Introduction and history

Introduction

This book differs essentially from the existing monographs on countable Markov chains. It intends to be, on the one hand, much *more constructive* than books similar to, for example Chung's [Chu67] and, on the other hand, much *less constructive* than some elementary monographs on queueing theory, where the emphasis is mainly put on the derivation of explicit expressions. The method of generating functions, which is to be sure the most constructive approach, is not included, since the dimension of the problems it can solve is small (in general ≤ 2). Our book could equally be called *Constructive use of Lyapounov functions method*. Here the term *constructive* is taken in the sense close to the one widely accepted in constructive mathematical physics. One can say that the objects considered have a sufficiently rich structure to be *concrete*, although the results may not always be *explicit* enough, as commonly understood. Semantically, it is permissible to say that our methods are more *qualitative constructive* than *quantitative constructive*.

The main goal of the book is to provide methods allowing a complete classification (necessary and sufficient conditions) or, in other words, allowing us to say when a Markov chain is ergodic, null recurrent or transient. Moreover, it turns out that, without doing much additional work, it is possible to study the *stability* (continuity or even analyticity) with respect to parameters, the rate of convergence to equilibrium,..., etc. by using the same Lyapounov functions.

Our primary concern with necessary *and* sufficient conditions is crucial, since in many cases it is indeed trivial to get explicit necessary *or* sufficient conditions. Another peculiarity of our approach is that we do not pursue generalizations, which could be easily done by any expert in standard classical probability theory. For example, in many places, we restrict ourselves to bounded jumps, whenever the formulation would remain unchanged in the case of unbounded jumps.

The various sections of chapter 1 give only exact definitions and some results taken from countable Markov chains that we use. To render the book accessible for the beginner, we also present section 1.4, to demonstrate the possibilities of perhaps more exact, but also more restrictive, elementary methods.

In chapter 2, we present the main classification criteria for general countable Markov chains, which are needed in the following chapters. Further far reaching martingale criteria are presented. Also we obtain some *exponential* bounds, which imply nice properties for the corresponding Markov chains.

The rest of the monograph is devoted to the so-called *deflected* random walks in $\mathbf{Z}_+^N$. The reader might wonder why random walks in $\mathbf{Z}_+^N$ are of primary interest. There are several striking reasons. First, they describe many networks of practical interest (e.g. see section 3.2) and the methods presented here could also be useful for more general networks, for instance with non-identical customers. Secondly, the problems involved not only are of probabilistic interest, but they also produce a large store of examples and, moreover, are closely connected with other branches of mathematics. In fact the classification problem for random walks in R_+^N is a probabilistic version of a well known question in functional analysis and partial differential equations: When is a multidimensional Toeplitz (or any general elliptic) operator in $\mathbf{Z}_+^N$ invertible? It also has much in common with the problem of the behaviour of diffusion processes near non-smooth boundaries of large codimension. The ideas and methods exhibited here are, in our opinion, useful for attacking problems of very different nature.

Chapter 3 gives techniques for an explicit geometrical construction of Lyapounov functions. They apply to random walks in $\mathbf{Z}_+^2$, as well as to the famous Jackson networks in $\mathbf{Z}_+^N$. The zero drift case in $\mathbf{Z}_+^2$ and *almost* zero drift one-dimensional examples of sections 3.6 and 3.7 constitute new directions of development, initiated by Lamperti [Lam60] thirty years ago. They are directly related to several works of R. Williams and others [VW85, Wil85].

The central method of *induced* chains and vector fields is presented in sections 4.1 and 4.2. In section 4.3, general results pertaining to the construction of Lyapounov functions in a uniformly bounded number of steps are given. Using these results, we obtain the complete classification in $\mathbf{Z}_+^3$.

Completely new phenomena appear in chapter 5: *scattering*, null recurrence for a positive Lebesgue measure in the parameter space, constants L and M in the simplest situation.

General criteria (some of them using Lyapounov functions), concerning conditions ensuring the continuity of stationary probabilities with respect to the parameters, are given in chapter 6.

Finally, chapter 7 offers a probabilistic criterion, again using the Lyapounov functions and Foster's theorems, for a family of Markov chains to be an *analytic Lyapounov family.* In particular, this property leads to analytic dependence on the parameters, as well as exponential convergence to equilibrium and exponential decrease of stationary probabilities.

Historical comments

Chapter 1. For the contents of this chapter we refer the reader to any standard textbooks on countable Markov chains, for example [Chu67, Kar68].

Chapter 2. The notion of *Lyapounov function* or *test function* similar to the well known Lyapounov functions for ordinary differential equations goes back to Foster [Fos53], as far as we know. Although his examples are now trivial, his ideas and criteria for ergodicity and for transience became basic for later extensions. There exist now many technical generalizations of these criteria, some of which we give in this chapter. Generalized Foster criteria for ergodicity were given in [Mal93]. In [Fil89] a new martingale proof is proposed with an important extension to random times. We have summarized and simplified all these results in theorems 2.1.1, 2.1.2 and 2.1.3. Theorems 2.2.1 and 2.2.2 extend, with new proofs, results contained in [MSZ78] and [Fos53]. Theorems 2.2.2 and 2.2.3 are the famous Foster criterion itself, with a slight modification and modern proofs. Theorem 2.2.6 generalizes some corresponding results of [Mal73] (given for a.s. uniformly bounded jumps). Theorem 2.2.8 is contained in [FMM92] and seems to be the unique constructive result allowing us to prove non-ergodicity by means of non-piecewise-linear Lyapounov functions. Theorems 2.1.1 and 2.1.10 are fundamental tools for proving all criteria we need and they also provide exponential estimates which are used in various parts of the book.

Chapter 3. Section 3.1 shows an elementary example. The results have been partially known for 20 years already. The proofs given in the book are pedagogic. Section 3.2 contains definitions taken from [MM79] and [Mal93]. Most of the theorems of sections 3.3 to 3.7 are new. The idea of using quadratic forms and functionals of quadratic forms is original, and appeared, as far as we know, for the first time in [Fay89, FB88, FMM92]. They are used in connection with the principle of almost linearity introduced in [Mal72a].

Chapter 4. Section 4.1, 4.3, 4.4 are taken, with some improvements, from [MM79]. Section 4.2 is basically contained in [Mal93].

The results of chapter 5 were first published in [FIVM91].

The content of chapters 6 and 7 is a substantial revision of the results in [MM79].

1

Preliminaries

In sections 1.1, 1.2 and 1.3 of this chapter, we briefly introduce basic notions and some results borrowed from the theory of discrete time homogeneous countable Markov chains (MC).

In section 1.4, some well known examples of MCs are given, for which a complete classification can be obtained by elementary methods: simple probabilistic arguments in 1.4.1, explicit solution of recurrent equations in 1.4.2, generating functions in 1.4.3.

It is not our intention to devote a detailed section to the fundamentals of probability theory, which are presented in a plethora of excellent textbooks. Thus, we only introduce in fact the minimal basic notions and notation useful for our purpose.

- The events are the subsets of some abstract set Ω, which belong to Σ, the σ-algebra defined on Ω.
- The couple (Ω, Σ) is a *measurable space* and the sets belonging to Σ are
  Σ-*measurable sets.*
- The triple (Ω, Σ, μ), where μ is a positive measure defined on Σ, is a *measure space.* A probability space is a measure space of total measure 1, i.e. $\mu(\Sigma) = 1$, and in this case most of the time we shall write (Ω, Σ, P).
- A Σ-measurable real-valued function f with domain Ω is called a *random variable.* More generally a *random element* φ with values in a measurable space $(X, \mathcal{B})$ is a measurable mapping of (Ω, Σ, P) into $(X, \mathcal{B})$. For $X = \mathbf{R}^N$ or $\mathbf{Z}^N$, $\mathcal{B}$ being the σ-algebra of Borel sets, we shall speak of random vectors.

1.1 Irreducibility and aperiodicity

Let $\mathcal{A}$ be a denumerable set and $\mathbf{P}$ a stochastic (transition) matrix such that

$$\mathbf{P} = (p_{\alpha\beta})_{\alpha,\beta\in\mathcal{A}}$$

and, for any $\alpha \in \mathcal{A}$, $P_\alpha = (p_{\alpha\beta})_{\beta\in\mathcal{A}}$ is a probability vector, that is

$$\sum_{\beta\in\mathcal{A}} p_{\alpha\beta} = 1 \ , \ p_{\alpha\beta} \geq 0 \, .$$

Definition 1.1.1 *The pair* $(\mathcal{A}, \mathbf{P})$ *is called a discrete time homogeneous Markov chain* (MC).

A *path* ω is any sequence

$$\omega = (\omega_0, \omega_1, \omega_2, \ldots),$$

where

$$\omega_i \in \mathcal{A} \, , \, \forall i \geq 0.$$

The *path space* $\Omega = \mathcal{A}^{\mathbf{N}}$ is the set of all paths and Σ is the standard σ-algebra generated by the cylinder sets

$$(\alpha_0 \, , \, \alpha_1, \ldots, \alpha_n) = \{\omega : \omega_i = \alpha_i \, , \, 0 \leq i \leq n\}, \; n > 0, \; \alpha_i \in \mathcal{A}.$$

Occasionally, it will be necessary to consider MC with a fixed initial distribution. Therefore we give the following

Definition 1.1.2 *We call an MC with initial distribution* $p_0(\alpha), \alpha \in \mathcal{A}$, $\sum_\alpha p_0(\alpha) = 1, p_0(\alpha) \geq 0$, *a probability measure* P *defined on* (Ω, Σ) *such that, for all cylinder sets* $(\alpha_0, \alpha_1, \ldots, \alpha_n)$,

$$P(\alpha_0, \alpha_1, \ldots, \alpha_n) = p_0(\alpha_0) p_{\alpha_0\alpha_1} \cdots p_{\alpha_{n-1}\alpha_n} \, . \tag{1.1}$$

The random variable $\xi_n(\omega) = \omega_n$, defined on (Ω, Σ, P) and taking its values in $\mathcal{A}$, will be called *the value of the chain at time n*, or *the position of the chain at time n*, etc. We shall simply write ξ_n, *ad libitum* and whenever unambiguous; ξ_0 is called an initial state. If there exists a sequence $\alpha_1, \alpha_2, \ldots, \alpha_{n-1}$ such that $p_{\alpha\alpha_1} p_{\alpha_1\alpha_2} \cdots p_{\alpha_{n-1}\beta} > 0$, we shall write $\alpha \rightsquigarrow \beta$.
Let us denote by $p^{(k)}_{\alpha\beta}$ the k-step transition probabilities, i.e. the elements of the matrix $\mathbf{P}^k$.

Definition 1.1.3 *The point α is called an inessential state of the MC, iff there exists a point β such that $\alpha \rightsquigarrow \beta$ but $\beta \not\rightsquigarrow \alpha$. All other states are called essential.*

It is easy to show that, for any initial state and any inessential state α, there exists a random time $N(\omega) < \infty$ such that ξ_n never equals α for $n > N(\omega)$, a.s. We shall write $\alpha \Longleftrightarrow \beta$ iff $\alpha \rightsquigarrow \beta$ and $\beta \rightsquigarrow \alpha$. The operation '$\Longleftrightarrow$' is obviously transitive. Sometimes we shall also say that α and β *communicate.*

Definition 1.1.4 *An equivalence class with respect to the operation '$\Longleftrightarrow$' is called an* essential class. *A Markov chain is called* irreducible *iff every state can be reached from any other state or, equivalently, if, and only if, $\mathcal{A}$ forms a single class of communicating states, which then are all essential.*

It is not difficult to prove that, for any initial distribution, there exists $N(\omega)$ such that all ξ_n's belong to the same essential class, for $n > N(\omega)$, almost surely (a.s.). As we shall be mainly interested in the long run behaviour of all random processes which will be encountered, from now on and for the rest of the book, *the Markov chain* $(\mathcal{A}, \mathbf{P})$ *will be assumed to be irreducible.*

Choose now $\alpha \in \mathcal{A}$. Let $n_1(\alpha) < n_2(\alpha) < \ldots$ be all the positive integers for which $p^{(n_i)}(\alpha, \alpha) > 0$, $i = 1, 2 \ldots$.

Definition 1.1.5 (Theorem) *Let us denote by $d(\alpha)$ the greatest common divisor of the $n_i(\alpha)$, $i \geq 1$. Then $d(\alpha)$ indeed does not depend on α and is called the* period *of the (irreducible) chain $\mathcal{A}$. If $d = 1$, the chain is called* aperiodic.

In the sequel, we shall consider only aperiodic chains, but all the theory can easily be transcribed with minor modifications to include the periodic case. In fact, it suffices to consider ξ_n at embedded instants $n = k + dm$, for some fixed k. It is also useful to keep in mind that, if for some α, $p_{\alpha\alpha} > 0$, then the chain is aperiodic. Unless otherwise stated, all the chains studied hereafter will be assumed to be *irreducible and aperiodic.*

1.2 Classification

Let $\alpha, \beta \in \mathcal{A}$. We define now, for $n \geq 1$,

$$f_n(\alpha, \beta) = \mathbf{P}(\xi_k(\omega) \neq \beta, 0 < k < n; \xi_n(\omega) = \beta / \xi_0(\omega) = \alpha) ,$$

the probability that the MC first enters into state β at time n, given that it starts from the state α. Then

$$Q(\alpha, \beta) = \sum_{n=1}^{\infty} f_n(\alpha, \beta)$$

is the probability that, starting at α, the MC ever visits β. Accordingly,

$$m_{\alpha\beta} = \sum_{n=1}^{\infty} n f_n(\alpha, \beta)$$

is the mean time of first reaching β when starting at α. Clearly $m_{\alpha\beta} = \infty$ if $Q(\alpha, \beta) < 1$.

Theorem 1.2.1 *If $Q(\alpha, \beta) = 1$ for some pair (α, β), then $Q(\alpha, \beta) = 1$ for all (α, β). Similarly, if $m_{\alpha\beta} + m_{\beta\alpha} = \infty$ for some (α, β), then $m_{\alpha\beta} + m_{\beta\alpha} = \infty$, for all (α, β) (in either instance, α and β need not be distinct).* ■

Definition 1.2.2 *An irreducible aperiodic MC is called*

(i) recurrent *if $Q(\alpha, \beta) = 1$, at least for one pair (α, β);*
(ii) non recurrent *or* transient *if $Q(\alpha, \beta) < 1$, $\forall (\alpha, \beta)$;*
(iii) positive recurrent *or* ergodic, *if $m_{\alpha\beta} + m_{\beta\alpha} < \infty$, at least for one pair (α, β);*
(iv) null recurrent *if $Q(\alpha, \beta) = 1$ and $m_{\alpha,\beta} = \infty$, at least for one pair (α, β).*
(v) non ergodic *if $m_{\alpha,\beta} = \infty$, at least for one pair (α, β).*

The purpose of the next theorems is to give other useful (equivalent) criteria for an MC to be ergodic. We consider the equation

$$\pi = \pi \mathbf{P} \text{ or, equivalently, } \pi_\beta = \sum_\alpha \pi_\alpha p_{\alpha\beta} , \tag{1.2}$$

where π is the unknown vector

$$\pi = (\pi_\alpha, \alpha \in \mathcal{A}) .$$

Theorem 1.2.3 *The limits*

$$v_\beta = \lim_{n\to\infty} p^{(n)}_{\alpha\beta} \ , \quad \forall \beta \in \mathcal{A} \, , \tag{1.3}$$

exist and are independent of the initial state α. Futhermore, when the MC is non-ergodic, $v_\beta = 0 \, , \ \forall\beta$.
When the MC is ergodic, then we have

$$v_\beta > 0 \ , \quad \sum_\beta v_\beta = 1$$

and

$$v_\beta = \sum_\alpha v_\alpha p_{\alpha\beta} \ ,$$

i.e. the vector v is a probabilistic solution of (1.2). ■

Theorem 1.2.4 *The following conditions are equivalent:*

(i) *the MC is ergodic;*
(ii) *there exists a unique* l^1*-solution of the equation (1.2), up to a multiplicative factor;*
(iii) *there exists a unique stationary distribution* $(\pi_\alpha, \alpha \in \mathcal{A})$*, i.e. a solution of (1.2) such that* $\pi_\alpha \geq 0$, $\sum_\alpha \pi_\alpha = 1$. *In this case* $\pi_\alpha > 0 \, , \ \forall \alpha \in \mathcal{A}$,

and

$$\pi_\alpha = \lim_{n\to\infty} p^{(n)}_{\gamma\alpha} \ , \ \forall \gamma \in \mathcal{A} \, . \tag{1.4}$$

■

Theorem 1.2.5 *For an ergodic MC, the invariant distribution is given by*

$$\pi_\alpha = \frac{1}{m_{\alpha\alpha}} \ , \quad \forall \alpha \in \mathcal{A} \, . \tag{1.5}$$

■

1.3 Continuous time

Many examples seem more natural in continuous time. Later on we introduce the necessary notation to the extent we need. But we want to stress immediately that all results concerning the classification in discrete time are automatically transposed into continuous time and vice versa.

There are two main definitions of a continuous time homogeneous countable MC (the set of states is still denoted by $\mathcal{A}$). In both cases, the *intensity matrix* $H = (\lambda_{\alpha\beta})$ is given by

$$\begin{cases} \lambda_{\alpha\beta} \geq 0, \ \alpha \neq \beta \,, \\ \lambda_{\alpha\alpha} = - \sum_{\beta:\beta\neq\alpha} \lambda_{\alpha\beta} \,. \end{cases} \tag{1.6}$$

For the examples we shall consider, it suffices to assume the existence of a constant $C > 0$ such that, for all α, β,

$$|\lambda_{\alpha\alpha}| < C \,. \tag{1.7}$$

Then the matrix

$$\| \, p_{\alpha\beta}(t) \, \| = P(t) = e^{Ht} = \sum_{n=0}^{\infty} H^n \, \frac{t^n}{n!} \,, \ t \geq 0 \,, \tag{1.8}$$

is defined by the convergent series (1.8) .

Definition 1.3.1 *The MC ξ_t, with initial distribution $p_\alpha(0)$, is defined by the following finite-dimensional distributions, for all $0 < t_1 < \ldots < t_n$:*

$$P(\xi_0 = \alpha_0, \ldots, \xi_{t_n} = \alpha_n) = p_{\alpha_0}(0) p_{\alpha_0\alpha_1}(t_1) \ldots p_{\alpha_{n-1}\alpha_n}(t_n). \tag{1.9}$$

This definition does not depend on the choice of the probability space. The next one uses a concrete choice. We define Ω to be the set of right-continuous piecewise constant mappings $\omega : [0, \infty) \to \mathcal{A}$, i.e. ω is given by a sequence $(\alpha_0, 0), (\alpha_1, \tau_1), \ldots$, such that

$$\omega(t) = \alpha_i, \ t \in [\tau_i, \tau_{i+1}), \tau_0 = 0 \,,$$

where $\tau_1, \tau_2, \ldots$, are the *jump times*. The measure on Ω, corresponding to the MC (using the standard canonical σ-algebra, see for example [Chu67, GS74]), is defined by the following conditions:

(i) Given $\alpha_0, \alpha_1, \ldots$, the random variables $\tau_{i+1} - \tau_i$ are mutually independent and have an exponential distribution with parameters $-\lambda_{\alpha_i\alpha_i}$;

(ii) $\alpha_0, \alpha_1, \ldots, \alpha_n, \ldots$, are distributed as an *embedded* discrete time homogeneous MC, with parameters

$$p_{\alpha\beta} = \frac{\lambda_{\alpha\beta}}{(-\lambda_{\alpha\alpha})}. \tag{1.10}$$