Perfect simulation of a coupling achieving the $\bar{d}$-distance between ordered pairs of binary chains of infinite order

Antonio Galves Nancy L. Garcia Clémentine Prieur

September 20, 2010

Abstract

We explicitly construct a stationary coupling attaining Ornstein's $\bar{d}$-distance between ordered pairs of binary chains of infinite order. Our main tool is a representation of the transition probabilities of the coupled bivariate chain of infinite order as a countable mixture of Markov transition probabilities of increasing order. Under suitable conditions on the loss of memory of the chains, this representation implies that the coupled chain can be represented as a concatenation of iid sequences of bivariate finite random strings of symbols. The perfect simulation algorithm is based on the fact that we can identify the first regeneration point to the left of the origin almost surely.

Key words: Ornstein's $\bar{d}$-distance, chains of infinite order, ordered binary chains, regenerative scheme.

1 Introduction

Let $\mathbf{X} = (X_n)_{n\in\mathbb{Z}}$ and $\mathbf{Y} = (Y_n)_{n\in\mathbb{Z}}$ be two stationary chains of infinite order on the alphabet $\mathcal{A} = \{0, 1\}$. The $\bar{d}$-distance between $\mathbf{X}$ and $\mathbf{Y}$ is defined as

$$\overline{d}(\mathbf{X},\mathbf{Y}) = \inf\left\{\mathbb{P}(\widetilde{X}_0 \neq \widetilde{Y}_0) \;:\; (\widetilde{\mathbf{X}},\widetilde{\mathbf{Y}}) \text{ stationary coupling of } \mathbf{X} \text{ and } \mathbf{Y}\right\}. \tag{1.1}$$

The $\bar{d}$-distance was introduced by Ornstein in several papers and summarized in an invited article in the first issue of *The Annals of Probability* (Ornstein 1973).

The existence of a stationary coupling attaining the $\bar{d}$-minimum follows from following basic topological considerations.

(i) The product space $(\mathcal{A} \times \mathcal{A})^{\mathbb{Z}}$ is compact by Tychonov's Theorem.

(ii) By Prohorov's Theorem, any sequence of probability measures on $(\mathcal{A}\times\mathcal{A})^{\mathbb{Z}}$ has a convergent subsequence in the weak*-topology.

(iii) Also, the set of all stationary couplings of $\mathbf{X}$ and $\mathbf{Y}$ is a closed subset of the set of all probability measures on $\mathcal{A}^{\mathbb{Z}} \times \mathcal{A}^{\mathbb{Z}}$.

(iv) Finally, the Boolean function $\mathbf{1}_{\{x_0 \neq y_0\}}$ that defines the $\bar{d}$-distance is continuous and bounded.

From (i)–(iv) it follows that there exists at least a coupling which attains the $\bar{d}$-distance. For more details we refer the reader to Theorem 4.1 in Villani (2009).

Obviously this general reasoning does not enable us to explicitly construct a coupling attaining the $\bar{d}$-minimum. In spite the large literature which has been concentrated to this area, as far as we know the problem of finding explicit solutions was addressed only for finite alphabet Markov chains and for finite volume Gibbs measures. To give a further step in this direction is exactly the goal and the novelty of this paper. We solve in a constructive way the problem of finding a coupling attaining the $\bar{d}$-distance between ordered pairs of binary chains of infinite order. First, using basic stationarity arguments, we prove that the $\bar{d}$-distance is bounded below by $|\mathbb{P}(Y_0 = 1) - \mathbb{P}(X_0 = 1)|$. Next, we present an explicit construction of a stationary coupling achieving the infimum (1.1) for stationary chains which are stochastically ordered. This construction can be effectively implemented in an algorithmic way to perfectly sample from this minimal $\bar{d}$-coupling.

This article is organized as follows. In Section 2 we introduce the notation and basic definitions. One coupling that attains the $\bar{d}$-distance is presented in Section 3. The perfect sampling algorithm is described in Section 4 and a pseudo-code implementing it is given by Algorithm 1. The proofs of the theorems are presented in Sections 5 and 6. We conclude the paper with a final discussion and some bibliographic remarks (see Section 7).

2 Basic definitions

In what follows all the processes and sequences of random variables are defined on the same probability space $(\Omega, \mathcal{B}, \mathbb{P})$.

Let $\mathbf{X} = (X_n)_{n\in\mathbb{Z}}$ and $\mathbf{Y} = (Y_n)_{n\in\mathbb{Z}}$ be two stationary chains of infinite order (in the sense of Harris 1955) on the alphabet $\mathcal{A} = \{0, 1\}$. Let p^X and p^Y respectively be the transition probabilities of these chains. This means that for any infinite sequence $x_{-\infty}^{-1} \in \mathcal{A}_{-\infty}^{-1}$ and any symbol $a \in \mathcal{A}$ we have

$$\mathbb{P}(X_0 = a|X_{-\infty}^{-1} = x_{-\infty}^{-1}) \;=\; p^X(a|x_{-\infty}^{-1})\,,$$

$$\mathbb{P}(Y_0 = a|Y_{-\infty}^{-1} = x_{-\infty}^{-1}) \;=\; p^Y(a|x_{-\infty}^{-1})\,.$$

In the above formula $x_{-\infty}^{-1}$ denotes the sequence $(x_i)_{i\leq -1}$ and $\mathcal{A}_{-\infty}^{-1}$ the set of all such sequences. These sequences will be called *pasts*. Given two integers $m \leq n$ we will also use the notation x_m^n to denote the sequence $(x_m, \ldots, x_n)$, and $\mathcal{A}_m^n$ to denote the set of such sequences.

In other terms p^X and p^Y are regular versions of the conditional expectation of X_0 and Y_0 with respect to the σ-algebra generated by $X_{-\infty}^{-1}$ and $Y_{-\infty}^{-1}$ respectively.

Given two pasts $x_{-\infty}^{-1}$ and $y_{-\infty}^{-1}$, we will say that $x_{-\infty}^{-1} \leq y_{-\infty}^{-1}$, if $x_n \leq y_n$ for all $n \leq -1$. This defines a *partial order* on $\mathcal{A}_{-\infty}^{-1}$.

Condition 1: Ordering condition We assume that the chains $\mathbf{X}$ and $\mathbf{Y}$ are stochastically ordered in the following sense

$$p^X(1|x_{-\infty}^{-1}) \leq p^Y(1|y_{-\infty}^{-1})\,, \;\; \text{whenever}\, x_{-\infty}^{-1} \leq y_{-\infty}^{-1}\,. \tag{2.1}$$

The stochastic order between p^X and p^Y makes it possible to construct a stationary coupling between $\mathbf{X}$ and $\mathbf{Y}$ in such a way that for all $n \in \mathbb{Z}$, $X_n \leq Y_n$ with probability 1. This coupling is a stationary chain taking values in the set

$$\mathcal{S} = \{(0, 0), (0, 1), (1, 1)\}\,.$$

The transition probabilities $P : \mathcal{S} \times \mathcal{S}_{-\infty}^{-1} \to [0, 1]$ of this chain are defined as follows: for any pair of ordered pasts $(x_{-\infty}^{-1}, y_{-\infty}^{-1}) \in S_{-\infty}^{-1}$ we have

$$\begin{aligned} P\left((1, 1)|(x_{-\infty}^{-1}, y_{-\infty}^{-1})\right) &= p^X(1|x_{-\infty}^{-1})\,, \\ P\left((0, 0)|(x_{-\infty}^{-1}, y_{-\infty}^{-1})\right) &= p^Y(0|y_{-\infty}^{-1}) \\ P\left((0, 1)|(x_{-\infty}^{-1}, y_{-\infty}^{-1})\right) &= p^X(0|x_{-\infty}^{-1}) - p^Y(0|y_{-\infty}^{-1})\,. \end{aligned} \tag{2.2}$$

We observe that for each pair of ordered pasts $(x_{-\infty}^{-1}, y_{-\infty}^{-1}) \in S_{-\infty}^{-1}$, $P\left((\cdot,\cdot)|(x_{-\infty}^{-1}, y_{-\infty}^{-1})\right)$ is the optimal coupling between $p^X(\cdot|x_{-\infty}^{-1})$ and $p^Y(\cdot|y_{-\infty}^{-1})$.

We want to construct a chain of infinite order on $\mathcal{S}$ invariant with respect to P. This can be done using a regenerative construction of the chain. This regenerative construction is based on a decomposition theorem which states that the stationary chain with infinite memory can be constructed by choosing at each step, in an iid way, the length of the suffix of the string of past symbols we need to look in order to sample the next symbol.

The above mentioned results will follow under certain conditions on the transition probabilities:

Condition 2: Continuity condition The transition probabilities p^X and p^Y on $\mathcal{A}$ are *continuous*, that is, the continuity rates satisfy

$$\max\left\{\beta^X(k), \beta^Y(k)\right\} \to 0 \text{ as } k \to \infty\,,$$

where the continuity rate $\beta^X(k)$ is defined as

$$\beta^X(k) \;=\; \max_{a\in\mathcal{A}} \sup\{|p^X(a|x_{-\infty}^{-1}) - p^X(a|y_{-\infty}^{-1})|,\ \text{for all } x_{-\infty}^{-1},\, y_{-\infty}^{-1} \text{ with } \ x_{-k}^{-1} = y_{-k}^{-1}\}\,, \tag{2.3}$$

and similarly for $\beta^Y(k)$.

To state our third condition we need some extra notation. For each pair $(a,b) \in \mathcal{S}$ and each fixed ordered pair of pasts $(x_{-\infty}^{-1}, y_{-\infty}^{-1}) \in \mathcal{S}_{-\infty}^{-1}$, we define a non-decreasing sequence $r_k((a,b)|(x_{-k}^{-1}, y_{-k}^{-1}))$ such that

$$r_0(a,b) = \inf\left\{P((a,b)\,|\,(u_{-\infty}^{-1}, v_{-\infty}^{-1})) :\ (u_{-\infty}^{-1}, v_{-\infty}^{-1}) \in \mathcal{S}_{-\infty}^{-1}\right\} \tag{2.4}$$

and for $k \geq 1$, $r_k((a,b)|(x_{-k}^{-1}, y_{-k}^{-1}))$ is defined by

$$\inf\left\{P((a,b)\,|\,(u_{-\infty}^{-1}, v_{-\infty}^{-1})) : (u_{-\infty}^{-1}, v_{-\infty}^{-1}) \in \mathcal{S}_{-\infty}^{-1}, u_{-k}^{-1} = x_{-k}^{-1}, v_{-k}^{-1} = y_{-k}^{-1}\right\} \tag{2.5}$$

We then define the non-decreasing sequence $(\alpha_k, k \in \mathbb{N})$

$$\alpha_0 = \sum_{(a,b)\in\mathcal{S}} r_0((a,b)) \tag{2.6}$$

and for $k \geq 1$

$$\alpha_k((x_{-k}^{-1}, y_{-k}^{-1})) = \sum_{(a,b)\in\mathcal{S}} r_k((a,b)|(x_{-k}^{-1}, y_{-k}^{-1})), \tag{2.7}$$

and

$$\alpha_k = \inf\left\{\alpha_k((x_{-k}^{-1}, y_{-k}^{-1})) : (x_{-k}^{-1}, y_{-k}^{-1}) \in \mathcal{S}_{-k}^{-1}\right\}. \tag{2.8}$$

Condition 3:

$$\prod_{k\geq 0} \alpha_k > 0\,. \tag{2.9}$$

To better understand Conditions 2 and 3 we will look at an interesting class of examples which are the renewal processes that forget the past every time they meet the symbol 1. Take $p^X(1|x_{-\infty}^{-1}) = q^X_{\ell(x_{-\infty}^{-1})}$ and $p^Y(1|y_{-\infty}^{-1}) = q^Y_{\ell(y_{-\infty}^{-1})}$ where $\ell(u_{-\infty}^{-1}) = \inf\{n \geq 1 : u_{-n} = 1\}$. We will consider the case when the expectation of the distance between two successive renewal points in the X process and the Y process are finite. That is,

$$\sum_{k=0}^{\infty}\prod_{j=0}^{k}(1-q_j^X) < \infty \quad \text{and} \quad \sum_{k=0}^{\infty}\prod_{j=0}^{k}(1-q_j^Y) < \infty.$$

The divergent case corresponds to the degenerated case in which the only stationary process with these transition probabilities is the zero sequence. Notice that $\sum_{k=0}^{\infty}\prod_{j=0}^{k}(1-q_j) = \infty$ is weaker than $\sum_{k=0}^{\infty} q_k < \infty$.

Example 1: If $\lim_{k\to\infty} q_k^X \searrow q_\infty^X > 0$ and $\lim_{k\to\infty} q_k^Y \searrow q_\infty^X > 0$ exist, then Condition 2 is satisfied. On the other hand, if we take $q_0^X = q_{2k}^X \neq q_{2k+1}^X = q_1^X$ with $0 < q_0^X < q_1^X < 1$, Condition 2 is not satisfied.

Example 2: If $\lim_{k\to\infty} q_k^X \searrow q_\infty^X > 0$ and $\lim_{k\to\infty} q_k^Y \searrow q_\infty^X > 0$, Condition 3 is equivalent to

$$\prod_n (1 - q_n^X + q_\infty^X)(1 - q_n^Y + q_\infty^Y) > 0.$$

For instance, it is enough to have $\sum_n (q_n^X - q_\infty^X) = +\infty$ or $\sum_n (q_n^Y - q_\infty^Y) = +\infty$ to break Condition 3.

3 Construction of our coupling

The goal of this section is to present a coupling between the chains $(X)_n$ and $(Y)_n$ that attains the $\bar{d}$-distance given by $|\mathbb{P}(Y_0 = 1) - \mathbb{P}(X_0 = 1)|$. To obtain such a coupling Conditions 1–3 are required. Therefore, we assume from now on that they are satisfied.

To start the construction we first decompose the transition probability P given by (2.2) as a convex combination of increasing order finite Markov kernels P_k on $\mathcal{S} \times \mathcal{S}_{-k}^{-1}$ for $k \geq 1$.

Let us define a probability distribution $(\lambda_k, k \in \mathbb{N})$ as follows.

$$\lambda_0 = \alpha_0 \tag{3.1}$$

and for $k \geq 1$

$$\lambda_k = \alpha_k \; - \; \alpha_{k-1}. \tag{3.2}$$

The fact that $(\lambda_k, k \in \mathbb{N})$ is a probability distribution follows from the fact that $\alpha_k \to 1$ as k diverges. Obviously this follows from Condition 2.

Theorem 3.3 *There exists a sequence of transition probabilities P_k on $\mathcal{S} \times \mathcal{S}_{-k}^{-1}$ for $k \geq 1$ and a probability measure P_0 on $\mathcal{S}$ such that for any pair of symbols (a, b) in $\mathcal{S}$ and any ordered pair of pasts $(x_{-\infty}^{-1}, y_{-\infty}^{-1}) \in \mathcal{S}_{-\infty}^{-1}$ we have*

$$P\left((a,b)|(x_{-\infty}^{-1}, y_{-\infty}^{-1})\right) = \lambda_0 P_0\left((a,b)\right) + \sum_{k=1}^{\infty} \lambda_k P_k\left((a,b)\,|\,(x_{-k}^{-1}, y_{-k}^{-1})\right) \,. \tag{3.4}$$

This decomposition allows us to construct simultaneously the pair of chains $(X_n, Y_n)_{n\in\mathbb{Z}}$ taking values in $\mathcal{S}$ by concatenating bivariate iid strings. This is done as follows.

Let now $\mathbf{L} = \{L_n, n \in \mathbb{Z}\}$ be an iid sequence of random variables such that $\mathbb{P}(L_n = k) = \lambda_k$ where $(\lambda_k, k \in \mathbb{N})$ is given by (3.1) and (3.2). Define also

$$T_0 \;=\; \sup\{z \leq 0; L_{z+m} \leq m, \text{ for all } m \geq 0\}$$

and for $n \geq 1$

$$T_{-n} \;=\; \sup\{z < T_{-n+1}; L_{z+m} \leq m, \text{ for all } m \geq 0\}$$

and

$$T_n \;=\; \inf\{z > T_{n-1}; L_{z+m} \leq m, \text{ for all } m \geq 0\}.$$

Given the random variables $\mathbf{L} = \{L_n, n \in \mathbb{Z}\}$ and $\mathbf{T} = \{T_j, j \in \mathbb{Z}\}$, we construct the bivariate chain $\{(X_n, Y_n), n \in \mathbb{Z}\}$ by concatenating the bivariate strings $(X_{T_j}^{T_{j+1}-1}, Y_{T_j}^{T_{j+1}-1})$. Each one of these strings is constructed as follows.

1. Choose $(X_{T_j}, Y_{T_j}) \in \mathcal{S}$ with probability P_0 independently of the past.

2. For any $T_j < n \leq T_{j+1} - 1$ choose $(X_n, Y_n) \in \mathcal{S}$ with probability

$$P_{L_n}\left((\cdot,\cdot)|(X_{n-L_n}^{n-1} = x_{n-L_n}^{n-1}, Y_{n-L_n}^{n-1} = y_{n-L_n}^{n-1})\right).$$

Observe that if $T_j \leq n < T_{j+1}$ then $n - L_n \geq T_j$ and therefore the choice of the pair (X_n, Y_n) is made independently of the choice of the symbols $(X_{-\infty}^{T_j-1}, Y_{-\infty}^{T_j-1})$. In this construction, the transition probabilities P_k are those appearing in Expression (3.4).

The existence of infinitely many finite renewal points T_n is given in the next theorem.

Theorem 3.5 *The sequence of random times* $\mathbf{T} = (T_n, n \in \mathbb{Z})$ *with* $\ldots, T_{-1} < T_0 \leq 0 < T_1 < T_2 < \ldots$ *satisfies*

(i) $\mathbb{P}$*-almost surely, all the random times* $\ldots T_{-1} < T_0 \leq 0 < T_1 < T_2 < \ldots$ *are finite.*

(ii) The random pairs of strings $(X_{T_i}^{T_{i+1}-1}, Y_{T_i}^{T_{i+1}-1}),\ i \neq 0$ *are mutually independent and identically distributed. The pair of strings* $(X_{T_0}^{T_1-1}, Y_{T_0}^{T_1-1})$ *is independent from the others.*

We can now present a stationary coupling attaining the $\bar{d}$-distance. This coupling is obtained concatenating the independent strings $(X_{T_i}^{T_{i+1}-1}, Y_{T_i}^{T_{i+1}-1}),\ i \in \mathbb{Z}$. For this coupling we have the following theorem.

Theorem 3.6 *The coupling obtained by concatenating the independent strings* $(X_{T_i}^{T_{i+1}-1}, Y_{T_i}^{T_{i+1}-1}),\ i \in \mathbb{Z}$ *attains the* $\bar{d}$*-distance between* $\mathbf{X}$ *and* $\mathbf{Y}$*.*

4 Perfect simulation algorithm

Given two fixed times $m \leq n$, we want to perfectly sample (X_m^n, Y_m^n) according to our minimal $\bar{d}$-coupling between the chains $\mathbf{X}$ and $\mathbf{Y}$ described in Section 3.

There is an obvious difficulty: we cannot identify a regeneration point experimentally. This follows from the fact that, for any $j \in \mathbb{Z}$ the event "j is a regeneration point" is measurable with respect to the σ-algebra generated by the random variables $L_{j+k}, k \geq 0$.

This difficult will be overcome by Algorithm 1 whose pseudo-code is given below. Algorithm 1 will produce a sequence $(\tilde{X}_m^n, \tilde{Y}_m^n)$ as follows. We sequentially choose iid random variables $L_s, s = n, n-1, \ldots,$ with distribution $\mathbb{P}(L_s = k) = \lambda_k$ given by (3.1) and (3.2). The algorithm checks every time $t \leq m$, until it finds the first one which has the property that

$$L_s \leq s - t \ \text{, for all} \ \ s = t, \ldots, n\,.$$

Call $T[m,n]$ the first $t \leq m$ which has this property:

$$T[m,n] = \sup\{t \leq m; L_s \leq s-t, \text{ for all } s = t, \ldots, n\}.$$

The random time $T[m,n]$ indicates how far back into the past we have to look in order to construct $(\tilde{X}_m^n, \tilde{Y}_m^n)$.

In other terms, if $T[m,n] = t$ then we can choose $(\tilde{X}_t, \tilde{Y}_t)$ independently of the past with distribution P_0. Moreover, the next pair $(\tilde{X}_{t+1}, \tilde{Y}_{t+1})$ can be chosen using distribution either P_0 or $P_1(\cdot|(\tilde{X}_t, \tilde{Y}_t))$ and recursively we can choose all the sequence $(\tilde{X}_t^n, \tilde{Y}_t^n)$ without knowledge of the symbols occurring before time $T[m,n]$. The kernels P_0 and P_k are defined as in Theorem 3.3.

The sequence $(\tilde{X}_m^n, \tilde{Y}_m^n)$ produced by Algorithm 1 in a finite number of steps depends on the particular choice of the random variables $L_j, j = T[m,n], \ldots, n$. Let us call this choice $\tilde{l}_j, j = T[m,n], \ldots, n$. On the other hand, the sequence (X_m^n, Y_m^n) produced by the theoretical construction presented in Section 3 depends on the choice of $L_j, j \in \mathbb{Z}$. Let us call $l_j, j \in \mathbb{Z}$ this choice. The important point to stress is that if $\tilde{l}_j = l_j, j = T[m,n], \ldots, n$ then $(\tilde{X}_m^n, \tilde{Y}_m^n) = (X_m^n, Y_m^n)$. This is the content of the following theorem.

We will prove the following theorem.

Theorem 4.1 *Under Conditions 1–3, for the decomposition given by (3.4), for every pair of integers $m \leq n$, we have:*

(a) $T[m,n]$ is a.s. finite.

(b) The event $\{T[m,n] = t\}$ is measurable with respect to the σ-algebra generated by the random variables $L_s, t \leq s \leq n$.

(c) Algorithm 1 stops almost surely after a finite number of steps.

(d) The sequence $(\tilde{X}_m^n, \tilde{Y}_m^n)$ produced by Algorithm 1 is a perfect sample of the minimal $\bar{d}$-coupling between the chains $\mathbf{X}$ and $\mathbf{Y}$ described in Section 3.

5 Proofs of Theorems 3.3 and 3.5

Proof of Theorem 3.3

Algorithm 1 Perfect simulation for a minimal $\bar{d}$-coupling

Require: Two integers $m \leq n$.

Ensure: The bivariate string $(\tilde{X}_m^n, \tilde{Y}_m^n)$ and the past time $T[m, n]$.

$B \leftarrow \emptyset$ $\{B$ is the set of time positions s for which the pair $(\tilde{X}_s, \tilde{Y}_s)$ has already been chosen$\}$
$t \leftarrow m$
$s \leftarrow t$
while $s \leq n$ **do**
    if $s \notin B$ **then**
        choose L_s with distribution $\mathbb{P}(L_s = k) = \lambda_k$ independently of everything
        if $L_s > s - t$ **then**
            $t \leftarrow t - 1$
            $s \leftarrow t$
        end if
    else
        choose $(\tilde{X}_s, \tilde{Y}_s)$ with distribution $P_{L_s}((\cdot, \cdot)|(\tilde{X}_{s-L_s}^{s-1}, \tilde{Y}_{s-L_s}^{s-1}))$
        $B \leftarrow B \cup \{s\}$
        $s \leftarrow s + 1$
    end if
end while
$T[m, n] \leftarrow t$
return $(\tilde{X}_m^n, \tilde{Y}_m^n)$, $T[m, n]$

Before starting the proof let us sketch its main ideas. Given an ordered pair of "past" strings $(x_{-\infty}^{-1}, y_{-\infty}^{-1})$, we want to randomly choose a new random pair of symbols $(a, b) \in \mathcal{S}$ according to $P(\cdot|(x_{-\infty}^{-1}, y_{-\infty}^{-1}))$. This random choice can be performed as follows. First make a partition $\{I((a,b)|(x_{-\infty}^{-1}, y_{-\infty}^{-1})), (a,b) \in \mathcal{S}\}$ of the interval $[0,1]$ where the length of $I((a,b)|(x_{-\infty}^{-1}, y_{-\infty}^{-1})$ is equal to $P((a,b)|(x_{-\infty}^{-1}, y_{-\infty}^{-1}))$. Then, choose a random element ξ uniformly distributed in $[0,1]$. If $\xi \in I((a,b)|(x_{-\infty}^{-1}, y_{-\infty}^{-1})$, then choose (a,b) as the new pair of symbols. It turns out that $I((a,b)|(x_{-\infty}^{-1}, y_{-\infty}^{-1})$ can be decomposed as the following disjoint union

$$I((a,b)|(x_{-\infty}^{-1}, y_{-\infty}^{-1}) \;=\; I_0((a,b)) \cup \cup_{k\geq 1} I_k((a,b)|(x_{-k}^{-1}, y_{-k}^{-1})), \tag{5.1}$$

where the length of $I_0((a,b))$ and $I_k((a,b)|(x_{-k}^{-1}, y_{-k}^{-1})$ are suitably chosen. Loosely speaking, the length of the interval $I_k((a,b)|(x_{-k}^{-1}, y_{-k}^{-1})$ is the smallest probability to choose (a,b) for any pair of ordered pasts having $(x_{-k}^{-1}, y_{-k}^{-1})$ as ending sequence.

We can consider a second different partition of $[0,1]$ by using the increasing sequence $0 < \alpha_0 \leq \alpha_1 \leq \ldots$. The length of the kth element of this partition is precisely λ_k. Loosely speaking, if ξ falls on this interval, then we only need to look at the last k symbols of the past.

Formally this is done as follows. Let us define a partition of the interval $[0,1]$ formed by the disjoint intervals

$$I_0((0,0)), I_0((0,1)), I_0((1,1)),$$

and for $k \geq 1$,

$$I_k((0,0)|(x_{-k}^{-1}, y_{-k}^{-1})), I_k((0,1)|(x_{-k}^{-1}, y_{-k}^{-1})), I_k((1,1)|(x_{-k}^{-1}, y_{-k}^{-1})), \ldots$$

disposed in the above order in such a way that the left extreme of one interval coincides with the right extreme of the precedent. These intervals have length

$$|I_0((a,b))| = r_0((a,b)) \tag{5.2}$$

and for $k \geq 1$,

$$|I_k((a,b)\,|\,(x_{-k}^{-1}, y_{-k}^{-1}))| = r_k((a,b)|(x_{-k}^{-1}, y_{-k}^{-1})) - r_{k-1}((a,b)|(x_{-(k-1)}^{-1}, y_{-(k-1)}^{-1})). \tag{5.3}$$

Notice that the continuity of transition probabilities p^X and p^Y implies that

$$r_k((a,b)|(x_{-k}^{-1}, y_{-k}^{-1})) \to P((a,b)|(x_{-\infty}^{-1}, y_{-\infty}^{-1})) \tag{5.4}$$

as k diverges.

By construction,

$$P((a,b)|(x_{-\infty}^{-1}, y_{-\infty}^{-1})) = |I_0((a,b))| + \sum_{k\geq 1} |I_k((a,b)|(x_{-k}^{-1}, y_{-k}^{-1}))|. \tag{5.5}$$

Therefore, we can simulate $P((a,b)|(x_{-\infty}^{-1}, y_{-\infty}^{-1}))$ by using an auxiliary random variable ξ uniformly distributed on $[0,1]$ as

$$P((a,b)|(x_{-\infty}^{-1}, y_{-\infty}^{-1})) \;=\; \mathbb{P}\left(\xi \in I_0((a,b)) \cup \cup_{k\geq 1} I_k((a,b)|(x_{-k}^{-1}, y_{-k}^{-1}))\right). \tag{5.6}$$

Observe that the right hand side of this equality can be rewritten as

$$\sum_{k\geq 0} \mathbb{P}\left(\xi \in [\alpha_{k-1}, \alpha_k)\right) \mathbb{P}\left(\xi \in I_0((a,b)) \cup \bigcup_{j\geq 1} I_j((a,b)|(x_{-j}^{-1}, y_{-j}^{-1})) \,|\, \xi \in [\alpha_{k-1}, \alpha_k)\right) \tag{5.7}$$

where $\alpha_{-1} = 0$.

By construction,

$$[0, \alpha_k) \cap \bigcup_{j>k} I_j((a,b)|(x_{-j}^{-1}, y_{-j}^{-1})) = \emptyset.$$

In other terms, for each k, the conditional probabilities on the right hand side of (5.7) depend on the suffix $(x_{-k}^{-1}, y_{-k}^{-1})$ and not on the remaining terms $(x_{-\infty}^{-(k+1)}, y_{-\infty}^{-(k+1)})$. Moreover,

$$\sum_{(a,b)\in\mathcal{S}} \mathbb{P}\left(\xi \in I_0((a,b)) \cup \bigcup_{j\geq 1} I_j((a,b)|(x_{-j}^{-1}, y_{-j}^{-1})) \,|\, \xi \in [\alpha_{k-1}, \alpha_k)\right) = 1.$$

Therefore, we are entitled to define the order k Markov probability transitions P_k as

$$P_k((a,b)|(x_{-k}^{-1}, y_{-k}^{-1})) \;=\; \mathbb{P}\left(\xi \in I_0((a,b)) \cup \bigcup_{j\geq 1} I_j((a,b)|(x_{-j}^{-1}, y_{-j}^{-1})) \,|\, \xi \in [\alpha_{k-1}, \alpha_k)\right). \tag{5.8}$$

Finally we define the probability distribution $(\lambda_k, k \in \mathbb{N})$ as follows.

$$\lambda_0 \;=\; \mathbb{P}(\xi \in [0, \alpha_0)) \;=\; \alpha_0 \tag{5.9}$$

and for $k \geq 1$

$$\lambda_k \;=\; \mathbb{P}(\xi \in [\alpha_{k-1}, \alpha_k)) \;=\; \alpha_k - \alpha_{k-1}. \tag{5.10}$$

This concludes the proof. $\square$

Proof of Theorem 3.5

Define the event B_n as "n is a regeneration point". Formally,

$$B_n = \bigcap_{m\geq 0}\{L_{n+m}\leq m\}. \tag{5.11}$$

Observe that

$$\left(\bigcap_{N\geq 1}\bigcup_{n\geq N} B_n\right)\cap\left(\bigcap_{N\leq 0}\bigcup_{n\leq N} B_n\right) = \bigcap_{k\geq 1}\{T_k<+\infty\}\cap\bigcap_{k\leq 0}\{T_k>-\infty\}. \tag{5.12}$$

Therefore, the existence of infinitely many regeneration times T_n will follow from the following lemma.

Lemma 5.13 *Assume that* $\alpha=\prod_{j=0}^{+\infty}\alpha_j>0$. *Then, for any* $N\in\mathbb{Z}$,

$$\mathbb{P}\left(\bigcup_{n=N}^{\infty} B_n\right) = 1\,.$$

Proof. For any $n\in\mathbb{Z}$ define

$$F_n^0=\{L_n>0\}$$

and $m\geq 1$

$$F_n^m=\bigcap_{j=0}^{m-1}\{L_{n+j}\leq j\}\cap\{L_{n+m}>m\}\,.$$

Define

$$D_1^N=B_N\,,$$

and for $k\geq 2$

$$D_k^N=\bigcup_{n_1=N+1}^{+\infty}\cdots\bigcup_{n_{k-1}=n_{k-2}+1}^{+\infty}\left(F_N^{n_1-N-1}\cap\ldots\cap F_{n_{k-2}}^{n_{k-1}-n_{k-2}-1}\cap B_{n_{k-1}}\right)\,.$$

How to interpret F_N^m? Assume $L_N=0$ and therefore, we can choose (X_N,Y_N) independently of the past symbols $(X_{-\infty}^{N-1},Y_{-\infty}^{N-1})$. From this point on, we look at the values of L_{N+j} and we can choose (X_{N+j},Y_{N+j}) using only the knowledge of $(X_N^{N+j-1},Y_N^{N+j-1})$. This sequence breaks down at $j=m$, since $L_{N+m}>m$ and therefore, the choice of (X_{N+m},Y_{N+m}) depends on the knowledge of symbols occurring before time N.

Therefore, D_k^N is the event in which the trials, described above, starting from time N fail exactly $k-1$ times before finally we find the starting point of a string which is entirely independent of the past symbols. Therefore, the events D_k^N, $k = 1, 2, \ldots$ are disjoint and

$$\bigcup_{n=N}^{+\infty} B_n = \bigcup_{k=1}^{+\infty} D_k^N \,.$$

Therefore

$$\mathbb{P}\left(\bigcup_{n=N}^{+\infty} B_n\right) = \sum_{k=1}^{+\infty} \mathbb{P}(D_k^N)\,.$$

Since the random lengths $\{L_n, n \in \mathbb{Z}\}$ are identically distributed, the probabilities computed above do not depend on the specific choice of N. By definition

$$\mathbb{P}(D_k^N) = \sum_{n_1=N+1}^{+\infty} \cdots \sum_{n_{k-1}=n_{k-2}+1}^{+\infty} \mathbb{P}(F_N^{n_1-N-1} \cap \ldots \cap F_{n_{k-2}}^{n_{k-1}-n_{k-2}-1} \cap B_{n_{k-1}})\,.$$

Using the independence of $F_N^{n_1-N-1}, \ldots, F_{n_{k-1}}^{n_k-n_{k-1}-1}$ and B_{n_k} whenever $N < n_1 < \ldots < n_k$ we can rewrite the right hand side of the last expression as

$$\mathbb{P}(D_k^N) = \sum_{n_1=N+1}^{+\infty} \cdots \sum_{n_k=n_{k-1}+1}^{+\infty} \mathbb{P}(F_N^{n_1-N-1}) \ldots \mathbb{P}(F_{n_{k-1}}^{n_k-n_{k-1}-1})\mathbb{P}(B_{n_k})\,.$$

Since $L_n, n \in \mathbb{Z}$ are iid random variables with $\mathbb{P}(L_0 \le m) = \alpha_m$, for any n, we have

$$\begin{aligned} \mathbb{P}(B_n) &= \mathbb{P}(\cap_{m\ge 0}\{L_{n+m} \le m\}) \\ &= \prod_{m\ge 0} \alpha_m = \alpha \end{aligned}$$

and

$$\sum_{l=n+1}^{+\infty} \mathbb{P}(F_n^{l-n-1}) = 1 - \alpha\,.$$

Therefore, for any $k \ge 1$ we have

$$\mathbb{P}(D_k^N) = \alpha(1-\alpha)^{k-1}$$

and

$$\mathbb{P}\left(\bigcup_{n=N}^{+\infty} B_n\right) = \sum_{k=1}^{+\infty} \alpha(1-\alpha)^{k-1} = 1\,.$$

This concludes the proof of the lemma. $\square$

Lemma 5.13 and the stationarity of the events B_n imply that

$$\mathbb{P}\left(\bigcap_{n=-\infty}^{0} B_n^c\right) = 0.$$

Observe that for each n, if B_n occurs, then (X_n^∞, Y_n^∞) can be chosen independently from from the past symbols $(X_{-\infty}^{n-1}, Y_{-\infty}^{n-1})$. This concludes the proof of Theorem 3.5. □

6 Proof of Theorems 3.6 and 4.1

We begin with a lemma giving a lower bound for the $\bar{d}$-distance between stationary binary chains. For this lemma we are not assuming that the chains are ordered.

Lemma 6.1 *Let $\mathbf{X} = (X_n)_{n\in\mathbb{Z}}$ and $\mathbf{Y} = (Y_n)_{n\in\mathbb{Z}}$ be any two stationary chains on $\{0,1\}$. Then*

$$\bar{d}(\mathbf{X}, \mathbf{Y}) \geq |\mathbb{P}(Y_0 = 1) - \mathbb{P}(X_0 = 1)|\,.$$

Proof. The set of all stationary chains $(X_n', Y_n')_{n\in\mathbb{Z}}$ taking values on $\{0,1\}^2$ such that $\mathbb{P}(X_n' = 1) = \mathbb{P}(X_n = 1)$ and $\mathbb{P}(Y_n' = 1) = \mathbb{P}(Y_n = 1)$ contains the set of all stationary couplings between $(X_n)_n$ and $(Y_n)_n$. Therefore,

$$\inf\left\{\mathbb{P}(\widetilde{X}_0 \neq \widetilde{Y}_0) \;:\; (\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}) \text{ stationary coupling of } \mathbf{X} \text{ and } \mathbf{Y}\right\}$$

is greater than

$$\inf\left\{\mathbb{P}(\overline{X}_0 \neq \overline{Y}_0) \text{ for all } (\overline{X}_0, \overline{Y}_0) \text{ such that } \overline{X}_0 \stackrel{\mathcal{D}}{=} X_0 \text{ and } \overline{Y}_0 \stackrel{\mathcal{D}}{=} Y_0\right\}\ .$$

It is a straightforward computation to check that this last term reaches its minimum with the following optimal coupling between X_0 and Y_0. For any $a \in \{0,1\}$, take

$$\begin{aligned} \mathbb{P}((X_0', Y_0') = (a,a)) &= \min\{\mathbb{P}(X_0 = a), \mathbb{P}(Y_0 = a)\}\,, \\ \mathbb{P}((X_0', Y_0') = (a, 1-a)) &= \mathbb{P}(X_0 = a) - \mathbb{P}((X_0', Y_0') = (a,a))\,. \end{aligned}$$

□

Now we show that for ordered binary stationary chains $|\mathbb{P}(Y_0 = 1) - \mathbb{P}(X_0 = 1)|$ is also an upper bound for $\bar{d}(\mathbf{X}, \mathbf{Y})$.

Consider the coupling obtained by concatenating the independent strings as described in Section 3. Theorems 3.3 and 3.5 imply that the process $(X_n, Y_n)_{n\in\mathbb{Z}}$ taking values in $\mathcal{S}$ is stationary. As a consequence

- the chains $(X_n)_{n\in\mathbb{Z}}$ and $(Y_n)_{n\in\mathbb{Z}}$ constructed simultaneously by the algorithm are also stationary,
- (X_0, Y_0) is a coupling of the probabilities $\mathbb{P}(X_0 = \cdot)$ and $\mathbb{P}(Y_0 = \cdot)$,
- moreover by construction $X_0 \le Y_0$.

There exists a unique optimal coupling between $\mathbb{P}(X_0 = \cdot)$ and $\mathbb{P}(Y_0 = \cdot)$, satisfying the order condition $X_0 \le Y_0$:

$$\mathbb{P}\{(X_0, Y_0) = (0,0)\} = \mathbb{P}(Y_0 = 0)\,,$$

$$\mathbb{P}\{(X_0, Y_0) = (1,1)\} = \mathbb{P}(X_0 = 1)\,,$$

$$\mathbb{P}\{(X_0, Y_0) = (0,1)\} = \mathbb{P}(X_0 = 0) - \mathbb{P}(Y_0 = 0)\,.$$

With this coupling we have

$$\mathbb{P}\{X_0 \neq Y_0\} = \mathbb{P}(Y_0 = 1) - \mathbb{P}(X_0 = 1). \tag{6.2}$$

Equality (6.2) together with Lemma 6.1 concludes the proof of Theorem 3.6. □

To prove Theorem 4.1 let us assume without loss of generality that $m = 0$.

Assertion (a) follows from the fact that for any $n \ge 0$, $T[0,n] \ge T_0$ and by Theorem 3.5, T_0 is finite almost surely.

The proof of (b) follows from the definition of $T[0,n]$.

We want to prove that the number of steps Algorithm 1 makes before stopping is finite. Observe that for each t between $T[0,n]$ and 0, the algorithm must do at most $C(|t| + n)$ steps

- to check if $L_s \le s - t$ for any $t \le s \le n$
- and to assign a value to X_s if this is possible.

In the expression $C(|t| + n)$, C is a fixed positive constant which bounds above the number of operations we need to perform at each single step.

Therefore the total number of steps Algorithm 1 must do before it stops is bounded above by

$$C.\sum_{k=0}^{-T[0,n]} (k+n) = C\left[(-T[0,n]+1).n + \frac{-T[0,n](-T[0,n]+1)}{2})\right]\,.$$

This concludes the proof of (c).

Finally, to prove (d) let us suppose that for $t \leq 0$ we have

$$L_t = 0,\, L_{t+1} \leq 1,\, \ldots, L_n \leq n - t. \tag{6.3}$$

Then, the choice of (X_t^n, Y_t^n), according to the theoretical construction of Section 3, is independent of $L_s, s < t$.

By definition, $T[0,n] = \sup\{t \leq 0; L_t = 0,\, L_{t+1} \leq 1,\, \ldots, L_n \leq n-t\}$. By (a) $T[0,n]$ is almost surely finite. By construction, if $T[0,n] = t$ then

$$(\tilde{X}_t^n, \tilde{Y}_t^n) = (X_t^n, Y_t^n).$$

□

7 Final comments and reference remarks

The main contribution of this article is to present an explicit construction of a stationary coupling between ordered binary chains of infinite order achieving the minimal $\bar{d}$-distance. Moreover, we show that this explicit construction is feasible, in the sense that it can be realized by a perfect simulation algorithm which stops almost surely after a finite number of steps.

Theorem 3.6 can be seen as a generalization to the infinite volume setting of results of Kirillov *et al.* (1989) who show that the classical coupling introduced by Holley (1974) attains $\bar{d}$-distance for finite volume Gibbs states. Besides Kirillov *et al.* (1989) the only other constructive results on this field are Ellis (1976, 1978, 1980a, 1980b) which consider the case of Markov chains on a finite alphabet. Ours seems to be the first constructive solution for chains of infinite order. Several challenges lay ahead. For instance the problem of finding a constructive solution for non-binary chains and/or non-ordered pairs of chains as well as infinite volume Gibbs measures.

Our results can be presented as a constructive solution for the Monge-Kantorovich problem with additive cost function on $C : \mathcal{A}^{\mathbb{Z}} \times \mathcal{A}^{\mathbb{Z}} \to [0,1]$ defined as follows. For any pair of sequences $x_{-\infty}^{+\infty}$ and $y_{-\infty}^{+\infty}$

$$C(x_{-\infty}^{+\infty}, y_{-\infty}^{+\infty}) = \sum_{n \in \mathbb{Z}} c_n |x_n - y_n|\,,$$

where $(c_n)_{n \in \mathbb{Z}}$ is a sequence of positive real numbers, with $\sum_{n \in \mathbb{Z}} c_n = 1$. This follows straightforward from the following observation.

$$d_{MK}(\mathbf{X}, \mathbf{Y})$$

$$
\begin{aligned}
&= \inf\left\{\sum_{n\in\mathbb{Z}} c_n \mathbb{P}(\widetilde{X}_n \neq \widetilde{Y}_n) \;:\; (\widetilde{\mathbf{X}},\widetilde{\mathbf{Y}}) \text{ stationary coupling of } \mathbf{X} \text{ and } \mathbf{Y}\right\} \\
&= \inf\left\{\mathbb{P}(\widetilde{X}_0 \neq \widetilde{Y}_0)\sum_{n\in\mathbb{Z}} c_n \;:\; (\widetilde{\mathbf{X}},\widetilde{\mathbf{Y}}) \text{ stationary coupling of } \mathbf{X} \text{ and } \mathbf{Y}\right\} \\
&= \inf\left\{\mathbb{P}(\widetilde{X}_0 \neq \widetilde{Y}_0) \;:\; (\widetilde{\mathbf{X}},\widetilde{\mathbf{Y}}) \text{ stationary coupling of } \mathbf{X} \text{ and } \mathbf{Y}\right\} \\
&= \overline{d}(\mathbf{X},\mathbf{Y}).
\end{aligned}
$$

The Monge-Kantorovich problem has attracted lots of attention recently. However, to the best of our knowledge, ours are the first results in this direction. The literature on MKP is very extensive. We let the interested reader to find his way starting with the classical reference Rachev (1984) up to the last Villani (2009).

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*). The name chains of infinite order was coined by Harris (1955). We refer the reader to Iosifescu and Grigorescu (1990) for a presentation of the classical material. We refer the reader to Fernández, Ferrari and Galves (2001) for a self contained presentation of chains of infinite order including the representation of chains of infinite order as a countable mixture of finite order Markov chains.

Our Theorem 3.5 is an application to pairs of chains of the results in Comets, Fernández and Ferrari (2002). However, our proof of the result is new and we believe it is simpler than theirs. The representation of chains of infinite order as a countable mixture of Markov chains of increasing order appears explicitly in Kalikow (1990) and implicitly in Ferrari *et al.* (2000) and Comets *et al.* (2002). Regeneration schemes for chains of infinite order have been obtained by Berbee (1987) and by Lalley (1986, 2000).

In the literature, the stochastically order between stochastic chains we considered here is also called *domination*. We refer the reader to the book of Lindvall (1992) for more on the subject.

To assure that Algorithm 1 stops after a finite number of steps we need weaker conditions than our Conditions 1 and 3. This follows from the fact that our Algorithm 1 is inspired by the one proposed in Comets *et al.*(2002) in a different context. For details, we refer the reader to the original article. However, our goal was to sample from a minimal $\bar{d}$-coupling. It is an open issue if this can be done under weaker conditions.

Acknowledgments Thanks to Claire Chauvin, Pierre Collet, Dàvide Gabrielli and Roberto Imbuzeiro Oliveira for discussions and critical remarks. This research was partially supported by PRONEX/ FAPESP's Project 03/09930-9, FAPESP/CNRS project 2006/50339-0 and CNPq grants 305447/2008-4 (AG) and 301530/2007-6 (NLG).

References

[1] Berbee, H. (1987) Chains with infinite connections: uniqueness and Markov representation. *Probab. Theory Related Fields*, **76**(2), 243–253.

[2] Comets, F., Fernández, R., Ferrari, P. A. (2002). Processes with Long Memory: Regenerative Construction and Perfect Simulation. *Ann. Appl. Probab.* **12**(3) 921-943.

[3] Ellis, M. H. (1976) The $\overline{d}$-distance between two Markov processes cannot always be attained by a Markov joining. *Israel J. Math.* **24**(3-4), 269–273.

[4] Ellis, M. H. (1978) Distances between two-state Markov processes attainable by Markov joinings. *Trans. Amer. Math. Soc.* **241**, 129–153.

[5] Ellis, M. H. (1980a) Conditions for attaining $\bar{d}$ by a Markovian joining. *Ann. Probab.* **8**(3), 431–440.

[6] Ellis, M. H. (1980b) On Kamae's conjecture concerning the $\bar{d}$-distance between two-state Markov processes. *Ann. Probab.* **8**(2), 372–376.

[7] Fernández, R., Ferrari, P.A., Galves, A. (2001). Coupling, renewal and perfect simulation of chains of infinite order. *Notes for a mini-course presented in Vth Brazilian School of Probability, 2001.* Can be downloaded from `http://www.ime.usp.br/~pablo/abstracts/vebp.html`

[8] Ferrari, P.A., Maass, A., Martinez, S., Ney, P. (2000). Cesàro mean distribution of group automata starting from measures with summable decay. *Ergodic Theory and Dynamical Systems*,**20**(6), 1657–1670.

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

[10] Holley, R. (1974) Remarks on the FKG Inequalities. *Commun. math. Phys.*, **36**, 227–231.

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

[12] Kalikow, S. (1990) Random Markov processes and uniform martingales., *Isr. J. Math.*, **71**(1), 33-54.

[13] Kirillov, A.B., Radulescu, D.C., and Styer, D.F. (1989). Vasserstein Distances in Two-States Systems. *Journal of Statistical Physics*, **56** (5/6), 931–937.

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

[15] Lalley, S. P. (2000) Regeneration in one-dimensional Gibbs states and chains with complete connections. *Resenhas*, **4**(3), 249–281.

[16] Lindvall, T. (1992) *Lectures on the coupling method.* Wiley, New York.

[17] Ornstein, D. (1973) An application of ergodic theory to probability theory. *Ann. Probab.*, **1**(1), 43–65.

[18] Onicescu, O. and Mihoc, G., (1935) Sur les chaînes statistiques, *C. R. Acad. Sci. Paris*, **200**, pp. 511–512.

[19] Rachev, S. T. (1984) The Monge-Kantorovich problem on mass transfer and its applications in stochastics. (Russian) *Teor. Veroyatnost. i Primenen.*, **29**(4), 625–653. English translation: Theory Probab. Appl. **29**(4), 647–676.

[20] Villani, C. (2009) *Optimal transport. Old and new.* Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], **338**. Springer-Verlag, Berlin.