

CHAINS WITH UNBOUNDED VARIABLE LENGTH MEMORY: PERFECT SIMULATION AND A VISIBLE REGENERATION SCHEME

SANDRO GALLO,* *University of Campinas*

Abstract

We present a new perfect simulation algorithm for stationary chains having unbounded variable length memory. This is the class of infinite memory chains for which the family of transition probabilities is represented by a *probabilistic context tree*. We do not assume any continuity condition: our condition is expressed in terms of the structure of the context tree. More precisely, the length of the contexts is a deterministic function of the distance to the last occurrence of some determined string of symbols. It turns out that the resulting class of chains can be seen as a natural extension of the class of chains having a renewal string. In particular, our chains exhibit a visible regeneration scheme.

Keywords: Variable length memory chain; probabilistic context tree; perfect simulation; regeneration scheme

2010 Mathematics Subject Classification: Primary 60G10
 Secondary 60G99

1. Introduction

We introduce a new class of discrete time stochastic chains $X = (X_n)_{n \in \mathbb{Z}}$, taking values in a countable alphabet A . These chains have unbounded variable length memory. This means that the state of the chain at time 0 depends on an unbounded suffix of the past $\cdots X_{-2}X_{-1}$, whose length depends on the values assumed by the chain in the past. In the present case, the length of this suffix depends on the distance to the last occurrence of a given finite reference string $a_{-i} \cdots a_{-1}$ of symbols of A . More precisely, there exists a function $f: \mathbb{N} \rightarrow \mathbb{N}$ such that if the last occurrence of $a_{-i} \cdots a_{-1}$ is at distance k in the past, that is, if $X_{-k-i} \cdots X_{-k-1} = a_{-i} \cdots a_{-1}$, and, for $j = i, \dots, k + i - 1$, we have $X_{-j} \cdots X_{-j+i-1} \neq a_{-i} \cdots a_{-1}$, then we need to know $X_{-f(k)-i-k} \cdots X_{-k-i-1} X_{-k-i} \cdots X_{-1}$ in order to decide the state of the chain at time 0:

$$\cdots \overbrace{X_{-f(k)-i-k} \cdots X_{-k-i-1} X_{-k-i} \cdots X_{-k-1}}^{\text{suffix of the past we need to know to decide } X_0} X_{-k} \cdots X_{-1}.$$

length = $f(k)$
last occurrence of a_{-i}^{-1}

In other words, the family of transition probabilities P for these chains is such that

$$P(\cdot \mid \cdots b_{-2}b_{-1}) = P(\cdot \mid \cdots c_{-2}c_{-1})$$

Received 22 July 2010; revision received 3 April 2011.

* Postal address: University of Campinas, Rua Sérgio Buarque de Holanda 651, CEP 13083-859 Campinas, São Paulo, Brazil. Email address: gsandro@ime.unicamp.br

This work forms part of the USP Project ‘Mathematics, Computation, Language and the Brain’. Research supported by an FAPESP fellowship (grant number 2006/57387-0).

whenever the last occurrence of $a_{-i} \cdots a_{-1}$ is at distance k in $\cdots b_{-2}b_{-1}$ and

$$b_{-f(k)-1-k} \cdots b_{-2}b_{-1} = c_{-f(k)-1-k} \cdots c_{-2}c_{-1}.$$

Observe that, on $A = \{1, 2\}$, if the reference string is the symbol 2, and the function f is identically 0, we obtain the renewal chain with symbol 2 as the renewal symbol. For this reason, we say that this class of stochastic chains generalizes the class of chains having a renewal string.

We highlight three main parameters for the study of this class of chains: the size of the reference string, the set of transition probabilities to the symbols of this reference string, and the deterministic function f .

We ask the following questions. (i) What assumptions should we place on these parameters in order to guarantee that there exists a stationary chain compatible with such a family of transition probabilities? (ii) Is this stationary chain unique? (iii) What are the statistical properties of this chain? (iv) Does the chain exhibit a regeneration scheme, as in the renewal case?

In order to avoid confusion, we insist on the fact that the denomination ‘variable length’ does not refer to the case of ‘random length’ or ‘random Markov’ as considered in [7] or [17]. In the present case, the length of the past we need to know in order to decide the next symbol depends on the past itself, whereas in the ‘random Markov’ models, the length is decided randomly and independently of the past.

It is important to observe that the results in the literature on chains of infinite order cannot answer these questions which, in our view, are quite natural. The main reason for this is the fact that, since the seminal paper [19], the literature has focused on the so-called continuity assumption, which is not assumed here. In fact, the way we describe the family of transition probabilities of our chains fits exactly into the notion of *probabilistic context trees*, introduced in [21]. It follows that the best framework for our study is probabilistic context trees and not a continuous family. This also implies that the method of ‘random Markov’ representation, used implicitly in [7], cannot be used here, since it was proved in [17] that such a representation is possible if and only if the family is continuous. Moreover, there is, so far, no ‘well-adapted’ (in a sense that we will make clear later) criteria for the existence and uniqueness of the stationary chain compatible with a given probabilistic context tree.

Consequently, the main method we use in order to answer the above questions is the constructive method, that is, we give sufficient conditions on our parameters which ensure that we can perfectly simulate the chain from the stationary distribution. This is our first main result (Algorithms 1 and 2 and Theorem 1). To the author’s knowledge, to date, the only perfect simulation algorithm for chains of infinite order was presented in [7], which applies in the continuous framework. The theory of renovating events, introduced by Borovkov (see, for example, [2]) for stochastic recursive sequences and extended in [11] to generic stochastic chains, gives an interesting framework, closely related to the present work. It is based on the existence of an event which, when it occurs with positive probability, can be used to perfectly simulate the chain. However, the definition of this event, its positiveness, and the way we can detect it, are issues we have to address for each particular class of processes. Also, instead of using the results of this theory, and in order to be self-contained, we explain precisely how our perfect simulation algorithm works. This algorithm shares several features with the algorithm of [7], mainly due to the fact that both algorithms use the *coupling-from-the-past method* (CFTP method) introduced in [20] to perfectly simulate Markov chains.

As a byproduct of Theorem 1, we have sufficient conditions for the existence and the uniqueness of the stationary chain (Corollary 1). We also show that this stationary chain has a *hidden* regeneration scheme, and that the expected size between two consecutive regeneration

times is finite (Corollary 2). The denomination *hidden* means that we cannot detect it on the realization of the chain. This regeneration scheme arises from the perfect simulation algorithm, and is similar to the scheme introduced in [7].

The last main result of this paper is the existence, under the same conditions, of a *visible* regeneration scheme (Theorem 2). This regeneration scheme can be detected directly on the realization of the chain. But since our chains are not necessarily renewal, detecting the regeneration scheme means, in general, knowing the entire future of the chain.

We would like to emphasize that the continuity assumption was originally introduced by Doebelin and Fortet [8] as a technical assumption (see the discussion therein). As they state, it is quite natural in this optic to assume that the probability transition from $a_{-\infty}^{-1}$ to a does not depend too much on the remote symbols of $a_{-\infty}^{-1}$. However, continuity is *one* way to mathematically interpret this assumption. The present paper gives a different interpretation. The success of the continuity assumption appeared three decades later (in the 1960s), in part because it is related to some well-behaved dynamical systems and statistical mechanic models, through the Gibbs formalism (see, for example, [3, Chapter 1]). However, from an application point of view, it is not clear that the real phenomena have to be represented through the continuous framework. It seems to us quite natural (and also of mathematical interest) to explore the noncontinuous world.

The aim of the present work is threefold. First, we extend the class of renewal chains to a class of stochastic chains having a visible regeneration scheme which is not a renewal scheme. Second, we give an appropriate condition on the form of the context tree to guarantee the possibility to make a perfect simulation of the unique stationary chain compatible. Finally, we make what seems to be the first attempt in the literature of chains of infinite memory to study the noncontinuous case.

This paper is organized as follows. In Section 2 we give the basic definitions and notation, introducing in particular the context tree framework. In Section 3 we give an example which motivates the above discussion and explains why we consider such a class of stochastic chains. In Section 4 we explain more precisely our assumptions using the context tree framework. In Section 5 we sketch the perfect simulation algorithm and state the results of this paper. Sections 6, 7, and 8 are dedicated to the proofs of the results. In Section 9 we present the complete perfect simulation algorithm, plus some simulations of the example of Section 3. We finish the paper with some references on the involved areas.

2. Basic definitions

Let A be a countable alphabet. Given two integers $m \leq n$, we denote by a_m^n the string $a_m \cdots a_n$ 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 in which v denotes a semi-infinite sequence, that is, $v = v_{-\infty}^{-1}$. If n is a positive integer and v a finite string of symbols in A , we denote by $v^n = vv \cdots v$ the concatenation of n times the string v . We define

$$A^{-\mathbb{N}} = A^{\{\dots, -2, -1\}} \quad \text{and} \quad A^{\star} = \bigcup_{j=0}^{+\infty} A^{\{-j, \dots, -1\}},$$

which respectively denote the set of all infinite strings of past symbols and the set of all finite strings of past symbols. The case in which $j = 0$ corresponds to the empty string \emptyset .

2.1. Probabilistic context tree

We say that a string s is a *suffix* or *prefix* of another string v if $|s| \leq |v|$ and $v_{-|s|}^{-1} = s$ or, respectively, $v_{-|v|+|s|}^{-1} = s$.

Definition 1. A subset τ of $A^* \cup A^{-\mathbb{N}}$ is a *tree* if no string $s \in \tau$ is a suffix of another string $v \in \tau$. This property is called the *suffix property*.

Definition 2. A tree τ is *complete* if any element $a_{-\infty}^{-1}$ of $A^{-\mathbb{N}}$ has a suffix belonging to τ . The suffix property implies that this suffix is unique. We call it the *context* of the sequence $a_{-\infty}^{-1}$ and it is denoted by $c_\tau(a_{-\infty}^{-1})$. A complete tree is called a *context tree*. When we have $\sup\{|v| : v \in \tau\} = +\infty$, we say that the tree τ is *unbounded*.

We also extend the notion of context for finite strings: for any $a_m^n \in A^*$, $m \leq n$, we set $c_\tau(a_m^n) = v$ if v is a suffix of a_m^n belonging to τ . If no context of τ is a suffix of a_m^n , we use the convention $c_\tau(a_m^n) = \emptyset$. In particular, $c_\tau(\emptyset) = \emptyset$.

Definition 3. A *probabilistic context tree on A* is an ordered pair (τ, p) such that

1. τ is a context tree;
2. $p = \{p(\cdot | v); v \in \tau\}$ is a family of transition probabilities over A .

Examples of probabilistic context trees are shown in Figure 1(a) (for the bounded case) and 1(b) (for the unbounded case).

We call the attention of the reader to the following notation: if $v = v_{-|v|}, \dots, v_{-1}$ is a context of a context tree τ , then v_{-i} , $i = 1, \dots, |v|$, denotes the path from the root to the leaf in the tree representation of τ . In the conditional probability $p(a | v)$, we will swap the order of the symbols of the context v to keep the overall temporal order, i.e.

$$p(a | v) = p(a | v_{-1} \cdots v_{-|v|})$$

is the probability of having the symbol a at time n , say, given that at time $n - 1$ we have v_{-1} , at time $n - 2$ we have v_{-2} , etc.

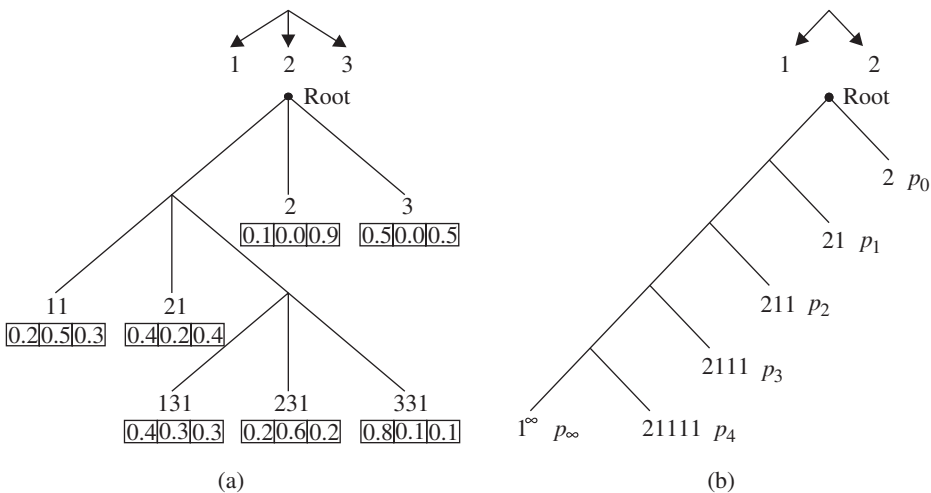


FIGURE 1: Examples of probabilistic context trees.

The context tree illustrated in Figure 1(a) is defined on $\{1, 2, 3\}$, and to each context v , we assign in the boxes the transition probabilities $p(1 | v)$, $p(2 | v)$, and $p(3 | v)$. In this paper we only consider unbounded context trees, but we mention this example to highlight that the model has been introduced in [21] in the bounded case. The context tree illustrated in Figure 1(b) is defined on $\{1, 2\}$; it corresponds to the discrete renewal chain, with renewal symbol 2. This means that the successive occurrences of 2 ‘split’ the realization of the chain into independent and identically distributed (i.i.d.) blocks. For any $i \geq 0$, p_i denotes the transition probability $p(2 | 1^i 2)$. The transition probability $p(1 | 1^i 2)$ is $1 - p_i$.

More general examples of unbounded context trees (without specifying the transition probabilities) are given in Figures 2, 3(a), and 3(b) in Sections 3.2 and 4.2.

Definition 4. We say that a probabilistic context tree (τ, p) is *weakly non-null* if

$$\sum_{a \in A} \inf_{v \in \tau} p(a | v) > 0.$$

We define $\alpha(a) := \inf_{v \in \tau} p(a | v)$ for any $a \in A$, and $\alpha := \sum_{a \in A} \alpha(a)$. For any finite size string $w := w_{-|w|}, \dots, w_{-1} \in A^*$, we also define $\alpha(w) := \prod_{i=1}^{|w|} \alpha(i)$. Finally, if there exists $\varepsilon > 0$ such that $\alpha(w) = \varepsilon$, we say that w is ε -regular.

2.2. Unbounded variable length memory chains

Definition 5. We say that a stationary stochastic chain $X = (X_n)_{n \in \mathbb{Z}}$ of law μ is compatible with a probabilistic context tree (τ, p) if, for μ -almost every pasts $a_{-\infty}^{-1} \in A^{-\mathbb{N}}$ and any $a \in A$, we have

$$\mu(X_0 = a | X_{-\infty}^{-1} = a_{-\infty}^{-1}) = p(a | c_\tau(a_{-\infty}^{-1})). \tag{1}$$

We call these chains *variable length memory chains*. If τ is unbounded, we call them *unbounded variable length memory chains*.

It remains to define what we call the ‘reference’ string in this framework. This will be done in Section 4, but first, let us give an example.

3. Discussion and examples

3.1. Motivation for the present work

Let us explain why the existing results of the literature cannot help us in our study. First, let us show that our chains need not be continuous. A family of transition probabilities P on a countable alphabet (equipped with the discrete topology) is continuous if

$$\beta_k := \sup\{|P(a | a_{-\infty}^{-1}) - P(a | b_{-\infty}^{-1})| : a \in A, a_{-\infty}^{-1}, b_{-\infty}^{-1} \in A^{-\mathbb{N}}, a_{-k}^{-1} = b_{-k}^{-1}\}$$

converges to 0 when k diverges. Thus, it is enough to consider the probabilistic context tree on $\{1, 2\}$ illustrated in Figure 1(b), with the probability transitions

$$p_i = \varepsilon_1 \mathbf{1}\{i \text{ is odd}\} + \varepsilon_2 \mathbf{1}\{i \text{ is even}\} \quad \text{and} \quad p_\infty := p(2 | 1^{+\infty}) = \varepsilon_3,$$

where $\varepsilon_1, \varepsilon_2$, and ε_3 are different real numbers in $(0, 1)$. Then it is straightforward to check that

$$\beta_k = \sup\{|\varepsilon_1 - \varepsilon_2|, |\varepsilon_2 - \varepsilon_3|, |\varepsilon_1 - \varepsilon_3|\}$$

for any $k \geq 0$. It follows from this simple observation that none of the chains in the class we consider have to be continuous.

The second motivation for the present work is that the perfect simulation algorithm given in [7] is not well adapted to probabilistic context trees. This follows from the fact that it does not use the information of the context tree. In order to illustrate this fact, let us consider the simplest possible case: (τ, p) is such that

- τ is the context tree illustrated in Figure 1(b);
- p satisfies the continuity condition of [7]; and
- symbol 2 is such that $\alpha(2) = \varepsilon > 0$.

In this case, there exists a very simple procedure to construct a sample X_0^n of the chain compatible with (τ, p) from the stationary measure. We use the fact that $\alpha(2) = \varepsilon > 0$ to couple the constructed chain with an i.i.d. sequence U of random variables uniformly distributed in $[0, 1)$ in such a way that we set $X_i = 2$ whenever $U_i < \varepsilon$. We generate backward in time, i.e. U_0, U_{-1}, \dots , and stop the procedure at the first time $-k$ such that $U_{-k} < \varepsilon$. We write $-k = \theta[0, n]$, a regeneration time for $[0, n]$, and we set $X_{-k} = 2$. Then, we construct the sample recursively from time $-k + 1$ up to n : for any $i \geq -k + 1$, we sample X_i from the distribution $(p(\cdot | X_{i-1} \cdots X_k) - \varepsilon)/(1 - \varepsilon)$. Observe that $p(\cdot | X_{i-1} \cdots X_k)$ is always well defined since $X_k = 2$. The constructed sample is stationary and compatible with (τ, p) . The way we defined $\theta[0, n]$ implies that $-\theta[0, n] + 1$ has a geometric distribution with parameter ε . This procedure is well known in the perfect simulation literature; such a chain is said to be uniformly minorized by ε (see, for example, [12]). The perfect simulation algorithm we present in this paper works for much more general chains, and is an extension of the procedure we just described.

Now, suppose that we perform the above algorithm and the algorithm of Comets *et al.* [7] at the same time, using the same sequence U , to perfectly simulate a window X_0^n . Let $\theta^{\text{CFF}}[0, n]$ be the regeneration time obtained using Comets *et al.*'s algorithm. The most objective way to compare both algorithms is to check which algorithm is faster. The random variable $\theta^{\text{CFF}}[0, n]$ is defined by

$$\theta^{\text{CFF}}[0, n] = \max\{k \leq 0 : U_i < a_{i-k}, i = k, \dots, n\},$$

where $(a_j)_{j \geq 0}$ is a $[0, 1]$ -valued sequence increasing to 1. The way it increases depends on the continuity assumption Comets *et al.* make. For a comparison with the above algorithm, we need to compare $a_0 := \inf_{a_{-\infty}^{-1}} p(1 | a_{-\infty}^{-1}) + \inf_{a_{-\infty}^{-1}} p(2 | a_{-\infty}^{-1})$ with ε . So, defining $\varepsilon := \inf_{a_{-\infty}^{-1}} p(2 | a_{-\infty}^{-1})$, we obtain $a_0 = \varepsilon + \inf_{a_{-\infty}^{-1}} p(1 | a_{-\infty}^{-1})$. The only way the algorithm of [7] could be faster than the above algorithm would be if its regeneration time occurs before the first time $U_{-k} < \varepsilon$ (which is the definition of $\theta[0, n]$). Denoting by Pr the law of U , we have to estimate

$$\text{Pr}(\theta[0, n] < \theta^{\text{CFF}}[0, n]).$$

It is difficult to find a good upper bound for this probability in general. We just mention two simple cases. In the case where $\inf_{a_{-\infty}^{-1}} p(1 | a_{-\infty}^{-1}) = 0$, it is clear that we have $\text{Pr}(\theta[0, n] < \theta^{\text{CFF}}[0, n]) = 0$. The other case we can study easily is when $p_i \searrow p_\infty = \varepsilon$. It is an exercise (see Section 7.2 of [10]) to show that in this case the context tree (τ, p) admits the following ‘random Markov’ representation:

$$p(a | c_\tau(a_{-\infty}^{-1})) = a_0 p_0(a) + \sum_{k \geq 1} (a_k - a_{k-1}) p^{[k]}(a | a_{-k}^{-1}) \quad \text{for all } a \in A \text{ and } a_{-\infty}^{-1} \in A^{-\mathbb{N}}.$$

Here $p_0(\cdot)$ is a probability distribution over A and, for any $k \geq 1$, the Markov kernel $p^{[k]}$ is defined by $p^{[k]}(2 | a_{-k}^{-1}) = 0$ if $a_{-k}^{-1} = 1^k$ and $p^{[k]}(2 | a_{-k}^{-1}) = 1$ otherwise. It follows that if

$\theta[0, n] < \theta^{\text{CFE}}[0, n]$ then the reconstructed sample is all 1: $X_{\theta[0, n]}^n = 1^{|\theta[0, n]|+n+1}$. Let us compute

$$\begin{aligned} \Pr(\theta[0, n] < \theta^{\text{CFE}}[0, n]) &= \sum_{i \geq 0} \Pr(\theta[0, n] = -i, \theta^{\text{CFE}}[0, n] > -i) \\ &= \sum_{i \geq 0} \sum_{j=0}^{i-1} \Pr(\theta[0, n] = -i, \theta^{\text{CFE}}[0, n] = -j). \end{aligned}$$

Since the event $\{\theta^{\text{CFE}}[0, n] = -j\}$ depends only on U_{-j}^n , it follows that the latter term is bounded above by

$$\sum_{i \geq 0} \sum_{j=0}^{i-1} \varepsilon(1 - \varepsilon)^{i-j+1} \Pr(X_{-j}^n = 1^{n+j+1}).$$

Using the fact that $\alpha(2) = \varepsilon > 0$, the probability $\Pr(X_{-j}^n = 1^{n+j+1})$ is bounded above by $(1 - \varepsilon)^{n+j+1}$. It follows that, for some constant $C > 0$,

$$\Pr(\theta[0, n] < \theta^{\text{CFE}}[0, n]) \leq C(1 - \varepsilon)^n,$$

which goes very fast to 0. These facts are a consequence of the following general remark. Since Comets *et al.*'s algorithm does not use the form of the context tree, it leads to regeneration times that have, a priori, nothing to do with the *natural regeneration times* one could expect: the successive occurrences of 2 along the realization of the chain. This leads to another misleading situation: the regeneration times of Comets *et al.*'s algorithm cannot be seen on the realization of the chain.

3.2. Example

As mentioned previously, the symbol 2 is renewal for the chain compatible with the context tree of Figure 1(b). Therefore, an example of an extension of this model is the following. If the last 2 occurred a distance i in the past (that is, if the last i symbols are all 1s) then look back i sites before this occurrence. In other words, the finite size contexts have the form

$$a_{-2i-1}^{-i-2} 2 1^i \quad \text{for all } i \geq 0 \text{ and } a_{-2i-1}^{-i-2} \in A^i,$$

and the context tree is

$$\tau = 1^{-\mathbb{N}} \cup \bigcup_{i \geq 0} \bigcup_{c \in A^i} c 2 1^i; \tag{2}$$

see Figure 2, for a representation.

Our results state that, assuming that $\inf_{v \in \tau} p(2 | v) \geq \varepsilon > 0$, we can perfectly simulate the unique stationary chain X compatible with (τ, p) (Theorem 1). The perfect simulation algorithm extends the algorithm we briefly described in Section 3.1 for the renewal chain. The main difference is in the definition of the regeneration time. In Section 9 we will give an explicit perfect simulation of this chain. We also show in this work that, almost surely, infinitely many occurrences of 2 split the realization of X into i.i.d. strings (Theorem 2). However, since any occurrence of 2 can be bypassed by a context at a future time index with positive probability, this 'regeneration scheme' differs substantially from the 'renewal scheme'.

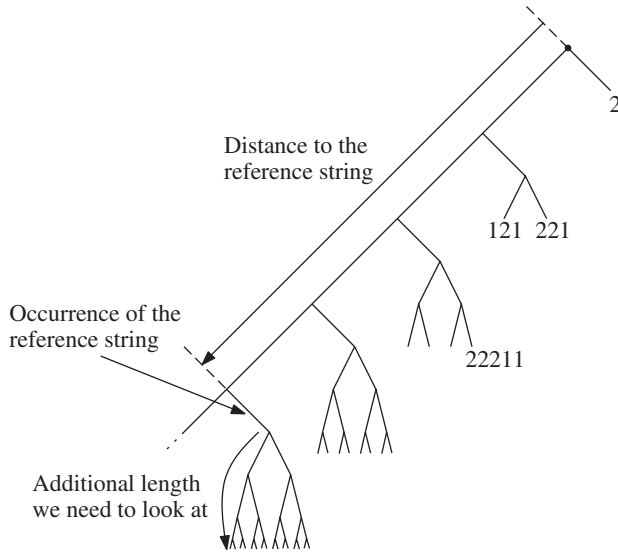


FIGURE 2: The upper part of the context tree τ defined by (2). We only specify some of the contexts.

4. On the form of our context trees

The examples of the preceding section gave us an idea of how our families can be described in terms of probabilistic context trees. The aim of the following definitions is to define, using the probabilistic context tree framework, what we called a ‘reference string’. At the end of this section, we give several examples explaining these definitions.

Suppose that we are given an unbounded context tree τ . For any finite string w of A^* , we define the function m^w which associates to any context $v \in \tau$ the integer number

$$m^w(v) := \inf\{j : 0 \leq j \leq |v| - |w| \text{ such that } v_{-j-|w|}^{-j-1} = w\}, \tag{3}$$

with the convention that $m^w(v) = +\infty$ if the set of indices is empty. In the context tree, $m^w(v)$ is the distance between the root and the first occurrence of w in the context v . If a context v is such that $m^w(v) = k$ then it can be written as the concatenation

$$v = v_{-|v|} \cdots v_{-k-|w|-1} w v_{-k} \cdots v_{-1},$$

where $v_{-j}^{-j+|w|-1} \neq w$ for $j = |w|, \dots, k + |w| - 1$. The context trees considered in the present work have the following form. There exist a finite string w of A^* and, related to this string, a function $\ell^w : \mathbb{N} \rightarrow \mathbb{N}$ satisfying $\ell^w(k) < +\infty$ for any $k \geq 0$ such that, for any $v \in \tau$,

$$|v| = m^w(v) + |w| + \ell^w(m^w(v)).$$

The string w is the *reference string* for the context tree τ . The function ℓ^w tells us how much further back from the last occurrence of w we need to look. It is precisely the notion of a reference string which generalizes the notion of a renewal string.

4.1. Example of Figure 1(b)

The reference string is the symbol 2. The function ℓ^2 is identically 0, which is equivalent to stating that 2 is a renewal symbol.

4.2. Example of Figures 2 and 3(a)

The reference string is also the symbol 2 in Figures 2 and 3(a). In Figure 2, ℓ^2 is the identity function. Comparing Figures 1(b) and 2, it is clear that the notion of a reference string generalizes the notion of a renewal symbol.

To simplify Figure 3, we used small triangles to depict subtrees. These subtrees are context trees of finite height since 2 is a reference string. Suppose that the context tree illustrated in Figure 3(a) is such that $\ell^w(k) = 1 + k^2$. It follows that any context of the context tree of $\tau_{\text{Figure 2}}$ is a suffix of a context of $\tau_{\text{Figure 3(a)}}$. In this case, we write $\tau_{\text{Figure 2}} \leq \tau_{\text{Figure 3(a)}}$, and observe that

$$\tau_{\text{Figure 1(b)}} \leq \tau_{\text{Figure 2}} \leq \tau_{\text{Figure 3(a)}}.$$

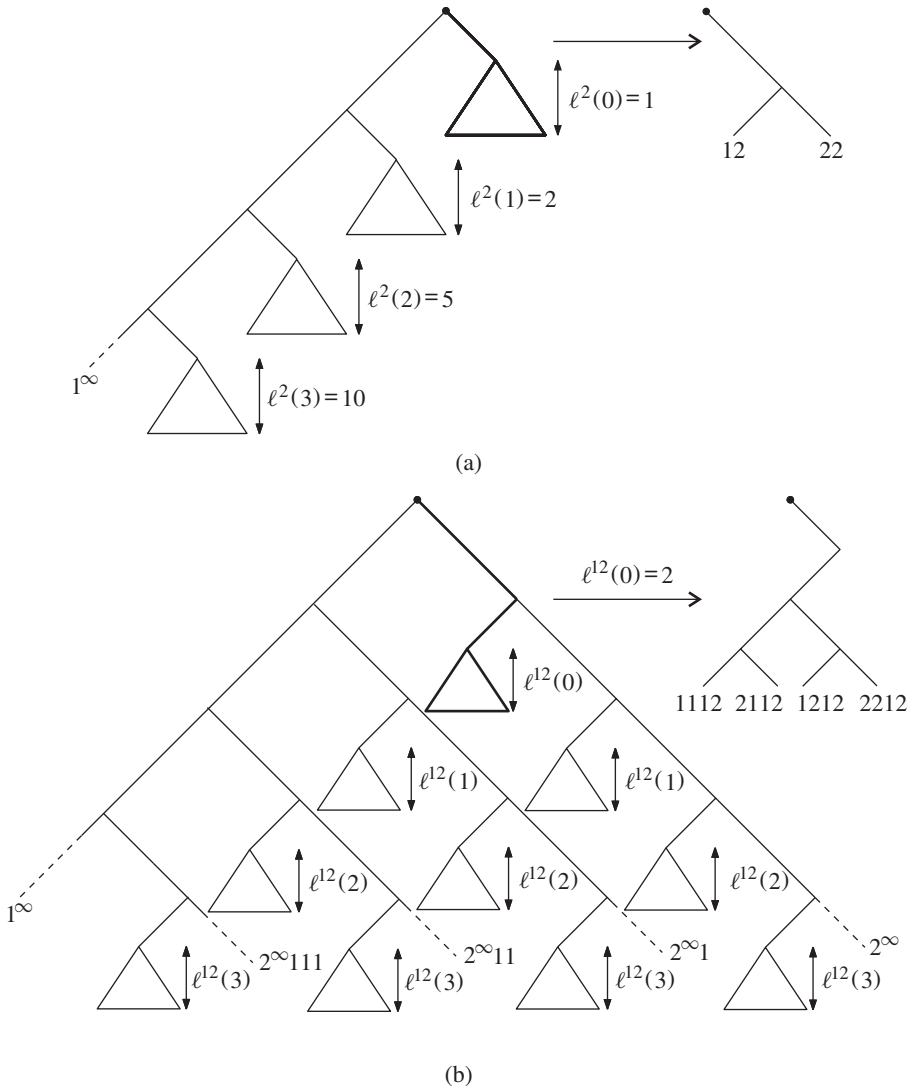


FIGURE 3: The upper parts of two unbounded context trees, with reference string (a) 2 and (b) 12.

There is a difference between $\tau_{\text{Figure 3(a)}}$ and the two others: the reference string 2 is not a context.

4.3. Example of Figure 3(b)

The reference string is 12 and it is not a context. We can see five infinite size contexts, but there are infinitely many of them.

5. Perfect simulation algorithm and statement of the results

We will assume without loss of generality that $A = \{1, 2, \dots\}$. We introduce a partition of $[0, 1)$, illustrated in Figure 4, which will be used for the construction of the chain. Let $J(1 | \emptyset) := [0, \alpha(1))$, and, for any $a \geq 2$, let

$$J(a | \emptyset) = \left[\sum_{i=1}^{a-1} \alpha(i), \sum_{i=1}^a \alpha(i) \right).$$

We also define, for any $a \in A$ and $v \in \tau$,

$$J(a | v) := \left[\alpha + \sum_{i=1}^{a-1} (p(i | v) - \alpha(i)), \alpha + \sum_{i=1}^a (p(i | v) - \alpha(i)) \right),$$

and the union $K(a | v) := J(a | \emptyset) \cup J(a | v)$. Note that if $\alpha(a) = 0$ then $K(a | v) = J(a | v)$ since $J(a | \emptyset) = \emptyset$. Moreover, for any $v \in \tau$,

$$J(1 | \emptyset), J(2 | \emptyset), \dots, J(1 | v), J(2 | v), \dots$$

defines a partition of $[0, 1)$ (see Figure 4), and, for any $a \in A$ and any $v \in \tau$,

$$\lambda(K(a | v)) = p(a | v),$$

where λ denotes the Lebesgue measure on $[0, 1)$.

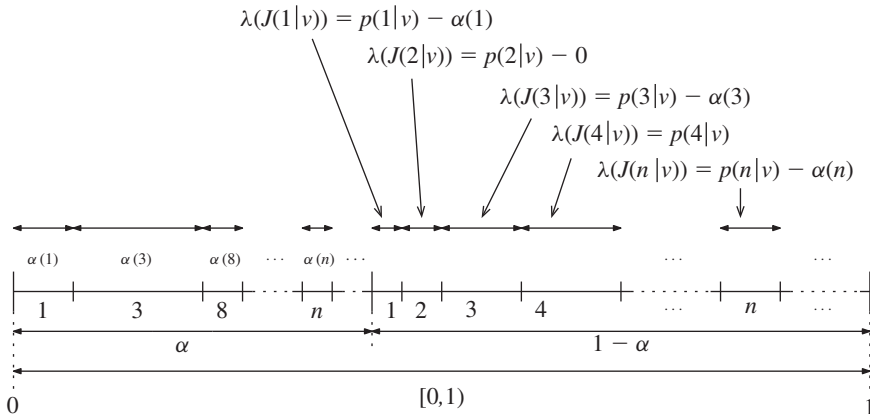


FIGURE 4: Illustration of the partition of $[0, 1)$ with the disjoint intervals $\{J(a | \emptyset)\}_{a \in A}$ and $\{J(a | v)\}_{a \in A}$ for some $v \in \tau$. In this particular case we have, for example, $\alpha(a) = 0$ for $a = 2, 4, 5, 6, 7$.

Let $U = (U_n)_{n \in \mathbb{Z}}$ be a sequence of i.i.d. random variables uniformly distributed in $[0, 1)$ and defined on some probability space $(\Omega, \mathcal{F}, \Pr)$. All the chains considered in what follows will be constructed using this sequence U .

We construct a deterministic measurable function $X: [0, 1)^{\mathbb{Z}} \rightarrow A^{\mathbb{Z}}$ such that the law $\Pr(X(U) \in \cdot)$ is compatible with (τ, p) . The construction of this function is carried out in such a way that, for any $n \in \mathbb{Z}$, $[X(U)]_n = a$ whenever $U_n \in J(a \mid \emptyset)$. Suppose that, for some time index $n \in \mathbb{Z}$, there exists a string $a_{-k}^{-1} \in A^k$ such that $U_{n-i} \in J(a_{-i} \mid \emptyset)$, $i = 1, \dots, k$ (it is understood that $|J(a_{-i} \mid \emptyset)| > 0$ for these symbols). In this case, we set

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

This is a sample that has been *spontaneously* constructed. We have three situations at this point: (i) U_n belongs to $[0, \alpha)$, (ii) U_n belongs to $[\alpha, 1)$ and $c_\tau(a_{-k}^{-1}) = v \in \tau$, and (iii) U_n belongs to $[\alpha, 1)$ and $c_\tau(a_{-k}^{-1}) = \emptyset$. In situation (iii), we are not able to construct $[X(U)]_n$ knowing *only* $[X(U)]_{n-k}^{n-1}$, and, therefore, we need to determine more past symbols. In situations (i) and (ii), we can construct $[X(U)]_n$ independently of $[X(U)]_{-\infty}^{n-k-1}$ and we set, for any $a \in A$,

$$[X(U)]_n = a \quad \text{if } U_n \in K(a \mid v).$$

Observe that $[X(U)]_{n-k}^n$ has been constructed independently of $U_{-\infty}^{n-k-1}$ and $U_{n+1}^{+\infty}$. Suppose that we want to sample the value of the stationary chain at time 0. The idea of the algorithm is to generate the U_i s backward in time until the first time $k \leq 0$ such that we can carry out the above construction from k to 0, without using $U_{-\infty}^{k-1}$ and $U_1^{+\infty}$. This is a CFTP algorithm.

The size of the suffix of the past we need to know in order to construct the next symbol depends here on the previously constructed past itself (except for the time indices where the symbol is spontaneously constructed). In the CFTP algorithm introduced in [7], the size of this suffix of the past is defined by an i.i.d. random variable, totally independent of the values assumed by the chain. This is the main difference between both works from a technical point of view, and it is what makes our perfect simulation algorithm a little bit more complicated.

Situations (i), (ii), and (iii) given above are formally described by the measurable function $F: [0, 1) \times (A^* \cup A^{-\mathbb{N}}) \rightarrow A \cup \{\star\}$, which we define as follows. For any $a_m^n \in A^* \cup A^{-\mathbb{N}}$, $-\infty \leq m \leq n + 1$,

$$F(u, a_m^n) = \sum_{a \in A} a \mathbf{1}\{u \in K(a \mid c_\tau(a_m^n))\} + \star \mathbf{1}\{u \in [\alpha, 1), c_\tau(a_m^n) = \emptyset\},$$

with the conventions that $a_{n+1}^n = \emptyset$ and $c_\tau(\emptyset) = \emptyset$. When $c_\tau(a_m^n) \neq \emptyset$, we have

$$\Pr(F(U_{n+1}, a_m^n) = a) = \Pr(U_{n+1} \in K(a \mid c_\tau(a_m^n))) = \lambda(K(a \mid c_\tau(a_m^n))) = p(a \mid c_\tau(a_m^n)). \tag{4}$$

This function is an *update function*. Since we consider chains of infinite order, this update function may return the symbol \star , meaning that we do not have sufficient knowledge of the past to continue the construction.

We define, for any $m \leq n$, the $\mathcal{F}(U_m^n)$ -measurable function $\mathcal{L}: [0, 1)^{n-m+1} \rightarrow \{0, 1\}$, which takes value 1 if and only if we can construct $[X(U)]_m^n$ independently of $U_{-\infty}^{m-1}$ and $U_{n+1}^{+\infty}$ using the construction described above. Formally,

$$\{\mathcal{L}(U_m^n) = 1\} := \bigcup_{a_m^n \in A^{n-m+1}} \bigcap_{i=m}^n \{F(U_i, a_m^{i-1}) = a_i\}.$$

Finally, we define, for any $m \leq n \leq +\infty$,

$$\theta[m, n] := \max\{k \leq m : \mathcal{L}(U_k^n) = 1\} \tag{5}$$

with the convention that $\theta[m] := \theta[m, m]$. We call this time the *regeneration time* for the window $[m, n]$, and it is the first time before m such that the construction described above is successful until time n .

We now state our main results and present a ‘simplistic’ perfect simulation algorithm (Algorithm 1) for the construction of a sample $[X(U)]_m^n$. A more ‘realistic’ algorithm (Algorithm 2) is given in Section 9 together with an explicit perfect simulation in the particular case of the example given in Section 3.

Algorithm 1. (*‘Simplistic’ perfect simulation algorithm of the sample $[X(U)]_m^n$*)

1. Inputs: m and n . Outputs: $\theta[m, n]$ and $([X(U)]_{\theta[m, n]}, \dots, [X(U)]_n)$.
2. Sample U_m, \dots, U_n uniformly in $[0, 1)$.
3. $i \leftarrow m, \theta[m, n] \leftarrow m, [X(U)]_m^n \leftarrow \star^{n-m+1}, \mathcal{L}(U_m^n) \leftarrow 0$
4. **while** $\mathcal{L}(U_i^n) \neq 1$ **do**
5. $i \leftarrow i - 1$
 Choose U_i uniformly in $[0, 1)$
6. **end while**
7. $\theta[m, n] \leftarrow i$
8. **while** $[X(U)]_n = \star$ **do**
9. $[X(U)]_i \leftarrow F(U_i, [X(U)]_{\theta[m, n]}^{i-1})$
 $i \leftarrow i + 1$
10. **end while**
11. **return** $\theta[m, n], ([X(U)]_{\theta[m, n]}, \dots, [X(U)]_n)$

Theorem 1. (Perfect simulation.) *Consider a probabilistic context tree (τ, p) having a reference string w which is ε -regular for some $\varepsilon > 0$. If*

$$\limsup_{k \rightarrow \infty} \frac{\log(\ell^w(k))}{C_\varepsilon k} < 1 \quad \text{and} \quad C_\varepsilon := -\frac{1}{|w|} \log(1 - \varepsilon) > 0,$$

then Algorithm 1 (and Algorithm 2 in Section 9) stops almost surely after a finite number of steps, i.e. we have, for any $-\infty < m \leq n \leq +\infty$,

$$\Pr(\theta[m, n] > -\infty) = 1.$$

Moreover, $\Pr(\theta[0, n] < -l)$ is summable in l for any $0 \leq n < +\infty$, and, therefore, the number of steps has finite expectation. Under the stronger requirement that ℓ^w is linear, $\Pr(\theta[0, n] < -l)$ decreases exponentially fast for any $0 \leq n \leq +\infty$.

In the sequel, we will often write X_i for $[X(U)]_i$ (and X for $X(U)$) in order to avoid overloaded notation, keeping in mind the fact that, for any i , X_i is constructed as a deterministic function of U . Actually, by Theorem 1, X_i depends only on a Pr-almost surely finite part of this sequence: $X_i := [X(\dots, u_{\theta[i]-1}, U_{\theta[i]}, \dots, U_i, u_{i+1}, \dots)]_i$ for any $\mathbf{u} \in [0, 1)^{\mathbb{Z}}$.

Corollary 1. (Existence and uniqueness.) *The output of Algorithm 1 (and Algorithm 2 in Section 9) is a sample of the unique stationary chain compatible with (τ, p) . We will call μ the stationary measure of X :*

$$\mu := \Pr(X(\mathbf{U}) \in \cdot).$$

Moreover, this measure has support on the finite size contexts, that is,

$$\mu(\{|c_\tau([X(\mathbf{U})]_{-\infty}^{-1})| < +\infty\}) = 1.$$

This implies that the set of discontinuous pasts has μ -measure 0.

The assumption that w is ε -regular for some $\varepsilon > 0$ is weaker than the *regularity* (also called *strongly non-nullness*) assumption of the literature. Also, we know that this latter condition neither implies existence nor uniqueness of the stationary measure (see, for example, [4]). Our assumption is comparable to the weakly non-nullness assumption, which is also assumed, for example, in [7]. In fact, this condition is very useful for our construction: it allows us to have symbols which appear spontaneously, and makes the CFTP easier to perform.

The proof of the first statement of Corollary 1 using the CFTP algorithm and Theorem 1 can be found in [7] (Proposition 6.1 for the existence statement and Corollary 4.1 for the uniqueness statement). We omit these proofs in the present work in order to save space, but we mention the main lines. The existence statement follows once we observe that Theorem 1 implies that we can construct a bi-infinite sequence X verifying, for any $n \in \mathbb{Z}$, $X_n = F(U_n, X_{-\infty}^{n-1})$. By (4), this chain is therefore compatible in the sense of (1). 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 for finite length windows. The proof of the second statement of Corollary 1 follows simply from the fact that, in the constructed chain, the context observed at time 0 has a length which is a function of the distance to the last occurrence of w in the past, and this function is finite whenever the distance is finite. But this distance is clearly finite since it is bounded above by the smallest j such that $U_{-i} \in J(w_{-i} | \emptyset)$, $i = 1, \dots, |w|$.

We call time t a regeneration time for the chain X if $\theta[t, +\infty] = t$. Define the chain ξ on $\{0, 1\}$ by $\xi_j := \mathbf{1}\{j = \theta[j, +\infty]\}$. Then, consider the sequence of time indices T defined by $\xi_j = 1$ if and only if $j = T_l$ for some l in \mathbb{Z} , where $T_l < T_{l+1}$, and with the convention $T_0 \leq 0 < T_1$. We say that X has a regeneration scheme if the chain ξ is renewal (that is, if the increments $(T_{i+1} - T_i)_{i \in \mathbb{Z}}$ are independent, and are identically distributed for $i \neq 0$).

Corollary 2. (Regeneration scheme.) *Under the conditions of Theorem 1, the chain X has a regeneration scheme. The random strings $([X(\mathbf{U})]_{T_i}, \dots, [X(\mathbf{U})]_{T_{i+1}-1})_{i \neq 0}$ are i.i.d. and have finite expected size. Under the stronger requirement that ℓ^w is linear, the length of this strings has an exponential tail.*

In words, Corollary 2 states that the unique stationary chain compatible with (τ, p) under the conditions of Theorem 1 can be viewed as an i.i.d. concatenation of strings of symbols of A having finite expected size. A similar result was first obtained in [18] for one-dimensional Gibbs states under appropriate conditions on the continuity rate, and then in [7] under weaker conditions than those of [18]. It is a hidden regeneration scheme, because it uses the sequence U . The main reason why we give this result is that it arises naturally from our perfect simulation approach.

The visible regeneration scheme involves several technical complications, even if in spirit, it is similar to the preceding scheme. We postpone the precise definitions to Section 8, and give the following simplified statement.

Theorem 2. (Visible regeneration scheme.) *Suppose that (τ, p) satisfies the conditions of Theorem 1. Then, for μ -almost every realization of the chain X compatible with (τ, p) , there exists a sequence of random times T^X such that*

- *for any $i \in \mathbb{Z}$, the event $\{T_i^X = k\}$ is measurable with respect to the σ -algebra generated by $X_k^{+\infty}$; and*
- *conditionally on T^X , the strings $(X_{T_i^X}, \dots, X_{T_{i+1}^X-1})_{i \neq 0}$ are i.i.d. and have finite expected size.*

It is worth mentioning that, while there exists a visible regeneration scheme with probability 1, part of the regeneration times are invisible. In fact, the set of visible and invisible regeneration times are generally different, because $\{t \in T\} \not\Rightarrow \{t \in T^X\}$ and $\{t \in T^X\} \not\Rightarrow \{t \in T\}$. We conclude with the following important observation.

Observation 1. (Monotonicity.) *Suppose that the probabilistic context tree (τ, p) satisfies the conditions of the above results. Then, all the above results hold true for any probabilistic context tree (τ', p') such that $\tau' \leq \tau$ and for which w is ε -regular.*

6. Proof of Theorem 1

Suppose that we are given a probabilistic context tree (τ, p) having an ε -regular reference string w to which corresponds the function ℓ^w . Owing to Observation 1, there is no loss of generality in restricting the proof to the case where (i) only the branches having w as a subsequence have finite length and (ii) ℓ^w increases and goes to ∞ .

We present a detailed proof for the $|w| = 1$ case only because the $|w| \geq 1$ case presents several notational complications which could hinder the comprehension of the reader, whereas there is no conceptual difference with the $|w| = 1$ case. In Section 6.5 we explain the main differences between both cases.

A slight complication arises from the fact that the random variable θ depends on the values assumed by the chain X along its construction. We first introduce in Section 6.1 the random variable $\Theta[0, n]$ (see (7)), which has the following properties: (i) it depends only on the spontaneous occurrences of w along the construction; (ii) it can be used to define a lower bound for $\theta[0, n]$.

In Sections 6.2 and 6.3 we relate the distribution of $\Theta[0, n]$ to the probability of return to the state 0 for an \mathbb{N} -valued auxiliary process which also depends on the spontaneous occurrences of w . At this point, there is a clear similarity with the proof given in [7], the principal difference being that our auxiliary process is not the house-of-cards process, since it is not Markovian.

The proof of Theorem 1 for the $|w| = 1$ case is given in Section 6.4. In Section 6.5 we detail how the $|w| = 1$ proof is modified for the generic $|w| \geq 1$ case.

6.1. Definition of a new random variable $\Theta[0, n]$ for the $|w| = 1$ case

In order to simplify the notation, we will write ℓ and m for ℓ^w and m^w . We define a new stochastic chain Z : for any $i \in \mathbb{Z}$, $Z_i = w$ if U_i belongs to $J(w \mid \emptyset)$, and $Z_i = \star$ otherwise. This chain takes into account the spontaneous occurrences of w in X : $X_i = w$ whenever $Z_i = w$. We also define the \mathbb{N} -valued random variables $m_i(U) = m_i$ and $L_i(U) = L_i$ by

$$m_i = \inf\{k \geq 0: Z_{i-k-1} = w\},$$

which is the distance to the last occurrence of w in $Z_{-\infty}^{i-1}$, and

$$L_i = \begin{cases} 0 & \text{if } U_i \in [0, \alpha), \\ m_i + 1 + \ell(m_i) & \text{otherwise.} \end{cases} \tag{6}$$

We introduce these random variables because if we have $\mathcal{L}(U_m^n) = 1$ for $-\infty < m < n < +\infty$ and $0 < L_{n+1} \leq n - m + 1$, then L_{n+1} is an upper bound for the number of sites in the past that we need to know in order to decide the state at time $n + 1$ using the perfect simulation algorithm. To see this, observe that, since

$$Z_i = w \Rightarrow X_i = w \quad \text{for all } i \in \mathbb{Z},$$

it follows that the distance to the last occurrence of w in Z_m^n is larger than in X_m^n :

$$m_{n+1} \geq m(c_\tau(X_m^n)).$$

Recall that here m stands for m^w , defined in (3). We also recall that, for any $v \in \tau$ in which the string w appears, we have

$$|v| = m(v) + 1 + \ell(m(v));$$

therefore, by definition (6), whenever $L_{n+1} > 0$, L_{n+1} is an upper bound for the size of the context needed at time $n + 1$:

$$|c_\tau(X_m^n)| \leq L_{n+1}.$$

Observe also that $L_{n+1} = 0$ if and only if the symbol appears spontaneously at time $n + 1$. We are now ready to define the new random variable: for any $n \geq 0$,

$$\Theta[0, n] := \max\{j \leq 0 : L_i \leq i - j, i = j, \dots, n\}. \tag{7}$$

The following lemma justifies the introduction of this new random variable.

Lemma 1. *For any $n \geq 0$, we have, in the set $\{U : \Theta[0, n] > -\infty\}$,*

$$\Theta[0, n] \leq \theta[0, n]. \tag{8}$$

Proof. To each time $i \in \{\Theta[0, n], \dots, n\}$, we associate an arrow going from time i to time $i - L_i$. The definition of $\Theta[0, n]$ says that no arrows starting from $\{\Theta[0, n], \dots, n\}$ go beyond time $\Theta[0, n]$. This means that the construction of $X_{\Theta[0, n]}^n$ can be performed recursively from time $\Theta[0, n]$ to time n using only $U_{\Theta[0, n]}^n$, and, therefore, that $\mathcal{L}(U_{\Theta[0, n]}^n) = 1$. Since $\theta[0, n]$ is the maximum over $\{k \leq 0 : \mathcal{L}(U_k^n) = 1\}$, it follows that (8) holds.

6.2. An auxiliary chain to study $\Theta[0, n]$ for the $|w| = 1$ case

The definition of $\Theta[0, n]$ given in (7), which uses the chain L , is analogous to the definition of the regeneration time given in [7, Equation (3.5)], which uses another \mathbb{N} -valued chain K (which basically plays the same role as L). Thus, as in Comets *et al.* [7], we can define an auxiliary \mathbb{N} -valued stochastic chain $W^{(n)}$. We set $W_i^{(n)} = 0$ for any $i \leq n$ and

$$W_i^{(n)} = (W_{i-1}^{(n)} + 1)\mathbf{1}\{L_i < i - n\} \quad \text{for all } i \geq n + 1. \tag{9}$$

In our case, this chain is *not* exactly a *house-of-cards process* as it is not Markovian, whereas in the case of [7], it is Markovian. This difference is due to the fact that L is not i.i.d., whereas their chain K is i.i.d. Nevertheless, the sequence of equivalences given in Equation (5.6) of [7] also holds in our case, and we obtain

$$\Theta[0, n] < -l \iff \text{there exists } j \in \{0, \dots, n\} \text{ such that } W_j^{(-l-1)} = 0. \tag{10}$$

6.3. Study of $\Pr(W_k^{(0)} = 0)$ for the $|w| = 1$ case

Define the *inverse* function of ℓ by

$$\ell^{-1}(i) = \inf\{k \geq 1: \ell(k) > i\} \quad \text{for all } i \geq 0. \tag{11}$$

Our objective here is to prove the following proposition.

Proposition 1. *Let $W^{(0)}$ be the chain defined through (9) using an \mathbb{N} -valued function ℓ and the i.i.d. chain Z on $\{1, \star\}$ with distribution $(\varepsilon, 1 - \varepsilon)$. Then, the sequence $u_k := \Pr(W_k^{(0)} = 0)$*

1. *is summable when $(1 - \varepsilon)^{\ell^{-1}(i)}$ is summable;*
2. *decreases exponentially fast when $(1 - \varepsilon)^{\ell^{-1}(i)}$ decreases exponentially fast.*

Proof. Part of the present proof is inspired by the proof of Proposition 2 of [5] for the house-of-cards process. Denote by ζ the first time $W^{(0)}$ returns to state 0, and by f_k the probability $\Pr(\zeta = k)$. First, we observe that the state 0 is a renewal state for the chain $W^{(0)}$. It follows that the sequences $(u_k)_{k \geq 1}$ and $(f_k)_{k \geq 1}$ satisfy

$$u_k = \sum_{i=1}^k f_i u_{k-i}. \tag{12}$$

By (12), the series

$$F(s) := \sum_{n \geq 1} f_n s^n \quad \text{and} \quad U(s) := \sum_{n \geq 1} u_n s^n$$

are related through

$$U(s) = \frac{1}{1 - F(s)} \tag{13}$$

for $s \geq 1$ such that $F(s) < 1$ (see, for example, Theorem 1 of [9, Chapter XIII.10]). In order to prove statement 1, all we need to prove is that state 0 is transient, that is, $F(1) < 1$, whenever $(1 - \varepsilon)^{\ell^{-1}(i)}$ is summable. Suppose that, for some $M > 0$, we have $Z_1^M = 1^M$, so that in particular $W_M^{(0)} = M$. The first possible arrow which can go until or beyond time 0 could be that of $M + \ell^{-1}(M - 1)$. This follows from definition (11) of $\ell^{-1}(M - 1)$. Then, in order that the chain $W^{(0)}$ touches 0 at the first possible time after M , it is necessary that $\ell^{-1}(M - 1)$ stars appear in Z from time $M + 1$ to time $M + \ell^{-1}(M - 1)$. More specifically, we have, for any integer $M \geq 1$,

$$\bigcup_{i \geq 1} \{\zeta = i\} \cap \{Z_1^M = 1^M\} = \bigcup_{i \geq M} \{Z_{i+1}^{i+\ell^{-1}(i-1)} = \star^{\ell^{-1}(i-1)}\} \cap \{Z_1^M = 1^M\}.$$

It follows that, using the partition

$$\bigcup_{i \geq 1} \{\zeta = i\} = \bigcup_{i \geq 1} \{\zeta = i\} \cap \{Z_1^M = 1^M\} \cup \bigcup_{i \geq 1} \{\zeta = i\} \cap \{Z_1^M \neq 1^M\} \quad \text{for all } M \geq 1,$$

we obtain the simple upper bound

$$\Pr\left(\bigcup_{i \geq 1} \{\zeta = i\}\right) \leq \Pr\left(\bigcup_{i \geq M} \{Z_{i+1}^{i+\ell^{-1}(i-1)} = \star^{\ell^{-1}(i-1)}\} \cap \{Z_1^M = 1^M\}\right) + \Pr(Z_1^M \neq 1^M).$$

The events $\bigcup_{i \geq M} \{Z_{i+1}^{i+\ell^{-1}(i-1)} = \star^{\ell^{-1}(i-1)}\}$ and $\{Z_1^M = 1^M\}$ are independent since the Z_i s are i.i.d. Therefore, for any $M \geq 1$,

$$\Pr\left(\bigcup_{i \geq 1} \{\zeta = i\}\right) \leq \varepsilon^M \sum_{i \geq M-1} (1 - \varepsilon)^{\ell^{-1}(i)} + 1 - \varepsilon^M.$$

If $\sum_{i \geq 0} (1 - \varepsilon)^{\ell^{-1}(i)} < +\infty$, we can take M sufficiently large to ensure that $\sum_{i \geq M-1} (1 - \varepsilon)^{\ell^{-1}(i)} < 1$. Thus, $\sum_{i \geq 0} (1 - \varepsilon)^{\ell^{-1}(i)} < +\infty$ implies that $\sum_{i \geq 1} f_i < 1$, concluding the proof of statement 1.

For the proof of statement 2, let us suppose that $(1 - \varepsilon)^{\ell^{-1}(i)}$ decreases exponentially fast. Then, on the one hand, $W^{(0)}$ is transient and, therefore, F and U are related through (13); on the other hand, $\ell^{-1}(i) \sim i$. It follows that

$$f_n = \Pr\left(W_i^{(0)} > 0, i = 1, \dots, \frac{n}{2} - 1\right) \varepsilon (1 - \varepsilon)^{n/2},$$

and $f_n / (1 - \varepsilon)^{n/2} \rightarrow \Pr(\zeta = +\infty) > 0$ as $n \rightarrow +\infty$. Thus, the radius of convergence of F is

$$\lim_{n \rightarrow +\infty} ((1 - \varepsilon)^{n/2})^{-1/n} = (1 - \varepsilon)^{-1/2},$$

which is strictly larger than 1. Since $F(1) = \Pr(\zeta < +\infty) < 1$, it follows that, by continuity, there exists a real number $s_0 > 1$ such that $F(s_0) = 1$. By (13), this means that $U(s) < +\infty$ for $s < s_0$, and, by the definition of U , it implies that u_n decreases faster than r^n for $r \in (s_0^{-1}, 1)$.

6.4. Proof of Theorem 1 for the $|w| = 1$ case

The proof of the theorem is now simple. By Lemma 1 and (10), we have

$$\Pr(\theta[0, n] < -l) \leq \sum_{k=l+1}^{l+n+1} u_k. \tag{14}$$

Under the conditions of Theorem 1, when $|w| = 1$, we have

$$\limsup_{k \rightarrow \infty} \frac{\log \ell(k)}{\log(1/(1 - \varepsilon))^k} < 1;$$

therefore, there exists a real number $\gamma > 0$ such that, for any sufficiently large k , $\ell(k) \leq (1/(1 - \varepsilon))^{k/(1+\gamma)}$. It follows that $\ell(k) \leq (1/(1 - \varepsilon))^{(k+1)/(1+\gamma)}$ for sufficiently large k and, by definition (11) of ℓ^{-1} ,

$$\ell^{-1}(n) \geq \frac{1 + \gamma}{\log(1/(1 - \varepsilon))} \log n - 1$$

for any sufficiently large n . Therefore, there exists n^* such that

$$\sum_{n=0} (1 - \varepsilon)^{\ell^{-1}(n)} \leq \sum_{n=0}^{n^*-1} (1 - \varepsilon)^{\ell^{-1}(n)} + (1 - \varepsilon)^{-1} \sum_{n \geq n^*} n^{-1-\gamma},$$

which is finite since γ is strictly positive. By item 1 of Proposition 1, $(u_k)_{k \geq 1}$ is summable. Thus, by (14), we obtain

$$\sum_{l \geq 0} \Pr(\theta[0, n] < -l) \leq \sum_{l \geq 0} \sum_{k=l+1}^{l+n+1} u_k \leq (n + 1) \sum_{l \geq 1} u_l,$$

implying that $\Pr(\theta[0, n] < -l)$ is summable in l for any $n, 0 \leq n < +\infty$, and that $\Pr(\theta[0, +\infty] < -l)$ goes to 0 when l diverges. For the second statement of Theorem 1, if ℓ is linear then ℓ^{-1} is linear, and by item 2 of Proposition 1, u_k decreases exponentially fast to 0. By (14), this implies that $\Pr(\theta[0, n] < -l)$ decreases exponentially fast as l increases for any $n, 0 \leq n \leq +\infty$. This concludes the proof of Theorem 1.

6.5. Proof of Theorem 1 for the generic $|w| \geq 1$ case

To keep the same form as $\Theta[0, n]$ defined in (7), for the general case, $|w| \geq 1$, we introduce the time rescaled chain \bar{Z} and the function $\bar{\ell}$ which respectively play the roles of Z and ℓ in the preceding proof, and which are defined by

$$\bar{Z}_m = \begin{cases} 1 & \text{if } U_{m|w|-i+1} \in J(w_{-i} \mid \emptyset), i = 0, \dots, |w| - 1, \\ \star & \text{otherwise,} \end{cases}$$

and

$$\bar{\ell}(i) := \left\lceil \frac{\ell^w((i + 1)|w| - 1)}{|w|} \right\rceil.$$

For any $r \in \mathbb{R}$, $\lceil r \rceil$ denotes the smallest integer greater than or equal to r . Using these new definitions, we introduce the rescaled 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 $\bar{Z}_{-\infty}^{i-1}$, and

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

We are now able to define our new random time:

$$\bar{\Theta}[0, n] := \max\{j \leq 0: \bar{L}_i \leq i - j, i = j, \dots, n\}. \tag{15}$$

Note that $\Theta[0, n]$ defined by (7) and $\bar{\Theta}[0, n]$ defined by (15) have exactly the same form. We now state the following lemma, the analogue of Lemma 1, but for the generic case. Its proof is omitted, as the only complexity arises from the notation.

Lemma 2. *For any $n \geq 0$, we have, in the set $\{U: \bar{\Theta}[0, n] > -\infty\}$,*

$$\theta[0, n|w|] \geq |w|(\bar{\Theta}[0, n] - 1) + 1.$$

Now define a process $\bar{W}^{(0)}$ using \bar{L} and $\bar{\ell}$, defined in the same way as $W^{(0)}$ using L and ℓ . Equivalence (10) still holds, substituting $W^{(-l-1)}$ with $\bar{W}^{(-l-1)}$ and $\Theta[0, n]$ with $\bar{\Theta}[0, n]$. Note that \bar{Z} is i.i.d. on $\{1, \star\}$ with distribution $(\varepsilon, 1 - \varepsilon)$; thus, Proposition 1 also holds for $\bar{W}^{(0)}$ if we replace ℓ by $\bar{\ell}$ in the statement. To conclude the proof in the generic case, first observe that

$$\Pr(\theta[0, n] < -l) \leq \Pr\left(\theta\left[0, |w|\left\lceil \frac{n}{|w|} \right\rceil\right] < -|w|\left\lfloor \frac{l}{|w|} \right\rfloor\right),$$

where $\lfloor r \rfloor$ denotes the integer part of r . By Lemma 2 and the transformed version of (10), we obtain

$$\Pr(\theta[0, n] < -l) \leq \sum_{k=\lfloor l/|w| \rfloor}^{\lfloor l/|w| \rfloor + \lceil n/|w| \rceil} \bar{u}_k$$

for any l such that $\lfloor l/|w| \rfloor \geq 1$, where $\bar{u}_k := \Pr(\bar{W}_k^{(0)} = 0)$. The rest of the proof is identical

to the proof given in Section 6.4, upon making the necessary changes to the notation and using the transformed version of Proposition 1.

7. Proof of Corollary 2

We refer the reader to Section 8 of [7] for a complete proof of the statements of Corollary 2. The lines of the proof given therein are as follows.

7.1. Existence of a regeneration scheme

On the one hand, we have $\Pr(\theta[0, +\infty] = 0) > 0$, which follows from the fact that

$$\Pr(\theta[0, +\infty] = 0) \geq \Pr\left(\bigcap_{i \geq 1} \{W_i^{(0)} \geq 1\}\right),$$

which is strictly positive, since the state 0 is transient in the conditions of Theorem 1 (this is shown in the proof of Proposition 1). On the other hand, we have to check that the chain ξ , defined by $\xi_j := \mathbf{1}\{j = \theta[j, +\infty]\}$, is renewal. This follows from the fact that, by definition (5) of the random variable θ , we have

$$\bigcap_{l=1}^n \{\theta[t_l, +\infty] = t_l\} = \bigcap_{l=1}^n \{\theta[t_l, t_{l+1} - 1] = t_l\}$$

(where we have used the convention $t_{n+1} = +\infty$), which is an intersection of independent events, since $\{\theta[t_l, t_{l+1} - 1] = t_l\}$ is $\mathcal{F}(U_{t_l}^{t_{l+1}-1})$ -measurable for $l = 1, \dots, n$. To conclude, the fact that the random strings $([X(\mathbf{U})]_{T_l}, \dots, [X(\mathbf{U})]_{T_{l+1}-1})_{l \neq 0}$ are i.i.d. follows from the construction using Algorithm 1 or Algorithm 2 (see Section 9.1).

7.2. On the tail distribution of the length of the i.i.d. strings

To show that the expected size between two consecutive 1s in ξ is finite, we observe that, by stationarity and definition (5) of the random variable θ ,

$$\Pr(T_{l+1} - T_l \geq m) = \Pr(\theta[1, +\infty] \leq -m \mid \theta[0, +\infty] = 0) = \Pr(\theta[0] < -m + 1),$$

which has been proved to be summable in Section 6.4. It has an exponential tail when ℓ^w is linear.

8. Proof of Theorem 2

Suppose that (τ, p) satisfies the conditions of Theorem 1. Denote by X the unique stationary chain which has been constructed with Algorithm 1 or Algorithm 2 (see Section 9.1). Define the random variable $L_i^X := |c_\tau(X_{-\infty}^{i-1})|$, and define $\sigma := \lceil \ell^w(|w| - 1)/|w| \rceil + 1$. For any integers m and n such that $-\infty < m + \sigma|w| \leq n \leq +\infty$, the visible regeneration time of the window $[m, n]$ is

$$\theta^X[m, n] := \max\{k \leq m : X_k^{k+\sigma|w|-1} = w^\sigma \text{ and } L_i^X \leq i - k, i = k + \sigma|w|, \dots, n\}. \quad (16)$$

Observe that, although $L_i^X := |c_\tau(X_{-\infty}^{i-1})|$, the event $\{\theta^X[m, n] = k\}$ is measurable with respect to the σ -algebra generated by X_k^n . To require that $X_k^{k+\sigma|w|-1} = w^\sigma$ ensures that there exist realizations of X such that $\theta^X[m, n] > -\infty$. To see this, observe that we can concatenate

one more w to these σ consecutive w s without needing to know more than w^σ : for $i = 0, \dots, |w| - 2$,

$$\begin{aligned} |c_\tau(w^{\sigma-1}ww_{-|w|}\cdots w_{-|w|+i})| &= \ell^w(i+1) + |w| + i + 1 \\ &\leq \ell^w(|w| - 1) + |w| + i + 1 \\ &\leq \sigma|w| + i + 1. \end{aligned}$$

We say that the time t is a *visible regeneration time* for the chain X if $\theta^X[t, +\infty] = t$. Finally, define ξ^X and T^X using $\theta^X[t, +\infty]$, in the same way ξ and T using $\theta[t, +\infty]$ were defined. We want to show that (i) $\theta^X[0, +\infty]$ is almost surely finite and (ii) ξ^X is renewal, with finite expected distance between two consecutive 1s.

We use the sequence Z , which tells us where we are sure that w occurs in X . The proof of item (i) is quite similar to the proof of Theorem 1. The main difference is in the following new definitions:

$$L'_i := m_i + |w| + \ell^w(m_i),$$

which is always strictly larger than 0, and, for any $n \geq \sigma|w|$,

$$\theta'[0, n] := \max\{k \leq 0: Z_k^{k+\sigma|w|-1} = w^\sigma \text{ and } L'_i \leq i - k \text{ for } i = k + \sigma|w|, \dots, n\}.$$

We can use the proofs given in Section 6 to show that $\theta'[0, n] \leq \theta^X[0, n]$, that $\Pr(\theta'[0, n] < -l)$ goes to 0 as l diverges for any $n \leq +\infty$, and that $\Pr(\theta'[0] < -l)$ is summable in l . In order to prove item (ii), we can adapt the proof of [7]. Define $v := w^\sigma$, and define, for any $-\infty < m \leq n < +\infty$, the events

$$h[m, n] := \bigcap_{i=0}^{|w|-1} \{|c_\tau(X_m^n w_{-|w|} \cdots w_{-|w|+i})| \leq n + i + 1 - m\},$$

which says that one more w can be concatenated to X_m^n without needing to look back before time m , and

$$H[m, n] := \{X_m^{m+\sigma|w|-1} = v, |c_\tau(X_m^i)| \leq i - m, i \in \{m + \sigma|w|, \dots, n\}\} \cap h[m, n].$$

Both events are measurable with respect to the σ -algebra generated by X_m^n . Finally, define

$$H[m, +\infty] := \{X_m^{m+\sigma|w|-1} = v, |c_\tau(X_m^i)| \leq i - m, i \geq m + \sigma|w|\},$$

which is measurable with respect to the σ -algebra generated by $X_m^{+\infty}$. By definition (16) of $\theta^X[m, n]$, $-\infty < m \leq n \leq +\infty$, we know that if

$$t_1 + \sigma|w| \leq t_2 + \sigma|w| \leq \cdots \leq t_{n-1} + \sigma|w| \leq t_n \tag{17}$$

then

$$\bigcap_{l=1}^n \{\theta^X[t_l, +\infty] = t_l\} = \bigcap_{l=1}^n H[t_l, t_{l+1} - 1], \tag{18}$$

where $t_{n+1} := +\infty$. This is an intersection of independent events. Then, we observe that, by stationarity,

$$\Pr(H[j, +\infty]) = \Pr(H[0, +\infty])$$

and

$$\Pr(H[-j, -1]) = \Pr(H[-j, +\infty] \mid H[0, +\infty]) \quad \text{for all } j \geq \sigma|w|.$$

Together with (18), this yields, for any sequence of integers t_1, \dots, t_n verifying (17),

$$\Pr(\xi_{t_l}^X = 1, l = 1, \dots, n) = \Pr(\xi_0^X = 1) \prod_{l=1}^{n-1} \Pr(\xi_{-(t_{l+1})-t_l}^X = 1 \mid \xi_0^X = 1),$$

and, therefore, the chain ξ^X is renewal, proving the existence of the visible regeneration scheme. By stationarity and definition (16) of the random variable θ^X , we have, for any $m \geq \sigma|w|$,

$$\Pr(T_{l+1}^X - T_l^X \geq m) = \Pr(\theta^X[1, +\infty] \leq -m \mid \theta^X[0, +\infty] = 0) = \Pr(\theta^X[0] < -m + 1),$$

which is summable in m , concluding the proof of Theorem 2.

9. The complete perfect simulation algorithm: simulation and discussion

9.1. The algorithm

Algorithm 1 is ‘simplistic’ in the sense that in order to compute $\theta[m, n]$ it uses the function \mathcal{L} , which is not explicit. A more complete algorithm is given below (Algorithm 2). We recall that, for any $a_m^n \in A^{n-m+1}$ and $u \in [0, 1)$,

$$F(u, a_m^n) := \sum_{a \in A} a \mathbf{1}\{u \in K(a \mid c_\tau(a_m^n))\} + \star \mathbf{1}\{u \in [\alpha, 1), c_\tau(a_m^n) = \emptyset\},$$

where if $m = n + 1$ then $c_\tau(a_m^n) = c_\tau(\emptyset) = \emptyset$. This function contains all the information we need about the probabilistic context tree (τ, p) , and we suppose that it is already implemented in the software used for programming the algorithm.

Algorithm 2. (*‘Explicit’ perfect simulation algorithm of the sample $[X(\mathbf{U})]_m^n$.*)

1. Inputs: m, n , and F . Outputs: $\theta[m, n]$ and $([X(\mathbf{U})]_{\theta[m, n]}, \dots, [X(\mathbf{U})]_n)$.
2. Sample U_m, \dots, U_n uniformly in $[0, 1)$.
3. $i \leftarrow m, B = \{m, \dots, n\}, \theta[m, n] \leftarrow m, [X(\mathbf{U})]_m^n \leftarrow \star^{n-m+1}$
4. **while** $F(U_i, [X(\mathbf{U})]_m^{i-1}) \in A$ and $B \neq \emptyset$ **do**
5. $[X(\mathbf{U})]_i \leftarrow F(U_i, [X(\mathbf{U})]_m^{i-1})$
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** $U_i \in [\alpha, 1)$ **do**
15. $i \leftarrow i - 1$
16. $B \leftarrow B \cup \{i\}$
17. Sample U_i uniformly in $[0, 1)$
18. **end while**
19. $[X(\mathbf{U})]_i \leftarrow F(U_i, \emptyset)$
20. $B \leftarrow B \setminus \{i\}$
21. $t \leftarrow \min B$
22. **while** $F(U_t, [X(\mathbf{U})]_i^{t-1}) \in A$ and $B \neq \emptyset$ **do**
23. $[X(\mathbf{U})]_t \leftarrow F(U_t, [X(\mathbf{U})]_i^{t-1})$
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** $\theta[m, n], ([X(\mathbf{U})]_{\theta[m, n]}, \dots, [X(\mathbf{U})]_n)$

Algorithm 2 uses two variables: i , which is a time index, and B , which is a set of time indices. The set B keeps track of the set of sites which have to be constructed. It is initialized with $B = \{m, \dots, n\}$, which is the set of time indices to be constructed, and the algorithm terminates when $B = \{\emptyset\}$. In the first ‘while’ loop (lines 2 to 8), we sample U_m^n , and directly attempt to construct $[X(\mathbf{U})]_m^n$ using this information. If the algorithm manages to do this, it returns $\theta[m, n] = m$ and the constructed sample $[X(\mathbf{U})]_m^n$. Otherwise, it enters the second ‘while’ loop (lines 10 to 27). 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 uniforms that have been previously generated.

9.2. Simulation

We will say that the algorithm *makes a step* each time it ‘enters’ a ‘while’ loop. In Figure 5, this corresponds to the number of arrows, plus 1. The total number of steps $N[m, n]$ needed for the construction of a sample X_m^n is

$$N[m, n] = (n - m + 1) + 2(m - \theta[m, n]).$$

Let us denote by C the maximum number of operations necessary to make a step. Suppose that we want to construct a sample X_0^{n-1} . Then the expected number of operations is bounded above by

$$C(n + 2 \times \mathbb{E}[\theta[0, n - 1]]).$$

Figure 5 illustrates an explicit perfect simulation of this chain using a finite sample of \mathbf{U} , in the case where $\alpha(1) = \alpha(2) = 0.2$.

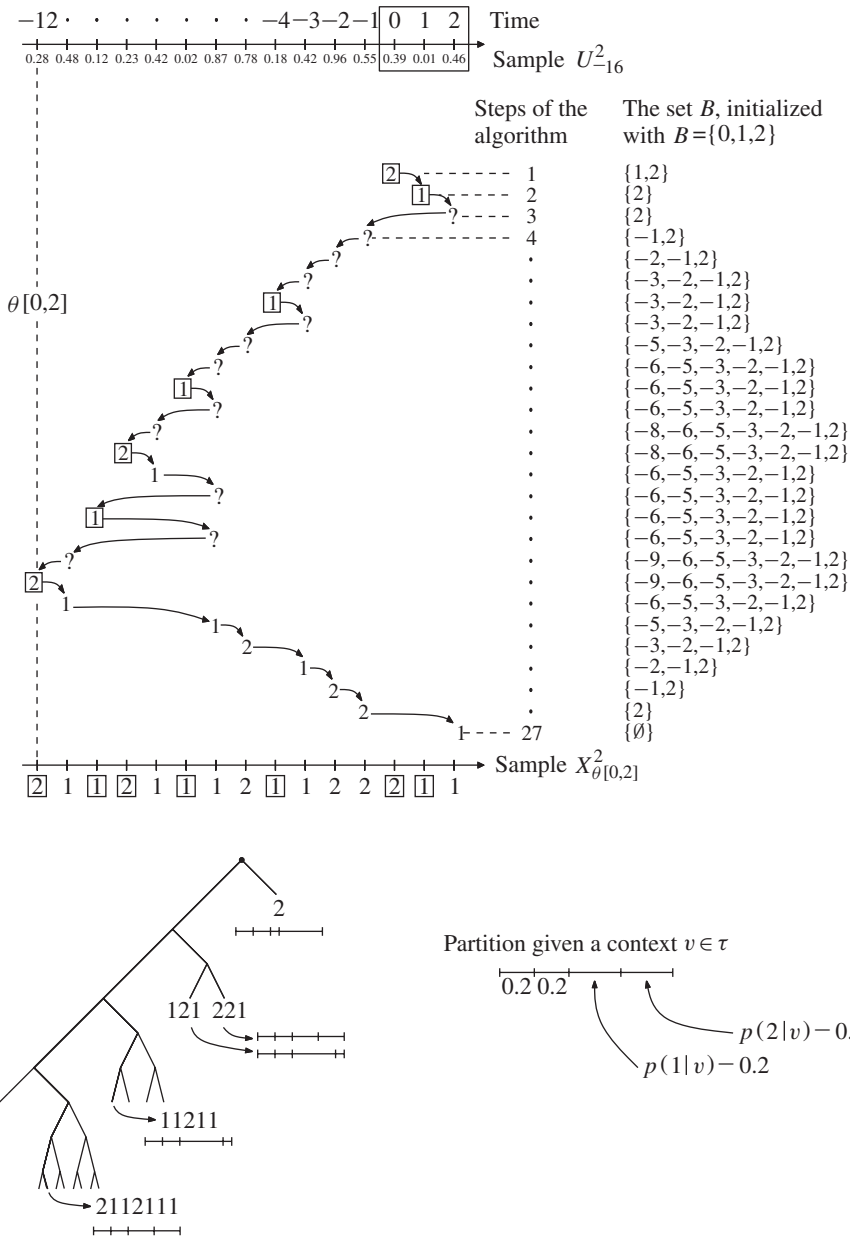


FIGURE 5: An explicit perfect simulation, using a sample U_{-12}^2 . The transition probabilities of the context tree are given in the bottom diagram. To recall what the small intervals represent, we refer the reader to Figure 4. Both symbols 1 and 2 are 0.2-regulars. The probability transitions are $p(2 | 2) = 0.7$, $p(2 | 121) = 0.3$, $p(2 | 122) = 0.5$, $p(2 | 11211) = 0.3$, and $p(2 | 1112112) = 0.5$.

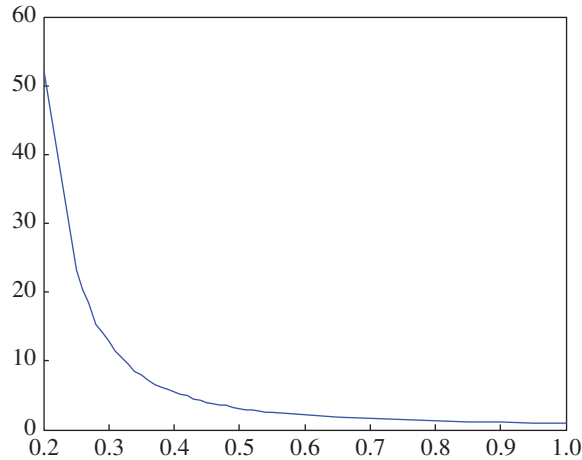


FIGURE 6: Graph representing the influence of the value of ε on the quantity $E|\theta[0]|$. The x -axis represents the successive values of ε , from 0.2 to 1. The y -axis represents $E|\theta_\varepsilon[0]|$, the expected value of $\theta[0]$, when this latter value is computed using ε .

The results of the present paper tell us that, under the conditions of Theorem 1, this expectation is finite. However, no insight is given into how large it can be. This is due to the fact that we did not manage to obtain sufficiently good explicit bounds for the probability of return to 0 of the chain $W^{(0)}$. This is also the case in [7]. However, this should strongly depend on the parameter ε , as in [7] (see Equations (2.4) and (2.5) therein, where their a_0 corresponds basically to our α). In the absence of such bounds, we implemented the above pseudocode in the case of the context tree of Section 3. We assumed that $p(2 | v) = \varepsilon$ for any $v \in \tau$. Note that this case corresponds to the i.i.d. chain on $\{1, 2\}$, where the symbol 2 has probability ε of occurring. This assumption considerably simplifies the implementation of the algorithm and gives us the largest possible regeneration times within the class of probabilistic context trees for which the symbol 2 is ε -regular. We used increasing values of ε , from 0.2 to 1, and, for each value, we took the mean over 10 000 iterations of Algorithm 2. The resulting graph is given in Figure 6. We can derive the corresponding expected number of steps $E N[0]$ realized by the algorithm, and the expected number of operations too.

10. Final comments and references

The first study of chains of infinite order seems to have been by Onicescu and Mihoc [19]. They called these chains '*chaînes à liaison complètes*' (chains with complete connections). Then, in [8], the authors proved the results on the speed of convergence towards the invariant measure under the continuity conditions. Other relevant works include [1], [5], [14], [16], and [18], among others. We refer the reader to [15] for a complete review of this topic, and to [10] for an introduction to the constructive approach.

Chains with variable length were introduced in [21] as a universal model for data compression. This model has been shown to have a great applicability in statistical inference and modeling. For a review and further references, we refer the reader to [13].

On perfect simulation using the CFTP method, we refer the reader to the webpage of Prof. David Bruce Wilson (<http://dbwilson.com/exact/>). The reader will find therein an extensive list of publications in this area.

A very interesting issue related to our work is whether or not the ε -regularity (or the weakly non-nullness) assumption is necessary for the existence of a (not necessarily practical) CFTP algorithm. We mention that necessary conditions exist for Markov chains: in [12], it was shown that such a coupling exists if and only if the Markov chain is geometrically ergodic.

A new and interesting direction of study has been opened in [6]: the interpretation into dynamical system terms (using mapping of the interval) of the probabilistic context tree sources. In [6], the authors considered two particular cases which are also particular cases of the context trees considered in the present work. It seems to be natural to try to extend such results to the new class of processes introduced here.

Acknowledgement

I am thankful to my PhD advisor Prof. A. Galves for his invaluable remarks and help.

References

- [1] BERBEE, H. (1987). Chains with infinite connections: uniqueness and Markov representation. *Prob. Theory Relat. Fields* **76**, 243–253.
- [2] BOROVKOV, A. A. (1998). *Ergodicity and Stability of Stochastic Processes*. John Wiley, Chichester.
- [3] BOWEN, R. (2008). *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms* (Lecture Notes Math. **470**), 2nd revised edn. Springer, Berlin.
- [4] BRAMSON, M. AND KALIKOW, S. (1993). Nonuniqueness in g -functions. *Israel J. Math.* **84**, 153–160.
- [5] BRESSAUD, X., FERNÁNDEZ, R. AND GALVES, A. (1999). Decay of correlations for non-Hölderian dynamics. A coupling approach. *Electron. J. Probab.* **4**, 19pp.
- [6] CÉNAC, P., CHAUVIN, B., PACCAUT, F. AND POUYANNE, N. (2010). Variable length markov chains and dynamical sources. Preprint, Université de Bourgogne.
- [7] COMETS, F., FERNÁNDEZ, R. AND FERRARI, P. A. (2002). Processes with long memory: regenerative construction and perfect simulation. *Ann. Appl. Probab.* **12**, 921–943.
- [8] DOEBLIN, W. AND FORTET, R. (1937). Sur des chaînes à liaisons complètes. *Bull. Soc. Math. France* **65**, 132–148.
- [9] FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications*, Vol. I, 3rd edn. John Wiley, New York.
- [10] FERNÁNDEZ, R., FERRARI, P. A. AND GALVES, A. (2001). *Coupling, Renewal and Perfect Simulations of Chains of Infinite Order*. Lecture Notes for the Vth Brazilian School of Probability.
- [11] FOSS, S. AND KONSTANTOPOULOS, T. (2003). Extended renovation theory and limit theorems for stochastic ordered graphs. *Markov Process. Relat. Fields* **9**, 413–468.
- [12] FOSS, S. G. AND TWEEDIE, R. L. (1998). Perfect simulation and backward coupling. *Commun. Statist. Stoch. Models* **14**, 187–203.
- [13] GALVES, A. AND LÖCHERBARCH, E. (2008). Stochastic chains with memory of variable length. *TICSP series* **38**, 117–133.
- [14] HARRIS, T. E. (1955). On chains of infinite order. *Pacific J. Math.* **5**, 707–724.
- [15] IOSIFESCU, M. AND GRIGORESCU, Ș. (1990). *Dependence with Complete Connections and Its Applications* (Camb. Tracts Math. **96**). Cambridge University Press.
- [16] JOHANSSON, A. AND ÖBERG, A. (2003). Square summability of variations of g -functions and uniqueness of g -measures. *Math. Res. Lett.* **10**, 587–601.
- [17] KALIKOW, S. (1990). Random Markov processes and uniform martingales. *Israel J. Math.* **71**, 33–54.
- [18] LALLEY, S. P. (1986). Regenerative representation for one-dimensional Gibbs states. *Ann. Probab.* **14**, 1262–1271.
- [19] ONICESCU, O. AND MIHOC, G. (1935). Sur les chaînes de variables statistiques. *Bull. Sci. Math* **59**, 174–192.
- [20] PROPP, J. G. AND WILSON, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms* **9**, 223–252.
- [21] RISSANEN, J. (1983). A universal data compression system. *IEEE Trans. Inform. Theory* **29**, 656–664.