D_{T_i^{(0)}+1} = \infty) = \mathbb{P}_i(D_1 = \infty) = 1$, so (5.3) also holds in this case. Hence, in both cases, we have

$$\zeta = \sum_{n\geq 1} D_n \geq \sum_{m\in\mathbb{N}} D_{T_i^{(m)}+1} = \infty \quad \mathbb{P}_i\text{-a.s.}$$

which implies that $\mathbb{P}_i(\zeta = \infty) = 1$. Consequently, since $(J_{T_i^{(m)}})_{m\in\mathbb{N}}$ is a subsequence of the increasing sequence $(J_n)_{n\in\mathbb{N}}$, we have $\mathbb{P}_i$-a.s.

$$\lim_{m\to\infty} J_{T_i^{(m)}} = \lim_{n\to\infty} J_n = \sup_{n\in\mathbb{N}} J_n = \zeta = \infty.$$

Thus by (5.2) the set $\{t \in \mathbb{R}_+ : X_t = i\}$ is $\mathbb{P}_i$-a.s. unbounded, and hence i is recurrent for $\mathbf{X}$.

(ii) Now suppose that i is transient for $\mathbf{Y}$. This implies that $\mathbf{X}$ is not absorbed in i. Indeed, assume by contradiction that $\mathbf{X}$ is absorbed in i. Then, under $\mathbb{P}_i$, the process $\mathbf{X}$ never leaves i and we have $Y_n = i$ for all $n \in \mathbb{N}$. Hence

$$N_i(\mathbf{Y}) = \sum_{n\geq 1} \mathbf{1}_{\{Y_n = i\}} = \infty \quad \mathbb{P}_i\text{-a.s.}$$

By Corollary I.3.3 of [Duq23], this implies that i is recurrent for $\mathbf{Y}$, which contradicts the assumption that i is transient for $\mathbf{Y}$. Therefore $\mathbf{X}$ is not absorbed in i, and hence $q(i) > 0$.

Moreover, $\mathbb{P}_i$-a.s., there exists a (random) $M \in \mathbb{N}$ such that $T_i^{(M)} < \infty$ and $T_i^{(M+1)} = \infty$. Indeed, set $M := N_i(\mathbf{Y}) = \sup\{m \in \mathbb{N} : T_i^{(m)} < \infty\}$, which is finite $\mathbb{P}_i$-a.s. by Corollary I.3.3 of [Duq23]. Therefore

$$\{t \in [0, \zeta) : X_t = i\} = \bigsqcup_{m=0}^{M} [J_{T_i^{(m)}}, J_{T_i^{(m)}+1}) \quad \mathbb{P}_i\text{-a.s.} \tag{5.4}$$

Again, by Theorem 3.38, under $\mathbb{P}_i$ the random variable $D_{T_i^{(M)}+1} = J_{T_i^{(M)}+1} - J_{T_i^{(M)}}$ follows an (extended) exponential distribution with parameter $q(i) > 0$ (since $Y_{T_i^{(M)}} = i$ $\mathbb{P}_i$-a.s.). This implies that $\mathbb{P}_i(D_{T_i^{(M)}+1} < \infty) = 1$. Moreover, we have $J_{T_i^{(M)}} < \infty$ $\mathbb{P}_i$-a.s. Indeed, assume by contradiction that $\mathbb{P}_i(J_{T_i^{(M)}} = \infty) > 0$. On the event $\{J_{T_i^{(M)}} = \infty\}$, the skeleton $\mathbf{Y}$ is eventually constant, and since $Y_{T_i^{(M)}} = i$, this constant value must be i. This contradicts the fact that $\mathbf{X}$ is not absorbed in i and therefore $J_{T_i^{(M)}} < \infty$ $\mathbb{P}_i$-a.s. It follows that $J_{T_i^{(M)}+1} = D_{T_i^{(M)}+1} + J_{T_i^{(M)}} < \infty$ $\mathbb{P}_i$-a.s., and by (5.4) we obtain $\{t \in \mathbb{R}_+ : X_t = i\} \subseteq [0, J_{T_i^{(M)}+1})$ $\mathbb{P}_i$-a.s. Thus i is transient for $\mathbf{X}$.

(iii) By Definition I.3.1 of [Duq23], every state $i \in E$ is either recurrent or transient for the discrete-time Markov chain $\mathbf{Y}$. If i is recurrent for $\mathbf{Y}$, then (i) implies that i is recurrent for $\mathbf{X}$. If i is transient for $\mathbf{Y}$, then (ii) implies that i is transient for $\mathbf{X}$. Therefore every state $i \in E$ is either recurrent or transient for $\mathbf{X}$.

(iv) By Proposition 5.2, communication between i and j for the process $\mathbf{X}$ is equivalent to communication between i and j for the chain $\mathbf{Y}$. Therefore if $i \leftrightarrow j$ for $\mathbf{X}$, then $i \leftrightarrow j$ for $\mathbf{Y}$ and

by Lemma I.3.4 of [Duq23], i and j are either both recurrent or both transient for $\mathbf{Y}$. Applying (i) and (ii) then shows that i and j are either both recurrent or both transient for $\mathbf{X}$.

(v) Assume that G is irreducible and fix $i \in E$. By (iii), i is either recurrent or transient for $\mathbf{X}$. For any $j \in E$, irreducibility gives $i \leftrightarrow j$, and then (iv) implies that j has the same classification (recurrent or transient) as i. Hence either all states are recurrent or all states are transient for $\mathbf{X}$. □

Using (iii) together with the analogous statement for the chain $\mathbf{Y}$ (see Definition I.3.1 of [Duq23]), the contrapositive of (ii) implies that if $i \in E$ is recurrent (i.e. not transient) for $\mathbf{X}$, then i is recurrent for $\mathbf{Y}$. Similarly, the contrapositive of (i) implies that if $i \in E$ is transient (i.e. not recurrent) for $\mathbf{X}$, then i is transient for $\mathbf{Y}$. This yields, for any $i \in E$, the equivalence

$$i \text{ is recurrent for } \mathbf{Y} \iff i \text{ is recurrent for } \mathbf{X}$$

and likewise for transience. Thus the classification of states for the Markov process $\mathbf{X}$ and for the skeleton chain $\mathbf{Y}$ coincides, and we simply say that a state is recurrent or transient, without specifying whether this refers to $\mathbf{X}$ or to $\mathbf{Y}$.

Next, for $i \in E$, define

$$T_i^+(\mathbf{X}) := \inf\{t \geq J_1 : X_t = i\}$$

with the usual convention $\inf \emptyset := \infty$. Under $\mathbb{P}_i$, this is called the *first return time* of $\mathbf{X}$ to i, since the condition $t \geq J_1$ ensures that at least one jump of the process has occurred. If no jump occurs (i.e. if the initial state i is absorbing), then $T_i^+(\mathbf{X}) = \infty$. For simplicity, we write $T_i^+ := T_i^+(\mathbf{X})$. The following proposition characterizes recurrence and transience through the return time T_i^+ and provides a useful analytic formulation in terms of the semigroup $(P_t)_{t\in\mathbb{R}_+}$.

Proposition 5.5. *Let $i \in E$. We distinguish two cases:*

(I) If $q(i) = 0$ or $\mathbb{P}_i(T_i^+ < \infty) = 1$, then i is recurrent and

$$\mathbb{E}_i\Big[\int_0^\infty \mathrm{d}t\, \mathbf{1}_{\{X_t=i\}}\Big] = \int_0^\infty \mathrm{d}t\, p_t(i,i) = \infty.$$

(II) If $q(i) > 0$ and $\mathbb{P}_i(T_i^+ < \infty) < 1$, then i is transient and

$$\mathbb{E}_i\Big[\int_0^\infty \mathrm{d}t\, \mathbf{1}_{\{X_t=i\}}\Big] = \int_0^\infty \mathrm{d}t\, p_t(i,i) < \infty.$$

Proof. (I) If $q(i) = 0$, then i is absorbing and therefore $\mathbb{P}_i(\forall\, t \in \mathbb{R}_+, X_t = i) = 1$. Hence the set $\{t \in \mathbb{R}_+ : X_t = i\}$ is $\mathbb{P}_i$-a.s. unbounded and i is recurrent.

If $\mathbb{P}_i(T_i^+ < \infty) = 1$, then i cannot be absorbing (since in that case $T_i^+ = \infty$) and thus $q(i) > 0$. We define the first return time to i for the skeleton chain $\mathbf{Y}$ by $T_i^{(1)}(\mathbf{Y}) := \inf\{n > 0 : Y_n = i\}$. Since i is not absorbing, we have $\mathbb{P}_i$-a.s. that $T_i^{(1)}(\mathbf{Y}) < \infty$ if and only if $T_i^+ < \infty$, and therefore $\mathbb{P}_i(T_i^{(1)}(\mathbf{Y}) < \infty) = 1$. By Definition I.3.1 of [Duq23], this implies that i is recurrent.

Moreover, we have

$$\int_0^\infty \mathrm{d}t\, \mathbf{1}_{\{X_t=i\}} = \sum_{n\in\mathbb{N}} D_{n+1} \mathbf{1}_{\{Y_n=i\}}.$$

Taking expectation of the left-hand side and applying Tonelli's theorem, we obtain

$$\mathbb{E}_i\Big[\int_0^\infty \mathrm{d}t\, \mathbf{1}_{\{X_t=i\}}\Big] = \int_0^\infty \mathrm{d}t\, \mathbb{E}_i[\mathbf{1}_{\{X_t=i\}}] = \int_0^\infty \mathrm{d}t\, p_t(i,i).$$

Taking expectation of the right-hand side and noting that $D_{n+1} \mid \{Y_n = i\} \sim \mathrm{Exp}(q(i))$ by Theorem 3.38, we get

$$\sum_{n\in\mathbb{N}} \mathbb{E}_i[D_{n+1}\mathbf{1}_{\{Y_n=i\}}] = \sum_{n\in\mathbb{N}} \mathbb{E}_i[D_{n+1} \mid Y_n = i]\mathbb{P}_i(Y_n = i) = \tfrac{1}{q(i)} \sum_{n\in\mathbb{N}} \mathbb{P}_i(Y_n = i) = \tfrac{1}{q(i)} \sum_{n\in\mathbb{N}} [Q^n](i,i)$$

where the last equality follows from Section I.2.a of [Duq23]. Putting everything together and applying Corollary I.3.3 of [Duq23], we conclude that

$$\int_0^\infty \mathrm{d}t\, p_t(i,i) = \tfrac{1}{q(i)} \sum_{n\in\mathbb{N}} [Q^n](i,i) = \infty.$$

(II) Assume that $q(i) > 0$ (i.e. i is not absorbing) and that $\mathbb{P}_i(T_i^+ < \infty) < 1$. As shown in (I), we have $\mathbb{P}_i$-a.s. that $T_i^{(1)}(\mathbf{Y}) < \infty$ if and only if $T_i^+ < \infty$. Hence $\mathbb{P}_i(T_i^{(1)}(\mathbf{Y}) < \infty) < 1$ and by Definition I.3.1 of [Duq23], the state i is transient.

Moreover, by the same computations as in (I), we have

$$\int_0^\infty \mathrm{d}t\, p_t(i,i) = \tfrac{1}{q(i)} \sum_{n\in\mathbb{N}} [Q^n](i,i)$$

which is finite by Corollary I.3.3 of [Duq23] (since $Q^0 = \mathrm{Id}_E$). □

Observe that (I) shows that if a state $i \in E$ is absorbing, then it is recurrent. The converse does not hold since the process may leave i after each visit and still return to i infinitely often, in which case i is recurrent but not absorbing. Taking the contrapositive of (I), we see that if a state $i \in E$ is transient, then it cannot be absorbing. Again, the converse fails since transience requires not only $q(i) > 0$ but also $\mathbb{P}_i(T_i^+ < \infty) < 1$, as shown in (II).

5.2 Criteria for Non-Explosion

We now continue our qualitative study of Markov processes by investigating explosion. We first establish simple sufficient conditions ensuring that the process does not explode and then derive a necessary and sufficient criterion characterizing non-explosion.

Throughout this section, we work in the same setting and with the same notation as in the previous section. The space E is a countable space equipped with the discrete topology, and hence endowed with the σ-algebra $\mathcal{P}(E)$. $G = (q_{i,j})_{i,j\in E}$ is an infinitesimal generator on E and $\mathbf{X}$ denotes the minimal right-continuous Markov process on E^* associated with G.

Proposition 5.6. *Let $\mu \in \mathcal{M}_1(E)$, so that the process starts in E and not at the cemetery state ∂. If one of the following conditions holds, then $\mathbb{P}_\mu(\zeta < \infty) = 0$, that is, the process $\mathbf{X}$ with initial distribution μ does not explode $\mathbb{P}_\mu$-a.s.*

(a) E is finite.

(b) The quantity $c := \sup_{i\in E} q(i)$ is finite.

(c) $\mu = \delta_i$ and i is recurrent.

Proof. We first show that (a) implies (b). If E is finite, then $\{q(i) : i \in E\}$ is a finite set. Note that although holding times are allowed to have extended exponential distributions with parameter $q(i) = \infty$, which would correspond to zero holding times and hence instantaneous jumps, the right-continuity of the minimal process excludes such behavior before explosion. This implies that $q(i) < \infty$ for all $i \in E$, and therefore $c = \sup_{i\in E} q(i) = \max_{i\in E} q(i) < \infty$.

Next, we prove that (b) implies $\mathbb{P}_\mu(\zeta < \infty) = 0$. Assume $c = \sup_{i\in E} q(i) < \infty$. By the construction of the minimal process associated with G (see Section 4.1), there exists a sequence of random variables $(\mathbf{e}_n)_{n\geq 1}$ taking values in $(0,\infty)$ which, under $\mathbb{P}_\mu$, are i.i.d. exponentially distributed with parameter 1 and independent of the skeleton chain $\mathbf{Y}$. Moreover, $D_n = \frac{\mathbf{e}_n}{q(Y_{n-1})}$ for $n \geq 1$, with the convention $\frac{\mathbf{e}_n}{0} := \infty$. Then $\mathbb{P}_\mu$-a.s.

$$\zeta = \sum_{n\geq 1} D_n = \sum_{n\geq 1} \tfrac{\mathbf{e}_n}{q(Y_{n-1})} \geq \tfrac{1}{c} \sum_{n\geq 1} \mathbf{e}_n = \infty$$

where the last step follows from the law of large numbers. Indeed, since $(\mathbf{e}_n)_{n\geq 1}$ are i.i.d. with finite mean $\mathbb{E}_\mu[\mathbf{e}_1] = 1$, we have

$$\tfrac{1}{n}\sum_{k=1}^{n} \mathbf{e}_k \xrightarrow[n\to\infty]{} 1 \quad \mathbb{P}_\mu\text{-a.s.}$$

which implies that $\sum_{k\geq 1} \mathbf{e}_k = \infty$ $\mathbb{P}_\mu$-a.s.

For the implication from (c) to $\mathbb{P}_\mu(\zeta < \infty) = 0$, we refer to the proof of point (i) of Theorem 5.4, where we assumed that i is recurrent for $\mathbf{Y}$ and then showed that $\mathbb{P}_i(\zeta = \infty) = 1$. □

We now give a complete characterization of non-explosion for Markov processes. While the previous proposition provides simple sufficient conditions ensuring non-explosion, we next establish a criterion that is both necessary and sufficient and is formulated directly in terms of the infinitesimal generator. We first introduce the terminology used in this criterion.

An infinitesimal generator G is said to be *explosive* if the associated minimal Markov process $\mathbf{X}$ explodes, that is, if there exists a state $i \in E$ such that $\mathbb{P}_i(\zeta < \infty) > 0$. Conversely, a generator G is said to be *non-explosive* if $\mathbb{P}_i(\zeta = \infty) = 1$ for all $i \in E$. We also recall that a function $f : E \to \mathbb{R}$ is bounded if and only if $\|f\|_\infty = \sup_{i\in E} |f(i)| < \infty$.

Theorem 5.7 (Reuter's criterion)**.** *The generator $G = (q_{i,j})_{i,j\in E}$ is non-explosive if and only if, for every real $c > 0$, the only bounded nonnegative solution of the equation*

$$\sum_{j\in E} q_{i,j} f(j) = c f(i), \quad i \in E \tag{5.5}$$

is the zero function (i.e. $f(i) = 0$ for all $i \in E$).

Proof. For every real $c > 0$ and every $i \in E$, define $g_c(i) := \mathbb{E}_i[e^{-c\zeta}]$ with the convention $e^{-\infty} := 0$. Since $\zeta \in [0, \infty]$, we have $0 \leq g_c(i) \leq 1$ for all $i \in E$, so the function g_c is bounded and nonnegative. We first show that g_c is a solution of the equation (5.5). Observe that $\zeta = \zeta(\mathbf{X}) = J_1 + \zeta(\theta_{J_1}\mathbf{X})$, so that for all $i \in E$

$$\begin{aligned} g_c(i) &= \mathbb{E}_i[e^{-c(J_1+\zeta(\theta_{J_1}\mathbf{X}))}] \\ &= \mathbb{E}_i[e^{-cJ_1} e^{-c\zeta(\theta_{J_1}\mathbf{X})}] \\ &= \mathbb{E}_i[e^{-cJ_1}\mathbb{E}_i[e^{-c\zeta(\theta_{J_1}\mathbf{X})} \mid \mathcal{F}_{J_1}]] \end{aligned}$$

where we used the weak tower property and the fact that J_1 is $\mathcal{F}_{J_1}$-measurable. By the strong Markov property at J_1, we have

$$\begin{aligned} \mathbb{E}_i[e^{-cJ_1}\mathbb{E}_i[e^{-c\zeta(\theta_{J_1}\mathbf{X})} \mid \mathcal{F}_{J_1}]] &= \mathbb{E}_i[e^{-cJ_1}\mathbb{E}_i[\mathbf{1}_{\{J_1<\infty\}} e^{-c\zeta(\theta_{J_1}\mathbf{X})} \mid \mathcal{F}_{J_1}]] \\ &= \mathbb{E}_i[e^{-cJ_1}\mathbf{1}_{\{J_1<\infty\}}\mathbb{E}_{X_{J_1}}[e^{-c\zeta(\mathbf{X})}]] \\ &= \mathbb{E}_i[e^{-cJ_1}\mathbb{E}_{Y_1}[e^{-c\zeta}]] \end{aligned}$$

where we used that $e^{-cJ_1} = 0$ on the event $\{J_1 = \infty\}$, since $c > 0$. Using the independence of J_1 and Y_1 under $\mathbb{P}_i$ (see Proposition 3.36), we obtain

$$\begin{aligned} \mathbb{E}_i[e^{-cJ_1}\mathbb{E}_{Y_1}[e^{-c\zeta}]] &= \mathbb{E}_i[e^{-cJ_1}]\mathbb{E}_i[\mathbb{E}_{Y_1}[e^{-c\zeta}]] \\ &= \mathbb{E}_i[e^{-cJ_1}]\sum_{j\in E} \mathbb{P}_i(Y_1 = j)\mathbb{E}_j[e^{-c\zeta}] \\ &= \mathbb{E}_i[e^{-cJ_1}]\sum_{j\in E} \pi_{i,j} g_c(j). \end{aligned}$$

Moreover, since J_1 has an (extended) exponential distribution with parameter $q(i)$ under $\mathbb{P}_i$ (see

Proposition 3.36), we have

$$\begin{aligned}\mathbb{E}_i[e^{-cJ_1}] &= \int_{-\infty}^{\infty} \mathbb{P}_i(J_1 \in \mathrm{d}t) e^{-ct} \\ &= q(i) \int_0^{\infty} \mathrm{d}t\, e^{-(q(i)+c)t} \\ &= q(i) \Big(-\tfrac{e^{-(q(i)+c)t}}{q(i)+c} \Big) \Big|_{t=0}^{t=\infty} \\ &= \tfrac{q(i)}{q(i)+c}. \end{aligned} \tag{5.6}$$

Combining the previous displays yields

$$(q(i)+c) g_c(i) = \sum_{j \in E} q(i) \pi_{i,j} g_c(j) \quad \text{for all } i \in E. \tag{5.7}$$

If $q(i) > 0$, then by Definition 4.1 we have $\pi_{i,j} = \frac{q_{i,j}}{q(i)}$ for all $j \neq i$ and $\pi_{i,i} = 0$. Substituting this into (5.7) yields

$$(q(i)+c) g_c(i) = \sum_{j \in E \setminus \{i\}} q_{i,j} g_c(j).$$

Using that $q(i) = -q_{i,i}$ for all $i \in E$, we obtain

$$\sum_{j \in E} q_{i,j} g_c(j) = c g_c(i) \tag{5.8}$$

which shows that in this case g_c is a solution of (5.5).

If $q(i) = 0$, then $\mathbb{P}_i(J_1 = \infty) = 1$ and therefore $\zeta = \sup_{n \in \mathbb{N}} J_n = \infty$ $\mathbb{P}_i$-a.s. This implies that $g_c(i) = \mathbb{E}_i[e^{-c\zeta}] = 0$. Moreover, by Definition 3.37, we have $q_{i,j} = 0$ for all $j \in E$. Consequently, both sides of (5.8) vanish, and the equation holds in this case as well. Hence g_c is a solution of (5.5) for all $i \in E$.

Now we prove that if the only bounded nonnegative solution of (5.5) is the zero function, then G is non-explosive. We argue by contraposition and assume that G is explosive. Then there exists $i \in E$ such that $\mathbb{P}_i(\zeta < \infty) > 0$, which implies that $g_c(i) > 0$. Indeed, since $c > 0$, we have $e^{-c\zeta} = 0$ on the event $\{\zeta = \infty\}$, and thus $g_c(i) = \mathbb{E}_i[e^{-c\zeta} \mathbf{1}_{\{\zeta < \infty\}}]$. Moreover, because $\mathbb{P}_i(\zeta < \infty) > 0$, we may choose $M > 0$ such that $\mathbb{P}_i(\zeta \leq M) > 0$. Then

$$g_c(i) \geq \mathbb{E}_i[e^{-c\zeta} \mathbf{1}_{\{\zeta \leq M\}}] \geq e^{-cM} \mathbb{P}_i(\zeta \leq M) > 0.$$

Thus there exists $i \in E$ such that $g_c(i) > 0$, so g_c is a bounded nonnegative solution of (5.5) which is not identically zero.

Finally, we show that if G is non-explosive, then the only bounded nonnegative solution of (5.5) is the zero function. We again argue by contraposition and assume that (5.5) admits a bounded nonnegative solution f which is not identically zero. Rewriting (5.5) using the relations $q_{i,i} = -q(i)$ for all $i \in E$ and $q_{i,j} = q(i)\pi_{i,j}$ for $j \neq i$ in E (see Definition 3.37) yields

$$(q(i)+c) f(i) = q(i) \sum_{j \in E \setminus \{i\}} \pi_{i,j} f(j) \quad \text{for all } i \in E.$$

Since by Definition 4.1 we have $\pi_{i,i} = 0$ whenever $q(i) > 0$, we can extend the sum in the last display to the whole set E and obtain

$$f(i) = \tfrac{q(i)}{q(i)+c} \sum_{j \in E} \pi_{i,j} f(j).$$

Using (5.6) and the independence of J_1 and Y_1 under $\mathbb{P}_i$, we get for all $i \in E$

$$\tfrac{q(i)}{q(i)+c} \sum_{j \in E} \pi_{i,j} f(j) = \mathbb{E}_i[e^{-cJ_1}] \mathbb{E}_i[f(Y_1)] = \mathbb{E}_i[e^{-cJ_1} f(Y_1)].$$

Next, we want to prove by induction that $f(i) = \mathbb{E}_i[e^{-cJ_n} f(Y_n)]$ for all $i \in E$ and all $n \geq 1$. We already proved the base case $f(i) = \mathbb{E}_i[e^{-cJ_1} f(Y_1)]$. Now we assume that the statement holds for some $n \geq 1$ and we prove it for $n+1$. Applying the base case at the state Y_n yields $f(Y_n) = \mathbb{E}_{Y_n}[e^{-cJ_1} f(Y_1)]$. Multiplying by the indicator $\mathbf{1}_{\{J_n<\infty\}}$ and using the strong Markov property at J_n, we obtain

$$\begin{aligned}\mathbf{1}_{\{J_n<\infty\}} f(Y_n) &= \mathbf{1}_{\{J_n<\infty\}} \mathbb{E}_{Y_n}[e^{-cJ_1} f(Y_1)] \\ &= \mathbb{E}_i[\mathbf{1}_{\{J_n<\infty\}} e^{-c(J_{n+1}-J_n)} f(Y_{n+1}) \mid \mathcal{F}_{J_n}].\end{aligned}$$

Multiplying both sides by e^{-cJ_n} and then taking expectations, the indicator becomes redundant, since, for $c > 0$, we have $e^{-cJ_n} = 0$ on the event $\{J_n = \infty\}$. Hence

$$\begin{aligned}\mathbb{E}_i[e^{-cJ_n} f(Y_n)] &= \mathbb{E}_i[e^{-cJ_n} \mathbb{E}_i[e^{-c(J_{n+1}-J_n)} f(Y_{n+1}) \mid \mathcal{F}_{J_n}]] \\ &= \mathbb{E}_i[\mathbb{E}_i[e^{-cJ_{n+1}} f(Y_{n+1}) \mid \mathcal{F}_{J_n}]].\end{aligned}$$

By the weak tower property and the induction assumption, we thus have $f(i) = \mathbb{E}_i[e^{-cJ_{n+1}} f(Y_{n+1})]$, which completes the induction. Because f is bounded, we have for all $i \in E$ and all $n \geq 1$

$$f(i) = \mathbb{E}_i[e^{-cJ_n} f(Y_n)] \leq \|f\|_\infty \mathbb{E}_i[e^{-cJ_n}].$$

Since $e^{-cJ_n} \to e^{-c\zeta}$ as $n \to \infty$ and $0 \leq e^{-cJ_n} \leq 1$ for all $n \geq 1$, dominated convergence yields

$$f(i) \leq \|f\|_\infty \mathbb{E}_i[e^{-c\zeta}] = \|f\|_\infty g_c(i) \quad \text{for all } i \in E.$$

Because f is nonnegative and not identically zero, there exists $i_0 \in E$ such that $f(i_0) > 0$, and the last display then implies that $g_c(i_0) > 0$. Since $c > 0$, we thus have $g_c(i_0) = \mathbb{E}_{i_0}[e^{-c\zeta} \mathbf{1}_{\{\zeta<\infty\}}] > 0$, which yields $\mathbb{P}_{i_0}(\zeta < \infty) > 0$. Indeed, by contradiction, if $\mathbb{P}_{i_0}(\zeta = \infty) = 1$, then, since $c > 0$, we have $g_c(i_0) = \mathbb{E}_{i_0}[e^{-c\zeta}] = 0$. This is a contradiction, and therefore G is explosive. □

Using the action of G on $f : E \to \mathbb{R}$ defined by $(Gf)(i) := \sum_{j \in E} q_{i,j} f(j)$ (i.e. the usual matrix-vector multiplication extended to the case where E is countably infinite), equation (5.5) can be written as

$$(Gf)(i) = cf(i), \quad i \in E$$

which is the eigenvalue equation for the generator G. In particular, Reuter's criterion states that non-explosion is equivalent to the absence of bounded nonnegative eigenfunctions of G that are not identically zero and correspond to positive eigenvalues.

Example 5.8. Using Reuter's criterion from Theorem 5.7, we derive a necessary and sufficient condition for the non-explosion of birth-death processes. A *birth-death process* is a continuous-time Markov process $\mathbf{X} = (X_t)_{t \in \mathbb{R}_+}$ with state space $E := \mathbb{N}$ which evolves by nearest-neighbor jumps only. More precisely, when the process is in state i, it may jump to $i+1$ (a birth) at rate $\lambda_i \geq 0$ or, if $i \geq 1$, to $i-1$ (a death) at rate $\mu_i \geq 0$, and no other transitions are possible.

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{R}_+}, \mathbf{X} = (X_t)_{t \in \mathbb{R}_+}, G, \mathbb{P}_\mu, \mu \in \mathcal{M}_1(\mathbb{N}))$ be a birth-death process with skeleton chain $\mathbf{Y} = (Y_n)_{n \in \mathbb{N}}$ and family of holding time parameters $(q(i))_{i \in \mathbb{N}}$. We first want to determine its infinitesimal generator $G = (q_{i,j})_{i,j \in \mathbb{N}}$. Since starting from i the chain $\mathbf{Y}$ can only move to a nearest neighbor, we have $\pi_{i,j} := \mathbb{P}_i(Y_1 = j) = 0$ whenever $|j - i| > 1$, and by Definition 3.37 this implies $q_{i,j} = 0$ for all such $i, j \in \mathbb{N}$ with $|j - i| > 1$. For the diagonal entries, we have $q_{i,i} = -q(i) = -(\lambda_i + \mu_i)$ for $i \geq 1$ and $q_{0,0} = -q(0) = -\lambda_0$, since by Theorem 3.38 $q(i)$ is the total rate at which the process leaves state i, which equals the sum of the rates of the only possible transitions out of i, namely λ_i and, for $i \geq 1$, μ_i.

Moreover, starting from i, the probability that the next jump is upward (resp. downward) is equal to the corresponding rate divided by the total rate $q(i)$. That is, $\pi_{i,i+1} = \mathbb{P}_i(Y_1 = i+1) = \lambda_i / q(i)$ for $i \in \mathbb{N}$ and $\pi_{i,i-1} = \mathbb{P}_i(Y_1 = i-1) = \mu_i / q(i)$ for $i \geq 1$. Hence, again by Definition 3.37, we obtain $q_{i,i+1} = q(i)\pi_{i,i+1} = \lambda_i$ for $i \in \mathbb{N}$ and $q_{i,i-1} = q(i)\pi_{i,i-1} = \mu_i$ for $i \geq 1$. Therefore the generator matrix has the form

$$G = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \cdots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & \cdots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots \end{pmatrix}.$$

We now assume that the generator G is irreducible, that is, $i \leftrightarrow j$ for all $i, j \in \mathbb{N}$. By point (iii) of Proposition 5.2, this is equivalent to requiring that for any $i, j \in \mathbb{N}$ there exists a finite path from i to j along which all corresponding generator coefficients are strictly positive. In the case of a birth-death process, where only transitions between nearest neighbors are allowed, this condition reduces to $q_{i,i+1}q_{i+1,i} > 0$ for all $i \in \mathbb{N}$.

Fix $c > 0$ and assume that there exists a nonnegative function $f : \mathbb{N} \to \mathbb{R}_+$ such that $Gf = cf$ (i.e. f is a solution of (5.5) for G). This is equivalent to the system

$$\begin{aligned} cf(0) &= (Gf)(0) = q_{0,1}f(1) + q_{0,0}f(0) = q_{0,1}(f(1) - f(0)), \\ cf(i) &= (Gf)(i) = q_{i,i+1}f(i+1) + q_{i,i}f(i) + q_{i,i-1}f(i-1) \\ &= q_{i,i+1}f(i+1) - (q_{i,i+1} + q_{i,i-1})f(i) + q_{i,i-1}f(i-1), \quad i \geq 1. \end{aligned} \tag{5.9}$$

Notice the following consequences:

- If $f(0) = 0$, then by the first identity in (5.9), we have $f(1) = 0$ since $q_{0,1} > 0$. Applying the second identity in (5.9) for $i = 1$ and the fact that $q_{1,2} > 0$, we obtain that $f(2) = 0$. Iterating this argument yields $f(i) = 0$ for all $i \geq 1$.
- If there exists $i \geq 1$ such that $f(i) > 0$ (i.e. f is not identically zero), then by contraposition of the first point we have $f(0) > 0$.
- Since f solves (5.5) (i.e. $Gf = cf$), it follows that af also solves (5.5) (i.e. $Gaf = caf$) for any $a \geq 0$.
- If f is not identically zero, then we may assume without loss of generality that $f(0) = 1$. Indeed, by the second point we have $f(0) > 0$, and by the third point replacing $f(i)$ with $f(i)/f(0)$ for all $i \in \mathbb{N}$ yields another solution of (5.5) satisfying $f(0) = 1$.

For $i \in \mathbb{N}$, define $g(i) := f(i+1) - f(i)$. By the second identity in (5.9), we have for all $i \geq 1$,

$$\begin{aligned} & cf(i) = q_{i,i+1}f(i+1) - q_{i,i+1}f(i) - q_{i,i-1}f(i) + q_{i,i-1}f(i-1) \\ \iff \quad & \tfrac{c}{q_{i,i+1}}f(i) = f(i+1) - f(i) - \tfrac{q_{i,i-1}}{q_{i,i+1}}f(i) + \tfrac{q_{i,i-1}}{q_{i,i+1}}f(i-1) \\ \iff \quad & f(i+1) - f(i) = \tfrac{c}{q_{i,i+1}}f(i) + \tfrac{q_{i,i-1}}{q_{i,i+1}}(f(i) - f(i-1)) \\ \iff \quad & g(i) = \tfrac{c}{q_{i,i+1}}f(i) + \tfrac{q_{i,i-1}}{q_{i,i+1}}g(i-1). \end{aligned} \tag{5.10}$$

Then we can show by induction on i that for all $i \geq 1$,

$$\begin{aligned} g(i) = {} & \tfrac{c}{q_{i,i+1}}f(i) + \tfrac{cq_{i,i-1}}{q_{i,i+1}q_{i-1,i}}f(i-1) + \tfrac{cq_{i,i-1}q_{i-1,i-2}}{q_{i,i+1}q_{i-1,i}q_{i-2,i-1}}f(i-2) + \cdots \\ & + \tfrac{cq_{i,i-1}\cdots q_{2,1}}{q_{i,i+1}q_{i-1,i}\cdots q_{1,2}}f(1) + \tfrac{cq_{i,i-1}\cdots q_{1,0}}{q_{i,i+1}\cdots q_{0,1}}f(0). \end{aligned} \tag{5.11}$$

Indeed, from the first identity in (5.9), we have $g(0) = f(1) - f(0) = \frac{c}{q_{0,1}}f(0)$. Using this and (5.10), we obtain the base case $i = 1$:

$$g(1) = \tfrac{c}{q_{1,2}}f(1) + \tfrac{q_{1,0}}{q_{1,2}}g(0) = \tfrac{c}{q_{1,2}}f(1) + \tfrac{cq_{1,0}}{q_{1,2}q_{0,1}}f(0).$$

For the induction step, assume that the expansion holds for $g(i-1)$, i.e.

$$\begin{aligned} g(i-1) = {} & \tfrac{c}{q_{i-1,i}}f(i-1) + \tfrac{cq_{i-1,i-2}}{q_{i-1,i}q_{i-2,i-1}}f(i-2) + \cdots + \tfrac{cq_{i-1,i-2}\cdots q_{2,1}}{q_{i-1,i}q_{i-2,i-1}\cdots q_{1,2}}f(1) \\ & + \tfrac{cq_{i-1,i-2}\cdots q_{1,0}}{q_{i-1,i}q_{i-2,i-1}\cdots q_{0,1}}f(0). \end{aligned}$$

Substituting this into (5.10) and distributing the factor $\frac{q_{i,i-1}}{q_{i,i+1}}$ yields exactly (5.11).

Since all fractions in (5.11) are positive and all values of f are nonnegative, (5.11) implies that $g(i) \geq 0$ for all $i \geq 1$. Moreover, from the first identity in (5.9) we have $g(0) = \frac{c}{q_{0,1}}f(0) \geq 0$. Hence $g(i) \geq 0$ for all $i \in \mathbb{N}$. We can therefore deduce that every nonnegative solution f of (5.5)

is increasing, i.e. $f(i+1) \geq f(i)$ for all $i \in \mathbb{N}$. Consequently, $f(i) \geq f(0)$ for all $i \in \mathbb{N}$ and thus (5.11) yields

$$\begin{aligned} g(i) \geq cf(0)\Big(&\tfrac{1}{q_{i,i+1}} + \tfrac{q_{i,i-1}}{q_{i,i+1}q_{i-1,i}} + \tfrac{q_{i,i-1}q_{i-1,i-2}}{q_{i,i+1}q_{i-1,i}q_{i-2,i-1}} + \cdots \\ &+ \tfrac{q_{i,i-1}\cdots q_{2,1}}{q_{i,i+1}q_{i-1,i}\cdots q_{1,2}} + \tfrac{q_{i,i-1}\cdots q_{1,0}}{q_{i,i+1}\cdots q_{0,1}}\Big) \end{aligned} \tag{5.12}$$

for all $i \geq 1$. Define

$$S := \sum_{i=1}^{\infty}\Big(\tfrac{1}{q_{i,i+1}} + \tfrac{q_{i,i-1}}{q_{i,i+1}q_{i-1,i}} + \cdots + \tfrac{q_{i,i-1}\cdots q_{2,1}}{q_{i,i+1}q_{i-1,i}\cdots q_{1,2}} + \tfrac{q_{i,i-1}\cdots q_{1,0}}{q_{i,i+1}\cdots q_{0,1}}\Big) \in (0,\infty].$$

Now assume that $S = \infty$. Suppose that there exists a bounded, nonnegative solution f of (5.5) which is not identically zero. By the fourth consequence listed above, we may assume without loss of generality that $f(0) = 1$. Since f is bounded, there exists $M > 0$ such that $f(i) \leq M$ for all $i \in \mathbb{N}$, and therefore

$$\sum_{i=0}^{n-1} g(i) = f(n) - f(0) \leq M - 1.$$

Since $g(i) \geq 0$ for all $i \in \mathbb{N}$ and the partial sums $\sum_{i=0}^{n-1} g(i)$ are uniformly bounded, it follows that $\sum_{i=0}^{\infty} g(i) < \infty$. On the other hand, by (5.12) and since $f(0) = 1$ we have

$$\sum_{i=0}^{\infty} g(i) \geq \sum_{i=1}^{\infty} g(i) \geq cS = \infty$$

which gives a contradiction. Hence, if $S = \infty$, there cannot exist a bounded, nonnegative solution of (5.5) which is not identically zero. Thus, by Reuter's criterion (Theorem 5.7), the generator G does not explode.

Conversely, let $f : \mathbb{N} \to \mathbb{R}_+$ satisfy (5.9) (or equivalently (5.5)) with $f(0) = 1$. Such a function can always be defined, since fixing $f(0) = 1$ and $c > 0$, the system (5.9) uniquely determines $f(1)$ from the equation at 0, then $f(2)$ from the equation at 1, and so on, defining $f(i)$ recursively for all $i \in \mathbb{N}$. The question is whether f can be bounded. The fact that f is increasing (as shown above), combined with (5.11), implies for all $i \geq 1$ that

$$f(i+1) - f(i) = g(i) \leq f(i)\Big(\tfrac{c}{q_{i,i+1}} + \tfrac{cq_{i,i-1}}{q_{i,i+1}q_{i-1,i}} + \cdots + \tfrac{cq_{i,i-1}\cdots q_{2,1}}{q_{i,i+1}q_{i-1,i}\cdots q_{1,2}} + \tfrac{cq_{i,i-1}\cdots q_{1,0}}{q_{i,i+1}\cdots q_{0,1}}\Big).$$

This yields for all $i \geq 1$

$$\begin{aligned} f(i+1) &\leq f(i)\Big(1 + \tfrac{c}{q_{i,i+1}} + \tfrac{cq_{i,i-1}}{q_{i,i+1}q_{i-1,i}} + \cdots + \tfrac{cq_{i,i-1}\cdots q_{2,1}}{q_{i,i+1}q_{i-1,i}\cdots q_{1,2}} + \tfrac{cq_{i,i-1}\cdots q_{1,0}}{q_{i,i+1}\cdots q_{0,1}}\Big) \\ &\leq f(i)e^{\frac{c}{q_{i,i+1}} + \frac{cq_{i,i-1}}{q_{i,i+1}q_{i-1,i}} + \cdots + \frac{cq_{i,i-1}\cdots q_{2,1}}{q_{i,i+1}q_{i-1,i}\cdots q_{1,2}} + \frac{cq_{i,i-1}\cdots q_{1,0}}{q_{i,i+1}\cdots q_{0,1}}} \end{aligned}$$

using the inequality $1 + x \leq e^x$ for all $x \in \mathbb{R}$, which follows by noting that the function $x \mapsto e^x - (1+x)$ is convex on $\mathbb{R}$, attains its minimum at 0 and satisfies $e^0 - (1+0) = 0$. Iterating the above inequality yields for all $i \in \mathbb{N}$

$$\begin{aligned} f(i) &\leq f(0)e^{\frac{c}{q_{0,1}}} \prod_{k=1}^{i-1} e^{\frac{c}{q_{k,k+1}} + \frac{cq_{k,k-1}}{q_{k,k+1}q_{k-1,k}} + \cdots + \frac{cq_{k,k-1}\cdots q_{2,1}}{q_{k,k+1}q_{k-1,k}\cdots q_{1,2}} + \frac{cq_{k,k-1}\cdots q_{1,0}}{q_{k,k+1}\cdots q_{0,1}}} \\ &= e^{\frac{c}{q_{0,1}}} e^{\sum_{k=1}^{i-1}\left(\frac{c}{q_{k,k+1}} + \frac{cq_{k,k-1}}{q_{k,k+1}q_{k-1,k}} + \cdots + \frac{cq_{k,k-1}\cdots q_{2,1}}{q_{k,k+1}q_{k-1,k}\cdots q_{1,2}} + \frac{cq_{k,k-1}\cdots q_{1,0}}{q_{k,k+1}\cdots q_{0,1}}\right)} \\ &\leq e^{\frac{c}{q_{0,1}}} e^{cS} \end{aligned}$$

where we used $f(0) = 1$ and the convention that empty products are equal to 1. Hence, if S is finite, then f is a bounded, nonnegative solution of (5.5) which is not identically zero (since $f(0) = 1$). Thus, by Reuter's criterion (Theorem 5.7), the generator G explodes.

In summary, we have shown that the irreducible generator G does not explode if and only if

$$\sum_{i=1}^{\infty}\Big(\tfrac{1}{q_{i,i+1}} + \tfrac{q_{i,i-1}}{q_{i,i+1}q_{i-1,i}} + \cdots + \tfrac{q_{i,i-1}\cdots q_{2,1}}{q_{i,i+1}q_{i-1,i}\cdots q_{1,2}} + \tfrac{q_{i,i-1}\cdots q_{1,0}}{q_{i,i+1}\cdots q_{0,1}}\Big) = \infty.$$

5.3 Invariant Measures and Positive Recurrence

The qualitative study of Markov processes naturally leads to the question of their long-term behavior. Rather than focusing on individual sample paths, it is often more informative to study how the process behaves on average. One way to do this is to look for measures on the state space that describe how the process distributes its time among the states. Such measures are called invariant measures. They express a balance in the jump dynamics of the process and remain unchanged as time evolves.

Throughout this section, we work in the same setting and with the same notation as in the previous two sections. The space E is a countable space equipped with the discrete topology, and hence endowed with the σ-algebra $\mathcal{P}(E)$. $\mathbf{X}$ denotes the minimal right-continuous Markov process on E^* associated with an infinitesimal generator G on E.

Definition 5.9 (Invariant measure)**.** Let $G = (q_{i,j})_{i,j\in E}$ be an infinitesimal generator on E. A measure $m = (m(i))_{i\in E}$ on E with values in $[0,\infty)$ is said to be *G-invariant* if it is not the zero measure and if $mG = 0$, i.e.

$$\sum_{i\in E} m(i)q_{i,j} = 0 \quad \text{for all } j \in E.$$

Since infinite sums mixing positive and negative terms may be ill defined (for instance expressions of the form $+\infty - \infty$), it is important to check that the condition $mG = 0$ is well defined. This is indeed the case here, since $mG = 0$ is equivalent to

$$m(j)q(j) = \sum_{i\in E\setminus\{j\}} m(i)q_{i,j} \quad \text{for all } j \in E$$

where both sides involve only nonnegative terms.

We have already seen that for $i \neq j$ in E, $q_{i,j}$ represents the instantaneous rate at which the process jumps from state i to state j. If $m(i)$ is interpreted as the mass located at state $i \in E$, then $m(i)q_{i,j}$ represents the rate at which mass is transferred from i to j. With this interpretation, the condition $mG = 0$ expresses a balance of flows. For each state $j \in E$, the total rate at which mass arrives at j, that is $\sum_{i\in E\setminus\{j\}} m(i)q_{i,j}$, is equal to the rate at which mass leaves j, namely $m(j)q(j)$.

The notion of G-invariance is closely connected to invariance for the transition matrix Q of the skeleton chain $\mathbf{Y}$. The following lemma makes this connection precise.

Lemma 5.10. *Let $G = (q_{i,j})_{i,j\in E}$ be an infinitesimal generator on E, which we assume to be irreducible. This implies that $q(i) > 0$ for all $i \in E$. Let $\mu \in \mathcal{M}_1(E)$ be related to a measure m on E by*

$$q(i)m(i) = \mu(i) \quad \textit{for all } i \in E. \tag{5.13}$$

Then m is G-invariant if and only if μ is Q-invariant, i.e. $\mu Q = \mu$ or equivalently

$$\sum_{i\in E} \mu(i)\pi_{i,j} = \mu(j) \quad \textit{for all } j \in E.$$

Proof. Since G is irreducible, there is no absorbing state and therefore $q(i) > 0$ for all $i \in E$. By Definition 4.1, it follows that $\pi_{i,i} = 0$ for all $i \in E$ and $q(i)\pi_{i,j} = q_{i,j}$ for all $i \neq j$ in E. Let $\mu \in \mathcal{M}_1(E)$ and let m be a measure on E such that they are related by (5.13). Using this relation and the identities established above, we obtain

$$\begin{aligned}
mG = 0 &\iff q(j)m(j) = \sum_{i\in E\setminus\{j\}} m(i)q_{i,j} \quad \text{for all } j \in E \\
&\iff q(j)m(j) = \sum_{i\in E} m(i)q(i)\pi_{i,j} \quad \text{for all } j \in E \\
&\iff \mu(j) = \sum_{i\in E} \mu(i)\pi_{i,j} \quad \text{for all } j \in E \\
&\iff \mu Q = \mu
\end{aligned}$$

which concludes the proof of the lemma. □

The previous lemma shows that G-invariant measures of the process $\mathbf{X}$ are in one-to-one correspondence with Q-invariant measures of the skeleton chain $\mathbf{Y}$. This allows us to transfer results on invariant measures for discrete-time Markov chains to the continuous-time setting, in particular the existence and uniqueness (up to multiplication by a constant) of invariant measures, as stated in the following proposition.

Proposition 5.11. *Let $G = (q_{i,j})_{i,j\in E}$ be an irreducible generator on E which is recurrent, in the sense that every state in E is recurrent for the minimal Markov process $\mathbf{X}$ associated with G. Then there exists a G-invariant measure m on E. Moreover, if m' is another G-invariant measure on E, then there exists a constant $c > 0$ such that $m'(i) = cm(i)$ for all $i \in E$. In other words, there exists a unique G-invariant measure on E up to multiplication by a positive constant. Furthermore, for all $i \in E$, $0 < m(i) < \infty$, and for any $i_0 \in E$,*

$$m(i) = m(i_0)q(i_0)\mathbb{E}_{i_0}\Big[\int_0^{T^+_{i_0}} \mathrm{d}t\, \mathbf{1}_{\{X_t=i\}}\Big] \quad \text{for all } i \in E \tag{5.14}$$

where we recall that $T^+_{i_0} = \inf\{t \geq J_1 : X_t = i_0\}$.

Proof. Since G is irreducible, all states in E communicate for the minimal Markov process $\mathbf{X}$ associated with G. By Proposition 5.2, this is equivalent to the fact that all states in E communicate for the discrete-time skeleton chain $\mathbf{Y}$ associated with Q. Thus the transition matrix Q is irreducible.

Moreover, since G is recurrent, every state in E is recurrent for the minimal Markov process $\mathbf{X}$ associated with G. By Theorem 5.4, this implies that every state in E is recurrent for the discrete-time skeleton chain $\mathbf{Y}$. Consequently, the transition matrix Q is also recurrent.

We first show the existence of a G-invariant measure on E. By Theorem I.4.2 in Section I.4.a of [Duq23], there exists a Q-invariant measure μ on E. Defining $m(i) := \frac{\mu(i)}{q(i)}$ for all $i \in E$, Lemma 5.10 implies that the measure m is G-invariant.

To prove uniqueness of m up to multiplication by a positive constant, let m' be another G-invariant measure on E and define $\mu'(i) := q(i)m'(i)$. By Lemma 5.10, the measure μ' is Q-invariant. By point (ii) of Theorem I.4.3 in Section I.4.a of [Duq23], there exists a constant $c > 0$ such that $\mu'(i) = c\mu(i)$ for all $i \in E$. This is equivalent to $q(i)m'(i) = cq(i)m(i)$ for all $i \in E$, and dividing by $q(i)$ yields $m'(i) = cm(i)$ for all $i \in E$. This proves uniqueness of m up to multiplication by a positive constant.

We now prove (5.14). Define $T^{(1)}_{i_0}(\mathbf{Y}) := \inf\{n > 0 : Y_n = i_0\}$, which represents the first return time to i_0 for the skeleton chain $\mathbf{Y}$ under $\mathbb{P}_{i_0}$. Since G is recurrent, Proposition 5.6 implies that the associated Markov process $\mathbf{X}$ does not explode. Fix $i_0 \in E$. Because G is irreducible, the state i_0 is not absorbing, which implies $\mathbb{P}_{i_0}$-a.s. that $T^{(1)}_{i_0}(\mathbf{Y}) < \infty$ if and only if $T^+_{i_0} < \infty$. Moreover, since Q is recurrent, the state i_0 is recurrent for the skeleton chain $\mathbf{Y}$, and therefore

$$\mathbb{P}_{i_0}(T^{(1)}_{i_0}(\mathbf{Y}) < \infty) = \mathbb{P}_{i_0}(T^+_{i_0} < \infty) = 1.$$

We then observe that $\mathbb{P}_{i_0}$-a.s. for all $i \in E$

$$\begin{aligned}
\int_0^{T^+_{i_0}} \mathrm{d}t\, \mathbf{1}_{\{X_t=i\}} &= \sum_{n\geq 0} \int_{J_n}^{J_{n+1}} \mathrm{d}t\, \mathbf{1}_{\{X_t=i\}} \mathbf{1}_{\{J_{n+1}\leq T^+_{i_0}\}} \\
&= \sum_{n\geq 0} \int_{J_n}^{J_{n+1}} \mathrm{d}t\, \mathbf{1}_{\{Y_n=i\}} \mathbf{1}_{\{J_{n+1}\leq J_{T^{(1)}_{i_0}(\mathbf{Y})}\}} \\
&= \sum_{n=0}^{T^{(1)}_{i_0}(\mathbf{Y})-1} \int_{J_n}^{J_{n+1}} \mathrm{d}t\, \mathbf{1}_{\{Y_n=i\}} \\
&= \sum_{n=0}^{T^{(1)}_{i_0}(\mathbf{Y})-1} (J_{n+1} - J_n)\mathbf{1}_{\{Y_n=i\}} \\
&= \sum_{n\geq 0} D_{n+1} \mathbf{1}_{\{Y_n=i, n<T^{(1)}_{i_0}(\mathbf{Y})\}}
\end{aligned}$$

where we used the identity $T_{i_0}^+ = J_{T_{i_0}^{(1)}(\mathbf{Y})}$. Taking expectation on both sides and applying Tonelli's theorem on the right-hand side, we obtain for all $i \in E$

$$\mathbb{E}_{i_0}\Big[\int_0^{T_{i_0}^+} \mathrm{d}t\, \mathbf{1}_{\{X_t=i\}}\Big] = \sum_{n\geq 0} \mathbb{E}_{i_0}[D_{n+1}\mathbf{1}_{\{Y_n=i, n<T_{i_0}^{(1)}(\mathbf{Y})\}}].$$

Theorem 3.38 implies that under $\mathbb{P}_{i_0}$, conditionally on $\mathbf{Y}$, the random variables $(D_n)_{n\geq 1}$ are exponentially distributed with respective parameters $q(Y_{n-1})$. Hence, using the weak tower property, we get for all $i \in E$

$$\begin{aligned}
\sum_{n\geq 0} \mathbb{E}_{i_0}[D_{n+1}\mathbf{1}_{\{Y_n=i, n<T_{i_0}^{(1)}(\mathbf{Y})\}}] &= \sum_{n\geq 0} \mathbb{E}_{i_0}[\mathbb{E}_{i_0}[D_{n+1}\mathbf{1}_{\{Y_n=i, n<T_{i_0}^{(1)}(\mathbf{Y})\}} \mid \mathbf{Y}]] \\
&= \sum_{n\geq 0} \mathbb{E}_{i_0}[\mathbf{1}_{\{Y_n=i, n<T_{i_0}^{(1)}(\mathbf{Y})\}} \mathbb{E}_{i_0}[D_{n+1} \mid \mathbf{Y}]] \\
&= \sum_{n\geq 0} \mathbb{E}_{i_0}[\mathbf{1}_{\{Y_n=i, n<T_{i_0}^{(1)}(\mathbf{Y})\}} \tfrac{1}{q(Y_n)}] \\
&= \tfrac{1}{q(i)} \sum_{n\geq 0} \mathbb{E}_{i_0}[\mathbf{1}_{\{Y_n=i, n<T_{i_0}^{(1)}(\mathbf{Y})\}}].
\end{aligned}$$

Putting the last two displays together and using Tonelli's theorem once again, we deduce for all $i \in E$

$$\mathbb{E}_{i_0}\Big[\int_0^{T_{i_0}^+} \mathrm{d}t\, \mathbf{1}_{\{X_t=i\}}\Big] = \tfrac{1}{q(i)} \sum_{n\geq 0} \mathbb{E}_{i_0}[\mathbf{1}_{\{Y_n=i, n<T_{i_0}^{(1)}(\mathbf{Y})\}}] = \tfrac{1}{q(i)} \mathbb{E}_{i_0}\Big[\sum_{n=0}^{T_{i_0}^{(1)}(\mathbf{Y})-1} \mathbf{1}_{\{Y_n=i\}}\Big].$$

Notice that the expectation $\mathbb{E}_{i_0}\Big[\sum_{n=0}^{T_{i_0}^{(1)}(\mathbf{Y})-1} \mathbf{1}_{\{Y_n=i\}}\Big]$ is exactly $\nu_{i_0}(i)$, where ν_{i_0} denotes the Q-invariant measure introduced in Theorem I.4.2 in Section I.4.a of [Duq23]. The condition $\nu_{i_0}(i_0) = 1$ uniquely determines this measure among all Q-invariant measures. Indeed, if ν'_{i_0} is another Q-invariant measure such that $\nu'_{i_0}(i_0) = 1$, then by point (ii) of Theorem I.4.3 in Section I.4.a of [Duq23] there exists $c > 0$ such that $\nu'_{i_0}(i) = c\nu_{i_0}(i)$ for all $i \in E$. Evaluating this identity at $i = i_0$ yields $c = 1$ and hence $\nu'_{i_0} = \nu_{i_0}$.

Define $\tilde{m}(i) := \mathbb{E}_{i_0}\Big[\int_0^{T_{i_0}^+} \mathrm{d}t\, \mathbf{1}_{\{X_t=i\}}\Big]$ for all $i \in E$. Then $\nu_{i_0}(i) = q(i)\tilde{m}(i)$ for all $i \in E$, and by Lemma 5.10 the measure $\tilde{m}$ is G-invariant. Since G-invariant measures are unique up to multiplication by a positive constant, there exists $c > 0$ such that $m(i) = c\tilde{m}(i)$ for all $i \in E$. In particular,

$$m(i_0) = c\tilde{m}(i_0) = c\tfrac{\nu_{i_0}(i_0)}{q(i_0)} = \tfrac{c}{q(i_0)}$$

where we used that $q(i_0) > 0$ since G is irreducible. It follows that $c = m(i_0)q(i_0)$ and therefore

$$m(i) = m(i_0)q(i_0)\tilde{m}(i) = m(i_0)q(i_0)\mathbb{E}_{i_0}\Big[\int_0^{T_{i_0}^+} \mathrm{d}t\, \mathbf{1}_{\{X_t=i\}}\Big] \quad \text{for all } i \in E.$$

It remains to show that $0 < m(i) < \infty$ for all $i \in E$. For this, we first prove that $0 < \tilde{m}(i) < \infty$ for all $i \in E$. Recall that $\nu_{i_0}(i) = q(i)\tilde{m}(i)$ for all $i \in E$. By Theorem I.4.2 of [Duq23], we have $0 < \nu_{i_0}(i) < \infty$ for all $i \in E$. Since G is irreducible, $q(i) > 0$ for all $i \in E$. In addition, $q(i) < \infty$ for all $i \in E$. Indeed, although holding times may formally be allowed to have extended exponential distributions with parameter $q(i) = \infty$, which would correspond to zero holding times and hence instantaneous jumps, the right-continuity of the minimal process excludes such behavior before explosion. Since no explosion occurs here, it follows that $q(i) < \infty$ for all $i \in E$. Consequently, $0 < \tilde{m}(i) < \infty$ for all $i \in E$. Finally, since $m(i) = c\tilde{m}(i)$ with $0 < c < \infty$, we conclude that $0 < m(i) < \infty$ for all $i \in E$. □

The previous proposition gives a concrete interpretation of invariant measures in terms of occupation times. The formula (5.14) shows that the mass $m(i)$ assigned to a state $i \in E$ is proportional

to the expected amount of time the process spends in i during one excursion away from a reference state $i_0 \in E$. In this sense, an invariant measure describes how the process distributes its time among the states in the long run.

A particularly important class of generators for which invariant measures can be constructed explicitly is given by *reversible* generators. Reversibility expresses a balance condition on the jump rates, meaning that transitions from a state i to a state j are compensated by transitions from j to i.

Definition 5.12 (Reversibility)**.** An infinitesimal generator $G = (q_{i,j})_{i,j\in E}$ on E is said to be *reversible* if there exists a non-identically zero measure m such that

$$m(i)q_{i,j} = m(j)q_{j,i} \quad \text{for all } i, j \in E.$$

We say that the measure m makes G reversible.

Reversibility is useful because any measure making a generator reversible is necessarily invariant. In other words, balancing the transitions between each pair of states ensures that the total incoming and outgoing mass rates at every state coincide. This is stated in the following lemma.

Lemma 5.13. *Let $G = (q_{i,j})_{i,j\in E}$ be an infinitesimal generator on E made reversible by a measure m. Then $mG = 0$, that is, m is G-invariant.*

Proof. For every $j \in E$, we have

$$(mG)(j) = \sum_{i\in E} m(i)q_{i,j} = \sum_{i\in E} m(j)q_{j,i} = m(j)\sum_{i\in E} q_{j,i} = 0$$

where we used that $-q_{j,j} = \sum_{i\in E\setminus\{j\}} q_{j,i}$. □

The following example shows how reversibility can be used to construct invariant measures for birth and death processes.

Example 5.14. Let $G = (q_{i,j})_{i,j\in\mathbb{N}}$ be the generator of a birth-death process (see Example 5.8). The birth rates are denoted $(\lambda_i)_{i\in\mathbb{N}}$ and the death rates $(\mu_i)_{i\geq 1}$. We assume that G is irreducible, that is, $i \leftrightarrow j$ for all $i, j \in \mathbb{N}$. By point (iii) of Proposition 5.2, this is equivalent to requiring that for any $i, j \in \mathbb{N}$ there exists a finite path from i to j along which all corresponding generator coefficients are strictly positive. In the case of a birth-death process, where only transitions between nearest neighbors are allowed, this condition reduces to $q_{i,i+1}q_{i+1,i} > 0$ for all $i \in \mathbb{N}$ or equivalently $\lambda_i\mu_{i+1} > 0$ for all $i \in \mathbb{N}$.

Suppose that G is made reversible by a measure m, which is equivalent to

$$m(i)q_{i,i+1} = m(i+1)q_{i+1,i} \quad \text{for all } i \in \mathbb{N}$$

since $q_{i,j} = 0$ for all $i, j \in \mathbb{N}$ with $|i-j| > 1$. This yields the recurrence relation $m(i+1) = m(i)\frac{q_{i,i+1}}{q_{i+1,i}}$ for all $i \in \mathbb{N}$. Thus the system in the previous display is equivalent to

$$m(i) = m(0)\prod_{\ell=1}^{i} \frac{q_{\ell-1,\ell}}{q_{\ell,\ell-1}} = m(0)\prod_{\ell=1}^{i} \frac{\lambda_{\ell-1}}{\mu_\ell} \quad \text{for all } i \geq 1.$$

The following theorem shows that a measure which is invariant for the generator G is also invariant for the associated minimal semigroup $(P_t)_{t\in\mathbb{R}_+}$. Since the P_t are the transition matrices of the process $\mathbf{X}$, the notion of invariant measure in Definition 5.9 coincides with the usual discrete-time definition of invariance expressed in terms of the transition matrix Q.

Theorem 5.15. *Let G be an irreducible and recurrent infinitesimal generator on E. Denote by $(P_t)_{t\in\mathbb{R}_+}$ the associated minimal semigroup (see Corollary 4.6). Let m be a G-invariant measure. Then, for all $t \in \mathbb{R}_+$, m is P_t-invariant, that is, $mP_t = m$ or equivalently*

$$m(j) = \sum_{i\in E} m(i)p_t(i,j) \quad \textit{for all } j \in E.$$

Proof. The case $t = 0$ is immediate since $P_0 = \mathrm{Id}_E$ and hence $mP_0 = m$. Now fix $t > 0$ and $i \in E$. Recall that $T_i^+ = \inf\{t \geq J_1 : X_t = i\}$ and that $\mathbb{P}_i(T_i^+ < \infty) = 1$, since G is irreducible and recurrent. By Proposition 5.11, for all $j \in E$,

$$m(j) = m(i)q(i)\mathbb{E}_i\Big[\int_0^{T_i^+} \mathrm{d}t\, \mathbf{1}_{\{X_t=j\}}\Big]. \tag{5.15}$$

Fix $j \in E$. Using the change of variables $u := s - T_i^+$ and the weak tower property, we have

$$\mathbb{E}_i\Big[\int_{T_i^+}^{t+T_i^+} \mathrm{d}s\, \mathbf{1}_{\{X_s=j\}}\Big] = \mathbb{E}_i\Big[\int_0^t \mathrm{d}u\, \mathbf{1}_{\{X_{T_i^+ + u}=j\}}\Big] = \mathbb{E}_i\Big[\mathbb{E}_i\Big[\int_0^t \mathrm{d}u\, \mathbf{1}_{\{X_{T_i^+ + u}=j\}} \mid \mathcal{F}_{T_i^+}\Big]\Big].$$

The fact that $T_i^+ < \infty$ $\mathbb{P}_i$-a.s. together with the strong Markov property at T_i^+ yield

$$\begin{aligned}
\mathbb{E}_i\Big[\int_0^t \mathrm{d}u\, \mathbf{1}_{\{X_{T_i^+ + u}=j\}} \mid \mathcal{F}_{T_i^+}\Big] &= \mathbb{E}_i[\mathbf{1}_{\{T_i^+<\infty\}} \int_0^t \mathrm{d}u\, \mathbf{1}_{\{(\theta_{T_i^+}\mathbf{X})_u=j\}} \mid \mathcal{F}_{T_i^+}] \\
&= \mathbf{1}_{\{T_i^+<\infty\}}\mathbb{E}_{X_{T_i^+}}\Big[\int_0^t \mathrm{d}u\, \mathbf{1}_{\{X_u=j\}}\Big] \\
&= \mathbb{E}_i\Big[\int_0^t \mathrm{d}u\, \mathbf{1}_{\{X_u=j\}}\Big].
\end{aligned}$$

Putting the last two displays together, we obtain

$$\mathbb{E}_i\Big[\int_{T_i^+}^{t+T_i^+} \mathrm{d}s\, \mathbf{1}_{\{X_s=j\}}\Big] = \mathbb{E}_i\Big[\int_0^t \mathrm{d}s\, \mathbf{1}_{\{X_s=j\}}\Big]$$

and consequently

$$\begin{aligned}
&m(i)q(i)\mathbb{E}_i\Big[\int_t^{t+T_i^+} \mathrm{d}s\, \mathbf{1}_{\{X_s=j\}}\Big] \\
&= m(i)q(i)\mathbb{E}_i\Big[\int_0^{t+T_i^+} \mathrm{d}s\, \mathbf{1}_{\{X_s=j\}}\Big] - m(i)q(i)\mathbb{E}_i\Big[\int_0^t \mathrm{d}s\, \mathbf{1}_{\{X_s=j\}}\Big] \\
&= m(i)q(i)\mathbb{E}_i\Big[\int_0^{t+T_i^+} \mathrm{d}s\, \mathbf{1}_{\{X_s=j\}}\Big] - m(i)q(i)\mathbb{E}_i\Big[\int_{T_i^+}^{t+T_i^+} \mathrm{d}s\, \mathbf{1}_{\{X_s=j\}}\Big] \\
&= m(i)q(i)\mathbb{E}_i\Big[\int_0^{T_i^+} \mathrm{d}s\, \mathbf{1}_{\{X_s=j\}}\Big] \\
&= m(i)
\end{aligned}$$

where we used (5.15) for the last equation. Using the change of variables $u := s - t$ and Tonelli's theorem, we get

$$\begin{aligned}
m(j) &= m(i)q(i)\mathbb{E}_i\Big[\int_t^{t+T_i^+} \mathrm{d}s\, \mathbf{1}_{\{X_s=j\}}\Big] \\
&= m(i)q(i)\mathbb{E}_i\Big[\int_0^{T_i^+} \mathrm{d}u\, \mathbf{1}_{\{X_{t+u}=j\}}\Big] \\
&= m(i)q(i)\mathbb{E}_i\Big[\int_0^{\infty} \mathrm{d}u\, \mathbf{1}_{\{X_{t+u}=j\}}\mathbf{1}_{\{u<T_i^+\}}\Big] \\
&= m(i)q(i)\int_0^{\infty} \mathrm{d}u\, \mathbb{E}_i[\mathbf{1}_{\{X_{t+u}=j\}}\mathbf{1}_{\{u<T_i^+\}}].
\end{aligned}$$

Since $\{u < T_i^+\} = \{T_i^+ \leq u\}^c \in \mathcal{F}_u$ (as T_i^+ is a stopping time), the weak tower property and the

simple Markov property at time u yield for all $u > 0$,

$$\begin{aligned}
\mathbb{E}_i[\mathbf{1}_{\{X_{t+u}=j\}}\mathbf{1}_{\{u<T_i^+\}}] &= \sum_{k\in E} \mathbb{E}_i[\mathbf{1}_{\{X_{t+u}=j\}}\mathbf{1}_{\{X_u=k,u<T_i^+\}}] \\
&= \sum_{k\in E} \mathbb{E}_i[\mathbf{1}_{\{X_u=k,u<T_i^+\}}\mathbb{E}_i[\mathbf{1}_{\{X_{t+u}=j\}} \mid \mathcal{F}_u]] \\
&= \sum_{k\in E} \mathbb{E}_i[\mathbf{1}_{\{X_u=k,u<T_i^+\}}\mathbb{E}_i[\mathbf{1}_{\{T_i^+<\infty\}}\mathbf{1}_{\{(\theta_u\mathbf{X})_t=j\}} \mid \mathcal{F}_u]] \\
&= \sum_{k\in E} \mathbb{E}_i[\mathbf{1}_{\{X_u=k,u<T_i^+\}}\mathbf{1}_{\{T_i^+<\infty\}}\mathbb{E}_{X_u}[\mathbf{1}_{\{X_t=j\}}]] \\
&= \sum_{k\in E} \mathbb{E}_i[\mathbf{1}_{\{X_u=k,u<T_i^+\}}]\mathbb{P}_k(X_t = j) \\
&= \sum_{k\in E} \mathbb{E}_i[\mathbf{1}_{\{X_u=k,u<T_i^+\}}]p_t(k,j).
\end{aligned}$$

Combining the last two displays and applying Tonelli's theorem twice, together with (5.15), we obtain

$$\begin{aligned}
m(j) &= m(i)q(i)\int_0^\infty \mathrm{d}u \sum_{k\in E} \mathbb{E}_i[\mathbf{1}_{\{X_u=k,u<T_i^+\}}]p_t(k,j) \\
&= m(i)q(i)\sum_{k\in E} p_t(k,j)\int_0^\infty \mathrm{d}u\, \mathbb{E}_i[\mathbf{1}_{\{X_u=k,u<T_i^+\}}] \\
&= m(i)q(i)\sum_{k\in E} p_t(k,j)\mathbb{E}_i\Big[\int_0^\infty \mathrm{d}u\, \mathbf{1}_{\{X_u=k,u<T_i^+\}}\Big] \\
&= m(i)q(i)\sum_{k\in E} p_t(k,j)\mathbb{E}_i\Big[\int_0^{T_i^+} \mathrm{d}u\, \mathbf{1}_{\{X_u=k\}}\Big] \\
&= \sum_{k\in E} m(k)p_t(k,j)
\end{aligned}$$

which concludes the proof. □

We now turn to the existence of an invariant probability measure (stationary distribution). As in the case of discrete-time Markov chains (see Definition I.4.6 of [Duq23]), this notion is closely related to the classification of states according to their recurrence properties. We therefore begin by introducing the notions of positive recurrence and null recurrence for continuous-time Markov processes.

Definition 5.16. We introduce the following notions.

(a) (Positive recurrent state) A state $i \in E$ is said to be *positive recurrent* if and only if

$$\mathbb{E}_i[T_i^+] < \infty.$$

(Note that this implies $\mathbb{P}_i(T_i^+ < \infty) = 1$ and thus, by point (I) of Proposition 5.5, the state i is recurrent.)

(b) (Null recurrent state) A state $i \in E$ is said to be *null recurrent* if and only if

$$\mathbb{P}_i(T_i^+ < \infty) = 1 \quad \text{but} \quad \mathbb{E}_i[T_i^+] = \infty.$$

(Note that by point (I) of Proposition 5.5, this implies that the state i is recurrent.)

(c) If

$$\mathbb{P}_i(T_i^+ < \infty) < 1$$

then the state $i \in E$ is transient. (Indeed, since $T_i^{(1)}(\mathbf{Y}) < \infty$ implies $T_i^+ < \infty$, we have $\mathbb{P}_i(T_i^{(1)}(\mathbf{Y}) < \infty) \le \mathbb{P}_i(T_i^+ < \infty) < 1$. By Definition I.3.1 of [Duq23], this implies that i is transient.)

We quickly verify by contraposition the claim made above that $\mathbb{E}_i[T_i^+] < \infty$ implies $\mathbb{P}_i(T_i^+ < \infty) = 1$. Indeed, suppose that $\mathbb{P}_i(T_i^+ = \infty) > 0$. Then, for every $n \in \mathbb{N}$, we have $T_i^+ \geq n\mathbf{1}_{\{T_i^+ \geq n\}}$, and taking expectations gives $\mathbb{E}_i[T_i^+] \geq n\mathbb{P}_i(T_i^+ \geq n) \geq n\mathbb{P}_i(T_i^+ = \infty)$. Letting $n \to \infty$, since $\mathbb{P}_i(T_i^+ = \infty) > 0$, we obtain $\mathbb{E}_i[T_i^+] = \infty$.

As in the case of discrete-time Markov chains (see Theorem I.4.7 of [Duq23]), the existence of an invariant probability measure is closely linked to positive recurrence. The following theorem makes this connection precise in the continuous-time setting.

Theorem 5.17. *Let G be an irreducible (not necessarily recurrent) infinitesimal generator on E. The following assertions are equivalent.*

(i) There exists a positive recurrent state in E.

(ii) All states in E are positive recurrent.

(iii) G is non-explosive and admits a G-invariant probability measure.

If one of these three equivalent conditions is satisfied, then the unique G-invariant probability measure is given by

$$m(i) = \tfrac{1}{q(i)\mathbb{E}_i[T_i^+]} \quad \text{for all } i \in E.$$

Proof. Observe that since G is irreducible, there is no absorbing state and thus $q(i) > 0$ for all $i \in E$. We first show that (i) implies (ii). Let $i_0 \in E$ be positive recurrent, which implies that it is recurrent. Since G is irreducible, by point (v) of Theorem 5.4, all states are recurrent. Now fix $i \in E$ and define the measure $m_i = (m_i(j), j \in E)$ by

$$m_i(j) := q(i)\mathbb{E}_i\Big[\int_0^{T_i^+} \mathrm{d}t\, \mathbf{1}_{\{X_t = j\}}\Big] \quad \text{for all } j \in E. \tag{5.16}$$

Recall $\mathbb{P}_i(T_i^+ < \infty) = \mathbb{P}_i(T_i^{(1)}(\mathbf{Y}) < \infty) = 1$, since G is irreducible and i is recurrent. For $t \in \mathbb{R}_+$ and $j \in E$, observe that $\{X_t = j\} = \{X_t = j\} \cap \{t < \zeta\}$, since from the explosion time onward the minimal process $\mathbf{X}$ remains in the cemetery state $\partial \notin E$. Hence, for all $t \in \mathbb{R}_+$, we have

$$\sum_{j \in E} \mathbf{1}_{\{X_t = j, t < T_i^+\}} = \sum_{j \in E} \mathbf{1}_{\{X_t = j, t < T_i^+, t < \zeta\}} = \mathbf{1}_{\{t < \zeta \wedge T_i^+\}}.$$

Summing over $j \in E$ in (5.16) and applying Tonelli's theorem then yields

$$\begin{aligned}
\sum_{j \in E} m_i(j) &= \sum_{j \in E} q(i)\mathbb{E}_i\Big[\int_0^{T_i^+} \mathrm{d}t\, \mathbf{1}_{\{X_t = j\}}\Big] \\
&= q(i)\mathbb{E}_i\Big[\int_0^{\infty} \mathrm{d}t \sum_{j \in E} \mathbf{1}_{\{X_t = j, t < T_i^+\}}\Big] \\
&= q(i)\mathbb{E}_i\Big[\int_0^{\infty} \mathrm{d}t\, \mathbf{1}_{\{t < \zeta \wedge T_i^+\}}\Big] \\
&= q(i)\mathbb{E}_i\Big[\int_0^{\zeta \wedge T_i^+} \mathrm{d}t\Big] \\
&= q(i)\mathbb{E}_i[\zeta \wedge T_i^+].
\end{aligned}$$

Since G is recurrent, point (c) of Proposition 5.6 implies that $\mathbb{P}_i(\zeta = \infty) = 1$. Consequently, $\mathbb{E}_i[\zeta \wedge T_i^+] = \mathbb{E}_i[T_i^+]$ and thus

$$\sum_{j \in E} m_i(j) = q(i)\mathbb{E}_i[T_i^+]. \tag{5.17}$$

By Proposition 5.11, m_i is the unique G-invariant measure such that $m_i(i) = 1$ and

$$m_i(j) = m_i(i_0)q(i_0)\mathbb{E}_{i_0}\Big[\int_0^{T_{i_0}^+} \mathrm{d}t\, \mathbf{1}_{\{X_t = j\}}\Big] = m_i(i_0)m_{i_0}(j) \quad \text{for all } i, j \in E. \tag{5.18}$$

Applying (5.17) with $i = i_0$ gives

$$\sum_{j\in E} m_{i_0}(j) = q(i_0)\mathbb{E}_{i_0}[T_{i_0}^+] < \infty \tag{5.19}$$

since i_0 is positive recurrent and $q(i_0) < \infty$ (by right-continuity of the minimal process, which excludes instantaneous jumps before explosion). By (5.18), we have

$$\sum_{j\in E} m_i(j) = m_i(i_0) \sum_{j\in E} m_{i_0}(j) \quad \text{for all } i \in E$$

which is finite by (5.19) and because $m_i(i_0) < \infty$ (see Proposition 5.11). Then since $q(i) < \infty$ for all $i \in E$, by (5.17) we obtain that $\mathbb{E}_i[T_i^+] < \infty$ for all $i \in E$, which proves (ii).

Assume (i). We have shown that G is recurrent and that the measures m_i are G-invariant and have finite total mass for all $i \in E$. By Proposition 5.11, all G-invariant measures are proportional to each other, hence there exists a unique G-invariant probability measure m. Indeed, fix $i \in E$ and define $m := \frac{m_i}{m_i(E)}$, which is well defined since $m_i(E) \in (0, \infty)$. Then m is a G-invariant probability measure since m_i is G-invariant and

$$m(E) = \sum_{j\in E} \tfrac{m_i(j)}{m_i(E)} = \tfrac{m_i(E)}{m_i(E)} = 1.$$

For uniqueness, let m' be another G-invariant probability measure. By Proposition 5.11, there exists $c > 0$ such that $m' = cm$. Since $1 = m'(E) = cm(E) = c$, we conclude that $m' = m$.

Now fix $i \in E$. By Proposition 5.11, there exists a constant $c(i) \in (0, \infty)$ such that $m(j) = c(i)m_i(j)$ for all $j \in E$. Since $m_i(i) = 1$, we have $m(i) = c(i)$. By (5.17), we have

$$1 = \sum_{j\in E} m(j) = c(i) \sum_{j\in E} m_i(j) = c(i)q(i)\mathbb{E}_i[T_i^+]$$

and since $q(i) > 0$ and $\mathbb{E}_i[T_i^+] > 0$ (as $T_i^+ \geq J_1 > 0$), we obtain

$$m(i) = c(i) = \tfrac{1}{q(i)\mathbb{E}_i[T_i^+]} \quad \text{for all } i \in E.$$

We now show that (ii) implies (iii). Since all states in E are positive recurrent, this implies that all states in E are recurrent. Thus G is irreducible and recurrent, and point (c) of Proposition 5.6 implies that $\mathbb{P}_i(\zeta < \infty) = 0$ for all $i \in E$. Using the weak tower property and the simple Markov property at time 0, for every $\mu \in \mathcal{M}_1(E)$ we obtain

$$\begin{aligned}
\mathbb{P}_\mu(\zeta < \infty) &= \mathbb{E}_\mu[\mathbf{1}_{\{\zeta(\mathbf{X})<\infty\}}] \\
&= \mathbb{E}_\mu[\mathbb{E}_\mu[\mathbf{1}_{\{\zeta(\theta_0\mathbf{X})<\infty\}} \mid \mathcal{F}_0]] \\
&= \mathbb{E}_\mu[\mathbb{E}_{X_0}[\mathbf{1}_{\{\zeta(\mathbf{X})<\infty\}}]] \\
&= \mathbb{E}_\mu[\mathbb{P}_{X_0}(\zeta < \infty)] \\
&= \sum_{i\in E} \mathbb{P}_i(\zeta < \infty)\mathbb{P}_\mu(X_0 = i) \\
&= \sum_{i\in E} \mathbb{P}_i(\zeta < \infty)\mu(i) \\
&= 0
\end{aligned}$$

and thus G is non-explosive. Next, fix $i_0 \in E$ and define

$$m(j) := \tfrac{1}{q(i_0)\mathbb{E}_{i_0}[T_{i_0}^+]} m_{i_0}(j) \quad \text{for all } j \in E.$$

Observe that m is G-invariant, since m_{i_0} is. Moreover, by (5.19)

$$\sum_{j\in E} m(j) = \tfrac{1}{q(i_0)\mathbb{E}_{i_0}[T_{i_0}^+]} \sum_{j\in E} m_{i_0}(j) = \tfrac{1}{q(i_0)\mathbb{E}_{i_0}[T_{i_0}^+]} q(i_0)\mathbb{E}_{i_0}[T_{i_0}^+] = 1.$$

Thus m is a probability measure, which proves (iii).

Finally, we show that (iii) implies (i). Assume that G is non-explosive and that there exists a G-invariant probability measure m. Fix $i_0 \in E$ such that $m(i_0) > 0$, which exists since m is not identically zero. Define the measure ν by

$$\nu(i) := \tfrac{q(i)m(i)}{q(i_0)m(i_0)} \quad \text{for all } i \in E.$$

Then since m is G-invariant, Lemma 5.10 implies that ν is Q-invariant. Fix $j \in E$. Since ν is Q-invariant and $\nu(i_0) = 1$, we have

$$\nu(j) = \sum_{i \in E} \nu(i)\pi_{i,j} = \pi_{i_0,j} + \sum_{i \in E \setminus \{i_0\}} \nu(i)\pi_{i,j}. \tag{5.20}$$

Applying (5.20) with j replaced by $i_1 \in E \setminus \{i_0\}$ yields

$$\nu(i_1) = \pi_{i_0,i_1} + \sum_{i_2 \in E \setminus \{i_0\}} \nu(i_2)\pi_{i_2,i_1}.$$

Substituting this into (5.20), we obtain

$$\nu(j) = \pi_{i_0,j} + \sum_{i_1 \in E \setminus \{i_0\}} \pi_{i_0,i_1}\pi_{i_1,j} + \sum_{i_1,i_2 \in E \setminus \{i_0\}} \nu(i_2)\pi_{i_2,i_1}\pi_{i_1,j}.$$

Iterating this argument, we obtain that for all $n \in \mathbb{N}$,

$$\begin{aligned}\nu(j) = \pi_{i_0,j} + \sum_{i_1 \in E \setminus \{i_0\}} \pi_{i_0,i_1}\pi_{i_1,j} + \cdots + \sum_{i_1,\dots,i_n \in E \setminus \{i_0\}} \pi_{i_0,i_1}\cdots\pi_{i_n,j} \\ + \sum_{i_1,\dots,i_{n+1} \in E \setminus \{i_0\}} \nu(i_{n+1})\pi_{i_{n+1},i_n}\cdots\pi_{i_2,i_1}\pi_{i_1,j}.\end{aligned}$$

Since the last sum is nonnegative, this implies for all $n \in \mathbb{N}$ that

$$\nu(j) \geq \pi_{i_0,j} + \sum_{i_1 \in E \setminus \{i_0\}} \pi_{i_0,i_1}\pi_{i_1,j} + \cdots + \sum_{i_1,\dots,i_n \in E \setminus \{i_0\}} \pi_{i_0,i_1}\cdots\pi_{i_n,j}.$$

We now interpret each summand on the right-hand side in terms of the skeleton chain $\mathbf{Y}$. In the case $j \neq i_0$, we have

$$\pi_{i_0,j} = \mathbb{P}_{i_0}(Y_1 = j) = \mathbb{P}_{i_0}(Y_1 = j, T^{(1)}_{i_0}(\mathbf{Y}) > 1),$$

where $T^{(1)}_{i_0}(\mathbf{Y}) := \inf\{n > 0 : Y_n = i_0\}$ with the convention $\inf \emptyset := \infty$. By Lemma I.1.4 in Section I.1.a of [Duq23], we obtain

$$\begin{aligned}\sum_{i_1 \in E \setminus \{i_0\}} \pi_{i_0,i_1}\pi_{i_1,j} &= \sum_{i_1 \in E \setminus \{i_0\}} \mathbb{P}_{i_0}(Y_1 = i_1, Y_2 = j) \\ &= \mathbb{P}_{i_0}(Y_2 = j, Y_1 \neq i_0) \\ &= \mathbb{P}_{i_0}(Y_2 = j, T^{(1)}_{i_0}(\mathbf{Y}) > 2)\end{aligned}$$

where in the last equality we used that on $\{Y_2 = j\}$ with $j \neq i_0$, the condition $T^{(1)}_{i_0}(\mathbf{Y}) > 2$ is equivalent to $Y_1 \neq i_0$. Iterating this argument yields that for all $1 \leq k \leq n$,

$$\sum_{i_1,\dots,i_k \in E \setminus \{i_0\}} \pi_{i_0,i_1}\cdots\pi_{i_k,j} = \mathbb{P}_{i_0}(Y_{k+1} = j, T^{(1)}_{i_0}(\mathbf{Y}) > k+1).$$

Hence, if $j \neq i_0$, we have

$$\begin{aligned}\nu(j) &\geq \sum_{k=0}^{n} \mathbb{P}_{i_0}(Y_{k+1} = j, T^{(1)}_{i_0}(\mathbf{Y}) > k+1) \\ &= \sum_{k=1}^{n+1} \mathbb{P}_{i_0}(Y_k = j, T^{(1)}_{i_0}(\mathbf{Y}) > k) \\ &= \sum_{k=0}^{n+1} \mathbb{P}_{i_0}(Y_k = j, T^{(1)}_{i_0}(\mathbf{Y}) > k)\end{aligned}$$

where the last equality follows from the fact that $\mathbb{P}_{i_0}(Y_0 = j, T_{i_0}^{(1)}(\mathbf{Y}) > 0) = 0$ for $j \neq i_0$. In the case $j = i_0$, we have $\mathbb{P}_{i_0}(Y_k = j, T_{i_0}^{(1)}(\mathbf{Y}) > k) = 0$ for all $1 \leq k \leq n+1$ and $\mathbb{P}_{i_0}(Y_0 = j, T_{i_0}^{(1)}(\mathbf{Y}) > 0) = 1$. Since $\nu(j) = \nu(i_0) = 1$, the inequality also holds in this case, namely

$$\nu(j) \geq \sum_{k=0}^{n+1} \mathbb{P}_{i_0}(Y_k = j, T_{i_0}^{(1)}(\mathbf{Y}) > k).$$

Letting $n \to \infty$ in the previous display, we obtain

$$\nu(j) \geq \sum_{k \geq 0} \mathbb{P}_{i_0}(Y_k = j, T_{i_0}^{(1)}(\mathbf{Y}) > k).$$

By Theorem 3.38, conditionally on Y_k, the holding time D_{k+1} is exponentially distributed with parameter $q(Y_k) > 0$ and is independent of the past $(Y_0, \dots, Y_{k-1})$. Hence

$$\begin{aligned}
\tfrac{m(j)}{q(i_0)m(i_0)} &= \tfrac{\nu(j)}{q(j)} \\
&\geq \sum_{k \geq 0} \tfrac{1}{q(j)} \mathbb{P}_{i_0}(Y_k = j, T_{i_0}^{(1)}(\mathbf{Y}) > k) \\
&= \sum_{k \geq 0} \mathbb{E}_{i_0}\big[\tfrac{1}{q(Y_k)} \mathbf{1}_{\{Y_k = j, T_{i_0}^{(1)}(\mathbf{Y}) > k\}}\big] \\
&= \sum_{k \geq 0} \mathbb{E}_{i_0}\big[\mathbb{E}_{i_0}[D_{k+1} \mid Y_k] \mathbf{1}_{\{Y_k = j, T_{i_0}^{(1)}(\mathbf{Y}) > k\}}\big] \\
&= \sum_{k \geq 0} \mathbb{E}_{i_0}\big[\mathbb{E}_{i_0}[D_{k+1} \mid \mathcal{F}_k^{\mathbf{Y}}] \mathbf{1}_{\{Y_k = j, T_{i_0}^{(1)}(\mathbf{Y}) > k\}}\big] \\
&= \sum_{k \geq 0} \mathbb{E}_{i_0}\big[\mathbb{E}_{i_0}[D_{k+1} \mathbf{1}_{\{Y_k = j, T_{i_0}^{(1)}(\mathbf{Y}) > k\}} \mid \mathcal{F}_k^{\mathbf{Y}}]\big]
\end{aligned}$$

where the last equality follows since $\{Y_k = j\} \in \mathcal{F}_k^{\mathbf{Y}}$ and $\{T_{i_0}^{(1)}(\mathbf{Y}) > k\} = \{Y_1 \neq i_0, \dots, Y_k \neq i_0\} \in \mathcal{F}_k^{\mathbf{Y}}$. Applying the weak tower property and Tonelli's theorem yields

$$\begin{aligned}
\tfrac{m(j)}{q(i_0)m(i_0)} &\geq \sum_{k \geq 0} \mathbb{E}_{i_0}[D_{k+1} \mathbf{1}_{\{Y_k = j, T_{i_0}^{(1)}(\mathbf{Y}) > k\}}] \\
&= \mathbb{E}_{i_0}\Big[\sum_{k \geq 0} D_{k+1} \mathbf{1}_{\{Y_k = j, T_{i_0}^{(1)}(\mathbf{Y}) > k\}}\Big].
\end{aligned}$$

Since D_{k+1} is the holding time at Y_k and the indicator $\mathbf{1}_{\{Y_k = j, T_{i_0}^{(1)}(\mathbf{Y}) > k\}}$ selects visits to j before the first return of the skeleton chain $\mathbf{Y}$ to i_0, the sum inside the expectation equals the occupation time of j by the continuous-time process $\mathbf{X}$ before $T_{i_0}^+ = J_{T_{i_0}^{(1)}(\mathbf{Y})}$. Therefore

$$\tfrac{m(j)}{q(i_0)m(i_0)} \geq \mathbb{E}_{i_0}\Big[\int_0^{T_{i_0}^+} \mathrm{d}t \, \mathbf{1}_{\{X_t = j\}}\Big]. \tag{5.21}$$

Summing over j in (5.21) and using the same argument as in the proof of the implication from (i) to (ii), we obtain

$$\sum_{j \in E} \tfrac{m(j)}{q(i_0)m(i_0)} \geq \sum_{j \in E} \mathbb{E}_{i_0}\Big[\int_0^{T_{i_0}^+} \mathrm{d}t \, \mathbf{1}_{\{X_t = j\}}\Big] = \mathbb{E}_{i_0}[\zeta \wedge T_{i_0}^+].$$

Since G is non-explosive, we have $\mathbb{P}_{i_0}(\zeta = \infty) = 1$, and hence the right-hand side is equal to $\mathbb{E}_{i_0}[T_{i_0}^+]$. Moreover, since m is a probability measure and both $q(i_0)$ and $m(i_0)$ are strictly positive, the left-hand side is equal to $\frac{1}{q(i_0)m(i_0)} < \infty$. Thus

$$\mathbb{E}_{i_0}[T_{i_0}^+] \leq \tfrac{1}{q(i_0)m(i_0)} < \infty$$

and i_0 is positive recurrent, which proves (i). □

Contrary to the case of discrete-time Markov chains, there exist continuous-time Markov processes whose infinitesimal generator is irreducible and not positive recurrent, but nevertheless admit an invariant probability measure. By the previous theorem, such processes are necessarily explosive, as illustrated in the next example.

Example 5.18 ([Ami], Example 6.4.11). Let $E := \mathbb{N}$ and let $G = (q_{i,j})_{i,j\in\mathbb{N}}$ be the infinitesimal generator of a birth-death process $\mathbf{X}$ on $\mathbb{N}$ (see Example 5.8). Fix $a, b > 0$ such that $a + b = 1$ and $a > b$. Let $(q(i))_{i\in\mathbb{N}}$ be a sequence of strictly positive real numbers, which we interpret as the holding time parameters of the process $\mathbf{X}$. Define the birth and death rates by

$$\lambda_i := aq(i) \text{ for } i \geq 1, \quad \lambda_0 := q(0), \quad \mu_i := bq(i) \text{ for } i \geq 1.$$

Therefore, by Example 5.8 and using that $a + b = 1$, the generator matrix has the form

$$G = \begin{pmatrix} -q(0) & q(0) & 0 & 0 & \cdots \\ bq(1) & -q(1) & aq(1) & 0 & \cdots \\ 0 & bq(2) & -q(2) & aq(2) & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots \end{pmatrix}.$$

Note that G is irreducible since $q_{i,i+1}q_{i+1,i} > 0$ for all $i \in \mathbb{N}$.

Since $q_{i,i} = -q(i) < 0$ for all $i \in \mathbb{N}$, Definition 4.1 implies that the skeleton chain $\mathbf{Y}$ associated with $\mathbf{X}$ has transition matrix $Q = (\pi_{i,j})_{i,j\in\mathbb{N}}$ given by

$$\pi_{i,i+1} = -\tfrac{q_{i,i+1}}{q_{i,i}} = a \text{ for } i \geq 1, \quad \pi_{0,1} = -\tfrac{q_{0,1}}{q_{0,0}} = 1, \quad \pi_{i,i-1} = -\tfrac{q_{i,i-1}}{q_{i,i}} = b \text{ for } i \geq 1$$

and

$$\pi_{i,j} = -\tfrac{q_{i,j}}{q_{i,i}} = 0 \text{ for } i, j \in \mathbb{N} \text{ with } |j - i| > 1, \quad \pi_{i,i} = 0 \text{ for } i \in \mathbb{N}.$$

We now show that G is transient. Observe that

$$\sum_{n\geq 1} \prod_{i=1}^{n} \tfrac{\pi_{i,i-1}}{\pi_{i,i+1}} = \sum_{n\geq 1} (\tfrac{b}{a})^n < \infty$$

since $0 < \frac{b}{a} < 1$. By Example I.3.13 in Section I.3.b of [Duq23], this implies that Q is transient. By point (ii) of Theorem 5.4, it follows that G is transient and hence not positive recurrent.

We now construct a G-invariant measure, that is, a non-identically zero measure m on $\mathbb{N}$ satisfying $mG = 0$ (see Definition 5.9). In component form, the condition $mG = 0$ can be written as

$$\begin{cases} -q(0)m(0) + bq(1)m(1) = 0, \\ q(0)m(0) - q(1)m(1) + bq(2)m(2) = 0, \\ aq(i-1)m(i-1) - q(i)m(i) + bq(i+1)m(i+1) = 0, \quad i \geq 2. \end{cases}$$

We claim that the measure $m = (m(i))_{i\in\mathbb{N}}$ defined by

$$m(0) := \tfrac{a}{q(0)}, \quad m(i) := \tfrac{1}{q(i)}(\tfrac{a}{b})^i \text{ for } i \geq 1$$

satisfies the above system of equations. Indeed, for $i = 0$ we have

$$-q(0)m(0) + bq(1)m(1) = -q(0)\tfrac{a}{q(0)} + bq(1)\tfrac{1}{q(1)}\tfrac{a}{b} = -a + a = 0.$$

For $i = 1$,

$$\begin{aligned} q(0)m(0) - q(1)m(1) + bq(2)m(2) &= q(0)\tfrac{a}{q(0)} - q(1)\tfrac{1}{q(1)}\tfrac{a}{b} + bq(2)\tfrac{1}{q(2)}(\tfrac{a}{b})^2 \\ &= a - \tfrac{a}{b} + \tfrac{a^2}{b} \\ &= \tfrac{a}{b}(b - 1 + a) \\ &= 0 \end{aligned}$$

and for $i \geq 2$,

$$\begin{aligned}
&aq(i-1)m(i-1) - q(i)m(i) + bq(i+1)m(i+1) \\
&= aq(i-1)\tfrac{1}{q(i-1)}(\tfrac{a}{b})^{i-1} - q(i)\tfrac{1}{q(i)}(\tfrac{a}{b})^{i} + bq(i+1)\tfrac{1}{q(i+1)}(\tfrac{a}{b})^{i+1} \\
&= \tfrac{a^i}{b^{i-1}} - \tfrac{a^i}{b^i} + \tfrac{a^{i+1}}{b^i} \\
&= (\tfrac{a}{b})^i(b-1+a) \\
&= 0.
\end{aligned}$$

Assume now that $1 < \frac{a}{b} < 2$ and choose $q(i) := 2^i$. Then

$$m(E) = \sum_{i\in\mathbb{N}} m(i) = a + \sum_{i\geq 1}(\tfrac{a}{2b})^i = a + \tfrac{a}{2b-a} < \infty$$

since $\frac{1}{2} < \frac{a}{2b} < 1$. Hence m can be normalized, that is, divided by $m(E)$, to obtain a G-invariant probability measure.

By Theorem 5.17, since G is irreducible, not positive recurrent and admits a G-invariant probability measure, it follows that G must be explosive. This provides a rigorous justification of the explosive behavior observed in the simulation in Example 4.9.

5.4 Ergodic Theorems

We have seen in the previous section that invariant measures describe a balance in the dynamics of the minimal process, but they do not yet explain how the process behaves at large times. In this section, we study the long-term behavior of the process and establish convergence of its distribution toward an equilibrium, independently of the initial distribution. We also analyze the asymptotic behavior of time averages along a single trajectory of the process.

Throughout this section, we work in the same setting and with the same notation as in the previous three sections. The space E is a countable space equipped with the discrete topology, and hence endowed with the σ-algebra $\mathcal{P}(E)$. $\mathbf{X}$ denotes the minimal right-continuous Markov process on E^* associated with an infinitesimal generator G on E. Recall the following elementary lemma concerning convergence of probability measures on E.

Lemma 5.19. *Let $\mu, \mu_n \in \mathcal{M}_1(E)$, where $n \in \mathbb{N}$. The following assertions are equivalent:*

(a) $\mu_n \to \mu$ in distribution on E, that is, for all bounded and continuous functions $f : E \to \mathbb{R}$, $\int_E \mathrm{d}\mu_n\, f \to \int_E \mathrm{d}\mu\, f$.

(b) For every $i \in E$, $\mu_n(i) \to \mu(i)$.

(c) $\sum_{i\in E} |\mu_n(i) - \mu(i)| \to 0$.

Proof. We start by proving the implication from (a) to (b). Since we are working with the discrete topology, every function $f : E \to \mathbb{R}$ is continuous. Indeed, for every open set $U \subseteq \mathbb{R}$, the preimage $f^{-1}(U) \subseteq E$ is open because all subsets of E are open in the discrete topology. Now fix $i \in E$ and define $f_i := \mathbf{1}_{\{i\}}$. Then f_i is bounded and continuous. By assumption (a), it follows that

$$\int_E \mathrm{d}\mu_n\, f_i \to \int_E \mathrm{d}\mu\, f_i.$$

Since E is countable, this can be written as

$$\sum_{j\in E} f_i(j)\mu_n(j) \to \sum_{j\in E} f_i(j)\mu(j).$$

By the definition of f_i, this reduces to $\mu_n(i) \to \mu(i)$, which proves (b).

Next, we prove the implication from (b) to (a). Let $f : E \to \mathbb{R}$ be bounded (hence bounded and continuous) and let $M > 0$ be such that $\|f\|_\infty := \sup_{i\in E} |f(i)| \leq M$. Enumerate $E = \{e_1, e_2, \dots\}$. Then

$$\int_E \mathrm{d}\mu_n\, f = \sum_{k\geq 1} f(e_k)\mu_n(e_k) \quad \text{and} \quad \int_E \mathrm{d}\mu\, f = \sum_{k\geq 1} f(e_k)\mu(e_k).$$

Fix $\varepsilon > 0$. Since μ is a probability measure on E, we have $\mu(E) = \sum_{k\geq 1} \mu(e_k) = 1$. Hence there exists $K \geq 0$ such that $\mu(\{e_{K+1}, e_{K+2}, \ldots\}) = \sum_{k>K} \mu(e_k) < \varepsilon$. Then

$$\begin{aligned}\Big|\int_E \mathrm{d}\mu_n\, f - \int_E \mathrm{d}\mu\, f\Big| &= \Big|\sum_{k\geq 1} f(e_k)(\mu_n(e_k) - \mu(e_k))\Big| \\ &= \Big|\sum_{k=1}^{K} f(e_k)(\mu_n(e_k) - \mu(e_k)) + \sum_{k>K} f(e_k)(\mu_n(e_k) - \mu(e_k))\Big| \\ &\leq \Big|\sum_{k=1}^{K} f(e_k)(\mu_n(e_k) - \mu(e_k))\Big| + \Big|\sum_{k>K} f(e_k)(\mu_n(e_k) - \mu(e_k))\Big| \qquad (5.22)\end{aligned}$$

where we used the triangle inequality in the last step. By (b), we have $\mu_n(e_k) - \mu(e_k) \to 0$ for all $k \geq 1$. Since the first sum is finite, it follows that

$$\Big|\sum_{k=1}^{K} f(e_k)(\mu_n(e_k) - \mu(e_k))\Big| \to 0.$$

Hence, for n large enough,

$$\Big|\sum_{k=1}^{K} f(e_k)(\mu_n(e_k) - \mu(e_k))\Big| < \varepsilon.$$

For the second sum, using the triangle inequality and the boundedness of f, we obtain

$$\begin{aligned}\Big|\sum_{k>K} f(e_k)(\mu_n(e_k) - \mu(e_k))\Big| &\leq \sum_{k>K} |f(e_k)||\mu_n(e_k) - \mu(e_k)| \\ &\leq M \sum_{k>K} |\mu_n(e_k) - \mu(e_k)| \\ &\leq M \sum_{k>K} (\mu_n(e_k) + \mu(e_k)) \\ &= M\Big(\sum_{k>K} \mu_n(e_k) + \sum_{k>K} \mu(e_k)\Big). \qquad (5.23)\end{aligned}$$

Observe that by assumption (b),

$$\sum_{k>K} \mu_n(e_k) = 1 - \sum_{k=1}^{K} \mu_n(e_k) \to 1 - \sum_{k=1}^{K} \mu(e_k) = \sum_{k>K} \mu(e_k).$$

Hence, for n large enough,

$$\Big|\sum_{k>K} \mu_n(e_k) - \sum_{k>K} \mu(e_k)\Big| < \varepsilon$$

which implies that

$$\sum_{k>K} \mu_n(e_k) \leq \sum_{k>K} \mu(e_k) + \varepsilon.$$

Substituting this estimate into (5.23), we obtain for n large enough,

$$\Big|\sum_{k>K} f(e_k)(\mu_n(e_k) - \mu(e_k))\Big| \leq M(2\sum_{k>K} \mu(e_k) + \varepsilon) < 3M\varepsilon.$$

Combining everything with (5.22), we conclude that for n large enough,

$$\Big|\int_E \mathrm{d}\mu_n\, f - \int_E \mathrm{d}\mu\, f\Big| \leq \varepsilon + 3M\varepsilon = (3M+1)\varepsilon.$$

Since ε was arbitrary, it follows that

$$\int_E \mathrm{d}\mu_n\, f \to \int_E \mathrm{d}\mu\, f$$

which proves (a).

The implication from (c) to (b) is immediate, since for all $i \in E$,

$$|\mu_n(i) - \mu(i)| \leq \sum_{j \in E} |\mu_n(j) - \mu(j)| \to 0$$

and hence $\mu_n(i) \to \mu(i)$.

Finally, we show the implication from (b) to (c). Enumerate $E = \{e_1, e_2, \dots\}$ and define $S_n := \sum_{k \geq 1} |\mu_n(e_k) - \mu(e_k)|$. Fix $\varepsilon > 0$ and choose K such that $\sum_{k>K} \mu(e_k) < \varepsilon$. Then

$$S_n = \sum_{k=1}^{K} |\mu_n(e_k) - \mu(e_k)| + \sum_{k>K} |\mu_n(e_k) - \mu(e_k)|.$$

By (b), the first sum converges to 0, and hence is smaller than ε for n large enough. Using the triangle inequality and the same arguments as above, we obtain for n large enough

$$\sum_{k>K} |\mu_n(e_k) - \mu(e_k)| \leq \sum_{k>K} \mu_n(e_k) + \sum_{k>K} \mu(e_k) < 3\varepsilon.$$

Consequently, $S_n < 4\varepsilon$ for all sufficiently large n. Since ε was arbitrary, it follows that $S_n \to 0$, i.e. $\sum_{i \in E} |\mu_n(i) - \mu(i)| \to 0$, which proves (c). □

The previous lemma shows that, on the state space E, all standard notions of convergence of probability measures coincide. In particular, convergence in distribution (point (a)) is equivalent to convergence in total variation (point (c)). We now use this equivalence to establish convergence in distribution of the minimal Markov process $\mathbf{X}$ toward the stationary distribution (i.e. the unique invariant probability measure) in the positive recurrent case.

Theorem 5.20 (Convergence to the stationary distribution)**.** *Let G be an infinitesimal generator, assumed to be irreducible and positive recurrent (i.e. all states in E are positive recurrent). Denote by $\mathbf{X} = (X_t)_{t \in \mathbb{R}_+}$ the associated minimal process. By Theorem 5.17, it admits a unique G-invariant probability measure m, given by*

$$m(i) = \tfrac{1}{q(i)\mathbb{E}_i[T_i^+]} \quad \textit{for all } i \in E.$$

For any initial distribution $\mu \in \mathcal{M}_1(E)$, we have

$$\lim_{t \to \infty} \sum_{j \in E} |\mathbb{P}_\mu(X_t = j) - m(j)| = 0.$$

Proof. Since G is recurrent, point (c) of Proposition 5.6 implies that G does not explode, that is, $\mathbb{P}_i(\zeta = \infty) = 1$ for all $i \in E$. Recall from Corollary 4.6 that the minimal semigroup $(P_t)_{t \in \mathbb{R}_+}$ is sub-Markovian. Since G is non-explosive, we have $p_t(i,j) = \mathbb{P}_i(X_t = j, \zeta > t) = \mathbb{P}_i(X_t = j)$ for all $t \in \mathbb{R}_+$ and all $i, j \in E$. Consequently,

$$\sum_{j \in E} p_t(i,j) = \sum_{j \in E} \mathbb{P}_i(X_t = j) = 1 \quad \text{for all } t \in \mathbb{R}_+, i \in E$$

which shows that $(P_t)_{t \in \mathbb{R}_+}$ is a Markovian semigroup.

Fix $h > 0$ and define $Z_n^h := X_{nh}$ for $n \in \mathbb{N}$. Taking $t_k = kh$ for $0 \leq k \leq n$ in Definition 3.16, we obtain that for all $n \in \mathbb{N}$, $i_0, \dots, i_n \in E$ and $\mu \in \mathcal{M}_1(E)$,

$$\mathbb{P}_\mu(Z_0^h = i_0, \dots, Z_n^h = i_n) = \mu(i_0) p_h(i_0, i_1) \cdots p_h(i_{n-1}, i_n).$$

Hence, by Lemma I.1.4 in Section I.1.a of [Duq23], $(Z_n^h)_{n \in \mathbb{N}}$ is a Markov chain with transition matrix P_h. Since G is irreducible, point (iv) of Proposition 5.2 ensures that $p_h(i,j) > 0$ for all $i, j \in E$, which by Definition I.2.9 in Section I.2.b of [Duq23] implies that P_h is irreducible and aperiodic. Since G is irreducible and recurrent and m is a G-invariant probability measure, it is also P_h-invariant by Theorem 5.15. Therefore, by Theorem I.4.7 in Section I.4.a of [Duq23], the matrix

P_h is positive recurrent. Since P_h is irreducible, positive recurrent and aperiodic, Theorem I.4.17 in Section I.4.c of [Duq23] yields

$$\lim_{n\to\infty} \sum_{j\in E} |\mathbb{P}_\mu(Z_n^h = j) - m(j)| = 0 \quad \text{for all } \mu \in \mathcal{M}_1(E). \tag{5.24}$$

Fix $t \in \mathbb{R}_+$. Choose $n \in \mathbb{N}$ such that $nh \le t < (n+1)h$, that is, $n := \lfloor \frac{t}{h} \rfloor$. Using the semigroup property and the triangle inequality, we obtain for all $i, j \in E$,

$$\begin{aligned}
|p_t(i,j) - p_{nh}(i,j)| &= \Big| \sum_{k\in E} p_{t-nh}(i,k) p_{nh}(k,j) - p_{nh}(i,j) \Big| \\
&= |(p_{t-nh}(i,i) - 1) p_{nh}(i,j) + \sum_{k\in E\setminus\{i\}} p_{t-nh}(i,k) p_{nh}(k,j)| \\
&\le (1 - p_{t-nh}(i,i)) p_{nh}(i,j) + \sum_{k\in E\setminus\{i\}} p_{t-nh}(i,k) p_{nh}(k,j) \\
&\le 1 - p_{t-nh}(i,i) + \sum_{k\in E\setminus\{i\}} p_{t-nh}(i,k) \\
&= 1 - p_{t-nh}(i,i) + \Big(\sum_{k\in E} p_{t-nh}(i,k) - p_{t-nh}(i,i) \Big) \\
&= 2(1 - p_{t-nh}(i,i)).
\end{aligned}$$

Note that for all $i \in E$, under $\mathbb{P}_i$, we have $\{J_1 > t - nh\} \subseteq \{X_{t-nh} = i\}$. Moreover, since $n = \lfloor \frac{t}{h} \rfloor$, we have $nh \le t < (n+1)h$, hence $0 \le t - nh < h$ and therefore $\{J_1 > h\} \subseteq \{J_1 > t - nh\}$. Consequently, for all $i \in E$,

$$p_{t-nh}(i,i) = \mathbb{P}_i(X_{t-nh} = i) \ge \mathbb{P}_i(J_1 > t - nh) \ge \mathbb{P}_i(J_1 > h) = e^{-q(i)h}$$

where the last equality follows from Proposition 3.36. Hence, for all $i, j \in E$,

$$|p_t(i,j) - p_{nh}(i,j)| \le 2(1 - e^{-q(i)h}).$$

Fix $\mu \in \mathcal{M}_1(E)$. Then using Definition 3.16, the triangle inequality and the last display, we obtain for all $j \in E$,

$$\begin{aligned}
|\mathbb{P}_\mu(X_t = j) - \mathbb{P}_\mu(Z_n^h = j)| &= \Big| \sum_{i\in E} \mu(i)(p_t(i,j) - p_{nh}(i,j)) \Big| \\
&\le \sum_{i\in E} \mu(i) |p_t(i,j) - p_{nh}(i,j)| \\
&\le 2 \sum_{i\in E} \mu(i)(1 - e^{-q(i)h}).
\end{aligned}$$

Hence, using the triangle inequality and the last display, we get for all $j \in E$,

$$\begin{aligned}
|\mathbb{P}_\mu(X_t = j) - m(j)| &\le |\mathbb{P}_\mu(X_t = j) - \mathbb{P}_\mu(Z_n^h = j)| + |\mathbb{P}_\mu(Z_n^h = j) - m(j)| \\
&\le |\mathbb{P}_\mu(Z_n^h = j) - m(j)| + 2 \sum_{i\in E} \mu(i)(1 - e^{-q(i)h}).
\end{aligned}$$

Using (5.24) (with $n = \lfloor \frac{t}{h} \rfloor$, so that $n \to \infty$ as $t \to \infty$), this implies for all $j \in E$ that

$$\limsup_{t\to\infty} |\mathbb{P}_\mu(X_t = j) - m(j)| \le 2 \sum_{i\in E} \mu(i)(1 - e^{-q(i)h}).$$

Note that for all $i \in E$ we have $0 \le \mu(i)(1 - e^{-q(i)h}) \le \mu(i)$ since $e^{-q(i)h} \in [0,1]$. Moreover, $\sum_{i\in E} \mu(i) = 1 < \infty$ and $\lim_{h\to 0} \mu(i)(1 - e^{-q(i)h}) = 0$ for all $i \in E$. By the dominated convergence theorem, the right-hand side in the last display converges to 0 as $h \to 0$. Since the left-hand side does not depend on h, we deduce for all $j \in E$ that

$$\lim_{t\to\infty} |\mathbb{P}_\mu(X_t = j) - m(j)| = 0.$$

Using the equivalence of points (b) and (c) in Lemma 5.19, it follows that

$$\lim_{t\to\infty} \sum_{j\in E} |\mathbb{P}_\mu(X_t = j) - m(j)| = 0.$$

□

Applying Lemma 5.19, the previous theorem shows that, in the positive recurrent case, the distribution of the minimal process $\mathbf{X}$ at a fixed time converges toward the stationary distribution, independently of the initial distribution. In other words, the law of X_t approaches the invariant probability measure as $t \to \infty$. However, convergence in distribution describes the global behavior of the process at large times and does not describe the long-term behavior of a single trajectory. We now conclude this qualitative study with the ergodic theorems, which address the convergence of time averages along individual sample paths.

Recall that for a measure ν on a countable space E and a function $h : E \to \mathbb{R}$, we use the notation

$$\langle \nu, h\rangle := \int_E \mathrm{d}\nu\, h = \sum_{i\in E} \nu(i)h(i) \quad \text{and} \quad \langle \nu, |h|\rangle := \int_E \mathrm{d}\nu\, |h| = \sum_{i\in E} \nu(i)|h(i)|.$$

Theorem 5.21 (Ergodic theorems)**.** *Let G be an infinitesimal generator, assumed to be irreducible and recurrent (not necessarily positive recurrent). Denote by $\mathbf{X} = (X_t)_{t\in\mathbb{R}_+}$ the associated minimal process. By Proposition 5.11, it admits a G-invariant measure m. Let $f, g : E \to \mathbb{R}$ be two functions such that*

$$\langle m, |f|\rangle < \infty, \quad \langle m, |g|\rangle < \infty \quad \textit{and} \quad \langle m, g\rangle \neq 0.$$

(i) (Ratio ergodic theorem) For any initial distribution $\mu \in \mathcal{M}_1(E)$, we have

$$\frac{\int_0^t \mathrm{d}s\, f(X_s)}{\int_0^t \mathrm{d}s\, g(X_s)} \xrightarrow[t\to\infty]{} \frac{\langle m, f\rangle}{\langle m, g\rangle} \quad \mathbb{P}_\mu\text{-a.s.}$$

(ii) (Ergodic theorem) Suppose that G is positive recurrent. Assume that m is the unique G-invariant probability measure (see Theorem 5.17). Then, for any initial distribution $\mu \in \mathcal{M}_1(E)$, we have

$$\frac{1}{t}\int_0^t \mathrm{d}s\, f(X_s) \xrightarrow[t\to\infty]{} \langle m, f\rangle \quad \mathbb{P}_\mu\text{-a.s.}$$

Proof. (i) Fix $\mu \in \mathcal{M}_1(E)$ and choose a state $i_0 \in E$ such that $m(i_0) > 0$. Since G is irreducible and recurrent, the transition matrix Q of the skeleton chain $\mathbf{Y}$ is irreducible and recurrent. By Corollary I.3.8 in Section I.3.b of [Duq23], $\mathbb{P}_\mu$-a.s. the chain $\mathbf{Y}$ visits i_0 infinitely often. Moreover, since G is recurrent, point (c) of Proposition 5.6 implies that the minimal Markov process $\mathbf{X}$ does not explode. Hence, each visit of $\mathbf{Y}$ to i_0 yields a visit of $\mathbf{X}$ to i_0, and thus $\mathbf{X}$ also visits i_0 infinitely often $\mathbb{P}_\mu$-a.s.

Next, define the successive return times

$$\tau_0 := 0, \quad \tau_1 := T^+_{i_0} \quad \text{and} \quad \tau_{p+1} := \inf\{t > \tau_p : X_t = i_0\} \text{ for all } p \geq 1$$

which are all finite $\mathbb{P}_\mu$-a.s. as we have just shown. For $p \geq 1$, define the p-th excursion away from i_0 by

$$\mathscr{E}^{(p)} := (X_{\tau_{p-1}+s})_{0\leq s<\tau_p-\tau_{p-1}}$$

and for any function $h : E \to \mathbb{R}$, set

$$V_p(h) := \int_{\tau_{p-1}}^{\tau_p} \mathrm{d}s\, h(X_s).$$

Using the strong Markov property at times τ_p, we now show that the excursions $(\mathscr{E}^{(p)})_{p\geq 2}$ are i.i.d. under $\mathbb{P}_\mu$, with the same law as $\mathscr{E}^{(1)}$ under $\mathbb{P}_{i_0}$. Let $p \geq 2$ and let $F : (E^*)^{\mathbb{R}_+} \to \mathbb{R}$ be bounded and

measurable. Applying the strong Markov property at time τ_{p-1} and using that $\mathbb{P}_\mu(\tau_{p-1} < \infty) = 1$, we obtain $\mathbb{P}_\mu$-a.s.

$$\mathbb{E}_\mu[F(\theta_{\tau_{p-1}}\mathbf{X}) \mid \mathcal{F}_{\tau_{p-1}}] = \mathbb{E}_{X_{\tau_{p-1}}}[F(\mathbf{X})] = \mathbb{E}_{i_0}[F(\mathbf{X})].$$

Now choose F of the form $F(\mathbf{X}) := H((X_s)_{0\le s<\tau_1})$, where H is bounded and measurable. Then $F(\theta_{\tau_{p-1}}\mathbf{X}) = H(\mathscr{E}^{(p)})$, because the p-th excursion of $\mathbf{X}$ is exactly the segment of the shifted path $\theta_{\tau_{p-1}}\mathbf{X}$ up to its first return to i_0. With this choice, the identity in the previous display becomes

$$\mathbb{E}_\mu[H(\mathscr{E}^{(p)}) \mid \mathcal{F}_{\tau_{p-1}}] = \mathbb{E}_{i_0}[H(\mathscr{E}^{(1)})] \quad \mathbb{P}_\mu\text{-a.s.}$$

This identity implies that, conditionally on $\mathcal{F}_{\tau_{p-1}}$, the law of $\mathscr{E}^{(p)}$ under $\mathbb{P}_\mu$ coincides with the law of $\mathscr{E}^{(1)}$ under $\mathbb{P}_{i_0}$ and is independent of the past. Therefore, the excursions $(\mathscr{E}^{(p)})_{p\ge2}$ are i.i.d. under $\mathbb{P}_\mu$.

By Proposition 5.11, for all $j \in E$,

$$m(j) = m(i_0)q(i_0)\mathbb{E}_{i_0}\Big[\int_0^{T^+_{i_0}} \mathrm{d}t\, \mathbf{1}_{\{X_t=j\}}\Big].$$

Hence, for any $h : E \to \mathbb{R}$ such that $\langle m, |h|\rangle < \infty$, multiplying both sides by $h(j)$, summing over $j \in E$ and using Tonelli's theorem (applied to h^+ and h^-) yields

$$\begin{aligned}
\langle m, h\rangle &= \sum_{j\in E} m(j)h(j) \\
&= \sum_{j\in E} m(i_0)q(i_0)\mathbb{E}_{i_0}\Big[\int_0^{T^+_{i_0}} \mathrm{d}t\, \mathbf{1}_{\{X_t=j\}}\Big]h(j) \\
&= m(i_0)q(i_0)\mathbb{E}_{i_0}\Big[\int_0^{T^+_{i_0}} \mathrm{d}t \sum_{j\in E} \mathbf{1}_{\{X_t=j\}}h(j)\Big] \\
&= m(i_0)q(i_0)\mathbb{E}_{i_0}\Big[\int_{\tau_0}^{\tau_1} \mathrm{d}t\, h(X_t)\Big] \\
&= m(i_0)q(i_0)\mathbb{E}_{i_0}[V_1(h)].
\end{aligned}$$

Since $m(i_0)q(i_0) > 0$, $\langle m, |f|\rangle < \infty$ and $\langle m, |g|\rangle < \infty$, we obtain

$$\mathbb{E}_{i_0}[|V_1(f)|] \le \mathbb{E}_{i_0}[V_1(|f|)] = \tfrac{\langle m,|f|\rangle}{m(i_0)q(i_0)} < \infty \quad \text{and} \quad \mathbb{E}_{i_0}[|V_1(g)|] \le \mathbb{E}_{i_0}[V_1(|g|)] = \tfrac{\langle m,|g|\rangle}{m(i_0)q(i_0)} < \infty$$

so $V_1(f)$ and $V_1(g)$ are integrable under $\mathbb{P}_{i_0}$. Moreover,

$$\mathbb{E}_{i_0}[V_1(f)] = \tfrac{\langle m,f\rangle}{m(i_0)q(i_0)} \quad \text{and} \quad \mathbb{E}_{i_0}[V_1(g)] = \tfrac{\langle m,g\rangle}{m(i_0)q(i_0)}.$$

Since $V_p(f)$ (resp. $V_p(g)$) depends only on the p-th excursion $\mathscr{E}^{(p)}$, the i.i.d. property of $(\mathscr{E}^{(p)})_{p\ge2}$ immediately implies that $(V_p(f))_{p\ge2}$ (resp. $(V_p(g))_{p\ge2}$) are i.i.d. under $\mathbb{P}_\mu$, with the same law as $V_1(f)$ (resp. $V_1(g)$) under $\mathbb{P}_{i_0}$. Hence, by the law of large numbers, $\mathbb{P}_\mu$-a.s.

$$\tfrac{1}{p}\sum_{k=2}^{p+1} V_k(f) \underset{p\to\infty}{\longrightarrow} \mathbb{E}_{i_0}[V_1(f)] = \tfrac{\langle m,f\rangle}{m(i_0)q(i_0)} \quad \text{and} \quad \tfrac{1}{p}\sum_{k=2}^{p+1} V_k(g) \underset{p\to\infty}{\longrightarrow} \mathbb{E}_{i_0}[V_1(g)] = \tfrac{\langle m,g\rangle}{m(i_0)q(i_0)}.$$

Since $\frac{1}{p}V_1(f) \to 0$ (resp. $\frac{1}{p}V_1(g) \to 0$) and $\frac{1}{p}V_{p+1}(f) \to 0$ (resp. $\frac{1}{p}V_{p+1}(g) \to 0$) as $p \to \infty$, $\mathbb{P}_\mu$-a.s., the same limit holds with $\sum_{k=1}^{p}$ in place of $\sum_{k=2}^{p+1}$.

Let $N_{i_0}(t) := \sup\{p \ge 0 : \tau_p \le t\}$ be the number of completed excursions away from i_0 up to time t. Since $\mathbf{X}$ returns to i_0 infinitely often, $N_{i_0}(t) \to \infty$ as $t \to \infty$, $\mathbb{P}_\mu$-a.s., and $\tau_{N_{i_0}(t)} \le t < \tau_{N_{i_0}(t)+1}$ for all $t \in \mathbb{R}_+$ $\mathbb{P}_\mu$-a.s. Hence, for any $h : E \to \mathbb{R}$ such that $\langle m, |h|\rangle < \infty$, $\mathbb{P}_\mu$-a.s.

$$\int_0^t \mathrm{d}s\, h(X_s) = \sum_{p=1}^{N_{i_0}(t)} V_p(h) + \int_{\tau_{N_{i_0}(t)}}^t \mathrm{d}s\, h(X_s).$$

Again, by the law of large numbers and since $N_{i_0}(t) \to \infty$ as $t \to \infty$, the first term satisfies

$$\tfrac{1}{N_{i_0}(t)} \sum_{p=1}^{N_{i_0}(t)} V_p(h) \underset{t\to\infty}{\longrightarrow} \tfrac{\langle m,h\rangle}{m(i_0)q(i_0)} \quad \mathbb{P}_\mu\text{-a.s.}$$

For the second term, we have $\mathbb{P}_\mu$-a.s.

$$\Big|\int_{\tau_{N_{i_0}(t)}}^{t} \mathrm{d}s\, h(X_s)\Big| \le \int_{\tau_{N_{i_0}(t)}}^{\tau_{N_{i_0}(t)+1}} \mathrm{d}s\, |h(X_s)| = V_{N_{i_0}(t)+1}(|h|)$$

and since $\frac{1}{N_{i_0}(t)} V_{N_{i_0}(t)+1}(|h|) \to 0$ as $t \to \infty$ $\mathbb{P}_\mu$-a.s., we obtain

$$\tfrac{1}{N_{i_0}(t)} \int_{\tau_{N_{i_0}(t)}}^{t} \mathrm{d}s\, h(X_s) \underset{t\to\infty}{\longrightarrow} 0 \quad \mathbb{P}_\mu\text{-a.s.}$$

Consequently, for any $h : E \to \mathbb{R}$ such that $\langle m, |h|\rangle < \infty$,

$$\tfrac{1}{N_{i_0}(t)} \int_0^t \mathrm{d}s\, h(X_s) \underset{t\to\infty}{\longrightarrow} \tfrac{\langle m,h\rangle}{m(i_0)q(i_0)} \quad \mathbb{P}_\mu\text{-a.s.}$$

Applying this to f and g, we get in particular $\mathbb{P}_\mu$-a.s.

$$\tfrac{1}{N_{i_0}(t)} \int_0^t \mathrm{d}s\, f(X_s) \underset{t\to\infty}{\longrightarrow} \tfrac{\langle m,f\rangle}{m(i_0)q(i_0)} \quad \text{and} \quad \tfrac{1}{N_{i_0}(t)} \int_0^t \mathrm{d}s\, g(X_s) \underset{t\to\infty}{\longrightarrow} \tfrac{\langle m,g\rangle}{m(i_0)q(i_0)}.$$

Since $\langle m, g\rangle \neq 0$, it follows that $\int_0^t \mathrm{d}s\, g(X_s) \neq 0$ for all sufficiently large t $\mathbb{P}_\mu$-a.s. and therefore

$$\frac{\int_0^t \mathrm{d}s\, f(X_s)}{\int_0^t \mathrm{d}s\, g(X_s)} \underset{t\to\infty}{\longrightarrow} \tfrac{\langle m,f\rangle}{\langle m,g\rangle} \quad \mathbb{P}_\mu\text{-a.s.}$$

which proves (i).

(ii) Fix $\mu \in \mathcal{M}_1(E)$. Assume that G is positive recurrent and that m is the unique G-invariant probability measure. Apply (i) with $g \equiv 1$, for which $\langle m, 1\rangle = \sum_{i\in E} m(i) = 1$, to obtain

$$\tfrac{1}{t} \int_0^t \mathrm{d}s\, f(X_s) \underset{t\to\infty}{\longrightarrow} \langle m, f\rangle \quad \mathbb{P}_\mu\text{-a.s.}$$

which concludes the proof. □

The ratio ergodic theorem states that, in the recurrent case, the ratio of two time averages computed along a single realization of the minimal process converges almost surely to the corresponding ratio of integrals with respect to the invariant measure, independently of the initial distribution. Thus, even when time averages themselves do not stabilize, the invariant measure still determines their relative long-term behavior.

In the positive recurrent case, the ergodic theorem yields a stronger result. It states that the time average along a single trajectory converges almost surely, independently of the initial distribution, to the corresponding integral with respect to the invariant probability measure. In particular, for a given state $i \in E$, taking $f = \mathbf{1}_{\{i\}}$ in the ergodic theorem yields, for any initial distribution $\mu \in \mathcal{M}_1(E)$,

$$\frac{1}{t} \int_0^t \mathrm{d}s\, \mathbf{1}_{\{X_s=i\}} \underset{t\to\infty}{\longrightarrow} \int_E \mathrm{d}m\, \mathbf{1}_{\{i\}} = m(i) \quad \mathbb{P}_\mu\text{-a.s.}$$

This shows that the fraction of time a single trajectory spends in state i converges almost surely to $m(i)$. This explains why the stationary distribution describes the long-term behavior observed along typical realizations of the minimal process.

6 Conclusion and Outlook

This thesis has presented a self-contained introduction to the theory of continuous-time Markov processes on discrete state spaces. The guiding idea throughout was to understand how the global and long-term behavior of such processes emerges from their local jump dynamics.

We began by defining continuous-time Markov processes on discrete state spaces through their transition semigroups and by establishing the simple Markov property. In order to avoid irregular examples, we restricted our attention to right-continuous processes. This made it possible to describe the process at the level of individual trajectories, in terms of jump times, the explosion time, the skeleton chain and holding times. This decomposition provides an intuitive picture of how the process evolves over time and serves as the basis for the proof of the strong Markov property, which is essential for many of the results developed later.

A key object in the theory is the infinitesimal generator, which encodes the jump rates and transition probabilities of the process. We studied its relation to the transition semigroup and derived the backward and forward Chapman-Kolmogorov equations, linking the probabilistic description of the process to systems of linear differential equations.

We then reversed the point of view and showed how a minimal right-continuous Markov process can be constructed starting from a given infinitesimal generator. Rather than beginning with a process and extracting its generator, this approach starts from prescribed jump rates and transition probabilities and builds the process pathwise by combining a discrete-time skeleton chain with exponentially distributed holding times. We showed that the resulting process has a sub-Markovian transition semigroup that is the minimal nonnegative solution of the Chapman-Kolmogorov equations associated with the generator. The construction of the minimal Markov process was then illustrated by simulation examples.

The final part of the thesis was dedicated to qualitative long-term properties. We investigated recurrence and transience and showed that these notions coincide with those of the associated discrete-time skeleton chain. This provided a natural connection between the two settings and allowed results from the discrete-time framework to be carried over to the continuous-time framework. We then established criteria for non-explosion and introduced invariant measures, which intuitively describe how the process distributes its occupation time among the states. This led to a characterization of positive recurrence via the existence of an invariant probability measure. Finally, we established ergodic theorems, showing how invariant measures describe the long-term behavior of the process, both in terms of convergence to equilibrium and almost sure convergence of time averages along single trajectories.

The scope of this thesis was intentionally limited to the main theoretical foundations developed in Chapter II of [Duq23]. Several topics treated later in that chapter were not considered, including stopping and time-change techniques, Dynkin's formula and the detailed study of concrete model classes such as Poisson processes, random walks, birth-death processes, queueing models and branching processes. These topics rely on the framework developed here and would be a natural continuation of this thesis.

In addition, the material of Chapter III of [Duq23], which introduces Poisson point processes and Poissonian clouds, was not included, since these notions lie outside the semigroup-based approach followed in this thesis. Nonetheless, they are essential for more explicit, pathwise constructions of jump processes and branching models and would provide a natural next step toward more concrete applications.

Bibliography

[Ami] Gideon Amir. *Continuous Time Markov Chains*. Lecture notes. Bar-Ilan University. URL: https://u.math.biu.ac.il/~amirgi/CTMCnotes.pdf (visited on 01/22/2026).

[Duq23] Thomas Duquesne. *Processus markoviens 1. Chaînes de Markov; Processus de Markov à temps continu et espace d'état discret; Nuages poissonniens*. Lecture notes. Sorbonne Université. 2023–2024. URL: https://xiaoyuguang.fr/PMA/Markov-I/Poly_Markov_I.pdf (visited on 03/10/2025).

[Sha22] Shannaliu. *Pure Birth Process and Yule Process*. RPubs (RStudio). 2022. URL: https://rpubs.com/Shannaliu/878513 (visited on 11/19/2025).

[Szn17] Alain-Sol Sznitman. *Applied Stochastic Processes*. Lecture notes. ETH Zurich. Aug. 2017. URL: https://metaphor.ethz.ch/x/2021/fs/401-3602-00L/sc/lecture-notes-2017.pdf (visited on 11/16/2025).

Declaration on the use of generative AI. GPT-5 (ChatGPT, OpenAI, 2025) was employed as an auxiliary tool in the formulation of certain proofs and in improving the clarity of the text. The author retains full responsibility for all mathematical content.