Department of
Mathematics

Quasi-Stationary Distributions for Epidemic Modelling

August 29, 2018

Candidate Number: 87846

Supervisor: Graham Brightwell

A Dissertation submitted to the
Department of Mathematics of the
London School of Economics and Political Science
for the degree of Master of Science

Summary

Under certain conditions, stochastic processes with an absorbing set allow the existence of so-called quasi-stationary distributions. In the models where they can be found, those quasi-stationary distributions can be useful to understand the long-term behaviour of the process. It turns out that the distribution of such processes will also converge toward such quasi-stationary distributions as $t \to \infty$. Therefore, it can also be useful to predict the time of absorption from any point.

The topic has been extensively studied in some of its applications, in particular birth-and-death processes, which require only a restricted framework. Yet, there is a small number of ressources in the literature with a more theoretic and general approach of quasi-stationary distributions. Moreover, the treatment given in those publications is not always satisfying.

Therefore, we will first try to provide a rigorous description of quasi-stationary distributions and their properties in general a setting. We will then study applied examples, in particular in epidemics which contain a large variety of models. In the literature, the connection between the general theory on quasi-stationary distributions and those models is generally not made. Thus, we will discuss how general results can help us study those cases and eventually get insights in the spread of epidemics.

Contents

Chapter 1

Introduction

Many mathematical models describing natural phenomena are balanced in a way such that they are unlikely to explode and will eventually collapse. But the reasonable long time behaviour that they will exhibit beforehand is something that we would like to predict. Quasi-stationary distributions are useful to provide information about this behaviour as well as to help predict when the collapse would happen. The underlying theory helps us understand why in many cases a situation which is not an equilibrium will eventually display a form of stability.

1.1 An intuitive description of quasi-stationary distributions

We provide here a basic explanation of quasi-stationary distributions and their use, with a simple approach and without the theoretical framework which we use in the following chapter. Historical elements presented in this section come from [19] and [28].

We consider a random Markov process $\{Y_t\}$ taking values in a discrete space. For instance, Y_t can be the number of people in a population, in which case Y_t takes values in the set of positive integers and is said to follow a *birth-and-death process*.

The particularity of the processes we are interested in is that some of their possible states are traps, called *absorbing states*. When the process ends up falling into one of those states we say that it has been *killed*. We are interested in the study of those processes before they get killed. In our example, 0 is an absorbing state: when there is no one left in the population nothing can happen anymore.

We generally have an origin of time $t = 0$, and we have some information regarding the value Y_0, in our example the number of people at time 0 in the population. This information comes as a probability distribution ν. Thus, when the states are described by positive integers as for birth-and-death processes, the distribution is determined by the sequence of probabilities $\nu(\{0\}), \nu(\{1\}), \ldots$ such that Y_0 is equal to n with probability $\nu(\{n\})$.

Then, this probability distribution evolves as time goes by. Take for example a birth-and-death process starting from a fixed point, which is equivalent to say that Y_0 follows a Dirac distribution. Possible paths for this process are represented in figure 1.1. In the absence of any additional information, what we know of the random variable Y_t for $t > 0$ is that it follows a probability distribution which depends on two things: the initial distribution ν of Y_0 and the

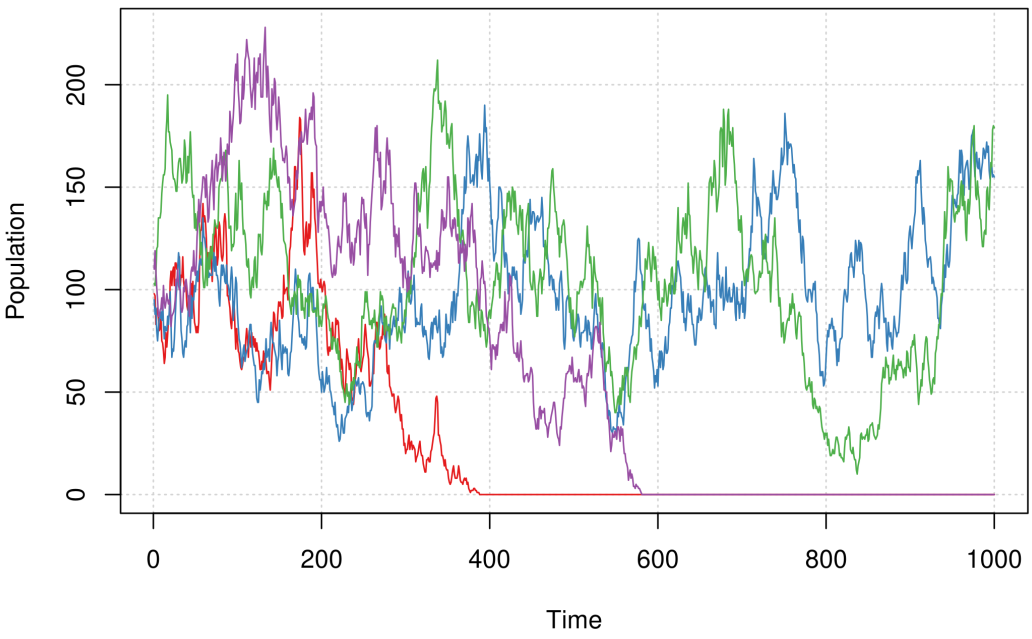

Figure 1.1: Several birth-and-death processes starting from the same point

transition probabilities characterizing the process.

We add one information to this: the fact that, at time $t > 0$, we did not reach an absorbing state yet. Then, we can describe the behaviour of the process conditioned on it being still "alive". This yields a probability distribution on the set of non-absorbing states.

Suppose now that the process never starts in an absorbing state, so that the initial distribution ν can be described as a distribution on the transient states only. From the previous paragraph, at any moment t we have a distribution, call it ν_t, on the transient states which describes Y_t conditioned upon the fact that it is not in an absorbing state, i.e. $Y_t \neq 0$ in our example.

A quasi-stationary distribution, which we will formally define in section 3.2, is an initial distribution ν such that any subsequent conditional distribution $\nu_t, t > 0$ is still equal to ν. This means that, if the process is still alive at time t, then it is expected to take the same values with the same probabilities than at time 0.

Depending on the model, there can be only one possible quasi-stationary distribution, a finite number, an infinity or none at all. Therefore, a large part of the literature in the field is about counting and identifying those distributions in particular settings.

Moreover, in practical applications, we don't normally start following a quasi-stationary distribution. But what often happens is that the sequence $(\nu_t)_{t \geqslant 0}$ of conditional distributions converges to a distribution ν_∞ on the transient states, which happens to be a quasi-stationary distribution. This notion is also formalized in section 3.2. Therefore, provided that the convergence to ν_∞ is fast enough compared to the expected lifetime of the process, the knowledge of quasi-stationary distributions provides a good description of the long-term behaviour of a system prior to killing.

This is summarized by Bartlett in [5], who later coined the term "quasi-stationary distribution":

> "It still may happen that the time to extinction is so long that it is still of more relevance to consider the effectively ultimate distribution [of the process]"

Earlier observations about quasi-stationarity, in the absence of theoretical formalism, were made in the middle of the XX$^{\text{th}}$ century in the fields of genetics or biological systems. In addition to the fields already mentioned, and epidemic models that we will develop here, quasi-stationary distributions can be used in a wide range of applications. A list of other possible uses, established in [28] includes cellular automata, immunology, medical decision making, physical chemistry, queues, telecommunications, etc.

1.2 Epidemic models

We are interested here in the application of quasi-stationary distributions to *compartemental models* of epidemic propagation. The essential simplification made by such models is that we can divide the population into separate compartments and individuals have homogeneous characteristics inside each of those. For example, the category of infectious people is one such compartment.

A classic reference regarding epidemic models is the book by Andersson and Britton [3] which describes various kinds of models. In particular, it distinguishes the deterministic and stochastic models. We give here a basic description of deterministic models in order to introduce the main ideas but we will later focus exclusively on stochastic models in chapter 5.

Those models hold in the case of viral or bacterial infection with a person-to-person transmission mechanism. This includes notably sexually transmitted diseases and childhood diseases such as measles, chickenpox and rubella. It is less adapted to host-vector and parasitic infections. However, some non-medical applications of the model exist, for example the spread of rumours.

One of the most basic model is *SIS*, which stands for the following mechanism:

$$\text{Susceptible} \rightarrow \text{Infectious} \rightarrow \text{Susceptible.}$$

This means that there are two compartments, infectious and susceptible, and that any person who recovers can get sick again anytime.

We denote by N the total size of the population, by S the number of susceptible people and by I the number of infectious ones, so that $N = S + I$. In a stochastic setting those values are random variables. In the deterministic version they are continuous functions of time governed by the following differential equations:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta \frac{SI}{N} + \gamma I \quad \text{and} \quad \frac{\mathrm{d}I}{\mathrm{d}t} = \beta \frac{SI}{N} - \gamma I$$

where $\beta \in \mathbb{R}_+$ is called *contact rate* and $\gamma \in \mathbb{R}_+$ is the *recovery rate*. The equations come from the representation of the N individuals living together in some restricted space. Thus, the term $\beta \frac{SI}{N}$ comes from the consideration that every susceptible person will meet every day other individuals, some of which may be infected and will infect the susceptible one with a certain probability, which we can compound in a factor $\beta I/N$. When summing over all susceptible we then get that $\beta\frac{SI}{N}$ is proportionate to the number of susceptible people getting infected at every step.

It is generally assumed that there is no vital dynamics, meaning that N is constant. This is justified by the fact that the spread of a disease is generally much faster than demographic evolutions. However, in some situations this hypothesis does not hold, which requires more complex models and is studied in chapter 6.

The second classic model is *SIR*, with a mechanism similar to the first:

$$\text{Susceptible} \rightarrow \text{Infectious} \rightarrow \text{Recovered}.$$

The difference is that recovery confers a long-term resistance to the disease, and thus creates a third compartment whose size is denoted by R. The deterministic model is then

$$N = S + I + R, \quad \frac{\mathrm{d}S}{\mathrm{d}t} = -\beta\frac{SI}{N}, \quad \frac{\mathrm{d}I}{\mathrm{d}t} = \beta\frac{SI}{N} - \gamma I \quad \text{and} \quad \frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I\ .$$

The stochastic models that we will review in chapter 5 are morally similar to the deterministic one. They use the same rates, and in both cases they are very dependent on the value called *basic reproduction number* or *basic reproductive ratio* which is defined as

$$R_0 = \frac{\beta}{\gamma}\ .$$

Typically, this ratio can go from 1.5 for diseases like Ebola during the 2014 outbreak and in the absence of control measures [1], to about 18 for measles which is airborne [10].

1.3 Organisation of the dissertation

Apart from the first description of quasi-stationary distributions in section 1.1, this dissertation is written with a top-down approach and a gradual degree of specialization. In order to be accessible to people who are not familiar with Markov processes, but who have a graduate-level knowledge in continuous stochastic processes, chapter 2 provides an introduction to the key notions that we use. It focuses on definitions and properties which are of use in the rest of the dissertation. It also sets the notation.

In chapter 3 we use this theoretical background to build the general setting of quasi-stationary distributions. Most of the content of this chapter comes from some key books and articles, but we focus here on three goals:

- to make it more accessible than the original content by developing the proofs,
- to use a notation which enables a continuity of the theory and that can be used in its applications,
- to smoothly specialize our content.

In chapter 4 we restrict slightly our approach. As practical applications are usually on countable spaces we study how the general setting used before translates in that case. We particularly care to explain how every new step that we take is still relying on the theory preceding it, which often lacks in the literature. For instance, we progressively add assumptions so that every result given at some point can be used in the following parts.

Then, in chapter 5 we describe several stochastic models of epidemic modelling using the notations and the setting of the previous chapter. We review the main results obtained from those models, using the theory developed in the previous chapter.

Chapter 6 is again dedicated to epidemic modelling, but focuses more specifically on models in which the overall population is not constant.

Finally, we review some techniques in chapter 7 to model those processes and examine some results obtained on quasi-stationary distributions from our simulations.

Chapter 2

Markov processes

This chapter provides theoretical background on Markov processes, mostly taken from [22] and [15], with elements from [2] and [6]. It can be skipped by any reader already at ease with this topic.

Although they are not at the core of this dissertation's topic, the following definitions and theorems serve as a basis for the theory of quasi-stationary distributions developed in chapter 3. We only give an introduction to the topic, with the notions necessary to understand the rest of the dissertation. Since those results are very well-known we state the theorems without copying the proofs.

2.1 Notation

Before introducing the theory of Markov processes we need some notations which will be of use in all the dissertation.

We denote by $\mathbb{N}$ the set of nonnegative integers, $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$, $\mathbb{R}$ the set of real numbers and $\mathbb{Q}$ the set of rational numbers. Moreover, $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$, $\mathbb{R}_+ = [0, +\infty[$ and $\bar{\mathbb{R}}_+ = \bar{\mathbb{R}}_+ \cup \{+\infty\}$.

The characteristic function of a set A is written $\mathbf{1}_A$. If A is a set with a topological structure, $\mathscr{B}(A)$ denotes the collection of all Borel sets on A.

Let $(\mathsf{X}, \mathscr{X})$ be a measurable space. We denote by $\mathscr{M}(\mathsf{X})$ the set of measures on $(\mathsf{X}, \mathscr{X})$ and by $\mathscr{P}(\mathsf{X})$ the set of probability measures.

For any measurable function $f : \mathsf{X} \to \mathbb{R}$ we denote $\|f\|_1 := \int |f|$ and $\|f\|_\infty := \sup_{x \in \mathsf{X}} |f(x)|$. We define $L^1(\mathsf{X})$ (resp. $L^\infty(\mathsf{X})$) as the set of real measurable functions whose norms for $\|\cdot\|_1$ (resp. $\|\cdot\|_\infty$) are finite. When no set is specified, we assume that L^1 and L^∞ contain functions on the state space X. Then, we define the scalar product $\langle \cdot \mid \cdot \rangle$ by

$$\forall f \in L^1, \forall g \in L^\infty, \langle f \mid g \rangle := \int f g \ .$$

If $\alpha \in \mathscr{P}(\mathsf{X})$ and $f \in L^\infty$ then we denote $\alpha(f) := \int_{\mathsf{X}} f(x) \alpha(\mathrm{d}x)$.

Furthermore, we denote by $(\theta_s)_{s \in \mathbb{R}_+}$ the family of *shift operators*. That is, if Ω is the set of sample paths with $\Omega \subset \mathsf{X}^{\mathbb{R}_+}$, then for any $s \in \mathbb{R}_+$, θ_s is defined by

$$\forall \omega = (\omega_t)_{t \geqslant 0} \in \Omega, \quad \theta_s((\omega_t)_{t \geqslant 0}) := (\omega_{t+s})_{t \geqslant 0} \ . \tag{2.1}$$

2.2 Transition functions and Markov processes

Let $(\mathsf{X}, \mathscr{X})$ be a measurable space.

Definition 2.1. A *kernel* N on X is a map from $\mathsf{X} \times \mathscr{X}$ into $\bar{\mathbb{R}}_+$ such that

(i) for every $x \in \mathsf{X}$, the map $A \mapsto N(x,A)$ is a positive measure on $\mathscr{X}$,

(ii) for every $A \in \mathscr{X}$, the map $x \mapsto N(x,A)$ is $\mathscr{X}$-measurable.

If, in addition, $\forall x \in \mathsf{X}, N(x,\mathsf{X}) = 1$ then N is called a *transition probability*.

Let $f : \mathsf{X} \to \mathbb{R}$ be a bounded (resp. nonnegative) and $\mathscr{X}$-measurable function. We define $Nf : \mathsf{X} \to \mathbb{R}$ by

$$Nf(x) := \int_{\mathsf{X}} N(x, \mathrm{d}y) f(y) \tag{2.2}$$

and Nf is also bounded (resp. nonnegative) and $\mathscr{X}$-measurable. If M is another kernel, then MN defined by

$$MN(x,A) := \int_{\mathsf{X}} M(x, \mathrm{d}y) N(y,A)$$

is again a kernel.

Definition 2.2. A *transition function* on $(\mathsf{X}, \mathscr{X})$ is a family $(P_{s,t})_{0 \leq s < t}$ of transition probabilities on $(\mathsf{X}, \mathscr{X})$ such that they satisfy the Chapman-Kolmogorov equation:

$$\forall (s,t,v) \in \mathbb{R}_+^3, s < t < v \implies \forall x \in \mathsf{X}, \forall A \in \mathscr{X}, \int_{\mathsf{X}} P_{s,t}(x, \mathrm{d}y) P_{t,v}(y,A) = P_{s,v}(x,A) \ .$$

The transition function is said to be homogeneous if $P_{s,t}$ depends only on $t-s$, in which case we write P_t for $P_{0,t}$.

If the transition function is homogeneous then the Chapman-Kolmogorov equation reads

$$\forall s,t \geqslant 0, \quad P_{t+s}(x,A) = \int_{\mathsf{X}} P_s(x, \mathrm{d}y) P_t(y,A) \ . \tag{2.3}$$

In other words, the family $(P_t)_{t \geqslant 0}$ forms a semigroup with $\forall t,s \geqslant 0, P_{t+s} = P_t P_s$.

In the case of quasi-stationary distributions and their applications in epidemic modelling, the transition functions that we use are generally homogeneous, i.e. only the difference $t-s$ matters. It means here that the transition rates do not evolve over time. But if we were to study for example seasonal changes in the propagation of an epidemy, then the use of a more general transition function could be a better model.

Definition 2.3. Let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_t, \mathbb{P})$ be a filtered probability space. An adapted process Y with values in X is a *Markov process* with respect to $(\mathscr{F}_t)$, with transition function $(P_{s,t})_{0 \leq s < t}$ if for any $\mathscr{X}$-measurable function $f : \mathsf{X} \to \mathbb{R}_+$ and any pair $(s,t) \in \mathbb{R}_+^2$ with $s < t$,

$$\mathbb{E}[f(Y_t) \mid \mathscr{F}_s] = P_{s,t} f(Y_s) \quad \mathbb{P}-\text{a.s.}$$

The law of Y_0 is called the *initial distribution* of Y. The process is said to be *homogeneous* if the transition function is homogeneous, in which case the above equality reads

$$\mathbb{E}[f(Y_t) \mid \mathscr{F}_s] = P_{t-s} f(Y_s) \ .$$

We sometimes extend this definition by allowing a transition function to be composed of kernels $(P_t)_t$ such that $P_t(x,\mathsf{X}) < 1$ for some $t \in \mathbb{R}_+, x \in \mathsf{X}$ while the Chapman-Kolmogorov equation still holds. In that case we say that it is a *dishonest* or *submarkovian* transition function. On the contrary, a transition function is sometimes said to be *honest* when $\forall t \in \mathbb{R}_+, \forall x \in \mathsf{X}, P_t(x,\mathsf{X}) = 1$.

We suppose now that $(\mathsf{X},\mathscr{X})$ is a Polish space, that is, a separable and completely metrizable topological space. Moreover, it is endowed with the σ-field of Borel subsets. In the context of the previous definition we set $\Omega = \mathsf{X}^{\mathbb{R}_+}$, $\mathscr{F}_\infty = \mathscr{X}^{\otimes\mathbb{R}_+}$ and $\mathscr{F}_t = \sigma(Y_u, u \leq t)$. The process Y verifies $\forall \omega \in \Omega, \forall t \in \mathbb{R}_+, Y_t(\omega) = \omega_t$ and is called the *coordinate process* or *canonical process*.

The following theorem gives a crucial property of this kind of processes: they are always Markovian under a certain probability.

Theorem 2.4. *Given a transition function $(P_{s,t})$ on $(\mathsf{X},\mathscr{X})$, for any probability measure ν on $(\mathsf{X},\mathscr{X})$, there is a unique probablity measure $\mathbb{P}_\nu$ on $(\Omega,\mathscr{F}_\infty)$ such that Y is Markov with respect to $(\mathscr{F}_t)$ with transition function $(P_{s,t})$ and initial measure ν.*

In the remaining part we consider only homogeneous processes.

By the previous theorem, for any $x \in \mathsf{X}$, denoting by δ_x the corresponding Dirac distribution, there exists a probability measure $\mathbb{P}_{\delta_x}$ such that Y is Markov and $Y_0 = x$ $\mathbb{P}_{\delta_x}$-a.s. We denote it more simply by $\mathbb{P}_x$.

Similarly, if Z is an $\mathscr{F}_\infty$-measurable and positive random variable we denote by $\mathbb{E}_\nu[Z]$ (resp. $\mathbb{E}_x[Z]$) its expectation under $\mathbb{P}_\nu$ (resp. $\mathbb{P}_x$). In the case when $Z = \mathbf{1}_{\{Y_t \in A\}}$ for some $A \in \mathscr{X}$ we then have

$$\mathbb{P}_x(Y_t \in A) = P_t(x,A) .$$

The following proposition is a generalization of this result.

Proposition 2.5. *If Z is $\mathscr{F}_\infty$-measurable and positive or bounded, the map $x \mapsto \mathbb{E}_x[Z]$ is $\mathscr{X}$-measurable and*

$$\mathbb{E}_\nu[Z] = \int_{\mathsf{X}} \nu(\mathrm{d}x)\mathbb{E}_x[Z] .$$

This proposition and the following are some of the most useful tools that we have to handle Markov processes. It is necessary to be familiar with them in order to understand the proofs given in the next chapters.

Proposition 2.6 (Markov property)**.** *If Z is $\mathscr{F}_\infty$-measurable and positive (or bounded), for every $t > 0$ and starting measure ν,*

$$\mathbb{E}_\nu[Z \circ \theta_t \mid \mathscr{F}_t] = \mathbb{E}_{Y_t}[Z] \quad \mathbb{P}_\nu - \text{a.s.}$$

Finally, we shall introduce the resolvent of the transition function. Although it is of major importance in the theory of Markov processes, we will need it explicitely only in some proofs of chapter 6, so the reader can skip it at first.

Definition 2.7. Let $\lambda > 0$. We call *λ-resolvent* of the transition semigroup $(P_t)_{t\geqslant 0}$ the linear operator $R_\lambda : L^\infty \to L^\infty$ defined by

$$\forall f \in L^\infty, \forall x \in \mathsf{X}, \quad R_\lambda f(x) := \int_0^\infty e^{-\lambda t} P_t f(x) \, \mathrm{d}t \ .$$

Proposition 2.8. *Let $\lambda > 0$. Then, the λ-resolvent has the following properties:*

(i) $\|\lambda R_\lambda\|_\infty \leq 1$,

(ii) $\forall \lambda, \mu > 0, R_\lambda - R_\mu + (\lambda - \mu) R_\lambda R_\mu = 0$ *(resolvent equation)*.

2.3 Feller processes

Feller semigroups are an important category of transition functions that we always use in the case of quasi-stationary distributions. We can see them as a family of positive linear operators and use (2.2) to associate a transition function to it. We will use here the definitions from [15] which are more directly useful for us.

We assume from now on that X is a metrizable locally compact topological space, countable at infinity, i.e. X is a countable union of compact sets. We endow X with its Borel σ-field.

We denote by $C_0(\mathsf{X})$ the set of continuous functions on X that tend to 0 at infinity, meaning that, for every $\varepsilon > 0$, there exists $K \subset \mathsf{X}$ compact such that $\forall x \in \mathsf{X} \setminus K, |f(x)| \leqslant \varepsilon$. This is a Banach space for the supremum norm $\|\cdot\|_\infty$.

Definition 2.9. Let $(P_t)_{t\geqslant 0}$ be a homogeneous transition semigroup on X. We say that (P_t) is a *Feller semigroup* if

(i) $\forall f \in C_0(\mathsf{X}), \forall t \geqslant 0, P_t f \in C_0(\mathsf{X})$,

(ii) $\forall f \in C_0(\mathsf{X}), \lim_{t\to 0} \|P_t f - f\| = 0$.

A Markov process with values in X is a *Feller process* if its semigroup is Feller.

Moreover, condition *(ii)* can be replaced by $\forall f \in C_0(\mathsf{X}), \forall x \in \mathsf{X}, \lim_{t\downarrow 0} P_t f(x) = f(x)$.

In the following, we fix a Feller semigroup $(P_t)_{t\geqslant 0}$ on X and denote by Y the canonical version of a Feller process with semigroup $(P_t)_{t\geqslant 0}$.

Theorem 2.10. *The process Y admits a càdlàg modification (right-continuous with left limits).*

This statement means that there exists a càdlàg process $\tilde{Y}$ on $(\Omega, \mathscr{F})$ such that $\tilde{Y}_t = Y_t$ $\mathbb{P}_\nu$—a.s. for any t and every initial distribution ν, and which is a Feller process with semigroup $(P_t)_t$. This is a major advantage of working with Feller semigroups: it is not actually restrictive to consider only càdlàg sample paths.

Then, we shall introduce the generator of the transition function, sometimes also called *infinitesimal generator* in the case of a discrete state space. It is a key tool to describe processes and work on quasi-stationary distributions.

Definition 2.11. We set $D(L) = \left\{ f \in C_0(\mathsf{X}) \mid \frac{P_t f - f}{t} \text{ converges in } C_0(\mathsf{X}) \text{ when } t \downarrow 0 \right\}$ and, for every $f \in D(L)$,

$$Lf = \lim_{t\downarrow 0} \frac{P_t f - f}{t} \ .$$

Then $D(L)$ is a linear subspace of $C_0(\mathsf{X})$ called the *domain* of L and $L : D(L) \to C_0(\mathsf{X})$ is a linear operator called the *generator* of the semigroup (P_t).

The following propositions explain then why the generator is particularly useful in describing a Markov process.

Proposition 2.12. *Let $f \in D(L)$ and $s > 0$. Then, $Q_s f \in D(L)$ and $L(Q_s f) = Q_s(Lf)$.*

Proposition 2.13. *The semigroup $(P_t)_{t\geqslant 0}$ is determined by the generator L and its domain.*

Moreover, in the case of a Feller process the resolvent has additional properties that we will exploit.

Proposition 2.14. *Let $\lambda > 0$ and set $\mathscr{R} = \{R_\lambda f \mid f \in C_0(\mathsf{X})\}$. Then, $\mathscr{R}$ does not depend on the choice of λ. Furthermore, $\mathscr{R}$ is a dense subset of $C_0(\mathsf{X})$.*

Proposition 2.15. *We have $D(L) = \mathscr{R}$. Moreover, for any $\lambda > 0$, the operator $R_\lambda \colon C_0(\mathsf{X}) \to \mathscr{R}$ and $\lambda - L \colon D(L) \to C_0(\mathsf{X})$ are the inverse of each other, i.e.*

$$\forall g \in C_0(\mathsf{X}),\ (\lambda - L)R_\lambda g = g, \quad \textit{and} \quad \forall f \in D(L),\ R_\lambda(\lambda - L)f = f\ .$$

Finally, we extend proposition 2.6 to stopping times. This proposition is used many times as well in the context of this dissertation.

Proposition 2.16 (Strong Markov property)**.** *If Z is a $\mathscr{F}_\infty$-measurable and positive (or bounded) random variable and T is a stopping time, for any initial measure ν,*

$$\mathbb{E}_\nu[\mathbf{1}_{\{T<\infty\}} Z \circ \theta_T \mid \mathscr{F}_T] = \mathbf{1}_{\{T<\infty\}} \mathbb{E}_{Y_T}[Z] \quad \mathbb{P}_\nu - \text{a.s.}$$

Chapter 3

General theory on quasi-stationary distributions

We first give the setting of quasi-stationary distributions and their formal definitions, as well as that of other key concepts such as quasi-limiting distributions and Yaglom limits.

3.1 Assumptions in a general setting

This first section develops the necessary theoretical setting and is mainly based on the book by Collet, Martinez and San Martin [9].

We consider a Polish space X, called *state space*, endowed with its Borel σ-field $\mathscr{B}(\mathsf{X})$. In further applications X is generally discrete, which satisfies this setting as any discrete space is a Polish space (Cf [18]).

Let $\Omega \subset \mathsf{X}^{\mathbb{R}_+}$ be the set of càdlàg trajectories on X, indexed by $\mathbb{R}_+$ which is the "time index".

Let $(P_t)_{t\in\mathbb{R}_+}$ be a Feller semigroup on X. Then, using theorem 2.10, we define Y to be the càdlàg Feller process with semigroup $(P_t)_t$. Thus, Y is a canonical process with $\forall s \in \mathbb{R}_+, Y_s(\omega) = \omega_s$, which can be seen as a family of projections on X.

We denote by $(\mathscr{F}_t)_{t\geqslant 0}$ the right-continuous filtration such that Y is adapted to $(\mathscr{F}_t)$ and $\mathscr{F}_0$ contains the negligible of X. We set $\mathscr{F} := \mathscr{F}_\infty$ which then contains the σ-fields generated by all the $X_t, t \in \mathbb{R}_+$.

Following theorem 2.4 and the subsequent remark, we denote by $\mathbb{P}_x$ the distribution on $(\Omega, \mathscr{F})$ with the initial conditon $Y_0 = x \in \mathsf{X}$. More generally, we denote by $\mathbb{P}_\nu$ the distribution on $(\Omega, \mathscr{F})$ with $Y_0 \sim \nu$.

Now that we have set the notation, we give the two main assumptions necessary to have quasi-stationary distributions.

Definition 3.1. Let N be a Markov Kernel on $(\mathsf{X}, \mathscr{B}(\mathsf{X}))$. A non-empty set $B \in \mathscr{B}(\mathsf{X})$ is called *absorbing* if $\forall x \in B,\ N(x,B) = 1$.

Assumption 1. There is a non-empty set $\mathsf{X}_{\text{ab}} \in \mathscr{B}(\mathsf{X})$ of values called *absorbing states*, such that $\mathsf{X}_{\text{ab}} \neq \mathsf{X}$ and X_{ab} is an absorbing set for any $P_t, t \geqslant 0$.

Remark. In general, when we say that $x \in \mathsf{X}$ is an absorbing state it does not mean that $\{0\}$ is an absorbing set but simply that $x \in \mathsf{X}_{\text{ab}}$, and there is no equivalence *a priori* between the two.

We denote the complement of X_{ab} by X_{tr}. The subscript comes from the fact that we will later restrict this setting to countable spaces so that X_{tr} will include only states which are called transient. In some practical applications given here, X_{tr} will even be exactly the set of transient states for the related Markov chain.

We allow ourselves to implicitely extend distributions on X_{tr} to use them on X. This is the case in particular with quasi-stationary distributions, which are distributions on X_{tr}. Since $\mathsf{X}_{\text{tr}} \in \mathscr{B}(\mathsf{X})$ this extension is straightforward: for any $\alpha \in \mathscr{P}(\mathsf{X}_{\text{tr}})$ and $A \in \mathscr{B}(\mathsf{X})$ we set $\alpha(A) := \alpha(A \cap \mathsf{X}_{\text{tr}})$.

We denote by T the hitting time of X_{ab}, also called *killing time*. The state of X_{ab} which is reached at T is also sometimes called *exit state*. We denote by Y^T the stopped process $(Y_{t\wedge T})_{t\geqslant 0}$.

Assumption 2. There is sure killing at X_{ab}:

$$\forall x \in \mathsf{X}_{\text{tr}}, \quad \mathbb{P}_x(T < \infty) = 1 \tag{3.1}$$

which is equivalent to $\forall \rho \in \mathscr{P}(\mathsf{X}_{\text{tr}}), \mathbb{P}_\rho(T < \infty) = 1$ where we have $\mathbb{P}_\rho = \int_{\mathsf{X}_{\text{tr}}} \mathbb{P}_x \, \mathrm{d}\rho(x)$.

One implication of (3.1) is

$$\forall \rho \in \mathscr{P}(\mathsf{X}_{\text{tr}}),\ \exists t_\rho \in \mathbb{R}_+,\ \forall t > t_\rho,\ \mathbb{P}_\rho(T < t) > 0\ . \tag{3.2}$$

In particular, there can be no stationary distribution for Y on X_{tr}, that is a distribution $\rho \in \mathscr{P}(\mathsf{X}_{\text{tr}})$ such that $\forall B \in \mathscr{B}(\mathsf{X}_{\text{tr}}), \forall t \geqslant 0, \mathbb{P}_\rho(Y_t \in B) = \rho(B)$. Indeed, we would have $\forall t \geqslant 0, \mathbb{P}_\rho(Y_t \in \mathsf{X}_{\text{tr}}) = \rho(\mathsf{X}_{\text{tr}}) = 1$, which contradicts (3.2).

3.2 Definitions

Using the setting of the previous section, we can now define quasi-stationarity and some other crucial notions in this topic. In the 1990's the definition of a quasi-stationary distribution was still unclear and depended on the author. The definitions of quasi-limiting distributions and Yaglom limits were also changing and sometimes assimilated.

The definition of a QSD given by Ferrari et al. [12] in 1995, in a more restrictive setting, was the basis of the following works. The definitions and results given here will be based on [9], [17] and [19], which are all relatively recent. Notably, it turns out that few people wrote about quasi-stationary distributions in a setting as general as this one. Most of the papers in this field focus on more applied results.

Definition 3.2. A probability measure ν on X_{tr} is said to be a *quasi-stationary distribution* (QSD) for Y if

$$\forall B \in \mathscr{B}(\mathsf{X}_{\text{tr}}),\ \forall t \geqslant 0, \quad \mathbb{P}_\nu(Y_t \in B \mid T > t) = \nu(B)\ .$$

Another name which is sometimes used instead of QSD is stationary conditional distribution. The definition implies that if ν is a QSD we have

$$\forall B \in \mathscr{B}(\mathsf{X}_{\text{tr}}),\ \forall t \geqslant 0, \quad \mathbb{P}_\nu(Y_t \in B, T > t) = \nu(B)\mathbb{P}_\nu(T > t)\ .$$

Then, since $B \cap \mathsf{X}_{\text{ab}} = \emptyset$, $\{Y_t \in B\}$ is included in $\{T > t\}$ and we can simplify the last equation as

$$\forall B \in \mathscr{B}(\mathsf{X}_{\text{tr}}),\ \forall t \geqslant 0, \quad \mathbb{P}_\nu(Y_t \in B) = \nu(B)\mathbb{P}_\nu(T > t)\ . \tag{3.3}$$

Definition 3.3. A probability measure ν on X_{tr} is a *quasi-limiting distribution* (QLD) for Y if there exists a probability measure α on X_{tr} such that

$$\forall B \in \mathscr{B}(\mathsf{X}_{\text{tr}}), \quad \lim_{t\to\infty} \mathbb{P}_\alpha(Y_t \in B \mid T > t) = \nu(B) .$$

Remark. Quasi-limiting distributions are also sometimes named limiting conditional distributions in the literature.

If ν is a QLD for some initial distribution α, we say that α is in the *domain of attraction* of ν. A problem is then to identify those domains of attraction. The problem is more simple in the presence of a Yaglom limit, which is stronger.

Definition 3.4. We say that Y has a *Yaglom limit* if there exists a probability measure ν on X_{tr} such that

$$\forall x \in \mathsf{X}_{\text{tr}}, \quad \forall B \in \mathscr{B}(\mathsf{X}_{\text{tr}}), \quad \lim_{t\to\infty} \mathbb{P}_x(Y_t \in B \mid T > t) = \nu(B) .$$

So, clearly, a Yaglom limit is a QLD since it is the QLD of any initial Dirac distribution δ_x with $x \in \mathsf{X}_{\text{tr}}$. From its definition, the Yaglom limit is necessarily unique when it exists, whereas there is no limit in general to the number of quasi-limiting distributions, so that not every QLD is a Yaglom limit.

We can actually prove a stronger relation between the three types of distributions that we introduced here, given schematically by

$$\text{Yaglom limit} \implies \text{QSD} \iff \text{QLD}.$$

This equivalence is presented in the following proposition, with a proof first given by Meleard and Villimonais [17] in the general setting, and earlier in the case of countable spaces by Vere-Jones [29].

Proposition 3.5. *Let $\nu \in \mathscr{P}(\mathsf{X}_{\text{tr}})$. Then, ν is a QLD for Y if and only if it is a QSD for Y.*

First, we state a property of quasi-limiting distributions, with a proof adapted from [14].

Lemma 3.6. *Let ν be a QLD for Y for some initial distribution $\alpha \in \mathscr{P}(\mathsf{X}_{\text{tr}})$. Then, we have*

$$\forall f \in L^\infty, \ \nu(f) = \lim_{t\to\infty} \mathbb{E}_\alpha[f(Y_t) \mid T > t] .$$

Proof. Let $f \in L^\infty$ and $\varepsilon > 0$. Then, there exist $N \in \mathbb{N}$ and $y_0, y_1, \ldots, y_N \in \mathbb{R}$ such that

$$y_0 \leqslant -\|f\|_\infty < y_1 < \ldots < y_{N-1} < \|f\|_\infty < y_N$$

and $\forall i \in \{1,\ldots,N\}, |y_i - y_{i-1}| < \varepsilon$. For every $i \in \{1,\ldots,N\}$, let $E_i = f^{-1}([y_{i-1}, y_i[)$.

We define in $\mathscr{P}(\mathsf{X}_{\text{tr}})$ the distributions $\nu_t : A \mapsto \mathbb{P}_\alpha(Y_t \in A \mid T > t)$ for all $t \in \mathbb{R}$ where $A \in \mathscr{B}(\mathsf{X}_{\text{tr}})$. Thus, (ν_t) converges to ν in the sense that $\forall A \in \mathscr{B}(\mathsf{X}_{\text{tr}}), \lim_{t\to\infty} \nu_t(A) = \nu(A)$. Therefore,

$$\limsup_{t\to\infty} \nu_t(f) \ \leqslant\ \limsup_{t\to\infty} \sum_{i=1}^{N} \nu_t(E_i)\cdot y_i \ =\ \sum_{i=1}^{N} \nu(E_i)\cdot y_i \ \leqslant\ \varepsilon + \nu(f)$$

When we let $\varepsilon \downarrow 0$ we get $\limsup_{t\to\infty} \nu_t(f) \leqslant \nu(f)$. If we consider $-f$ we obtain similarly the reverse inequality $\liminf_{t\to\infty} \nu_t(f) \geqslant \nu(f)$. We conclude that

$$\nu(f) = \lim_{t\to\infty} \nu_t(f) = \lim_{t\to\infty} \int_{\mathsf{X}_{\mathrm{tr}}} f(x)\mathbb{P}_\alpha(Y_t \in \mathrm{d}x \mid T > t) = \lim_{t\to\infty} \mathbb{E}_\alpha[f(Y_t) \mid T > t]$$

with a change of variable in the integral to find the expectation. ■

Proof of proposition 3.5. The second implication is direct: if ν is a QSD then it is a QLD for itself.

Assume now that ν is a QLD for Y for an initial distribution $\alpha \in \mathscr{P}(\mathsf{X}_{\mathrm{tr}})$. We use lemma 3.6 with the function $f : x \mapsto \mathbb{P}_x(T > s)$ for some $s \in \mathbb{R}_+$. We get

$$\mathbb{P}_\nu(T > s) = \lim_{t\to\infty} \mathbb{E}_\alpha[\mathbb{P}_{Y_t}(T > s) \mid T > t] = \lim_{t\to\infty} \frac{\mathbb{E}_\alpha[\mathbf{1}_{\{T>t\}}\mathbb{P}_{Y_t}(T > s)]}{\mathbb{P}_\alpha(T > t)}$$

Moreover, using the Markov property (proposition 2.6) and tower property,

$$\begin{aligned}
\mathbb{E}_\alpha[\mathbf{1}_{\{T>t\}}\mathbb{P}_{Y_t}(T > s)] &= \mathbb{E}_\alpha[\mathbf{1}_{\{T>t\}}\mathbb{E}_{Y_t}[\mathbf{1}_{\{T>s\}}]] \\
&= \mathbb{E}_\alpha[\mathbf{1}_{\{T>t\}}\mathbb{E}_\alpha[\mathbf{1}_{\{T>s\}} \circ \theta_t \mid \mathscr{F}_t]] \\
&= \mathbb{E}_\alpha[\mathbb{E}_\alpha[\mathbf{1}_{\{T>t\}}\mathbf{1}_{\{T>s+t\}} \mid \mathscr{F}_t]] \\
&= \mathbb{E}_\alpha[\mathbf{1}_{\{T>s+t\}}] \\
&= \mathbb{P}_\alpha(T > s + t)\ .
\end{aligned}$$

Thus, we have $\mathbb{P}_\nu(T > s) = \lim_{t\to\infty} \dfrac{\mathbb{P}_\alpha(T > s+t)}{\mathbb{P}_\alpha(T > t)}$.

Now, we take $f : x \mapsto \mathbb{P}_x(Y_s \in A, T > s)$ with $A \subset \mathsf{X}_{\mathrm{tr}}$. With the same logic as before we get

$$\begin{aligned}
\mathbb{P}_\nu(Y_s \in A, T > s) &= \lim_{t\to\infty} \frac{\mathbb{P}_\alpha(Y_{t+s} \in A, T > s + t)}{\mathbb{P}_\alpha(T > t)} \\
&= \lim_{t\to\infty} \frac{\mathbb{P}_\alpha(T > s+t)}{\mathbb{P}_\alpha(T > t)} \cdot \mathbb{P}_\alpha(Y_{t+s} \in A \mid T > s + t) \\
&= \mathbb{P}_\nu(T > s) \cdot \lim_{t\to\infty} \mathbb{P}_\alpha(Y_{t+s} \in A \mid T > s + t)
\end{aligned}$$

using the finiteness of the limits and the previous result. Then, by the definition of ν as the QLD of α we get

$$\mathbb{P}_\nu(Y_s \in A \mid T > s) = \nu(A)$$

which proves that ν is a QSD. ■

3.3 Exponential killing and exit state

A crucial observation in the case of quasi-stationary distributions is that they have a constant rate of survival and death. This implies that the killing time is exponentially distributed as stated in the next theorem. This property proves particularly useful in practical applications to answer questions about survival time. We review here some major results on that point, mostly from [9], [16] and [17]. In what follows we use the notation of section 3.1, where in particular T denotes the killing time of the process.

Theorem 3.7. *Let $(Y_t)_{t\geqslant 0}$ be a Feller process satisfying assumptions 1 and 2. If ν is a QSD, then there exists $\vartheta \in]0,+\infty[$ such that*

$$\forall t \geqslant 0, \quad \mathbb{P}_\nu(T > t) = e^{-\vartheta t}$$

meaning that, with initial distribution ν, T is exponentially distributed with parameter ϑ.

Proof. From equation (3.3) we have that, for any measurable function $g\colon \mathsf{X} \to \mathbb{R}$ positive or bounded we have

$$\forall t \geqslant 0, \quad \mathbb{E}_\nu[\mathbf{1}_{\{T>t\}} g(Y_t)] = \nu(g)\mathbb{P}_\nu(T > t)\ .$$

We take $g\colon x \mapsto \mathbb{P}_x(T > s)$, so that $\nu(g) = \mathbb{P}_\nu(T > s)$ by proposition 2.5, and then

$$\forall t \geqslant 0, \quad \mathbb{E}_\nu[\mathbf{1}_{\{T>t\}} \mathbb{P}_{Y_t}(T > s)] = \mathbb{P}_\nu(T > s)\mathbb{P}_\nu(T > t)\ .$$

Using the Markov property (proposition 2.6) we get, for every $t, s \in \mathbb{R}_+$,

$$\begin{aligned}\mathbb{P}_\nu(T > t+s) = \mathbb{E}_\nu[\mathbf{1}_{\{T>t+s\}}] &= \mathbb{E}_\nu[\mathbf{1}_{\{T>t\}} \cdot \mathbb{E}[\mathbf{1}_{\{T>s\}} \circ \theta_t \mid \mathscr{F}_t]] \\ &= \mathbb{E}_\nu[\mathbf{1}_{\{T>t\}} \cdot \mathbb{E}_{Y_t}[\mathbf{1}_{\{T>s\}}]] \\ &= \mathbb{P}_\nu(T > t)\mathbb{P}_\nu(T > s)\end{aligned}$$

which implies that there exists $\vartheta \in \mathbb{R}_+$ such that T is exponentially distributed with parameter ϑ. Moreover, using assumption 2 we get $\vartheta > 0$. ■

Notation. Let ν be a QSD, then $\vartheta(\nu)$ denotes the parameter given by theorem 3.7.

In particular, we can rewrite equation (3.3) so that $\nu \in \mathscr{P}(\mathsf{X}_{\mathrm{tr}})$ is a QSD if and only if there exists $\vartheta > 0$ such that

$$\forall B \in \mathscr{B}(\mathsf{X}_{\mathrm{tr}}),\ \forall t \geqslant 0, \quad \mathbb{P}_\nu(Y_t \in B) = \nu(B)e^{-\vartheta t}\ . \tag{3.4}$$

in which case $\vartheta = \vartheta(\nu)$.

Proposition 3.8. *If ν is a QSD then, for any $\gamma \in]0, \vartheta(\nu)[$, $\mathbb{E}_\nu[e^{\gamma T}] < \infty$ and there exists $x \in \mathsf{X}_{\mathrm{tr}}$ such that $\mathbb{E}_x[e^{\gamma T}] < \infty$.*

Proof. Let $\gamma \in]0, \vartheta(\nu)[$. Using theorem 3.7 we can calculate the first moment of $e^{\gamma T}$:

$$\mathbb{E}_\nu[e^{\gamma T}] = \int_0^\infty e^{\gamma t}\vartheta(\nu)e^{-\vartheta(\nu)t}\,\mathrm{d}t = \frac{\vartheta(\nu)}{\vartheta(\nu) - \gamma} < +\infty\ .$$

Moreover, we have $\mathbb{E}_\nu[e^{\gamma T}] = \int_{\mathsf{X}_{\mathrm{tr}}} \mathbb{E}_x[e^{\gamma T}]\nu(\mathrm{d}x)$ (see proposition 2.5) and the finiteness of the integral proves the second part of the proposition. ■

Remark. So far there is no equivalence between exponential killing and the existence of a QSD. This result requires more assumptions and we will give one in particular in proposition 4.15.

Then, we prove the independence between the exit time T and the exit state Y_T. This property is not actually useful for the practical models presented in chapters 4 and 5 where there is only one possible exit state. But it is relevant for the study of the models presented in chapter 6 where the processes are in several dimensions and admit multiple absorbing states. The original proof given in [9] is only true for the stopped process, so we give a new proof below, which is more general and shorter.

Theorem 3.9. *Let ν be a QSD for the process Y satisfying assumptions 1 and 2. Then T and Y_T are independent variables under $\mathbb{P}_\nu$.*

Proof. Let $B \in \mathscr{B}(\mathsf{X}_{\mathrm{ab}})$ and $a > 0$. We notice that $Y_T \circ \theta_a$ is the same as Y_T when $T > a$, with θ_a the shift operator. Thus, using the strong Markov property 2.16 we get

$$\begin{aligned}\mathbb{P}_\nu(Y_T \in B \mid T > a) &= \mathbb{P}_\nu(Y_T \circ \theta_a \in B \mid Y_a \in \mathsf{X}_{\mathrm{tr}}) \\ &= \mathbb{E}_\nu[\mathbb{E}_\nu[\mathbf{1}_B(Y_T) \circ \theta_a \mid Y_a = x \in \mathsf{X}_{\mathrm{tr}}] \mid T > a] \\ &= \mathbb{E}_\nu[\mathbb{E}_{Y_a}[\mathbf{1}_B(Y_T)] \mid T > a] \,.\end{aligned}$$

Since ν is a QSD, Y_a has distribution ν conditioned on $T > a$, by proposition 2.5 we have

$$\mathbb{P}_\nu(Y_T \in B \mid T > a) = \int_{\mathsf{X}} \nu(\mathrm{d}x)\mathbb{E}_x[\mathbf{1}_B(Y_T)] = \mathbb{P}_\nu(Y_T \in B)$$

which proves the independence. ∎

3.4 Characterization by the semigroup

There are several conditions related to the semigroup $(P_t)_{t\geqslant 0}$ and to P_1 in particular which enable us to conclude that there exists a QSD, while staying in a rather general setting. The results exposed here come from [9].

Let $\alpha \in \mathscr{M}(\mathsf{X}_{\mathrm{tr}})$. For any $t \geqslant 0$, the mapping $B \mapsto \alpha(P_t \mathbf{1}_B)$ on $\mathscr{B}(\mathsf{X}_{\mathrm{tr}})$ defines a measure on X_{tr}. We denote this measure by $P_t^\dagger \alpha$. By linearization, this action can be defined for all finite signed measures on X_{tr}, so that we have

$$\forall f \in L^\infty, \quad (P_t^\dagger \alpha)(f) := \alpha(P_t f)\,. \tag{3.5}$$

We notice that $(P_t^\dagger)_t$ is also a semigroup, acting in the space of measures, i.e. $\forall t, s \geqslant 0, P_{t+s}^\dagger = P_t^\dagger \circ P_s^\dagger$. The verification is straightforward: for any $f \in L^\infty$ we have

$$P_t^\dagger((P_s^\dagger \alpha)f) = (P_s^\dagger \alpha)(P_t^\dagger f) = \alpha(P_s P_t f) = \alpha(P_{t+s} f) = (P_{t+s}^\dagger \alpha)f \,.$$

Then, the result of equation (3.4) can be written as: $\nu \in \mathscr{P}(\mathsf{X}_{\mathrm{tr}})$ is a QSD if and only if there exists $\vartheta(\nu) > 0$ such that

$$\forall t \geqslant 0, \quad P_t^\dagger \nu = e^{-\vartheta(\nu)t}\,\nu\,. \tag{3.6}$$

This is called the eigenmeasure equation, which is consistent with the discrete case studied in chapters 4 and 5 where the QSD are given by eigenvectors of the transition function. As stated by the following lemma, using $P_1^\dagger$ we can exploit a weaker condition for the existence of a quasi-stationary distribution. We will use more practical results in applied cases, but this lemma is interesting in its similarity to those stronger results, as it is essentially a more abstract phrasing of it.

Lemma 3.10. *Let Y satisfy assumption 1 and 2. Let $\tilde{\nu} \in \mathscr{P}(\mathsf{X}_{\mathrm{tr}})$ and $\beta > 0$ be such that $P_1^\dagger \tilde{\nu} = \beta \tilde{\nu}$. Then $\beta < 1$ and there exists a QSD ν whose exponential rate of survival is $\vartheta = -\log \beta > 0$.*

Proof. From equation (3.2) we know that there exists $n \in \mathbb{N}$ such that $\mathbb{P}_{\tilde{\nu}}(T > n) < 1$. Therefore, $\beta^n = (P_n^\dagger \tilde{\nu})\mathbf{1}_{\{\mathsf{X}_{\mathrm{tr}}\}} = \mathbb{P}_{\tilde{\nu}}(T > n) < 1$ which proves that $\beta < 1$. Thus, taking $\vartheta := -\log \beta$ we have $\vartheta > 0$.

We know that $P_t \mathbf{1}_B$ is measurable for any $t \geqslant 0$ and $B \in \mathscr{B}(\mathsf{X}_{\mathrm{tr}})$. Thus, we can define

$$\nu \colon B \mapsto \int_0^1 e^{\vartheta s}\, \tilde{\nu}(P_s \mathbf{1}_B)\, \mathrm{d}s = \int_0^1 e^{\vartheta s} (P_s^\dagger \tilde{\nu})(B)\, \mathrm{d}s \ .$$

By linearity and monotonicity, it comes that ν is a finite measure on X_{tr}. Let $t \in]0,1]$. We have,

$$\begin{aligned}
P_t^\dagger \nu(B) = \nu(P_t \mathbf{1}_B) &= \int_0^1 e^{\vartheta s} P_s^\dagger \tilde{\nu}(P_t \mathbf{1}_B)\, \mathrm{d}s \\
&= \int_0^1 e^{\vartheta s} P_{t+s}^\dagger \tilde{\nu}(B)\, \mathrm{d}s = \int_0^{1-t} e^{\vartheta s} P_{t+s}^\dagger \tilde{\nu}(B)\, \mathrm{d}s + \int_{1-t}^1 e^{\vartheta s} P_{t+s}^\dagger \tilde{\nu}(B)\, \mathrm{d}s \\
&= \int_t^1 e^{\vartheta(u-t)} P_u^\dagger \tilde{\nu}(B)\, \mathrm{d}u + \int_1^{1+t} e^{\vartheta(u-t)} P_u^\dagger \tilde{\nu}(B)\, \mathrm{d}u \\
&= e^{-\vartheta t} \int_t^1 e^{\vartheta u} P_u^\dagger \tilde{\nu}(B)\, \mathrm{d}u + e^{-\vartheta t} \int_0^t e^{\vartheta u} e^{\vartheta} (P_u^\dagger P_1^\dagger \tilde{\nu})(B)\, \mathrm{d}u \\
&= e^{-\vartheta t} \nu(B) \ .
\end{aligned}$$

Then, let $t > 1$. We can choose $\tau \in]0,1]$ and $k \in \mathbb{N}$ such that $k\tau = t$ and using the previous result we get

$$P_t^\dagger \nu = \underbrace{P_\tau^\dagger \cdots P_\tau^\dagger}_{k \text{ times}} \nu = (e^{-\tau\vartheta})^k \nu = e^{-\vartheta t} \nu$$

which proves that ν is a QSD using equation (3.6), after that we normalize it. ∎

There are some other conditions that can guarantee the existence of a quasi-stationary distribution. For example, if X_{tr} is a compact Hausdorff set and P_1 preserves the set of continuous functions. A couple of other interesting results of that sort are given in [9]. Nervertheless, we will be interested in more practical cases from now on.

Chapter 4

Quasi-stationary distributions on countable spaces

We use the notations defined in sections 2 and 3 but we suppose now that X is a discrete space, not necessarily finite.

Remark. We already assumed that X is separable as part of being a Polish space. Since a discrete space is separable if and only if it is countable, both terms, discrete and countable, are equivalent here.

4.1 Introduction to jump processes

Let $(P_t)_{t\geqslant 0}$ be a Feller semigroup on X. As in 3.1, we define Y as the càdlàg canonical process of $(P_t)_t$. In the discrete setting, and because of its regularity, Y is a *jump process*, meaning that it has all its sample paths constant except for isolated jumps. The study of its specificities as a jump process is the objective of this section.

The results given here are based mostly on [15] and [24]. They are general properties and we don't use assumptions 1 or 2 in this section.

The literature on QSDs with a discrete state space uses mathematical objects specific to this discrete setting but seldomly explains how they sufficiently define the Markov process which is studied and relate to quasi-stationary distributions. The purpose of this section is to explain how we go consistently from the notions used in the general setting of chapter 3 to those that we will use later in the applications.

As we will see, the regularity of Y allows us to consider the existence of quasi-stationary distributions using various results. Then, in some applications, the burden becomes to prove that the characterization of the process is that of a Feller process.

Since the sample paths of Y are càdlàg, it comes that there exists a sequence $(T_i)_{i\in\mathbb{N}}$ of random variables such that, for any $\omega \in \Omega$

$$T_0(\omega) = 0 < T_1(\omega) \leqslant T_2(\omega) \leqslant T_3(\omega) \leqslant \ldots \leqslant \infty$$

with $\forall t \in [0, T_1(\omega)[, Y_t(\omega) = Y_0(\omega)$ and for any $i \geqslant 1$, provided that $T_i(\omega) < \infty$, $Y_{T_i}(\omega) \neq Y_{T_{i-1}}(\omega)$ and $\forall t \in [T_i(\omega), T_{i+1}(\omega)[$, $Y_t(\omega) = Y_{T_i}(\omega)$. We can verify that they are stopping times as follows:

$$\{T_i < t\} = \bigcup_{q\in[0,t[\cap\mathbb{Q}} \left(\bigcap_{0\leqslant j\leqslant i-1} \{Y_q \neq Y_{T_j}\} \right).$$

We first state some general results on jump state processes.

Proposition 4.1. *Let Y be a càdlàg Feller process and $x \in \mathsf{X}$. There exists $q(x) \in \mathbb{R}_+$ called the* jump rate *of x such that the random variable T_1 is exponentially distributed with parameter $q(x)$ under $\mathbb{P}_x$. Moreover, if $q(x) > 0$, T_1 and Y_{T_1} are independent under $\mathbb{P}_x$.*

This theorems translates as

$$\forall t \in \mathbb{R}_+, \quad \mathbb{P}_x(T_1 > t) = e^{-q(x)t} \tag{4.1}$$

Proof. Let $s, t \in \mathbb{R}_+$. Let $Z = \mathbf{1}_{\{\forall r\in[0,t], Y_r = Y_0\}}$, i.e. $Z(\omega) = 1$ if and only if $\forall r \in [0,t], \omega_t = \omega_0$. Using tower property we have,

$$\begin{aligned}\mathbb{P}_x(T_1 > s+t) &= \mathbb{E}_x[\mathbf{1}_{\{T_1>s\}} \times Z\circ\theta_s] \\ &= \mathbb{E}_x[\mathbf{1}_{\{T_1>s\}} \times \mathbb{E}_x[Z\circ\theta_s \mid \mathscr{F}_s]]\end{aligned}$$

Then, we can use the Markov property 2.6, and we get

$$\begin{aligned}\mathbb{P}_x(T_1 > s+t) &= \mathbb{E}_x[\mathbf{1}_{\{T_1>s\}} \times \mathbb{E}_{Y_s}[Z]] \\ &= \mathbb{E}_x[\mathbf{1}_{\{T_1>s\}} \times \mathbb{P}_x(T_1 > t)] \\ &= \mathbb{P}_x(T_1 > s)\cdot\mathbb{P}_x(T_1 > t)\end{aligned}$$

using that $Y_s = x$ whenever $\mathbf{1}_{\{T_1>s\}}$ is non-zero. This implies that T_1 is exponentially distributed.

Assume now $q(x) > 0$ so that $T_1 < \infty$ $\mathbb{P}_x-$a.s.. Let $t \geqslant 0$ and $y \in E$. Let F be the random variable on Ω such that $F(\omega) = 1$ if $T_1 < \infty$ and $Y_{T_1}(\omega) = y$, and 0 otherwise. With arguments similar to the first part of the proof we get

$$\begin{aligned}\mathbb{P}_x(T_1 > t, Y_{T_1} = y) &= \mathbb{E}_x[\mathbf{1}_{\{T_1>t\}} \times F\circ\theta_t] \\ &= \mathbb{E}_x[\mathbf{1}_{\{T_1>t\}} \times \mathbb{E}_{Y_t}[F]] \\ &= \mathbb{E}_x[\mathbf{1}_{\{T_1>t\}} \times \mathbb{P}_x(Y_{T_1} = y)] \\ &= \mathbb{P}_x(T_1 > t)\cdot\mathbb{P}_x(Y_{T_1} = y)\end{aligned}$$

so we can conclude that T_1 and Y_{T_1} are independent under $\mathbb{P}_x$. ■

Remark. If assumption 2 of sure killing holds, we can observe that $x \in \mathsf{X}_{\mathrm{tr}}$ implies $q(x) > 0$. In that case, the only states such that $q(x) = 0$ are in X_{ab}.

In the literature, a state x is sometimes qualified as *permanent* if $q(x) = 0$, *stable* if $0 < q(x) < \infty$ and instantaneous if $q(x) = \infty$. However, the latter is not compatible with the assumption of right continuity, and it should not be confused with the notion of instantaneous transition rate that we will define.

The transition function (P_t) is called stable itself when it has no instantaneous jump rate, which is the case here.

For every $x, y \in \mathsf{X}$, we set the *conditional transitions*,

$$\Pi(x,y) := \begin{cases} \mathbb{P}_x(Y_{T_1} = y) & \text{if} \quad q(x) > 0 \\ 0 & \text{if} \quad q(x) = 0 \end{cases} \tag{4.2}$$

Note that $\forall x \in \mathsf{X}, \Pi(x,x) = 0$.

Definition 4.2. For every $x, y \in \mathsf{X}$, we define the *(instantaneous) transition rate* from x to y

$$q(x,y) := \begin{cases} q(x)\Pi(x,y) & \text{if} \quad y \neq x \\ -q(x) & \text{if} \quad y = x \end{cases} .$$

The matrix $Q := (q(x,y))_{x,y\in\mathsf{X}}$ is called the *transition rate matrix*.

In the literature we normally start from those transition rates to then identify the quasi-stationary distributions when they exist. Moreover, it is easy to show that, in a discrete space, when the jump rates are bounded the process is automatically Feller. In particular, that is the case when X is finite. But in general there is no bound so it must be proved that the process is Feller.

Remark. The matrix Q is said to be *conservative* because its rows sum to zero:

$$\forall x \in \mathsf{X}_{\text{tr}}, \sum_{y\in\mathsf{X}} q(x,y) = q(x)\left(-1 + \sum_{y\neq x} \Pi(x,y)\right) = q(x)(-1 + \Pi(x,\mathsf{X})) = 0 .$$

Proposition 4.4 shows how this relates to the transition function through its generator (see 2.11). But before that we need an additional assumption for some crucial results to hold.

For any $x \in \mathsf{X}$ we denote by $\mathsf{A}(x)$ the set of states accessible in one jump from x, i.e.

$$\mathsf{A}(x) = \{y \in \mathsf{X} \mid q(x,y) > 0\} \tag{4.3}$$

Assumption 3. The càdlàg Feller process Y on X with transition rate matrix Q satisfies

$$\forall x \in \mathsf{X}, \sup_{y\in\mathsf{A}(x)} q(y) < \infty .$$

We can see that this assumption is clearly verified when $\mathsf{A}(x)$ is finite for all x, which will be the case of all the models that we will study. As a subcase of this, the assumption is satisfied when X itself is finite.

Lemma 4.3. *Let Y satisfy assumption 3, and let $x \in \mathsf{X}$ such that $q(x) > 0$. Then,*

$$\mathbb{P}_x(T_2 \leqslant t) \underset{0}{=} O(t^2) .$$

Proof. First, we use the strong Markov property 2.16 at T_1:

$$\begin{aligned} \mathbb{P}_x(T_2 \leqslant t) \leqslant \mathbb{P}_x(T_1 \leqslant t, T_2 \leqslant T_1 + t) &= \mathbb{E}_x[\mathbf{1}_{\{T_1\leqslant t\}} \cdot \mathbf{1}_{\{T_2\leqslant T_1+t\}}] \\ &\leqslant \mathbb{E}_x[\mathbf{1}_{\{T_1\leqslant t\}} \cdot \mathbb{E}_x[\mathbf{1}_{\{T_1\leqslant t\}} \circ \theta_{T_1} \mid \mathscr{F}_{T_1}]] \\ &\leqslant \mathbb{E}_x[\mathbf{1}_{\{T_1\leqslant t\}} \cdot \mathbb{P}_{Y_{T_1}}(T_1 \leqslant t)] . \end{aligned}$$

Then, we bound using proposition 4.1,

$$\mathbb{P}_{Y_{T_1}}(T_1 \leqslant t) \leqslant \sup_{y\in\mathsf{A}(x)} \mathbb{P}_y(T_1 \leqslant t) \leqslant t \sup_{y\in\mathsf{A}(x)} q(y)$$

and similarly $\mathbb{P}_x(T_1 \leqslant t) \leqslant q(x)t$, which gives $\mathbb{P}_x(T_2 \leqslant t) \leqslant t^2 q(x) \sup_{y\in\mathsf{A}(x)} q(y) = O(t^2)$. ■

Proposition 4.4. *Let Y satisfy assumption 3, and let L denote the generator of $(P_t)_{t\geqslant 0}$, the transition function of Y. Then $D(L) = C_0(\mathsf{X}) \subset L^\infty$, where $D(L)$ is the domain of L and $C_0(\mathsf{X})$ is the set of continuous functions on X that tend to 0 at infinity, as defined in section 2.3.*

Moreover, for every $\varphi \in D(L)$ and every $x \in \mathsf{X}$,

$$L\varphi(x) = q(x)\sum_{y\neq x}\Pi(x,y)(\varphi(y)-\varphi(x)) = \sum_{y\in\mathsf{A}(x)} q(x,y)\varphi(y)$$

and we can extend L to L^∞ so that this formula still holds and there is convergence in L^∞.

Proof. Let $\varphi \in L^\infty$ and $x \in \mathsf{X}$. If $q(x) = 0$, then from 2.2 and the definition 3.1 of absorbing states we get $\forall t \in \mathbb{R}_+, P_t\varphi(x) = \varphi(x)$ and so

$$L\varphi(x) = \lim_{t\downarrow 0}\frac{P_t\varphi(x)-\varphi(x)}{t} = 0\,.$$

Suppose now that $q(x) > 0$. We have

$$\begin{aligned}P_t\varphi(x) &= \mathbb{E}_x[\varphi(Y_t)]\\ &= \mathbb{E}_x[\varphi(Y_t)\mathbf{1}_{\{T_1>t\}}] + \mathbb{E}_x[\varphi(Y_{T_1})\mathbf{1}_{\{T_1\leqslant t\}}] + \mathbb{E}_x[(\varphi(Y_t)-\varphi(Y_{T_1}))\mathbf{1}_{\{T_2\leqslant t\}}]\end{aligned}$$

Since φ is bounded we have

$$\begin{aligned}\left|\mathbb{E}_x\left[(\varphi(Y_t)-\varphi(Y_{T_1}))\mathbf{1}_{\{T_2\leqslant t\}}\right]\right| &\leqslant \mathbb{E}_x\left[\left|\varphi(Y_t)-\varphi(Y_{T_1})\right|\mathbf{1}_{\{T_2\leqslant t\}}\right]\\ &\leqslant 2\,\|\varphi\|_\infty\cdot\mathbb{P}_x(T_2\leqslant t)\\ &= O(t^2)\end{aligned}$$

so that, using lemma 4.3 and the independence of T_1 and Y_{T_1} we get

$$P_t\varphi(x) = \varphi(x)e^{-q(x)t} + (1-e^{-q(x)t})\sum_{y\in\mathsf{A}(x)}\Pi(x,y)\varphi(y) + O(t^2)\,.$$

We complete the proof with

$$L\varphi(x) = \lim_{t\to 0}\frac{P_t\varphi(x)-\varphi(x)}{t} = -q(x)\varphi(x) + q(x)\sum_{y\in\mathsf{A}(x)}\Pi(x,y)\varphi(y)\,.$$

■

We still need to introduce an additional concept, adapted from [24], which is very common in the literature about jump processes and quasi-limiting distribution.

Proposition 4.5. *Let Y be a jump process and define $p : \mathbb{R}_+\times\mathsf{X}\times\mathsf{X}\to\mathbb{R}_+$ by*

$$p(t,x,y) := \mathbb{P}_x(Y_t = y) = P_t(x,\{y\})\,.$$

Then, for any $t\in\mathbb{R}_+$, the transition probability P_t can be represented by the matrix $(p(t,x,y))_{x,y\in\mathsf{X}}$ and we have, for any measurable function $f : \mathsf{X}\to\mathbb{R}$ bounded or nonnegative,

$$\forall x\in\mathsf{X},\quad P_tf(x) = \sum_{y\in\mathsf{X}} p(t,x,y)f(y)\,.$$

By extension we also often call p the transition function.

Proof. The second equality in the definition of p is given using definition 2.3:

$$\mathbb{P}_x(Y_t = y) \;=\; \mathbb{E}_x[\mathbf{1}_{\{y\}}(Y_t)] \;=\; P_t\mathbf{1}_{\{y\}}(x) \;=\; \sum_{z\in\mathsf{X}} P_t(x,\{z\})\mathbf{1}_{\{y\}}(z) \;=\; P_t(x,\{y\})\,.$$

The second result is a direct application of equation (2.2) since we have a discrete topology. ■

When (P_t) is taken as a matrix, the generator can then be described in the same way. From proposition 4.4 it is clear that L is described by $(q(x,y))_{x,y\in\mathsf{X}}$ and we have

$$Q = \lim_{t\downarrow 0} \frac{P_t - I}{t} \tag{4.4}$$

where I is the identity matrix. Thus, by proposition 2.13, Q completely describes the transition function. The interpretation of $q(x,y)$ as an instantaneous rate of transition is then given by the following proposition.

Proposition 4.6. *Let Y satisfy assumption 3. Then, for any $x,y\in\mathsf{X}$ we have,*

$$q(x,y) = \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}_x(Y_t = y)_{|t=0} = p'(0,x,y)\,.$$

Proof. When we apply proposition 4.4 with $\varphi = \mathbf{1}_{\{y\}}$ we get $L\mathbf{1}_{\{y\}}(x) = q(x,y)$. Moreover, using definition 2.11 of the generator we have

$$L\mathbf{1}_{\{y\}} = \lim_{t\downarrow 0}\frac{1}{t}\left(P_t\mathbf{1}_{\{y\}} - \mathbf{1}_{\{y\}}\right) = \lim_{t\downarrow 0}\frac{1}{t}\left(P_t\mathbf{1}_{\{y\}} - P_0\mathbf{1}_{\{y\}}\right)$$

and we conclude using that $\mathbb{P}_x(Y_t = y) = \mathbb{E}_x[\mathbf{1}_{\{y\}}(Y_t)\mid\mathscr{F}_0] = P_t\mathbf{1}_{\{y\}}(x)$. ■

More generally, a well-known property (Cf [24] for example) is that, provided that there is no instantaneous jump rate, the transition function satisfies the *Kolmogorov's backward equation*,

$$\forall t\geqslant 0,\quad P'(t) = QP(t) \tag{4.5}$$

and when we have $\forall t\geqslant 0, \forall x\in\mathsf{X}, \sum_{y\in\mathsf{X}} p(t,x,y)q(y) < \infty$ then the *Kolmogorov's forward equation* applies too:

$$\forall t\geqslant 0,\quad P'(t) = P(t)Q\,. \tag{4.6}$$

4.2 Useful results in discrete space

We can now give more specific results about jump processes in the setting given before, featuring an absorbing set with sure killing. This means that Y, still defined as the canonical càdlàg Feller process with semigroup (P_t), satisfies now assumptions 1 and 2.

Recall that we denote by T the hitting time of the absorbing set X_{ab} and by Y^T the stopped process $(Y_{t\wedge T})_{t\geqslant 0}$. To use Y^T here makes sense in the context of quasi-stationary distributions since we are essentially interested in what happens before T. In particular, we can rewrite $\Pi(x,y)$ as equal to $\mathbb{P}_x(Y^T_{T_1} = y)$ for any $x\in\mathsf{X}$.

We create, for the following proposition a transition probability K (as defined in 2.1) by

$$\forall x \in \mathsf{X}, \forall A \in \mathscr{B}(\mathsf{X}), \quad K(x,A) := \begin{cases} \mathbb{P}_x(Y_{T_1} \in A) & \text{if} \quad x \in \mathsf{X}_{\mathrm{tr}} \\ \mathbf{1}_A(x) & \text{if} \quad x \in \mathsf{X}_{\mathrm{ab}} \end{cases} \tag{4.7}$$

which on X_{tr} can be seen as an extension of Π.

Proposition 4.7. *Let $x \in \mathsf{X}$. The sequence $Y_0, Y^T_{T_1}, Y^T_{T_2}, \ldots$ is under $\mathbb{P}_x$ a discrete Markov chain with transition kernel K started from x. Furthermore, for every $i \in \mathbb{N}^*$, conditionaly on $T_i < \infty$ and $(Y_0, Y_{T_1}, \ldots, Y_{T_i})$ the random variables $T_1 - T_0, \ldots, T_i - T_{i-1}$ are independent and the conditional distribution of $T_i - T_{i-1}$ is exponential with parameter $q(Y_{T_{i-1}})$.*

Proof. Let $x \in \mathsf{X}$, $k \in \mathbb{N}$ and $A \in \mathscr{B}(\mathsf{X})$. Using the strong Markov property 2.16 at T_k we get

$$\mathbb{P}_x(Y^T_{T_{k+1}} \in A \mid \mathscr{F}_{T_k}) = \mathbb{E}_{Y^T_{T_k}}[\mathbf{1}_{\{Y^T_{T_1} \in A\}}] = \mathbb{P}_{Y^T_{T_k}}(Y^T_{T_1} \in A) = K(Y^T_{T_k}, A)$$

which proves that this is a Markov chain. Then, let $x, y \in \mathsf{X}_{\mathrm{tr}}$, $z \in \mathsf{X}$ and $f_1, f_2 \in L^\infty(\mathbb{R}_+)$. Again, using the strong Markov property at T_1 we get

$$\begin{aligned} &\mathbb{E}_x[\mathbf{1}_{\{T_2<\infty\}}\mathbf{1}_{\{Y_{T_1}=y\}}f(T_1)\mathbf{1}_{\{Y_{T_2}=z\}}f(T_2-T_1)] \\ &\quad = \mathbb{E}_x[\mathbf{1}_{\{T_2<\infty\}}\mathbf{1}_{\{Y_{T_1}=y\}}f_1(T_1)\mathbb{E}_{Y_{T_1}}[\mathbf{1}_{\{Y_{T_1}=z\}}f_2(T_1)]] \\ &\quad = \mathbb{E}_x[\mathbf{1}_{\{Y_{T_1}=y\}}f_1(T_1)]\mathbb{E}_y[\mathbf{1}_{\{Y_{T_1}=z\}}f_2(T_1)] \\ &\quad = q(x,y)q(y,z)\int_0^\infty e^{-q(x)s_1}f_1(s_1)\,\mathrm{d}s_1 \int_0^\infty e^{-q(y)s_2}f_2(s_2)\,\mathrm{d}s_2\,. \end{aligned}$$

Arguing by induction, we have similarly, for every $y_0, y_1, \ldots, y_{p-1} \in \mathsf{X}_{\mathrm{tr}}$, $y_p \in \mathsf{X}$ and $f_1, \ldots, f_p \in L^\infty(\mathbb{R}_+)$,

$$\begin{aligned} &\mathbb{E}_{y_0}[\mathbf{1}_{\{T_p<\infty\}}\mathbf{1}_{\{Y_{T_1}=y_1\}}\mathbf{1}_{\{Y_{T_2}=y_2\}}\cdots\mathbf{1}_{\{Y_{T_p}=y_p\}}f_1(T_1)f_2(T_2-T_1)\cdots f_p(T_p-T_{p-1})] \\ &= \prod_{i=1}^{p}\left(q(y_{i-1},y_i)\int_0^\infty e^{-q(y_{i-1})s}f_i(s)\,\mathrm{d}s\right) \end{aligned}$$

The remaining assertions of the proposition follow. ∎

This is what justifies the notation of X_{tr}, its elements are the transient states of the discrete-time Markov chain $(Y^T_{T_i})$. Indeed, for a discrete-time Markov chain, states are either transient or recurrent. If $x \in \mathsf{X}$ is transient, then, when the Markov chain starts at x, there is a positive probability that it will never return to that state. On the contrary, if x is recurrent, then the Markov chain will take the value x again with probability 1. Clearly, a recurrent state in X_{tr} would break assumption 2, so all the states in X_{tr} are necessarily transient.

It is now necessary to introduce the concept of explosiveness, to better rule out this possibility after.

Definition 4.8. A jump process $(Y_t)_{t\geqslant 0}$ with jump times denoted by $T_1, T_2, \ldots,$ is said to be *explosive* if there exists a state $x \in \mathsf{X}$ such that $\mathbb{P}_x(T_\infty < \infty) > 0$, with $T_\infty = \lim_{n\to\infty} T_n$.

As a consequence, a process is *non-explosive* if it satisfies

$$\forall \alpha \in \mathscr{P}(\mathsf{X}),\ \mathbb{P}_\alpha(T_\infty = \infty) = 1\ . \tag{4.8}$$

A non-explosive right-continuous jump process is also said to be *regular*. It makes sense for us to look for non-explosive processes in order to study birth-and-death processes and epidemics: the number of people born, deceased or infected in a given period of time is naturally finite. As we will show now, this is always the case with the assumptions that we already have.

Proposition 4.9. *Let Y be a càdlàg Feller process. Then Y is non-explosive. Moreover, if Y satisfies assumptions 1 and 2, for any $x \in \mathsf{X}_{\mathrm{tr}}$, the number of jumps before T under $\mathbb{P}_x$ is almost surely finite.*

Proof. In the countable setting, X_{ab} is reached at the first jump to one of its states and we will show that there can only be a finite number of jumps before. Indeed, if there was an infinite number of jumps in any $[0,t]$ with $t \in \mathbb{R}_+$ then, using the Bolzano-Weierstraß property, we would find a point τ limit of a subsequence of $(T_i)_{i\in\mathbb{N}}$. But then the sample path at τ couldn't have both a left limit and a right limit in the discrete topology, breaking the definition of Y as càdlàg. Thus, the process is non-explosive.

In particular, with any $\omega \in \{T < \infty\}$ we know that there are only finitely many jumps in $[0, T(\omega)]$. We can then conclude using assumption 2. ■

Remark. When X_{ab} is a singleton, then $\forall \omega \in \Omega, \exists p \in \mathbb{N}^*, Y_T(\omega) = Y_{T_p}(\omega)$ and $T_{p+1}(\omega) = \infty$, implying that the following terms are infinite as well.

Proposition 4.10. *Let Y satisfy assumptions 1 and 2. Then there exists $x \in \mathsf{X}_{\mathrm{tr}}$ such that $\mathbb{P}_x(T_1 = T) > 0$.*

This result can also be written as $\mathbb{P}_x(Y_{T_1} \in \mathsf{X}_{\mathrm{ab}}) = K(x, \mathsf{X}_{\mathrm{ab}}) > 0$ where K is the kernel defined by (4.7), or equivalently $\sum_{y\in\mathsf{X}_{\mathrm{ab}}} q(x,y) > 0$. This result is intuitive and key to understand our processes, but not formally proved in our sources. Therefore, we give an original proof of it.

Proof. Suppose by contradiction that $\forall x \in \mathsf{X}_{\mathrm{tr}}, \mathbb{P}_x(T_1 = T) = \mathbb{P}_x(Y_{T_1} \in \mathsf{X}_{\mathrm{ab}}) = 0$.

Let $x \in \mathsf{X}_{\mathrm{tr}}$ and $n \in \mathbb{N}$. Using the strong Markov property we get

$$\begin{aligned}
\mathbb{P}_x(Y_{T_n} \in \mathsf{X}_{\mathrm{ab}}) &= \mathbb{E}_x\left[\mathbb{E}_x[\mathbf{1}_{\mathsf{X}_{\mathrm{ab}}}(Y_{T_n}) \mid \mathscr{F}_{n-1}]\right] \\
&= \mathbb{E}_x\left[\mathbb{E}_{Y_{T_{n-1}}}[\mathbf{1}_{\mathsf{X}_{\mathrm{ab}}}(Y_1)]\right] \\
&= \mathbb{P}_x(Y_{T_{n-1}} \in \mathsf{X}_{\mathrm{tr}}) \cdot \mathbb{E}_x\left[\mathbb{E}_{Y_{T_{n-1}}}[\mathbf{1}_{\mathsf{X}_{\mathrm{ab}}}(Y_1)] \mid Y_{T_{n-1}} \in \mathsf{X}_{\mathrm{tr}}\right] \\
&\quad + \mathbb{P}_x(Y_{T_{n-1}} \in \mathsf{X}_{\mathrm{ab}}) \cdot \mathbb{E}_x\left[\mathbb{E}_{Y_{T_{n-1}}}[\mathbf{1}_{\mathsf{X}_{\mathrm{ab}}}(Y_1)] \mid Y_{T_{n-1}} \in \mathsf{X}_{\mathrm{ab}}\right] \\
&= 0 + \mathbb{P}_x(Y_{T_{n-1}} \in \mathsf{X}_{\mathrm{ab}})
\end{aligned}$$

and arguing by induction from $\mathbb{P}_x(Y_{T_0} \in \mathsf{X}_{\mathrm{ab}}) = 0$ we get $\mathbb{P}_x(Y_{T_n} \in \mathsf{X}_{\mathrm{ab}}) = 0$. Thus, X_{ab} cannot be reached in a finite number of jumps, but non-explosiveness implies that $T_\infty = \infty$ almost surely so this would contradict the sure killing assumption. ■

Let $Q = (q(x,y))_{x,y\in\mathsf{X}}$ be the transition rate matrix and $Q_{\mathrm{tr}} = (q(x,y))_{x,y\in\mathsf{X}_{\mathrm{tr}}}$ its restriction to X_{tr}. The matrix Q_{tr} is said to be a defective generator since its entries sum to a negative value, as a consequence of proposition 4.10.

Definition 4.11. Let Z be a subset of X. A matrix $M = (m_{x,y})_{x,y\in\mathsf{Z}}$ is *irreducible* if for any pair $(x, y) \in \mathsf{Z}^2, x \neq y$, there exists an integer n and a sequence $z_1, z_2, \ldots, z_{n-1}$ in $\mathsf{Z}\setminus\{x, y\}$ such that all z_i are distincts and we have

$$m(x,z_1) > 0, \quad m(z_{n-1}, y) > 0 \quad \text{and} \quad \forall i \in \{1,\ldots,n-2\}, m(z_i,z_{i+1}) > 0 \; .$$

The process Y is said to be irreducible if Q_{tr} is irreducible.

For Y to be irreducible means that there always exists a "path" between two transient states.

Proposition 4.12. *If X is discrete and Y is irreducible then $\forall t > 0, \forall x, y \in \mathsf{X}_{\text{tr}}, p(t, x, y) > 0$.*

We sometimes say that Y is *aperiodic* because of that property, by similarity with the discrete-time case. This result is the reason why it is often easier to work with Q: for each $x \in \mathsf{X}$ there are often only a finite number of y such that $q(x, y) > 0$. So the action of Q as an operator can be described simply. On the contrary, for each P_t with $t > 0$, if Q is irreducible then P_1 has all its terms positive, except those of some absorbing states.

The following proof is also an original work.

Proof. Let $t > 0$ and $x, y \in \mathsf{X}_{\text{tr}}$. Then, by definition 4.11 there exist $n \in \mathbb{N}^*$ and states $z_1, z_2, \ldots, z_{n-1}$ such that $q(x,z_1) > 0$, $q(z_{n-1}, y) > 0$ and $\forall i \in \{1, \ldots, n-2\}, q(z_i, z_{i+1}) > 0$. Let

$$\begin{aligned} & A = \left\{Y_{T_1} = z_1, \ldots, Y_{T_{n-1}} = z_{n-1}, Y_{T_n} = y\right\} \\ \text{and} \quad & B = \left\{T_1 \leqslant \frac{t}{n}, T_2 - T_1 \leqslant \frac{t}{n}, \ldots, T_n - T_{n-1} \leqslant \frac{t}{n}, T_{n+1} - T_n > t\right\} \; . \end{aligned}$$

Then, using proposition 4.7 we have

$$p(t, x, y) = \mathbb{P}_x(Y_t = y) \geqslant \mathbb{P}_x(A \cap B) = \mathbb{P}_x(A) \cdot \mathbb{P}_x(B) \; .$$

Since all the states are in X_{tr} we get

$$\mathbb{P}_x(A) = \frac{q(x,z_1)}{q(x)} \times \frac{q(z_1,z_2)}{q(z_1)} \times \cdots \times \frac{q(z_{n-1},y)}{q(z_{n-1})} > 0$$

and

$$\mathbb{P}_x(B) = \mathbb{P}_x(T_{n+1} - T_n > t) \cdot \prod_{i=1}^{n} \mathbb{P}_x(T_i - T_{i-1}) > 0$$

which completes the proof. ∎

In some applications in section 6, we will be interested in state spaces whose absorbing set contains an infinity of states. As it is easier to work with processes which admit only one absorbing state, we will show that the dishonest transition function obtained by restricting $(P_t)_t$ to X_{tr} can be made into an honest one by adding only one state. This new state and can then be used as an equivalent of X_{ab} for the analysis of what happens before T. This proposition is taken from [2].

Proposition 4.13. *Let $(R_t)_t$ be a dishonest transition function on a countable set E. Let ϕ be a point not in E and define $E_\phi = E \cup \{\phi\}$ and*

$$\forall (x,y) \in E_\phi, \forall t \in \mathbb{R}_+, \quad R_t^\phi(x,\{y\}) = \begin{cases} R_t(x,\{y\}) & \text{if} \quad x,y \in E \\ 1 - R_t(x,E) & \text{if} \quad x \in E, y = \phi \\ 0 & \text{if} \quad x = \phi, y \in E \\ 1 & \text{if} \quad x = y = \phi \end{cases} .$$

Then, $(R_t^\phi)_t$ is an honest transition function on E_ϕ.

Proof. It is straightforward from its definition that $(R_t^\phi)_t$ is a collection of transition probabilities, so we only need to check that it satisfies the Chapman-Kolmogorov equation. Moreover, in the discrete topology it is enough to verify it on singletons. Let $s,t \in \mathbb{R}_+$ and $x,y \in E_\phi$.

If $x,y \in E$ we have

$$\begin{aligned} R_s^\phi R_t^\phi(x,\{y\}) &= \sum_{z \in E} R_s(x,\{z\}) R_t(z,\{y\}) + R_s^\phi(x,\{\phi\}) R_t^\phi(\phi,\{y\}) \\ &= R_{s+t}(x,\{y\}) + 0 = R_{s+t}^\phi(x,\{y\}) \end{aligned}$$

If $x \in E$ and $y = \phi$ then

$$\begin{aligned} R_s^\phi R_t^\phi(x,\{y\}) &= \sum_{z \in E} R_s(x,\{z\})(1 - R_t(z,E)) + (1 - R_t(x,E)) \\ &= 1 - \sum_{z \in E} R_s(x,\{z\}) R_t(z,E) \\ &= 1 - R_{s+t}(x,E) = R_{s+t}^\phi(x,\phi) . \end{aligned}$$

If $x = \phi$ and $y \in E$ then $R_s^\phi R_t^\phi(x,\{y\}) = 0 = R_{s+t}^\phi(x,\{y\})$, and finally if $x = y = \phi$ then $R_s^\phi R_t^\phi(x,\{y\}) = 1 = R_{s+t}^\phi(x,\{y\})$. ∎

4.3 Existence of a QSD

Results of this section come from [17] but seemed to be originally incorrect on some points. They were corrected and adapted to the discrete case.

Let Y be a càdlàg Feller process satisfying assumptions 1, 2 and 3, with semigroup $(P_t)_t$ and generator L. Let T be the hitting time of the absorbing set X_{ab}. We denote by $D(L)$ the domain of L (Cf definition 2.11) and

$$\Delta = \{\mathbf{1}_B \mid B \subset \mathsf{X}_{\text{tr}}, |B| < \infty\} . \tag{4.9}$$

Lemma 4.14. *We have $\Delta \subset D(L)$ and, for any set $A \subset \mathsf{X}_{\text{tr}}$, there exists a uniformly bounded sequence $(f_n)_n$ in Δ converging point wisely to $\mathbf{1}_A$.*

Proof. By proposition 4.4 we have $D(L) = C_0(\mathsf{X})$. The functions in Δ are trivially continuous since we are in a discrete space, and tend to 0 at infinity because they have a compact support. Thus, $\Delta \subset D(L)$.

Now, let $A \subset \mathsf{X}_{\text{tr}}$. If A is finite we can put $\forall n \in \mathbb{N}, f_n := \mathbf{1}_A$. If A is not finite then, since the space is countable, we can write $A = \{a_1, a_2, a_3, \dots\}$ for some distinct $a_1, a_2, a_3, \dots \in \mathsf{X}_{\text{tr}}$. For every $n \in \mathbb{N}$ we set f_n as the indicator function of $\{a_1, \dots, a_n\}$. It is direct to verify that those functions are uniformly bounded (by 1) and converge pointwise to $\mathbf{1}_A$. ∎

Proposition 4.15. *Let $\nu \in \mathscr{P}(\mathsf{X}_{\text{tr}})$. Then ν is a quasi-stationary distribution if and only if*

$$\exists \vartheta(\nu) > 0, \forall f \in D(L),\ \nu(Lf) = -\vartheta(\nu)\nu(f)\ .$$

Proof. 1. Let ν be a QSD for Y. Using the definition 2.3 of homogeneous Markov processes in equation (3.4) we get

$$\forall t \in \mathbb{R}_+, \forall A \in \mathscr{B}(\mathsf{X}_{\text{tr}}),\ \nu(P_t \mathbf{1}_A) = \nu(A)e^{-\vartheta(\nu)t}$$

where $\vartheta(\nu) \in \mathbb{R}_+$ is given by theorem 3.7. In the discrete setting it is then straightforward that

$$\forall f \in D(L),\ \nu(P_t f) = e^{-\vartheta(\nu)t}\nu(f)\ .$$

Then, we use Kolmogorov's forward equation (4.6) to get

$$\forall f \in D(L), \forall x \in \mathsf{X}, \left|\frac{\partial P_t f}{\partial t}(x)\right| = |P_t Lf(x)| \leqslant \|Lf\|_\infty < +\infty\ .$$

This implies that we can differenciate $\nu(P_t f) = \sum_{y \in \mathsf{X}_{\text{tr}}} P_t f(x)\nu(\{x\})$ under the sign sum, and this yields $\forall f \in D(L), \nu(Lf) = -\vartheta(\nu)\nu(f)$.

2. Assume now that $\forall f \in D(L), \nu(Lf) = -\vartheta(\nu)\nu(f)$. Using Kolmogorov's backward equation and the same derivation argument we get

$$\forall f \in \Delta, \frac{\partial\, \nu(P_t f)}{\partial t} = \nu(LP_t f) = -\vartheta(\nu)\nu(P_t f)$$

since $P_t f$ is still in $D(L)$ by proposition 2.12. We deduce that

$$\forall f \in D(L),\ \nu(P_t f) = e^{-\vartheta(\nu)t}\nu(f)\ .$$

Let $A \subset \mathsf{X}$. From lemma 4.14, there exists a uniformly bounded sequence (f_n) in $D(L)$ which converges point-wisely to $\mathbf{1}_A$. Therefore, by dominated convergence we get

$$\nu(P_t \mathbf{1}_A) = e^{-\vartheta(\nu)t}\nu(A)$$

and as stated in section 3.3 this is equivalent to ν being a QSD.

■

4.4 Some results on birth-and-death processes

Before going further into details about epidemic models, we review some results on birth-and-death processes. This is useful in so far as it is one of the most studied application of quasi-stationary distributions and some results can be applied without much change to epidemic models. This section is mostly based on [17].

We assume now $\mathsf{X} = \mathbb{N}$ and $\mathsf{X}_{\text{ab}} = \{0\}$, meaning that the system is isolated: there is no immigration so when we reach 0 the system does not evolve anymore. We still denote by $(Y_t)_{t \geqslant 0}$ our Markov process, which gives here the size of the population.

As in the previous section we denote by $q(x, y)$ the transition rate from x to y. The standard condition characterizing birth-death processes is

$$\forall i, j \in \mathbb{N}, \quad |i - j| > 1 \implies q(i, j) = 0 \tag{4.10}$$

meaning that there cannot be several births or deaths at the exact same time.

If we exclude $q(0,0)$ which is trivial, then it remains a defective transition matrix. This matrix can be described more simply by a *birth rate* sequence $(\lambda_i)_{i\geqslant 1}$ and a *death rate* sequence $(\mu_i)_{i\geqslant 1}$ such that $q(0,1)=0$ and

$$\forall i \in \mathbb{N}^*, \quad q(i,i+1)=\lambda_i>0, \quad q(i,i-1)=\mu_i>0, \quad q(i,i)=-(\lambda_i+\mu_i)\;. \tag{4.11}$$

Remark. From the positivity conditions above we notice that Q_{tr} is irreducible by construction. We can also see that assumption 1 is a consequence of the definition of the birth and death rates, so it is always verified by a birth-and-death process.

A first classic model is that of *linear birth-and-death process*, for a case when individuals reproduce and die independently from the number of other people. Formally, there exist $\lambda,\mu\geqslant 0$ such that for all $i\in\mathbb{N}^*$, $\lambda_i=i\lambda$ and $\mu_i=i\mu$. In that case, λ and μ are simply named birth rate and death rate respectively.

A more evolved model is the *logistic birth-and-death process* where individuals compete to share ressouces in a finite environment. Thus, each individual j creates a competition pressure on any individual $k\neq j$, with a factor $c>0$ such that the individual death rate due to competition is $c(i-1)$ when the population's size is i. Overall we have

$$\forall i\in\mathbb{N}^*, \quad \lambda_i=i\lambda, \quad \mu_i=i\mu+i(i-1)c$$

where λ and μ are still positive real numbers.

Note that we should verify that these processes are Feller. We will skip this matter for now and study it in chapter 6. In particular, the fact that the two models above are Feller is a consequence of proposition 6.5. In the rest of this section we just assume that the model considered is Feller.

The following result is a classic condition to ensure that the process considered through its birth rate and death rate sequences does not explode. Thus, it is a limitation that we need to consider when we are looking for quasi-stationary distributions.

Proposition 4.16. *A birth-and-death process Y with birth rates (λ_i) and death rates (μ_i) satisfies assumption 2, i.e. goes almost surely to extinction, if and only if*

$$\sum_{k=1}^{\infty}\frac{\mu_1\cdots\mu_k}{\lambda_1\cdots\lambda_k}=+\infty\;.$$

Proof. Let (u_i) be the sequence given by $\forall i\in\mathbb{N}, u_i=\mathbb{P}_i(T<\infty)$. We have

$$\begin{aligned}\mathbb{P}_i(T<\infty)&=\mathbb{P}_i(T<\infty,Y_{T_1}=i-1)+\mathbb{P}_i(T<\infty,Y_{T_1}=i+1)\\&=\mathbb{P}_i(Y_{T_1}=i-1)\cdot\mathbb{P}_i(T<\infty\mid Y_{T_1}=i-1)\\&\quad+\mathbb{P}_i(Y_{T_1}=i+1)\cdot\mathbb{P}_i(T<\infty\mid Y_{T_1}=i+1)\;.\end{aligned}$$

So that, with the strong Markov property and the notation that we introduced we get

$$\forall i\in\mathbb{N}^*,\ \lambda_i u_{i+1}-(\lambda_i+\mu_i)u_i+\mu_i u_{i-1}=0\;. \tag{4.12}$$

For every $j\in\mathbb{N}$ we denote by τ_j the hitting time of the state j and $\tau_\infty=\lim_{j\to\infty}\tau_j$. Moreover, for all $i,j\in\mathbb{N}$ we define $u_i^{(j)}:=\mathbb{P}_i(T<\tau_j)$. In particular, we have $\lim_{j\to\infty}u_i^{(j)}\leqslant u_i$. We assumed that the process is Feller, so by proposition 4.9 we get $\forall i,\mathbb{P}_i(\tau_\infty=\infty)=1$, and then $\lim_{j\to\infty}u_i^{(j)}=u_i$.

We also set $\forall i \in \mathbb{N}, U_i := \sum_{k=1}^{i-1} \frac{\mu_1 \cdots \mu_k}{\lambda_1 \cdots \lambda_k}$. From the definition of $u_i^{(j)}$ and since we can not "skip" states, we have $\forall i \geqslant j, u_i^{(j)} = 0$ and similarly to the result above we get

$$\forall i \in \{1, \dots, j-1\},\ \lambda_i u_{i+1}^{(j)} - (\lambda_i + \mu_i) u_i^{(j)} + \mu_i u_{i-1}^{(j)} = 0\ .$$

It is then direct to verify that the solution to these equations is

$$\forall i \in \{1, \dots, j-1\},\ u_i^{(j)} = \frac{1}{1+U_j} \sum_{k=i}^{j-1} \frac{\mu_1 \cdots \mu_k}{\lambda_1 \cdots \lambda_k}\ .$$

In particular, $u_1^{(j)} = \frac{U_j}{1+U_j}$, hence $u_1 = \frac{U_\infty}{1+U_\infty}$. Thus, $U_\infty = \infty$ implies $u_1 = 1$ and using equation (4.12) it follows that $\forall i \in \mathbb{N}, \mathbb{P}_i(T < \infty) = 1$. Conversely, if Y satisfies assumption 2, then $u_1 = 1$, which implies $U_\infty = \infty$. ■

Corollary 1. 1. *The linear birth-and-death process with rates λi and μi satisfies assumption 2 if and only if $\lambda \leqslant \mu$.*

2. *The logistic birth-and-day process goes almost surely to extinction.*

This is a direct application of the previous proposition. We give an original proof of it: in [17] the first case is justified differently and the second has no proper proof at all.

Proof. 1. The series $\sum_{k=1}^{\infty} \frac{\mu_1 \cdots \mu_k}{\lambda_1 \cdots \lambda_k}$ is now that of a geometric sequence with ratio $\frac{\mu}{\lambda}$. Therefore, it is infinite if and only if $\frac{\mu}{\lambda} \geqslant 1$ and we conclude by proposition 4.16.

2. In the case of the logistic birth-and-day process we have, for any $k \in \mathbb{N}^*$,

$$\frac{\mu_1 \cdots \mu_k}{\lambda_1 \cdots \lambda_k} = \frac{\prod_{i=1}^{k} (\mu i + c i (i-1))}{k! \lambda^k} \leqslant \mu \left(\frac{c}{\lambda}\right)^k (k-1)!$$

which gives us $\lim_{k\to\infty} \frac{\mu_1 \cdots \mu_k}{\lambda_1 \cdots \lambda_k} = +\infty$. Therefore its series sums up to $+\infty$ and we conclude by theorem 4.16.

■

We are generaly interested in the *extinction rate* and its evolution, defined by

$$\forall \alpha \in \mathscr{P}(\mathsf{X}), \forall t \geqslant 0,\ r_\alpha(t) = -\frac{\frac{\partial}{\partial t} \mathbb{P}_\alpha(T > t)}{\mathbb{P}_\alpha(T > t)} \tag{4.13}$$

and from theorem 3.7 it is clear that if α is a QSD then r_α is constant and equal to $\vartheta(\alpha)$.

Consistently with this, a result from [17] that we do not reproduce here shows that if α admits a QLD ν then there exists a long term mortality plateau such that $\lim_{t\to\infty} r_\alpha(t) = r_\nu(0)$.

Now that we have characterized birth-and-death processes and their relation to the previous results, we can focus on quasi-stationary distributions.

Notation. When talking about a distribution α over X we will now use equivalently the sequence $(\alpha_i)_{i\in\mathbb{N}}$ where $\forall i \in \mathbb{N}, \alpha_i = \alpha(\{i\})$. A sufficient and necessary condition on $(\alpha_i)_i$ to represent a distribution is that all its terms are non-negative and sum up to 1. Moreover, if $\alpha = (\alpha_i)_{i\geqslant 1}$ is a distribution on X_{tr}, we implicitely extend it to X by the use of $\alpha_0 = 0$.

Theorem 4.17. *Let Y be a Feller birth-and-death process satisfying assumptions 1 and 2, with birth rates (λ_i) and death rates (μ_i). A distribution $\nu \in \mathscr{P}(\mathsf{X}_{\mathrm{tr}})$ is a QSD for Y if and only if*

$$\forall i \geqslant 1, \quad \lambda_{i-1}\nu_{i-1} - (\lambda_i + \mu_i)\nu_i + \mu_{i+1}\nu_{i+1} = -\mu_1\nu_1\nu_i$$

and $-(\lambda_1 + \mu_1)\nu_1 + \mu_2\nu_2 = -\mu_1\nu_1^2$.

Proof. Let ν be a QSD. Then from proposition 4.15 we know that there exists $\vartheta > 0$ such that $\forall f \in D(L)$, $\nu(Lf) = -\vartheta\nu(f)$. The matrix interpretation of this property is

$$\left(\sum_j \nu_j q(j,i)\right)_i = -\vartheta \times (\nu_i)_i$$

and, in the case of birth-and-death processes, element by element this gives us

$$\forall i \in \mathbb{N}^*, \quad \nu_{i-1}q(i-1,i) + \nu_i q(i,i) + \nu_{i+1}q(i+1,i) = -\vartheta\nu_i \ .$$

With the specific notation we get

$$\forall i \geqslant 2, \quad \nu_{i-1}\lambda_{i-1} - \nu_i(\lambda_i + \mu_i) + \nu_{i+1}\mu_{i+1} = -\vartheta\nu_i$$

and $-(\lambda_1 + \mu_1)\nu_1 + \mu_2\nu_2 = -\vartheta\nu_1$. Then, summing over all elements,

$$\sum_{i\in\mathbb{N}} \lambda_i\nu_i + \sum_{i\in\mathbb{N}}(\lambda_i + \mu_i)\nu_i + \sum_{i\in\mathbb{N}}\mu_i\nu_i - \mu_1\nu_1 = -\vartheta \ .$$

Therefore, $\vartheta = \mu_1\nu_1$, which concludes the proof. ■

Following this theorem, we can find in the literature (in particular [27]) that there exist a parameter ξ_1 and a series (S) on which depend the existence and number of quasi-stationary distributions.

Chapter 5

The case of epidemic modelling

We will now study more specifically the case of epidemics and the quasi-stationary distributions of their related models. This is generally more complex than the modelling of a population and we need to make some practical assumptions in addition to mathematical considerations.

In particular, we will suppose that there is *homogeneous mixing*, meaning that all individuals have the same probability of meeting and infecting each other. In some papers, e.g. [11], the case of *heterogeneous mixing* is studied as well.

In this chapter we will consider directly stochastic models and not deterministic ones. Those stochastic models are often introduced through infinitesimal transition probabilities of the form $p(\delta t, x, y)$ with δt a small interval of time. Instead, we will present the models directly from transition rates, which provide a more rigorous description and allow us to use results from the previous sections.

5.1 Results in a finite state space

The most basic models in epidemic modelling often occur in a finite state space. In that setting all the results given in chapter 4 still hold but we have access to stronger results regarding quasi-stationary distributions. We will review some of them from [17].

To this end we consider a Feller process Y on the state space $\mathsf{X} = \{0, \dots, N\}$ and satisfying assumption 1 with $\mathsf{X}_{\mathrm{ab}} = \{0\}$, meaning that $\mathsf{X}_{\mathrm{tr}} = \{1, \dots, N\}$. Its transition function is still denoted by (P_t) and the killing time by T.

Remark. A Markov process with bounded jump rates $q(i)$ corresponds to a Feller process. This is clearly verified in a finite state space as long as no jump rate is infinite.

The first proposition is an original work and provides a link with the assumptions of the previous chapters. It shows that, in the finite case, it is unnecessary to make assumption 2 or prove it since it is a direct consequence of the model.

Proposition 5.1. *Let Y be an irreducible càdlàg Feller process on $\{0, \dots, N\}$ with absorbing state 0, satisfying assumption 1. Then, Y satisfies assumption 2 of sure killing if and only if there exists $x \in \{1, \dots, N\}$ such that $q(x, 0) > 0$.*

Proof. The first implication is given by proposition 4.10. Suppose now that we have x such that

$q(x,0) > 0$ and let $y \in \{1,\dots,N\}$. Using the Markov property and proposition 4.1 we get

$$\begin{aligned}\mathbb{P}_y(Y_1 = 0) &\geqslant \mathbb{P}_y(Y_{1/2} = x, Y_1 = 0)\\ &= p(1/2, y, x) \cdot \mathbb{P}_y(Y_1 = 0 \mid Y_{1/2} = x)\\ &= p(1/2, y, x) \cdot \mathbb{P}_x(Y_{1/2} = 0)\\ &\geqslant p(1/2, y, x) \cdot \mathbb{P}_x(Y_{T_1} = 0, T_1 \leqslant 1/2)\\ &= p(1/2, y, x) \cdot \frac{q(x,0)}{q(x)} \cdot (1 - e^{1/2})\end{aligned}$$

From proposition 4.12 we conclude that there exists $\varepsilon_y \in \,]0,1]$ such that $p(1,y,0) = \varepsilon_y$. The same is true for any state in X_{tr} so we can take $\varepsilon := \min_{y\in\mathsf{X}_{\text{tr}}} \varepsilon_y > 0$ such that $\forall y \in \mathsf{X}_{\text{tr}}, p(1,y,0) \geqslant \varepsilon$. Let α be an initial distribution over X and $(u_n)_n$ the sequence given by $\forall n \in \mathbb{N}, u_n = \mathbb{P}_\alpha(T < n)$. From the previous calculations it comes that

$$\forall n \in \mathbb{N},\ u_n \geqslant u_{n-1} + \varepsilon(1 - u_{n-1})\ .$$

Therefore, the sequence $(u_n)_n$ converges to 1, which proves that there is sure killing. ∎

The distributions, that we wrote as sequences in chapter 4 can now be represented as vectors in $\mathbb{R}^{N+1}$, and we use them as column matrices. Then, a straightforward calcultation from proposition 2.5 gives us that, with an initial distribution $\alpha = (\alpha_i)_{0\leqslant i\leqslant N}$, the distribution of Y_t is given by $\alpha^\mathsf{T} P_t$ where the transition probability P_t is taken as the matrix $(p(t,i,j))_{0\leqslant i,j\leqslant N}$ and α^T is the transpose of α, that is α taken as a line matrix.

We truncate (P_t) of $\{0\}$ to create the submarkovian semigroup $(R_t)_{t\in\mathbb{R}_+}$ given by

$$\forall t \in \mathbb{R}_+, \quad R_t := (p(t,i,j))_{1\leqslant i,j\leqslant N} \tag{5.1}$$

where p is the transition function defined in 4.5. Then, equation (3.4) gives us that a distribution $\nu \in \mathscr{P}(\mathsf{X}_{\text{tr}})$ is a QSD if and only if there exists $\vartheta(\nu) > 0$ such that

$$\forall t \in \mathbb{R}_+, \quad \nu^\mathsf{T} R_t = e^{-\vartheta(\nu)t}\,\nu \tag{5.2}$$

meaning that, for any t, ν is a left eigenvector of R_t with eigenvalue $e^{-\vartheta(\nu)t}$.

Additionaly, in the finite case, the forward and backward Kolmogorov's equations happen to have a simple solution, which is

$$\forall t \in \mathbb{R}_+, \quad P_t = e^{tQ}, \quad R_t = e^{tQ_{\text{tr}}}, \tag{5.3}$$

where $e^A = \sum_{k=0}^\infty \frac{A^k}{k!}$.

The following theorem extends on those considerations.

Theorem 5.2. *Let Y be an irreducible càdlàg Feller process on $\{0,\dots,N\}$ satisfying assumptions 1 and 2. Then the Yaglom limit exists and is the unique QSD of Y.*

Moreover, if we denote by ν this QSD and by $\vartheta(\nu)$ its extinction rate, as given by theorem 3.7, there exists $\pi \in \mathscr{P}(\mathsf{X}_{\text{tr}})$ such that for any $i,j \in \mathsf{X}_{\text{tr}}$,

$$\lim_{t\to\infty} \mathbb{P}_i(Y_t = j) = e^{-\vartheta(\nu)t}\pi_i \nu_j \quad \text{and} \quad \lim_{t\to\infty} \frac{\mathbb{P}_i(T > t+s)}{\mathbb{P}_j(T > t)} = \frac{\pi_i}{\pi_j} e^{-\vartheta(\nu)s}\ .$$

The proof of this theorem relies essentially on the Perron-Frobenius theorem that follows. We give it as stated in [17], which is more specific than the well-known usual theorem. The proof is not given in that paper either and the references given do not seem to contain the complete proof. Nonetheless, this result is covered completely across several theorems in the book [25] by Seneta, when we recall equation (5.3) and use its theorem 2.7 to extend the usual Perron-Frobenius theorem.

Theorem 5.3 (Perron-Frobenius Theorem). *Let (R_t) be a submarkovian semigroup on $\{1,\dots,N\}$ such that P_1 is positive in the sense that all its entries are. Then, there exists a unique simple positive eigenvalue ρ of R_1, which is the maximum of the modulus of the eigenvalues. There exists a unique left-eigenvector $\nu \in \mathbb{R}^N$ such that $\forall i, \nu_i > 0$ and $\sum_i \nu_i = 1$, and there exists a unique right-eigenvector $\pi \in \mathbb{R}^N$ such that $\forall i, \pi_i > 0$ and $\sum_i \alpha_i \pi_i = 1$, satisfying*

$$\nu^\mathsf{T} R_1 = \rho\,\nu, \quad R_1 \pi = \rho\,\pi\ .$$

Moreover, $\rho < 1$, so there exists $\vartheta > 0$ such that $\rho = e^{-\vartheta}$ and we have

$$\forall t > 0,\ R_t = e^{-\vartheta t}\pi\nu^\mathsf{T} + M(e^{-\xi t})$$

where $\xi > 0$ and $M(e^{-\xi t})$ denotes a matrix such that none of the entries exceeds $Ce^{-\xi t}$ for some constant $C > 0$.

Proof of theorem 5.2. Since Y is irreducible we know from proposition 4.12 that R_1 has only positive entries. Therefore, the Perron-Frobenius theorem applies. We get that there exist $\vartheta > 0$ and $\pi, \nu \in \mathscr{P}(\mathsf{X}_{\mathrm{tr}})$ such that

$$\forall i,j \in \mathsf{X}_{\mathrm{tr}}, \quad e^{\vartheta t}\mathbb{P}_i(Y_t = j) = e^{\vartheta t}[P_t]_{i,j} = \pi_i \nu_j + O(e^{-\xi t})$$

so that, summing over $j \in \mathsf{X}_{\mathrm{tr}}$, we get

$$\forall i \in \mathsf{X}_{\mathrm{tr}}, \quad e^{\vartheta t}\mathbb{P}_i(T > t) = \pi_i + O(e^{-\xi t})\ .$$

From the two previous equations it follows that

$$\forall i,j \in \mathsf{X}_{\mathrm{tr}}, \quad \lim_{t\to\infty} \mathbb{P}_i(Y_t = j \mid T > t) = \lim_{t\to\infty} \frac{\mathbb{P}_i(Y_t = j)}{\mathbb{P}_i(T > t)} = \nu_j$$

which proves that the Yaglom limit exists and is equal to ν. The last result of the theorem is also a direct consequence of the second equation. Since X is finite it also comes that

$$\forall \alpha \in \mathscr{P}(\mathsf{X}_{\mathrm{tr}}),\ \lim_{t\to\infty} \mathbb{P}_\alpha(Y_t = j \mid T > t) = \sum_{i\in\mathsf{X}_{\mathrm{tr}}} \alpha_i \lim_{t\to\infty} \mathbb{P}_i(Y_t = j \mid T > t) = \sum_{i\in\mathsf{X}_{\mathrm{tr}}} \alpha_i \nu_j = \nu_j$$

implying that the Yaglom limit is the only QLD of Y and by proposition 3.5 it is the only QSD. Lastly, by the Perron-Frobenius theorem and equation (5.2) it comes that $\vartheta = \vartheta(\nu)$. ■

In particular, we notice that a QSD allocates a positive weight to every state in X_{tr}.

5.2 A stochastic SIS model

We will now consider a basic SIS model with a fixed population. The results of this part are mostly from [7].

We denote by $S = (S_t)_{t\in\mathbb{R}_+}$ the process of the number of susceptible people and by $I = (I_t)_{t\in\mathbb{R}_+}$ the process of the number of infected (and infectious). They both take values in $\mathbb{N}$ and we assume that there exists $N \in \mathbb{N}^*$ such that

$$\forall t \in \mathbb{R}_+, S_t + I_t = N \ . \tag{5.4}$$

Therefore, it is sufficient to study I only to completely describe the mechanism. Both are Feller processes and if we consider I then it verifies assumption 1: the set $\{0\}$ is absorbing. Indeed, since there are no non-human hosts allowed, then when there is no infected left the spread ceases. The instantaneous transition rates (Cf definition 4.2) are given by

$$\forall i \in \{1,\dots,N\}, \quad q(i,i+1) = \beta i \frac{N-i}{N}, \quad q(i,i-1) = \gamma i, \quad q(i,i) = -i\left(\gamma + \beta\frac{N-i}{N}\right) \tag{5.5}$$

where $\beta, \gamma > 0$. Moreover, equation (4.10) still holds: all the other transition rates are zero. The logic behind those equations is the same as already explained in 1.2.

From theorem 5.2 we know that there is a single QSD ν which is the Yaglom limit and that the convergence of the conditional distribution toward ν is geometric. The question is then to describe ν. When N is small, its computation as a left-eigenvector is tractable, but in a larger population we have to approximate it.

In that case we can still apply theorem (4.17) that we had for birth-and-day processes, which gives us the equations

$$\forall i \in \{1,\dots,N\}, \quad \beta(i-1)\frac{N-i+1}{N}\nu_{i-1} - i\left(\gamma + \beta\frac{N-i}{N}\right)\nu_i + \gamma(i+1)\nu_{i+1} = -\gamma\nu_1\nu_i$$

with the convention that $\nu_0 = \nu_{N+1} = 0$. Approximations given in the literature often start from those equations and there are sometimes several alternative approximations for the same model. Crucially, we make use of the *basic reproductive ratio* $R_0 = \beta/\gamma$. Any approximation is given for a certain interval of R_0. When R_0 is significantly smaller than 1 it is said to be *subcritical*, *supercritical* when $R_0 \gg 1$, and the interval of values close to 1 is a transition region.

The previous equations written as a function of R_0 gives us

$$\forall i \in \{1,\dots,N\}, \quad R_0(i-1)\frac{N-i+1}{N}\nu_{i-1} - i\left(1 + R_0\frac{N-i}{N}\right)\nu_i + (i+1)\nu_{i+1} = -\nu_1\nu_i \tag{5.6}$$

so that the dependence on the basic reproductive ratio appears clearly.

Thus, in [7] it is found that a geometric distribution is the best fit for a QSD in the subcritical region, and a beta-binomial distribution in the supercritical region. We do not review the techniques which are used to obtain those analytical approximations, but in chapter 7 we will conduct simulations on this model in the transition region.

Furthermore, notice that results regarding processes on finite spaces still hold if we consider processes on a two-dimension bounded discrete space. This is the case, in particular, of the SEIS model analysed in [13]. In this model there is a transition phase when people are “exposed”, that is, after that they get infected but before that they become infectious themselves. If we still have N constant then there are three random variables and one linear equation, which leaves us with a two-dimension state space to describe the process.

Chapter 6

Endemic diseases in a dynamic population

In the case of long infectious periods, the dynamic of the population can become significant in the model that we use. These are called *endemic diseases*. Notable examples are measles and HIV. Depending on the disease and the situation, we might want to consider either the inner dynamic demography of the population, or migrations.

Naturally, in the literature, the models are not only studied through quasi-stationary distributions. In particular, the analysis of the equivalent deterministic model is generally done as it also yields some valuable information. In this dissertation though we will focus on the stochastic part. The deterministic models, which have been introduced in section 1.2 are kept out of the scope of the dissertation.

6.1 Introduction to dynamic models

As the mechanisms are more complex now, a model too specific is likely to be intractable, but simple models can give good insights on the evolution of a disease. A notable difficulty, especially when it comes to computations, is that the description of the process requires a space with two dimensions. Indeed, the number I of infected and S of susceptibles are not linearly related anymore.

Of course, results given in section 4 are still applicable since the state space $\mathsf{X} = \mathbb{N}^2$ is countable. However, we need to adapt the notation to the form (S, I). For clarity we will denote $\forall s_1, i_1, s_2, i_2 \in \mathbb{N}$, $q_{s_1,i_1}(s_2, i_2) := q((s_1, i_1), (s_2, i_2))$ where the first entry of each pair is the number of susceptible people and the second is the number of infectious ones.

This dimensionality issue can get even worse. For instance, the study of the spread of HIV in [23] distinguishes between male and female among both susceptible and infectious people, which amounts to four different categories. As they also consider the case of an open population, there is no linear relation between them, so that it is a problem in four dimensions.

A major reference in the literature about epidemic models is [4] by Bartlett, written in 1956. In this paper he studies a kind of SIR model with immigration, known as the Hamer-Soper model.

The transition rates for all $i, j \in \mathbb{N}$ are given by

$$q_{s,i}(s-1, i+1) = \beta i s, \quad q_{s,i}(s, i-1) = \gamma i, \quad q_{s,i}(s+1, i) = \eta \tag{6.1}$$

with $-q_{s,i}(s,i)$ is the sum of those three and other transition rates are equal to zero.

There are several differences with the model given in 5.2. One of the most important is that after infection people don't become susceptible again, which we see from the fact that $q_{s,i}(s+1, i-1) = 0$ but $q_{s,i}(s, i-1) > 0$. They either die, or recover and develop a permanent immunity. In both cases we don't need to count them in I or S anymore.

On the other hand we have $q_{s,i}(s+1, i) > 0$ provided that $\eta > 0$. Therefore, the population is likely to grow and since the rate is independent from the number of people already in the population it can be interpreted as immigration. Thus, the three transitions given correspond respectively to an infection, a removal and the arrival of someone.

We can observe that this model doesn't take into account people in the susceptible category who die naturally. That is, the rate $q_{s,i}(s-1, i)$ could be set positive in a more elaborated model.

The absorbing set X_{ab} is given by $\mathbb{N} \times \{0\}$, when there is no infective left, and it is clear from the transition rates that Q_{tr} is irreducible. However, the absorbing set differs in other models as we will see in section 6.4.

The approximation of a quasi-limiting distribution in that case has been later studied in several papers. Nevertheless, the existence of a QSD has not been formally proved. In spite of this, we can at least ensure that the process satisfies some of our basic assumptions, which is already useful information in practical cases, and allow us to look for quasi-stationary distributions. This is the purpose of the next two sections, which exploit the peculiarities of the processes similar to the Hamer-Soper model.

6.2 Determining Feller processes

We have seen in chapter 4 that when Y is a Feller process we can draw various results which then enable us to look for quasi-stationary distributions. In the case of our models on $\mathbb{N}^2$, defined through their transition rates, we need to prove first that the resulting process is Feller.

Recall that, for any operator A represented by a matrix $(a_{x,y})_{x,y\in\mathsf{X}}$ we denote,

$$Af(x) = \sum_{y\in\mathsf{X}} a_{x,y} f(y) \quad \text{and} \quad fA(x) = \sum_{y\in\mathsf{X}} a_{y,x} f(y)$$

for any $x \in \mathsf{X}$ and function f such that those quantities are well-defined.

The following propositions from [2] gives us a way to verify that a process defined by its transition rate matrix is Feller. We give the essential arguments of the proofs without going into much details. Indeed, we need theorems which rely on a deeper analysis of the resolvents but there is no point in providing this here. As in chapter 2, the λ-resolvent of (P_t) is denoted by R_λ.

Proposition 6.1. *Let $(P_t)_t$ be a transition function on X. Then (P_t) is Feller if and only if $\forall \lambda > 0, \forall f \in C_0(\mathsf{X}), R_\lambda f \in C_0(\mathsf{X})$.*

The interpretation of this proposition is that, if there is some kind of order on X, e.g. $\mathsf{X} = \mathbb{N}$, then with $\lambda > 0$ and $y \in \mathsf{X}$ fixed we have $\lim_{x\to\infty} \varphi_{x,y}(\lambda) = 0$ where $\varphi_{x,y}$ is the Laplace transform of $t \mapsto p(t,x,y)$.

Proof. The first implication is given in general by proposition 2.15. For the second implication, suppose that R_λ preserves $C_0(\mathsf{X})$. Define S as the range of R_λ on $C_0(\mathsf{X})$.

First, we show that S is dense in $C_0(\mathsf{X})$. Suppose by contraction that it's not. Then there is a $f \in C_0(\mathsf{X})$ which is at distance $\varepsilon > 0$ of $S \subset L^\infty$. Using a result of functional analysis from [26] (theorem 3.4) we know that there exists $g \in L^1$ such that $\langle f \mid g \rangle = \varepsilon$ and $\forall h \in S, \langle g \mid h \rangle = 0$. The latter can be written as gR_λ. But R_λ is a one-to-one operator on L^1 so this would mean that $g = 0$, which is impossible.

By the Hille-Yosida Theorem (theorem 4.1 in [2]), there exists a unique continuous contraction semigroup on $C_0(\mathsf{X})$ corresponding to the collection $\{R_\lambda, \lambda > 0\}$. The uniqueness of the Laplace transform gives us that this semigroup is equal to $(P_t)_t$ and therefore $(P_t)_t$ is Feller. ■

Theorem 6.2. *Let $(P_t)_t$ be a transition function with transition rate matrix Q satisfying the forward equation (4.6). Suppose that Q satisfies the following conditions:*

(i) *for every $y \in \mathsf{X}$, the function $Q\mathbf{1}_{\{y\}} : x \mapsto q(x,y)$ is in $C_0(\mathsf{X})$,*

(ii) *for any $\lambda > 0$, the equation $\lambda f = fQ$, with $f \in L^1$ has no solution other than $f = 0$.*

Then, $(P_t)_t$ is a Feller transition function.

Proof. Let $\mathscr{D}$ be the set of functions in $C_0(\mathsf{X})$ with a finite support. It is straightforward to see that $\mathscr{D}$ is dense in $C_0(\mathsf{X})$. Moreover, it is clear from condition *(i)* that $f \mapsto Qf$ maps $\mathscr{D}$ into $C_0(\mathsf{X})$. Then, let S denote the range of the operator $f \mapsto (\lambda I - Q)f$ on $\mathscr{D}$. We will show that S is also dense in $C_0(\mathsf{X})$.

Suppose that it is not the case. With the same arguments as in the proof of proposition 6.1 it comes that there exists $g \in L^1 \setminus \{0\}$ such that $g(\lambda I - Q) = 0$ where I denotes the identity operator. But then condition *(ii)* implies that $h = 0$, which contradicts $\langle f \mid g \rangle = \varepsilon$.

Let now R_λ be the λ-resolvent of $(P_t)_t$. Then, using the forward equation (4.6) we get

$$R_\lambda(\lambda I - Q) = \int_0^\infty e^{-\lambda t} P_t(\lambda I - Q)\,\mathrm{d}t = \int_0^\infty \frac{\mathrm{d}}{\mathrm{d}t}\left(e^{-\lambda t} P_t\right)\mathrm{d}t = I$$

which implies that, $\forall s \in S, R_\lambda s \in \mathscr{D}$. We also know that R_λ is a continous operator since it is linear and bounded. Therefore, since S is dense in $C_0(\mathsf{X})$, R_λ maps $C_0(\mathsf{X})$ into $C_0(\mathsf{X})$. Then, the result follows from proposition 6.1. ■

Moreover, it turns out that condition *(ii)* guarantees the uniqueness of the transition function, which is also known as the "minimal transition function" and satisfies the forward equation (Cf theorems 2.2.7 and 2.2.8 in [2]).

6.3 Processes with sure killing

With the results of the previous section and of section 4.2 we can assume that our model describes a process with a regular transition rate matrix Q. Then, it remains to verify that the assumption 2 of sure killing is satisfied (as assumption 1 is generally obvious from the transition rates).

The dynamic models that we use for epidemics fall in the category of competition models, which have a two dimensional state space. Competition models are themselves a subspace of *multidimensional population model* which occur in a d-dimensional state space to describe the evolution of a population with d compartments, or equivalently of d populations at different places with immigration between one another.

The theorems used to assess the sure killing condition can be found in [20] and [21] by Reuter. He considers competition processes, but does not always state and prove very clearly his results. The more recent book [2] by Anderson gives a more satisfying version of it, in the case of multidimensional population models which, have we have seen, are a bit broader but very similar theoretically.

The formal definition of a competition process as found in [20] is the following.

Definition 6.3. Let $\mathsf{X} = \mathbb{N}^2$ and $Q = (q_{i,j}(k,l))_{i,j,k,l\in\mathbb{N}}$ be a transition rate matrix on X. Then the corresponding Markov process is said to be a *competition process* if there exist functions a, b, c, d, e, f on $\mathbb{N}^2$ such that, for every $(i, j) \in \mathbb{N}^2$,

$$\begin{aligned} q_{i,j}(i+1,j) &= a(i,j) & q_{i,j}(i,j+1) &= b(i,j) \\ q_{i,j}(i-1,j) &= c(i,j) & q_{i,j}(i,j-1) &= d(i,j) \\ q_{i,j}(i-1,j+1) &= e(i,j) & q_{i,j}(i+1,j-1) &= f(i,j) \end{aligned}$$

with $-q_{i,j}(i,j)$ equal to the sum of those rates and other transition rates from (i,j) equal to zero.

Remark. Since there are no states with negative coordinates, for any competition process we have

$$\forall j \in \mathbb{N}, c(0,j) = e(0,j) = 0 \quad \text{and} \quad \forall i \in \mathbb{N}, d(i,0) = f(i,0) = 0\,.$$

A consequence of this is that, for any $x = (i, j) \in \mathsf{X}$ we have $|\mathsf{A}(x)| \leqslant 6$ where $\mathsf{A}(x)$ is the set of states accessible in one jump from x defined by (4.3). In particular, assumption 3 is always satisfied.

We can then extend the domain of the generator using the formula of proposition 4.4:

$$\forall x \in \mathsf{X}, \quad Lf(x) := \sum_{y\in\mathsf{A}(x)} q(x,y)f(y) \tag{6.2}$$

for any function $f : \mathsf{X} \to \mathbb{R}$. Thus, Lf is clearly well defined everywhere.

The general theorem that we can use to assess whether there is sure killing of a process is the following.

Theorem 6.4. *Let Y satisfy assumption 1 and suppose its transition rate matrix Q to be regular. If, for some function $\varphi : \mathsf{X} \to \mathbb{R}_+$, we have $L\varphi + 1 \leqslant 0$ on X_{tr} then, Y satisfies assumption 2 of sure killing and $\forall x \in \mathsf{X}_{\mathrm{tr}}, \mathbb{E}_x[T] \leqslant \varphi(x) < \infty$.*

We omit the proof of this theorem, which is quite long and is given by both Reuter and Anderson.

As an example, we will now use the previous propositions to the subcase of competition processes where $\mathsf{X}_{\mathrm{ab}} = \{0\}$, although the results can also be applied to any X_{ab} given by a triangle, i.e. the pairs of integer whose sum of coordinates is less than l for some $l \in \mathbb{N}^*$.

We denote $\forall n \in \mathbb{N}, \mathsf{X}_{\text{tr}}^{(n)} := \{(i,j) \in \mathsf{X}_{\text{tr}} \mid i+j=n\}$. Then, we use the following notation, which is morally similar to birth rate and a death rate as we used earlier: for any $n \in \mathbb{N}^*$,

$$\lambda_n := \max\left\{a(i,j)+b(i,j) \mid (i,j) \in \mathsf{X}_{\text{tr}}^{(n)}\right\}$$
$$\text{and} \quad \mu_n := \min\left\{c(i,j)+d(i,j) \mid (i,j) \in \mathsf{X}_{\text{tr}}^{(n)}\right\} .$$

Moreover, we shall assume that for any $n \geqslant 1$ we have $\lambda_n > 0$ and $\mu_n > 0$.

We can then state under which conditions those processes are regular and absorbed with probability one. We will skip the proofs as well, which can be found in [2], as those propositions are essentially consequences of theorems 6.2 and 6.4 respectively.

Proposition 6.5. *Let Q be the transition rate matrix of a competition process absorbed at* $\{0\}$. *If,*

$$\sum_{n=1}^{\infty}\left(\frac{1}{\lambda_n}+\frac{\mu_n}{\lambda_n\lambda_{n-1}}+\ldots+\frac{\mu_n\cdots\mu_2}{\lambda_n\cdots\lambda_1}\right)=+\infty$$

then Q is Feller.

Proposition 6.6. *Let Q be the transition rate matrix of a competition process and suppose that it is Feller. If*

$$\sum_{n=1}^{\infty}\frac{\lambda_n\cdots\lambda_1}{\mu_{n+1}\cdots\mu_2}<+\infty$$

then assumption 2 is satisfied and $\forall x \in \mathsf{X}_{\text{tr}}, \mathbb{E}_x[T] < \infty$.

6.4 Model with inner dynamic demography

We will be interested here in a model described in [8] and still similar to the one given by (6.1), but now the people added to the population come result from birth in the population. Its transition rates are given by

$$q_{s,i}(s-1,i+1)=\beta is, \quad q_{s,i}(s,i-1)=\gamma i, \quad q_{s,i}(s+1,i)=\eta s \tag{6.3}$$

where, again, $-q_{i,s}(i,s)$ is equal to the sum of the three transitions given and all other transition rates are equal to zero. We suppose that β,γ,η are all positive. Since the birth mechanism is dependent on s only, it is more logical to assume that people removed from the infected population are dead and did not recover, to explain that they play no part in the breeding.

The state space is again $\mathsf{X}=\mathbb{N}\times\mathbb{N}$ but we have two absorbing sets now: $\{0\}\times\mathbb{N}$ and $\mathbb{N}\times\{0\}$. The difference compared to the model of section 6.1 is that, before, even when everyone was infected, new susceptible people could arrive through immigration. But in the new model, since infected people don't give birth, when everyone is infected the population can only shrink with the progressive removal of individuals until there is no one is left.

Therefore we set $\mathsf{X}_{\text{ab}}=(\{0\}\times\mathbb{N})\cup(\mathbb{N}\times\{0\})$ and $\mathsf{X}_{\text{tr}}=\mathbb{N}^*\times\mathbb{N}^*$. It is then straightforward to verify that Q_{tr} is irreducible. Using the results of sections 6.2 and 6.3, we can show that the process is Feller and goes almost surely to extinction. However, the authors of [8] identify no analytical solution to the QSD equation. Therefore, we need to approximate the quasi-stationary distributions, using numerical methods in particular.

Chapter 7

Numerical approximations

We present some of the most usual computational techniques that we use to work on quasi-stationary distributions.

There are several kinds of computations that we want to perform. One is to simply simulate a jump process given its transition rates, which is addressed in 7.1 and is not specific to quasi-stationary distributions, another is to find or approximate quasi-stationary distributions for those processes, which is addressed in 7.2.

7.1 Simulating jumps

Let $\tau > 0$. We want to simulate a jump process Y with jump rates denoted by $q(x)$ and transition rates denoted by $q(x, y)$ for x, y in the state space. To be represented by the computer we need to restrict the simulation to a certain set of points in time. The most natural way is to choose a linear scale, that is points $k\tau$ with $k \in \{1, \dots, M\}$ for some $M \in \mathbb{N}$. This section is essentially original work.

Let $\mathscr{U}([0,1])$ denote the uniform distribution on $[0,1]$. A well-known property often used to model such processes is that, if we denote by X a random variable and by F_X its cumulative distribution function, we can take $U \sim \mathscr{U}([0,1])$ and simulate X with

$$X = F_X^{-1}(U) \; . \tag{7.1}$$

Applying equation (7.1) to a random variable following $\mathscr{E}(\lambda)$, the exponential distribution with parameter $\lambda > 0$, we get

$$-\frac{1}{\lambda}\ln(U) \sim \mathscr{E}(\lambda) \tag{7.2}$$

where $U \sim \mathscr{U}([0,1])$.

This yields a first method to achieve a numerical simulation of a jump process. At every step k we draw two uniform random numbers. The first is for the time $T_k - T_{k-1}$, using equation (7.2) with parameter $q(Y_{T_{k-1}})$. The second one is for the change of state, using equation (7.1) with the distribution given by $K(Y_{T_{k-1}}, \cdot)$, where K is the kernel defined by equation (4.7).

This last distribution can be written in terms of transition rates

$$\forall x,y \in \mathsf{X}, \quad K(x,\{y\}) = \begin{cases} \mathbf{1}_{\{x\}}(y) & \text{if} \quad x \in \mathsf{X}_{ab} \\ 0 & \text{if} \quad x \in \mathsf{X}_{tr} \quad \text{and} \quad x = y \\ q(x,y)/q(x) & \text{if} \quad x \in \mathsf{X}_{tr} \quad \text{and} \quad x \neq y \end{cases}.$$

Note that this cuts the process at the killing time T. So, what we are really simulating is Y^T, considering that we are only interested in what happens before the killing time.

Then, knowing what all the jumps are, for any given $k \in \mathbb{N}$ we can find in which state the process is at time $k\tau$. The plot of figure 1.1 was obtained by this method and represents a logistic birth-and-death model with parameters $\lambda = \frac{1.05}{4}$, $\mu = \frac{1}{4}$, $c = \frac{1}{8000}$. We give below the code of the two functions in R giving respectively the time before the next jump, and the next state. The call to `br(n)` returns λ_n while `dr(n)` returns μ_n.

```
timeNext <- function(n) {
    if(n == 0) {
        return(1)
    }
    return(rexp(1, br(n) + dr(n)))
}

nextState <- function(n) {
    if(n == 0) {
        return(0)
    }
    if ( runif(1) < 1/(1 + br(n)/dr(n)) ) {
        return(n - 1)
    } else {
        return(n + 1)
    }
}
```

We also used this method to plot simulations of the SIS process with a dynamic population described in section 6.4. We used the same parameters as in [8]: $\beta = 0.01$, $\gamma = 0.5$ and $\eta = 0.5$ but with a different starting point which is now $(100, 10)$, i.e. 100 susceptible people and 10 infectious ones. Figure 7.1 represents four simulations of this process, where we only show the number of infectives, but the total size of the population changes as well. The cyclical patterns that we observe are similar to those of [8].

A second method, that we can use in the finite case, is to use the equation $P_\tau = e^{Q\tau}$ where $(P_t)_t$ is the transition function of Y. At the first step, we can draw Y_0 using equation (7.1) if we know the initial distribution. Then, denoting by δ_{Y_0} the Dirac distribution in Y_0 we can draw Y_τ with equation (7.1) knowing that its distribution is given by $\delta_{Y_0}^{\mathsf{T}} P_\tau$. We can then iterate the process: at each step, knowing $Y_{k\tau}$ we draw $Y_{(k+1)\tau}$ following distribution $\delta_{Y_{k\tau}}^{\mathsf{T}} P_\tau$.

In case of an infinite state space we could use this with a truncature of Q. This would presumably return a decent approximation (only extreme cases are rejected) provided that τ is small and that the size of the troncature is big enough.

Using this second technique we modelled the SIS process described in section 5.2. Figure 7.1 features four simulated sample paths of this process. The parameters used are: $N = 400$,

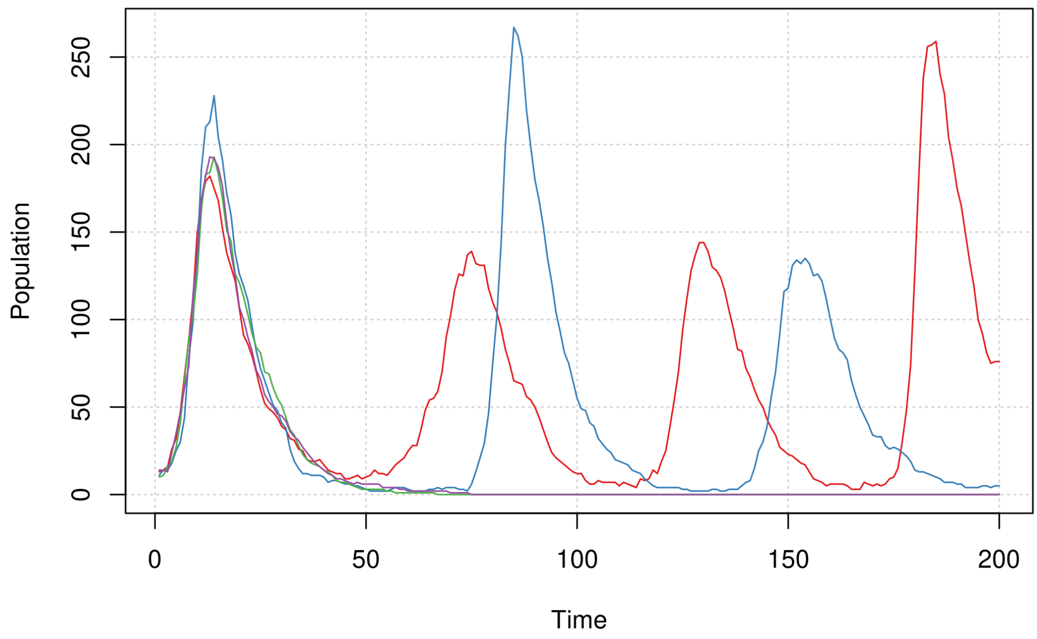

Figure 7.1: Simulations of the SIS process of equation (6.3)

$\beta = 0.55$, $\gamma = 0.5$ and the disease starts with ten infectives at time 0.

7.2 Finding quasi-stationary distributions

We consider that the matrix Q_{tr} of transition rates between transient states is irreducible, which holds in the applications presented here. In cases when it is not, we have to consider several classes of states to exclude those which cannot be accessible from a QSD, which is a more complex problem. The sources used in this section are [28] and [8].

If the state space is finite equal to $\{0, \ldots, N\}$, then we can directly compute the eigenvector of R_1^{T}, where R_1 is the submarkovian restriction of P_1 to X_{tr}. As we have seen, this eigenvector should have all its entries of the same sign and non-zero since Q_{tr} is supposed irreducible. Then, we normalize it so that the eigenvector $\nu \in \mathbb{R}^N$ satisfies $\sum_i \nu_i = 1$.

This is the easiest case and we can compute it in R by:

```
R1 = expm(Qtr)
ev <- eigen(t(R1))
theta = -log(Re(ev$values[1]))
nu <- Re(ev$vectors[,1])
nu <- abs(nu / sum(nu))
```

where `theta` is the extinction rate and `nu` is the QSD.

Figure 7.2 represents the QSD that we get with the SIS model defined in section 5.2, with the same parameters that we used in the previous section. In particular, we can note that the path of the one process which survives for a long time in figure 7.1 takes the values in the range predicted by the QSD.

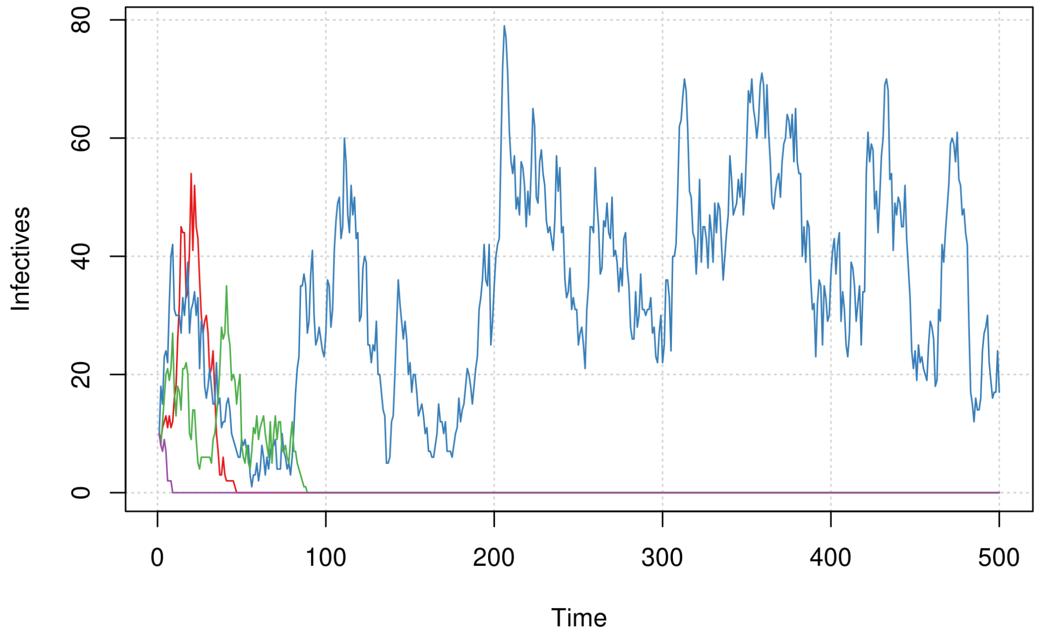

Figure 7.2: Simulations of the SIS process of equation (5.5)

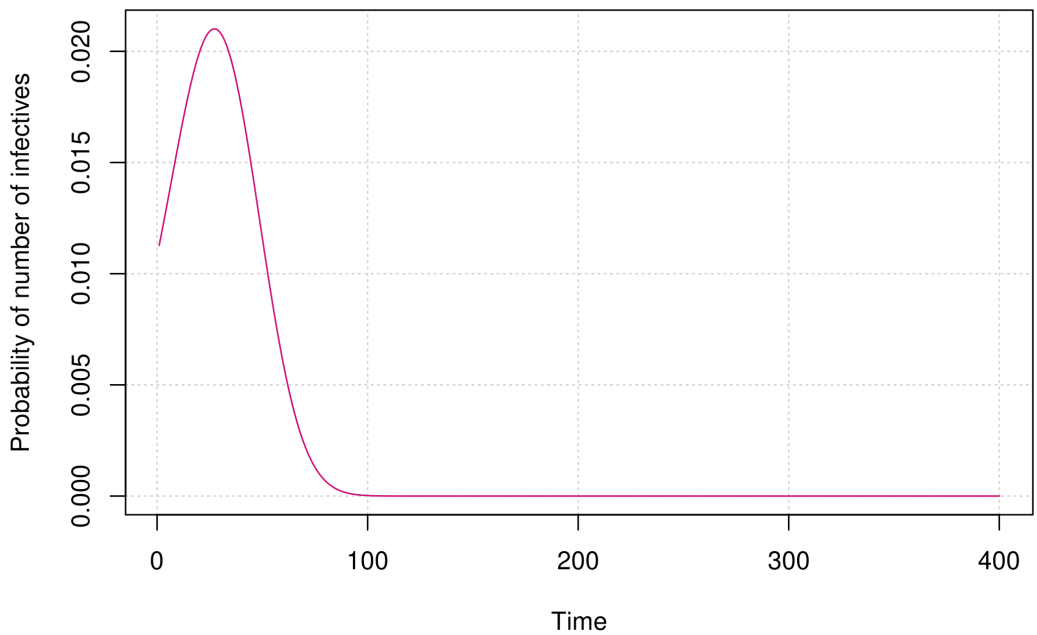

Figure 7.3: A QSD for a model of the SIS process defined by (5.5)

If the state space is infinite and so is Q, we use a truncation procedure, approximating Q by a sequence $(Q^{(k)})_{k\in\mathbb{N}^*}$ of irreducible finite square matrices. We exploit the corresponding R_1 matrices to compute a sequence $(\nu^{(k)})$ of normalized left eigenvectors and we take their limit.

Thus, some previous analytical approximation can guide us in order to find the size at which we truncate the matrix. Then, the decay parameters $(\vartheta(\nu^{(k)}))_k$ will converge to $\vartheta(\nu)$ but the convergence of $(\nu^{(k)})$ is not guaranteed.

Finally, a solution is to look for quasi-limiting distributions by simulating the process many times, following the first method given in section 7.1. We have seen in section 5.1 that the convergence is fast in the finite case, and most of the models studied in the literature have a reasonably fast convergence as well.

The algorithm goes as follows: we generate the process n times with the same starting point. We get rid of those which are absorbed before the end of the time window. For those which remain, we may cut a first period of time for each in order to keep only the part where the conditional distribution should be close to the QLD. Then, for every state $x \in \mathsf{X}_{\text{tr}}$, we count how many times over all the simulations this state is reached. Then, we simply have to normalize in order to get the quasi-limiting distribution if it exists.

Note that the second method, by calculation or approximation of P_1, would be correct as well. But in that case we are interested in every value that the processes take, and not just in where they are at some points of time. Therefore, it is better to calculate every jump, all the more that it requires less computation for the same number of samples.

A clear limitation of that method is that, when we have a multi-dimensional process this stastistical can require many simulations to give a satisfying result. For instance, with a two-dimensional epidemic model (S, I), the number of simulations to maintain a precision will grow as N^2 when we look at patches from 1 to N for each component.

Chapter 8

Conclusion

In this dissertation, we summarised results from scattered sources with the objective of providing an accessible text for non-specialists. We did so through a gradual specialization meant to provide a theoretical environment consistent from the beginning to the end of this work.

The publications [9], [17] which were our main sources on the general theory of quasi-stationary distributions can prove difficult for the novice. Therefore, we started by placing quasi-stationary distributions in the theory of Markov processes whose most useful elements were introduced in chapter 2. This allowed the dissertation to be self-contained in its most crucial parts. We can also noted that those two sources given here suffer from some flaws that we corrected here.

Then, we proceeded to the transition to discrete spaces. Again, as a complement to papers on quasi-stationary distributions, which naturally do not cover the theory of jump processes, we explain how they can be built from the theory already developed in the previous chapters. A higher level of exposition is possible on the construction of those processes, but the one given here is sufficient to understand them in the context of quasi-stationary distributions. In particular, we made the choice to consider Feller processes all along, and see how they satisfy basic assumptions. In applied papers, processes are almost never identified as Feller and various regularity properties are considered instead, which are often similar. On the contrary, we tried to highlight the fact that the theorems used in our applications are still strongly connected to the theory of Feller processes and to the more abstract properties of the first chapters.

Thus, starting from basic assumptions we reviewed some interesting results on the existence of quasi-stationary distributions and the properties of corresponding processes. The process was then reversed when we considered birth-and-deaths processes and epidemics. We enhanced the fact that it was necessary to then check that the assumptions of regularity and sure killing were satisfied. A significant part of chapters 5 and 6 was dedicated to this, although it is a mere preliminary to the study of quasi-stationary distributions and it could have been more interesting to favor the core of the subject.

Nevertheless, those verifications are necessary in the context of this dissertation to give a good sense of what a rigorous study of quasi-stationary distributions is. Moreover, we can note that there is a lack of considerations for this in the literature. Papers which stay theoretical make those assumptions directly while oftentimes the more applied ones do not give satisfying proofs of it. Yet, there are some clear ways to carry on those proofs, and the book [2] by Anderson in particular contains many useful theorems and calculation techniques to do so.

Overall, we provided in chapters 5 and 6 a review of results regarding epidemics and theorems which can be used generically to study competition processes which admit quasi-stationary distributions. Nonetheless, we did not, as was initially considered, bring new results about those models. This is mostly because the in-depth study of any model does not rely only on our theory in quasi-stationary distributions. It is generally paired with a larger analysis, including notably the deterministic model, which would have brought us too far from the subject of this dissertation.

Yet, the results that we gave are certainly of use to proceed to the analysis of a stochastic model. There are other ways that we could have treated this part, in adapting them to other sub-cases for example.

Finally, we reviewed different techniques used to simulate processes and approximate quasi-stationary distributions. We could also have been further in this direction in a longer work, and compare the efficiency of different methods on a range of competition processes. The performance of algorithms are rarely studied in the literature on quasi-stationary distributions. Yet, we can tell that it matters, as the approximations made by researchers are most of the time at a small scale, because it is computationally difficult to do bigger. Therefore, an improvement on that matter would probably allow more accurate simulations of epidemics.

Bibliography

[1] Christian L Althaus. “Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa”. In: *PLoS currents* 6 (2014).

[2] William J Anderson. *Continuous-time Markov chains: An applications-oriented approach*. Springer-Verlag, 1991.

[3] Hakan Andersson and Tom Britton. *Stochastic epidemic models and their statistical analysis*. Vol. 151. Springer Science & Business Media, 2012.

[4] Maurice S Bartlett. “Deterministic and stochastic models for recurrent epidemics”. In: *Proceedings of the third Berkeley symposium on mathematical statistics and probability*. Vol. 4. 81. 1956, p. 109.

[5] Maurice S Bartlett. *Stochastic population models; in ecology and epidemiology*. Tech. rep. 1960.

[6] Richard F Bass. *Jump Processes*. 2014. url: http://bass.math.uconn.edu/jump-processes.pdf (visited on 07/14/2018).

[7] Damian Clancy and Sang Taphou Mendy. “Approximating the quasi-stationary distribution of the SIS model for endemic infection”. In: *Methodology and Computing in Applied Probability* 13.3 (2011), pp. 603–618.

[8] Damian Clancy, Philip D O’Neill, and Philip K Pollett. “Approximations for the long-term behavior of an open-population epidemic model”. In: *Methodology and computing in Applied Probability* 3.1 (2001), pp. 75–95.

[9] Pierre Collet, Servet Martínez, and Jaime San Martín. *Quasi-stationary distributions: Markov chains, diffusions and dynamical systems*. Springer Science & Business Media, 2012.

[10] Centers for Disease Control, Prevention, et al. *History and epidemiology of global smallpox eradication*. 2012. url: https://stacks.cdc.gov/view/cdc/27929.

[11] Athina Economou, A Gómez-Corral, and M López-García. “A stochastic SIS epidemic model with heterogeneous contacts”. In: *Physica A: Statistical Mechanics and its Applications* 421 (2015), pp. 78–97.

[12] Pablo A Ferrari et al. “Existence of quasi-stationary distributions. A renewal dynamical approach”. In: *The annals of probability* (1995), pp. 501–521.

[13] Carlos M Hernández-Suárez et al. “An application of queuing theory to SIS and SEIS epidemic models”. In: *Math. Biosci. Eng* 7.4 (2010), pp. 809–823.

[14] Achim Klenke. *Probability theory: a comprehensive course*. Springer Science & Business Media, 2013.

[15] Jean-François Le Gall. *Brownian motion, martingales, and stochastic calculus*. Vol. 274. Springer, 2016.

[16] Servet Martínez. "Notes and a remark on quasi-stationary distributions". In: *Monogr. semin. mat. García Galdeano* 34 (2008), pp. 61–80.

[17] Sylvie Méléard and Denis Villemonais. "Quasi-stationary distributions and population processes". In: *Probability Surveys* 9 (2012), pp. 340–410.

[18] Sidney A Morris. *Topology without tears*. University of New England, 1989.

[19] Philip K Pollett. *Quasi-stationary distributions: then and now.*

[20] Gerd EH Reuter. "Competition processes". In: *Proc. 4th Berkeley Symp. Math. Statist. Prob.* Vol. 2. 1961, pp. 421–430.

[21] Gerd EH Reuter. "Denumerable Markov processes and the associated contraction semigroups onl". In: *Acta Mathematica* 97.1-4 (1957), p. 1.

[22] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*. Vol. 293. Springer Science & Business Media, 2013.

[23] A Sani, DP Kroese, and Philip K Pollett. "Stochastic models for the spread of HIV in a mobile heterosexual population". In: *Mathematical biosciences* 208.1 (2007), pp. 98–124.

[24] Christof Schuette and Philipp Metzner. *Markov Chains and Jump Processes. An Introduction to Markov Chains and Jump Processes on Countable State Spaces.* Freie Universitat Berlin & DFG Research Center Matheon, Berlin, July 2009.

[25] Eugene Seneta. *Non-negative matrices and Markov chains*. Springer Science & Business Media, 2006.

[26] Angus E Taylor and David C Lay. *Introduction to functional analysis*. 2nd ed. 1980.

[27] Erik A Van Doorn. "Quasi-stationary distributions and convergence to quasi-stationarity of birth-death processes". In: *Advances in Applied Probability* 23.4 (1991), pp. 683–700.

[28] Erik A Van Doorn and Philip K Pollett. "Quasi-stationary distributions". In: (2011).

[29] D Vere-Jones. "Some limit theorems for evanescent processes". In: *Australian Journal of Statistics* 11.2 (1969), pp. 67–78.

Index