Coupling, renewal and perfect simulation of chains of infinite order

R. Fernández P. A. Ferrari A. Galves

August 10, 2001

Contents

Chapter 1

Introduction

These notes are dedicated to Ted Harris who 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.

The purpose of these notes is to explain in an elementary way how coupling and regeneration can be used to construct and study chains of infinite order. These are stochastic processes taking values on a finite alphabet in which the choice of each new symbol depends on the whole past history. This is in contrast with Markov chains, in which the choice depends on only a fixed finite number of preceding values. Our 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.

Chains of infinite order seem to have been first studied by Onicescu and Mihoc (1935a) who called them *chains with complete connections* (*chaînes à liaisons complètes*). Their study was soon taken up by Doeblin and Fortet (1937) who proved the first results on speed of convergence towards the invariant measure. The name chains of infinite order was coined by Harris (1955) . We refer the reader to Iosifescu and Grigorescu (1990) for a complete survey.

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. For a description of Doeblin's contributions to probability theory we refer the reader to Lindvall (1991).

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), Harris (1972) and the sovietic school of Dobrushin, Toom, Piatevsky-Shapiro, Vaserstein and others. This names gave rise to a new area in stochastic processes developed extensively by Harris, Holley, Liggett, Durrett, Griffeath, Kipnis and others. We refer the interested reader to the books by Liggett (1985), (1999) and Kipnis and Landim (1999) for recent developments in the field. Liggett (1994) reviews the use of the coupling technique for interacting Markov systems.

Our constructive approach comes directly from the *graphical construction* of interacting particle systems introduced by Harris (1972, 1978). The way we couple chains can be traced back to Dobrushin (1956), even when there is no coupling in his paper. A coupling approach related to what to do in Chapter 8 has been used by Marton (1996).

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.

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

$$\begin{array}{rcl} \mathbb{P}(X = 1) & = & p \;=\; 1 - \mathbb{P}(X = 0) \\ \mathbb{P}(Y = 1) & = & q \;=\; 1 - \mathbb{P}(Y = 0) \\ X & \leq & Y. \end{array}$$

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

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

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

$$X_n := \mathbf{1}\{U_n > h(X_{n-1})\} \tag{1.1}$$

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

$$\mathbb{P}(X_n = 1 \mid U_n \in [0, h(0)] \cup [h(1), 1]) = \frac{1 - h(1)}{h(0) + 1 - h(1)} \tag{1.2}$$

independently of the past. Definition (1.1) 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 $\tau(n) := \max\{k \leq n : U_k \in [0, h(0)] \cup [h(1), 1]\}$, 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:

$$\mathbb{P}(X_n = y \mid X_{n-1} = x) = Q(x, y) \tag{1.3}$$

where

$$\begin{array}{ll} Q(0,0) = h(0) & Q(0,1) = 1 - h(0) \\ Q(1,0) = h(1) & Q(1,1) = 1 - h(1)\,. \end{array} \tag{1.4}$$

Processes with this kind of property are called *Markov chains*. The principal consequence of construction (1.1) is that the pieces of the process between two regeneration times are independent random vectors (of random length). We use this approach to construct *perfect simulation* algorithms not only for Markov chain but, more generally, for chains of infinite order, with a suitable memory-loss rate.

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) .

In these notes we adopt the *graphic construction* philosophy introduced by Ted Harris to deal with interacting particle systems. Our first elementary systematic presentation of Harris' point of view is contained in the booklet *Acoplamento em Processos Estocásticos* in Portuguese for a mini-course tow of us offered at the XXI Coloquio Brasileiro de Matemática, held in Rio de Janeiro in July of 1997 (Ferrari and Galves 1997), followed by *Construction of Stochastic processes, Coupling and Regeneration* (Ferrari and Galves 2000), notes for the *XIII Escuela Venezolana de Matemáticas.* In these references, Markov processes were the main concern. In the present set of lectures we focus instead on recent results on chains of infinite order presented in Bressaud, Fernández and Galves (1999a, 1999b) and Comets, Fernández and Ferrari (2000). We refer the reader to these papers for further technical details and more extensive references.

Acknowledgements

A number of collaborators and colleagues helped us to learn and understand many of the issues of thee notes. We like to express our warm thanks to Miguel Abadi, Xavier Bressaud, Jean-René Chazottes, Pierre Collet, Francis Comets, Denise Duarte, Davide Gabriele, Jesús García, Daniela Guiol, Nancy Lopes Garcia, Alejandro Maass, Grégory Maillard, Servet Martinez, Peter Ney Bernard Schmitt and Paul Shields.

This final version of the notes has benefited from the sharp comments, questions and criticism of the audience of the V EBP. We thank specially Christian Borgs, David Brillinger, Jennifer Chayes, Tom Kurtz, Steve Lalley, Jesper Møller, Errico Presutti, ... We also thank the enthusiastic students that pointed out a number of errors and imprecisions.

The authors thank FINEP, FAPESP, CNPq and an agreement USP-COFECUB for support during the writing of these notes.

We owe special thanks to the Zentrum für interdisziplinäire Forschung (ZiF) of the University of Bielefeld who offered us support and an extraordinary scientific and human environment during the completion of these notes.

Chapter 2

Basic definitions

2.1 Simulation algorithms

We consider an evolution in discrete time $\mathbb{Z}$ taking values in a finite *alphabet* $\mathcal{A}$. The evolution is random, that is, its possible realizations are described by a family of random variables $(X_n)_{n\in\mathbb{Z}}$, with images in $\mathcal{A}$, defined on a certain measure space.

The existence of a probability space in which a given stochastic process can be defined is a basic issue in probability theory. One of the advantages of Harris' constructive approach is that it shows that the processes considered in these notes, and many others, can be rigorously constructed *using only a double infinite sequence of independent random variables uniformly distributed in* $[0,1]$. The existence of such a sequence is the only measure-theoretical fact we will need in these lectures. This sequence will be denoted $(U_n,\, n\in\mathbb{Z})$. In some applications the uniform variables will be relabelled so that each U_n will in fact correspond to a N-tuple of independent random variables $U_n^{(1)},\cdots,U_n^{(N)}$, with N fixed. People that do not feel comfortable with measure theory should simply think these variables as the outcomes of a random number generator in a computer simulation.

The only probability space we shall be concern with is the one in which the variables (U_n) are defined. Let us call it $(\Omega,\mathcal{F},\mathbb{P})$ and use $\mathbb{E}$ for the corresponding expectation.

The value of X_n is interpreted as the "state" of the process at "time" n. The outcomes of such an evolution corresponds to strings of symbols $x = (x_n)_{n\in\mathbb{Z}} \in \mathcal{A}^{\mathbb{Z}}$ which we shall call a *path* of the process. The theory is developed purely in terms of the path space $\mathcal{A}^{\mathbb{Z}}$, and the specific choice for the space of definition of the X_n plays no role. Formally this is because evolutions are described in terms of joint *laws* of the variables X_n.

The traditional way to introduce a *stochastic process* is starting from the family of *joint probability distributions* or, equivalently, by a probability measure on $\mathcal{A}^{\mathbb{Z}}$ corresponding to the joint laws. This has two drawbacks. First, the existence of a process so defined is not an easy matter. Second, these measures are seldomly directly accesible. Rather, the starting objects are conditional probabilities of the form

$$P(X_{n+\ell} = x_{n+\ell}, \cdots, X_n = x_n | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \cdots) \;, \tag{2.1}$$

which are either explicitly defined from modelling considerations or estimated from actual outputs. For process of truly infinite order, some care must be taken to give a rigorous meaning to (2.1) because the conditionning usually refers to an event of probability zero. A possible formalization, adopted for instance by Lalley (1986) , is to define a process as a measure for which the conditional probabilities on finite pasts, $P(X_{n+\ell} = x_{n+\ell}, \cdots, X_n = x_n | X_{n-1} = x_{n-1}, \cdots, X_{n-s} = x_{n-s})$ have a well defined limit as $s \to -\infty$. This is a natural setup when describing experiences started at some initial time before which there is no meaningful past [eg. in Onicescu and Mihoc (1935a)]. The limit $s \to -\infty$ corresponds to pushing this initial time to the remote past. The only conceptual disadvantage of this approach is that the limit $s \to -\infty$ depends, in principle, on the process considered, that is on the inaccessible joint measure of the random variables. This makes the approach less direct from the computational point of view.

We introduce now the formal definitions necessary for Harris' approach. We leave for Section 2.2 the presentation of the "traditional" formalism in terms of objects like (2.1). The relation between both appraoches is discussed in Section 2.3.

To state the necessary definitions we need some notation. For $k \leq n \in \mathbb{Z}$ let x_k^n denote the sequence $x_k, \cdots, x_n$, and let $\mathcal{A}_k^n$ denote the set of such sequences. Likewise, let $x_{-\infty}^n$ denote the sequence $(x_i)_{i \leq n}$ —*histories* up to time n— and $\mathcal{A}_{-\infty}^n$ the corresponding space. Full sequences will be denoted without sub or superscripts, $x \in \mathcal{A}$. The notation $y_{n+1}^m x_k^n$ indicates the sequence that takes values $x_k, \cdots, x_n, y_{n+1}, \cdots, y_m$.

Remark 2.2 Before starting with the definitions, let us insist that in these notes we do not wish to make an issue of measurability. We shall mention the word "measurable" only sparingly and always in a context such that: (i) the σ-algebra in question is the natural one, and (ii) the measurability requirement is practically a formality, as every function used for the corresponding application will invariably be measurable. Readers can safely ignore measurability issues, and concentrate instead on the algorithmic aspects of our constructions and proofs.

Let us now define the central objects of our approach.

Definition 2.3 *A* **simulation algorithm** *is a family of measurable functions $(f_n)_{n\in\mathbb{Z}}$, where $f_n : [0,1] \times \mathcal{A}_{-\infty}^n \longrightarrow \mathcal{A}$.*

For completeness (but see Remark 2.2), let us state for the first and last time that the σ-algebra

- of $[0,1]$ is the Lebesgue σ-algebra,
- of (finite or infinite) products of $\mathcal{A}$ is the product of the discrete σ-algebra of $\mathcal{A}$,
- of products of these spaces is the corresponding product σ-algebra.

Definition 2.4 *A simulation algorithm (f_n) is* **time-homogeneous** *if the functions f_n coincide up to a shift. That is, if $x_{-\infty}^n \in \mathcal{A}_{-\infty}^n$ and $y_{-\infty}^{n+1} \in \mathcal{A}_{-\infty}^{n+1}$ are such that $x_i = y_{i+1}$ for $i \leq n$, then*

$$f_{n+1}(u, y_{-\infty}^{n+1}) \;=\; f_n(u, x_{-\infty}^n)\;. \tag{2.5}$$

In this case, we will eliminate the subscript from f_n.

Definition 2.6 *A* **stochastic process** *with alphabet $\mathcal{A}$ is a sequence of $\mathcal{A}$-valued random variables (=measurable functions) $(X_n)_{n\in\mathbb{Z}}$ defined in our one and only space $(\Omega, \mathcal{F}, \mathbb{P})$. The process is* **stationary** *if $\mathbb{P}(X_n^{n+k} = x_0^k)$ is independent of n for each $k \in \mathbb{N}$ and each $x_0^k \in \mathcal{A}_0^k$.*

Definition 2.7 *A* **stochastic process defined by the simulation algorithm** *$(f_n)_{n\in\mathbb{Z}}$ is a sequence of random variables $(X_n)_{n\in\mathbb{Z}}$ such that*

$$X_n \;=\; f_n(U_n, X_{-\infty}^{n-1}) \; . \tag{2.8}$$

The variable U_n is, in general, a finite family $U_n^{(1)}, \cdots, U_n^{(N)}$ of uniform random variables, with N fixed. All the random variables $U_j^{(i)}$ are independent.

Definition 2.9 *A stochastic process is a* **Markov chain** *if the f_n are local in their second coordinate, that is if there exist a fixed k, such that $f_n(u, y_{-\infty}^n) = f_n(u, x_{-\infty}^n)$ whenever $x_{n-k}^n = y_{n-k}^n$. The integer k is called the* **order** *of the Markov chain.*

Chains of infinite order are more general process for which there may exist no such k. They are usually required to satisfy some continuity and non-nullness hypotheses. We defer formal definitions to Chapter 3.

Prescription (2.8) is not enough to construct the process. We need a starting past from which to apply it iteratively.

Definition 2.10 *For $\ell \in \mathbb{Z}$ and $z_{-\infty}^\ell \in \mathcal{A}_{-\infty}^\ell$, the* **stochastic process with fixed past $z_{-\infty}^\ell$ defined by the simulation algorithm** *(f_n) is the sequence of random variables $(X_n[z_{-\infty}^\ell])_{n\in\mathbb{Z}}$ defined by*

$$\begin{aligned} X_n[z_{-\infty}^\ell] &= z_n && \text{for } n \le \ell \;, \\ X_{\ell+1}[z_{-\infty}^\ell] &= f_{\ell+1}(U_{\ell+1}, z_{-\infty}^\ell) && \text{and} \\ X_n[z_{-\infty}^\ell] &= f_n(U_n, X_{\ell+1}^{n-1}[z_{-\infty}^\ell]\, z_{-\infty}^\ell) && \text{for } n > \ell + 1 \;. \end{aligned} \tag{2.11}$$

While these fixed-past processes $(X_n[z_{-\infty}^\ell])_{n>\ell}$ are always well defined, they are not processes in the sense of Definition 2.7 because they verify

(2.8) only for times larger than ℓ. The *existence problem* of the theory of stochastic processes is, precisely, to obtain process without a fixed past for the given algorithm (f_n). That is, to determine variables $(X_n)_{n\in\mathbb{Z}}$ such that $X_n = f_n(U_n, X_{-\infty}^{n-1})$ *for all* $n \in \mathbb{Z}$. A second central issue in the theory of stochastic processes, is the *uniqueness problem*, namely whether there exist a unique such process $(X_n)_{n\in\mathbb{Z}}$ or several (phase transitions!). The approach we shall use here to solve the existence problem is to construct the function f in such a way that for each realization of the uniform random variables (U_n) a realization of the process can be constructed in any finite interval. We will show that this construction coincides with the process obtained as a limit of fixed-past processes. Furthermore, for the processes considered in these notes, uniqueness corresponds to such a limit being independent of the fixed past chosen. We shall use two main tools to analyze limits of fixed-past processes and their insensitivity to the past: (1) regeneration schemes, and (2) coupling techniques. The former schemes are the subject of next chapter. The formal definition of the notion of coupling will be discussed in Section 2.4.

2.2 Transition probabilities

To make the connection with the traditional approach, based on the objects (2.1), let us briefly formalize the basic definitions on which the latter relies.

Definition 2.12 *A* **system of transition probabilities** *is a family* $\{P_n(\,\cdot\,|\,\cdot\,) : n \in \mathbb{Z}\}$ *of functions* $P_n : \mathcal{A} \times \mathcal{A}_{-\infty}^{n-1} \longrightarrow [0,1]$*, such that the following conditions hold for each* $n \in \mathbb{Z}$*:*

(i) Measurability*: For each* $x_n \in \mathcal{A}$ *the function* $P_n(x_n|\,\cdot\,)$ *is measurable with respect to the product* σ*-algebra.*

(ii) Normalization*: For each* $x_{-\infty}^{n-1} \in \mathcal{A}_{-\infty}^{n-1}$

$$\sum_{x_n\in\mathcal{A}} P_n(x_n|x_{-\infty}^{n-1}) \;=\; 1 \;. \tag{2.13}$$

In the following definition we consider *excepcionally* an abstract probability space that nevertheless we denote $(\Omega, \mathcal{F}, \mathbb{P})$ as before.

Definition 2.14 *A stochastic process defined on* $(\Omega, \mathcal{F}, \mathbb{P})$ *is* **consistent** *with a system of transition probabilities* (P_n) *if*

$$\mathbb{P}(X_n = x_n | X_{-\infty}^{n-1} = x_{-\infty}^{n-1}) \;=\; P_n(x_n | x_{-\infty}^{n-1}) \tag{2.15}$$

for all $n \in \mathbb{Z}$, $x \in \mathcal{A}^{\mathbb{Z}}$.

Equation (2.15) means that the functions P_n are *regular versions of the conditional probabilities with respect to the natural filtration* $\mathcal{F}_n = \sigma(X_{-\infty}^{n})$. Equivalently, a stochastic process is consistent with a system of transition probabilities (P_n) iff

$$\mathbb{E}\Big[g(X_{-\infty}^{n})\Big] \;=\; \mathbb{E}\Big[\sum_{y_n \in \mathcal{A}} g(y_n\, X_{-\infty}^{n-1})\, P_n(y_n | X_{-\infty}^{n-1})\Big] \tag{2.16}$$

for every $n \in \mathbb{Z}$ and g measurable with respect to $\mathcal{F}_n$.

The transition probabilities of Definition 2.12 can be thought as next-move transition probabilities. They can be used to construct the ℓ-move transitions ($\ell \geq 1$) probabilities

$$P_{[n,n+\ell]}(x_n^{n+\ell} | x_{-\infty}^{n-1}) \;:=\; \prod_{i=1}^{\ell} P_{n+i}(x_{n+i} | x_{-\infty}^{n+i-1})\;. \tag{2.17}$$

[We adopt the convention $P_{[n,n]} := P_n$.] These transitions satisfy the consistency condition

$$\sum_{x_n^{n+\ell} \in \mathcal{A}_n^{n+\ell}} P_{[n,n+\ell]}(x_n^{n+\ell} | x_{-\infty}^{n-1}) \sum_{y_{n+i}^{n+j} \in \mathcal{A}_{n+i}^{n+j}} g(y_{n+i}^{n+j} x_{-\infty}^{n+i-1})\, P_{[n+i,n+j]}(y_{n+i}^{n+j} | x_{-\infty}^{n+i-1})$$

$$= \sum_{x_n^{n+\ell} \in \mathcal{A}_n^{n+\ell}} g(x_{-\infty}^{n+j}) P_{[n,n+\ell]}(x_n^{n+\ell} | x_{-\infty}^{n-1}) \tag{2.18}$$

for all $n \in \mathbb{Z}$, $i, j, \ell \in \mathbb{N}$, $0 \le i \le j \le \ell$, $x \in \mathcal{A}$ and all $\mathcal{F}_{n+j}$-measurable functions f. Furthermore, (2.16) implies that

$$\mathbb{E}\Big[g(X_{-\infty}^{n+\ell})\Big] \;=\; \mathbb{E}\Big[\sum_{y_n^{n+\ell} \in \mathcal{A}_n^{n+\ell}} g(y_n^{n+\ell} X_{-\infty}^{n-1})\, P_{[n,n+\ell]}(y_n^{n+\ell} | X_{-\infty}^{n-1})\Big] \tag{2.19}$$

for every $n \in \mathbb{Z}$, $\ell \in \mathbb{N}$, and g measurable with respect to $\mathcal{F}_{n+\ell}$. The verification of formulas (2.17)–(2.19) is left as an straightforward exercise to the reader. Condition (2.18) implies that the kernels $P_{[n,n+\ell]}(\,\cdot\,|\,\cdot\,)$ constitute the one-sided analogous of a statistical mechanical *specification*, while identities (2.19) are the analogous of the DLR equations [see, for instance, Georgii (1988) for the statistical mechanical framework].

The basic mathematical problem of the theory of stochastic processes is, precisely, to construct and characterize the processes consistent with a given system of transition probabilities. The comments of the end of Section 2.1 can be transcribed in this framework in a natural way. In particular, we can transcribe notions related with fixed pasts.

Definition 2.20 *Given a system of transition probabilities $\{P_n(\,\cdot\,|\,\cdot\,) : n \in \mathbb{Z}\}$, an $\ell \in \mathbb{Z}$ and a $z_{-\infty}^{\ell} \in \mathcal{A}_{-\infty}^{\ell}$, the* **system of transition probabilities with fixed past** $z_{-\infty}^{\ell}$ *is the family of functions*

$$P_n^{z_{-\infty}^{\ell}} : \mathcal{A} \times \mathcal{A}_{-\infty}^{n-1} \longrightarrow [0,1]\,, \tag{2.21}$$

defined as

$$P_n^{z_{-\infty}^{\ell}}(x_n | x_{\ell+1}^{n-1}) \;=\; \begin{cases} P_n(x_n | x_{\ell+1}^{n-1}\, z_{-\infty}^{\ell})\, \mathbf{1}[x_{-\infty}^{\ell} = z_{-\infty}^{\ell}] & \text{for } n \ge \ell+1 \\ \mathbf{1}[x_{-\infty}^{n} = z_{-\infty}^{n}] & \text{for } n \le \ell\,. \end{cases} \tag{2.22}$$

It is simple to check that these functions qualify as transition probabilities, as they satisfy requirements (i) and (ii) of Definition 2.12. It is also easy to verify that such a system defines a unique process.

Definition 2.23 *For $\ell \in \mathbb{Z}$ and $z_{-\infty}^{\ell} \in \mathcal{A}_{-\infty}^{\ell}$, the process consistent with the system (2.22) is called the* **stochastic process with fixed past** $z_{-\infty}^{\ell}$ **consistent with a system of transition probabilities** (P_n).

2.3 Simulation algorithms and transition probabilities

Let us now establish the equivalence of the simulation-oriented (Harris') approach of Section 2.1 with the transition-probability ("traditional") approach of Section 2.2.

Given a simulation algorithm $(f_n)_{n\in\mathbb{Z}}$, the prescription

$$P_n(a|x_{-\infty}^{n-1}) \;=\; \mathbb{P}\{f_n(U_n, x_{-\infty}^{n-1}) = a\} \;, \tag{2.24}$$

defines a system of transition probabilities (Exercise 2.48).

Now let $\{P_n(\,\cdot\,|\,\cdot\,) \;:\; n \in \mathbb{Z}\}$ be a system of transition probabilities. We construct a simulation algorithm by mimicking the way such transition probabilities would be simulated in a computer, namely by partitioning the interval $[0,1]$ into intervals of length equal to the probabilities. For each $x_{-\infty}^{n-1}$ let us consider a partition of $[0,1]$

$$\mathcal{P}^{x_{-\infty}^{n-1}} \;=\; \{I_a^{x_{-\infty}^{n-1}} : a \in \mathcal{A}\} \;, \tag{2.25}$$

each of the sets $I_a^{x_{-\infty}^{n-1}}$ being a union of intervals, such that

$$\mathrm{length}\Big(I_a^{x_{-\infty}^{n-1}}\Big) \;=\; P_n(a|x_{-\infty}^{n-1}) \;. \tag{2.26}$$

The prescription

$$f_n(u, x_{-\infty}^{n}) \;=\; a \text{ iff } u \in I_a^{x_{-\infty}^{n-1}} \tag{2.27}$$

defines a simulation algorithm.

The previous considerations amount to a procedure to transcribe simulation algorithms into transition probabilities and viceversa. The following proposition summarizes its main features. It proofs is basically contained in the preceding discussion, except for some minor mathematical details left to the reader.

Proposition 2.28 *(i) For a given a simulation algorithm $(f_n)_{n\in\mathbb{Z}}$, prescription (2.24) defines a system of transition probabilities $\{P_n(\,\cdot\,|\,\cdot\,) :$ $n \in \mathbb{Z}\}$*

(ii) *For a given system of transition probabilities* $\{P_n(\,\cdot\,|\,\cdot\,) : n \in \mathbb{Z}\}$, *each choice of partitions* $\{\mathcal{P}^{x_{-\infty}^{n-1}} : n \in \mathbb{Z}, x_{-\infty}^{n-1} \in \mathcal{A}_{-\infty}^{n-1}\}$ *satisfying (2.26) defines, through (2.27), a simulation algorithm* $(f_n)_{n\in\mathbb{Z}}$ *such that:*

   (ii.a) *every process consistent with* (P_n) *is defined by* (f_n), *and*

   (ii.b) *the system of transition probabilities constructed from such* (f_n) *by the procedure of part (i) is the original* (P_n).

Remark 2.29 In part (i) there is no claim that every process defined by $(f_n)_{n\in\mathbb{Z}}$ be consistent with

$$P_n(a|x_{-\infty}^{n-1}) := \mathbb{P}\{f_n(U_n, x_{-\infty}^{n-1}) = a\} \tag{2.30}$$

By (2.16), this would require that such a process verify

$$\mathbb{E}\Big[g(X_{-\infty}^{n})\Big] = \mathbb{E}\Big[g\Big(f_n(U_n, X_{-\infty}^{n-1})\,,\, X_{-\infty}^{n-1})\Big)\Big] \tag{2.31}$$

for $n \in \mathbb{Z}$ and g measurable with respect to $\mathcal{F}_n$. This may fail to be true unless the algorithm (f_n) satisfy some suitable properties. However, we remark that, by construction, **the consistency (2.31) holds for fixed-past processes**.

Proposition 2.28 allows the transcription of properties defined for simulation algorithms to properties of transition probabilities and viceversa. For instance, the **system of transition probabilities** is **Markovian of order** k if for each $x_n \in \mathcal{A}$ the function $P_n(x_n|\,\cdot\,)$ depends only on the k preceding symbols, that is if

$$P_n(x_n|x_{n-k}^{n-1}\, y_{-\infty}^{n-k}) = P_n(x_n|x_{n-k}^{n-1}\, z_{-\infty}^{n-k}) =: P_n(x_n|x_{n-k}^{n-1}) \tag{2.32}$$

for every $y_{-\infty}^{n-k}, z_{-\infty}^{n-k} \in \mathcal{A}_{-\infty}^{n-k}$. We leave to the reader the exercise of defining a time-homogeneous system, transcribing property (2.5) (Exercise 2.49).

In the sequel we shall only consider time-homogeneous chains and denote simply f the function in (2.8). In this case, it is enough to work with the transitions at time zero, i.e.

$$P(a|x_{-\infty}^{-1}) = \mathbb{P}(X_0 = a|X_{-\infty}^{-1} = a_{-\infty}^{-1})\,. \tag{2.33}$$

To simplify we shall denote $\underline{x} = x_{-\infty}^{-1}$, $\underline{X} = X_{-\infty}^{-1}$ and $\underline{\mathcal{A}} = \mathcal{A}_{-\infty}^{-1}$.

2.4 Coupling and coupling algorithms

Let us now present the main tool used in these notes.

Definition 2.34 *A* **coupling** *of the stochastic processes*

$$(X_n^{[1]})_{n\in\mathbb{Z}}, \cdots, (X_n^{[k]})_{n\in\mathbb{Z}}$$

is a stochastic process $(\widetilde{X}_n)_{n\in\mathbb{Z}}$ *with alphabet* $\mathcal{A}^k$ *whose marginal distributions are those of the processes* $(X_n^{[i]})$*. That is, such that for each* $i = 1, \ldots, k$ *and each* $x_l^m \in \mathcal{A}_l^m$*,* $l \leq m \in \mathbb{Z}$*, the probabilities of cylinders satisfy*

$$\mathbb{P}\Big(i\text{-th component of } \widetilde{X}_l^m = x_l^m\Big) \;=\; \mathbb{P}\Big((X^{[i]})_l^m = x_l^m\Big)\,. \tag{2.35}$$

Couplings will be defined via simulation algorithms.

Definition 2.36 *A* **coupling algorithm** *of stochastic processes*

$$(X_n^{[1]})_{n\in\mathbb{Z}}, \cdots, (X_n^{[k]})_{n\in\mathbb{Z}}$$

is a simulation algorithm $(\widetilde{f}_n)_{n\in\mathbb{Z}}$ *for the process* $(X_n^{[1]}, \cdots, X_n^{[k]})_{n\in\mathbb{Z}}$*. Explicitly,* $\widetilde{f}_n$ *is a function of the form* $(f_n^{[1]}, \cdots, f_n^{[k]})$*, with each* $f^{[i]} : [0,1] \times (\mathcal{A}^k)_{-\infty}^n \to \mathcal{A}$*, such that*

$$X_n^{[i]} \;=\; f_n^{[i]}\Big(U_n, (X^{[1]}, \cdots, X^{[k]})_{-\infty}^{n-1}\Big) \tag{2.37}$$

for $i = 1, \cdots, k$*, for the common (vector) independent uniform variables* U_n*.*

Thus, a coupling algorithm produces at time n simultaneously the time-n state of *all* the processes $(X_n^{[i]})$ using the *same* random number U_n for all of them. There is considerable freedom and some potential danger, in the construction of coupling algorithms. On the one hand condition (2.37) leaves plenty of room for designing algorithms with features suited to each particular application. These notes will repeatedly illustrate this fact. On the other hand, the algorithm $(\widetilde{f}_n)$ may define *several* processes in $\mathcal{A}^k$, and some of them may fail to be a coupling of the target processes $(X^{[i]})$ —that is, (2.35) may not hold.

¿From a constructive point of view Definition 2.36 does not look very informative. Indeed, processes are seldomly given directly. Rather, in these lectures they are constructed starting from simulation algorithms. What we would need, then, are prescriptions on how to construct a coupling algorithm starting from the simulation arguments of the individual processes.

Let us settle these issues while doing at the same time the connection with the transition-probability framework.

Definition 2.38 *A* **coupling of the systems of transition probabilities** $P_n^{[1]}(\cdot\,|\,\cdot), \cdots, P_n^{[k]}(\cdot\,|\,\cdot)$ *is a system of transition probabilities* $\widetilde{P}_n : \mathcal{A}^k \times (\mathcal{A}_{-\infty}^{n-1})^k \longrightarrow [0,1]$ *such that*

$$\sum_{\substack{x_n^{[1]},\cdots,x_n^{[j-1]}\in\mathcal{A}\\ x_n^{[j+1]},\cdots,x_n^{[k]}\in\mathcal{A}}} \widetilde{P}_n\Big(x_n^{[1]},\ldots,x_n^{[k]}\;\Big|\;(x^{[1]})_{-\infty}^{n-1},\ldots,(x^{[k]})_{-\infty}^{n-1}\Big) \;=\; P_n^{[j]}\Big(x_n^{[j]}\;\Big|\;(x^{[j]})_{-\infty}^{n-1}\Big) \tag{2.39}$$

for all $j=1,\ldots,k$, *all* $x_n^{[j]}\in\mathcal{A}$ *and all* $(x^{[1]})_{-\infty}^{n-1},\ldots,(x^{[k]})_{-\infty}^{n-1}\in\mathcal{A}_{-\infty}^{n-1}$.

[This definition is, in fact, a particular instance of the notion of coupling among probability measures.]

Every coupling of transition probabilities produces a coupling algorithm through the prescription (2.26)–(2.27). First one must choose partitions of $[0,1]$ in Lebesgue measurable sets

$$\Big\{I_{a^{[1]}\cdots a^{[k]}}^{(x^{[1]})_{-\infty}^{n-1}\cdots(x^{[k]})_{-\infty}^{n-1}} : a^{[i]}\in\mathcal{A}, (x^{[i]})_{-\infty}^{n-1}\in\mathcal{A}_{-\infty}^{n-1}, i=1,\ldots,k\Big\}\,, \tag{2.40}$$

such that

$$\mathrm{length}\Big(I_{a^{[1]}\cdots a^{[k]}}^{(x^{[1]})_{-\infty}^{n-1}\cdots(x^{[k]})_{-\infty}^{n-1}}\Big) \;=\; \widetilde{P}_n\Big(a^{[1]},\ldots,a^{[k]}\;\Big|\;(x^{[1]})_{-\infty}^{n-1},\ldots,(x^{[k]})_{-\infty}^{n-1}\Big)\,. \tag{2.41}$$

The coupling algorithm is then defined by

$$\widetilde{f}_n\Big(u, (x^{[1]})_{-\infty}^{n-1}, \cdots, (x^{[k]})_{-\infty}^{n-1}\Big) \;=\; (a^{[1]}, \cdots, a^{[k]}) \qquad \text{iff } u \in I_{a^{[1]}\cdots a^{[k]}}^{(x^{[1]})_{-\infty}^{n-1}\cdots(x^{[k]})_{-\infty}^{n-1}} \,. \tag{2.42}$$

A possible strategy to construct a coupling algorithm would, in principle, involve two steps:

Step 1: Construct a coupled transition $(\widetilde{P}_n)$ starting from the individual transition probabilities $(P^{[i]})$ (or, equivalently, from the individual simulation algorithms)

Step 2: Take the algorithm defined in (2.42).

We shall adopt, however, a more economical *graphical procedure* which yield *directly the partitions* (2.40), hence the coupling algorithm, bypassing the definition of coupling transitions [which, of course, can be obtained from the coupling algorithm by (2.24)]. Furthermore, the coupling algorithms "factor" in the sense that *each component* $(f_n^{[i]})$ *is itself a simulation algorithm of* $(X_n^{[i]})$. That is, relation (2.37) is satisfied in the particular form

$$X_n^{[i]} \;=\; f_n^{[i]}\Big(U_n, X^{[i]}\Big) \,. \tag{2.43}$$

This can be achieved in the following fashion.

First: First, for each $j = 1, \ldots, k$ and each $(x^{[1]})_{-\infty}^{n-1}, \ldots, (x^{[k]})_{-\infty}^{n-1} \in \mathcal{A}_{-\infty}^{n-1}$, find partitions

$$\Big\{ I_a^{(x^{[j]})_{-\infty}^{n-1} \,|\, (x^{[1]})_{-\infty}^{n-1}\cdots(x^{[k]})_{-\infty}^{n-1}} : a \in \mathcal{A} \Big\} \,, \tag{2.44}$$

formed by unions of intervals such that

$$\mathrm{length}\Big(I_a^{(x^{[j]})_{-\infty}^{n-1} \,|\, (x^{[1]})_{-\infty}^{n-1}\cdots(x^{[k]})_{-\infty}^{n-1}}\Big) \;=\; P_n^{[j]}\Big(a \;\Big|\; (x^{[j]})_{-\infty}^{n-1}\Big) \tag{2.45}$$

whatever the choice of $(x^{[i]})_{-\infty}^{n-1}$ for $i \neq j$.

Second: Take the algorithm defined, by (2.42), by the sets

$$I_{a^{[1]}\cdots a^{[k]}}^{(x^{[1]})_{-\infty}^{n-1}\cdots(x^{[k]})_{-\infty}^{n-1}} = \bigcap_{j=1}^{k} I_{a^{[j]}}^{(x^{[j]})_{-\infty}^{n-1} \,|\, (x^{[1]})_{-\infty}^{n-1}\cdots(x^{[k]})_{-\infty}^{n-1}} . \tag{2.46}$$

Notice that condition (2.45) implies (2.43).

The sets of the partitions (2.44) can be visualized as obtained by "cutting and pasting" parts of intervals of length $P_n^{[j]}(a|(x^{[j]})_{-\infty}^{n-1})$ in a manner that depends on the other transitions. We observe that the coupling of transition probabilities obtained from the intersections (2.46) by the prescription (2.24) is in general different from the mere product of the individual transitions. In particular it gives probability zero to states $a^{[1]}\cdots a^{[k]}$ for which the intersections (2.46) are empty.

Definition 2.47 *Partitions defined by (2.44)–(2.46) are called a* **graphical procedure** *to construct a coupling algorithm among processes consistent with transition probabilities* $(P_n^{[1]}), \cdots, (P_n^{[k]})$.

As commented above, the graphical procedure does not, in general, settle the issue of finding an actual coupling among the target processes. We must actually *construct* a process defined by the coupling algorithm. Furthermore, we may have to choose properly if there are several such processes. This choice is actually unnecessary if *each of the transitions* $(P_n^{[i]})$ *admits a* **unique** *consistent process.* This will be the situation for all the processes studied in these notes.

It is apparent that there is considerable freedom in the choice of the partitions defining a simulation algorithms. This freedom can be exploited to design partitions adapted to particular mathematical or numerical purposes.

The above technique can be applied, without modification, to countable families of processes $(X_n^{[i]})$ (and countable alphabets).

2.5 Exercises

Exercise 2.48 Verify that, if $(f_n)_{n\in\mathbb{Z}}$ is a simulation algorithm, prescription (2.30) indeed defines a system of transition probabilities.

Exercise 2.49 Define a time-homogeneous system of transition probabilities. Establish the relation with property (2.5) and explain why it is enough to consider the objects $P(a|x_{-\infty}^{-1})$ defined in (2.33).

Exercise 2.50 Consider a time-homogeneous system of transition probabilities. Show that (2.16) is equivalent to the existence of measures π_n on $\mathcal{A}_{-\infty}^n$ such that

$$\int_{\mathcal{A}_{-\infty}^n} \pi_n(dx_{-\infty}^n)\, P_{n+1}(\,\cdot\,|x_{-\infty}^n) \;=\; \pi_{n+1}(\,\cdot\,)\,. \tag{2.51}$$

Exercise 2.52 (a) Check that the fixed-past transitions (2.22) verify conditions (i) and (ii) of Definition 2.12.

(b) Show that they define a unique consistent process.

Chapter 3

Types of chains of infinite order. Examples

Before passing to examples, let us spell out the different types of hypotheses we will be demanding for the processes studied in these notes. These hypotheses are best expressed in terms of transition probabilities and they refer to (i) continuity with respect to histories, and (ii) strict positivity. In turns, suitable combinations of these hypotheses give rise to three standard notions of chains of infinite order.

3.1 Continuity hypotheses

Definition 3.1 *A system of transition probabilities is* **continuous** *if the functions $P_n(x_n | \cdot)$ are continuous for each $n \in \mathbb{Z}$ and each $x_n \in \mathcal{A}$ or, equivalently, if*

$$\begin{aligned} \beta_s \quad &:= \quad \sup_{n\in\mathbb{Z}} \sup_{x,y} \Big| P_n(x_n | x_{-\infty}^{n-1}) - P_n(x_n | x_{n-s}^{n-1}\, y_{-\infty}^{n-s-1}) \Big| \\ &\underset{s\to\infty}{\longrightarrow} \quad 0 \,. \end{aligned} \tag{3.2}$$

The sequence $(\beta_s)_{s\in\mathbb{N}}$ is called the **continuity rate***.*

The existence problem is not a problem for continuous transitions:

Proposition 3.3 *A system of continuous transition probabilities has at least one stochastic process consistent with it.*

Proof. To be written (uses compactness). □

The following stronger notion of continuity has also been introduced:

Definition 3.4 *A system of transition probabilities is* **log-continuous** *if*

$$\begin{aligned} \gamma_s \quad &:= \quad \sup_{n\in\mathbb{Z}} \sup_{x,y} \left| \frac{P_n(x_n | x_{-\infty}^{n-1})}{P_n(x_n | x_{n-s}^{n-1}\, y_{-\infty}^{n-s-1})} \;-\; 1 \right| \\ &\underset{s\to\infty}{\longrightarrow} \quad 0\,. \end{aligned} \tag{3.5}$$

The sequence $(\gamma_s)_{s\in\mathbb{N}}$ *is called the* **log-continuity rate***.*

The strongest notion of continuity refers to the ℓ-move transitions (2.17):

Definition 3.6 *A system of transition probabilities is* **multiple-move log-continuous** *if*

$$\begin{aligned} \alpha_s \quad &:= \quad \sup_{n\in\mathbb{Z},\ell\in\mathbb{N}} \sup_{x,y} \left| \frac{P_{[n,n+\ell]}(x_n^{n+\ell} | x_{-\infty}^{n-1})}{P_{[n,n+\ell]}(x_n^{n+\ell} | x_{n-s}^{n-1}\, y_{-\infty}^{n-s-1})} \;-\; 1 \right| \\ &\underset{s\to\infty}{\longrightarrow} \quad 0\,. \end{aligned} \tag{3.7}$$

The sequence $(\alpha_s)_{s\in\mathbb{N}}$ *is called the* **multiple-move log-continuity rate***.*

3.2 Non-nullness hypotheses and types of chains

Two kinds of non-nullness hypotheses are used.

Definition 3.8 *A system of transition probabilities is* **weakly non-null** *if*

$$\inf_{n\in\mathbb{Z}} \sum_{y_n\in\mathcal{A}} \inf_{x} P_n(y_n | x_{-\infty}^{n-1}) \;>\; 0\,. \tag{3.9}$$

Definition 3.10 *A system of transition probabilities is* **strongly non-null** *if*

$$\inf_{n\in\mathbb{Z}} \inf_{x} P_n(y_n|x_{-\infty}^{n-1}) \;>\; 0\;. \tag{3.11}$$

We are finally ready to define the different types of chains to be discussed in the sequel.

Definition 3.12 *A stochastic process is a* **chain of infinite order**

(i) **of type A** *if it is consistent with a system of transition probabilities that is continuous and weakly non-null.*

(ii) **of type B** *if it is consistent with a system of transition probabilities that is log-continuous and strongly non-null.*

(iii) **of type C** *if it is consistent with a system of transition probabilities that is multiple-move log-continuous and strongly non-null.*

Types A and B were already considered by Doeblin and Fortet (1937). Type C was introduced, as far as we know, by Lalley (1986).

3.3 Examples

The following two examples are more than just illustrations. In fact, a central aspect of these lectures is to show that large families of chains can be written in any of these forms.

Countable mixtures of Markov chains (CMMC) These are chains whose transition probabilities are countable convex combinations of Markov transitions of increasing order. That is, they are of the form

$$P(a|\underline{x}) \;=\; \lambda_0\, P^{(0)}(a) + \sum_{k=1}^{\infty} \lambda_k\, P^{(k)}(a|x_{-k}^{-1}) \tag{3.13}$$

where $\lambda_k \geq 0$, $\sum_{k=0}^{\infty} \lambda_k = 1$, and each $P^{(k)}(a|x_{-k}^{-1})$ is a Markov transition of order k for $k \geq 1$, while $P^{(0)}$ is a probability measure. The transitions (3.13) can be thought as resulting of two independent random steps. First, an integer $k \geq 0$ is chosen with probability λ_k, and, second, a symbol is chosen with the order-k transition probability $P^{(k)}$. Thus, each transition actually depends on a finite, but random, number of preceding states. To our knowledge, an expression like (3.13) —but with k ranging over finitely many values and $P^{(k)}(a|x_{-k}^{-1}) = g^{(k)}(a, x_{-k})$— was first studied by Raftery (1985a, 1985b) under the name of mixture transition distribution (MTD) model (see also Raftery and Tavaré, 1994)).

As we shall see in Chapter 7 that, under suitable hypotheses on the family (λ_k), a chain consistent with transitions of the form (3.13) has the *renewal property*: There exists a sequence of random times $(t_i)_{i\in\mathbb{Z}}$, with independent increments $t_{i+1} - t_i$, such that for each $i \in \mathbb{Z}$ the distribution of the variables $\{X_n : n \geq t_i\}$ is independent of the variables $\{X_n : n < t_i\}$. This is an example of a *regeneration scheme*. At the same Chapter 7 we shall show that any chain of infinite order with not-too-slow continuity rates [see (3.2)] is actually a CMMC.

Variable-length Markov chains (VLMC) The transition probabilities of these chains also depend on a finite number of preceding states, but this number is determined by the past history. More precisely, there exists a *lag function*

$$\ell : \underline{A} \longrightarrow \{0, -1, -2, \cdots, \infty\} \tag{3.14}$$

such that

$$P(a|\underline{x}) = P(a|x_{\ell(\underline{x})}^{-1}) \tag{3.15}$$

with the convention that when $\ell(\underline{x}) = 0$, the transition probability is actually independent of the past.

This type of processes was introduced by Bühlman and Wyner (1999), albeit for bounded functions ℓ. In Chapter 7 we shall show that chains of infinite order with not-too-slow continuity rates can be embedded into a VLMC.

The following example illustrates the differences between the different types of chains introduced above.

Sparse VLMC This is an infinite-order version of example (M5) of Bühlman and Wyner (1999). It has a two-symbol alphabet, for instance $\mathcal{A} = \{0, 1\}$, and a lag function

$$\ell(\underline{x}) \;=\; \ell \text{ if } x_{-1} = 0 = \cdots = x_{-\ell}, x_{-\ell-1} = 1\;. \tag{3.16}$$

The transition probabilities are defined by

$$P(1|\underline{x}) \;=\; q_{\ell(\underline{x})} \tag{3.17}$$

with $0 < q_k < 1$. We leave to the reader (Exercise 3.24 the verification of the following facts:

(a) If $\lim_k q_k$ does not exists or it is different from q_∞, the system is not continuous.

(b) If $\lim_k q_k = q_\infty$ and there exist constants $0 < c \leq d < 1$ such that $q_k \in [c, d]$ for all k, then the system is log-continuous and strictly non-null.

(c) If $\lim_k q_k = q_\infty = 0$ then the system is continuous but not log-continuous. Furthermore, it is weakly but not strongly non-null.

Sparse VLMC are closely related to *renewal processes* on $\mathbb{Z}$. In fact, let us define $(T_k)_{k\in\mathbb{Z}}$ the succesive times in which the sparse VLMC (X_n) takes the value 1, i.e.

$$\begin{array}{rcl} & \vdots & \\ T_0 &=& \sup\{n \leq 0 : X_n = 1\} \\ T_1 &=& \inf\{n > 0 : X_n = 1\} \\ T_2 &=& \inf\{n > T_1 : X_n = 1\} \\ & \vdots & \end{array} \tag{3.18}$$

Then, the point process $(T_k)_{k\in\mathbb{Z}}$ is a renewal process, that is,

(i) the random increments $(T_k - T_{k-1})_{k\in\mathbb{Z}}$ are independent, and

(ii) the random increments $(T_k - T_{k-1})_{k\neq 0}$ are identically distributed.

We conclude with some well known families of processes that fit in our framework.

Hidden Markov models (HMM) These models refer to a process (X_n), with values on an alphabet $\mathcal{A}$, which is defined in terms of a Markov process (S_n) with values in a finite set of states $\mathcal{S}$ —the hidden process. This models situations in which there is a simple but inaccesible process containing all the information about the problem, and the observer has access only to an impoverished *ersatz* of it.

Examples of processes (X_n) of this type were introduced by Shannon (1948) under the name *Markov sources.* These processes are defined by the coordinate-by-coordinate transformation $X_n = f(S_n)$ of an order-1 Markov chain (S_n). We leave to the reader the verification that such a process may not be a Markov chain (Exercise 3.25).

The processes were reintroduced, with a different flavor, by Baum and Petrie (1966) and were later intensively used in the theory of speech recognition (see, for instance Jelinek, 1999). In this formulation, there is a family of probability measures $\{\mu_s : s \in \mathbb{S}\}$ on $\mathcal{A}$ establishing the relation between the processes (X_n) and (S_n) through the relations

$$\mathbb{P}(X_m^n = x_m^n | S_m^n = s_m^n) \;=\; \prod_{i=m}^{n} \mu_{s_i}(x_i) \tag{3.19}$$

valid for each choice of $x_m^n \in \mathcal{A}_m^n$ and $s_m^n \in \mathbb{S}_m^n$, for each $m, n \in \mathbb{Z}$, $m \leq n$. Therefore the observable process (X_n) is a coordinate-by-coordinate *random* transformation of the hidden Markov chain (S_n).

In fact, Markov sources and hidden Markov models are equivalent notions. The proof of this fact is left as an exercise to the reader (Exercise 3.26 below). The fact that hidden Markov models are chains of infinite order with continuous transition probabilities is also left as an exercise (Exercise 7.22). The proof uses a regeneration property of Markov chains. This property follows from exercise 6.40 or as a particular isntance of the much more general theory developed in Chapters 6–4 for chains of infinite orders.

Binary autoregressions Let G be the two-points set, for instance $G = \{-1,+1\}$, θ_0 a real number and $(\theta_k; k \geq 1)$ a summable real sequence. Let $q : \mathbb{R} \mapsto]0,1[$ be strictly increasing and continuously differentiable. Define

$$P(\,\cdot\,|\underline{w}) \text{ is the Bernoulli law on } \{-1,+1\} \text{ with parameter } q\Big(\theta_0+\sum_{k\geq 1}\theta_k w_{-k}\Big)\ , \tag{3.20}$$

i.e., $P(+1|\underline{w}) = q(\theta_0 + \sum_{k\geq 1}\theta_k w_{-k}) = 1 - P(-1|\underline{w})$. Such a process is the binary version of autoregressive (long memory) processes used in statistics and econometrics. It describes binary responses when covariates are historical values of the process (see McGullagh and Nelder, 1989, Sect. 4.3). A popular choice for q is the logistic function

$$q(x) = \frac{\exp x}{2\cosh x} = \frac{1}{2(1+\exp -2x)}\ . \tag{3.21}$$

Random systems with complete connections These are processes formed by pairs of chains evolving in an inter-related manner, used to model a number of practical problems. Applications include urn models, the theory of continuous functions, learning models, etc. We refer the reader to Iosifescu and Grigorescu (1990) for a survey. Of the two chains, one is Markov, but in a complicated "alphabet", or with complicated transition functions, while the other is of infinite order in a simpler alphabet. The latter chain is, in practice, used to infer properties of the complicated Markov chain. As an example, let us present the *Markov chains defined by D-ary expansions*

These are process having the unit interval as "alphabet", $I = [0,1]$, and defined through another, auxiliary, process with a finite alphabet. Formally, a family of maps is established between sequences of a finite alphabet $G =$

$\{0, 1, \cdots, D-1\}$ and real numbers in I via D-ary expansions: For each $n \in \mathbb{Z}$

$$\begin{array}{rcl} X_n : G^{\mathbb{Z}} & \longrightarrow & I \\ (\eta(i) : i \in \mathbb{Z}) & \mapsto & x_n = \sum_{j=1}^{\infty} \eta(n-j)/D^j \;. \end{array} \tag{3.22}$$

This map induces a natural map from probability kernels $P : G \times G^{-\mathbb{N}^*} \mapsto [0, 1]$ to probability kernels $F : I \times I^{-\mathbb{N}^*} \mapsto [0, 1]$: For each $x \in I$, given an $\underline{w} \in G^{-\mathbb{N}^*}$ with $x = X_0(\underline{w})$

$$F\Big(X_1 = \frac{g + x}{D} \;\Big|\; X_0 = x\Big) \;=\; P(g|\underline{w}) \;. \tag{3.23}$$

Interest focuses on the existence and properties of measures on the Borelians of $I^{\mathbb{Z}}$ compatible with such a probability kernel F.

Maps (3.22)–(3.23) have been already introduced by Borel in 1909 for i.i.d. $\eta(i)$. The general case in which the $\eta(i)$ form a chain with long memory is the object of Harris (1955) seminal paper.

3.4 Exercises

Exercise 3.24 Verify facts (a), (b) and (c) for the sparce VLMC defined by (3.16)–(3.17).

Exercise 3.25 Consider a Markov source, that is a process (X_n) defined by the coordinate-by-coordinate transformation $X_n = f(S_n)$ of an order-1 Markov chain (S_n) taking values in a finite set of states $\mathbb{S}$. Show that, in general, such a process is not a Markov chain.

Exercise 3.26 Observe that every Markov source is trivially a HMM. Conversely, prove that every HMM can be written as a Markov source. Hint: Consider the process $Z_n = (S_n, X_n)$.

Exercise 3.27 Prove that for a sparse VLMC (X_n), the times (T_k) defined in (3.18) form a renewal process. Hint:

$$\mathbb{P}(T_k - T_{k-1} = \ell) \;=\; q_\ell \prod_{i=1}^{\ell-1} (1 - q_i) \;. \tag{3.28}$$

Chapter 4

A regeneration scheme for CMMC

4.1 Random orders and regeneration times

Let us recall that a CMMC is defined by a system of transition probabilities which can be decomposed as

$$P(a|\underline{x}) \;=\; \lambda_0 \, P^{(0)}(a) + \sum_{k=1}^{\infty} \lambda_k \, P^{(k)}(a|x_{-k}^{-1}) \tag{4.1}$$

where each $P^{(k)}(a|x_{-k}^{-1})$ is a Markov transition of order k for $k \geq 1$, $P^{(0)}$ is a probability measure, and the λ_k are non-negative real numbers with $\sum_{k=0}^{\infty} \lambda_k = 1$.

We shall use a simulation algorithm for these transitions constructed on the basis of a double sequence of uniform random variables $(U_n^{(1)}, U_n^{(2)})$ which we simply denote$(U_i, V_i)_{i\in\mathbb{Z}}$.

Definition 4.2 *A* **CMMC simulation algorithm** *is an algorithm of the form*

$$X_n \;=\; \sum_{k=0}^{\infty} \mathbf{1}\{\alpha_{k-1} \leq U_n \leq \alpha_k\} \, f^{(k)}(V_n, X_{n-k}^{n-1}) \; . \tag{4.3}$$

where the $f^{(k)}$ are simulation algorithms of order-k Markov chains and (α_k) is an increasing non-negative sequence with $\alpha_k \nearrow 1$. (By convention $\alpha_{-1} = 0$.)

We leave the reader the task of verifying that (4.4) is a simulation algorithm for a process (X_n) consistent with (4.1) if

(i) $f^{(k)}$ are the simulation algorithms of the Markov chains with transitions $P^{(k)}$ [defined, for instance, as in (2.26)–(2.27)], and

(ii)

$$\alpha_k \;=\; \sum_{i=0}^{k} \lambda_i \; . \tag{4.4}$$

(Exercise 4.49).

In this section we shall study properties of processes defined by this simulation algorithm. In Section 4.2 we discuss the existence problem and the (non-trivial) issue of whether such processes are in fact consistent with (4.1).

In fact, the variables (U_n) define in (4.3) an auxiliary process which plays a key role in the sequel.

Definition 4.5 *Let us call* **random orders**, *or* **random-order process** *to the independent random variables $(L_n)_{n\in\mathbb{Z}}$, defined as*

$$L_n \;=\; \sum_{k=0}^{\infty} k\,\mathbf{1}\{\alpha_{k-1} \le U_n \le \alpha_k\} \; . \tag{4.6}$$

It is crucial to observe that the random orders are constructed with total independence of the rest of the procedure. The variable L_n indicates how many instants in the past are actually used to determine X_n: substituting the definition of L_n in (4.3), the simulation algorithm reads

$$X_n \;=\; \sum_{k=0}^{\infty} \mathbf{1}\{L_n = k\}\, f^{(k)}(V_n, X_{n-k}^{n-1}) \; . \tag{4.7}$$

In other words,

$$L_n = k \quad \text{implies} \quad X_n \;=\; f^{(k)}(V_n, X_{n-k}^{n-1}) \tag{4.8}$$

if $k \geq 1$, while for $k = 0$

$$L_n = 0 \quad \text{implies} \quad X_n \;=\; f^{(0)}(V_n) \tag{4.9}$$

which is independent of the past. The variables L_n can be visualized as arrows pointing from the instant n to the instant $n - L_n$. Each realization of random orders determines the "genealogy" of the state at each instant. The state at time n is determined by the configuration on the interval $[n - L_n, n]$; each $i \in [n - L_n, n]$ is in turns determined by the states at the interval $[i - L_i, i]$ and so on. This back-referencing procedure can lead us to one of two situations:

(i) the procedure continues forever and take us to $-\infty$,

(ii) the procedure actually stops at a time $\tau[n]$ such that no arrow starting from n or its "ancestors" crosses it. In particular, the configuration at $\tau[n]$ must be independent of the past, that is $L_{\tau[n]} = 0$.

In the second case, the values assumed by the process before $\tau[n]$ are irrelevant for the determination of X_n. This time $\tau[n]$ is a *regeneration time* for the instant n.

More generally, we can consider *windows* $(X_l, \ldots, X_m)$, for two integers $l < m$ and analyze the possibility of constructing it knowing only a finite part of the past history of the process. In other words, we want to find the closest past time $\tau[l, m]$ such that the window $(X_l, \ldots, X_m)$ is independent of the variables $\{X_i : i < \tau[l, m]\}$. This random time can be bounded through the random-order process (L_n).

Definition 4.10 *The* **regeneration time for the window** $(X_l, \ldots, X_m)$ *is*

$$\tau[l, m] \;:=\; \max\Big\{t \leq l : t \leq n - L_n, \;\textit{for all } n \in [t, m]\Big\} \tag{4.11}$$

with the convention $\tau[l, m] = -\infty$ *if the set in the right-hand side is empty. In case* $l = m$ *we write* $\tau[l] := \tau[l, l]$.

Notice that, by the definition (4.6) of the variables L_n,

$$\tau[l,m] \;=\; \max\Big\{t \le l : U_n \le \alpha_{n-t},\ \text{for all } n \in [t,m]\Big\}\,. \tag{4.12}$$

To be sure, definition (4.11) refers to the worst-case scenario, where each order-k Markov transition probability depends on all the k preceding times. For less drastic dependences, the actual regeneration times can be closer to the window than the one defined by (4.11). An extreme example is when the different Markov transitions depend on only one site in the past — $P^{(k)}(a|x_{-k}^{-1}) = g^{(k)}(a, x_{-k})$. In this case, the state at each time depends of exactly one ancestor and regeneration can take place at times much closer than (4.11).

For fixed l, the sequence of regeneration times $\tau[l,m]$ for $m \ge l$ is decreasing. In particular, a regeneration time for a given interval $[l,m]$ is not, in general, a regeneration time for a larger interval $[l,m']$ with $m' > m$.

The monotonicity of the sequence $(\tau[l,m])_m$ implies the existence of the limit

$$\tau[l,+\infty[\;:=\; \lim_{m\to\infty} \tau[l,m]\,. \tag{4.13}$$

Definition 4.14 *If* $\tau[l,+\infty[< \infty$ *we call it a* **renewal time** *for the CMMC algorithm.*

We remark that

$$\tau[l,m] \;=\; \min_{l\le i\le m} \tau[i] \tag{4.15}$$

and

$$\tau[l,+\infty[\;=\; \inf_{l\le i} \tau[i]\,. \tag{4.16}$$

The considerations of this section clearly indicate the strategy to follow for the study of CMMC:

(1) Determine the distribution of regeneration times. This depends *only* on the random-order process, that is on the parameters $(\lambda_k)_{k\ge 0}$ in (4.1).

(2) Study the properties of the process in terms of this distribution.

We stress, however, that the decomposition (4.1) of a CMMC is not unique. See the discussion in Section 4.3 below and Exercise 7.24.

We develop this strategy in Chapter 6. We emphasize that this approach is based on the simulation algorithm (4.3). We still have to relate process defined by these algorithms with process consistent with CMMC decompositions (4.1). This is done in next section.

4.2 Existence, uniqueness and loss of memory of CMMC

4.2.1 Main results

This section is devoted to the proof of the following theorems.

Theorem 4.17 (Existence and uniqueness) *Consider a CMMC system of transition probabilities as in (4.1), and the related CMMC simulation algorithm (4.3). If*

$$\mathbb{P}(\tau[0] > -\infty) = 1 \tag{4.18}$$

then

(i) There exists exactly one stochastic process $(X_n)_{n\in\mathbb{Z}}$ defined by the algorithm. The process can be defined almost surely in the following way. To define X_n, start from $\tau[n]$ and determine first

$$X_{\tau[n]} \;=\; f^{(0)}(V_{\tau[n]}) \tag{4.19}$$

and then, inductively,

$$X_i \;=\; f^{(L_i)}(V_i, X_{L_i}^{i-1}) \tag{4.20}$$

for $i \in [\tau[n]+1, n]$. [The functions $f^{(i)}$ are the simulation algorithms of the Markov components of the CMMC used in the algorithm (4.3).]

(ii) *For any* $z_{-\infty}^{-p} \in \mathcal{A}_{-\infty}^{-p}$

$$\lim_{p\to\infty} X_n[z_{-\infty}^{-p}] \;=\; X_n \tag{4.21}$$

$\mathbb{P}$*-almost surely for all* $n \in \mathbb{Z}$. *[The left-hand side is the process with fixed past defined in (2.11).]*

(iii) *This process* (X_n) *is the only process consistent with the CMMC transition probabilities.*

We remark that by (4.15) and translation invariance

$$\mathbb{P}(\tau[0] > -\infty) = 1 \quad \Longleftrightarrow \quad \mathbb{P}(\tau[l,m] > -\infty) = 1 \quad \forall\, l \le m \in \mathbb{Z}\,. \tag{4.22}$$

Theorem 4.23 (Loss of memory) (i) *If* (X_n) *is consistent with a CMMC system of transition probabilities,*

$$\Big|\mathbb{P}\Big(X_0^j = a_0^j\Big) - \mathbb{P}\Big(X_0^j[z_{-\infty}^{-p}] = a_0^j\Big)\Big| \;\le\; \mathbb{P}\Big(\tau[0,j] \le -p\Big) \tag{4.24}$$

for each $j,k \in \mathbb{N}$ *and each past* $z_{-\infty}^{-p} \in \mathcal{A}_{-\infty}^{-p}$.

(ii) *If* $(\widetilde{X}_n)$ *and* $(\widehat{X}_n)$ *are two processes consistent with a CMMC system of transition probabilities,*

$$\Big|\widetilde{\mathbb{P}}\Big(\widetilde{X}_0^j = a_0^j\Big) - \widehat{\mathbb{P}}\Big(\widehat{X}_0^j = a_0^j\Big)\Big| \;\le\; \mathbb{P}\Big(\tau[0,j] > -\infty\Big) \tag{4.25}$$

for each $j,k \in \mathbb{N}$.

Inequality (4.24) bounds the speed at which the process is "lossing memory" from the original history $z_{-\infty}^{-k}$. This bound will be exploited in Chapter 6. Inequality (4.25) could be useful for CMMC exhibiting phase coexistence, i.e. with more than one consistent process.

The proof of these theorems is presented in the next sections.

4.2.2 Existence

Proof of part (i) of Theorem 4.17 The process is defined through (4.19) and (4.20).

Proof of part (ii) of Theorem 4.17 (convergence of fixed-past processes) The process $(X_n[z_{-\infty}^{-p}])$ is defined by the fixed-past version of the algorithm (4.3):

$$\begin{aligned} X_n[z_{-\infty}^{-p}] \;&=\; \sum_{k=0}^{n+p-1} \mathbf{1}\{\alpha_{k-1} \le U_n \le \alpha_k\}\, f^{(k)}\Big(V_n\,,\, X_{n-k}^{n-1}[z_{-\infty}^{-p}]\Big) \\ &\quad + \sum_{k=n+p}^{\infty} \mathbf{1}\{\alpha_{k-1} \le U_n \le \alpha_k\}\, f^{(k)}\Big(V_n\,,\, X_{-p+1}^{n-1}[z_{-\infty}^{-p}]\, z_{n-k}^{p}\Big)\,. \end{aligned} \tag{4.26}$$

If $\tau[n] > -p$, the last sum dissapears and we recover the same recursive equations (**??**), which are, in fact, equivalent to (4.19)–(4.20). We conclude that

$$X_n[z_{-\infty}^{-p}]\; \mathbf{1}\{\tau[n] > -p\} \;=\; X_n\; \mathbf{1}\{\tau[n] > -p\} \tag{4.27}$$

with X_n defined by (4.19)–(4.20). Furthermore, via (4.15) this identity generalizes to

$$X_l^m[z_{-\infty}^{-p}]\; \mathbf{1}\{\tau[l,m] > -p\} \;=\; X_l^m\; \mathbf{1}\{\tau[l,m] > -p\}\,. \tag{4.28}$$

In particular identity (4.27) proves part (ii).

Proof of consistency in part (iii) of Theorem 4.17 We show now that the process of part (i) is consistent with the transition probabilities

$$P(a|x_{-\infty}^{n-1}) \;:=\; \mathbb{P}\{F(U_n, V_n, x_{-\infty}^{n-1}) = a\} \tag{4.29}$$

where F is the function in the right-hand side of (4.3). For this we must verify (2.31) for $f_n = F$ for any *cylindrical* g. Let us consider $g = g(X_l^n)$.

Our starting point is the consistency of the fixed-past processes. Indeed, by the remark following (2.31), we have that

$$\mathbb{E}\Big[g(X_l^n[z_{-\infty}^{-p}])\Big] \;=\; \mathbb{E}\Big[g\Big(F(U_n,V_n,X_{-\infty}^{n-1}[z_{-\infty}^{-p}])\,,\,X_{-l}^{n-1}[z_{-\infty}^{-p}]\Big)\Big] \tag{4.30}$$

for any past $z_{-\infty}^{-p} \in \mathcal{A}_{-\infty}^{-p}$. We shall take the limit $p \to \infty$ of this expression.

By part (ii) of the theorem and dominated convergence

$$\mathbb{E}\Big[g(X_l^n[z_{-\infty}^{-p}])\Big] \;\underset{p\to\infty}{\longrightarrow}\; \mathbb{E}\Big[g(X_l^n)\Big]\;. \tag{4.31}$$

We now insert inside the expectation in the right-hand side of (4.30)

$$1 \;=\; \mathbf{1}\{\tau[l,n] > -p\} + \mathbf{1}\{\tau[l,n] \le -p\}\;. \tag{4.32}$$

By (4.28)

$$\begin{aligned} & g\Big(F(U_n,V_n,X_{-\infty}^{n-1}[z_{-\infty}^{-p}])\,,\,X_{-l}^{n-1}[z_{-\infty}^{-p}]\Big)\;\mathbf{1}\{\tau[l,n] > -p\} \\ &= \quad g\Big(F(U_n,V_n,X_{-\infty}^{n-1})\,,\,X_{-l}^{n-1}\Big)\;\mathbf{1}\{\tau[l,n] > -p\} \\ &\underset{p\to\infty}{\longrightarrow} \quad g\Big(F(U_n,V_n,X_{-\infty}^{n-1})\,,\,X_{-l}^{n-1}\Big) \quad \mathbb{P}\text{-a.s.} \end{aligned} \tag{4.33}$$

The last convergence is due to hypothesis (4.18) (plus translation invariance). The same hypothesis implies that

$$\mathbb{E}\Big[g\Big(F(U_n,V_n,X_{-\infty}^{n-1}[z_{-\infty}^{-p}])\,,\,X_{-l}^{n-1}[z_{-\infty}^{-p}]\Big)\;\mathbf{1}\{\tau[l,n] \le -p\}\Big] \;\underset{p\to\infty}{\longrightarrow}\; 0\;. \tag{4.34}$$

¿From (4.30)–(4.34) we conclude that

$$\mathbb{E}\Big[g(X_l^n)\Big] \;=\; \mathbb{E}\Big[g\Big(F(U_n,V_n,X_{-\infty}^{n-1})\,,\,X_{-l}^{n-1}\Big)\Big]\;. \tag{4.35}$$

This proves consistency. The uniqueness statement in part (iii) is a particular case of part (ii) of Theorem 4.23. This theorem is proved below.

4.2.3 Loss of memory and uniqueness

Let us consider any process $(\widehat{X}_n)$ consistent with the transition probabilities (4.1). Let's denote $(\widehat{\Omega}, \widehat{\mathcal{F}}, \widehat{\mathbb{P}})$ the corresponding probability space. Consistency means the validity of (2.19) for the corresponding expectation $\widehat{\mathbb{E}}$ and the CMMC transition probabilities (P_n). Applied to $g(X_{-\infty}^j) = I[X_0^j = a_0^j]$, the consistency condition implies that

$$\widehat{\mathbb{P}}\Big(\widehat{X}_0^j = a_0^j\Big) \;=\; \widehat{\mathbb{E}}\Big[P(a_0^j | \widehat{X}_{-\infty}^{-1})\Big] \tag{4.36}$$

for each $j \in \mathbb{N}$. To prove uniqueness we must condition further the left-hand side with respect to a remote past $z_{-\infty}^{-p} \in \mathcal{A}_{-\infty}^{-p}$, $p \in \mathbb{N}$. That is, we write

$$\widehat{\mathbb{P}}\Big(\widehat{X}_0^j = a_0^j\Big) \;=\; \int \widehat{\mu}(dz)\, \widehat{\mathbb{E}}\Big[P\Big(a_0^j \;\Big|\; \widehat{X}_{-p+1}^{-1}[z_{-\infty}^{-p}]\Big)\Big] \;, \tag{4.37}$$

where $\widehat{\mu}$ is the law of the process $\widehat{X}$, that is, μ is the measure defined by $\int \widehat{\mu}(dz) f(z) = \widehat{\mathbb{E}} f(\widehat{X})$ for cylinder functions $f : \mathcal{A}_{-\infty}^{\infty} \to \mathbb{R}$. For each past $z_{-\infty}^{-p}$, however, there is only one process consistent with the fixed-past version of the CMMC, and it is the process defined by the corresponding CMMC algorithm (Remark 2.29). We can therefore remove the innermost "hats" and write

$$\widehat{\mathbb{P}}\Big(\widehat{X}_0^j = a_0^j\Big) \;=\; \int \widehat{\mu}(dz)\, \mathbb{E}\Big[P\Big(a_0^j \;\Big|\; X_{-p+1}^{-1}[z_{-\infty}^{-p}]\Big)\Big] \;, \tag{4.38}$$

where now $\mathbb{E}$ is our usual expectation on the variables (U_n, V_n) and $(X_n[z_{-\infty}^{-p}])$ the fixed-past process defined by (4.26). We can now use the results of our previous sections. In particular, by (4.28)

$$\Big| P\Big(a_0^j \;\Big|\; X_{-p+1}^{-1}[z_{-\infty}^{-p}]\Big) - P\Big(a_0^j \;\Big|\; X_{-p+1}^{-1}[w_{-\infty}^{-p}]\Big) \Big| \;\leq\; \mathbf{1}\{\tau[0,j] \leq -p\} \;, \tag{4.39}$$

uniformly in the pasts $z_{-\infty}^{-p}, w_{-\infty}^{-p}$. All the uniqueness results follow from this formula and (4.38):

(i) To obtain (4.24) we just need to write

$$\begin{aligned}
&\mathbb{P}\Big(X_0^j = a_0^j\Big) - \mathbb{P}\Big(X_0^j[z_{-\infty}^{-p}] = a_0^j\Big) \\
&\quad = \int \mu(dz)\,\mathbb{E}\Big[P\Big(a_0^j \;\Big|\; X_{-p+1}^{-1}[w_{-\infty}^{-p}]\Big) - P\Big(a_0^j \;\Big|\; X_{-p+1}^{-1}[z_{-\infty}^{-p}]\Big)\Big] \qquad (4.40)
\end{aligned}$$

and use (4.39). Here μ is the law of the process X.

(ii) To obtain (4.25) we write

$$\begin{aligned}
&\widetilde{\mathbb{P}}\Big(\widetilde{X}_0^j = a_0^j\Big) - \widehat{\mathbb{P}}\Big(\widehat{X}_0^j = a_0^j\Big) \\
&\quad = \int\!\!\int \widetilde{\mu}(dz)\widehat{\mu}(dw)\,\mathbb{E}\Big[P\Big(a_0^j \;\Big|\; X_{-p+1}^{-1}[z_{-\infty}^{-p}]\Big) - P\Big(a_0^j \;\Big|\; X_{-p+1}^{-1}[w_{-\infty}^{-p}]\Big)\Big]\;, \qquad (4.41)
\end{aligned}$$

use (4.39) and take the limit $p \to \infty$. Here $\widehat{\mu}$ is the law of the process $\widehat{X}$.

4.3 Finiteness of regeneration times

To finish this chapter let us state sufficient conditions for the regeneration and renewal times to be finite.

Theorem 4.42 *If*

$$\sum_{m\geq 0}\prod_{k=0}^{m}\alpha_k \;=\; \infty \qquad (4.43)$$

then for each finite interval $[l,m]$,

$$\mathbb{P}(\tau[l,m] > -\infty) \;=\; 1\;. \qquad (4.44)$$

Furthermore, if

$$\lim_{m\to\infty}\prod_{k=0}^{m}\alpha_k \;>\; 0 \qquad (4.45)$$

then for each $l \in \mathbb{Z}$,

$$\mathbb{P}(\tau[l,\infty[> -\infty) \;=\; 1 \; . \tag{4.46}$$

Conditions (4.45) and (4.43) impose lower bounds on the speed of the convergence $\alpha_k \nearrow 1$. In particular both conditions require $\lambda_0 > 0$ [see (4.1)–(4.4)]. In Exercise 4.50 the reader is asked to show that as a result a CMMC with λ_j decreasing at least as $1/j^{2+\delta}$ has finite renewal times if $\delta > 0$. In constrast, if $\lambda_j \sim 1/j^2$ Theorem 4.42 guarantees only the finiteness of the regeneration times for *finite* windows.

It is clear that CMMC transition probabilities admit infinitely many decompositions of the type (4.1). For instance, if the parameters $(\lambda_k)_{k\in\mathbb{N}}$ define such a decomposition with Markovian transitions $P^{(k)}$, then the parameters $\lambda_0/2, (\lambda_k + \lambda_0/2^{k+1})_{k\in\mathbb{N}^*}$ define another decomposition with Markovian transitions $[\lambda_0 P^{(0)}/2^{k+1} + \lambda_k P^{(k)}]/(\lambda_0/2^{k+1} + \lambda_k)$. A more drastic manifestation of this fact is shown in Exercise 7.24. It is natural to wonder as to whether there is an "optimal" such decomposition, at least from the point of view of Theorem 4.42. It is clear that this sense of optimality is related to the fastest possible convergence $\alpha_k \nearrow 1$. In turns, this corresponds to choosing distributions (λ_k) that put as much weight as possible in the lowest values of k. A quick look to the combination (4.1) reveals that λ_0 can not exceed

$$\lambda_0 \;\le\; \sum_{a\in\mathcal{A}} \inf_{\underline{x}} P(a|\underline{x}) \; . \tag{4.47}$$

Furthermore, proceeding inductively,

$$\lambda_0 + \cdots + \lambda_k \;\le\; \inf_{x_{-k}^{-1}} \sum_{a\in\mathcal{A}} \inf_{y_{-\infty}^{-k-1}} P(a|x_{-k}^{-1}\, y_{-\infty}^{-k-1}) \; . \tag{4.48}$$

In Chapter 6 we shall explicitely determine, for large families of chains of infinite orders, CMMC decompositions that *saturate* these inequalities.

The proof of Theorem 4.42, and of other consequences of the regeneration scheme, will be given in Chapter 6. It uses a very simple instance of coupling technique and it relies on an auxiliary Markov chain called the *house-of-cards process*. The relevant properties of this chain are derived in an "intermezzo" chapter, Chapter 5.

4.4 Exercises

Exercise 4.49 Show that the prescription $X_n = F(U_n, V_n, X_{n-1})$ given in (4.3) is indeed a simulation algorithm for the CMMC with transition probabilities (4.1). That is, show that

$$P(a|x_{-\infty}^{n-1}) \;=\; \mathbb{P}\{F(U_n, V_n, x_{-\infty}^{n-1}) = a\}$$

where the left-hand side is given by (4.1) and the function F is defined by the right-hand side of (4.3) for the choices discussed after Definition 4.2.

Exercise 4.50 Let $\alpha_k \in [0,1]$ form a sequence such that $\alpha_k \nearrow 1$. Write $\alpha_k =: 1 - \varepsilon_k$.

(a) Show that

$$\lim_{m\to\infty} \prod_{k=0}^{m} \alpha_k \;>\; 0 \;\Longleftrightarrow\; \sum_{k=0}^{\infty} \varepsilon_k \;<\; \infty\,. \tag{4.51}$$

(b) Show that

$$\exp\Bigl\{-\sum_{k=0}^{\infty}\varepsilon_k - \sum_{k=0}^{\infty}\varepsilon_k^2/2\Bigr\} \;\le\; \prod_{k=0}^{m}\alpha_k \;\le\; \exp\Bigl\{-\sum_{k=0}^{\infty}\varepsilon_k\Bigr\}\,. \tag{4.52}$$

(c) Applying (a) and (b) to the case

$$\varepsilon_k \;=\; \sum_{j=k+1}^{\infty} \lambda_j\,, \tag{4.53}$$

conclude that a CMMC with λ_j decreasing at least as $1/j^{2+\delta}$ has finite renewal times if $\delta > 0$, and finite finite-window regeneration times if $\delta = 0$.

Exercise 4.54 Prove the bounds (4.47)–(4.48). Any idea about the Markovian transitions $P^{(k)}$ that lead to a saturation of these bounds?

Exercise 4.55 Prove that (2.16) implies (4.36).

Chapter 5

Intermezzo: the house-of-cards process

5.1 Recurrence and transience

Given a set of parameters $\alpha_0, \alpha_1, \ldots \in [0,1]$, we define the associated *house-of-cards system of transition probabilities* as the order-1 Markovian system on $\mathcal{A} = \mathbb{N}$ such that

$$\begin{aligned} P(x+1|x) &= \alpha_x \\ P(0|x) &= 1 - \alpha_x \end{aligned} \tag{5.1}$$

and $P(x|x_{-1}) = 0$ otherwise. Thus processes consistent with these transitions climb in a staircase-like fashion and at some instants fall abruptly to the ground. Let us now consider a chain $(W_n)_{n\geq 0}$ starting from 0 and evolving with (5.1). A simulation algorithm for such a chain is:

$$W_n = \begin{cases} 0 & n \leq 0 \\ (W_{n-1}+1)\,\mathbf{1}\{U_n < \alpha_{-W_{n-1}}\} & n \geq 1\,. \end{cases} \tag{5.2}$$

The property of interest for our purposes is that this chain is not positive-recurrent.

Lemma 5.3 *The chain* $(W_n : n \geq 0)$ *is*

(a) *null-recurrent if, and only if,* $\sum_{n\geq 0}\prod_{k=0}^{n}\alpha_k = \infty$, *and*

(b) *transient if, and only if,* $\prod_{k=0}^{\infty}\alpha_k > 0$.

Proof. It is a direct computation, using the definiton of null-recurrence and transience of a Markov chain.

5.2 Return times

As we shall see, [see formula (6.3) below], the distribution of the regeneration times of a CMMC is related to the return-time probabilities of the house-of-cards process

$$\rho_n := \mathbb{P}(W_n = 0) \tag{5.4}$$

for all $s \in \mathbb{Z}$. The following proposition collects a number of useful properties of these quantities.

Proposition 5.5 *Let* $(\alpha_k)_{k\in\mathbb{N}}$ *be an increasing non-negative sequence with* $\alpha_k \nearrow 1$, *and consider the associated house-of-cards process* (W_n) *defined by (5.2). Let* $(\rho_n)_{n\in N}$ *be the return-time probabilities (5.4). Then*

(i) $\sum_{n\geq 0}\prod_{k=0}^{n}\alpha_k = \infty$ *if, and only if,* $\rho_n \to 0$.

(ii) $\prod_{k=0}^{\infty}\alpha_k > 0$ *if, and only if,* $\sum_{n\geq 0}\rho_n < \infty$.

(iii) *If* $(1-\alpha_n)$ *decreases exponentially, so does* ρ_n.

(iv) *If* $\prod_{k=0}^{\infty}\alpha_k > 0$ *and*

$$\limsup_{k\to\infty}\sup_i\left(\frac{1-\alpha_i}{1-\alpha_{ki}}\right)^{1/k} \leq 1\ , \tag{5.6}$$

then $\rho_n \leq const\,(1-\alpha_n)$. *Condition (5.6) holds, for instance, when* $\alpha_n \sim 1-(\log n)^b n^{-\gamma}$ *for* $\gamma > 1$.

In fact, we shall proof a statement slightly stronger than (iv) (Lemma 5.18 below)

Proof of (i)–(iii). Statement *(i)* is just part (a) of Lemma 5.3. To prove parts *(ii)* and *(iii)* we introduce the first-return time

$$\tau \;=\; \inf\{n>0; W_n=0\}\;. \tag{5.7}$$

We see that

$$\mathbb{P}(\tau=1) \;=\; 1-\alpha_0\;, \tag{5.8}$$

$$\mathbb{P}(\tau=n) \;=\; (1-\alpha_{n-1})\prod_{k=0}^{n-2}\alpha_k \quad \text{for } n\geq 2, \tag{5.9}$$

$$\mathbb{P}(\tau=+\infty) \;=\; \prod_{k=0}^{+\infty}\alpha_k\;. \tag{5.10}$$

As the house-of-card process is Markovian,

$$\rho_n \;=\; \sum_{k=1}^{n}\mathbb{P}(\tau=k)\,\rho_{n-k}\;. \tag{5.11}$$

Let us now consider the generating functions

$$F(s) \;=\; \sum_{n=1}^{+\infty}\mathbb{P}(\tau=n)\,s^n \tag{5.12}$$

and

$$G(s) \;=\; \sum_{n=0}^{+\infty}\rho_n\,s^n\;. \tag{5.13}$$

Formula (5.11) implies that these series are related in the form

$$G(s)=\frac{1}{1-F(s)}\;, \tag{5.14}$$

for all $s\geq 0$ such that $F(s)<1$.

It is clear that the radius of convergence of F is at least 1. In fact,

$$F(1) \;=\; \mathbb{P}(\tau<+\infty)\;. \tag{5.15}$$

Moreover, if $\prod_{k\geq 1}\alpha_k > 0$, the radius of convergence of F is

$$\lim_{n\to\infty}[1-\alpha_n]^{-1/n} . \tag{5.16}$$

This follows from the fact that $\mathbb{P}(\tau = n)/(1-\alpha_{n-1}) \to \mathbb{P}(\tau = +\infty) > 0$, by (5.9)–(5.10).

Statement *(ii)* follows from the chain of equivalences:

$$\prod_{k=0}^{\infty}\alpha_k > 0 \iff \mathbb{P}(\tau < +\infty) < 1 \iff G(1) < \infty \iff \sum_{n\geq 0}\rho_n < \infty . \tag{5.17}$$

The first equivalence is part (b) of Lemma 5.3, the second one follows from (5.14) and (5.15), and the last one from the definition (5.13) of G.

To prove statement *(iii)* let us assume that $1-\alpha_m \leq C\gamma^m$ for some constants $C < +\infty$ and $0 < \gamma < 1$. In particular this implies that $\prod_{k=0}^{\infty}\alpha_k > 0$ [Exercise 4.50 (a)] and, hence, by (5.16), that the radius of convergence of F is at least $\gamma^{-1} > 1$. Moreover, by (5.15) and the first equivalence in (5.17) we conclude that $F(1) < 1$. By continuity it follows that there exists $s_0 > 1$ such that $F(s_0) = 1$ and, hence, by (5.14), $G(s) < +\infty$ for all $s < s_0$. By definition of G, this implies that ρ_n decreases faster than ζ^n for any $\zeta \in (s_0^{-1}, 1)$. $\square$

The proof of (iv) is a consequence of the following lemma.

Lemma 5.18 *If* $\prod_{k=0}^{\infty}\alpha_k > 0$ *and*

$$\limsup_{k\to\infty}\sup_i\left(\frac{\mathbb{P}(\tau = 1)}{\mathbb{P}(\tau = ki)}\right)^{1/k} < \frac{1}{\mathbb{P}(\tau < +\infty)} , \tag{5.19}$$

then $\rho_n \leq C\,\mathbb{P}(\tau = n)$ *for some constant* C.

To see how (iv) follows from this lemma, observe that hypotesis (5.6) implies that the left-hand side of (5.19) does not exceed 1 [see (5.9)–(5.10)]. This guarantees the validity of (5.19) because of the first equivalence in (5.17).

Proof of the lemma. We start with the following explicit relation between the coefficients of F and G.

$$\rho_n \;=\; \sum_{k=1}^{n} \sum_{\substack{i_1,\ldots,i_k \geq 1 \\ i_1+\cdots+i_k=n}} \prod_{m=1}^{k} \mathbb{P}(\tau = i_m) \;, \tag{5.20}$$

for $n \geq 1$. This relation can be obtained directly from (5.14) or, alternatively, by decomposing each return time as a sum of k times of first return and using Markovianness. Multiplying and dividing each factor in the rightmost product by $\mathbb{P}(\tau < +\infty)$, this formula can be rewritten as

$$\rho_n \;=\; \sum_{k=1}^{n} \mathbb{P}(\tau < +\infty)^k \sum_{\substack{i_1,\ldots,i_k \geq 1 \\ i_1+\cdots+i_k=n}} \prod_{m=1}^{k} \mathbb{P}(\tau = i_m \,|\, \tau < +\infty). \tag{5.21}$$

At this point we observe the following. If $i_1 + \cdots + i_k = n$, then $\max_{1\leq m\leq k} i_m \geq n/k$ and thus, for g increasing

$$g(n) \;\leq\; g\left(k\, i_{\max}\right) \;,$$

where $i_{\max} = \max_{1\leq m\leq k}\, i_m$. If we apply this to $g(n) = 1/P(\tau = n)$, which is increasing by (5.9), we obtain

$$1 \;\leq\; \frac{\mathbb{P}(\tau = n)}{\mathbb{P}(\tau = k\, i_{\max})} \;. \tag{5.22}$$

This inequality, inserted in (5.21), yields the inequality

$$\begin{aligned} \rho_n \;\leq\;& \mathbb{P}(\tau = n) \sum_{k=1}^{n} \mathbb{P}(\tau < +\infty)^k \\ &\times \sum_{\substack{i_1,\ldots,i_k \geq 1 \\ i_1+\cdots+i_k=n}} \frac{\prod_{m=1}^{k} \mathbb{P}(\tau = i_m \,|\, \tau < +\infty)}{\mathbb{P}(\tau = k\, i_{\max})} \;, \end{aligned} \tag{5.23}$$

We now single out a factor $\mathbb{P}(\tau = i_{\max}|\tau < +\infty) = \mathbb{P}(\tau = i_{\max})/\mathbb{P}(\tau < +\infty)$ from the rightmost product of (5.23). If there are several $i_j = i_{\max}$ we choose the smaller j. We then use (5.19) plus (5.9)–(5.10) to obtain a bound of the form

$$\frac{\mathbb{P}(\tau = i_{\max})}{\mathbb{P}(\tau = k\, i_{\max})} \;\leq\; \delta^k\;, \tag{5.24}$$

valid for k sufficiently large, where

$$\delta \;<\; \frac{1}{\mathbb{P}(\tau < +\infty)}\;. \tag{5.25}$$

Expressions (5.23)–(5.25) imply the inequality

$$\rho_n \;\leq\; C\,\mathbb{P}(\tau = n) \sum_{k=1}^{n} \delta^k\, \mathbb{P}(\tau < +\infty)^{k-1}\, S_k\;, \tag{5.26}$$

for some constant $C > 0$, where

$$\begin{aligned} S_k \;:=\; & \sum_{M=1}^{k} \sum_{\substack{i_M \geq 1, \ell_1 \geq 0, \ell_2 \geq 0\\ i_M + \ell_1 + \ell_2 = n}} \sum_{\substack{1 \leq i_1, \ldots, i_{M-1} < i_M\\ i_1 + \cdots + i_{M-1} = \ell_1}} \prod_{1 \leq m \leq M-1} \mathbb{P}(\tau = i_m \,|\, \tau < +\infty)\\ & \times \sum_{\substack{1 \leq i_{M+1}, \ldots, i_k \leq i_M\\ i_{M+1} + \cdots + i_k = \ell_2}} \prod_{M+1 \leq m \leq k} \mathbb{P}(\tau = i_m \,|\, \tau < +\infty)\;. \end{aligned} \tag{5.27}$$

[M is the smallest j for which $i_j = i_{\max}$ in each summand of (5.23).]

To bound this sum we introduce a sequence of independent random variables $(\tau^{(i)})_{i\in\mathbb{N}}$ with common distribution

$$\mathbf{P}(\tau^{(i)} = j) \;=\; \mathbb{P}(\tau = j \,|\, \tau < +\infty)\;. \tag{5.28}$$

With this probabilistic interpretation we see that

$$S_k \;\leq\; \sum_{i_M=1}^{k} \sum_{j=1}^{n-k+1} \mathbf{P}\Big(\sum_{\substack{1 \leq s \leq k-1\\ s \neq M}} \tau^{(s)} = n - j \Big) \;\leq\; k\;. \tag{5.29}$$

Hence, (5.26) implies

$$
\begin{aligned}
\rho_n &\leq C\,\delta \left[\sum_{k=1}^{\infty} k \; \left[\delta\,\mathbb{P}(\tau < +\infty)\right]^{k-1}\right] \mathbb{P}(\tau = n) \\
&\leq \text{const}\,\mathbb{P}(\tau = n)\;.\;\Box
\end{aligned}
\tag{5.30}
$$

Chapter 6

Mixing properties and perfect simulations for CMMC

6.1 Houses of cards and regeneration

In this chapter we shall use the results of the previous chapter to prove a number of properties of CMMC, including the promised Theorem 4.42. The analysis is based on the following graphical procedure. Consider a fixed window corresponding to the interval $[l, m]$. We first check whether the left endpoint l is a regeneration time for this window. This would be the case if, first of all, $L_l = 0$ (the state at l is independent of the past) and, furthermore, $L_i < i - l$ for $i \in]l, m]$ (the states at times in $]l, m]$ depend only on times not earlier than l). Equivalently, *l is a regeneration time for the window $X_l, \cdots, X_m$ if, and only if, a house-of-cards process starting at the origin at time l does not return to the origin in the interval $]l, m]$.* If this house-of-cards process does visit the origin inside the interval, then we rule out l as a regeneration time and perform a similar test to a house-of-cards process starting at $l - 1$. We continue this way until we find the first $s \leq l$ such that the house-of-cards process starting there manages to pass over the whole interval $[s, m]$ without visiting the origin.

To formalize this argument, let us consider a *coupled* family of house-of-cards processes $((W_n^s : n \ge s) : s \in \mathbb{Z})$, all defined by (5.1) with the same sequence $\alpha_k \nearrow 1$ but started at the origin at different times $s \in \mathbb{Z}$. We couple them by running them with the same common uniform variables (U_n), that is, through a coupling algorithm (Definition 2.36)

$$W_n^s \;=\; \begin{cases} 0 & n \le s \\ (W_{n-1}^s + 1)\,\mathbf{1}\{U_n < \alpha_{-W_{n-1}^s}\} & n \ge s+1\,. \end{cases} \tag{6.1}$$

The process (5.2) is $(W_n) = (W_n^0)$. Given a CMMC, we shall call the *associated house of cards*, the family of processes (6.1) constructed with the (α_k) given in (4.4).

We start with our key identity.

Lemma 6.2 *The following identity holds between the random-order process of a CMMC and its associated house of cards:*

$$\Big\{\tau[l,m] < s\Big\} = \bigcup_{i\in[l,m]} \Big\{W_i^{s-1} = 0\Big\} \tag{6.3}$$

for $s \le l$.

Proof. The asumed monotonicity of the α_k's implies that

$$W_n^s \;\ge\; W_n^t \quad \text{for all } s < t \le n\,. \tag{6.4}$$

Hence, $W_n^s = 0$ implies that $W_n^t = 0$ for $s < t \le n$ and, therefore, all these chains *coalesce* at time n:

$$W_n^s = 0 \implies W_k^s = W_k^t,\; s \le t\; k \ge n\,. \tag{6.5}$$

Expression (4.12) tells us that, if $s \le l$,

$$\tau[l,m] < s \iff \forall j \in [s,l], \exists n \in [j,m] : W_n^{j-1} = 0\,. \tag{6.6}$$

By the coalescing property (6.5), the statement on the right-hand-side is true if, and only if, the same statement is true but with $n \in [l,m]$. By the monotonicity property (6.4) we then conclude

$$\begin{aligned} \tau[l,m] < s \quad &\iff \quad \max\Big\{m < s : \forall n \in [s,t], W_n^m > 0\Big\} < j-1 \\ &\iff \quad \exists n \in [s,t] : W_n^{j-1} = 0\,. \;\square \end{aligned} \tag{6.7}$$

As an immediate corollary of the key identity (6.3), plus time homogeinity, we obtain the following bound on the distribution of regeneration times.

Corollary 6.8 *For a CMMC*

$$\mathbb{P}(\tau[l,m] < s) \;\leq\; \sum_{i=l}^{m} \rho_{i-s+1} \tag{6.9}$$

for $s \leq l \leq m$, where ρ_j are the return times (5.4) of the associated house-of-card process started at time 0 [defined in (5.2)). Estimations for ρ_j are given in Proposition 5.5.

6.2 Finiteness of renewal and regeneration times

As a first application of the key identity we show now how it yields a proof of Theorem 4.42. In view of Lemma 5.3, the following lemma yields such a proof.

Lemma 6.10 *The chain $(W_n : n \geq m)$ [thus, by translation invariance, all the chains $(W_n^s : n \geq s)$] is*

- *(a) null-recurrent if, and only if, $\mathbb{P}(\tau[l,m] > -\infty) = 1$ for each finite interval $[l,m]$, and*
- *(b) transient if, and only if, $\mathbb{P}(\tau[l,\infty[> -\infty) = 1$ for each $l \in \mathbb{Z}$.*

Proof. By translation invariance, the probability of the right-hand side of (6.3) coincides with

$$\mathbb{P}\Big(\bigcup_{i\in[l,m]} \{W_{-s+i+1} = 0\}\Big)\;. \tag{6.11}$$

Therefore, by the monotonicity property (6.4) we have that

$$\mathbb{P}(\tau[l,m] < s) \;\in\; \Big[\mathbb{P}(W_{m-s+1} = 0)\;,\; \sum_{i=1}^{m-l+1} \mathbb{P}(W_{l-s+i} = 0)\Big]\;. \tag{6.12}$$

As $s \to -\infty$ this interval remains bounded away from 0 in the positive-recurrent case, but shrinks to 0 otherwise. Part (a) of the lemma follows from the fact that

$$\mathbb{P}(\tau[l,m] = -\infty) \;=\; \lim_{s\to-\infty} \mathbb{P}(\tau[l,m] < s)\;. \tag{6.13}$$

The proof of part (b) is analogous but simpler. By translation invariance and (6.3) we have that

$$\mathbb{P}(\tau[l,\infty] < s) \;=\; \mathbb{P}\Big(\bigcup_{i\in[l-s+1,\infty]} \{W_i = 0\}\Big) \tag{6.14}$$

which goes to zero as $s \to -\infty$ if, and only if, (W_n) is transient. □

6.3 Mixing properties

Another immediate application of the key identity (6.3) is to obtain relaxation properties, also known as mixing properties, of CMMC. The procedure used in Section 4.2 to construct a CMMC can be thought as a simulation prescription: An initial history is chosen and subsequent states are generated through transition probabilities (through appropriate simulation algorithms. Theorem 4.42 gives conditions guaranteeing that asymptotically this procedure yields the process we are after. Two questions arise naturally at this point:

(1) Can we estimate how far we are from the equilibrium? That is, how long we have to wait to see the influence of the original history become smaller than some acceptable level?

(2) Can we design an alternative procedure with *faster* relaxation times?

Both questions will be studied in these notes. Here we shall use expression (4.24) to give estimates related with the first question. In Section 6.5 below we shall show that it is possible to give the best conceivable answer to question (2): The regeneration scheme off CMMC provides a way to simulate these chains *without* relaxation errors.

Let us now state the estimations that follow from our previous work.

Proposition 6.15 *For CMMC a (X_n),*

$$\left|\mathbb{P}\Big(X_\ell^{m+\ell} = a_0^m\Big) - \mathbb{P}\Big(X_\ell^{m+\ell}[\underline{z}] = a_0^m\Big)\right| \;\le\; \sum_{i=0}^{m} \rho_{i+\ell}\;, \tag{6.16}$$

where ρ_j is the return-time probability (5.4) of the associated house-of-card process started at time 0. Estimations for ρ_j are given in Proposition 5.5.

This proposition follows immediately from the loss-of-memory inequality (4.24) and the bound (6.9) on the distribution of regeneration times.

6.4 Regeneration scheme

As a consequence of Theorem 4.42 and part (b) of Lemma 6.10 we see that if $\prod_{k=0}^{\infty}\alpha_k > 0$, almost all realizations of the CMMC exhibit a strictly increasing sequence (s_i) of renewal times. In this case, the process may be visualized as a sequence of independent blocks, of random length $s_{i+1}-s_i$. This defines a *regeneration scheme*. The formal statement of this property is as follows.

Let $\mathbf{N} \in \{0,1\}^{\mathbb{Z}}$ be the random Boolean variables defined by

$$\mathbf{N}(j) := \mathbf{1}\{\tau[j,\infty] = j\}\,. \tag{6.17}$$

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 ℓ, $T_\ell < T_{\ell+1}$ and $T_0 \le 0 < T_1$.

Corollary 6.18 *Let us consider a CMMC. If $\prod_{k=0}^{\infty}\alpha_k > 0$, then the process* $\mathbf{N}$ *defined in (6.17) is a stationary renewal process with renewal distribution*

$$\mathbb{P}(T_{\ell+1} - T_\ell \ge m) \;=\; \rho_m \tag{6.19}$$

for $m > 0$ and $\ell \neq 0$, where ρ_m is the return time defined in (5.4). Furthermore, the random vectors $\xi_\ell \in \cup_{n\geq 1}\mathcal{A}^n$, $\ell \in \mathbb{Z}$, defined by $\xi_\ell = (X_{T_\ell}, \cdots X_{T_{\ell+1}-1})$ are mutually independent and identically distributed with conditional distribution

$$\mathbb{P}\Big(\xi_\ell = (a_{T_\ell}, \ldots, a_{T_{\ell+1}-1}) \;\Big|\; (U_n)\Big) \;=\; P^0(a_{T_\ell}) \;\cdots\; P^{(L_{T_{\ell+1}}-1)}(a_{T_{\ell+1}-1} | a_{T_\ell}^{T_{\ell+1}-2}) \;. \tag{6.20}$$

Schemes of this nature have been obtained by Berbee (1987), in the context of chains of Type B (see Definition 3.12), and by Lalley (1986, 2000) for chains of Type C. The present construction, valid for the more general Type A chaines, was was done by Ferrari et al (2000).

Proof. The stationarity of $\mathbf{N}$ follows immediately from the construction. Let

$$f(j) \;:=\; \mathbb{P}\Big(\mathbf{N}(-j) = 1 \,|\, \mathbf{N}(0) = 1\Big) \tag{6.21}$$

for $j \in \mathbb{N}^*$. To see that $\mathbf{N}$ is a renewal process it is sufficient to show that

$$\mathbb{P}\Big(\mathbf{N}(s_\ell) = 1\,;\, \ell = 1, \ldots, n\Big) \;=\; \beta \prod_{\ell=1}^{n-1} f(s_{\ell+1} - s_\ell) \tag{6.22}$$

for arbitrary integers $s_1 < \cdots < s_k$. [From Poincaré's inclusion-exclusion formula, a measure on $\{0,1\}^{\mathbb{Z}}$ is characterized by its value on cylinder sets of the form $\{\zeta \in \{0,1\}^{\mathbb{Z}} : \zeta(s) = 1,\, s \in S\}$ for all finite $S \subset \mathbb{Z}$. For $S = \{s_1, \ldots, s_k\}$, a renewal process must satisfy (6.22).] For $j \in \mathbb{Z}$, $j' \in \mathbb{Z} \cup \{\infty\}$, define

$$H[j, j'] \;:=\; \begin{cases} \{U_{j+\ell} < \alpha_\ell, \ell = 0, \ldots, j' - j\}, & \text{if } j \leq j' \\[2ex] \text{“full event”}, & \text{if } j > j' \end{cases} \tag{6.23}$$

With this notation,

$$\mathbf{N}(j) = \mathbf{1}\{H[j, \infty]\}, \quad j \in \mathbb{Z}. \tag{6.24}$$

and

$$\mathbb{P}\Big(\mathbf{N}(s_\ell) = 1\,;\, \ell = 1, \ldots, n\Big) \;=\; \mathbb{P}\Big\{\bigcap_{\ell=1} H[s_\ell, \infty]\Big\} \tag{6.25}$$

¿From monotonicity we have for $j < j' < j'' \leq \infty$,

$$H[j,j''] \cap H[j',j''] \;=\; H[j,j'-1] \cap H[j',j''], \tag{6.26}$$

and then, with $s_{n+1} = \infty$ we see that (6.25) equals

$$\prod_{i=1}^{n} \mathbb{P}\Big\{H[s_\ell, s_{\ell+1} - 1]\Big\}\,, \tag{6.27}$$

which equals the right hand side of (6.22). Hence $\mathbf{N}$ is a renewal process.

On the other hand, by stationarity,

$$\mathbb{P}(T_{\ell+1} - T_\ell \geq m) \;=\; \mathbb{P}\Big(\tau[-1,\infty] < -m+1 \;\Big|\; \tau[0,\infty] = 0\Big) \tag{6.28}$$

and, hence, by the key identity (6.3)

$$\mathbb{P}(T_{\ell+1} - T_\ell \geq m) \;=\; \mathbb{P}(W_{-1}^{-m+1} = 0) \;=\; \rho_m \;\;, \tag{6.29}$$

proving (6.19).

The independence of the random vectors ξ_ℓ follows from the definition of T_ℓ. □

6.5 Perfect simulation

To explain what is a perfect-simulation algorithm we start with the important definition of *stopping time*.

Definition 6.30 (Stopping time) Let (U_n) be a sequence of random variables on some set $\mathbf{U}$. 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 $U_1, \ldots, U_j$. That is, if there exist events $A_j \subset \mathbf{U}^j$ such that

$$\{T \leq j\} \;=\; \{(U_1, \ldots, U_j) \in A_j\} \tag{6.31}$$

Example 6.32 Let $c \in (0,1)$, $\mathbf{U} = [0,1]$, (U_n) be a sequence of random variables uniformly distributed in $\mathbf{U}$ and $T :=$ first time a U_n is less than c:

$$T := \min\{n \geq 1 \,:\, U_n < c\} \tag{6.33}$$

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

$$A_j \;=\; \{U_1 > c, \ldots, U_{j-1} > c, U_j < c\} \tag{6.34}$$

and the law of T is geometric with parameter c:

$$\mathbb{P}(T > n) \;=\; (1-c)^n \tag{6.35}$$

In contrast, variables whose definition involves the *last* time in which a certain condition is satisfied are *not* stopping times.

Definition 6.36 *A* **perfect simulation** *for a process* (X_n) *is a family* $\{(T_{[l,m]}, F_{[l,m]}) : l \leq m \in \mathbb{Z}\}$, *where for each* $l \leq m \in \mathbb{Z}$

(i) $T_{[l,m]}$ *is a stopping time on the variables* $(U_{m-n})_{n\geq 0}$,

(ii) $\mathbb{P}(T_{[l,m]} < \infty) = 1$, *and*

(iii) $F_{[l,m]} : (U_{l-T_{[l,m]}}, \ldots, U_m) \to \mathcal{A}_l^m$ *is such that*

$$\mathbb{P}\Big(X_l^m = a_l^m\Big) \;=\; \mathbb{P}\Big((F_{[l,m]})_l^m = a_l^m\Big) \tag{6.37}$$

for each $a_l^m \in \mathcal{A}_l^m$.

Perfect simulations, therefore, allow to obtain, in a finite time, samples of windows distributed *exactly* as the process, without relaxation errors. The regeneration scheme provides a natural perfect-simulation algorithm for CMMC.

Proposition 6.38 *For CMMC with* $\sum_{m\geq 0}\prod_{k=0}^{m}\alpha_k = \infty$ *there exist a perfect simulation. The stopping times are* $T_{[l,m]} = m - \tau[l,m]$ *and*

$$\begin{aligned} (F_{[l,m]})_{\tau[l,m]} &= f^{(0)}(U_{\tau[l,m]}) \\ &\vdots \\ (F_{[l,m]})_m &= f^{(L_m)}(U_m, (F_{[l,m]})_{m-L_m}^{m-1}) \end{aligned} \tag{6.39}$$

The order-variables (L_n) *are defined in (4.6), and the* $f^{(k)}$ *are the simulation algorithms (4.3).*

6.6 Exercises

Exercise 6.40 Consider a CMMC defined by

$$P(a|\underline{x}) = \lambda_0\, P^{(0)}(a) + \lambda_1\, P^{(1)}(a|x_{-1}) \tag{6.41}$$

with $\lambda_0 + \lambda_1 = 1$.

(a) Show that for any $l \in \mathbb{Z}$, $\tau[l]$ has a geometric distribution and determine its parameters. Hint: show that

$$\tau[l] = \max\{n \leq l : L_n = 0\}\ . \tag{6.42}$$

(b) Conclude that for all $n \geq l$, $(X_{n+\tau[l]})_{n\geq 0}$ and $(X_{\tau[l]-n})_{n\geq 0}$ are independent.

Exercise 6.43 Consider now a CMMC

$$P(a|\underline{x}) = \lambda_0\, P^{(0)}(a) + \sum_{i=1}^{k}\lambda_i\, P^{(i)}(a|x_{-i}^{-1}) \tag{6.44}$$

with $\lambda_0 + \cdot + \lambda_k = 1$ and $2 \leq k < \infty$.

(a) Show that formula 6.42 is no longer valid.

(b) Show that

$$\mathbb{P}(\tau[l] \geq l - s) \geq \lambda_0(\lambda_0 + \lambda_1)\cdots(\lambda_0 + \cdots + \lambda_{\min\{k-1,s\}})\ . \tag{6.45}$$

Chapter 7

Every chain of infinite order is a CMMC and a VLMC

7.1 Chains as CMMC

The overall goal of this section is summarized in the following theorem. [For the definition of continuity and other hypotheses of chains of infinite order see Definition 3.1. For the defintion of CMMC and related notation see Section 4.]

Theorem 7.1 *Every chain of infinite order with a continuous system of transition probabilities is a CMMC.*

The method of proof is of interest in itself. It is based on a rather general prescription to decompose conditional probabilities as convex combinations of Markovian processes. This prescription, in fact, is very flexible and leaves room for user-defined choices. Our presentation is organized so to clerly exhibit this flexibility, with the hope that readers will put it to good use in specific applications.

Definition 7.2 *A* **CMMC partition** *is a pair* $(\{\mathcal{P}^{\underline{x}} : \underline{x} \in \underline{\mathcal{A}}\}, \{B_k : k \in \mathbb{N}\})$ *where:*

(i) *Each* $\mathcal{P}^{\underline{x}}$ *is a partition of the interval* $[0,1]$ *of the form*

$$[0,1] \;=\; \bigcup_{\substack{a\in\mathcal{A}\\ n\in\mathbb{N}}} I_a^{x_1^n}\;, \tag{7.3}$$

with sets $I_a^{x_1^n}$ *formed by unions of intervals. These sets may be different for different* $\underline{x}$, *except those for* $n=0$ *for which we use the abusive notation* $I_a^{x_1^0}$.

(ii) *The sets* B_k *form a partition of* $[0,1]$

(iii) *The partitions* $\{\mathcal{P}^{\underline{x}}\}$ *and* (B_k) *are such that*

$$\Bigl(\bigcup_{\substack{a\in\mathcal{A}\\ 0\le k\le n}} I_a^{x_1^k}\Bigr) \;\supset\; \bigcup_{k=0}^{n} B_k \tag{7.4}$$

for each $n\in\mathbb{N}$ *and* $\underline{x}\in\underline{\mathcal{A}}$.

Proposition 7.5 *A CMMC decomposition defines an algorithm for a CMMC.*

Proof. We have to define λ_k and $f^{(k)}$ in (4.3). For the former we take $\lambda_k = \text{length}(B_k)$. We then consider, for each $\underline{x}\in\mathcal{A}$ and $n\in\mathbb{N}$, the sets

$$J_a^{x_1^n} \;:=\Bigl(\bigcup_{0\le k\le n} I_a^{x_1^k}\Bigr)\cap B_n\;. \tag{7.6}$$

Condition (7.4) implies that the sets $\{J_a^{x_1^n}\;:\;a\in\mathcal{A}\}$ form a partition of $[\alpha_{n-1},\alpha_n]$. Finally we define

$$f^{(k)}(V_n, x_{n-k}^{n-1}) \;=\; a \qquad \text{if} \quad \lambda_k V_k \in J_a^{x_{n-k}^{n-1}}\;.\;\square \tag{7.7}$$

Theorem (7.1) follows from the previous and the following propositions.

Proposition 7.8 *Every chain of infinite order with continuous transition probabilities defines a CMMC partition.*

Proof. For each $\underline{x} \in \underline{\mathcal{A}}$, the partition $\mathcal{P}^{\underline{x}}$ is defined as follows. We first determine numbers

$$\begin{aligned} r_0(a) &:= \inf_{\underline{z} \in \underline{\mathcal{A}}} P(a|\underline{z}) \\ &\vdots \\ r_k(a|x_{-1}^{-k}) &= \inf_{\underline{z} \in \underline{\mathcal{A}}} P(a|x_{-k}^{-1}\,\underline{z})\;, \quad k \geq 1\,, \end{aligned} \tag{7.9}$$

defined for each $k \in \mathbb{N}$, $g \in \mathcal{A}$ and $x_{-1}^{-k} \in \mathcal{A}_{-k}^{-1}$. [These functions are denoted $g(i_0|i_{-1}, \ldots, i_{-k})$ by Berbee (1987)]. Then we take the differences

$$\begin{aligned} \Delta_0(a) &:= r_0(g) \\ \Delta_k(a|x_{-k}^{-1}) &:= r_k(a|x_{-k}^{-1}) - r_{k-1}(a|x_{-k+1}^{-1})\;, \quad \text{for } k \geq 1 \end{aligned} \tag{7.10}$$

for $a \in \mathcal{A}$. We take now a partition of $[0,1]$ formed by sets $I_a^{x_{-n}^{-1}}$ such that:

(i) For $a \in \mathcal{A}$, $k \geq 0$,

$$\text{length}\Big(I_a^{x_{-k}^{-1}}\Big) \;=\; \Delta_k(a|x_{-k}^{-1})\;. \tag{7.11}$$

(ii) These intervals are disposed in increasing lexicographic order with respect to a and k in such a way that the left extreme of one interval coincides with the right extreme of the precedent.

That is, the intervals are disposed, along the interval $[0,1]$ in the form

$$I_{a_1}^0, I_{a_2}^0, \ldots, I_{a_{|\mathcal{A}|}}^0, I_{a_1}^{x_{-1}}, I_{a_2}^{x_{-1}}, \ldots, I_{a_{|\mathcal{A}|}}^{x_{-1}}, I_{a_1}^{x_{-2}^{-1}}, \ldots$$

($|\mathcal{A}|$ is the cardinality of the alphabet). To complete the algorithm, we consider the numbers

$$\alpha_k := \min_{x_{-k}^{-1} \in \mathcal{A}_{-1}^{k}} \sum_{a \in \mathcal{A}} r_k(a|x_{-k}^{-1}) \quad , \tag{7.12}$$

$k \in \mathbb{N}$. By the continuity of the chain, $\alpha_k \nearrow 1$. Finally, we take the sets

$$B_k = [\alpha_{k-1}, \alpha_k] \tag{7.13}$$

for $k \geq 1$ and $a_{-1} = 0$. □

We observe that the decomposition just obtained saturates the inequalities (4.47)–(4.48).

7.2 Chains with a regeneration scheme as VLMC

It is almost obvious that a chain (X_n) with a regeneration scheme can be embedded in a VLMC. Indeed, let for instance $\mathbf{N}_n$ be the random Boolean variables defined in equation (6.17). We introduce the process $(Z_n) = (X_n, \mathbf{N}_n)$ taking values in $\mathcal{A} \times \{0, 1\}$. We then have

$$\mathbb{P}\Big(Z_0 = (a, \kappa) \;\Big|\; (\underline{X}, \underline{\mathbf{N}}) = (\underline{x}, \underline{\mathbf{n}})\Big) = \mathbb{P}\Big(Z_0 = (a, \kappa) \;\Big|\; X_{\ell(\underline{\mathbf{N}})}^{-1} = x_{\ell(\underline{\mathbf{n}})}^{-1}\Big) \tag{7.14}$$

with lag function defined by

$$\ell(\underline{\mathbf{n}}) = \sup\{s \leq 0 : \mathbf{n}_s = 1\} \tag{7.15}$$

with the convention that when $\ell(\underline{x}) = 0$, the transition probability is actually independent of the past.

The observation that a chain with regeneration can be thought as a VLMC is, however, of little practical value. The extra "flag" variables $\mathbf{N}_n$ needed for the embedding can not be deduced from the values taken by the variables X_n. They are part of the simulation machinery, exactly as the uniform random variables (U_n).

Let us conclude with an example showing how tricky the relation between VLMC and CMMC can be. Let us consider the sparse VLMC introduced in Section 3.3. This is in fact one of the simplest non-trivial possible VLMC. We recall the reader that this VLMC takes values in $\mathcal{A} = \{0, 1\}$, and its lag function is $\ell(\underline{x}) = \ell$ if $x_{-1} = 0 = \cdots = x_{-\ell}, x_{-\ell-1} = 1$. Its transition probabilities are defined by $P(1|\underline{x}) = q_{\ell(\underline{x})}$ with $0 < q_k < 1$. Let assume in addition that

$$1 \; > \; q_n \searrow q_\infty \; > \; 0\,. \tag{7.16}$$

We construct the associated CMMC using the prescription given in the proof of Proposition 7.8. The results (whose verification is left to the reader) are the following. The parameters of the convex combination are

$$\lambda_k \;=\; \begin{cases} 1 - q_1 + q_\infty & k = 0 \\ q_k - q_{k+1} & k \geq 1\,. \end{cases} \tag{7.17}$$

The Markovian transition probabilities for $k = 0$ are defined by

$$p^{(0)}(1) \;=\; q_\infty/\lambda_0\,, \tag{7.18}$$

while for $k \geq 1$

$$p^{(k)}(1|x_{-k-1}^{-1}) \;=\; \begin{cases} 0 & \text{if } x_{-k-1}^{-1} = 0_{-k-1}^{-1} \\ 1 & \text{otherwise}\,. \end{cases} \tag{7.19}$$

In particular, we notice that the decomposition of the transition probabilities $p(1|0_{-n}^{-1}\,1\,x_{-\infty}^{-n-2})$ involve all Markovian orders, despite the fact that they do not depend on $x_{-\infty}^{-n-2}$.

7.3 Exercises

Exercise 7.20 Prove that every CMMC is a chain with complete connections with continuous transition probabilities.

Exercise 7.21 Prove that every hidden Markov model is a chain of infinite order with continuous transition probabilities. More specifically, let (X_n) be the observable chain and (S_n) the hidden Markov chain. Denote τ_0^S the regeneration time for S_0. Then prove that

$$\sup_{x,y}\Big|P(a|\underline{x}) - P(a|x_s^{-1}\,y_{-\infty}^{s-1})\Big| \;\leq\; \mathbb{P}(\tau_0^S < s) \tag{7.22}$$

for every $a \in \mathring{A}$ and $s \leq 0$. This issue was already discussed in Exercise 6.40. What else is needed to make the HMM a chain of type A?

Exercise 7.23 Verify that for the sparse VLMC satisfying (7.16), the partition on the proof of Proposition 7.8 yields (7.17)–(7.19).

Exercise 7.24 Consider a CMMC defined on $\mathcal{A} = \{0,1\}$ by

$$P(1|\underline{x}) \;=\; \sum_{k=1}^{\infty} \eta_k g(a, x_{-k}) \;, \tag{7.25}$$

with $0 \le \eta_k \le 1$, $\sum_k \eta_k = 1$ and

$$g(a,x) \;=\; (1-\varepsilon)\,\mathbf{1}\{x=1\} + \varepsilon\,\mathbf{1}\{x=0\}\;. \tag{7.26}$$

(i) Write the decomposition given in Proposition 7.5.

(ii) Calling λ_k the coefficients or the decomposition obtained in (i), show that $\lambda_0 \ge \varepsilon$. Observe that this is true even if there exists an $\ell \in \mathbb{N}$ such that $\eta_k = 0$ for $0 \le k \le \ell$.

Chapter 8

Markov approximations for chains of infinite order

8.1 Introduction

This chapter addresses the following question: How well can we approximate an infinite-order chain by Markov chains? This leads to a second, technical, question: Which distance should we use to measure the quality of an approximation? We adopt here Ornstein's $\overline{d}$-distance.

The main result of this chapter is an estimation of the speed of convergence —in the $\overline{d}$-distance— of the *canonical* Markov approximation of chains of infinite order. If the continuity rates of the chain are summable, we show that the speed of convergence is at worst proportional to these rates. Our result applies to Type A chains with summable continuity rates. This is a slight improvement of the result in Bressaud, Fernández and Galves (1999a), which holds for chains of type B with summable log-continuity rates.

It is known that type B chains with summable log-continuity rates are weak Bernoulli (Ledrappier 1974). This implies, by Ornstein theorem (Ornstein 1974), that the process is the $\overline{d}$-limit of its canonical k-step Markov approximations. Curiously, this indirect argument appears to be the only published proof of such $\overline{d}$-convergence. In contrast, our construction below yields an explicit and direct proof. Ornstein and Weiss (1990) have constructed a remarkable "guessing scheme" for $\overline{d}$-limits of aperiodic Markov processes, based on observed data. Nevertheless, these approaches do not shed light on how well the chains can be appoximated by Markov processes.

In this chapter we analyze precisely this issue for the chains with complete connections and the less sophisticated of the approximation schemes: the canonical k-step Markov. Our results show that the continuity rates of the chain directly determine —in the summable case— the speed of convergence of the approximation. Our method is constructive and straightforward. We exhibit explicit couplings between the original chain and each of its k-step approximations. The couplings are such that: (i) if the two component processes have been equal for a certain number of steps, there is a large probability that they will remain so in the next step [formula (8.40)], and (ii) if the components fail to be equal at some step there is a nonzero probability that they will become equal at the next one [formula (8.41)]. As a consequence, the coupled processes tend to coicide most of the time, and separations do not last too long [formula (8.48)]. This yields a small $\overline{d}$-distance between the original process and its k-step approximations.

8.2 Definitions and main result

The first definition follows Ornstein (1974).

Definition 8.1 *The* canonical Markov approximation of order $k \in \mathbf{N}$ *of a process* $(X_n)_{n\in\mathbb{Z}}$ *is the stationary Markov chain of order* k *having as transition probabilities,*

$$P^{[k]}(b \mid a_1, \ldots, a_k) := \mathbb{P}(X_{k+1} = b | X_j = a_j, 1 \le j \le k) \tag{8.2}$$

for all integer $k \geq 1$ *and* $a_1, \ldots, a_k, b \in A$.

Definition 8.3 *The* distance $\overline{d}$ *between two stationary processes* X *and* Y *is defined as*

$$\overline{d}(X,Y) = \inf \left\{ \mathbb{P}(\widetilde{X}_0 \neq \widetilde{Y}_0) \ : \ (\widetilde{X},\widetilde{Y}) \textit{ stationary coupling of } X \textit{ and } Y \right\}.$$

We now state our main result.

Theorem 8.4 *Let $X = (X_n)_{n\in\mathbb{Z}}$ be a chain of infinite order of type A with summable continuity rate $(\beta_s)_{s\geq 1}$. Then there is a constant $K > 0$ such that, for all $k \geq 1$,*

$$\bar{d}(X, X^{[k]}) \leq K\ \beta_k\ ,$$

where $X^{[k]} = (X_n^{[k]})_{n\in\mathbb{Z}}$ is the canonical Markov approximation of order k of the process X.

8.3 Construction of the coupling

Consider two time-homogeneours systems of transition probabilities $P(\,\cdot\,|\,\cdot\,)$ and $Q(\,\cdot\,|\,\cdot\,)$. We want to construct a coupling algorithm for them, with the following properties:

(a) it loads the diagonal as much as possible, and

(b) each step of the coupling depends only on the past.

This will be done through a graphical procedure (cf. Definition 2.47).

Given two pasts $\underline{x}, \underline{y}$ and an element a of the alphabet $\mathcal{A}$, let us define

$$\begin{aligned} t_a(\underline{x},\underline{y}) &:= P(a\,|\,\underline{x}) \wedge Q(a\,|\,\underline{y}) \\ r_a(\underline{x},\underline{y}) &:= (P(a\,|\,\underline{x}) - Q(a\,|\,\underline{y})) \vee 0 \\ s_a(\underline{x},\underline{y}) &:= (Q(a\,|\,\underline{y}) - P(a\,|\,\underline{x})) \vee 0\ . \end{aligned} \tag{8.5}$$

Notice that

$$\begin{array}{rl} \text{either} & r_a(\underline{x},\underline{y}) = 0 \text{ and } s_a(\underline{x},\underline{y}) > 0 \\ \text{or} & r_a(\underline{x},\underline{y}) > 0 \text{ and } s_a(\underline{x},\underline{y}) = 0 \end{array} \tag{8.6}$$

and that

$$t_a(\underline{x},\underline{y}) + r_a(\underline{x},\underline{y}) = P(a|\underline{x}) \tag{8.7}$$

$$t_a(\underline{x},\underline{y}) + s_a(\underline{x},\underline{y}) = Q(a|\underline{y})\ . \tag{8.8}$$

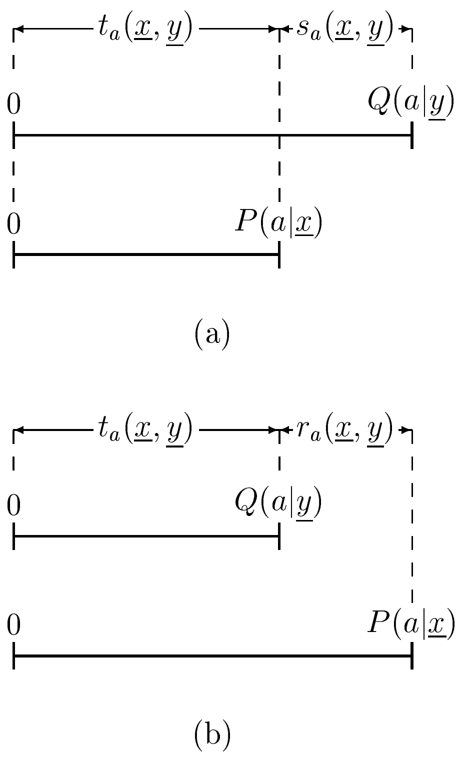

Figure 8.1: Graphic representation of Definition (8.5). (a) Case with $r_a(\underline{x}, \underline{y}) = 0$. (b) Case with $s_a(\underline{x}, \underline{y}) = 0$

Figure 8.1 gives a graphic representation of these identities.

As a consequence,

$$\sum_{a\in A} t_a(\underline{x},\underline{y}) + \sum_{a\in A} r_a(\underline{x},\underline{y}) \;=\; 1 \tag{8.9}$$
$$\sum_{a\in A} t_a(\underline{x},\underline{y}) + \sum_{a\in A} s_a(\underline{x},\underline{y}) \;=\; 1\,. \tag{8.10}$$

Identities (8.9)/(8.10) enable us to define two partitions of $[0,1]$, each one formed by the non-empty sets of the following $2|\mathcal{A}|$ intervals:

$$\{T_1^{\underline{x},\underline{y}},\ldots,T_{|\mathcal{A}|}^{\underline{x},\underline{y}},R_1^{\underline{x},\underline{y}},\ldots,R_{|\mathcal{A}|}^{\underline{x},\underline{y}}\} \;\text{ and }\; \{T_1^{\underline{x},\underline{y}},\ldots,T_{|\mathcal{A}|}^{\underline{x},\underline{y}},S_1^{\underline{x},\underline{y}},\ldots,S_{|\mathcal{A}|}^{\underline{x},\underline{y}}\} \tag{8.11}$$

These are intervals of lengths

$$|T_a^{\underline{x},\underline{y}}| = t_a(\underline{x},\underline{y})\,, \quad |R_a^{\underline{x},\underline{y}}| = r_a(\underline{x},\underline{y}) \quad \text{and} \quad |S_a^{\underline{x},\underline{y}}| = s_a(\underline{x},\underline{y})\,,$$

for all $a \in A$

We define the transition probabilities $\widetilde{P}((a,b) \mid (\underline{x},\underline{y}))$ as

$$\widetilde{P}((a,b) \mid (\underline{x},\underline{y})) := \begin{cases} |T_a^{\underline{x},\underline{y}}| & \text{if } a=b, \\ |R_a^{\underline{x},\underline{y}} \cap S_b^{\underline{x},\underline{y}}| & \text{if } a\neq b \end{cases} \tag{8.12}$$

(see figure 8.2). The corresponding simulation algorithm is

$$f(u,\underline{x},\underline{y}) \;=\; (a,a) \quad \text{if } u \in T_a^{\underline{x},\underline{y}}\,, \tag{8.13}$$
$$f(u,\underline{x},\underline{y}) \;=\; (a,b) \quad \text{if } u \in R_a^{\underline{x},\underline{y}} \cap S_b^{\underline{x},\underline{y}}\,, \tag{8.14}$$

with $a \neq b$ in the second line.

The properties of this coupling are summarized in the following theorem

Theorem 8.15 *If the chains with transition probabilities P and Q are both of type A, so is the coupling defined through (8.12)–(8.14). More explicitly,*

$$\widetilde{\beta}_s \;\leq\; \text{const}\;(\beta_s^P \vee \beta_s^Q)\,, \tag{8.16}$$

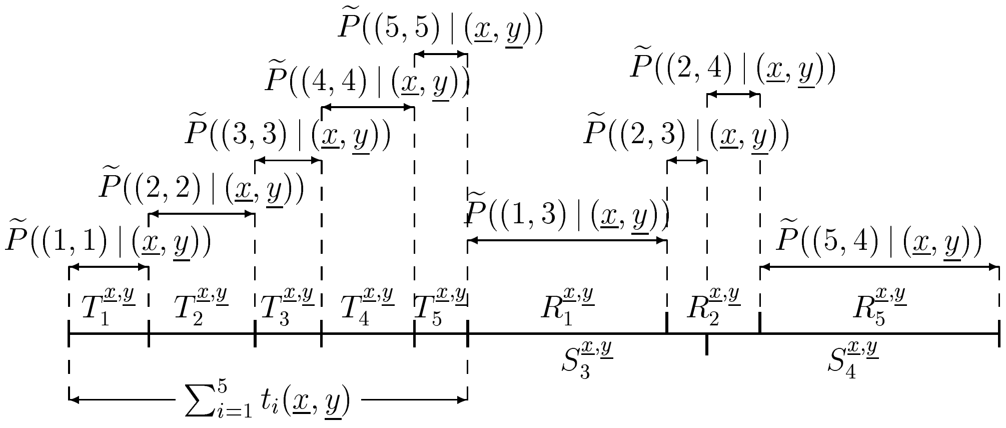

Figure 8.2: Case $|\mathcal{A}| = 5$, $P(a|\underline{x}) > Q(a|\underline{y})$ for $a = 1, 2, 5$ and $P(a|\underline{x}) < Q(a|\underline{y})$ for $a = 3, 4$

and

$$\sum_{a,b\in\mathcal{A}} \inf_{\underline{x},\underline{y}} \widetilde{P}\Big((a,b)\ \Big|\ (\underline{x},\underline{y})\Big) \ \geq$$

$$\geq\ \Big[\sum_{a\in\mathcal{A}} \inf_{\underline{x}} P(a|\underline{x})\Big] \ \wedge\ \Big[\sum_{a\in\mathcal{A}} \inf_{\underline{x}} Q(a|\underline{x})\Big]\ . \tag{8.17}$$

We remark that, even if the transitions P and Q are chains of type B, this coupling is not in general a chain of type B, because all pairs (a, b) with

$$\inf_{\underline{x},\underline{y}} \widetilde{P}\Big((a,b)\ \Big|\ (\underline{x},\underline{y})\Big) \ =\ 0\ .$$

This happens whenever $R_a^{\underline{x},\underline{y}} \cap S_b^{\underline{x},\underline{y}} = \emptyset$.

Proof.

Non-nullness

$$\sum_{a,b\in\mathcal{A}} \inf_{\underline{x},\underline{y}} \widetilde{P}\Big((a,b)\ \Big|\ (\underline{x},\underline{y})\Big) \ \geq\ \sum_{a\in\mathcal{A}} \inf_{\underline{x},\underline{y}} \widetilde{P}\Big((a,a)\ \Big|\ (\underline{x},\underline{y})\Big) \tag{8.18}$$

But the right-hand side is

$$\sum_{a\in\mathcal{A}} \inf_{\underline{x},\underline{y}} \Big[P(a|\underline{x}) \ \wedge\ Q(a|\underline{y}) \Big] \tag{8.19}$$

$$\geq\ \Big[\sum_{a\in\mathcal{A}} \inf_{\underline{x}} P(a|\underline{x}) \Big] \ \wedge\ \Big[\sum_{a\in\mathcal{A}} \inf_{\underline{x}} Q(a|\underline{x}) \Big]\ . \tag{8.20}$$

Continuity Let us denote

$$\Delta_m(a,b) \ = \ \sup_{\underline{x},\underline{y},\underline{u},\underline{w}} \Big| \widetilde{P}\Big((a,b) \ \Big|\ (\underline{x},\underline{y}) \Big) - \widetilde{P}\Big((a,b) \ \Big|\ (x_{-m}^{-1} u_{-\infty}^{-m-1}, y_{-m}^{-1} w_{-\infty}^{-m-1}) \Big) \Big|\ . \tag{8.21}$$

Case $a=b$**:**

$$\Delta_m(a,a) \ = \ \sup_{\underline{x},\underline{y},\underline{u},\underline{w}} \Big| t_a(\underline{x},\underline{y}) - t_a(x_{-m}^{-1} u_{-\infty}^{-m-1}, y_{-m}^{-1} w_{-\infty}^{-m-1}) \Big| \tag{8.22}$$

Using $|\alpha\wedge\beta - \alpha'\wedge\beta'| \leq |\alpha-\alpha'| \vee |\beta-\beta'|$ we get

$$\Delta_m(a,a) \ \leq \ \sup_{\underline{x},\underline{y},\underline{u},\underline{w}} \Big[|P(a|\underline{x}) - P(a|x_{-m}^{-1} u_{-\infty}^{-m-1})| \ \vee\ |Q(a|\underline{y}) - Q(a|y_{-m}^{-1} w_{-\infty}^{-m-1})| \Big]\ . \tag{8.23}$$

Hence,

$$\Delta_m(a,a) \ \leq\ \beta_m^P \vee \beta_m^Q\ . \tag{8.24}$$

Case $a\neq b$**:** Computations are similar but longer. □

8.4 Proof of the theorem

We are ready to prove Theorem (8.4).

8.4.1 Bound among transition probabilities

Let $P^{[k]}$ be the transition probability defined by (8.2). We shall abbreviate our notation and write $P^{[k]}(a \,|\, \underline{y})$ instead of $P^{[k]}(a \,|\, y_{-k}, \ldots, y_{-1})$. We also denote $\underline{x} \stackrel{k}{=} \underline{y}$ to indicate that $x_{-k}^{-1} = y_{-k}^{-1}$. In particular

$$\underline{x} \stackrel{k}{=} \underline{y} \implies P^{[k]}(a \,|\, \underline{y}) = P^{[k]}(a \,|\, \underline{x}) \quad \forall\, a \in \mathcal{A}\,. \tag{8.25}$$

The following proposition contains the only property of the canonical approximation needed for the result.

Proposition 8.26

$$\inf_{\underline{u}\,:\,\underline{u} \stackrel{k}{=} \underline{y}} P(a \,|\, \underline{u}) \;\leq\; P^{[k]}(a \,|\, \underline{y}) \;\leq\; \sup_{\underline{u}\,:\,\underline{u} \stackrel{k}{=} \underline{y}} P(a \,|\, \underline{u}). \tag{8.27}$$

Remark 8.28 In fact, (8.27) is the only property of the Markov transitions used in the sequel. Thus, our results apply to any Markov approximation scheme, not necessarily the canonical one, satisfying (8.27).

8.4.2 The proof

Positive probability of coincidence

By the definition of the coupling,

$$\mathbb{P}\Big(\widetilde{X}_0 = \widetilde{X}_0^{[k]} \;\Big|\; (\underline{x}, \underline{y})\Big) \;=\; \sum_a t_a(\underline{x}, \underline{y})\;. \tag{8.29}$$

By (8.17)

$$\sum_a t_a(\underline{x}, \underline{y}) \;\geq\; \sum_{a \in \mathcal{A}} \inf_{\underline{x}} P(a|\underline{x}) \;=:\; \lambda_0 \tag{8.30}$$

which is positive because the chain (X_n) is weak non-null.

Probability of remaining coincident

Let us introduce the following notation

$$D_{m,n} \;:=\; \bigcap_{p=m}^{n} \{\widetilde{X}_j = \widetilde{Y}_j\}\,. \tag{8.31}$$

As a consequence of (8.27)

$$\sup_{a,\underline{x},\underline{y}} \left| P(a \,|\, \underline{x}) - P^{[k]}(a \,|\, x_{-m}^{-1} y_{-\infty}^{-m-1}) \right| \;\leq\; \beta_{m\wedge k} \tag{8.32}$$

Lemma 8.33 *If* $\underline{x} \overset{m}{=} \underline{y}$ *then*

$$\mathbb{P}\Big(\widetilde{X}_0 \neq \widetilde{X}_0^{[k]} \;\Big|\; (\underline{x},\underline{y})\Big) \;\leq\; |\mathcal{A}|\,\beta_{k\wedge m}\,. \tag{8.34}$$

Proof. By definition of the coupling

$$\mathbb{P}\Big(\widetilde{X}_0 \neq \widetilde{X}_0^{[k]} \;\Big|\; (\underline{x},\underline{y})\Big) \;=\; \sum_a r_a(\underline{x},\underline{y}) \tag{8.35}$$

But the right-hand side is

$$\sum_{a\in A} \left| P(a \,|\, \underline{x}) - P^{[k]}(a \,|\, \underline{y}) \right| \;\leq\; |\mathcal{A}|\,\beta_{k\wedge m} \tag{8.36}$$

by (8.32). □

Let us denote

$$\left\{ \begin{array}{lcl} \beta_0^* & = & 1-\lambda_0 \\ \beta_n^* & = & \min\left(\beta_0^*, |\mathcal{A}|\,\beta_n\right), \end{array} \right. \tag{8.37}$$

The previous lemma yields, by straightforward manipulations, the following bounds:

Lemma 8.38 *(i) For all integers* $m,n \geq 0$ *and* $(\underline{x},\underline{y})$ *with* $\underline{x} \overset{m}{=} \underline{y}$,

$$\mathbb{P}(D_{0,n} \mid (\underline{x},\underline{y})) \geq \prod_{p=0}^{n} \Big(1 - \beta_{k\wedge(m+p)}^*\Big)\,. \tag{8.39}$$

(ii) For all integers $k \geq 1$,

$$\mathbb{P}(D_{0,k-1} \mid D_{-k,-1}) \;\geq\; \Big(1-\beta_k^*\Big)^k . \tag{8.40}$$

(iii) For all integers $k \geq 1$,

$$\mathbb{P}(D_{0,k-1} \mid D^c_{-k,-1}) \geq \prod_{p=0}^{+\infty}\Big(1-\beta_p^*\Big) . \tag{8.41}$$

Lemma 8.42

$$\mathbb{P}\Big(\widetilde{X}_0 \neq \widetilde{X}_0^{[k]}\Big) \;\leq\; \frac{\mathbb{P}(D^c_{0,k-1})}{\sum_{j=1}^{k-1}\prod_{m=0}^{k-1}(1-\beta_m^*)} \tag{8.43}$$

Proof.

$$\begin{aligned}\mathbb{P}(D^c_{0,k-1}) \;&=\; \mathbb{P}\Big(\widetilde{X}_{k-1} \neq \widetilde{X}^{[k]}_{k-1}\Big)\\ &+\sum_{\ell=0}^{k-2}\mathbb{P}\Big(D_{\ell+1,k-1}\;\Big|\;\widetilde{X}_\ell \neq \widetilde{X}_\ell^{[k]}\Big)\,\mathbb{P}\Big(\widetilde{X}_\ell \neq \widetilde{X}_\ell^{[k]}\Big) .\end{aligned} \tag{8.44}$$

By translation invariance:

$$\mathbb{P}\Big(\widetilde{X}_0 \neq \widetilde{X}_0^{[k]}\Big) \;=\; \frac{\mathbb{P}(D^c_{0,k-1})}{1\;+\;\sum_{j=1}^{k-1}\mathbb{P}\Big(D_{1,j}\;\Big|\;\widetilde{X}_0 \neq \widetilde{X}_0^{[k]}\Big)} . \tag{8.45}$$

Now the conclusion is straightforward, and there is room for fantasy. Inequality (8.43) follows by bounding

$$\mathbb{P}\Big(D_{1,j}\;\Big|\;\widetilde{X}_0 \neq \widetilde{X}_0^{[k]}\Big) \;\geq\; \prod_{m=1}^{j-1}(1-\beta_m^*) . \square \tag{8.46}$$

To conclude, we observe that

$$\begin{aligned}\mathbb{P}(D^c_{0,k-1}) \;&=\; \mathbb{P}(D^c_{0,k-1}|D_{-k,-1})\,\mathbb{P}(D_{-k,-1})\\ &\qquad + \mathbb{P}(D^c_{0,k-1}|D^c_{-k,-1})\,\mathbb{P}(D^c_{-k,-1})\\ &\leq\; [1-(1-\beta_k^*)^k] + \Big[1-\prod_{p=0}^{+\infty}(1-\beta_p^*)\Big]\mathbb{P}(D^c_{0,k-1}) .\end{aligned} \tag{8.47}$$

Hence

$$\mathbb{P}(D^c_{0,k-1}) \;\leq\; \frac{1-(1-\beta^*_k)^k}{\prod_{p=0}^{+\infty}(1-\beta^*_p)}\,. \tag{8.48}$$

Plugging (8.48) into (8.43) we finally get

$$\mathbb{P}\Big(\widetilde{X}_0 \neq \widetilde{X}_0^{[k]}\Big) \;\leq\; \frac{1-(1-\beta^*_k)^k}{\prod_{p=0}^{+\infty}(1-\beta^*_p)\,\sum_{j=1}^{k-1}\prod_{m=0}^{k-1}(1-\beta^*_m)}\,.\;\square \tag{8.49}$$

Bibliography

[1] K. Athreya and P. Ney. A new approach to the limit theory of recurrent Markov chains. *Trans. Am. Math. Soc.*, 245:493–501, 1978.

[2] H. Berbee. Chains with complete connections: Uniqueness and Markov representation. *Prob. Th. Rel. Fields*, 76:243–53, 1987.

[3] X. Bressaud, R. Fernández, and A. Galves. Speed of $\overline{d}$-convergence for Markov approximations of chains with complete connections. a coupling approach. *Stoch. Proc. and Appl.*, 83:127–38, 1999a.

[4] X. Bressaud, R. Fernández, and A. Galves. Decay of correlations for non Hölderian dynamics. a coupling approach. *Elect. J. Prob.*, 4, 1999b. (`http://www.math.washington.edu/~ejpecp/`).

[5] P. Bühlmann and A. J. Wyner. Variable length markov chains. *Ann. Statist.*, 27(2):480–513, 1999.

[6] F. Comets, R. Fernández, and P. A. Ferrari. Processes with long memory: Regenerative construction and perfect simulation. Preprint, can be retrieved from `http://xxx.lanl.gov/abs/math.PR/0009204`, 2000.

[7] W. Doeblin and R. Fortet. Sur les chaînes à liaisons complétes. *Bull. Soc. Math. France*, 65:132–148, 1937.

[8] P. A. Ferrari and A. Galves. *Acoplamentos e processos estocásticos.* IMPA, Rio de Janeiro, Brazil, 1997. A preliminary updated English version of these notes is available from the page of P. Ferrari, `http://www.ime.usp.br/~pablo/index.html`.

[9] P. A. Ferrari, A. Maass, S. Martíinez, and P. Ney. Cesàro mean distribution of group automata starting from measures with summable decay, 2000. To be published in *Ergodic Th. Dyn. Syst.*

[10] H.-O. Georgii. *Gibbs Measures and Phase Transitions.* Walter de Gruyter (de Gruyter Studies in Mathematics, Vol. 9), Berlin–New York, 1988.

[11] T. E. Harris. On chains of infinite order. *Pacific J. Math.*, 5:707–24, 1955.

[12] M. Iosifescu and S. Grigorescu. *Dependence with Complete Connections and its Applications.* Cambridge University Press, Cambridge, UK, 1990.

[13] S. P. Lalley. Regeneration representation for one-dimensional Gibbs states. *Ann. Prob.*, 14:1262–71, 1986.

[14] S. P. Lalley. Regeneration in one-dimensional Gibbs states and chains with complete connections. *Resenhas IME-USP*, 4:249–80, 2000.

[15] T. M. Liggett. *Interacting particle systems.* Springer-Verlag, Berlin, 1985.

[16] T. Lindvall. *Lectures on the Coupling Method.* Wiley, New York, 1992.

[17] E. Nummelin. A splitting technique for Harris recurrent Markov chains. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 43:309–18, 1978.

[18] O. Onicescu and G. Mihoc. Sur les chaînes statistiques. *C. R. Acad. Sci. Paris*, 200:511—12, 1935a.

[19] D. S. Ornstein. *Ergodic theory, randomness and dynamical systems.* Yale University Press (Yale Mathematical Monographs 5), 1974.

[20] J. G. Propp and D. B. Wilson. 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, 1996.

[21] A. E. Raftery. *A new model for discrete-valued time series: autocorrelations and extensions.* Pitagora, Bologna, 1985.

[22] A. E. Raftery. Some properties of a new continuous bivariate exponential distribution. *Statist. Decisions*, (suppl. 2):53–58, 1985. Selected papers presented at the 16th European meeting of statisticians (Marburg, 1984).

[23] A. E. Raftery and S. Tavaré. Estimation and modelling repeated patterns in high order markov chains with the mtd model. *Appl. Statist.*, 43:179–199, 1994.

[24] D. B. Wilson. 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, 1998. Updated versions can be found at `http://dimacs.rutgers.edu/~dbwilson/exact`.