Probability and Its Applicati

Pierre Collet
Servet Martínez
Jaime San Martín

Quasi-St
Distribu

Probability and Its Applications

Published in association with the Applied Probability Trust

Editors: S. Asmussen, J. Gani, P. Jagers, T.G. Kurtz

Probability and Its Applications

The *Probability and Its Applications* series publishes research monographs, with the expository quality to make them useful and accessible to advanced students, in probability and stochastic processes, with a particular focus on:

- Foundations of probability including stochastic analysis and Markov and other stochastic processes
- Applications of probability in analysis
- Point processes, random sets, and other spatial models
- Branching processes and other models of population growth
- Genetics and other stochastic models in biology
- Information theory and signal processing
- Communication networks
- Stochastic models in operations research

For further volumes:
www.springer.com/series/1560

Pierre Collet • Servet Martínez • Jaime San Martín

Quasi-Stationary Distributions

Markov Chains, Diffusions and Dynamical Systems

Springer

Pierre Collet
Centre de Physique Théorique, UMR7644
CNRS, École Polytechnique
Palaiseau, France

Servet Martínez
CMM-DIM, Faculty of Mathematical
and Physical Sciences
University of Chile
Santiago, Chile

Jaime San Martín
CMM-DIM, Faculty of Mathematical
and Physical Sciences
University of Chile
Santiago, Chile

ISSN 1431-7028 Probability and Its Applications
ISBN 978-3-642-33130-5 ISBN 978-3-642-33131-2 (eBook)
DOI 10.1007/978-3-642-33131-2
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012951442

Mathematics Subject Classification: 60J70, 60JXX, 60J10, 60J60, 60J65, 92Dxx, 37DXX, 35PXX

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

This book is dedicated to all surviving small communities and the endangered species.

The authors P.C., S.M. and J.S.M. are indebted to Sonia Lund, María Angélica Salazar and Valeska Amador for their patience, support and encouragements along all these years.

Preface

This book is devoted to the study of the main concepts on survival for killed Markov processes and dynamical systems. The first computations on the number of survivals after a long time in branching process was done by Kolmogorov (1938). Later on Yaglom (1947) showed that the limit behavior of sub-critical branching processes conditioned to survival was given by a proper distribution. This pioneer work triggered an important activity in this field.

Quasi-stationary distributions (QSD) capture the long term behavior of a process that will be surely killed when this process is conditioned to survive. A basic and useful property is that the time of killing is exponentially distributed when starting from a QSD, which implies that the rate of survival of a process must be at most exponential in order that QSD can exist.

The study of QSD on finite state irreducible Markov chains started with the pioneering work Darroch and Seneta (1965). For these processes many of the fundamental ideas in the topic can be easily developed. In particular, the Perron–Frobenius theory for finite positive matrices gives all the required information. Thus, the exponential rate of survival is the Perron–Frobenius eigenvalue. The QSD is the Perron–Frobenius normalized eigenmeasure, and the chain of trajectories that are never killed (Q-process) is governed by an h-process where h is the Perron–Frobenius eigenfunction. This relation between a QSD and the kernel of the Q-process is also encountered in many other processes. In the finite case, the fact that the eigenvalues are isolated simplifies the study. Since the dominant eigenvalue is simple, the QSD attracts all the conditioned measures, and is the quasi-limiting distribution, or the Yaglom limit. When the transition kernel is symmetric, the spectral decomposition contains all the data that allow one to understand the survival phenomenon of the chain.

For general countable state Markov chains, this study started in Seneta and Vere-Jones (1996). We give a detailed proof of the characterization of QSDs as a finite mass eigenmeasure due to Pollett, and we give a more general result on the existence of QSDs, which states that if there is no entrance from infinity then exponential survival is a necessary and sufficient condition to have a QSD. We follow the original proof done in Ferrari et al. (1995a). We study also in details the symmetric chains and monotone chains. We also give conditions in order that the exponential decay rate be equal to the Kingman decay parameter of the transition probability.

An important literature has been devoted to time-continuous birth-and-death chains in the last 25 years, many of these works take their basis in the spectral representation given in McGregor (1957a, 1957b). We revisit some of these results. An important problem appears because under exponential survival, the set of QSDs can be a continuum or a singleton. We give the criterion, due to Van Doorn (1991), to identify which case occurs. This is done in terms of the parameters of the chain. When there is a continuum of QSDs one can identify the Yaglom limit, it corresponds to the minimal QSD This result is due to Good (1968). On the other hand, the spectrum can be continuous, and the fact that, in general, there is no spectral gap raises a delicate technical problem. But a useful point is that the bottom of the spectrum is the exponential rate of survival. For birth-and-death chains, we derive the classification of the associated Q-process (as transient, recurrent or positive recurrent) by studying the exponential asymptotic rate function of the survival probability when the chain is restricted to be above some barrier. For some particular birth-and-death chains, we identify explicitly the QSD and other significant parameters and properties.

We present some results for one dimensional diffusions on the half-line killed at the origin. We study the problem of QSD for diffusions that have nice behavior at 0 and at ∞. In this framework, we also study the Q-process and its recurrence classification in relation with spectral properties of the generator of the initial diffusion. We also consider one dimensional diffusions that present singularities at 0 and/or at ∞, and we study for them the problems of existence and properties of QSD. This generalization is motivated by some models in mathematical ecology. These chapters on one dimensional diffusions find their source of inspiration in the work of Mandl (1961). The main tools come from the spectral theory of the Sturm–Liouville operators. The study of QSD in population dynamics is a very active topic nowadays, we include some of the results in the chapters on birth-and-death chains and diffusions.

We also consider the case of dynamical systems (with discrete time evolution) with a trap in the phase space. A QSD in this context deals with trajectories which do not fall into the trap. There are many analogies with the stochastic case concerning the questions, the results and their proofs. The QSDs for dynamical systems have been first studied in the context of expanding systems in Pianigiani and Yorke (1979). We discuss in particular Gibbs QSDs for symbolic systems and absolutely continuous QSDs for repellers. For both cases, we use a similar technique of proof based on quasi-compact operators analogous to those used in the thermodynamic formalism.

One of the main objects in our study is the associated Q-process, that is, the process having trajectories that survive forever, as well as its relations with the QSD and the spectral properties. For a branching process, the Q-process was introduced by Spitzer (unpublished) and in Lamperti and Ney (1968). In this book, we study the Q-process and describe it as an h-process with respect to the eigenfunction h having the same eigenvalue as the minimal QSD. This is done for Markov chains, one-dimensional diffusions and expanding dynamical systems, giving a unified approach. Thus, it appears that the Gibbs invariant measure for topological Markov chains described by Bowen–Ruelle–Sinai can be interpreted as the stationary measure of a Q-process associated to a process that is killed in a region having the Pianigiani–Yorke as its QSD.

The problem of existence of a QSD is in a sense more difficult than the problem of proving the existence of an invariant measure since one has to determine also the rate of decay of the probability of survival. This also implies that the equation for the QSD is nonlinear.

Several techniques have been used in the literature to prove existence and some of them will be explained in details in the present book. Without trying to be exhaustive, one can mention the abstract techniques relying on the Tychonov fixed point theorem (see Tychonov 1935), Krein's lemma (Oikhberg and Troitsky 2005), and the Krein–Rutman Theorem (Dunford and Schwartz 1958). Birkhoff's Theorem on Hilbert's projective metric can also be useful (see Eveson and Nussbaum 1995). Many results are based on spectral techniques, and in particular on various extensions of the Perron–Frobenius Theorem like in the theory of the Ruelle–Perron–Frobenius operator Ruelle (1978), Bowen (1975), the spectral theory of quasi-compact positive operators (see, for example, Ionescu Tulcea and Marinescu 1950 and Nussbaum 1970), and the uniform ergodic theorem (Yosida and Kakutani 1941).

The literature on QSDs is quite large and covers many fundamental and applied domains. This book does not pretend to cover them all. We have not included the developments of QSDs in all directions, for instance, we do not treat the case of branching processes where several seminal results can be found in the book Athreya and Ney (1972) and where there is a lot of active research nowadays. We have selected some of the topics which either have called the attention of specialists, or where the authors of this book have made some contributions.

This book is not intended to be a complete exposition of the theory of QSDs. It mostly reflects the (present) interest of the authors in some parts of this vast field.

In Chap. 2 dealing with general aspects of QSDs, we have mainly gathered some of our own research. We also point out that only Sects. 2.1, 2.2, 2.3 are needed to read independently the other chapters.

Paris, France — Pierre Collet
Santiago, Chile — Servet Martínez
Jaime San Martín

Acknowledgements

The authors acknowledge their institutions for the facilities given to write this book: the Center of Theoretical Physics at the Ecole Polytechnique, the Center of Mathematical Modeling and the Department of Engineering Mathematics of the Universidad de Chile, the first two being units of the CNRS. We have had the support of various institutions and programs: Fondap, Conicyt, Fondecyt, Presidential Chair, Nucleus Millennium, Basal program, Franco–Chilean cooperation, Guggenheim fellowship, Chair MMB, MathAmsud program.

Besides the common work among the authors, the origins of the book can be found at the end of the 1980s in collaborations with Pablo Ferrari and Pierre Picco, at University Sao Paulo and CPT Marseille, where we first learned on QSD and collaborated with Harry Kesten from Cornell University. Later on the collaboration with Bernard Schmitt and Antonio Galves from University of Bourgogne and University of Sao Paulo triggered our interest in the case of dynamical systems. We are deeply indebted to all of them and also to the collaboration with Patrick Cattiaux, Andrew Hart, Amaury Lambert, Véronique Maume-Deschamps, Maria Eulalia Vares and Bernard Ycart. Finally, we acknowledge Sylvie Méléard from CMAP at Ecole Polytechnique for our collaboration on QSD in population dynamics.

Contents

Chapter 1
Introduction

1.1 Quasi-Stationary Distributions

In the framework of this theory, there is a Markov process evolving in a domain where there is a set of forbidden states that constitutes a trap. The process is said to be killed when it hits the trap and it is assumed that this happens almost surely. We investigate the behavior of the process before being killed, more precisely we study what happens when one conditions the process to survive for a long time.

Let us fix some notation. The state space is $\mathcal{X}$ endowed with a measurable structure. We consider a Markov process $Y = (Y_t : t \geq 0)$ taking values in $\mathcal{X}$ and $(\mathbb{P}_x : x \in \mathcal{X})$ is the family of distributions, $\mathbb{P}_x$, with the initial condition $x \in \mathcal{X}$. As said, there is a measurable set of forbidden states which we denote by $\partial\mathcal{X}$ (sometimes we will denote this set by $\mathcal{X}^{\text{tr}}$ because it does not necessarily coincides with the boundary of $\mathcal{X}$). Its complement $\mathcal{X}^{\mathbf{a}} := \mathcal{X} \setminus \partial\mathcal{X}$ is the set of allowed states. Let $T = T_{\partial\mathcal{X}}$ be the hitting time of $\partial\mathcal{X}$. We will assume that there is sure killing at $\partial\mathcal{X}$, so $\mathbb{P}_x(T < \infty) = 1$ for all $x \in \mathcal{X}^{\mathbf{a}}$. We will be mainly concerned with the trajectory before extinction $(Y_t : t < T)$, so without loss of generality, we can assume $Y_t = Y_T$ when $t \geq T$, that is, the set of forbidden states is absorbing, or a trap. A sketch of various trajectories is shown in Fig. 1.1.

Among the main concepts in our study are the quasi-stationary distributions (QSDs, for short), these are distributions that are invariant under time evolution when the process is conditioned to survive. More precisely, a probability measure ν on $\mathcal{X}^{\mathbf{a}}$ is called a QSD (for the process killed at $\partial\mathcal{X}$) if for every measurable set B contained in $\mathcal{X}^{\mathbf{a}}$,

$$\mathbb{P}_\nu(Y_t \in B | T > t) = \nu(B), \quad t \geq 0.$$

Here, as usual, $\mathbb{P}_\nu = \int_{\mathcal{X}^{\mathbf{a}}} \mathbb{P}_x \, d\nu(x)$.

The existence of a QSD for killed processes, its description, the convergence to it of conditioned processes, its role in the process conditioned to never be killed, and the behavior of the killing time constitute the core of this work. On the other hand, since only few results exist for general processes, we prefer to introduce these problems in concrete contexts, as discrete Markov chains and dynamical systems.

P. Collet et al., *Quasi-Stationary Distributions*, Probability and Its Applications,
DOI 10.1007/978-3-642-33131-2_1,

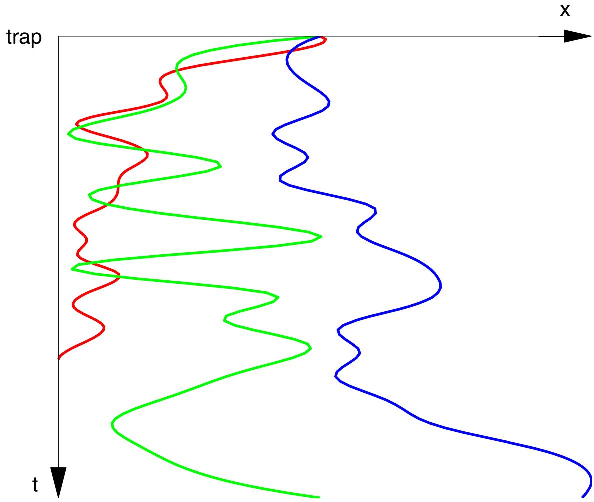


Fig. 1.1 Trajectories killed at the trap $x = 0$

1.2 Markov Chains

The study of QSDs on Markov chains have been extensively developed since the pioneering work (Yaglom, 1947) on branching process, and the classification of killed processes introduced in Vere-Jones (1962). The description of QSDs for finite state Markov chains was done in Darroch and Seneta (1965).

Below we describe some of the main problems of the theory on long time survival and QSDs on Markov chains. To avoid technical difficulties in this section, we prefer to introduce the ideas for discrete time Markov chains. But, in the next chapters, except the one devoted to QSDs of dynamical systems, we will only consider continuous time because this is the case of the major part of the bibliography devoted to QSDs of Markov processes. Some of the questions we set will be answered in the following chapters where we will also give precise references.

So, let us assume $\mathcal{X}$ is countable and $Y = (Y_t : t \in \mathbb{Z}_+ = \{0, 1, \ldots\})$ is a discrete time Markov chain taking values in $\mathcal{X}$. We assume that $(Y_t : t < T)$ is irreducible on $\mathcal{X}^{\mathbf{a}}$ and that the process is surely killed, $\mathbb{P}_x(T < \infty) = 1$ for $x \in \mathcal{X}^{\mathbf{a}}$. In this context, a QSD is a probability measure ν on $\mathcal{X}^{\mathbf{a}}$ that satisfies

$$\mathbb{P}_\nu(Y_t = x | T > t) = \nu(x) \quad \forall x \in \mathcal{X}^{\mathbf{a}}, t \in \mathbb{Z}_+.$$

From the irreducibility condition, one obtains that every QSD ν charges all points in $\mathcal{X}^{\mathbf{a}}$. On the other hand, since Y is absorbed at time T, the condition of a QSD takes the form

$$\mathbb{P}_\nu(Y_t = x) = \nu(x)\mathbb{P}_\nu(T > t) \quad \forall x \in \mathcal{X}^{\mathbf{a}}, t \in \mathbb{Z}_+.$$

Let ν be a QSD. We proceed to prove that when starting from ν, the killing time T at $\partial\mathcal{X}$ is geometrically distributed, that is,

$$\exists \alpha = \alpha(\nu) \in [0, 1] \quad \text{such that} \quad \mathbb{P}_\nu(T > t) = \alpha^t \quad \forall t \in \mathbb{Z}_+.$$

Since ν is a QSD, the Markov property gives the following equalities for all $t, s \geq 0$:

$$\begin{aligned}\mathbb{P}_\nu(T > t+s) &= \sum_{x\in\mathcal{X}^{\mathbf{a}}} \mathbb{P}_\nu(T > t+s, Y_s = x)\\ &= \sum_{x\in\mathcal{X}^{\mathbf{a}}} \mathbb{P}_\nu(T > t+s | Y_s = x)\mathbb{P}_\nu(Y_s = x)\\ &= \sum_{x\in\mathcal{X}^{\mathbf{a}}} \mathbb{P}_x(T > t)\nu(Y_s = x)\mathbb{P}_\nu(T > s) = \mathbb{P}_\nu(T > s)\mathbb{P}_\nu(T > t).\end{aligned}$$

Then, the geometrical distribution of T starting from ν follows. We avoid trivial situations, and so $\alpha(\nu) \in (0, 1)$. We denote by $\theta(\nu) := -\log \alpha(\nu) \in (0, \infty)$ the exponential rate of survival of ν, it verifies $\mathbb{P}_\nu(T > t) = e^{-\theta(\nu)t}$ for all $t \in \mathbb{Z}_+$. Furthermore, when ν is a QSD, we get

$$\forall \theta < \theta(\nu): \quad \mathbb{E}_\nu\left(e^{\theta T}\right) < \infty.$$

Hence, a necessary condition for the existence of a QSD is that some exponential moments be finite:

$$\forall x \in \mathcal{X}^{\mathbf{a}} \exists \theta > 0: \quad \mathbb{E}_x\left(e^{\theta T}\right) < \infty.$$

A solidarity argument, based upon irreducibility, shows that the quantity

$$\theta^* = \sup\left\{\theta : \mathbb{E}_x\left(e^{\theta T}\right) < \infty\right\}$$

does not depend on $x \in \mathcal{X}^{\mathbf{a}}$, we call it the exponential rate of survival of the process Y (killed at $\mathcal{X}^{\mathbf{a}}$). The above discussion shows that $\theta^* > 0$ is a necessary condition for the existence of a QSD. When this condition holds, we say that the process Y is exponentially killed.

Let $P = (p(x, y) : x, y \in \mathcal{X})$ be the transition matrix of Y, $P_{\mathbf{a}} = (p(x, y) : x, y \in \mathcal{X}^{\mathbf{a}})$ be the transition kernel restricted to the allowed states, and denote by ${P_{\mathbf{a}}}^{(t)} = (p^{(t)}(x, y) : x, y \in \mathcal{X}^{\mathbf{a}})$ the tth power of $P_{\mathbf{a}}$ for $t \in \mathbb{Z}_+$.

Let ν be a probability distribution on $\mathcal{X}^{\mathbf{a}}$, we denote by $\nu' = (\nu(x) : x \in \mathcal{X}^{\mathbf{a}})$ the associated row probability vector indexed by $\mathcal{X}^{\mathbf{a}}$. We have $\mathbb{P}_\nu(Y_t = y) = \nu' {P_{\mathbf{a}}}^{(t)}(y)$. In the case ν is a QSD, there exists an $\alpha \in (0, 1)$ such that $\mathbb{P}_\nu(T > t) = \alpha^t$ for all $t \in \mathbb{Z}_+$. So, if ν is a QSD, ν' must verify

$$\forall t \in \mathbb{Z}_+: \quad \nu' {P_{\mathbf{a}}}^{(t)} = \alpha^t \nu'.$$

This is equivalent to $\nu' P_{\mathbf{a}} = \alpha \nu'$ and so ν' must be a left-eigenvector of $P_{\mathbf{a}}$ with the eigenvalue $\alpha \in (0, 1)$. Observe that the eigenvalue is related to ν' by the equality $\alpha = \nu' P_{\mathbf{a}} \mathbf{1} = \mathbb{P}_\nu(T > 1)$, where $\mathbf{1}$ is the unit vector (all its components are 1). Note that the equation of a QSD is nonlinear in ν because α also depends on ν.

Let $\mathcal{X}^{\mathbf{a}}$ be finite; the sure killing condition is equivalent to $P_{\mathbf{a}}\mathbf{1} \leq \mathbf{1}$ and at some $x_0 \in \mathcal{X}^{\mathbf{a}}$ there is a loss of mass, that is, $\sum_{y\in\mathcal{X}^{\mathbf{a}}} p(x_0, y) < 1$. In this case, the existence and uniqueness of a QSD follow straightforwardly from the Perron–Frobenius theory. Indeed, since $P_{\mathbf{a}}$ is a positive, irreducible, strictly substochastic matrix, this theory ensures the existence of a simple eigenvalue $\alpha \in (0,1)$ of $P_{\mathbf{a}}$ that has left- and right-eigenvectors ν and ψ, respectively, that can be chosen to be strictly positive. Moreover, the associated eigenvectors of any other eigenvalue cannot be positive. We can impose the following two normalizing conditions:

$$\nu'\psi = 1, \qquad \nu'\mathbf{1} = \mathbf{1},$$

where $\nu'\psi = \sum_{x\in\mathcal{X}^{\mathbf{a}}} \psi(x)\nu(x)$. In particular, ν is a probability vector. Let us see that ν is a QSD. From $\nu'\mathbf{1} = 1$, we get

$$\mathbb{P}_\nu(T > t) = \nu' P^{(t)}\mathbf{1} = \alpha^t,$$

which shows that ν is a QSD, and it is clearly the unique one. Also $\theta = -\log\alpha$ is the exponential rate of survival.

The Perron–Frobenius Theorem also gives

$$\exists\beta \in [0,\alpha) \quad \text{such that} \quad P_{\mathbf{a}}{}^{(t)} = \alpha^t\psi\nu' + O\big(\beta^t\big),$$

or equivalently, $p^{(t)}(x,y) = \alpha^t\psi(x)\nu(y) + O(\beta^t)$. So, in the finite case, one has

$$\forall x \in \mathcal{X}^{\mathbf{a}}\colon \quad \mathbb{P}_x(T > t) = \sum_{y\in I^{\mathbf{a}}} p^{(t)}(x,y) = \alpha^t\psi(x) + O\big(\beta^t\big).$$

When a probability measure ν defined on $\mathcal{X}^{\mathbf{a}}$ satisfies the property

$$\forall y \in \mathcal{X}^{\mathbf{a}}\exists\nu(y) := \lim_{t\to\infty} \mathbb{P}_x(Y_t = y | T > t),$$

it is called a quasi-limiting distribution, or a Yaglom limit. As defined, it is independent of $x \in \mathcal{X}^{\mathbf{a}}$.

Notice that

$$\forall x \in \mathcal{X}^{\mathbf{a}}, \quad \alpha = -\lim_{t\to\infty} \frac{1}{t}\mathbb{P}_x(T > t),$$

so the Perron–Frobenius eigenvalue is the exponential rate of survival of the process starting from any initial distribution.

When the chain evolves in an infinite countable set, the following problems on QSDs can be set:

Problem 1.1 The problem of existence of a QSD. From the discussion, a necessary condition for the existence of a QSD is that the process is exponentially killed. Under this condition, we will give the proof that a sufficient condition for the existence of a QSD is that the process does not come from infinity in finite time (that is, for all fixed positive time, it is not possible to be killed before this time when the process starts arbitrarily far away). As shown for birth-and-death chains, this condition on non-coming from infinity is clearly unnecessary for the existence of a QSD.

Problem 1.2 The problem of uniqueness of a QSD. As we shall see for exponentially killed time-continuous birth-and-death chains, there are two possibilities: (i) if the chain comes from infinity in finite time, there is a unique QSD, and (ii), in the opposite case, there is a continuum of QSDs, in fact, there is one QSD ν_θ for each $\theta \in (0, \theta^*]$, where θ^* is the exponential rate of survival of the process. The problem of the interpretation of the extremal QSD ν_{θ^*} appears.

Problem 1.3 The problem of the domains of attraction. If ρ is a probability distribution on $\mathcal{X}^{\mathbf{a}}$ and we consider the conditional evolution $\mathbb{P}_\rho(Y_t \in \bullet | T > t)$, we would like to know if it converges; and when the answer is positive and there are several QSDs, we would like to know to which of the QSDs it converges.

For birth-and-death chains, we will give the proof that this limit exists for all Dirac measures $\rho = \delta_x$ and it coincides with the extremal QSD, that is,

$$\forall y \in \mathcal{X}^{\mathbf{a}}: \quad \nu_{\theta^*}(y) = \lim_{t\to\infty} \mathbb{P}_x(Y_t = y | T > t).$$

When there is a unique QSD (for instance, for birth-and-death chains that are exponentially absorbed and coming from infinity in finite time), a problem is when all distributions ρ belong to the domain of attraction of the QSD.

For initial distributions different from the Dirac measures, a complete answer to the question of domains of attraction is very hard. In fact, it is answered in a satisfactory way only (in the context of diffusions) for the Brownian motion with strictly negative constant drift.

Let us study the process of trajectories that survive forever. From the Markov property, we get

$$\mathbb{P}_x(Y_1 = x_1, \ldots, Y_k = i_k, T > t) = \mathbb{P}_x(Y_1 = i_1, \ldots, Y_k = i_k)\mathbb{P}_{x_k}(T > t - k).$$

In the finite case, we can go further. In fact, the computations already done give the following ratio limit result on the survival probabilities:

$$\forall x, y \in \mathcal{X}^{\mathbf{a}}: \quad \lim_{t\to\infty} \frac{\mathbb{P}_x(T > t - k)}{\mathbb{P}_y(T > t)} = \alpha^k \frac{\psi(x)}{\psi(y)}.$$

Hence,

$$\lim_{t\to\infty} \mathbb{P}_{x_0}(Y_1 = x_1, \ldots, Y_k = x_k | T > t) = \prod_{l=1}^{k} \left(\alpha \frac{\psi(x_l)}{\psi(x_{l-1})} p(x_{l-1}, x_l) \right).$$

Therefore, in the finite case, the process of trajectories that survive forever, $Z = (Z_t : t \geq 0)$, whose law is given by

$$\mathbb{P}_x(Z_1 = x_1, \ldots, Z_k = x_k) := \lim_{t\to\infty} \mathbb{P}_x(Y_1 = x_1, \ldots, Y_k = x_k | T > t), \quad x \in \mathcal{X}^{\mathbf{a}},$$

is a Markov chain taking values on $\mathcal{X}^{\mathbf{a}}$, with transition probability

$$\tilde{p}(x, y) := \mathbb{P}_x(Z_1 = y) = \alpha \frac{\psi(y)}{\psi(x)} p(x, y).$$

Observe that $\widetilde{P} = (\tilde{p}(x, y) : x, y \in \mathcal{X}^{\mathbf{a}})$ is a stochastic matrix because

$$\sum_{y \in \mathcal{X}^{\mathbf{a}}} \tilde{p}(x, y) = \alpha \sum_{y \in \mathcal{X}^{\mathbf{a}}} p(x, y) \frac{\psi(y)}{\psi(x)} = 1.$$

In this finite case, and under irreducibility, the chain Z is obviously positive recurrent, and from the equalities

$$(\nu\psi)' P(y) = \sum_{x \in I^{\mathbf{a}}} \nu(x)\psi(x)\tilde{p}(x, y) = \alpha\psi(y) \sum_{x \in I^{\mathbf{a}}} \nu(x) p(x, y) = \nu(y)\psi(y),$$

$\nu\psi = (\nu(x)\psi(x) : x \in I^{\mathbf{a}})$ is the stationary distribution for Z. To avoid confusion, we denote by $\widetilde{\mathbb{P}}$ the law of the process Z given by the transition matrix $\widetilde{P}$.

For a general Markov chain, the following problem is put:

Problem 1.4 Existence and classification of the process Z of trajectories that survive forever. For birth-and-death chains, we show the existence of the process Z. To do it, we extend to birth-and-death chains the result showing that the ratio between the components of the right-eigenvector correspond to the limit ratios of the survival probabilities. In relation with the classification of the process Z (following Vere-Jones classification), we supply a criterion based on the exponential rate of survival which ensures when the process Z is positive recurrent, null-recurrent or transient.

We have discussed the convergence in distribution of the initial piece of trajectory,

$$\lim_{t \to \infty} \mathbb{P}_\rho(Y_0 = x_0, \ldots, Y_k = x_k | T > t) = \widetilde{\mathbb{P}}_\rho(Z_0 = x_0, \ldots, Z_k = x_k).$$

In the finite case, the quasi-limiting behavior allows getting the convergence of the final piece of trajectory. It is given by

$$\lim_{t \to \infty} \mathbb{P}_\rho(Y_t = x_0, \ldots, Y_{t-l} = x_l | T > t) = \mathbb{P}_\nu(Y_t = x_0, \ldots, Y_{t-l} = x_l | T > l),$$

where ν is the quasi-limiting distribution.

Note that this limit distribution on the last piece of the trajectory requires the existence of the quasi-limiting distribution and the limit ratio of the survival probabilities. Hence, by previous discussion, this limit behavior can be also retrieved for the birth-and-death chains that are exponentially killed.

1.3 Diffusions

In this section, we give a simple example of a QSD for a diffusion. We refer to Pinsky (1985) for a more general case and to Chaps. 6 and 7 for other diffusions in dimension one.

In this section, we denote by $\mathcal{X}$ a compact, connected domain of $\mathbb{R}^n$ with a regular boundary, and the process $Y_\cdot$ will be the n-dimensional Brownian motion killed on $\partial\mathcal{X}$. We will prove the existence of a QSD for this process. When necessary, we will use the notation $Y_\cdot^x$ to denote the process starting at x.

Theorem 1.1 *Under the above hypothesis, the Brownian motion killed on $\partial\mathcal{X}$ has a QSD which is absolutely continuous with respect to the Lebesgue measure.*

Proof From the spectral theory of elliptic partial differential operators (see, for example, Davies 1989), it is well known that the Dirichlet Laplacian in $\mathcal{X}$ has a largest eigenvalue $-\lambda < 0$ with an associated eigenfunction u which is nonnegative and bounded, namely

$$\Delta u = -\lambda u,$$

$u \geq 0$ and $u_{|\partial\mathcal{X}} = 0$.

From Ito's formula (see, for example, Gikhman and Skorokhod 1996) we have for any function f which is twice continuously differentiable and with compact support contained in the interior of $\mathcal{X}$ that

$$e^{\lambda t/2} f\left(Y_{t\wedge T}^x\right) = f(x) + \frac{1}{2}\int_0^{t\wedge T} \left(\Delta f\left(Y_{s\wedge T}^x\right) + \lambda f\left(Y_{s\wedge T}^x\right)\right) ds + \mathscr{M}_t$$

where $(\mathscr{M}_\cdot)$ is a martingale. Therefore,

$$\begin{aligned} e^{\lambda t/2}\mathbb{E}_x\left(f(Y_{t\wedge T})\right) &= f(x) + \frac{1}{2}\mathbb{E}\left(\int_0^{t\wedge T} \left(\Delta f\left(Y_{s\wedge T}^x\right) + \lambda f\left(Y_{s\wedge T}^x\right)\right) ds\right) \\ &= f(x) + \frac{1}{2}\mathbb{E}\left(\int_0^{t\wedge T} \left(\Delta f\left(Y_{s\wedge T}^x\right) + \lambda f\left(Y_{s\wedge T}^x\right)\right) ds\right). \end{aligned}$$

Using integration by parts, we get

$$\mathbb{E}\left(\int_{\mathcal{X}} u(x)\left(\Delta f\left(Y_{s\wedge T}^x\right) + \lambda f\left(Y_{s\wedge T}^x\right)\right) dx\right) = 0.$$

Therefore, by Fubini's Theorem, we obtain

$$e^{\lambda t/2}\int_{\mathcal{X}} u(x)\mathbb{E}_x\left(f(Y_{t\wedge T})\right) dx = \int_{\mathcal{X}} u(x) f(x)\, dx.$$

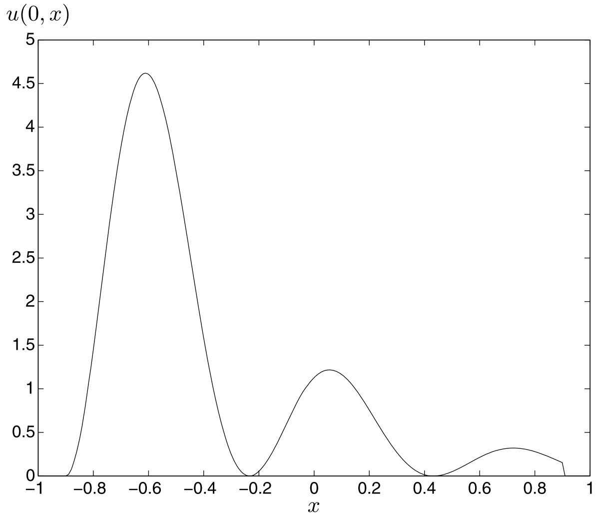


Fig. 1.2 The density of $u(0,x)$ of the initial distribution

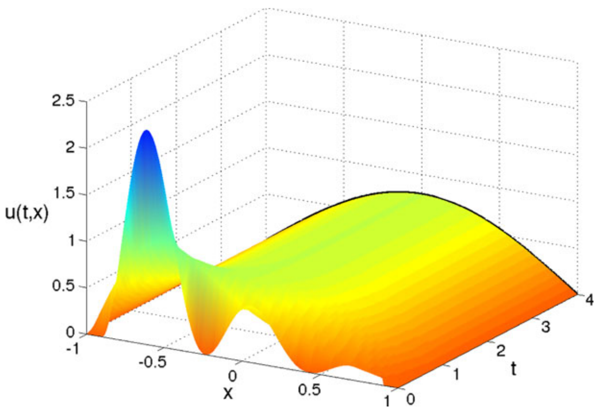


Fig. 1.3 Time evolution $u(t,x)$ of the density

Let ν denote the positive measure with density u ($d\nu = u\,dx$), the above formula can be rewritten

$$\mathbb{E}_\nu\big(f(Y_{t\wedge T})\big) = e^{-\lambda t/2}\int_{\mathcal{X}} f\,d\nu.$$

Since this formula holds for any function f which is twice continuously differentiable and with compact support contained in the interior of $\mathcal{X}$, and since u vanishes on the boundary, we conclude that ν is a QSD. Note also that, since u is bounded and hence integrable, we can normalize ν to a probability measure. $\square$

One can give other proofs of this theorem avoiding the use of the spectral theory of elliptic operators, using, for example, the spectral theory. Results for more general diffusions in dimension one will be proved in Chaps. 6 and 7.

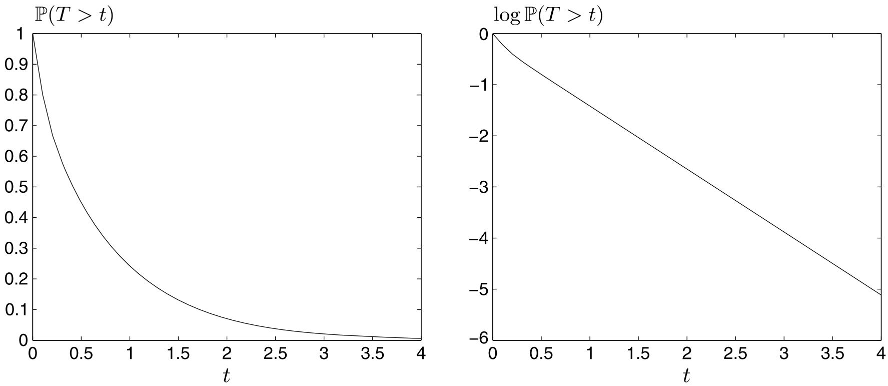


Fig. 1.4 (*Left*) $\mathbb{P}(T > t)$ as a function of t. (*Right*) $\log \mathbb{P}(T > t)$ as a function of t

Spectral theory allows, in particular, one to prove the existence of a Yaglom limit and to show that the rate of convergence is exponential (we will see more on this in later chapters). Here we illustrate these properties using a simulation for the one dimensional Brownian motion on the interval $[-1, 1]$. We start with the initial condition which is absolutely continuous with respect to the Lebesgue measure with density drawn in Fig. 1.2.

In Fig. 1.3, we show the time evolution of the density, normalized to have integral one. We clearly see convergence to the density of the Yaglom limit ($\pi \cos(\pi x/2)/4$ in this case) which is drawn in black at the end of the surface.

One can also look at the time evolution of the survival probability $\mathbb{P}(T > t)$. This is shown in Fig. 1.4 on the left. On the right, we plotted the logarithm of this quantity as a function of time, clearly showing convergence to a straight line, namely that $\mathbb{P}(T > t)$ decays asymptotically exponentially.

1.4 Dynamical Systems

A classical system (in physics, chemistry, biology, etc.) is described by the set $\mathcal{X}$ of all its possible states, often called the phase space of the system. At a given time, all the properties of the system can be recovered from the knowledge of the instantaneous state $x \in \mathcal{X}$. The system is observed using the so-called observables which are real-valued functions on $\mathcal{X}$. Most often the space of states $\mathcal{X}$ is a metric space (so we can speak of nearby states). In many physical situations, there are even more structures on $\mathcal{X}$: $\mathbb{R}^d$, Riemannian manifolds, Banach or Hilbert spaces, etc.

As a simple example one can consider a mechanical system with one degree of freedom. The state of the system at a given time is given by two real numbers: the position q and momentum p. The state space is therefore $\mathbb{R}^2$. A continuous material (solid, fluid, etc.) is characterized by the field of local velocities, pressure, density, temperature, etc. In that case, one often uses phase spaces which are Banach spaces.

As time goes on, the instantaneous state changes (unless the system is in a situation of equilibrium). The time evolution is a rule giving the change of the state with time. It comes in several flavors and descriptions summarized below:

(i) Discrete time evolution. This is a map f from the state space $\mathcal{X}$ into itself producing a new state from an old one after one unit of time. If x_0 is the state of the system at time zero, the state at time one is $x_1 = f(x_0)$, and more generally, the state at time n is given recursively by $x_n = f(x_{n-1})$. This is often written $x_n = f^n(x_0)$ with $f^n = f \circ f \circ \cdots \circ f$ (n times). The sequence $(f^n(x_0))$ is called the trajectory or the orbit of the initial condition x_0.

(ii) Continuous-time semi-flow. This is a family $(\varphi_t : t \in \mathbb{R}^+)$ of maps of $\mathcal{X}$ satisfying

$$\varphi_0 = \mathrm{Id}, \qquad \varphi_s \circ \varphi_t = \varphi_{s+t}.$$

The set $\{\varphi_t(x_0) : t \in \mathbb{R}^+\}$ is called the trajectory (orbit) of the initial condition x_0. Note that if we fix a time step $\tau > 0$, and observe only the state at times $n\tau$ ($n \in \mathbb{N}$), we obtain a discrete time dynamical system given by the map $f = \varphi_\tau$. A flow is a family $(\varphi_t : t \in \mathbb{R})$ of maps of $\mathcal{X}$ satisfying the above two conditions for any t and s in $\mathbb{R}$. Note in particular that φ_{-t} is the inverse of φ_t.

(iii) A differential equation on a manifold associated to a vector field $\mathbf{F}$

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}).$$

This is, for example, the case of a mechanical system in the Hamiltonian formalism. Under regularity conditions on $\mathbf{F}$, the integration of this equation leads to a flow.

(iv) There are other more complicated situations like nonautonomous systems (in particular, stochastically forced systems), systems with memory, systems with delay, etc., but we will not consider them below.

A dynamical system is a set of states $\mathcal{X}$ equipped with a time evolution. If there is more structure on $\mathcal{X}$, one can put more structure on the time evolution itself. For example, in the case of discrete time, the map f may be measurable, continuous, differentiable, etc.

Assume now that the phase space $\mathcal{X}$ is equipped with a Borel structure, and let ν_0 be a Borel probability measure on $\mathcal{X}$. To any dynamical system on $\mathcal{X}$, we can now associate a stochastic process. We explain how this is done for the case of a discrete time evolution, the case of continuous time is similar. Let f be a measurable map of the phase space $\mathcal{X}$ as above, and let ν_0 be a probability measure on $\mathcal{X}$. Let $\mathcal{Y}$ be a measurable space, and let Y be a measurable map from $\mathcal{X}$ to $\mathcal{Y}$. We can consider Y as a $\mathcal{Y}$-valued random variable by the usual construction. Namely, for any measurable $B \in \mathcal{Y}$, we define

$$\mathbb{P}(Y \in B) = \nu_0\big(Y^{-1}(B)\big).$$

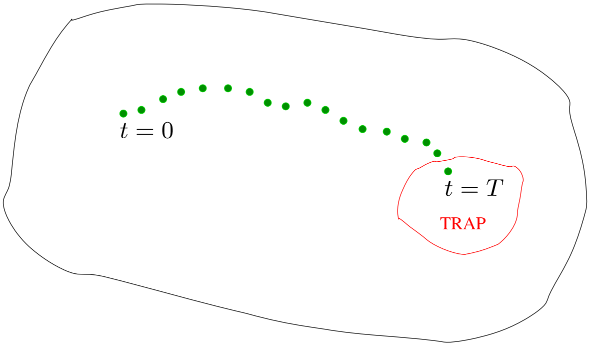


Fig. 1.5 Time evolution with a trap

More generally, we can define a discrete time stochastic process on $\mathscr{Y}$ by the sequence $(Y_n : n \in \mathbb{Z}_+)$ of random variables defined by

$$Y_n = Y \circ f^n.$$

In these processes defined by dynamical systems, the time evolution is deterministic, but the initial condition is chosen at random. This randomness is propagated by the time evolution. There is no loss of generality if we assume that $\mathcal{X} = \mathscr{Y}$ and Y is the identity mapping.

A measure ν on $\mathcal{X}$ is called an invariant measure for the map f if for any measurable set $B \in \mathcal{B}(\mathcal{X})$ we have

$$\nu\big(f^{-1}(B)\big) = \nu(B).$$

We recall that f^{-1} denotes here the set-theoretic inverse, or in other words, $f^{-1}(B)$ is the set of preimages of points in B

$$f^{-1}(B) = \big\{x : f(x) \in B\big\}.$$

The definition is the same for a semi-flow.

As explained above, to the map f and the measure ν, we can associate stochastic processes. If the measure ν is invariant, it is easy to verify that these processes are stationary.

We now consider the particular setting corresponding to quasi-stationary measures. We assume that a measurable subset $\mathcal{X}^{\text{tr}}$ of the phase space $\mathcal{X}$ is given, which is a hole or a trap. Namely, when an orbit reaches $\mathcal{X}^{\text{tr}}$, the time evolution stops (although the dynamical system may be well defined on $\mathcal{X}^{\text{tr}}$). See Fig. 1.5 for a sketch of this evolution.

Let $T = T_{\mathcal{X}^{\text{tr}}}$ be the first time an initial condition x reaches the trap $\mathcal{X}^{\text{tr}}$, namely, $f^j(x) \notin \mathcal{X}^{\text{tr}}$ for $0 \le j < T(x)$ and $f^{T(x)}(x) \in \mathcal{X}^{\text{tr}}$. It is convenient to define $T = 0$ on $\mathcal{X}^{\text{tr}}$.

In this situation, it is natural to define the sequence $(\mathcal{X}_n)$ of sets of initial conditions which do not reach $\mathcal{X}^{\text{tr}}$ before time n, namely $\mathcal{X}_n = \{T > n\}$. The sets $\mathcal{X}_n$ can also be defined recursively by $\mathcal{X}_0 = \mathcal{X} \backslash \mathcal{X}^{\text{tr}}$, and

$$\mathcal{X}_{n+1} = \mathcal{X}_n \cap \big\{x : f^{n+1}(x) \notin \mathcal{X}^{\text{tr}}\big\}.$$

Note that we also have

$$\mathcal{X}_n = \bigcap_{j=0}^{n} f^{-j}(\mathcal{X}_0).$$

We recall that there are several natural questions associated to this situation.

Problem 1.5 Given an initial distribution ν_0 on $\mathcal{X}$, what is the probability that a trajectory has survived up to time $n > 0$? (For example, if ν_0 is the Dirac measure on one point.)

In other words, what is the behavior of $\nu_0(\mathcal{X}_n)$. Often one can say something only for large n.

Problem 1.6 Assume a trajectory initially distributed with ν_0 has survived up to time $n > 0$. What is the distribution at that time (of the surviving trajectory)?

In other words, can we say something about

$$\frac{\nu_0(f^{-n}(B) \cap \mathcal{X}_n)}{\nu_0(\mathcal{X}_n)},$$

B a measurable subset of $\mathcal{X}$.

We recall that the Yaglom limit of the measure ν_0 is defined as the limit of the above measure when n tends to infinity (if the limit exists). It may depend on the initial distribution ν_0.

A related object is a quasi-stationary measure (QSD).

We recall that ν (a probability measure on $\mathcal{X}$) is quasi-stationary if for any $n \geq 0$

$$\frac{\nu(f^{-n}(B) \cap \mathcal{X}_n)}{\nu(\mathcal{X}_n)} = \nu(B)$$

for any measurable set B. If there is no trap, this is a stationary measure.

Problem 1.7 Are there trajectories which never reach the trap A? ($T = \infty$).

If so, how are they distributed?

How is this related to Problem 1.6?

To get some intuition about these questions in the case of dynamical systems, we now discuss a simple example where explicit computations can be performed.

The phase space $\mathcal{X}$ is the unit interval $[0, 1]$. Consider the map f given by $f(x) = 3x \pmod 1$. In Fig. 1.6 we show the graph of f, and in Fig. 1.7 the geometrical construction of one iteration.

It is easy to verify that the Lebesgue measure Leb is invariant, namely, for any Borel set B

$$\mathrm{Leb}\big(f^{-1}(B)\big) = \mathrm{Leb}(B). \tag{1.1}$$

Indeed, it is enough to verify (1.1) for a finite union of disjoint sub-intervals of [0, 1], and in fact for each interval separately. By a simple computation, one obtains

Fig. 1.6 The map $f(x) = 3x$ (mod 1)

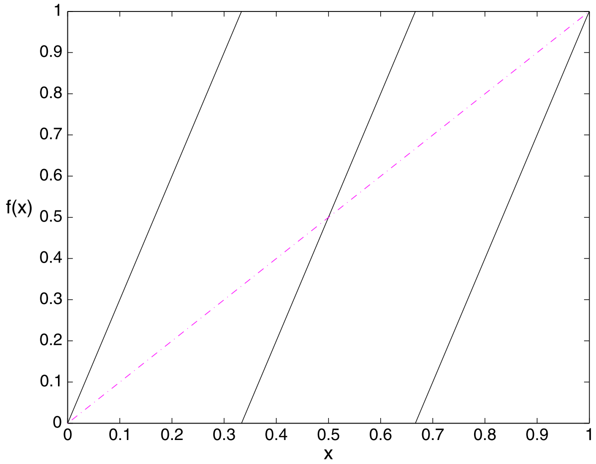
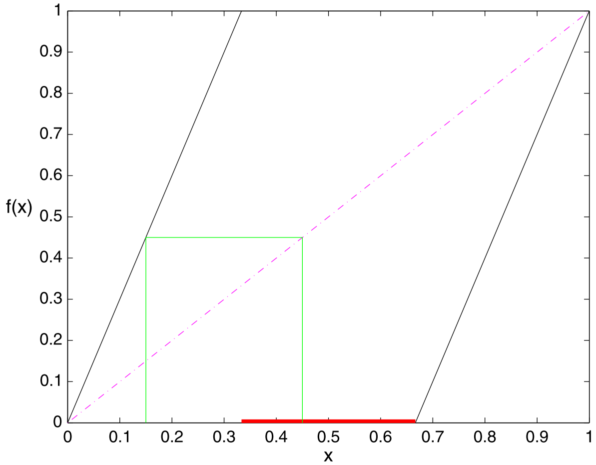


Fig. 1.7 The map $f(x) = 3x$ (mod 1) with the trap $]1/3, 2/3[$ (*in red*), geometrical construction of an iterate (*in green*) falling into the trap after one step



immediately for the interval $I = (a, b) \subset [0, 1]$

$$f^{-1}(I) = (a/3, b/3) \cup \big((a+1)/3, (b+1)/3\big) \cup \big((a+2)/3, (b+2)/3\big),$$

and therefore,

$$\text{Leb}\big(f^{-1}(I)\big) = 3(b/3 - a/3) = b - a = \text{Leb}(I).$$

As explained above, given an initial distribution, we have a discrete time stochastic process.

The probability space is $[0, 1]$ and the process is defined recursively by $Y_{n+1} = f(Y_n)$.

We use the interval $]1/3, 2/3[$ as a trap.

The normalized Lebesgue measure $\nu = 3\text{Leb}/2$ on $\mathcal{X}_0 = [0, 1/3] \cup [1/3, 1]$ is quasi-stationary.

The proof is an easy computation from the definition in the discrete time case. As we will see later (see Lemma 8.2), it is enough to check that

$$\frac{\nu(f^{-1}(B) \cap \mathcal{X}_1)}{\nu(\mathcal{X}_1)} = \nu(B),$$

for any Borel set $B \subseteq \mathcal{X}_0$. Moreover, by completion, it is enough to check this identity for a finite union of disjoint intervals, and this follows if we can check it for any interval $(a, b) \subset \mathcal{X}_0$. A simple computation (as above for the invariant measure) leads to

$$\mathcal{X}_1 = [0, 1/9] \cup [2/9, 1/3] \cap [2/3, 7/9] \cap [8/9, 1].$$

Therefore,

$$\text{Leb}(\mathcal{X}_1) = \frac{4}{9} \quad \text{and} \quad \nu(\mathcal{X}_1) = \frac{2}{3}.$$

As above, for $I = (a, b)$, we have

$$f^{-1}(I) = (a/3, b/3) \cup \big((a+1)/3, (b+1)/3\big) \cup \big((a+2)/3, (b+2)/3\big),$$

and therefore, if $I \subset \mathcal{X}_0$,

$$f^{-1}(I) \cap \mathcal{X}_1 = (a/3, b/3) \cup \big((a+2)/3, (b+2)/3\big).$$

This implies

$$\text{Leb}\big(f^{-1}(I) \cap \mathcal{X}_1\big) = \frac{2}{3}(b - a).$$

Therefore,

$$\frac{\nu(f^{-1}(I) \cap \mathcal{X}_1)}{\nu(\mathcal{X}_1)} = \frac{3\text{Leb}(f^{-1}(I) \cap \mathcal{X}_1)/2}{2/3} = \frac{3}{2}(b - a) = \nu(I).$$

It is easy to verify inductively that

$$\nu(\mathcal{X}_n) = \left(\frac{2}{3}\right)^n.$$

Namely, we have the answer to Problem 1.5: the probability of surviving up to time n when the initial condition is distributed according to the QSD ν.

The set of initial conditions which never die is

$$K = \bigcap_n \mathcal{X}_n$$

which is the Cantor set. K is of zero Lebesgue measure. Moreover, during the recursive construction, if we start with the Lebesgue measure (or the QSD ν), $\mathcal{X}_n$ is a union of triadic intervals which have the same length and hence the same weight. We get at the end the Cantor measure which is singular with respect to the Lebesgue measure (or the QSD ν). So there is little connection with the nice QSD. Intuitively, a QSD (or a Yaglom limit) describes the distribution of trajectories which have survived for a large time but most of them are on the verge of falling in the trap.

Problem 1.7 deals with trajectories that will never see the trap, this is very different. In the case of dynamical systems, these trajectories concentrate on a very small set which is invariant and disjoint from the trap.

Problem 1.6 (distribution of survivors at large time) and Problem 1.7 (eternal life) have very different answers.

Chapter 2
Quasi-Stationary Distributions: General Results

In this chapter, we introduce the main concepts in a general context of killed processes. Thus, in Sect. 2.2, we give the definition of quasi-stationary distributions (QSDs). In Theorem 2.2 of Sect. 2.3, we show that starting from a QSD the killing time is exponentially distributed, and in Theorem 2.6 of Sect. 2.4, we show that the killing time and the state of killing are independent random variables. In Theorem 2.11 of Sect. 2.7, we give a theorem of existence of a QSD in a topological setting without any assumption on compactness or spectral properties.

2.1 Notation

We will consider a Markov process $Y = (Y_t : t \geq 0)$ taking values in a state space $\mathcal{X}$ which is endowed with a σ-field $\mathcal{B}(\mathcal{X})$. Let us fix some notation:

- $\mathcal{M}(\mathcal{X})$ is the set of real measurable functions defined on $\mathcal{X}$; $\mathcal{M}_+(\mathcal{X})$ (respectively, $\mathcal{M}_b(\mathcal{X})$) is the set of positive (respectively, bounded) elements in $\mathcal{M}(\mathcal{X})$;
- $\mathcal{P}(\mathcal{X})$ is the set of probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.

When $\mathcal{X}$ has a topological structure, $\mathcal{B}(\mathcal{X})$ denotes its Borel σ-field. In this case, we denote by $\mathcal{C}(\mathcal{X})$ the set of continuous functions and by $\mathcal{C}_b(\mathcal{X})$ its bounded elements.

Let v be a measure on $\mathcal{X}$ and $f \in \mathcal{M}(\mathcal{X})$. If $\int f\,dv$ is well defined then we set $v(f) = \int f\,dv$. This is the case if $f \in L^1(v)$, or when $f \in \mathcal{M}_+(\mathcal{X})$, or when $v \in \mathcal{P}(\mathcal{X})$ and $f \in \mathcal{M}_b(\mathcal{X})$.

Let us fix $\mathcal{X}$ a topological space endowed with its Borel σ-field $\mathcal{B}(\mathcal{X})$. Let Y be a Markov process taking values in $\mathcal{X}$ satisfying the conditions we fix below.

Let Ω be the set of right continuous trajectories on $\mathcal{X}$, indexed by points in $[0, \infty)$. So $\omega = (\omega_s : s \in [0, \infty)) \in \Omega$ means that $\omega_s \in \mathcal{X}$ and $\omega_s = \lim_{h \to 0^+} \omega_{s+h}$ for all $s \in [0, \infty)$. Let $(\Omega, \mathcal{F})$ be a measurable space where $\mathcal{F}$ contains the σ-field generated by all the projections $\text{pr}_s : \Omega \to \mathcal{X}, \omega \to \text{pr}_s(\omega) = \omega_s, s \geq 0$. For $t \geq 0$ let $\Theta_t : \Omega \to \Omega$ be the operator shifting the trajectories by t, so $\Theta_t(\omega) = (\omega_s : s \geq t)$.

P. Collet et al., *Quasi-Stationary Distributions*, Probability and Its Applications,
DOI 10.1007/978-3-642-33131-2_2,