

BACKWARD COALESCENCE TIMES FOR PERFECT SIMULATION OF CHAINS WITH INFINITE MEMORY

EMILIO DE SANTIS* ** AND

MAURO PICCIONI,* *** *Sapienza Università di Roma*

Abstract

This paper is devoted to the perfect simulation of a stationary process with an at most countable state space. The process is specified through a kernel, prescribing the probability of the next state conditional to the whole past history. We follow the seminal work of Comets, Fernández and Ferrari (2002), who gave sufficient conditions for the construction of a perfect simulation algorithm. We define *backward coalescence times* for these kind of processes, which allow us to construct perfect simulation algorithms under weaker conditions than in Comets, Fernández and Ferrari (2002). We discuss how to construct backward coalescence times (i) by means of *information depths*, taking into account some a priori knowledge about the histories that occur; and (ii) by identifying suitable *coalescing events*.

Keywords: Perfect simulation; coupling; chains with complete connections

2010 Mathematics Subject Classification: Primary 60G99; 68U20; 60J10

1. Introduction and main definitions

Perfect simulation algorithms for stochastic processes have been developed mostly for Markov chains, starting from the original coupling-from-the-past algorithm presented in the founding paper [14]. Later, the fundamental role of the so-called stochastic recursive sequences for perfect simulation was recognized [8]. In [13] a stochastic recursive sequence for perfect simulation was constructed, called the gamma coupler, for transition kernels satisfying a minorization condition. We also mention the area of perfect simulation devoted to spatial contexts, such as stochastic geometry [6], [12] and random fields [3], [11].

In [2] the ideas of coupling from the past have been extended to processes constructed through a transition kernel depending on the whole past, leading to the construction of a perfect simulation algorithm. A somewhat more general presentation was given in [5]. We need a greater generality to cover a number of different transition kernels. We also mention the paper [7], which does not refer to a kernel, but only to a process and its ‘shifted’ versions, the shift being associated to a ‘stationary environment’ modeled in the probability space.

Throughout the paper, we consider stochastic processes defined on \mathbb{Z} , with values in a finite or countable alphabet G . The law of the process is specified through a transition kernel, prescribing the probability that each individual letter of the alphabet occurs, conditional to the whole history preceding it.

Received 22 February 2011; revision received 13 September 2011.

* Postal address: Dipartimento di Matematica, Sapienza Università di Roma, Piazzale Aldo Moro 2, 00185 Roma, Italy.

** Email address: desantis@mat.uniroma1.it

*** Email address: piccioni@mat.uniroma1.it

For notational convenience, we use the following conventions: for $m, n \in \mathbb{Z}$ with $m \leq n$, we write *words* as

$$s_m^n = (s_n, \dots, s_m) \in G^{n-m+1}.$$

We also allow $m = -\infty$ and $n = +\infty$, and we write $s = s_{-\infty}^{+\infty}$. For $m = -\infty$ and finite n , a word will be called a *history*. Histories are identified with elements of $G^{-\mathbb{N}_+}$, where \mathbb{N}_+ is the set of positive integers. We can concatenate any finite word s_m^n with a history $w_{-\infty}^{m-1}$, obtaining another history

$$(s_m^n, w_{-\infty}^{m-1}) = (s_n, \dots, s_m, w_{m-1}, w_{m-2}, \dots).$$

The same notational conventions are used for sequences with values in the interval $[0, 1]$. The set $G^{-\mathbb{N}_+}$ is equipped with the product topology and the product σ -algebra.

Definition 1. A (transition) kernel is a function $p: G \times G^{-\mathbb{N}_+} \rightarrow [0, 1]$, whose values are denoted by $p(g|w_{-\infty}^{-1})$, such that, for any $g \in G$, $p(g|\cdot)$ is a measurable function in $G^{-\mathbb{N}_+}$ such that $\sum_{g \in G} p(g|w_{-\infty}^{-1}) = 1$. A process $Y = \{Y_n, n \in \mathbb{Z}\}$ on a probability space (Ω, \mathcal{F}, P) is compatible with the kernel p if, for any $m \in \mathbb{Z}$ and $g \in G$,

$$P(Y_m = g \mid Y_{m-i}, i \in \mathbb{N}_+) = p(g|Y_{-\infty}^{m-1}) \quad \text{almost surely (a.s.).} \tag{1}$$

Starting from a kernel p , perfect simulation constructs algorithms for sampling a compatible process Y , under conditions that guarantee its uniqueness. The main ingredients to construct a perfect simulation algorithm are given below.

Definition 2. A set of *admissible histories* is a measurable subset \mathcal{H} of $G^{-\mathbb{N}_+}$ with the following properties.

Remove invariance: $w_{-\infty}^{-1} \in \mathcal{H}$ implies that $w_{-\infty}^{-2} \in \mathcal{H}$.

Add invariance: $w_{-\infty}^{-1} \in \mathcal{H}$ and $g \in G$ such that $p(g|w_{-\infty}^{-1}) > 0$ implies that $(g, w_{-\infty}^{-1}) \in \mathcal{H}$.

The simplest nontrivial set of admissible histories can be defined in the following way. A letter g of the alphabet G is *forbidden* if $p(g|w_{-\infty}^{-1}) = 0$ for any $w_{-\infty}^{-1} \in G^{-\mathbb{N}_*}$ (in this case it could be directly excluded from G). A word $s_{-n}^0 = (s_0, s_{-n}^{-1})$ is *forbidden* if either s_{-n}^{-1} is forbidden or

$$p(s_0|w_{-\infty}^{-1}) = 0 \quad \text{for any } w_{-\infty}^{-1} \in G^{-\mathbb{N}_+} \text{ such that } w_{-n}^{-1} = s_{-n}^{-1}.$$

For any $n \in \mathbb{N}_+$, let \mathcal{A}_n be the set of words of length n that are not forbidden. The set

$$\mathcal{H}^* = \{w_{-\infty}^{-1} \in G^{-\mathbb{N}_+} : w_{-n}^{-1} \in \mathcal{A}_n, n \in \mathbb{N}_+\} \tag{2}$$

is a set of admissible histories, as it is easily checked that the required add and remove invariance properties hold.

In practical applications there is usually a natural choice for the set of admissible histories. The choice of admissible histories excludes irrelevant histories from the domain of the coupling function, which we are going to define.

Definition 3. Let \mathcal{H} be a set of admissible histories. A *coupling function* f for the kernel p is a function $f: [0, 1] \times \mathcal{H} \rightarrow G$ of the form

$$f(u|w_{-\infty}^{-1}) = \sum_{g \in G} g \sum_{n \in \mathbb{N}} \mathbf{1}_{I_n(g|w_{-n}^{-1})}(u), \quad w_{-\infty}^{-1} \in \mathcal{H}. \tag{3}$$

Here, for each $w_{-\infty}^{-1} \in \mathcal{H}$, $\{I_n(g|w_{-\infty}^{-1}), g \in G, n \in \mathbb{N}\}$ is a partition of $[0, 1)$ and each $I_n(g|w_{-\infty}^{-1})$ is either empty or it is a disjoint union of intervals (closed to the left, open to the right), with $\sum_n |I_n(g|w_{-\infty}^{-1})| = p(g|w_{-\infty}^{-1})$ (here $|\cdot|$ stands for the Lebesgue measure in $[0, 1)$).

By definition, if f is a coupling function for p and U is a random variable uniformly distributed in $[0, 1)$, then

$$P(f(U|w_{-\infty}^{-1}) = g) = p(g|w_{-\infty}^{-1}) \quad \text{for each } g \in G, w_{-\infty}^{-1} \in \mathcal{H}.$$

Note that the assumptions imply that f is jointly measurable and, for any $u \in [0, 1)$, the function $f(u|\cdot)$ is continuous on \mathcal{H} , since it depends on a finite portion of the history. Moreover, the add invariance property of \mathcal{H} and Definition 3 imply that

$$w_{-\infty}^{-1} \in \mathcal{H} \text{ and } u \in [0, 1) \implies (f(u|w_{-\infty}^{-1}), w_{-\infty}^{-1}) \in \mathcal{H}.$$

Coupling functions of this form appeared in [5]. In that paper (as in [2]) only the simplest choice $\mathcal{H} = G^{-\mathbb{N}^+}$ was considered.

For the construction of a perfect simulation algorithm, we require that the coupling function f admits a backward coalescence time, which we now define. Starting from $f^{(1)} = f$, we recursively define $f^{(n)} : [0, 1)^n \times \mathcal{H} \rightarrow G$ for any $n \in \mathbb{N}$, in the following natural way:

$$f^{(n+1)}(u_0^n|w_{-\infty}^{-1}) := f(u_n|f^{(n)}(u_0^{n-1}|w_{-\infty}^{-1}), \dots, f^{(1)}(u_0|w_{-\infty}^{-1}), w_{-\infty}^{-1}). \tag{4}$$

The source of randomness for the construction of the simulation process is a sequence $U = \{U_i, i \in \mathbb{Z}\}$ of independent and identically distributed (i.i.d.) random variables, with uniform distribution in the interval $[0, 1)$. For any $m, n \in \mathbb{Z}$ with $m \leq n$, we define the σ -algebra $\mathcal{F}_m^n = \sigma(U_i, i = m, \dots, n)$. For ease of notation, we specify an arbitrary reference admissible history $g_{-\infty}^{-1} \in \mathcal{H}$.

Definition 4. We say that a measurable function $\tau_0(U_{-\infty}^0)$ with nonpositive values is a backward coalescence time if it satisfies the following properties:

- (H1) $-\tau_0$ is an a.s. finite stopping time with respect to the filtration $\{\mathcal{F}_{-n}^0 : n \in \mathbb{N}\}$, i.e. $\{\tau_0 = -l\} \in \mathcal{F}_{-l}^0$ for any $l \in \mathbb{N}$;
- (H2) if $\tau_0 = -l$ then, for any $w_{-\infty}^{-(l+1)} \in \mathcal{H}$,

$$f^{(l+1)}(U_{-l}^0|w_{-\infty}^{-(l+1)}) = f^{(l+1)}(U_{-l}^0|g_{-\infty}^{-(l+1)}). \tag{5}$$

This definition is interpreted as follows: by simulating backward in time the sequence (U_0, U_{-1}, \dots) , an instant τ_0 is identified with the property that the value of the coupling function at time 0 does not depend on samples of U at times prior than τ_0 . Obviously, property (5) remains true for any $l > -\tau_0$.

Under suitable conditions on \mathcal{H} , the a.s. finiteness of a backward coalescence time τ_0 allows us to construct a process which is compatible with the kernel. First, define the shifted backward coalescence times

$$\tau_n(U_{-\infty}^n) = n + \tau_0(U_{-\infty}^n) \quad \text{for all } n \in \mathbb{Z}.$$

If τ_0 is a.s. finite, by translation invariance, τ_n is also a.s. finite. Likewise, for any $m < n \in \mathbb{N}$,

$$\tau[m, n] = \inf\{\tau_i : i \in [m, n]\} \tag{6}$$

is a.s. finite. Finally, define the process $X = \{X_n, n \in \mathbb{Z}\}$ as

$$X_n(U_{-\infty}^n) = f^{(l+1)}(U_{n-l}^n | \mathbf{g}_{-\infty}^{n-l-1}) \quad \text{if } \tau_n(U_{-\infty}^n) = n - l,$$

which is stationary by construction.

The following proposition explains our interest in the process X . Its proof is deferred to Appendix A.

Proposition 1. *Let f be a coupling function for the kernel p and the set of admissible histories \mathcal{H} . If*

- (i) τ_0 is a backward coalescence time for f ;
- (ii) $X_{-\infty}^0 \in \mathcal{H}$ with probability 1,

then X is the unique (in law) process compatible with the kernel p verifying property (ii).

The concept of backward coalescence time is similar to that of weak verifiable time presented in Definition 4.1 of [7]. Corollary 2.2 therein is somewhat analogous to our Proposition 1.

Remark 1. It is clear that uniqueness holds without further constraints whenever it can be shown that any compatible process Y has the property that $Y_{-\infty}^0 \in \mathcal{H}$ with probability 1, as will be the case in all the examples that will be presented. This property is automatically satisfied by the set \mathcal{H}^* defined in (2), which also fulfills condition (ii) of Proposition 1. In this case, to apply Proposition 1, it is enough to prove the existence of a backward coalescence time.

If the assumptions of Proposition 1 hold, the process X can be *perfectly simulated* in any finite window $[m, n]$, with $m \leq n \in \mathbb{N}$, starting from time $\tau[m, n]$. By construction, $m - \tau[m, n]$ is a stopping time with respect to the filtration $\{\mathcal{F}_{m-k}^n, k \in \mathbb{N}\}$. This implies that a *perfect simulation algorithm* can be realized in the following way.

- (i) Inspect the U_i s backward from U_n , until $\tau[m, n]$ is located.
- (ii) Compute the X_i s forward, starting from $\tau[m, n]$, using the coupling function.

In (6) we also allow $n = +\infty$. If $\tau[0, +\infty]$ is a.s. finite, the perfect simulability of X on any finite window is guaranteed.

The main idea in [5] is to construct a backward coalescence time for a coupling function through the computation of an ‘instantaneous information depth’ (called a ‘random order’ in that paper). This is a bound on the number of past values of the process needed to compute the current one, taking into account *only* the current value of the process U (the *input*). Here we give a more general definition, in which the previous input values can also be taken into account.

First note that the U_{-h}^{-1} observation constrains the set of possible histories to the random subset

$$M_h(U_{-h}^{-1}) = \{(f^{(h)}(U_{-h}^{-1} | \mathbf{w}_{-\infty}^{-h-1}), \dots, f(U_{-h} | \mathbf{w}_{-\infty}^{-h-1}), \mathbf{w}_{-\infty}^{-h-1}) : \mathbf{w}_{-\infty}^{-h-1} \in \mathcal{H}\}$$

for $h \in \mathbb{N}_+$. For consistency, we also define $M_0 = \mathcal{H}$. Then we can give the following definition.

Definition 5. A general information depth $L(U_{-\infty}^0)$ for a coupling function f is a stopping time with respect to the filtration $\{\mathcal{F}_{-n}^0, n \in \mathbb{N}\}$ such that, for any $h \in \mathbb{N}$,

$$L(U_{-\infty}^0) = h \implies \text{for all } \mathbf{w}_{-\infty}^{-1} \in M_h(U_{-h}^{-1}), \text{ there exist } n \leq h, g \in G \text{ such that } U_0 \in I_n(g|\mathbf{w}_{-n}^{-1}). \tag{7}$$

The information depth is *instantaneous* when $L(U_{-\infty}^{-1}) = K(U_0)$ for some measurable K .

Note that property (7) continues to hold for any $h \geq L(U_{-\infty}^0)$, the sequence of subsets $\{M_h(U_{-h}^{-1}), h \in \mathbb{N}_+\}$ being nonincreasing. Since $M_0 = \mathcal{H}$, the event $\{L = 0\}$ means that $f(u_0|\mathbf{w}_{-\infty}^{-1})$ does not depend at all on the history $\mathbf{w}_{-\infty}^{-1} \in \mathcal{H}$.

Given an information depth L for a coupling function f , we define the whole sequence of *shifted information depths*

$$\mathbf{L} = \{L_j = L(U_{-\infty}^j), j \in \mathbb{Z}\}.$$

Finally, given a sequence of shifted information depths \mathbf{L} , define

$$\tau_0^L(U_{-\infty}^0) = \sup\{s \leq 0 : L_j \leq j - s, s \leq j \leq 0\}.$$

The following proposition clarifies our interest in τ_0^L ; for the reader's convenience, its simple proof is given in Appendix A.

Proposition 2. *If τ_0^L is a.s. finite, it is a backward coalescence time.*

If K is an instantaneous information depth then the sequence of shifted information depths is obviously i.i.d. The main result in [2] concerns this case. It can be rephrased in the following way.

Proposition 3. *Let $\mathbf{K} = \{K_i, i \in \mathbb{Z}\}$ be a sequence of i.i.d. information depths.*

- (i) *If $\sum_{k=0}^{\infty} \prod_{j=0}^k P(K_0 \leq j) = \infty$ then $\tau_0^K > -\infty$ a.s.*
- (ii) *If $\prod_{j=0}^{\infty} P(K_0 \leq j) > 0$ then $\tau^K[0, +\infty) > -\infty$ a.s.*

The above result suggests constructing a coupling function with the ‘smallest’ possible distribution of an instantaneous information depth K_0 . In fact, in [2] such a kind of coupling function was defined, which in the sequel will be called a *canonical coupling function*. This is certainly the best choice within the instantaneous framework. Its definition will be reviewed in the next section.

As seen from the conditions of Proposition 3, this framework necessarily requires two kinds of condition: (a) K_0 has to be finite, and, furthermore, the tail of its distribution function has to fall quickly to 0; (b) $P(K_0 = 0) > 0$.

In Section 2 we first discuss how to relax condition (a) by using information depths for the canonical coupling function which are not instantaneous. Instantaneous information depths generate i.i.d. sequences, but represent only worst case bounds. However, if some past values of the states can be located, much less information on past inputs could be required. Exploiting this idea, we establish some general results, namely Theorem 1 and Corollary 1, and we discuss some applications in which the instantaneous framework does not work. Starting from similar considerations, in [9] and [10] different coupling functions were proposed, based on probabilistic context trees.

In Section 3 we deal with condition (b). If $\inf\{p(g|\mathbf{w}_{-\infty}^{-1}) : \mathbf{w}_{-\infty}^{-1} \in \mathcal{H}\} = 0$ for any $g \in G$ then $K_0 = 0$ with probability 1. Therefore, there is no hope of defining a backward

coalescence time through some information depth and we have to consider more general backward coalescence times. The problem is already present for ergodic Markov chains with zeros in the transition matrix, but it does not prevent the existence of backward coalescence times. More generally, we assume that $K_0 < r$ with probability 0, but $K_0 = r$ with positive probability. In this case a backward coalescence time can be defined if a suitable ‘coalescing event’ can be identified and the asymptotic behavior of the distribution of K_0 verifies the conditions of Proposition 3. As will be shown by an example, the identification of a coalescing event may require the modification of the canonical coupling function. When $r = 1$, under suitable assumptions, these ideas suggest combining the algorithm in [2] for long-memory processes with Markovian coupling-from-the-past algorithms, as in [14].

2. Information depths depending on the whole past

In this section we work with the canonical coupling functions introduced in [2], with the natural adjustment that allows choices of the set of admissible histories \mathcal{H} to be different from $G^{-\mathbb{N}_+}$.

Let us first define some relevant quantities needed for the construction, namely,

$$a_0(g) = \inf\{p(g|z_{-\infty}^{-1}) : z_{-\infty}^{-1} \in \mathcal{H}\}, \quad g \in G, \tag{8}$$

$$a_k(g|\mathbf{w}_{-k}^{-1}) = \inf\{p(g|z_{-\infty}^{-1}) : z_{-\infty}^{-1} \in \mathcal{H}, z_{-k}^{-1} = \mathbf{w}_{-k}^{-1}\}, \quad g \in G, \mathbf{w}_{-\infty}^{-1} \in \mathcal{H}. \tag{9}$$

It is easily seen that, for any $g \in G$, the function $p(g|\cdot)$ is continuous in $\mathbf{w}_{-\infty}^{-1} \in \mathcal{H}$ if and only if

$$a_n(g|\mathbf{w}_{-n}^{-1}) \uparrow p(g|\mathbf{w}_{-\infty}^{-1}). \tag{10}$$

Defining, for any $\mathbf{w}_{-\infty}^{-1} \in \mathcal{H}$ and $k \in \mathbb{N}$,

$$a_k(\mathbf{w}_{-k}^{-1}) = \sum_{g \in G} a_k(g|\mathbf{w}_{-k}^{-1}), \tag{11}$$

we observe that (10) is equivalent to $a_k(\mathbf{w}_{-k}^{-1}) \uparrow 1$.

A canonical coupling function f is specified through the sets $I_n(g|\mathbf{w}_{-n}^{-1})$ appearing in (3). Starting from $a_{-1} = 0$, for any $k \in \mathbb{N}$, the intervals $[a_{k-1}(\mathbf{w}_{-k+1}^{-1}), a_k(\mathbf{w}_{-k}^{-1})]$ are partitioned into the subintervals $I_k(g|\mathbf{w}_{-k}^{-1})$ of total length $a_k(g|\mathbf{w}_{-k}^{-1}) - a_{k-1}(g|\mathbf{w}_{-k+1}^{-1})$ for $g \in G$ (if this length is 0 then $I_k(g|\mathbf{w}_{-k}^{-1})$ is empty). If (10) is satisfied, a function f of this form is a coupling function for the kernel p .

Next define $a_0 = \sum_{g \in G} a_0(g)$ and

$$a_k = \inf\{a_k(\mathbf{w}_{-k}^{-1}) : \mathbf{w}_{-\infty}^{-1} \in \mathcal{H}\}, \quad k \in \mathbb{N}_+.$$

The sequence $\{a_k\}$ is nondecreasing, so we let $a_\infty = \lim_k a_k$. Since $a_k(\mathbf{w}_{-k}^{-1}) \geq a_k$ for any $\mathbf{w}_{-\infty}^{-1} \in \mathcal{H}$, the random variable

$$K(u) = \inf\{k \in \mathbb{N} : a_k > u\} = \sum_{k=0}^{\infty} k \mathbf{1}_{[a_{k-1}, a_k)}(u) \tag{12}$$

is an instantaneous information depth with $P(K(U) \leq k) = a_k$ for any $k \in \mathbb{N}$, U being uniformly distributed in $[0, 1)$. As observed in [5], it has the minimality property cited in the introduction, once the set of histories \mathcal{H} has been specified. The a.s. finiteness of $K(U)$ is

equivalent to the fact that $a_\infty = 1$, and it is clearly a necessary condition for the a.s. finiteness of any instantaneous information depth for a specific choice of \mathcal{H} .

We now use an information depth L for a canonical coupling function, which is not instantaneous, defined by

$$J_h(\mathbf{U}_{-h}^{-1}) = \{\mathbf{w}_{-\infty}^{-1} \in \mathcal{H} : w_{-k} = g \text{ if } u_{-k} \in I_0(g), k \leq h\}, \quad h \in \mathbb{N}_+.$$

We set $J_0 = \mathcal{H}$, and observe that $M_h(\mathbf{U}_{-h}^{-1}) \subset J_h(\mathbf{U}_{-h}^{-1})$ for any $h \in \mathbb{N}$. Then

$$L(\mathbf{U}_{-\infty}^0) = \inf\{h \in \mathbb{N} : U_0 < a_h(\mathbf{w}_{-h}^{-1}) \text{ for all } \mathbf{w}_{-\infty}^{-1} \in J_h(\mathbf{U}_{-h}^{-1})\} \tag{13}$$

is an information depth. Out of a null set of $\mathbf{U}_{-\infty}^0$ s, (13) is equal to

$$L(\mathbf{U}_{-\infty}^0) = \inf\{h \in \mathbb{N} : U_0 < A_h(\mathbf{U}_{-h}^{-1})\}, \tag{14}$$

where $A_0 = a_0$ and

$$A_h(\mathbf{U}_{-h}^{-1}) = \inf\{a_h(\mathbf{w}_{-h}^{-1}) : \mathbf{w}_{-\infty}^{-1} \in J_h(\mathbf{U}_{-h}^{-1})\}, \quad h \in \mathbb{N}_+.$$

It is clear that L is finite a.s. if and only if $\lim_h A_h(\mathbf{U}_{-h}^{-1}) = 1$ a.s.

We now give a sufficient condition for the a.s. finiteness of $\tau^L[0, +\infty]$.

Theorem 1. *If*

$$\mathbb{E}\left(\prod_{h=0}^{\infty} \frac{1}{A_h(\mathbf{U}_0^{h-1})}\right) < \infty \tag{15}$$

then $\tau^L[0, +\infty] > -\infty$ a.s.

Proof. The sequence

$$Y_n = \frac{\prod_{h=0}^n \mathbf{1}_{\{U_h < A_h(\mathbf{U}_0^{h-1})\}}}{\prod_{h=0}^n A_h(\mathbf{U}_0^{h-1})}, \quad n \in \mathbb{N},$$

is uniformly integrable, because it is bounded by the integrable random variable

$$\left\{ \prod_{h=0}^{\infty} A_h(\mathbf{U}_0^{h-1}) \right\}^{-1}.$$

Moreover, it is a martingale with respect to the filtration $\{\mathcal{F}_0^n\}$. In fact, since U_n is independent of \mathcal{F}_0^{n-1} , we have

$$\mathbb{E}(\mathbf{1}_{\{U_n < A_n(\mathbf{U}_0^{n-1})\}} \mid \mathcal{F}_0^{n-1}) = A_n(\mathbf{U}_0^{n-1});$$

hence,

$$\mathbb{E}(Y_n \mid \mathcal{F}_0^{n-1}) = \frac{\prod_{h=0}^{n-1} \mathbf{1}_{\{U_h < A_h(\mathbf{U}_0^{h-1})\}}}{\prod_{h=0}^n A_h(\mathbf{U}_0^{h-1})} \mathbb{E}(\mathbf{1}_{\{U_n < A_n(\mathbf{U}_0^{n-1})\}} \mid \mathcal{F}_0^{n-1}) = Y_{n-1}.$$

Furthermore, since $Y_0 = a_0^{-1} \mathbf{1}_{\{U_0 < a_0\}}$, it follows that $\mathbb{E}(Y_n) = \mathbb{E}(Y_0) = 1$.

Since $\{Y_n\}$ is uniformly integrable, from a well-known result (see [15, p. 134]), the limit Y_∞ is finite a.s. with $E(Y_\infty) = 1$. As a consequence, $Y_\infty > 0$ with positive probability. But, clearly,

$$\{Y_\infty > 0\} = \bigcap_{h=0}^\infty \{U_h < A_h(\mathbf{U}_0^{h-1})\} = \{L_h \leq h, h \in \mathbb{N}\}.$$

From (6), observe that

$$\tau^L[0, +\infty] = \sup\{s \leq 0: L_j \leq j - s, s \leq j\},$$

so $\{\tau^L[0, +\infty] = 0\}$ coincides with the event $\{Y_\infty > 0\}$, which has positive probability. Therefore, $P(\tau^L[0, +\infty] = 0) > 0$. We end the proof by observing that in this case $\tau^L[0, +\infty]$ is a.s. finite. To this end, define the sequence of events

$$R_m = \{\tau^L[m, +\infty] = m\}, \quad m \in -\mathbb{N}_+.$$

Since $\mathbf{1}_{R_m}(U) = \mathbf{1}_{R_0}(U_m^{+\infty})$, by using the ergodic theorem, we deduce that there exists an a.s. finite $Y_0 \leq 0$ that realizes R_{Y_0} . Since $\tau^L[0, \infty] \geq \tau^L[Y_0, \infty] = Y_0$, the former is a.s. finite.

In the following corollary we present a sufficient condition for the application of Theorem 1, which will be used throughout the examples presented in this section.

Corollary 1. *For any $h \in \mathbb{N}_+$, let C_h be a Borel subset of $[0, 1)^h$ with the following properties:*

(i) *the sequence*

$$\xi_h = \inf\{A_h(\mathbf{u}_0^{h-1}): \mathbf{u}_0^{h-1} \in C_h\} \tag{16}$$

is such that $\prod_{h=0}^\infty \xi_h > 0$ (in particular, $\xi_0 = a_0 > 0$);

(ii) *the random variable*

$$N = N(U_0^{+\infty}) = \inf\{m: U_0^{n-1} \in C_n \text{ for all } n \geq m\} \tag{17}$$

has probability generating function $E(s^N) < \infty$ for some $s > 1/a_\infty$.

Then $\tau^L[0, +\infty] > -\infty$ a.s.

Proof. By the definition of N , $A_h \geq \xi_h > 0$ for $h \geq N$. Moreover, since $A_h \geq a_h$ for each integer h , we have

$$\frac{1}{\prod_{h=0}^\infty A_h} \leq \frac{1}{\prod_{h=0}^N a_h} \frac{1}{\prod_{h=N+1}^\infty \xi_h} \leq \frac{1}{\prod_{h=0}^N a_h} \frac{1}{\prod_{h=0}^\infty \xi_h}.$$

By taking the expectation of both sides, we obtain

$$E\left(\frac{1}{\prod_{h=0}^\infty A_h}\right) \leq E\left(\frac{1}{\prod_{h=0}^N a_h}\right) \frac{1}{\prod_{h=0}^\infty \xi_h}.$$

The second factor on the right-hand side is finite by property (i). Now choose s as in (ii): by assumption, there exists an integer k such that $1/a_h < s$ for $h \geq k$. Consequently, we have the bound

$$E\left(\frac{1}{\prod_{h=0}^N a_h}\right) \leq \frac{1}{\prod_{h=0}^{k-1} a_h} E(s^N) < +\infty,$$

from which (15) follows. Application of Theorem 1 completes the proof.

We now present a first class of models that satisfy the conditions of Corollary 1, but not those of Proposition 3.

Example 1. Consider the alphabet $G = \{-1, 1\}$, with kernel

$$p(1|\mathbf{w}_{-\infty}^{-1}) = \frac{1}{2} + \sum_{k=1}^{\infty} \theta_k \exp(-\beta S_k(\mathbf{w}_{-k}^{-1}))w_{-k}, \quad \mathbf{w}_{-\infty}^{-1} \in G^{\mathbb{N}^+} = \mathcal{H},$$

where $\sum_{k=1}^{\infty} |\theta_k| < \frac{1}{2}$, $\beta > 0$, and

$$S_k(\mathbf{w}_{-k}^{-1}) = \sum_{i=1}^{k-1} \delta_{w_{-i}w_{-i-1}, -1}$$

counts the number of sign changes in the history within the interval $[-k, -1]$. Each of these changes reduces the dependence of the current state from the values prior to the change. We can immediately verify that

$$a_0(-1), \quad a_0(1) \geq \frac{1}{2} - \sum_{k=1}^{\infty} |\theta_k| > 0.$$

In the definition of the canonical coupling function, set $I_0(-1) = [0, a_0(-1))$ and $I_0(1) = [a_0(-1), a_0(-1) + a_0(1))$. Next, for any $n \in \mathbb{N}_+$, define

$$C_{2n} = \left\{ \mathbf{u}_0^{2n-1} \in [0, 1)^{2n} : \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_{I_0(-1)}(u_{2k}) \mathbf{1}_{I_0(1)}(u_{2k+1}) > \sigma \right\},$$

where $\sigma < a_0(-1)a_0(1)$; moreover, define $C_{2n+1} = C_{2n} \times [0, 1)$. From Chernoff's bound, there exists $K > 0$ such that

$$P(\mathbf{U}_0^{n-1} \notin C_n) \leq e^{-Kn};$$

therefore, defining N as in (17),

$$P(N \leq m) = P\left(\bigcap_{n=m}^{\infty} \{\mathbf{U}_0^{n-1} \in C_n\}\right) \geq 1 - \sum_{n=m}^{\infty} e^{-Kn} = 1 - \frac{e^{-Km}}{1 - e^{-K}},$$

from which we deduce the existence of the probability generating function of N for some $s > 1$. Moreover,

$$\begin{aligned} a_k(\mathbf{w}_{-k}^{-1}) &= 1 - \sup_{\mathbf{v}_{-h}^{-k-1}} \sum_{h=k+1}^{\infty} \theta_h \exp(-\beta S_h(\mathbf{w}_{-k}^{-1}, \mathbf{v}_{-h}^{-k-1}))v_h \\ &\quad + \inf_{\mathbf{z}_{-h}^{-k-1}} \sum_{h=k+1}^{\infty} \theta_h \exp(-\beta S_h(\mathbf{w}_{-k}^{-1}, \mathbf{z}_{-h}^{-k-1}))z_h, \end{aligned}$$

from which

$$a_k = \inf_{\mathbf{w}_{-\infty}^{-1} \in \mathcal{H}} a_k(\mathbf{w}_{-k}^{-1}) \geq 1 - 2 \sup_{\mathbf{w}_{-\infty}^{-1} \in \mathcal{H}} \exp(-\beta S_k(\mathbf{w}_{-k}^{-1})) \sum_{h=k+1}^{\infty} |\theta_h| \geq 1 - 2 \sum_{h=k+1}^{\infty} |\theta_h|,$$

so, the θ_k s being summable, we conclude that $a_\infty = 1$, which proves that Corollary 1(ii) holds. On the other hand, observe that, whenever $U_0^{2k-1} \in C_{2k}$,

$$A_{2k}(U_0^{2k-1}) \geq 1 - 2e^{-\beta\sigma k} \sum_{l>2k} |\theta_l| \geq 1 - e^{-\beta\sigma k};$$

therefore, by definition (16), we obtain $\xi_{2k} \geq 1 - e^{-\beta\sigma k}$, and, thus, $\prod_{h=0}^\infty \xi_h > 0$, which proves Corollary 1(i). As a consequence, the backward coalescence time for the canonical coupling function τ_0^L , with L defined as in (14), is finite a.s. On the other hand,

$$a_k \leq 1 - |p(1|\mathbf{1}_{-\infty}^{-1}) - p(1|\mathbf{1}_{-k}^{-1}, -\mathbf{1}_{-\infty}^{-(k+1)})| \leq 1 - (1 - e^{-\beta}) \sum_{l>k} |\theta_l|;$$

thus, by letting the θ_k s approach 0 slowly enough, Proposition 3(i) is not satisfied. For example, taking $\theta_k = c_1/k^{2-\varepsilon}$ with $c_1 > 0$ sufficiently small and $\varepsilon \in (0, 1)$, we have $a_k \leq 1 - c_2/k^{1-\varepsilon}$ for some $c_2 > 0$, from which

$$\sum_{n=1}^\infty \prod_{k=1}^n a_k \leq \sum_{n=1}^\infty \prod_{k=1}^n \left(1 - \frac{c_2}{k^{1-\varepsilon}}\right) \leq \sum_{n=1}^\infty \exp\left(-\sum_{k=1}^n \frac{c_2}{k^{1-\varepsilon}}\right) \leq \sum_{n=1}^\infty \exp(-c_3 n^\varepsilon) < \infty,$$

where c_3 is a positive constant.

In the previous example the gain obtained by using the information depth constructed through the sequence $\{A_h, h \in \mathbb{N}\}$ is due to a lower bound on the number of sign changes in a finite window, as its width grows to ∞ . In the next example, which is inspired by [9], a single appearance of the word controls the decay of the ‘memory’ of the process. It serves also to present a situation in which \mathcal{H} differs from $G^{-\mathbb{N}_+}$.

Example 2. Consider again the alphabet $G = \{-1, 1\}$, and choose the set of admissible histories

$$\mathcal{H} = \left\{ \mathbf{w}_{-\infty}^{-1} \in G^{-\mathbb{N}_+} : \sum_{i \in \mathbb{N}_+} \mathbf{1}_{\{w_{-i}=1\}} = +\infty \right\}. \tag{18}$$

In order to specify the transition kernel, define the quantity $l(\mathbf{w}_{-\infty}^{-1}) = \inf\{i \in \mathbb{N}_+ : w_{-i} = 1\}$, which is finite for any $\mathbf{w}_{-\infty}^{-1} \in \mathcal{H}$. Now arbitrarily choose $p(1|\mathbf{1}_{-\infty}^{-1}) > 0$ and set

$$p(g|\mathbf{w}_{-\infty}^{-1}) = a_0(g) + (1 - a_0) \sum_{n \in \mathbb{N}_+} \delta_{g, w_{-l-n}} q_l(\mathbf{w}_{-\infty}^{-1}, n), \quad \mathbf{w}_{-\infty}^{-1} \neq -\mathbf{1}_{-\infty}^{-1}, \tag{19}$$

where $a_0(-1)$ and $a_0(1)$ are positive constants, $a_0 = a_0(-1) + a_0(1)$, and $\{q_{l,n}, n \in \mathbb{N}\}$ is a family of probability mass functions on the positive integers, depending on the parameter l . We assume that the distribution function $F_l(s) = \sum_{n \leq s} q_{l,n}$ is nonincreasing in l for any $s \in \mathbb{N}_+$. In this model the state of the process at any time is either the result of the toss of a coin, with probabilities $a_0(-1)/a_0$ and $a_0(1)/a_0$, or a value selected randomly from the past, according to F_l , among the values which appear after position l of the first 1. We again take $I_0(-1) = [0, a_0(-1))$ and $I_0(1) = [a_0(-1), a_0(-1) + a_0(1))$.

Let $\beta \in (0, 1)$, and, for any $h \in \mathbb{N}_+$, define

$$C_h^{(\beta)} = \{\mathbf{u}_0^{h-1} \in [0, 1)^h : u_j \in I_0(1) \text{ for some } j \in (\beta h, h - 1]\}. \tag{20}$$

If $w_s = 1$ for some $s \in [0, h - 1]$ then

$$a_h(\mathbf{w}_0^{h-1}) \geq a_0 + (1 - a_0)F_{h-s}(s).$$

Since the quantity on the right-hand side is monotone nondecreasing in s , if $\mathbf{U}_0^{h-1} \in C_h^{(\beta)}$ then

$$A_h(\mathbf{U}_0^{h-1}) \geq a_0 + (1 - a_0)F_{h-\lceil\beta h\rceil}(\lceil\beta h\rceil).$$

As a consequence, Corollary 1(i) is satisfied, for example, when

$$F_{h-\lceil\beta h\rceil}(\lceil\beta h\rceil) \geq 1 - Kh^{-c_\beta}$$

with $K_\beta > 0$ and $c_\beta > 1$. In particular, this holds when F_l is a geometric distribution function with success probability $p_l \asymp l^{-a}$ with $a < 1$.

Next observe that

$$\sum_h \mathbb{P}(\mathbf{U}_0^{h-1} \notin C_h^{(\beta)}) = \sum_h (1 - a_0(1))^{h-\lceil\beta h\rceil} < \infty,$$

which implies, by Borel–Cantelli’s lemma, that N is finite a.s. Moreover, since $\mathbb{P}(N = h + 1) \leq \mathbb{P}(\mathbf{U}_0^{h-1} \notin C_{h-1}^{(\beta)})$, $\mathbb{E}(s^N)$ is finite for $s < (1 - a_0(1))^{\beta-1}$. Therefore, Corollary 1(ii) holds provided that $a_\infty > (1 - a_0(1))^{1-\beta}$. Taking β arbitrarily close to 0, we prove the a.s. finiteness of $\tau^L[0, +\infty]$ for

$$a_\infty + a_0(1) > 1. \tag{21}$$

Finally, it is clear that \mathcal{H} in (18) satisfies Proposition 1(ii); as a consequence, we can perfectly simulate the process \mathbf{X} , compatible with kernel (19), which is also globally unique, since any compatible process \mathbf{Y} has $\mathbf{Y}_{-\infty}^0 \in \mathcal{H}$ with probability 1.

Observe that, for any positive integer k , $a_k = a_k((-\mathbf{1})_{-k+1}^0) = a_0$; hence, it is impossible to construct a backward coalescence time from an instantaneous information depth.

Remark 2. The above example allows us to compare the conditions in Corollary 1 with those presented in [10] for perfect simulation. For the sake of simplicity, we consider only the case in which the word which ‘controls’ the memory of the process is the digit 1 in a binary alphabet. The sufficient conditions in [10] for perfect simulation are thus

$$\inf_{\mathbf{w}_{-\infty}^{-1} \in G^{-\mathbb{N}_+}} p(1|\mathbf{w}_{-\infty}^{-1}) > 0, \quad \prod_{k=1}^\infty \xi_k^{(1)} > 0, \tag{22}$$

where, for $k \in \mathbb{N}_+$,

$$\xi_k^{(1)} = \inf_{i \in \mathbb{N}} \inf_{\mathbf{w}_{-i-k-1}^{-i-2} \in G^k} a_{i+k+1}(-\mathbf{1}_{-i}^{-1}, 1, \mathbf{w}_{-i-k-1}^{-i-2}),$$

$a_k(\mathbf{w}_{-k}^{-1})$ being defined as in (9) and (11), where \mathcal{H} is taken equal to $G^{-\mathbb{N}_+}$. Since $\xi_k^{(1)}$ is monotone nondecreasing in k , the fulfillment of the second condition in (22) depends only on the asymptotic behavior of the $\xi_k^{(1)}$ s. In the above example, in the interesting case when $\lim_{l \rightarrow \infty} F_l(s) = 0$ for each $s \in \mathbb{N}$, all the $\xi_k^{(1)}$ s are equal to a_0 ; hence, (22) is not satisfied (neglecting the trivial case in which $a_0 = 1$).

In general, hypotheses (i) and (ii) of Corollary 1, with the choices of \mathcal{H} and $C_k^{(\beta)}$ made in (18) and (20), are weaker than those in (22). Indeed, if $U_0^{k-1} \in C_k^{(\beta)}$ then

$$\begin{aligned}
 A_k(U_0^{k-1}) &\geq \xi_k \\
 &\geq \inf\{a_k(\mathbf{w}_{-k}^{-1}) : \mathbf{w}_{-\infty}^{-1} \in \mathcal{H}, w_{-i} = 1 \text{ for some } i \in [1, (1-\beta)k]\} \\
 &= \inf_{i \in [1, (1-\beta)k]} \inf_{\mathbf{w}_{-k}^{-i-2} \in G^{k-i-1}} a_k(-\mathbf{1}_{-i}^{-1}, 1, \mathbf{w}_{-k}^{-i-2}) \\
 &\geq \inf_{i \in [1, (1-\beta)k]} \xi_{k-i-1}^{(1)} \\
 &\geq \xi_{k-\lceil(1-\beta)k\rceil}^{(1)} \\
 &= \xi_{\lfloor \beta k \rfloor}^{(1)}.
 \end{aligned} \tag{23}$$

If (22) holds then $\prod_{k=0}^{\infty} (\xi_k^{(1)})^{1/\beta} > 0$ and $\prod_{k=0}^{\infty} \xi_{\lfloor \beta k \rfloor}^{(1)} > 0$ too. Then, by (23), we also have $\prod_{k=0}^{\infty} \xi_k > 0$, which proves Corollary 1(i). Moreover, $\prod_{k=0}^{\infty} \xi_k^{(1)} > 0$ implies that $a_{\infty} = \lim_{k \rightarrow \infty} \xi_k^{(1)} = 1$; as a consequence, inequality (21) holds, Corollary 1(ii) is satisfied. Thus, Proposition 1 and Corollary 1 imply the sufficiency of (22) for existence, uniqueness, and perfect simulability, which is obviously not necessary, as shown in the previous example.

3. Algorithms without the minorization condition

In this section we explore the possibility of defining a backward coalescence time when $a_i = 0$ for $i = 0, \dots, r-1$ with some positive r . Our technique is based on the determination of *coalescing events*, which possibly requires the modification of the canonical coupling function, as will be clarified by the next example. Within this section we always consider the choice $\mathcal{H} = \mathcal{H}^*$ given in (2). Then, by Proposition 1, the existence of a backward coalescence time will prove the existence and uniqueness of a compatible process for a kernel p , together with its perfect simulability.

We illustrate the basic idea through the following example.

Example 3. Let $G = \{-1, 1\}$, and consider the probability kernel

$$p(g|\mathbf{w}_{-\infty}^{-1}) = \frac{1}{2} - g \sum_{k=1}^{\infty} \frac{w_{-k}}{2^{k+1}}, \quad g \in G, \mathbf{w}_{-\infty}^{-1} \in G^{-\mathbb{N}_+}.$$

We immediately see that, for this kernel, $\mathcal{H}^* = G^{-\mathbb{N}_+}$. For simplicity of notation, we introduce the history $\hat{\mathbf{g}}_{-\infty}^{-1}$ which is constantly equal to $g \in \{-1, 1\}$. Since

$$\begin{aligned}
 p(g|\mathbf{w}_{-k}^{-1}, \hat{\mathbf{g}}_{-\infty}^{-(k+1)}) &\leq p(g|\mathbf{w}_{-\infty}^{-1}) \leq p(g|\mathbf{w}_{-k}^{-1}, -\hat{\mathbf{g}}_{-\infty}^{-(k+1)}), \\
 a_0(g) = p(g|\hat{\mathbf{g}}_{-\infty}^{-1}) &= 0 \text{ for any } g \in G, a_k(g|\mathbf{w}_{-k}^{-1}) = p(g|\mathbf{w}_{-k}^{-1}, \hat{\mathbf{g}}_{-\infty}^{-(k+1)}), \text{ and}
 \end{aligned}$$

$$a_k(g|\mathbf{w}_{-k}^{-1}) - a_{k-1}(g|\mathbf{w}_{-(k-1)}^{-1}) = \begin{cases} 0 & \text{if } w_{-k} = g, \\ \frac{1}{2^k} & \text{if } w_{-k} = -g. \end{cases}$$

The canonical coupling function is given by

$$f(u|\mathbf{w}_{-\infty}^{-1}) = - \sum_{k=1}^{\infty} w_{-k} \mathbf{1}_{[1-1/2^{k-1}, 1-1/2^k)}(u), \quad \mathbf{w}_{-\infty}^{-1} \in G^{-\mathbb{N}_+}, u \in [0, 1),$$

with the instantaneous information depth defined in (12):

$$K(u) = - \sum_{k=1}^{\infty} k \mathbf{1}_{[1-1/2^{k-1}, 1-1/2^k)}(u), \quad \mathbf{w}_{-\infty}^{-1} \in G^{-\mathbb{N}_+}, \quad u \in [0, 1).$$

Therefore, for any $u \in [0, 1)$ and any history $\mathbf{w}_{-\infty}^{-1} \in G^{-\mathbb{N}_+}$, we have

$$f(u | \mathbf{w}_{-\infty}^{-1}) = -f(u | \mathbf{w}_{-\infty}^{-1}).$$

This property is clearly inherited by any composition of the form (4) and this makes the existence of a backward coalescence time impossible. Indeed, since $a_0 = 0$, this model is not covered by either Proposition 3 or the material presented in the previous section.

This unfavorable property can be overcome by making the following modification to the canonical coupling function. The modified coupling function $\tilde{f}(\cdot | \mathbf{w}_{-\infty}^{-1})$ is still equal to $f(\cdot | \mathbf{w}_{-\infty}^{-1})$, unless $w_{-1} = +1$, $w_{-2} = -1$, and $w_{-3} = +1$, in which case

$$\tilde{f}(u | \mathbf{w}_{-\infty}^{-1}) = \begin{cases} -1, & \frac{1}{2} \leq u < \frac{5}{8}, \\ 1, & \frac{5}{8} \leq u < \frac{7}{8}, \\ f(u | \mathbf{w}_{-\infty}^{-1}), & \text{otherwise.} \end{cases}$$

In other words, the interval $[\frac{3}{4}, \frac{7}{8})$ where f takes the value -1 is shifted to $[\frac{1}{2}, \frac{5}{8})$, and the interval $[\frac{1}{2}, \frac{3}{4})$ where f takes the value 1 is shifted to $[\frac{5}{8}, \frac{7}{8})$. It is readily seen that \tilde{K} is an information depth for the coupling function \tilde{f} , given by

$$\tilde{K}(u) = \begin{cases} 3, & \frac{1}{2} \leq u < \frac{3}{4}, \\ K(u), & \text{otherwise.} \end{cases}$$

Note that $\tilde{K}(u)$ adds 1 to $K(u)$ for $\frac{1}{2} \leq u < \frac{3}{4}$.

Next define

$$E = \{\mathbf{u}_{-3}^0 \in [0, 1)^4 : u_i < \frac{1}{2}, i = -3, -2, -1, \frac{1}{2} \leq u_0 < \frac{5}{8}\},$$

and observe that, whenever $\mathbf{u}_{-3}^0 \in E$, then, necessarily, for any history $\mathbf{w}_{-\infty}^{-4} \in G^{-\mathbb{N}_+}$,

$$\tilde{f}^{(4)}(\mathbf{u}_{-3}^0 | \mathbf{w}_{-\infty}^{-4}) = -1. \tag{24}$$

In fact, $u_{-i} < \frac{1}{2}$ for $i = 1, 2, 3$ implies that the vector

$$(\tilde{f}^{(3)}(\mathbf{u}_{-3}^{-1} | \mathbf{w}_{-\infty}^{-4}), \tilde{f}^{(2)}(\mathbf{u}_{-3}^{-2} | \mathbf{w}_{-\infty}^{-4}), \tilde{f}(\mathbf{u}_{-3} | \mathbf{w}_{-\infty}^{-4}))$$

takes either the value $(-1, 1, -1)$ or $(1, -1, 1)$. By the definition of \tilde{f} we also have, for $\frac{1}{2} \leq u_0 < \frac{5}{8}$,

$$\tilde{f}(u_0 | -1, 1, -1, \mathbf{w}_{-\infty}^{-4}) = \tilde{f}(u_0 | 1, -1, 1, \mathbf{w}_{-\infty}^{-4}) = -1$$

for any $\mathbf{w}_{-\infty}^{-4} \in G^{-\mathbb{N}_+}$, which proves (24).

It is clear that the subsequent iteration of \tilde{f} leads to alternate signs, as long as $u_i < \frac{1}{2}$. Thus, if $u_1, u_2 < \frac{1}{2}$, we have

$$\tilde{f}^{(5)}(\mathbf{u}_{-3}^1 | \mathbf{w}_{-\infty}^{-4}) = 1, \quad \tilde{f}^{(6)}(\mathbf{u}_{-3}^2 | \mathbf{w}_{-\infty}^{-4}) = -1.$$

Next we can compute another iterate if $u_3 < \frac{7}{8}$, and continue indefinitely if $u_k < 1 - 1/2^k$ for any $k \geq 3$. The probability that $\mathbf{U} = (U_j, j \in \mathbb{Z})$ realizes all these events is

$$\frac{1}{2^8} \prod_{k=3}^{\infty} \left(1 - \frac{1}{2^k}\right) > 0.$$

Arguing as at the end of the proof of Theorem 1, it is then possible to construct a backward coalescence time for the coupling function \tilde{f} , as will be stated in the forthcoming Theorem 2. This example does not satisfy the sufficient condition for uniqueness of a compatible process given in Proposition 4.2 of [4].

In Example 3, the fact that the information depth $\tilde{K} \geq 1$ is circumvented by the fact that the coalescing event $U_{-3}^0 \in E$ has positive probability. When this event is realized, the process has the constant value 1 at time 0, which does not depend on the history $\mathbf{w}_{-\infty}^{-4} \in \mathcal{H}$. To keep a constant value at time l also, it is enough to have $\tilde{K}(U_i) \leq i$ for $i \leq l$. With the next result we proceed to generalize this example, allowing coalescing events which determine the state process for r instants of time, as appropriate when the minimum value of the information depth is r .

Theorem 2. *Let f be a coupling function, and let K be an instantaneous information depth. Let $r \in \mathbb{N}_+$ be such that $\mathbb{P}(K(U) \leq r) > 0$. Suppose that, for some $n \geq r$, there exists a Borel set $E_{n,r} \subset [0, 1]^n$ with positive n -dimensional Lebesgue measure such that*

$$\mathbf{u}_{-n+1}^0 \in E_{n,r} \implies f^{(n-i)}(\mathbf{u}_{-n+1}^{-i} | \mathbf{w}_{-\infty}^{-n}) = f^{(n-i)}(\mathbf{u}_{-n+1}^{-i} | \mathbf{g}_{-\infty}^{-n})$$

for $i = 0, \dots, r - 1$ and all $\mathbf{w}_{-\infty}^{-n} \in \mathcal{H}$. (25)

(i) *If $\sum_{k=r}^{\infty} \prod_{j=r}^k \mathbb{P}(K(U) \leq j) = \infty$ then*

$$\tau_0 = \sup\{m \leq -n + 1 : \mathbf{U}_m^{m+n-1} \in E_{n,r} \text{ and } K(U_j) - r \leq j - m - n, \\ j \in [m + n, 0], \text{ if } m < -n + 1\}$$

(26)

is a backward coalescence time.

(ii) *If $\prod_{j=r}^{\infty} \mathbb{P}(K(U) \leq j) > 0$ then*

$$\tau[0, \infty] = \inf\{\tau_i : i \in \mathbb{N}\}$$

is a.s. finite.

Proof. Let us start by proving (i). First note that $-\tau_0$ is $\{\mathcal{F}_{-n}^0\}$ -adapted. For property (H2) to hold, note that, if $\tau_0 = m$ then $\mathbf{U}_m^{m+n-1} \in E_{n,r}$. By (25), this implies that

$$f^{(n-j+1)}(\mathbf{U}_m^{m+n-j} | \mathbf{w}_{-\infty}^{m-1}) = f^{(n-j+1)}(\mathbf{U}_m^{m+n-j} | \mathbf{g}_{-\infty}^{m-1}), \quad j = 1, \dots, r \text{ for all } \mathbf{w}_{-\infty}^{m-1} \in \mathcal{H},$$

which means that the value of the process within the interval $[m+n-r, m+n-1]$ is determined. If $m = -n + 1$, (H2) immediately holds; otherwise, the fact that $K(U_j) \leq j - m - n + r$ for $j \in [m+n, 0]$ guarantees that the same property is kept within the interval $[m+n, 0]$, which proves (H2).

The main problem is to establish that τ_0 is a.s. finite, which will ensure property (H1). Indeed, we will define a backward coalescence time τ' such that $\tau'(\mathbf{u}_{-\infty}^0) \leq \tau_0(\mathbf{u}_{-\infty}^0) \leq -n + 1$

for each $u_{-\infty}^0 \in [0, 1)^{-\mathbb{N}}$. We start by defining the sequence $\{Y_i, i = 0, 1, \dots\}$, whose elements we prove to be a.s. finite, too. Let $Y_0 = 1$, and inductively define, for $i \in \mathbb{N}$,

$$Y_{i+1} = \sup\{m \leq Y_i - n : K(U_j) - r \leq j - m, j \in [m, Y_i - 1]\}. \tag{27}$$

Replacing $K(U_j)$ with $\max\{K(U_j) - r, 0\}$ and applying Proposition 3(a), we see that condition (i) guarantees that Y_1 is a.s. finite. Moreover, $-Y_1$ is adapted with respect to the filtration $\{\mathcal{F}_{-k}^0 : k \in \mathbb{N}\}$. Therefore, $\{-Y_i, i \in \mathbb{N}_+\}$ forms a renewal process and the sequence $\{U_{-n}, n \in \mathbb{N}\}$ is regenerative with respect to it, since $Y_1 = \hat{Y}(U_{-\infty}^0)$ and $Y_{i+1} = \hat{Y}(U_{-\infty}^{Y_i-1})$ for $i \in \mathbb{N}_+$, where \hat{Y} is a suitable measurable function.

Let

$$I = \inf\{i \in \mathbb{N} : U_{Y_i-n}^{Y_i-1} \in E_{n,r}\}.$$

From the previous result we deduce that I has a geometric distribution with success probability equal to

$$P(U_{-n+1}^0 \in E_{n,r}, Y_1 > -\infty) = P(U_{-n+1}^0 \in E_{n,r}) = \lambda_n(E_{n,r}), \tag{28}$$

where λ_n is the Lebesgue measure of dimension n . Note that, since $Y_1 \leq -n + 1$, the probability on the left-hand side of (28) has the form $E(g(U_{Y_1}^0))$, which, by the regenerative property, is also equal to $E(g(U_{Y_{i+1}}^{Y_i-1}))$ for any $i \geq 2$. Finally, define the random variable $\tau' = Y_I - n$, which is a.s. finite.

In order to prove that $\tau' \leq \tau_0$, we have to verify that Y_{i+1} can be rewritten as

$$Y_{i+1} = \sup\{m \leq Y_i - n : K(U_j) - r \leq j - m, j \in [m, 0]\}. \tag{29}$$

For $i = 0$, expressions (27) and (29) coincide. By induction, suppose that

$$Y_i = \sup\{m \leq Y_{i-1} - n : K(U_j) - r \leq j - m, j \in [m, 0]\}$$

is true. Then, for $j \in [Y_i, 0]$, it holds that

$$K(U_j) - r \leq j - Y_i \leq j - Y_{i+1},$$

proving, together with (27), that (29) holds. Finally, to prove that $\tau' \leq \tau_0$, it is enough to check that τ' belongs to the set defined in (26).

In order to prove (ii), we observe that

$$\tau[0, \infty] = -n \iff U_{-n+1}^0 \in E_{n,r} \text{ and } \{K(U_j) \leq j + r - 1, j \in \mathbb{N}_+\}.$$

Note that the probability of the event on the right-hand side is equal to

$$\lambda_n(E_{n,r}) \prod_{k=r}^{\infty} P(K(U_0) \leq k) > 0.$$

Arguing as at the end of the proof of Theorem 1, there exists an a.s. finite $T_0 \leq 0$ such that

$$\tau[T_0, \infty] = T_0 - n,$$

and since $\tau[T_0, \infty] \leq \tau[0, \infty] \leq 0$, we obtain property (ii).

The simplest situation in which Theorem 2 is of interest is when $a_0 = 0$ and $a_1 > 0$; thus, $r = 1$. In this situation a more explicit construction of the coalescing event can be given. For the sake of simplicity, we restrict attention to the canonical coupling function, together with the information depth given in (12) for which

$$\{u \in [0, 1): K(u) = 1\} = [0, a_1).$$

We say that the simulation process is in the *Markovian regime* at time n whenever $U_n < a_1$, which means that the information needed to compute the state of the process at time n concerns only the state at time $n - 1$. We define the *Markovian coupling function*

$$\tilde{f}(u|w) = f(a_1u|w_{-\infty}^{-1}), \quad u \in [0, 1), w \in G, \tag{30}$$

for any choice of $w_{-\infty}^{-1} \in \mathcal{H}$ with $w_{-1} = w$.

If a backward coalescence time $\tilde{\tau}$ is available for the Markovian coupling function \tilde{f} then, for some $n \in \mathbb{N}_+$,

$$E_{n,1} = \{u_{-n+1}^0 \in [0, a_1)^n : \tilde{\tau}(a_1^{-1}u_{-n+1}^0) = -n\} \tag{31}$$

has positive n -dimensional Lebesgue measure. Since $E_{n,1}$ satisfies property (25), with $r = 1$, it could be used to construct τ_0 as in (26). In this case, however, rather than checking coalescence in the Markovian regime only for a fixed value of n , we can continue checking coalescence as long as the Markovian regime continues. Thus, we are led to the backward coalescence time candidate

$$\tau_0^* = \sup\{m < 0 : \text{there exists } l \in [m, 0] \text{ such that } U_m^l \in E_{l-m+1,1}, K(U_j) \leq j - l, \\ j \in [l + 1, 0] \text{ if } l < 0\},$$

whose a.s. finiteness is discussed in the following theorem. Its proof proceeds along lines similar to those of Theorem 2, and is thus omitted for the sake of brevity.

Theorem 3. *Let f be a canonical coupling function, and suppose that $a_0 = 0$ and $a_1 > 0$. Let $\tilde{\tau}$ be a backward coalescence time for the Markovian coupling function (30), and assume that the coalescing event (31) has positive n -dimensional Lebesgue measure.*

- (i) *If $\sum_{n=1}^{\infty} \prod_{j=1}^n a_j = +\infty$ then τ_0^* is a backward coalescence time.*
- (ii) *If $\prod_{j=1}^{\infty} a_j > 0$, $\tau^*[0, \infty]$ is a.s. finite.*

Necessary conditions for the existence of a backward coalescence time $\tilde{\tau}$ for a Markovian coupling function \tilde{f} can be given in terms of the kernel induced by \tilde{f} . Recall that \tilde{f} induces the Markov kernel

$$M(g|w) = |\{u \in [0, 1): \tilde{f}(u|w) = g\}|, \quad g, w \in G.$$

Since the existence of a backward coalescence time implies the convergence in law of the chain as time increases, for any starting state, it is necessary that M has a single positive recurrent irreducible class which is aperiodic. When G is finite, which is assumed from now on, this condition can be directly related to the oriented graph induced by M . If \tilde{f} is deduced from a canonical coupling function f then, for any arc (w, g) of this graph, $a_1(g|w) > 0$. However, the converse is not true, since $\sum_{g \in G} a_1(g|w) = a_1(w)$ in general exceeds a_1 . But, through a

convenient rearrangement and possibly a splitting of the subintervals of $[0, a_1(w))$ assigned to the various $g \in G$, we can define a coupling function \bar{f} such that $a_1(g|w) > 0$ implies that $M(w, g) > 0$.

Backward coalescence with probability 1 cannot be ensured by the properties of the Markov kernel, without reference to the coupling function. However, when the state space is finite, there is a universal construction of a Markov coupling function, which ensures backward coalescence with probability 1 under the only assumption that the induced kernel has a unique irreducible class which is aperiodic (see Proposition 8.1 of [1, p. 122]). The construction consists in letting the different trajectories move independently before merging, which is easily incorporated into our framework, if coupling functions depending on a vector of uniform random variables are allowed.

As a corollary to Theorem 3, by taking into account the last two previous remarks, we have the following result.

Corollary 2. *Suppose that $p: G \times G^{-\mathbb{N}_+} \rightarrow [0, 1]$ is a kernel on the finite state space G . Define the oriented graph \mathcal{G} with the set of vertices G and the set of arcs*

$$\mathcal{A} = \{(w, g) \in G^2: a_1(g|w) > 0\}. \tag{32}$$

Suppose that

- (i) $\sum_{n=1}^\infty \prod_{m=1}^n a_m = \infty$;
- (ii) \mathcal{G} has a single irreducible class which is aperiodic;
- (iii) for any $g \in G$, there exists $w \in G$ such that $(w, g) \notin \mathcal{A}$.

Then it is possible to construct a backward coalescence time for the coupling function \hat{f} .

Applying Proposition 1 with $\mathcal{H} = \mathcal{H}^*$ and assuming the existence of a backward coalescence time, we obtain the existence and uniqueness of a process compatible with p . Note that assumption (iii) of Corollary 2 has the unique purpose of ensuring that $a_0 = 0$.

Example 4. Corollary 2 covers *generalized random walks* on a finite directed graph $\mathcal{G} = (G, \mathcal{A})$. Before defining this kind of process, we define the set of one-sided infinite paths on the oriented graph \mathcal{G} as

$$\mathcal{C} = \{\mathbf{w}_{-\infty}^{-1} \in G^{-\mathbb{N}_+}: (w_{-(k+1)}, w_{-k}) \in \mathcal{A}, k \in \mathbb{N}_+\}.$$

Generalized random walks on $\mathcal{G} = (G, \mathcal{A})$ are processes compatible with a kernel p over the alphabet G with the properties

- if $(g, w) \notin \mathcal{A}$ then, for all $\mathbf{w}_{-\infty}^{-1} \in G^{-\mathbb{N}_+}$ with $w_{-1} = w$, we have $p(g|\mathbf{w}_{-\infty}^{-1}) = 0$;
- if $(g, w) \in \mathcal{A}$, there exists $\varepsilon > 0$ such that, for all $\mathbf{w}_{-\infty}^{-1} \in \mathcal{C}$ with $w_{-1} = w$, we have $p(g|\mathbf{w}_{-\infty}^{-1}) > \varepsilon$.

The first property implies that $\mathcal{H}^* \subset \mathcal{C}$, whereas the second property ensures the opposite inclusion. Moreover,

$$a_1(g|w) = \inf\{p(g|\mathbf{w}_{-\infty}^{-1}): \mathbf{w}_{-\infty}^{-1} \in \mathcal{C}, w_{-1} = w\} \geq \varepsilon > 0$$

if $(g, w) \in \mathcal{A}$ is an arc of \mathcal{G} ; otherwise, clearly, $a_1(g|w) = 0$. Thus, the set \mathcal{A} is related to the kernel p as in (32). If the graph \mathcal{G} satisfies Corollary 2(ii) and (iii), and the sequence $\{a_k, k \in \mathbb{N}_+\}$ satisfies Corollary 2(i), then a backward coalescence time exists.

Appendix A. Proofs of Propositions 1 and 2

Proof of Proposition 1. By stationarity, for compatibility, it is enough to prove (1) for $m = 0$. Given the realization $\mathbf{U}_{-\infty}^0$, define the nonempty random subsets of \mathcal{H} by

$$\mathcal{I}_{n+1}(\mathbf{U}_{-n}^0) = \{(f^{(n+1)}(\mathbf{U}_{-n}^0 | \mathbf{w}_{-\infty}^{-(n+1)}), \dots, f^{(1)}(U_{-n} | \mathbf{w}_{-\infty}^{-(n+1)}), \mathbf{w}_{-\infty}^{-(n+1)}) : \mathbf{w}_{-\infty}^{-(n+1)} \in \mathcal{H}\}$$

for $n \in \mathbb{N}$. By (4), this subset consists of admissible histories in \mathcal{H} , obtained by varying in all possible ways the initial history prior to time $-n$, and then applying the coupling function with the fixed values U_{-n}, \dots, U_0 , until time 0. The sequence $\mathcal{I}_{n+1}(\mathbf{U}_{-n}^0)$ is nonincreasing in $n \in \mathbb{N}$.

It is now convenient to consider on \mathcal{H} the ultrametric distance

$$d(\mathbf{w}_{-\infty}^{-1}, \mathbf{w}_{-\infty}^{-1}) = 2^{-\inf\{i \in \mathbb{N}_+ : v_i \neq w_i\}},$$

which induces the product topology. The first n elements in the histories $\mathcal{I}_{-\tau[-n,0]+1}(\mathbf{U}_0^{\tau[-n,0]})$ have the same values, namely, \mathbf{X}_{-n}^0 . Therefore, the ultrametric distance from $\mathbf{X}_{-\infty}^0$ is smaller than 2^{-n} . Therefore, $\mathbf{X}_{-\infty}^0$ is at distance 0 from

$$\mathcal{I}_1(U_0) = \{(f(U_0 | \mathbf{w}_{-\infty}^{-1}), \mathbf{w}_{-\infty}^{-1}) : \mathbf{w}_{-\infty}^{-1} \in \mathcal{H}\};$$

thus, it belongs to its closure. So there exists a sequence $\{(\mathbf{w}^{(n)})_{-\infty}^{-1}\} \subset \mathcal{H}$ such that

$$f(U_0 | (\mathbf{w}^{(n)})_{-\infty}^{-1}) \rightarrow X_0, \quad (\mathbf{w}^{(n)})_{-\infty}^{-1} \rightarrow \mathbf{X}_{-\infty}^{-1}.$$

Finally, since $f(U_0|\cdot)$ is continuous in \mathcal{H} , and $\mathbf{X}_{-\infty}^{-1} \in \mathcal{H}$ with probability 1, we have, a.s.,

$$X_0 = f(U_0 | \mathbf{X}_{-\infty}^{-1}),$$

and since $\mathbf{X}_{-\infty}^{-1}$ is $\mathcal{F}_{-\infty}^{-1}$ -adapted, it is independent of U_0 , from which the compatibility of X follows.

The proof of uniqueness is essentially the same as in [2, p. 935]. The variational distance between two distributions indexed by initial histories in \mathcal{H} differing only before time $-i$, obtained by means of the kernel p for the interval $[0, n]$, is bounded by the tail probability $P(\tau[0, n] \leq -i)$. By the a.s. finiteness of $\tau[0, n]$, this tail probability goes to 0 as $i \rightarrow \infty$.

Proof of Proposition 2. Let us observe that, for any $m \in \mathbb{N}$,

$$\{-\tau_0^K \leq m\} = \bigcup_{i=0}^m F_i,$$

where

$$F_i = \{K_{-i} = 0, K_{-i+1} \leq 1, \dots, K_{-1} \leq i - 1, K_0 \leq i\}.$$

Since $F_i \in \mathcal{F}_{-i}^0$, (H1) is proved.

It is now easy to check that if F_i is realized then

$$f^{(i)}(U_0, \dots, U_{-i+1}, U_{-i} | \mathbf{w}_{-\infty}^{-(i+1)}) = f^{(i)}(U_0, \dots, U_{-i+1}, U_{-i} | \mathbf{g}_{-\infty}^{-(i+1)})$$

for any $\mathbf{w}_{-\infty}^{-(i+1)} \in \mathcal{H}$, which establishes (H2).

Acknowledgements

We would like to warmly thank an anonymous referee for bringing to our attention reference [7], and Davide Gabrielli for pointing out the papers [9] and [10].

References

- [1] ASMUSSEN, S. AND GLYNN, P. W. (2007). *Stochastic Simulation: Algorithms and Analysis*. Springer, New York.
- [2] COMETS, F., FERNÁNDEZ, R. AND FERRARI, P. A. (2002). Processes with long memory: regenerative construction and perfect simulation. *Ann. Appl. Prob.* **12**, 921–943.
- [3] DE SANTIS, E. AND PICCIONI, M. (2008). Exact simulation for discrete time spin systems and unilateral fields. *Methodology Comput. Appl. Prob.* **10**, 105–120.
- [4] FERNÁNDEZ, R. AND MAILLARD, G. (2005). Chains with complete connections: general theory, uniqueness, loss of memory and mixing properties. *J. Statist. Phys.* **118**, 555–588.
- [5] 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, 86pp.
- [6] FERRARI, P. A., FERNÁNDEZ, R. AND GARCIA, N. L. (2002). Perfect simulation for interacting point processes, loss networks and Ising models. *Stoch. Process. Appl.* **102**, 63–88.
- [7] FOSS, S. AND KONSTANTOPOULOS, T. (2003). Extended renovation theory and limit theorems for stochastic ordered graphs. *Markov Process. Relat. Fields* **9**, 413–468.
- [8] FOSS, S. AND TWEEDIE, R. L. (1998). Perfect simulation and backward coupling. *Commun. Statist. Stoch. Models* **14**, 187–203.
- [9] GALLO, S. (2011). Chains with unbounded variable length memory: perfect simulation and a visible regeneration scheme. *Adv. Appl. Prob.* **43**, 735–759.
- [10] GALLO, S. AND GARCIA, N. L. (2010). Perfect simulation for stochastic chains of infinite memory: relaxing the continuity assumption. Preprint. Available at <http://arXiv.org/abs/1005.5459v1>.
- [11] HÄGGSTRÖM, O. AND STEIF, J. E. (2000). Propp-Wilson algorithms and finitary codings for high noise Markov random fields. *Combinatorics Prob. Comput.* **9**, 425–439.
- [12] MØLLER, J. (2001). A review of perfect simulation in stochastic geometry. In *Selected Proceedings of the Symposium on Inference for Stochastic Processes* (Athens, GA, 2000; IMS Lecture Notes Monogr. Ser. **37**), Institute of Mathematical Statistics, Beachwood, OH, pp. 333–355.
- [13] MURDOCH, D. J. AND GREEN, P. J. (1998). Exact sampling from a continuous state space. *Scand. J. Statist.* **25**, 483–502.
- [14] PROPP, D. B. AND WILSON, J. G. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures Algorithms* **9**, 223–252.
- [15] WILLIAMS, D. (1991). *Probability with Martingales*. Cambridge University Press.