Pablo A. Ferrari
Antonio Galves

# Construction of Stochastic Processes, Coupling and Regeneration

Pablo A. Ferrari, Antonio Galves
Instituto de Matemática e Estatística,
Universidade de São Paulo,
Caixa Postal 66281,
05315-970 - São Paulo,
BRAZIL
email: pablo@ime.usp.br, galves@ime.usp.br
Homepage: http://www.ime.usp.br/~pablo

# Contents

# Preface

These notes present in an elementary way a set of notions, results and basic models useful to work with stochastic processes. The approach does not use Measure Theory as it adopts a *constructive* point of view inherent to the notion of *simulation* and *coupling* of random variables or processes.

To couple two random variables means to construct them simultaneously using the same random mechanism. More informally: coupling is just to simulate two random variables using the same random numbers. The first coupling was introduced by Doeblin (1938) to show the convergence to equilibrium of a Markov chain. Doeblin considered two independent trajectories of the process, one of them starting with an arbitrary distribution and the other with the invariant measure and showed that the trajectories meet in a finite time.

Perhaps due to the premature and tragical death of Doeblin and the extreme originality of his ideas, the notion of coupling only come back to the literature with Harris (1955). Coupling become a central tool in Interacting particle systems, subject proposed by Spitzer (1970) and the sovietic school of Dobrushin, Toom, Piatvisky-Shapiro, Vaserstein and others. These names give rise to a new area in stochastic processes, the so called *Interacting Particle Systems*, then developed extensively by Holley, Liggett, Durrett, Griffeath and others. We refer the interested reader to the books by Liggett (1985), (1999) and Durrett (1988) and (1997) and Kipnis and Landim (1999) for recent developments in the field. We learned and used coupling and constructive techniques from those authors when working in particle systems as the exclusion and contact processes. Our constructive approach comes directly from the *graphical construction* of interacting particle systems introduced by

Harris (1972).

Coupling techniques had a somehow independent development for "classical" processes. The books of Lindvall (1992) and the recent book of Thorisson (2000) are excellent sources for these developments.

The art of coupling consists in looking for the best way to simultaneously construct two processes or, more generally, two probability measures. For instance, to study the convergence of a Markov chain, we construct simultaneously two trajectories of the same process starting at different states and estimate the time they need to meet. This time depends on the joint law of the trajectories. The issue is then to find the construction "minimizing" the meeting time. In the original Doeblin's coupling the trajectories evolved independently. This coupling is *a priori* not the best one in the sense that it is not aimed to reduce the meeting time. But once one realizes that coupling is useful, many other constructions are possible. We present some of them in these notes. A discussion about the velocity of convergence and of the so called Dobrushin ergodicity coefficient is presented in Chapter 3.

The central idea behind coupling can be presented through a very simple example. Suppose we toss two coins, and that the probability to obtain a "head" is $p$ for the first coin and $q$ for the second coin with $0 < p < q < 1$. We want to construct a random mechanism simulating the simultaneous tossing of the two coins in such a way that when the coin associated to the probability $p$ shows "head", so does the other (associated to $q$). Let us call $X$ and $Y$ the results of the first and second coin, respectively; $X, Y \in \{0, 1\}$, with the convention that "head" $= 1$. We want to construct a random vector $(X, Y)$ in such a way that

$$\begin{aligned} \mathbb{P}(X = 1) &= p = 1 - \mathbb{P}(X = 0) \\ \mathbb{P}(Y = 1) &= q = 1 - \mathbb{P}(Y = 0) \\ X &\le Y. \end{aligned}$$

The first two conditions just say that the marginal distribution of $X$ and $Y$ really express the result of two coins having probabilities $p$ and $q$ of being "head". The third condition is the property we want the coupling to have. This condition implies in particular that the event

$$\{X = 1,\ Y = 0\},$$

corresponding to a head for the first coin and a tail for the second, has probability zero.

To construct such a random vector, we use an auxiliary random variable $U$, uniformly distributed in the interval $[0,1]$ and define

$$X := \mathbf{1}\{U \le p\} \quad \text{and} \quad Y := \mathbf{1}\{U \le q\}.$$

where $\mathbf{1}A$ is the indicator function of the set $A$. It is immediate that the vector $(X,Y)$ so defined satisfies the three conditions above. This coupling is a prototype of the couplings we use in this notes.

With the same idea we construct stochastic processes (sequences of random variables) and couple them. One important product of this approach is the *regenerative construction* of stochastic processes. For instance, suppose we have a sequence $(U_n : n \in \mathbb{Z})$ of independent, identically distributed uniform random variables in $[0,1]$. Then we construct a process $(X_n : n \in \mathbb{Z})$ on $\{0,1\}^{\mathbb{Z}}$, using the rule

$$X_n := \mathbf{1}\{U_n > h(X_{n-1})\} \tag{1}$$

where $h(0) < h(1) \in (0,1)$ are arbitrary. We say that there is a *regeneration time* at $n$ if $U_n \in [0,h(0)] \cup [h(1),1]$. Indeed, at those times the law of $X_n$ is given by

$$\mathbb{P}(X_n = 1 \mid U_n \in [0,h(0)] \cup [h(1),1]) = \frac{1-h(1)}{h(0)+1-h(1)} \tag{2}$$

independently of the past. Definition (1) is incomplete in the sense that we need to know $X_{n-1}$ in order to compute $X_n$ using $U_n$. But, if we go back in time up to $\tau(n) := \max\{k \le n : U_k \in [0,h(0)] \cup [h(1),1]\}$, then we can construct the process from time $\tau(n)$ on. Since this can be done for all $n \in \mathbb{Z}$, we have constructed a stationary process satisfying:

$$\mathbb{P}(X_n = y \mid X_{n-1} = x) = Q(x,y) \tag{3}$$

where $Q(0,0) = h(0)$, $Q(0,1) = 1-h(0)$, $Q(1,0) = h(1)$ and $Q(1,1) = 1-h(1)$.

Processes with this kind of property are called *Markov chains*. The principal consequence of construction (1) is that the pieces of the process between two regeneration times are independent random vectors (of random length). We use this idea to construct *perfect simulation* algorithms of the Markov chain.

Regenerative schemes have a long history, starting with Harris (1956) approach to recurrent Markov chains in non countable state-spaces passing by the basic papers by Athreya and Ney (1978) and Nummelin (1978). We refer the reader to Thorisson (2000) for a complete review. Perfect simulation was recently proposed by Propp and Wilson (1996) and become very fast an important issue of research. See Wilson (1998).

Markov chains are introduced in Chapter 1, further properties are proven in Chapter 2. Coupling is discussed in Chapter 3 and the regeneration scheme in Chapter 4.

A Markov chain is a process with short memory. It only "remembers" last state in the sense of (1). In processes with "long memory" the value the process assumes at each step depends on the entire past. These kind of processes has been introduced in the literature by Onicescu and Mihoc (1935) with the name *chains with complete connections* (*chaînes à liaisons complètes*), then studied by Doeblin and Fortet (1937), Harris (1955) and the Rumanian school. We refer the reader to Iosifescu and Grigorescu (1990) for a complete survey. In Chapter 6 we show a regeneration scheme and a stationary construction of these processes.

In Chapter 8 we treat *continuous time* Markov processes. Here the role of the uniform random variables $U_n$ is played by a bi-dimensional Poisson process. The study of Poisson processes is done in Chapter 7.

We conclude the description of contents of the book with an important remark. In this text we tried to remain at an elementary level. We assume without further discussion that there exists a double infinite sequence of independent random variables uniformly distributed in $[0,1]$. This is all we need to construct all the processes studied here.

In these notes we adopt the *graphic construction* philosophy introduced by Ted Harris to deal with interacting particle systems. He taught us how to construct particle systems using random graphs, cutting and pasting pieces so that to put in evidence, in the most elementary way, the properties of the process. For all this influence and inspiration, we dedicate these notes to him.

These notes are originated in part from the courses in Stochastic Processes we give in the Instituto de Matemática e Estatística da Universidade de São Paulo. Part of these notes appeared as the booklet *Acoplamento em processos estocásticos* in Portuguese for a mini-course we offered in the XXI Coloquio Brasileiro de Matemática, held in Rio de Janeiro in July of 1997 [Ferrari and Galves (1997)]. Besides a revision of errors and misprints, these notes contain two new chapters: Chapter 4, Regeneration and perfect simulation and Chapter 6, Chains with complete connections. To keep the focus we omitted two chapters of the Portuguese version: one about Queueing Theory and the other about Interacting Particle Systems.

## Acknowledgements

These notes have been are written for the XIII Escuela Venezolana de Matemática. We thank Stella Brassesco and Carlos de Prisco, for the invitation to present a course in this school.

We thank Lane Pereira Alencar, Hervé Guiol, Ricardo Maronna and Tomas Tetzlaff for comment and pointing out errors in the Portuguese version of this notes. We thank Xavier Bressaud, Francis Comets, Davide Gabriele, Nancy Lopes Garcia, Alejandro Maass, Servet Martinez and Peter Ney with whom we have learned together many of the issues of this book. A special thanks to Roberto Fernández for his permanent inquiring and collaboration in the elaboration and presentation of the ideas cristalized in these notes.

PAF thanks support from the Guggenheim Foundation. PAF and AG thank FINEP, FAPESP, CNPq for support during the writing of these notes. PAF thanks Joachim Krug and DAAD for providing the computer where these notes have being written.

# Chapter 1

# Construction of Markov chains

## 1.1 Markov chains.

A stochastic process is a family $(X_t : t \in \mathbb{T})$ of random variables. The label $t$ represents time. Informally speaking, a stochastic process describes the history of some random variable which evolves in time. In the first chapters of this book we concentrate in discrete time, *i.e.* $\mathbb{T} = \mathbb{N}$ or $\mathbb{T} = \mathbb{Z}$. We assume that the process takes its values in the set $\mathcal{X}$, called *state space*. We assume $\mathcal{X}$ finite or countable.

When $\mathbb{T}$ is discrete, the *law* of a stochastic process is characterized by the finite-dimensional distributions

$$\mathbb{P}(X_t = x_t \, : \, t \in T) \tag{1.1}$$

for arbitrary $T \subset \mathbb{T}$ and $x_t \in \mathcal{X}$, $t \in T$.

Let $\underline{U} = (U_n, \, n \in \mathbb{Z})$ be a double infinite sequence of independent random variables with uniform distribution in the interval $[0, 1[$. From now on we refer to $\mathbb{P}$ and $\mathbb{E}$ for the probability and the expectation with respect to the sequence $(U_n)$.

**Definition 1.2 (Markov Chain)** A process $(X_n^a)_{n \in \mathbb{N}}$ with state-space $\mathcal{X}$ is a *Markov chain* with initial state $a \in \mathcal{X}$ if there exists a function $F$ :

$\mathcal{X} \times [0,1] \to \mathcal{X}$ such that $X_0^a = a$ and for all $n \geq 1$,

$$X_n^a = F(X_{n-1}^a; U_n)\,. \tag{1.3}$$

In the sequel we may drop the super label $a$ when the initial state is not relevant for the discussion. The above definition refers to *time homogeneous* Markov chains. In non homogeneous Markov chains the transitions depend on time: these chains would be characterized by a family $(F_n)_{n\in\mathbb{T}}$ and the definition would be $X_n^a = F_n(X_{n-1}^a; U_n)$. But in these notes we will treat only the homogeneous case.

**Example 1.4** Let $\mathcal{X} = \{0,1\}$ and

$$F(x;u) = \mathbf{1}\{u > h(x)\}\,, \tag{1.5}$$

where $h(0)$ and $h(1)$ are arbitrary numbers in $[0,1]$. Informally speaking, at each instant $n$ the process updates its value to 0 or 1, accordingly to $U_n$ being less or bigger than $h(X_{n-1})$.

**Example 1.6** Let $\mathcal{X} = \{0,1\}$,

$$F(x;u) \;=\; \begin{cases} 1-x, & \text{if } u > g(x) \\ x & \text{otherwise} \end{cases} \tag{1.7}$$

where $g(0)$ and $g(1)$ are arbitrary numbers in $[0,1]$. Informally speaking, at each instant $n$ the process decides to change or keep the current value according to $U_n$ being bigger or smaller than $g(X_{n-1})$.

**Basic properties.** These examples will serve as a laboratory to present some of the central results of the theory. We present now some of their basic properties. Let

$$Q(x,y) = \mathbb{P}(X_n = y \mid X_{n-1} = x)\;, \tag{1.8}$$

for $n \geq 1$ and $x, y \in \{0,1\}$.

Let $Q_h$ and $Q_g$ be the conditional probabilities corresponding to the first and second example respectively. If $h(0) = h(1)$, then

$$Q_h(0,y) = Q_h(1,y). \tag{1.9}$$

In this case the Markov chain corresponding to the function $h$ is just a sequence of independent and identically distributed (*iid* in the sequel) random variables.

On the other hand, if $h(0) = g(0)$ and $1 - h(1) = g(1)$, then

$$Q_g(x,y) = Q_h(x,y) \tag{1.10}$$

for all $x$ and $y$. This identity means that if we choose the same initial state for the chains defined by (1.5) and (1.7), then the processes have the same joint law: calling $X_t^a$ and $Y_t^b$ the processes constructed with the functions (1.5) and (1.9) and initial states $a$ and $b$ respectively,

$$\mathbb{P}(X_t^a = x_t\,,\, t \in T) \;=\; \mathbb{P}(Y_t^a = x_t\,,\, t \in T) \tag{1.11}$$

for any finite set $T \subset \mathbb{N}$ and arbitrary values $x_t \in \mathcal{X}$. From this point of view, the constructions are equivalent. Sometimes definition (1.5) is more useful because it is related to the notion of *coupling*. We give the formal definition of coupling in Chapter 3, but to give the reader the taste of it, we present a simple example.

**Example of Coupling.** Let $(X_n^0)_{n\in\mathbb{N}}$ and $(X_n^1)_{n\in\mathbb{N}}$ be two chains constructed with the function (1.3) using the *same* sequence $U_1, U_2, \ldots$ of uniform random variables with initial states 0 and 1 respectively. That is

$$\begin{aligned} (X_0^0, X_0^1) &= (0,1) \\ (X_n^0, X_n^1) &= (F(X_{n-1}^0; U_n), F(X_{n-1}^1; U_n)) \end{aligned} \tag{1.12}$$

Let $\tau$ be the meeting time of the chains. That is,

$$\tau := \min\{n \geq 1 : X_n^0 = X_n^1\} \tag{1.13}$$

**Lemma 1.14** *Assume the chains $(X_n^0, X_n^1)$ are constructed with the procedure described by* (1.12) *using the function $F$ defined in* (1.5) *with $0 < h(0) < h(1) < 1$ and denote $\rho := h(1) - h(0)$. Then*

$$\mathbb{P}(\tau > n) = \rho^n\ , \tag{1.15}$$

*that is, $\tau$ has geometric distribution with parameter $\rho$. Furthermore*

$$n \geq \tau \ \ \textit{implies}\ \ X_n^0 = X_n^1 \tag{1.16}$$

In other words, after a random time with geometric distribution the chains $(X_n^0)_{n\in\mathbb{N}}$ and $(X_n^1)_{n\in\mathbb{N}}$ are indistinguishable. This result is the prototype of *loss of memory.*

**Proof.** It is left as an exercise for the reader. □

We computed above the conditional probability for the process to be at some state at time $n$ given it was at another state at the previous time. We obtained a function $Q : \mathcal{X} \times \mathcal{X} \to [0,1]$. This type of function is useful to characterize Markov chains.

**Definition 1.17 (Transition matrix)** A function $Q : \mathcal{X} \times \mathcal{X} \to [0,1]$ is called a *transition matrix* if for all element $x \in \mathcal{X}$, $\sum_{y\in\mathcal{X}} Q(x,y) = 1$. In other words, if the sum of the entries of each row of the matrix equals one.

In the examples (1.5) and (1.7) the property $\sum_{y\in\mathcal{X}} Q(x,y) = 1$ for all fixed $x$ is a consequence of the fact that $Q(x,\cdot)$ was a conditional probability.

We use the matrix representation of the function $Q$. For instance, if $\mathcal{X} = \{1,2\}$ and $Q(1,1) = p$, $Q(1,2) = 1-p$, $Q(2,1) = 1-q$ and $Q(2,2) = q$, for some $p$ and $q$ in $[0,1]$, then it is natural the representation

$$Q = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix} \tag{1.18}$$

**Proposition 1.19** *Each transition matrix $Q$ on $\mathcal{X}$ and $a \in \mathcal{X}$ defines a Markov chain $(X_n^a)_{n\in\mathbb{N}}$ with transition probabilities given by $Q$ and initial state $a$. That is, there exists a function $F$ such that the Markov chain* (1.3) *satisfies* (1.8).

**Proof.** We need to exhibit the function $F : \mathcal{X} \times [0,1] \to \mathcal{X}$ of definition (1.3) with the property

$$\mathbb{P}(F(x;u) = y) = Q(x,y) \tag{1.20}$$

We saw in the examples that there are various different manners of constructing such a $F$. We propose now a general construction.

For each $x \in \mathcal{X}$ we construct a partition of $[0,1]$: let $(I(x,y) : y \in \mathcal{X})$ be a family of Borel sets (in these notes $I(x,y)$ will always be a finite or countable union of intervals) satisfying

$$\begin{aligned} & I(x,y) \cap I(x,z) = \emptyset \text{ if } y \neq z\,, \\ & \bigcup_{y\in\mathcal{X}} I(x,y) = [0,1] \end{aligned} \tag{1.21}$$

Calling $|I|$ the Lebesgue measure (length) of the set $I$, we ask

$$|I(x,y)| \;=\; Q(x,y) \tag{1.22}$$

For instance, in (1.5) the partition is $I(x,0) = [0,h(x))$ and $I(x,1) = [h(x),1]$.

There are many ways of defining the partitions. The simplest is just to order the states of $\mathcal{X}$ —we can identify $\mathcal{X} = \{1,2,\ldots\}$— and concatenate intervals of length $Q(x,y)$.

With the partition in hands we define the function $F$ as follows.

$$F(x;u) = \sum_{y\in\mathcal{X}} y\,\mathbf{1}\{u \in I(x,y)\}\;. \tag{1.23}$$

In other words, $F(x;u) = y$, if and only if $u \in I(x,y)$. We construct the chain $(X^a_n)_{n\in\mathbb{N}}$ using definition (1.3) with the function defined in (1.23). To see that this chain has transition probabilities $Q$, compute

$$\mathbb{P}(X_n = y \mid X_{n-1} = x) = \mathbb{P}(F(x;U_n) = y) \tag{1.24}$$

By construction, this expression equals

$$\mathbb{P}(U_n \in I(x,y)) = |I(x,y)| = Q(x,y).\,\square \tag{1.25}$$

Proposition 1.19 says that for any process $(Y_n)_{n\in\mathbb{N}}$ satisfying (1.27) it is possible to construct another process $(\bar{Y}_n)_{n\in\mathbb{N}}$ with the same law using the algorithm (1.3). Many of the results in the following chapters rely on a smart construction of the partitions $((I(x,y) : y \in \mathcal{X}) : x \in \mathcal{X})$. Proposition 1.19 motivates the following theorem.

**Theorem 1.26 (Markov Chain)** *A stochastic process* $(Y_n)_{n\in\mathbb{N}}$ *with state space* $\mathcal{X}$ *is a Markov chain with transition matrix* $Q$*, if for all* $n \geq 1$ *and every finite sequence* $x_0, x_1, \cdots, x_n$ *contained in* $\mathcal{X}$ *such that* $\mathbb{P}(Y_0 = x_0, \cdots, Y_{n-1} = x_{n-1}, Y_n = x_n) > 0$*, it holds*

$$\begin{aligned} \mathbb{P}(Y_n = x_n \mid Y_0 = x_0, \cdots, Y_{n-1} = x_{n-1}) & \qquad (1.27) \\ = \mathbb{P}(Y_n = x_n \mid Y_{n-1} = x_{n-1}) & \qquad (1.28) \\ = Q(x_{n-1}, x_n). & \qquad (1.29) \end{aligned}$$

**Proof.** Follows from Proposition 1.19. □

The theorem says that, in a Markov chain, the forecast of the next step knowing all the past is as good as when one knows only the current value of the process. The statement of the theorem is what most books take as definition of Markov chain.

As a consequence of Theorem 1.26, the joint law of a Markov chain is given by:

$$\begin{aligned} & \mathbb{P}(Y_0 = x_0, \cdots, Y_{n-1} = x_{n-1}, Y_n = x_n) \\ = \;& \mathbb{P}(Y_0 = x_0)\,\mathbb{P}(Y_1 = x_1, \cdots, Y_{n-1} = x_{n-1}, Y_n = x_n \,|\, Y_0 = x_0) \\ = \;& \mathbb{P}(Y_0 = x_0)\,Q(x_0, x_1)\ldots Q(x_{n-1}, x_n) \qquad (1.30) \end{aligned}$$

## 1.2 Examples

We finish this chapter with some important examples.

**Example 1.31 (Random walk in the hypercube)** Let $N$ be a positive integer and $\mathcal{X} = \{0,1\}^N$. If $N = 2$ we can think of $\mathcal{X}$ as being the set of vertices of a square. When $N = 3$ we can think of $\mathcal{X}$ as the set of vertices of a cube. When $N \geq 4$ we think of $\mathcal{X}$ as the set of vertices of a hypercube. Let $\xi = (\xi(1), \cdots, \xi(N))$ be an element of $\mathcal{X}$. The *neighbors* of $\xi$ are those elements of $\mathcal{X}$ having all but one coordinates equal to the coordinates of $\xi$. If $j$ is an integer in $[1, n]$, we call $\xi^j$ the element of $\mathcal{X}$ having all coordinates equal to $\xi$ but the $j$-th:

$$\xi^j(i) = \begin{cases} \xi(i) & \text{if } i \neq j, \\ 1 - \xi(j) & \text{if } i = j. \end{cases} \tag{1.32}$$

In that way the neighbors of $\xi$ are the elements $\xi^1, \ldots, \xi^N$.

This induces a natural notion of distance between elements of $\mathcal{X}$:

$$d(\xi, \zeta) = \sum_{i=1}^{N} |\xi(i) - \zeta(i)|$$

for $\xi$, $\zeta \in \mathcal{X}$. The distance between $\xi$ and $\zeta$ is the number of coordinates for which $\xi$ and $\zeta$ are different. This is know as *Hamming's distance.* Two elements of $\mathcal{X}$ are neighbors when they are at distance one.

Now we want to construct a Markov chain on $\mathcal{X}$ with the following behavior. At each time the process decides to change state (or not) according to the result of a fair coin. If it decides to change, then it jumps to one of its neighbors with equal probability. This process can be constructed in the following way:

For $a \in \{0, 1\}$ let

$$\xi^{j,a}(i) = \begin{cases} \xi(i) & \text{if } i \neq j, \\ a & \text{if } i = j, \end{cases} \tag{1.33}$$

be a configuration with value $a$ in the $j$-th coordinate and equal to $\xi$ in the other coordinates. Let

$$I(\xi, \zeta) := \begin{cases} \Big[\frac{j-1}{N}, \frac{j-1}{N} + \frac{1}{2N}\Big) & \text{if } \zeta = \xi^{j,0},\ j = 1, \ldots, N \\ \Big[\frac{j-1}{N} + \frac{1}{2N}, \frac{j}{N}\Big) & \text{if } \zeta = \xi^{j,1},\ j = 1, \ldots, N \\ \emptyset & \text{if } d(\xi, \zeta) > 1 \end{cases} \tag{1.34}$$

Then use the definitions (1.3) and (1.23). The probability of trial site $j$ at time $n$ is

$$\mathbb{P}\Big(U_n \in \Big[\frac{j-1}{N}, \frac{j}{N}\Big)\Big) = 1/N$$

and the probability to assign the value 0 to the chosen site is

$$\mathbb{P}\Big(U_n \in \Big[\frac{j-1}{N}, \frac{j-1}{N} + \frac{1}{2N}\Big) \;\Big|\; U_n \in \Big[\frac{j-1}{N}, \frac{j}{N}\Big)\Big) = 1/2$$

Analogously, the probability to assign the value 1 to the chosen site is 1/2. With these expressions in hand we can compute $Q(\xi,\zeta)$. We know that this is zero if $\zeta$ is not neighbor of $\xi$. Assume $\xi(j)=1$ and compute

$$Q(\xi,\xi^j) \;=\; \mathbb{P}\Big(U_1 \in \Big[\frac{j-1}{N},\frac{j-1}{N}+\frac{1}{2N}\Big)\Big) \;=\; \frac{1}{2N} \tag{1.35}$$

Analogously, assuming $\xi(j)=0$,

$$Q(\xi,\xi^j) \;=\; \mathbb{P}\Big(U_1 \in \Big[\frac{j-1}{N}+\frac{1}{2N},\frac{j}{N},\Big)\Big) \;=\; \frac{1}{2N} \tag{1.36}$$

Now we want to use this construction for two chains $\xi_t$ and $\xi'_t$ with different initial configurations $\xi$ and $\xi'$ and compute the meeting time $\tau$ as defined in (1.13). The point is that if $U_n$ belongs to the interval $[\frac{j-1}{N},\frac{j}{N}[$, then the $j$th coordinate of the two processes coincide: $U_n \in [\frac{j-1}{N},\frac{j}{N}[$ implies $\xi_t(j)=\xi'_t(j)$ for all $t\ge n$. Hence defining

$$\tau_j \;:=\; \min\{n \ge N \,:\, U_n \in [\tfrac{j-1}{N},\tfrac{j}{N})\} \tag{1.37}$$

we have

$$\tau \;\le\; \max_j \tau_j\,. \tag{1.38}$$

It is ease to see that $\max_j \tau_j$ is a sum of geometric random variables of varying parameter:

$$\max_j \tau_j \;=\; \sigma_1+\ldots,+\sigma_N \tag{1.39}$$

where $\sigma_1=1$, and $\sigma_i$ is a geometric random variable of parameter $(N-i+1)/N$:

$$\mathbb{P}(\sigma_i=k) = \Big(\frac{N-i+1}{N}\Big)\Big(1-\Big(\frac{N-i+1}{N}\Big)\Big)^{k-1};\quad i\ge 1 \tag{1.40}$$

The variables $\sigma_i$ represent the number of new uniform random variables one needs to generate to visit one of the intervals $[\frac{j-1}{N},\frac{j}{N}[$ not previously visited. In this way

$$\mathbb{E}\tau \;\le\; \mathbb{E}(\min_j \tau_j) \;=\; 1+\frac{N}{N-i+1}+\ldots+N \;\sim\; N\log N \tag{1.41}$$

**Example 1.42 (Ehrenfest model)** The random walk on the hypercube can be seen as a caricature of the evolution of a gas between two containers. Initially all the molecules of the gas are confined in one of the containers which are labeled 0 and 1. The experiment starts when a valve intercommunicating the containers is open and the molecules start passing from one container to the other. We have $N$ molecules of gas, each one belonging to one of the containers. The number of molecules is of the order of $10^{23}$. The random walk describes the position of each one of the $N$ molecules at each instant of time: the number $\xi_n^\zeta(i)$ is the label of the container the molecule $i$ belongs to at time $n$. The initial configuration $\zeta$ describes the initial position of the molecules, that is, $\zeta(i) = 0$ for all $i = 1, \ldots, N$ (all molecules belong to container 0).

Sometimes one is interested only on the total number of molecules in each container at each time. This process has state-space $\{0, 1, \ldots, N\}$ and is defined by

$$Z_n = \sum_{i=1}^{N} \xi_n^\zeta(i) \ . \qquad (1.43)$$

$Z_n$ indicates the number of molecules in container 1 at time $n$. We can say that the Ehrenfest model is a *macroscopic* description of the *microscopic* model corresponding to the random walk in the hypercube. This process can be seen as a random walk in $\{0, \ldots, N\}$.

**Example 1.44 (The-house-of cards process)** For this process the state space is $\mathcal{X} = \mathbb{N}$. Let $(a_k : k \in \mathbb{N})$ be a sequence of numbers in $[0, 1]$. Informally speaking when the process is at site $k$ it jumps with probability $a_k$ to site $k + 1$ and with probability $1 - a_k$ to site 0. For this process the partitions are given by

$$I(k, k+1) \,=\, [0, a_k) \ ; \qquad I(k, 0) \,=\, [a_k, 1] \qquad (1.45)$$

for $k \geq 0$. The name house of cards reflects the following interpretation: suppose we are constructing a house of cards, additioning one by one a new card. The house has probability $(1 - a_k)$ of falling down after the addition of the $k$th card.

**Example 1.46 (Polya urn)** The state-space is $\mathbb{N}^2$. The state $(W_n, B_n)$ at time $n$ indicates the number of white, respectively black balls in an urn. At each time a ball is chosen at random from the urn and it is replaced by two balls of the same color. The partition of state $(k, \ell)$ has only two intervals given by

$$\begin{aligned} I((k,\ell),(k+1,\ell)) &= \Big[0, \frac{k}{k+\ell}\Big[ \\ I((k,\ell),(k,\ell+1)) &= \Big[\frac{k}{k+\ell}, 1\Big] \end{aligned} \tag{1.47}$$

The process that keeps track of the number of balls of each color is a Markov chain. In contrast, the process

$$s_n = \frac{W_n}{B_n + W_n} \tag{1.48}$$

which keeps track of the proportion of white balls in the urn is not Markov. We ask the reader to prove this in the exercises.

## 1.3 Exercises

**Exercise 1.1** Consider a Markov chain in the state-space $\mathcal{X} = \{1, 2, 3\}$ and transition matrix

$$Q = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{pmatrix} \tag{1.49}$$

(a) Construct a family of partitions of $[0, 1]$ such that the function $F$ defined as in (1.23) permits to construct the Markov chain with transition matrix (1.49).

(b) Compute $\mathbb{P}(X_2 = 1 \mid X_1 = 2)$

(c) Compute $\mathbb{P}(X_3 = 2 \mid X_1 = 1)$

(d) Using the function of item (a) and a sequence of random numbers in $[0, 1]$ simulate 10 steps of the above chain with initial state $X_0 = 1$.

**Exercise 1.2** Consider two chains $(X_n^0)_{n\in\mathbb{N}}$ and $(Y_n^0)_{n\in\mathbb{N}}$ in $\mathcal{X} = \{0,1\}$. The chain $X_n$ is defined by the function $F(x;u) = \mathbf{1}\{u > h(x)\}$ (for fixed values $h(0)$ and $h(1)$); the chain $Y_n$ is defined by the function $F(x;u) = 1 - x$, if $u > g(x)$ and $F(x;u) = x$ otherwise (for fixed values $g(0)$ and $g(1)$).

(a) Compute

$$\begin{aligned} Q_h(x,y) &= \mathbb{P}(X_n = y \mid X_{n-1} = x\} \\ Q_g(x,y) &= \mathbb{P}(Y_n = y \mid Y_{n-1} = x\} \end{aligned} \tag{1.50}$$

where $n \geq 1$, and $x$ and $y$ assume values in $\{0,1\}$.

(b) Show that if $h(0) = h(1)$, then $Q_h(x,y)$ depends only on $y$ and it is constant in $x$. In this case the corresponding Markov chain is just a sequence of *iid* random variables.

(c) Simulate 10 instants of the chain $X_n$ with $h(0) = 1/3$, $h(1) = 1/2$.

(d) Simulate 10 instants of the chain $Y_n$ with $g(0) = 1/3$, $g(1) = 1/2$.

**Exercise 1.3 (Coupling)** Let $(X_n^0, X_n^1)_{n\in\mathbb{N}}$ be the joint construction of the chains of Lemma 1.14. That is, the two chains are constructed as the chain $(X_n)$ of the previous exercise with the *same* sequence $U_1, U_2, \ldots$ of uniform random variables and initial states 0 and 1 respectively, that is, $X_0^0 = 0$ and $X_0^1 = 1$. Let $\tau$ be the first time the chains meet. That is,

$$\tau := \inf\{n \geq 1 : X_n^0 = X_n^1\} \tag{1.51}$$

Assuming $0 < h(0) < h(1) < 1$, show that

$$\mathbb{P}(\tau > n) = \rho^n \ , \tag{1.52}$$

where $\rho = h(1) - h(0)$.

**Exercise 1.4** Compute the transition matrix of the process $(Z_n)_{n\in\mathbb{N}}$ defined in (1.43) (Ehrenfest model).

**Exercise 1.5** Compute the transition matrix of the Polya's urn process $(X_n) = (B_n, V_n)$ defined with the partitions (1.47). Show that the process $(B_n/(B_n + V_n))$ is not Markov.

## 1.4 Comments and references

Markov chains were introduced by A. A. Markov at the beginning of the XX-th century as a linguistic model. The construction of Proposition 1.19 is inspired in the graphic construction of Harris (1972) for interacting particle systems. The notion of coupling was introduced by Doeblin (1938). The Ehrenfest model was introduced by P. and T. Ehrenfest in a famous paper in defense of the ideas of Boltzmann in 1905. In particular, this model illustrates how it is possible to have irreversibility (at macroscopic level) and reversibility (at microscopic level). The example of Polya urn shows the fact that the image of a Markov chain through some function is not always Markov.

# Chapter 2

# Invariant measures

In this chapter we present the notions of invariant measure and reversibility of Markov chains. We also show Kăc's Lemma which establishes a rapport between the invariant measure and the mean return time to a given arbitrary state.

## 2.1 Transition matrices

We start with some linear algebra. As we saw in the previous chapter, the one-step transition probabilities can be thought of as a matrix. In the next proposition we see that this is a convenient way to represent them. The state space $\mathcal{X}$ is always a countable space.

**Proposition 2.1** *Let $(X_n^a : n \in \mathbb{N})$ be a Markov chain on $\mathcal{X}$ with transition matrix $Q$ and initial state $a$. Then, for all time $n \geq 1$ and every state $b \in \mathcal{X}$ we have*

$$\mathbb{P}(X_n^a = b) = Q^n(a, b). \tag{2.2}$$

*where $Q^n(a, b)$ is the element of row $a$ and column $b$ of the matrix $Q^n$.*

**Proof.** By recurrence. The identity (2.2) is immediate for $n = 1$, by definition of $Q$.

Assume it is true for some $n \geq 1$, then

$$\mathbb{P}(X_{n+1}^a = b) = \sum_{z \in \mathcal{X}} \mathbb{P}(X_n^a = z, X_{n+1}^a = b) \tag{2.3}$$

because the family of sets $(\{X_n^a = z\} : z \in \mathcal{X})$ is a partition of $\Omega$. We write then the joint probability, conditioning to the position of the chain at time $n$, and get:

$$\sum_{z \in \mathcal{X}} \mathbb{P}(X_n^a = z, X_{n+1}^a = b) = \sum_{z \in \mathcal{X}} \mathbb{P}(X_n^a = z)\mathbb{P}(X_{n+1}^a = b \mid X_n^a = z). \tag{2.4}$$

By induction hypothesis

$$\mathbb{P}(X_n^a = z) = Q^n(a, z). \tag{2.5}$$

On the other hand, by definition,

$$\begin{aligned} \mathbb{P}(X_{n+1}^a = b \mid X_n^a = z) &= \mathbb{P}(F(X_n^a; U_{n+1}) = b \mid X_n^a = z) \\ &= \mathbb{P}(F(z; U_{n+1}) = b). \end{aligned} \tag{2.6}$$

Since the random variables $U_1, U_2, \ldots$ are *iid* uniform in $[0, 1]$,

$$\mathbb{P}(F(z; U_{n+1}) = b) = \mathbb{P}(F(z; U_1) = b). \tag{2.7}$$

By definition

$$\mathbb{P}(F(z; U_1) = b) = \mathbb{P}(X_1^z = b) = Q(z, b). \tag{2.8}$$

Using (2.5) and (2.8) we can rewrite (2.3) as

$$\mathbb{P}(X_{n+1}^a = b) = \sum_{z \in \mathcal{X}} Q^n(a, z)Q(z, b) \tag{2.9}$$

but this is exactly the definition of $Q^{n+1}(a, b)$:

$$Q^{n+1}(a, b) = \sum_{z \in \mathcal{X}} Q^n(a, z)Q(z, b), \tag{2.10}$$

and this finishes the proof. □

**Example 2.11** Let $(X_n : n \in \mathbb{N})$ be a Markov chain on $\mathcal{X} = \{1, 2\}$ and transition matrix

$$Q = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix} \tag{2.12}$$

To compute $\mathbb{P}(X_2^1 = 1)$ we need to sum the probabilities along the paths $(X_1^1 = 1, X_2^1 = 1)$ and $(X_1^1 = 2, X_2^1 = 1)$. That is,

$$\mathbb{P}(X_2^1 = 1) = Q(1,1)Q(1,1) + Q(1,2)Q(2,1) = p^2 + (1-p)(1-q), \tag{2.13}$$

which is exactly the element in the first row and first column of the matrix

$$Q^2 = \begin{pmatrix} p^2 + (1-p)(1-q) & p(1-p) + (1-p)q \\ (1-q)p + q(1-q) & q^2 + (1-q)(1-p) \end{pmatrix} \tag{2.14}$$

For larger values of $n$ the direct computation is messy. To simplify the notation we write $P(n) = \mathbb{P}(X_n^1 = 1)$. Then

$$\begin{aligned} P(n) &= P(n-1)Q(1,1) + (1 - P(n-1))Q(2,1) \\ &= P(n-1)p + (1 - P(n-1))(1-q). \end{aligned} \tag{2.15}$$

This system of finite-differences equations has a simple solution. Indeed, (2.15) can be written as

$$P(n) = P(n-1)(p + q - 1) + (1 - q). \tag{2.16}$$

Substituting $P(n-1)$ for the equivalent expression given by (2.16) for $n-1$ we get

$$\begin{aligned} P(n) &= [P(n-2)(p+q-1) + (1-q)](p+q-1) + (1-q) \\ &= P(n-2)(p+q-1)^2 + (1-q)(p+q-1) + (1-q). \end{aligned} \tag{2.17}$$

Iterating this procedure we obtain

$$P(n) = P(0)(p+q-1)^n + (1-q)\sum_{k=0}^{n-1}(p+q-1)^k. \tag{2.18}$$

By definition, $P(0) = 1$. The sum of a geometric series is

$$\sum_{k=0}^{n-1} \theta^k = \frac{1-\theta^n}{1-\theta}. \tag{2.19}$$

Hence we can write (2.17) as

$$\begin{aligned} P(n) &= (p+q-1)^n + \frac{(1-q)[1-(p+q-1)^n]}{(1-p)+(1-q)} \\ &= \frac{1-q}{(1-p)+(1-q)} + (p+q-1)^n \frac{1-p}{(1-p)+(1-q)}. \end{aligned} \tag{2.20}$$

This finishes the computation. Observe that $P(n)$ converges exponentially fast to

$$\frac{1-q}{(1-p)+(1-q)}. \tag{2.21}$$

Analogous computations show that $Q^n$ converges exponentially fast in each entry to the matrix

$$\begin{pmatrix} \dfrac{1-p}{(1-p)+(1-q)} & \dfrac{1-q}{(1-p)+(1-q)} \\ \dfrac{1-p}{(1-p)+(1-q)} & \dfrac{1-q}{(1-p)+(1-q)} \end{pmatrix} \tag{2.22}$$

The situation is in fact harder than it looks. The computation above is one of the few examples for which the transition probabilities at time $n$ can be explicitly computed. On the other hand, even in the most complicated cases, the matrix $Q^n$ converges, as $n \to \infty$, to a object that in many cases can be explicitly characterized. We postpone for the next chapter the proof of convergence. Next we define the limiting object.

## 2.2 Invariant measures

Let $\mathcal{X}$ be a finite or countable set and $Q$ a transition matrix on $\mathcal{X}$. A probability measure on $\mathcal{X}$ is called *invariant* with respect to $Q$ if for each element $x \in \mathcal{X}$,

$$\mu(x) = \sum_{a \in \mathcal{X}} \mu(a) Q(a, x) \tag{2.23}$$

This is just an algebraic definition. In other words, it says that an invariant probability is a left eigenvector of $Q$ with eigenvalue 1. However it has a deep statistical meaning, as we will see.

The very first question about this definition is: *are there invariant measures for a given transition matrix $Q$?* As we see in the next result, the answer is positive if the state space $\mathcal{X}$ is finite. This is a particular case of a general theorem for finite positive matrices called *Theorem of Perron-Frobenious.*

**Theorem 2.24 (Perron-Frobenious)** *If $Q$ is a transition matrix on a finite state space $\mathcal{X}$ then there is at least one invariant distribution $\mu$ for $Q$.*

**Proof.** Postponed. The traditional proof of this theorem uses the *fixed point theorem* which is highly non constructive. A probabilistic proof is proposed in Exercise 2.7 below. We provide another proof after Theorem 4.45 of Chapter 4; by explicitly constructing an invariant measure. □

Besides its natural algebraic meaning, the definition of invariant measure has a important probabilistic significance, as we see in the next proposition which says roughly the following. Assume that the initial state of a Markov chain is chosen randomly using an invariant probability with respect to the transition matrix of the chain. In this case, the law of the chain at each instant will be the same as the law at time zero. In other words, the probability to find the chain at time $n$ in position $a$ will be exactly $\mu(a)$, for all $n$ and $a$, where $\mu$ is the invariant probability with respect to $Q$ used to choose the initial position.

We introduce now a notation to describe a Markov chain with random initial state. Let $(X_n^a : n \in \mathbb{N})$ be a Markov chain on $\mathcal{X}$ with initial state $a \in \mathcal{X}$. Let $X_n^\mu$ be the process with law

$$\mathbb{P}(X_n^\mu = b\} = \sum_{a \in \mathcal{X}} \mu(a) \mathbb{P}(X_n^a = b\}. \tag{2.25}$$

**Proposition 2.26** *Let* $(X^\mu_n : n \in \mathbb{N})$ *be a Markov chain on* $\mathcal{X}$ *with transition matrix* $Q$ *and initial state randomly chosen according to the probability* $\mu$. *If* $\mu$ *is invariant with respect to* $Q$, *then*

$$\mathbb{P}(X^\mu_n = b\} = \mu(b), \tag{2.27}$$

*for any* $n$ *and* $b$.

**Proof.** By recurrence. The identity holds for $n = 1$, by hypothesis. Assume it holds for some $n \geq 1$. Then

$$\mathbb{P}(X^\mu_{n+1} = b) = \sum_{z\in\mathcal{X}}\sum_{a\in\mathcal{X}} \mu(a)\mathbb{P}(X^a_n = z)\mathbb{P}(X^a_{n+1} = b \mid X^a_n = z)\,. \tag{2.28}$$

(In the countable case the interchange of sums is justified by the Theorem of Fubini; all summands are positive) Identity (2.8) says

$$\mathbb{P}(X^a_{n+1} = b \mid X^a_n = z) = Q(z, b). \tag{2.29}$$

By induction hypothesis, for all $z$,

$$\sum_{a\in\mathcal{X}} \mu(a)\mathbb{P}(X^a_n = z) = \mu(z)\,. \tag{2.30}$$

Using (2.29) and (2.30) we can rewrite (2.28) as

$$\mathbb{P}(X^\mu_{n+1} = b) = \sum_{z\in\mathcal{X}} \mu(z)Q(z, b)\,. \tag{2.31}$$

Since, by hypothesis, $\mu$ is invariant with respect to $Q$,

$$\sum_{z\in\mathcal{X}} \mu(z)Q(z, b) = \mu(b)\,, \tag{2.32}$$

which finishes the proof. □

## 2.3 Reversibility

Let $Q$ be a transition matrix on $\mathcal{X}$. A probability measure $\mu$ on $\mathcal{X}$ is called *reversible* with respect to $Q$ if for all elements $x$, $y$ of $\mathcal{X}$

$$\mu(x)Q(x,y) = \mu(y)Q(y,x)\,. \tag{2.33}$$

**Proposition 2.34** *If $\mu$ is reversible for $Q$ then it is also invariant for $Q$.*

**Proof.** Just sum over $x$ in (2.33) and use that $\sum_x Q(y,x) = 1$ to obtain (2.23). □

One advantage of chains with reversible measures is that the system of equations (2.33) is simpler than the one stated by (2.23). On the other hand, if one has a probability measure, it is easy to construct a Markov chain having this measure as reversible. This second property is largely exploited in the so called Markov Chain Monte Carlo (MCMC) method to simulate probability distributions.

Consider a Markov chain with transition matrix $Q(x,y)$ and invariant measure $\mu$. Define the *reverse* matrix of $Q$ with respect to $\mu$ by

$$Q^*(x,y) = \frac{\mu(y)}{\mu(x)}Q(y,x). \tag{2.35}$$

It is easy to see that $Q^*$ is also a transition matrix for a Markov chain. This chain is called *reverse chain* (with respect to $\mu$). An alternative definition of reversibility is to say that $\mu$ is reversible for $Q$ if $Q = Q^*$, where $Q^*$ is defined in (2.35).

**Lemma 2.36** *If $\mu$ is invariant for $Q$ and $Q^*$ is the reverse matrix of $Q$ with respect to $\mu$ if and only if for all $x_1, \ldots, x_n \in \mathcal{X}$*

$$\mu(x_0)Q(x_0,x_1)\ldots Q(x_{n-1},x_n) \;=\; \mu(x_n)Q^*(x_n,x_{n-1})\ldots Q^*(x_1,x_0). \tag{2.37}$$

**Proof.** Follows from definition (2.35). □

Lemma 2.36 says that under the invariant measure, if we look at the process backwards in time, then we see a Markov chain with transition matrix $Q^*$.

**Lemma 2.38** *Let $Q$ be a transition matrix on $\mathcal{X}$ with invariant measure $\mu$. Let $Q^*$ be the reverse matrix of $Q$ with respect to $\mu$ defined by* (2.35)*. Then*

$$Q(x_0, x_1) \dots Q(x_{n-1}, x_n) = Q^*(x_n, x_{n-1}) \dots Q^*(x_1, x_0). \tag{2.39}$$

*for any cycle $x_0, \dots, x_n \in \mathcal{X}$ with $x_n = x_0$.*

**Proof.** If $\mu$ is invariant for $Q$, then (2.37) implies immediately (2.39). □

This lemma says that the probability of seeing the chain going along a closed cycle is the same as to see the reverse chain to make the cycle in the opposite sense.

When looking for the invariant measure of a transition matrix $Q$ sometimes (but now always) it is easier to guess the measure and the reverse process with respect to this measure than to directly solve the set of equations satisfied by the invariant measure. This is because each equation in (2.35) involves only two states. The following lemma whose proof is immediate indicates when this is possible.

**Lemma 2.40** *Let $Q$ be a transition matrix on $\mathcal{X}$. Let $\widetilde{Q}$ be a matrix and $\mu$ a probability measure on $\mathcal{X}$ such that for all $x, y \in \mathcal{X}$,*

$$\mu(x)Q(x, y) = \mu(y)\widetilde{Q}(y, x) \tag{2.41}$$

*Then $\mu$ is invariant for $Q$ if and only if $\widetilde{Q}$ is a transition matrix, that is*

$$\sum_y \widetilde{Q}(x, y) = 1 \tag{2.42}$$

*for all $x \in \mathcal{X}$. In this case $\mu$ is also invariant for $\widetilde{Q}$ and $Q^* = \widetilde{Q}$, that is, $\widetilde{Q}$ is the reverse matrix of $Q$ with respect to $\mu$.*

The message of the above lemma is: "do not forget to check that the guessed reverse matrix is a transition matrix".

**Example 2.43 (The asymmetric random walk in the circle)** Consider an arbitrary natural number $n$ and let the state space be $\mathcal{X} = \{1,\dots,n\}$. Let $a_1,\dots,a_n$ be arbitrary numbers in $[0,1]$ and define the transition matrix $Q$ by define

$$Q(x,y) \quad = \quad \begin{cases} a(x) & \text{if } x = 1,\dots,n-1 \text{ and } y = x+1 \\ a(n) & \text{if } x = n \text{ and } y = 1 \\ 1-a(x) & \text{if } x = y \\ 0 & \text{ortherwise} \end{cases} \tag{2.44}$$

## 2.4 Irreducible chains

We study now two questions: When a chain admits an invariant measure? When a chain admits a *unique* invariant measure? The following example shows a chain with more than one invariant measure.

**Example 2.45** Let $\mathcal{X} = \{1,2\}$ and the transition matrix $Q$ be given by

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

The chain corresponding to this matrix is the constant chain: $X_n \equiv X_0$ for all $n \geq 0$. Any measure on $\mathcal{X}$ is invariant for this matrix. The point is that the states 1 and 2 do not communicate, that is $Q^n(1,0) = Q^n(0,1) = 0$ for all $n \geq 0$.

**Example 2.46** Let $\mathcal{X} = \{1,2,3,4\}$ and $Q$ be the transition matrix defined by:

$$Q = \begin{pmatrix} p & 1-p & 0 & 0 \\ 1-p & p & 0 & 0 \\ 0 & 0 & q & 1-q \\ 0 & 0 & 1-q & q \end{pmatrix}$$

with $p, q \in (0,1)$. In this case states 1 and 2 communicate but those states do not communicate with states 3 and 4. These also communicate, In fact we have two separate chains, one with state space $\mathcal{X}_1 = \{1, 2\}$ and transition matrix

$$Q_1 = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}$$

and the other with state space $\mathcal{X}_2 = \{3, 4\}$ and transition matrix

$$Q_2 = \begin{pmatrix} q & 1-q \\ 1-q & q \end{pmatrix}$$

The uniform distribution in $\mathcal{X}_i$ is invariant for the transition matrix $Q_i$, $i = 1, 2$. Any convex combination of those invariant measures, when considered as measures on $\mathcal{X} = \{1, 2, 3, 4\}$, is an invariant measure for $Q$: calling $\mu_1(x) = 1/2$ for $x = 1, 2$, $\mu_1(x) = 0$ otherwise, and $\mu_2(x) = 1/2$ for $x = 3, 4$, $\mu_1(x) = 0$ otherwise, we have that any measure $\mu$ defined by

$$\mu(x) = \alpha\mu_1(x) + (1-\alpha)\mu_2(x) \tag{2.47}$$

with $\alpha \in [0, 1]$ is invariant for $Q$.

The above examples motivate the following definition.

**Definition 2.48** A transition matrix $Q$ defined on a finite or countable set $\mathcal{X}$ is called *irreducible* if for any couple $x, y \in \mathcal{X}$ there exists a number $n = n(x, y)$, such that

$$Q^n(x, y) > 0.$$

## 2.5 Kăc's Lemma

The question of existence of invariant measures is delicate. Perron Frobenious Theorem 2.24 says that if $\mathcal{X}$ is finite, then the system of equations (2.23) has at least one solution which is a probability. However, if $\mathcal{X}$ is infinite, then there are transition matrices with no invariant probabilities. The next proposition gives a probabilistic characterization of this algebraic fact. This is a particular case of a general result proven by Mark Kăc in the fifties.

Given a Markov chain $(X_n : n \in \mathbb{N})$ on $\mathcal{X}$ let the hitting time of a state $a \in \mathcal{X}$ for the chain starting at $b \in \mathcal{X}$ be defined by

$$T^{b\to a} := \inf\{n \geq 1 : X_n^b = a\}. \tag{2.49}$$

**Theorem 2.50 (Kăc's Lemma)** *Let $(X_n^a : n \in \mathbb{N})$ be a irreducible Markov chain on a finite or countable set $\mathcal{X}$ with transition matrix $Q$ and initial state $a$. Assume that $\mu$ is an invariant probability for $Q$ and that $\mu(a) > 0$. Then*

$$\mathbb{E}(T^{a\to a}) = \frac{1}{\mu(a)}. \tag{2.51}$$

Kăc's Lemma says that the mean return time to a given state is inversely proportional to the invariant probability of this state. The proof of the theorem is based in three lemmas.

We need the following notation

$$\mathbb{P}(T^{\mu\to a} = n) := \begin{cases} \sum_{b\in\mathcal{X}\setminus\{a\}} \mu(b)\mathbb{P}(T^{b\to a} = n), & \text{if } n > 0 \\ \mu(a), & \text{if } n = 0 \end{cases} \tag{2.52}$$

**Lemma 2.53** *Let $a \in \mathcal{X}$ satisfying $\mu(a) > 0$. Then*

$$\mathbb{P}(T^{\mu\to a} = n) = \mu(a)\mathbb{P}(T^{a\to a} > n). \tag{2.54}$$

**Proof.** For $n = 0$ the identity holds by definition. For $n > 0$ write

$$\begin{aligned} \mathbb{P}(T^{\mu\to a} = n) &= \sum_{x_0\in\mathcal{X}\setminus\{a\}} \mu(x_0)\mathbb{P}(X_1^{x_0} \neq a, \cdots, X_{n-1}^{x_0} \neq a, X_n^{x_0} = a) \\ &= \sum_{x_0\neq a}\sum_{x_1\neq a}\cdots\sum_{x_{n-1}\neq a} \mu(x_0)Q(x_0,x_1)\ldots Q(x_{n-1},a). \end{aligned} \tag{2.55}$$

Since by hypothesis $\mu$ is invariant for $Q$, we can use Lemma 2.36 to rewrite (2.55) as

$$\sum_{x_0\neq a}\sum_{x_1\neq a}\cdots\sum_{x_{n-1}\neq a} \mu(a)Q^*(a,x_{n-1})\ldots Q^*(x_1,x_0), \tag{2.56}$$

where $Q^*$ is the reverse matrix of $Q$. The expression (2.56) equals

$$\mu(a)\mathbb{P}((T^*)^{a\to a} > n), \tag{2.57}$$

where $(T^*)^{b\to a}$ is the hitting time of $a$ starting from $b$ for the reverse chain. By (2.39), $(T^*)^{a\to a}$ and $T^{a\to a}$ have the same law. This finishes the proof of the lemma. □

**Lemma 2.58** *Let $a \in \mathcal{X}$ satisfying $\mu(a) > 0$. Then $\mathbb{P}(T^{a\to a} > n) \to 0$, as $n \to \infty$. Consequently $\mathbb{P}(T^{a\to a} < \infty) = 1$*

**Proof.** It is clear that

$$1 \ge \sum_{n=0}^{\infty} \mathbb{P}(T^{\mu\to a} = n) = \sum_{n=0}^{\infty} \mu(a)\mathbb{P}(T^{a\to a} > n)\,, \tag{2.59}$$

by Lemma 2.53. Since by hypothesis, $\mu(a) > 0$, (2.59) implies

$$\sum_{n=0}^{\infty} \mathbb{P}(T^{a\to a} > n) < \infty. \tag{2.60}$$

Hence

$$\lim_{n\to\infty} \mathbb{P}(T^{a\to a} > n) = 0, \tag{2.61}$$

which proves the first part of the Lemma. Now notice that

$$\mathbb{P}(T^{a\to a} = \infty) \le \mathbb{P}(T^{a\to a} > n) \to 0 \quad (\text{ as } n \to \infty) \tag{2.62}$$

which finishes the proof. □

**Lemma 2.63** *Assume $\mu(a) > 0$. Then the probability $\mathbb{P}(T^{\mu\to a} > n)$ goes to zero as $n \to \infty$.*

**Proof.** Initially we verify that the measure $\mu$ gives positive mass to any element of $\mathcal{X}$. Indeed, as the matrix $Q$ is irreducible, for any $b \in \mathcal{X}$, there exists a $k$ such that $Q^k(a,b) > 0$. By invariance of $\mu$ we have

$$\mu(b) = \sum_{x\in\mathcal{X}} \mu(x)Q^k(x,b) \ge \mu(a)Q^k(a,b) > 0. \tag{2.64}$$

Since $\mu(b) > 0$, Lemmas 2.53 and 2.58 also apply to $b$. Hence, by Lemma (2.58),

$$\mathbb{P}(T^{b\to b} < \infty) = 1\,. \tag{2.65}$$

Since the process starts afresh each time it visits $b$, a simple application of the Borel-Cantelli lemma shows that the chain starting in $b$ comes back to $b$ infinitely often with probability one. Let $T_n^{b\to b}$ be the time of the $n$-th return to $b$. Consider the trajectories of the process between two consecutive visits to $b$. Each of these trajectories includes or not the state $a$. Since $Q$ is irreducible for each pair $a, b \in \mathcal{X}$, there exist a $k \geq 1$, a $j \geq 1$ and finite sequences of states $(x_0, \dots, x_k)$, $(y_0, \dots, y_j)$ such that $x_0 = b = y_j$ and $x_k = a = y_0$ such that

$$Q(b, x_1)\dots Q(x_{k-1}, a)Q(a, y_1)\dots Q(y_{j-1}, b) \;=\delta\;>\;0, \tag{2.66}$$

for some $\delta > 0$. By the Markov property (1.27)–(1.29), the events $A_k :=$ {the chain visits $a$ in the $k$-th excursion from $b$ to $b$} are independent and $\mathbb{P}(A_k) \geq \delta$. This implies

$$\mathbb{P}(T^{b\to a} > T_k^{b\to b}) = \mathbb{P}(\;A_1^c \cap \dots \cap A_k^c) \leq (1-\delta)^k \tag{2.67}$$

Since $\delta > 0$, inequality (2.67) implies

$$\lim_{n\to\infty} \mathbb{P}(T^{b\to a} > n) = 0. \tag{2.68}$$

To see that, for any $k \geq 0$, write

$$\begin{aligned} \mathbb{P}(T^{b\to a} > n) \;&=\; \mathbb{P}(T^{b\to a} > n, T_k^{b\to b} < n) \;+\; \mathbb{P}(T^{b\to a} > n, T_k^{b\to b} \geq n) \\ &\leq\; \mathbb{P}(T^{b\to a} > T_k^{b\to b}) \;+\; \mathbb{P}(T_k^{b\to b} \geq n) \\ &\leq\; (1-\delta)^k \;+\; \mathbb{P}(T_k^{b\to b} \geq n) \end{aligned} \tag{2.69}$$

By the Markov property $T_k^{b\to b}$ is a sum of $k$ independent random variables each one with the same law as $T^{b\to b}$. By (2.65) $\mathbb{P}(T_k^{b\to b} \geq n)$ goes to zero as $n \to \infty$. Hence

$$\lim_{n\to\infty} \mathbb{P}(T^{b\to a} > n) \;\leq\; (1-\delta)^k \tag{2.70}$$

for all $k \geq 0$. This shows (2.68).

To conclude the proof write

$$\lim_{n\to\infty} \mathbb{P}(T^{\mu\to a} > n) = \lim_{n\to\infty} \sum_{b\in\mathcal{X}} \mu(b)\mathbb{P}(T^{b\to a} > n). \tag{2.71}$$

Taking the limit inside the sum we get

$$\lim_{n\to\infty} \mathbb{P}(T^{\mu\to a} > n) = \sum_{b\in\mathcal{X}} \mu(b) \lim_{n\to\infty} \mathbb{P}(T^{b\to a} > n) = 0, \tag{2.72}$$

We need to justify the interchange of sum and limit in (2.71). For finite $\mathcal{X}$ this is authorized by continuity of the sum. If the space is countable, we need to call the *Monotone Convergence Theorem*, a classic theorem in Integration Theory. □

**Proof of Theorem 2.50.** Lemma 2.63 guarantees

$$\mathbb{P}(T^{\mu\to a} < \infty) = 1. \tag{2.73}$$

Hence, by Lemma 2.53,

$$1 = \sum_{n=0}^{\infty} \mathbb{P}(T^{\mu\to a} = n) \ = \mu(a) \sum_{n=0}^{\infty} \mathbb{P}(T^{a\to a} > n).$$

This concludes the proof because

$$\sum_{n=0}^{\infty} \mathbb{P}(T^{a\to a} > n) \ = \mathbb{E}(T^{a\to a})\,. \ \square$$

## 2.6 Exercises

**Exercise 2.1** Show that if $\mu$ is reversible for $Q$, then $\mu$ is invariant for $Q$.

**Exercise 2.2** Show that if $\mu$ is invariant for $Q$, then the matrix $Q^*$ defined by

$$Q^*(x,y) = \frac{\mu(x)}{\mu(y)} Q(y,x). \tag{2.74}$$

is the transition matrix of a Markov chain.

**Exercise 2.3** Show that the identity (2.39) implies that $T^{a\to a}$ and $(T^*)^{a\to a}$ have the same law.

**Exercise 2.4** Show that the events $A_n$ defined after (2.66) are independent. Show that (2.67) implies (2.68).

**Exercise 2.5** Compute the invariant probability measures for the Markov chains presented in Examples 1.4, 1.6 and in Exercise 2.2.

**Exercise 2.6** Give an example of a Markov chain with an invariant measure which is not reversible.

**Exercise 2.7** Let $Q$ be an irreducible matrix on a finite state space $\mathcal{X}$. Let

$$\mu(y) := \frac{1}{\mathbb{E}T^{x\to x}} \sum_{n\geq 0} \mathbb{P}(X_n^x = y, T^{x\to x} > n) \tag{2.75}$$

In other words, for the chain starting at $x$, the probability $\mu(y)$ is proportional to the expected number of visits to $y$ before the chain return to $x$. Show that $\mu$ is invariant for $Q$. This is an alternative proof of Theorem 2.24.

**Exercise 2.8 (Random walk in $\mathbb{Z}$)** Let $\mathcal{X} = \mathbb{Z}$ and $U_1, U_2, \dots$ be a sequence of *iid* random variables on $\{-1,+1\}$ with law

$$\mathbb{P}(U_n = +1) = p = 1 - \mathbb{P}(U_n = -1),$$

where $p \in [0,1]$. Let $a$ be an arbitrary fixed point of $\mathbb{Z}$. We define the random walk $(S_n^a)_{n\in\mathbb{N}}$, with initial state $a$ as follows:

$$S_0^a = a$$

and

$$S_n^a = S_{n-1}^a + U_{n-1}, \text{ se } n \geq 1.$$

i) Show

$$\mathbb{P}(S_n^0 = x) = \begin{cases} \binom{n}{\frac{x+n}{2}} p^k (1-p)^{n-k}, & \text{if } x+n \text{ is even and } |x| \leq n \ ; \\ 0 & \text{otherwise.} \end{cases}$$

ii) Assume $p = \frac{1}{2}$. Compute $\lim_{n\to\infty} \mathbb{P}(S_n^0 = 0)$. Hint: use Stirling's formula.
iii) Assume $p > \frac{1}{2}$. Use the law of large numbers and the Borel-Cantelli Lemma to show that

$$\mathbb{P}(S_n^0 = 0, \text{ for infinitely many } n\ ) = 0.$$

iv) Is there an invariant measure for the random walk? Establish a connection among items (iii) and (i)–(ii).

**Exercise 2.9 (Birth and death process)** The birth and death process is a Markov chain on $\mathbb{N}$ with transition matrix:

$$Q(x, x+1) = q(x) = 1 - Q(x, x-1), \text{ se } x \geq 1,$$

$$Q(0, 1) = q(0) = 1 - Q(0, 0),$$

where $(q(x))_{x\in\mathbb{N}}$ is a sequence of real numbers contained in $(0, 1)$. Under which condition on $(q(x))_{x\in\mathbb{N}}$ the process accepts an invariant measure? Specialize to the case $q(x) = p$ for all $x \in \mathbb{N}$.

**Exercise 2.10 (Monte Carlo method)** One of the most popular Monte Carlo methods to obtain samples of a given probability measure consists in simulate a Markov chain having the target measure as invariant measure. To obtain a sample of the target measure from the trajectory of the Markov chain, one needs to let the process evolve until it "attains equilibrium". In the next chapter we discuss the question of the necessary time for this to occur. Here we propose a chain for the simulation.

i) Let $\mathcal{X}$ be a finite set and $\mu$ a probability measure on $\mathcal{X}$. Consider the following transition probabilities on $\mathcal{X}$: for $y \neq x$,

$$Q_1(x,y) = \frac{1}{N-1}\frac{\mu(y)}{\mu(x)+\mu(y)} \tag{2.76}$$

$$Q_1(x,x) = 1 - \sum_{y\neq x} Q_1(x,y) \tag{2.77}$$

(Choose uniformly a state $y$ different of $x$ and with probability $\mu(y)/(\mu(x)+\mu(y))$ jump to $y$; with the complementary probability stay in $x$.)

$$Q_2(x,y) = \frac{1}{N-1}\left[\mathbf{1}\{\mu(x)\leq\mu(y)\} + \frac{\mu(y)}{\mu(x)}\mathbf{1}\{\mu(x)>\mu(y)\}\right] \tag{2.78}$$

$$Q_2(x,x) = 1 - \sum_{y\neq x} Q_2(x,y). \tag{2.79}$$

(Choose uniformly a state $y$ different of $x$ and if $\mu(x) < \mu(y)$, then jump to $y$; if $\mu(x) \geq \mu(y)$, then jump to $y$ with probability $\mu(y)/\mu(x)$; with the complementary probability stay in $x$.)

Check if $\mu$ is reversible with respect to $Q_1$ and/or $Q_2$.

ii) Let $\mathcal{X} = \{0,1\}^N$ and $\mu$ be the uniform distribution on $\mathcal{X}$ (that is, $\mu(\zeta) = 2^{-N}$, for all $\zeta \in \mathcal{X}$. Define a transition matrix $Q$ on $E$ satisfying the following conditions:

a) $Q(\xi,\zeta) = 0$, if $d(\xi,\zeta) > 2$, where $d$ is the Hamming's' distance, introduced in the previous chapter;

b) $\mu$ is reversible with respect to $Q$.

iii) Simulate the Markov chain corresponding to the chosen matrix $Q$. Use the algorithm used in Definition 1.2. How would you determine empirically the moment when the process attains equilibrium? Hint: plot the relative frequency of visit to each site against time and wait this to stabilize. Give an empiric estimate of the density of ones. Compare with the true value given by $\mu(1)$, where $\mu$ is the invariant measure for the chain.

iv) Use the Ehrenfest model to simulate the Binomial distribution with parameters $\frac{1}{2}$ and $N$.

**Exercise 2.11** (i) Verify directly that the limit as $n \to \infty$ of $P(n)$, given by (2.20) exists and compute it explicitly. Compute $\lim_{n\to\infty} \mathbb{P}(X_n^2 = 1)$ and identify the limits.

ii) For the same chain compute $\mathbb{E}(T^{1\to 1})$, where

$$T^{1\to 1} = \inf\{n \geq 1 \ : \ X_n^1 = 1\}.$$

iii) Establish a relationship between items (i) and (ii).

**Exercise 2.12** Let $\mathcal{X} = \mathbb{N}$ and $Q$ be a transition matrix defined as follows. For all $x \in \mathbb{N}$

$$Q(0, x) = p(x) \text{ and}$$

$$Q(x, x-1) = 1 \text{ if } x \geq 1,$$

where $p$ is a probability measure on $\mathbb{N}$. Let $(X_n^0)_{n\in\mathbb{N}}$ be a Markov chain with transition matrix $Q$ and initial state 0.

i) Give sufficient conditions on $p$ to guarantee that $Q$ has at least one invariant measure.

ii) Compute $\mathbb{E}(T^{1\to 1})$.

iii) Establish a relationship between items (i) and (ii).

**Exercise 2.13 (Stirring process)** The stirring process is a Markov chain on the hypercube $\mathcal{X} = \{0, 1\}^N$ defined by the following algorithm. Let $\mathcal{P}_N$ be the set of all possible permutations of the sequence $(1, 2, \cdots, N)$, that is, the set of bijections from $\{1, 2, \cdots, N\}$ into itself. Let $\pi$ be an element of $\mathcal{P}_N$. Let $F_\pi \colon \mathcal{X} \to \mathcal{X}$ be the function defined as follows. For all $\xi \in \mathcal{X}$

$$F_\pi(\xi)(i) = \xi(\pi(i)).$$

In other words, $F_\pi$ permutates the values of each configuration $\xi$ assigning to the coordinate $i$ the former value of the coordinate $\pi(i)$.

Let $(\Pi_1, \Pi_2, \cdots)$ be a sequence of *iid* random variables on $\mathcal{P}_N$. The stirring process $(\eta_n^\zeta)_{n\in\mathbb{N}}$ with initial state $\zeta$ is defined as follows:

$$\eta_n^\zeta = \begin{cases} \zeta, & \text{se } n = 0; \\ F_{\Pi_n}(\eta_{n-1}^\zeta), & \text{se } n \geq 1. \end{cases} \tag{2.80}$$

i) Show that the stirring process is not irreducible (it is *reducible*!).

ii) Assume that the random variables $\Pi_n$ have uniform distribution on $\mathcal{P}_N$. Which are all the invariant measures for the stirring process in this case?

iii) Let $\mathcal{V}_N$ be the set of permutations that only change the respective positions of two neighboring points of $(1, 2, \cdots, N)$. A typical element of $\mathcal{V}_N$ is the permutation $\pi^k$, for $k \in \{1, 2, \cdots, N\}$, defined by:

$$\pi^k(i) = \begin{cases} i, & \text{if } i \neq k,\ i \neq k+1, \\ k+1, & \text{if } i = k, \\ k, & \text{if } i = k+1. \end{cases} \tag{2.81}$$

In the above representation, the sum is done "module $N$", that is, $N+1 = 1$. Assume that the random variables $\Pi_n$ are uniformly distributed in the set $\mathcal{V}_N$. Compute the invariant measures of the stirring process in this case.

iv) Compare the results of items (ii) and (iii).

## 2.7 Comments and references

The proof of Theorem 2.24 proposed in Exercise 2.7 can be found in Thorisson (2000) and Häggtröm (2000), Theorem 5.1.

# Chapter 3

# Convergence and loss of memory

In this chapter we discuss the notions of *coupling* between two Markov chains and *meeting time* of coupled chains. The knowledge of asymptotic properties of the distribution of the meeting time permits to obtain bounds for the speed of convergence of the chain to its invariant distribution.

## 3.1 Coupling

In this Section we first give a formal definition of coupling between Markov chains. Then we use it as a tool to find conditions on the probability transition matrix to guarantee convergence to the invariant probability. We show two different approaches.

A coupling between two Markov chains is defined as follows.

**Definition 3.1** Let $Q$ and $Q'$ be probability transition matrices for processes $X_t$ and $X'_t$ on state spaces $\mathcal{X}$ and $\mathcal{X}'$ respectively. A *coupling* between $(X_t : t \geq 0)$ and $(X'_t : t \geq 0)$ is characterized by a function $\widetilde{F} : \mathcal{X} \times \mathcal{X}' \times \mathcal{U} \to \mathcal{X} \times \mathcal{X}'$ satisfying

$$\sum_{y \in \mathcal{X}} \mathbb{P}(\widetilde{F}(x, x'; U) = (y, y')) \quad = \quad Q(x', y') \text{ for all } x \in \mathcal{X},\ x', y' \in \mathcal{X}'$$

$$\sum_{y'\in\mathcal{X}'} \mathbb{P}(\widetilde{F}(x,x';U) = (y,y')) \;=\; Q(x,y) \text{ for all } x,y\in\mathcal{X},\ x'\in\mathcal{X}' \quad (3.2)$$

where $U$ is a random variable or vector on a set $\mathcal{U}$. The coupled process with initial configuration $(a,a')\in\mathcal{X}\times\mathcal{X}'$ is defined as follows. Let $(U_n)$ be a sequence of *iid* random variables with the same law as $U$, then define $(X_0,X_0') = (a,a')$ and for $n\geq 1$,

$$(X_n,X_n') := \widetilde{F}(X_{n-1},X_{n-1}';U_n) \quad (3.3)$$

In other words, a coupling is the simultaneous construction of the two Markov chains in the same probability space. In this case the space is the one induced by the sequence $(U_n)$. Usually we take $U$ as a random variable uniformly distributed in $\mathcal{U}=[0,1]$ but in some cases we need vectors (this is not really a constraint, as one can always construct a random vector using a unique uniform random variable, but to use vectors may facilitate notation). Conditions (3.2) just say that the marginal law of the first (respectively second) coordinate is exactly the law of the process with transition matrix $Q$ (respectively $Q'$).

In these notes we will only couple two (and later on, more) versions of the same process. That is, $\mathcal{X}=\mathcal{X}'$ and $Q=Q'$ in Definition 3.1.

**Example 3.4 (Free coupling)** Let $\mathcal{U}=[0,1]$, $U$ be a uniform random variable in $\mathcal{U}$ and $F$ be a function satisfying (1.20) (corresponding to a Markov chain). The *free coupling* is the one defined by (3.3) with

$$\widetilde{F}(x,x';u) \;:=\; (F(x;u),F(x';u)). \quad (3.5)$$

In words, the marginals use the same $U_n$ to compute the value at each time.

The following is a coupling recently proposed to simulate the invariant measure of a Markov chain. It exploits the fact that if $U$ is a random variable uniformly distributed in $[0,1]$, then so is $1-U$.

**Example 3.6 (Antithetic coupling)** Let $\mathcal{U}$, $U$ and $F$ be as in Example 3.4. The *antithetic coupling* is the one defined by (3.3) with

$$\widetilde{F}(x,x';u) \;:=\; (F(x;u),F(x';1-u)). \quad (3.7)$$

In this coupling the marginals use the same uniform random variable but each marginal uses it in a different way.

**Example 3.8 (Doeblin coupling)** The *Doeblin coupling* is the first coupling appeared in the literature. It was introduced by Doeblin. Consider $\mathcal{U} = [0,1] \times [0,1]$, $U = (V, V')$, a bi-dimensional vector of independent random variables uniformly distributed in $[0,1]$. The coupling is defined by (3.3) with

$$\widetilde{F}(x, x'; v, v') \ := \ (F(x; v), F(x'; v')) \tag{3.9}$$

This coupling consists just on two independent chains.

**Example 3.10 (Independent coalescing coupling)** Consider $\mathcal{U} = [0,1] \times [0,1]$, $U = (V, V')$, a bi-dimensional vector of independent random variables uniformly distributed in $[0,1]$. The coupling is defined by (3.3) with

$$\widetilde{F}(x, x'; v, v') \ := \ \begin{cases} (F(x; v), F(x'; v')) & \text{if } x \neq x' \\ (F(x; v), F(x'; v)) & \text{if } x = x' \end{cases} \tag{3.11}$$

In this case if the marginals are different, then they evolve independently. If the marginals coincide, then they evolve together. The first time the marginals coincide is an important object:

**Definition 3.12** The *meeting time* $\tau^{a,b}$ of the coupling $\widetilde{F}$ is defined by

$$\tau^{a,b} = \begin{cases} +\infty, & \text{if } X_n^a \neq X_n^b, \text{ for all } n \geq 0; \\ \min\{n \geq 1 : X_n^a = X_n^b\}, & \text{otherwise.} \end{cases} \tag{3.13}$$

where $(X_n^a, X_n^b)$ is the coupling (3.2) constructed with the function $\widetilde{F}$ and initial states $(X_0^a, X_0^b) = (a, b)$.

The free and the independent-coalescing couplings have an important property: the marginal processes coalesce after the meeting time. This is the result of next lemma.

**Lemma 3.14** *Let $\widetilde{F}$ be the function corresponding either to the free coupling or to the independent-coalescing coupling and $(X_n^a, X_n^b)$ be the corresponding process with initial states $a$ and $b$ respectively. Let $\tau^{a,b}$ be the meeting time of the coupling $\widetilde{F}$. Then*

$$n \geq \tau^{a,b} \quad \textit{implies} \quad X_n^a = X_n^b \tag{3.15}$$

*for all $a, b \in \mathcal{X}$.*

**Proof.** Left to the reader. □

## 3.2 Loss of memory

In this section we propose a free coupling between two trajectories of the same chain, each trajectory having different starting point. We choose a function $F$ which helps the trajectories to meet as soon as possible. By Lemma 3.14 we know that after meeting the trajectories coalesce into the same trajectory. When the meeting time is "small", we say that *loss of memory* occurs. This terminology makes sense because after the meeting time one cannot distinguish the initial states. A first simple version of this is given by Theorem 3.19 below. Theorem 3.38 shows then how to use a loss-of-memory result to deduce the convergence in law of the chain and the uniqueness of the invariant measure.

Later on, in Theorem 3.54 we propose a more refined coupling to obtain an improved version of loss of memory. We introduce the *Dobrushin ergodicity coefficient* of the chain. Then we present the Convergence result in Theorem 3.63. Before it we introduce the notion of aperiodic chain, needed to state the hypotheses of the Theorem of Convergence in greater generality.

The first version of the result controls the speed of loss of memory with the *ergodicity coefficient* $\beta$ defined by

**Definition 3.16 (Coefficient of ergodicity)** The *ergodicity coefficient* of a transition matrix $Q$ on the state space $\mathcal{X}$ is defined by

$$\beta(Q) := \sum_{x \in \mathcal{X}} \min_{a \in \mathcal{X}} Q(a, x). \tag{3.17}$$

**Theorem 3.18** *Let $Q$ be a transition matrix on a countable state space $\mathcal{X}$. Then there exists a function $F : \mathcal{X} \times [0,1] \to \mathcal{X}$ such that for the free coupling defined by* (3.3) *with the function $\widetilde{F}$ given by* (3.5)*,*

$$\mathbb{P}\Big( \sup_{a,b \in \mathcal{X}} \tau^{a,b} > n \Big) \;\le\; (1 - \beta(Q))^n. \tag{3.19}$$

*where $\tau^{a,b}$ is the meeting time of the coupling $\widetilde{F}$.*

**Proof.** We recall the proof of Proposition 1.19. The construction of the function $F$ only requires a family of partitions of $[0,1]$ $((I(x,y) : y \in \mathcal{X}) : x \in \mathcal{X})$, such that $|I(x,y)| = Q(x,y)$, where $|A|$ is the length (Lebesgue measure) of the set $A \subset \mathbb{R}$. In these notes $A$ will always be a union of intervals and its length will be the sum of the lengths of those intervals.

The key of the proof is a smart definition of the partitions. The (union of) intervals $I(x,y)$ must be chosen in such a way that for all $x, x', y$ the sets $I(x,y) \bigcap I(x',y)$ be as big as possible.

Since $\mathcal{X}$ is countable, we can assume $\mathcal{X} = \{1, 2, \ldots\}$.

For each $y \in \mathcal{X}$ let

$$J(y) \;:=\; [l(y-1), l(y)) \tag{3.20}$$

where

$$l(y) \;:=\; \begin{cases} 0, & \text{if } y = 0; \\ l(y-1) + \min_{a \in \mathcal{X}} Q(a,y), & \text{if } y \ge 1. \end{cases} \tag{3.21}$$

Let $l(\infty) := \lim_{y \to \infty} l(y)$. Displays (3.20) and (3.21) define a partition of the interval $[0, l(\infty)]$.

We now define a family of partitions of the complementary interval $(l(\infty), 1]$, indexed by the elements of $\mathcal{X}$. For each $x \in \mathcal{X}$ let

$$\widetilde{J}(x,y) \;:=\; [\tilde{l}(x, y-1), \tilde{l}(x,y)) \tag{3.22}$$

where

$$\tilde{l}(x,y) \;:=\; \begin{cases} l(N), & \text{if } y = 0; \\ \tilde{l}(x,y) + Q(x,y) - \min_{z \in \mathcal{X}} Q(z,y) & \text{if } y \ge 1. \end{cases} \tag{3.23}$$

Define

$$I(x,y) \;:=\; J(y) \cup \widetilde{J}(x,y). \tag{3.24}$$

It is easy to see that

$$|I(x,y)| \;=\; Q(x,y). \tag{3.25}$$

Finally we define the function $F : \mathcal{X} \times [0,1] \to \mathcal{X}$ as in 3.1:

$$F(x;u) \;:=\; \sum_{y\in\mathcal{X}} y\,\mathbf{1}\{u \in I(x,y)\}. \tag{3.26}$$

That is, $F(x;u) = y$, if and only if $u \in I(x,y)$. Notice that for all $y \in \mathcal{X}$,

$$\bigcap_{x\in\mathcal{X}} I(x,y) = J(y). \tag{3.27}$$

Hence,

$$\text{if } u < l(N), \text{ then } \; F(x;u) = F(z;u), \;\; \text{for all } \; x,z \in \mathcal{X}. \tag{3.28}$$

This is the key property of $F$.

Let $\widetilde{F}$ be the free coupling function defined by (3.2) with the function $F$ defined in (3.26). Let $((X_n^a, X_n^b) \,:\, n \geq 0)$ be the coupled process constructed with this $\widetilde{F}$ according to (3.3). Define

$$\widetilde{\tau} \;:=\; \inf\{n \geq 1 : U_n < l(\infty)\}\;, \tag{3.29}$$

Property (3.28) implies

$$\tau^{a,b} \;\leq\; \widetilde{\tau}, \tag{3.30}$$

for all $a,b \in \mathcal{X}$. Notice that $\widetilde{\tau}$ has geometric distribution:

$$\mathbb{P}(\widetilde{\tau} > n) \;=\; \mathbb{P}(U_1 > l(\infty), \cdots, U_n > l(\infty)) \tag{3.31}$$

$$=\; \prod_{i=1}^{n} \mathbb{P}(U_i > l(\infty)) = (1 - l(\infty))^n. \tag{3.32}$$

To conclude observe that

$$l(\infty) \;=\; \beta(Q). \;\; \square$$

We now deduce convergence theorems from the loss-of-memory result of Theorem 3.18. The idea is the following. Choose randomly the starting point of one of the chains of the coupling accordingly to the invariant measure. The marginal distribution of this chain will be (always) in equilibrium. Then use this information and the coupling to find an upper bound for the distance between the invariant measure and the law of the process at time $t$.

**Lemma 3.33** *Let $Q$ be a transition matrix on a countable $\mathcal{X}$. For each pair $(a,b)\in\mathcal{X}\times\mathcal{X}$, let $((X_n^a,X_n^b)\,:\,n\geq 1)$ be a coupling of the chains with initial states $a$ and $b$, respectively. Let $\tau^{a,b}$ be the meeting time for the coupling. Then for all $x,z\in\mathcal{X}$,*

$$|\mathbb{P}(X_n^x=y)-\mathbb{P}(X_n^z=y)|\;\leq\;\sup_{a,b}\mathbb{P}(\tau^{a,b}>n) \tag{3.34}$$

**Proof.** Rewriting the difference of probabilities in (3.34) as the expectation of the difference of indicator functions we get:

$$\begin{aligned}|\mathbb{P}(X_n^a=y)-\mathbb{P}(X_n^b=y)| &= \left|\mathbb{E}[\mathbf{1}\{X_n^a=y\}-\mathbf{1}\{X_n^b=y\}]\right| \\ &\leq \mathbb{E}\left|\mathbf{1}\{X_n^a=y\}-\mathbf{1}\{X_n^b=y\}\right|. \end{aligned} \tag{3.35}$$

The identity in the above display is the crucial point. It is true because the chains are constructed in the same probability space. Now,

$$\left|\mathbf{1}\{X_n^a=y\}-\mathbf{1}\{X_n^b=y\}\right|\;\leq\;\mathbf{1}\{X_n^a\neq X_n^b\}. \tag{3.36}$$

But Lemma 3.14 implies

$$\mathbf{1}\{X_n^a\neq X_n^b\}\;=\;\mathbf{1}\{\tau^{a,b}>n\}. \tag{3.37}$$

which finishes the proof. □

We now use the bounds on the tails of the distribution of the meeting time to obtain bounds in the speed of convergence to the invariant measure.

**Theorem 3.38** *Let $Q$ be the transition matrix of a Markov chain with countable state space $\mathcal{X}$. Assume $\beta(Q)>0$. Then the chain has a unique invariant measure $\mu$ and*

$$\sup_{a,y}|\mathbb{P}(X_n^a=y)-\mu(y)|\;\leq\;(1-\beta(Q))^n. \tag{3.39}$$

**Proof.** We postpone the proof of existence until Theorem 4.45 of Chapter 4. Since $\mu$ is invariant, we can write the modulus in the left hand side of (3.39) as

$$\Big|\mathbb{P}(X_n^a = y) - \sum_{b\in\mathcal{X}} \mu(b)\mathbb{P}(X_n^b = y)\Big|. \tag{3.40}$$

Since $\sum_b \mu(b) = 1$, this is bounded from above by

$$\sum_{b\in\mathcal{X}} \mu(b) \left|\mathbb{P}(X_n^a = y) - \mathbb{P}(X_n^b = y)\right|. \tag{3.41}$$

By Lemma 3.33 this is bounded by

$$\sum_{b\in\mathcal{X}} \mu(b)\mathbb{P}(\tau^{a,b} > n) \;\le\; \sup_{a,b} \mathbb{P}(\tau^{a,b} > n) \;\le\; (1-\beta(Q))^n. \tag{3.42}$$

by Lemma 3.18. This finishes the proof of (3.39).

Let $\mu$ and $\nu$ be two invariant measures for $Q$. As we did in Lemma 3.33, we construct two chains $X_n^\mu$ and $X_n^\nu$ with initial states randomly chosen according with $\mu$ and $\nu$, respectively. Then,

$$\begin{aligned}
|\nu(y) - \mu(y)| &= \Big|\sum_x \nu(x)\mathbb{P}(X_n^x = y) - \sum_z \mu(z)\mathbb{P}(X_n^z = y)\Big| \\
&= \Big|\sum_x\sum_z \nu(x)\mu(z)\left(\mathbb{P}(X_n^x = y) - \mathbb{P}(X_n^z = y)\right)\Big| \\
&\le \sum_x\sum_z \nu(x)\mu(z) \sup_{a,b}\mathbb{P}(\tau^{a,b} > n) \\
&\le (1-\beta(Q))^n\,, 
\end{aligned} \tag{3.43}$$

using Lemma 3.33 in the first inequality and (3.19) in the second. Since the bound (3.43) holds for all $n$ and $\beta(Q) > 0$ by hypothesis, $\mu = \nu$. This shows uniqueness. □

The assumption $\beta(Q) > 0$ is of course very restrictive. The following corollary allows us to get a result as Theorem 3.38 in a more general case.

**Corollary 3.44** *Let $Q$ be an aperiodic transition matrix on a countable state space $\mathcal{X}$ satisfying $\beta(Q^k) > 0$ for some $k \geq 1$. Then the chain has a unique invariant measure $\mu$ and*

$$\sup_{a,y} |\mathbb{P}(X_n^a = y) - \mu(y)| \leq (1 - \beta(Q^k))^{n/k}. \tag{3.45}$$

**Proof.** Left to the reader. □

A natural question arises: which are the transition matrices on a countable state space $\mathcal{X}$ having $\beta(Q^j)$ strictly positive for some $j \geq 1$. One example for the infinite countable case is the house-of-cards process of Example 1.44. If for this process there exists an $\varepsilon > 0$ such that $a_k < 1 - \varepsilon$ for all $k$, then $\beta(Q) > \varepsilon$. We do not discuss further the infinite countable case.

A related question in the finite case is the following: which are the transition matrices on a finite state space $\mathcal{X}$ having *all* entries positive starting from some power? The examples of irreducible matrices proposed in the previous section show that the *irreducibility* condition is necessary. However, as it is shown in the next section, it is not sufficient.

## 3.3 Periodic and aperiodic chains

Let us start with an example.

**Example 3.46** Let $\mathcal{X} = \{1, 2\}$ and the transition matrix $Q$ be given by

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

It is clear that this matrix corresponds to an irreducible process. However, any power has null entries. Indeed, for all $k \geq 0$, we have

$$Q^{2k} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and

$$Q^{2k+1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

The problem of this matrix is that the transitions from 1 to 2 or from 2 to 1 can only occur in an odd number of steps, while the transitions from 1 to 1, or from 2 to 2, are only possible in an even number of steps. Anyway this matrix accepts a unique invariant measure, the uniform measure in $\mathcal{X}$. This type of situation motivates the the notion of *periodicity* which we introduce in the next definition.

**Definition 3.47** Assume $Q$ to be the transition matrix of a Markov chain on $\mathcal{X}$. An element $x$ of $\mathcal{X}$ is called *periodic* of period $d$, if

$$\text{mcd}\ \{n \geq 1 : Q^n(x,y) > 0\} = d.$$

The element will be called *aperiodic* if $d = 1$.

For example, in the matrix of the Example 3.46 both states 1 and 2 are periodic of period 2.

We omit the proof of next two propositions. They are elementary and of purely algebraic character and can be found in introductory books of Markov chains. The first one says that for irreducible Markov chains the period is a *solidarity property*, that is, all states have the same period.

**Proposition 3.48** *Let $Q$ be a transition matrix on $\mathcal{X}$. If $Q$ is irreducible, then all states of $\mathcal{X}$ have the same period.*

The proposition allows us to call irreducible matrices of *chain of period* $d$ or *aperiodic chain*

**Proposition 3.49** *Let $Q$ be an irreducible transition matrix on a finite set $\mathcal{X}$. If $Q$ is irreducible and aperiodic, then there exists an integer $k$ such that $Q^j$ has all entries positive for all $j \geq k$.*

**Proof.** Omitted. It can be found in Häggström (2000), Theorem 4.1, for instance. □

Irreducible periodic matrices induce a partition of the state space in classes of equivalence. Let $Q$ be a matrix with period $d$ on $\mathcal{X}$. We say that $x$

is *equivalent* to $y$ if there exists a positive integer $k$ such that $Q^{kd}(x,y) > 0$. Then $\mathcal{X} = \mathcal{X}_1, \ldots, \mathcal{X}_d$, where $\mathcal{X}_i$ contains equivalent states and are called *equivalent classes.*

**Proposition 3.50** *Let $Q$ be a irreducible matrix with period $d$. Then $Q^d$ is aperiodic in each one of the classes of equivalence $\mathcal{X}_1, \ldots, \mathcal{X}_d$. Let $\mu_1, \ldots, \mu_d$ be invariant measures for $Q^d$ on $\mathcal{X}_1, \ldots, \mathcal{X}_d$, respectively. Then the measure $\mu$ defined by*

$$\mu(x) := \frac{1}{d} \sum_i \mu_i(x) \tag{3.51}$$

*is an invariant measure for $Q$.*

**Proof.** It is left as an exercise for the reader. □

The last result of this section relates the positivity of all elements of a power of an irreducible matrix $Q$ with the positivity of $\beta(Q)$.

**Lemma 3.52** *Let $Q$ be a transition matrix on a finite set $\mathcal{X}$. If there exists an integer $k$ such that $Q^j$ has all entries positive for all $j \geq k$ then $\beta(Q^j) > 0$.*

**Proof.** It is clear that $Q^j(x,y) > 0$ for all $x, y \in \mathcal{X}$ implies $\beta(Q) > 0$. □

## 3.4 Dobrushin's ergodicity coefficient

We present another coupling to obtain a better speed of loss of memory of the chain.

**Definition 3.53** The *Dobrushin's ergodicity coefficient* of a transition matrix $Q$ on $\mathcal{X}$ is defined by

$$\alpha(Q) = \min_{a,b} \sum_{x \in \mathcal{X}} \min\{Q(a,x), Q(b,x)\}.$$

**Theorem 3.54** *If $Q$ is a transition matrix on a* finite *state space $\mathcal{X}$, then there exists a coupling (joint construction of the chains) $(X_n^a, X_n^b\,:\, n \in \mathbb{N})$ constructed with a function $\widetilde{F}$ such that the meeting time of the coupling $\widetilde{F}$ satisfies*

$$\mathbb{P}(\tau^{a,b} > n) \leq (1 - \alpha(Q))^n. \tag{3.55}$$

To prove this theorem we use the *Dobrushin coupling.*

**Definition 3.56 (Dobrushin coupling)** Let $Q$ be a transition probability matrix on a countable state space $\mathcal{X}$. We construct a family of partitions of $[0,1]$ as in Theorem 3.18. But now we double label each partition, in such a way that the common part of the families $I^a(b,y) \cap I^b(a,y)$ be as large as possible.

We assume again that $\mathcal{X} = \{1, 2, \ldots\}$. For each fixed elements $a$ and $b$ of $\mathcal{X}$ define

$$J^{a,b}(y) \;:=\; [l^{a,b}(y-1), l^{a,b}(y)) \tag{3.57}$$

where

$$l^{a,b}(y) \;:=\; \begin{cases} 0, & \text{if } y = 0; \\ l^{a,b}(y-1) + \min\{Q(a,y), Q(b,y)\} & \text{if } y \geq 1. \end{cases} \tag{3.58}$$

Let $l^{a,b}(\infty) := \lim_{y\to\infty} l^{a,b}(y)$.

Displays (3.57) and (3.58) define a partition of the interval $[0, l^{a,b}(\infty)]$. We need to partition the complementary interval $(l^{a,b}(\infty), 1]$. Since the common parts have already been used, we need to fit the rests in such a way that the total lengths equal the transition probabilities. We give now an example of this construction. Define

$$\widetilde{J}^b(a,y) = [\tilde{l}^b(a, y-1), \tilde{l}^b(a,y))$$

where

$$\tilde{l}^b(a,y) = \begin{cases} l^{a,b}(N), & \text{if } y = 0; \\ \tilde{l}^b(a, y-1) + \max\{0, (Q(a,y) - Q(b,y))\}, & \text{if } y \geq 1. \end{cases}$$

Finally we define

$$I^b(a,y) = J^{a,b}(y) \cup \widetilde{J}^b(a,y).$$

It is easy to see that for all $b$ the following identity holds

$$|I^b(a, y)| = Q(a, y).$$

Define the function $\widetilde{F} : \mathcal{X} \times \mathcal{X} \times [0, 1] \to \mathcal{X} \times \mathcal{X}$ as in Definition 3.1:

$$\widetilde{F}(a, b; u) = \sum_{y=1}^{N} \sum_{z=1}^{N} (y, z) \mathbf{1}\{u \in I^b(a, y) \cap I^a(b, z)\}. \tag{3.59}$$

In other words, $\widetilde{F}(a, b; u) = (y, z)$, if and only if $u \in I^b(a, y) \cap I^a(b, z)$. Notice that

$$I^b(a, y) \cap I^a(b, z) = \begin{cases} J^{a,b}(y), & \text{if } y = z \\ \emptyset, & \text{otherwise} \end{cases}$$

Hence, for any $a$ and $b$,

$$\text{if } \ u < l^{a,b}(N), \ \text{ then } \ \widetilde{F}(a, b; u) = (x, x) \ \text{ for some } \ x \in \mathcal{X}. \tag{3.60}$$

We construct the coupling $(X_n^a, X_n^b)_{n \geq 0}$ as follows:

$$(X_n^a, X_n^b) = \begin{cases} (a, b), & \text{if } n = 0 \ ; \\ \widetilde{F}(X_{n-1}^a, X_{n-1}^b; U_n), & \text{if } n \geq 1, \end{cases} \tag{3.61}$$

where $(U_1, U_2, \cdots)$ is a sequence of *iid* uniformly distributed in $[0, 1]$. The process so defined will be called *Dobrushin coupling*.

**Proof of Theorem 3.54.** With Dobrushin coupling in hands, the rest of the proof follows those of Theorem 3.18 with the only difference that now the coincidence interval changes from step to step, as a function of the current state of the marginal chains. Let

$$\widetilde{\tau}^{a,b} \ = \ \inf\{n \geq 1 : U_n < l^{X_{n-1}^a, X_{n-1}^b}(\infty)\} \ .$$

Property (3.60) implies

$$\tau^{a,b} \leq \widetilde{\tau}^{a,b}.$$

The law of $\widetilde{\tau}^{a,b}$ is stochastically dominated by a geometric random variable:

$$\begin{aligned}
\mathbb{P}(\widetilde{\tau}^{a,b} > n) &= \mathbb{P}(U_1 > l^{X_0^a, X_0^b}(\infty), \cdots, U_n > l^{X_{n-1}^a, X_{n-1}^b}(\infty)) \\
&\leq \mathbb{P}(U_1 > \min_{x,y} l^{x,y}(\infty), \cdots, U_n > \min_{x,y} l^{x,y}(\infty)) \\
&= \prod_{i=1}^{n} \mathbb{P}(U_i > \min_{x,y} l^{x,y}(\infty)) \\
&= (1 - \min_{x,y} l^{x,y}(\infty))^n.
\end{aligned}$$

To conclude observe that

$$\min_{x,y} l^{x,y}(\infty) \; = \; \alpha(Q). \; \square$$

Finally we can state the convergence theorem.

**Theorem 3.62** *If $Q$ is an irreducible aperiodic transition matrix on a countable state space $\mathcal{X}$, then*

$$sup_{(a,b)} |\mathbb{P}(X_n^a = b) - \mu(b)| \leq (1 - \alpha(Q^k))^{\frac{n}{k}}, \tag{3.63}$$

*where $\mu$ is the unique invariant probability for the chain and $k$ is the smallest integer for which all elements of $Q^k$ are strictly positive.*

**Proof.** Follows as in Theorem 3.38, substituting the bound (3.19) by the bound (3.55). $\square$

## 3.5 Recurrence and transience

We finish this chapter with some concepts useful when dealing with countable state spaces.

**Definition 3.64** We say that a state $x$ is

$$\text{transient, if } \mathbb{P}(T^{x\to x} = \infty) > 0; \tag{3.65}$$

$$\text{null recurrent, if } \mathbb{P}(T^{x\to x} = \infty) = 0 \text{ and } \mathbb{E}T^{x\to x} = \infty; \tag{3.66}$$

$$\text{positive recurrent, if } \mathbb{E}T^{x\to x} < \infty. \tag{3.67}$$

If the state space is finite, there are no null recurrent states. In words, a state is transient if the probability that the chain never visit it is positive. A state is recurrent if it is visited infinitely many often with probability one. A state is positive recurrent if the expected return time has finite expectation.

For irreducible chains recurrence and transience are solidarity properties. For this reason we can talk of (irreducible) recurrent or transient chains.

**Example 3.68 (House-of-cards process)** This process was introduced in Example 1.44. Let $\mathcal{X} = \mathbb{N}$ and $Q$ be the transition matrix on $\mathcal{X}$ defined by

$$Q(x,y) = \begin{cases} a_x & \text{if } y = x+1 \\ 1-a_x & \text{if } y = 0 \end{cases} \tag{3.69}$$

**Lemma 3.70** *Let $(W_n^a : n \geq 0)$ be a Markov chain on $\mathbb{N}$ with the transition matrix $Q$ defined in* (3.69) *and initial state $a$. Then*

*(a) The chain is non positive-recurrent if and only if*

$$\sum_{n\geq 0} \prod_{k=1}^{n} a_k = \infty \tag{3.71}$$

*(b) The chain is transient if and only if*

$$\prod_{k=1}^{\infty} a_k \;>\; 0 \tag{3.72}$$

**Proof.** Omitted in this notes. The reader can find it in Bressaud *et al.* (1999). □

Observe that the condition (3.71) is weaker than the condition $\beta(Q) > 0$ (this last is equivalent to $\sup_x a_x < 1$).

## 3.6 Exercises

**Exercise 3.1** Prove Lemma 3.14.

**Exercise 3.2** *Doeblin's coupling.* Let $Q$ be a transition matrix on the finite or countable set $\mathcal{X}$. Define the matrix $\bar{Q}$ on $\mathcal{X} \times \mathcal{X}$ as follows

$$\bar{Q}((a,b),(x,y)) = \begin{cases} Q(a,x)Q(b,y), & \text{if } a \neq b; \\ Q(a,x), & \text{if } a = b \text{ and } x = y; \\ 0, & \text{if } a = b \text{ and } x \neq y. \end{cases} \tag{3.73}$$

Verify that $\bar{Q}$ is a transition matrix. In other words, verify that for all $(a,b) \in \mathcal{X} \times \mathcal{X}$,

$$\sum_{(x.y) \in \mathcal{X} \times \mathcal{X}} \bar{Q}((a,b),(x,y)) = 1.$$

Observe that the chain corresponding to $\bar{Q}$ describes two Markov chains of transition matrix $Q$ which evolve independently up to the first time both visit the same state. From this moment on, the two chains continue together for ever.

**Exercise 3.3** *Doeblin's coupling.* Show that the process defined in Example 3.4 has transition matrix $Q$ defined by (3.73).

**Exercise 3.4** Prove Corollary 3.44.

**Exercise 3.5** Determine if the chains presented in Examples 1.42, 1.45, 1.31, 1.44, 2.46 and in Exercises 1.1, 2.8 and 2.13 are periodic and determine the period. For those matrices that are aperiodic and irreducible, determine the smallest power $k$ satisfying that all the entries of $Q^k$ are strictly positive.

**Exercise 3.6** Determine $\beta(Q)$ and $\alpha(Q)$ for all aperiodic and irreducible chains $Q$ of Exercise 3.5. In case the computations become complicate, try to find bounds for $\alpha(Q)$ and $\beta(Q)$. When $\alpha(Q)$ gives a better convergence velocity than $\beta(Q)$?

**Exercise 3.7** Let $\mathcal{X} = \{1, 2\}$ and $Q$ be the following transition matrix

$$Q = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix}$$

(a) Show that there exists $\bar{n}$, such that, for all $n \geq \bar{n}$,

$$0,45 \leq Q^n(1,2) \leq 0,55 \quad \text{and}$$

$$0,45 \leq Q^n(2,2) \leq 0,55.$$

Find bounds for $\bar{n}$.

(b) Obtain similar results for $Q^n(1,1)$ and $Q^n(2,1)$.

## 3.7 Comments and references

The coupling technique was introduced by Doeblin (1938a), who constructed two chains which evolved independently up to the first moment they meet, as in exercise 3.2. The coefficient $\beta(Q)$ was introduced by Dobrushin (1956) and then used by Dobrushin (1968a,b) to show the existence of a unique Gibbs state.

# Chapter 4

# Regeneration and perfect simulation

In this chapter we propose a somewhat different way of constructing the process. The construction permits to prove the existence of *regeneration times*, that is, random times such that the process starts afresh. It also induces a *perfect simulation* algorithm of stationary Markov chains. That is, an algorithm to obtain an exact sample of the invariant measure for the transition matrix $Q$.

## 4.1 Stopping time

We start with the important definition of *stopping time*.

**Definition 4.1 (Stopping time)** Let $(U_n)$ be a sequence of random variables on some set $\mathcal{U}$. We say that $T$ is a *stopping time* for $(U_n : n \geq 0)$ if the event $\{T \leq j\}$ depends only on the values of $U_1, \dots, U_j$. That is, if there exist events $A_j \subset \mathcal{U}^j$ such that

$$\{T \leq j\} = \{(U_1, \dots, U_j) \in A_j\} \tag{4.2}$$

**Example 4.3** Let $c \in (0,1)$, $\mathcal{U} = [0,1]$, $(U_n)$ be a sequence of random variables uniformly distributed in $\mathcal{U}$ and $T :=$ first time a $U_n$ is less than $c$:

$$T := \min\{n \geq 1 \,:\, U_n < c\} \tag{4.4}$$

Then $T$ is a stopping time, the sets $A_j$ are defined by

$$A_j \;=\; \{U_1 > c, \ldots, U_{j-1} > c, U_j < c\} \tag{4.5}$$

and the law of $T$ is geometric with parameter $c$:

$$\mathbb{P}(T > n) \;=\; (1-c)^n \tag{4.6}$$

## 4.2 Regeneration

In this section we show that if there is a constant lower bound for the probability of passing from any state to any other, then for any measure $\nu$ on $\mathcal{X}$, there exists a random time such that the chain has law $\nu$ at that time.

**Theorem 4.7** *Let $Q$ be the transition matrix of a Markov chain on a finite state space $\mathcal{X}$ such that there exist $c > 0$ with $Q(x,y) \geq c$ for all $x, y \in \mathcal{X}$. Let $\mu$ be an arbitrary probability measure on $\mathcal{X}$. Let $U_1, U_2, \ldots$ be a sequence of uniform random variables in $[0,1]$. Then there exists a function $F : \mathcal{X} \times [0,1] \to \mathcal{X}$ and a random stopping time $T$ for $(U_n)$ such that*

1. *The chain defined by $X_n = F(X_{n-1}; U_n)$ has transition matrix $Q$.*
2. *$X_T$ has law $\mu$*
3. *$T$ has geometric distribution with parameter $c$: $\mathbb{P}(T > n) = (1-c)^n$*
4. *Given the event $\{T = t; X_i = x_i, \text{ for } i < t\}$, the chain $(X_{t+s} : s \geq 0)$ has the same law as $(X_n \,:\, n \geq 0)$ with initial distribution $\mu$.*

**Remark 4.8** The last item in the theorem is the *regeneration* statement: at time $T$ the chain starts afresh with initial distribution $\mu$ independently of the past.

**Proof.** The point is to construct a family of partitions of $[0,1]$ with suitable properties.

We first construct a partition of the interval $[0,c]$: let $l(0)=0$, $l(y)=l(y-1)+c\mu(y)$ and

$$J(y) := [l(y-1), l(y)) \tag{4.9}$$

The length of $J(y)$ is

$$|J(y)| = c\mu(y) \tag{4.10}$$

for all $y \in \mathcal{X}$.

Then construct a partition of $[c,1]$ as follows. Define $k(x,0) := c$,

$$k(x,y) := k(x,y-1) + Q(x,y) - c\mu(y),$$

$$K(x,y) \ := \ [k(x,y-1), k(x,y)) \tag{4.11}$$

and

$$I(x,y) \ := \ J(y) \cup K(x,y) \tag{4.12}$$

Then we have

$$|I(x,y)| = Q(x,y)\,. \tag{4.13}$$

We define

$$F(x;u) \ := \ \sum_{y\in\mathcal{X}} y\,\mathbf{1}\{u \in I(x,y)\} \tag{4.14}$$

and

$$X_n \ := \ F(X_{n-1};U_n) \tag{4.15}$$

With this function, the chain $X_n$ has transition matrix $Q$. This implies that $X_n$ satisfies the first item of the theorem.

Let

$$T := \min\{n \geq 1 \ : \ U_n \leq c\} \tag{4.16}$$

We saw in Example 4.3 that $T$ is a stopping time for $(U_n)$ and that $T$ has geometric distribution with parameter $c$. That is, $T$ satisfies the third item of the theorem.

Compute

$$\begin{aligned}\mathbb{P}(X_T = y) &= \sum_n \mathbb{P}(X_n = y, T = n) \\ &= \sum_n \mathbb{P}(X_n = y, U_1 > c, \ldots, U_{n-1} > c, U_n \le c). \quad (4.17)\end{aligned}$$

By Definition 4.15 of $X_n$,

$$\{X_n = y\} = \{U_n \in J(y) \cup K(X_{n-1}, y)\}.$$

Since $J(y) \subset [0, c)$ for all $y \in \mathcal{X}$ and $K(x, y) \subset [c, 1]$ for all $x, y \in \mathcal{X}$, we have

$$\{X_n = y, U_n < c\} = \{U_n \in J(y)\}. \quad (4.18)$$

Hence, (4.17) equals

$$\begin{aligned}&\sum_{n\ge 1} \mathbb{P}(U_1 > c, \ldots, U_{n-1} > c,\, U_n \in J(y)) \\ &= \sum_{n\ge 1} \mathbb{P}(U_1 > c) \ldots \mathbb{P}(U_{n-1} > c)\, \mathbb{P}(U_n \in J(y)) \\ &= \sum_{n\ge 1} (1-c)^{n-1}\, c\, \mu(y), \quad (4.19)\end{aligned}$$

by (4.10). Since (4.19) equals $\mu(y)$, this shows the second item of the theorem. The last item is left to the reader. □

The following lemma shows that if the measure at the regeneration time $T$ is chosen to be the invariant measure for $Q$, then, after the regeneration time the chain is always distributed according to the invariant measure.

**Lemma 4.20** *Assume the conditions of Theorem 4.7 and that $\mu$ is the invariant measure for $Q$. Then*

$$\mathbb{P}(X^a_t = b \mid t \ge T) = \mu(b) \quad (4.21)$$

**Proof.** Partitioning first on the possible values of $T$ and then on the possible values of $X_n$ we get

$$\begin{aligned}\mathbb{P}(X_t^a = b,\, t \geq T) &= \sum_{n=1}^{t} \mathbb{P}(X_t^a = b,\, T = n) \\ &= \sum_{n=1}^{t}\sum_{x\in\mathcal{X}} \mathbb{P}(X_n^a = x,\, X_t^a = b,\, T = n) \qquad (4.22)\\ &= \sum_{n=1}^{t}\sum_{x\in\mathcal{X}} \mathbb{P}(X_{t-s}^x = b)\, \mathbb{P}(X_n^a = x,\, T = n) \qquad (4.23)\end{aligned}$$

The last step requires some care. We ask the reader to show it in the exercises. By item (2) of Theorem 4.7, the above equals

$$\begin{aligned} &= \sum_{n=1}^{t}\sum_{x\in\mathcal{X}} \mathbb{P}(X_{t-s}^x = b)\, \mu(x)\mathbb{P}(T = n) \\ &= \sum_{x\in\mathcal{X}} \mathbb{P}(X_{t-s}^x = b)\, \mu(x) \sum_{n=1}^{t} \mathbb{P}(T = n) \\ &= \mu(b)\mathbb{P}(t \geq T) \qquad (4.24)\end{aligned}$$

because $\mu$ is invariant for $Q$. □

**Proposition 4.25** *Assume the conditions of Theorem 4.7 and that $\mu$ is the invariant measure for $Q$. Then for any initial state $a$*

$$\sup_x |\mathbb{P}(X_n^a = x) - \mu(x)| \;\leq\; (1-c)^n. \qquad (4.26)$$

**Proof.** Partition according to the values of $T$:

$$\begin{aligned}\mathbb{P}(X_n^a = x) &= \mathbb{P}(X_n = x, t \geq T) \,+\, \mathbb{P}(X_n^a = x, t < T) \\ &= \mathbb{P}(t \geq T)\, \mu(x) \,+\, \mathbb{P}(t < T)\, \mathbb{P}(X_n^a = x \mid t < T) \\ &= (1 - \mathbb{P}(t < T))\, \mu(x) \,+\, \mathbb{P}(t < T)\, \mathbb{P}(X_n^a = x \mid t < T)\end{aligned}$$

by Lemma 4.20. Hence,

$$
\begin{aligned}
|\mathbb{P}(X_n^a = x) - \mu(x)| &= \mathbb{P}(t < T)\,|\mu(x) \,-\, \mathbb{P}(X_n^a = x \,|\, t < T)| \\
&\leq \mathbb{P}(t < T) \qquad (4.27)
\end{aligned}
$$

because the difference of two probabilities is always bounded above by 1. The proposition follows now from item (3) of Theorem 4.7. □

## 4.3 Coupling and regeneration

We compare here regeneration and coupling. Then we extend Theorem 4.7 to the case of countable state space. Let us first define the *regeneration coefficient* of a chain with respect to a measure.

**Definition 4.28** Let the *regeneration coefficient* of a transition matrix $Q$ with respect to a probability measure $\mu$ be

$$C(\mu, Q) := \sup\{c > 0 \,:\, \min_y [\min_x Q(x,y) - c\mu(y)] \,\geq\, 0\} \qquad (4.29)$$

In the next theorem we show that the function $F$ of Theorem 4.7 can be constructed with the regeneration coefficient $C(\mu, Q)$ instead of $c$ and the rate of geometric decay of $T$ will be $C(\mu, Q)$. The regeneration coefficient $C(\mu, Q)$ is the maximal geometric rate we can get when the law of the process with transition matrix $Q$ at regeneration times is $\mu$.

**Theorem 4.30** *Let $Q$ be the transition matrix of a Markov chain on a countable state space $\mathcal{X}$. Assume there exists a measure $\mu$ on $\mathcal{X}$ satisfying $c := C(\mu, Q) > 0$. Then the conclusions of Theorem 4.7 hold with this $\mu$ and this $c$.*

**Proof.** We need to check that the construction of the intervals $J(y)$ in (4.9) and $I(x, y)$ of (4.12) be such that (4.13) hold. For that, it is sufficient that $\min_x Q(x, y) \geq c\mu(y)$ for all $x$. This is guaranteed by the definition of $C(\mu, Q)$. □

By conveniently defining the measure $\mu$ at regeneration times we can get an alternative proof to Theorem 3.38.

**Theorem 4.31** *Let $Q$ be a transition matrix on a countable state space $\mathcal{X}$. Assume $\beta(Q) > 0$, where $\beta(Q)$ is defined in* (3.17). *Define*

$$\mu(y) := \frac{\min_x Q(x,y)}{\sum_x \min_x Q(x,y)} \tag{4.32}$$

*Then*

1. *The regeneration coefficient of $Q$ with respect to this $\mu$ is the same as the ergodicity coefficient $\beta(Q)$:*

$$C(\mu, Q) = \beta(Q)\,; \tag{4.33}$$

2. *The time $T$ constructed in* (4.16) *with $c = C(\mu, Q)$ has geometric law with parameter $\beta(Q)$.*

3. *The following bounds in the loss of memory hold:*

$$|\mathbb{P}(X_n^a = x) - \mathbb{P}(X_n^b = x)| \;\le\; (1 - \beta(Q))^n \tag{4.34}$$

*for any $a, b, x \in \mathcal{X}$.*

**Proof.** Notice first that $\mu$ is well defined because the denominator in (4.32) is exactly $\beta(Q) > 0$ by hypothesis.

Identity (4.33) follows from the definition of $C(\mu, Q)$. $T$ is geometric by construction. Inequality (4.34) holds because the left hand side is bounded above by $\mathbb{P}(T > n)$ which by item (2) is bounded by $(1 - \beta(Q))^n$. □

The measure $\mu$ maximizing the regeneration coefficient of a matrix $Q$ is the one given by (4.32). In particular we get that the coupling time is less than the regeneration time:

$$\tau^{a,b} \;\le\; T \tag{4.35}$$

where $\tau^{a,b}$ is the meeting time of the free coupling with the function $F$ defined in the proof of Theorem 3.18 and $T$ is the regeneration time of the chain with respect to the measure $\mu$ defined by (4.32).

If $\alpha(Q) < \beta(Q)$, then the meeting time of the Dobrushin coupling is less than the regeneration time of Theorem (4.31).

## 4.4 Construction of the invariant measure

In this section we show how to use the regeneration ideas to construct directly the invariant measure. As a corollary we get a perfect simulation algorithm for the invariant measure.

Consider a double infinite sequence of uniform random variables $(U_n : n \in \mathbb{Z})$.

**Definition 4.36 (Regeneration times)** Let $Q$ be an irreducible transition matrix on a countable state space $\mathcal{X}$. Let $\mu$ be a measure on $\mathcal{X}$ satisfying $c := C(\mu, Q) > 0$. Define the sequence of random variables

$$N(i) \;:=\; \mathbf{1}\{U_i \le c\} \tag{4.37}$$

Let the *regeneration times* $(\tau(n))$ with respect to the measure $\mu$ be defined by

$$\tau(n) := \max\{i \le n \,:\, U_i < c\} \tag{4.38}$$

The sequence $N(i)$ is a sequence of *iid* Bernoulli random variables of parameter $c$. Observe that

$$\tau(j) \;=\; \tau(n) \;\text{ for all }\; j \in [\tau(n), n]. \tag{4.39}$$

**Lemma 4.40** *For all $n \in \mathbb{Z}$ there exists an $i \le n$ such that $N(i) = 1$. Furthermore*

$$\mathbb{P}(n - \tau(n) > k) \;\le\; (1-c)^k \tag{4.41}$$

**Proof.** Immediate. □

**Definition 4.42 (The stationary process)** Let $Q$, $\mu$, $c$, $(U_i)$ and $N(\cdot)$ be as in Definition 4.37. For those $i$ such that $N(i) = 1$ define

$$X_i \;=\; \sum_y y\, \mathbf{1}\{U_i \in J(y)\} \tag{4.43}$$

where $(J(y))$ are defined in (4.9) and satisfy (4.10). The values (4.43) are well defined because in this case $U_i < c$ and $J(y) \subset [0, c]$ for all $y \in \mathcal{X}$.

To define $X_i$ for those $i$ such that $N(i) = 0$ we use the function $F$ of the Theorem 4.7: inductively let

$$X_j = F(X_{j-1}; U_j) \tag{4.44}$$

for $j \in [\tau(n), n]$. Property (4.39) guarantees that this construction does not depend on the starting point.

**Theorem 4.45** *Let $Q$ be a transition matrix on a countable state space $\mathcal{X}$. Assume there exists a measure $\mu$ on $\mathcal{X}$ such that $C(\mu, Q) > 0$. Then the process $(X_n : n \in \mathbb{Z})$ defined in Definition 4.42 is a stationary Markov process with transition matrix $Q$. The marginal law $\nu$ on $\mathcal{X}$ defined by*

$$\nu(x) := \mathbb{P}(X_0 = x) \tag{4.46}$$

*is the unique invariant measure for $Q$.*

**Proof of Theorem 2.24** Let $\widetilde{\mathcal{X}}$ be an irreducible class of states of $\mathcal{X}$. Assume first that $\widetilde{\mathcal{X}}$ is aperiodic. Then, by Proposition 3.49, there exists a positive $k$ such that $\widetilde{Q}$, the transition matrix restricted to $\widetilde{\mathcal{X}}$ satisfies

$$\widetilde{Q}^k(x, y) > 0 \quad \text{for all} \quad x, y \in \widetilde{\mathcal{X}} \tag{4.47}$$

Hence $\beta(\widetilde{Q}^k) > 0$ and we can apply Theorem 4.45 to show that the law $\widetilde{\nu}$ of $\widetilde{X}_0$, the value at time zero of the process restricted to $\widetilde{\mathcal{X}}$ given in Definition 4.42 is an invariant measure for $\widetilde{Q}$. Hence the measure $\nu$ defined by

$$\nu(x) := \begin{cases} \widetilde{\nu}(x) & \text{if } x \in \widetilde{Q} \\ 0 & \text{otherwise} \end{cases} \tag{4.48}$$

is invariant for $Q$.

If $\widetilde{\mathcal{X}}$ has period $d$, then by Proposition 3.50, $Q^d$ is aperiodic in each of the equivalence classes $\widetilde{\mathcal{X}}_1, \dots, \widetilde{\mathcal{X}}_d$. Use the argument above to construct invariant measures $\mu_1, \dots, \mu_d$ for $Q^d$ on $\widetilde{\mathcal{X}}_1, \dots, \widetilde{\mathcal{X}}_d$, respectively. Then, $\mu := (1/d) \sum_i \mu_i$ is invariant for $Q$, by Proposition 3.50. □

**Proof of the existence part of Theorem 3.38.** If $\beta(Q) > 0$ it suffices to choose $\mu$ as in (4.32) and apply Theorem 4.45. □

**Proof of Theorem 4.45.** Let $(X_n : n \in \mathbb{Z})$ (Notice: $n \in \mathbb{Z}$) be the process constructed in (4.38)–(4.44). The construction is translation invariant. That is

$$\mathbb{P}(X_{t+1} = x_1, \dots, X_{t+k} = x_k) \;=\; \mathbb{P}(X_1 = x_1, \dots, X_k = x_k) \tag{4.49}$$

for all $t \in \mathbb{Z}$. This implies the process is stationary. The Markov property follows from (4.44). Hence

$$\mathbb{P}(X_1 = y) \;=\; \sum_x \mathbb{P}(X_1 = y \mid X_0 = x)\mathbb{P}(X_0 = x) \tag{4.50}$$

That is,

$$\nu(y) \;=\; \sum_x \mathbb{P}(X_1 = y \mid X_0 = x)\nu(x) \tag{4.51}$$

which implies $\nu$ is invariant for $Q$.

To show uniqueness let $(X_n^{-k,\mu} : n \geq -k)$ be a process starting at time $-k$ with the invariant measure $\mu$, also constructed using the function (4.44). Then

$$\begin{aligned}
&|\nu(y) - \mu(y)| && (4.52)\\
&= \Big|\sum_a \mathbb{P}(X_{-n} = a)\mathbb{P}(X_0 = y \mid X_{-n} = a) - \sum_b \mu(b)\mathbb{P}(X_0 = y \mid X_{-n} = b)\Big| \\
&\leq \sum_a \sum_b \mathbb{P}(X_{-n} = a)\mu(b)\,|\mathbb{P}(X_0 = y \mid X_{-n} = a) - \mathbb{P}(X_0 = y \mid X_{-n} = b)| \\
&\leq \sum_a \sum_b \mathbb{P}(X_{-n} = a)\mu(b)\mathbb{P}(\tau(0) < -n) \\
&= (1-c)^{n+1} && (4.53)
\end{aligned}$$

Since (4.52) is independent of $n$ and (4.53) goes to zero as $n \to \infty$, we conclude that (4.52) must vanish. □

Notice that Theorem 4.45 holds for *any* measure $\mu$ in the regeneration times of Theorem 4.45. To optimize the bounds in (4.26) and (4.53) we can choose a measure $\mu$ which maximizes $C(\mu, Q)$. This is

$$\mu(y) := \frac{\min_x Q(x,y)}{\sum_y \min_x Q(x,y)} \tag{4.54}$$

and for this $\mu$

$$C(\mu, Q) = \beta(Q) = \sum_y \min_x Q(x, y) \tag{4.55}$$

## 4.5 Perfect simulation

The above construction naturally gives us an algorithm to perfect simulate the invariant measure in a minimal number of steps. Let $Q$ be a transition matrix in a finite probability state space $\mathcal{X}$ such that $\beta(Q) > 0$.

**Algorithm 4.56 (Sample the invariant measure)** Perform the following steps

1. Choose $\mu$ as in (4.54) (This can be done because $\beta(Q) > 0$).
2. Simulate uniform random variables $U_0, U_{-1}, \dots$ up to $\tau(0)$, the first time $U_{-n} < \beta(Q)$. (Here we only need a random geometrically distributed number of steps).
3. Compute $X_{\tau(0)}$ using (4.43):

$$X_{\tau(0)} \leftarrow \sum_y y\, \mathbf{1}\{U_{\tau(0)} \in J(y)\} \tag{4.57}$$

   and $X_{\tau(0)+1}, \dots, X_0$ using (4.44):

$$X_j \leftarrow F(X_{j-1}; U_j) \tag{4.58}$$

   for $j \in [\tau(0)+1, 0]$. Important: in both (4.57) and (4.58) use the *same* uniform random variables generated in (2).
4. Output the value $X_0$. End.

**Theorem 4.59** *The law of $X_0$ generated by Algorithm 4.56 is exactly the law of the unique invariant measure $\nu$ for $Q$:*

$$\mathbb{P}(X_0 = x) = \nu(x) \tag{4.60}$$

*for all $x \in \mathcal{X}$.*

**Proof.** Immediate. □

If one wants to generate a piece $(X_0, \dots, X_n)$ of the stationary process, the algorithm is as follows

**Algorithm 4.61 (Sample the stationary process)** Perform the following steps

1. Use Algorithm 4.56 above to generate $X_0$.
2. Generate uniform random variables $U_1, \dots, U_n$
3. Set $X_1, \dots, X_n$ using (4.58).
4. Output the vector $(X_0, \dots, X_n)$. End.

**Theorem 4.62** *The law of the random vector $(X_0, \dots, X_n)$ generated by Algorithm 4.61 is the following*

$$\mathbb{P}(X_i = x_i\,, i \in \{0, \dots, n\}) \;=\; \nu(x_0)Q(x_0, x_1)\dots Q(x_{n-1}, x_n) \qquad (4.63)$$

*where $\nu$ is the unique invariant measure for $Q$.*

**Proof.** Immediate. □

## 4.6 Coupling from the past

Let $Q$ be an irreducible transition probability matrix on a finite state space $\mathcal{X} = \{1, \dots, N\}$. We start with the definition of multi coupling. It is essentially the same as the definition of coupling given in Definition 3.1, but now we allow as many marginals as elements are in $\mathcal{X}$.

**Definition 4.64 (Multi coupling)** Let a function $\widetilde{F} : \mathcal{X}^N \times \mathcal{U} \to \mathcal{X}^N$ be such that for all $x_1, \dots, x_N \in \mathcal{X}$ , all $i = 1, \dots, N$ and all $y_i \in \mathcal{X}$:

$$\sum_i \mathbb{P}(\widetilde{F}(x_1, \dots, x_N; U) = (y_1, \dots, y_N)) \;=\; Q(x_i, y_i) \qquad (4.65)$$

where $\sum_i$ is the sum over all $(N-1)$-tuples $(y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_N)$ and $U$ is a random variable in $\mathcal{U}$. Such an $\widetilde{F}$ will be called a *multi coupling function*. A *multi coupling process* in $\mathcal{X}^N$ is defined by

$$(X_n^1, \ldots, X_n^N) \;:=\; \begin{cases} (1, \ldots, N) & \text{if } n = 0 \\ \widetilde{F}(X_{n-1}^1, \ldots, X_{n-1}^N; U_n) & \text{if } n > 0 \end{cases} \tag{4.66}$$

where $(U_n)$ is a sequence of *iid* random variables defined on $\mathcal{U}$.

One naturally can extend Definitions 3.4 and 3.8 to the notion of free multi coupling and independent multi coupling. There is no simple way to extend the coupling (3.61) to a Dobrushin multi coupling. The reason is that Dobrushin coupling is a genuine two-coordinates coupling in the sense that the way one coordinate uses the variable $U$ to compute its value depends on the value of the other coordinate.

**Definition 4.67** Let $k \in \mathbb{Z}$ be an arbitrary time. Let $(X_n^{1;k}, \ldots, X_n^{N;k})$ be a multi coupling of the Markov chain in $\mathcal{X}$ with transition matrix $Q$, starting at time $k$ with states $(1, \ldots, N)$, defined as follows:

$$(X_n^{1;k}, \ldots, X_n^{N;k}) \;:=\; \begin{cases} (1, \ldots, N) & \text{if } n = k \\ \widetilde{F}(X_{n-1}^{1;k}, \ldots, X_{n-1}^{N;k}; U_n) & \text{if } n > k \end{cases} \tag{4.68}$$

Define

$$\kappa^k \;:=\; \min\{\ell \geq k \;:\; X_\ell^{a;k} = X_\ell^{b;k} \text{ for all } a, b \in \{1, \ldots, N\}\} \tag{4.69}$$

In words, $\kappa^k$ is the first time all the marginals of the multi coupling started at time $k$ with all possible states meet. Define

$$\tau(n) \;:=\; \max\{k \leq n \;:\; \kappa^k \leq n\} \tag{4.70}$$

That is, $\tau(n)$ is the last time prior to $n$ that the configuration at $n$ does not depend on the state the process assumes at times $k \leq \tau(n)$.

**Theorem 4.71 (Coupling from the past)** *Let $U_n$ be a sequence of* iid *random variables in $\mathcal{U}$. Let $\widetilde{F}$ be a multi coupling function satisfying* (4.65).

*Let* $(((X_n^{1;k},\dots,X_n^{N;k}) : n \ge k) : k \in \mathbb{Z})$ *be a family of multi couplings of the Markov chain in* $\mathcal{X}$ *with transition matrix* $Q$*, starting at all possible times* $k \in \mathbb{Z}$ *with states* $(1,\dots,N)$ *defined in* (4.68) *(and with the same sequence* $(U_n)$*). Assume* $\tau(0)$ *finite, that is,* $\mathbb{P}(\tau(0) > -\infty) = 1$*. Then the matrix* $Q$ *admits a unique invariant measure* $\nu$*, the law of any marginal of the process at time zero starting at time* $\tau(0)$*:*

$$\mathbb{P}(X_0^{a;\tau(0)} = y) \;=\; \nu(y) \tag{4.72}$$

*for all* $a, y \in \mathcal{X}$*.*

Before proving the theorem we make a comment and then show how the result is used for perfect simulation applications. In view of Definition 4.36 one is tempted to guess that $\tau(0)$ is a regeneration time for the chains. But this is not true, because, for the multi coupling time $\tau(n)$ (4.39) does not hold.

The original perfect simulation algorithm proposed by Propp and Wilson is the following:

**Algorithm 4.73 (Coupling from the past)** Perform the following steps

1. Fix a moderate time $-t < 0$ and generate independent random variables $U_{-t}, U_{-t+1}, \dots, U_0$.

2. With the random variables $U_{-t},\dots,U_0$ previously generated, construct a sample of the multi coupling $(X_\ell^{1;-t},\dots,X_\ell^{N;-t})$ for $\ell = -t,\dots,0$, using a multi coupling function $\widetilde{F}$.

3. If
$$X_0^{a;-t} = X_0^{b;-t} \;\text{ for all }\; a, b \in \{1,\dots,N\} \tag{4.74}$$
then $X_0^{1;-t}$ is a perfect sample of the invariant measure $\nu$. Print the result and finish. If (4.74) does not hold, then generate independent random variables $U_{-2t},\dots,U_{-t-1}$, set
$$t \leftarrow 2t \tag{4.75}$$
and go to (2).

The meaning of "moderate time" is something only experience can give. If no a priori estimation exists, just start with an arbitrary value.

The algorithm avoids the explicit computation of $\tau(0)$. It starts multi coupling from time $-t$ to time 0. If all marginals coincide at time 0, then the algorithm finishes and the value of a(ny) marginal at time zero is a perfect sample of the invariant measure. If some marginals disagree then, generate new $U$ random variables from time $-2t$ up to time $-t-1$ and multi couple again from time $-2t$, etc. Use the $U$ variables already generated to construct the multi coupling in the time interval going from $-t$ to 0: The crucial point is that once the variable $U_n$ is generated, it is used in all future steps to generate the multi coupling at time $n$.

**Proof of Theorem 4.71** Observe first that $\mathbb{P}(\tau(0) = -\infty) = 0$, and

$$X_0^{a,\tau(0)} \, \mathbf{1}\{\tau(0) > -t\} \;=\; X_0^{a,-t} \, \mathbf{1}\{\tau(0) > -t\}, \tag{4.76}$$

imply

$$\mathbb{P}(X_0^{a,\tau(0)} = y) \;=\; \lim_{t\to\infty} \mathbb{P}(X_0^{a,-t} = y) \tag{4.77}$$

The fact that $\nu$ is invariant follows from the construction: $\mathbb{P}(X_n^{a,\tau(n)} = y)$ does not depend on $n$. Furthermore, by construction, the process $(X_n^{a,\tau(n)})$ is a Markov process with transition matrix $Q$. Since $\nu$ is invariant,

$$\begin{aligned} & |\mathbb{P}(X_0^{a,-t} = y) - \nu(y)| && (4.78) \\ =\; & \Big|\mathbb{P}(X_0 = y \,|\, X_{-t} = a) - \sum_b \nu(b) \mathbb{P}(X_0 = y \,|\, X_{-t} = b)\Big| \\ \le\; & \sum_b \nu(b) \, |\mathbb{P}(X_0 = y \,|\, X_{-t} = a) - \mathbb{P}(X_0 = y \,|\, X_{-t} = b)| \\ \le\; & \sum_b \nu(b) \mathbb{P}(\tau(0) < -t) \\ \to\; & 0 && (4.79) \end{aligned}$$

as $t \to \infty$. This and (4.77) show the theorem. □

## 4.7 Exercises

**Exercise 4.1** Show that (4.22) equals (4.23).

**Exercise 4.2** Compare $\tau(n)$ defined by (4.70) for the free coupling with the $\tau(n)$ defined in (4.38) for the same coupling.

**Exercise 4.3** Show the last item of Theorem 4.7. That is, show that

$$\begin{aligned}\mathbb{P}(X_{t+s} = y_s, s = 0, \dots, \ell \,|\, T = t;\, X_i = x_i, i < t) \\ = \mu(y_0)Q(y_0, y_1)\dots Q(y_{\ell-1}, y_\ell) \qquad (4.80)\end{aligned}$$

for arbitrary $\ell, t \geq 0$ and $x_i, y_j$, $i \leq t$, $j \leq \ell$.

**Exercise 4.4** Show (4.35), that is, show that the the meeting time of the free coupling defined with the function $F$ defined in the proof of Theorem 3.18 is less than the regeneration time of the chain constructed with $\mu$ defined by (4.32).

**Exercise 4.5** Under the conditions of Exercise 4.4, give an example where the meeting time of $\tau^{a,b}$ may be strictly smaller than the regeneration time.

**Exercise 4.6** Show that the construction proposed in (4.43)-(4.44) is translation invariant. That is, show (4.49).

**Exercise 4.7** Extend Definitions 3.4 and 3.8 to the notion of free multi coupling and independent multi coupling. How would you define a Dobrushin multi coupling?

## 4.8 Comments and references

Theorem 4.7 which uses the fact that there are regeneration times in Markov chains is due to Athreya and Ney (1978) and Nummelin (1878). Coupling from the past is an idea introduced by Propp Wilson (1996), see Wilson (1998) for references. Theorem 4.71 is proven by Foss and Tweedie (2000). See an exhaustive discussions about background on regeneration times in Thorisson (2000) and Comets, Fernández and Ferrari (2000).

# Chapter 5

# Renewal Processes.

## 5.1 Renewal processes

We start with the definition of renewal processes

**Definition 5.1 (Renewal process)** A *renewal process* is a strictly increasing sequence of random variables

$$0 \leq T_1 < T_2 < \cdots < T_k < \cdots$$

with values in $\mathbb{N}$ or $\mathbb{R}$ satisfying the following conditions:

1. The random variables $T_1$, $T_2 - T_1$,...,$T_{k+1} - T_k$, $k \geq 1$ are mutually independent.

2. The random variables $T_{k+1} - T_k$, $k \geq 1$ are identically distributed.

The variables $(T_k)_{k \geq 1}$ model the successive occurrence times of some phenomenon which repeats independently from the past history. For instance the random variables $T_k$ may be interpreted as the successive instants of replacement of an electric bulb. Condition (1) somehow express the fact that the lifetime of each bulb is independent of the others. Condition (2) is verified only if the conditions of occurrence of the phenomenon are unaltered.

This means that we replace always with bulbs with the same specifications and that the conditions of the electric network do not change with time.

The reason why we excluded $T_1$ from condition (2) is that the time up to the first occurrence may depend on other factors related to the observation. For instance, if we model the arrival times of a “perfect metro”, for which the time interval between two trains is (for instance) exactly 3 minutes, then $T_{k+1} - T_k \equiv 3$, for all $k \geq 1$. However, $T_1$ depends on the moment we choose as observation starting time. For example, if one arrives to the metro station unaware of the schedule and starts counting the time intervals, we can assume that $T_1$ has uniform distribution in $[0, 3]$. We present now some examples.

In this chapter we consider *discrete time* renewal processes, that is, $T_n \in \mathbb{N}$.

**Example 5.2** Let $U_1, U_2, \ldots$ be a sequence of *iid* random variables with values in $\{-1, +1\}$ and law

$$\mathbb{P}(U_n = +1) = p = 1 - \mathbb{P}(U_n = -1),$$

where $p \in [0, 1]$. Define

$$T_1 = \inf\{n \geq 1 \,:\, U_n = -1\} \;\; \text{e}$$

$$T_k = \inf\{n \geq T_{k-1} \,:\, U_n = -1\}, \text{ for all } k \geq 2.$$

In this case, the independence of the random variables $U_n$ implies immediately the independence condition (1) of the definition and also the fact that the increments $T_{k+1} - T_k$ are identically distributed, for all $k \geq 1$. In this case it is easy to see that the increments, as well as $T_1$, have geometric distribution.

**Example 5.3** Let $(X^a_n)_{n\in\mathbb{N}}$ be an irreducible Markov chain on $\mathcal{X}$ with initial state $a$. Let $b \in \mathcal{X}$. Define the successive visits to $b$:

$$T^{a\to b}_1 = \inf\{n \geq 1 : X^a_n = b\} \;\; \text{and}$$

$$T^{a\to b}_k = \inf\{n \geq T^{a\to b}_{k-1} : X^a_n = b\}, \text{ for all } k \geq 2.$$

The Markov property implies that the increasing sequence $(T_n^{a\to b})$ satisfies conditions (1) and (2). Notice that in this case, if $a = b$, then $T_1^{a\to b}$ and $T_{k+1}^{a\to b} - T_k^{a\to b}$ have the same law for all $k \geq 1$.

**Definition 5.4** Given a renewal process $(T_n)_{n\geq 1}$, for all pair of times $s \leq t$, define the *counting measure* $\mathbf{N}[s,t]$ as follows

$$\mathbf{N}[s,t] = \sum_{k\geq 1} \mathbf{1}\{s \leq T_k \leq t\}.$$

The measure $\mathbf{N}[s,t]$ counts the number of events of the renewal process between times $s$ and $t$. We also use the notation

$$\mathbf{N}\{t\} = \mathbf{N}[t,t] = \sum_{k\geq 1} \mathbf{1}\{T_k = t\}.$$

**Lemma 5.5** *Let $(T_n)_{n\geq 1}$ be a renewal process with values in $\mathbb{N}$ and let $t \in \mathbb{N}$. Then*

$$\sum_{k\geq 1} \mathbf{1}\{T_k = t\} = \mathbf{1}\left\{t \in \{T_k : k \geq 1\}\right\},$$

*and hence*

$$\mathbb{P}(\mathbf{N}\{t\} = 1) = \mathbb{P}\left(t \in \{T_k : k \geq 1\}\right). \tag{5.6}$$

**Proof.** This is just a simple exercise left to the reader. □

## 5.2 Basic questions

In this chapter we study the following basic questions for renewal processes:

**Question 5.7 (Stationarity)** Determine the law of $T_1$ under which the law of the counting measure $\mathbf{N}[s, s+t]$ is independent of $s$.

**Question 5.8 (Law of Large Numbers)** Study the following limit in mean, in probability and almost sure

$$\lim_{t\to+\infty} \frac{\mathbf{N}[0,t]}{t}.$$

**Question 5.9 (Key Theorem)** Determine the limit

$$\lim_{s \to +\infty} \mathbb{P}(\mathbf{N}[s, s+t] = n).$$

for any fixed $t$ and $n$.

Notice that in Questions 5.8 and 5.9 the existence of the limits must also be proven.

## 5.3 Relation with the flea-jump process

In this section we introduce an associated Markov chain with state space $\mathcal{X} = \mathbb{N}$ related with the renewal process. For the Markov chain the answers to Questions relatec to 5.7, 5.8 and 5.9 are just applications of results established in previous chapters.

**Lemma 5.10 (Translation lemma)** *Let $(T_n)_{n\geq 1}$ be a renewal process on $\mathbb{N}$ and $\nu$ the common law of the increments $T_{k+1} - T_k$, that is*

$$\mathbb{P}(T_{k+1} - T_k = n) = \nu(n),$$

*for all $n \in \mathbb{N}$. Then*

$$\mathbf{N}[m, n] = \sum_{m \leq t \leq n} \mathbf{1}\{X_t^a = 0\} \quad \textit{and}$$

$$\mathbb{P}(\mathbf{N}\{t\} = 1) \ = \ \sum_a \mathbb{P}(X_t^a = 0)\, \mathbb{P}(T_1 = a)$$

*where $(X_n^a)_{n\in\mathbb{N}}$ is the Markov chain on $\mathbb{N}$ with initial state $a$ and transition matrix $Q_\nu$ defined by:*

$$Q_\nu(0, x) = \nu(x+1), \ \textit{for all } x \in \mathbb{N}; \tag{5.11}$$

*and*

$$Q_\nu(x, x-1) = 1, \ \textit{for all } x \geq 1. \tag{5.12}$$

**Proof.** This is a graphic proof. At time 0, we mark the point with ordinate $a = T_1$. Starting from the point $(0, a)$ we draw a $45^0$ line up to the point $(a, 0)$. Then mark the point $(a+1, D_1)$, where $D_k = T_{k+1} - T_k - 1$, for all $k \geq 1$. Repeat the procedure, drawing a $45^0$ line linking the points $(a+1, D_1)$ and $(a+1+D_1, 0)$. Then mark the point $(a+2+D_1, D_2)$ and repeat the procedure, drawing a line up to $(a+2+D_1+D_2, 0)$. In general, we mark the point $(a+k+\sum_{j=1}^{k-1} D_j, D_k)$ and draw a line perpendicular to the diagonal up to the point $(a+k+\sum_{j=1}^{k} D_j, 0)$.

For each $t \in \mathbb{N}$, define $X_t^a$ as the ordinate of abscise $t$ in the graphic constructed above. The reader will be able to prove that the process $(X_t^a)_{t\in\mathbb{N}}$ so defined is a Markov chain with transition matrix $Q_\nu$.

This construction also proves that the points $T_k$ of the renewal process are exactly the times for which the chain $(X_t^a)_{t\in\mathbb{N}}$ visits state 0. This and Lemma 5.5 finishes the proof. □

## 5.4 Stationarity

In this section we answer Question 5.7. From now on we use the letter $\nu$ to call the common law of the increments $T_{k+1}-T_k$ and $\mathcal{X}_\nu := \{i \in \mathbb{N} : \nu(i) > 0\}$ is the support of the measure $\nu$.

**Definition 5.13** Let $\nu$ be a probability distribution on $\{1, 2, \cdots\}$ with finite mean $\theta$, that is:

$$\theta = \sum_{n\geq 1} n\nu(n) < +\infty.$$

Define the probability measure $G^\nu$ on $\mathbb{N}$ as follows. For all $x \in \mathbb{N}$

$$G^\nu(x) = \frac{\nu(x, +\infty)}{\theta}, \tag{5.14}$$

where

$$\nu(x, +\infty) = \sum_{y>x} \nu(y).$$

**Remark 5.15** The identity

$$\theta = \sum_{n\geq 1} n\nu(n) = \sum_{x\geq 0} \nu((x,+\infty)) \tag{5.16}$$

is proven as the integration by parts formula and is left as an easy exercise to the reader. The identity (5.16) guarantees that $G^\nu$ is a probability distribution when $\theta < +\infty$.

**Proposition 5.17** *Let $(T_n)_{n\geq 1}$ be a renewal process on $\mathbb{N}$ and let $\nu$ be the common law of the increments $T_{k+1}-T_k$. Assume $\theta$ finite. Then $\mathbb{P}(\mathbf{N}\{t\}=1)$ is constant in $t$ if and only if $T_1$ has law $G^\nu$.*

**Proof.** The Translation Lemma 5.10 implies that

$$\mathbb{P}(\mathbf{N}\{t\}=1) \;=\; \sum_{k\geq 0} \mathbb{P}(T_1=k)\mathbb{P}(X_t^k=0), \tag{5.18}$$

where the super label $k$ indicates the starting state of the chain. On the other hand, the law of $N(t)$ does not depend on $t$ if and only if

$$\mathbb{P}(\mathbf{N}\{t\}=1) \;=\; \mathbb{P}(\mathbf{N}\{0\}=1) \;=\; \mathbb{P}(T_1=0) \tag{5.19}$$

for all $t\geq 0$. Hence the law of $N(t)$ does not depend on $t$ if and only if the law of $T_1$ satisfies

$$\mathbb{P}(T_1=0) \;=\; \sum_{k\geq 0} \mathbb{P}(T_1=k)\mathbb{P}(X_t^k=0) \tag{5.20}$$

which is exactly the equation an invariant measure for the chain $X_t$ must satisfy. We have proved that $\mathbb{P}(\mathbf{N}\{t\}=1)$ constant in $t$ is equivalent to the invariance of the law $\{\mathbb{P}(T_1=k), k\geq 0\}$ for the chain $(X_t)_{t\in\mathbb{N}}$.

This means that the stationarity of $N(t)$ is equivalent to say that the starting point of the chain, that is $T_1$, is chosen following the invariant measure of the matrix $Q_\nu$, defined in the Translation Lemma.

Notice that $Q_\nu$ is irreducible in the set

$$\{0, \sup\{x\geq 0 : \nu(x+1)>0\}\}$$

and that the condition $\theta < +\infty$ is equivalent to positive recurrence of $Q_\nu$. This ensures the existence of an invariant measure for $Q_\nu$. Finally, as the reader will easily check, the measure $G^\nu$ is invariant for the transition matrix $Q_\nu$. □

**Corollary 5.21** *Under the conditions of Proposition 5.17, if $T_1$ has law $G^\nu$, then*

$$\mathbb{P}(\mathbf{N}\{t\} = 1) = \frac{1}{\theta}.$$

*for all $t \geq 0$*

**Definition 5.22** When the law of $\mathbf{N}\{t\}$ is independent of $t$ we will say that the renewal process is *stationary.*

## 5.5 Law of large numbers

We now use the translation lemma to solve Question 5.8.

**Proposition 5.23 (Law of large numbers for the averages)** *Assume the conditions of Proposition 5.17. If*

$$\sum_{x \geq 1} x G^\nu(x) < +\infty \quad \textit{and} \quad \mathbb{E}(T_1) < +\infty,$$

*then*

$$\lim_{t \to \infty} \frac{\mathbb{E}(\mathbf{N}[0,t])}{t+1} = \frac{1}{\theta}.$$

**Proof.** The proof is based in a coupling of the renewal process $(T_k)_{k \geq 1}$ with another renewal process $(S_k)_{k \geq 1}$, whose increments $(S_{k+1} - S_k)_{k \geq 1}$ have the same law $\nu$ as the increments $(T_{k+1} - T_k)_{k \geq 1}$ of the original process but the initial time $S_1$ has law $G^\nu$. By proposition 5.17, $(S_k)$ is a stationary renewal process. The coupling is performed in such a way that both renewal processes have the same increments except for the first time interval.

Let $S_1$ be a random variable with law $G^\nu$, independent of the process $(T_k)_{k\geq 1}$. Then, for all $k \geq 2$, define

$$S_k = S_{k-1} + T_k - T_{k-1}.$$

Let $\mathbf{N}_T[0,t]$ and $\mathbf{N}_S[0,t]$ be the counting measures corresponding to the processes $(T_k)_{k\geq 1}$ and $(S_k)_{k\geq 1}$, respectively. That is,

$$\mathbf{N}_T[0,t] \;:=\; \sum_{k\geq 1} \mathbf{1}\{T_k \leq t\},$$

$$\mathbf{N}_S[0,t] \;:=\; \sum_{k\geq 1} \mathbf{1}\{S_k \leq t\}.$$

By construction, since we are working with discrete time, the following inequality holds

$$|\mathbf{N}_T[0,t] - \mathbf{N}_S[0,t]| \leq S_1 + T_1\,. \tag{5.24}$$

This implies

$$\frac{1}{t+1}\mathbb{E}|\mathbf{N}_T[0,t] - \mathbf{N}_S[0,t]| \leq \frac{1}{t+1}\mathbb{E}(S_1 + T_1). \tag{5.25}$$

By hypothesis,

$$\mathbb{E}(T_1) < +\infty \quad \text{and} \quad \mathbb{E}(S_1) = \sum_{x\geq 1} xG^\nu(x) < +\infty.$$

Under these conditions (5.24) implies

$$\lim_{t\to\infty} \frac{|\mathbb{E}(\mathbf{N}_T[0,t]) - \mathbb{E}(\mathbf{N}_S[0,t])|}{t+1} = 0.$$

It now suffices to observe that since the renewal process $(S_k)_{k\geq 1}$ is stationary,

$$\mathbb{E}\mathbf{N}_S[0,t] = \sum_{s=0}^{t} \mathbb{E}(\mathbf{N}_S\{s\}) = (t+1)\frac{1}{\theta},$$

which finishes the proof. $\square$

**Proposition 5.26 (Law of Large Numbers a.s.)** *If $T_1$ and $T_{k+1}-T_k$ assume values in a finite set $\{1,\dots,k\}$, then*

$$\mathbb{P}\left(\lim_{t\to\infty}\frac{\mathbf{N}[0,t]}{t}=\frac{1}{\theta}\right)=1. \tag{5.27}$$

In fact Proposition 5.26 holds under the hypothesis $\mathbb{E}T_1<\infty$ and $\mathbb{E}(T_k-T_{k-1})<\infty$. We prefer to stay in the finite case because the idea of the proof is the same but much more elementary.

**Definition 5.28** If 5.27 holds, we say that $\frac{\mathbf{N}[0,t]}{t}$ converges *almost surely* to $1/\theta$.

**Proof.** In this proof we will use the following law of large numbers. If $\{X_i : i\ge 0\}$ is a sequence of *iid* random variables, then

$$\mathbb{P}\left(\lim_{n\to\infty}\frac{\sum_{i=1}^n X_i}{n}=\mathbb{E}X_1\right)\;=\;1.$$

If we write $t=T_{\mathbf{N}[0,t]+1}-T_1+(T_{\mathbf{N}[0,t]+1}-t)+T_1$, we obtain

$$\frac{t}{\mathbf{N}[0,t]}=\frac{T_{\mathbf{N}[0,t]+1}-T_1}{N\mathbb{N}[0,t]}-\frac{T_{\mathbf{N}[0,t]+1}-t}{\mathbf{N}[0,t]}+\frac{T_1}{\mathbf{N}[0,t]}. \tag{5.29}$$

We treat each of these terms separately. The first term may be written as

$$\frac{\sum_{i=1}^{\mathbf{N}[0,t]}(T_{i+1}-T_i)}{\mathbf{N}[0,t]}.$$

Since $\nu(0)=0$, we have $\mathbf{N}[0,t]\ge t$, which implies in particular that $\mathbf{N}[0,t]$ goes to infinity as $t$ goes to infinity. (This can be proven even when $\nu(0)\in(0,1)$). Hence we can apply the law of large numbers for independent variables to the first term to prove that this term goes to $\theta=\mathbb{E}(T_i-T_{i-1})$.

Since the law of the inter-renewal times concentrates on a finite set, the numerator of the second term in (5.29) is bounded; since $\mathbf{N}[0,t]$ goes to infinity, this term converges to zero almost surely.

The last term goes also to zero because it is the quotient between a random variable and another one that goes to infinity. □

## 5.6 The Key Theorem

We now solve Question 5.9. This result is generally called *Key Theorem* or *Renewal theorem*. We assume that $\mathcal{X}_\nu$ finite.

**Theorem 5.30 (Key Theorem – finite version)** *Assume $\mathcal{X}_\nu$ finite and $Q_\nu$ an aperiodic transition matrix. Under the conditions of Proposition 5.17, for any law of $T_1$ on $\mathcal{X}_\nu$, there exists the limit*

$$\lim_{t\to+\infty} \mathbb{P}\left(\mathbf{N}\{t\}=1\right) = \frac{1}{\theta}.$$

*Furthermore, if $a_1,\dots,a_n \in \{0,1\}$ are arbitrary, then for any $k$*

$$\lim_{t\to+\infty} \mathbb{P}\left(\mathbf{N}\{t+i-k\}=a_i, i=1,\dots,n\right) = H(a_1,\dots,a_n), \tag{5.31}$$

*where*

$$H(a_1,\dots,a_n) = \mathbb{P}\left(\mathbf{N}_S\{i\}=a_i, i=1,\dots,n\right).$$

*where $\mathbf{N}_S$ is the counting measure for the stationary renewal process introduced in Proposition 5.23.*

**Proof.** The Translation Lemma 5.10 says that

$$\mathbb{P}\left(\mathbf{N}\{t\}=1\right) \;=\; \mathbb{P}(X_t=0). \tag{5.32}$$

Since $\mathcal{X}_\nu$ is finite and $Q_\nu$ is irreducible, there exists a $k$ such that $Q_\nu^k(x,y)>0$ for all $x,y\in\mathcal{X}_\nu$. Hence $\beta(Q_\nu^k)>0$ and we can apply Corollary 3.44 to conclude that $\mathbb{P}(X_t=0)$ converges to $G^\nu(0)$, the probability of 0 under the invariant measure. We have $G^\nu(0)=1/\theta>0$ because $\theta<+\infty$. To show (5.31) use again the translation Lemma to get

$$\mathbb{P}\left(\mathbf{N}\{t+i-k\}=a_i, i=1,\dots,n\right) \;=\; \sum_{b_i\in B_i} \mathbb{P}(X_{t+i-k}=b_i, i=1,\dots,n) \tag{5.33}$$

where

$$B_i \;=\; \begin{cases} \{0\} & \text{if } a_i=1 \\ \{1,\dots,|M_\nu|-1\} & \text{if } a_i=0 \end{cases} \tag{5.34}$$

with $M_\nu := \max\{n : \nu(n) > 0\}$. Since the sum in (5.33) is finite, we can pass the limit inside the sum. For each term, Corollary 3.44 and the Markov property imply

$$\lim_{t\to\infty} \mathbb{P}(X_{t+i-k} = b_i, i = 1,\dots,n) \;=\; G_\nu(b_1)\prod_{i=1}^{n-1} Q_\nu(b_i, b_{i+1})\,. \tag{5.35}$$

Summing on $b_i$ we get $H(a_1,\dots,a_n)$. □

Two important concepts in renewal theory are the age and the residual time of a process. We define the *age* $A(t)$ and the *residual time* $R(t)$ of the process at time $t$ as follows:

$$A(t) \;:=\; t - T_{\mathbf{N}[0,t]}; \qquad R(t) \;:=\; \begin{cases} T_{\mathbf{N}[0,t]+1} - t & \text{if } \mathbf{N}\{t\} = 0 \\ 0 & \text{if } \mathbf{N}\{t\} = 1 \end{cases}.$$

Intuitively, if $T_{i+1} - T_i$ represents the lifetime of the $i$th bulb, then there is a renewal each time that a bulb is changed. In this case $A(t)$ represents the age of the bulb currently in function at time $t$ and $R(t)$ represents the time this bulb will still be working. A consequence of the Key Theorem is that we can compute the asymptotic law of these variables as $t \to \infty$.

**Corollary 5.36** *Under the hypotheses of the Key Theorem, the laws of $A(t)$ and $R(t)$ converge to $G^\nu$ as $t \to \infty$.*

**Proof.** Notice that

$$\mathbb{P}(A(t) = k) \;=\; \mathbb{P}(\mathbf{N}\{t-k\} = 1, \mathbf{N}[t-k+1, t] = 0),$$

which, by the Key Theorem converges, as $t \to \infty$ to

$$\begin{aligned} \mathbb{P}(\mathbf{N}_S\{0\} = 1, \mathbf{N}_S[1,k] = 0) &= \mathbb{P}(S_1 = 0)\mathbb{P}(S_2 - S_1 > k) \\ &= \frac{1}{\theta}\nu(k,\infty) = G^\nu(k). \end{aligned} \tag{5.37}$$

On the other hand,

$$\mathbb{P}(R(t) = k) = \mathbb{P}(\mathbf{N}[t, t+k-1] = 0, \mathbf{N}\{t+k\} = 1),$$

which, by the Key Theorem, converges as $t \to \infty$ to

$$\mathbb{P}(\mathbf{N}_S[t, t+k-1] = 0, \mathbf{N}_S\{t+k\} = 1) = \mathbb{P}(S_1 = k) = G^\nu(k). \;\square$$

## Exponential convergence

A direct construction of the renewal process gives a proof of the Key Theorem, which holds also for infinite $\mathcal{X}_\nu$. This construction goes in the vein of the construction of Markov chains of Chapter 1. The proof is simple but requires uniform bounds on the failure rate of the inter-renewal distribution. Let $\nu$ be the common law of $T_{i+1} - T_i$ for $i \geq 1$ and $\nu'$ the law of $T_1$. Define

$$\rho_k := \frac{\nu(k)}{\sum_{i\geq k} \nu(i)}; \quad \rho'_k := \frac{\nu'(k)}{\sum_{i\geq k} \nu'(i)}.$$

The value $\rho_k$ can be interpreted as the *failure rate* of the distribution $\nu$ at time $k$. Let $(U_i : i \geq 1)$, be a family of *iid* random variables uniformly distributed in $[0,1]$. Define

$$T_1 := \min\{n : U_n \leq \rho'(n)\}$$

and for $k \geq 1$,

$$T_{k+1} := \min\{n > T_k : U_n \leq \rho(n - T_k)\}.$$

**Lemma 5.38** *The variables $T_1$ and $T_{k+1} - T_k$ have law $\nu'$ and $\nu$ respectively for $k \geq 1$. Furthermore they are independent. In other words, the process $(Y_n)_{n\geq 1}$ so constructed is a version of the renewal process with laws $\nu'$ and $\nu$ for $T_1$ and $T_{k+1} - T_k$ respectively.*

**Proof.** We prove that $\mathbb{P}(T_1 = k) = \nu'(k)$.

$$\mathbb{P}(T_1 = k) = \mathbb{P}(U_1 > \rho'_1, \ldots, U_{k-1} > \rho'_{k-1}, U_k \leq \rho'_k).$$

Since $U_i$ are *iid* random variables, the above expression equals

$$\mathbb{P}(U_1 > \rho'_1) \ldots \mathbb{P}(U_{k-1} > \rho'_{k-1})\mathbb{P}(U_k \leq \rho'_k) = (1 - \rho'_1) \ldots (1 - \rho'_{k-1})\rho'_k.$$

But $1 - \rho'_i = \nu'[i+1, \infty)/\nu'[i, \infty)$. Hence we get

$$\frac{\nu'[2,\infty)}{\nu'[1,\infty)} \cdots \frac{\nu'[k,\infty)}{\nu'[k-1,\infty)} \frac{\nu'(k)}{\nu'[k,\infty)} = \nu'(k).$$

(We have used $\nu'[1,\infty) = 1$.) The rest of the proof follows the same line and is left to the reader. $\square$

**Theorem 5.39 (Key Theorem with rate of convergence)** *If there exists a constant $\gamma \in (0,1)$ such that for all $n \geq 0$*

$$\mathbb{P}(T_1 = n \mid T_1 \geq n) \ \geq \ \gamma; \qquad \mathbb{P}(T_{k+1} - T_k = n \mid T_{k+1} - T_k \geq n) \ \geq \ \gamma, \quad (5.40)$$

*then*

$$|\mathbb{P}(\mathbf{N}\{t\} = 1) - (1/\theta)| \ \leq \ (1-\gamma)^t.$$

**Proof.** Under the hypotheses (5.40),

$$\rho'_n \geq \gamma \quad \text{and} \quad \rho_n \geq \gamma,$$

for all $n \geq 1$. On the other hand, for the variable $S_1$ (with law $G^\nu$), the following inequalities hold

$$\mathbb{P}(S_1 = n \mid S_1 \geq n) \ \geq \ \gamma. \qquad (5.41)$$

To prove them, notice first that they are equivalent to

$$\frac{\mathbb{P}(S_1 > n+1)}{\mathbb{P}(S_1 > n)} \ \leq \ 1 - \gamma. \qquad (5.42)$$

To show (5.42), notice that

$$\mathbb{P}(S_1 > n) \ = \ \sum_{i=n+1}^{\infty} \mathbb{P}(S_1 = i) \ = \ \frac{1}{\theta} \sum_{i=n+1}^{\infty} \nu[i, \infty). \qquad (5.43)$$

By definition,

$$\nu[i,\infty) \ = \ \mathbb{P}(T_{k+1} - T_k > i-1) \ \geq \ \frac{1}{1-\gamma} \mathbb{P}(T_{k+1} - T_k > i),$$

the inequality is true by hypothesis. On the other hand, using again the definition, the last expression equals

$$\frac{1}{1-\gamma} \nu[i+1, \infty).$$

Hence

$$\mathbb{P}(S_1 > n) \;\geq\; \frac{1}{1-\gamma}\frac{1}{\theta}\sum_{i=n+2}^{\infty}\nu[i,\infty) \;=\; \frac{1}{1-\gamma}\mathbb{P}(S_1 > n+1),$$

This shows (5.41). Defining

$$\rho''_n \;=\; \frac{\mathbb{P}(S_1 = n)}{\mathbb{P}(S_1 \geq n)},$$

inequalities (5.42) imply $\rho''_n \geq \gamma$.

We couple the process $\mathbf{N}[0,t]$ with initial distribution $\nu'$ with the process $\mathbf{N}_S[0,t]$ with initial time $S_1$ distributed according to $G^\nu$. We know that this second process is stationary. This implies that $\mathbb{P}(\mathbf{N}_S\{t\} = 1) = 1/\theta$ for all $t$. Define

$$\tau \;=\; \min\{n \geq 1 : \mathbf{N}\{n\} = 1, \mathbf{N}_S\{n\} = 1\}.$$

Then we have

$$\begin{aligned}
&|\mathbb{P}(\mathbf{N}\{t\} = 1) - (1/\theta)| && (5.44)\\
&= \; |\mathbb{P}(\mathbf{N}\{t\} = 1) - \mathbb{P}(\mathbf{N}_S\{t\} = 1)| \\
&= \; |\mathbb{E}(\mathbf{1}\{\mathbf{N}\{t\} = 1\} - \mathbf{1}\{\mathbf{N}_S\{t\} = 1\} \,|\, \tau \leq t)|\,\mathbb{P}(\tau \leq t) \\
&\qquad + |\mathbb{E}(\mathbf{1}\{\mathbf{N}\{t\} = 1\} - \mathbf{1}\{\mathbf{N}_S\{t\} = 1\} \,|\, \tau > t)|\,\mathbb{P}(\tau > t) \\
&\leq \; \mathbb{P}(\tau > t). && (5.45)
\end{aligned}$$

The last inequality follows because the the absolute value of the difference of indicator functions cannot exceed one. The expectation of the difference of indicator functions indicates that both processes are realized as a function of the same random variables $U_n$. Since for all $n$, $\rho_n > \gamma$, $\rho'_n > \gamma$ and $\rho''_n > \gamma$, we know that $\{U_n < \gamma\} \subset \{\mathbf{N}\{n\} = 1, \mathbf{N}_S\{n\} = 1\}$. This implies that $\tau$ is dominated by $\widetilde{\tau}$, a geometric random variable with parameter $\gamma$:

$$\widetilde{\tau} = \min\{n : U_n < \gamma\}.$$

Hence,

$$\mathbb{P}(\tau > t) \;\leq\; \mathbb{P}(\widetilde{\tau} > t) \;=\; (1-\gamma)^t$$

This shows the proposition. □

## 5.7 Exercises

**Exercise 5.1** Let $U_1, U_2, \dots$ be a sequence of *iid* random variables with values in the set $\{-1, +1\}$ with law

$$\mathbb{P}(U_n = +1) = p = 1 - \mathbb{P}(U_n = -1),$$

where $p \in [0, 1]$. Define

$$T_1 = \inf\{n \geq 1 : U_n = -1\} \text{ e}$$

$$T_k = \inf\{n > T_{k-1} : U_n = -1\} \quad \text{for all} \quad k \geq 2.$$

i) Show that the random variables $T_1, T_2 - T_1, \dots, T_{k+1} - T_k, \dots$ are *iid* with law

$$\mathbb{P}(T_1 = n) = \mathbb{P}(T_{k+1} - T_k = n) = p^{n-1}(1-p), \quad \text{for all} \quad n \geq 1.$$

ii) Let $m < n$ be arbitrary elements of $\mathbb{N}$. Compute

$$\mathbb{P}(\mathbf{N}[m, n] = k), \quad \text{for all} \quad k \in \mathbb{N}.$$

Compute

$$\mathbb{E}(\mathbf{N}[m, n]).$$

**Exercise 5.2** Let $(X_n^a)_{n \in \mathbb{N}}$ be an irreducible Markov chain on $\mathcal{X}$ with initial state $a$. Let $b$ be an arbitrary fixed element of $\mathcal{X}$. Define the successive passage times of the chain at $b$ by

$$T_1^{a \to b} := \inf\{n \geq 1 : X_n^a = b\} \text{ and}$$

$$T_k^{a \to b} := \inf\{n > \tau_{k-1}^{a,b} : X_n^a = b\} \text{ for all } k \geq 2.$$

i) Show that the increasing sequence $(T_n^{a \to b})_{n \geq 1}$ is a renewal process.

ii) Assume that the chain is positive recurrent and that the initial point is chosen according to the invariant measure $\mu$. Let $m < n$ be arbitrary elements of $\mathbb{N}$. Show that

$$\mathbb{E}(\mathbf{N}[m, n]) = (n - m + 1)\mu(b).$$

**Exercise 5.3** Let $\nu$ be a probability distribution on $\{1, 2, \cdots\}$. Show that

$$\theta = \sum_{n\geq 1} n\nu(n) = \sum_{x\geq 0} \nu((x, +\infty)).$$

which proves the identity (5.16).

**Exercise 5.4** Compute the exact form of the law $G^\nu$, defined by (5.14) in the following cases.

i) The distribution $\nu$ is degenerated and gives weight one to the point $a$, that is,

$$\nu(x) = \begin{cases} 1 & \text{if } x = a; \\ 0 & \text{if } x \neq a. \end{cases}$$

ii) For all $p \in (0, 1)$, $\nu$ is the geometric distribution in $\{1, 2, \cdots\}$ with mean $\frac{1}{1-p}$, that is,

$$\nu(n) = p^{n-1}(1-p), \text{ for all } n \geq 1.$$

**Exercise 5.5** Let $\nu$ be a probability distribution on $\{1, 2, \cdots\}$ with finite mean $\theta$, that is:

$$\theta = \sum_{n\geq 1} n\nu(n) < +\infty.$$

Let $G^\nu$ and $Q_\nu$ be the probability measure and the transition matrix defined by (5.14), (5.11) and (5.12).

i) Show that the matrix $Q_\nu$ is irreducible in the set

$$G_\nu = \{0, \sup\{x \geq 0 : \nu(x+1) > 0\}\}.$$

ii) Give examples of sufficient conditions for the aperiodicity of $Q_\nu$.

iii) Show that $G^\nu$ is invariant with respect to $Q_\nu$.

**Exercise 5.6** Under the conditions of Corollary 5.21 determine the law of the counting measure $\mathbf{N}[t, t+1]$, where $t$ is an arbitrary natural number.

**Exercise 5.7 (The inspection paradox)** Using Corollary 5.36 compute the law of the length of the interval containing the instant $t$, as $t \to \infty$. Show that the expectation of this length is $2\theta$.

## 5.8 Comments and References

The Key Theorem is a famous result of Blackwell (1953). Its proof using Markov chains is part of the probabilistic folklore but we do not know a precise reference. The Key Theorem with exponential decay when the failure rate is uniformly bounded below is clearly not optimal but of simple proof. Lindvall (1992) presents alternative proofs for the Key Theorem in the general case using coupling. In particular Lindvall proves that if $\nu$ has exponential decay, then the Key Theorem holds with exponentially fast convergence (as in our Theorem 5.39).

# Chapter 6

# Chains with complete connections

In this chapter we explain how to use the ideas developed in previous chapters to treat the case of non Markov measures. We start with the notion of *specification*, borrowed from Statistical Mechanics.

## 6.1 Specifications

The state space is finite and denoted by $\mathcal{X}$. Let $\mathbb{N}^* = \{1, 2, \ldots\}$ and $-\mathbb{N}^* = \{-1, -2, \ldots\}$ be the sets of positive, respectively negative integers. Instead of transition matrices we work with probability transition functions $P : \mathcal{X} \times \mathcal{X}^{-\mathbb{N}^*} \to [0, 1]$.

**Definition 6.1** We say that a function $P : \mathcal{X} \times \mathcal{X}^{-\mathbb{N}^*} \to [0, 1]$ is a *specification* if it satisfies the following properties:

$$P(a|\underline{w}) \geq 0 \text{ for all } a \in \mathcal{X} \tag{6.2}$$

and

$$\sum_{a \in \mathcal{X}} P(a|\underline{w}) = 1\,, \tag{6.3}$$

for each $\underline{w} \in \mathcal{X}^{-\mathbb{N}^*}$.

We start with an existence result analogous to Proposition 1.19 of Chapter 1.

**Proposition 6.4** *Given an arbitrary "past" configuration* $\underline{w} \in \mathcal{X}^{-\mathbb{N}^*}$ *it is possible to construct a stochastic process* $(X_t,\, t \geq 0)$ *on* $\mathcal{X}^{\mathbb{N}}$ *with the property that for any* $n \geq 0$ *and arbitrary values* $x_0, \ldots, x_n \in \mathcal{X}$,

$$\mathbb{P}(X_t = x_t\,,\, t \in [0,n] \,|\, X_i = w_i\,,\, i \in -\mathbb{N}^*) \;=\; \prod_{t\in[0,n]} P(x_t \,|\, x_{t-1}, \ldots, x_0, \underline{w}) \tag{6.5}$$

*where* $(x_{t-1}, \ldots, x_0, \underline{w}) = (x_{t-1}, \ldots, x_0, w_{-1}, w_{-2}, \ldots)$.

**Proof.** First construct a family of partitions of the interval $[0,1]$

$$((\mathbf{B}(y|\underline{w}) \,:\, y \in \mathcal{X}) \,:\, \underline{w} \in \mathcal{X}^{-\mathbb{N}^*}) \tag{6.6}$$

satisfying

$$|\mathbf{B}(y|\underline{w})| = P(y|\underline{w})\,; \quad \cup_y \mathbf{B}(y|\underline{w}) \;=\; [0,1] \tag{6.7}$$

for all $\underline{w} \in \mathcal{X}^{-\mathbb{N}^*}$. This is always possible because of property (6.3) of $P$. Then proceed as in the proof of Proposition 1.19: construct the function

$$F(\underline{w}; u) \;:=\; \sum_y y\, \mathbf{1}\{u \in \mathbf{B}(y|\underline{w})\} \tag{6.8}$$

and define

$$X_t = F(X_{t-1}, \ldots, X_0, \underline{w}; U_t) \tag{6.9}$$

where $(U_n \,:\, n \in \mathbb{Z})$ is a family of *iid* random variables uniformly distributed in $[0,1]$. $\square$

Our second goal is to give sufficient conditions for the existence of the limits

$$\lim_{t\to\infty} \mathbb{P}(X_{t+k} = x_k,\, k = 1, \ldots, n \,|\, X_{-1} = w_{-1}, X_{-2} = w_{-2}, \ldots) \tag{6.10}$$

for any $n$ and $x_1, \ldots, x_k \in \mathcal{X}$ and the independence of the limit on the "left boundary condition" $\underline{w}$. This question is analogous to the one answered by

Theorem 3.38. Before stating this result we introduce some notation and propose —in the next section— a construction of a process $X_k$ satisfying the specification $P$.

For $k \in \mathbb{N}$, $y \in \mathcal{X}$ and $\underline{w} \in \mathcal{X}^{-\mathbb{N}^*}$ define

$$a_k(y|\underline{w}) := \inf\{P(y|w_{-1},\ldots,w_{-k},\underline{z}) \ : \ \underline{z} \in \mathcal{X}^{-\mathbb{N}^*}\}, \tag{6.11}$$

where $(w_{-1},\ldots,w_{-k},\underline{z}) = (w_{-1},\ldots,w_{-k},z_{-1},z_{-2},\ldots)$. Notice that $a_0(y|\underline{w})$ does not depend on $\underline{w}$. For $k \in \mathbb{N}$ define

$$a_k := \min_{\underline{w}} \Big(\sum_{y\in\mathcal{X}} a_k(y|\underline{w})\Big) . \tag{6.12}$$

Define

$$\beta_m := \prod_{k=0}^{m} a_k \tag{6.13}$$

**Definition 6.14** A specification $P$ is called of *complete connections* if $a_0 > 0$.

**Definition 6.15** We say that a measure $\underline{\nu}$ on $\mathcal{X}^{\mathbb{Z}}$ is *compatible* with a specification $P$ if the one-sided conditional probabilities of $\underline{\nu}$ are given by

$$\underline{\nu}(\underline{X} \in \mathcal{X}^{\mathbb{Z}} \ : \ X_n = y \mid X_{n+j} \ = \ w_j,\, j \in -\mathbb{N}^*) = P(y|\underline{w}) \tag{6.16}$$

for all $n \in \mathbb{Z}$, $y \in \mathcal{X}$ and $\underline{w} \in \mathcal{X}^{-\mathbb{N}^*}$.

## 6.2 A construction

In this chapter we will show that if $\sum_{m\geq 0} \beta_m = \infty$, then the measure $\nu$ on $\mathcal{X}^{\mathbb{Z}}$ defined by 6.35 below is the unique measure compatible with $P$.

For $y \in \mathcal{X}$ let $b_0(y|\underline{w}) := a_0(y|\underline{w})$, and for $k \geq 1$,

$$b_k(y|\underline{w}) := a_k(y|\underline{w}) - a_{k-1}(y|\underline{w}).$$

For each $\underline{w} \in \mathcal{X}^{-\mathbb{N}^*}$ let $\{\mathbf{B}_k(y|\underline{w}) : y \in \mathcal{X}, k \in \mathbb{N}\}$ be a partition of $[0,1]$ with the following properties: (i) for $y \in \mathcal{X}$, $k \geq 0$, $\mathbf{B}_k(y|\underline{w})$ is an interval closed in the left extreme and open in the right one of Lebesgue measure $b_k(y|\underline{w})$; (ii) these intervals are disposed in increasing lexicographic order with respect to $y$ and $k$ in such a way that the left extreme of one interval coincides with the right extreme of the precedent:

$$\mathbf{B}_0(1|\underline{w}), \ldots, \mathbf{B}_0(q|\underline{w}), \mathbf{B}_1(1|\underline{w}), \ldots, \mathbf{B}_1(q|\underline{w}), \ldots$$

with no intersections. More precisely, calling $\mathrm{left}(A) = \inf\{x : x \in A\}$ and $\mathrm{right}(A) = \sup\{x : x \in A\}$, the above construction is required to satisfy

1. $\mathrm{left}[\mathbf{B}_0(1|\underline{w})] = 0\,;$
2. $\mathrm{right}[\mathbf{B}_k(y|\underline{w})] = \mathrm{left}[\mathbf{B}_k(y+1|\underline{w})]$, for $1 \leq y < q$
3. $\mathrm{right}[\mathbf{B}_k(q|\underline{w})] = \mathrm{left}[\mathbf{B}_{k+1}(1|\underline{w})]$, for $k \geq 0$

Define

$$\mathbf{B}(y|\underline{w}) \;:=\; \bigcup_{k\geq 0} \mathbf{B}_k(y|\underline{w}) \tag{6.17}$$

The above properties imply

$$\mathrm{right}[\mathbf{B}_k(q|\underline{w})] = \sum_y a_k(y,\underline{w}) \quad \text{and} \quad \lim_{k\to\infty} \mathrm{right}[\mathbf{B}_k(q|\underline{w})] = 1\,, \tag{6.18}$$

$$|\mathbf{B}(y|\underline{w})| \;=\; \Big|\bigcup_{k\geq 0} \mathbf{B}_k(y|\underline{w})\Big| \;=\; \sum_{k\geq 0} |\mathbf{B}_k(y|\underline{w})| \;=\; P(y|\underline{w}) \tag{6.19}$$

and

$$\Big|\bigcup_{y\in\mathcal{X}} \mathbf{B}(y|\underline{w})\Big| \;=\; \Big|\bigcup_{y\in\mathcal{X}} \bigcup_{k\geq 0} \mathbf{B}_k(y|\underline{w})\Big| \;=\; \sum_{y\in\mathcal{X}} \sum_{k\geq 0} |\mathbf{B}_k(y|\underline{w})| \;=\; 1. \tag{6.20}$$

All the unions above are disjoint. For $\ell \geq 0$ let

$$\mathbf{B}_\ell(\underline{w}) := \bigcup_{y\in\mathcal{X}} \mathbf{B}_\ell(y|\underline{w}).$$

Notice that neither $\mathbf{B}_0(y|\underline{w})$ nor $\mathbf{B}_0(\underline{w})$ depend on $\underline{w}$.

For any left boundary condition $\underline{w} \in \mathcal{X}^{-\mathbb{N}^*}$ define

$$F(\underline{w}; u) := \sum_y y \, \mathbf{1}\{u \in \mathbf{B}(y|\underline{w})\}, \tag{6.21}$$

Let $\underline{U} = (U_i : i \in \mathbb{Z})$ be a sequence of independent random variables with uniform distribution in $[0, 1]$. Define $X_0^{\underline{w}} := F(\underline{w}; U_0)$ and for $t \geq 1$,

$$X_t^{\underline{w}} := F(X_{t-1}^{\underline{w}}, \ldots, X_0^{\underline{w}}, \underline{w}; U_t) \tag{6.22}$$

**Lemma 6.23** *The process $X_t^{\underline{w}}$ defined by* (6.22) *has law* (6.5)*, that is, a distribution compatible with the specification $P$ and with left boundary condition $\underline{w}$.*

**Proof.** It is immediate. Just verify that the length of the intervals is the correct one. This is guaranteed by (6.19). □

**Remark 6.24** The construction of the intervals $\mathbf{B}(y|\underline{w})$ is so complicated for future applications. Any construction satisfying (6.19) would have the properties stated by Lemma (6.23).

We introduce extra notation which will be useful in the sequel. For $\ell \geq 0$ let

$$\mathbf{B}_\ell(\underline{w}) := \bigcup_{y \in \mathcal{X}} \mathbf{B}_\ell(y|\underline{w}). \tag{6.25}$$

Notice that neither $\mathbf{B}_0(y|\underline{w})$ nor $\mathbf{B}_0(\underline{w})$ depend on $\underline{w}$. Recalling the definition (6.13), we have

$$[0, a_k] \subset \bigcup_{\ell=0}^{k} \mathbf{B}_\ell(\underline{w}), \qquad \text{for all } \underline{w} \in \mathcal{X}^{-\mathbb{N}^*}. \tag{6.26}$$

Display (6.26) implies that in the computation of (6.21), if the event $\{U_n \leq a_k\}$ holds, then we only need to look at $x_{n-1}, \ldots, x_{n-k}$ to decide the value of

$x_n$. More precisely, it follows from (6.26) that for *any* $\underline{w}$, $\underline{v} \in \mathcal{X}^{\mathbb{Z}}$ such that $v_j = w_j$ for $j \in [-k, -1]$,

$$[0, a_k] \cap \mathbf{B}_\ell(\underline{w}) \;=\; [0, a_k] \cap \mathbf{B}_\ell(\underline{v})\,. \tag{6.27}$$

From this we have

$$u \le a_k \quad \text{implies} \quad F(x_{-1}, \ldots, x_{-k}, \underline{w}; u) = F(x_{-1}, \ldots, x_{-k}, \underline{v}; u) \tag{6.28}$$

for all $x_{-1}, \ldots, x_{-k} \in \mathcal{X}$, $\underline{w}$, $\underline{v} \in \mathcal{X}^{-\mathbb{N}^*}$, $u \in [0, 1]$.

## 6.3 Loss of memory

We show in this section that the construction of Section 6.2 gives a loss of memory result analogous to the one of Theorem 3.18.

For any $\underline{w}, \underline{v} \in \mathcal{X}^{-\mathbb{N}^*}$ let

$$\tau^{\underline{w},\underline{v}} \;:=\; \inf\{n \ge 0 \,:\, X_k^{\underline{w}} = X_k^{\underline{v}}\,, \text{ for all } k \ge n\} \tag{6.29}$$

Of course $\tau^{\underline{w},\underline{v}}$ could in principle be infinite. The next proposition shows that under the condition $\prod_k a_k > 0$, this time is almost surely finite.

**Proposition 6.30** *If* $\prod_k a_k > 0$, *then for any* $\underline{w}, \underline{v} \in \mathcal{X}^{-\mathbb{N}^*}$

$$\sum_n \mathbb{P}(\tau^{\underline{w},\underline{v}} = n) \;=\; 1 \tag{6.31}$$

**Proof.** (6.28) implies that

$$\tau^{\underline{w},\underline{v}} \;\le\; \min\{n \ge 0 \,:\, U_{n+k} \le a_k \text{ for all } k \ge 0\} \tag{6.32}$$

In other words, $\tau^{\underline{w},\underline{v}}$ is dominated stochastically by the *last* return time to the origin of the house-of-cards process with transitions

$$Q(k, k+1) = a_k \;\;;\;\; Q(x, 0) = 1 - a_k \tag{6.33}$$

and initial state 0. By Lemma 3.70, the condition $\prod_k a_k > 0$ is equivalent to the transience of the house-of-cards chain. Hence the chain may visit the origin only a finite number of times. This implies that the last return time to the origin is finite with probability one and so is $\tau^{\underline{w},\underline{v}}$. □

**Theorem 6.34** *If* $\sum_{m\geq 0}\beta_m = \infty$, *then for any* $\underline{w} \in \mathcal{X}^{-\mathbb{N}^*}$ *the following limits exist*

$$\lim_{t\to\infty} \mathbb{P}(X_{t+k} = x_k,\, k = 1,\ldots,n \,|\, X_{-1} = w_{-1}, X_{-2} = w_{-2},\ldots) \tag{6.35}$$

This theorem will be proven right after Theorem 6.58 below. The existence of the limit is a more delicate matter here than in the case of a Markov chain with finite state space. We cannot a priori guarantee that this limit exist. We show the existence by explicitly constructing a measure and then proving that it is time translation invariant. This construction —performed in the next section— is a particular case of the so called *thermodynamic limit* of Statistical Mechanics.

In the meanwhile we can prove that these limits, if exist, are independent of the left boundary condition. This is the contents of the next theorem.

**Theorem 6.36** *Assume* $\sum_{m\geq 0}\beta_m = \infty$ *and that the limits* (6.35) *exist. Then they are independent of* $\underline{w}$.

**Proof.** By (6.28),

$$\begin{aligned} &|\mathbb{P}(X_{t+k} = x_k,\, k = 1,\ldots,n \,|\, X_{-1} = w_{-1}, X_{-2} = w_{-2},\ldots) && (6.37)\\ &\qquad - \mathbb{P}(X_{t+k} = x_k,\, k = 1,\ldots,n \,|\, X_{-1} = v_{-1}, X_{-2} = v_{-2},\ldots)| \\ &\leq \quad \mathbb{P}(\tau^{\underline{w},\underline{v}} > t) && (6.38)\end{aligned}$$

which goes to zero by Proposition 6.30. □

## 6.4 Thermodynamic limit

We consider now a slightly different problem. Instead of fixing the left condition to the left of time zero, we fix the condition to the left of time $-n$ and compute the limiting distribution when $n \to \infty$. We give conditions which guarantee the existence of the limit and its independence of the boundary conditions. This is the so called (one sided) *thermodynamic limit*. Under those conditions we show that it suffices to look at a finite random number of uniform random variables in order to construct the thermodynamic limiting measure in any finite time interval.

Let $s \le t \le \infty$. For $-\infty < s < \infty$ and $s \le t \le \infty$, define

$$\tau[s,t] := \max\{m \le s \,:\, U_k \le a_{k-m} \text{ for all } k \in [m,t]\} \tag{6.39}$$

which may be $-\infty$. We use the notation $\tau[n] := \tau[n,n]$. Notice that for fixed $s$, $\tau[s,t]$ is non increasing in $t$:

$$t \le t' \quad \text{implies} \quad \tau[s,\infty] \;\le\; \tau[s,t'] \;\le\; \tau[s,t] \tag{6.40}$$

and non decreasing in $s$ in the following sense:

$$[s',t'] \subset [\tau[s,t],t] \quad \text{implies} \quad \tau[s',t'] \;\ge\; \tau[s,t]\,. \tag{6.41}$$

Notice that $\tau[s,t]$ is a *left stopping time* for $\underline{U}$ with respect to $[s,t]$, in the sense that

$$\{\tau[s,t] \le j\} \text{ depends only on the values of } (U_i : i \in [j,t]) \tag{6.42}$$

for $j \le s$.

**Lemma 6.43** *If* $\sum_{n\ge 0}\prod_{k=1}^{n} a_k \;=\; \infty$*, then for each* $-\infty < s \le t < \infty$*,* $\tau[s,t]$ *is a "honest" random variable:* $\sum_i \mathbb{P}(\tau[s,t] = i) = 1$*. If* $\prod_{k=1}^{\infty} a_k \;>\; 0$*, then for each* $-\infty < s$*,* $\tau[s,\infty]$ *is a "honest" random variable:* $\sum_i \mathbb{P}(\tau[s,\infty] = i) = 1$*.*

**Proof.** By the definition of $\tau$:

$$\tau[s,t] = \max\{m \le s : W_n^{0;m} > 0,\, \forall\, n \in [s,t]\} \tag{6.44}$$

where $W_n^{0;m}$ is the state of the house-of-cards process with transition matrix (3.69) starting at time $m$ in state 0. Hence, for $m \le s$,

$$\{\tau[s,t] < m\} \subset \bigcup_{i\in[s,t]} \{W_i^{0;m} = 0\} \tag{6.45}$$

By translation invariance, the probability of the rhs of (6.45) is

$$\mathbb{P}\Big(\bigcup_{i\in[s,t]}\{W^{0;0}_{-m+i}=0\}\Big) \;\le\; \sum_{i=1}^{t-s}\mathbb{P}(W^{0;0}_{s-m+i}=0) \tag{6.46}$$

Since by hypothesis $W^{0;0}_n$ is a non positive-recurrent chain, by Lemma 3.70, each of the $t-s$ terms in the rhs of (6.46) goes to zero as $m\to-\infty$. This shows the first part of the lemma. For the second part bound $\mathbb{P}(\tau[s,\infty]<m)$ by

$$\mathbb{P}\Big(\bigcup_{i\in[s-m,\infty]}\{W^{0;0}_i=0\}\Big) \tag{6.47}$$

which, by transience, goes to zero as $m\to-\infty$. □

Property (6.27) allows us to introduce the following definition

$$\mathbf{B}_{\ell,j}(w_{-1},\dots,w_{-j}) \;:=\; [0,a_j[\cap\mathbf{B}_\ell(\underline{w})\,, \tag{6.48}$$

$$\mathbf{B}_{\ell,j}(y|w_{-1},\dots,w_{-j}) \;:=\; [0,a_j[\cap\mathbf{B}_\ell(y|\underline{w}) \tag{6.49}$$

**Definition 6.50** Assume $\mathbb{P}(\tau[n]>-\infty)=1$ for all $n\in\mathbb{Z}$. Let

$$Y_n \;:=\; \sum_{y\in\mathcal{X}} y\,\mathbf{1}\Big\{U_n\in\bigcup_{k\ge0}\mathbf{B}_{k,\tau[n]}(y|Y_{n-1},\dots,Y_{\tau[n]})\Big\}. \tag{6.51}$$

This is well defined because by (6.41) $\tau[n]\le\tau[j]$ for $j\in[\tau[n],n]$ and (6.48). This allows to construct $Y_j$ for all $j\in[\tau[n],n]$ for all $n$, and in particular to construct $Y_n$.

**Theorem 6.52** *If* $\sum_{m\ge0}\beta_m=\infty$*, then the law of the vector* $(Y_j\,:\,j\in\mathbb{Z})$ *defined by* (6.51) *is the unique measure compatible with* $P$*. Furthermore the following* thermodynamic limits *exist: for any* $n\ge1$ *and arbitrary* $s\le t\in\mathbb{Z}$*,* $x_s,\dots,x_t\in\mathcal{X}$*,*

$$\lim_{t\to\infty}\mathbb{P}(X_k=x_k,\,k\in[s,t]\,|\,X_{-n}=w_{-n},X_{-n-1}=w_{-n-1},\dots) \\ =\;\mathbb{P}(Y_k=x_k,\,k\in[s,t]) \tag{6.53}$$

*for all* $\underline{w}$.

**Proof.** The fact that the law of $\underline{Y}$ is compatible with $P$ follows from (6.19).

For each $\underline{w} \in \mathcal{X}^{\mathbb{Z}}$ and $i \in \mathbb{Z}$ let $\underline{X}^{\underline{w},i}$ be defined as follows. Set $X_j^{\underline{w},i} = w_j$ for $j < i$ and for $n \geq i$,

$$X_n^{\underline{w},i} := \sum_{y \in \mathcal{X}} y\, \mathbf{1}\Big\{U_n \in \mathbf{B}(y|X_{n-1}^{\underline{w},i}, \ldots, X_i^{\underline{w},i}, w_{i-1}, w_{i-2} \ldots)\Big\}. \tag{6.54}$$

Then by (6.25) and (6.19),

$$\begin{aligned}
&\mathbb{P}(X_k^{\underline{w},i} = x_k,\, k \in [s,t]) \\
&\quad = \mathbb{P}(X_k = x_k,\, k \in [s,t] \,|\, X_{-n} = w_{-n}, X_{-n-1} = w_{-n-1}, \ldots)\,.
\end{aligned}$$

Hence

$$\begin{aligned}
|\mathbb{P}(X_k = x_k,\, k \in [s,t] \,&|\, X_{-i} = w_{-i}, X_{-i-1} = w_{-i-1}, \ldots) \\
&\qquad - \mathbb{P}(Y_k = x_k,\, k \in [s,t])| \\
&\leq |\mathbb{P}(X_k^{\underline{w},i} = x_k,\, k \in [s,t]) \,-\, \mathbb{P}(Y_k = x_k,\, k \in [s,t])]| \\
&= |\mathbb{E}(\mathbf{1}\{X_k^{\underline{w},i} = x_k,\, k \in [s,t]\} \,-\, \mathbf{1}\{Y_k = x_k,\, k \in [s,t]\}| \\
&\leq \mathbb{P}(\tau[s,t] < i)
\end{aligned} \tag{6.55}$$

which goes to zero as $i \to -\infty$ by hypothesis and Lemma 6.43. This shows that the thermodynamic limit converges to the law of $\underline{Y}$. □

**Definition 6.56** Let $\underline{\nu}$ be the law of the sequence $\underline{Y}$:

$$\underline{\nu}(\underline{X} \in \mathcal{X}^{\mathbb{Z}} \,:\, X_k = x_k\,,\, k = 1, \ldots, n) := \mathbb{P}(Y_k = x_k,\, k = 1, \ldots, n) \tag{6.57}$$

**Theorem 6.58** *If $\sum_{m \geq 0} \beta_m = \infty$, then $\underline{\nu}$ is the unique measure compatible with $P$.*

**Proof.** Call $\underline{\mu}(\cdot|\underline{w}, i)$ the law on $\mathcal{X}^{[i,\infty)}$ of the random sequence $\underline{X}^{\underline{w},i}$. By (6.19), $\underline{\mu}(\cdot|\underline{w}, i)$ is compatible with $P$. The measure $\underline{\mu}(\cdot|\underline{w}, i)$ is the unique measure compatible with $P$ on $\mathcal{X}^{[i,\infty)}$ with boundary conditions $\{X_j = w_j\,:$

$j < i\}$. Assume $\underline{\mu}$ and $\underline{\mu}'$ be two measures on $\mathcal{X}^{\mathbb{Z}}$ compatible with $P$. Let $f$ be a function depending on $X_j$ for $j$ in the interval $[s,t]$.

$$\begin{aligned} &|\underline{\mu} f - \underline{\mu}' f| \\ &= \left| \int \underline{\mu}(d\underline{w}) \underline{\mu}(f \,|\, X_j = w_j,\, j < i) \;-\; \int \underline{\mu}'(d\underline{w}') \underline{\mu}'(f \,|\, X_j = w'_j,\, j < i) \right| \\ &= \left| \int \underline{\mu}(d\underline{w}) \mathbb{E}[f(\underline{X}^{\underline{w},i})] \;-\; \int \underline{\mu}'(d\underline{w}') \mathbb{E}[f(\underline{X}^{\underline{w}',i})] \right| \\ &\le \mathbb{E} \int \underline{\mu}(d\underline{w})\, \underline{\mu}'(d\underline{w}') \left| f(\underline{X}^{\underline{w},i}) \;-\; f(\underline{X}^{\underline{w}',i}) \right| \end{aligned} \tag{6.59}$$

Since

$$X^{\underline{w},i}(n) \neq X^{\underline{w}',i}(n) \;\le\; \mathbf{1}\{\tau[n] \le i\} \tag{6.60}$$

the last term in (6.59) is bounded above by

$$2\, \|f\|_\infty \, \mathbb{P}(\tau[s,t] < i) \tag{6.61}$$

which goes to zero as $i \to -\infty$ by hypothesis and Lemma 6.43. □

**Proof of Theorem 6.34** It suffices to notice that

$$\begin{aligned} &\mathbb{P}(X_{t+k} = x_k,\, k = 1, \dots, n \,|\, X_{-1} = w_{-1}, X_{-2} = w_{-2}, \dots) \\ &\qquad = \;\mathbb{P}(X_k = x_k,\, k = 1, \dots, n \,|\, X_{t-1} = w_{t-1}, X_{t-2} = w_{t-2}, \dots) \end{aligned}$$

which converges to $\underline{\nu}(\underline{X} \in \mathcal{X}^{\mathbb{Z}} \,:\, X_k = x_k,\, k = 1, \dots, n)$ by Theorem 6.52. □

## 6.5 Bounds on the rate of convergence

We state without proof the following bound for the law of $\tau$:

$$\mathbb{P}(s - \tau[s,t] > m) \;\le\; \sum_{i=1}^{t-s} \rho_{m+i} \tag{6.62}$$

where $\rho_m$ is the probability of return to the origin at epoch $m$ of the Markov chain on $\mathbb{N}$ starting at time zero at the origin with transition probabilities

$p(x, x+1) = a_x$, $p(x,0) = (1 - a_x)$, $p(x,y) = 0$ otherwise. Furthermore, if $(1 - a_k)$ decreases exponentially fast with $k$, then so does $\rho_k$. If $(1 - a_k)$ decreases as a summable power, then $\rho_k$ decreases with the same power. These results can be found in Comets, Fernández and Ferrari (2000).

As a consequence we get the following bounds on the loss of memory:

**Theorem 6.63** *The following bounds hold for the rate of convergence in Theorem 6.52.*

$$\begin{aligned} |\mathbb{P}(X_k = x_k,\, k \in [s,t] \,|\, X_{-n} = w_{-n}, X_{-n-1} = w_{-n-1}, \ldots) \\ - \mathbb{P}(Y_k = x_k,\, k \in [s,t])| \\ \leq \sum_{i=1}^{t-s} \rho_{n+i} \end{aligned} \tag{6.64}$$

## 6.6 Regeneration

Let $\mathbf{N} \in \{0,1\}^{\mathbb{Z}}$ be the random counting measure defined by

$$\mathbf{N}(j) := \mathbf{1}\{\tau[j,\infty] = j\}\,, \tag{6.65}$$

Let $(T_\ell : \ell \in \mathbb{Z})$ be the ordered time events of $\mathbf{N}$ defined by $\mathbf{N}(i) = 1$ if and only if $i = T_\ell$ for some $\ell$, $T_\ell < T_{\ell+1}$ and $T_0 < 0 \leq T_1$.

**Theorem 6.66** *If $\beta > 0$, then the process $\mathbf{N}$ defined in (6.65) is a stationary renewal process. Furthermore, the random vectors $\xi_\ell \in \cup_{n\geq 1} \mathcal{X}^n$, $\ell \in \mathbb{Z}$, defined by*

$$\xi_\ell \;:=\; (Y_{T_\ell}, \ldots, Y_{T_{\ell+1}-1}) \tag{6.67}$$

*are mutually independent and $(\xi_\ell : \ell \neq 0)$ are identically distributed.*

The sequence $(T_\ell)$ correspond to the *regeneration times*: at times $T_\ell$ the process does not depend on the past.

**Proof.** Stationarity follows immediately from the construction. The density of $\mathbf{N}$ is positive:

$$\mathbb{P}(\mathbf{N}(j) = 1) = \mathbb{P}(\cap_{\ell \geq j}\{U_\ell < a_{\ell-j}\}) = \beta > 0 \tag{6.68}$$

by hypothesis. Let

$$f(j) := \mathbb{P}(\mathbf{N}(-j) = 1 \,|\, \mathbf{N}(0) = 1) \tag{6.69}$$

for $j \in \mathbb{N}^*$. To see that $\mathbf{N}$ is a renewal process it is sufficient to show that

$$\mathbb{P}(\mathbf{N}(s_\ell) = 1\,;\, \ell = 1, \dots, n) \;=\; \beta \prod_{\ell=1}^{n-1} f(s_{\ell+1} - s_\ell) \tag{6.70}$$

for arbitrary integers $s_1 < \dots < s_k$. For $j \in \mathbb{Z}$, $j' \in \mathbb{Z} \cup \{\infty\}$, define

$$H[j, j'] \;:=\; \begin{cases} \{U_{j+\ell} < a_{\ell-1}, \ell = 0, \dots, j' - j\}, & \text{if } j \leq j' \\ \\ \text{“full event”}, & \text{if } j > j' \end{cases} \tag{6.71}$$

With this notation,

$$\mathbf{N}(j) = \mathbf{1}\{H[j, \infty]\}, \quad j \in \mathbb{Z}. \tag{6.72}$$

and

$$\mathbb{P}(\mathbf{N}(s_\ell) = 1\,;\, \ell = 1, \dots, n) \;=\; \mathbb{P}\Big(\bigcap_{i=1}^{n} H[s_\ell, \infty]\Big) \tag{6.73}$$

From monotonicity we have for $j < j' < j'' \leq \infty$,

$$H[j, j''] \cap H[j', j''] \;=\; H[j, j' - 1] \cap H[j', j'']\,. \tag{6.74}$$

Then (6.73) equals

$$\prod_{i=1}^{n} \mathbb{P}(H[s_\ell, s_{\ell+1} - 1]), \tag{6.75}$$

where $s_{n+1} := \infty$. Since $\mathbb{P}(H[s_n, \infty]) = \beta$, (6.75) equals the right hand side of (6.70), implying that $\mathbf{N}$ is a renewal process. From stationarity we have

$$\begin{aligned} \mathbb{P}(T_{\ell+1} - T_\ell \geq m) \;&=\; \mathbb{P}(\tau[-1, \infty] < -m + 1 \mid \tau[0, \infty] = 0) \\ &=\; \mathbb{P}(W_{-1}^{-m+1} = 0) \\ &=\; \rho_m \;\;. \end{aligned}$$

The last statement (6.67) is clear by construction of $\underline{Y}$.

## 6.7 Perfect simulation

We propose another construction of $\tau[s,t]$ which is more convenient for perfect simulation algorithms. Let $a_{-1} = 0$ and define for $n \in \mathbb{Z}$,

$$K_n := \sum_{k\geq 0} k \,\mathbf{1}\{U_n \in [a_{k-1}, a_k)\}\,, \tag{6.76}$$

the number of sites backwards that we need to know (at most) to compute $Y_n$ using formula (6.51), see (6.42). For $s < \infty$ and $s \leq t \leq \infty$, define

$$Z[s,t] := \max\{K_n \,:\, n \in [s,t]\}. \tag{6.77}$$

Let $\theta_{-1} := t+1$, $\theta_0 := s$ and for $n \geq 1$, inductively

$$\theta_n \;:=\; \theta_{n-1} - Z[\theta_{n-1}, \theta_{n-2} - 1] \tag{6.78}$$

Then it is easy to see that

$$\tau[s,t] \;=\; \lim_{n\to\infty} \theta_n \;=\; \max\{\theta_n \,:\, \theta_n = \theta_{n+1}\} \;\text{ a.s..} \tag{6.79}$$

The construction (6.51) and (6.79) can be translated into the following *perfect simulation algorithm for* $\underline{\nu}$. Assume $s \leq t < \infty$.

**Algorithm 6.80 (Simulation of the stationary measure)** Perform

1. Set $\theta_{-1} = t+1$, $\theta_0 = s$ and iterate the following step up to the first moment $\theta_n = \theta_{n-1}$:
2. Generate $U_{\theta_\ell}, \ldots, U_{\theta_{\ell-1}-1}$. Use (6.76) to compute $K_{\theta_\ell}, \ldots, K_{\theta_{\ell-1}}$ and (6.77) and (6.78) to compute $\theta_{\ell+1}$.
3. Let $\tau = \theta_n$
4. For $k = \tau$ to $k = t$ define $Y_k$ using (6.51).
5. Print $(Y_j \,:\, j \in [s,t])$. End.

**Theorem 6.81** *Let* $(Y_j : j \in [s,t])$ *be the output of the above algorithm. Then*

$$\mathbb{P}(Y_j = x_j : j \in [s,t]) = \underline{\nu}(X_j = x_j : j \in [s,t]) \tag{6.82}$$

*for arbitrary* $x_s, \ldots, x_t \in \mathcal{X}$.

**Proof.** Follows from the construction. □

## 6.8 Exercises

**Exercise 6.1** Show that (6.70) suffices to characterize a renewal process. Hint: use the "inclusion-exclusion formula" to compute the probability of any set depending on a finite number of coordinates.

**Exercise 6.2 (The noisy voter model)** Let $\varepsilon > 0$ and $(\alpha_i)$ be a probability measure on $-\mathbb{N}^*$. Let $\mathcal{X}$ be a finite alphabet with $N$ elements and define the following specification:

$$P(x|\underline{w}) := (1-\varepsilon) \sum_{i \le -1} \alpha_i \mathbf{1}\{x = w_i\} + \varepsilon/N \tag{6.83}$$

In words, with probability $(1-\varepsilon)$ the origin applies the following "voter rule": with probability $\alpha_i$ choose coordinate $i$ and adopt its value; with probability $\varepsilon$ choose at random a value in $\mathcal{X}$. (i) Compute $a_k$ for this model. (ii) Give conditions on $(\alpha_i)$ guaranteeing that $\prod a_k > 0$. What happens if $\varepsilon = 0$?

**Exercise 6.3 (The noisy majority voter model)** Let $\varepsilon > 0$ and $(\alpha_i)$ be a probability measure on $-\mathbb{N}^*$. Let $\mathcal{X}$ be a finite alphabet with $N$ elements and define the following specification:

$$P(x|\underline{w}) := (1-\varepsilon) \sum_{i \le -1} \alpha_i \, \mathrm{maj}(w_{-1}, \ldots, w_i) + \varepsilon/N \tag{6.84}$$

where $\mathrm{maj}(w_{-1}, \ldots, w_i)$ is the value that appears more times in the vector $(w_{-1}, \ldots, w_i)$ (in case of equality, use any rule you like to decide). In words, with probability $(1-\varepsilon)$ the origin applies the following "majority rule": with probability $\alpha_i$ choose the first $i$ coordinates to the left of the origin and adopt the same value as the majority of them; with probability $\varepsilon$ choose at random a value in $\mathcal{X}$. (i) Compute $a_k$ for this model. (ii) Give conditions on $(\alpha_i)$ guaranteeing that $\prod a_k > 0$. What happens if $\varepsilon = 0$?

**Exercise 6.4 (One sided Potts model)** Let $\alpha_i$ be as in the previous exercises. Let $H(x|\underline{w}) = -\sum_{i\in -\mathbb{N}^*} \alpha_i \mathbf{1}\{x = w_i\}$. Let

$$P(x|\underline{w}) := \exp(-H(x|\underline{w}))/Z(\underline{w}) \tag{6.85}$$

where $Z(\underline{w})$ is the normalization (partition function). (i) Compute $a_k$ for this model. (ii) Give conditions on $(\alpha_i)$ guaranteeing that $\prod a_k > 0$.

**Exercise 6.5 (The contact process)** Let $\varepsilon$ and $\alpha_i$ be as in the previous exercises. Let $\mathcal{X} = \{0, 1\}$ and

$$\begin{aligned} P(1|\underline{w}) &:= (1-\varepsilon)\sum_i \alpha_i \mathbf{1}\Big\{\sum_{j=i}^{-1} w_j \geq 1\Big\} \\ P(0|\underline{w}) &:= 1 - P(1|\underline{w}) \end{aligned} \tag{6.86}$$

Compute $a_k$ for this model and give conditions on $(\alpha_i)$ and $\varepsilon$ guaranteeing that $\prod a_k > 0$.

**Exercise 6.6 (Renewal process)** Let $\mathcal{X} = \{0, 1\}$ and $\nu$ be the inter renewal probability of a stationary renewal process $S_n$. Compute the specifications of the renewal process and the relationship between $\nu$ and $(a_k)$.

## 6.9 Comments and references

This chapter is based in the construction of chains with complete connections proposed by Ferrari *et al.* (2000) and Comets *et al.* (2000). The regeneration for non Markov chains were proposed by Lalley (1986) for chains with exponential continuity rates (in which case the process is a Gibbs state of an exponentially summable interaction). Berbee (1987) proposed a very close approach for chains with a countable alphabet and summable continuity rates. Ney and Nummelin (1993) have extended the regeneration approach to chains for which the continuity rates depend on the history. Comets *et al.* (2000) proposed the perfect simulation algorithm we describe in Section 6.7. The introduction of this paper contains a detailed discussion of the previous literature in the field.

# Chapter 7

# Poisson processes

## 7.1 One dimensional Poisson processes

A one-dimensional point process on $\mathbb{R}$ is an increasing sequence of random variables $\dots S_{-1} \le S_0 \le 0 \le S_1, \dots$ in $\mathbb{R}$. These variables can be interpreted as the successive epochs of occurrence of a given event.

We construct now a particular process $S_n \in \mathbb{R}$, $n \in \mathbb{Z}$, which will be called Poisson Process.

Start by partitioning $\mathbb{R}$ in intervals $A_i$, $i \in \mathbb{Z}$, $A_i = [l_i, l_{i+1})$, with $l_i < l_{i+1}$. Call $|A_i|$ the length $l_{i+1} - l_i$ of the interval $A_i$. In this way $\dot\cup_{i\in\mathbb{Z}} A_i = \mathbb{R}$, where $\dot\cup$ means disjoint union.

The second step is to assign to each interval $A_i$ a random variable $Y_i$ with Poisson distribution of mean $\lambda|A_i|$. That is,

$$\mathbb{P}(Y_i = k) = \frac{e^{-\lambda|A_i|}(\lambda|A_i|)^k}{k!}.$$

Assume that $(Y_i : i \in \mathbb{Z})$ is a family of independent random variables.

To each $i \in \mathbb{Z}$ associate a sequence of *iid* random variables $\{U_{i,j} : j = 1, 2, \dots\}$, with uniform distribution in $A_i$:

$$\mathbb{P}(U_{i,j} \in A \cap A_i) = \frac{|A \cap A_i|}{|A_i|}.$$

Let $\mathbf{S}$ be the random set

$$\mathbf{S} = \bigcup_{i\in\mathbb{Z}} \mathbf{S}_i,$$

where

$$\mathbf{S}_i = \begin{cases} \{U_{i,j} : 1 \le j \le Y_i\}, & \text{if } Y_i \ge 1; \\ \emptyset, & \text{if } Y_i = 0. \end{cases}$$

In other words, put $Y_i$ points in each interval $A_i$, independently and with uniform distribution.

Finally reorder the points of $\mathbf{S}$ to obtain a point process. Let $\{S_n\}$ be the ordered sequence of points of $\mathbf{S}$, where $S_1$ is the first point to the right of the origin. We fix $S_1$ in this way just to be precise; any other convention would be just as good. More formally:

$$S_1 = \min\{s > 0 : s \in \mathbf{S}\}$$

and

$$S_n = \begin{cases} \min\{s > S_{n-1} : s \in \mathbf{S}\}, & \text{if } n \ge 2; \\ \max\{s < S_{n+1} : s \in \mathbf{S}\}, & \text{if } n \le 0. \end{cases} \tag{7.1}$$

For $A \subset \mathbb{R}$, define $\mathbf{N}_\mathbf{S}(A)$ = number of points of the set $\mathbf{S} \cap A$. It is clear that

$$\mathbf{N}_\mathbf{S}(A) = \sum_n \mathbf{1}\{S_n \in A\}. \tag{7.2}$$

When no confusions arise we will write just $\mathbf{N}(A)$ instead of $\mathbf{N}_\mathbf{S}(A)$.

**Definition 7.3** The process defined by (7.2) will be called one dimensional *Poisson process.*

## 7.2 Formal definition of point processes

The Poisson process we have just constructed is a particular case of point process. The renewal processes constructed in Chapter 5 is another example. The formal definition of point process in $\mathbb{R}$ is the following. Consider the set

$$\mathcal{M} = \{\mathbf{S} \subset \mathbb{R} : \mathbf{N}_\mathbf{S}(A) < \infty, \text{ for all interval } A \subset \mathbb{R}\}$$

This is the set of the possible realizations of the point process that do not have points of accumulation. A point process is characterized by the definition of a probability measure on $\mathcal{M}$. Since $\mathcal{M}$ is not countable this involves a non trivial notion of *event* in $\mathcal{M}$ which uses Measure Theory; this is beyond the scope of this notes. We limit ourselves to present the essential events which take the form

$$\{\mathbf{S} \in \mathcal{M} \,:\, \mathbf{N_S}(B_i) = b_i, i = 1, \dots, \ell\} \tag{7.4}$$

for arbitrary $\ell \in \mathbb{N}$, $b_i \in \mathbb{N}$ and finite intervals $B_i$. Let $\mathcal{A}$ be the family of events of the form (7.4).

**Theorem 7.5** *A probability measure on $\mathcal{M}$ is totally determined by the probabilities of the events in $\mathcal{A}$.*

Theorem 7.5 is due to Kolmogorov and gives an operative way of dealing with measures in non countable spaces.

## 7.3 Properties

In this section we discuss the basic properties of the one dimensional Poisson process. A corollary of the following three lemmas is that the way the partition $(A_i)$ is chosen does not influence the law of the process.

**Lemma 7.6** *For each interval $A$, the random variable $\mathbf{N}(A)$ has Poisson law of mean $\lambda|A|$.*

**Proof.** Notice first that since $A \cap A_i$ are disjoint events, the random variables $\mathbf{N}(A \cap A_i)$ are independent. Compute the law of $\mathbf{N}(A \cap A_i)$. By construction,

$$\begin{aligned} \mathbb{P}(\mathbf{N}(A \cap A_i) = k) &= \sum_{h \geq k} \mathbb{P}(Y_i = h, \mathbf{N}(A \cap A_i) = k) && (7.7) \\ &= \sum_{h \geq k} \mathbb{P}\left(Y_i = h, \sum_{j=0}^{h} \mathbf{1}\{U_{i,j} \in A \cap A_i\} = k\right). && (7.8) \end{aligned}$$

Since the variables $U_{i,j}$ are independent of $Y_i$, we can factorize the last probability to obtain

$$= \sum_{h\geq k} \mathbb{P}(Y_i = h)\mathbb{P}\left(\sum_{j=0}^{h} \mathbf{1}\{U_{i,j} \in A\cap A_i\} = k\right). \tag{7.9}$$

But $U_{i,j}$ are *iid* random variables uniformly distributed in $A_i$. Hence, $\mathbf{1}\{U_{i,j} \in A \cap A_i\}$ are *iid* random variables with Bernoulli law of parameter $\frac{|A\cap A_i|}{|A_i|}$. The sum of these variables has Binomial distribution with parameters $n$ and $\frac{|A\cap A_i|}{|A_i|}$. In this way, the last expression equals

$$\sum_{h\geq k} \frac{e^{-\lambda|A_i|}(\lambda|A_i|)^h}{h!}\binom{h}{k}\left(\frac{|A\cap A_i|}{|A_i|}\right)^k\left(1-\frac{|A\cap A_i|}{|A_i|}\right)^{h-k} \tag{7.10}$$

$$= \frac{e^{-\lambda|A\cap A_i|}(\lambda|A\cap A_i|)^k}{k!}, \tag{7.11}$$

a Poisson distribution of mean $\lambda|A\cap A_i|$. To finish the proof observe that $\mathbf{N}(A) = \sum \mathbf{N}(A\cap A_i)$ is a sum of independent random variables with Poisson law with means $\lambda|A\cap A_i|$. Since $A_i$ are disjoint, $\sum |A\cap A_i| = |A|$. In this way, $\mathbf{N}(A)$ has Poisson law with mean $\lambda|A|$. □

**Lemma 7.12** *For each family of disjoint intervals $B_l$, $l = 1,\dots,L$, the random variables $\mathbf{N}(B_l)$ are independent and have Poisson law with mean $\lambda|B_l|$, respectively.*

**Proof.** The proof follows the pattern of the previous lemma. It is easy to verify that for fixed $i$, given $\mathbf{N}(A_i) = h_i$, the random variables

$$\{\mathbf{N}(B_l \cap A_i) : l = 1,\dots,L\}$$

have Multinomial distribution, that is, for arbitrary integers $k_{l,i} \geq 0$, such that $\sum_{l=1}^{L} k_{l,i} = h_i$,

$$\mathbb{P}(\mathbf{N}(B_l\cap A_i) = k_{l,i} \mid \mathbf{N}(A_i) = h_i) = \frac{h_i!}{\prod_{l=1}^{L+1} k_{l,i}!}\prod_{l=1}^{L+1}(b_{l,i})^{k_{l,i}}, \tag{7.13}$$

where

$$k_{L+1,i} = h_i - \sum_{l=1}^{L} k_{l,i}, \quad b_{l,i} = \frac{|B_l \cap A_i|}{|A_i|}$$

for $1 \le l \le L$ and $b_{L+1} = 1 - \sum_{l=1}^{L} b_l$.

Since $\{\mathbf{N}(A_i) : i \in \mathbb{Z}\}$ are independent random variables, it follows from (7.13) that $\{\mathbf{N}(B_l \cap A_i) : l = 1, \ldots, L, i \in \mathbb{Z}\}$ is a family of independent random variables with Poisson law with parameters $\lambda |B_l \cap A_i|$, respectively. To conclude the proof it suffices to sum over $i$ and to use the fact that sum of independent random variables with Poisson law is Poisson. □

**Lemma 7.14** *For any interval $A$, the conditional distribution of the points in $\mathbf{S} \cap A$ given $\mathbf{N}(A) = n$ is the same as the law of $n$ independent random variables uniformly distributed in $A$.*

**Proof.** We use Theorem 7.5. Let $B_1, \ldots, B_L$ be a partition of $A$ and $n_1, \ldots, n_L$ non negative integers such that $n_1 + \ldots + n_L = n$.

$$\mathbb{P}(\mathbf{N}(B_l) = n_l, l = 1, \ldots, L \mid \mathbf{N}(A) = n) \tag{7.15}$$

$$= \frac{\mathbb{P}(\mathbf{N}(B_l) = n_l, l = 1, \ldots, L)}{\mathbb{P}(\mathbf{N}(A) = n)} \tag{7.16}$$

$$= \frac{n!}{n_1! \ldots n_L!} \prod_{l=1}^{L} \left( \frac{|B_l|}{|A|} \right)^{n_l} \tag{7.17}$$

By Theorem 7.5 it is sufficient to show that if $U_1, \ldots, U_L$ are independent random variables uniformly distributed in $A$, and $\mathbf{M}(B) = \sum_i \mathbf{1}\{U_i \in B\}$, then

$$\mathbb{P}(\mathbf{M}(B_l) = n_l, l = 1, \ldots, L) = \frac{n!}{n_1! \ldots n_L!} \prod_{l=1}^{L} \left( \frac{|B_l|}{|A|} \right)^{n_l},$$

which is left as an exercise to the reader. □

**Corollary 7.18** *The conditioned distribution of the vector $(S_1, S_2, \ldots, S_n)$ given $\mathbf{N}([0, t]) = n$ is the same as the law of $(Y_1, Y_2, \ldots, Y_n)$, the order statis-*

*tics of the random variables* $(U_1, U_2, \dots, U_n)$ *uniformly distributed in the interval* $[0, t]$ *defined by:*

$$Y_1 = \min\{U_1, U_2, \dots, U_n\}; \tag{7.19}$$
$$Y_i = \min(\{U_1, U_2, \dots, U_n\} \setminus \{Y_1, \dots, Y_{i-1}\}), i = 2, \dots, n. \tag{7.20}$$

**Remark 7.21** Lemmas 7.12 and 7.14 show that the choice of the sets $A_i$ is not important for the construction of the process.

## 7.4 Markov property

We present an alternative construction of the one-dimensional Poisson process. Let $T_1$ be an exponential random variable with mean $1/\lambda$ and $\widetilde{\mathbf{N}}(\cdot)$ be a Poisson process with rate $\lambda$, independent of $T_1$.

Let $\mathbf{N}(\cdot)$ be the process defined by

$$\mathbf{N}(B) = \mathbf{1}\{T_1 \in B\} + \widetilde{\mathbf{N}}(B - T_1), \tag{7.22}$$

where

$$B - t = \{x \in \mathbb{R} : x - t \in B\}.$$

In other words, the process $\mathbf{N}(\cdot)$ is obtained by first fixing the first event with the random variable $T_1$ and then gluing after this instant an independent Poisson process.

**Theorem 7.23** *The point process defined by* (7.22) *is a Poisson process with rate* $\lambda$.

**Proof.** In view of Theorem 7.5 it suffices to prove that for the process $\mathbf{N}(\cdot)$ constructed above Lemma 7.12 holds. Hence we want to compute

$$\mathbb{P}(\mathbf{N}(B_l) = k_l, l = 1, \dots, L),$$

for arbitrary intervals $B_l$, $k_l \in \mathbb{N}$ and $L \geq 1$. To simplify the presentation of the proof we will consider $L = 1$ and $B_1 = [a, c]$. The extension to any $L$ is left as an exercise. We condition to the value of $T_1$.

$$\begin{aligned}
&\mathbb{P}(\mathbf{N}([a,c]) = k) \\
&\quad = \mathbb{P}(\mathbf{N}([a,c]) = k, T_1 < a) \\
&\qquad + \mathbb{P}(\mathbf{N}([a,c]) = k, T_1 \in [a,c]) + \mathbb{P}(\mathbf{N}([a,c]) = k, T_1 > c). \quad (7.24)
\end{aligned}$$

Assume first $k = 0$. In this case the central term is zero; conditioning to the value of $T$ we get that the first term equals

$$\begin{aligned}
\mathbb{P}(\mathbf{N}([a,c]) = 0, T_1 < a) &= \int_0^a \lambda e^{-\lambda t} \mathbb{P}(\widetilde{\mathbf{N}}([a-t, c-t]) = 0) dt \\
&= \int_0^a \lambda e^{-\lambda t} e^{-\lambda[c-t-(a-t)]} dt \\
&= e^{-\lambda[c-a]}(1 - e^{-\lambda a}). \quad (7.25)
\end{aligned}$$

On the other hand, the third term equals

$$\mathbb{P}(T_1 > c) = e^{-\lambda c}.$$

Hence, adding both terms we get

$$\mathbb{P}(\mathbf{N}([a,c]) = 0) = e^{-\lambda[c-a]},$$

which is the desired result for $k = 0$.

For $k > 0$ the third term vanishes and the first term is

$$\begin{aligned}
&\mathbb{P}(\mathbf{N}[a,c] = k, T_1 < a) \\
&\quad = \int_0^a \lambda e^{-\lambda t} \mathbb{P}(\widetilde{\mathbf{N}}([a-t, c-t]) = k) dt \\
&\quad = \int_0^a \lambda e^{-\lambda t} \frac{e^{-\lambda[c-t-(a-t)]} \lambda^k [c-t-(a-t)]^k}{k!} dt \\
&\quad = \frac{e^{-\lambda[c-a]} \lambda^k [c-a]^k}{k!} (1 - e^{-\lambda a}) \quad (7.26)
\end{aligned}$$

and the second term equals

$$\mathbb{P}(\mathbf{N}[a,c] = k, T_1 \in [a,c])$$

$$
\begin{aligned}
&= \int_a^c \lambda e^{-\lambda t}\mathbb{P}(\widetilde{\mathbf{N}}([0,c-t]) = k-1)dt \\
&= \int_a^c \lambda e^{-\lambda t}\frac{e^{-\lambda[c-t]}\lambda^{k-1}[c-t]^{k-1}}{(k-1)!}dt \\
&= \frac{e^{-\lambda c}\lambda^k[c-a]^k}{k!}. \qquad (7.27)
\end{aligned}
$$

The addition of those terms gives the desired

$$\mathbb{P}(\mathbf{N}[a,c] = k) = \frac{e^{-\lambda[c-a]}\lambda^k[c-a]^k}{k!},$$

which finishes the proof of the theorem. □

**Corollary 7.28** *Let $(S_i)$ be a Poisson process. Then the random variables $T_n = S_n - S_{n-1}$, $n \geq 2$, are independent, identically distributed with exponential law of parameter $\lambda$.*

**Proof.** Theorem 7.23 says that a Poisson process $\mathbf{N}_0(\cdot)$ can be constructed by fixing the first time-event according with an exponential random variable $T_1$ and then gluing a Poisson process $\mathbf{N}_1(\cdot)$ independent of $T_1$. The process $\mathbf{N}_1(\cdot)$, on the other hand, can be constructed using an exponential random variable $T_2$ and an independent Poisson process $\mathbf{N}_2(\cdot)$. Iterating this construction we have that the times between successive events is a sequence $T_1, T_2, \ldots$ of independent exponential random variables. □

## 7.5 Alternative definitions

At the beginning of this chapter we have given a constructive definition of Poisson process. We want now to compare ours with other definitions that can be found in books of Stochastic Processes. We consider point processes defined in the whole real line $\mathbb{R}$. We use the notation $\mathbf{N}(t)$ for $\mathbf{N}([0,t])$.

**Definition 7.29** A one-dimensional point process $\mathbf{N}(\cdot)$ has *stationary increments* if the law of $\mathbf{N}[s+r, t+r]$ does not depend on $r$. We say that $\mathbf{N}(\cdot)$ has *independent increments* if for any disjoint subsets of $\mathbb{R}$, $A$ and $B$ the variables $\mathbf{N}(A)$ and $\mathbf{N}(B)$ are independent.

We also need the definition of $o(h)$:

**Definition 7.30** We say that a function $f : \mathbb{R} \to \mathbb{R}$ is a $o(h)$ if

$$\lim_{h \to 0} \frac{f(h)}{h} = 0.$$

Let $\lambda > 0$. For a process $\mathbf{N}(\cdot)$ consider the following sets of conditions.

**Conditions 7.31**

1. The process $\mathbf{N}(\cdot)$ has independent and stationary increments;
2. The random variables $\mathbf{N}[s,t]$ have Poisson law with mean $\lambda(t-s)$, for any $s < t$

**Conditions 7.32**

1. The process $\mathbf{N}(\cdot)$ has independent and stationary increments;
2. $\mathbb{P}(\mathbf{N}[t, t+h] = 1) = \lambda h + o(h)$;
3. $\mathbb{P}(\mathbf{N}[t, t+h] \geq 2) = o(h)$.

Conditions 7.31 and 7.32 characterize univoquely a process as a consequence of Theorem 7.5.

**Proposition 7.33** *Conditions 7.31 and 7.32 are equivalent.*

**Proof.** It is easy to prove that Conditions 7.31 imply Conditions 7.32. For that, using that $\mathbf{N}[t, t+h]$ has Poisson law of parameter $\lambda h$, we can write

$$\begin{aligned} \mathbb{P}(\mathbf{N}[t, t+h] = 1) &= \lambda h e^{-\lambda h} \\ &= \lambda h \left(1 - \lambda h + \frac{(\lambda h)^2}{2} + \dots\right) \\ &= \lambda h + \lambda h \left(-\lambda h + \frac{(\lambda h)^2}{2} + \dots\right), \end{aligned} \tag{7.34}$$

using the series expansion of $e^{-\lambda h}$. But the second term in the last member is $o(h)$. This proves item 2 of Conditions 7.32. On the other hand,

$$\mathbb{P}(\mathbf{N}[t, t+h] \geq 2) = e^{-\lambda h}\left(\frac{(h\lambda)^2}{2!} + \frac{(h\lambda)^3}{3!} + \dots\right) = o(h).$$

This proves item 3 of Conditions 7.32.

The proof that Conditions 7.32 imply Conditions 7.31 is more complicated and involves the solution of differential equations. We omit this proof. It can be found, for instance, in Ross (1983). □

**Proposition 7.35** *The process $S_n$ defined by* (7.1) *satisfies both the Conditions 7.31 and 7.32*

**Proof.** The construction guarantees that the process has independent and stationary increments. By Lemma 7.6 the distribution of points in $[s, t]$ has Poisson law of parameter $\lambda(t - s)$. This implies that the process defined by (7.1) satisfies Conditions 7.31. By Proposition 7.33 the process also satisfies Conditions 7.32. □

## 7.6 Inspection paradox

Imagine a city where the bus service is perfectly regular: according to the schedule the time between two successive buses is exactly one hour or two hours. The schedule is 0, 1, 3, 4, 6, 7, 9, 10, 12, etc. Half of the time intervals between two successive buses has length 1 and the other half has length 2. If we arrive to the bus stop at a time randomly chosen in the interval 0:00 and 12:00, how long we need do wait in mean until the departure of the next bus?

One way to do that is just to compute the length of the average interval. This operation corresponds to choose an interval at random (from a urn, for instance) and measure its length $T$. Since there are as many intervals of length 1 as intervals of length 2, the probability of choosing an interval of length 1 is 1/2 and the same for an interval of length 2. The resulting average would be

$$\mathbb{E}T = \frac{1}{2}\frac{1}{2} + \frac{1}{2}1 \;=\; \frac{3}{4}.$$

This corresponds to compute the average length of half interval. This is *WRONG*.

The problem with the above reasoning is that when one chooses a point uniformly distributed in the interval $[0, 12]$, the probability that it belongs to an interval of length 2 is $8/12$ while the probability of belonging to an interval of length 1 is $4/12$. The average length is then

$$\mathbb{E}T = \frac{4}{12}\frac{1}{2} + \frac{8}{12}1 = \frac{5}{6} > \frac{3}{4}.$$

That is, the average waiting time is bigger than the average of half interval. The reason is that the probability of choosing a long interval is bigger than the probability of choosing a short one. The same happens when we have a stationary point process in $\mathbb{R}$. The length of the interval containing the origin is in general bigger than the length of the typical interval.

Set us see what happens with the Poisson process. Let $S_1$ be the epoch of the first Poisson event to the right of the origin of time and $S_0$ the time of the first event to the left of the origin. We have seen that $S_1$ has exponential distribution with rate $\lambda$. The same argument shows that $S_0$ has also exponential distribution with the same rate. Since the process has independent increments, $S_0$ and $S_1$ are independent random variables. Hence, the law of $S_1 - S_0$ is the law of the sum of two independent random variables exponentially distributed with rate $\lambda$. This corresponds to a Gamma distribution with parameters 2 and $\lambda$:

$$\mathbb{P}(S_1 - S_0 > t) \;=\; \int_t^\infty \lambda x e^{-\lambda x} dx.$$

and the average value of this interval is

$$\mathbb{E}(S_1 - S_0) \;=\; \int_0^\infty \lambda x^2 e^{-\lambda x} dx = \frac{2}{\lambda}.$$

which is exactly twice the average value of the typical interval; the typical interval has exponential distribution with rate $\lambda$.

Notice also that the interval containing an arbitrary time $t$ also has average length equal to $2/\lambda$. For $t \geq 0$, this interval is $S_{\mathbf{N}(t)+1} - S_{\mathbf{N}(t)}$. By the same reasoning, $S_{\mathbf{N}(t)+1} - t$ and $t - S_{\mathbf{N}(t)}$ are independent random variables exponentially distributed. Hence $S_{\mathbf{N}(t)+1} - S_{\mathbf{N}(t)}$ has Gamma distribution with parameters 2 and $\lambda$.

The only process for which the length of the interval containing a point is the same as the length of the typical interval is the process for which the inter arrival times are constant.

## 7.7 Poisson processes in $d \geq 2$

In this section we construct a Poisson process in two dimensions. The resulting process will be used in the following sections to construct one-dimensional non-homogeneous processes and superposition of Poisson processes.

We construct a random subset of $\mathbb{R}^2$. The same construction can be performed in $\mathbb{R}^d$ for $d \geq 2$, but for simplicity we stay in $d = 2$. The construction is almost the same we did in $d = 1$.

We start with a partition of $\mathbb{R}^2$ in finite rectangles $A_i$. For instance $A_i$ can be the squares determined by the lattice $\mathbb{Z}^2$. Denote $|A_i|$ the area (Lebesgue measure) of $A_i$. We have

$$\dot{\cup}_{i\in\mathbb{Z}} A_i = \mathbb{R}^2,$$

where $\dot{\cup}$ means disjoint union.

For each $i$, let $Y_i$ be a Poisson random variable with mean $\lambda|A_i|$. That is,

$$\mathbb{P}(Y_i = k) = \frac{e^{-\lambda|A_i|}(\lambda|A_i|)^k}{k!}.$$

Assume the random variables $Y_i$ are independent.

Finally, for each $i$ consider a sequence of *iid* random variables $(U_{i,j})_{j\geq 1}$ uniformly distributed in $A_i$:

$$\mathbb{P}(U_{i,j} \in A \cap A_i) = \frac{|A \cap A_i|}{|A_i|}.$$

The point process is the following (random) set:

$$\mathbf{S} = \cup_{i \in \mathbb{Z}} \cup_{j=1}^{Y_i} \{U_{i,j}\},$$

where we used the convention $\cup_{j=1}^{0}\{U_{i,j}\} = \emptyset$. In other words, we put $Y_i$ independent points uniformly distributed in the rectangle $A_i$.

Up to now we have repeated the procedure of $d = 1$. The difference is that there is no satisfactory way to order the points of the random set $\mathbf{S}$. But this is not important.

**Definition 7.36** For each measurable set $A \subset \mathbb{R}^2$, define

$$\mathbf{M}(A) \ := \ \text{number of points of the set } \mathbf{S} \cap A.$$

To avoid confusions we use the letter $\mathbf{M}$ for bi-dimensional Poisson processes and the letter $\mathbf{N}$ for one dimensional processes.

The following properties are proven in the same way that in the one-dimensional case. We leave the details to the reader.

**Lemma 7.37** *For each finite set $A$ the random variable $\mathbf{M}(A)$ has Poisson distribution with mean $\lambda|A|$.*

**Lemma 7.38** *For each finite family of measurable sets $B_l$, the random variables $\mathbf{M}(B_l)$ have Poisson law with mean $\lambda|B_l|$.*

**Lemma 7.39** *For each measurable set $A \subset \mathbb{R}^2$, the conditional distribution of the points of $\mathbf{S} \cap A$ given that $\mathbf{M}(A) = n$ is the same as the $n$ independent random variables uniformly distributed in $A$.*

**Example 7.40** Given the two-dimensional construction, we can compute the law of the random variable which measures the distance to the origin of the point closer to the origin. Let $V = \inf\{|x| : x \in \mathbf{S}\}$.

$$\mathbb{P}(V > b) = \mathbb{P}(\mathbf{M}(B(0,b)) = 0) = e^{-\lambda \pi b^2},$$

where $B(0, b)$ is the circle centered at the origin with radius $b$.

## 7.8 Projections

Suppose the random measure $\mathbf{M}(\cdot)$ describes the bi-dimensional Poisson process of parameter 1. Now we want to construct a one-dimensional Poisson process $\mathbf{N}(\cdot)$ with parameter $\lambda$ as a function of $\mathbf{M}(\cdot)$. For each interval $I \subset \mathbb{R}$ define

$$\mathbf{N}(I) = \mathbf{M}(I \times [0, \lambda]) \tag{7.41}$$

that is, the number of points of $\mathbf{N}(\cdot)$ in the interval $I$ will be the same as the number of points in the rectangle $I \times [0, \lambda]$ for $\mathbf{M}(\cdot)$. This is the same as to project the points of $\mathbf{M}(\cdot)$ of the strip $\mathbb{R} \times [0, \lambda]$ on $\mathbb{R}$.

**Lemma 7.42** *The process* $\mathbf{N}(\cdot)$ *defined by* (7.41) *is a one-dimensional Poisson process of parameter* $\lambda$*.*

**Proof.** Since the fact that the increments are independent is immediate, it suffices to prove that for disjoint intervals $I_1, \dots, I_n$:

$$\mathbb{P}(\mathbf{N}(I_i) = k_i\,,\, i = 1, \dots, n) = \prod_{i=1}^{n} \frac{e^{-\lambda |I_i|}(\lambda |I_i|)^{k_i}}{k_i!},$$

which follows immediately from the definition (7.41). □

The reader may ask if two points of the bi dimensional process could project to a unique point of the one-dimensional one. The following lemma answers negatively this question.

**Lemma 7.43** *Let* $I$ *be a finite interval. The event "two points of the bi-dimensional process are projected over a unique point of* $I$*" has probability zero.*

**Proof.** Without loss of generality we may assume that $I = [0, 1]$. Partition $I$ in small intervals of length $\delta$: $I_n^\delta = (n\delta, (n+1)\delta]$. The probability that two points are projected in one is bounded above by the probability that two points belong to the same interval, which is given by

$$\mathbb{P}(\cup_{n=1}^{|I|/\delta}\{M(I_n^\delta \times [0, \lambda]) \geq 2\}) \leq \sum_{n=1}^{|I|/\delta} \mathbb{P}(M(I_n^\delta \times [0, \lambda]) \geq 2) \leq \frac{|I|}{\delta} o(\delta).$$

This goes to zero as $\delta \to 0$. □

The advantage of the projection method to construct one-dimensional processes is that we can simultaneously construct processes with different rates in such a way that the number of points of the process with bigger rate always dominate the number of points of the process with smaller rate.

We are going to construct a coupling between two Poisson processes $\mathbf{N}_1(\cdot)$ and $\mathbf{N}_2(\cdot)$ with parameters $\lambda_1$ and $\lambda_2$, respectively. For $i = 1, 2$ let

$$\mathbf{N}_i(I) = \mathbf{M}(I \times [0, \lambda_i]). \tag{7.44}$$

**Lemma 7.45** *Assume $\lambda_1 \geq \lambda_2$. Then, for the coupled process defined by* (7.44) *we have*

$$\mathbf{N}_1(I) \geq \mathbf{N}_2(I),$$

*for all interval $I \subset \mathbb{R}$.*

**Proof.** Follows from the definition that $\mathbf{N}_2$ projects less points than $\mathbf{N}_1$. □

## 7.9 Superposition of Poisson processes

Men arrive to a bank accordingly to a Poisson process of parameter $p\lambda$, women do so at rate $(1-p)\lambda$.

A way to model the arrival process including the attribute "sex" to each arrival is to construct a bi-dimensional process $\mathbf{M}(\cdot)$ as we did in the previous section and to define $\mathbf{N}_1(\cdot)$ and $\mathbf{N}_2(\cdot)$ as the projection of strips of width $p\lambda$ and $(1-p)\lambda$:

$$\mathbf{N}_1(I) = \mathbf{M}(I \times [0, p\lambda]); \quad \mathbf{N}_2(I) = \mathbf{M}(I \times [p\lambda, \lambda]).$$

That is, the projected points coming from the strip $[0, \lambda] \times \mathbb{R}$ indicate the arrival times of clients disregarding sex. The points coming from the strip $[0, \lambda p] \times \mathbb{R}$ are marked 1 (corresponding to men) and the points coming from the strip $[p\lambda, \lambda] \times \mathbb{R}$ are marked 2 (corresponding to women).

We have seen in the previous section that the process $\mathbf{N}(\cdot)$ defined by $\mathbf{N}(I) = \mathbf{N}_1(I) + \mathbf{N}_2(I)$ is a Poisson process with rate $\lambda$ because we can express $\mathbf{N}(I) = \mathbf{M}([0,\lambda] \times I)$.

Let $S_n$, $n \geq 1$ be the arrival times of the clients disregarding sex; that is, the arrival times of the process $\mathbf{N}(\cdot)$. Each point has a mark 1 or 2 according to the strip it is projected from. Let

$$G(S_i) = \begin{cases} 1 & \text{if } S_i \text{ is marked } 1 \\ 2 & \text{if } S_i \text{ is marked } 2. \end{cases}$$

**Proposition 7.46** *The random variables $G(S_i)$ are independent, identically distributed with marginals*

$$\mathbb{P}(G(S_i) = 1) = p\,; \qquad \mathbb{P}(G(S_i) = 2) = 1 - p$$

**Proof.** We start constructing $n$ *iid* random variables uniformly distributed in the rectangle $[0,t] \times [0,\lambda]$. Let $V_1, V_2, \ldots$ and $W_1, W_2, \ldots$ be two independent sequences of *iid* random variables uniformly distributed in $[0,1]$.

For each fixed $n$ let $\pi_n$ be the random permutation of the set $\{1,\ldots.n\}$ (that is a bijection of this set onto itself) defined by: for $i = 1,\ldots,n-1$,

$$V_{\pi_n(i)} \leq V_{\pi_n(i+1)}.$$

That is, $\pi_n(i)$ is the label of the $i$-th variable when the first $n$ $V_j$ are ordered from the smallest to the biggest.

For each fixed $n$ construct a sequence of random variables in $[0,t] \times [0,\lambda]$ as follows:

$$U_i = (tV_{\pi_n(i)}, \lambda W_i). \tag{7.47}$$

The family $(U_i \,:\, i = 1,\ldots,n)$ consists on $n$ *iid* random variables uniformly distributed in $[0,t] \times [0,\lambda]$. The advantage of this construction is that $W_i$ is the ordinate of the $i$-th $U_j$, when the $n$ first $U_i$ are ordered from smallest to the biggest.

Fix an arbitrary $L$ and for $1 \leq k \leq L$ consider arbitrary $a_k \in \{0,1\}$. Define

$$A_L = \{G(S_1) = a_1, \ldots, G(S_L) = a_L\}.$$

By a discrete version of Theorem 7.5, to prove that $(G(S_i))$ are *iid* Bernoulli random variables with parameter $p$ it suffices to prove that

$$\mathbb{P}(A_L) = p^{\sum a_k}(1-p)^{\sum(1-a_k)}. \tag{7.48}$$

For each fixed $t$,

$$\mathbb{P}(A_L) \;=\; \sum_{n=L}^{\infty} \mathbb{P}(A_L, \mathbf{N}(0,t) = n) + \mathbb{P}(A_L, \mathbf{N}(0,t) < L). \tag{7.49}$$

Let us compute the first term:

$$\mathbb{P}(A_L, \mathbf{N}(0,t) = n) = \mathbb{P}(A_L \mid \mathbf{N}(0,t) = n)\, \mathbb{P}(\mathbf{N}(0,t) = n). \tag{7.50}$$

But $\mathbf{N}(0,t) = \mathbf{M}([0,t]\times[0,\lambda])$ and given $\mathbf{M}([0,t]\times[0,\lambda]) = n$, the law of the points in this rectangle is the same as the law of $n$ *iid* random variables uniformly distributed in the rectangle. Let

$$I_k = \begin{cases} [0,\lambda p] & \text{if } a_k = 1 \\ [\lambda p, \lambda] & \text{if } a_k = 0. \end{cases} \tag{7.51}$$

We can then use the construction (7.47) to obtain

$$\begin{aligned} \mathbb{P}(A_L \mid \mathbf{N}(0,t) = n) &= \mathbb{P}(A_L \mid \mathbf{M}([0,t]\times[0,\lambda]) = n) \\ &= \mathbb{P}(U_{\pi_n(k)} \in [0,t]\times I_k, 1 \le k \le n) \\ &= \mathbb{P}(\lambda W_k \in I_k, 1 \le k \le n) \\ &= p^{\sum a_k}(1-p)^{\sum(1-a_k)}. \end{aligned} \tag{7.52}$$

Since this identity does not depend on $n$, using (7.52) and (7.50) in (7.49) we get

$$\begin{aligned} \mathbb{P}(A_L) &= p^{\sum a_k}(1-p)^{\sum(1-a_k)} \sum_{n=L}^{\infty} \mathbb{P}(\mathbf{N}(0,t) = n) + \mathbb{P}(A_L, \mathbf{N}(0,t) < L). \\ &= p^{\sum a_k}\,(1-p)^{\sum(1-a_k)}\, \mathbb{P}(\mathbf{N}(0,t) \ge L) \;+\; \mathbb{P}(A_L, \mathbf{N}(0,t) < L). \end{aligned}$$

By the law of large numbers for one dimensional Poisson processes (which can be proven as the law of large numbers for renewal processes of Chapter 5) $\mathbf{N}(0,t)/t$ converges to $\lambda$; hence

$$\lim_{t\to\infty} \mathbb{P}(\mathbf{N}(0,t) \ge L) \;=\; 1 \quad \text{and} \quad \lim_{t\to\infty} \mathbb{P}(A_L, \mathbf{N}(0,t) \ge L) \;=\; 0 \tag{7.53}$$

This shows the Proposition. □

The proof above induces the following alternative construction of a bi-dimensional Poisson process $\mathbf{M}(\cdot)$ with rate 1 in the strip $[0,\infty]\times[0,\lambda]$. Let $T_1$ be an exponential random variable with rate $\lambda$, $W_1$ be a random variable uniformly distributed in $[0,\lambda]$ and $\mathbf{M}_1(\cdot)$ be a one dimensional process with rate 1 in the strip $[0,\infty]\times[0,\lambda]$. Assume independence among $T_1$, $W_1$ e $\mathbf{M}_1(\cdot)$.

Define $\mathbf{M}(\cdot)$ as the process

$$\mathbf{M}(A) = \mathbf{1}\{(T_1, W_1)\in A\} + \mathbf{M}_1(A - T_1),$$

where

$$A - t = \{(x,y)\in\mathbb{R}^2 : (x+t, y)\in A\}.$$

Arguments very similar to those of Theorem7.23, Corollary 7.28 and Proposition 7.46 show the following

**Theorem 7.54** *Let $\mathbf{M}(\cdot)$ be a bi-dimensional Poisson process with rate 1. Let $S_1, S_2, \dots$ be the ordered times of occurrence of events in the strip $[0,\lambda]$. Let $W_1, W_2, \dots$ be the second coordinates of those event times. Then, $(S_{i+1} - S_i)_{i\geq 1}$ are* iid *random variables exponentially distributed with rate $\lambda$ and $(W_i)_{i\geq 1}$ are* iid *random variables with uniform distribution in $[0,\lambda]$. Furthermore $(S_{i+1} - S_i)_{i\geq 1}$ and $(W_i)_{i\geq 1}$ are independent.*

## 7.10 Non homogeneous processes.

Let $\lambda : \mathbb{R}\to\mathbb{R}^+$ be a nonnegative piecewise continuous function. Assume that for any finite interval $I\subset\mathbb{R}$, the number of discontinuities of $\lambda(t)$ is finite.

We want to construct a point process with independent increments and "instantaneous rate" $\lambda(t)$. That is, a process $\mathbf{N}(\cdot)$ such that, for the continuity points of $\lambda(t)$,

$$\mathbb{P}(\mathbf{N}([t,t+h]) = 1) = h\lambda(t) + o(h) \quad (7.55)$$

$$\mathbb{P}(\mathbf{N}([t,t+h]) \geq 2) = o(h). \quad (7.56)$$

(See Definition 7.32.)

We consider a bi-dimensional process $\mathbf{M}(\cdot)$ and define

$$\mathbf{N}(I) = \mathbf{M}(\Lambda(I)), \tag{7.57}$$

where $\Lambda(I) = \{(x, y) \in \mathbb{R}^2 : x \in I \text{ and } y \leq \lambda(x)\}$. That is, $\mathbf{N}(\cdot)$ is the process obtained when we project the points of $\mathbf{M}(\cdot)$ lying below the function $\lambda(t)$.

**Lemma 7.58** *The process defined by* (7.57) *satisfies conditions*(7.56) *and* (7.55).

**Proof.** By Definition 7.57,

$$\mathbb{P}(\mathbf{N}([t, t+h]) = 1) = \mathbb{P}(\mathbf{M}(\Lambda[t, t+h]) = 1).$$

Let $y_0 = y_0(t, h)$ and $y_1 = y_1(t, h)$ be respectively, the infimum and the supremum of $\lambda(t)$ in the interval $[t, t+h]$. In this way we obtain the following bounds

$$\mathbf{M}([t, t+h] \times [0, y_0]) \leq \mathbf{M}(\Lambda[t, t+h]) \leq \mathbf{M}([t, t+h] \times [0, y_1]).$$

Since the function $\lambda(t)$ is continuous in $t$,

$$\lim_{h \to 0} y_0(t, h) = \lim_{h \to 0} y_1(t, h) = \lambda(t)$$

and $y_1(t, h) - y_0(t, h) = O(h)$, by continuity, where $O(h)$ is a notation to indicate a function of $h$ that stays bounded above and below when divided by $h$ as $h \to 0$. Hence,

$$\mathbb{P}(\mathbf{M}([t, t+h] \times [0, y_1]) - \mathbf{M}([t, t+h] \times [0, y_0]) \geq 1) = o(h).$$

So that,

$$\begin{aligned}
&\mathbb{P}(\mathbf{M}(\Lambda[t, t+h]) = 1) \\
&\qquad = \mathbb{P}(\mathbf{M}(\Lambda[t, t+h]) = 1, \\
&\qquad\qquad \mathbf{M}([t, t+h] \times [0, y_1]) - \mathbf{M}([t, t+h] \times [0, y_0]) = 0) \\
&\qquad\quad + \mathbb{P}(\mathbf{M}(\Lambda[t, t+h]) = 1, \mathbf{M}([t, t+h] \times [0, y_1]) \\
&\qquad\qquad\qquad - \mathbf{M}([t, t+h] \times [0, y_0]) \geq 1).
\end{aligned} \tag{7.59}$$

The first term equals

$$\mathbb{P}(\mathbf{M}([t, t+h] \times [0, y_0)]) = 1, \\ \mathbf{M}([t, t+h] \times [0, y_1]) - \mathbf{M}([t, t+h] \times [0, y_0]) = 0),$$

which, using independence of the process $\mathbf{M}(\cdot)$, gives

$$\begin{aligned} &\mathbb{P}(\mathbf{M}([t, t+h] \times [0, y_0)]) = 1) \\ &\quad \times \mathbb{P}(\mathbf{M}([t, t+h] \times [0, y_1]) - \mathbf{M}([t, t+h] \times [0, y_0]) = 0) \\ &= h\lambda(t) + o(h). \end{aligned}$$

The second term is bounded by

$$\mathbb{P}(\mathbf{M}([t, t+h] \times [0, y_1]) - \mathbf{M}([t, t+h] \times [0, y_0]) \geq 1) = o(h). \square$$

Definition (7.57) allows us to show immediately that the number of points of the projected process in a set has Poisson law:

**Proposition 7.60** *The non homogeneous Poisson process constructed in display*

(7.57) *has the property that the number of points in an arbitrary interval has Poisson law with mean equal to the area below the function $\lambda(t)$ in that interval. That is,*

$$\mathbb{P}(\mathbf{N}([0, t]) = n) = \frac{e^{\mu(t)}(\mu(t))^n}{n!},$$

*where $\mu(t) = \int_0^t \lambda(s) ds$.*

Definition (7.57) of non-homogeneous process has the advantage that we can couple two or more processes with different rates with the following properties.

**Proposition 7.61** *Let $\lambda_1(t)$ and $\lambda_2(t)$ be two continuous functions satisfying*

$$\lambda_1(t) \leq \lambda_2(t) \text{ for all } t$$

*Then it is possible to couple the non-homogeneous Poisson process $\mathbf{N}_1(\cdot)$ and $\mathbf{N}_2(\cdot)$ with rates $\lambda_1(t)$ and $\lambda_2(t)$ respectively such that for all interval $I$,*

$$\mathbf{N}_1(I) \leq \mathbf{N}_2(I).$$

**Proof.** The proof of this statement is immediate from the definition. $\square$

## 7.11 Exercises

**Exercise 7.1** Let $X$ and $Y$ be *iid* exponential random variables with parameters $\lambda_1$ and $\lambda_2$ respectively,

1. Compute the law of $Z = \min(X, Y)$.
2. Compute the conditional law of $Z$ given $X = x$?

**Exercise 7.2** Show that the sum of $n$ *iid* random variables with Poisson distribution with means $\lambda_1, \ldots, \lambda_n$, respectively has Poisson law with mean $\lambda_1 + \ldots + \lambda_n$.

**Exercise 7.3** In one urn we put $N_1$ balls type 1, $N_2$ balls type 2 and $N_3$ balls type 3, where $N_i$, $i = 1, 2, 3$ are independent Poisson random variables with expectations $\mathbb{E}N_i = \lambda_i$.
(a) Choose a ball at random from the urn. Show that given the event $\{N_1 + N_2 + N_3 \geq 1\}$, the probability to have chosen a ball type i is $\lambda_i/(\lambda_1+\lambda_2+\lambda_3)$.
(b) Show that the result of (a) is the same if we condition to the event $\{N_1 + N_2 + N_3 = n\}$ for a fixed arbitrary $n \geq 1$.
(c) Show that, given the event $\{N_1+N_2+N_3 = n \geq 1\}$, the law of the type of the balls in the urn is a trinomial with parameters $n$ and $p_i = \lambda_i/(\lambda_1+\lambda_2+\lambda_3)$.
(d) Generalize item (c) for $k \geq 1$ different types of balls:

$$\mathbb{P}(N_i = n_i \mid N = n) = \frac{n!}{n_1! \ldots n_k!} \lambda_1^{n_1} \ldots \lambda_k^{n_k} (\lambda_1 + \ldots + \lambda_k)^{-n},$$

for $n_1 + \ldots + n_k = n \geq 1$.

**Exercise 7.4** Let $\mathbf{N}(t)$ be a Poisson process with parameter $\lambda$. For $s < r < t$ compute

$$P(\mathbf{N}(s) = k, \mathbf{N}(r) - \mathbf{N}(s) = j | \mathbf{N}(t) = n).$$

**Exercise 7.5** Show that the random variables $U_k$ defined in (7.47) are independent and uniformly distributed in $[0, t] \times [0, \lambda]$.

**Exercise 7.6** Show Theorem 7.54.

**Exercise 7.7** (i) Let $N$ be a random variable with Poisson law of mean $\lambda$. Let $B = X_1 + \ldots + X_N$, where $X_i$ is a Bernoulli random variable with parameter $p$ independent of $N$. Prove that $B$ is Poisson with parameter $\lambda p$. Hint: construct a process $\mathbf{M}(\cdot)$ in $[0,1] \times [0,\lambda]$ and mark the points accordingly to the fact that they are above or below the line $y = \lambda p$.

(ii) Prove that if $T_n$ are the successive instants of a Poisson process with parameter $\lambda$, then the process defined by the times

$$S_n = \inf\{T_l > S_{n-1} : X_l = 1\},$$

with $S_1 > 0$, is a Poisson process with parameter $\lambda p$. Hint: construct simultaneously $(S_n)$ and $(T_n)$ using Definition 7.41.

**Exercise 7.8** Assume that a random variable $X$ satisfies the following property. For all $t \geq 0$

$$\mathbb{P}(X \in (t, t+h) \mid X > t) = \lambda h + o(h).$$

Show that $X$ is exponentially distributed with rate $\lambda$.

**Exercise 7.9** For a Poisson process of rate $\lambda$,

(a) Compute the law of $T_2$, the instant of the second event.

(b) Compute the law of $T_2$ given $\mathbf{N}(0,t) = 4$.

**Exercise 7.10** Let $\mathbf{N}(0,t)$ be a Poisson process with parameter $\lambda$. Compute

(a) $\mathbb{E}\mathbf{N}(0,2)$, $\mathbb{E}\mathbf{N}(4,7)$,

(b) $\mathbb{P}(\mathbf{N}(4,7) > 2 \mid \mathbf{N}(2,4) \geq 1)$,

(c) $\mathbb{P}(\mathbf{N}(0,1) = 2 \mid \mathbf{N}(0,3) = 6)$,

(d) $\mathbb{P}(\mathbf{N}(0,t) = \text{odd})$,

(e) $\mathbb{E}(\mathbf{N}(0,t)\mathbf{N}(0,t+s))$.

**Exercise 7.11** For a Poisson process $\mathbf{N}(0,t)$ compute the joint distribution of $S_1, S_2, S_3$, the arrival instants of the first three events.

**Exercise 7.12** Compute the law of the age $A(t)$ and residual time $R(t)$ of a Poisson process with parameter $\lambda$, where

$$A(t) := t - S_{\mathbf{N}(t)} \qquad R(t) := S_{\mathbf{N}(t)+1} - t \tag{7.62}$$

**Exercise 7.13** Let $\mathbf{N}(t)$ be a stationary Poisson process. Show that $\mathbf{N}(t)/t \to 1/\lambda$, as $t \to \infty$. Hint: use the ideas of Proposition 5.26 and the laws of the age $A(t)$ and the residual time $R(t)$ computed in Exercise 7.12.

**Exercise 7.14** Women arrive to a bank according to a Poisson process with parameter 3 per minute and men do so according to a Poisson process with parameter 4 per minute. Compute the following probabilities:

(a) The first person to arrive is a man.

(b) 3 men arrive before the fifth woman.

(c) 3 clients arrive in the first 3 minutes.

(d) 3 men and no woman arrive in the first 3 minutes.

(e) Exactly three women arrive in the first 2 minutes, given that in the first 3 minutes 7 clients arrive.

**Exercise 7.15** Assume that only two types of clients arrive to a supermarket counter. Those paying with credit card and those paying cash. Credit card holders arrive according to a Poisson process $\{\mathbf{N}_1(t), t \geq 0\}$ with parameter 6 per minute, while those paying cash arrive according to a Poisson process $\{\mathbf{N}_2(t), t \geq 0\}$ with rate 8 per minute.

(a) Show that the process $\mathbf{N}(t) = \{\mathbf{N}_1(t) + \mathbf{N}_2(t), t \geq 0\}$ of arrivals of clients to the supermarket is a Poisson process with rate 14.

(b) Compute the probability that the first client arriving to the supermarket pays cash.

(c) Compute the probability that the first 3 clients pay with credit card, given that in the first 10 minutes no client arrived.

**Exercise 7.16** Vehicles arrive to a toll according to a Poisson process with parameter 3 per minute (180 per hour). The probability each vehicle to be a car depends on the time of the day and it is given by the function $p(s)$:

$$p(s) = \begin{cases} s/12 & \text{if } 0 < s < 6 \\ 1/2 & \text{if } 6 \leq s < 12 \\ 1/4 & \text{if } 12 \leq s < 18 \\ (1/4) - s/24 & \text{if } 18 \leq s < 24 \end{cases}.$$

The probability each vehicle to be a truck is $1 - p(s)$.

(a) Compute the probability that at least a truck arrives between 5:00 and 10:00 in the morning.

(b) Given that 30 vehicles arrived between 0:00 and 1:00, compute the law of the arrival time of the first vehicle of the day.

**Exercise 7.17** Construct a non homogeneous Poisson process with rate $\lambda(t) = e^{-t}$. Compute the mean number of events in the interval $[0, \infty]$.

**Exercise 7.18** Show that for the non homogeneous Poisson process defined in (7.57), the law of $\{S_1, \ldots, S_n\}$ in $[0, t]$ given $\mathbf{N}(0, t) = n$ is the same as the law of $n$ independent random variables with density

$$\frac{\lambda(r)}{\int_0^t \lambda(s) ds}; \quad r \in [0, t].$$

## 7.12 Comments and references

The construction of Poisson process given in this chapter is a particular case of the general construction proposed by Neveu (1977). The construction of processes in a lower dimension as a projection of a process in a higher dimension was introduced by Kurtz (1989) and used by Garcia (1995).

# Chapter 8

# Continuous time Markov processes

## 8.1 Pure jump Markov processes

The crucial difference between the Markov chains of Chapter 2 and the continuous time Markov processes of this chapter is the following. A Markov chain jumps from one state to another one at integer times: $1, 2, \dots$. Pure jump processes jump at random times $\tau_1, \tau_2, \dots$.

We now define a process $X_t \in \mathcal{X}$, $t \in \mathbb{R}$. We assume $\mathcal{X}$ countable, consider *transition rates* $q(x, y) \geq 0$ for $x, y \in \mathcal{X}$ such that $x \neq y$ and want to construct a process with the property that the rate of jumping from state $x$ to state $y$ be $q(x, y)$. In other words, the process must satisfy

$$\mathbb{P}(X_{t+h} = y \mid X_t = x) = hq(x, y) + o(h) \quad \text{for all } x \neq y. \tag{8.1}$$

Let

$$q(x) = \sum_y q(x, y),$$

be the *exit rate* from state $x$.

To construct the process we introduce a bi-dimensional Poisson process $\mathbf{M}(\cdot)$. For each state $x$ partition the interval $I^x = [0, q(x)]$ in intervals $I(x, y)$ of length $q(x, y)$.

Let $x_0 \in \mathcal{X}$ and suppose $X_0 = x_0$. Let $\tau_1$ be the first time an event of the process $\mathbf{M}(\cdot)$ appears in the interval $I^{x_0}$:

$$\tau_1 = \inf\{t > 0 : \mathbf{M}(I^{x_0} \times [0,t]) > 0\}.$$

Define $x_1 = y$, where $y$ is the unique state satisfying

$$\begin{aligned} & \inf\{t > 0 : \mathbf{M}(I^{x_0} \times [0,t]) > 0\} \\ = \ & \inf\{t > 0 : \mathbf{M}(I(x_0, y) \times [0,t]) > 0\} \end{aligned} \tag{8.2}$$

that is, $x_1$ is determined by the interval $I(x_0, x_1)$ which realizes the infimum.

Assume now that $\tau_{n-1}$ and $x_{n-1}$ are already determined. Define inductively

$$\tau_n = \inf\{t > \tau_{n-1} : \mathbf{M}(I^{x_{n-1}} \times (\tau_{n-1}, t]) > 0\}.$$

and $x_n = y$ if and only if

$$\inf\{t > \tau_{n-1} : \mathbf{M}(I^{x_{n-1}} \times (\tau_{n-1}, t]) > 0\} \tag{8.3}$$

$$= \ \inf\{t > 0 : \mathbf{M}(I(x_{n-1}, y) \times (\tau_{n-1}, t] > 0)\}. \tag{8.4}$$

**Definition 8.5** Define $\tau_\infty = \sup_n \tau_n$ and

$$X_t = x_n, \ \text{ if } t \in [\tau_n, \tau_{n+1}). \tag{8.6}$$

for all $t \in [0, \infty)$

In this way, for each (random) realization of the bi-dimensional Poisson process $\mathbf{M}(\cdot)$, we construct (deterministically) a realization of the process $X_t$ for $t \in [0, \tau_\infty)$. Observe that $\tau_n$ is the $n$-th jump time of the process and $x_n$ is the $n$-th state visited by the process.

**Proposition 8.7** *The process* $(X_t : t \in [0, \tau_\infty))$ *defined above satisfies the properties* (8.1).

**Proof.** By definition,

$$\mathbb{P}(X_{t+h} = y \mid X_t = x) \tag{8.8}$$

$$= \ \mathbb{P}(\mathbf{M}(I(x,y) \times (t, t+h]) = 1) + \mathbb{P}(\text{other things}), \tag{8.9}$$

where the event {other things} is contained in the event

$$\{\mathbf{M}([0, q(x)] \times (t, t+h]) \geq 2\}$$

(two or more Poisson events occur in the rectangle $[0, q(x)] \times (t, t+h]$). It is clear by Definition 7.36 of $\mathbf{M}(\cdot)$ that

$$\mathbb{P}(\mathbf{M}([0, q(x)] \times (t, t+h]) \geq 2) = o(h) \quad \text{and} \tag{8.10}$$

$$\mathbb{P}(\mathbf{M}(I(x, y) \times (t, t+h]) = 1) = hq(x, y) + o(h). \tag{8.11}$$

This finishes the proof of the proposition. □

**Example 8.12** In this example we show how the construction goes for a process with 3 states: $\mathcal{X} = \{0, 1, 2\}$. This may model a system with two servers but with no waiting place for clients that are not being served: clients arriving when both servers are occupied are lost. If the client arrivals occur according to a Poisson process with rate $\lambda^+$ and the services have exponential distribution with rate $\lambda^-$, then the process has rates:

$$q(0, 1) = q(1, 2) = \lambda^+ \tag{8.13}$$

$$q(1, 0) = \lambda^-; \quad q(2, 1) = 2\lambda^- \tag{8.14}$$

$$q(x, y) = 0, \text{ in the other cases.} \tag{8.15}$$

The construction described in the proposition can be realized with the following intervals:

$$I(0, 1) = I(1, 2) = [0, \lambda^+] \tag{8.16}$$

$$I(1, 0) = [\lambda^+, \lambda^+ + \lambda^-]; \; I(2, 1) = [0, 2\lambda^-]. \tag{8.17}$$

In this case all rates are bounded by $\max\{\lambda^+ + \lambda^-, 2\lambda^-\}$.

**Example 8.18 (Pure birth process)** We construct a process $X_t$ in the state space $\mathcal{X} = \{0, 1, 2, \ldots\}$ with the following properties.

$$\mathbb{P}(X_{t+h} = x + 1 \mid X_t = x) = \lambda x h + o(h) \tag{8.19}$$

$$\mathbb{P}(X_{t+h} - X_t \geq 2 \mid X_t = x) = o(xh). \tag{8.20}$$

That is, the arrival rate at time $t$ is proportional to the number of arrivals occurred up to $t$. Given the bi-dimensional process $\mathbf{M}(\cdot)$ with rate one, we construct the jump times as follows. Let

$$\tau_1 = \inf\{t > 0 : \mathbf{M}([0,t] \times [0,\lambda]) = 1\}$$

and in general

$$\tau_n = \inf\{t > \tau_{n-1} : \mathbf{M}((\tau_{n-1},t] \times [0,(n-1)\lambda]) = 1\}.$$

Then the process $X_t$ is defined by $X_0 = 1$ and

$$X_t := n+1 \quad \text{if} \quad t \in [\tau_n, \tau_{n+1}], \tag{8.21}$$

for $n \geq 0$. It is easy to see that this process satisfies the conditions (8.19) and (8.20).

## 8.2 Explosions

It is easy to see that if the state space $\mathcal{X}$ is finite, then $\tau_\infty$ is infinity. That is, in any finite time interval there are a finite number of jumps. The situation is different when the state space is infinite.

Consider a process with state space $\mathcal{X} = \mathbb{N}$ and rates

$$q(x,y) = \begin{cases} 2^x, & \text{if } y = x+1 \\ 0, & \text{otherwise} \end{cases} \tag{8.22}$$

The construction using a Poisson process goes as follows. Define $\tau_0 = 0$ and inductively $\tau_n$ by

$$\tau_n = \inf\{t > \tau_{n-1} : \mathbf{M}((\tau_{n-1},t] \times [0,2^n]) = 1\} \tag{8.23}$$

Then define $X_t$ by

$$X_t := n, \quad \text{for } t \in [\tau_n, \tau_{n+1}]. \tag{8.24}$$

If $x_0 = 0$, the $n$-th jump occurs at time

$$\tau_n = \sum_{i=0}^{n} T_i,$$

where $(T_i : i \geq 0)$ is a family of independent random variables with exponential distribution with $\mathbb{E}T_i = 2^{-i}$, for $i \geq 0$. Hence,

$$\mathbb{E}\tau_n = \sum_{i=0}^{n} 2^{-i} \leq 2, \quad \text{for all } n. \tag{8.25}$$

Define $\tau_\infty = \sup_n \tau_n$. We prove now that $\tau_\infty$ is a finite random variable. Since $\tau_n$ is an increasing sequence,

$$\mathbb{P}(\tau_\infty > t) \leq \mathbb{P}(\cup_n \{\tau_n > t\}) = \lim_n \mathbb{P}(\tau_n > t) \tag{8.26}$$

because $\{\tau_n > t\}$ is an increasing sequence of events. We use now the Markov inequality to dominate the last expression with

$$\lim_n \frac{\mathbb{E}\tau_n}{t} \leq \frac{2}{t} \tag{8.27}$$

by (8.25). We have proved that $\mathbb{P}(\tau_\infty > t)$ goes to zero as $t \to \infty$, which implies that $\tau_\infty$ is a finite random variable.

Let $K(t)$ be the number of jumps of the process up to time $t$, that is,

$$K(t) := \sup\{n : \tau_n < t\}.$$

By definition,

$$K(t) > n \quad \text{if and only if} \quad \tau_n < t.$$

Hence, for all $t > \tau_\infty$, the process performs infinitely many jumps before time $t$.

**Definition 8.28** We say that the process $X_t$ explodes if

$$\mathbb{P}(\lim_{n\to\infty} \tau_n < \infty) > 0.$$

After a finite random time $\tau_\infty$ the process is not formally defined. But we can define an explosive process by adding a new state called $\infty$ with transition rates $q(\infty, x) = 0$ for all $x \in \mathcal{X}$.

If there are no explosions, that is, if

$$\mathbb{P}(\lim_{n\to\infty} \tau_n < \infty) = 0,$$

then, the rates $q(x,y)$ define univoquely a process which can be constructed as in Proposition 8.7.

A necessary and sufficient condition for the absence of explosions is the following.

**Proposition 8.29** *The process $X_t$ has no explosions if and only if*

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

*where $q(x) = \sum_y q(x,y)$ is the exit rate from state $x$.*

**Proof.** Omitted. □

The processes we study in this notes have no explosions.

## 8.3 Properties

Next result says that the process constructed in Proposition 8.7 is Markov. In other words, we prove that given the present, the future and the past are independent.

**Proposition 8.30** *The process $(X_t)$ defined by (8.6) satisfies*

$$\mathbb{P}(X_{t+u} = y \mid X_t = x, X_s = x_s, 0 \leq s < t) = P(X_{t+u} = y \mid X_t = x).$$

**Proof.** Indeed, in our construction, the process in the times after $t$ depend only on the bi-dimensional Poisson process $M$ in the region $\mathbf{M}((t,\infty) \times \mathbb{R}^+)$ and on the value assumed by $X_t$ at time $t$. Given $X_t$, this is independent of what happened before $t$. □

Given that at time $\tau_n$ the process $X_t$ is at state $x$, the time elapsed up to the next jump is an exponentially distributed random variable with mean $1/q(x)$; when the process decides to jump, it does so to state $y$ with probability

$$p(x,y) = \frac{q(x,y)}{q(x)}. \tag{8.31}$$

These properties are proven in the next two theorems.

**Theorem 8.32** *For a continuous time Markov process $X_t$,*

$$\mathbb{P}(\tau_{n+1} - \tau_n > t \mid X_{\tau_n} = x) \quad = \quad e^{-tq(x)}. \tag{8.33}$$

**Proof.** We prove the theorem for the case when the rates $q(x)$ are uniformly bounded by $\lambda \in (0, \infty)$. The general case can be proven using finite approximations. We use the representation of $\mathbf{M}(\cdot)$ in the strip $[0, \infty] \times [0, \lambda]$ of Theorem 7.54.

It is clear that the set $(\tau_n)_n$ is contained in the set $(S_n)_n$ defined in Theorem 7.54. Indeed, given $x_0 \in \mathcal{X}$, we can define

$$\begin{aligned} \tau_n &= \min\{S_k > \tau_{n-1} : W_k < q(x_{n-1})\} \\ K_n &= \{k : S_k = \tau_n\} \\ x_n &= \{y : W_{K_n} \in I(x_{n-1}, y)\}. \end{aligned} \tag{8.34}$$

This definition is a consequence of the representation of the bi-dimensional Poisson process of Theorem (7.54) and the construction of the Markov process using the Poisson process summarized in Definition 8.6.

The distribution of $\tau_{n+1} - \tau_n$, conditioned to $X_{\tau_n} = x$ is given by

$$\begin{aligned} &\mathbb{P}(\tau_{n+1} - \tau_n > t \mid X_{\tau_n} = x) \\ &\quad = \frac{\mathbb{P}(\tau_{n+1} - \tau_n > t, X_{\tau_n} = x)}{\mathbb{P}(X_{\tau_n} = x)}. \end{aligned} \tag{8.35}$$

Conditioning on the possible values $K_n$ may assume, the numerator can be written as

$$\sum_k \mathbb{P}(\tau_{n+1} - \tau_n > t, X_{\tau_n} = x, K_n = k)$$

$$
\begin{aligned}
&= \sum_k \mathbb{P}(\tau_{n+1} - S_k > t, X_{S_k} = x, K_n = k) \\
&= \sum_k \sum_\ell \mathbb{P}(S_{k+\ell} - S_k > t, W_{k+1} > q(x), \dots, W_{k+\ell-1} > q(x), \\
&\qquad\qquad\qquad W_{k+\ell} < q(x))\, \mathbb{P}(X_{S_k} = x, K_n = k) \\
&= e^{-q(x)} \sum_k \mathbb{P}(X_{S_k} = x, K_n = k).
\end{aligned}
$$

by Exercise 7.7.ii. Hence,

$$\frac{\mathbb{P}(\tau_{n+1} - \tau_n > t, X_{\tau_n} = x)}{\mathbb{P}(X_{\tau_n} = x)} = e^{-q(x)},$$

which finishes the proof. □

**Definition 8.36** The discrete process $(Y_n : n \in \mathbb{N})$ defined by $Y_n = X_{\tau_n}$ is called *skeleton* of the (continuous time) process $(X_t : t \in \mathbb{R}^+)$.

**Theorem 8.37** *The skeleton $(Y_n)$ of a continuous time process $(X_t)$ is a Markov chain with probability transitions $\{p(x,y), x, y \in \mathcal{X}\}$ given by* (8.31).

**Proof.** We want to prove that

$$\mathbb{P}(X_{\tau_{n+1}} = y \mid X_{\tau_n} = x) = p(x,y). \tag{8.38}$$

We use again the construction (8.34). Partitioning according with the possible values of $K_n$:

$$\mathbb{P}(X_{\tau_{n+1}} = y, X_{\tau_n} = x) = \sum_k \mathbb{P}(X_{\tau_{n+1}} = y, X_{\tau_n} = x, K_n = k) \tag{8.39}$$

By construction, the event $\{X_{\tau_{n+1}} = y, X_{\tau_n} = x, K_n = k\}$ is just

$$\bigcup_{l\ge 1}\{W_{k+1} > q(x), \dots, W_{k+\ell-1} > q(x), W_{k+\ell} \in I(x,y), X_{S_k} = x, K_n = k\},$$

where we used the convention that for $l = 1$ the event $\{W_{k+1} > q(x), \dots, W_{k+\ell-1} > q(x), W_{k+\ell} \in I(x,y)\}$ is just $\{W_{k+1} \in I(x,y)\}$. By independence between $(W_k)$ and $(S_k)$, expression (8.39) equals

$$\sum_k \sum_{l\ge 1} \mathbb{P}(W_{k+1} > q(x), \dots, W_{k+\ell-1} > q(x), W_{k+\ell} \in I(x,y)) \times \mathbb{P}(X_{S_k} = x, K_n = k)$$

But,

$$\sum_{l\geq 1} \mathbb{P}(W_{k+1} > q(x), \ldots, W_{k+\ell-1} > q(x), W_{k+\ell} \in I(x,y)) = \frac{q(x,y)}{q(x)}. \quad (8.40)$$

Hence, (8.39) equals

$$\frac{q(x,y)}{q(x)} \sum_k \mathbb{P}(X_{S_k} = x, K_n = k) \ = \ \frac{q(x,y)}{q(x)} \mathbb{P}(X_{\tau_n} = x), \quad (8.41)$$

which implies (8.38). □

## 8.4 Kolmogorov equations

It is useful to use the following matrix notation. Let $Q$ be the matrix with entries

$$q(x,y) \qquad \text{if } x \neq y \quad (8.42)$$

$$q(x,x) \ = \ -q(x) \ = \ -\sum_{y\neq x} q(x,y). \quad (8.43)$$

and $P_t$ be the matrix with entries

$$p_t(x,y) \ = \ \mathbb{P}(X_t = y \mid X_0 = x).$$

**Proposition 8.44 (Chapman-Kolmogorov equations)** *The following identities hold*

$$P_{t+s} \ = \ P_t P_s. \quad (8.45)$$

*for all* $s,t \geq 0$.

**Proof.** Compute

$$\begin{aligned} p_{t+s}(x,y) &= \mathbb{P}(X_{t+s} = y \mid X_0 = x) \\ &= \sum_z \mathbb{P}(X_s = z \mid X_0 = x)\mathbb{P}(X_{t+s} = y \mid X_s = z) \\ &= \sum_z p_s(x,z)p_t(z,y). \ \square \quad (8.46) \end{aligned}$$

**Proposition 8.47 (Kolmogorov equations)** *The following identities hold*

$$P_t' = QP_t \quad \textit{(Kolmogorov Backward equations)}$$

$$P_t' = P_tQ \quad \textit{(Kolmogorov Forward equations)}$$

*for all $t \geq 0$, where $P_t'$ is the matrix having as entries $p_t'(x,y)$ the derivatives of the entries of the matrix $P_t$.*

**Proof.** We prove the backward equations. Using the Chapman-Kolmogorov equations we have that for any $h > 0$:

$$\begin{aligned} p_{t+h}(x,y) - p_t(x,y) &= \sum_z p_h(x,z)p_t(z,y) - p_t(x,y) \\ &= (p_h(x,x) - 1)p_t(x,y)) + \sum_{z \neq x} p_h(x,z)p_t(z,y). \end{aligned}$$

Dividing by $h$ and taking $h$ to zero we obtain $p_t'(x,y)$ in the left hand side. To compute the right hand side, observe that

$$p_h(x,x) = 1 - q(x)h + o(h).$$

Hence

$$\lim_{h \to 0} \frac{p_h(x,x) - 1}{h} = -q(x) = q(x,x).$$

Analogously, for $x \neq y$

$$p_h(x,y) = q(x,y)h + o(h)$$

and

$$\lim_{h \to 0} \frac{p_h(x,y)}{h} = q(x,y).$$

This shows the Kolmogorov Backward equations. The forward equations are proven analogously. The starting point is the following way to write $p_{t+h}(x,y)$:

$$p_{t+h}(x,y) = \sum_z p_t(x,z)p_h(z,y). \; \square$$

## 8.5 Recurrence and transience

We start with a definition or hitting time analogous to the one for discrete chains.

**Definition 8.48** Let $T^{x\to y} = \inf\{t > \tau_1 : X_t^x = y\}$, be the first time the process starting at $x$ hits $y$.

The exigency $t > \tau_1$ is posed to avoid $T^{x\to x} \equiv 0$.

**Definition 8.49** We say that a state $x$ is

$$\text{transient, if } \mathbb{P}(T^{x\to x} = \infty) > 0; \tag{8.50}$$

$$\text{null recurrent, if } \mathbb{P}(T^{x\to x} = \infty) = 0 \text{ and } \mathbb{E}T^{x\to x} = \infty; \tag{8.51}$$

$$\text{positive recurrent, if } \mathbb{E}T^{x\to x} < \infty. \tag{8.52}$$

$$\text{recurrent, if it is null recurrent or positive recurrent.} \tag{8.53}$$

If the state space is finite, there are no null recurrent states.

**Definition 8.54** We say that a process is irreducible if for all states $x, y$, the probability to hit $y$ starting from $x$ in a finite time is positive:

$$P(T^{x\to y} < \infty) > 0.$$

**Theorem 8.55** *A process $(X_t)$ is irreducible if and only if its skeleton $(Y_n)$ is irreducible.*

*A state $x$ is recurrent (respectively transient) for the process $(X_t)$ if and only if $x$ is recurrent (respectively transient) for the skeleton $(Y_n)$.*

*If the exit rates are uniformly bounded from below and above, that is, if*

$$0 < \inf_x q(x); \ \sup_x q(x) < \infty,$$

*then $x$ is null recurrent for $(X_t)$, if and only if $x$ is null recurrent for the skeleton $(Y_n)$.*

**Proof.** The proof is straightforward. □

**Remark 8.56** It is clear that in the finite state space case the hypotheses of Theorem (8.55) are automatically satisfied.

The first part of the theorem says that a state is recurrent (respectively transient) for the continuous time process if and only if it is recurrent (respectively transient) for its skeleton. Hence, by Remark 8.56 recurrence and transience are equivalent for a process and its skeleton when the state space is finite.

When the state space is infinite it is possible to find examples of continuous time processes (violating the conditions of Theorem 8.55) and its skeleton with qualitative different behavior. We see now some of these examples.

**Example 8.57** In this example we present a process having null recurrent states which are positive recurrent for the skeleton. Consider the rates $q(x,0) = 1/2^x$, $q(x,x+1) = 1/2^x$. Hence $p(x,0) = 1/2$, $p(x,x+1) = 1/2$ and the skeleton is positive recurrent, because the return time to the origin is given by a geometric random variable with parameter $1/2$. On the other hand, since the mean jump time of each state $x$ is $2^x$,

$$\mathbb{E}T^{0\to 0} = \sum 2^x 1/2^{x+1} = \infty.$$

**Example 8.58** A simple (however explosive) example for which the states are positive recurrent for the continuous time process but null recurrent for its skeleton is given by the following rates: $q(x,0) = 1$, $q(x,x+1) = x^2$, $q(x,y) = 0$ otherwise. The transition probabilities of the skeleton are given by $p(x,0) = 1/(1+x^2)$ and $p(x,x+1) = x^2/(1+x^2)$. The mean return time to the origin of the skeleton is given by

$$\sum_x x \left(\frac{1}{1+x^2}\right) \prod_{y=1}^{x-1} \left(\frac{y^2}{1+y^2}\right).$$

We let the reader the proof that this sum is infinity. The mean return time to the origin for the continuous process is given by

$$\sum_x \left(\sum_{y=1}^{x} \frac{1}{y^2}\right) \left(\frac{1}{1+x^2}\right) \left(\prod_{y=1}^{x-1} \left(\frac{y^2}{1+y^2}\right)\right).$$

We let the reader the proof that this sum is finite.

## 8.6 Invariant measures

**Definition 8.59** We say that $\pi$ is an *invariant measure* for $(X_t)$ if

$$\sum_x \pi(x) p_t(x,y) = \pi(y) \tag{8.60}$$

$$\sum_x \pi(x) = 1 \tag{8.61}$$

that is, if the distribution of the initial state is given by $\pi$, then the distribution of the process at time $t$ is also given by $\pi$ for any $t \geq 0$. Sometimes we also use the term *stationary measure* to refer to an invariant measure.

**Theorem 8.62** *A measure $\pi$ is invariant for a process with rates $q(x,y)$ if and only if*

$$\sum_x \pi(x) q(x,y) = \pi(y) \sum_z q(y,z). \tag{8.63}$$

Condition (8.63) can be interpreted as a *flux condition*: the entrance rate under $\pi$ to state $y$ is the same as the exit rate from $y$. For this reason the equations (8.63) are called *balance equations.*

**Proof.** In matrix notation, a stationary measure can be seen as a row vector satisfying

$$\pi P_t = \pi.$$

We can differentiate this equation to obtain

$$\sum_x \pi(x) p'_t(x,y) = 0.$$

Applying Kolmogorov backward equations we get (8.63).

Reciprocally, equations (8.63) can be read as

$$\pi Q = 0.$$

Applying Kolmogorov backwards equations we get

$$(\pi P_t)' = \pi P_t' = \pi Q P_t = 0;$$

In other words, if the initial state of a process is chosen accordingly to the law $\pi$, the law of the process at any future time $t$ is still $\pi$. This is because $P_0$ is the identity matrix and $\pi P_0 = \pi$. □

The following result is analogous to Theorem 3.54. The novelty is that the result holds even when the skeleton is periodic. The conclusion is that continuous time processes are more "mixing" than discrete time processes.

**Theorem 8.64** *An irreducible process $(X_t)$ in a finite state space has a unique invariant measure $\nu$. Furthermore,*

$$\sup_{x,y} |\nu(y) - P_t(x,y)| < e^{-\gamma t},$$

*where*

$$\gamma = \min_{x,y} \left( \sum_{z \notin \{x,y\}} \min\{q(x,z), q(y,z)\} + q(x,y) + q(y,x) \right). \tag{8.65}$$

**Proof.** We couple two processes with the same transition rates matrix: $X_t$ with arbitrary initial state and $Y_t$ with initial state chosen according to the invariant measure $\nu$.

Assume first that the states are well ordered. Then, for each pair of disjoint states $x < y$, construct two families of disjoint subsets of $\mathbb{R}$, $\{I^x(y,z) : z \in \mathcal{X}\}$ and $\{I^y(x,z) : z \in \mathcal{X}\}$ such that $|I^x(y,z)| = q(y,z)$ and $|I^y(x,z)| = q(x,z)$.

Starting from the origin, construct a family of successive intervals $J^{x,y}(z)$ with lengths

$$|J^{x,y}(z)| = \min\{q(x,z), q(y,z)\}; \quad z \neq x, y.$$

After those, put an interval $I(x,y)$ with length $q(x,y)$, after it put an interval $I(y,x)$ with length $q(y,x)$. After this, put intervals $J^y(x,z)$ with length

$$|J^y(x,z)| = q(x,z) - \min\{q(x,z), q(y,z)\}; \quad z \neq x, y.$$

After this, put intervals $J^x(y,z)$ with lengths

$$|J^x(y,z)| = q(y,z) - \min\{q(x,z), q(y,z)\}; \quad z \neq x, y.$$

Now call

$$I^x(y,z) = J^x(y,z) \cup J^{x,y}(z); \quad I^y(x,z) = J^y(x,z) \cup J^{x,y}(z)$$

for $z \neq x, y$.

If $x = y$, just define successive intervals $I(x,z)$ with lengths $q(x,z)$; that is, $I^x(x,x) = \emptyset$ for all $x \in \mathcal{X}$.

Let

$$I^{x,y} = \left[\bigcup_{z \neq x,y} (I^x(y,z) \cup I^y(x,z))\right] \cup I(x,y) \cup I(y,x).$$

Assume

$$(X_0, Y_0) \;=\; (x_0, y_0) \tag{8.66}$$

Let $\tau_1$ be the first time the process $\mathbf{M}(\cdot)$ has a point with the first coordinate in the interval $I^{x_0,y_0}$:

$$\tau_1 = \inf\{t > 0 : \mathbf{M}(I^{x_0,y_0} \times [0,t]) > 0\}.$$

Let

$$x_1 = \begin{cases} z & \text{if } \inf\{t > 0 : \mathbf{M}(I^{x_0,y_0} \times [0,t]) > 0\} \\ & \quad = \inf\{t > 0 : \mathbf{M}(I^{y_0}(x_0,z) \times [0,t]) > 0\} \\ x_0 & \text{otherwise} \end{cases}$$

that is, $x_1$ is determined by the interval $I^{y_0}(x_0,z)$ realizing the infimum or it stays equal to $x_0$ if none of those intervals realize the infimum. Analogously,

$$y_1 = \begin{cases} z & \text{if } \inf\{t > 0 : \mathbf{M}(I^{x_0,y_0} \times [0,t]) > 0\} \\ & \quad = \inf\{t > 0 : \mathbf{M}(I^{x_0}(y_0,z) \times [0,t]) > 0\} \\ y_0 & \text{otherwise} \end{cases}$$

that is, $y_1$ is determined by the interval $I^{x_0}(y_0,z)$ realizing the infimum, or stays equal to $y_0$ if none of those intervals realizes the infimum. Let $\tau_n$ be the first time a point of the process $\mathbf{M}(\cdot)$ appears in the interval $I^{x_{n-1},y_{n-1}}$:

$$\tau_n = \inf\{t > \tau_{n-1} : \mathbf{M}(I^{x_{n-1},y_{n-1}} \times (\tau_{n-1}, t]) > 0\}.$$

Let

$$x_n = \begin{cases} z & \text{if } \inf\{t > \tau_{n-1} : \mathbf{M}(I^{x_{n-1},y_{n-1}} \times (\tau_{n-1}, t]) > 0\} \\ & \quad = \inf\{t > \tau_{n-1} : \mathbf{M}(I^{y_{n-1}}(x_{n-1}, z) \times (\tau_{n-1}, t]) > 0\} \\ x_{n-1} & \text{otherwise} \end{cases}$$

that is, $x_n$ is determined by the interval $I^{y_{n-1}}(x_{n-1}, z)$ realizing the infimum, or stays equal to $x_{n-1}$ if none of those intervals realizes the infimum. Analogously,

$$y_n = \begin{cases} z & \text{if } \inf\{t > \tau_{n-1} : \mathbf{M}(I^{x_{n-1},y_{n-1}} \times (\tau_{n-1}, t]) > 0\} \\ & \quad = \inf\{t > \tau_{n-1} : \mathbf{M}(I^{x_{n-1}}(y_{n-1}, z) \times (\tau_{n-1}, t]) > 0\} \\ y_{n-1} & \text{otherwise} \end{cases}$$

that is, $y_n$ is determined by the interval $I^{x_{n-1}}(y_{n-1}, z)$ realizing the infimum or stays equal to $y_{n-1}$ if none of those intervals realizes the infimum.

Now define

$$(X_t, Y_t) := (x_n, y_n) \ \text{ if } \ t \in [\tau_n, \tau_{n+1}). \tag{8.67}$$

**Proposition 8.68** *The process defined by* (8.66) *and* (8.67) *is a coupling of two Markov processes with transition rates $Q$ and initial state $(x_0, y_0)$.*

**Proof.** It is left as an exercise to the reader. □

An important fact of this construction is that when the coupled process is in the state $(x, y)$ and a point of the process $\mathbf{M}(\cdot)$ appears in the interval

$$(\cup_z (J^{x,y}(z)) \cup I(y, x) \cup I(x, y),$$

then both processes jump to the same state and from this moment on continue together for ever (*coalesce*). The length of this interval is exactly

$$\sum_{z \notin \{x,y\}} \min\{q(x, z), q(y, z)\} + q(x, y) + q(y, x).$$

Since for any couple $x, y$ these intervals have the origin as left point, the processes coalesce the first time a point of the process $\mathbf{M}(\cdot)$ appears in the intersection of those intervals. This intersection has length $\gamma$ defined by (8.65). □

The following is the convergence theorem under weaker hypotheses. It does not have speed of convergence.

**Theorem 8.69** *If the continuous time Markov process $(X_t)$ is positive recurrent and irreducible, then it admits a unique invariant measure $\pi$. Furthermore, for any initial state $x$,*

$$\lim_{t\to\infty} p_t(x,y) = \pi(y).$$

**Proof.** omitted. □

## 8.7 Skeletons

Assume the process $(X_t)$ and its discrete skeleton $(Y_n)$ defined by $Y_n = X_{\tau_n}$ are irreducible and positive recurrent. In Chapter 2 we saw that an invariant measure $\nu$ for the discrete skeleton $Y_n$ must satisfy the following system of equations

$$\sum_x \nu(x)p(x,y) = \nu(y)$$

On the other hand, the invariant measure $\pi$ for the process $X_t$ must satisfy

$$\sum_x \pi(x)q(x,y) = \pi(y)q(y)$$

This implies that $\nu$ is the invariant measure for $Y_n$ if and only if the measure $\pi$ defined by

$$\pi(x) = \frac{\nu(x)}{q(x)}\left(\sum_z \frac{\nu(z)}{q(z)}\right)^{-1} \tag{8.70}$$

is invariant for $X_t$. Intuitively, the different waiting time between jumps in the continuous time process require a correction in the invariant measure for the discrete time process which takes into account these differences. The second factor in the right hand side of (8.70) is just normalization to guarantee that $\pi$ is a probability.

As a corollary we have that if the exit rates do not depend on the state, then a measure is invariant for the continuous time process if and only if it is invariant for the discrete time one. That is, if $q(x) = q(y)$ for all $x, y \in \mathcal{X}$, then $\pi(x) = \nu(x)$ for all $x \in \mathcal{X}$.

## 8.8 Birth and death process

A birth and death process represents the growth (or extinction) of a population. The value $X_t$ represents the number of alive individuals of the population at time $t$. The rates of birth and death depend only on the number of alive individuals. That is,

$$q(x, x+1) \; = \; \lambda^+_x \quad \text{and} \quad q(x, x-1) = \lambda^-_x, \tag{8.71}$$

where $\lambda^+_x$, $\lambda^-_x$ are families of non-negative parameters. We use the balance equations (8.63) to look for conditions under which the process admits an invariant measure. We look for a vector $\pi$ satisfying the equations

$$\pi(0)q(0,1) \; = \; \pi(1)q(1,0) \tag{8.72}$$

$$\pi(x)q(x, x-1) + q(x, x+1)) \tag{8.73}$$

$$= \; \pi(x-1)q(x-1, x) + \pi(x+1)q(x+1, x), \tag{8.74}$$

for $x \geq 1$.

It is not difficult to conclude that for all $x \geq 0$,

$$\pi(x+1)\lambda^-_{x+1} - \pi(x)\lambda^+_x \; = \; 0, \tag{8.75}$$

where

$$\pi(x+1) \; = \; \frac{\lambda^+_x}{\lambda^-_{x+1}}\pi(x), \quad x \geq 0.$$

Hence, for all $x \geq 1$

$$\pi(x) \; = \; \frac{\lambda^+_0 \ldots \lambda^+_{x-1}}{\lambda^-_1 \ldots \lambda^-_x}\pi(0). \tag{8.76}$$

It is clear that $\pi(x)$ so constructed satisfies (8.63). To satisfy (8.61) we need $\pi(0) > 0$. Hence, there will be a solution if

$$\sum_{x\geq 1} \frac{\lambda_0^+ \dots \lambda_{x-1}^+}{\lambda_1^- \dots \lambda_x^-} < \infty. \tag{8.77}$$

If (8.77) is satisfied, then we can define

$$\pi(0) \;=\; \left( \sum_{x\geq 1} \frac{\lambda_0^+ \dots \lambda_{x-1}^+}{\lambda_1^- \dots \lambda_x^-} \right)^{-1},$$

and $\pi(x)$ inductively by (8.76).

## 8.9 Exercises

**Exercise 8.1** Let $X_t$ be a continuous time process in $\mathcal{X} = \{0, 1\}$ with rates $q(0,1) = 1$, $q(1,0) = 2$. Compute the Kolmogorov equations and find $p_t(1,0)$.

**Exercise 8.2** Prove that the variable $\tau_\infty$ defined in the Example (8.12) is finite with probability one. Hint: observe that, by the Markov inequality

$$\mathbb{P}(\tau_\infty > K) \leq \frac{\mathbb{E}\tau_\infty}{K}.$$

Then use the fact that

$$\lim_n \mathbb{P}(\tau_n > K) = \mathbb{P}(\tau_\infty > K).$$

**Exercise 8.3** Show that the pure birth process constructed in (8.21) satisfies conditions (8.19) and (8.20).

**Exercise 8.4** Coupling of pure birth processes. Prove that it is possible to couple two pure birth processes $(X_t^1)$ and $(X_t^2)$ with rates $\lambda_1^- \geq \mu_2$, respectively, in such a way that if $X_0^1 \geq X_0^2$, then $X_t^1 \geq X_t^2$.

**Exercise 8.5** Show identity(8.40).

**Exercise 8.6** Prove that the process presented in Example (8.58) is positive recurrent and that its skeleton is null recurrent.

**Exercise 8.7** Prove that the marginals of the joint process defined in (8.67) have the right distribution of the process. That is, prove that

$$\mathbb{P}(X_{t+h} = y \mid X_t = x) = q(x,y)h + o(h) \quad (8.78)$$

$$\mathbb{P}(Y_{t+h} = y \mid Y_t = x) = q(x,y)h + o(h) \quad (8.79)$$

for all $t \geq 0$.

**Exercise 8.8** Let $X_t$ be the process in $\mathcal{X} = \{1,2,3\}$ with transition rates matrix $Q$ defined by

$$Q = \begin{pmatrix} -4 & 1 & 3 \\ 2 & -3 & 1 \\ 1 & 1 & -2 \end{pmatrix} \quad (8.80)$$

(The diagonal is just the sum of the exit rates of the corresponding states with a minus sign.) Construct the coupling $(X_t, Y_t)$ given by (8.67). Compute the value of $\gamma$.

**Exercise 8.9** For the process introduced in Exercise (8.8) compute the transition probabilities of the skeleton, the invariant measure for the continuous time process and for the skeleton and verify (8.70).

**Exercise 8.10** Consider a system consisting of one queue and one server. The clients arrive at rate $\lambda^+$ in groups of two. The system serves one client at a time at rate $\lambda^-$. Assume the system has a maximal capacity of 4 clients. That is, if at some moment there are 3 clients and a group of two arrives, then only one of those stays in the system and the other is lost. The space state is $\mathcal{X} = \{0,1,2,3,4\}$ and the transition rate matrix of such a system is given by

$$Q = \begin{pmatrix} -\lambda^+ & 0 & \lambda^+ & 0 & 0 \\ \lambda^- & -\lambda^- + \lambda^+ & 0 & \lambda^+ & 0 \\ 0 & \lambda^- & -\lambda^- + \lambda^+ & 0 & \lambda^+ \\ 0 & 0 & \lambda^- & -\lambda^- + \lambda^+ & \lambda^+ \\ 0 & 0 & 0 & \lambda^- & -\lambda^- \end{pmatrix} \quad (8.81)$$

(a) Establish the balance equations and find the invariant measure.

(b) Compute the probability that a group of two people arrive to the system and none of them can stay in it.

(c) Compute the mean number of clients in the system when the system is in equilibrium.

(d) Compute the mean time a client stays in the system.

**Exercise 8.11** Consider a queue system M/M/$\infty$, that is, arrivals occur according to a Poisson process of rate $\lambda^+$ and service times are exponentially distributed with rate $\lambda^-$ but now the system has infinitely many servers (that is all clients start service upon arrival). Solve items of Exercise (8.10) in this case.

**Exercise 8.12** Consider a population with $m$ individuals. At time zero there are $k$ "infected" individuals and $m - k$ non infected. An infected individual heals after an exponentially distributed time with parameter $\lambda^-$. If there are $k$ infected individuals, the rate for each one of the remained $m-k$ non infected individuals to get infected is $\lambda^+(k + 1)$.

(a) Establish the balance equations.

(b) For $m = 4$, $\lambda^+ = 1$ and $\lambda^- = 2$ compute the invariant measure.

(c) Compute the average number of infected individuals under the invariant measure.

(d) Compute the probability of the event "all individuals are infected" under the invariant measure.

## 8.10 Comments and references

In this chapter we have used the bi-dimensional Poisson process to construct a continuous time Markov process in a countable state space. This is a natural extension of the projection method used in the previous chapter. The method allows to couple as in discrete time.

# Bibliography

[1] K. B. Athreya, P. Ney (1978) A new approach to the limit theory of recurrent Markov chains. *Trans. Amer. Math. Soc.* **245** 493–501.

[2] H. Berbee (1987) Chains with complete connections: Uniqueness and Markov representation. *Prob. Th. Rel. Fields*, 76:243–53.

[3] D. Blackwell (1953) Extension of a renewal theorem. *Pacific J. Math.* 3. 315–320.

[4] X. Bressaud, R. Fernández, and A. Galves (1999a) Speed of $\bar{d}$-convergence for Markov approximations of chains with complete connections. a coupling approach. *Stoch. Proc. and Appl.*, 83:127–38.

[5] X. Bressaud, R. Fernández, and A. Galves (1999b) Decay of correlations for non Hölderian dynamics. a coupling approach. *Elect. J. Prob.*, 4. (`http://www.math.washington.edu/~ejpecp/`).

[6] F. Comets, R. Fernández, and P. A. Ferrari (2000) Processes with long memory: regenerative construction and perfect simulation. Preprint. `http://www.ime.usp.br/~pablo/abstracts/cff.html`

Dobrushin, R. L. (1971) Markov processes with a large number of locally interacting components: Existence of a limit process and its ergodicity. (Russian) *Problemy Peredači Informacii* **7**, no. 2, 70–87.

[7] W. Doeblin (1938a) Esposé de la théorie des chaînes simple constantes de Markov à un nombre fini d'états *Rev. Math. Union Interbalkan.* **2**, 77–105.

[8] W. Doeblin (1938b) Sur deux problémes de M. Kolmogoroff concernant les chaînes dénombrables. *Bull. Soc. Math. Fr.*, 66:210–20.

[9] R. Dobrushin (1956) Central limit theorem for non-stationary Markov chains. I. (Russian) *Teor. Veroyatnost. i Primenen. 1*, 72–89.

[10] R. L. Dobrushin (1968a) Gibbsian random fields for lattice systems with pairwise interactions. *Funct. Anal. Appl.* **2** 4:292–301.

[11] R. L. Dobrushin (1968b) The description of random fields by means of conditional probabilities and conditions of its regularity. *Theory Prob. Appl.* **13** 2:197–224.

[12] Durrett, Richard (1988) *Lecture notes on particle systems and percolation*. The Wadsworth & Brooks/Cole Statistics/Probability Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA.

[13] Durrett, Rick (1995) *Ten lectures on particle systems*. Lectures on probability theory (Saint-Flour, 1993), 97–201, Lecture Notes in Math., 1608, Springer, Berlin.

[14] W. Feller (1968) *An introduction ot Probability Theory and its Applications*. John Wiley & Sons, Ltd.

[15] P. A. Ferrari and A. Galves (1997) *Acoplamento em processos estocásticos* (Portuguese). SBM, IMPA, Rio de Janeiro.
`http://www.ime.usp.br/~pablo/abstracts/libro.html`

[16] P. A. Ferrari, A. Maass, S. Martínez, and P. Ney (2000) Cesàro mean distribution of group automata starting from measures with summable decay. To be published in *Ergodic Th. Dyn. Syst.*
`http://xxx.lanl.gov/abs/math.PR/9912135`

[17] Foss and Tweedie (2000) Perfect simulation and backward coupling. *Stochastic Models*. `http://www.stats.bris.ac.uk/MCMC/`

[18] A. Frigessi, J. Gasemyrand and H. Rue (1999) Antithetic Coupling of Two Gibbs Sampler Chains. Preprint.
http://www.math.ntnu.no/~hrue/anti-mcmc/abstract/abstract.html

[19] N. L. Garcia (1995) Birth and death processes as projections of higher dimensional Poisson processes. *Adv. Appl. Probab.* **27**, 911-930.

[20] O. Häggström (2000) *Finite Markov chains and algorithmic applications* Preprint.

[21] T. E. Harris (1955) On chains of infinite order. *Pacific J. Math.*, 5:707–24.

[22] T. E. Harris (1956) The existence of stationary measures for certain Markov processes. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability,* 1954–1955, vol. II, pp. 113–124. University of California Press, Berkeley and Los Angeles.

[23] T. E. Harris (1972) Nearest-neighbor Markov interaction processes on multidimensional lattices. *Advances in Math. 9, 66–89.*

[24] Harris, T. E. (1974) Contact interactions on a lattice. *Ann. Probability* **2**, 969–988.

[25] M. Iosifescu (1961) On the asymptotic behaviour of chains with complete connections. *Comunicǎrile Acad. RPR*, 11:619–24.

[26] M. Iosifescu (1992) A coupling method in the theory of dependence with complete connections according to Doeblin. *Rev. Roum. Math. Pures et Appl.*, 37:59–65.

[27] M. Iosifescu, S. Grigorescu, Dependence with complete connections and its applications. *Cambridge Tracts in Mathematics*, **96**. Cambridge University Press, Cambridge, 1990.

[28] F. P. Kelly (1979) *Reversibility and stochastic networks.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Ltd., Chichester.

[29] T. G. Kurtz (1989) Stochastic processes as projections of Poisson random measures. Special invited paper at IMS meeting. Washington D.C.. Unpublished.

[30] S. P. Lalley (1986) Regeneration representation for one-dimensional Gibbs states. *Ann. Prob.*, 14:1262–71.

[31] C. Kipnis and C. Landim (1999) *Scaling limits of interacting particle systems.* Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 320. Springer-Verlag, Berlin.

[32] F. Ledrappier (1974) Principe variationnel et systèmes dynamiques symboliques. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 30:185–202.

[33] T.M.Liggett (1985) *Interacting particle systems*. Springer-Verlag, New York-Berlin.

[34] T.M.Liggett (1999) *Stochastic interacting systems: contact, voter and exclusion processes.* Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 324. Springer-Verlag, Berlin.

[35] T. Lindvall (1992) *Lectures on the coupling method.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York.

[36] J. Neveu (1977) *Processus ponctuels*. École d'Été de Probabilités de Saint-Flour, VI—1976, pp. 249–445. Lecture Notes in Math., Vol. 598, Springer-Verlag, Berlin.

[37] E. Nummelin and P. Ney (1993) Regeneration for chains with infinite memory. *Prob. Th. Rel. Fields*, 96:503–20.

[38] E. Nummelin (1978) A splitting technique for Harris recurrent Markov chains. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 43:309–18.

[39] J. G. Propp and D. B. Wilson (1996) Exact sampling with coupled Markov chains and applications to statistical mechanics. In *Proceedings of the Seventh International Conference on Random Structures and Algorithms (Atlanta, GA, 1995)*, volume 9, pages 223–252.

[40] S. M. Ross (1983) *Stochastic processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. Lectures in Mathematics, 14. John Wiley & Sons, Inc., New York.

[41] F. Spitzer (1970) Interaction of Markov processes. *Advances in Math.* **5** 246–290.

[42] H. Thorisson (2000) *Coupling, stationarity and regeneration* Springer.

[43] L. N. Vasershtein (1969) Markov processes over denumerable products of spaces describing large system of automata. *Problems of Information Transmission* **5**, no. 3, 47–52; translated from *Problemy Peredači Informacii* **5**, no. 3,64–72 (Russian).

[44] D. B. Wilson (1998) Annotated bibliography of perfectly random sampling with Markov chains. In D. Aldous and J. Propp, editors, *Microsurveys in Discrete Probability*, volume 41 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 209–220. American Mathematical Society. Updated versions can be found at `http://dimacs.rutgers.edu/~dbwilson/exact`.