BULLETIN BRAZILIAN MATHEMATICAL SOCIETY

Bull Braz Math Soc, New Series 33(3), 295-306

Markov approximations of chains of infinite order

R. Fernández and A. Galves

Abstract. We consider chains whose transition probabilities depend on the whole past, with summable continuity rates. We show that Ornstein's $\bar{d}$-distance between one such chain and its canonical Markov approximations of different orders is at worst proportional to the continuity rate of the chain. The result generalizes previous bounds obtained by X. Bressaud and ourselves, while relying on a similar coupling argument.

Keywords: Chains of infinite order, chains with complete connections, Ornstein distance, coupling.

Mathematical subject classification: Primary 60G10; secondary 60F15, 60K35.

1 Introduction

This article addresses the following question: How well can we approximate a chain of infinite order by a Markov chain of order k? This leads to a second, technical, question: Which distance should we use to measure the quality of an approximation?

It is natural to use as Markov approximation a Markov chain whose transition probabilities can be estimated from a sample of the infinite-order chain. This is the so-called *canonical* Markov approximation. The conditional probabilities of the canonical approximation of order k coincide with the order-k conditional probabilities of the original infinite-order chain.

The main result of the present article is an upper bound of Ornstein's $\bar{d}$-distance between a chain of infinite order and its canonical Markov approximation of any given order. In fact, the bound is proportional to the continuity rate of the chain of infinite order, whenever the sequence formed by these rates is summable.

The present result applies to a more general type of chains than those covered by Bressaud, Fernández and Galves [2] (see Remark 4 below). The result actually

Received 9 April 2002.

applies to any Markov approximation whose transition probabilities satisfy a suitable sandwich inequality [(25) below]. In addition, in this work we do not assume stationarity of the chain. Thus, in particular, our result applies to chains starting at a finite time from a fixed past.

Our proof is constructive. Following the spirit of the *graphical construction* of interacting particle systems introduced by Harris [5, 7], we construct an explicit coupling between the original chain and its k-step approximation. In this coupling, the probability of coincidence of the chains at a given time is an increasing function of the length of the immediately preceding period of coincidence. Furthermore, if the chains differ at some time, there is a nonzero probability that they will become equal at the next instant. Therefore, the coupled processes tend to coincide most of the time, and separations do not last too long.

Chains of infinite order seem to have been first studied by Onicescu and Mihoc [9] who called them *chains with complete connections* (*chaînes à liaisons complètes*). Their study was soon taken up by Doeblin and Fortet [3] who proved the first results on speed of convergence towards the invariant measure. The name chains of infinite order was coined by Harris [6]. We refer the reader to Iosifescu and Grigorescu [8] for a complete survey, and to our notes with Pablo Ferrari for the Vth Brazilian School of Probability [4] for an elementary presentation of the subject from a constructive point of view.

2 Definitions and main result

We consider stochastic processes $(X_n)_{n\in\mathbb{Z}}$ taking values in a finite *alphabet* $\mathcal{A}$ and defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We adopt the following notation. For $k \le n \in \mathbb{Z}$, x_k^n denotes the sequence $x_k, \ldots, x_n$, and $\mathcal{A}_k^n$ the set of such sequences. Likewise, $x_{-\infty}^n$ denotes the sequence $(x_i)_{i\le n}$ and $\mathcal{A}_{-\infty}^n$ the corresponding space. Full sequences will be denoted without sub or superscripts, $x \in \mathcal{A}^{\mathbb{Z}}$. The notation $y_{n+1}^m x_k^n$ indicates the sequence that takes values $x_k, \ldots, x_n, y_{n+1}, \ldots, y_m$.

Definition 2.1. *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* $x_{-\infty}^{n-1} \in \mathcal{A}_{-\infty}^{n-1}$,

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

Definition 2.2. *A stochastic process* $(X_n)_{n\in\mathbb{Z}}$ *defined on* $(\Omega, \mathcal{F}, \mathbb{P})$ *is* **consistent** *with a system of transition probabilities* (P_n) *if*

$$\mathbb{P}(X_n = x_n | X_{-\infty}^{n-1} = x_{-\infty}^{n-1}) \;=\; P_n(x_n | x_{-\infty}^{n-1}) \tag{2}$$

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

Remark 1. Note that we do *not* assume stationarity (i.e. P_n may depend on n).

Definition 2.3. *A system of transition probabilities is* **continuous** *if for each* $n \in \mathbb{Z}$ *and each* $x_n \in \mathcal{A}$

$$\begin{aligned} \beta(s) \quad &:= \quad \sup_{n\in\mathbb{Z}} \sup_{x,y} \left| P_n(x_n | x_{-\infty}^{n-1}) - P_n(x_n | x_{n-s}^{n-1}\, y_{-\infty}^{n-s-1}) \right| \\ \underset{s\to\infty}{\longrightarrow} \quad & \quad 0\;. \end{aligned} \tag{3}$$

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

A compactness argument shows that every system of continuous transition probabilities has at least one stochastic process consistent with it.

Remark 2. A stronger notion of continuity, often used in the literature involves the *log-continuity rates*

$$\gamma(s) \;:=\; \sup_{n\in\mathbb{Z}} \sup_{x,y} \left| \frac{P_n(x_n | x_{-\infty}^{n-1})}{P_n(x_n | x_{n-s}^{n-1}\, y_{-\infty}^{n-s-1})} \;-\; 1 \right| \tag{4}$$

A system of transition probabilities is *log-continuous* if $\gamma(s) \longrightarrow 0$ as $s \to \infty$.

Definition 2.4. *A system of transition probabilities is* **weakly non-null** *if*

$$\inf_{n\in\mathbb{Z}} \sum_{y_n \in \mathcal{A}} \inf_{x} P_n(y_n | x_{-\infty}^{n-1}) \;>\; 0\;. \tag{5}$$

Remark 3. The stronger requirement

$$\inf_{n\in\mathbb{Z}} \inf_{x} P_n(y_n|x_{-\infty}^{n-1}) \;>\; 0 \tag{6}$$

is instead often assumed in the literature.

Definition 2.5. *A stochastic process is a* **chain of infinite order of type A** *if it is consistent with a system of transition probabilities that is continuous and weakly non-null.*

Remark 4. This type of chains was already considered by Doeblin and Fortet [3]. They also considered the *chains of type B,* defined by transition probabilities that are log continuous [see (4)] and satisfy the more restrictive non-nullness condition (6). The approximation result of Bressaud, Fernández and Galves [2] only applies to chains of type B while the result here applies to the more general class of chains of type A.

Definition 2.6. *The* **canonical Markov approximation of order** $k \in \mathbf{N}$ *of a process* $(X_n)_{n\in\mathbb{Z}}$ *is the Markov chain of order* k $X^{[k]} = (X_n^{[k]})_{n\in\mathbb{Z}}$ *having as transition probabilities,*

$$P_n^{[k]}(a \mid x_{n-k}^{n-1}) := \mathbb{P}(X_n = a \mid X_{n-k}^{n-1} = x_{n-k}^{n-1}) \tag{7}$$

for all integer $n, k \geq 1$ *and all* $a \in A$ *and* $x_{n-k}^{n-1} \in \mathcal{A}_{n-k}^{n-1}$.

Definition 2.7. *The* distance $\overline{d}$ *between two processes* $X = (X_n)$ *and* $Y = (Y_n)$ *is defined as*

$$\overline{d}(X, Y) = \inf \left\{ \sup_{n\in\mathbb{Z}} \mathbb{P}(\widetilde{X}_n \neq \widetilde{Y}_n) \;:\; (\widetilde{X}, \widetilde{Y}) \text{ coupling of } X \text{ and } Y \right\} .$$

This definition naturally extends Ornstein's to non-stationary chains.

We now state our main result.

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

$$\bar{d}(X, X^{[k]}) \leq C\ \beta(k)\ ,$$

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

3 Construction of the coupling

The proof of our theorem is based on the construction of a suitable coupling between the transition probabilities $P_n(\,\cdot\,|\,\cdot\,)$ of the original chain and the probabilities $P_n^{[k]}(\,\cdot\,|\,\cdot\,)$ of its Markov approximation.

In general, a coupling of two systems of transition probabilities $P_n(\,\cdot\,|\,\cdot\,)$ and $Q_n(\,\cdot\,|\,\cdot\,)$ is a system of transition probabilities $\widetilde{P}_n : \mathcal{A}^2 \times (\mathcal{A}_{-\infty}^{n-1})^2 \longrightarrow [0,1]$ such that

$$\begin{aligned} \sum_{y_n\in\mathcal{A}} \widetilde{P}_n\Big(x_n, y_n \;\Big|\; x_{-\infty}^{n-1}, y_{-\infty}^{n-1}\Big) &= P_n\Big(x_n \;\Big|\; x_{-\infty}^{n-1}\Big) \\ \sum_{x_n\in\mathcal{A}} \widetilde{P}_n\Big(x_n, y_n \;\Big|\; x_{-\infty}^{n-1}, y_{-\infty}^{n-1}\Big) &= Q_n\Big(y_n \;\Big|\; y_{-\infty}^{n-1}\Big) \end{aligned} \tag{8}$$

for all $n \in \mathbb{Z}$, all $x_n, y_n \in \mathcal{A}$ and all $x_{-\infty}^{n-1}, y_{-\infty}^{n-1} \in \mathcal{A}_{-\infty}^{n-1}$. [This definition is, in fact, a particular instance of the notion of coupling among probability measures.]

We resort to a coupling 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.

In fact, we shall use the well known *maximal coupling* (see, for instance, Appendix A.1 in Barbour, Holst, and Janson [1]). We present here a graphical construction of this coupling which is convenient for our purposes.

Given two trajectories $x = (x_n)$, $y = (y_n)$ and an element a of the alphabet $\mathcal{A}$, let us define

$$\begin{aligned} t_{a,n}(x,y) &:= P_n(a \,|\, x_{-\infty}^{n-1}) \wedge Q_n(a \,|\, y_{-\infty}^{n-1}) \\ r_{a,n}(x,y) &:= [P_n(a \,|\, x_{-\infty}^{n-1}) - Q_n(a \,|\, y_{-\infty}^{n-1})] \vee 0 \\ s_{a,n}(x,y) &:= [Q_n(a \,|\, y_{-\infty}^{n-1}) - P_n(a \,|\, x_{-\infty}^{n-1})] \vee 0 \,. \end{aligned} \tag{9}$$

Notice that

$$\begin{aligned} \text{either} \quad & r_{a,n}(x,y) = 0 \text{ and } s_{a,n}(x,y) > 0 \\ \text{or} \quad & r_{a,n}(x,y) > 0 \text{ and } s_{a,n}(x,y) = 0 \end{aligned} \tag{10}$$

and that

$$t_{a,n}(x,y) + r_{a,n}(x,y) = P_n(a|x_{-\infty}^{n-1}) \tag{11}$$

$$t_{a,n}(x,y) + s_{a,n}(x,y) = Q_n(a|y_{-\infty}^{n-1}) \,. \tag{12}$$

As a consequence,

$$\sum_{a\in A} t_{a,n}(x,y) + \sum_{a\in A} r_{a,n}(x,y) \;=\; 1 \tag{13}$$

$$\sum_{a\in A} t_{a,n}(x,y) + \sum_{a\in A} s_{a,n}(x,y) \;=\; 1\,. \tag{14}$$

Identities (13)/(14) enable us to define two partitions of [0, 1], each one formed by the non-empty sets of the following $2|\mathcal{A}|$ intervals:

$$\{T^{x,y}_{1,n},\ldots,T^{x,y}_{|\mathcal{A}|,n}, R^{x,y}_{1,n},\ldots,R^{x,y}_{|\mathcal{A}|,n}\} \text{ and } \{T^{x,y}_{1,n},\ldots,T^{x,y}_{|\mathcal{A}|,n}, S^{x,y}_{1,n},\ldots,S^{x,y}_{|\mathcal{A}|,n}\} \tag{15}$$

These are intervals of lengths

$$|T^{x,y}_{a,n}| = t_{a,n}(x,y)\,,\quad |R^{x,y}_{a,n}| = r_{a,n}(x,y) \quad\text{and}\quad |S^{x,y}_{a,n}| = s_{a,n}(x,y)\,,$$

for all $a \in A$

We define the transition probabilities

$$\widetilde{P}_n((a,b) \mid (x^{n-1}_{-\infty}, y^{n-1}_{-\infty})) := \begin{cases} |T^{x,y}_{a,n}| & \text{if } a=b,\\ |R^{x,y}_{a,n}\cap S^{x,y}_{b,n}| & \text{if } a\neq b \end{cases} \tag{16}$$

The properties of this coupling are summarized in the following theorem

Theorem 2. *If the chains with transition probabilities P and Q are both of type A, so is the coupling defined by (16). More explicitly, we have*

$$\widetilde{\beta}(s) \;\le\; const\;[\beta^P(s)\vee\beta^Q(s)]\,, \tag{17}$$

and

$$\sum_{a,b\in\mathcal{A}} \inf_{x,y} \widetilde{P}_n\Big((a,b)\;\Big|\;(x^{n-1}_{-\infty}, y^{n-1}_{-\infty})\Big) \;\ge\; \\ \ge\; \Big[\sum_{a\in\mathcal{A}} \inf_x P_n(a|x^{n-1}_{-\infty})\Big] \;\wedge\; \Big[\sum_{a\in\mathcal{A}} \inf_x Q_n(a|x^{n-1}_{-\infty})\Big]\,. \tag{18}$$

We remark that, even if the transitions P_n and Q_n are of type B, this coupling is not in general of type B, because there may be pairs (a,b) with

$$\inf_{x,y} \widetilde{P}_n\Big((a,b)\;\Big|\;(x^{n-1}_{-\infty}, y^{n-1}_{-\infty})\Big) \;=\; 0\,.$$

This happens whenever $R^{x,y}_{a,n}\cap S^{x,y}_{b,n} = \emptyset$.

Proof.

Non-nullness.

$$\sum_{a,b\in\mathcal{A}} \inf_{x,y} \widetilde{P}_n\Big((a,b) \;\Big|\; (x_{-\infty}^{n-1}, y_{-\infty}^{n-1})\Big) \;\geq\; \sum_{a\in\mathcal{A}} \inf_{x,y} \widetilde{P}_n\Big((a,a) \;\Big|\; (x_{-\infty}^{n-1}, y_{-\infty}^{n-1})\Big) \tag{19}$$

But the right-hand side is

$$\begin{aligned} &\sum_{a\in\mathcal{A}} \inf_{x,y}\Big[P_n(a|x_{-\infty}^{n-1}) \;\wedge\; Q(a|y_{-\infty}^{n-1})\Big] \;\geq \\ &\geq\; \Big[\sum_{a\in\mathcal{A}} \inf_{x} P_n(a|x_{-\infty}^{n-1})\Big] \;\wedge\; \Big[\sum_{a\in\mathcal{A}} \inf_{x} Q_n(a|x_{-\infty}^{n-1})\Big]\,. \end{aligned} \tag{20}$$

Continuity. Let us denote

$$\begin{aligned} \Delta_{m,n}(a,b) \;=\; \sup_{x,y,u,w}\Big|\widetilde{P}_n\Big((a,b) \;\Big|\; (x_{-\infty}^{n-1}, y_{-\infty}^{n-1})\Big) - \\ \widetilde{P}_n\Big((a,b) \;\Big|\; (x_{n-m}^{n-1}u_{-\infty}^{n-m-1}, y_{n-m}^{n-1}w_{-\infty}^{n-m-1})\Big)\Big|\,. \end{aligned} \tag{21}$$

Case $a = b$**:**

$$\Delta_{m,n}(a,a) \;=\; \sup_{x,y,u,w}\Big|t_{a,n}(x,y) - t_{a,n}(x_{n-m}^{n-1}u_{-\infty}^{n-m-1}, y_{n-m}^{n-1}w_{-\infty}^{n-m-1})\Big|\,. \tag{22}$$

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

$$\begin{aligned} \Delta_{m,n}(a,a) \;\leq\; \sup_{x,y,u,w}\Big[|P_n(a|x_{-\infty}^{n-1}) - P_n(a|x_{n-m}^{n-1}u_{-\infty}^{n-m-1})| \;\vee \\ |Q_n(a|y_{-\infty}^{n-1}) - Q_n(a|y_{n-m}^{n-1}w_{-\infty}^{n-m-1})|\Big]\,. \end{aligned} \tag{23}$$

Hence,

$$\Delta_{m,n}(a,a) \;\leq\; \beta^P(m) \vee \beta^Q(m) \tag{24}$$

uniformly in n.

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

4 Proof of the theorem

We are ready to prove Theorem 1.

4.1 Bounds for the transition probabilities

Let $P_n^{[k]}$ be the transition probability defined by (7). The following proposition contains the only property of the canonical approximation needed for the result.

Proposition 3. *For each trajectory $y \in \mathcal{A}^{\mathbb{Z}}$,*

$$\inf_{u\in\mathcal{A}^{\mathbb{Z}}} P_n(a \mid y_{n-k}^{n-1} u_{-\infty}^{n-k-1}) \;\le\; P_n^{[k]}(a \mid y_{n-k}^{n-1}) \;\le\; \sup_{u\in\mathcal{A}^{\mathbb{Z}}} P_n(a \mid y_{n-k}^{n-1} u_{-\infty}^{n-k-1}). \tag{25}$$

Proof. First observe that, by definition, the transition probabilities of the canonical Markov approximations satisfy

$$P_n^{[k]}(a|x_{n-k}^{n-1}) \;=\; P_n(a|x_{n-k}^{n-1}) \;. \tag{26}$$

The conclusion of the proof is an exercise on conditional probabilities. Indeed, each conditional probability $P_n(a|x_{n-k}^{n-1})$ can be written as an integral of the function $u_{-\infty}^{n-1} \mapsto P_n(a|u_{-\infty}^{n-1})$ with respect to a probability measure conditioned on $\{X_{n-k}^{n-1} = x_{n-k}^{n-1}\}$. Inequalities (25) follow by taking the corresponding supremum and infimum of the integrand. □

Remark 5. In fact, (25) 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 (25).

4.2 The proof

Positive probability of coincidence

By the definition of the coupling,

$$\mathbb{P}\Big(\widetilde{X}_n = \widetilde{X}_n^{[k]} \;\Big|\; (x_{-\infty}^{n-1}, y_{-\infty}^{n-1})\Big) \;=\; \sum_a t_{a,n}(x, y) \;. \tag{27}$$

We define λ_0 as

$$\lambda_0 \;:=\; \inf_n \sum_{a\in\mathcal{A}} \inf_x P_n(a|x_{-\infty}^{n-1}) \tag{28}$$

and observe that, by (18),

$$\sum_a t_{a,n}(x, y) \;\geq\; \lambda_0 \tag{29}$$

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

Probability of remaining coincident

In the sequel we use the short-hand notation

$$\mathbb{P}\Big(B \;\Big|\; (x_{-\infty}^{n-1}, y_{-\infty}^{n-1})\Big) \;:=\; \mathbb{P}\Big(B \;\Big|\; (\widetilde{X}_{-\infty}^{n-1}, (\widetilde{X}^{[k]})_{-\infty}^{n-1}) = (x_{-\infty}^{n-1}, y_{-\infty}^{n-1})\Big)\;, \tag{30}$$

for B an event measurable with respect to the variables $(\widetilde{X}_n^{\infty}, (\widetilde{X}^{[k]})_n^{\infty})$.

Lemma 4. *If $x_{n-m}^{n-1} = y_{n-m}^{n-1}$ then*

$$\mathbb{P}\Big(\widetilde{X}_n \neq \widetilde{X}_n^{[k]} \;\Big|\; (x_{-\infty}^{n-1}, y_{-\infty}^{n-1})\Big) \;\leq\; |\mathcal{A}|\, \beta(k \wedge m)\;. \tag{31}$$

Proof. By definition of the coupling

$$\mathbb{P}\Big(\widetilde{X}_n \neq \widetilde{X}_n^{[k]} \;\Big|\; (x_{-\infty}^{n-1}, y_{-\infty}^{n-1})\Big) \;=\; \sum_a r_{a,n}(x, y)\;. \tag{32}$$

Inequalities (25) imply

$$\sup_{a,x,y} \Big| P_n(a \mid x_{-\infty}^{n-1}) - P_n^{[k]}(a \mid x_{n-m}^{n-1}\, y_{-\infty}^{n-m-1}) \Big| \;\leq\; \beta(m \wedge k)\;. \tag{33}$$

Hence, in (32),

$$\sum_{a \in A} \Big| P_n(a \mid x_{-\infty}^{n-1}) - P_n^{[k]}(a \mid y_{n-k}^{n-1}) \Big| \;\leq\; |\mathcal{A}|\, \beta(k \wedge m) \qquad \square \tag{34}$$

Let us denote

$$\begin{cases} \beta^*(0) &=& 1 - \lambda_0 \\ \beta^*(n) &=& \min\left(\beta^*(0), |\mathcal{A}|\, \beta(n)\right), \end{cases} \tag{35}$$

Let us introduce the following notation, for integers $m \leq n$

$$D_m^n \;:=\; \bigcap_{j=m}^{n} \{\widetilde{X}_j = \widetilde{X}_j^{[k]}\}\;. \tag{36}$$

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

Lemma 5.

(i) For all integers $m \leq n$ *and* $\ell > 0$, *and all* (x, y) *with* $x_{m-\ell}^{m-1} = y_{m-\ell}^{m-1}$,

$$\mathbb{P}\Big(D_m^n \;\Big|\; (x_{-\infty}^{m-1}, y_{-\infty}^{m-1})\Big) \geq \prod_{j=0}^{n-m}\Big(1 - \beta^*(k \wedge (\ell + j))\Big)\,. \tag{37}$$

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

$$\mathbb{P}\Big(D_n^{n+k-1} \;\Big|\; D_{n-k}^{n-1}\Big) \;\geq\; \Big(1 - \beta^*(k)\Big)^k\,. \tag{38}$$

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

$$\mathbb{P}\Big(D_n^{n+k-1} \;\Big|\; [D_{n-k}^{n-1}]^c\Big) \geq \prod_{j=0}^{+\infty}\Big(1 - \beta^*(j)\Big)\,. \tag{39}$$

Lemma 6.

$$\sup_n \mathbb{P}\Big(\widetilde{X}_n \neq \widetilde{X}_n^{[k]}\Big) \;\leq\; \frac{\sup_n \mathbb{P}([D_n^{n+k-1}]^c)}{\sum_{j=1}^{k-1}\prod_{m=1}^{k-j}(1 - \beta^*(m))} \tag{40}$$

Proof.

$$\begin{aligned}\mathbb{P}([D_n^{n+k-1}]^c) \;=\; & \mathbb{P}\Big(\widetilde{X}_{n+k-1} \neq \widetilde{X}_{n+k-1}^{[k]}\Big) + \\ & \sum_{\ell=n}^{n+k-2} \mathbb{P}\Big(D_{\ell+1}^{n+k-1} \;\Big|\; \widetilde{X}_\ell \neq \widetilde{X}_\ell^{[k]}\Big)\,\mathbb{P}\Big(\widetilde{X}_\ell \neq \widetilde{X}_\ell^{[k]}\Big)\,.\end{aligned} \tag{41}$$

By (37),

$$\mathbb{P}\Big(D_{\ell+1}^{\ell+j} \;\Big|\; \widetilde{X}_\ell \neq \widetilde{X}_\ell^{[k]}\Big) \;\geq\; \prod_{m=1}^{j}\Big(1 - \beta^*(m)\Big)\,. \tag{42}$$

Replacing this in (41) and taking supremum over n in both sides, we obtain (40). □

To conclude the proof of the theorem, we observe that

$$\begin{aligned}\mathbb{P}\Big([D_n^{n+k-1}]^c\Big) = & \mathbb{P}\Big([D_n^{n+k-1}]^c \;\Big|\; D_{n-k}^{n-1}\Big)\,\mathbb{P}\Big(D_{n-k}^{n-1}\Big) \\ & + \mathbb{P}\Big([D_n^{n+k-1}]^c \;\Big|\; [D_{n-k}^{n-1}]^c\Big)\,\mathbb{P}\Big([D_{n-k}^{n-1}]^c\Big)\,.\end{aligned} \tag{43}$$

Using parts (ii) and (iii) of Lemma 5, the right-hand side of (43) can be bounded above by

$$[1-(1-\beta^*(k))^k]+\Big[1-\prod_{j=0}^{+\infty}(1-\beta^*(j))\Big]\mathbb{P}\Big([D_n^{n+k-1}]^c\Big)\ . \tag{44}$$

Hence

$$\sup_n \mathbb{P}\Big([D_n^{n+k-1}]^c\Big) \ \leq\ \frac{1-(1-\beta^*(k))^k}{\prod_{j=0}^{+\infty}(1-\beta^*(j))}\ . \tag{45}$$

Plugging (45) into (40) we finally get

$$\begin{aligned}\sup_n \mathbb{P}\Big(\widetilde{X}_n \neq \widetilde{X}_n^{[k]}\Big) \ &\leq\ \frac{1-(1-\beta^*(k))^k}{\prod_{j=0}^{+\infty}(1-\beta^*(j))\ \sum_{i=1}^{k-1}\prod_{m=1}^{k-i}(1-\beta^*(m))} \\ &\leq\ \frac{1-(1-\beta^*(k))^k}{k\Big[\prod_{j=0}^{+\infty}(1-\beta^*(j))\Big]^2}\ .\end{aligned} \tag{46}$$

By definition, $\beta^*(k)$ is equal to $|\mathcal{A}|\beta(k)$ except, may be, for the first k's. By assumption, $\beta(k)$ is summable, thus $k\beta(k)\to 0$ and

$$1-(1-\beta^*(k))^k \leq \text{const}\ k\,\beta(k)\ . \tag{47}$$

To conclude, we observe that the product in the denominator is strictly positive, again by the summability of the $\beta(k)$. □

Acknowledgements. We are indebted to Xavier Bressaud whose contribution was crucial for the present results and to Pablo Ferrari for years of illuminating discussions about chains of infinite order. It is a pleasure to thank the enthusiastic audience of the Vth Brazilian School of Probability for challenging questions and important comments. The authors thank FAPESP, FINEP, CNPq and an agreement USP-COFECUB for support. RF thanks the University of São Paulo and AG thanks the University of Rouen and the Institut Henri Poincaré for continuous hospitality.

References

[1] A. D. Barbour, L. Holst and S. Janson. *Poisson Approximations*. Oxford University Press, Oxford, 1992.

[2] 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** (1999), 127–38.

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

[4] R. Fernández, P. A. Ferrari and A. Galves. Coupling, renewal and perfect simulation of chains of infinite order, 2001. Notes for a minicourse at the Vth Brazilian School of Probability, available from `http://www.ime.unicamp.br/~ebp5/`.

[5] T. Harris. Nearest-neighbor Markov interaction processes on multidimensional lattices. *Advanves in Math.*, **9** (1972), 66–89.

[6] T. E. Harris. On chains of infinite order. *Pacific J. Math.*, **5** (1955), 707–724.

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

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

[9] O. Onicescu and G. Mihoc. Sur les chaînes statistiques. *C. R. Acad. Sci. Paris,* **200** (1935), 511–12.

Roberto Fernández
Laboratoire de Mathématiques Raphael SALEM
UPRES-A CNRS 6085
Mathématiques, site Colbert
Université de Rouen
F 76821 Mont Saint Aignan
FRANCE
e-mail: Roberto.Fernandez@univ-rouen.fr

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