PERFECT SIMULATION FOR STOCHASTIC CHAINS OF INFINITE MEMORY: RELAXING THE CONTINUITY ASSUMPTION

SANDRO GALLO & NANCY L. GARCIA

Abstract. This paper is composed of two main results concerning chains of infinite order which are not necessarily continuous. The first one is a decomposition of the transition probability kernel as a countable mixture of unbounded probabilistic context trees. This decomposition is used to design a simulation algorithm which works as a combination of the algorithms given by Comets et al. (2002) and Gallo (2009). The second main result gives sufficient conditions on the kernel for this algorithm to stop after an almost surely finite number of steps. Direct consequences of this last result are existence and uniqueness of the stationary chain compatible with the kernel.

1. Introduction

The goal of this paper is to construct a perfect simulation scheme for chains of infinite order on a countable alphabet, compatible with a transition probability kernel which is not necessarily continuous. By a perfect simulation algorithm we mean an algorithm which samples precisely from the stationary law of the process.

Perfect simulation for chains of infinite order was first done by Comets et al. (2002) under the continuity assumption. They used the fact (observed earlier by Kalikow (1990)) that under this assumption, the transition probability kernel can be decomposed as a countable mixture of Markov kernels. Then, Gallo (2009) obtained a perfect simulation algorithm for chains compatible with a class of unbounded probabilistic context trees where each infinite size branch can be a discontinuity point.

In this paper, we consider a class of transition probability kernels which are neither necessarily continuous nor necessarily probabilistic context trees. In fact, the same infinite size branches as in the context trees considered by Gallo (2009) are allowed to be discontinuity points, and the other branches must have a certain localized-continuity assumption. Under these new assumptions, we obtain a Kalikow-type decomposition of our kernels as a mixture of unbounded probabilistic context trees where each infinite size branch can be a discontinuity point.

In this paper, we consider a class of transition probability kernels which are neither necessarily continuous nor necessarily probabilistic context trees. In fact, the same infinite size branches as in the context trees considered by Gallo (2009) are allowed to be discontinuity points, and the other branches must have a certain localized-continuity assumption. Under these new assumptions, we obtain a Kalikow-type decomposition of our kernels as a mixture of unbounded probabilistic context trees where each infinite size branch can be a discontinuity point.

As a consequence of this decomposition and some minimum extra condition, we can show that there exists at least one stationary chain compatible with our kernels, extending the existing result stating that continuity was sufficient. A perfect simulation is then constructed using this decomposition together with the coupling from the past (CFTP) method introduced in the seminal paper of Propp & Wilson (1996). One of the main consequence of the existence of a perfect simulation algorithm is the fact that there exists a unique stationary chain compatible with our kernels.

2000 Mathematics Subject Classification. Primary 60G10; Secondary 60G99.
Key words and phrases. Perfect simulation, regeneration scheme, chains of infinite order.

SG is supported by a FAPESP fellowship (grant 2009/09809-1), NG is supported by CNPq grants 475504/2008-9 and 301530/2007-6.
More precise explanations of what is done here are postponed to Section 3, since we need the notation and definitions given in Section 2. Our first main result, Theorem 4.1, which is stated and proved in Section 4, gives the decomposition which holds without the continuity condition. In Section 5, we explain how our perfect simulation works using Theorem 4.1, and we present it under the form of the pseudo-code, Algorithm 1. After that, we state our second main theorem, Theorem 5.1 which says that Algorithm 1 stops almost surely after a finite number of steps. Section 6 is dedicated to the proof of Theorem 5.1. We finish this paper with some comments and further questions.

2. Notation and definitions

Let $A$ be a countable alphabet. Given two integers $m \leq n$, we denote by $a_m^n$ the string $a_m \ldots a_n$ of symbols in $A$. For any $m \leq n$, the length of the string $a_m^n$ is denoted by $|a_m^n|$ and is defined by $|a_m^n| = n - m + 1$. For any $n \in \mathbb{Z}$, we will use the convention that $a_{n+1}^n = \emptyset$, and naturally $|a_{n+1}^n| = 0$. Given two strings $v$ and $v'$, we denote by $vv'$ the string of length $|v| + |v'|$ obtained by concatenating the two strings. The concatenation of strings is also extended to the case where $v$ denotes a semi-infinite sequence, that is $v = \ldots a_{-2} a_{-1}, a_{-1} \in A$ for $i \geq 1$. If $n$ is a positive integer and $v$ a finite string of symbols in $A$, we denote by $v^n = v \ldots v$ the concatenation of $n$ times the string $v$. We denote

$$A^{-N} = A^{(-\ldots,-2,-1)} \quad \text{and} \quad A^* = \bigcup_{j=0}^{+\infty} A^{(-j, \ldots, -1)},$$

which are, respectively, the set of all infinite strings of past symbols and the set of all finite strings of past symbols. The case $j = 0$ corresponds to the empty string $\emptyset$. Finally, we denote by $a = \ldots a_{-2} a_{-1}$ the elements of $A^{-N}$.

2.1. Standard definitions. A transition probability kernel (or simply kernel in the sequel) on an alphabet $A$ is a function

$$P : A \times A^{-N} \rightarrow [0,1] \quad \text{(1)}$$

such that

$$\sum_{a \in A} P(a|a) = 1, \quad \forall a \in A^{-N}.$$ 

In this paper, we consider kernels $P$ which depends on an unbounded part of the past, unlike the markovian case. A stationary stochastic chain $X = (X_n)_{n \in \mathbb{Z}}$ on $A$ having law $\mu$ is said to be compatible with a kernel $P$ if the later is a regular version of the conditional probabilities of the former, that is

$$\mu(X_0 = a | X_{-\infty} = b) = P(a|b) \quad \text{(2)}$$

for every $a \in A$ and $\mu$-almost every $b$ in $A^{-N}$. We call these chains chains of infinite memory.

2.2. Probabilistic context tree. We say that a kernel $P$ has a probabilistic context tree representation if there exists a function $d : A^{-N} \rightarrow \mathbb{N} \cup \{+\infty\}$ such that for any two infinite sequences of past symbols $a$ and $b$

$$a^{-1}_{d(a)} = b^{-1}_{d(b)} \Rightarrow P(a|b) = P(a|b).$$

It follows that the length $d(a)$ only depends on the suffix $a^{-1}_{d(a)}$ of $a$. This allows us to identify the set $\tau := \{a^{-1}_{d(a)}\}_{a \in A^{-N}}$ with the set of leaves of a rooted tree where
each node has either $|A|$ sons (internal node) or 0 sons (leaf). The set $\tau$ is called the context tree, “context” being the original name Rissanen (1983) gave to the strings $c_\tau(a) := a_{-d(a)}$ when he introduced this model. A probabilistic context tree is an ordered pair $(\tau, p)$ where $\tau$ is a context tree and $p := \{p(a|v)\}_{a \in A, v \in \tau}$ is a set of transition probabilities associated to each element of $\tau$. Thus, the probabilistic context tree $(\tau, p)$ represents the kernel $P$ if for any $a \in A^{-N}$ and any $a \in A$

$$P(a|a) = p(a|c_\tau(a)).$$

Examples of probabilistic context trees are shown in Figures 1(a) (for the bounded case) and 1(b) (for the unbounded case). In the first one, at each leaf (context) of the tree we associate three boxes in which are given the transition probabilities to each symbols of $A$ given this context. In the second one, we only specify the probability $p_i := p(2|0^i2)$ (observe that we swap the order when we write the context in a conditioning), the transition probabilities to 1 are simply $1 - p_i$.

![Image of probabilistic context trees](image)

**Figure 1.** Examples of probabilistic context trees.

A stochastic chain $X$ compatible, in the sense of (2), with a probabilistic context tree is called a chain with variable length memory.

3. Motivation

The aim of this section is twofold: it motivates and explains at the same time, the present work.

3.1. Countable mixture of Markov kernels under the continuity assumption. We say that a point (an infinite sequence of past symbols) $a$ is a continuity point for a given transition probabilities kernel $P$ if

$$\beta_k(a) := \sup_{a \in A} \sup_{y, z} |P(a|a_{-k}^y) - P(a|a_{-k}^z)|^{k \to +\infty} 0.$$ 

Otherwise, we say that $a$ is a discontinuity point for $P$. $P$ is said to be continuous if

$$\beta_k := \sup_{a_{-k}} \beta_k(a)^{k \to +\infty} 0. \tag{3}$$

Kalikow (1990) showed that continuous transition probability kernels $P$ can be represented through the form of a countable mixture of Markov kernels, that is, there
exist two probability distributions \( \{p^C_F(a)\}_{a \in A} \) and \( \{\lambda^C_F\}_{k \geq 0} \) and a sequence of Markov kernels \( \{p^C_k\}_{k \geq 1} \) such that for any \( a \in A \) and \( \bar{z} \in A^{-N} \)
\[
P(a|\bar{z}) = \lambda_0^C p_0^C(a) + \sum_{k \geq 1} \lambda_k^C p_k^C(a|z_{-k}^{-1}).
\tag{4}
\]

The superscript “CFF” refers to the fact that we will use the definitions from [Comets et al. (2002)]

Define an \( \mathbb{N} \)-valued random variable \( K^{\text{CFF}} \) taking value \( k \geq 0 \) w.p. \( \lambda_k^C \). Decomposition (4) means the following. To choose the next symbol looking at the whole past \( \bar{z} \) using the distribution \( \{P(a|\bar{z})\}_{a \in A} \) is equivalent to the following two steps procedure:

1. choose \( K^{\text{CFF}} \),
2. (a) if \( K^{\text{CFF}} = 0 \), then choose the next symbol w.p. \( \{p_0(a)\}_{a \in A} \),
   (b) if \( K^{\text{CFF}} = k > 0 \) then choose the next symbol looking at \( z_{-k}^{-1} \) and using
   \( \{p_k^C(a|z_{-k}^{-1})\}_{a \in A} \).

Observe that \( K^{\text{CFF}} \) is independent of everything (in particular, it does not depend on \( \bar{z} \)).

To clarify the parallel between this decomposition and the decomposition presented in Theorem 4.1, we explain how [Comets et al. (2002)]

\[
\alpha_0^C(a) = \inf_{\bar{z}} P(a|\bar{z}) \quad \text{and} \quad \alpha_k^C(a|a_{-k}^{-1}) = \inf_{\bar{z}} P(a|a_{-k}^{-1} \bar{z})
\]

and the sequence \( \{\alpha_k^C\}_{k \geq 0} \) defined by \( \alpha_0^C = \sum_{a \in A} \alpha_0^C(a) \) and for any \( k \geq 1 \)
\[
\alpha_k^C = \inf_{a_{-1}^{-1} \in A^A} \sum_{a \in A} \alpha_k^C(a|a_{-k}^{-1}).
\]

These are, as they say, “probabilistic threshold for memories limited to \( k \) preceding instants.” Taking the infimum over every \( a_{-k}^{-1} \) is related to the continuity assumption (3). In fact, to assume continuity is equivalent to assume that \( \alpha_k^C \) goes to 1 as \( k \) diverges, and to assume punctual continuity in \( \bar{a} \) is equivalent to assume that
\[
\alpha_k^C(a) := \sum_{a \in A} \alpha_k^C(a|a_{-k}^{-1})
\]
goes to 1 as \( k \) diverges. Under the continuity assumption (3), the probability distribution \( \{\lambda_k^C\}_{k \geq 0} \) used in (4) is defined as follows: \( \lambda_k^C = \alpha_k^C - \alpha_k^{CFF} \) for \( k \geq 1 \) and \( \lambda_0^C = \alpha_0^{CFF} \).

3.2. **Without the continuity assumption.** To fix ideas, in the remaining of this section, assume \( P \) is a transition probability kernel on \( A = \{1, 2\} \) which has a single discontinuity point which is \( 1^{-N} \). Then \( \alpha_k^C(a) \) goes to 1 as \( k \) diverges if and only if \( a \neq 1^{-N} \). In this case, \( \alpha_k^C \) does not converge to 1 and the above result does not apply.

3.2.1. **The context tree assumption.** [Gallo (2009)]

assumed that \( P \) is represented by the probabilistic context tree \((\tau, p)\), where
\[
\tau = 1^{-N} \cup \bigcup_{k \geq 0} \bigcup_{a_{-\ell(i)}^{-1} \in A^{\ell(i)}} a_{-\ell(i)}^{-1} 21^1,
\]
\( \ell : \mathbb{N} \to \mathbb{N} \) being a deterministic function. This context tree is represented in Figure 2

Observe that, under this assumption, for any \( i \geq 0 \) and \( k > \ell(i) \) we have
\[
P(a|1^i 2a_{-k}^{-1} \bar{z}) = P(a|1^i 2a_{-k}^{-1} y), \quad \text{for any } \bar{z} \text{ and } y \in A^{-N}.
\tag{5}
\]
It follows that
\[
\inf_{a^{-1} \in A^k} \sum_{a \in A} \alpha_k(a|1^i2a^{-1}_k) = 1
\]
whenever \( k > \ell(i) \). Therefore \( \sum_{a \in A} \alpha_k(a|a^{-1}_k) \) goes to 1 as \( k \) diverges for any \( a \neq 1^{-N} \). We do not specify what happens for the point \( a = 1^{-N} \). Making a parallel with the above case, we can decompose such a kernel as follows: for any \( a \in A \) and \( z \)
\[
P(a|z) = \lambda_0^{\text{CFF}} p_0^{CFF}(a) + (1 - \lambda_0^{\text{CFF}}) p'(a|c_T(z))
\]
where
\[
p'(a|c_T(z)) := \frac{p(a|c_T(z)) - \lambda_0^{\text{CFF}} p_0^{\text{CFF}}(a)}{1 - \lambda_0^{\text{CFF}}}
\]
Define an \( \mathbb{N} \)-valued random variable \( K^G \) which takes value 0 w.p. \( \lambda_0^{\text{CFF}} \) or \( |c_T(z)| \) w.p. \( 1 - \lambda_0^{\text{CFF}} \). The context tree assumption for \( P \) means the following. To choose the next symbol looking at the whole past \( z \) using the distribution \( \{P(a|z)\}_{a \in A} \) is equivalent to the following two steps procedure:

1. choose \( K^G \),
2. (a) if \( K^G = 0 \), choose the next symbol w.p. \( \{p_0^{\text{CFF}}(a)\}_{a \in A} \),
   (b) if \( K^G = |c_T(z)| \), choose the next symbol looking at \( c_T(z) \) and using \( \{p'_0(a|c_T(z))\}_{a \in A} \).

Observe that the random variable \( K^G \) is a deterministic function of the past \( z \) whenever its value is not 0: \( K^G = |c_T(z)| \).

3.2.2. The countable mixture of probabilistic context trees. So far, two extreme cases have been considered: \( K^G \) is a deterministic function of the past, and \( K^{\text{CFF}} \) is a random variable totally independent of the past. In the present work, we introduce a way to combine these two approaches. It allows us to consider kernels \( P \) which are neither necessarily represented by a probabilistic context tree, nor necessarily continuous. This new approach is based on the assumption that
\[
\alpha_k := \inf_{i \geq 0} \inf_{a^{-1} \in A^k} \sum_{a \in A} \alpha_k^{\text{CFF}}(a|1^i2a^{-1}_k) \xrightarrow{k \to \infty} 1.
\]
The \( \alpha_k \)’s are “probabilistic threshold for memories going until the \( k^{\text{th}} \) instant preceding the last occurrence of symbol 2 in the past.” In this case also, we have that \( \sum_{a \in A} \alpha_k^{\text{CFF}}(a|a^{-1}_k) \) goes to 1 as \( k \) diverges for any \( a \neq 1^{-N} \) and not necessarily for
1−N. Notice that the probabilistic context tree assumption we introduced in Section 3.2.1 only satisfies
\[ \inf_{a_{i-j} \in A^k} \sum_{a \in A} \lambda_i^{CFF} p_i^{CFF} (a) \xrightarrow{k \to \infty} 1. \]
Under assumption (6), it will be shown in the next section that there exists a probability distribution \( \{ \lambda_k \}_{k \geq 0} \), and a sequence of probabilistic context trees \( \{ (\tau_k, p_k) \}_{k \geq 0} \) such that
\[ P(a \mid z) = \lambda_0^{CFF} p_0^{CFF} (a) + \sum_{k \geq 0} \lambda_k p_k (a \mid c_{\tau_k} (z)). \quad (7) \]
The \( k \)-th context tree of decomposition (11) is given by
\[ \tau_k := 1^{-N} \cup \bigcup_{i \geq 0} \bigcup_{a_{i-k} \in A^k} a_{i-k}^{-1} 2 1^i. \quad (8) \]
The sequence of context trees \( \{ \tau_k \}_{k \geq 0} \) for the present particular case is illustrated in Figure 3. Define a random variable \( K^{GG} \) taking values 0 w.p. \( \lambda_0^{CFF} \) and \( |c_{\tau_k} (z)| \) w.p. \( \lambda_k \) for \( k \geq 0 \). One more time, let us translate this decomposition into a two steps procedure:
1. choose \( K^{GG} \),
2. (a) if \( K^{GG} = 0 \), choose the next symbol w.p. \( \{ p_0^{CFF} (a) \}_{a \in A} \),
   (b) if \( K^{GG} = |c_{\tau_k} (z)| \) for some \( k \geq 0 \), choose the next symbol looking at \( c_{\tau_k} (z) \) and using \( \{ p_k (a \mid c_{\tau_k} (z)) \}_{a \in A} \).
Observe that this time, the random variable \( K^{GG} \) depends on the past \( z \), but through a random mechanism using the distribution \( \{ \lambda_k \}_{k \geq 0} \).

In the next section, we state our first main result in a general framework. The alphabet can be countable and the role which is played above by symbol 2 can be played by any finite string \( w \in A^* \). In this case, we allow \( P \) to have discontinuities at every point \( z \in A^{-1} \) which does not have \( w \) as subsequence.

4. First main result: a countable mixture of unbounded probabilistic context trees

4.1. Some more definitions and statement of the first results. Fix a finite size string \( w \in A^* \) and define the function \( m^w \) which associates to any string \( a_{-m}^{-1} \in A^m \), \( |w| \leq m \leq +\infty \) the distance to the first occurrence of \( w \) when we look backward in \( a_{-m}, \ldots, a_{-1} \), that is
\[ m^w (a_{-m}) = \inf \{ k \geq 0 : a_{-k-1}^{-1} = w \}, \quad (9) \]
Theorem 4.1. Consider a transition probability kernel $P$ such that

$$a_k^w := \inf_{\bar{a} \in \mathbb{A}} \inf_{c \in \mathbb{C}} \inf_{\bar{z} \in \mathbb{Z}} \sum_{a \in \mathbb{A}} P(a | b_{-1}^w c_{-1}^\bar{a}) \frac{k^{|w|}}{1}. \tag{10}$$

Then, there exist two probability distributions \{\lambda_k^w\}_{k \geq 1} and \{p_{-1}^w(a)\}_{a \in A}, and a sequence of probabilistic context trees \{(\tau_k^w, p_k^w)\}_{k \geq 0} such that

$$P(a | \bar{z}) = \lambda_k^w p_{-1}^w(a) + \sum_{k \geq 0} \lambda_k^w p_k^w(a | c_{-1}^\bar{a}(\bar{z})). \tag{11}$$

Corollary 4.1. Under the same condition of Theorem 4.1 if

$$\inf_{a \in \mathbb{A}} P(a | \bar{a}) > 0, \tag{12}$$

then there exists at least one stationary chain compatible with $P$ in the sense of \cite{4}. \hfill \square

Corollary 4.1 follows from Theorem 4.1 by the same arguments used in proof of Theorem 11 in \cite{Kalikow1990}. The decomposition of $P$ as mixture of probabilistic context trees together with assumption \cite{12} provide the necessary features to obtain that the limit of the empirical measures is in fact compatible with $P$.

Since the assumption of Theorem 4.1 is not intuitive, let us give sufficient conditions for Theorem 4.1 to hold.

Proposition 4.1.

$$\inf_{a_{-1}^k \in \mathbb{A}^k} \sum_{a \in \mathbb{A}} P(a | a_{-1}^k \bar{z}) \frac{k^{+ \infty}}{1} \tag{13}$$

implies that

$$\inf_{a_{-1}^k \in \mathbb{A}^k} \sum_{a \in \mathbb{A}} P(a | a_{-1}^k \bar{z}) \frac{k^{+ \infty}}{1} \tag{14}$$

which implies that $\alpha_k^w$ converges to 1 as $k$ diverges.

A consequence of this proposition is that the assumption of Theorem 4.1 is weaker than the continuity assumption. It follows, in particular, that the existence result of Corollary 4.1 extends the well-known fact that existence hold whenever the transition probability kernel is continuous.

The proofs of Theorem 4.1 and Proposition 4.1 are given in Section 4.3.

Theorem 4.1 is based on the existence of a triplet of parameters (which is not unique): two probability distributions \{\lambda_k^w\}_{k \geq 1} and \{p_{-1}^w(a)\}_{a \in A}, and a sequence of probabilistic context trees \{(\tau_k^w, p_k^w)\}_{k \geq 0}. What follows is dedicated to the definition of such a triplet of parameters.
4.2. A triplet of parameters \( \{(A_k^w)_{k \geq -1}, \{p_{k-1}^w(a)\}_{a \in A}, \{(\tau_k^w, p_k^w)\}_{k \geq 0}\} \). We fix a string \( w \) of \( A^* \). The definition of our triplet is based on two partitions of \([0,1]\) inspired by Comets et al. (2002). Let us first show that for any \( a \in A, i \geq 0, b_i^{-1} \in I_i(\tilde{w}) \) and \( \varepsilon \in A^{-N} \),

\[
\inf_{\tilde{z}} P(a|b_i^{-1} w c_{-k}^{-1} z) \xrightarrow{k \to +\infty} P(a|b_i^{-1} w \varepsilon).
\]

Observe that

\[
0 \leq \sum_{a \in A} \left( P(a|b_i^{-1} w \varepsilon) - \inf_{\tilde{z}} P(a|b_i^{-1} w c_{-k}^{-1} z) \right) = 1 - \sum_{a \in A} \inf_{\tilde{z}} P(a|b_i^{-1} w c_{-k}^{-1} z).
\]

Moreover,

\[
\sum_{a \in A} \inf_{\tilde{z}} P(a|b_i^{-1} w c_{-k}^{-1} z) \geq \inf_{i \geq 0} \inf_{b_i^{-1} \in I_i(\tilde{w})} \inf_{c_{-k}^{-1} \in A^k} \sum_{a \in A} \inf_{\tilde{z}} P(a|b_i^{-1} w c_{-k}^{-1} z)
\]

therefore, under assumption (10), \( \sum_{a \in A} \inf_{\tilde{z}} P(a|b_i^{-1} w c_{-k}^{-1} z) \) goes to 1 and

\[
\sum_{a \in A} \left( P(a|b_i^{-1} w \varepsilon) - \inf_{\tilde{z}} P(a|b_i^{-1} w c_{-k}^{-1} z) \right) \xrightarrow{k \to +\infty} 0.
\]

Since all the terms in the sum over \( A \) are positive the convergence of (15) holds.

4.2.1. Definition of the first partition of \([0,1]\). We will denote for any \( v \in A^{-N} \cup A^* \) such that \( m^w(v) < +\infty \)

\[
\alpha^w(a, v, k) := \inf_{\tilde{z}} P\left(a|v_{-m^w(v)} w v_{-(m^w(v)+|v|+1)} z\right)
\]

for \( k = 0, \ldots, |v| - m^w(v) - |w| \). This notation is not ambiguous since once we fix \( v \) and \( k \), we automatically fix \( v_{-m^w(v)} \) and \( v_{-(m^w(v)+|v|+1)} \). Now, let us introduce the partition which is illustrated in the upper part of Figure 4.

Define for any \( a \in A \)

\[
\alpha(a) := \inf_{\tilde{z}} P(a|\tilde{z}),
\]

and the collection of intervals \( \{I(a)\}_{a \in A} \), each one having length \( |I(a)| = \alpha(a) \). For any \( v \in A^* \cup A^{-N} \) such that \( m^w(v) < +\infty \) we define the collection of intervals

\[
I^w(a, v, k) = a \in A, k = 0, \ldots, |v| - m^w(v) - |w|,
\]

each one having length

\[
|I^w(a, v, k)| = \alpha^w(a, v, k) - \alpha^w(a, v, k - 1)
\]

for \( k \geq 1 \), and

\[
|I^w(a, v, 0)| = \alpha^w(a, v, 0) - \alpha(a).
\]

Suppose now we are given an entire past \( \tilde{z} \in A^{-N}(w) \), and glue these intervals in the following order

\[I(1), I(2), \ldots, I^w(1, \tilde{z}, 0), I^w(2, \tilde{z}, 0), \ldots, I^w(1, \tilde{z}, 1), I^w(2, \tilde{z}, 1) \ldots\]

in such a way that the left extreme of \( I(1) \) coincides with 0, and the left extreme of each intervals coincides with the right extreme of the preceding interval. What we obtain, by the convergence (15), is a partition of \([0,1]\) such that for any \( a \in A \)

\[
\text{Leb}\left(I(a) \cup \bigcup_{k \geq 0} I^w(a, \tilde{z}, k)\right) = P(a|\tilde{z}),
\]

where \( \text{Leb} \) denote the Lebesgue measure on \([0,1]\). It is important to notice that for any \( a, k \) and \( \tilde{z} \in A^{-N}(w) \), we can construct the interval \( I^w(a, \tilde{z}, k) \) knowing only the suffix \( z_{-(k+|w|+m^w(\tilde{z}))} \).
Proof of Theorem 4.1.

What we have to prove is that equality (11) holds with the probabilities of the first partition stated in Theorem 4.1. Define

\[ P(a|z) = \mathbb{P}(U_0 \in I(a)) + \mathbb{P} \left( U_0 \in \bigcup_{k \geq 0} I^w(a, z, k) \right), \tag{20} \]

4.2.2. Definition of the second partition of \([0,1]\). Observe that \(\{\alpha_k^w\}_{k \geq 0}\) is a \([0,1]\)-valued non-decreasing sequence which converges to 1 as \(k\) diverges. It follows that denoting \(\alpha_{-1}^w := \sum_{a \in A} \alpha(a)\), and using the convention that \(\alpha_0^w = 0\), the sequence of intervals \(\{[\alpha_{-1}^w, \alpha_k^w]\}_{k \geq 1}\) constitutes a partition of \([0,1]\). This partition is illustrated in the lower part of Figure 4.

4.2.3. Definition of the triplet. Let us introduce an i.i.d. chain \(\mathbf{U} = (U_t)_{t \in \mathbb{Z}}\) of random variables uniformly distributed in \([0,1]\). We denote by \((\Omega, \mathcal{F}, \mathbb{P})\) the corresponding probability space. It is the only probability space we will consider all along this paper.

We now introduce one triplet \(\{(\lambda_k^w)_{k \geq 1}, \{p_{w-1}^w(a)\}_{a \in A}, \{(\tau_k^w, p_k^w)\}_{k \geq 0}\}\) that will give the decomposition stated in Theorem 4.1. Define

- For any \(k \geq -1\)
  \[ \lambda_k^w := \mathbb{P}(U_0 \in [\alpha_{k-1}^w, \alpha_k^w]). \tag{16} \]
- For any \(a \in A\)
  \[ p_{w-1}^w(a) := \mathbb{P}(U_0 \in I(a) | U_0 \in [0, \alpha_{w-1}^w]) = \alpha(a) / \alpha_{w-1}^w. \tag{17} \]
- For any \(k \geq 0\), let
  \[ \tau_k^w := A^{-N}(w) \cup \bigcup_{i \geq 0} \bigcup_{b_i^w \in T(w)} \bigcup_{c_i^w \in A^v} e_i^{-1} w b_i^{-1}, \tag{18} \]
  and for \(v \in \tau_k^w\) (that is, \(|v| = m^w(v) + |w| + k\)) we put
  \[ p_k^w(a|v) := \begin{cases} \mathbb{P}(U_0 \in I(a) \cup \bigcup_{l=0}^{\infty} I^w(a, v, l) | U_0 \in [\alpha_{k-1}^w, \alpha_k^w]) & \text{if } m^w(v) < +\infty, \\ \frac{P(a|v) - \lambda_k^w p_{w-1}^w(a)}{1 - \lambda_{k-1}^w} & \text{otherwise.} \end{cases} \tag{19} \]

Two examples of sequences of context trees \(\{\tau_k^w\}_{k \geq 0}\) on \(A = \{1, 2\}\) are given in Figures 3 and 5. The first one with \(w = 2\), and the second one with \(w = 12\).

4.3. Proofs of the results of this section.

Proof of Theorem 4.1. What we have to prove is that equality (11) holds with the triplet \(\{(\lambda_k^w)_{k \geq 1}, \{p_{w-1}^w(a)\}_{a \in A}, \{(\tau_k^w, p_k^w)\}_{k \geq 0}\}\) introduced above. On the one hand, using (16), for any \(a \in A\) and \(z \in A^{-N}(w)\) we have

\[ P(a|z) = \mathbb{P}(U_0 \in I(a)) + \mathbb{P} \left( U_0 \in \bigcup_{k \geq 0} I^w(a, z, k) \right), \tag{20} \]
where the second term can be rewritten
\[
\sum_{k \geq 0} P(U \in [\alpha_{k-1}^w, \alpha_k^w]) P\left(U \in \bigcup_{l \geq 0} I^w(a, z, l) \bigg| U \in [\alpha_{k-1}^w, \alpha_k^w]\right).
\]

On the other hand, by the definition of $\alpha_k^w, k \geq -1$, we have $I(a) \subset [0, \alpha_{-1}^w]$ and for any $k \geq 0$
\[
[0, \alpha_k^w] \subset \bigcup_{a \in A_{l=0}} \bigcup_{l=0}^k I^w(a, z, l),
\]
it follows that for any $\hat{z} \in A^{-N}(w)$,
\[
P(a|\hat{z}) = \lambda_{-1}^w p_{-1}(U_0 \in I(a)|U_0 \in [0, \alpha_{-1}^w])
\]
+ \sum_{k \geq 0} \lambda_k^w P\left(U \in I(a) \cup \bigcup_{l=0}^k I^w(a, z, l) \bigg| U \in [\alpha_k^w, \alpha_{k-1}^w]\right).
\]

It follows from (19) that for any $\hat{z} \in A^{-N}$ and $a \in A$
\[
P(a|\hat{z}) = \lambda_{-1}^w p_{-1}(a|\tau_{-1}(\hat{z})) + \sum_{k \geq 0} \lambda_k^w p_k^w(a|\tau_k(\hat{z})).
\]
\]
Proof of Proposition 4.1. The fact that (13) implies (14) is clear since $A^k(w) \subset A^k$. Now let us show that (14) implies (10). First, define for any $k$ to be finite and observe that

$$
\alpha_{k,i}^w := \inf_{k^{-1} \in \mathcal{I}(w)} \inf_{c^{-1} \in A^k} \sum_{a \in A} \inf_{\hat{z}} P(\{ a | b^{-1}_{-i} w c^{-1}_{-k} \hat{z} \})
$$

and observe that

$$
\inf_{k^{-1} \in \mathcal{I}(w)} \inf_{c^{-1} \in A^k} \sum_{a \in A} \inf_{\hat{z}} P(\{ a | b^{-1}_{-i} w c^{-1}_{-k} \hat{z} \}) \leq \alpha_{k,i}^w.
$$

Thus, condition (10) implies at the same time that $\alpha_{k,i}^w \Rightarrow 1$ for any fixed $i$ and $\alpha_{k,i}^w \Rightarrow 1$ for any fixed $k$. Since $\alpha_{k,i}^w$ belongs to $[0,1]$ for any $k$ and $i$, it follows that

$$
\inf_{i \geq 0} \alpha_{k,i}^w (:= \alpha^E) \text{ also goes to 1 as } k \text{ diverges.}
$$

\[ \square \]

5. Second main result: perfect simulation

In this section, we present the perfect simulation algorithm and state the second main result of this paper, which gives sufficient conditions for the algorithm to stop after a $\mathbb{P}$-a.s. finite number of steps.

5.1. Explaining how our algorithm works. The algorithm works as a mixture of the algorithm presented in [Gallos (2009)] and the one of [Comets et al. (2002)].

Assume that the set

$$
\mathcal{E} := \{ a \in A : \inf_{\hat{z}} P(a | \hat{z}) > 0 \}
$$

is not empty, and let $\mathcal{E}^*$ denotes the set of finite strings of symbols of $\mathcal{E}$.

Consider a transition probability kernel $P$ satisfying the condition of Theorem 4.1 with reference string $w \in \mathcal{E}^*$. In order to simplify the notation, we will omit the superscript $w$ in most of the quantities that depend on this string.

We want to get a deterministic measurable function $X : [0,1]^2 \to \mathbb{R}$, $\mathbf{U} \mapsto X(\mathbf{U})$ such that the law $\mathbb{P}(X(U) \in \cdot)$ is compatible with $P$ in the sense of (2). The idea is to use the sequence $\mathbf{U}$ together with the partitions of $[0,1]$ introduced before (and illustrated in Figure 3.2.2) to mimic the two steps procedure we described in Section 3.2.2.

In particular, for any $n \in \mathbb{Z}$, we put $[X(\mathbf{U})]_n = a$ whenever $U_n \in I(a)$. Suppose that for some time index $n \in \mathbb{Z}$ there exists a string $a^{-1}_{-k} \in A^k$ such that $U_{n-i} \in I(a_{-i})$, $i = 1, \ldots, k$, in this case, we put

$$
[X(\mathbf{U})]_{n-k} = a^{-1}_{-k}.
$$

We say that this sample has been spontaneously constructed. Now suppose $U_n \in [\alpha_{l-1}, \alpha_l]$ for some $l \geq 0$. This means that we pick up the context tree $\tau_l$ in the countable mixture representation of $P$, and look whether or not there exists a context in $\tau_l$ which is suffix of $[X(\mathbf{U})]_{n-k}^{-1} = a^{-1}_{-k}$. If there exists such a context, then we put

$$
[X(\mathbf{U})]_n = \sum_{a \in A} a \mathbf{1} \left\{ U_n \in \bigcup_{j=0}^l I(a, a_{-k-j}) \right\}.
$$

If there is no such context (we will write $c_{\tau_l}(a^{-1}_{-k}) = \emptyset$) we cannot construct the state $[X(\mathbf{U})]_n$: we need more knowledge of the past. In the first case, $[X(\mathbf{U})]_{n-k}$ has been constructed independently of $U^{-k}_{-n-1}$ and $U^{+\infty}_{n+1}$. Now suppose we want to construct $[X(\mathbf{U})]_0$. We generate backward in time the $U_i$’s until the first time $k \leq 0$ such that we can perform the above construction from time $k$ up to time 0 using only $U^k_0$. A priori, there is no reason for $k$ to be finite. Theorem 5.1 gives sufficient conditions for $k$ to be finite $\mathbb{P}$-almost surely.
To formalize what we just said, let us define for any \( u \in [0, 1] \)
\[
\ell(u) = \sum_{k \geq 1} k.1\{u \in [\alpha_{k-1}, \alpha_k]\}.
\]

By Theorem 4.1, \( \ell(U_i) = -1 \) means that we can choose the state of \( X(U)_i \) according to distribution \( p_{-1}(\cdot) \), and independently of everything else. On the other hand, \( \ell(U_i) = l \geq 0 \) means that we have to use the context tree \((\tau_i, p_i)\) in order to construct the state of \( X(U)_i \). In particular, we recall that for any \( l \geq 0 \) the size of the context \( c_{\tau_i}(a^n_m) \) is \( m^w(a^n_m) + |w| + l \).

One of the inputs for Algorithm 1 is the update function \( F \). It is a measurable function \( F : [0, 1] \times (\emptyset \cup A^* \cup A^{-N}) \rightarrow A \cup \{\star\} \) which uses the part of the past we already know and the uniform random variable to compute the present state. It is defined as follows:

\[
F(u, a^n_m) := \begin{cases} 
\sum_{a \in A} a.1\{u \in I(a)\} & \text{if } \ell(u) = -1 \\
\sum_{a \in A} a.1\{u \in \bigcup_{k \geq 0} I(a, a^n_m, k)\} & \text{if } \ell(u) \geq 0 \text{ and } c_{\tau_i(u)}(a^n_m) \neq \emptyset \\
\star & \text{otherwise}
\end{cases}
\]

with the conventions that \( a^n_{m+1} = \emptyset \) and for any context tree \( \tau \), \( c_\tau(\emptyset) = \emptyset \). When we consider an infinite past \( z \in A^{-N}(w) \), we have by (20), for any \( u \in [0, 1] \)
\[
P(F(u, z) = a) = P \left( u \in I(a) \cup \bigcup_{k \geq 0} I^w(a, z, k) \right) = P(a|z). \tag{22}
\]

When the update function returns the symbol \( \star \), it means that we do not have sufficient knowledge of the past to compute the present state.

We define, for any \( m \leq n \), the \( F(U^m_n) \)-measurable function \( \mathcal{L} : [0, 1]^{[n-m+1]} \rightarrow \{0, 1\} \) which takes value 1 if, and only if, we can construct \( [X(U)]^n_m \) independently of \( U^{-\infty}_m \) and \( U^+_{n+1} \) using the construction described above. Formally
\[
\{\mathcal{L}(U^m_n) = 1\} := \bigcup_{a^n_m \in A^{n-m+1}} \bigcap_{i=m}^n \{F(U_i, a^{i-1}_m) = a_i\}.
\]

Finally, for any \( -\infty < m \leq n \leq +\infty \), we define the regeneration time for the window \([m, n]\) as the first time before \( m \) such that the construction described above is successful until time \( n \), that is
\[
\theta[m, n] := \max\{k : \text{m} \leq m : \mathcal{L}(U^m_k) = 1\} \tag{23}
\]
with the convention that \( \theta[m] := \theta[m, m] \).

5.2. **The algorithm.** This algorithm takes as “input” two integers \( -\infty < m \leq n < +\infty \) and the update function \( F \), and returns as “output” the regeneration time \( \theta[m, n] \) and the constructed sample \( [X(U)]^n_m \). The function \( F \) contains all the information we need about \( P \), and we suppose that it is already implemented in the software used for programming the algorithm.

At each time, the set \( B \) contains the sites that remains to be constructed. At first \( B = \{m, \ldots, n\} \) and a forward procedure (lines 2–8) tries to construct \( [X(U)]^n_m \) using \( U_m, \ldots, U_n \). If it succeeds, then the algorithm stops and returns \( \theta[m, n] = m \) and the constructed sample. If it fails, \( B \) is not empty and a backward procedure (“while loop” lines 10–27) begins. In this loop, each time the algorithm cannot construct the next site of \( B \), it generates a new uniform random variable backward in time. At each new generated random variable, the algorithm attempts to go as far as possible in the construction of the remaining sites of \( B \) using the uniform that have been previously
Algorithm 1 Perfect simulation algorithm of the sample \([X(U)]_m^n\)

1: Input: \(m, n, F\); Output: \(\theta[m, n]\), \([X(U)]_{\theta[m,n]}, \ldots, [X(U)]_n\)
2: Sample \(U_m, \ldots, U_n\) uniformly in \([0, 1]\)
3: \(i \leftarrow m, B = \{m, \ldots, n\}, \theta[m, n] \leftarrow m, \ [X(U)]_{m}^{n} \leftarrow \ast^{n-m+1}\)
4: while \(F(U_i, [X(U)]_{i}^{i-1}) \in \mathcal{A}\) and \(B \neq \emptyset\) do
5: \([X(U)]_{i} \leftarrow F(U_i, [X(U)]_{i}^{i-1})\)
6: \(B \leftarrow B \setminus \{i\}\)
7: \(i \leftarrow i + 1\)
8: end while
9: \(i \leftarrow m\)
10: while \(B \neq \emptyset\) do
11: \(i \leftarrow i - 1\)
12: \(B \leftarrow B \cup \{i\}\)
13: Sample \(U_i\) uniformly in \([0, 1]\)
14: while \(U_i \in \#E, 1]\) do
15: \(i \leftarrow i - 1\)
16: \(B \leftarrow B \cup \{i\}\)
17: Sample \(U_i\) uniformly in \([0, 1]\)
18: end while
19: \([X(U)]_{i} \leftarrow F(U_i, \emptyset)\)
20: \(B \leftarrow B \setminus \{i\}\)
21: \(t \leftarrow \min B\)
22: while \(F(U_t, [X(U)]_{t}^{t-1}) \in \mathcal{A}\) and \(B \neq \emptyset\) do
23: \([X(U)]_{t} \leftarrow F(U_t, [X(U)]_{t}^{t-1})\)
24: \(B \leftarrow B \setminus \{t\}\)
25: \(t \leftarrow \min B\)
26: end while
27: end while
28: \(\theta[m, n] \leftarrow i\)
29: return \(\theta[m, n], ([X(U)]_{\theta[m,n]}, \ldots, [X(U)]_n)\)

generated. Theorem 5.1 gives sufficient conditions for this procedure to stop after a finite number of steps.

5.3. Statement of the second main theorem.

Theorem 5.1. Consider a kernel \(P\) satisfying the conditions of Theorem 4.1 for some string \(w \in \mathcal{E}^\ast\). If the sequence \((\alpha_{k}^w)_{k \geq 0}\) defined by (10) satisfies

\[
\sum_{k \geq 0} (1 - \alpha_{k}^w) < +\infty \quad \text{or, equivalently} \quad \prod_{k \geq 0} \alpha_{k}^w > 0
\]

then Algorithm 1 stops after a \(P\)-a.s. finite number of steps for any \(-\infty < m \leq n \leq +\infty\). Moreover, for any \(n \in \mathbb{Z}\)

\[
\sum_{l \geq 0} P(\theta[0] < -l) < +\infty.
\]

Corollary 5.1. The output of Algorithm 1 is a sample of the unique stationary chain compatible with \(P\). Moreover, there exists a sequence of random times \(T = T(U)\) which splits the realization \(X\) into i.i.d. pieces. More specifically, the random strings \([X(U)]_{T_i}, \ldots, [X(U)]_{T_{i+1} - 1}, i \neq 0\) are i.i.d. and have finite expected size.
The proof of Corollary 5.1 using the CFTP algorithm and Theorem 5.1 is essentially the same as [Comets et al. (2002)] (Proposition 6.1, Corollary 4.1 and Corollary 4.3). We omit these proofs in the present work and just mention the main ideas. The existence statement follows once we observe that Theorem 5.1 implies that one can construct a bi-infinite sequence $X$ verifying for any $n \in \mathbb{Z}$, $X_n = F(U_n, X_{n-\infty}^n)$. By (22), this chain is therefore compatible in the sense of (2). It is stationary by construction. The uniqueness statement follows from the loss of memory the chain inherits because of the existence of almost surely finite regeneration times. The regeneration scheme follows from (23).

6. Proof of Theorem 5.1

Let us explain what are the main steps of this proof. To study directly the random variable $\theta[m, n]$ is complicated, because it depends on the construction of the states of $X(U)$: in order to construct the next state of the chain, we may need to know the distance to the last occurrence of $w$ in the constructed sample. The idea is to introduce, first, a new random variable $\bar{\theta}[m, n]$, defined by (29), which can be used to define a lower bound for $\theta[m, n]$. The advantage of $\bar{\theta}[m, n]$ is that its definition depends on the reconstructed sample only through the spontaneous occurrences of $w$. Section 6.1 is dedicated to the definition of this new random variable. After that, the main problem is transformed into the problem of showing that $\bar{\theta}[\theta[0, n], \theta[0, n]]$ through equation (31). The conclusion of the proof is done in Section 6.3, by studying the chain $D^{(0)}$ (Lemma 6.1).

6.1. Definition of a new random variable $\bar{\theta}[m, n]$. Define the i.i.d. stochastic chain $Z$ which takes value $Z_i = a$ if $U_i$ belongs to $I(a)$, and $Z_i = \ast$ otherwise. This chain takes in account only the symbols which appear spontaneously in $X(U)$: $[X(U)]_i = a$ whenever $Z_i = a$, and in particular $[X(U)]_i^{-|w|+1} + 1 = w$ whenever $Z_i^{-|w|+1} = w$, for any $i \in \mathbb{Z}$. We also define the distance to the last spontaneous occurrence of $w$ in $Z$ before time $i$ as

$$m_i = \inf\left\{k \geq 0 : Z_i^{-k-1} = w\right\}.$$  


Suppose we already constructed a sample $[X(U)]_{-k}^{-1}$. Since for any $n$, $[X(U)]_{n-|w|+1}^n = w$ whenever $Z_{n-|w|+1}^n = w$, it follows that $m_0$ is larger or equal than $m^w([X(U)]_{-k}^{-1})$.

Denote $\ell_i := \ell(U_i)$ and define the random variable

$$L_i = \begin{cases} 0 & \text{if } Z_i \in \mathcal{E}, \\ m_i + |w| + \ell_i & \text{otherwise.} \end{cases}$$  

Then, whenever $L_0 > 0$, it is larger or equal than $m^w([X(U)]_{-k}^{-1}) + |w| + \ell_0$ which is the size of the suffix of $[X(U)]_{-k}$ we need to know in order to construct $X(U)_0$. Before we define $\bar{\theta}[m, n]$, let us introduce an intermediary random variable $\theta'[m, n]$ which depends on the spontaneous occurrences of $w$. For any $-\infty \leq \ell \leq n \leq +\infty$

$$\theta'[m, n] = \max\{k \leq m : L_i \leq \ell - k, \ i = k, \ldots, n\}. $$  

Associate to each site $i \in \{\theta'[m, n], \ldots, n\}$ an arrow going from time $i$ to time $i - L_i$. Definition (26) says that no arrow will pass time $\theta'[m, n]$, meaning that we can construct $[X(U)]_{\theta'[m, n]}^{[\theta'[m, n]}$ knowing only $U^n_{\theta'[m, n]}$. Therefore, $L(U^n_{\theta'[m, n]}) = 1$. Since $\theta[m, n]$ is the maximum over all time indexes $k \leq m$ such that $L(U^n_k) = 1$, it follows that

$$\theta'[m, n] \leq \theta[m, n].$$
The definition of $\hat{\theta}[m, n]$ is done using the following rescaled quantities. Consider the chain $\bar{Z}$ defined by

$$\bar{Z}_m = \begin{cases} 1 & \text{if } U_{m+i+1} \in I(w) \text{, } i = 0, \ldots, |w|-1 \\ * & \text{otherwise.} \end{cases}$$

and the rescaled function

$$\ell_i := \left[ \sup \{\ell_j : j = (i-1)|w| + 1, \ldots, i|w|\} \right] / |w|$$

where for any $r \in \mathbb{R}$, $[r]$ denotes the smaller integer which is larger or equal to $r$. Using these rescaled quantities, we define the corresponding random variables

$$\bar{m}_i = \inf \{k \geq 0 : \bar{Z}_{i-k-1} = 1\}$$

which is the distance to the last occurrence of 1 in $\bar{Z}^{i-1}_{-\infty}$ and

$$\bar{L}_i = \begin{cases} 0 & \text{if } \bar{Z}_1 = 1, \\ \bar{m}_i + 1 + \ell_i & \text{otherwise.} \end{cases}$$

The utility of all these new definitions lays in the fact (which is proven in details in [Gallo (2009)], the only difference being the definition of the function $\ell_i$) that the rescaled random variable

$$\bar{\theta}[0, n] := \max\{k \leq 0 : \bar{L}_i \leq i - k, \ i = k, \ldots, n\}$$

satisfies the inequality

$$(\bar{\theta}[0, n] - 1)|w| + 1 \leq \theta'[0, n|w|] \leq \theta[0, n|w|]$$

for any $n \geq 0$. All we need to study now is the distribution of $\bar{\theta}[0, n]$. This is done in Sections 6.2 and 6.3.

To clarify the relationship between $\bar{\theta}[0, n]$, $\theta'[0, n|w|]$ and $\theta[0, n|w|]$ let us give a concrete example.

**Example.** Consider a kernel $P$ satisfying the conditions of Theorem 5.1 with a string $w$ having length $|w| = 3$. Assume we are given a sample $U^{0,38}_6$ (which we do not specify) to which correspond two samples $Z^{0,38}_6$ and $Z^{2,12}_6$, with two sequences of arrows $L^{0,38}_6$ and $L^{2,12}_6$. The sample $Z^{0,38}_6$ together with the sequence of arrows $L^{0,38}_6$ are illustrated in the lower part of Figure 6, and the sample $Z^{2,12}_6$ together with the sequence of arrows $L^{2,12}_6$ are illustrated in the middle part of Figure 6. The loops mean that $L_i = 0$ or $\bar{L}_i = 0$.

We have sufficient information to determine lower bounds for $\bar{\theta}[0, 6]$. In fact, we can see on the lower sequence that no arrow merging from $i \in \{-29, \ldots, 6\}$ go further time $-29$, and that $-29$ is the first time in the past satisfying this. Therefore $\theta'[0, 6] = -29$. This is a first lower bound for $\theta[0, 6]$. Another lower bound can be obtained looking at the sequence in the middle of the figure, no arrow goes further time $-12$, meaning that $\bar{\theta}[0, 2] = -12$. Then, as we said, $\bar{\theta}[0, 2]$ satisfies inequality $-29$, this allows us to use the lower bound $(\bar{\theta}[0, 2] - 1)|w| + 1 = -38$ for $\bar{\theta}[0, 6]$.

**6.2. A new auxiliary chain for the study of $\bar{\theta}[0, n]$.** For any $n \in \mathbb{Z}$, the chain $D^{(n)}$ takes values $D^{(n)}_i = 0$ for any $i \leq n$ and

$$D^{(n)}_i = (i - i^{(n)} - \bar{L}_i) \vee 0, \ \forall i \geq n + 1,$$

where $i^{(n)} := \max\{l < i : D^{(n)}_l = 0\}$. The behavior of this chain is explained in the upper part of Figure 6. But it is clear from its definition that if $D^{(n)}_i > 0$ for $n = i + 1, \ldots, k$ for some $k \geq i + 1$, then, in the process $L$, no arrow merging from
Figure 6. Illustration of the inequalities of (29) (samples of $Z$, $L$, $\bar{Z}$ and $\bar{L}$) and of the behavior of the chain $D^{(i)}$, for $i = -13, -11, -8$ constructed using the samples of $Z$ and $L$.

$\{i + 1, \ldots, k\}$ passes time $i + 1$, meaning that $\bar{\theta}[i + 1, k] = i + 1$. More generally, the sequence of chains $\{D^{(i)}\}_{n \in \mathbb{Z}}$ satisfies the equation

$$\{\bar{\theta}[0, n] < -l\} = \bigcap_{i=-l-1}^{-1} \bigcup_{k=i+1}^{n} \{D^{(i)}_k = 0\}.$$ 

In fact, we can show in a similar way as in Section 6 of Gallo (2009), the only difference being the definition of the function $\bar{\ell}_i$, that

$$\mathbb{P}(\theta[0, n] < -l) \leq \sum_{k = \left\lfloor \frac{n}{l} \right\rfloor}^{\left\lfloor \frac{n}{l} \right\rfloor + \left\lceil \frac{n}{l} \right\rceil} \mathbb{P}(D^{(0)}_k = 0)$$

(31)

where $\lfloor r \rfloor$ denotes the integer part of $r$. This inequality relates the distribution we are interested in with the probability of return to 0 of the chain $D^{(0)}$. 
6.3. **Finishing the proof of Theorem 5.1.** Owing to inequality (31), the proof of Theorem 5.1 is done if we prove the following lemma.

**Lemma 6.1.** *Under the conditions of Theorem 5.1*

\[
\sum_{i \geq 1} \mathbb{P}(D_i^{(0)} = 0) < +\infty.
\]

**Proof.** For the clarity of the presentation, let us consider the chain \( E^{(0)} \) which is defined using \( D^{(0)} \) as follows.

\[
E^{(0)}_i = \begin{cases} 
D^{(0)}_i & \text{whenever } D^{(0)}_i = i - i^{(0)} \text{ or } 0 \\
E^{(0)}_{i-1} & \text{otherwise.}
\end{cases}
\]  

(32)

The behavior of \( E^{(0)} \) is easier to understand than \( D^{(0)} \) and their relationship is illustrated in Figure 7 for given samples \( \bar{Z} \) and \( \bar{L} \). At time \( j > 0 \), supposing that \( E^{(0)}_{j-1} = n \geq 0 \),

\[
E^{(0)}_j = \begin{cases} 
 j - j^{(0)} & \text{if } U_j \mid_{|w|-i+1} \in I(w-1) \text{ for any } i = 0, \ldots, |w| - 1, \text{ (i.e. } \bar{Z}_j = 1) \\
n & \text{if } \alpha_{n-1} \leq U_j \mid_{|w|-i+1} < \alpha_{n-1} \text{ for any } i = 0, \ldots, |w| - 1 \\
0 & \text{if } U_j \mid_{|w|-i+1} \geq \alpha_{n-1} \text{ for some } i = 0, \ldots, |w| - 1.
\end{cases}
\]  

(33)

It is clear that \( \mathbb{P}(D^{(0)}_i = 0) = \mathbb{P}(E^{(0)}_i = 0) \) and that the state 0 is renewal for \( E^{(0)} \). It follows from Feller [1968, Chapter XIII.10, Theorem 1] that \( \mathbb{P}(E^{(0)}_k = 0) \) is summable in \( k \) if and only if the state 0 is transient. Denote by \( \zeta \) the first time after time 0 that the chain \( E^{(0)} \) returns to the state 0, and for \( k \geq 1 \) we put \( f_k = \mathbb{P}(\zeta = k) \).

![Figure 7](image)

**Figure 7.** Figure illustrating the behavior of the chains \( E^{(0)} \) and \( D^{(0)} \) together, both using the samples of \( \bar{Z} \) and \( \bar{L} \).
We want to show that under the assumption of Theorem 5.1, the state 0 is transient for $E^{(0)}$. Let us denote by $G_{i,M}$ the event $\{\zeta = i\} \cap \{Z_1^M = 1\}$ for $l \geq 1$ and $i \geq l + 1$.

If $Z_i^M = 1^M$ for some $M \geq 1$, then $E^{(0)}_{i,M} = M$, $G_{i,M} = \emptyset$ for $i \leq M$ and for any $i \geq M + 1$

$$G_{i,M} = \{Z_1^M = 1^M\} \cap \{E_j^{(0)} \geq M : j = M+1, \ldots, i-2\} \cap \bigcup_{k=M}^{i-1} \{E_{i-1}^{(0)} = k\} \cap \{E_i^{(0)} = 0\}.$$  

The definition of $E^{(0)}$ implies that whenever $\{\zeta = i\}$ for $i \geq 1,$

$$\{E_{i-1}^{(0)} = k\} = \{E_j^{(0)} = k \text{ for } j = k, \ldots, i-1\}$$

for $1 \leq k \leq i - 1$. It follows that

$$G_{i,M} = \{Z_1^M = 1^M\} \cap \bigcup_{k=M}^{i-1} \{E_j^{(0)} \geq M : j = M+1, \ldots, i-1\} \cap \{E_i^{(0)} = 0\}.$$  

Using (33) and the chain $U$, one obtains that for $m \geq 1$ and $i \geq M + 1$

$$G_{i,M} = \bigcup_{k=M}^{i-1} \left( \{Z_1^M = 1^M\} \cap \{E_j^{(0)} \geq M : j = M+1, \ldots, k-1\} \right) \cap \{E_k^{(0)} = k \text{ for } j = k, \ldots, i-1\} \cap \{E_i^{(0)} = 0\}.$$  

(34)

where the event $B$ is $\mathcal{F}(U_{(k-1)}^{(i-1)|w|})$-measurable, the event $C$ is $\mathcal{F}(U_{(k-1)}^{(i-1)|w|+1})$-measurable and the event $D$ is $\mathcal{F}(U_{(k-1)}^{(i-1)|w|+1})$-measurable. Therefore, they are independents. Recall that $w \in \mathcal{E}^*$ and assume that

$$\inf_{i=1,\ldots,|w|} \inf_{z} P(w_{-i}|z) = \epsilon > 0.$$  

Using the partition

$$\bigcup_{i \geq 1} \{\zeta = i\} = \left( \bigcup_{i \geq 1} \{\zeta = i\} \cap \{Z_1^M = 1^M\} \right) \cup \left( \bigcup_{i \geq 1} \{\zeta = i\} \cap \{Z_1^M \neq 1^M\} \right), \quad \forall M \geq 1$$

one obtains the following upper bound (recall that $G_{i,M} = \emptyset$ for $i \leq M$):

$$\sum_{i \geq 1} f_i \leq \sum_{i \geq M+1} P(G_{i,M}) + (1 - \epsilon|w|M),$$

which holds for any $M \geq 1$. Using the fact that $\alpha_k \leq 1$ for any $k$, we have

$$P(B) \leq \epsilon|w|M,$$

$$P(C) = (\alpha|w|(k-1) - \alpha_{i-1})|w|(i-k-1) \leq (1 - \alpha_{i-1})|w|(i-k-1)$$

$$P(D) \leq |w|(1 - \alpha|w|(k-1)).$$

Therefore, equality (34) gives us the following upper bound for any $i \geq M + 1$

$$P(G_{i,M}) \leq |w|\epsilon|w|M \sum_{k=M}^{i-1} (1 - \alpha_{i-1})|w|(i-k-1)(1 - \alpha_{(k-1)|w|}).$$
We have
\[
\sum_{i \geq M+1} \sum_{k=M}^{i-1} (1-\alpha_{-1})|w|(i-k-1)(1-\alpha_{(k-1)|w|}) = \sum_{k \geq M} (1-\alpha_{(k-1)|w|}) \sum_{i \geq k+1} (1-\alpha_{-1})(i-k-1)|w|
\]
where we interchanged the order of the sums, this last equation yields
\[
\sum_{i \geq 1} f_i \leq \frac{\sum_{i \geq 1} \sum_{k \geq M} (1-\alpha_{(k-1)|w|}) + (1-\epsilon|w|_M}). \tag{35}
\]
Under the conditions of Theorem 5.1, \(\sum_{k \geq M} (1-\alpha_{(k-1)|w|})\) goes to 0 as \(M\) increases. This means that the right hand side of (35) is strictly smaller than 1 for some sufficiently large \(M\), and it follows that \(\sum_{i \geq 1} f_i < 1\). This finishes the proof of Lemma 6.1. \(\square\)

7. Conclusion

Comets et al. (2002) use the uniform continuity assumption \(\alpha_k^{CFF} \to 1\). Perfect simulation under a weaker condition was done recently by De Santis & Piccioni (2010) requiring only punctual continuity, ie, \(\alpha_k^{CFF}(\bar{w}) \to 1\) for any \(\bar{w}\) in the set of “admissible histories” (see Section 2 therein). Our extension allows to consider kernels \(P\) having discontinuities along all the points \(\bar{a} \in A^{-\infty}(\bar{w})\), for any \(w \in \mathcal{E}^*\), and a priori, no assumption is made on the set of “admissible histories”, so that it is generically the set \(A^{-\infty}\). More specifically, consider a transition probability kernel \(P\) such that \(\alpha_k^{CFF}\) satisfies \(\sum_{k \geq 0} (1-\alpha_k^{CFF}) < +\infty\). It follows that, for any \(w \in \mathcal{E}^*\),
\[
\sum_{k \geq 0} (1-\alpha_k^w) < +\infty.
\]

Now consider any \(\bar{P}\) satisfying that \(\bar{a}_k^w = \alpha_k^w\) and allowing discontinuities along branches not containing the string \(w\). Theorem 5.1 says that we still can make a perfect simulation of the unique stationary chain compatible with \(\bar{P}\). This shows that our result is a strict generalization of the work of Comets et al. (2002) whenever we are in the regime \(\sum_{k \geq 0} (1-\alpha_k^{CFF}) < +\infty\).

Also, our condition does not necessarily fit under the conditions of De Santis & Piccioni (2010). It is possible to see this checking Equation (35) of Example 1 in their work. Their notation corresponds to
\[
a_0(-1) = \alpha(-1), \quad a_0(1) = \alpha(1) \quad \text{and} \quad a_\infty = \lim_{k \to +\infty} \alpha_k^{CFF}.
\]

Taking \(w = (-1)(1)\) or \(w = (1)(-1)\), Theorem 5.1 says that we only need \(a_0(-1)\) and \(a_0(1)\) to be strictly positive without any assumption on \(a_\infty\). We also mention that this particular example was already handled by the results of Gallo (2009).

Let us finish with some questions. The condition \(\sum_{k \geq 0} (1-\alpha_k^w) < +\infty\) guarantees that the perfect simulation scheme stops at a finite time \(\theta\) which has finite expected value. Can we find weaker conditions such that \(\theta\) is finite a.s. but has infinite expectation? Is the minorization assumption on our reference string \((w \in \mathcal{E}^*)\) necessary to obtain a practical coupling from the past algorithm for our class of (non-necessarily continuous) chains of infinite memory?

References

Comets, F., Fernandez, R. & Ferrari, P. (2002). Processes with long memory: regenerative construction and perfect simulation. Ann. Appl. Prob. 12(3), 921–943.

De Santis, E. & Piccioni, M. (2010). A general framework for perfect simulation of long memory processes. arXiv:1004.0113v1.
Feller, W. (1968). *An introduction to probability theory and its applications. Vol. I*. Third edition. New York: John Wiley & Sons Inc.

Gallo, S. (2009). Chains with unbounded variable length memory: perfect simulation and visible regeneration scheme. *Submitted, arXiv: 0907.2150*.

Kalikow, S. (1990). Random Markov processes and uniform martingales. *Israel J. Math.* 71(1), 33–54. URL [http://dx.doi.org/10.1007/BF02807249](http://dx.doi.org/10.1007/BF02807249).

Propp, J. G. & Wilson, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. In: *Proceedings of the Seventh International Conference on Random Structures and Algorithms (Atlanta, GA, 1995)*, vol. 9.

Rissanen, J. (1983). A universal data compression system. *IEEE Trans. Inform. Theory* 29(5), 656–664.

Sandro Gallo, Nancy L. Garcia
Instituto de Matemática, Estatística e Computação Científica
Universidade Estadual de Campinas
Rua Sergio Buarque de Holanda, 651
13083-859 Campinas, Brasil
e-mail: nancy@ime.unicamp.br, gsandro@ime.unicamp.br