Ergod. Th. & Dynam. Sys. (2014), **34**, 1567–1586 © *Cambridge University Press, 2013*
doi:10.1017/etds.2013.7

Attractive regular stochastic chains: perfect simulation and phase transition

SANDRO GALLO† and DANIEL Y. TAKAHASHI‡

† *Departamento de Métodos Estatísticos, Instituto de Matemática, Universidade Federal de Rio de Janeiro, Caixa Postal 68350, CEP 21941-909, Rio de Janeiro, Brasil*
(e-mail: sandro@im.ufrj.br)

‡ *Neuroscience Institute and Psychology Department, Princeton University, Greenhall, Princeton, NJ 08540, USA*
(e-mail: takahashiyd@gmail.com)

(*Received* 24 *November* 2011 *and accepted in revised form* 20 *December* 2012)

Abstract. We prove that uniqueness of the stationary chain, or equivalently, of the g-measure, compatible with an attractive regular probability kernel is equivalent to either one of the following two assertions for this chain: (1) it is a finitary coding of an independent and identically distributed (i.i.d.) process with countable alphabet; (2) the concentration of measure holds at exponential rate. We show in particular that if a stationary chain is uniquely defined by a kernel that is continuous and attractive, then this chain can be sampled using a coupling-from-the-past algorithm. For the original Bramson–Kalikow model we further prove that there exists a unique compatible chain if and only if the chain is a finitary coding of a finite alphabet i.i.d. process. Finally, we obtain some partial results on conditions for phase transition for general chains of infinite order.

1. *Introduction*

In this work we consider chains of infinite order, or equivalently g-measures, on a finite alphabet. These are processes specified by kernels of transition probabilities that can depend on the whole past and include as special cases the finite order Markov chains, stochastic models that exhibit phase transitions [**1, 2, 15**], and models that are not Gibbsian [**8**]. An important question for this class of models is what properties distinguish kernels exhibiting phase transition from kernels satisfying uniqueness. This work gives necessary and sufficient conditions for the existence of phase transition for an important class of chains of infinite order, namely, the attractive regular chains.

A probability kernel is called regular when it is strongly non-null and continuous with respect to the past. The regularity of the kernel guarantees the existence of at

https://doi.org/10.1017/etds.2013.7 Published online by Cambridge University Press

least one chain compatible with the kernel. Attractiveness means that the transition probabilities exhibit a monotonicity property and it is analogous to the attractiveness of the specifications considered in statistical mechanics [**24**].

It was quite unexpected when Bramson and Kalikow [**2**] showed an example of a family of regular and attractive kernels with more than one compatible chain. Some interesting results are based on this example (which we refer to as the *BK example*). For instance, [**26**] used the BK example to construct a C^1 expanding map of the circle which preserves Lebesgue measure and such that the system is ergodic, but not weak-mixing. Also using the BK example, Stenflo [**30**] showed a counterexample to a conjecture raised by Karlin [**18**]. Lacroix [**20**] simplified the Bramson and Kalikow [**2**] proof of phase transition and Hulse [**15**] showed a different example of a regular and attractive kernel exhibiting phase transition using similar ideas to those in [**20**]. To the best of our knowledge, Berger *et al* [**1**] exhibited the only non-attractive example of phase transition that is not based on the BK example. Despite the importance of these works, none of them give general sufficient conditions for the existence of phase transition, even for the special (but important) case of attractive kernels.

In the present work, we prove that for attractive regular kernels, uniqueness of the stationary chain compatible with a kernel is equivalent to either one of the following two assertions: (1) the compatible chain is a finitary coding of a countable alphabet independent and identically distributed (i.i.d.) process; (2) the concentration of measure holds at exponential rate. Condition (1) means that the compatible chain is a factor of a countable alphabet i.i.d. process whose mapping depends almost surely on a finite number of coordinates with respect to the measure of the i.i.d. process. Condition (2) means that phase transition yields loss of 'good' concentration of measure. We show that the regularity of the kernel is essential for the three equivalences to hold and, in general, cannot be relaxed. We also obtain some partial results for non-regular and non-attractive cases, which are of independent interest.

The main ingredient for the proof of the existence of a finitary coding is the proof that uniqueness of the compatible chain for regular and attractive kernels is equivalent to the existence of a coupling-from-the-past (CFTP) perfect simulation algorithm. This class of simulation algorithms was first introduced for Markov chains by Propp and Wilson [**25**] and then generalized to several other stochastic models. Despite its simplicity, our algorithm can generate samples of continuous and attractive chains under regimes in which it was previously not known to be possible.

Finally, from a coding point of view, it is interesting to have a finitary coding from a finite alphabet i.i.d. process. We show that this is possible in the original Bramson and Kalikow [**2**] example if and only if there is a unique compatible chain, that is, choosing the parameters of the model in such a way that uniqueness holds.

It is worth mentioning that Steif and van den Berg [**29**] obtained similar results for a certain class of d-dimensional probabilistic cellular automata. Specifically, they proved that if a random field is obtained as invariant measures of monotonic, exponentially ergodic probabilistic cellular automata, then the random field is a finitary coding of a finite alphabet i.i.d. random field. They used this result to show that there exists a finitary coding from a finite alphabet i.i.d. process to the plus phase of a ferromagnetic Ising model strictly

https://doi.org/10.1017/etds.2013.7 Published online by Cambridge University Press

below the critical temperature. They also proved that for the plus phase of an Ising model strictly above the critical temperature, there is no finitary coding from a finite alphabet i.i.d. process. The situation for the critical temperature remains open. For the BK example, it is not clear what is the quantity equivalent to the statistical mechanics notion of temperature, and therefore we cannot make a one-to-one comparison with [**29**, Theorem 1.1]. Nevertheless, we observe that our result for the original BK example does not have *a priori* restriction on the parameters of the model or on the loss of memory of the process.

This paper is organized as follows. In §2 we introduce the notation, definitions, the necessary background, and two important examples. In §3 we state the main results. In §4 we introduce the attractive sampler, which is used to prove the theorems of §3. Finally, we prove the results in §5.

2. *Notation, standard definitions, and examples*

2.1. *Notation.* For any set $\mathcal{U}$ we denote the sets of bi-infinite, right-infinite and finite sequences of symbols of $\mathcal{U}$ by $\mathcal{U}^{-\mathbb{Z}} = \mathcal{U}^{\{\dots,2,1,0,-1,-2,\dots\}}$, $\mathcal{U}^{-\mathbb{N}} = \mathcal{U}^{\{-1,-2,\dots\}}$ and $\mathcal{U}^{\star} = \bigcup_{j\geq 1} \mathcal{U}^{\{-1,\dots,-j\}}$, respectively. The elements of these sets will be denoted, respectively, $\mathbf{u} = \cdots u_2u_1u_0u_{-1}u_{-2}\cdots$, $\underline{u} = u_{-1}u_{-2}\cdots$ and $u_{-k}^{-1} = u_{-1}u_{-2}\cdots u_{-k}$ for any $1 \leq k < +\infty$. We also use the notation $\mathcal{U}^{-n} = \mathcal{U}^{\{-1,\dots,-n\}}$ and $\mathcal{U}^{n} = \mathcal{U}^{\{n,\dots,1\}}$ for any $n \geq 1$. Finally, if $u \in \mathcal{U}^n$ and $v \in \mathcal{U}^m$ we will use the notation uv for the string of length $n+m$ belonging to $\mathcal{U}^{n+m}$, obtained by concatenating both strings. This notation is extended to the case where one of the strings is a right-infinite sequence, for example $v\underline{u}$. Note that we are using the convention that the past (negative indices) is on the right and the future (positive indices) is on the left.

In the present paper $\mathcal{U}$ is some Polish space and A is the finite ordered set $\{1, 2, \dots, s\}$ unless specified. A is called the *alphabet*. We define a partial order on $A^{-\mathbb{N}}$ by saying that $\underline{a} \leq \underline{b}$ whenever $a_{-i} \leq b_{-i}$ for every $i \geq 1$. In $A^{-\mathbb{N}}$, the maximal element is $\underline{s} := sss\cdots$ and the minimal element is $\underline{1} := 111\cdots$.

2.2. *Chains of infinite order.* A probability kernel, or simply a *kernel*, P on the alphabet A is a function

$$\begin{aligned} P: \quad A \times A^{-\mathbb{N}} &\to [0,1] \\ (a, \underline{x}) &\mapsto P(a|\underline{x}) \end{aligned}$$

such that

$$\sum_{a\in A} P(a|\underline{x}) = 1 \quad \text{for all } \underline{x} \in A^{-\mathbb{N}}.$$

We say that a stationary stochastic chain $\mathbf{X} = \{X_j\}_{j\in\mathbb{Z}}$ (of stationary law μ) on A is *compatible* with a kernel P if the latter is a regular version of the conditional probabilities of the former, that is,

$$\mu(X_0 = a | X_{-\infty}^{-1} = \underline{x}) = P(a|\underline{x})$$

for every $a \in A$ and μ-almost every $\underline{x}$ in $A^{-\mathbb{N}}$. When there is more than one stationary chain compatible with P, we say that there is *phase transition*, otherwise we say that the chain is *unique*. We follow the Harris nomenclature [**13**] and call the chains compatible

https://doi.org/10.1017/etds.2013.7 Published online by Cambridge University Press

with kernels *chains of infinite order*. These were first introduced in [**23**] under the name *chaînes à liaisons complètes*. The existence of an invariant measure for these chains was first studied by Doeblin and Fortet [**7**]. In ergodic theory, it was rediscovered by Keane [**19**] who gave the name g-function to the kernel and g-measure to the chain compatible with the kernel. For a comprehensive historical account and recent developments we refer the reader to [**9**].

2.3. *Non-nullness, continuity rate, oscillations and attractiveness.* We say that a kernel P is *strongly non-null* if

$$\inf_{a\in A,\underline{x}\in A^{-\mathbb{N}}} P(a|\underline{x}) > 0.$$

The *continuity rate* (or *variation*) *of order k* of a kernel P is

$$\mathrm{var}_k := \sup_{b\in A} \sup_{a_{-k}^{-1}\in A^{-k}} \sup_{\underline{x},\underline{y}\in A^{-\mathbb{N}}} |P(b|a_{-k}^{-1}\underline{x}) - P(b|a_{-k}^{-1}\underline{y})|.$$

We say that P is continuous if $\lim_{k\to\infty} \mathrm{var}_k = 0$. A compactness argument shows that if the kernel is continuous, at least one compatible stationary chain exists (see, for example, [**19**]). If P is strongly non-null and continuous, we say that P is a *regular kernel*.

Another characterization of kernels is given by the *oscillation rate*

$$\mathrm{osc}_n := \sum_{a\in A} \mathrm{osc}_n(a),$$

where

$$\mathrm{osc}_n(a) := \sup\{|P(a|\underline{x}) - P(a|\underline{y})| : \underline{x}, \underline{y} \in A^{-\mathbb{N}}, x_{-i} = y_{-i}\ \forall\, i \neq n\}.$$

The sequences $\{\mathrm{var}_k\}_{k\geq 0}$ and $\{\mathrm{osc}_k\}_{k\geq 0}$ are related to the uniqueness of the compatible stationary chain, as we will see in the examples below.

Finally, we say that a kernel P on A is *attractive* if for all $a \in A$ the value of $\sum_{j\geq a} P(j|\underline{x})$ is increasing on $\underline{x} \in A^{-\mathbb{N}}$.

Let us give two important examples taken from the literature, which we will repeatedly use in the following to illustrate our assertions.

2.4. *Binary autoregressive models.* These models are extensively used in the statistical literature [**22**], and are defined through the following parameters: a continuously differentiable and increasing function $\psi : \mathbb{R} \to]0, 1[$ such that $\psi(t) + \psi(-t) = 1$, a summable sequence of non-negative real numbers $\{\xi_n\}_{n\geq 1}$, and a non-negative real parameter $\gamma \geq 0$. Consider the class of kernels P on the alphabet $\{-1, +1\}$ such that

$$P(a|\underline{x}) := \psi\left(a \sum_{n\geq 1} \xi_n x_{-n} + a\gamma\right).$$

A straightforward computation shows that these kernels are attractive and regular. Moreover, if ψ is Lipschitz continuous, then one immediately obtains $\mathrm{var}_k \leq C \sum_{n>k} \xi_n$ and $\mathrm{osc}_n \leq C\xi_n$ for some positive constant C. In this case, possible criteria for uniqueness are $\xi_n = cn^{-\alpha}$ with $\alpha > 3/2$ and c any constant [**16**], or $\mathrm{osc}_n \leq C \sum_{n\geq 1} \xi_n < 1$ [**9**].

https://doi.org/10.1017/etds.2013.7 Published online by Cambridge University Press

An important example of binary autoregressive models is when $\psi(t) = e^{-t}(e^{-t} + e^{t})$. The resulting kernel is called the *logit model* in the statistics literature, and the *one-sided one-dimensional long-range Ising model* in the statistical physics literature. For instance, Hulse [**15**] used this model to give an example of phase transition in chains of infinite order.

2.5. *The Bramson and Kalikow example* [**2**]. Consider an increasing function $\phi : [-1, +1] \to]0, 1[$ such that $\phi(t) + \phi(-t) = 1$, an increasing sequence of odd positive integers $\{m_j\}_{j\geq 1}$ and a sequence $\{\lambda_j\}_{j\geq 1}$ such that $\lambda_j \geq 0$ and $\sum_{j\geq 1} \lambda_j = 1$. We call the class of kernels defined on the alphabet $A = \{-1, +1\}$ by

$$P(+1|\underline{x}) = \sum_{j\geq 1} \lambda_j \phi\left(\frac{1}{m_j} \sum_{i=1}^{m_j} x_{-i}\right)$$

the *BK example* [**10**]. The kernels of this class are attractive and regular. Attractiveness and strong non-nullness follow directly from the definition of ϕ and simple calculations yield $\mathrm{var}_n \leq \sum_{\{j:m_j>n\}} \lambda_j$, showing that P is continuous. When $\phi(t) = (1-\epsilon)\mathbf{1}\{t > 0\} + \epsilon\mathbf{1}\{t < 0\}$ for some $\epsilon \in (0, 1/2)$ and $\lambda_k = cr^k$ for some $r \in (2/3, 1)$ and $c = (1-r)/r$, we call this model the *original BK example*, as it is precisely the model introduced in [**2**], where it is proved that, taking the sequence $\{m_k\}_{k\geq 1}$ increasing sufficiently fast, the kernel P exhibits phase transition.

2.6. *Maximum and minimum phases for attractive kernels.* Define, for any $\underline{x} \in A^{-\mathbb{N}}$, the fixed past chain $\mathbf{X}^{\underline{x}}$ by

$$X_n^{\underline{x}} = \begin{cases} x_n & \text{if } n \leq 0, \\ a \text{ with probability } P(a|X_{n-1}^{\underline{x}} \dots X_1^{\underline{x}}\underline{x}) & \text{otherwise.} \end{cases}$$

$\mathbf{X}^{\underline{x}}$ is the non-stationary chain obtained by fixing the past $\underline{x}$ from time 0 to $-\infty$, and 'running P from this past'. For any $\underline{x}$, $j \geq 1$, and $n \in \mathbb{Z}$, let $\mathbf{X}^{\underline{x},j}$ be the process defined by $X_n^{\underline{x},j} := X_{n+j}^{\underline{x}}$. The attractiveness of P implies that for $\underline{x} = \underline{1}$ and $\underline{s}$, the sequence of processes $\{\mathbf{X}^{\underline{x},j}\}_{j\geq 1}$ is stochastically non-decreasing and non-increasing, respectively [**14**], and therefore, the weak limits

$$\mathbf{X}^{\min} := \lim_{j\to+\infty} \mathbf{X}^{\underline{1},j} \quad \text{and} \quad \mathbf{X}^{\max} := \lim_{j\to+\infty} \mathbf{X}^{\underline{s},j}$$

exist and are stationary. If P is continuous, then $\mathbf{X}^{\min}$ and $\mathbf{X}^{\max}$ are compatible with P. We call $\mathbf{X}^{\min}$ the *minimum phase* and $\mathbf{X}^{\max}$ the *maximum phase*.

2.7. *Finitary process and B-process.* Let $T_{\mathcal{U}}$ and T_A be the shift operators that act respectively on $\mathcal{U}^{\mathbb{Z}}$ and $A^{\mathbb{Z}}$ shifting the sequences by one unit: $T_{\mathcal{U}}(\mathbf{u}) = \{u_{i+1}\}_{i\in\mathbb{Z}}$ and $T_A(\mathbf{a}) = \{a_{i+1}\}_{i\in\mathbb{Z}}$. A stationary process $\mathbf{X}$ (with stationary law μ) on the alphabet A is a *stationary coding* of a stationary process $\mathbf{U}$ (with stationary law $\mathbb{P}$) on $\mathcal{U}$ if there exists a measurable function $\Phi : \mathcal{U}^{\mathbb{Z}} \to A^{\mathbb{Z}}$ which is *translation equivariant* (that is $\Phi(T_{\mathcal{U}}(\mathbf{U})) = T_A\Phi(\mathbf{U})$) and such that $\mu = \mathbb{P} \circ \Phi^{-1}$. We follow the nomenclature given in [**28**] and call a process that is a stationary coding of an i.i.d. process a *B-process*. A stationary

https://doi.org/10.1017/etds.2013.7 Published online by Cambridge University Press

coding is called a *finitary coding* if there exist *stopping times* $\theta_1 : \mathcal{U}^{\mathbb{Z}} \to \mathbb{N} \cup \{\infty\}$ and $\theta_2 : \mathcal{U}^{\mathbb{Z}} \to \mathbb{N} \cup \{\infty\}$, both $\mathbb{P}$-almost surely (a.s.) finite, such that

$$[\Phi(\mathbf{U})]_0 = [\Phi(\mathbf{V})]_0 \quad \text{whenever } U_{-\theta_1(\mathbf{U})}^{+\theta_2(\mathbf{U})} = V_{-\theta_1(\mathbf{U})}^{+\theta_2(\mathbf{U})}. \tag{1}$$

This last assumption means that the event $\{\theta_1(\mathbf{U}) = k\} \cap \{\theta_2(\mathbf{U}) = l\}$ is $\mathcal{F}(U_{-k}^{+l})$-measurable. In other words, the stopping times $\theta_1(\mathbf{U})$, $\theta_2(\mathbf{U})$ are checkable by looking only at an a.s. finite number of U_i. We call a process that is a finitary coding of an i.i.d. process a *finitary process* (FP). We will often use the simplified notation θ_1 and θ_2 respectively for $\theta_1(\mathbf{U})$ and $\theta_2(\mathbf{U})$.

The notion of stationary coding comes from ergodic theory, and has a one-sided analogue in the literature of stochastic processes, called the *coupling-from-the past* algorithm. Such algorithms, which were first introduced in [**25**] for Markov chains, aim to construct the function Φ using the kernel P and an a.s. finite number of past values of an i.i.d. process $\mathbf{U}$. If a CFTP algorithm is *feasible* for a given kernel P, then the constructed stationary measure μ is an FP, because it is a particular finitary coding of $\mathbf{U}$, for which we can take $\theta_2 = 0$ and θ_1 finite $\mathbb{P}$-a.s. For simplicity, in this case, we use the notation $\theta := \theta_1$ and say that θ is the stopping time of the CFTP algorithm.

In the literature, sometimes a process is called finitary only if the set $\mathcal{U}$ is finite. We do not assume this. In the special case of $\mathcal{U}$ being finite (or countable), we say that the process is a *finitary coding of a finite (countable) alphabet i.i.d. process*.

2.8. *Exponential concentration of measure.* Let $f : A^n \to \mathbb{R}$ be measurable. Define

$$\delta_j f = \sup\{|f(a_1^n) - f(b_1^n)| : a_i = b_i, \forall i \neq j\}$$

and let δf be the vector with jth coordinate given by $\delta_j f$. We say that the *concentration of measure holds at exponential rate* for a stationary process $\mathbf{X}$ if, for all $n > 0$, $\epsilon > 0$, and functions f,

$$\mathbb{P}(|f(X_1^n) - \mathbb{E}[f(X_1^n)]| > \epsilon) \leq C \exp\left\{-\frac{g(\epsilon, \|\delta f\|_{\ell_1(\mathbb{N})})}{\|\delta f\|_{\ell_2(\mathbb{N})}^2}\right\}, \tag{2}$$

where $g(\epsilon, \|\delta f\|_{\ell_1(\mathbb{N})}) > 0$ and C is a numerical constant.

In particular, inequality (2) implies that for all $k \in \mathbb{N}$, $n \geq k$, and $h : A^k \to [0, 1]$,

$$\mathbb{P}\left(\left|\frac{1}{n-k+1}\sum_{j=0}^{n-k} h(X_{j+1}^{j+k}) - \mathbb{E}[h(X_1^k)]\right| > \epsilon\right) \leq Ce^{-ng_k(\epsilon)}, \tag{3}$$

where $g_k(\epsilon) > 0$ and C is a numerical constant. We say that the *ergodic theorem holds at exponential rate* for a stationary process $\mathbf{X}$ if it satisfies (3).

3. *Main results*

THEOREM 1. *Let P be an attractive regular kernel. The following are equivalent:*

(1) *there exists a unique stationary chain compatible with P;*
(2) $\mathbf{X}^{\max}$ *is a finitary coding of a countable alphabet i.i.d. process;*
(3) *the concentration of measure holds at exponential rate for* $\mathbf{X}^{\max}$.

https://doi.org/10.1017/etds.2013.7 Published online by Cambridge University Press

From [**14**] we know that the maximum phase (respectively, the minimum phase) is always a B-process regardless of being equal or different to the minimum phase (respectively, the maximum phase). Therefore, the fact of being a B-process does not distinguish the presence or not of phase transition. Theorem 1 shows that, for a regular attractive kernel P, being a finitary coding of an i.i.d. process or having concentration of measure at exponential rate is what distinguishes between existence or not of phase transition.

We now show that Theorem 1 is optimal in the class of attractive chains, in the sense that if we relax either continuity or strong non-nullness, we can find examples of stationary chains that are FPs and with the ergodic theorem holding at exponential rate, although they are not uniquely determined by their conditional probabilities.

3.1. *Relaxing the strong non-nullness assumption.* The following example shows that in general we cannot relax the strong non-nullness condition. Before giving our example, we need some more definitions. Let $A = \{-1, +1\}$. For any $\underline{x} \in A^{-\mathbb{N}}$, we define

$$\ell(\underline{x}) := \min\{j \geq 0 : x_{-j-1} = -1\}.$$

When we look into the past in $\underline{x}$, $\ell(\underline{x})$ counts the number of $+1$ before finding the first -1. Let $+\underline{1}$ be the pasts such that $+\underline{1}_j = +1$ for all j. We use the convention that $\ell(+\underline{1}) = \infty$. Let $\{p_i\}_{i\geq 0}$ be a monotonically decreasing sequence of [0, 1]-valued real numbers, and let $p_\infty = \lim_{i\to+\infty} p_i$. The kernel P is defined on $\{-1, +1\}$ by $P(-1|\underline{x}) = p_{\ell(\underline{x})}$ for any $\underline{x} \neq +\underline{1}$ and $P(-1|+\underline{1}) := p_\infty$. It is clear that this example is attractive. It is also continuous. To see this, observe that

$$\sup_{\underline{y},\underline{z}\in A^{-\mathbb{N}}} |P(a|a_{-k}^{-1}\underline{y}) - P(a|a_{-k}^{-1}\underline{z})| = 0$$

for any $a_{-k}^{-1} \in A^k$, except for $(+1)^k$. Hence,

$$\sup_{a_{-k}^{-1}\in A^k} \sup_{\underline{y},\underline{z}\in A^{-\mathbb{N}}} |P(a|a_{-k}^{-1}\underline{y}) - P(a|a_{-k}^{-1}\underline{z})| = \sup_{\underline{y},\underline{z}\in A^{-\mathbb{N}}} |P(a|(+1)^k\underline{y}) - P(a|(+1)^k\underline{z})|,$$

and thus we obtain that

$$\sup_{a_{-k}^{-1}\in A^k} \sup_{\underline{y},\underline{z}\in A^{-\mathbb{N}}} |P(a|a_{-k}^{-1}\underline{y}) - P(a|a_{-k}^{-1}\underline{z})| = p_k - p_\infty,$$

which goes to 0 by the definition of the sequence $\{p_k\}_{k\geq 0}$. If $p_\infty = 0$, the chain is not strongly non-null, and the degenerate chain with all symbols equal to $+1$ is trivially stationary and compatible with P. If we further assume that $\sum_{k\geq 1}\prod_{i=0}^{k-1}(1 - p_i) < +\infty$, there exists another class of stationary chains compatible with the kernel P. These are the so-called *renewal chains*, obtained by concatenating i.i.d. blocks of the form $(-1, +1, \ldots, +1, +1)$ of random length. These blocks have length $k + 1$ with probability $\prod_{i=0}^{k-1}(1 - p_i)p_k$, and therefore have finite expected length. The existence of several compatible chains is due to the fact that this kernel is *not irreducible* [**3**]. Therefore, this chain has a degenerate type of phase transition. Nevertheless, the compatible chain with probability one of having $+1$ is obviously a finitary coding of an i.i.d. process with

https://doi.org/10.1017/etds.2013.7 Published online by Cambridge University Press

a concentration of measure at exponential rate. This shows that if the strong non-nullness assumption is removed, the existence of a finitary coding for the maximum phase does not imply uniqueness of the compatible chain.

3.2. *Relaxing the continuity assumption.* For a discontinuous attractive kernel P, the maximum and minimum phases are always distinct and not consistent with the kernel P [**14**]. Hence, strictly speaking, we do not have a phase transition where there is more than one chain compatible with P. In fact, this means that considering discontinuous attractive chains does not make much sense from the point of view of non-uniqueness. Nevertheless, we can still ask if the maximum and minimum phases of a discontinuous attractive kernel can be finitary codings of i.i.d. processes. The example below shows that, in general, this could happen.

Let $\underline{a} \in \{-1, 1\}^{-\mathbb{N}}$ and let $-\underline{1}$ be the pasts such that $-\underline{1}_j = -1$ for all j. More generally, for any $\underline{z} \in \{-1, 1\}^{-\mathbb{N}}$ (respectively, $u \in \{-1, 1\}^\star$) we denote by $-\underline{z}$ (respectively, $-u$) the sequence of symbols obtained by switching the signal of each coordinate. Let $\{\epsilon_n\}_{n\in\mathbb{N}}$ be a non-increasing sequence of positive numbers such that $\lim_{n\to\infty} \epsilon_n = \epsilon$ with $0 < \epsilon < 1/2$. We define P by

$$P(1|a_{-n}^{-1}(+\underline{1})) = 1 - \epsilon_n = P(-1|-a_{-n}^{-1}(-\underline{1})),$$
$$P(1|a_{-n}^{-1}(-\underline{1})) = \epsilon_n = P(-1|-a_{-n}^{-1}(+\underline{1})),$$

and put $P(1|\underline{x}) = 1/2$ for all the remaining pasts $\underline{x}$. Clearly P is strongly non-null, attractive, and non-continuous. Also, let

$$P_+(1|\underline{a}) = \lim_{n\to\infty} P(1|a_{-n}^{-1}(+\underline{1})) = 1 - \epsilon,$$
$$P_-(-1|\underline{a}) = \lim_{n\to\infty} P(-1|a_{-n}^{-1}(-\underline{1})) = 1 - \epsilon.$$

By [**14**, Lemma 2.3], the maximum phase $\mathbf{X}^{(+\underline{1})}$ is consistent with P_+ and therefore it is the i.i.d. process with probability $1 - \epsilon$ for 1. Analogously, the minus phase $\mathbf{X}^{(-\underline{1})}$ is the i.i.d. process with probability $1 - \epsilon$ for -1.

Theorem 1 is a direct consequence of Theorems 2–4 below.

THEOREM 2. *Let P be an attractive continuous kernel. If there exists a unique stationary chain compatible with P then there exists a feasible CFTP algorithm using a countable alphabet i.i.d. process that simulates this chain.*

Observe that for this theorem we do not require strong non-nullness of the kernel. As an immediate application of Theorem 2, our *attractive sampler* given in §4 perfectly simulates binary autoregressive and BK processes introduced in §2 in their uniqueness regime. Notice that for the binary autoregressive and BK processes, for any $\eta > 0$, we can exhibit kernels having continuity rate $\mathrm{var}_k = O(1/k^\eta)$ for which the unique compatible stationary chain can be perfectly simulated. In particular, in the BK example var_k can be taken so that it converges arbitrarily slowly to 0. As a comparison, in the work of Comets *et al* [**5**] the condition $\sum_{k\geq 0} \prod_{i=0}^{k-1} (1 - \mathrm{var}_k) = +\infty$ is assumed to guarantee that their CFTP algorithm is feasible, that is, the stopping time is a.s. finite. This condition does not

https://doi.org/10.1017/etds.2013.7 Published online by Cambridge University Press

hold if $\mathrm{var}_k = O(1/k^\eta)$ with sufficiently small η. In other words, in the class of regular attractive chains, our perfect simulation algorithm (attractive sampler) is optimal. This is particularly clear for the binary autoregressive model with $\phi(r) = (1+r)/2$, which is also considered in [**5**]. In this case, when $\sum_{n\geq 1} \xi_n + \gamma < 1$, the criterion of Fernández and Maillard [**9**] implies uniqueness, and therefore, our attractive sampler works whereas the algorithm in Comets *et al* [**5**] is not guaranteed to work in general.

The following theorem relates the stopping times of an FP to its concentration property.

THEOREM 3. *If* $\mathbf{X}$ *is an FP then the concentration of measure holds at exponential rate for* $\mathbf{X}$. *Explicitly, using the same notation as in (1) and (2), if* θ_1 *and* θ_2 *are the stopping times of the FP and* r_1 *and* r_2 *are positive numbers such that*

$$\mathbb{P}(\{\theta_1 > r_1\} \cup \{\theta_2 > r_2\}) \leq \epsilon/(6\|\delta f\|_{\ell_1(\mathbb{N})}),$$

then

$$\mathbb{P}(|f(X_1^n) - \mathbb{E}[f(X_1^n)]| > \epsilon) \leq 4\exp\left\{-\frac{2\epsilon^2}{9(1+r_2+\sum_{j=1}^{r_1}\mathbb{P}(\theta_1 \geq j))^2\|\delta f\|^2_{\ell_2(\mathbb{N})}}\right\}.$$

The above theorem holds for any FP (we assume neither regularity nor attractiveness) and it is of independent interest. We note that, instead of Theorem 3, the blowing-up property of FPs proved in [**21**] could be used together with our Theorems 2 and 4 to prove the equivalence between the uniqueness of a compatible chain and the existence of a finitary coding from a countable alphabet i.i.d. process to the compatible chain. This is because in [**21**] it is proved that if a process is FP then it satisfies the blowing-up property, which implies that the ergodic theorem holds at exponential rate. Nevertheless, we think that Theorem 3 gives us more explicit information about the process relating the stopping time to the concentration of measure. For instance, if a process can be sampled using a CFTP algorithm and the stopping time has finite expectation, we have the following useful corollary, which is obtained simply taking $r_1 = \infty$ and $r_2 = 0$.

COROLLARY 1. *Let* $\mathbf{X}$ *be a process that can be simulated by a CFTP algorithm with a stopping time* θ. *If* $\mathbb{E}[\theta] < \infty$, *then for all* $\epsilon > 0$ *and all functions* $f : A^n \to \mathbb{R}$,

$$\mathbb{P}(|f(X_1^n) - \mathbb{E}[f(X_1^n)]| > \epsilon) \leq 4\exp\left\{-\frac{2\epsilon^2}{9(1+\mathbb{E}[\theta])^2\|\delta f\|^2_{\ell_2(\mathbb{N})}}\right\}. \tag{4}$$

As an example of application of the above result, if a probability kernel has summable continuity rate var_k, we can construct a CFTP algorithm with stopping time θ such that $\mathbb{E}[\theta] \leq C\sum_{k=1}^{\infty}\mathrm{var}_k$, where C is a numerical constant [**5**]. We refer the reader to [**6, 11, 12**] for bounds on probability of θ for different conditions and not necessarily regular kernels.

We now come to the last ingredient for the proof of Theorem 1.

THEOREM 4. *Let* P *be a regular kernel and* $\mathbf{X}$ *a process compatible with* P *that satisfies the concentration of measure at exponential rate. Then* $\mathbf{X}$ *is the unique stationary process compatible with* P.

https://doi.org/10.1017/etds.2013.7 Published online by Cambridge University Press

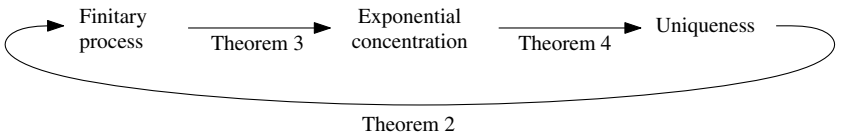


FIGURE 1: The chain of implications stated by Theorem 1.

Note that, for this result, we do not assume that the kernel is attractive, and therefore Theorem 4 constitutes an interesting characterization of uniqueness for chains of infinite order. We cannot, in general, relax the strong non-nullness condition in this theorem, because by the example given just after Theorem 1, where the kernel has only one null transition probability, there exists a chain that satisfies the ergodic theorem at exponential rate but it is not the unique chain compatible with the kernel.

Now, Theorem 1 follows from the sequence of implications shown in Figure 1 that holds when the kernel is attractive and regular.

From the coding point of view, it is natural to ask if Theorem 1 can be strengthened to a finitary coding from a finite alphabet i.i.d. process. This is indeed the case for the original BK example.

THEOREM 5. *Let P be the original BK example. Then there exists a unique chain compatible with P if and only if the compatible chain is a finitary coding of a finite alphabet i.i.d. process.*

4. *The attractive sampler*

Assume that we are given an attractive continuous kernel P for which there exists a unique compatible stationary chain. The alphabet is $A = \{1, \dots, s\}$. As stated in Theorem 2, there exists a CFTP algorithm that samples from its stationary law. Here we construct one such CFTP algorithm and call it the *attractive sampler*. First, let us consider the kernel $\tilde{P}$ on $A \times A$ introduced in [**14**] and defined by

$$\tilde{P}((x_0 \geq a,\, y_0 \geq b)|(\underline{x},\, \underline{y})) := \sum_{i=a}^{s} P(i|\underline{x}) \wedge \sum_{i=b}^{s} P(i|\underline{y}), \tag{5}$$

for any pair of pasts $\underline{x}$ and $\underline{y}$ in $A^{-\mathbb{N}}$ and any pair of symbols a and b in A. Note that, for any a, b in A,

$$\begin{aligned}\tilde{P}((a,\, b)|(\underline{x},\, \underline{y})) &= \tilde{P}((x_0 = a,\, y_0 = b)|(\underline{x},\, \underline{y})) \\ &= \tilde{P}((x_0 \geq a,\, y_0 \geq b)|(\underline{x},\, \underline{y})) - \tilde{P}((x_0 \geq a+1,\, y_0 \geq b)|(\underline{x},\, \underline{y})) \\ &\quad - \tilde{P}((x_0 \geq a,\, y_0 \geq b+1)|(\underline{x},\, \underline{y})) \\ &\quad + \tilde{P}((x_0 \geq a+1,\, y_0 \geq b+1)|(\underline{x},\, \underline{y})).\end{aligned}$$

This kernel defines a coupling between the kernel $\{P(a|\underline{x})\}_{a\in A}$ and $\{P(a|\underline{y})\}_{a\in A}$. To see this, first observe that

$$\sum_{a=1}^{s} \tilde{P}((a,\, s)|(\underline{x},\, \underline{y})) = \tilde{P}((a' \geq 1,\, b' \geq s)|(\underline{x},\, \underline{y})) = P(s|\underline{y}),$$

https://doi.org/10.1017/etds.2013.7 Published online by Cambridge University Press

and then observe that

$$\begin{aligned}\sum_{a=1}^{s} \tilde{P}((a, s-1)|(\underline{x}, \underline{y})) &= \tilde{P}((a' \geq 1, b' \geq s-1)|(\underline{x}, \underline{y})) - \tilde{P}((a' \geq 1, b' \geq s)|(\underline{x}, \underline{y})) \\ &= [P(s-1|\underline{y}) + P(s|\underline{y})] - P(s|\underline{y}) \\ &= P(s-1|\underline{y}).\end{aligned}$$

Now, continuing the recursion, we show that $\sum_{a=1}^{s} \tilde{P}((a, b)|(\underline{x}, \underline{y})) = P(b|\underline{y})$ for any $b \in A$. The same holds for the sum over b, that is, $\sum_{b=1}^{s} \tilde{P}((a, b)|(\underline{x}, \underline{y})) = P(a|\underline{y})$ for any $a \in A$. This means that $\tilde{P}$ defines a coupling between the chains with respective fixed pasts.

A straightforward but tedious computation shows that $\tilde{P}$ is indeed continuous and we can use the result of Kalikow [**17**] stating that continuous kernels can be written as a countable mixture of Markov kernels of increasing order. Formally, for $\tilde{P}$, there exists a sequence of non-negative numbers $\{\lambda_k\}_{k\geq 0}$ with $\sum_{k\geq 0} \lambda_k = 1$, a sequence of Markov kernels $\{P^{[k]}\}_{k\geq 1}$, where $P^{[k]}$ is a k-step Markov kernel, and a probability distribution $P^{[0]}$ on $A \times A$ such that

$$\tilde{P}((a, b)|(\underline{x}, \underline{y})) = \lambda_0 P^{[0]}((a, b)) + \sum_{k\geq 1} \lambda_k P^{[k]}((a, b)|(x_{-k}^{-1}, y_{-k}^{-1})).$$

We will now use this representation of $\tilde{P}$ to define our algorithm. First, we introduce a series of partitions of $[0, 1]$ that will be used to define the finitary coding.

For any $k \in \mathbb{N}$, $(x_{-k}^{-1}, y_{-k}^{-1}) \in A^{-k} \times A^{-k}$ and $a, b \in \{1, \ldots, s\}$, define

$$\begin{aligned}r_{a,b}^{[k]}(x_{-k}^{-1}, y_{-k}^{-1}) = &\sum_{c\leq a-1} \sum_{d\leq s} P^{[k]}((x_0 = c, y_0 = d)|(x_{-k}^{-1}, y_{-k}^{-1})) \\ &+ \sum_{d\leq b} P^{[k]}((x_0 = a, y_0 = d)|(x_{-k}^{-1}, y_{-k}^{-1}))\end{aligned}$$

and

$$r_{a,b}^{[0]} = \sum_{c\leq a-1} \sum_{d\leq s} P^{[0]}((x_0 = c, y_0 = d)) + \sum_{d\leq b} P^{[0]}((x_0 = a, y_0 = d)).$$

Let $r_{1,0}^{[k]} := 0$ and, for $a \in \{2, \ldots, s\}$, $r_{a,0}^{[k]} := r_{a-1,s}^{[k]}$. Notice that

$$r_{a,b}^{[k]}(x_{-k}^{-1}, y_{-k}^{-1}) - r_{a,b-1}^{[k]}(x_{-k}^{-1}, y_{-k}^{-1}) = P^{[k]}((x_0 = a, y_0 = b)|(x_{-k}^{-1}, y_{-k}^{-1})).$$

For any integer $k \geq 0$, we set

$$\mathcal{R}^{[k]} = \{r_{a,b}^{[k]}(x_{-k}^{-1}, y_{-k}^{-1}) \in [0, 1] : (a, b) \in A^2, (x_{-k}^{-1}, y_{-k}^{-1}) \in A^{-k} \times A^{-k}\}. \tag{6}$$

Denote the elements of $\mathcal{R}^{[k]}$ by $t_1^{[k]} < t_2^{[k]} < \cdots < t_{|\mathcal{R}^{[k]}|}^{[k]} = 1$ and define $t_0^{[k]} := 0$. Now, for all $k \in \mathbb{N}$, define the following set of intervals:

$$\mathcal{I}^{[k]} = \left\{ \left[\sum_{i=0}^{k-1} \lambda_i + \lambda_k t_{j-1}^{[k]}, \sum_{i=0}^{k-1} \lambda_i + \lambda_k t_j^{[k]} \right[: j \in \{1, \ldots, |\mathcal{R}^{[k]}|\} \right\}$$

and

$$\mathcal{I}^{[0]} = \{[\lambda_0 t_{j-1}^{[0]}, \lambda_0 t_j^{[0]}[: j \in \{1, \ldots, |\mathcal{R}^{[0]}|\}\}.$$

https://doi.org/10.1017/etds.2013.7 Published online by Cambridge University Press

Finally, let $\mathcal{I} = \bigcup_{k\geq 0} \mathcal{I}^{[k]}$. Observe that the set $\mathcal{I}$ has a countable number of elements; therefore we can identify its elements by I_j, $j \in \mathbb{N}$. We define the *update* function $F : A^{-\mathbb{N}} \times A^{-\mathbb{N}} \times \mathbb{N} \to A \times A$ by

$$F((\underline{x}, \underline{y}), j) = \sum_{(a,b)\in A^2} (a, b)\mathbf{1}\left\{ I_j \subset \bigcup_{k\in\mathbb{N}} [r^{[k]}_{a,b-1}(x^{-1}_{-k}, y^{-1}_{-k}), r^{[k]}_{a,b}(x^{-1}_{-k}, y^{-1}_{-k})[\cup [r^{[0]}_{a,b-1}, r^{[0]}_{a,b}[\right\}.$$

Now, let $\mathbf{L} = \{L_j\}_{j\in\mathbb{Z}}$ be an i.i.d. process with values on $\mathbb{N}$ such that

$$\mathbb{P}(L_0 = j) = \sup I_j - \inf I_j. \tag{7}$$

We observe by construction that

$$\mathbb{P}(F((\underline{x}, \underline{y}), L_0) = (a, b)) = \tilde{P}((a, b)|(x, y)),$$

which justifies the name 'update function' for F.

Let the concatenation of pairs of symbols be understood coordinatewise, that is, $(a, b)(c, d) = (ac, bd)$ whenever (ac, bd) is well defined. Using this notation, we define, for any $-\infty < k \leq l < +\infty$, the successive iterations of F as

$$F_{[k,l]}((\underline{x}, \underline{y}), L^l_k) = F(F_{[k,l-1]}((\underline{x}, \underline{y}), L^{l-1}_k) \cdots F_{[k,k]}((\underline{x}, \underline{y}), L_k)(\underline{x}, \underline{y}), L_l),$$

where $F_{[k,k]}((\underline{x}, \underline{y}), L_k) := F((\underline{x}, \underline{y}), L_k)$. Notice that for any $(\underline{x}, \underline{y})$, we can obtain a coupling $\{(X^{\underline{x}}_i, X^{\underline{y}}_i)\}_{i\geq 0}$ by observing that

$$F_{[0,i]}((\underline{x}, \underline{y}), L^i_0) \stackrel{d}{=} (X^{\underline{x}}_i, X^{\underline{y}}_i).$$

Furthermore, we define, for any $i \in \mathbb{Z}$, the random variable

$$\theta[i] := \min\{j \geq 0 : F_{[i-j,i]}((\underline{a}, \underline{b}), L^i_{i-j}) \in \{(1, 1), \ldots, (s, s)\} \text{ for all } (\underline{a}, \underline{b}) \in A^{-2\mathbb{N}}\}.$$

An important observation is that in the particular case of attractive chains,

$$\theta := \theta[0] = \min\{j \geq 0 : F_{[-j,0]}((\underline{1}, \underline{s}), L^0_{-j}) \in \{(1, 1), \ldots, (s, s)\}\}.$$

Finally, define the value of the *coding function* Φ at time $i \in \mathbb{Z}$ by

$$[\Phi(\mathbf{L})]_i = F_{[-\theta[i],i]}((\underline{1}, \underline{s}), L^i_{-\theta[i]}).$$

The attractive sampler algorithm is defined by the pseudocode on the next page.

Observe that definitions above of θ and Φ satisfy the requirements of a stationary coding from $\mathbf{L}$. In particular, it can be shown in a standard way (see, for example, [**25**] for the Markovian case, or [**5**] for chains of infinite order) that if θ is $\mathbb{P}$-a.s. finite (that is, the CFTP algorithm is feasible), then the coding is finitary and the sample $[\Phi(\mathbf{L})]_0$ is constructed according to the unique stationary measure compatible with P. The compatibility and the stationarity follow from the construction and the property of the update function. The uniqueness follows from the loss of memory the chain inherits because of the existence of a.s. finite stopping time $\theta[i]$ for any $i \in \mathbb{Z}$.

In the proof of Theorem 2 in §5.1 we will prove that for attractive continuous P, the attractive sampler is feasible if P has a unique compatible chain.

https://doi.org/10.1017/etds.2013.7 Published online by Cambridge University Press

Attractive Sampler

Input: F; *Output:* $\theta[0]$, $[\Phi(\mathbf{L})]_0$
Sample L_0 with distribution $\mathbb{P}$
$i \leftarrow 0$, $\theta[0] \leftarrow 0$, $[\Phi(\mathbf{L})]_0 \leftarrow 0$
while $F_{[-i,0]}((\underline{1}, \underline{s}), L^0_{-i}) \notin \{(1, 1), \ldots, (s, s)\}$ **do**
  $i \leftarrow i + 1$
  Sample L_{-i} with distribution $\mathbb{P}$
end while
$\theta[0] \leftarrow i$
$[\Phi(\mathbf{L})]_0 \leftarrow F_{[-i,0]}((\underline{1}, \underline{s}), L^0_{-i})$
return $\theta[0]$, $[\Phi(\mathbf{L})]_0$.

5. *Proof of the results*

From §3 it is clear that Theorem 1 follows directly from Theorems 2–4.

5.1. *Proof of Theorem 2.* First we need the following lemma proved for the regular and attractive kernels by Hulse [**14**]. Here, we drop the unnecessary non-nullness condition.

LEMMA 1. *Let P be attractive and continuous and consider the update function F defined in §4. If there exists a unique chain compatible with P, then for all $i \in A = \{2, \ldots, s\}$,*

$$\lim_{n\to\infty} \mathbb{P}(F_{[0,n]}((\underline{1}, \underline{s}), L^n_0) \geq (i, 1)) - \mathbb{P}(F_{[0,n]}((\underline{1}, \underline{s}), L^n_0) \geq (1, i)) = 0. \tag{8}$$

OBSERVATION 1. *Remember that $F_{[0,n]}((\underline{1}, \underline{s}), L^n_0) \stackrel{d}{=} (X^{\underline{1}}_n, X^{\underline{s}}_n)$, and therefore we have that*

$$\mathbb{P}(F_{[0,n]}((\underline{1}, \underline{s}), L^n_0) \geq (i, 1)), \quad i \in A,$$

gives the law of $X^{\underline{1}}_n$.

Proof. For $a, b, c, d \in A$, we write $(a, b) \geq (c, d)$ if $a \geq c$ and $b \geq d$.

Because P is attractive, for all $l_1, \ldots, l_k, k \in \mathbb{N}$, and $a_1, \ldots, a_k \in A$, we have that

$$\mathbb{P}(F_{[0,n+l_1]}((\underline{1}, \underline{s}), L^{n+l_1}_0) \geq (1, a_1), \ldots, F_{[0,n+l_k]}((\underline{1}, \underline{s}), L^{n+l_k}_0) \geq (1, a_k))$$

and

$$\mathbb{P}(F_{[0,n+l_1]}((\underline{1}, \underline{s}), L^{n+l_1}_0) \geq (a_1, 1), \ldots, F_{[0,n+l_k]}((\underline{1}, \underline{s}), L^{n+l_k}_0) \geq (a_k, 1))$$

are respectively non-increasing and non-decreasing in $n \in \mathbb{N}$. Therefore, both sequences are convergent in $n \in \mathbb{N}$ and their limits when n diverges define stationary chains. By construction, for all $b \in A$ and $\underline{a} \in A^{-\mathbb{N}}$ the respective chains are compatible with the kernels $P^{\max}$ and $P^{\min}$ given by

$$P^{\max}(b|\underline{a}) = \lim_{n\to\infty} P(b|a^{-1}_{-n}\underline{s}) \quad \text{and} \quad P^{\min}(b|\underline{a}) = \lim_{n\to\infty} P(b|a^{-1}_{-n}\underline{1}).$$

If the kernel P is continuous, $P^{\max} = P^{\min} = P$, and therefore both chains are compatible with P. This implies that if there exists only one chain compatible with P,

https://doi.org/10.1017/etds.2013.7 Published online by Cambridge University Press

then we have, for $i = 2, \ldots, s$,

$$\lim_{n\to\infty} \mathbb{P}(F_{[0,n]}((\underline{1}, \underline{s}), L_0^n) \geq (i, 1)) - \mathbb{P}(F_{[0,n]}((\underline{1}, \underline{s}), L_0^n) \geq (1, i)) = 0,$$

as we wanted. □

We now prove that, for attractive regular chains, convergence (8) implies that the unique stationary chain compatible with P can be sampled by the attractive sampler algorithm. In other words, we need to prove that uniqueness of compatible chain implies that $\theta[0]$ is $\mathbb{P}$-a.s. finite. We have

$$\mathbb{P}(\theta[0] > n) = \mathbb{P}\Bigg(\bigcap_{j=-n}^{0} \{F_{[j,0]}((\underline{1}, \underline{s}), L_j^0) \notin \{(1, 1), \ldots, (s, s)\}\}\Bigg),$$

which yields, using first the attractiveness and then the translation invariance of $\mathbf{L}$,

$$\begin{aligned}\mathbb{P}(\theta[0] > n) &= \mathbb{P}(F_{[-n,0]}((\underline{1}, \underline{s}), L_{-n}^0) \notin \{(1, 1), \ldots, (s, s)\}) \\ &= \mathbb{P}(F_{[0,n]}((\underline{1}, \underline{s}), L_0^n) \notin \{(1, 1), \ldots, (s, s)\}).\end{aligned}$$

Thus, we want to prove that

$$\lim_{n\to+\infty} \mathbb{P}(F_{[0,n]}((\underline{1}, \underline{s}), L_0^n) \notin \{(1, 1), \ldots, (s, s)\}) = 0.$$

Due to the attractiveness, our coupling guarantees that for $a \in A$,

$$\mathbb{P}(F_{[0,n]}((\underline{1}, \underline{s}), L_0^n) \geq (a, a)) = \mathbb{P}(F_{[0,n]}((\underline{1}, \underline{s}), L_0^n) \geq (a, 1)).$$

Taking the limit and using (8), we have for any $a \in A$,

$$\begin{aligned}\lim_{n\to+\infty} \mathbb{P}(F_{[0,n]}((\underline{1}, \underline{s}), L_0^n) \geq (a, a)) &= \lim_{n\to+\infty} \mathbb{P}(F_{[0,n]}((\underline{1}, \underline{s}), L_0^n) \geq (a, 1)) \\ &= \lim_{n\to+\infty} \mathbb{P}(F_{[0,n]}((\underline{1}, \underline{s}), L_0^n) \geq (1, a)).\end{aligned}$$

Now, for any $a \in A$, let $\Gamma(a) = \{(i, j) \in A^2 : i \geq a \text{ and } j < a\}$. From the last equation, we have that

$$\lim_{n\to+\infty} \mathbb{P}(F_{[0,n]}((\underline{1}, \underline{s}), L_0^n) \in \Gamma(a)) = 0,$$

and this implies that

$$\lim_{n\to+\infty} \mathbb{P}(F_{[0,n]}((\underline{1}, \underline{s}), L_0^n) \notin \{(1, 1), \ldots, (s, s)\}) = 0,$$

which concludes the proof.

5.2. *Proof of Theorem 3.* Let $\mathbf{X}$ be any process with alphabet A. For all $\underline{x} \in A^{-\mathbb{N}}$, let $\mathbf{X}^{\underline{x}}$ be the process with fixed past $\underline{x}$. Let $\underline{y}, \underline{z} \in A^{-\mathbb{N}}$ and $(\mathbf{X}^{\underline{y}}, \mathbf{X}^{\underline{z}})$ be a coupling between the processes $\mathbf{X}^{\underline{y}}$ and $\mathbf{X}^{\underline{z}}$. The following lemma, which we state without proof, is a direct consequence of [**4**, Theorem 1].

https://doi.org/10.1017/etds.2013.7 Published online by Cambridge University Press

LEMMA 2. (Chazottes *et al* [**4**]) *Let* $(\mathbf{X}^{\underline{y}}, \mathbf{X}^{\underline{z}})$ *be couplings for each pair* $\underline{y}, \underline{z} \in A^{-\mathbb{N}}$. *If*

$$\sup_{\underline{y},\underline{z}} \sum_{j=1}^{\infty} \mathbb{P}(X_j^{\underline{y}} \neq X_j^{\underline{z}}) \leq \Delta < \infty,$$

then for all integers $n \geq 1$, *functions* $f : A^n \to \mathbb{R}$, *and* $\epsilon > 0$,

$$\mathbb{P}(|f(X_1^n) - \mathbb{E}[f(X_1^n)]| > \epsilon) \leq 2 \exp\left\{-\frac{2\epsilon^2}{(1+\Delta)^2 \|\delta f\|_{\ell_2(\mathbb{N})}^2}\right\}.$$

Assume that the process $\mathbf{X}$ is a finitary coding of a sequence $\mathbf{U}$, where $U_i \in \mathcal{U}$ for any $i \in \mathbb{Z}$. Let θ_1, θ_2, and Φ be the quantities involved in the finitary coding as defined generically in §2 (and not necessarily as in §4, which is specific to the attractive sampler algorithm). Let $\epsilon > 0$. Take two positive numbers r_1 and r_2 such that

$$\mathbb{P}(\{\theta_1 > r_1\} \cup \{\theta_2 > r_2\}) \leq \epsilon/(6\|\delta f\|_{\ell_1(\mathbb{N})}).$$

Now, we want to show that we can approximate the FP by two $(r_1 + r_2)$-dependent processes, for which we have good control of the concentration of measure property.

For $j \in \mathbb{Z}$, let $\mathbf{V}^{[j]}$ be a family of i.i.d. processes with values in $\mathcal{U}$ independent of each other and of $\mathbf{U}$. We define a process $\mathbf{Y}$ as

$$Y_j = [\Phi(V_{j+r_2}^{[j],\infty} U_{j-r_1}^{j+r_2} V_{-\infty}^{[j],j-r_1})]_j \quad \text{for any } j \in \mathbb{Z},$$

where, for any i and j in $\mathbb{Z}$, we use the notation $V_i^{[j],\infty}$ for the sequence $\cdots V_{i+1}^{[j]} V_i^{[j]}$ and the notation $V_{-\infty}^{[j],i}$ for the sequence $V_{i-1}^{[j]} V_{i-2}^{[j]} \cdots$. Clearly, $\mathbf{Y}$ is stationary and if $\theta_1 \leq r_1$ and $\theta_2 \leq r_2$, then $Y_0 = X_0$. Moreover, $\mathbf{Y}$ is an $(r_1 + r_2)$-dependent process, that is, for all $l, m > 1$ and $\mathbf{y} \in A^{\mathbb{Z}}$,

$$\mathbb{P}(Y_1^l = y_1^l, Y_{l+r_1+r_2+1}^{l+r_1+r_2+m} = y_{l+r_1+r_2+1}^{l+r_1+r_2+m}) = \mathbb{P}(Y_1^l = y_1^l)\mathbb{P}(Y_{l+r_1+r_2+1}^{l+r_1+r_2+m} = y_{l+r_1+r_2+1}^{l+r_1+r_2+m}).$$

We will now use the equality

$$\begin{aligned} f(X_1^n) - \mathbb{E}[f(X_1^n)] = {} & f(X_1^n) - f(Y_1^n) - \mathbb{E}[f(X_1^n) - f(Y_1^n)] \\ & + f(Y_1^n) - \mathbb{E}[f(Y_1^n)]. \end{aligned} \tag{9}$$

Consider the events

$$O_1^j = \{\theta_1(T_{\mathcal{U}}^j \mathbf{U}) > r_1\} \quad \text{and} \quad O_2^j = \{\theta_2(T_{\mathcal{U}}^j \mathbf{U}) > r_2\}.$$

From the definition of δf, we have that

$$|f(X_1^n) - f(Y_1^n)| \leq \sum_{j=1}^{n} \mathbf{1}\{X_j \neq Y_j\}\delta_j f \leq \sum_{j=1}^{n} \mathbf{1}\{O_1^j \cup O_2^j\}\delta_j f \tag{10}$$

and

$$|\mathbb{E}[f(X_1^k) - f(Y_1^k)]| \leq \mathbb{E}[|f(X_1^k) - f(Y_1^k)|] \leq \mathbb{P}(O_1^0 \cup O_2^0)\|\delta f\|_{\ell_1(\mathbb{N})} \leq \epsilon/6. \tag{11}$$

Observe that

$$\begin{aligned} \mathbb{P}(|f(X_1^n) - \mathbb{E}[f(X_1^n)]| > \epsilon) \leq {} & \mathbb{P}(|f(X_1^n) - f(Y_1^n) - \mathbb{E}[f(X_1^n) - f(Y_1^n)]| > 2\epsilon/3) \\ & + \mathbb{P}(|f(Y_1^n) - \mathbb{E}[f(Y_1^n)]| > \epsilon/3). \end{aligned}$$

https://doi.org/10.1017/etds.2013.7 Published online by Cambridge University Press

Therefore, collecting (9)–(11), we have

$$\begin{aligned}\mathbb{P}(|f(X_1^n)-\mathbb{E}[f(X_1^n)]|>\epsilon)&\\ \le \mathbb{P}\left(\left|\sum_{j=1}^{n}\mathbf{1}\{O_1^j\cup O_2^j\}\delta_j f-\mathbb{P}(O_1^0\cup O_2^0)\|\delta f\|_{\ell_1(\mathbb{N})}\right|>\epsilon/3\right)&\\ +\mathbb{P}(|f(Y_1^n)-\mathbb{E}[f(Y_1^n)]|>\epsilon/3).&\end{aligned}\tag{12}$$

We will use Lemma 2 to obtain upper bounds for the two terms of the right-hand side of (12).

Let us begin with the second term, and we will use the fact that $\mathbf{Y}$ is an (r_1+r_2)-dependent process. First we define a coupling $(\tilde{\mathbf{Y}},\hat{\mathbf{Y}})$, where $\tilde{\mathbf{Y}}$ and $\hat{\mathbf{Y}}$ are copies of $\mathbf{Y}$, by

$$(\tilde{Y}_j,\hat{Y}_j):=([\Phi(V_{j+r_2}^{[j],\infty}U_{j-r_1}^{j+r_2}V_{-\infty}^{[j],j-r_1})]_j,[\Phi(V_{j+r_2}^{\prime[j],\infty}U_{j-r_1}^{\prime j+r_2}V_{-\infty}^{\prime[j],j-r_1})]_j)\quad\text{for all } j\in\mathbb{Z},$$

where $\mathbf{V}^{[j]}$, $\mathbf{V}^{\prime[j]}$ and $\mathbf{U}$, $\mathbf{U}'$ are i.i.d. processes satisfying the following properties. For $j>0$, $\mathbf{V}^{[j]}=\mathbf{V}^{\prime[j]}$. For $j\le 0$, $\mathbf{V}^{[j]}$ is independent of $\mathbf{V}^{\prime[j]}$, and both are independent of the rest. The processes $\mathbf{U}$, $\mathbf{U}'$ are independent of $\mathbf{V}^{[j]}$, $\mathbf{V}^{\prime[j]}$ for all $j\in\mathbb{Z}$. Also $\{U_j\}_{j\le r_2}$ and $\{U_j'\}_{j\le r_2}$ are independent and for $j>r_2$, $U_j=U_j'$. From the construction, for all $\underline{y},\underline{z}\in A^{-\mathbb{N}}$ and $j>r_1+r_2$,

$$\mathbb{P}(\tilde{Y}_j=\hat{Y}_j|\tilde{Y}_{-\infty}^0=\underline{y},\hat{Y}_{-\infty}^0=\underline{z})=\mathbb{P}(\tilde{Y}_j=\hat{Y}_j)=1.$$

Therefore, using the above coupling and the fact that $\mathbf{Y}$ is an (r_1+r_2)-dependent process, we have $\Delta=r_2+\sum_{j=1}^{r_1}\mathbb{P}(\theta_1\ge j)$ in Lemma 2. Hence,

$$\mathbb{P}(|f(Y_1^n)-\mathbb{E}[f(Y_1^n)]|>\epsilon/3)\le 2\exp\left\{-\frac{g(\epsilon,\|\delta f\|_{\ell_1(\mathbb{N})})}{\|\delta f\|_{\ell_2(\mathbb{N})}^2}\right\},\tag{13}$$

where

$$g(\epsilon,\|\delta f\|_{\ell_1(\mathbb{N})})=\frac{2}{9}\epsilon^2\left(1+r_2+\sum_{j=1}^{r_1}\mathbb{P}(\theta_1\ge j)\right)^{-2}.$$

Note that r_1 and r_2 depend on ϵ and $\|\delta f\|_{\ell_1(\mathbb{N})}$.

For the first term of the right-hand side of (12), let $Z_j=\mathbf{1}\{O_1^j\cup O_2^j\}$. Observe that the process $\mathbf{Z}=\{Z_j\}_{j\in\mathbb{Z}}$ is stationary. Also, because the event $O_1^j\cup O_2^j$ is $\mathcal{F}(U_{-r_1+j}^{r_2+j})$ measurable, the process $\mathbf{Z}$ is an (r_1+r_2)-dependent process. We define a coupling $(\tilde{\mathbf{Z}},\hat{\mathbf{Z}})$, where $\tilde{\mathbf{Z}}$ and $\hat{\mathbf{Z}}$ are copies of $\mathbf{Z}$, by

$$\begin{aligned}(\tilde{Z}_j,\hat{Z}_j)=(&\mathbf{1}\{\{\theta_1(T_{\mathcal{U}}^j\mathbf{U})\ge r_1\}\cup\{\theta_2(T_{\mathcal{U}}^j\mathbf{U})\ge r_2\}\},\\ &\mathbf{1}\{\{\theta_1(T_{\mathcal{U}}^j\mathbf{U}')\ge r_1\}\cup\{\theta_2(T_{\mathcal{U}}^j\mathbf{U}')\ge r_2\}\}),\end{aligned}$$

where $\mathbf{U}$ and $\mathbf{U}'$ are i.i.d. processes with $U_j=U_j'$ for $j>r_2$ and $\{U_j\}_{j\le r_2}$ and $\{U_j'\}_{j\le r_2}$ are independent. Therefore, in the same way as (13), by Lemma 2 we have that

$$\mathbb{P}\left(\left|\sum_{j=1}^{n}\mathbf{1}\{O_1^j\cup O_2^j\}\delta_j f-\mathbb{P}(O_1^0\cup O_2^0)\|\delta f\|_{\ell_1(\mathbb{N})}\right|>\frac{\epsilon}{3}\right)\le 2\exp\left\{-\frac{g(\epsilon,\|\delta f\|_{\ell_1(\mathbb{N})})}{\|\delta f\|_{\ell_2(\mathbb{N})}^2}\right\},$$

https://doi.org/10.1017/etds.2013.7 Published online by Cambridge University Press

where g was defined in (13). Finally, we have

$$\mathbb{P}(|f(X_1^n) - \mathbb{E}[f(X_1^n)]| > \epsilon) \le 4\exp\left\{-\frac{g(\epsilon, \|\delta f\|_{\ell_1(\mathbb{N})})}{\|\delta f\|^2_{\ell_2(\mathbb{N})}}\right\},$$

which proves the theorem.

5.3. *Proof of Theorem 4.* We say that a stationary process $\mathbf{X}$ has the *positive divergence property* if

$$\liminf_{n\to+\infty} \frac{1}{n+1} \sum_{a\in A^{n+1}} \mathbb{P}(Y_{-n}^0 = a_{-n}^0) \log \frac{\mathbb{P}(Y_{-n}^0 = a_{-n}^0)}{\mathbb{P}(X_{-n}^0 = a_{-n}^0)} > 0$$

for any ergodic process $\mathbf{Y}$ different of $\mathbf{X}$.

The proof of Theorem 4 is based on the following lemmas. The first result is from [**21**, Theorem 1].

LEMMA 3. (Marton and Shields [**21**]) *Let* $\mathbf{X}$ *be a stationary process with ergodic theorem holding at exponential rate. Then* $\mathbf{X}$ *has the positive divergence property.*

We will now prove that two stationary chains compatible with the same regular probability kernel cannot be distinguished by divergence rate.

LEMMA 4. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be two stationary processes compatible with a continuous kernel* P. *Let* $\inf_{\underline{a}\in A^{\mathbb{Z}}} P(a_0|a_{-\infty}^{-1}) = \delta > 0$. *Then the relative entropy rate*

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{a\in A^{n+1}} \mathbb{P}(X_{-n}^0 = a_{-n}^0) \log \frac{\mathbb{P}(X_{-n}^0 = a_{-n}^0)}{\mathbb{P}(Y_{-n}^0 = a_{-n}^0)}$$

exists and is zero.

Proof. Let $\mathbf{Z}$ represent $\mathbf{X}$ or $\mathbf{Y}$. Define, for $k \in \mathbb{N}$,

$$H_{\mathbf{X}}(Z_{-k}^0) = -\sum_{a\in A^{k+1}} \mathbb{P}(X_{-k}^0 = a_{-k}^0) \log \mathbb{P}(Z_{-k}^0 = a_{-k}^0).$$

Now, we can rewrite the relative entropy rate as

$$\lim_{n\to\infty} \frac{1}{n+1}\{H_{\mathbf{X}}(Y_{-n}^0) - H_{\mathbf{X}}(X_{-n}^0)\}.$$

Define also, for $k \in \mathbb{N}$,

$$H_{\mathbf{X}}(Z_0|Z_{-k}^{-1}) = -\sum_{a\in A^{k+1}} \mathbb{P}(X_{-k}^0 = a_{-k}^0) \log \mathbb{P}(Z_0 = a_0|Z_{-k}^{-1} = a_{-k}^{-1}).$$

By the chain rule and the stationarity of the processes,

$$H_{\mathbf{X}}(Z_{-k}^0) = H_{\mathbf{X}}(Z_0) + \sum_{k=1}^{n} H_{\mathbf{X}}(Z_0|Z_{-k}^{-1}).$$

Therefore, the relative entropy rate is a difference between two Cesàro sums. To prove that the relative entropy exists, it is enough to show that the limit

$$\lim_{n\to\infty} H_{\mathbf{X}}(Z_0|Z_{-n}^{-1})$$

https://doi.org/10.1017/etds.2013.7 Published online by Cambridge University Press

exists. To see that $H_{\mathbf{X}}(Z_0|Z_{-n}^{-1})$ converges, let $\mu_{\mathbf{Z}}$ be the measure associated with $\mathbf{Z}$; then

$$H_{\mathbf{X}}(Z_0|Z_{-n}^{-1}) = -\mathbb{E}_{\mathbf{X}}(\log(\mu_{\mathbf{Z}}(X_0|X_{-n}^{-1}))).$$

By assumption, for all $a \in A^{\mathbb{Z}}$,

$$-\log(\mu_{\mathbf{Z}}(a_0|a_{-n}^{-1})) \leq -\log \delta,$$

and, therefore, by the dominated convergence theorem,

$$\lim_{n\to\infty} -\mathbb{E}_{\mathbf{X}}(\log(\mu_{\mathbf{Z}}(X_0|X_{-n}^{-1}))) = -\mathbb{E}_{\mathbf{X}}(\lim_{n\to\infty} \log(\mu_Z(X_0|X_{-n}^{-1}))).$$

By continuity of P we have that, for all $a \in A^{\mathbb{Z}}$,

$$\lim_{n\to\infty} \log(\mu_{\mathbf{Z}}(a_0|a_{-n}^{-1})) = \log P(Y_0 = a_0|Y_{-\infty}^{-1} = a_{-\infty}^{-1}).$$

Hence,

$$\lim_{n\to\infty} H_{\mathbf{X}}(Z_0|Z_{-n}^{-1}) = -\mathbb{E}_{\mathbf{X}}(\log(P(X_0|X_{-\infty}^{-1}))),$$

which concludes the proof. □

Proof of Theorem 4. If $\mathbf{X}$ has the concentration of measure holding at exponential rate, then it satisfies the ergodic theorem at exponential rate (see equation (3)). By Lemma 3, $\mathbf{X}$ has the positive divergence property. By Lemma 4, if $\mathbf{X}$ has the positive divergence property, there is no other ergodic process compatible with P. Therefore, we conclude that $\mathbf{X}$ is the unique stationary process compatible with P. □

5.4. *Proof of Theorem 5.* The kernel of the original BK example (see §2) is defined through

$$P(+1|\underline{x}) = \sum_{k\geq 1} \lambda_k \phi\left(\frac{1}{m_k}\sum_{i=1}^{m_k} x_{-i}\right)$$

with $\phi(t) = (1-\epsilon)\mathbf{1}\{t > 0\} + \epsilon\mathbf{1}\{t < 0\}$ for some $\epsilon \in (0, 1/2)$, an increasing sequence of odd positive integers $\{m_j\}_{j\in\mathbb{N}}$, and $\lambda_k = cr^k$ for some $r \in (2/3, 1)$ and $c = (1-r)/r$. In other words, the BK example is given in the form of a countable mixture of Markov kernels $P^{[m_k]}$, $k \geq 1$, of lacunary ranges. When uniqueness holds, we denote by $\mathbf{X}^{\mathrm{BK}}$ the stationary chain compatible with P.

The proof of Theorem 5 is based on Lemma 5 below, which implies that the $\mathbf{X}^{\mathrm{BK}}$ is a finitary coding of a *finite entropy* i.i.d. process with countable alphabet. Then we can use [**27**, Theorem 8] which states that countable state mixing Markov process having an exponentially decaying return times state is finitarily Bernoulli. This implies, in particular, that i.i.d. chains with finite entropy are finitary codings of i.i.d. chains on a finite alphabet. We conclude the proof of the theorem by observing that if a process $\mathbf{X}$ is a finitary coding of a process $\mathbf{Y}$, which is itself a finitary coding of a process $\mathbf{Z}$, then $\mathbf{X}$ is also a finitary coding of $\mathbf{Z}$.

LEMMA 5. *The sequence* $\mathbf{L}$ *used in the attractive sampler for* $\mathbf{X}^{\mathrm{BK}}$ *can be chosen to have finite entropy.*

https://doi.org/10.1017/etds.2013.7 Published online by Cambridge University Press

Proof. The sequence $\mathbf{L}$ used in the attractive sampler (see §4) derives from the choice of a countable decomposition of the (maximal) coupling kernel $\tilde{P}$. In the present case, the special form of the BK example will naturally yield the decomposition, as it is itself defined through such decomposition.

Let us first rewrite the kernel P as

$$P(+1|\underline{x}) = \epsilon + \sum_{k\geq 1}(1-2\epsilon)\lambda_k \mathbf{1}\{\mathrm{maj}(x_{-m_k}^{-1}) = 1\}, \tag{14}$$

where

$$\mathrm{maj}(x_{-j}^{-1}) = \mathbf{1}\Big\{\sum_{i=1}^{j} x_{-i} \geq 0\Big\} - \mathbf{1}\Big\{\sum_{i=1}^{j} x_{-i} < 0\Big\}.$$

Representation (14) motivates the following way to write the coupling kernel:

$$\tilde{P}((a,b)|(\underline{x},\underline{y})) = \tilde{\lambda}_0 \tilde{P}^{[0]}((a,b)) + \sum_{k\geq 1}\tilde{\lambda}_k \tilde{P}^{[k]}((a,b)|(x_{-k}^{-1}, y_{-k}^{-1})),$$

where:

- $\tilde{\lambda}_0 = 2\epsilon$, $\tilde{\lambda}_{m_k} = (1-2\epsilon)\lambda_k$ for $k \geq 1$ and $\tilde{\lambda}_i = 0$ for any $i \notin \{0, m_1, m_2, \ldots\}$;
- $\tilde{P}^{[0]}((a,b)) := \frac{1}{2}\mathbf{1}\{a = b\}$; and
- $\tilde{P}^{[k]}((a,b)|(x_{-k}^{-1}, y_{-k}^{-1})) = \mathbf{1}\{a = \mathrm{maj}(x_{-k}^{-1}) \text{ and } b = \mathrm{maj}(y_{-k}^{-1})\}$.

We observe that this kernel satisfies (5), and that the Markovian kernels of the decomposition are deterministic with respect to the pasts (only 0s and 1s in the transition probabilities). It follows that, for any $k \geq 1$, the set $\mathcal{R}^{[m_k]}$ (see (6)) consists of only two values, $\mathcal{R}^{[m_k]} = \{0, 1\}$, and

$$\mathcal{I}^{[m_k]} = \Big\{\Big[\sum_{j=1}^{k-1}\tilde{\lambda}_{m_j}, \sum_{j=1}^{k}\tilde{\lambda}_{m_j}\Big[\Big\}.$$

We also observe that

$$\mathcal{I}^{[0]} = \{[0, \epsilon[; [\epsilon, 2\epsilon[\}.$$

Therefore, the entropy of the i.i.d. process $\mathbf{L}$ is $\epsilon \log \epsilon + \sum_{k=1}^{\infty}\tilde{\lambda}_{m_k}\log\tilde{\lambda}_{m_k}$, which is finite since $\tilde{\lambda}_{m_k} = (1-2\epsilon)(1-r)r^{k-1}$ for some $r \in (2/3, 1)$. □

Acknowledgements. We gratefully acknowledge A. Galves, R. Fernández, K. Marton and S. Friedli for many discussions during the writing of this paper. We thank the referees for their careful reading of the manuscript and for comments which improved the presentation of the paper. This work is part of the USP project 'Mathematics, Computation, Language and the Brain', FAPESP project 'NeuroMat' (grant 2011/51350-6), CNPq projects 'Stochastic Modeling of the Brain Activity' (grant 480108/2012-9), and Sistemas Estoásticos: equilíbrio e não equilíbrio, limites em escala e percolaç ão (grant 474233/2012-0), and was partially completed during the 'Jorma's Razor 2' meeting supported by CAPES grant AUXPE-PAE-598/2011. SG was supported by FAPESP grant 2009/09809-1. DYT was partially supported by FAPESP grant 2008/08171-0 and a Pew Latin American Fellowship.

https://doi.org/10.1017/etds.2013.7 Published online by Cambridge University Press

References

[1] N. Berger, C. Hoffman and V. Sidoravicius. Nonuniqueness for specifications in $\ell^{2+\epsilon}$. *Preprint*, 2005, arXiv:math/0312344.

[2] M. Bramson and S. Kalikow. Nonuniqueness in g-functions. *Israel J. Math.* **84**(1–2) (1993), 153–160.

[3] P. Cénac, B. Chauvin, F. Paccaut and N. Pouyanne. Variable length Markov chains and dynamical sources. *Séminaire de Probabilités XLIV (Lecture Notes in Mathematics, 2046)*. Eds. C. Donati-Martin, A. Lejay and A. Rouault. Springer, Berlin, 2012, pp. 1–39.

[4] J.-R. Chazottes, P. Collet, C. Külske and F. Redig. Concentration inequalities for random fields via coupling. *Probab. Theory Related Fields* **137**(1–2) (2007), 201–225.

[5] F. Comets, R. Fernández and P. A. Ferrari. Processes with long memory: regenerative construction and perfect simulation. *Ann. Appl. Probab.* **12**(3) (2002), 921–943.

[6] E. De Santis and M. Piccioni. Backward coalescence times for perfect simulation of chains with infinite memory. *J. Appl. Probab.* **49**(2) (2012), 319–337.

[7] W. Doeblin and R. Fortet. Sur des chaînes à liaisons complètes. *Bull. Soc. Math. France* **65** (1937), 132–148.

[8] R. Fernández, S. Gallo and G. Maillard. Regular g-measures are not always Gibbsian. *Electron. Commun. Probab.* **16** (2011), 732–740.

[9] R. Fernández and G. Maillard. Chains with complete connections: general theory, uniqueness, loss of memory and mixing properties. *J. Stat. Phys.* **118**(3–4) (2005), 555–588.

[10] S. Friedli. A note on the Bramson–Kalikow process. Unpublished, 2010, http://www.mat.ufmg.br/~sacha/textos/BK/plateaux.pdf.

[11] S. Gallo. Chains with unbounded variable length memory: perfect simulation and a visible regeneration scheme. *Adv. in Appl. Probab.* **43**(3) (2011), 735–759.

[12] S. Gallo and N. L. Garcia. Perfect simulation for locally continuous chains of infinite order. *Preprint*, 2011, arXiv:1103.2058v4.

[13] T. E. Harris. On chains of infinite order. *Pacific J. Math.* **5** (1955), 707–724.

[14] P. Hulse. Uniqueness and ergodic properties of attractive g-measures. *Ergod. Th. & Dynam. Sys.* **11**(1) (1991), 65–77.

[15] P. Hulse. An example of non-unique g-measures. *Ergod. Th. & Dynam. Sys.* **26**(2) (2006), 439–445.

[16] A. Johansson and A. Öberg. Square summability of variations of g-functions and uniqueness of g-measures. *Math. Res. Lett.* **10**(5–6) (2003), 587–601.

[17] S. Kalikow. Random Markov processes and uniform martingales. *Israel J. Math.* **71**(1) (1990), 33–54.

[18] S. Karlin. Some random walks arising in learning models. I. *Pacific J. Math.* **3** (1953), 725–756.

[19] M. Keane. Strongly mixing g-measures. *Invent. Math.* **16** (1972), 309–324.

[20] Y. Lacroix. A note on weak-$\star$ perturbations of g-measures. *Sankhyā A* **62**(3) (2000), 331–338.

[21] K. Marton and P. C. Shields. The positive-divergence and blowing-up properties. *Israel J. Math.* **86**(1–3) (1994), 331–348.

[22] P. McCullagh and J. A. Nelder. *Generalized Linear Models (Monographs on Statistics and Applied Probability)*. Chapman & Hall, London, 1983.

[23] O. Onicescu and G. Mihoc. Sur les chaînes de variables statistiques. *Bull. Sci. Math.* **59**(2) (1935), 174–192.

[24] C. Preston. *Random Fields (Lecture Notes in Mathematics, 534)*. Springer, Berlin, 1976.

[25] J. G. Propp and D. B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Proceedings of the Seventh International Conference on Random Structures and Algorithms (Atlanta, GA, 1995)*. Vol. 9. Eds. M. Karoński, J. Spencer and A. Ruciński. Wiley, New York, 1996.

[26] A. N. Quas. Non-ergodicity for C^1 expanding maps and g-measures. *Ergod. Th. & Dynam. Sys.* **16**(3) (1996), 531–543.

[27] D. J. Rudolph. A mixing Markov chain with exponentially decaying return times is finitarily Bernoulli. *Ergod. Th. & Dynam. Sys.* **2**(1) (1982), 85–97.

[28] P. C. Shields. *The Ergodic Theory of Discrete Sample Paths (Graduate Studies in Mathematics, 13)*. American Mathematical Society, Providence, RI, 1996.

[29] J. Steif and J. van den Berg. On the existence and nonexistence of finitary codings for a class of random fields. *Ann. Probab.* **11** (1999), 1501–1522.

[30] Ö. Stenflo. A note on a theorem of Karlin. *Statist. Probab. Lett.* **54**(2) (2001), 183–187.

https://doi.org/10.1017/etds.2013.7 Published online by Cambridge University Press