

PROBABILITY AND ALGORITHMICS:
A FOCUS ON SOME RECENT DEVELOPMENTS

PEGGY CÉNAC¹, IRÈNE MARCOVICI², CHRISTELLE ROVETTA³, MATHIEU SABLİK⁴ AND
RÉMI VARLOOT⁵

Abstract. This article presents different recent theoretical results illustrating the interactions between probability and algorithmics. These contributions deal with various topics: cellular automata and calculability, variable length Markov chains and persistent random walks, perfect sampling via coupling from the past. All of them involve discrete dynamics on complex random structures.

Résumé. Cet article présente différents résultats récents de nature théorique illustrant les interactions entre probabilités et algorithmique. Ces contributions traitent de sujets variés : automates cellulaires et calculabilité, chaînes de Markov à mémoire variable et marches aléatoires persistantes, échantillonnage parfait par la méthode de couplage par le passé. Leur point commun est de faire intervenir des dynamiques discrètes sur des structures aléatoires complexes.

Probability and algorithmics are two domains that mutually enrich each other. On the one hand, the analysis of algorithms naturally involves the study of random discrete objects. On the other hand, in order to study and generate complex combinatorial structures, one often needs to develop sophisticated algorithms. In that context, a topical question is also the classification of probability distributions that can be described with the help of different models of computation.

This article presents some recent theoretical results illustrating the interactions between probability and algorithmics. It is the reflect of the session of the *Journées MAS 2016* entitled “Probability and Algorithmics”. The three sections are based on the talks given in Grenoble on this occasion by Mathieu Sablik, Peggy Cénac, Christelle Rovetta, and Rémi Varloot. Irène Marcovici, the organizer of that session of the *Journées MAS 2016*, has written the introduction and coordinated the present article.

Section 1 is an introduction to the advances of Mathieu Sablik on the possible asymptotic measures that can be reached by a cellular automaton.^a Cellular automata are used to model all kinds of phenomena with local interactions that appear in biology, physics, or theoretical computer science. S. Wolfram [52] was the first to propose a classification of cellular automata, based on the observation of space-time diagrams. His empirical classification was then followed by many other attempts, the aim being to understand the complex phenomena that can appear. Here, we focus on the typical asymptotic configurations observed from a random initial configuration. They are described by the set of limit measures. From a modelisation point of view, this

¹ Institut de Mathématiques de Bourgogne, UMR 5584, Université de Bourgogne, CNRS, F-21078 Dijon Cedex, France

² Institut Élie Cartan de Lorraine, UMR 7502, Université de Lorraine, CNRS, F-54506 Vandoeuvre-lès-Nancy Cedex, France

³ Inria Paris and LIP6, UMR 7606, Université Pierre et Marie Curie, CNRS, F-75252 Paris Cedex 05, France

⁴ Institut de Mathématiques de Toulouse, UMR 5219, Université de Toulouse, CNRS, Université Paul Sabatier, F-31062 Toulouse Cedex 9, France

⁵ Microsoft Research - Inria Joint Centre, F-91120 Palaiseau, France

^aThis work was partially supported by the ANR project QuasiCool ANR-12-JS02-011-01

set corresponds to the set of measures that are “physically” relevant, starting from a simple measure such as the uniform Bernoulli measure. From a computer science point of view, cellular automata can be seen as a model of computation with high parallelism and it is natural to study the possible complexities of the asymptotic behavior, and its dependence on the initial distribution.

Section 2, presented by Peggy Cénac, provides a short zoology of variable length Markov chains, with necessary and sufficient conditions for existence and uniqueness of invariant measure. Such conditions can be related to recurrence/transience properties of random walks built from Markov chains with unbounded memory. We will follow some examples illustrating the unstable nature of the random walk behavior under slightly disturbed settings. The works described in this section are extracted from two reference articles [10, 11].

Section 3 is based on works of Christelle Rovetta and Rémi Varloot on perfect sampling. In [42], Propp and Wilson have introduced the coupling from the past method to generate random variables according to the stationary distribution of a given Markov chain. The advantage over the classical Monte Carlo approach is that it provides a perfect sampling algorithm: it gives an actual algorithm that terminates in finite expected time, as opposed to a converging algorithm, and the output distribution is unbiased, rather than arbitrarily close to the target distribution. This initially comes at an important cost in complexity, which becomes proportional to the size of the state space. Different strategies have been introduced so as to remove this dependency and achieve acceptable complexity. The basic building blocks for such strategies are presented, alongside two recent applications stemming from collaborations with Anne Bouillard and Ana Bušić. The first is a novel means of sampling from the stationary distribution of closed queueing networks in an efficient way [4]. The second is an approach to speed up the simulation of “very lazy” random walks, *i.e.* random walks for which the probability of changing states during any given transition is small [47], presented here in the context of sampling random independent sets from a graph.

1. LIMIT MEASURES OF CELLULAR AUTOMATA: A COMPUTATIONAL APPROACH

In this section, we present (in a simplified version) some new advances on the study of typical asymptotic measures obtained when iterating a cellular automaton from some initial measure. There are two possible approaches for the study of the sets of limit measures of cellular automata:

- taking a class of cellular automata having empirically the same typical asymptotic behavior, and trying to determine their sets of limit measures;
- characterizing sets of measures that can be reached as sets of limit measures for some cellular automata.

In Subsection 1.1, we recall the main definitions and present some results about the asymptotic measures for some classes of cellular automata. Generally, one observes a behavior on space-time diagrams, and then tries to determine the set of limit measures for this class. This type of study can be difficult and often requires to develop new tools for each class that is considered.

In Subsection 1.2, we are interested in characterizing the limit measures that can be reached asymptotically when iterating a cellular automaton from some initial measure. A first step is to determine some obstructions, that is, some necessary conditions that these measures must satisfy. As cellular automata are model of computation, in addition to the classical topological obstructions, some necessary computational obstructions appear. The main problem is to prove a reverse statement: given a measure satisfying the computational obstructions, we want to construct a cellular automaton which, starting from any simple initial measure, reaches this measure asymptotically. Similar computational obstructions appear in topological dynamics, when characterizing possible properties of subshifts of finite type or cellular automata: possible entropies [29], possible growth-type invariants [38], possible sub-actions [1, 28]... However, the construction is quite different here, since starting from a random configuration, the construction requires an ability to self-organize the space. The results presented here are part of a collaboration with Benjamin Hellouin de Menibus [25].

In the last subsection we give some perspectives and open questions.

1.1. Definitions and study of some classes of cellular automata

1.1.1. Problematic

A *cellular automaton* is a complex system defined by a local rule which acts synchronously and uniformly on the configuration space $\mathcal{A}^{\mathbb{Z}}$, where \mathcal{A} is a finite alphabet. Formally, it is a function $F : \mathcal{A}^{\mathbb{Z}} \rightarrow \mathcal{A}^{\mathbb{Z}}$ defined by a local rule $\bar{F} : \mathcal{A}^{[-r,r]} \rightarrow \mathcal{A}$, where $r \in \mathbb{N}$ is called the radius, such that:

$$F(x)_i = \bar{F}(x_{[i-r,i+r]}) \text{ for all } i \in \mathbb{Z}.$$

Equivalently, cellular automata are exactly continuous functions which commute with the shift map $\sigma : \mathcal{A}^{\mathbb{Z}} \rightarrow \mathcal{A}^{\mathbb{Z}}$, defined by $\sigma(x)_i = x_{i+1}$ for all $x \in \mathcal{A}^{\mathbb{Z}}$ and $i \in \mathbb{Z}$.

Example 1.1. Let $\mathcal{A} = \{0, 1\}$ and consider the cellular automaton defined by

$$F(x)_i = x_{i-1} + x_i + x_{i+1} \pmod{2} \text{ for all } i \in \mathbb{Z}.$$

Given an initial configuration, a representation of the behavior of this cellular automaton is given by the superposition of the different configurations obtained by iteration of the dynamics. Figure 1.1 shows an example of such a space-time diagram, starting from a random initial configuration (represented at the bottom).

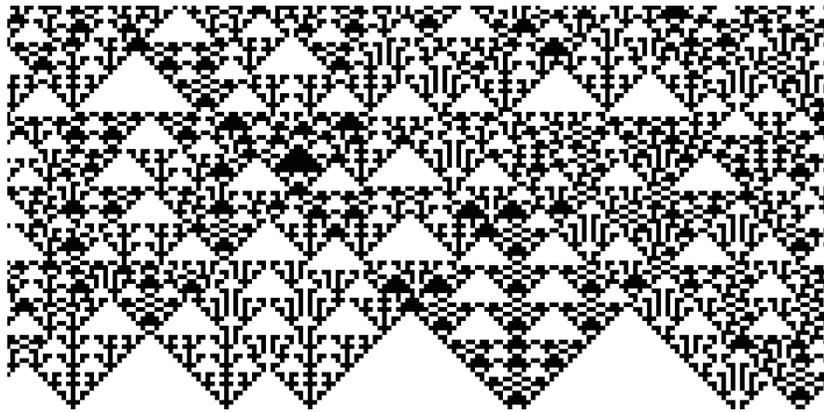


FIGURE 1.1. Space time diagram of the cellular automaton defined by $F(x)_i = x_{i-1} + x_i + x_{i+1} \pmod{2}$ for all $i \in \mathbb{Z}$.

Cellular automata are simple models but have a wide variety of different dynamical behaviors. We are interested in their typical asymptotic behavior starting from a random configuration, as observed on simulations. We consider that the space of configurations is homogeneous. This is rendered by an initial configuration chosen randomly according to a σ -invariant measure μ , meaning that $\mu(B) = \mu(\sigma^{-1}(B))$ for any Borel set B . Denote by $\mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$ the set of σ -invariant measures. This space is compact and metrizable for the weak* topology. We consider here the following metric:

$$d_{\mathcal{M}}(\mu, \nu) = \sum_{n \in \mathbb{N}} \frac{1}{2^n} \max_{u \in \mathcal{A}^n} |\mu([u]) - \nu([u])|,$$

where $[u] = \{x \in \mathcal{A}^{\mathbb{Z}} : x_{[0,n-1]} = u\}$ is the cylinder centered on the word $u \in \mathcal{A}^n$.

For simulations, the initial configuration is generally chosen randomly according to a *Bernoulli measure*: the probability that a cell has the value $a \in \mathcal{A}$ is $p_a \in [0, 1]$, independently for different cells. In other words,

the Bernoulli measure associated to $(p_a)_{a \in \mathcal{A}}$ with $\sum_{a \in \mathcal{A}} p_a = 1$ is defined by $\lambda_{(p_a)_{a \in \mathcal{A}}}([u]) = p_{u_0} p_{u_1} \cdots p_{u_{n-1}}$ for all $u \in \mathcal{A}^n$. For the *uniform Bernoulli measure*, all the parameters are chosen to have the same value $p_a = \frac{1}{\text{Card} \mathcal{A}}$, we denote it $\lambda_{\mathcal{A}^{\mathbb{Z}}}$. It is also possible to define measures with finite ranges of spatial dependences, thanks to Markov processes.

Another important class of measures consists in the *σ -invariant measures supported by periodic orbits*: for $w \in \mathcal{A}^n$, we denote by ${}^\infty w {}^\infty$ the σ -periodic configuration associated to w , and we define:

$$\widehat{\delta}_w = \frac{1}{|w|} \sum_{i \in [0, |w| - 1]} \delta_{\sigma^i({}^\infty w {}^\infty)}, \quad \text{where } \delta_x \text{ is the Dirac measure in } x \in \mathcal{A}^{\mathbb{Z}}.$$

Let us recall that \mathcal{A}^* denotes the set of words on the alphabet \mathcal{A} . Even if the measure $\widehat{\delta}_w$ is algorithmically simple, the set $\{\widehat{\delta}_w : w \in \mathcal{A}^*\}$ is dense in $\mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$.

Most of the results we obtain are not restricted to Bernoulli initial measures. A measure is *σ -ergodic* if σ -invariant sets have measure 0 or 1. We denote by $\mathcal{M}_{\sigma\text{-erg}}(\mathcal{A}^{\mathbb{Z}})$ the set of σ -ergodic measures. By Birkhoff's theorem, the frequency of a word $u \in \mathcal{A}^n$ in a random configuration chosen according to a σ -ergodic measure is equal to the measure of the cylinder $[u]$ (see [46] for details). Sometimes, we also need to evaluate the independence of cells in a random configuration. A measure is *full support* if all cylinders are charged, and we denote by $\mathcal{M}_\sigma^{\text{full}}(\mathcal{A}^{\mathbb{Z}})$ the set of σ -invariant measures of full support.

Since a cellular automaton F commutes with the shift, it acts naturally on $\mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$ by:

$$F_* : \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}}) \longrightarrow \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}}) \\ \mu \longmapsto F_* \mu \quad \text{such that } F_* \mu(B) = \mu(F^{-1}(B)) \text{ for any Borel set } B.$$

The measure $F_*^t \mu$ represents the measure obtained at time t when iterating the cellular automaton F from the initial measure μ . For an initial measure $\mu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$, we are interested in the asymptotic behavior of the sequence $(F_*^t \mu)_{t \in \mathbb{N}}$, and of its Cesàro mean, defined by $\varphi_t^F \mu = \frac{1}{t+1} \sum_{i=0}^t F_*^i \mu$. In particular, we focus on the following sets of limit points, which are non-empty, by compactness of $\mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$:

- the *μ -limit measures set* $\mathcal{V}(F, \mu)$, defined as the set of **limit points** of the sequence $(F_*^t \mu)_{t \in \mathbb{N}}$;
- the *Cesàro mean μ -limit measures set* $\mathcal{V}'(F, \mu)$, defined as the set of limit points of $(\varphi_t^F \mu)_{t \in \mathbb{N}}$.

The union of the support of the μ -limit measures set corresponds to the μ -limit sets introduced in [34] in a different way. Very complex μ -limit sets can be constructed [6, 7], and the construction of Subsection 1.2 is partly inspired from these works.

1.1.2. Equicontinuous dynamics

The notions of equicontinuity and equicontinuous points in some directions, introduced in [45], allow to detect some rigidity in a particular direction. In that case, it is easy to obtain informations on the μ -limit measures sets, which are σ -invariant (see Figure 1.2 for some examples). Given a direction $\alpha \in \mathbb{R}$, we say that

- F is equicontinuous if for all $e \in \mathbb{N}$ there exists $d \in \mathbb{N}$ such that $x_{[-d, d]} = y_{[-d, d]} \implies F^n(\sigma^{\lfloor n\alpha \rfloor}(x))_{[-e, e]} = F^n(\sigma^{\lfloor n\alpha \rfloor}(y))_{[-e, e]}$ for all $n \in \mathbb{N}$,
- $x \in \mathcal{A}^{\mathbb{Z}}$ is an equicontinuous point of F if for all $e \in \mathbb{N}$, there exists $d \in \mathbb{N}$ such that $x_{[-d, d]} = y_{[-d, d]} \implies F^n(\sigma^{\lfloor n\alpha \rfloor}(x))_{[-e, e]} = F^n(\sigma^{\lfloor n\alpha \rfloor}(y))_{[-e, e]}$ for all $n \in \mathbb{N}$. In this case, the word $x_{[-d, d]}$ is said a blocking word.

Nilpotence. A cellular automaton F is nilpotent if there exists $k \in \mathbb{N}$ and $a \in \mathcal{A}$ such that $F^k(x) = {}^\infty a {}^\infty$ for all $x \in \mathcal{A}^{\mathbb{Z}}$. This is equivalent to having two different directions of equicontinuity. In this case, $\mathcal{V}(F, \mu) = \{\widehat{\delta}_a\}$ for all $\mu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$.

Equicontinuity. If a cellular automaton F is equicontinuous of direction α , then it is ultimately periodic in this direction. Thus, there exist $k, p \in \mathbb{N}$ such that

$$\mathcal{V}(F, \mu) = \{F_*^{k+i} \mu : i \in [0, p - 1]\}.$$

Equicontinuous points in two directions. If a cellular automaton F has equicontinuous points in two directions, there exists $a \in \mathcal{A}$ such that $\mathcal{V}(F, \mu) = \{\widehat{\delta}_a\}$ for all $\mu \in \mathcal{M}_{\sigma\text{-erg}}^{\text{full}}(\mathcal{A}^{\mathbb{Z}})$ (see Theorem 4.8 of [45]).

Equicontinuous points in one direction. If a cellular automaton F has equicontinuous points in one direction, there exists a word u , called a blocking word, such that for any word v , all the configurations in the cylinder $[uvu]$ have the same ultimate periodic behavior in direction α . In this case, $\mathcal{V}(F_*, \mu)$ is a singleton when $\mu \in \mathcal{M}_{\sigma\text{-erg}}^{\text{full}}(\mathcal{A}^{\mathbb{Z}})$, and we have some information on the limit measure [3, 21].

Remark 1.1. *In the last two classes, the condition of full support can be weakened by considering an initial measure charging a blocking word. Moreover, the notion of equicontinuous points can be relaxed by the notion of μ -equicontinuous points, which takes into account the initial measure [45].*

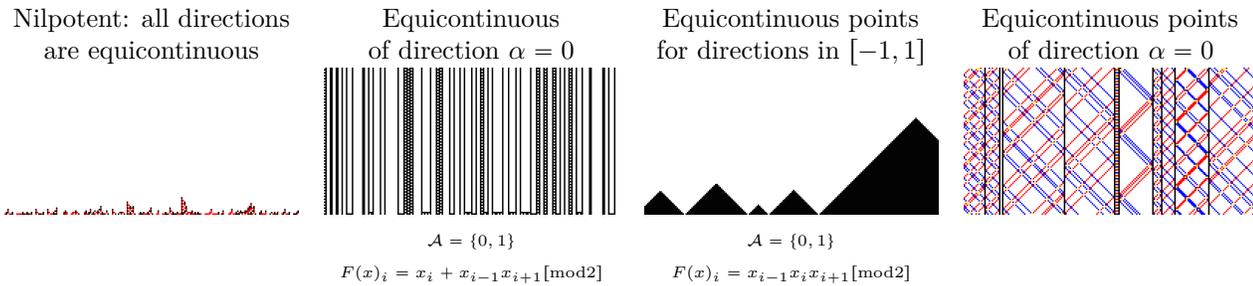


FIGURE 1.2. Examples of cellular automata that are equicontinuous or have equicontinuous points.

1.1.3. Cellular automata with particles

For some cellular automata, starting from a random configuration, a system of interacting particles emerges. Particles correspond to interfaces between homogenous zone in the space time diagram. Generally, in this case, there is convergence towards a “simple” measure (see Figure 1.3 for some examples).

An interesting example is given by the following gliders cellular automaton, defined on the alphabet $\{0, L, R\}$. Under the action of the gliders cellular automaton, the letter L is shifted to the left, the letter R is shifted to the right, and two letters disappear if they intersect. It is shown in [2] that for $\mu \in \mathcal{M}_{\sigma\text{-erg}}(\mathcal{A}^{\mathbb{Z}})$, one has $\mathcal{V}(F, \mu) = \{\nu\}$, with: $\nu = \widehat{\delta}_0$ if $\mu([L]) = \mu([R])$; $\nu \in \mathcal{M}_{\sigma}(\{0, L\}^{\mathbb{Z}})$ if $\mu([L]) > \mu([R])$; and $\nu \in \mathcal{M}_{\sigma}(\{0, R\}^{\mathbb{Z}})$ if $\mu([L]) < \mu([R])$.

Another interesting example is the n -states cyclic cellular automaton on the alphabet $\{0, \dots, n - 1\}$, defined by $F_n(x)_i = x_i + 1[\text{mod } n]$ if x_{i-1} or x_{i+1} is equal to $x_i + 1[\text{mod } n]$, and $F_n(x)_i = x_i$ otherwise. For the 3-states cyclic cellular automaton, if μ is a Bernoulli measure, one has $\mathcal{V}(F_3, \mu) = \{\mu([2])\widehat{\delta}_0 + \mu([0])\widehat{\delta}_1 + \mu([1])\widehat{\delta}_2\}$ (see [24] Section 2.3 for a proof). For the 4-states cyclic cellular automaton, one has $\mathcal{V}(F_4, \mu) = \{\alpha_0\widehat{\delta}_0 + \alpha_1\widehat{\delta}_1 + \alpha_2\widehat{\delta}_2 + \alpha_3\widehat{\delta}_3\}$ but the value of the parameters α_i are unknown. For $n \geq 5$, n -states cyclic cellular automata have μ -equicontinuous points.

We refer to [26] for more examples and references. In this article, we focus on qualitative and quantitative properties of the set of μ -limit measures for cellular automata having a system of particles with good collision properties.

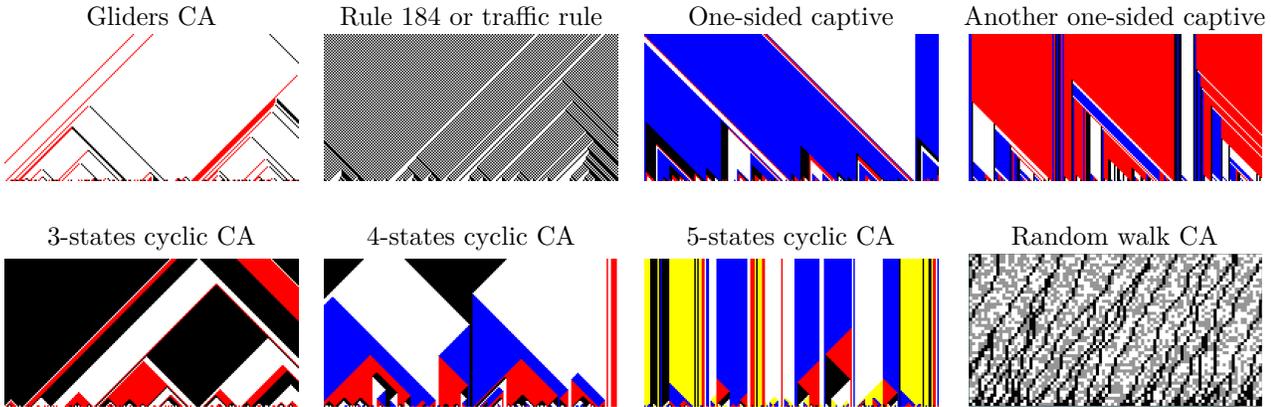


FIGURE 1.3. Examples of cellular automata with emergence of particles (the typical asymptotic behavior is described in [26]).

1.1.4. Algebraic cellular automata

When \mathcal{A} is a finite group, if the cellular automaton F is a group morphism for the product group $\mathcal{A}^{\mathbb{Z}}$, we say that F is *linear* (the cellular automaton of Example 1.1 was such a linear cellular automaton). In this case, a randomisation phenomenon appears: as soon as the initial measure is in a large class which contains Markov measures, the Cesàro mean sequence converges to the uniform Bernoulli measure [18, 36, 37, 41]. In other words, $\mathcal{V}(F, \mu) = \{\lambda_{\mathcal{A}^{\mathbb{Z}}}\}$. In some simple cases, it is possible to describe $\mathcal{V}(F, \mu)$ with the help of the convolution measure of μ (see [24], Chapter III).

1.2. Possible convergence towards a unique measure

The first natural question is to determine what are the measures that can be reached asymptotically by the iteration of a cellular automaton from an initial measure. Given a measure ν , this measure is clearly reached asymptotically if we iterate the identity cellular automaton from this measure. In fact, the question is not precise enough: we want to determine the measures that are reached asymptotically starting from a large class of measures, for example, starting from all Bernoulli measures.

Computability obstruction. As cellular automata can be seen as models of computation, computability obstructions naturally appear. A function $f : \mathcal{A}^* \rightarrow \mathcal{B}^*$ is *computable* if there exists an algorithm which, for a given word u over the alphabet \mathcal{A} given as input, returns the word $f(u)$ over the alphabet \mathcal{B} . This definition can be generalized to \mathbb{N} , \mathbb{Q} or $\mathcal{A}^* \times \mathbb{N}$, by the introduction of a natural encoding. We distinguish two notions of computable measure, depending on whether we know or not the speed of convergence of the algorithm approximating the measure.

Definition 1.1. A measure $\mu \in \mathcal{M}_{\sigma}(\mathcal{A}^{\mathbb{Z}})$ is

- *computable* if there exists a computable function $f : \mathcal{A}^* \times \mathbb{N} \rightarrow \mathbb{Q}$ such that

$$|\mu([u]) - f(u, n)| < 2^{-n};$$

- *limit-computable* if there exists a computable function $f : \mathcal{A}^* \times \mathbb{N} \rightarrow \mathbb{Q}$ such that

$$\lim_{n \rightarrow \infty} f(u, n) = \mu([u]).$$

The set of computable (resp. limit-computable) measures is very large, even if it is countable, since the number of programs is countable. Let us present some examples:

- any measure supported by a periodic orbit is computable;
- any Bernoulli measure or Markov measure with computable (resp. limit-computable) parameters is computable (resp. limit-computable);
- if an effective subshift (*i.e.* a subset of $\mathcal{A}^{\mathbb{Z}}$ for which the set patterns which does not appear in a configuration can be enumerated by a Turing machine) has a unique σ -ergodic measure μ , then μ is computable.

The image of a computable measure by a cellular automaton F of radius r is computable. Indeed, given a word $u \in \mathcal{A}^n$, one can compute the set of words in \mathcal{A}^{n+2r} which are pre-images of u , and for each of these words, one can compute an approximation of the measure of the corresponding cylinder. One thus obtains an approximation of $F_*\mu([u])$, with a bound on the error. As a consequence, we obtain the following obstruction for the limit.

Proposition 1.1. *Let $F : \mathcal{A}^{\mathbb{Z}} \rightarrow \mathcal{A}^{\mathbb{Z}}$ be a cellular automaton. If $\mu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$ is computable and $F_*^n \mu \xrightarrow[n \rightarrow \infty]{} \nu$, then ν is limit-computable.*

Theorem of realization. In fact, there is no other obstruction: given a limit-computable measure ν , it is always possible to “program” a cellular automaton forcing the convergence to the limit ν starting from any σ -ergodic measure of full support. This is given by the following realization theorem.

Theorem 1.1 (see [25]). *Let $\nu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$ be a limit-computable measure, there exists a cellular automaton $F : \mathcal{B}^{\mathbb{Z}} \rightarrow \mathcal{B}^{\mathbb{Z}}$ with $\mathcal{A} \subset \mathcal{B}$ such that for all $\mu \in \mathcal{M}_{\sigma\text{-erg}}(\mathcal{B}^{\mathbb{Z}})$ of full support, one has*

$$\lim_{n \rightarrow \infty} F_*^n \mu = \nu.$$

The property for a measure $\nu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$ to be limit-computable is equivalent to the existence of a sequence $(w_n)_{n \in \mathbb{N}}$ of words, enumerated by an algorithm, such that $\widehat{\delta_{w_n}}_{n \in \mathbb{N}} \rightarrow \nu$. The idea of the proof is thus to construct a cellular automaton enumerating the words $(w_n)_{n \in \mathbb{N}}$ and copying them successively on the space. In order to do so, we extend the initial alphabet with auxiliary states. The dynamics of this cellular automaton, described in [25], is very simple. It consists in three different phases, which coexist:

Formatting: Computation takes place on *segments*, defined as regions of \mathbb{Z} located between two symbols $\boxed{\text{W}}$, that play the role of walls, and persist in time except under special circumstances. Since we have no control over the initial content of each segment, we first want to format them, that is, to erase uninitialized content of auxiliary states which can perturb the computation. A segment is said to be initialized if the walls come from a special state $\boxed{\text{I}}$, which can only appear in the initial configuration, and send a signal to the right. This signal deletes all auxiliary states until finding another initialized wall $\boxed{\text{W}}$. The difficult task is to distinguish uninitialized walls from initialized walls. To do that, the right signal sent by $\boxed{\text{I}}$ keeps track of its age using a binary counter. Meanwhile, each initialized wall also keeps track of its age under the form of a binary counter to its left, incrementing at each step. Counters already present in the initial configuration (uninitialized) have a nonnegative value at time 0, whereas those created by a symbol $\boxed{\text{I}}$ (initialized) have value 0 at time 1, and they increment at the same rate. Thus, uninitialized walls have older time counters, and by comparing time counters and formatting counters as they cross, we can erase older counters and uninitialized walls. This version with binary counter is detailed in [25], there exists also a geometric version of this process of formatting, described in [5, 6, 15]

Computation and copy: A Turing machine is simulated in a small part of the segment. Precisely, if the segment has size k , the Turing machine needs a length $o(k)$ of space for the computation. In [25], the space is delimited by the time counter, whereas in [5, 6], the size of the segment is measured by a signal, and a small part is used to simulate the Turing machine, whose evolution can be seen as a particular cellular automaton. The machine considered computes successively each w_n on the part allowed for the computation. When a word w_n is produced, concatenated copies are written progressively on all the

segment. By the process of copying, the cellular automaton produces a sample of $\widehat{\delta_{w_n}}$ on the segment associated.

Merging: In order to have larger and larger domains of computation with a low density of auxiliary states, a segments of size k merges with a neighboring segment, when we need more than $o(k)$ cells for auxiliary states. Different processes of merging can be proposed, either unsynchronized as in [6], or synchronized as in [25]. When walls are destroyed in the merging process, only the computation of destroyed walls are stopped. The sample written on the newly merged segment can be only deleted by a copy of a new sample computed with more space.

It remains to prove the convergence towards ν . In short, for a given ϵ and a word $u \in \mathcal{A}^l$, there exists K such that $|\widehat{\delta_{w_k}}([u]) - \nu([u])| < \epsilon$ for all $k \geq K$. By the merging process, for a time n large enough, with probability larger than $1 - \epsilon$, the pattern $F^n(x)_{[0, l-1]}$ is in a sufficiently large segment which allows to have enough space to compute w_K , and where the density of auxiliary states is less than ϵ (this is possible since auxiliary states occur in a negligible part of the segment compared to its size). Thus, with probability $1 - \epsilon$, the configuration $F^n(x)_{[0, l-1]}$ contains auxiliary states, or is a part of a sample composed of successions of w_k for some $k \geq K$. We deduce that for sufficiently large n , one has $\max_{k \geq K} |F_*^n \mu([u]) - \widehat{\delta_{w_k}}([u])| < 2\epsilon$. Thus, $|F_*^n \mu([u]) - \nu([u])| \xrightarrow{n \rightarrow \infty} 0$.

1.3. Some further directions of research

1.3.1. What are the possible sets of limit measures?

Generally, the sequence $(F_*^n \mu)_{n \in \mathbb{N}}$ does not converge. So, the second step in the problematic is to determine the set of limit points, denoted by $\mathcal{V}(F, \mu)$. By compactness of $\mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$, it is a non-empty set. Some necessary conditions can be described using the setting of computable analysis on metric spaces (the formalism is introduced in [48]). In [25], the authors give a partial characterization when the set $\mathcal{V}(F, \mu)$ is connected.

Open question 1.1. *What are the sets of measures that can be reached asymptotically iterating a measure by a cellular automaton?*

1.3.2. What happens if we do not allow additional letters?

The theorem of realization of Section 1.2 uses additional states to encode the different processes. In particular, we need a special symbol, denoted by $\boxed{\mathbb{I}}$, that only appears in the initial configuration. What can we do if we do not allow additional states in the construction? Given a limit-computable measure ν , we want to construct a cellular automaton defined on $\mathcal{A}^{\mathbb{Z}}$, which realizes ν as limit measure. If ν does not charge a word u , it is possible to code auxiliary states using this word and we obtain the following corollary.

Corollary 1.1 (see [25]). *Let $\nu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$ be such that there exists a word u which is not charged by ν . Then, there exists a cellular automaton $F : \mathcal{A}^{\mathbb{Z}} \rightarrow \mathcal{A}^{\mathbb{Z}}$ such that for all σ -ergodic measure $\mu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$ of full support, one has*

$$\lim_{n \rightarrow \infty} F_*^n \mu = \nu.$$

If μ is full support, the cellular automaton constructed must be surjective, since every word has pre-images. In fact, there is no hope having a result similar to Corollary 1.1: indeed, the uniform Bernoulli measure is invariant by surjective cellular automata, so it is the only limit measure that can be reached starting from the uniform Bernoulli measure. The following question remains.

Open question 1.2. *What are the limit measures that can be reached by surjective cellular automata?*

The difficulty is that we cannot initialize the computation by a special pattern that would appear only in the initial configuration. In fact, there are some other obstructions that are not well understood yet. For example, the action of a surjective cellular automaton on random configurations preserves the quantity of information. Formally, this means that the metric entropy of the shift according to $\mu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$, denoted by $h_\mu(\sigma)$, is equal to the metric entropy according to $F_* \mu$ (see [32]). By the semi-continuity of the entropy, the limit measures can only have a greater entropy than the initial measure. The following obstruction follows.

Proposition 1.2. *Let $F : \mathcal{A}^{\mathbb{Z}} \rightarrow \mathcal{A}^{\mathbb{Z}}$ be a surjective cellular automaton and let $\mu \in \mathcal{M}_{\sigma}(\mathcal{A}^{\mathbb{Z}})$. If $F_{*}^n \mu \xrightarrow[n \rightarrow \infty]{} \nu$, one has $h_{\nu}(\sigma) \geq h_{\mu}(\sigma)$.*

We remark that this obstruction implies the obstruction given by the uniform Brenoulli measure since it is the unique measure of maximal entropy.

1.3.3. *What is the speed of convergence?*

In Theorem 1.1, we can estimate the speed of the convergence to ν . The merging process used in [25] is very slow and might be improved for some classes of measures. A natural question is to determine what are the classes of measures that can be quickly reached, and we can wonder if this property is related to the algorithmic complexity of the measures.

Open question 1.3. *What is the link between the algorithmic complexity of a measure and the best possible speed of convergence towards this measure?*

2. BESTIARY OF VARIABLE LENGTH MARKOV CHAINS AND PERSISTENT RANDOM WALKS

2.1. Variable Length Markov Chains

2.1.1. Definition

Let us recall briefly the probabilistic presentation of Variable Length Markov Chains (VLMC), following [10] (cf [44] and [14, pp. 117-134]). Introduce the set $\mathcal{L} = \mathcal{A}^{-\mathbb{N}}$ of left-infinite words on the alphabet $\mathcal{A} := \{\mathbf{d}, \mathbf{u}\}$ or $\mathcal{A} := \{0, 1\}$ and consider a complete tree on this alphabet, *i.e.* a tree such that each node has 0 or 2 children, whose leaves \mathcal{C} are (possibly infinite) words on \mathcal{A} . The set \mathcal{C} is supposed finite or countable. To each leaf $c \in \mathcal{C}$, called a context, is attached a probability distribution q_c on \mathcal{A} . Endowed with this probabilistic structure, such a tree is named a probabilized context tree. The related VLMC is defined as the Markov Chain on \mathcal{L} whose transitions are given by

$$\mathbb{P}(U_{n+1} = U_n \ell | U_n) = q_{\overline{\text{pref}}(U_n)}(\ell), \tag{2.1}$$

where $\overline{\text{pref}}(w) \in \mathcal{C}$ is defined as the shortest prefix of $w \in \mathcal{L}$, when w is read from right to left, appearing as a leaf of the context tree. In the example of Figure 2.4, on the left side, the context tree is finite of height 4 and, for instance,

$$\mathbb{P}(U_{n+1} = U_n \mathbf{d} | U_n = \dots \mathbf{d} \mathbf{u} \mathbf{d} \mathbf{u} \mathbf{u} \mathbf{u} \mathbf{d}) = q_{\mathbf{d} \mathbf{u} \mathbf{u}}(\mathbf{d})$$

because $\overline{\text{pref}}(\dots \mathbf{d} \mathbf{u} \mathbf{d} \mathbf{u} \mathbf{u} \mathbf{u} \mathbf{d}) = \mathbf{d} \mathbf{u} \mathbf{u}$ (read the word $\dots \mathbf{d} \mathbf{u} \mathbf{d} \mathbf{u} \mathbf{u} \mathbf{u} \mathbf{d}$ right-to-left and stop when finding a context). The rightmost letter of the sequence $U_n \in \mathcal{L}$ will be denoted by X_n so that

$$\forall n \geq 0, U_{n+1} = U_n X_{n+1}.$$

The final letter process $(X_n)_{n \geq 0}$ is *not* Markov as soon as the context tree has at least one infinite context. When the tree is finite, $(X_n)_{n \geq 0}$ is a Markov chain whose order is the height of the tree, *i.e.* the length of its longest branch. The vocable VLMC is somehow confusing but commonly used.

Consider the probabilized context tree given on the right side of Figure 2.4, called *Infinite Comb* probabilized context tree. For this model, [13] established a sufficient condition for the existence of a stationary probability measure. In this case, there is one infinite leaf 0^{∞} and countably many finite leaves $0^n 1, n \in \mathbb{N}$. The data of a corresponding VLMC consists thus in probability measures on $\mathcal{A} = \{0, 1\}$:

$$q_{0^{\infty}} \text{ and } q_{0^n 1}, n \in \mathbb{N}.$$

Proposition 2.1. (Stationary probability measures for an infinite comb) *Let $(U_n)_{n \geq 0}$ be a VLMC defined by a probabilized infinite comb. Assume that $q_{0^{\infty}}(0) \neq 1$ (irreducible case). The Markov process $(U_n)_{n \geq 0}$*

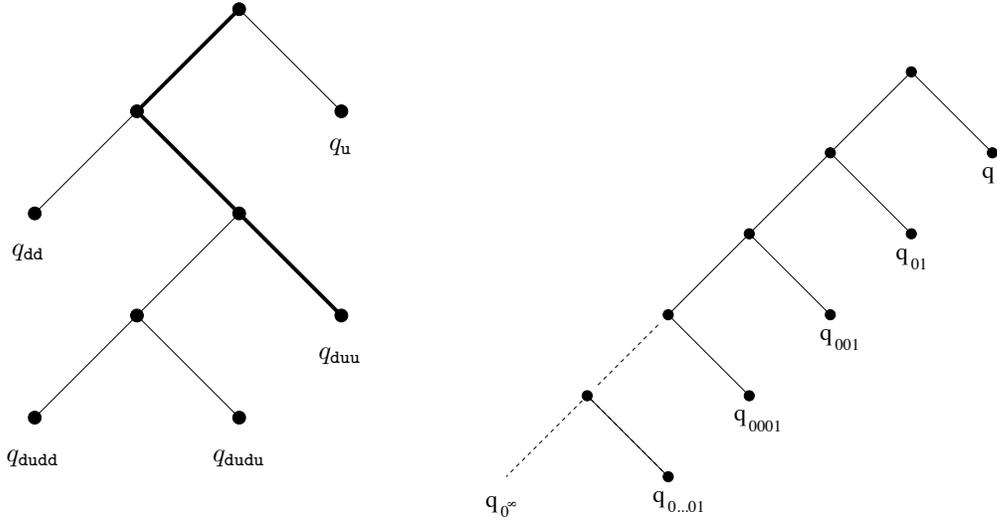


FIGURE 2.4. Examples of probabilized context tree.

admits a stationary probability measure on \mathcal{L} if and only if the numerical series $\sum c_n$ defined by

$$c_n = \prod_{k=0}^{n-1} q_{0^k 1}(0)$$

converges. Assuming the series $\sum c_n$ converges, the stationary probability measure π on \mathcal{L} is unique.

2.1.2. Towards the bestiary

The following propositions give some examples of VLMC admitting a stationary probability measure without necessary and sufficient condition of series convergence.

Proposition 2.2 (Gallo Paccaut [19]). *Let $(U_n)_{n \geq 0}$ be a VLMC. Assume there is a finite number of infinite contexts. Moreover, assume there exists $\varepsilon > 0$ such that*

$$\forall c \in \mathcal{C}, \quad \forall \alpha \in \mathcal{A}, \quad \varepsilon < q_c(\alpha) < 1 - \varepsilon.$$

Then the Markov process $(U_n)_{n \geq 0}$ admits a unique invariant measure.

Proposition 2.3 (Meyn and Tweedies [39]). *Let $(U_n)_{n \geq 0}$ be a VLMC defined on $\mathcal{A} = \{0, 1\}$. Assume that*

$$\forall c \in \mathcal{C}, \quad \forall \alpha \in \mathcal{A}, \quad q_c(\alpha) \neq 0, 1.$$

If $1 \in \mathcal{C}$ and for some integer $a \geq 1$, $0^a \in \mathcal{C}$ (or $0, 1^a$ are in \mathcal{C}), then the Markov process $(U_n)_{n \geq 0}$ admits a unique invariant measure.

2.2. Persistent random walks

Classical random walks are usually defined from a sequence of independent and identically distributed (*i.i.d.*) increments $(X_n)_{n \geq 1}$ by

$$S_0 := 0 \quad \text{and} \quad S_n := \sum_{k=1}^n X_k \quad \text{for all integers } n \geq 1. \tag{2.2}$$

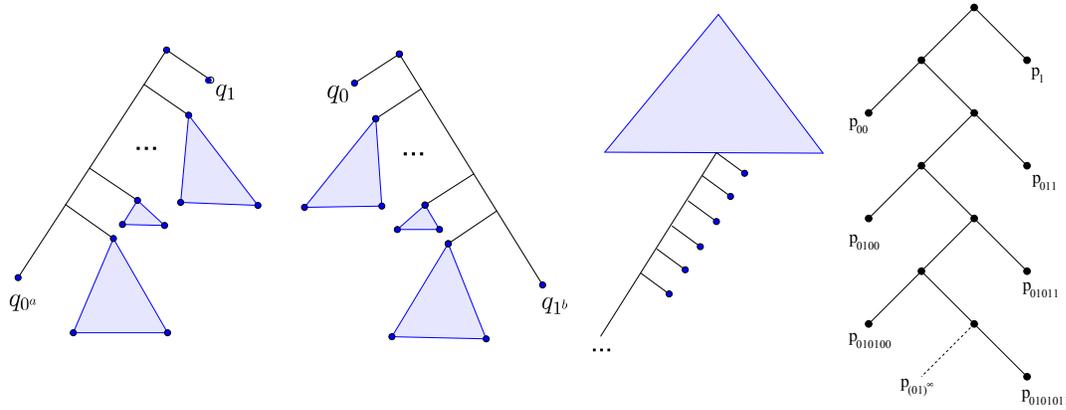


FIGURE 2.5. Assume that $\forall c \in \mathcal{C}, \forall \alpha \in \mathcal{A}, q_c(\alpha) \neq 0, 1$. For all these context trees, the Markov process $(U_n)_{n \geq 0}$ admits a unique invariant measure without any condition. For the first and the second tree, the existence and unicity is a consequence of Proposition 2.3. For the last tree, called the *bamboo blossom*, an explicit formula for the invariant measure can be found in [10]. The existence and unicity can also be viewed as a corollary of Proposition 2.3. For the third tree, a pedestrian and quite long calculation leads to an explicit formula as for the bamboo.

When the jumps are defined as a finite-order Markov chain, a short memory in the dynamics of the stochastic paths is introduced and the random walk $(S_n)_{n \geq 0}$ itself is no longer Markovian. Such a process is called in the literature a Persistent Random Walk (PRW), a Goldstein-Kac random walk or also a correlated random walk. Concerning the genesis of the theory, we allude to [16, 23, 31, 43, 49, 50] as regards the discrete-time situation but also its connections with the continuous-time telegraph process and we refer to [27, 35] concerning recurrence and transience features. Here, we aim at investigating the asymptotic behavior of one-dimensional PRW for which the increments are driven by a VLMC (an infinite-order Markov chain) built from a probabilized context tree. This construction furnishes an extended model for the dependence of the increments of PRWs which can be easily adapted to various situations.

Let $(U_n)_{n \geq 0}$ be a VLMC defined on the alphabet $\mathcal{A} = \{\mathbf{d}, \mathbf{u}\}$. The n^{th} increment X_n of the corresponding PRW is given as the latest letter of U_n , with the one-to-one correspondance $\mathbf{d} = -1$ (for a descent) and $\mathbf{u} = 1$ (for a rise). $X_n := +1$ if $U_n = U_{n-1}\mathbf{u}$ whereas $X_n := -1$ if $U_n = U_{n-1}\mathbf{d}$. Different probabilized context trees lead to different probabilistic impacts on the asymptotic behavior of resulting PRWs. Besides, the characterization of the recurrent *versus* transient behavior, the so-called type problem, is difficult in general. Here, we state exhaustive criteria for PRWs defined from a double-infinite comb context tree introduced in [9]. Roughly speaking, the leaves – coding for the memory – are the words on $\{\mathbf{d}, \mathbf{u}\} \simeq \{-1, 1\}$ of the form $\mathbf{d}^n \mathbf{u}$ and $\mathbf{u}^n \mathbf{d}$. In other words, the probability to invert the current direction depends only on its length. In addition, we derive sufficient conditions for the type of PRWs built from a larger class of context trees obtained by some grafting of the original double-infinite comb model. Foremost, we refer to Figure 2.7 that illustrates our notations and assumptions by a realization of S , for the so called double-infinite comb PRW. In order to avoid trivial cases, we assume that S cannot be frozen in one of the two directions with a positive probability – it makes infinitely many U-turns almost surely. Besides, for the sake of simplicity, we deal throughout this paper with the conditional probability with respect to the event $(X_0, X_1) = (\mathbf{u}, \mathbf{d})$ – the initial time is suppose to be an up-to-down turn. Obviously, there is no loss of generality supposing this and the long time behavior of S is not affected as well. Therefore, we assume the following.

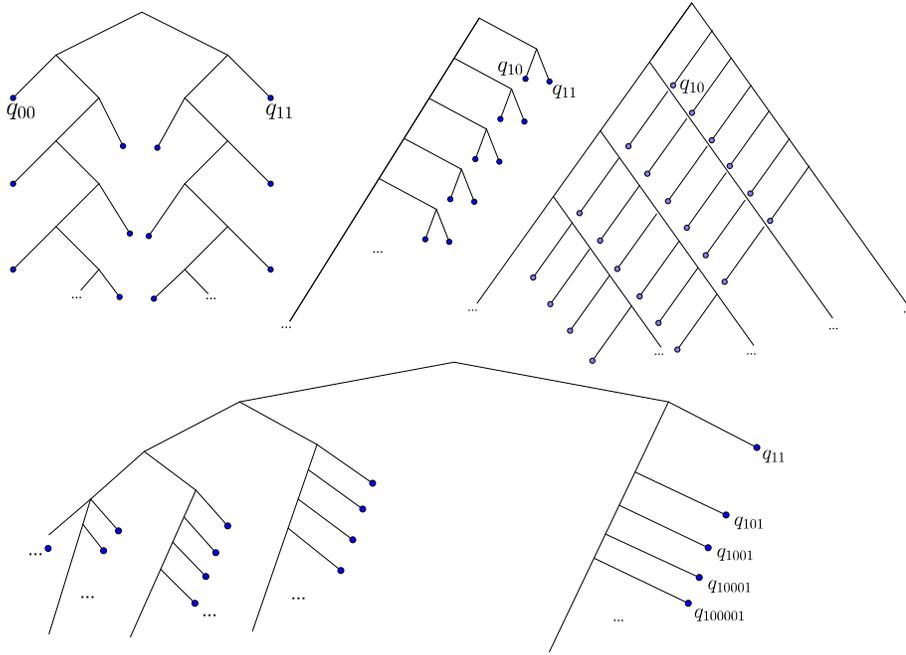


FIGURE 2.6. Assume that $\forall c \in \mathcal{C}, \forall \alpha \in \mathcal{A}, q_c(\alpha) \neq 0, 1$. For all these context trees, with a bit calculations, we prove that the Markov process $(U_n)_{n \geq 0}$ admits a unique invariant measure if and only if a specific series converges.

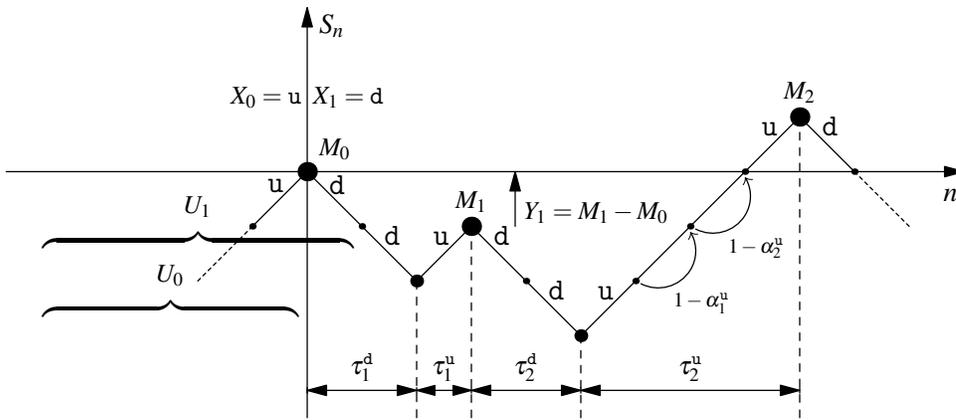


FIGURE 2.7. Persistent random walk

Assumption 2.1 (finiteness of the length of runs).

$$\prod_{k=1}^{\infty} (1 - q_{d^k u}(u)) = 0 \tag{2.3}$$

and symmetrically

$$\prod_{k=1}^{\infty} (1 - q_{u^k d}(d)) = 0. \tag{2.4}$$

Let τ_n^u and τ_n^d be respectively the length of the n^{th} rise and of the n^{th} descent, that is

$$\tau_n^u = \inf \{j \geq \tau_n^d | X_j = \mathbf{d}\} - 1, \quad \tau_n^d = \inf \{j \geq \tau_{n-1}^u | X_j = \mathbf{u}\} - 1.$$

Then, by the renewal property, (τ_n^d) and (τ_n^u) are independent sequences of *i.i.d.* random variables. In order to deal with a more tractable random walk built with possibly unbounded but *i.i.d.* increments, we introduce the underlying skeleton random walk $(M_n)_{n \geq 0}$ associated with the even U-turns – the original walk observed at the random times of up-to-down turns. Note that the expectation \mathbf{d}_M of an increment $Y_n := \tau_n^u - \tau_n^d$ is meaningful whenever one of the persistence times (at least) is integrable. In this situation, we can set (extended by continuity whenever one of the persistence times is not integrable)

$$\mathbf{d}_S := \frac{\mathbb{E}[\tau_1^u] - \mathbb{E}[\tau_1^d]}{\mathbb{E}[\tau_1^u] + \mathbb{E}[\tau_1^d]} \in [-1, 1]. \tag{2.5}$$

In regards to the convergence (2.6) below, the latter quantity is naturally called the almost sure drift of the double-infinite comb PRW S when it exists.

Proposition 2.4 (well-defined drift case (see [12])). *Assuming 2.1, the double-infinite comb PRW S is recurrent if and only if $\mathbf{d}_S = 0$ and transient otherwise. Furthermore, one has*

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = \mathbf{d}_S \quad \text{a.s.} \tag{2.6}$$

We consider the remaining case in which both $\mathbb{E}[\tau_1^u]$ and $\mathbb{E}[\tau_1^d]$ are infinite. Let us define

$$J_{\ell_1 | \ell_2} := \sum_{n=1}^{\infty} \frac{n \mathbb{P}(\tau^{\ell_1} = n)}{\sum_{k=1}^n \mathbb{P}(\tau^{\ell_2} \geq k)}, \tag{2.7}$$

for any $\ell_1 \neq \ell_2 \in \{\mathbf{u}, \mathbf{d}\}$.

Theorem 2.1 (undefined drift case (see [12])). *Assuming 2.1, the double-infinite comb PRW S is recurrent if and only if $J_{\mathbf{u}|\mathbf{d}}$ and $J_{\mathbf{d}|\mathbf{u}}$ are both infinite. Moreover, if $J_{\mathbf{u}|\mathbf{d}} = \infty$ (resp. $J_{\mathbf{d}|\mathbf{u}} = \infty$), then*

$$\limsup_{n \rightarrow \infty} \frac{S_n}{n} = 1 \quad \text{a.s.} \quad \left(\liminf_{n \rightarrow \infty} \frac{S_n}{n} = -1 \quad \text{a.s.} \right). \tag{2.8}$$

In particular, it is transient to ∞ (resp. $-\infty$) if and only if $J_{\mathbf{u}|\mathbf{d}}$ (resp. $J_{\mathbf{d}|\mathbf{u}}$) is infinite.

The case when the characteristics $J_{\mathbf{u}|\mathbf{d}}$ and $J_{\mathbf{d}|\mathbf{u}}$ are both finite do not appear in the latter theorem since in that situation it follows from [17] that the persistence times are both integrable and hence it reduces to the well-defined drift case in Proposition 2.4.

Consider a double-infinite comb and attach to each finite leaf c another context tree \mathbb{T}_c (possibly trivial) as in Figure 2.8. The leaves of the related graft are denoted by \mathcal{C}_c and this one is endowed with Bernoulli distributions $\{q_\ell : \ell \in \mathcal{C}_c\}$ on $\{\mathbf{u}, \mathbf{d}\}$. Note that any probabilized context tree on $\{\mathbf{u}, \mathbf{d}\}$ can be constructed this way. We denote by $S^{\mathfrak{g}}$ the corresponding PRW. In that case, the random walk is particularly persistent, in the sense that the rises and descents are no longer independent. A renewal property may persist but it is heavier

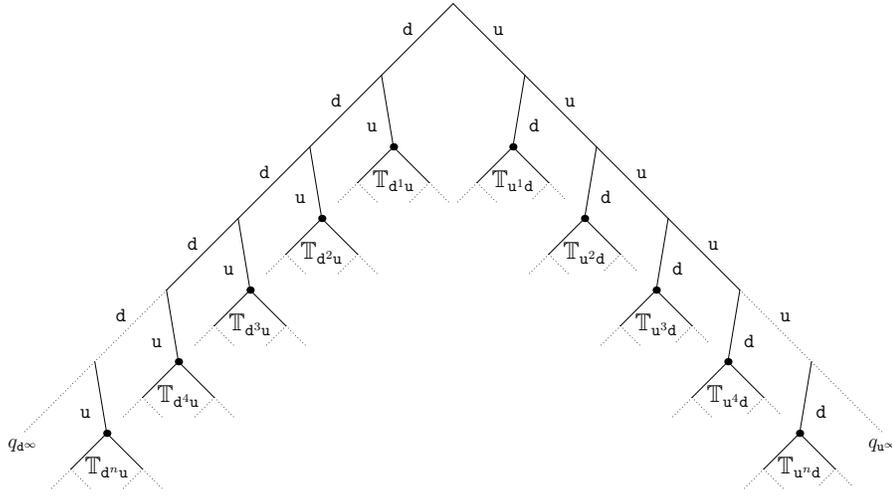


FIGURE 2.8. Grafting of the double-infinite comb

to write in general. Let \underline{S} and \overline{S} be the double-infinite comb PRWs with respective transition

$$\alpha_n^u := \sup\{q_c(d) : c \in \mathcal{C}_{u^n d}\}, \quad \alpha_n^d := \inf\{q_c(u) : c \in \mathcal{C}_{d^n u}\},$$

$$\text{and } \overline{\alpha}_n^u := \inf\{q_c(d) : c \in \mathcal{C}_{u^n d}\}, \quad \overline{\alpha}_n^d := \sup\{q_c(u) : c \in \mathcal{C}_{d^n u}\}. \quad (2.9)$$

Lemma 2.1 (comparison of grafts). *With probability one, one has*

$$\limsup_{n \rightarrow \infty} \underline{S}_n = \infty \implies \limsup_{n \rightarrow \infty} S_n^g = \infty \quad \text{and} \quad \liminf_{n \rightarrow \infty} \overline{S}_n = -\infty \implies \liminf_{n \rightarrow \infty} S_n^g = -\infty.$$

As a consequence, if \underline{S} and \overline{S} are of the same recurrent or transient type, then S is recurrent or transient accordingly. We can extend our results to some probabilized context trees satisfying Assumption 2.1, mainly when the persistence times of \underline{S} or \overline{S} are both non-integrable.

Corollary 2.1 (extended criterion). *Assume that there are a finite number of grafts to the double-infinite-comb such that either \underline{S} or \overline{S} has running times of infinite mean. Then the general PRW S^g together with the double-infinite comb PRWs \underline{S} and \overline{S} are of the same type.*

3. THE COUPLING FROM THE PAST ALGORITHM AND TWO RECENT APPLICATIONS

3.1. Perfect sampling

Perfect sampling was introduced in 1996 by Propp and Wilson [42]. This technique allows one to generate unbiased samples from the stationary distribution of an ergodic Markov chain over a finite state space. The key feature is the use of coupled Markov chains originating from all possible states.

3.1.1. Markov automata

For the purpose of simulations, a Markov chain $(X_n)_{n \in \mathbb{N}}$ with transition matrix P is represented by means of a *Markov automaton* \mathbf{A} (not to be confused with the cellular automata discussed in Section 1). In addition to the state space \mathcal{S} , \mathbf{A} comprises a finite event space \mathcal{A} , a distribution ρ and a *transition function* $\cdot : \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{S}$. The Markov chain $(X_n)_{n \in \mathbb{N}}$ is obtained by generating an i.i.d. sequence $(A_n)_{n \in \mathbb{N}}$ according to ρ , and recursively

setting $X_{n+1} = X_n \cdot A_{n+1}$. The sequence $(X_n)_{n \in \mathbb{N}}$ is clearly a Markov chain, and if the Markov automaton is chosen such that $\mathbf{P}(i \cdot A_1 = j) = P_{i,j}$ for all $i, j \in \mathcal{S}$, then its transition matrix is P .

3.1.2. *Coupling From The Past*

The algorithm described here is based on the following intuition: a Markov chain that has been running “infinitely long” is now distributed according to its stationary distribution π . Starting this Markov chain *arbitrarily far back in time* and determining its state at time 0 gives a sample distributed according to π . The proof underlying this method can be found in [42].

The objective is to determine this final state using only a finite amount of information. The key idea is to consider only the last k_1 steps of the Markov chain, generated by the sequence of events $A_{-(k_1-1)}, \dots, A_0 \sim \rho^{\otimes k_1}$. As we have no means of determining the value of X_{-k_1} , we take into account all possibilities, i.e. we compute X_0 for all possible starting states at time $-k_1$. This gives us a set of plausible end states

$$S_0 = \{s \cdot A_{-(k_1-1)} \cdot \dots \cdot A_0, s \in \mathcal{S}\}.$$

If S_0 is a singleton, then regardless of X_{-k_1} , X_0 is the unique element of S_0 , which is therefore distributed according to π . Otherwise, it simply means we have not considered a large enough portion of the past; we therefore take a $k_2 > k_1$, generate new events $A_{-(k_2-1)}, \dots, A_{-k_1} \sim \rho^{\otimes k_2-k_1}$ independently from the previous ones, and compute $S_2 = \{s \cdot A_{-(k_2-1)} \cdot \dots \cdot A_0, s \in \mathcal{S}\}$. By reiterating for increasing values of k_i until S_i is a singleton, we have an algorithm which returns a state distributed according to π .

Algorithm 1: Coupling From The Past [42]

```

Data:  $\mathcal{S}, A, \rho$  et .
1 begin
2    $i \leftarrow 0;$ 
3    $A \leftarrow ();$ 
4    $S \leftarrow \mathcal{S};$ 
5   while  $|S| > 1$  do
6      $i \leftarrow i + 1;$ 
7      $\text{prefix} \sim \rho^{\otimes k_i};$ 
8      $A \leftarrow \text{concatenate}(\text{prefix}, A);$ 
9      $S \leftarrow S \cdot A$  // Possibly computed using only prefix and  $S$ 
10  return  $S$ 

```

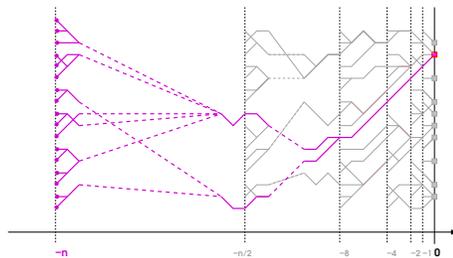


FIGURE 3.9. Successive iterations for Algorithm 1.

The exact choice of the $(k_i)_{i \in \mathbb{N}}$ does not change the distribution of the output, but can impact complexity. When taking $k_i = i$ deterministically, for example, one can maintain a unique vector \mathbf{s} in memory, updating it at each iteration by setting $\mathbf{s}_{i+1}(x) = \mathbf{s}_i(x \cdot A_{-i})$. We then have that $S_i = \{\mathbf{s}_i(x) | x \in \mathcal{S}\}$, i.e. the algorithm

terminates when all the entries in \mathbf{s} are equal. In the case of monotonicity (described below), it is no longer possible to work with this unique vector: the entire trajectories must be computed once more. Doubling the value of k between successive iterations, however, greatly reduces the execution time of the algorithm [42].

Furthermore, call τ the first k_i such that S_i is a singleton; regardless of the choice of $(k_i)_{i \in \mathbb{N}}$, so long as one can construct a finite sequence A_1, \dots, A_l with positive probability such that $\{s \cdot A_1 \cdot \dots \cdot A_l, s \in \mathcal{S}\}$ is a singleton, then τ has finite expectation, i.e. the algorithm 1 terminates in finite expected time. For bounds on this running time, see [42].

3.1.3. Reducing complexity through envelopes

The above algorithm can be refined in many ways, such as that described in [51], which does not require storing the $(A_{-n})_{n \in \mathbb{N}}$. In general, sampling techniques based on Markov chains, often referred to as Monte Carlo methods, are used when sampling from a large, multi-dimensional set \mathcal{S} for which computing the exact distribution is not an option. Coupling From The Past is interesting in that it generates unbiased samples, but in its raw form, its complexity is linear in the size of \mathcal{S} . Techniques have therefore been introduced to overcome this limitation, the most common of which use some form of *envelope*.

Informally, an envelope E implicitly or explicitly defines a subset of \mathcal{S} , comprising the states it encompasses. The idea is to define a low-dimension set of envelopes which allow us to keep track of the evolution of our Markov chains at a reduced cost: the transition function \cdot is generalized in order to apply to these envelopes, after which we run the simulation over the envelopes (S_i becomes $S'_i = S'_0 \cdot A_{-(k_i-1)} \cdot \dots \cdot A_0$, where S'_0 encompasses all of \mathcal{S}). The algorithm now terminates when S'_i comprises only one state.

Many envelope-based strategies exist [22, 30, 33, 40]. The most common examples rely on an ordering of the set space; this is the case for both the monotonous [42] and non-monotonous [8] examples described below.

3.1.4. Monotonous chains

Suppose that one can define a partial order \leq over \mathcal{S} such that there exists an upper bound s_{\max} and a lower bound s_{\min} for \leq , and such that the following holds:

$$\forall i, j \in \mathcal{S}, \forall A \in \mathcal{A}, \quad s_i \leq s_j \implies s_i \cdot A \leq s_j \cdot A.$$

Such a chain is said to be monotonous for \leq .

Consider the set of intervals $\{(m, M), m \leq M\}$, where $[m, M] = \{s \in \mathcal{S} \mid m \leq s \leq M\}$. Taking $S'_0 = (s_{\min}, s_{\max})$ and $(m, M) \cdot A = (m \cdot A, M \cdot A)$, one can verify that, for all $i \in \mathbb{N}$, S'_i contains all the states in S_i [42]. A direct consequence of this is that, if ever $S'_i = (m_i, M_i)$ is such that $m_i = M_i$, then S_i is a singleton, and its unique element, m_i , is distributed according to π .

3.1.5. Non-monotonous chains

Suppose now that the state space is still partially ordered, but that the dynamic is no longer monotonous. We still use envelopes of the form (m, M) , with $m_0 = s_{\min}$ and $M_0 = s_{\max}$, but the generalization of the transition function is altered:

$$(m, M) \cdot A \stackrel{(\text{def.})}{=} \left(\inf_{m \leq s \leq M} s \cdot A, \sup_{m \leq s \leq M} s \cdot A \right).$$

Like before, if this dynamic over the envelopes leads to a point where $m_i = M_i$, then m_i is distributed according to the stationary distribution π [8].

Notice that with this technique, (m_i) and (M_i) are *not* trajectories of the initial dynamic. Furthermore, one must verify that the algorithm converges; the fact that there exists a sequence A_1, \dots, A_l such that $\{s \cdot A_1 \cdot \dots \cdot A_l, s \in \mathcal{S}\}$ is a singleton does not imply that there exists a sequence A'_1, \dots, A'_k such that $(s_{\min}, s_{\max}) \cdot A'_1 \cdot \dots \cdot A'_k = (s, s)$ for some $s \in \mathcal{S}$.

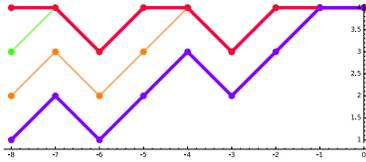


FIGURE 3.10. Monotonous chain.

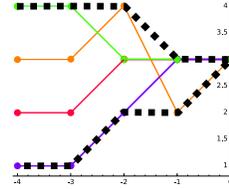


FIGURE 3.11. Non monotonous chain.

3.2. Coupling from the past of closed queuing networks

3.2.1. Closed queuing networks model

Consider a closed queuing network of K queues and M customers in total. Each queue k has a finite capacity $C_k \leq M$ and a single server with independent exponential services of rate μ_k . After a customer has been served in queue i , it is routed to queue j with probability $p_{i,j}$, independently of the current state and past evolution of the network. If queue j is full, then the customer remains in queue i and waits to be served again (a new destination will then be drawn, independently from the previous one). A configuration is represented by a vector $\mathbf{x} \in \mathcal{S}$ where x_i represents the number of customers in queue i . The state space \mathcal{S} being defined as

$$\mathcal{S} = \left\{ \mathbf{x} = (x_1, x_2, \dots, x_K) \in \mathbb{N}^K \mid \sum_{k=1}^K x_k = M \text{ and } 0 \leq x_k \leq C_k, \forall k \right\}.$$

If all queues have an infinite capacity (*i.e.* if for each $k \in \{1, \dots, K\}$, $C_k = M$) then $|\mathcal{S}| = \binom{K+M-1}{K-1}$. Thus for $M \gg K$ the complexity of the state space is $O(M^K)$.

The dynamics of the system defines a Markov chain $(X_n)_{n \in \mathbb{N}}$, represented by the Markov automaton composed of the event space $\{t_{i,j} \mid 1 \leq i, j \leq K\}$, the distribution ρ such that $\rho(t_{i,j}) = \frac{\mu_i p_{i,j}}{\sum_{k=1}^K \mu_k}$, and the transition function $t_{i,j}$ which describe routing from queue i to queue j is defined by

$$\mathbf{x} \cdot t_{i,j} = \mathbf{x} + \mathbf{1}_{x_i > 0, x_j < C_j} (\mathbf{e}_j - \mathbf{e}_i).$$

The aim is to sample the stationary distribution with perfect sampling techniques. The chain is non-monotonous and the cardinality of the state space is $O(M^K)$, however Bouillard et al. [4] developed a new envelope-based strategy to improve the perfect sampling algorithm for this model. The idea is to exploit the constraint of the total number of customers in the network to represent any set of states as a directed graph, called *diagram*.

3.2.2. Set of states as a diagram

In a diagram, a state is encoded by a path of length K starting at node $(0, 0)$ and ending at node (K, M) , such that the arc $((k-1, m), (k, m'))$ belongs to the path if and only if there are $m' - m$ customers in queue k . A given diagram envelopes all configurations for which the encoding is present. A diagram is said to be *complete* if it encodes the state space \mathcal{S} . A complete diagram contains at most $O(KM^2)$ arcs. In Figures 3.12 and 3.13, we illustrate the encoding of the configuration $(0, 0, 2, 0, 1)$, as well as the complete diagram for a network with $K = 5$ queues, $M = 3$ customers and a capacity vector $C = (2, 1, 3, 1, 2)$.

3.2.3. Transition in a diagram

The transition induced by event $t_{i,j}$ is applied directly on the diagram D . To compute this transition, we need to know what paths of D will remain unchanged, either because the number of customers in queue i is equal to 0 (set of arcs *Stay*), or because queue j is full (set of arcs *Full*). From there, we deduce which arcs will be modified (set of arcs *Transit*). In accordance with $t_{i,j}$, these arcs are modified as follows:

- In column i , the slopes of the arcs are decreased by one.

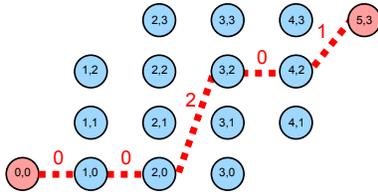


FIGURE 3.12. Encoding of $(0, 0, 2, 0, 1)$.

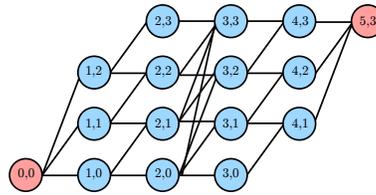


FIGURE 3.13. Complete diagram.

- In columns $k \in \{\min(i, j) + 1, \dots, \max(i, j) - 1\}$, the arcs are shifted up or down, depending whether $i < j$ or $i > j$.
- In column j , the slopes of the arcs are increased by one.

Figure 3.14 shows how the diagram D is modified by event $t_{4,2}$. First, the sets $Stay$, $Full$ and $Transit$ are determined according the values of arcs in columns 4 and 2. As an example, arcs in $Stay$ are along paths such that arcs in column 4 represent 0 customers. Afterward, arcs in $Transit$ are updated in accordance with $t_{4,2}$, that corresponds to the set $Transit'$. Finally, $T_{4,2}(D)$ is the diagram that contains arcs in $Stay \cup Full \cup Transit'$.

Each transition has complexity $O(KM^2)$, that corresponds to the number of arcs in a diagram.

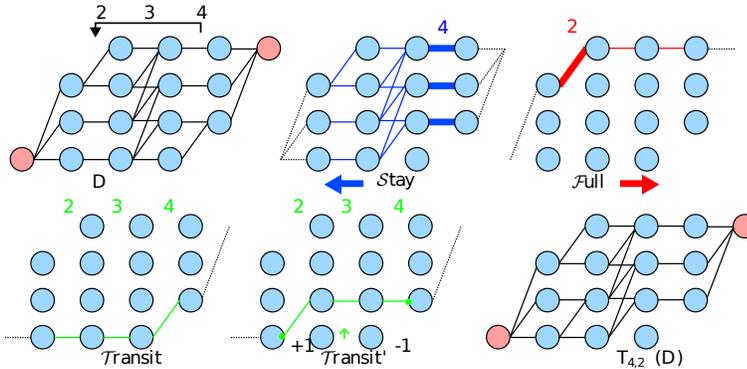


FIGURE 3.14. Transition induced by event $t_{4,2}$.

3.2.4. Perfect simulation

To use diagrams with Coupling From the Past algorithm (Algorithm 1), we start by replacing the set \mathcal{S} with the complete diagram, and apply the transitions $t_{i,j}$ on this diagram. The algorithm terminates when the final diagram contains only one path. This path corresponds to a state. Bouillard et al. [4] proved that there exists a finite sequence of events $(t_{i_1, j_1}, \dots, t_{i_N, j_N})$ that reduces the complete diagram into a diagram that contains only one path, thereby proving the following theorem:

Theorem 3.1. *The coupling from the past algorithm using diagrams terminates in finite expected time and produces an exact sample from the stationary distribution.*

3.3. Accelerating Coupling From The Past

3.3.1. The hard-core gas model

For a simple graph $G = (V, E)$, define an independent set as a subset I of V , such that no two vertices of I are connected by an edge in E . Denote \mathcal{I} the set of independent sets.

The hard-core gas model, originating from statistical physics [20], looks at the feasible configurations for a gas. In this model, particles occupy sites, represented by vertices, and incompatible sites are connected by

an edge. A configuration is considered feasible if no two incompatible sites are occupied, *i.e.* if it forms an independent set. The likelihood of each configuration is of the form

$$\pi(I) = \frac{\lambda^{|I|}}{Z_\lambda}.$$

The positive parameter λ is called the *fugacity* of the system, and Z_λ is a normalizing constant.

We are interested in generating random samples according to the distribution π . This cannot be done in a straightforward fashion, as that would require computing Z_λ , which is $\#P$ -complete^a. We therefore rely of Monte Carlo methods, and namely on Coupling From the Past.

A simple dynamic for defining a Markov chain with stationary distribution π is the following: let $\mathcal{A} = V \times \{+, -\}$ be the set of events, and ρ be the distribution over \mathcal{A} such that $\rho(u^+) = \frac{1}{|V|} \frac{\lambda}{1+\lambda}$ and $\rho(u^-) = \frac{1}{|V|} \frac{1}{1+\lambda}$ (u is a vertex chosen uniformly at random, after which we pick $+$ or $-$ independently with respective probabilities $\frac{\lambda}{1+\lambda}$ and $\frac{1}{1+\lambda}$). Consider the transition function \cdot defined by

$$I \cdot u^- = I \setminus \{u\}$$

and

$$I \cdot u^+ = \begin{cases} I \cup \{u\} & \text{if } I \cup \{u\} \in \mathcal{I}, \\ I & \text{otherwise.} \end{cases}$$

The resulting Markov automaton defines a Markov chain whose stationary distribution is π . An enhanced variant of this dynamic can be found in [30].

For coupling from the past, we build an envelope based upon the partial ordering induced by inclusion [30]: we maintain two sets $m, M \subseteq V$ such that, regardless of the initial state of a trajectory X , we have $m_i \subseteq X_i \subseteq M_i$ at each time step i . Note that m and M need not be independent sets themselves; indeed, to envelope trajectories emanating from all states, it is necessary to take $m_0 = \emptyset$ and $M_0 = V$. The transitions are then applied to the envelope as followed:

$$(m, M) \cdot u^- = (m \setminus \{u\}, M \setminus \{u\})$$

and

$$(m, M) \cdot a^+ = \begin{cases} (m \cup \{u\}, M \cup \{u\}) & \text{if } M \cup \{u\} \in \mathcal{I}, \\ (m, M \cup \{u\}) & \text{if } m \cup \{u\} \in \mathcal{I}, \\ (m, M) & \text{otherwise.} \end{cases}$$

This envelope can be used to run the Coupling From the Past algorithm for the hard-core gas model. Bounds on its complexity can be found in [30].

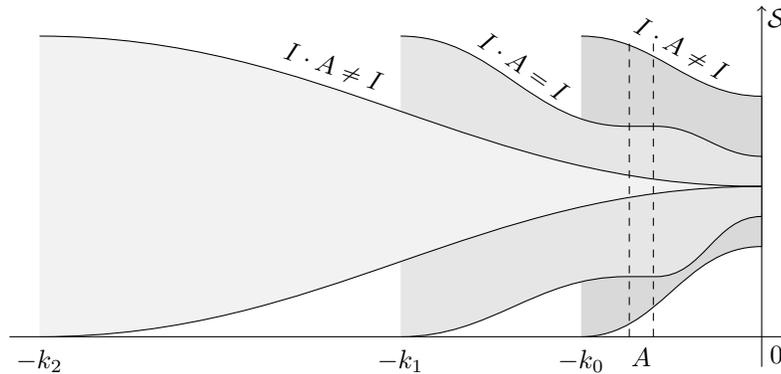
3.3.2. Event skipping

The overall Monte Carlo scheme for the hard-core gas model suffers from an inherent weakness: it can easily be verified that the probability of having $I \cdot a = I$ for $a \sim \rho$ is at least $\frac{1}{2}$, and can be arbitrarily close to 1 for eccentric values of λ (either very large or very small with respect to 1). For typical Monte Carlo, this is not necessarily an issue, as one can sample a conditioned to having $I \cdot a \neq I$. However, one must pay special attention to the time at which the simulation is stopped, in order to make sure that the output is not biased. Since the conditional distribution of a could be obtained through rejection sampling (though this method would not reduce the complexity, since “skipped” events are still sampled), a simple coupling argument suffices to show that the number of omitted transitions one should account for at each step of the algorithm is geometrically distributed. Counting these transitions allows one to stop the algorithm in such a way as to obtain an unbiased output.

For Coupling From the Past, this strategy can be adapted, but requires even more care. Indeed, this simple conditioning cannot be applied, as a given event A_{-i} does not always apply to the same state: it is first applied

^a $\#P$ is the complexity class of counting problems associated to decision problems in NP.

to $(m_0, M_0) \cdot A_{-(k_0-1)} \cdot \dots \cdot A_{-(i+1)}$, but then as we start the simulation from further back, it is also applied to $(m_0, M_0) \cdot A_{-(k_1-1)} \cdot \dots \cdot A_{-(i+1)}$, and so on (cf. Figure 3.15). The sequence of events cannot simply be re-sampled independently from the preceding ones every time we re-run the simulation from further back in time; it is therefore necessary to introduce an update scheme for the sequence of events. This scheme must both remove events which do not modify the “current” state of the envelope, and insert new events according to a precise distribution, so as to compensate any bias that might result from removals in previous states. Note that we cannot simply keep track of which events have been removed: this presents the same caveat as rejection sampling, where the overall complexity is unchanged, since all the events are still sampled. The exact details of this scheme, as well as the proof of its correctness, can be found in [47].



Notice that A modifies I for the first and third runs, but not for the second one. Removing A from the sequence of events during the second run requires compensating for this removal during the third run.

FIGURE 3.15. Events applied to different states

REFERENCES

- [1] N. Aubrun and M. Sablik. Simulation of effective subshifts by two-dimensional subshifts of finite type. *Acta Appl. Math.*, 126:35–63, 2013.
- [2] V. Belitsky and P. A. Ferrari. Invariant measures and convergence properties for cellular automaton 184 and related processes. *J. Stat. Phys.*, 118(3-4):589–623, 2005.
- [3] F. Blanchard and P. Tisseur. Some properties of cellular automata with equicontinuity points. *Ann. Inst. H. Poincaré Probab. Statist.*, 36(5):569–582, 2000.
- [4] A. Bouillard, A. Bušić, and C. Rovetta. Perfect sampling for closed queueing networks. *Performance Evaluation*, 79(0):146–159, 2014.
- [5] L. Boyer, M. Delacourt, V. Poupet, M. Sablik, and G. Theyssier. μ -limit sets of cellular automata from a computational complexity perspective. *J. Comput. System Sci.*, 81(8):1623–1647, 2015.
- [6] L. Boyer, M. Delacourt, and M. Sablik. Construction of μ -limit sets. In *JAC*, pages 76–87, 2010.
- [7] L. Boyer, V. Poupet, and G. Theyssier. On the complexity of limit sets of cellular automata associated with probability measures. In *MFCS*, pages 190–201, 2006.
- [8] A. Bušić, B. Gaujal, and J.-M. Vincent. Perfect Simulation and Non-monotone Markovian Systems. In *3rd International Conference Valuetools'08*, Athens, Greece, 2008. ICST.
- [9] P. Cénac, B. Chauvin, S. Herrmann, and P. Vallois. Persistent random walks, variable length Markov chains and piecewise deterministic Markov processes. *Markov Process. Related Fields*, 19(1):1–50, 2013.
- [10] P. Cénac, B. Chauvin, F. Paccaut, and N. Pouyanne. Context trees, variable length Markov chains and dynamical sources. In *Séminaire de Probabilités XLIV*, volume 2046 of *Lecture Notes in Math.*, pages 1–39. Springer, Heidelberg, 2012.
- [11] P. Cénac, B. De Loynes, A. Le Ny, and Y. Offret. Persistent random walks i. *accepted for publication in Journal of Theoretical Probability*, 2017. <https://hal.archives-ouvertes.fr/hal-01135794v2/document>.
- [12] P. Cénac, A. Le Ny, B. de Loynes, and Y. Offret. Persistent random walks. i. recurrence versus transience. *Journal of Theoretical Probability*, pages 1–12, 2016.

- [13] F. Comets, R. Fernández, and P. A. Ferrari. Processes with long memory: regenerative construction and perfect simulation. *Ann. Appl. Probab.*, 12(3):921–943, 2002.
- [14] I. Csiszár. Festschrift in honor of jorma rissanen on the occasion of his 75th birthday edited by peter grünwald, petri myllymäki, ioan tabus, marcelo weinberger, bin yu. *International Statistical Review*, 76(3):436–437, 2008.
- [15] M. Delacourt, V. Poupet, M. Sablik, and G. Theyssier. Directional dynamics along arbitrary curves in cellular automata. *Theor. Comput. Sci.*, 412(30):3800–3821, 2011.
- [16] E. C. Eckstein, J. A. Goldstein, and M. Leggas. The mathematics of suspensions: Kac walks and asymptotic analyticity. In *Proceedings of the Fourth Mississippi State Conference on Difference Equations and Computational Simulations (1999)*, volume 3 of *Electron. J. Differ. Equ. Conf.*, pages 39–50. Southwest Texas State Univ., San Marcos, TX, 2000.
- [17] K. B. Erickson and R. A. Maller. Drift to infinity and the strong law for subordinated random walks and Lévy processes. *J. Theoret. Probab.*, 18(2):359–375, 2005.
- [18] P. A. Ferrari, A. Maass, S. Martínez, and P. Ney. Cesàro mean distribution of group automata starting from measures with summable decay. *Ergodic Theory Dynam. Systems*, 20(6):1657–1670, 2000.
- [19] S. Gallo and F. Paccaut. On non-regular g -measures. *Nonlinearity*, 26(3):763–776, 2013.
- [20] D. Galvin and P. Tetali. Slow Mixing of Glauber Dynamics for the Hard-core Model on Regular Bipartite Graphs. *Random Struct. Algorithms*, 28(4):427–443, July 2006.
- [21] F. García-Ramos. Limit behaviour of μ -equicontinuous cellular automata. *Theoret. Comput. Sci.*, 623:2–14, 2016.
- [22] B. Gaujal, G. Gorgo, and J.-M. Vincent. *Perfect Sampling of Phase-Type Servers Using Bounding Envelopes*, pages 189–203. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.
- [23] S. Goldstein. On diffusion by discontinuous movements, and on the telegraph equation. *Quart. J. Mech. Appl. Math.*, 4:129–156, 1951.
- [24] B. Hellouin de Menibus. *Asymptotic behavior of cellular automata: computation and randomness*. PhD thesis, Université Aix-Marseille, 2014.
- [25] B. Hellouin de Menibus and M. Sablik. Characterization of sets of limit measures of a cellular automaton iterated on a random configuration. *Ergodic Theory and Dynamical Systems*, pages 1–50, 001 2016.
- [26] B. Hellouin de Menibus and M. Sablik. Self-organisation in cellular automata with coalescent particles: qualitative and quantitative approaches. *Prépublication (<https://arxiv.org/abs/1602.06093>)*, 2016.
- [27] R. Henderson, E. Renshaw, and D. Ford. A note on the recurrence of a correlated random walk. *J. Appl. Probab.*, 20(3):696–699, 1983.
- [28] M. Hochman. On the dynamics and recursive properties of multidimensional symbolic systems. *Invent. Math.*, 176(1):131–167, 2009.
- [29] M. Hochman and T. Meyerovitch. A characterization of the entropies of multidimensional shifts of finite type. *Ann. of Math. (2)*, 171(3):2011–2038, 2010.
- [30] M. Huber. Perfect sampling using bounding chains. *Ann Appl Probab*, 14(2):734–753, 2004.
- [31] M. Kac. A stochastic model related to the telegrapher’s equation. *Rocky Mountain J. Math.*, 4:497–509, 1974. Reprinting of an article published in 1956, Papers arising from a Conference on Stochastic Differential Equations (Univ. Alberta, Edmonton, Alta., 1972).
- [32] J. Kari and S. Taati. Statistical mechanics of surjective cellular automata. *J. Stat. Phys.*, 160(5):1198–1243, 2015.
- [33] W. Kendall and J. Møller. Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Advances in Applied Probability*, 32(3):844–865, 9 2000.
- [34] P. Kůrka and A. Maass. Limit sets of cellular automata associated to probability measures. *J. Statist. Phys.*, 100(5-6):1031–1047, 2000.
- [35] M. Lenci. Recurrence for persistent random walks in two dimensions. *Stoch. Dyn.*, 7(1):53–74, 2007.
- [36] D. A. Lind. The entropies of topological Markov shifts and a related class of algebraic integers. *Ergodic Theory Dynam. Systems*, 4(2):283–300, 1984.
- [37] A. Maass and S. Martínez. On Cesàro limit distribution of a class of permutative cellular automata. *J. Statist. Phys.*, 90(1-2):435–452, 1998.
- [38] T. Meyerovitch. Growth-type invariants for \mathbb{Z}^d subshifts of finite type and arithmetical classes of real numbers. *Inventiones Mathematicae*, 184:567–589, 2011.
- [39] S. Meyn and R. Tweedie. *Markov chains and stochastic stability*. Berlin: Springer-Verlag, 1993.
- [40] K. N. Olle Häggström. On Exact Simulation of Markov Random Fields Using Coupling from the Past. *Scandinavian Journal of Statistics*, 26(3):395–411, 1999.
- [41] M. Pivato and R. Yassawi. Limit measures for affine cellular automata. *Ergodic Theory Dynam. Systems*, 22(4):1269–1287, 2002.
- [42] J. G. Propp and D. B. Wilson. Exact Sampling with Coupled Markov Chains and Applications to Statistical Mechanics. *Random Struct. Algorithms*, 9(1-2):223–252, 1996.
- [43] E. Renshaw and R. Henderson. The correlated random walk. *J. Appl. Probab.*, 18(2):403–414, 1981.
- [44] J. Rissanen. A universal data compression system. *IEEE Trans. Inform. Theory*, 29(5):656–664, 1983.

- [45] M. Sablik. Directional dynamics for cellular automata: a sensitivity to initial condition approach. *Theoret. Comput. Sci.*, 400(1-3):1–18, 2008.
- [46] P. C. Shields. *The ergodic theory of discrete sample paths*, volume 13 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1996.
- [47] R. Varloot, A. Bušić, and A. Bouillard. Speeding up Glauber Dynamics for Random Generation of Independent Sets. *CoRR*, abs/1504.04517, 2015.
- [48] K. Weihrauch. *Computable analysis*. Texts in Theoretical Computer Science. An EATCS Series. Springer-Verlag, Berlin, 2000. An introduction.
- [49] G. H. Weiss. *Aspects and applications of the random walk*. Random Materials and Processes. North-Holland Publishing Co., Amsterdam, 1994.
- [50] G. H. Weiss. Some applications of persistent random walks and the telegrapher’s equation. *Phys. A*, 311(3-4):381–410, 2002.
- [51] D. B. Wilson. How to Couple from the Past Using a Read-once Source of Randomness. *Random Struct. Algorithms*, 16(1):85–113, Jan. 2000.
- [52] S. Wolfram. Universality and complexity in cellular automata. *Phys. D*, 10(1-2):1–35, 1984. Cellular automata (Los Alamos, N.M., 1983).