One-Dimensional Infinite Memory Imitation Models with Noise

Emilio De Santis & Mauro Piccioni

Springer

Springer

CrossMark

One-Dimensional Infinite Memory Imitation Models with Noise

Emilio De Santis[1] · Mauro Piccioni[1]

Abstract In this paper we study stochastic process indexed by $\mathbb{Z}$ constructed from certain transition kernels depending on the whole past. These kernels prescribe that, at any time, the current state is selected by looking only at a previous random instant. We characterize uniqueness in terms of simple concepts concerning families of stochastic matrices, generalizing the results previously obtained in De Santis and Piccioni (J Stat Phys 150(6):1017–1029, 2013).

1 Introduction and Main Definitions

This paper is concerned with stochastic processes indexed by $\mathbb{Z}$ taking values in a finite alphabet G, constructed from a *transition kernel* which depends on the whole past, i.e. a map $p : G \times G^{-\mathbb{N}_+} \to [0, 1]$ such that, for any choice of $\mathbf{w}_{-\infty}^{-1} = (w_{-1}, w_{-2}, \ldots) \in G^{-\mathbb{N}_+}$, $p(\cdot|\mathbf{w}_{-\infty}^{-1})$ is a probability measure on G. In the literature these models appear under various names, as *chains with complete connections* [18], *g-functions* [3] or *processes with long memory* [4].

Two alternative ways to associate a G-valued stochastic process, i.e. a probability measure on $G^{\mathbb{Z}}$, to a transition kernel are possible. The first deals with processes with a boundary condition. Let $\mathbf{w} = (w_n, n \in \mathbb{Z})$ be an arbitrary *configuration*, i.e. an element of $G^{\mathbb{Z}}$, possibly random, and let $r \in \mathbb{Z}$; we say that $\mathbf{X}^{r,\mathbf{w}} = (X_n^{r,\mathbf{w}}, n \in \mathbb{Z})$ is *governed by the kernel p with boundary condition $\mathbf{w}$ from the instant r*, if $X_n^{r,\mathbf{w}} = w_n$ for $n \le r$ and

✉ Emilio De Santis
desantis@mat.uniroma1.it

Mauro Piccioni
piccioni@mat.uniroma1.it

[1] Dipartimento di Matematica, Università di Roma La Sapienza, Rome, Italy

$$P\left(X_n^{r,\mathbf{W}} = g | X_{n-1}^{r,\mathbf{W}}, X_{n-2}^{r,\mathbf{W}}, \dots\right) = p\left(g | X_{n-1}^{r,\mathbf{W}}, X_{n-2}^{r,\mathbf{W}}, \dots\right) \text{ a.s.,} \tag{1}$$

for any $n > r$. It is clear that, given the law of $\mathbf{w}$ (in particular if it is a deterministic sequence), the law of this process is uniquely defined. If for some strictly decreasing sequence (r_n), $\mathbf{X}^{r_n,\mathbf{w}}$ converges weakly in the product topology of $G^{\mathbb{Z}}$, the limit is said to be an *infinite volume limit*. By weak compactness of the set of probability measures on a compact space, this set is non-empty: it reduces to a single element μ if and only if $\mathbf{X}^{r,\mathbf{w}}$ converges weakly to μ, as $r \to -\infty$, irrespectively of $\mathbf{w}$. Infinite volume limits are mainly considered in the theory of multi-dimensional random fields [16,17]. Here we find convenient to borrow the standard usage in the multi-dimensional framework to introduce boundary conditions from an entire configuration on $\mathbb{Z}$, rather than shifting to the left a configuration defined only on the half-line $-\mathbb{N}_+$.

The second construction, more often used in the one-dimensional *time-directional* context we are concerned with, is to declare directly a process $\mathbf{X} = (X_n, n \in \mathbb{Z})$, equivalently its law, to be *compatible* with p if (1) holds for any $n \in \mathbb{Z}$. This is analogous to the Dobrushin–Lanford–Ruelle definition in the theory of multi-dimensional random fields [16,17]. Compatible laws are immediately seen to be infinite volume limits; indeed if $\mathbf{W}$ is compatible with p and we choose it as a boundary condition, then $\mathbf{X}^{r,\mathbf{W}}$ has the same law of $\mathbf{W}$, for any $r \in \mathbb{Z}$: here we profit of having allowed random boundary conditions. Conversely, since (1) is equivalent to

$$E\left[\mathbf{1}_{\{g\}}(X_n^{r,\mathbf{W}})h(X_{n-1}^{r,\mathbf{W}}, \dots, X_{n-m}^{r,\mathbf{W}})\right] = E\left[p(g|X_{n-1}^{r,\mathbf{W}}, X_{n-2}^{r,\mathbf{W}}, \dots)h(X_{n-1}^{r,\mathbf{W}}, \dots, X_{n-m}^{r,\mathbf{W}})\right],$$

for any positive integer m and any real function h defined on G^m, this relation is maintained in the limit provided $p(g|\cdot)$ is continuous for any $g \in G$. In this paper only continuous kernels will be considered, therefore we will identify infinite volume limits with compatible laws, denoting their set with $\mathcal{G}(p)$. In the proofs both characterizations of $\mathcal{G}(p)$ will be found useful.

Notice that elements of $\mathcal{G}(p)$ are not necessarily stationary, i.e. translation invariant, but from a non-stationary element of $\mathcal{G}(p)$ one can produce a stationary one by performing Cesaro averages of shifts over a finite window increasing to $\mathbb{Z}$. Thus if $\mathcal{G}(p)$ reduces to a single element, it has to be stationary. On the contrary, it is possible that $|\mathcal{G}(p)| > 1$ but this set contains only one stationary element; indeed we will present later a situation in which this happens. Notice that, $\mathcal{G}(p)$ being convex, in case of non uniqueness $\mathcal{G}(p)$ has infinitely many elements.

Uniqueness conditions for general transition kernels of the form (1) are scattered in the literature for various decades. Some of these results refer to a dynamical systems setting, see e.g. [19,20,26]. The use of techniques of a more probabilistic flavor, in particular *coupling techniques*, has increased in time, see e.g. [1,21,25]. The work [4] has started a constructive approach, focused to the design of *perfect simulation schemes* for the unique compatible measure. In a number of cases this has allowed to prove not only the uniqueness, but also the existence of a compatible law, when G is countable. Finally, multi-dimensional statistical mechanics techniques, such as the Dobrushin criterion, have recently been used also in this setting [11,12]. For perfect simulation in the multi-dimensional case the reader is addressed to e.g. [5,7,15]; also the continuity assumption can be relaxed, as in [6].

The various sufficient conditions for uniqueness usually take a suitable positivity condition together with some regularity assumption on the kernel p. The latter allows to control the behavior of the range of the functions $p(g|\mathbf{w}_{-r}^{-1}\cdot)$ on $G^{-\mathbb{N}_+}$, for fixed $g \in G$ and $\mathbf{w}_{-\infty}^{-1} \in$

$G^{-\mathbb{N}_+}$, as r gets large. Regularity assumptions of some sort are actually needed for uniqueness, as shown in [3], where an example of a positive transition kernel has been given with a strong "dependence on the remote past" that gives rise to different infinite volume limits.

In order to motivate the class of kernels considered in the paper it is useful to recall the setting of [4]. In this paper they write down a decomposition of a continuous kernel of the following form:

$$p(g|\mathbf{w}_{-\infty}^{-1}) = \theta_0 \nu(g) + \sum_{k=1}^{\infty} \theta_k P_{(k)}(g; w_{-1}, \ldots w_{-k}), \tag{2}$$

where ν is a probability distribution on G, $\theta = (\theta_n, n \in \mathbb{N})$ is a probability distribution on the integers and for any $k \in \mathbb{N}_+$ $P_{(k)} : G \times G^k \to [0, 1]$ is a transition kernel depending only on the k-th most recent values. If $\theta_0 > 0$ and θ_n decays to zero fast enough they provide a perfect simulation algorithm for the unique compatible measure. The first assumption corresponds to positivity of $p(g|\cdot)$, for some $g \in G$, whereas the second amounts again to a regularity assumption on the kernel p.

The mixture decomposition presented in [4] is not unique. Other decompositions have been proposed to prove uniqueness [8,13,14], leading to relax not only the regularity but also the positivity assumption in [4].

In the present paper we consider general transition kernels of the following form

$$p(g|\mathbf{w}_{-\infty}^{-1}) = \sum_{k \in \mathcal{A}} \theta_k P_{(k)}(w_{-k}, g), \tag{3}$$

for some probability distribution θ supported by $\mathcal{A} \subset \mathbb{N}_+$ and $P_{(k)}$ is a stochastic matrix on G, for any $k \in \mathcal{A}$. Since $\sum_{k \in \mathcal{A}} \theta_k = 1$, any kernel of the form (3) is clearly continuous. A transition kernel of the above form will be called an *imitation kernel*.

When the $P_{(k)}$'s have the property that each row contains only a single positive entry, necessarily equal to 1 (as happens in particular for permutation matrices), the updating rule (1) means $X_n^{r,\mathbf{w}} = f_k(X_{n-k}^{r,\mathbf{w}})$ with probability θ_k, where f_k is a function on G obtained from $P_{(k)}$. We refer to these cases as imitation kernels *without noise*. Otherwise we speak about imitation kernels *with noise*. Imitation kernels without noise are in some sense, to be clarified later, the most interesting to consider.

For general kernels of the form (3) the value $X_{r+1}^{r,\mathbf{w}}$ can be drawn in the following way. An integer K_{r+1} is chosen at random according to the distribution θ, and the value of the boundary condition $w_{r+1-K_{r+1}}$ is read. Then $X_{r+1}^{r,\mathbf{w}}$ is drawn from the $w_{r+1-K_{r+1}}$-th row $P_{(K_{r+1})}(w_{r+1-K_{r+1}}, \cdot)$ of the matrix $P_{(K_{r+1})}$. To perform this step, it is convenient to make reference to a sequence $(f_{(k)}, k \in \mathbb{N}_+)$ of *coupling functions* $f_{(k)} : G \times [0, 1] \to G$, having the property that, for any $k \in \mathbb{N}_+$, $f_{(k)}(g, U)$ is distributed as $P_{(k)}(g, \cdot)$ whenever U is a random variable uniformly distributed in $[0, 1]$, for $g \in G$. So, if U_{r+1} is uniformly distributed in $[0, 1]$, $f_{(K_{r+1})}(w_{r+1-K_{r+1}}, U_{r+1})$ yields $X_{r+1}^{r,\mathbf{w}}$. This updating rule can be iterated to produce the values of the process $\mathbf{X}^{r,\mathbf{w}}$ at all sites $n > r$ by drawing a random sample $(K_n, n > r)$ from θ and an independent random sample $(U_n, n > r)$ from the uniform distribution on $[0, 1]$. It is clear that for imitation kernels without noise, the U_n's are not needed for the construction of $\mathbf{X}^{r,\mathbf{w}}$.

Rather than proceeding *forward* from the boundary sites, one can proceed *backwards* from any site of interest. In this case to produce the random variable $X_n^{r,\mathbf{w}}$, with $n > r$, we have to follow the random walk $\mathcal{T}^{(n)} = (T_k^{(n)}, k \in \mathbb{N})$

$$T_{k+1}^{(n)} = T_k^{(n)} - K_{T_k^{(n)}}, \text{ for } k \in \mathbb{N}, \tag{4}$$

with $T_0^{(n)} = n$, whose distribution of decrements is θ.

Let us define

$$M_r^{(n)} = \inf\{k : T_k^{(n)} \le r\}, \; V_r^{(n)} = T_{M_r^{(n)}}^{(n)} \tag{5}$$

which are the minimum number of steps leading the random walk to land on a site below the *threshold* r and the landing site, respectively. The information on the boundary condition $\mathbf{w}$ is propagated forward by applying recursively the coupling functions in the following way

$$X_{T_{k-1}^{(n)}}^{r,\mathbf{w}} = f_{(T_{k-1}^{(n)} - T_k^{(n)})}\left(X_{T_k^{(n)}}^{r,\mathbf{w}}, U_{T_k^{(n)}}\right), \quad k = M_r^{(n)}, \ldots, 1 \tag{6}$$

starting from $X_{T_{M_r^{(n)}}^{(n)}}^{r,\mathbf{w}} = w_{V_r^{(n)}}$. Thus, at the end of the recursion one has, for any $n > r$

$$X_n^{r,\mathbf{w}} = F_{r,n}(K_m, U_m, r < m \le n; w_{V_r^{(n)}}) \tag{7}$$

for some suitably defined function $F_{r,n}$.

One can appreciate here that, if an additional zero order term $\theta_0 \nu(g)$, with $\theta_0 > 0$ and ν probability measure on G, appears in the kernel (3), the K_n's can also assume the value 0 with probability θ_0. When this happens, one stops the random walk from going further in the past and reads directly the value at that site by sampling from ν. Since this event will happen a.s., uniqueness always holds in this case. Incidentally, $\theta_0 > 0$ means $\sum_{k=1}^{\infty} \theta_k < 1$, the Dobrushin sufficient criterion for uniqueness for this kind of kernels. It is not difficult to realize that a zero order term *cannot be singled out* when each of the columns of $P_{(k)}$ has a zero entry, for all $k \in \mathcal{A}$. In particular, this happens for imitation kernels without noise, except in the trivial case of some $P_{(k)}$ with all the rows equal to the same unit versor.

Whenever for some pair of distinct sites $m, n \in \mathbb{Z}$, it happens that $T_h^{(m)} = T_k^{(n)}$, for some positive integers h and k, we say that the two random walks started from the sites m and n *coalesce*. If this is the rightmost site in which this happens, we say that $T_h^{(m)} = T_k^{(n)}$ is the *coalescence point* of the two random walks. In this case one has $T_{h+l}^{(m)} = T_{k+l}^{(n)}$, for any $l \in \mathbb{N}$, hence for $r \le T_h^{(m)}$, it is $V_r^{(m)} = V_r^{(n)}$. As a consequence the values $w_{V_r^{(m)}}$ and $w_{V_r^{(n)}}$ coincide, conveying all the information about the boundary condition $\mathbf{w}$ needed to compute both $X_m^{r,\mathbf{w}}$ and $X_n^{r,\mathbf{w}}$, by means of the functions $F_{r,m}$ and $F_{r,n}$ defined in (7).

If the random walks $\mathcal{T}^{(n)}$, started from $n \in \Lambda$, with Λ arbitrary finite subset of $\mathbb{Z}$, coalesce a.s. we say that the distribution $\theta = (\theta_n, n \in \mathcal{A})$ is *coalescent*. In order to verify this property it is enough to check it for a window Λ made by two adjacent sites of $\mathbb{Z}$. If we let two particles perform two independent random walks started from these two sites, with the rule that it is always the rightmost that moves, the distance between the two particles is a Markov chain on the integers, the so called von Schelling process [10], up to coalescence. This is again a random walk with decrements following the law θ, but with a reflection around the origin once the negative half-line is hit. Thus coalescence of θ means that from any $n \in \mathbb{N}_+$ the return of this process to the origin is almost sure; this requires both some algebraic property for $\mathcal{A}$ and a control of the tail behavior of the θ_n's, see [23].

Back to imitation kernels, we mention that already in [4] a particular class of binary kernels of the form (3) was examined, in which $P_{(k)}$ took only the two possible values

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, J_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{8}$$

for $k \in \mathcal{A}$, the so-called *binary autoregressive kernels*. However the presence of a zero order term $\theta_0 \nu(g)$ made the uniqueness problem trivial. In our previous work [9] we have

considered binary autoregressive kernels with $\theta_0 = 0$, equivalently with the $P_{(k)}$'s equal to either I_2 or J_2, making a first step towards understanding the implications of the lack of positivity for imitation kernels. The main result of that paper is that for a coalescent θ uniqueness holds.

In the present paper the results are completely general, and not restricted to the binary case. The main result is that uniqueness for imitation kernels can be characterized completely in terms of the properties of what we call the *G-stochastic function* induced by the imitation kernel (3), namely the mapping

$$k \in \mathcal{A} \subset \mathbb{N}_+ \mapsto P_{(k)}.$$

The necessary and sufficient conditions generalize the well known concepts of *irreducibility* and *aperiodicity* for a single stochastic matrix.

Irreducibility is discussed in Sect. 2. Since the presence of two irreducible classes implies the existence of two different compatible laws (Proposition 2) and states not belonging to an irreducible class cannot appear in the support of a compatible law (Proposition 3), we are allowed to focus our further study to irreducible kernels.

Aperiodicity is the subject of Sect. 3. Here a difference with the case of a single stochastic matrix appears: the states are constrained to have a period which is a multiple of the gcd $d(\mathcal{A})$ of $\mathcal{A}$. But since any element of $\mathcal{G}(p)$ has independent marginals along the residual classes mod $d(\mathcal{A})$, the uniqueness problem is reduced to any of them, for which with an obvious rescaling $d(\mathcal{A}) = 1$. This is the content of Proposition 4, which allows to correct a mistake occurred in [9]. Furthermore, as it happens for finite Markov chains, the presence of several periodic classes implies the existence of different *non-stationary* elements of $\mathcal{G}(p)$, obtained one from the other by shifts (Theorem 1). Notice that this kind of *phase transition* is entirely different from the one in [3] that concerns a positive kernel. At the end of the section we prove an important lemma relating the stationary elements of $\mathcal{G}(p)$ with the invariant distributions of the stochastic matrix $\hat{P} = \sum_{k \in \mathcal{A}} \theta_k P_{(k)}$.

In Sect. 4 we prove our main result (Theorem 2), which is analogous to the ergodicity theorem for finite Markov chains: uniqueness holds for irreducible and aperiodic imitation kernels. The unique invariant distribution $\hat{\lambda}$ of $\hat{P}$ is identified as the single-site marginal of the unique compatible law.

The fact that the tail behavior of the θ_k's does not enter in the uniqueness conditions entails that, by keeping $\mathcal{A}$ fixed, but distributing enough mass to larger values of $k \in \mathcal{A}$, we can construct examples of uniqueness in which any of the general sufficient conditions appeared in the literature fails. The uniqueness theorem appearing in [9] is found as a particular case, without assuming coalescence (Theorem 3).

In Sect. 5 we propose two simulation algorithms of the CFTP type [24], to construct the unique compatible law on any finite window of $\mathbb{Z}$. The first, presented in Theorem 6, works when the distribution θ is known to be coalescent. When θ is not coalescent or at least this is unknown, a threshold has to be specified, introducing an error in the algorithm. In Theorem 7 we prove that the error introduced in this way can be made arbitrarily small pushing the threshold towards $-\infty$. For this reason we call it an ε-perfect simulation algorithm. In a situation of non uniqueness, the algorithms presented here can still be used to construct any stationary element of $\mathcal{G}(p)$. In the irreducible but non-aperiodic case, it can also be proved that there is only one stationary element.

Finally, in Sect. 6 we present a result for the case of countable G. A sufficient condition for the existence of a unique element in $\mathcal{G}(p)$, together with a perfect simulation algorithm is obtained. For finite G, such a condition reduces to irreducibility and aperiodicity. The

algorithm eliminates the approximation error for non-coalescent θ but it can be considerably more complicate for G large or infinite.

2 Irreducibility of G-Stochastic Functions and Uniqueness

We start with a brief discussion of mappings defined on some $\mathcal{A} \subset \mathbb{N}_+$ with values in the set of stochastic matrices over the set G. We call a mapping of this type a G-*stochastic function*.

Recall that the free semigroup generated by $\mathcal{A}$ is the set $\mathcal{A}^* = \bigcup_{n \in \mathbb{N}^+} \mathcal{A}^n$ of finite n-tuples with elements in $\mathcal{A}$, for all positive integers n, called *words* in the sequel. It is indeed a semigroup under concatenation, defined for $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{b} = (b_1, \ldots, b_m)$ by $\mathbf{ab} = (a_1, \ldots, a_n, b_1, \ldots, b_m)$. A G-stochastic function defined on $\mathcal{A}$ extends to a homomorphism of the semigroup $\mathcal{A}^*$ into the semigroup of stochastic matrices on G by associating to each $\mathbf{a} = (a_1, \ldots, a_n)$ the stochastic matrix

$$P_{\mathbf{a}} = P_{(a_n)} \cdots P_{(a_1)}. \tag{9}$$

Likewise, for any $\mathbf{a} = (a_1, \ldots, a_n) \in \mathcal{A}^*$, we define the composition of coupling functions $f_{\mathbf{a}} : G \times [0,1]^n \to G$ as

$$f_{\mathbf{a}}(g; u_1, \ldots, u_n) = f_{(a_1)}(f_{(a_2)}(\ldots f_{(a_{n-1})}(f_{(a_n)}(g; u_n), u_{n-1}), \ldots, u_2), u_1),$$

for $g \in G$ and $u_i \in [0,1]$, $i = 1, \ldots, n$. We define the *depth* of $\mathbf{a}$ as $s(\mathbf{a}) = \sum_{i=1}^{n} a_i$.

Let us define a directed graph Γ_θ with the sites of $\mathbb{Z}$ as vertices, and arcs joining $n \in \mathbb{Z}$ with $n - k$, whenever $k \in \mathcal{A}$. It is natural to visualize the elements $\mathbf{a} \in \mathcal{A}^*$ as paths of the graph Γ_θ. Once we have weighted the arc $(n, n - k)$ with the probability $\theta_k > 0$, we can assign a probability to any path, given by the product of the probabilities of the arcs belonging to the path. Now, for $\mathbf{a} = (a_1, \ldots, a_m) \in \mathcal{A}^*$, $P_{\mathbf{a}}$ is the stochastic matrix used to compute the value of $X_n^{r,\mathbf{w}}$ from the value $X_{n-s(\mathbf{a})}^{r,\mathbf{w}}$, for $n - s(\mathbf{a}) > r$, whenever $K_n = a_1, K_{n-a_1} = a_2, \ldots, K_{n-a_1-\cdots-a_{m-1}} = a_m$: an event which has probability $\theta_{\mathbf{a}} = \theta_{a_1} \cdot \ldots \cdot \theta_{a_m}$. Notice that infinitely many paths are associated to each $\mathbf{a} \in \mathcal{A}^*$, differing in the starting site in $\mathbb{Z}$. Also observe that the *depth* $s(\mathbf{a})$ is the distance of the last site of the path from the first one (see Fig. 1)

The sample $\mathbf{K} = (K_n, n \in \mathbb{Z})$ selects a particular random subgraph $\Gamma^{\mathbf{K}}$ of Γ_θ, made of the arcs $(m, m - K_m)$, for all $m \in \mathbb{Z}$. Likewise, the random walks $T^{(n)} = (T_k^{(n)}, k \in \mathbb{N})$, for $n \in \mathbb{Z}$, can be actually seen as random walks on the graph Γ_θ.

We can extend to a G-stochastic function $P_{(.)}$ a number of concepts which are well known for the "standard" case of a single stochastic matrix P, i.e. the G-stochastic function defined on $\mathcal{A} = \{1\}$, with $P_{(1)} = P$.

Definition 1 Let $P_{(.)}$ be a G-stochastic function defined on $\mathcal{A}$ and $i \neq j \in G$. We say that $i \in G$ $P_{(.)}$-communicates with $j \in G$ if there exists $\mathbf{a} \in \mathcal{A}^*$ such that $P_{\mathbf{a}}(i, j) > 0$. We say that i and j $P_{(.)}$-intercommunicate when i communicates with j and vice versa.

Fig. 1 The word $\mathbf{a} = (2, 4, 3)$ and the corresponding path

As usual for the standard case, we declare that each $i \in G$ intercommunicates with itself, so that intercommunication becomes an equivalence relation and G is partitioned in intercommunicating classes.

Definition 2 An intercommunicating class C is closed when $i \in C$ communicates with $j \in G$ implies $j \in C$ (therefore j communicates with i). We say that a G-stochastic function $P_{(.)}$ is irreducible if for any i and $j \in G$ there exists $\mathbf{a} \in \mathcal{A}^*$ such that $P_a(i, j) > 0$, that is the whole G is the only intercommunicating class. More generally, $P_{(.)}$ is essentially irreducible if there exists a single intercommunicating class.

It is easily verified that $P_{(.)}$ is irreducible if and only if $\hat{P}$ is irreducible. As in the standard case it is possible to decompose G in a rather familiar way.

Proposition 1 *Let $P_{(.)}$ be a G-stochastic function. There exists a unique partition $\{R_1, \ldots, R_s, T\}$ of G, with $s \geq 1$, where*

(1) R_h *is a closed intercommunicating class, thus $\{P_{(k)}|_{R_h}, k \in \mathcal{A}\}$ is irreducible, for $h = 1, \ldots, s$;*
(2) *for any $i \in T$ there exists $j \in R_1 \cup \ldots \cup R_s$ and $\mathbf{a} \in \mathcal{A}^*$ such that $P_{\mathbf{a}}(i, j) > 0$.*

Proof The proof is completely analogous to that for a single finite stochastic matrix. In that case the sets R_h, $h = 1, \ldots, s$ represent the recurrent states, whereas the set T, which is the union of the intercommunicating classes that are not closed, represents the remaining transient states.   □

The reader will notice that the above decomposition coincides with that concerning any convex combination of the stochastic matrices $P_{(k)}$, $k \in \mathcal{A}$ with positive weights, in particular the matrix $\hat{P}$. The following result allows to rule out a trivial case of non uniqueness for kernels of the form (3).

Proposition 2 *If the G-stochastic function $P_{(.)}$ has more than one closed intercommunicating class then for the kernel (3) one has $|\mathcal{G}(p)| > 1$.*

Proof The restriction of a kernel (3) to a closed intercommunicating class R_h is by itself a transition kernel p_{R_h} on R_h. Identifying $\mathcal{G}(p_{R_h})$ with a subset of $\mathcal{G}(p)$ in the natural way, and taking into account that $\mathcal{G}(p_{R_h})$ are non empty and disjoint, for $h = 1, \ldots, s$, the statement of the theorem is immediately obtained.   □

The simplest example of this sort is contained in [9]: if $P_{(k)} = I_2$ for all $k \in \mathcal{A}$, then the two states are two closed classes and the Dirac measures on the two constant sequences are two elements of $\mathcal{G}(p)$.

The following proposition ensures that, when there is only one communicating class, we can restrict the kernel to it.

Proposition 3 *For a transition kernel of the form (3), consider the corresponding G-stochastic function $P_{(.)}$ and suppose that R is the union of all closed intercommunicating classes. Let p_R be the restriction of p to R. Then $\mathcal{G}(p) = \mathcal{G}(p_R)$ for the kernel (3).*

Proof The main step is to prove that if $\mathbf{X} \sim \mu \in \mathcal{G}(p)$, then, for any $n \in \mathbb{Z}$, $P(X_n \in R^c) = 0$. For this it is enough to prove that

$$\lim_{r \to -\infty} P(X_n^{r, \mathbf{w}} \in R^c) = 0 \tag{10}$$

for any $\mathbf{w} \in G^{\mathbb{Z}}$.

Starting with $\mathcal{Z}_0 = R^c$, we define recursively a sequence of subsets $\mathcal{Z}_h \subset R^c$, for $h = 1, \ldots$, with strictly decreasing cardinality, until for some integer L it is $\mathcal{Z}_L = \emptyset$. During this construction we will define $\mathbf{a}_h \in \mathcal{A}^*$ of length n_h and Borel sets $\Delta_h \subset [0, 1]^{n_h}$ of positive Lebesgue measure Leb_{n_h}, for $h = 0, 1, \ldots, L - 1$. The recursive construction is given by

$$\mathcal{Z}_{h+1} = \{f_{\mathbf{a}_h}(i, \Delta_h), i \in \mathcal{Z}_h\} \cap R^c$$

and has the properties

1. $\{f_{\mathbf{a}_h}(i, \Delta_h), i \in \mathcal{Z}_h\} \cap R \neq \emptyset, h = 0, \ldots, L - 1$
2. For any $i \in \mathcal{Z}_h, |f_{\mathbf{a}_h}(i, \Delta_h)| = 1$.

Let us explain the generic step h of the construction. Fix an arbitrary state $j \in \mathcal{Z}_h$. By the definition of R there exists a word $\mathbf{a}_h \in \mathcal{A}^*$ of length n_h such that $P_{\mathbf{a}_h}(j, R) > 0$. This ensures the existence of Δ_h^* with a positive n_h-dimensional Lebesgue measure such that $f_{\mathbf{a}_h}(j, \Delta_h^*) \in R$, hence Property 1 is guaranteed. To obtain Property 2 one may need to reduce Δ_h^* to some smaller $\Delta_h \subset \Delta_h^*$ keeping positive Lebesgue measure, which is clearly always possible by finiteness of $\mathcal{Z}_h$.

Now we are in a position to prove (10). Define the word $\mathbf{a}$ of length $m = n_0 + n_1 + \cdots + n_{L-1}$ by the concatenation $\mathbf{a} = \mathbf{a}_{L-1} \ldots \mathbf{a}_0$ and the Borel set $\Delta = \Delta_{L-1} \times \cdots \times \Delta_0 \subset [0, 1]^m$. Let $c = \theta_{\mathbf{a}} \cdot \text{Leb}_m(\Delta) > 0$.

Recall the recursive construction of $X_n^{r,\mathbf{w}}$ in terms of the random walk $\mathcal{T}^{(n)} = \{T_k^{(n)}, k \in \mathbb{N}\}$ with the corresponding sequences $\{K_k = K_{T_k^{(n)}}, k \in \mathbb{N}\}$, obtained through the relation (4), and let $\{U_k = U_{T_k^{(n)}}, k \in \mathbb{N}\}$. If, for some integer l, a segment $(K_l, \ldots, K_{l+m}, U_l, \ldots, U_{l+m})$ belongs to $\mathbf{a} \times \Delta$ with $T_{l+m}^{(n)} > r$, then $X_n^{r,\mathbf{w}} \in R$, irrespectively of $\mathbf{w}$. Since a segment of this kind will eventually occur with probability 1, (10) holds.

As a consequence, if $\mu \in \mathcal{G}(p)$ then $\mu(R^{\mathbb{Z}}) = 1$. Moreover, being μ compatible and supported by $R^{\mathbb{Z}}$, it is actually in $\mathcal{G}(p_R)$. $\qquad\square$

3 Periodicity of *G*-Stochastic Functions and Uniqueness

In the previous section we have justified to restrict our attention to irreducible G-stochastic functions $P_{(\cdot)}$. In this section we turn our attention to the notion of periodicity of a state $i \in G$, which is slightly more delicate than in the standard case. We start by observing that the depth s is a homomorphism of the free semigroup $\mathcal{A}^*$ into the additive semigroup of positive integers $\mathbb{N}_+$. As a consequence $s(\mathcal{A}^*)$ is a sub-semigroup of $\mathbb{N}_+$. The period of $\mathcal{A}$ is defined as

$$d(\mathcal{A}) = gcd\{s(\mathcal{A})\} = gcd\{s(\mathcal{A}^*)\}. \tag{11}$$

If $1 \in \mathcal{A}$, as it happens in the standard case, then $s(\mathcal{A}^*)$ coincides with $\mathbb{N}_+$, and $d(\mathcal{A}) = 1$. Recall that, except for a finite number of elements, an additive semigroup of positive integers has always the form $\{n_0 d, (n_0 + 1)d, \ldots\}$, where d is the gcd of the semigroup and n_0 is a suitable positive integer.

Now suppose that the transition kernel p in (3) has $d(\mathcal{A}) > 1$. Then it is natural to move from p to

$$\bar{p}(g|\bar{\mathbf{w}}_{-\infty}^{-1}) = \sum_{l \in \bar{\mathcal{A}}} \theta_{ld(\mathcal{A})} P_{(ld(\mathcal{A}))}(\bar{w}_{-l}, g), \tag{12}$$

with $\bar{\mathcal{A}} = \{l : ld(\mathcal{A}) \in \mathcal{A}\}$, so that $d(\bar{\mathcal{A}}) = 1$. The following proposition allows us to set $d(\mathcal{A}) = 1$ in all the uniqueness proofs of the present paper, without restriction of generality.

Proposition 4 *For a transition kernel p of the form* (3) *let $d(\mathcal{A}) > 1$ and define $\bar{p}$ as in* (12). *Then uniqueness holds for p if and only if it holds for $\bar{p}$.*

Proof Let us consider the process $\mathbf{X}^{r,\mathbf{w}}$ governed by the transition kernel p. Then for any $h = 0, 1, \ldots, d(\mathcal{A}) - 1$ the $d(\mathcal{A})$-marginal process $\bar{\mathbf{X}}_{(h)}$, defined by

$$\bar{X}_{(h)k} = X^{r,\mathbf{w}}_{h+kd(\mathcal{A})}, \ k \in \mathbb{Z} \tag{13}$$

is governed by the kernel $\bar{p}$, with boundary conditions $\bar{\mathbf{w}}^{(h)} = (w_{h+kd(\mathcal{A})}, k \in \mathbb{Z})$, with $r^{(h)} = \max\{k : h + kd(\mathcal{A}) \le r\}$. If uniqueness holds for p, then $\mathbf{X}^{r,\mathbf{w}}$ converges weakly as $r \downarrow -\infty$ to the unique element μ of $\mathcal{G}(p)$, for any choice of $\mathbf{w} \in G^{\mathbb{Z}}$. Let $\mathbf{Y}$ be a process with distribution μ. Likewise the process $\bar{\mathbf{X}}_{(h)}$ converges weakly to $\mathbf{Y}^{(h)} = (Y_{h+kd(\mathcal{A})}, k \in \mathbb{Z})$, as $r^{(h)} \downarrow -\infty$ which proves that uniqueness holds also for $\bar{p}$.

For the converse notice that, for any r, conditionally to $\mathbf{w} \in G^{\mathbb{Z}}$, the $d(\mathcal{A})$-marginal processes $\bar{\mathbf{X}}_{(h)}$, defined in (13), are independent, for $h = 0, 1, \ldots, d(\mathcal{A}) - 1$. By consequence, if $\bar{\mu}$ is the unique element in $\mathcal{G}(\bar{p})$, each of the $d(\mathcal{A})$-marginal processes converges to it, so the whole process $\mathbf{X}^{r,\mathbf{w}}$ has a limit distribution with the $d(\mathcal{A})$-marginal processes $\bar{\mu}$ distributed and independent, which ends the proof. □

The previous proposition corrects the erroneous statement contained in our paper [9] (see Proposition 1 and Theorem 1) that $d(\mathcal{A}) = 1$ is necessary for uniqueness.

Next assume that $d(\mathcal{A}) = 1$ and define the period of $i \in G$ to be $d_i = gcd(s(\mathcal{A}_i^*))$, where $\mathcal{A}_i^* = \{\mathbf{a} \in \mathcal{A}^* : P_{\mathbf{a}}(i, i) > 0\}$. If $d_i = 1$ we say that the state $i \in G$ is *aperiodic* for $P_{(\cdot)}$. The following proposition guarantees that, for irreducible G-stochastic functions, we can refer the term to the whole function, since all states have the same period.

Proposition 5 *Let $P_{(\cdot)}$ be an irreducible G-stochastic function defined on $\mathcal{A}$, with $d(\mathcal{A}) = 1$. Then $d_i = \hat{d}$, for $i \in G$, for some $\hat{d} \in \mathbb{N}_+$. Moreover, there exists a partition of G in sets $\{G_h, h = 0, \ldots, \hat{d} - 1\}$ such that*

$$P_{\mathbf{b}}(i, j) > 0, i \in G_h \Rightarrow j \in G_{h+s(\mathbf{b})},$$

identifying $G_{h+k\hat{d}}$ with G_h.

Proof By irreducibility for any pair $i, j \in G$ there exists $\mathbf{a}_1, \mathbf{a}_2 \in \mathcal{A}^*$ such that $P_{\mathbf{a}_1}(i, j) P_{\mathbf{a}_2}(j, i) > 0$. This implies that $P_{\mathbf{a}_2\mathbf{a}_1}(i, i) \ge P_{\mathbf{a}_1}(i, j) P_{\mathbf{a}_2}(j, i) > 0$, so $s(\mathbf{a}_2\mathbf{a}_1) = s(\mathbf{a}_1) + s(\mathbf{a}_2) = kd_i$. Moreover if $P_{\mathbf{b}}(i, i) > 0$, then $P_{\mathbf{a}_1\mathbf{b}\mathbf{a}_2}(j, j) \ge P_{\mathbf{a}_2}(j, i) P_{\mathbf{b}}(i, i) P_{\mathbf{a}_1}(i, j) > 0$, thus $\mathbf{a}_1 \mathcal{A}_i^* \mathbf{a}_2 \subset \mathcal{A}_j^*$, from which

$$s(\mathbf{a}_1) + s(\mathcal{A}_i^*) + s(\mathbf{a}_2) = kd_i + s(\mathcal{A}_i^*) \subset s(\mathcal{A}_j^*).$$

This implies that $d_j \le d_i$. Exchanging the roles between i and j one gets $d_i = d_j$ as promised.

For the second statement let us fix some reference state $k \in G$, and, for any $h = 0, 1, \ldots, \hat{d} - 1$ define the subsets of G

$$G_h = \{i \in G : P_{\mathbf{a}}(k, i) > 0 \text{ for some } \mathbf{a} \in \mathcal{A}^* \text{ with } s(\mathbf{a}) = n\hat{d} + h, n \in \mathbb{N}\}.$$

By irreducibility the union of the G_h's is the whole G. Now suppose that $i \in G_{h_1} \cap G_{h_2}$. Then there exist $\mathbf{a}_1 \in \mathcal{A}^*$ with $s(\mathbf{a}_1) = n_1\hat{d} + h_1$ and $\mathbf{a}_2 \in \mathcal{A}^*$ with $s(\mathbf{a}_2) = n_2\hat{d} + h_2$ such

that $P_{\mathbf{a}_1}(k, i) P_{\mathbf{a}_2}(k, i) > 0$. We can safely assume that $n_1 = n_2 = n$, since $s(\mathcal{A}_k^*)$ contains all the multiples of $\hat{d}$ large enough. Next let $\mathbf{b} \in \mathcal{A}^*$ such that $P_{\mathbf{b}}(i, k) > 0$: we will have that

$$P_{\mathbf{ba}_1}(k, k) P_{\mathbf{ba}_2}(k, k) \geq P_{\mathbf{b}}(i, k)^2 P_{\mathbf{a}_1}(k, i) P_{\mathbf{a}_2}(k, i) > 0,$$

which implies that $s(\mathbf{ba}_1)$ and $s(\mathbf{ba}_2)$ are multiples of $\hat{d}$. Hence

$$s(\mathbf{ba}_1) - s(\mathbf{ba}_2) = s(\mathbf{a}_1) - s(\mathbf{a}_2) = h_1 - h_2$$

must be a multiple of $\hat{d}$. Since $|h_2 - h_1| < \hat{d}$ this happens only when $h_1 = h_2$. So we have proved that $\{G_h, h = 0, 1, \ldots, \hat{d} - 1\}$ is a partition.

Next assume $P_{\mathbf{b}}(i, j) > 0, i \in G_h, j \in G_l$. By assumption there exists $\mathbf{a} \in \mathcal{A}^*$ such that $P_{\mathbf{a}}(k, i) > 0$ with $s(\mathbf{a}) = n\hat{d} + h$. Then $P_{\mathbf{ba}}(k, j) \geq P_{\mathbf{a}}(k, i) P_{\mathbf{b}}(i, j) > 0$, so

$$s(\mathbf{ba}) = s(\mathbf{b}) + s(\mathbf{a}) = s(\mathbf{b}) + n\hat{d} + h = m\hat{d} + l$$

for some integer m, from which

$$s(\mathbf{b}) = (m - n)\hat{d} + (l - h) \Rightarrow l = h + s(\mathbf{b}), \mod \hat{d},$$

as desired.   □

We can immediately make profit of the previous proposition to establish the following result.

Theorem 1 *Let p be a transition kernel of the form* (3) *and let $d(\mathcal{A}) = 1$. If $P_{(.)}$ is irreducible but not aperiodic (i.e. $\hat{d} > 1$) then $|\mathcal{G}(p)| > 1$.*

Proof By Proposition 5 the classes $G_0, G_1, \ldots, G_{\hat{d}-1}$ are well defined. Let us select an element from each class, say $g_i \in G_i$, for $i = 0, \ldots, \hat{d} - 1$ and define $w_k = g_i$ if k is congruent to $i \; mod(\hat{d})$, for $k \in \mathbb{Z}$. Finally define $\mathbf{w} = (w_k : k \in \mathbb{Z})$ and the translated $\hat{\mathbf{w}}$, with $\hat{w}_k = w_{k+1}$, for $k \in \mathbb{Z}$.

Recall that any probability measure in $\mathcal{G}(p)$ can be obtained as a weak limit of the laws of $\mathbf{X}^{r,\mathbf{w}}$, for $r \downarrow -\infty$. These laws give probability one to the event $\{X_0 \in G_0\}$, so this remains true in the limit. On the other hand the weak limits of the laws of $\mathbf{X}^{r,\hat{\mathbf{w}}}$ give probability one to the event $\{X_0 \in G_1\}$, therefore the measure are necessarily distinct. This ends the proof.
□

An example of application of Theorem 1 is the situation examined in [9]. If the image of $\mathcal{A}$ under $P_{(.)}$ is made by the matrices I_2 and J_2, then $P_{(.)}$ is irreducible. This happens also if this image is the singleton $\{J_2\}$. Now suppose $d(\mathcal{A}) = 1$ and, for $k \in \mathcal{A}$ and odd, $P_{(k)} = J_2$ and for $k \in \mathcal{A}$ and even, $P_{(k)} = I_2$. Then, for any state $i \in G$, $s(\mathcal{A}_i^*)$ is made only by even numbers, thus d_i is a multiple of 2. It is precisely 2 since $d(\mathcal{A}) = 1$. Thus $P_{(.)}$ is not aperiodic. The two elements of $\mathcal{G}(p)$ constructed in Theorem 1 are Dirac measures supported by the two coherent sequences alternating the two states, as defined in [9]. Clearly these measures are not stationary: as a matter of fact the convex combination of them with equal weights is the unique stationary element in $\mathcal{G}(p)$. More generally, under the hypotheses of Theorem 1, we will prove in Sect. 5 that $\mathcal{G}(p)$ contains only one stationary measure.

We close this section focusing our attention on the set $\mathcal{I}$ of invariant distributions for the stochastic matrix $\hat{P} = \sum_{k \in \mathcal{A}} \theta_k P_{(k)}$.

Lemma 1 *The following statements hold:*

(1) *Let $\lambda \in \mathcal{I}$ and $\mathbf{w}$ be a configuration with λ-distributed single-site marginals. Then, for any $r \in \mathbb{Z}$, $\mathbf{X}^{r,\mathbf{w}}$ has the same property.*
(2) *For any $\lambda \in \mathcal{I}$ there is at least one element of $\mathcal{G}(p)$ with single-site marginals equal to λ. This element can be always chosen to be stationary.*
(3) *A stationary element of $\mathcal{G}(p)$ has all its single-site marginals equal to an element of $\mathcal{I}$.*
(4) *If there is a unique (stationary) element in $\mathcal{G}(p)$, there is a unique invariant measure for $\hat{P}$.*

Proof For (1) it is enough to notice that, by induction on $n > r$, and interchanging the two sums of positive terms,

$$P(X_n^{r,\mathbf{w}} = g) = E(P(X_n^{r,\mathbf{w}} = g \mid X_i^{r,\mathbf{w}}, \ i < n)) = \sum_{k \in \mathcal{A}} \theta_k(\lambda P_{(k)})(g) = (\lambda \hat{P})(g) = \lambda(g). \tag{14}$$

For (2) take $\mathbf{X}^{r,\mathbf{w}}$ as above and send r to $-\infty$: any weak limit point will be in $\mathcal{G}(p)$ and it will keep the single-site marginals equal to λ. By taking Cesaro averages on a window increasing to $\mathbb{Z}$ and going to the limit one gets at least one stationary process with single site marginals still equal to λ. For (3), let $\mathbf{X} \in \mathcal{G}(p)$ be stationary and call λ its single-site marginals: then, similarly to (14)

$$\lambda(g) = P(X_n = g) = E(P(X_n = g \mid X_i, \ i < n)) = \sum_{k \in \mathcal{A}} \theta_k(\lambda P_{(k)})(g) = (\lambda \hat{P})(g). \tag{15}$$

(4) is an immediate consequence of (3). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

By the remark following Proposition 1, it is immediately seen that $\mathcal{I}$ has a single element if and only if $P_{(\cdot)}$ is essentially irreducible. So this is a necessary condition for the existence of a unique stationary element in $\mathcal{G}(p)$. At the end of Sect. 5 we will able to prove that this condition is also sufficient.

4 Main Results

After the results of the previous section, it remains to consider transition kernels of the form (3) with a corresponding G-stochastic function $P_{(\cdot)}$ which is irreducible and aperiodic with $d(\mathcal{A}) = 1$. In this section we prove that uniqueness holds for all of them. Here is the main result.

Theorem 2 *Let p have the form (3) and assume that $P_{(\cdot)}$ is irreducible and aperiodic. Then $\mathcal{G}(p)$ has a unique element μ, with single-site marginals equal to $\hat{\lambda}$, the unique invariant distribution of $\hat{P}$.*

Before giving the proof of this result, we prove a weaker result in order to present the basic argument in a simpler context. Recall that under the above assumptions, $\hat{P}$ has a unique invariant distribution.

Lemma 2 *Let p have the form (3) and let $P_{(\cdot)}$ be irreducible and aperiodic. Then, for any $n \in \mathbb{Z}$ the distribution of $X_n^{r,\mathbf{w}}$ converges to $\hat{\lambda}$, as $r \to -\infty$, irrespectively of $\mathbf{w} \in G^{\mathbb{Z}}$.*

Proof Without loss of generality, we take $d(\mathcal{A}) = 1$ and $n = 0$. We are going to couple in a suitable way the random variables $X_0^{r,\mathbf{w}}$ for any $\mathbf{w} \in G^{\mathbb{Z}}$, in such a way that, as $r \to -\infty$, they all converge to the same random variable a.s.

Let i_0 be a fixed state in G. By irreducibility the additive semigroup $s(\mathcal{A}^*_{i_0})$ is non empty and, by aperiodicity, it has a gcd equal to 1: by the characterization of these semigroups, there exists $m_0 \in \mathbb{N}_+$ such that, for $m \geq m_0$, then $m \in s(\mathcal{A}^*_{i_0})$. Now use again irreducibility to prove that there exists $n_0 \in \mathbb{N}_+$ such that, for any $i, j \in G$ one can choose a word $\mathbf{b}_{i,j} \in \mathcal{A}^*$ with $P_{\mathbf{b}_{i,j}}(i,j) > 0$ and $s(\mathbf{b}_{i,j}) = n_0$. We call $\mathcal{B}$ the collection of these (distinct) words. These words can be identified with paths in Γ_θ with the same depth n_0. For any $\mathbf{b} \in \mathcal{B}$ we call $\theta_{\mathbf{b}}$ its probability and let $\rho = \sum_{\mathbf{b} \in \mathcal{B}} \theta_{\mathbf{b}}$ be the sum of the probabilities of these paths.

Next we are going to construct the random walk $\mathcal{T}^{(0)} = \{T_k^{(0)}, k \in \mathbb{N}\}$ by generating $\{K_{T_k^{(0)}}, k \in \mathbb{N}\}$, according to (4). Let $(l_h^{(0)}, u_h^{(0)}]$ be the interval between whose endpoints a path corresponding to a word in $\mathcal{B}$ appears for the h-th time in the random walk $\mathcal{T}^{(0)}$ (it is understood that such a word can vary with h). By definition $u_h^{(0)} - l_h^{(0)} = n_0$ (see Fig. 2). The independence of the decrements of the random walk ensures that, with probability 1, a sub-walk in $\mathcal{B}$ will appear infinitely many times. Once $\mathcal{T}^{(0)}$ reaches the site $V_r^{(0)}$, the value $X_0^{r,\mathbf{w}}$ can be obtained from $w_{V_r^{(0)}}$ by using the recursion in (6), with an exception for each of the sub-walks in $\mathcal{B}$ identified before. Suppose that one of these sub-walks joins $s = T_k^{(0)}$ with $s - n_0 = T_{k+m}^{(0)}$, with $s - n_0 > r$. Then, conditionally to this event, the transition from $X_{s-n_0}^{r,\mathbf{w}}$ to $X_s^{r,\mathbf{w}}$ follows the transition matrix

$$Q = \frac{1}{\rho} \sum_{\mathbf{b} \in \mathcal{B}} \theta_{\mathbf{b}} P_{\mathbf{b}}. \tag{16}$$

This matrix is positive by construction, therefore there exists a coupling function $f_* : G \times [0,1] \to G$ with the following property. If U is uniform in $[0,1]$, $f_*(g, U)$ has the law $Q(g, \cdot)$, and there exists $\epsilon > 0$ such that for $u < \epsilon$

$$f_*(g, u) = f_*(h, u), \forall g, h \in G. \tag{17}$$

It is enough to number the states and use the Skorohod representation. Thus any transition from the left endpoint to the right one of the segments $(l_h^{(0)}, u_h^{(0)}]$ can be performed by drawing some independent random variable U with uniform distribution in $[0,1]$. When $U < \epsilon$ occurs, then $X_s^{r,\mathbf{w}}$ does not depend on $X_{s-n_0}^{r,\mathbf{w}}$, and thus does not depend on $\mathbf{w}$. Once this *coupling* occurs, by following the recursion (6) one obtains that $X_0^{r,\mathbf{w}}$ does not depend on $\mathbf{w}$ as well.

Since each time a sub-walk in $\mathcal{B}$ occurs the U used by the coupling function f_* are independent, coupling will happen with probability 1 and thus, as $r \to -\infty$, $X_0^{r,\mathbf{w}}$ converges a.s. to a random variable which does not depend of $\mathbf{w}$. Finally take $\mathbf{w}$ with all the single-site marginal equal to $\hat{\lambda}$. Then, by Lemma 1, $X^{r,\mathbf{w}}$ has law $\hat{\lambda}$ for any r, and this law is kept by any weak limit point. This ends the proof. □

Proof of Theorem 2 Again, for any finite subset $\Lambda \subset \mathbb{Z}$, we have to couple all the processes $\mathbf{X}_\Lambda^{r,\mathbf{w}}$ for all $r \in \mathbb{Z}$ and $\mathbf{w} \in G^{\mathbb{Z}}$ in such a way that, as $r \to -\infty$, they converge a.s. to the same limit vector, hence they share the same limit in law.

Fig. 2 An example with $\mathcal{B} = \{(1,2), (2,1)\}$. Paths in $\mathcal{B}$ appear above the sites

Thus, we proceed first to the construction of the random walks $\mathcal{T}^{(n)} = \{T_k^{(n)}, k \in \mathbb{N}\}$, for $n \in \Lambda$, by using the i.i.d. θ-distributed random variables $(K_m, m \in \mathbb{Z})$. The intervals $(l_h^{(n)}, u_h^{(n)}]$ where a path in $\mathcal{B}$ occurs for the h-th time, $h = 1, 2, \ldots$ are located within the random walk $\mathcal{T}^{(n)}$, for any $n \in \Lambda$, with the constraint that *distinct* intervals which overlap are discarded.

We recall that the distance between two different random walks can be seen as a von Schelling process until the possible coalescence. The results in [2] imply that when $d(\mathcal{A}) = 1$ it is recurrent if and only if the distribution θ is coalescent. In this case any finite family of random walks coalesce a.s.; otherwise this process is transient and any pair of distinct random walks $\mathcal{T}^{(m)}$ and $\mathcal{T}^{(n)}$ either coalesce or their distance goes to infinity a.s.

The set Λ is partitioned according to the following equivalence relation: $m \equiv n$ whenever $\mathcal{T}^{(m)}$ and $\mathcal{T}^{(n)}$ coalesce. If this happens let $S_{m,n}$ denote their coalescence point, otherwise we set $S_{m,n} = -\infty$. For each of the equivalence classes $C_1, \ldots, C_l$, let

$$S_{C_h} = \begin{cases} \inf\{S_{m,n}, m, n \in C_h\}, & \text{if } |C_h| \geq 2, \\ n, & \text{if } C_h = \{n\}, \end{cases} \tag{18}$$

for $h = 1, \ldots, l$. For a coalescent distribution θ, $l = 1$ a.s. Otherwise the random walks $\mathcal{T}^{(S_{C_h})}$ do not coalesce, for $h = 1, \ldots, l$ and their distance goes to infinity with probability 1.

Therefore there are infinitely many non overlapping intervals where a path in $\mathcal{B}$ occurs, for each of these random walks. With the same argument used in the previous proof we get that, with probability 1, provided r is sufficiently close to $-\infty$, $X_{S_{C_h}}^{r, \mathbf{w}}$ converge to $\hat{\lambda}$, irrespectively of $\mathbf{w}$, independently for any $h = 1, \ldots, l$. Since any random walk $\mathcal{T}^{(m)}$, for $m \in \Lambda$, necessarily hits the set $\{S_{C_1}, \ldots, S_{C_l}\}$, by forward iteration of (6) it is then possible to obtain the required limiting values for $X_n^{r, \mathbf{w}}$, for all $n \in \Lambda$.   □

Next we apply the previous theorem to the class of binary autoregressive kernels examined in [9]. Theorem 3 in this reference proved uniqueness under conditions (a) and (b) therein. Now condition (a) is nothing but irreducibility and aperiodicity of the associated G-stochastic function. Condition (b) assumes coalescence so, as a consequence of Theorem 2, this condition appears to be unnecessary for uniqueness. Altogether we have the following uniqueness theorem.

Theorem 3 *Consider a transition kernel p of the form* (3) *with $|G| = 2$, $d(\mathcal{A}) = 1$ and the range of $P_{(\cdot)}$ contained in $\{I_2, J_2\}$, with I_2 and J_2 defined in* (8). *Then uniqueness holds for p if and only if there exists $m \in \mathcal{A}$ even with $P_{(m)} = J_2$, or at least $m_1, m_2 \in \mathcal{A}$, both odd, with $P_{(m_1)} = J_2$ and $P_{(m_2)} = I_2$.*

Proof It is clear that these conditions ensure that J_2 is contained in the range of $P_{(\cdot)}$ (which is equivalent to irreducibility) and that they exclude that the odd elements of $\mathcal{A}$ are sent by $P_{(\cdot)}$ into J_2 and the even into I_2 (which is the only periodic case).   □

By putting together all the results proved so far, we finally get the following

Theorem 4 *For a given kernel p of the form* (3) *one has uniqueness if and only if the corresponding G-stochastic function $P_{(\cdot)}$ is essentially irreducible and its single intercommunicating class is aperiodic.*

5 Perfect and ε-Perfect Simulation

In this section, under the conditions of Theorem 2 we construct simulation algorithms for the unique element μ of $\mathcal{G}(p)$ on a finite set of sites $\Lambda \subset \mathbb{Z}$. Recall that we have assumed without loss of generality that $d(\mathcal{A}) = 1$, so the coalescence property of the distribution θ is equivalent to the recurrence of the corresponding von Schelling process.

The problem of determining conditions for coalescence of θ has been recently addressed in [23], where it has been established that the following tail condition

$$\sum_{k=1}^{\infty} \left(\sum_{n=k}^{\infty} \theta_n \right)^2 < +\infty$$

is sufficient. Notice that this is weaker than the finiteness of the mean of θ, which has the same form but without the square. The latter is equivalent to the positive recurrence of the corresponding von Schelling process, which is in turn equivalent to the finiteness of the mean of the coalescence time of any two of the random walks $\mathcal{T}^{(n)}$, $n \in \mathbb{Z}$. In [23] an example of transient von Schelling process was also given.

In the uniqueness regime, if θ is coalescent a perfect simulation algorithm for the marginal distribution μ_Λ in any finite window $\Lambda \subset \mathbb{Z}$ can be designed. Indeed, in this case the coalescence point $S_{m,n}$ is finite a.s. for any $n, m \in \Lambda$, and so is the coalescence point

$$S_\Lambda = \inf\{S_{m,n}, m \neq n \in \Lambda\}. \tag{19}$$

Moreover these coalescence points are all adapted to the filtration

$$\mathcal{F}_s^{\max \Lambda} = \sigma(K_n, s < n \leq \max \Lambda), \ s < \min \Lambda, \tag{20}$$

which makes them accessible through sequential simulation. The following proposition essentially coincides with a result appearing in [9] in a particular case.

Theorem 5 *Consider a transition kernel p of the form* (3) *with $d(\mathcal{A}) = 1$, a coalescent distribution θ, and $P_{(\cdot)}$ irreducible and aperiodic. Let $\hat{\lambda}$ be the unique invariant distribution of the stochastic matrix $\hat{P}$. Then, for any finite $\Lambda \subset \mathbb{Z}$ the random vector $\mathbf{X}_\Lambda = (X_n, n \in \Lambda)$ given by Simulation Algorithm 1 is distributed as μ_Λ, the marginal on Λ of the unique element μ of $\mathcal{G}(p)$.*

Simulation Algorithm 1

1. *Construct the random walks $\mathcal{T}^{(n)}$, $n \in \Lambda$ up to their coalescence time S_Λ (see Fig. 3);*
2. *Sample $\tilde{X}_{S_\Lambda} \sim \hat{\lambda}$;*
3. *Keeping the K_m used in the first step and sampling U_m i.i.d. uniform in $(0, 1)$ for $S_\Lambda < m \leq \max \Lambda$, compute*

$$X_n = F_{S_\Lambda, n}\left(K_m, U_m, S_\Lambda < m \leq n; \tilde{X}_{S_\Lambda}\right), n \in \Lambda.$$
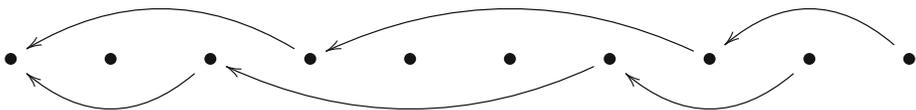

Fig. 3 An example of coalescing random walks, with $\mathcal{A} = \{2k, k \in \mathbb{N}_+\} \cup \{3\}$, for the simulation of the two rightmost adjacent sites. The leftmost site is drawn from the distribution $\hat{\lambda}$

Proof By assumption we are working in the uniqueness regime, so we can approximate μ_Λ in total variation norm with the law of $\mathbf{X}_\Lambda^{r,\mathbf{w}}$, with arbitrarily chosen boundary condition $\mathbf{w}$, as $r \to -\infty$. Choosing the components of $\mathbf{w}$ to be i.i.d. from the law $\hat{\lambda}$ we can make profit of Lemma 1. By the strong Markov property applied to the random walks $\mathcal{T}^{(n)}, n \in \Lambda$, conditionally to $\{S_\Lambda = s\}$ with $s \geq r$, $X_s^{r,\mathbf{w}}$ is independent of $\mathcal{F}_s^{\max \Lambda}$ and has the distribution of $\hat{\lambda}$. By consequence we can represent the random variable $\tilde{X}$ produced in Step 2 of the algorithm as $X_s^{r,\mathbf{w}}$, which implies that on the event $\{S_\Lambda = s\}$ with $s \geq r$,

$$X_n^{r,\mathbf{w}} = F_{s,n}(K_m, U_m, s < m \leq n; X_s^{r,\mathbf{w}}) = X_n, n \in \Lambda.$$

By the coupling inequality (see e.g. [22]) we have that the total variation distance between the law of $\mathbf{X}_\Lambda^{r,\mathbf{w}}$ and that of the output $\mathbf{X}_\Lambda$ of the algorithm is bounded by the probability that $S_\Lambda < r$. By sending r to $-\infty$ the proof is completed.   □

Notice that if the unique invariant distribution $\hat{\lambda}$ for the stochastic matrix $\hat{P}$ is difficult to compute, one can use a perfect simulation algorithm for finite Markov chains to obtain a sample from it.

In the non-coalescent case the simulation algorithm of the previous theorem is unfeasible since the coalescence point S_Λ of Λ is not finite with probability 1. Even if the distribution θ is coalescent we may need to stop the simulation when it goes beyond some large negative threshold because of memory and time limitations. A fortiori this needs to be done if we are not able to prove coalescence. In all these cases it is still possible to produce a sampling algorithm, provided a certain small error is accepted. In order to evaluate this error we need to introduce the following random time

$$\hat{S}_\Lambda = \inf\{S_{n,m} : n, m \in \Lambda, S_{n,m} > -\infty\}. \tag{21}$$

In case the set appearing at the r.h.s. of (21) is empty we define $\hat{S}_\Lambda = \min \Lambda$. As a consequence $\hat{S}_\Lambda$ is finite, but when θ is not coalescent, it is not adapted to the filtration (20).

Here is a ε-approximate simulation algorithm, where ε is an error which goes to zero as the *threshold site* $u \in \mathbb{Z}$ appearing in the algorithm decreases to $-\infty$. Indeed, $\hat{S}_\Lambda$ being finite, in principle it is possible to select u sufficiently close to $-\infty$ to make the r.h.s. of the forthcoming (23) smaller than any fixed $\varepsilon > 0$. In practice, the determination of a tail estimate on the distribution function of $\hat{S}_\Lambda$ can be extremely complicate.

Theorem 6 *Consider a transition kernel p of the form* (3) *with $d(\mathcal{A}) = 1$ and let $P_{(\cdot)}$ irreducible and aperiodic. Let $\hat{\lambda}$ be the unique invariant distribution of the stochastic matrix $\hat{P}$ and μ be the unique element of $\mathcal{G}(p)$. Let $\Lambda \subset \mathbb{Z}$ be finite, and let $u < \min \Lambda$. Let $\tilde{\mu}_\Lambda^u$ be the law of the random vector $\tilde{\mathbf{X}}_\Lambda^u = (\tilde{X}_n^u, n \in \Lambda)$defined by the following*
Simulation Algorithm 2

1. *Construct the random walk $\mathcal{T}^{(n)}$ until $V_{(n)}^u$ is reached, for $n \in \Lambda$;*
2. *For any $m \in \mathcal{V}^{u,\Lambda} = \{V_{(n)}^u : n \in \Lambda\}$ sample $\tilde{X}_m^u \sim \hat{\lambda}$, independently;*
3. *Keeping the K_m used in the step 1 and sampling U_m i.i.d. uniform in $(0, 1)$ for $u < m \leq \max \Lambda$, compute*

$$\tilde{X}_n^u = F_{u,n}\left(K_m, U_m, u < m \leq n; \tilde{X}_{V_u^{(n)}}^u\right), \quad n \in \Lambda. \tag{22}$$

Then, in total variation norm

$$||\tilde{\mu}_\Lambda^u - \mu_\Lambda|| \leq P(\hat{S}_\Lambda < u). \tag{23}$$

Proof Choosing the same boundary condition **w** as in the previous theorem we can decompose the formula (7) in the following two steps: first the boundary conditions are propagated up to $\mathcal{V}^{u,\Lambda}$

$$X_m^{r,\mathbf{w}} = F_{r,m}(K_l, U_l, r < l \le m; w_{V_{(m)}^r}), \quad m \in \mathcal{V}^{u,\Lambda}, \tag{24}$$

for any $u < \min \Lambda$. Next the random variables at Λ are constructed

$$X_n^{r,\mathbf{w}} = F_{u,n}\left(K_l, U_l, u < l \le n; X_{V_{(n)}^u}^{r,\mathbf{w}}\right), \quad n \in \Lambda. \tag{25}$$

Now we construct on the same probability space the random vector $\tilde{\mathbf{X}}_{\mathcal{V}^{u,\Lambda}}^u$ produced by step 2 of the algorithm. It will be shown that, when the event $\{\hat{S}_\Lambda \ge u\}$ occurs, $\tilde{\mathbf{X}}_{\mathcal{V}^{u,\Lambda}}^u$ will coincide with $\mathbf{X}_{\mathcal{V}^{u,\Lambda}}^{r,\mathbf{w}}$, as given by (24). Therefore, on this event also $\mathbf{X}_\Lambda^{r,\mathbf{w}}$, given by (25), will coincide with the output of the algorithm $\tilde{\mathbf{X}}_\Lambda^u$ given by (22). By applying again the coupling inequality and sending $r \to -\infty$ the proof will be concluded.

For the last step we define i.i.d. sequences $(K_l^{(m)}, U_l^{(m)}), r < l \le u,$, independently for any $m \in \mathcal{V}^{u,\Lambda}$. From each site m, by means of the $K_l^{(m)}$'s, independent random walks are started. As done in the case of two walks, we assume that the rightmost of them is updated first. The same (K_l, U_l)'s defined to construct $\mathbf{X}^{r,\mathbf{w}}$ are used, except at coalescence sites, where additional independent copies are sampled. Analogously to (5), let us define $\tilde{V}_r^{(m)}$ to be the site where the random walk starting from m lands under the threshold r. At this site we use the original boundary condition appearing in $\mathbf{X}^{r,\mathbf{w}}$. If more random walks land at the same site, additional independent copies are sampled as before.

These boundary conditions are propagated forward in time on each walk, using the $U_l^{(m)}$'s, until the starting points $m \in \mathcal{V}^{u,\Lambda}$ are reached. Following the notation of (7), we have

$$X_m^{r,\mathbf{w}^{(m)}} = F_{r,m}\left(K_l^{(m)}, U_l^{(m)}, r < l \le m; w_{V_r^{(m)}}^{(m)}\right), \quad m \in \mathcal{V}^{u,\Lambda} \tag{26}$$

By Lemma 1, $X_m^{r,\mathbf{w}^{(m)}} \sim \hat{\lambda}$, and they are independent by construction, so they are identical in law to the $\tilde{X}_m^u$'s generated in step 2 of the algorithm as promised.

Finally, observe that when $\hat{S}_\Lambda \ge u$, no site is visited twice by any of these random walks and therefore, for any $m \in \mathcal{V}^{u,\Lambda}$, the random variables $X_m^{r,\mathbf{w}}$ defined by (24) coincide with $X_m^{r,\mathbf{w}^{(m)}}$ in (26). Using the coupling inequality and sending r to $-\infty$ concludes the proof. □

As remarked by one of the referees, the previous two algorithms work also in case of non-uniqueness to construct any stationary element of $\mathcal{G}(p)$. Which element is picked up depends on the invariant distribution of $\hat{P}$ used in Step 2. In the essentially irreducible case the matrix $\hat{P}$ is itself essentially irreducible, so its unique invariant distribution is the only possible choice.

Corollary 1 *Consider a transition kernel p of the form (3) with $d(\mathcal{A}) = 1$ and let $P_{(\cdot)}$ irreducible but not aperiodic. Let $\hat{\lambda}$ be the unique invariant distribution of the stochastic matrix $\hat{P}$. Then $\mathcal{G}(p)$ has a unique stationary element μ_s. Let $\Lambda \subset \mathbb{Z}$ be finite. Simulation Algorithm 1 (for coalescent distributions θ) constructs a random vector $\tilde{\mathbf{X}}_\Lambda$ distributed as $\mu_{s,\Lambda}$. Simulation Algorithm 2 constructs a random vector $\hat{\mathbf{X}}_\Lambda^u$ converging to $\mu_{s,\Lambda}$ as $u \to -\infty$.*

Proof Consider any stationary element $\tilde{\mu}$ of $\mathcal{G}(p)$. By Lemma 1 it has necessarily $\hat{\lambda}$-distributed single-site marginals. Take the boundary condition **w** appearing in the Proof

of Theorems 5 and 6 to be $\tilde{\mu}$-distributed. Then $\mathbf{X}^{r,\mathbf{w}}$ has the distribution $\tilde{\mu}_\Lambda$. Since this is either the law or it is arbitrarily close in total variation to the random vectors constructed by these algorithms, which do not depend on the choice of $\tilde{\mu}$, one establishes both the statements of this corollary.   □

6 A Result with G Countable

In this section G is allowed to be countable. In this case the complete characterization presented in Theorem 4 fails, despite the fact that the conditions of essential irreducibility and aperiodicity continue to make sense. Indeed, by the lack of compactness of the sets of probability measures over $G^{\mathbb{Z}}$, existence is not guaranteed. On one side this prevents the construction of more than one compatible law when essential irreducibility and/or aperiodicity do not hold, and on the other requires to strengthen these assumptions to prove existence and uniqueness. In this section we are going to provide an assumption of Doeblin type that, in case G is countable, allows to prove existence and uniqueness of compatible laws. For definiteness assume that either $G = \mathbb{N}_+$ or $G = \{1, \ldots, |G|\}$.

Hypothesis D For a kernel p of the form (3), there exists a certain state, say 1 without loss of generality, and an integer $\bar{n}_0 \in \mathbb{N}_+$, with the following property

$$\forall i \in G, \ \exists \mathbf{b}_i \in \mathcal{A}^* \text{ with } s(\mathbf{b}_i) = \bar{n}_0 \text{ such that } \inf_{i \in G} P_{\mathbf{b}_i}(i, 1) =: \varepsilon > 0. \tag{27}$$

Whereas in the countable case this assumption is strictly stronger than essential irreducibility and aperiodicity, if G is finite it is actually equivalent for the following reason. First, the fact that for any $i \in G$ one has $P_{\mathbf{b}_i}(i, 1) > 0$ for some $\mathbf{b}_i \in \mathcal{A}^*$ it implies that two different irreducible classes cannot exist. Second, the fact that these words can be chosen with the same depth denies the existence of a non-trivial partition in periodic classes as in Proposition 5.

Before stating the result, define $\bar{\mathcal{B}} = \{\mathbf{b}_i, i \in G\}$, $\bar{\rho} = \sum_{\mathbf{b} \in \bar{\mathcal{B}}} \theta_{\mathbf{b}} > 0$ and

$$\bar{Q} = \frac{1}{\bar{\rho}} \sum_{\mathbf{b} \in \bar{\mathcal{B}}} \theta_{\mathbf{b}} P_{\mathbf{b}}.$$

These quantities will play in the forthcoming result the same role as played by $\mathcal{B}$, ρ and Q in the proof of Lemma 2. Under Hypothesis D, this matrix has all the entries of the first column not smaller than ε, so the Skorohod construction gives a coupling function $f_* : G \times [0, 1] \to G$ that, in addition to the property that $f_*(g, U)$ has the law $Q(g, \cdot)$ when U is uniform in $(0, 1)$, for any $g \in G$, satisfies

$$f_*(g, u) = 1, g \in G, 0 < u < \varepsilon. \tag{28}$$

We warn the reader that this coupling function enters explicitly in Simulation Algorithm 3 presented below, differently from what happened in the algorithms of the previous section. If the cardinality of G is large or infinite the computation of $\bar{Q}$, and consequently of the coupling function f^*, can give rise to accuracy and computational time problems. This is the reason for which, even if the following theorem applies to G finite as well, and as such it provides a perfect simulation algorithm under the uniqueness regime which is free of error, in practice it may be preferable to accept a small error introducing a truncation threshold in the simpler Simulation Algorithm 2, which, in addition to the (K_l, U_l)'s, requires only the computation of the invariant distribution $\hat{\lambda}$. Moreover Simulation Algorithm 3 does not apply in a situation of non uniqueness, as in the non aperiodic case.

Theorem 7 *Under Hypothesis D, there exists a unique element in $\mathcal{G}(p)$, whose restriction to any finite $\Lambda \subset \mathbb{Z}$ is the law of the random vector $\mathbf{X}_\Lambda$ given by Simulation Algorithm 3.*

Simulation Algorithm 3

1. *Set $\mathcal{V} = \Lambda$;*
2. *If $\mathcal{V} = \emptyset$ stop.*
3. *Otherwise set $m = \max \mathcal{V}$;*
4. *Construct the random walk $\mathcal{T}^{(m)}$ and check if a subpath corresponding to a word in $\bar{\mathcal{B}}$ appears before it lands on or below a site s in $\mathcal{V}$ (this is possible only if the distance between m and $\mathcal{V} \setminus \{m\}$ exceeds $\bar{n}_0$);*
5. *If such a subpath appears connecting, say, the sites u and l (with $u - l = n_0$), extract U^* uniform in $(0, 1)$, independent of all the variables generated previously: if $U^* < \varepsilon$, set $X_u = 1$, compute by forward simulation X_m, using (6) and the coupling functions $f^*(U^*; \cdot)$ on the previously located segments $(l, u]$, together with $X_n, n \in \Lambda$ for all n such that $\mathcal{T}^{(n)}$ hits m, delete m from $\mathcal{V}$ and go to 2.;*
6. *Otherwise replace m by s in $\mathcal{V}$ (if s is already in $\mathcal{V}$ this means that the two random walks have coalesced) and go to 2.*

Proof The main to prove is that the above algorithm stops in a random but a.s. finite time. We already know that any two walks $\mathcal{T}^{(m)}$ and $\mathcal{T}^{(n)}$, with $m \neq n$, either coalesce, or their distance go to infinity a.s. As a consequence it happens with probability zero that the rightmost walk, constructed from site m at step 3, only finitely many times is at distance larger than $\bar{n}_0$ from the current position of the other "surviving" walks, stored in the set $\mathcal{V}$. Each time it is at a distance larger than $\bar{n}_0$ there is a probability $\bar{\rho}\varepsilon$ to stop the walk with the determination of the value X_u at some site u, independently of the outcome of all previous simulations. Hence a finite number of walks will all be stopped in a finite time a.s. We call W the leftmost site where a random walk is stopped.

Next consider any process of the form $\mathbf{X}^{r,\mathbf{w}}$ constructed through the recursion (7). If this rule is modified by using on each interval $(l, u]$ where a path in $\bar{\mathcal{B}}$ has occurred the coupling function f^*, such process can be represented on the same probability space where the algorithm is constructed, without changing its law. Again by the coupling inequality the variation distance between $\mathbf{X}_\Lambda$ and $\mathbf{X}_\Lambda^{r,\mathbf{w}}$ will be bounded by the probability that $\{W \leq r\}$. The proof is completed by sending r to $-\infty$. $\square$

References

1. Berbee, H.: Chains with infinite connections: uniqueness and Markov representation. Probab. Theory Relat. Fields **76**(2), 243–253 (1987)
2. Boudiba, M.A.: La chaîne de Feller $X_{n+1} = |X_n - Y_{n+1}|$ et les chaînes associées. Ann. Sci. Univ. Clermont-Ferrand II Probab. Appl. **5**, 91–132 (1986)
3. Bramson, M., Kalikow, S.: Nonuniqueness in g-functions. Israel J. Math. **84**(1–2), 153–160 (1993)
4. Comets, F., Fernández, R., Ferrari, P.A.: Processes with long memory: regenerative construction and perfect simulation. Ann. Appl. Probab. **12**(3), 921–943 (2002)
5. De Santis, E., Lissandrelli, A.: Developments in perfect simulation of Gibbs measures through a new result for the extinction of Galton–Watson-like processes. J. Stat. Phys. **147**(2), 231–251 (2012)
6. De Santis, E., Maffei, A.: Perfect simulation for the infinite random cluster model, Ising and Potts models at low or high temperature. Probab. Theory Relat. Fields (2015). doi:10.1007/s00440-014-0608-2
7. De Santis, E., Piccioni, M.: Exact simulation for discrete time spin systems and unilateral fields. Methodol. Comput. Appl. Probab. **10**(1), 105–120 (2008)
8. De Santis, E., Piccioni, M.: Backward coalescence times for perfect simulation of chains with infinite memory. J. Appl. Probab. **49**(2), 319–337 (2012)

9. De Santis, E., Piccioni, M.: Perfect simulation of autoregressive models with infinite memory. J. Stat. Phys. **150**(6), 1017–1029 (2013)
10. Feller, W.: An Introduction to Probability Theory and Its Applications, vol. II, 2nd edn. Wiley, New York (1971)
11. Fernández, R., Maillard, G.: Chains and specifications. Markov Process. Relat. Fields **10**(3), 435–456 (2004)
12. Fernández, R., Maillard, G.: Chains with complete connections: general theory, uniqueness, loss of memory and mixing properties. J. Stat. Phys. **118**(3–4), 555–588 (2005)
13. Gallo, S.: Chains with unbounded variable length memory: perfect simulation and a visible regeneration scheme. Adv. Appl. Probab. **43**(3), 735–759 (2011)
14. Gallo, S., Garcia, N.L.: Perfect simulation for locally continuous chains of infinite order. Stoch. Process. Appl. **123**(11), 3877–3902 (2013)
15. Galves, A., Löcherbach, E., Orlandi, E.: Perfect simulation of infinite range Gibbs measures and coupling with their finite range approximations. J. Stat. Phys. **138**(1–3), 476–495 (2010)
16. Georgii, H.-O.: Gibbs Measures and Phase Transitions. de Gruyter Studies in Mathematics, vol. 9. Walter de Gruyter & Co., Berlin (1988)
17. Grimmett, G.: The random-cluster model. In: *Probability on Discrete Structures*, vol. 110. *Encyclopaedia Mathematical Sciences*, pp. 73–12. Springer, Berlin (2004)
18. Iosifescu, M., Grigorescu, Ş.: Dependence with Complete Connections and Its Applications. Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge (1990)
19. Johansson, A., Öberg, A.: Square summability of variations of g-functions and uniqueness of g-measures. Math. Res. Lett. **10**(5–6), 587–601 (2003)
20. Keane, M.: Strongly mixing g-measures. Invent. Math. **16**, 309–324 (1972)
21. Lalley, S.P.: Regenerative representation for one-dimensional Gibbs states. Ann. Probab. **14**(4), 1262–1271 (1986)
22. Lindvall, T.: Lectures on the Coupling Method. Wiley Series in Probability and Mathematical Statistics. Wiley, New York (1992)
23. Peigné, M., Woess, W.: Stochastic dynamical systems with weak contractivity properties I. Strong and local contractivity. Colloq. Math. **125**(1), 31–54 (2011)
24. Propp, J.G., Wilson, D.B.: Exact sampling with coupled Markov chains and applications to statistical mechanics. Random Struct. Algorithms **9**(1–2), 223–252 (1996)
25. Stenflo, Ö.: Uniqueness in g-measures. Nonlinearity **16**(2), 403–410 (2003)
26. Walters, P.: Ruelle's operator theorem and g-measures. Trans. Am. Math. Soc. **214**, 375–387 (1975)