Speed of $\overline{d}$-convergence for Markov approximations of chains with complete connections. A coupling approach*

Xavier Bressaud[†] Roberto Fernández[‡§] Antonio Galves[¶]

Universidade de São Paulo

November 14, 1997

Abstract

We compute the speed of convergence of the canonical Markov approximation of a chain with complete connections with summable decays. We show that in the $\overline{d}$-topology the approximation converges at least at a rate proportional to these decays. This is proven by explicitly constructing a coupling between the chain and each range-k approximation.

Running Head: $\overline{d}$-convergence of Markov approximations

1 Introduction

The main result of this paper is an estimation of the speed of convergence —in the $\overline{d}$-distance— of the canonical Markov approximation of chains with complete connections. If the continuity rates of the chain are summable, we show that the speed of convergence is at worst proportional to these rates.

Approximations schemes are essential for understanding and handling non-Markovian processes. The speed of convergence is perhaps the most important characterization of an approximation scheme. On the one hand it may carry information about regularity properties of the target process. On the other hand it can be used as a tool to design efficient numerical approaches, and to establish tests to determine whether a given process is of some particular type. These facts could be all the more relevant in relation with some strongly non-Markovian processes and fields of recent interest [see eg. van Enter, Fernández and Sokal (1993)]. Nevertheless, published results on non-Markovian random processes deal only with the issue of *existence* of Markov approximations, and properties inherited from this fact. There appears to be no result so far on speed of convergence.

The existence results apply to stationary processes that either

(a) are the $\overline{d}$-limit of k-step Markov processes, or

(b) have a continuous dependence on past history;

* Work done within the Projeto Temático "Fenômenos Críticos em Processos Evolutivos e Sistemas em Equilíbrio", supported by FAPESP (grant 95/0790-1), and is part of the activities of the Núcleo de Excelência "Fenômenos Críticos em Probabilidade e Processos Estocásticos" (grant 41.96.0923.00)

† Work partially supported by FAPESP (grant 96/04860-7).

‡ Researcher of the National Research Council (CONICET), Argentina.

§ Work partially supported by CNPq (grant 301625/95-6) and FAPESP (grant 95/8896-3).

¶ Work partially supported by CNPq (grant 301301/79).

0 *AMS 1991 Subject Classification. Primary* 60G10; *secondary* 60F15, 60K35.

0 *Key words and Phrases:* Chains with complete connections, Markov approximation, $\overline{d}$-convergence, coupling approac.h

where $\overline{d}$ is the distance introduced by Ornstein (see Definition 3 below).

Stationary process of type (a) inherit the property of being Bernoulli if the approximating Markov chains are aperiodic [Friedman and Ornstein (1970)]. The use of the distance $\overline{d}$ is definitory. Indeed, *every* process can be approximated in the *vague* topology by the so-called *canonical k-step Markov approximations*, defined so to have the same transitions from k to $k+1$ states as the original process (Definition 2 below). This fact, however, is of little use, because weak limits do not convey information on long-range properties. A more revealing issue is whether the canonically defined Markov chains provide also an approximation scheme in the finer $\overline{d}$ topology. The class of processes for which this is true has been completely characterized by Rudolph and Schwarz (1977). In particular, totally ergodic processes have this property if and only if they are Bernoulli [Friedman and Ornstein (1970)].

Stationary processes of type (b) have been studied under the stronger hypothesis of log-continuity. Following Lalley (1986), we shall call them *chains with complete connections* [Lalley's definition differs from the one introduced by Doeblin and Fortet (1937)]. Each process with exponential rates of (log-)continuity is in correspondence with the unique Gibbs measure of a one-dimensional system with an exponentially decaying interaction. If the continuity rates are summable, the process is 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.

We mention two further developments. Lalley (1986) has proposed a regenerative representation of chains with complete connections, in terms of what he calls *list processes*. These are processes which at some random times "forget the past" and "begin from scratch". The distribution of these random times depends on the continuity rates of the initial process: It has a finite exponential moment if the rates are exponential, and only moments up to some finite order if the continuity rates decay as a power-law. On the other hand, 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 paper 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 (23)], 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 (24)]. As a consequence, the coupled processes tend to conicide most of the time, and separations do not last too long [formula (28)]. This yields a small $\overline{d}$-distance between the original process and its k-step approximations.

The coupling concept was introduced by Doeblin in 1938 in a hardly known paper published at the Revue Mathématique de l'Union Interbalkanique. To study the convergence to equilibrium of a Markov chain, Doeblin let two independent trajectories of the process evolve simultaneously, one starting from the stationary measure, and the other from an arbitrary distribution. The convergence follows from the fact that both realizations meet at a finite time. For a description of Doeblin's contributions to probability theory we refer the reader to Lindvall (1991). The idea was only exploited much later, in the sixties, in papers by Athreya, Ney, Harris, Spitzer and Toom among others. Liggett (1994) reviews the use of the coupling technique for interacting Markov systems. The basic idea of our coupling can be traced back to Dobrushin (1956), even when there is no coupling in his paper. Other source of inspiration is Harris' graphical method [Harris (1978)]. For a pedestrian derivation of Dobrushin's ergodic coefficient using coupling we refer the reader to Ferrari and Galves (1997). A coupling approach related to ours has been used by Marton (1996).

A problem related to the discussion of the present paper is the determination of the relaxation rate of the chain. In a recent paper, Kondah, Maume and Schmitt (1996) have estimated this rate for one-dimensional Gibbsian systems, for non-Hölder potentials, using the technique of projective metrics. In a forthcoming paper [Bressaud, Fernández and Galves (1998)] we shall show that similar results can be obtained using our coupling approach.

The paper is organized as follows. The main result and relevant definitions are stated in Section 2 while the proof is developed Section 3. We include two appendices with results of a somehow general character. In Appendix A we present, for completeness, a concise proof of the existence of processes defined by transition probabilities of the type proposed for the coupling. In Appendix B we prove an inequality satisfied by measures associated to continuous conditional probabilities.

2 Definitions and main result

Let $X = (X_n)_{n\in\mathbf{Z}}$ be a stationary stochastic process, defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, taking values in a finite set A (the "alphabet").

Definition 1 *We shall say that the process $(X_n)_{n\in\mathbf{Z}}$ is a* chain with complete connections *if it satisfies the following three properties*

- *for all $a_1, \ldots, a_n \in A$,*
$$\mathbf{P}(X_1 = a_1, \ldots, X_n = a_n) > 0 \tag{1}$$
- *the limit*
$$\lim_{m\to\infty} \mathbf{P}(X_0 = a_0 | X_j = a_j, -m \le j \le -1) = \mathbf{P}(X_0 = a_0 | X_j = a_j, j \le -1) \tag{2}$$
exists for all $a_j, j \le -1$,
- *there is a sequence $(\gamma_m)_{m\ge1}$ with $\lim_{m\to\infty} \gamma_m = 0$, such that, for all $a_j, b_j \in A, j \le -1$ with $a_j = b_j$ for $-1 \ge j \ge -m$,*
$$\left| \frac{\mathbf{P}(X_0 = a_0 \mid X_j = a_j, j \le -1)}{\mathbf{P}(X_0 = a_0 \mid X_j = b_j, j \le -1)} - 1 \right| \le \gamma_m. \tag{3}$$

We shall say that the process has summable decay *if $\sum \gamma_m < +\infty$.*

The next definition follows Ornstein [see for example Ornstein (1974)].

Definition 2 *The* canonical Markov approximation of order $k \in \mathbf{N}$ *of a process $(X_n)_{n\in\mathbf{Z}}$ satisfying (1) is the stationary Markov chain of order k having as transition probabilities,*
$$P^{(k)}(b \mid a_1, \ldots, a_k) := \mathbf{P}(X_{k+1} = b | X_j = a_j, 1 \le j \le k) \tag{4}$$
for all integer $k \ge 1$ and $a_1, \ldots, a_k, b \in A$.

We recall that a *coupling* of two processes $X = (X)_{n\in\mathbf{Z}}$ and $Y = (Y)_{n\in\mathbf{Z}}$ taking values in the alphabet A is any process $(\widetilde{X}, \widetilde{Y}) = (\widetilde{X}_n, \widetilde{Y}_n)_{n\in\mathbf{Z}}$ taking values in $A \times A$ such that $\mathrm{Law}(\widetilde{X}) = \mathrm{Law}(X)$ and $\mathrm{Law}(\widetilde{Y}) = \mathrm{Law}(Y)$.

Definition 3 *The* distance $\overline{d}$ *between two stationary processes X and Y is defined as*
$$\overline{d}(X, Y) = \inf \left\{ \mathbf{P}(\widetilde{X}_0 \ne \widetilde{Y}_0) \ : \ (\widetilde{X}, \widetilde{Y}) \text{ stationary coupling of } X \text{ and } Y \right\}.$$

We now state our main result.

Theorem 4 *Let $X = (X_n)_{n\in\mathbf{Z}}$ be a chain with complete connections and summable decay $(\gamma_m)_{m\ge1}$. Then there is a constant $K > 0$ such that, for all $k \ge 1$,*
$$\bar{d}(X, Y^{(k)}) \le K\, \gamma_k\ ,$$
where $Y^{(k)} = (Y_n^{(k)})_{n\in\mathbf{Z}}$ is the canonical Markov approximation of order k of the process X.

3 Proof of the theorem

The proof of the theorem is decomposed in the following way.

- First we introduce some notation.
- In Section 3.1, we prove two lemmas showing that the transition probabilities of the approximating Markov chain are "close" to the transition probabilities of the original chain.
- In Section 3.2, we construct the coupling. We first define an appropriate system of transition probabilities $\widetilde{P}$. We then prove the existence of a stationary process $(\widetilde{X}_n, \widetilde{Y}_n)_{n\in\mathbf{Z}}$ with these transition probabilities.
- In Section 3.3, we obtain lower bounds for the probability of $\widetilde{X}$ being equal to $\widetilde{Y}$ during a certain number of steps given the history of the coupling. The more they have been equal in the past, the greater is this probability. If they were not equal at the previous step, they keep a positive (bounded away from 0) probability to become equal.
- The final estimation of $\mathbf{P}(\widetilde{X}_0 \neq \widetilde{Y}_0)$ is given in Section 3.4.

A sequence $\underline{x} = (x_j)_{j\le -1}$ of elements of the alphabet A will be called a *history*. We shall denote by $\underline{A}$ the set of all the histories. Given two histories $\underline{x}$ and $\underline{y}$, the notation $\underline{x} \overset{m}{=} \underline{y}$ indicates that $x_j = y_j$ for all $-m \le j \le -1$. For the sake of notational simplicity, we shall denote

$$P(a|\underline{x}) = \mathbf{P}(X_0 = a | X_j = x_j, \le j \le -1). \tag{5}$$

These objects exist for all $\underline{x} \in \underline{A}$ and $a \in A$ by virtue of (2). They admit three different interpretations. Firstly they can be seen as (a continuous version of) the conditional probabilities "knowing all the past" of the event $\{X_0 = a\}$. This motivates our notation. Secondly, they can be interpreted as transition probabilities that to each history associate (continuously) a law for the next step. Finally, one can think of them simply as functions from $A \times \underline{A}$ onto $[0,1]$. Property (2) says that these functions are continuous while property (3) implies that they are indeed log-continuous.

With notation (5), property (3) becomes

$$\sup\left\{\left|\frac{P(a|\underline{x})}{P(a|\underline{y})} - 1\right| \; ; \; \underline{x}, \underline{y} : \underline{x} \overset{m}{=} \underline{y}\right\} \;\le\; \gamma_m, \tag{6}$$

with $a \in A$, $\underline{x}, \underline{y} \in \underline{A}$.

3.1 Properties of the Markov approximation

Let $P^{(k)}$ be the transition probability defined by (4). It is natural to use the same notation for the map from $A \times \underline{A}$ to $[0,1]$ defined as

$$P^{(k)}(a \,|\, \underline{y}) = P^{(k)}(a \,|\, y_{-k}, \ldots, y_{-1}). \tag{7}$$

With this notation, as soon as $\underline{x} \overset{k}{=} \underline{y}$,

$$P^{(k)}(a \,|\, \underline{y}) = P^{(k)}(a \,|\, \underline{x}). \tag{8}$$

Notice that (for details, see Appendix B, Lemma 18),

$$\inf_{\underline{u}\,:\,\underline{u}\overset{k}{=}\underline{y}} P(a \,|\, \underline{u}) \;\le\; \mathbf{P}(X_0 = a \mid X_j = y_j, -k \le j \le -1) \;\le\; \sup_{\underline{u}\,:\,\underline{u}\overset{k}{=}\underline{y}} P(a \,|\, \underline{u}). \tag{9}$$

It follows from (4), (7) and (9), that

$$\inf_{\underline{u}\,:\,\underline{u}\overset{k}{=}\underline{y}} P(a \,|\, \underline{u}) \;\le\; P^{(k)}(a \,|\, \underline{y}) \;\le\; \sup_{\underline{u}\,:\,\underline{u}\overset{k}{=}\underline{y}} P(a \,|\, \underline{u}). \tag{10}$$

Remark 5 In fact, (10) 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 (10).

We now state the crucial consequences of property (6) for the transition probabilities of the canonical Markov approximation.

Lemma 6 *For all integer $m > 0$ and all $a \in A$,*

$$\sup\left\{\left|\frac{P^{(k)}(a\,|\,\underline{x})}{P^{(k)}(a\,|\,\underline{y})} - 1\right| \;:\; \underline{x},\underline{y} \in \underline{A},\ \underline{x} \overset{m}{=} \underline{y}\right\} \le \begin{cases} \gamma_m \ \text{if}\ m < k \\ 0 \ \text{if}\ m \ge k \end{cases} \tag{11}$$

and

$$\sup\left\{\left|\frac{P(a\,|\,\underline{x})}{P^{(k)}(a\,|\,\underline{y})} - 1\right| \;:\; \underline{x},\underline{y} \in \underline{A},\ \underline{x} \overset{m}{=} \underline{y}\right\} \le \begin{cases} \gamma_m \ \text{if}\ m < k \\ \gamma_k \ \text{if}\ m \ge k. \end{cases} \tag{12}$$

Proof Let us fix $a \in A$ and histories $\underline{x}$, $\underline{y}$ such that $\underline{x} \overset{m}{=} \underline{y}$ for some integer $m > 0$. First notice that according to (8) the case $m \ge k$ in (11) is indeed an equality. According to (10), we have,

$$\left.\begin{array}{l} \left|\dfrac{P^{(k)}(a\,|\,\underline{x})}{P^{(k)}(a\,|\,\underline{y})} - 1\right| \\[2ex] \left|\dfrac{P(a\,|\,\underline{x})}{P^{(k)}(a\,|\,\underline{y})} - 1\right| \end{array}\right\} \le \sup\left\{\left|\frac{P(a\,|\,\underline{u})}{P(a\,|\,\underline{v})} - 1\right| \;;\; \underline{u},\underline{v} : \underline{u} \overset{k}{=} \underline{x},\ \underline{v} \overset{k}{=} \underline{y}\right\},$$

Hence, to obtain the three remaining inequalities, it is enough to notice that $\underline{u} \overset{k}{=} \underline{x}$, $\underline{v} \overset{k}{=} \underline{y}$ and $\underline{x} \overset{m}{=} \underline{y}$ imply $\underline{u} \overset{k\wedge m}{=} \underline{v}$, and apply (6). $\square$

The following lemma is a reformulation of these results in the form needed for our purposes.

Lemma 7 *For all integer $m > 0$,*

$$\inf_{a\in A, \underline{x}\in\underline{A}} P^{(k)}(a\,|\,\underline{x}) \;\ge\; \inf_{a\in A, \underline{x}\in\underline{A}} P(a\,|\,\underline{x}) \;>\; 0 \tag{13}$$

and

$$\sup_{\underline{x},\underline{y}:\, \underline{x} \overset{m}{=} \underline{y}} \sum_{a\in A} \left|P(a\,|\,\underline{x}) - P^{(k)}(a\,|\,\underline{y})\right| \;\le\; \gamma_{m\wedge k}. \tag{14}$$

Proof For all $a \in A$, we have, according to (10),

$$P^{(k)}(a\,|\,\underline{x}) \ge \inf_{\underline{u}:\underline{u} \overset{k}{=} \underline{x}} (P(a\,|\,\underline{u})) \ge \inf_{\underline{u}\in\underline{A}} (P(a\,|\,\underline{u})).$$

This proves the leftmost inequality in (13). Property (6) guarantees that the functions $\underline{x} \to \log(P(a\,|\,\underline{x}))$ are continuous on the compact set $\underline{A}$. Hence, they are bounded for all a and the rightmost inequality in (13) follows.

According to (12), for all integer $m > 0$ and all histories $\underline{x},\underline{y}$ such that $\underline{x} \overset{m}{=} \underline{y}$, we have,

$$\left|P(a\,|\,\underline{x}) - P^{(k)}(a\,|\,\underline{y})\right| \;\le\; \gamma_{m\wedge k}\, P^{(k)}(a\,|\,\underline{y}).$$

By summing over all the possible a, we get (14). $\square$

3.2 Construction of the coupling

We first define coupled transition probabilities. These are laws on A^2 depending measurably on double histories, whose projections on each coordinate coincide, respectively, with the transition probabilities of the original and the approximating process. These transition probabilities are shown to be continuous and, hence, there exists a process compatible with them. This process is indeed a coupling of the original process and its canonical Markov approximation.

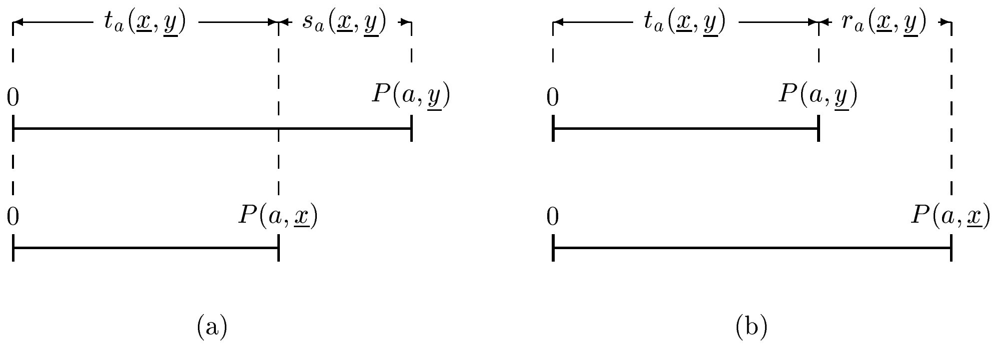

Figure 1: Graphic representation of Definition (15). (a) Case with $r_a(\underline{x},\underline{y}) = 0$. (b) Case with $s_a(\underline{x},\underline{y}) = 0$

3.2.1 Transition probabilities of the coupling

We first introduce some notation. Given the past, that is a double history $(\underline{x},\underline{y})$, we set,

$$\begin{cases} t_a(\underline{x},\underline{y}) & := & P(a\,|\,\underline{x}) \wedge P^{(k)}(a\,|\,\underline{y}) \\ r_a(\underline{x},\underline{y}) & := & (P(a\,|\,\underline{x}) - P^{(k)}(a\,|\,\underline{y})) \vee 0 \\ s_a(\underline{x},\underline{y}) & := & (P^{(k)}(a\,|\,\underline{y}) - P(a\,|\,\underline{x})) \vee 0 \end{cases} \tag{15}$$

Figure 1 gives a graphic representation of the two main possibilities. Notice that

$$\sum_{a\in A} t_a(\underline{x},\underline{y}) + \sum_{a\in A} r_a(\underline{x},\underline{y}) = 1$$

and

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

We now suppose without loss of generality that $A = \{1, 2, \ldots, |A|\}$. For each double history $(\underline{x},\underline{y})$, we can consider two partitions of $[0,1]$, each one made of $2|A|$ intervals,

$$\{T_1^{\underline{x},\underline{y}}, \ldots, T_{|A|}^{\underline{x},\underline{y}}, R_1^{\underline{x},\underline{y}}, \ldots, R_{|A|}^{\underline{x},\underline{y}}\} \ \text{ and } \ \{T_1^{\underline{x},\underline{y}}, \ldots, T_{|A|}^{\underline{x},\underline{y}}, S_1^{\underline{x},\underline{y}}, \ldots, S_{|A|}^{\underline{x},\underline{y}}\} \tag{16}$$

formed by 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 $a \in A$ (see figure 2). We define the transition probabilities $\widetilde{P}((a,b) \mid (\underline{x},\underline{y}))$ as

$$\widetilde{P}((a,b) \mid (\underline{x},\underline{y})) := \begin{cases} |T_a^{\underline{x},\underline{y}}| & \text{if } a = b, \\ |R_a^{\underline{x},\underline{y}} \cap S_b^{\underline{x},\underline{y}}| & \text{if } a \neq b. \end{cases} \tag{17}$$

Notice that

$$|R_a^{\underline{x},\underline{y}} \cap S_b^{\underline{x},\underline{y}}| = \left[\min\left(\sum_{u=1}^{a} r_u(\underline{x},\underline{y})\,;\, \sum_{u=1}^{b} s_u(\underline{x},\underline{y})\right) - \max\left(\sum_{u=1}^{a-1} r_u(\underline{x},\underline{y})\,;\, \sum_{u=1}^{b-1} s_u(\underline{x},\underline{y})\right)\right]^+ . \tag{18}$$

Remark 8 For each double history $(\underline{x},\underline{y})$ one can construct a one-sided process $Z^{\underline{x},\underline{y}} = (Z_n^{\underline{x},\underline{y}})_{n\geq 0}$ on $A\times A$ with the transition probabilities defined by (17). Let $(U_n)_{n\geq 0}$ be a sequence of independent random variables uniformly distributed on $[0,1]$. We define

$$Z_0^{\underline{x},\underline{y}} = \begin{cases} (a,a) & \text{if } U_0 \in T_a^{\underline{x},\underline{y}}, \\ (a,b) & \text{if } U_0 \in R_a^{\underline{x},\underline{y}} \cap S_b^{\underline{x},\underline{y}}. \end{cases}$$

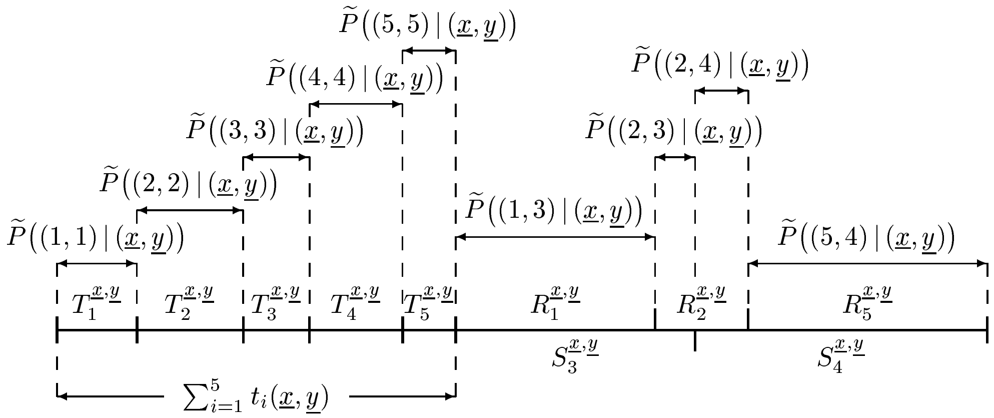

Figure 2: The partition (16) of the interval $[0,1]$ for $|A|=5$. The figure corresponds to histories such that $P(a|\underline{x}) > P^{(k)}(a|\underline{y})$ for $a=1,2,5$ and $P(a|\underline{x}) < P^{(k)}(a|\underline{y})$ for $a=3,4$. The coupling transition rates not depicted in the figure are zero

Observe that, in the last case, $a \neq b$ since $r_a(\underline{x},\underline{y}) > 0$ implies $s_a(\underline{x},\underline{y}) = 0$. We continue in the same manner to define $Z_n^{\underline{x},\underline{y}}$ for $n \geq 1$, using U_n and completing the double history $(\underline{x},\underline{y})$ with the terms $Z_0^{\underline{x},\underline{y}}, Z_1^{\underline{x},\underline{y}}, \ldots, Z_{n-1}^{\underline{x},\underline{y}}$. Notice that the stochastic process so defined is not stationary since the original double history is chosen in an arbitrary way.

3.2.2 Continuity of the transition probabilities

Lemma 9 *The function g defined on $(\underline{A} \times A)^2$ by $g : (\underline{x}, a, \underline{y}, b) \to \widetilde{P}((a,b) \mid (\underline{x},\underline{y}))$ is continuous (in the product topology). More precisely,*

$$\sup\left\{\left|\widetilde{P}((a,b)|(\underline{x},\underline{y})) - \widetilde{P}((a,b)|(\underline{x}',\underline{y}'))\right| \ ; \ a,b \in A, \ \textit{and}\ \underline{x},\underline{x}',\underline{y},\underline{y}' \in \underline{A}\ \underline{x} \stackrel{m}{=} \underline{x}',\ \underline{y} \stackrel{m}{=} \underline{y}'\right\} \leq 4\gamma_m$$

Proof Let m be an integer. Let us fix $a, b, \underline{x}, \underline{y}, \underline{x}', \underline{y}'$ such that $\underline{x} \stackrel{m}{=} \underline{x}'$ and $\underline{y} \stackrel{m}{=} \underline{y}'$.

If $a = b$, we have $|\widetilde{P}((a,a)|(\underline{x},\underline{y})) - \widetilde{P}((a,a)|(\underline{x}',\underline{y}'))| = |t_a(\underline{x},\underline{y}) - t_a(\underline{x}',\underline{y}')|$. Noticing that $|\alpha \wedge \beta - \alpha' \wedge \beta'| \leq |\alpha - \alpha'| \vee |\beta - \beta'|$, we see that

$$|t_a(\underline{x},\underline{y}) - t_a(\underline{x}',\underline{y}')| \leq \max\Big(|P(a|\underline{x}) - P(a|\underline{x}')| \ ; \ |P^{(k)}(a|\underline{y}) - P^{(k)}(a|\underline{y}')|\Big).$$

Applying (6) and (11), we conclude that

$$|\widetilde{P}((a,a)|(\underline{x},\underline{y})) - \widetilde{P}((a,a)|(\underline{x}',\underline{y}'))| \leq \gamma_m.$$

If a and b are different, we use (6), (11) and the inequality $||\alpha-\beta| - |\alpha'-\beta'|| \leq |\alpha-\alpha'| + |\beta-\beta'|$ to obtain

$$\Big||P(a|\underline{x}) - P^{(k)}(b|\underline{y})| - |P(a|\underline{x}') - P^{(k)}(b|\underline{y}')|\Big| \leq \gamma_m\,[P(a|\underline{x}) + P^{(k)}(b|\underline{y})].$$

Hence, for all $w \in A$,

$$\left.\begin{array}{l} \left|\sum_{u=1}^{w} r_u(\underline{x},\underline{y}) - \sum_{u=1}^{w} r_u(\underline{x}',\underline{y}')\right| \\ \left|\sum_{u=1}^{w} s_u(\underline{x},\underline{y}) - \sum_{u=1}^{w} s_u(\underline{x}',\underline{y}')\right| \end{array}\right\} \ \leq \ \sum_{u=1}^{|A|}[P(u|\underline{x}) + P^{(k)}(u|\underline{y})]\,\gamma_m \ \leq \ 2\,\gamma_m.$$

Therefore, the variation of $\left|\min\left(\sum_{u=1}^{a} r_u; \sum_{u=1}^{b} s_u\right) - \max\left(\sum_{u=1}^{a-1} r_u; \sum_{u=1}^{b-1} s_u\right)\right|$ between $(\underline{x}, \underline{y})$ and $(\underline{x}', \underline{y}')$ is less than $4\gamma_m$. Taking the positive part does not alter this fact. Thus,

$$\left|\widetilde{P}((a,b)|(\underline{x},\underline{y})) - \widetilde{P}((a,b)|(\underline{x}',\underline{y}'))\right| \leq 4\,\gamma_m. \quad \square$$

Remark 10 In general these transition probabilities do not define chains with complete connections because some of the transitions can be zero. In fact, in some situations, one can even find arbitrarily close pairs of histories $(\underline{x},\underline{y})$, $(\underline{x}',\underline{y}')$ such that $\widetilde{P}((a,b)|(\underline{x},\underline{y})) > 0$ but $\widetilde{P}((a,b)|(\underline{x}',\underline{y}')) = 0$.

3.2.3 Existence of a stationary coupling

We now can state,

Proposition 11 *There is a stationary process $(\widetilde{X}_n, \widetilde{Y}_n)_{n\in\mathbf{Z}}$ taking values in $A \times A$ whose conditional probabilities satisfy,*

$$\mathbf{P}\left((\widetilde{X}_0, \widetilde{Y}_0) = (a,b) \mid (\widetilde{X}_j, \widetilde{Y}_j) = (x_j, y_j), j \leq -1\right) = \widetilde{P}\left((a,b) \mid (\underline{x},\underline{y})\right). \tag{19}$$

Moreover, under this probability, Law$(\widetilde{X})$ = Law(X) and Law$(\widetilde{Y})$ = Law(Y)

Proof We consider the functions $\widetilde{P}$ as a system of transition probabilities and we ask whether there exists a stationary process compatible with them. This is a rather classical problem. As the transition probabilities are continuous, a result by Ledrappier (1974) or Keane (1971) (concerning the so-called g-measures) proves the existence of a process satisfying (19) (see Appendix A for a complete proof).

Let $(\widetilde{X}, \widetilde{Y}) = (\widetilde{X}_n, \widetilde{Y}_n)_{n\in\mathbf{Z}}$ be such a process. Indeed, it appears from the construction that its marginal transition probabilities are what we need.

$$\begin{aligned}
\mathbf{P}\left(\widetilde{X}_0 = a \mid (\widetilde{X}_j, \widetilde{Y}_j) = (x_j, y_j), j \leq -1\right) &= \sum_{b\in A} \mathbf{P}\left(((\widetilde{X}_0, \widetilde{Y}_0) = (a,b) \mid (\widetilde{X}_j, \widetilde{Y}_j) = (x_j, y_j), j \leq -1\right) \\
&= |T_a| + \sum_{b\in A} |R_a \cap S_b| \\
&= t_a(\underline{x},\underline{y}) + r_a(\underline{x},\underline{y}) \\
&= P(a\,|\,\underline{x}) \wedge P^{(k)}(a\,|\,\underline{y}) + \left(P(a\,|\,\underline{x}) - P^{(k)}(a\,|\,\underline{y})\right) \vee 0 \\
&= P(a|\underline{x}).
\end{aligned}$$

Hence,

$$\mathbf{P}(\widetilde{X}_0 = a \mid \widetilde{X}_j = x_j, j \leq -1) = P(a|\underline{x}).$$

The transition probabilities for $\widetilde{X}$ satisfy property (6). Hence, according to Ledrappier (1974) (for more details, see Appendix A, Remark 16), we have the unicity of the law of the processes compatible with these probabilities. We deduce that Law$(\widetilde{X})$ = Law(X). The proof that Law$(\widetilde{Y})$ = Law(Y) is even simpler: an analogous computation shows that $\mathbf{P}(\widetilde{Y}_0 = b \mid \widetilde{Y}_j = y_j, j \leq -1) = P^{(k)}(b|\underline{y})$. Hence $\widetilde{Y}$ is the only Markov chain compatible with the transition probabilities $P^{(k)}(b|y_{-k}, \ldots, y_{-1})$. $\square$

Let H be an event measurable with respect to the σ-algebra generated by $(\widetilde{X}_n, \widetilde{Y}_n)_{n\geq 0}$ and $(\underline{x},\underline{y})$ a double history. From now on, we shall use for short the following notation:

$$\mathbf{P}(H \mid (\underline{x},\underline{y})) = \mathbf{P}(H \mid (\widetilde{X}_j, \widetilde{Y}_j) = (x_j, y_j), j \leq -1).$$

3.3 Main estimates

Let $\underline{x}, \underline{y}$ be two histories with $\underline{x} \stackrel{m}{=} \underline{y}$. We want to obtain an estimation of the probability of $\widetilde{X}_0$ being different from $\widetilde{Y}_0$ given these histories. First notice that, according to the definition of the coupling,

$$\mathbf{P}(\widetilde{X}_0 \neq \widetilde{Y}_0 \,|(\underline{x}, \underline{y})) = 1 - \sum_a t_a(\underline{x}, \underline{y}) = \sum_a r_a(\underline{x}, \underline{y}).$$

Let us define the sequence $(\widetilde{\gamma}_n)_{n \in \mathbf{N}}$ by,

$$\begin{cases} \widetilde{\gamma}_0 &= 1 - \inf_{a \in A, \underline{x} \in \underline{A}} P(a|\underline{x}), \\ \widetilde{\gamma}_n &= \min\left(\widetilde{\gamma}_0, \gamma_n\right), \end{cases}$$

and let m_0 denote the first integer for which $\gamma_n \leq \widetilde{\gamma}_0$. If $m \leq m_0$, we use (13), to see that,

$$\sum_a t_a(\underline{x}, \underline{y}) \geq \inf_{a \in A} t_a(\underline{x}, \underline{y}) \geq \inf_{a \in A, \underline{x} \in \underline{A}} \min\left(P(a|\underline{x}) \ ; \ P^{(k)}(a|\underline{x})\right) \geq \inf_{a \in A, \underline{x} \in \underline{A}} P(a|\underline{x}) > 1 - \widetilde{\gamma}_0.$$

If $m > m_0$ (provided $k > m_0$), we have, by (14),

$$\sum_a r_a(\underline{x}, \underline{y}) \leq \sum_{a \in A} \left| P(a\,|\,\underline{x}) - P^{(k)}(a\,|\,\underline{y}) \right| \leq \gamma_{k \wedge m} \leq \widetilde{\gamma}_{k \wedge m}.$$

We have that, for all $m \in \mathbf{N}$, and for all histories $\underline{x}, \underline{y}$ with $\underline{x} \stackrel{m}{=} \underline{y}$,

$$\mathbf{P}(\widetilde{X}_0 \neq \widetilde{Y}_0 \,|(\underline{x}, \underline{y})) \leq \widetilde{\gamma}_{k \wedge m}. \tag{20}$$

Let us denote by $\Delta_{m,n}$ the sets $\Delta_{m,n} := \bigcap_{p=m}^{n} \{\widetilde{X}_j = \widetilde{Y}_j\} = \{\widetilde{X}_j = \widetilde{Y}_j, m \leq j \leq n\}$ and by $\Delta^c_{m,n}$ their complementary sets. Notice that $\Delta_{-m,-1}$ is the reunion over all the sequences $\underline{x}, \underline{y}$ with $\underline{x} \stackrel{m}{=} \underline{y}$ of the events $\{(\widetilde{X}_j, \widetilde{Y}_j) = (x_j, y_j); j \leq -1\}$.

Lemma 12 *For all integers m, n and all double histories $(\underline{x}, \underline{y})$ with $\underline{x} \stackrel{m}{=} \underline{y}$,*

$$\mathbf{P}(\Delta_{0,n} \mid (\underline{x}, \underline{y})) \geq \prod_{p=0}^{n} (1 - \widetilde{\gamma}_{k \wedge (m+p)}). \tag{21}$$

Proof Let $\underline{x}, \underline{y}$ be two histories with $\underline{x} \stackrel{m}{=} \underline{y}$. We write,

$$\begin{aligned} \mathbf{P}(\Delta_{0,n} \mid (\underline{x}, \underline{y})) &= \mathbf{P}(\widetilde{X}_0 = \widetilde{Y}_0 \mid (\underline{x}, \underline{y})) \prod_{p=1}^{n} \mathbf{P}(\widetilde{X}_p = \widetilde{Y}_p \mid \Delta_{0,p-1}, (\underline{x}, \underline{y})) \\ &= (1 - \mathbf{P}(\widetilde{X}_0 \neq \widetilde{Y}_0 \mid (\underline{x}, \underline{y})) \prod_{p=1}^{n} \left(1 - \mathbf{P}(\widetilde{X}_p \neq \widetilde{Y}_p \mid \Delta_{0,p-1}, (\underline{x}, \underline{y}))\right) \\ &= \prod_{p=0}^{n} \left(1 - \mathbf{P}(\widetilde{X}_0 \neq \widetilde{Y}_0 \mid H^{(\underline{x}, \underline{y})}_{m+p})\right), \end{aligned} \tag{22}$$

where $H^{(\underline{x}, \underline{y})}_{m+p}$ is the event corresponding to the set of double histories $(\underline{u}, \underline{v})$ with $\underline{u} \stackrel{p}{=} \underline{v}$ and $u_{-p+j} = x_j$, $v_{-p+j} = y_j$ for all $j \leq -1$. Notice that $\underline{u} \stackrel{m+p}{=} \underline{v}$ for all histories $(\underline{u}, \underline{v})$ corresponding to an element of $H^{(\underline{x}, \underline{y})}_{m+p}$. That is, $H^{(\underline{x}, \underline{y})}_{m+p} \subset \Delta_{-m-p,-1}$. Taking into account Lemma 18 (Appendix B), we can use (20) to obtain,

$$\mathbf{P}(\widetilde{X}_0 \neq \widetilde{Y}_0 \,|H^{(\underline{x}, \underline{y})}_{m+p}) \leq \sup_{(\underline{u}, \underline{v}) \in H^{(\underline{x}, \underline{y})}_{m+p}} \mathbf{P}(\widetilde{X}_0 \neq \widetilde{Y}_0 \,|(\underline{u}, \underline{v})) \leq \sup_{\underline{u} \stackrel{m+p}{=} \underline{v}} \mathbf{P}(\widetilde{X}_0 \neq \widetilde{Y}_0 \,|(\underline{u}, \underline{v})) \leq \widetilde{\gamma}_{k \wedge (m+p)}.$$

The lemma follows from this and (22). □

¿From this result, we easily deduce,

Lemma 13
$$\mathbf{P}(\Delta_{0,k-1} \mid \Delta_{-k,-1}) \geq (1-\widetilde{\gamma}_k)^k \tag{23}$$

and
$$\mathbf{P}(\Delta_{0,k-1} \mid \Delta^c_{-k,-1}) \geq \prod_{p=0}^{+\infty}(1-\widetilde{\gamma}_p). \tag{24}$$

Proof According to Lemma 18 (Appendix B), we have, for $H = \Delta_{-k,-1}$ and for $H = \Delta^c_{-k,-1}$,
$$\mathbf{P}(\Delta_{0,k-1} \mid H) \geq \inf_{(\underline{x},\underline{y})\in H} \mathbf{P}(\Delta_{0,k-1} \mid (\underline{x},\underline{y})).$$

Hence, using Lemma 12 for $n = k-1$, $m = k$, we obtain,
$$\mathbf{P}(\Delta_{0,k-1} \mid \Delta_{-k,-1}) \geq \prod_{p=0}^{k-1}(1-\widetilde{\gamma}_{k\wedge(k+p)}) = \prod_{p=0}^{k-1}(1-\widetilde{\gamma}_k) = (1-\widetilde{\gamma}_k)^k,$$

and, using Lemma 12 for $n = k-1$, $m = 0$,
$$\mathbf{P}(\Delta_{0,k-1} \mid \Delta^c_{-k,-1}) \geq \prod_{p=0}^{k-1}(1-\widetilde{\gamma}_{k\wedge p}) \geq \prod_{p=0}^{k-1}(1-\widetilde{\gamma}_p) \geq \prod_{p=0}^{+\infty}(1-\widetilde{\gamma}_p). \quad \square$$

Lemma 14
$$\mathbf{P}(\widetilde{X}_0 \neq \widetilde{Y}_0) \leq \frac{1}{\prod_{m=0}^{+\infty}(1-\widetilde{\gamma}_m)} \frac{\mathbf{P}(\Delta^c_{0,k-1})}{k}. \tag{25}$$

Proof For all finite family $(A_i)_{i=1..k}$ of measurable sets, we have the decomposition
$$\cup_{i=1}^k A_i = \cup_{i=1}^k \left(A_i \setminus \left(A_i \cap (\cup_{j=i+1}^k A_j)\right)\right).$$

Notice that the last element of this partition is exactly A_k. Hence,
$$\begin{aligned}
\mathbf{P}(\cup_{i=1}^k A_i) &= \mathbf{P}\Big(\cup_{i=1}^k \left(A_i \setminus \left(A_i \cap (\cup_{j=i+1}^k A_j)\right)\right)\Big) = \sum_{i=1}^k \mathbf{P}\Big(A_i \setminus \left(A_i \cap (\cup_{j=i+1}^k A_j)\right)\Big) \\
&= \sum_{i=1}^k \mathbf{P}(A_i) - \sum_{i=1}^{k-1} \mathbf{P}\left(A_i \cap (\cup_{j=i+1}^k A_j)\right).
\end{aligned}$$

We use this decomposition to compute the probability of $\Delta^c_{i,k-1} = \cup_{j=i}^{k-1}\{\widetilde{X}_j \neq \widetilde{Y}_j\}$,
$$\begin{aligned}
\mathbf{P}(\Delta^c_{0,k-1}) &= \sum_{i=0}^{k-1} \mathbf{P}(\widetilde{X}_i \neq \widetilde{Y}_i) - \sum_{i=0}^{k-2} \mathbf{P}\left(\{\widetilde{X}_i \neq \widetilde{Y}_i\} \cap \Delta^c_{i+1,k-1}\right) \\
&= \sum_{i=0}^{k-1} \mathbf{P}(\widetilde{X}_0 \neq \widetilde{Y}_0) - \sum_{i=0}^{k-2} \mathbf{P}\left(\Delta^c_{i+1,k-1} \mid \widetilde{X}_i \neq \widetilde{Y}_i\right) \mathbf{P}(\widetilde{X}_i \neq \widetilde{Y}_i) \\
&= k\mathbf{P}(\widetilde{X}_0 \neq \widetilde{Y}_0) - \sum_{i=0}^{k-2} \mathbf{P}\left(\Delta^c_{0,k-i-2} \mid \widetilde{X}_{-1} \neq \widetilde{Y}_{-1}\right) \mathbf{P}(\widetilde{X}_0 \neq \widetilde{Y}_0). \qquad (26)
\end{aligned}$$

Let us now notice that, according to Lemma 12,
$$\mathbf{P}\left(\Delta_{0,k-i-2} \mid \widetilde{X}_{-1} \neq \widetilde{Y}_{-1}\right) \geq \prod_{m=0}^{k-i-1}(1-\widetilde{\gamma}_{k\wedge m}) \geq \prod_{m=0}^{+\infty}(1-\widetilde{\gamma}_m). \tag{27}$$

Inequalities (26) and (27) yield the lemma. $\square$

3.4 Conclusion of the proof

We now have all the elements to prove the theorem. From

$$\begin{aligned}\mathbf{P}(\Delta_{0,k-1}^c) &= \mathbf{P}(\Delta_{0,k-1}^c \mid \Delta_{-k,-1})\mathbf{P}(\Delta_{-k,-1}) + \mathbf{P}(\Delta_{0,k-1}^c \mid \Delta_{-k,-1}^c)\mathbf{P}(\Delta_{-k,-1}^c) \\ &\leq (1-(1-\widetilde{\gamma}_k)^k) + \left(1-\prod_{p=0}^{+\infty}(1-\widetilde{\gamma}_p)\right)\mathbf{P}(\Delta_{0,k-1}^c),\end{aligned}$$

we deduce that

$$\mathbf{P}(\Delta_{0,k-1}^c) \leq \frac{1-(1-\widetilde{\gamma}_k)^k}{\prod_{p=0}^{+\infty}(1-\widetilde{\gamma}_p)}. \tag{28}$$

Using Lemma 14, we get,

$$\mathbf{P}(\widetilde{X}_0 \neq \widetilde{Y}_0) \leq \frac{1}{k\prod_{p=0}^{+\infty}(1-\widetilde{\gamma}_p)}\mathbf{P}(\Delta_{0,k-1}^c) \leq \frac{1}{\left(\prod_{p=0}^{+\infty}(1-\widetilde{\gamma}_p)\right)^2}\frac{1-(1-\widetilde{\gamma}_k)^k}{k}.$$

To conclude the proof we notice that, on the one hand,

$$1-(1-\widetilde{\gamma}_k)^k \sim 1-e^{k\log(1-\gamma_k)} \sim 1-e^{-k\gamma_k} \sim k\gamma_k,$$

because, as $(\gamma_m)_{m\geq 0}$ is deacreasing and summable, $k\gamma_k \to 0$, and, on the other hand,

$$\prod_{p=0}^{+\infty}(1-\widetilde{\gamma}_p) > 0,$$

because, $\log\prod_{p=0}^{n}(1-\widetilde{\gamma}_p) = \sum_{p=0}^{n}\log(1-\widetilde{\gamma}_p) \sim -\sum_{p=0}^{n}\gamma_p$ and $\sum_{p=0}^{+\infty}\gamma_p < +\infty$. □

Acknowledgements

It is a pleasure to thank P. Collet, P. Ferrari, F. Ledrappier and B. Schmitt for useful discussions.

A Invariant measure for the coupling

Lemma 15 *Let P be a a family of transition probabilities, i.e. a measurable application $\underline{A}\times A \to [0,1], (\underline{x},a) \to P(a|\underline{x})$ satisfying $\sum_{a\in A}P(a|\underline{x}) = 1$, for all $\underline{x}\in\underline{A}$. Assume that it is continuous. Then, there is a stationary process $(X_n)_{n\in\mathbf{Z}}$ taking values in A whose conditional probabilities satisfy,*

$$\forall(\underline{x},a)\in\underline{A}\times A, \quad \mathbf{P}(X_0 = a \mid X_j = x_j, j\leq -1) = P(a\mid\underline{x}) \tag{29}$$

Proof Let us fix a history $\underline{x}$ and consider the process $Z^{\underline{x}} = (Z_t^{\underline{x}})_{t\in\mathbf{Z}}$ equal to $\underline{x}$ for $t\leq -1$ and whose conditional probabilities satisfy for all $t\geq 0$ and all $a\in A$,

$$\mathbf{P}(Z_t^{\underline{x}} = a \mid Z_j^{\underline{x}}, 0\leq j\leq t-1) = P(a \mid \underline{(Z_{t+j}^{\underline{x}})_{j\leq -1}}),$$

where $\underline{(Z_{t+j}^{\underline{x}})_{j\leq -1}}$ is the sequence $(Z_{t-1}^{\underline{x}}, Z_{t-2}^{\underline{x}},\ldots,Z_0^{\underline{x}}, x_{-1},\ldots)$ seen as an element of $\underline{A}$. Notice that it is defined in the same way as the process in Remark 8. Let Π_0 denote the law of $Z^{\underline{x}}$ and, more generally, $(\Pi_n)_{n\geq 0}$ the sequence of the laws of the shifted processes $(Z_{t+n}^{\underline{x}})_{t\in\mathbf{Z}}$. The set of probability measures on $A^{\mathbf{Z}}$ is compact. Hence, the sequence of probability measures $(\frac{1}{n}\sum_{k=0}^{n-1}\Pi_k)_{n\geq 1}$ has an accumulation point, say Π_∞. Let $(X_t)_{t\in\mathbf{Z}}$ be a process with this law. We claim that this process is stationary and has the conditional probabilities (29).

Indeed, its stationarity is almost immediate. For each continuous function f defined on $A^{\mathbf{Z}}$ we have that $\Pi_n[f(z)] = \mathbf{P}[f((Z^{\underline{x}}_{t+n})_{t\in\mathbf{Z}})]$ and $(\frac{1}{n}\sum_{k=0}^{n-1}\Pi_k)[f(z)] = \frac{1}{n}\sum_{k=0}^{n-1}\mathbf{P}[f((Z^{\underline{x}}_{t+k})_{t\in\mathbf{Z}})]$. Hence,

$$\frac{1}{n}\sum_{k=0}^{n-1}\Pi_{k+1}[f(z)] = \frac{1}{n}\sum_{k=0}^{n-1}\Pi_k[f(z)] + \frac{\mathbf{P}[f((Z^{\underline{x}}_{t+n})_{t\in\mathbf{Z}})] - \mathbf{P}[f((Z^{\underline{x}}_{t})_{t\in\mathbf{Z}})]}{n}.$$

Letting n go to infinity along the converging subsequence, we see that $\mathbf{P}[f((X_{t+1})_{t\in\mathbf{Z}})] = \mathbf{P}[f((X_t)_{t\in\mathbf{Z}})]$. Hence $(X_n)_{n\in\mathbf{Z}}$ is stationary.

Let us now check (29). Fix an integer m, a function f defined on A^m and some $a\in A$. We know that, for all positive integer k,

$$\mathbf{P}[1_{\{Z^{\underline{x}}_{m+k}=a\}}f(Z^{\underline{x}}_k,\dots,Z^{\underline{x}}_{m+k-1})] = \mathbf{P}[P(a|\underline{(Z^{\underline{x}}_{t+m+k})_{t\le -1}})f(Z^{\underline{x}}_k,\dots,Z^{\underline{x}}_{m+k-1})]$$

Hence,

$$\Pi_k[1_{\{a\}}(z_m)f(z_0,\dots,z_{m-1})] = \Pi_k[P(a|\underline{(z_{m+t})_{t\le -1}})f(z_0,\dots,z_{m-1})]$$

Taking the mean and letting k go to infinity along the converging subsequence, the continuity of $\underline{z}\to P(a|\underline{z})$ implies,

$$\mathbf{P}[1_{\{X_m=a\}}f(X_0,\dots,X_{m-1})] = \mathbf{P}[P(a|\underline{(X_{m+t})_{t\le -1}})f(X_0,\dots,X_{m-1})]$$

Hence, by stationarity,

$$\mathbf{P}[1_{\{X_0=a\}}f(X_{-m},\dots,X_{-1})] = \mathbf{P}[P(a|\underline{(X_t)_{t\le -1}})f(X_{-m},\dots,X_{-1})]$$

We deduce that $P(a|\underline{(X_t)_{t\le -1}})$ is a version of the conditional expectation of $1_{\{X_0=a\}}$ given the past. □

Remark 16 The existence of the process has been proven by Keane (1971) and Ledrappier (1974). In fact, in this last reference it is shown that under an additional assumption such a measure is unique. The additional assumption corresponds exactly to the condition (6) with a summable γ_m. Let us formulate Ledrappier's result with our notation:

Lemma 17 *Let P be a a family of transition probabilities, i.e. an application $\underline{A}\times A\to[0,1]$, $(\underline{x},a)\to P(a|\underline{x})$ satisfying $\sum_{a\in A}P(a|\underline{x})=1$, for all $\underline{x}\in\underline{A}$. Assume that there is a sequence $(\gamma_m)_{m\in\mathbf{N}^*}$ with $\sum_{m=1}^{+\infty}\gamma_m<+\infty$, such that for all $a\in A$,*

$$\sup\left\{\left|\frac{P(a|\underline{x})}{P(a|\underline{y})}-1\right| \; ; \; \underline{x},\underline{y} : \underline{x}\overset{m}{=}\underline{y}\right\}\le\gamma_m$$

Then, there is a unique *(in law) stationary process $X=(X)_{n\in\mathbf{Z}}$ whose conditional probabilities satisfy,*

$$\forall(\underline{x},a)\in\underline{A}\times A,\quad \mathbf{P}(X_0=a\mid X_j=x_j, j\le -1) = P(a\mid\underline{x})$$

B Standard bounds

In this appendix, we prove a general result concerning stationary measures associated with a continuous system of transition probabilities.

Lemma 18 *Let $\mathbf{P}$ be the law of a process with contiuous transition probabilities, and, for some integer p, let H be an element of $\sigma(X_j, -p\le j\le -1)$ of positive measure (i.e. H has positive measure and is measurable with respect to a finite number of coordinates). Then, for all $a\in A$ we have,*

$$\inf_{\underline{u}\in H}(P(a\,|\,\underline{u}))\le\mathbf{P}(X_0=a\mid H)\le\sup_{\underline{u}\in H}(P(a\,|\,\underline{u}))$$

Proof Let $\mathbf{P}_H$ be the conditioned probability measure on the set $\underline{A}$, given the event H. It is caracterized by its values on the finite cylinders:

$$\mathbf{P}_H(u_{-n},\ldots,u_{-1}) := \frac{\mathbf{P}(\{X_j = u_j, -n \leq j \leq -1\} \cap H)}{\mathbf{P}(H)}$$

Given an element a of A, let us denote,

$$h_n(\underline{u}) = \mathbf{P}(X_0 = a \mid \{X_j = u_j, -n \leq j \leq -1\} \cap H)$$

The equality

$$\mathbf{P}(X_0 = a \mid H) = \sum_{u_{-n},\ldots,u_{-1} \in A} \mathbf{P}\left(X_0 = a \mid \{X_j = u_j, -n \leq j \leq -1\} \cap H\right) \frac{\mathbf{P}\left(\{X_j = u_j, -n \leq j \leq -1\} \cap H\right)}{\mathbf{P}(H)}$$

can then be written as,

$$\mathbf{P}(X_0 = a \mid H) = \int h_n(\underline{u})\, d\mathbf{P}_H(\underline{u})$$

We now notice that when $n > p$, $(u_{-n},\ldots,u_{-1}) \cap H$ is either reduced to $(u_{-n},\ldots,u_{-1})$, or empty, according to whether $\underline{u}$ belongs to H or not. Hence, by continuity [property (2)], $h_n(\underline{u})$ converges towards $1_H(\underline{u})P(a|\underline{u})$. By dominated convergence, the convergence occurs also in $L^1(\mathbf{P}_H)$. We deduce that,

$$\mathbf{P}(X_0 = a \mid H) = \int 1_H(\underline{u})\, P(a|\underline{u})\, d\mathbf{P}_H(\underline{u})$$

As $\mathbf{P}_H(H) = 1$, this identity implies,

$$\inf_{\underline{u} \in H} (P(a|\underline{u})) \leq \int_H P(a|\underline{u})\, d\mathbf{P}_H((\underline{u}) \leq \sup_{\underline{u} \in H} (P(a|\underline{u})). \quad \square$$

References

[BFG98] X. Bressaud, R. Fernández and A. Galves (1998). Relaxation speed for chains with complete connections. In preparation.

[DOB56] R. Dobrushin (1956). Central limit theorem for nonstationary Markov chains. *Theor. Prob. and its Appl.* 1:65–80 (Part I) and 329–383 (Part II).

[DF37] W. Doeblin and R. Fortet (1937). Sur les chaînes à liaisons complètes. *Bull. Soc. Math. France*, 65:132–148.

[EFS93] A. C. D. van Enter, R. Fernández and A. D. Sokal (1993). Regularity Properties and pathologies of Position-Space Renormalization-Group Transformations: Scope and Limitations of Gibbsian Theory *J. Stat. Phys.*, 72:879–1167.

[FG97] P. A. Ferrari and A. Galves (1997). *Acoplamento e processos estocásticos.* IMPA. (Postscript file available at URL http://www.ime.usp.br/ pablo/publications.html)

[FO70] N.A. Friedman and D.S. Ornstein (1970). On isomorphism of weak Bernoulli transformations. *Advances in Math.*, 5:365–394.

[HAR] T. E. Harris (1978). Additive set-valued Markov processes and graphical methods. *Ann. Probability*, 6:355–378.

[KEA71] Michael Keane (1971). Sur les mesures invariantes d'un recouvrement regulier. *Comptes Rendus de l'Académie des Sciences*, Volume 272, Part A:585–587.

[KMS96] A. Kondah, V. Maume and B. Schmitt (1996). Vitesse de convergence vers l'état d'équilibre pour des dynamiques markoviennes non höldériennes. Preprint, 89, Laboratoire de Topologie, Université de Borgogne.

[LAL86] S. P. Lalley (1986). Regenerative representation for one-dimensional Gibbs states. *Ann. Probab.*, 14(4):1262–1271.

[LED74] François Ledrappier (1974). Principe variationnel et systèmes dynamiques symboliques. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 30:185–202.

[LIG93] T. M. Liggett (1993). The coupling technique in interacting particle systems. Doeblin and modern probability (Blaubeuren, 1991), 73–83, *Contemp. Math.*, 149, Amer. Math. Soc., Providence, RI.

[LIN91] T. Lindvall (1991). W. Doeblin 1915–1940. *Ann. Probab.*, 19:929–934.

[MAR96] K. Marton (1996). Bounding $\overline{d}$-distance by informational divergence: a method to prove measure concentration. *Ann. Probab.*, 24(2):857–866.

[ORN74] D. S. Ornstein (1974). Ergodic theory, randomness, and dynamical systems. In *Yale Mathematical Monographs*, volume 5. Yale University Press.

[OW90] D. S. Ornstein and B. Weiss (1990). How sampling reveals a process. *Ann. Probability*, 18(3):905–930.

[RS77] D. J. Rudolph and G. Schwarz (1977). The limit in $\overline{d}$ of multi-step Markov chains. *Israel Journal of Math.*, 28(1-2):103–109.

Xavier Bressaud
Instituto de Matemática e Estatística
Universidade de São Paulo
Caixa Postal 66281
05315-970 São Paulo, Brazil
e-mail: bressaud@ime.usp.br

Robert Fernández
Instituto de Estudos Avançados
da Universidade de São Paulo
Av. Prof. Luciano Gualberto, Travessa J, 374 Térreo,
Cidade Universitaria
05508-900 São Paulo, Brazil
e-mail: rf@ime.usp.br

Antonio Galves
Instituto de Matemática e Estatística
Universidade de São Paulo
Caixa Postal 66281
05315-970 São Paulo, Brazil
e-mail: galves@ime.usp.br