

EXPLICIT ESTIMATES IN THE BRAMSON-KALIKOW MODEL

GALLESCO, C.¹, GALLO, S.², AND TAKAHASHI, D. Y.³

ABSTRACT. The aim of the present article is to explicitly compute parameters for which the Bramson-Kalikow model exhibits phase-transition. The main ingredient of the proof is a simple new criterion for non-uniqueness of g -measures. We show that the existence of multiple g -measures compatible with a function g can be proved by estimating the \bar{d} -distances between some suitably chosen Markov chains. The method is optimal for the important class of binary regular attractive functions, which includes the Bramson-Kalikow model.

1. INTRODUCTION

In this work we consider chains of infinite order, or equivalently g -measures, on a finite alphabet. They constitute an important class of stochastic models, which includes, for example, Markov chains, stochastic models that exhibit non-uniqueness and models that are not Gibbsian (Fernández *et al.*, 2011). The question of uniqueness of g -measures was extensively studied and important progresses have been obtained in several areas related to probability and ergodic theory, from the seminal works of Onicescu & Mihoc (1935); Doeblin & Fortet (1937) to recent advances in Johansson *et al.* (2012); Gallo & Paccaut (2013), and the contributions of Harris (1955); Keane (1972); Walters (1975); Lalley (1986); Stenflo (2003); Fernández & Maillard (2005) among many others. Notwithstanding, the problem of non-uniqueness is much less understood and the literature is still based on few examples (Bramson & Kalikow, 1993; Hulse, 2006; Berger *et al.*, 2005). As far as we know, general criteria

2000 *Mathematics Subject Classification.* Primary 60G10; Secondary 60G99.

Key words and phrases. Chains of infinite order, coupling, phase transition, Bramson-Kalikow, \bar{d} -distance.

This work is supported by 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 Estocásticos: equilíbrio e não equilíbrio, limites em escala e percolação” (grant 474233/2012-0). SG was partially supported by FAPESP (grant 2009/09809-1). CG was partially supported by FAPESP (grants 2009/51139-3 and 2013/10101-9). DYT was partially supported by FAPESP grant 2008/08171-0, Pew Latin American Fellowship, and Ciência sem Fronteiras fellowship (CNPq - 246778/2012-1).

for non-uniqueness have only been obtained for the class of regular attractive kernels (Gallo & Takahashi, 2013; Hulse, 1991).

In the present article we focus on the Bramson-Kalikow (BK) model (Bramson & Kalikow, 1993). It is the most well studied example of non-uniqueness (Lacroix, 2000; Friedli, 2010; Gallo & Takahashi, 2013), nevertheless our understanding of the model is still far from complete. For instance, to our knowledge, there is no explicit computation of the values of the parameters for which the BK model exhibits multiple g -measures (see Friedli (2010) for related discussion). Our main result (Theorem 1) gives such explicit relationship between the parameters in the case of non-uniqueness. Furthermore, we obtain an improvement on the range of parameters that imply non-uniqueness of the BK model. Corollaries 1 and 2 give numerical examples of choices for these parameters. The proof of this result is based on three ingredients:

- 1) a new and simple criterion for non-uniqueness of g -measures (Theorem 2),
- 2) a concentration of measure inequality for g -measures obtained using a result from Chazottes *et al.* (2007),
- 3) \bar{d} -distance estimates between Markov chains using a coupling from the past algorithm.

Theorem 2 has a life of its own and is a criterion for non-uniqueness of g -measures that in principle can be applied to other models. The motivation of Theorem 2 is to avoid the direct study of kernels g with multiple g -measures as these are objects that are generally difficult to analyze (Gallo & Takahashi, 2013). Instead, we study the properties of a sequence of suitably chosen Markov chains. Theorem 2 is inspired by the works of Bramson & Kalikow (1993), Lacroix (2000), and Hulse (1991), but has the advantage of being formulated using the \bar{d} -distance, which is key to our constructive proof of Theorem 1. Moreover, Theorem 3 states that our criterion (Theorem 2) is optimal in the important class of binary regular attractive kernels, giving a necessary and sufficient condition for non-uniqueness in this class, which includes the BK model.

The article is organized as follows. We state the main results and relevant definitions in Section 2. In Section 3 we introduce the couplings used to prove Theorem 1 and Section 4 contains the proofs of Theorem 1 and Corollaries 1 and 2. Finally, in Section 5, we prove Theorem 2 and Theorem 3.

2. NOTATION, DEFINITIONS AND MAIN RESULTS

Let A be a finite set we call *alphabet* and $\mathcal{X} = A^{\mathbb{Z}^-}$. We denote by x_i the i -th coordinate of $x \in \mathcal{X}$. For $i, j \in \mathbb{Z} \cup \{-\infty, +\infty\}$ and $i \leq j$ we write $x_i^j := (x_j \dots x_i)$.

Observe that we use the convention that smaller indices are on right hand side. For $x, y \in \mathcal{X}$, a *concatenation* $x_{-i}^0 y$ is a new sequence $z \in \mathcal{X}$ with $z_{-i}^0 = x_{-i}^0$ and $z_{-\infty}^{-i-1} = y$. We introduce on \mathcal{X} the metric $\rho(x, y) := \min\{\frac{1}{j+1} : x_{-j}^0 = y_{-j}^0\}$, which turns \mathcal{X} into a compact metric space. Denote by \mathcal{B} the Borel σ -algebra on \mathcal{X} . Let $T : \mathcal{X} \rightarrow \mathcal{X}$ be the *shift operator* such that for $x \in \mathcal{X}$ we have $(Tx)_i = x_{i-1}$. We denote by $\mathcal{C}(\mathcal{X})$ the space of continuous functions with norm $\|f\| := \sup_{x \in \mathcal{X}} |f(x)|$. Let also

$$\mathcal{G} := \{g \in \mathcal{C}(\mathcal{X}) : g(x) \in (0, 1) \text{ and } \sum_{a \in A} g(ax) = 1, \forall x \in \mathcal{X}\}.$$

In the literature (Bramson & Kalikow, 1993), a function in \mathcal{G} is called *regular kernel*. We denote by \mathcal{M}_k the set of regular k -th order *Markov kernels* on \mathcal{X} and by $\mathcal{M} = \bigcup_{k \geq 0} \mathcal{M}_k$ the set of regular Markov kernels. We have $\mathcal{M} \subset \mathcal{G}$. Sometimes, we consider a well ordered set A and then \mathcal{X} is endowed with partial order $x \geq y \Leftrightarrow x_i \geq y_i$ for all $i \in \mathbb{Z}_-$. A function $g \in \mathcal{G}$ is *attractive* if A is well ordered and for all $a \in A$, $\sum_{b \geq a} g(bx)$ is an increasing function of $x \in \mathcal{X}$.

Let $g \in \mathcal{G}$, following Walters (1975) we say that a probability measure μ on \mathcal{X} is a *g -measure* if it is T -invariant and, for all $a \in A$ and $x \in \mathcal{X}$,

$$\mu(\{x \in \mathcal{X} : x_0 = a\} | T^{-1}\mathcal{B})(x) = g(ax_{-\infty}^{-1})$$

or equivalently,

$$\int_{\mathcal{X}} f d\mu = \int_{\mathcal{X}} \sum_{a \in A} g(ax) f(ax) d\mu, \quad (1)$$

for all $f \in \mathcal{C}(\mathcal{X})$. A process $(X_n)_{n \in \mathbb{Z}}$ is said to be *compatible with g* if its law is a g -measure. In this article, we are interested on conditions for *non-uniqueness* of g -measures, *i.e.*, sufficient conditions for the existence of several g -measures with the same kernel g .

We will now define the model introduced by Bramson & Kalikow (1993). Let $A = \{-1, +1\}$, $\epsilon \in (0, 1/2)$, and $(m_j)_{j \geq 1}$ be an increasing sequence of positive odd numbers. Let $x \in \mathcal{X}$, we denote by $p_{[m_j]} \in \mathcal{M}_{m_j}$ the kernel

$$p_{[m_j]}(x) = \mathbf{1} \left\{ x_0 \sum_{l=1}^{m_j} x_{-l} > 0 \right\} (1 - \epsilon) + \mathbf{1} \left\{ x_0 \sum_{l=1}^{m_j} x_{-l} < 0 \right\} \epsilon. \quad (2)$$

Let $(\lambda_j)_{j \geq 1}$ be a sequence of positive numbers such that $\sum_{j=1}^{\infty} \lambda_j = 1$. Given $(m_j)_{j \geq 1}$ and $(\lambda_j)_{j \geq 1}$, the BK-model is given by the kernel $p \in \mathcal{G}$ such that, for all $x \in \mathcal{X}$,

$$p(x) = \sum_{j=1}^{\infty} \lambda_j p_{[m_j]}(x). \quad (3)$$

It is immediate that the BK-model p is attractive and regular. Bramson & Kalikow (1993) showed that if $\lambda_j = (1 - s)s^{j-1}$ for $s \in (2/3, 1)$, there exists a sequence $(m_j)_{j \geq 1}$ for which the BK model has multiple p -measures. However, it is not known how the sequence $(m_j)_{j \geq 1}$ should be explicitly chosen. Theorem 1 below exhibits an explicit relationship between sequences $(\lambda_j)_{j \geq 1}$ and $(m_j)_{j \geq 1}$ for which there are multiple p -measures.

Theorem 1. *Let $(\lambda_j)_{j \geq 1}$ and $(m_j)_{j \geq 1}$ be the sequences that define the BK model p in (3). Let $m_0 = 0$, $r : \{1, 2, \dots\} \rightarrow \mathbb{Z}_+$ be a function such that $r_k < k$, and $\alpha \in (0, \frac{1}{2} - \epsilon)$. If for all $k \geq 0$ we have $\sum_{j \geq k+2} \lambda_j > \sum_{j=r_{k+1}+1}^{k+1} \lambda_j$ and*

$$m_{k+1} \geq \frac{A_k}{\left(\sum_{j \geq k+2} \lambda_j - \sum_{j=r_{k+1}+1}^{k+1} \lambda_j\right)^2}, \quad (4)$$

where

$$A_k := 8(1 - 2\epsilon)^{-2} (1 + m_{r_{k+1}}(2\epsilon)^{-m_{r_{k+1}}})^2 \ln \left(2^{k+2} (1 + m_k(2\epsilon)^{-m_k}) \alpha^{-1} \right), \quad (5)$$

then the corresponding BK model p has multiple p -measures.

Let us now give two numerical examples of sequences $(\lambda_j)_{j \geq 1}$ and $(m_j)_{j \geq 1}$ for which there are multiple p -measures, illustrating the relationship between the sequences $(\lambda_j)_{j \geq 1}$ and $(m_j)_{j \geq 1}$ in Theorem 1.

Corollary 1. *Let $\epsilon = 1/4$ and for $j \geq 1$, $\lambda_j = \frac{1}{2}(\frac{2}{3})^j$. Let $m_1 = 217$, c be an odd positive integer, and for $j \geq 1$, $m_{j+1} = c^{m_j}$. If $c \geq 577$, then the associated BK model has multiple p -measures.*

The next corollary illustrates the improvement on the growth rate of $(m_j)_{j \geq 1}$ that we obtain due to a better understanding of its relationship with the rate of $(\lambda_j)_{k \geq 1}$ through the function r in Theorem 1.

Corollary 2. *Let $\epsilon = 1/4$, c be a positive integer and for $j \geq 1$, $m_j = 2^{c^j} - 1$. Let $b_1 = 1$ and for $l \geq 2$, $b_l = 2^{(c^{\sum_{j=1}^{l-1} b_j})^2}$. For $l \geq 1$ and $j \in \{\sum_{k=1}^{l-1} b_k + 1, \dots, \sum_{k=1}^l b_k\}$ we set $\lambda_j = (3/4)^{l-1}/(4b_l)$. If $c \geq 8$, then the associated BK model has multiple p -measures.*

To prove that our result is tight, we need a criterion for uniqueness of p -measures with conditions on the parameters comparable to Theorem 1. Known criteria for uniqueness (Johansson *et al.*, 2012; Fernández & Maillard, 2005) don't give such conditions. Therefore, the existence of a sharp transition from uniqueness to non-uniqueness regime for the BK model still remains an interesting open problem.

Before stating Theorem 2, we need to introduce Ornstein's \bar{d} -distance (Shields, 1996). We say that a measure ν on $\mathcal{X} \times \mathcal{X}$ is a coupling between μ and μ' if for all measurable subsets Γ of \mathcal{X} we have $\nu(\Gamma \times \mathcal{X}) = \mu(\Gamma)$ and $\nu(\mathcal{X} \times \Gamma) = \mu'(\Gamma)$. The set of all $T \otimes T$ -invariant couplings between μ and μ' is denoted by $C(\mu, \mu')$ and the \bar{d} -distance between μ and μ' is defined by

$$\bar{d}(\mu, \mu') = \inf_{\nu \in C(\mu, \mu')} \nu(\{(x, x') \in \mathcal{X} \times \mathcal{X} : x_0 \neq x'_0\}).$$

Let $\tilde{\mu}$ and \tilde{T} be respectively the natural extensions on $A^{\mathbb{Z}}$ of the g -measure μ and the shift operator T . We say that this natural extension $(\tilde{\mu}, \tilde{T})$ is *Bernoulli* if it is isomorphic to a Bernoulli shift.

Theorem 2. *Let $(g_j)_{j \geq 0}$ and $(g'_j)_{j \geq 0}$ be two sequences of kernels in \mathcal{M} both converging to $g \in \mathcal{G}$ in $\mathcal{C}(\mathcal{X})$. Let μ_j and μ'_j be the unique associated g_j and g'_j -measures. If there exists an integer $k \geq 0$ such that*

$$\sum_{j \geq k} \bar{d}(\mu_j, \mu_{j+1}) + \sum_{j \geq k} \bar{d}(\mu'_j, \mu'_{j+1}) < \bar{d}(\mu_k, \mu'_k), \quad (6)$$

then there exist at least two distinct g -measures μ and μ' . Moreover, the natural extensions of both g -measures are Bernoulli.

The main advantage of Theorem 2 is that we need to know nothing *a priori* about the non-unique g -measures. The only requirement is a good control of the coupling between the Markov approximations.

We state below that the converse of Theorem 2 holds for the important class of binary attractive kernels $g \in \mathcal{G}$, which includes for example the Bramson-Kalikow model (Bramson & Kalikow, 1993).

Theorem 3. *Let $A = \{-1, +1\}$. If $g \in \mathcal{G}$ is attractive, then there exist multiple g -measures if and only if there exist two sequences $(g_j)_{j \geq 0}$ and $(g'_j)_{j \geq 0}$ of functions in \mathcal{M} both converging to $g \in \mathcal{G}$ in $\mathcal{C}(\mathcal{X})$ such that the associated g_j and g'_j -measures μ_j and μ'_j satisfy, for some $k \geq 0$, the inequality (6).*

3. COUPLINGS AND PERFECT SIMULATIONS

The proof of Theorem 1 will use Theorem 2 which involves the \bar{d} -distance between Markov chains. Therefore, we will construct several couplings and Markov chains. The constructions are conceptually straightforward but tedious to write, thus for convenience of the reader we will define all the constructions in the present section.

All the stationary measures needed in the proof of Theorem 1 will be simultaneously constructed using only a single sequence $\mathbf{U} := (U_j)_{j \in \mathbb{Z}}$ of i.i.d. r.v.'s uniformly

distributed in $[0, 1)$. We let $(\Omega, \mathcal{F}, \mathbb{P})$ denote the probability space corresponding to the sequence \mathbf{U} , and \mathbb{E} the expectation under \mathbb{P} .

The first notion that we need to introduce is that of *coupling from the past* (CFTP) algorithm.

Coupling from the past. The idea of the construction is the following. First, for $g \in \mathcal{G}$, we associate an *update function* $F : [0, 1) \times A^{\mathbb{Z}^-} \rightarrow A$ that satisfies $\mathbb{P}(F(U_0, x) = a) = g(ax)$ for any $x \in \mathcal{X}$ and $a \in A$. For any pair of integers i, j such that $-\infty < i \leq j < +\infty$, let $F_{\{j,i\}}(U_i^j, x) \in A^{j-i+1}$ be the sample obtained by *applying recursively* F on the fixed past x , i.e, let $F_{\{i,i\}}(U_i, x) := F(U_i, x)$ and for any $j > i$

$$F_{\{j,i\}}(U_i^j, x) := F(U_j, F_{\{j-1,i\}}(U_i^{j-1}, x))F_{\{j-1,i\}}(U_i^{j-1}, x).$$

Secondly, define $F_{[i,i]}(U_i, x) := F(U_i, x)$ and

$$F_{[j,i]}(U_i^j, x) = F(U_j, F_{\{j-1,i\}}(U_i^{j-1}, x)x).$$

$F_{[j,i]}(U_i^j, x)$ is the last symbol of the sample $F_{\{j,i\}}(U_i^j, x)$.

With these definitions, for all $x \in \mathcal{X}$ we can construct the sequence $(X_j^{(x)})_{j \geq 1}$ defined by

$$X_j^{(x)} := F_{[j,1]}(U_1^j, x),$$

which is the stochastic process starting with a fixed past $x \in \mathcal{X}$ and updated according to g .

Now we can define the notion of perfect simulation by coupling from the past. Let θ be the *coalescence time* defined by

$$\theta := \min \{i \geq 0 : F_{[0,-i]}(U_{-i}^0, x) = F_{[0,-i]}(U_{-i}^0, y) \text{ for all } x, y \in \mathcal{X}\}.$$

It can be proved (see Propp & Wilson (1996); Comets *et al.* (2002); De Santis & Piccioni (2012) for instance) that if θ is \mathbb{P} -a.s. finite then there is a unique process $(X_j)_{j \in \mathbb{Z}}$ compatible with g , such that,

$$F_{[0,-\theta]}(U_{-\theta}^0, x) \stackrel{\mathcal{D}}{=} X_0, \quad \forall x \in \mathcal{X}.$$

It is clear from the above construction that we can repeat the procedure iteratively using the same sequence \mathbf{U} to obtain the process on any finite interval. Therefore, when an update function F and a \mathbb{P} -a.s. finite θ exist, we say that there exists a CFTP algorithm that *perfectly simulates* $(X_j)_{j \in \mathbb{Z}}$. Observe that we are considering the bi-infinite stationary process on \mathbb{Z} rather than the process restricted on \mathbb{Z}_+ , as this is more convenient for the proof of Theorem 1.

We note that Gallo & Takahashi (2013) proved that an attractive g -measure is unique if and only if it can be perfectly simulated through a CFTP algorithm. Therefore, the non-unique p -measures for the BK model considered in the present paper cannot be simulated through a CFTP. Instead, in the present article we use CFTP to make a simultaneous construction of all the Markov chains obtained by truncating the initial kernel p . These truncations are introduced in the next paragraph.

Update function for the truncating Markov kernels. We will consider different Markov kernels in the proof of Theorem 1. They are truncations of order m_k of the Bramson-Kalikow's kernel $p \in \mathcal{C}(\mathcal{X})$ defined in (3). Let $x \in \mathcal{X}$, $p_{[0]} \in \mathcal{M}_0$ be defined by $p_{[0]}(x) = (1 - \epsilon)\mathbf{1}\{x_0 > 0\} + \epsilon\mathbf{1}\{x_0 < 0\}$, and $p_{[m_j]} \in \mathcal{M}_{m_j}$ be defined as in (2). For $l > k \geq 0$, consider the following m_k -th order Markov kernels

$$p_k(x) = \sum_{j=1}^k \lambda_j p_{[m_j]}(x) + \sum_{j=k+1}^{\infty} \lambda_j p_{[0]}(x), \quad (7)$$

$$p'_k(x) = \sum_{j=1}^k \lambda_j p_{[m_j]}(x) + \sum_{j=k+1}^{\infty} \lambda_j (1 - p_{[0]}(x)), \quad (8)$$

$$q_{k,l}(x) = \sum_{j=1}^k \lambda_j p_{[m_j]}(x) + \sum_{j=k+1}^l \lambda_j (1 - p_{[0]}(x)) + \sum_{j=l+1}^{\infty} \lambda_j p_{[0]}(x), \quad (9)$$

$$q'_{k,l}(x) = \sum_{j=1}^k \lambda_j p_{[m_j]}(x) + \sum_{j=k+1}^l \lambda_j p_{[0]}(x) + \sum_{j=l+1}^{\infty} \lambda_j (1 - p_{[0]}(x)), \quad (10)$$

where $\sum_{j=1}^0 x_j$ means that the summand is zero.

Defining $\bar{\lambda}_0 := 2\epsilon$ and $\bar{\lambda}_j := \lambda_j(1 - 2\epsilon)$ for $j \geq 1$, we can respectively rewrite (7) and (9) as

$$p_k(x) = \bar{\lambda}_0 \frac{1}{2} + \sum_{j=1}^k \bar{\lambda}_j \mathbf{1} \left\{ x_0 \sum_{i=1}^{m_j} x_{-i} > 0 \right\} + \sum_{j \geq k+1} \bar{\lambda}_j \left(\frac{1 + x_0}{2} \right),$$

$$q_{k,l}(x) = \bar{\lambda}_0 \frac{1}{2} + \sum_{j=1}^k \bar{\lambda}_j \mathbf{1} \left\{ x_0 \sum_{i=1}^{m_j} x_{-i} > 0 \right\} + \sum_{j=k+1}^l \bar{\lambda}_j \left(\frac{1 + x_0}{2} \right) + \sum_{j \geq l+1} \bar{\lambda}_j \left(\frac{1 - x_0}{2} \right).$$

Similar equations hold for (8) and (10).

Now, for any past $x \in \mathcal{X}$, consider the intervals

$$I_0(-1) := [0, \epsilon[, \quad I_0(+1) = [\epsilon, 2\epsilon[\quad \text{and} \quad I_j = \left[\sum_{i=0}^{j-1} \bar{\lambda}_i, \sum_{i=0}^j \bar{\lambda}_i \right[, \quad j \geq 1. \quad (11)$$

We observe that the lengths $|I_0(-1)| = |I_0(+1)| = \epsilon$ and for $j \geq 1$, $|I_j| = \bar{\lambda}_j$.

It is natural to consider the following update functions for the Markov kernels p_k and $q_{k,l}$ respectively.

$$F^{p_k}(U_0, x) = \sum_{a \in A} a \mathbf{1}\{U_0 \in I_0(a)\} + \sum_{a \in A} \sum_{j=1}^k a \mathbf{1}\{U_0 \in I_j\} \mathbf{1} \left\{ a \sum_{i=1}^{m_j} x_{-i} > 0 \right\} \\ + \sum_{a \in A} \sum_{j \geq k+1} a \mathbf{1}\{U_0 \in I_j\} \left(\frac{1+a}{2} \right),$$

and

$$F^{q_{k,l}}(U_0, x) = \sum_{a \in A} a \mathbf{1}\{U_0 \in I_0(a)\} + \sum_{a \in A} \sum_{j=1}^k a \mathbf{1}\{U_0 \in I_j\} \mathbf{1} \left\{ a \sum_{i=1}^{m_j} x_{-i} > 0 \right\} \\ + \sum_{a \in A} \sum_{j=k+1}^l a \mathbf{1}\{U_0 \in I_j\} \left(\frac{1-a}{2} \right) + \sum_{a \in A} \sum_{j \geq l+1} a \mathbf{1}\{U_0 \in I_j\} \left(\frac{1+a}{2} \right).$$

We can define analogous update functions for p'_k and $q'_{k,l}$.

Let $\underline{+1}, \underline{-1} \in \mathcal{X}$ be defined by $\underline{+1}_j = 1$ and $\underline{-1}_j = -1$ for $j \leq 0$. We define the *coalescence time*

$$\theta^{p_k} := \min \left\{ i \geq 0 : F_{[0,-i]}^{p_k}(U_{-i}^0, x) = F_{[0,-i]}^{p_k}(U_{-i}^0, y) \text{ for all } x, y \in \mathcal{X} \right\} \\ = \min \left\{ i \geq 0 : F_{[0,-i]}^{p_k}(U_{-i}^0, \underline{+1}) = F_{[0,-i]}^{p_k}(U_{-i}^0, \underline{-1}) \right\},$$

where the last equality is a direct consequence of the attractiveness of p_k . We substitute in the above definitions p_k by p'_k , $q_{k,l}$, or $q'_{k,l}$ to define $\theta^{p'_k}$, $\theta^{q_{k,l}}$, and $\theta^{q'_{k,l}}$.

We also define, for any $i \in \mathbb{Z}$ and $k \geq 1$, the *regeneration time* of order k

$$\eta_k := \min \{ i \geq m_k - 1 : U_{-j} \in I_0(-1) \cup I_0(+1), j = i - m_k + 1, \dots, i \}.$$

Couplings between the chains and an upperbound for θ^{p_k} and η_k . We couple all the chains together constructing them simultaneously using the CFTP algorithm with same sequence \mathbf{U} and the respective update functions. Consequently, the coupling law is always \mathbb{P} , *i.e.*, the product law of \mathbf{U} . We also use the same symbol to indicate the marginal process and coupled process, when there is no ambiguity.

In what follows, we collect some lemmas that will be used in the proof Theorem 1. Let us give an upper bound on the expectation of the coalescence and regeneration times that hold for p_k , p'_k , q_k , and q'_k . First, observe that by construction,

$$F_{[0,-\eta_k]}^{p_k}(U_{-\eta_k}^0, \underline{+1}) = F_{[0,-\eta_k]}^{p_k}(U_{-\eta_k}^0, \underline{-1})$$

and, therefore,

$$\mathbb{P}(\eta_k \geq \theta^{p_k}) = 1.$$

The same holds for p'_k , $q_{k,l}$, and $q'_{k,l}$. Now, we have the following lemma.

Lemma 1. *Let η_k be the regeneration time of order k . We have that*

$$\mathbb{E}(\theta^{p_k}) \leq \mathbb{E}(\eta_k) \leq \frac{m_k}{(2\epsilon)^{m_k}}.$$

The same bound holds for $\theta^{p'_k}$, $\theta^{q_{k,l}}$, and $\theta^{q'_{k,l}}$.

Proof. By the definition of η_k we have

$$\mathbb{P}(\eta_k \geq n \cdot m_k) \leq \prod_{i=1}^n \mathbb{P} \left(\left\{ \bigcap_{j=-im_k+1}^{-(i-1)m_k} \left\{ U_j \in \bigcup_{l=0}^{im_k-1+j} I_l \right\} \right\}^c \right).$$

Using the stationarity and independence of \mathbf{U} , we have for $i = 1, \dots, n$

$$\begin{aligned} \mathbb{P} \left(\bigcap_{j=-im_k+1}^{-(i-1)m_k} \left\{ U_j \in \bigcup_{l=0}^{im_k-1+j} I_l \right\} \right) &= \prod_{j=1}^{m_k} \mathbb{P} \left(U_j \in \bigcup_{l=0}^j I_l \right) \\ &= \prod_{j=1}^{m_k} \mathbb{P} \left(U_0 \in \bigcup_{l=0}^j I_l \right). \end{aligned}$$

A simple upper bound is $\prod_{j=1}^{m_k} \mathbb{P}(U_0 \in \bigcup_{l=0}^j I_l) \leq (2\epsilon)^{m_k}$. This yields

$$\mathbb{E}(\eta_k) \leq m_k \sum_{n \geq 1} \mathbb{P}(\eta_k \geq n \cdot m_k) \leq m_k \sum_{n \geq 1} (1 - (2\epsilon)^{m_k})^n \leq \frac{m_k}{(2\epsilon)^{m_k}}.$$

□

Lemma 2. *Let $k < l$ and $(Y_j^{k,l})_{j \in \mathbb{Z}}$ be the stationary process compatible with $q_{k,l}$. If $\sum_{j \geq l+1} \lambda_j > \sum_{j=k+1}^l \lambda_j$ then*

$$\mathbb{E}(Y_0^{k,l}) \geq (1 - 2\epsilon) \left(\sum_{j \geq l+1} \lambda_j - \sum_{j=k+1}^l \lambda_j \right) > 0.$$

Proof. Let $(Z_j^{k,l})_{j \in \mathbb{Z}}$ be the stationary process compatible with $q'_{k,l}$ we observe that

$$\begin{aligned} \mathbb{E}(Y_0^{k,l}) &= \mathbb{P}(Y_0^{k,l} = 1) - \mathbb{P}(Y_0^{k,l} = -1) \\ &= \mathbb{P}(Y_0^{k,l} = 1) - \mathbb{P}(Z_0^{k,l} = 1). \end{aligned}$$

Now, we want to construct a maximal coupling between $(Y_j^{k,l})_{j \in \mathbb{Z}}$ and $(Z_j^{k,l})_{j \in \mathbb{Z}}$. For this we define an update function for $q'_{k,l}$ using a set of intervals slightly different from the intervals defined in (13). We have

$$I'_0(-1) := [0, \epsilon[, \quad I'_0(+1) = [\epsilon, 2\epsilon[\quad \text{and} \quad I'_j = \left[\sum_{i=0}^{j-1} \bar{\lambda}_i, \sum_{i=0}^j \bar{\lambda}_i \right], \quad \text{for } k \geq j \geq 1,$$

$$I'_j = \left[\sum_{i=0}^{j-1} \bar{\lambda}_i + \sum_{i \geq l+1} \bar{\lambda}_i, \sum_{i=0}^j \bar{\lambda}_i + \sum_{i \geq l+1} \bar{\lambda}_i \right], \quad \text{for } l \geq j \geq k+1, \quad (12)$$

and

$$I'_j = \left[\sum_{i=0}^k \bar{\lambda}_i + \sum_{i=l+1}^j \bar{\lambda}_i, \sum_{i=0}^k \bar{\lambda}_i + \sum_{i=l+1}^j \bar{\lambda}_i \right], \quad \text{for } j \geq l, \quad (13)$$

where $\sum_{i=l+1}^l \bar{\lambda}_i = 0$. The update function for $q'_{k,l}$ is then defined by

$$\begin{aligned} H^{q'_{k,l}}(U_0, x) &= \sum_{a \in A} a \mathbf{1}\{U_0 \in I'_0(a)\} + \sum_{a \in A} \sum_{j=1}^k a \mathbf{1}\{U_0 \in I'_j\} \mathbf{1} \left\{ a \sum_{i=0}^{m_j-1} x_{-i} > 0 \right\} \\ &+ \sum_{a \in A} \sum_{j=k+1}^l a \mathbf{1}\{U_0 \in I'_j\} \left(\frac{1-a}{2} \right) + \sum_{a \in A} \sum_{j \geq l+1} a \mathbf{1}\{U_0 \in I'_j\} \left(\frac{1+a}{2} \right). \end{aligned}$$

Observe that this update function is different from $F^{q'_{r,k+1}}$, which uses the intervals defined in (13).

By construction

$$\mathbb{P}(Y_0^{k,l} = 1) - \mathbb{P}(Z_0^{k,l} = 1) = \mathbb{P}(Y_0^{k,l} \neq Z_0^{k,l}).$$

Now, the following lower bound is an immediate consequence of the construction of the coupling

$$\mathbb{P}(Y_0^{k,l} \neq Z_0^{k,l}) \geq (1 - 2\epsilon) \left(\sum_{j \geq l+1} \lambda_j - \sum_{j=k+1}^l \lambda_j \right).$$

□

4. PROOF OF THEOREM 1 AND COROLLARIES 1, 2

The proof of Theorem 1 is based on the results of the last section, Theorem 2 which is proved in the next section, and the following lemma.

Lemma 3. *Let $k < l$ and $(Y_j^{k,l})_{j \in \mathbb{Z}}$ be the stationary process compatible with $q_{k,l}$. For all $l > k > 0$ and $j \geq 1$, we have*

$$\mathbb{P} \left(\left| \frac{1}{m_l} \sum_{i=1}^{m_l} Y_i^{k,l} - \mathbb{E}(Y_0^{k,l}) \right| \geq \frac{\mathbb{E}(Y_0^{k,l})}{2} \right) \leq 2 \exp \left(- \frac{m_l \mathbb{E}(Y_0^{k,l})^2}{8(1 + \mathbb{E}(\theta^{q_{k,l}}))^2} \right). \quad (14)$$

Proof. We will use Theorem 1 of Chazottes *et al.* (2007) to obtain an upper bound for the left-hand side of (14). Let $(Z_j)_{j \in \mathbb{Z}}$ be a canonical process on $\{-1, +1\}^{\mathbb{Z}}$ with law μ . Let $n \geq 1$, $\sigma \in \{-1, 1\}^n$, $(Z_j^{(+1\sigma)})_{j \geq 1}$ and $(Z_j^{(-1\sigma)})_{j \geq 1}$ be respectively the processes with laws defined by the conditional distributions $\mu((Z_j)_{j \geq i+1} = \cdot \mid Z_i = 1, Z_{i-1} =$

$\sigma_{i-1}, \dots, Z_1 = \sigma_1$) and $\mu((Z_j)_{j \geq i+1} = \cdot \mid Z_i = -1, Z_{i-1} = \sigma_{i-1}, \dots, Z_1 = \sigma_1)$. We denote by \mathbb{Q}_i^σ the maximal coupling between the conditional distributions. We introduce the upper-triangular matrix D^σ defined for $1 \leq i < j \leq n$ by

$$\begin{aligned} D_{i,i}^\sigma &:= 1 \\ D_{i,j}^\sigma &:= \mathbb{Q}_i^\sigma \left(Z_j^{(+1\sigma)} \neq Z_j^{(-1\sigma)} \right). \end{aligned}$$

Then, we define the matrix \bar{D} as $\bar{D}_{i,j} := \sup_{\sigma \in \{-1,1\}^n} D_{i,j}^\sigma$. For a given function $f : \{-1, 1\}^n \rightarrow \mathbb{R}$ we define the variation of f at site i with $1 \leq i \leq n$ by

$$\delta_i f := \sup_{\sigma_j = \sigma'_j, i \neq j} |f(\sigma) - f(\sigma')|.$$

Now, let $n \geq 1$ be arbitrary and assume that $\|\bar{D}\|_2 < \infty$ and $\|\delta f\|_2 < \infty$. Then, Theorem 1 in Chazottes *et al.* (2007) states that, for all functions $f : \{-1, 1\}^n \rightarrow \mathbb{R}$ and all $t > 0$, we have

$$\mu(|f - \mathbb{E}(f)| \geq t) \leq 2 \exp \left(-\frac{2t^2}{\|\bar{D}\|_2^2 \|\delta f\|_2^2} \right). \quad (15)$$

In our case, for the process $(Y_j^{k,l})_{j \in \mathbb{Z}}$ and measure \mathbb{P} , we observe that the elements of matrix $\bar{D}_{i,j}$ are bounded from above by the probabilities $\mathbb{P}(\theta^{q_{k,l}} > j)$ for all $j > i \geq 1$. To see this we note that

$$\begin{aligned} \mathbb{P}(\theta^{q_{k,l}} > j) &= \mathbb{P}(F^{q_{k,l}}(U_{-i}^0, \underline{+1}) \neq F^{q_{k,l}}(U_{-i}^0, \underline{-1}) \text{ for } i = 1, \dots, j) \\ &= \mathbb{P}(F^{q_{k,l}}(U_{-j}^0, \underline{+1}) \neq F^{q_{k,l}}(U_{-j}^0, \underline{-1})), \end{aligned}$$

where the last equality is a consequence of the attractiveness of q_r . Now, by the stationarity of \mathbf{U} we have

$$\begin{aligned} \mathbb{P}(\theta^{q_{k,l}} > j) &= \mathbb{P}\left(F_{[j,0]}^{q_{k,l}}(U_0^j, \underline{+1}) \neq F_{[j,0]}^{q_{k,l}}(U_0^j, \underline{-1})\right) \\ &\geq \bar{D}_{i,j}. \end{aligned}$$

By norm inequality, we obtain

$$\|\bar{D}\|_2 \leq \sum_{j=1}^n \bar{D}_{i,j}.$$

Taking $n = m_l$, we deduce that

$$\|\bar{D}\|_2^2 \leq \left(1 + \sum_{j=1}^{m_l} \mathbb{P}(\theta^{q_{k,l}} > j) \right)^2 \leq (1 + \mathbb{E}(\theta^{q_{k,l}}))^2. \quad (16)$$

Now, taking $f = f(x_1, \dots, x_{m_l}) = \frac{1}{m_l} \sum_{i=1}^{m_l} x_i$ we have $\delta_i f = \frac{2}{m_l}$ if $i \in \{1, \dots, m_l\}$. Thus, we obtain

$$\|\delta g\|_2^2 = \sum_{i=1}^{m_l} \left(\frac{2}{m_l}\right)^2 = \frac{4}{m_l}. \quad (17)$$

Applying (15) and using (16) and (17), we obtain

$$\mathbb{P} \left(\left| \frac{1}{m_l} \sum_{j=1}^{m_l} Y_j^{k,l} - \mathbb{E}(Y_0^{k,l}) \right| \geq \frac{\mathbb{E}(Y_0^{k,l})}{2} \right) \leq 2 \exp \left(-\frac{m_l \mathbb{E}(Y_0^{k,l})^2}{8(1 + \mathbb{E}(\theta^{q_{k,l}}))^2} \right).$$

□

Proof of Theorem 1.

We fix a sequence $(\lambda_j)_{j \geq 1}$ of positive real numbers such that $\sum_{j \geq 1} \lambda_j = 1$. Let $r : \{1, 2, \dots\} \rightarrow \mathbb{Z}_+$ such that $r_k < k$ and $\sum_{j \geq k+1} \lambda_j > \sum_{j=r_k+1}^k \lambda_j, \forall k \geq 1$. The sequence of odd positive integer numbers $(m_j)_{j \geq 1}$ will be chosen afterwards.

Clearly $(p_j)_{j \geq 1}$ and $(p'_j)_{j \geq 1}$ defined in (7) and (8) converge to the Bramson-Kalikow kernel p in $\mathcal{C}(\mathcal{X})$. For all $k \geq 0$, let μ_k (resp. μ'_k) be the unique stationary measure compatible with p_k (resp. p'_k). Observe that for $k = 0$, μ_0 (resp. μ'_0) is a Bernoulli process of parameter $1 - \epsilon$ (resp. ϵ).

We will apply Theorem 2 with $k = 0$. Since $\bar{d}(\mu_0, \mu'_0) = 1 - 2\epsilon$, we need to find an explicit sequence $(m_j)_{j \geq 1}$ such that

$$\sum_{k \geq 0} \bar{d}(\mu_k, \mu_{k+1}) + \sum_{k \geq 0} \bar{d}(\mu'_k, \mu'_{k+1}) < 1 - 2\epsilon. \quad (18)$$

By symmetry of the kernels p_k and p'_k , (18) is equivalent to

$$2 \sum_{k \geq 0} \bar{d}(\mu_k, \mu_{k+1}) < 1 - 2\epsilon. \quad (19)$$

Now, our task is to upper bound $\bar{d}(\mu_k, \mu_{k+1})$. For all $k \geq 0$, let $(X_j^k)_{j \in \mathbb{Z}}$ be the stationary process compatible with the measure μ_k . By definition of the \bar{d} -distance we have that

$$\bar{d}(\mu_k, \mu_{k+1}) \leq \mathbb{P}(X_0^k \neq X_0^{k+1}), \quad (20)$$

where \mathbb{P} is the coupling defined in Section 3. Define for all $i \in \mathbb{Z}_-$, the interval $H_i := [i - m_{k+1}, i - 1]$ and the events

$$S_i := \left\{ \sum_{j \in H_i} X_j^{k+1} > 0 \right\}.$$

As η_k (defined as in Lemma 1) is a stopping time for the filtration $(\mathcal{F}_i)_{i \geq 0} = (\sigma(U_0, U_{-1}, \dots, U_{-i}))_{i \geq 0}$ and the events S_i are independent of \mathcal{F}_i for all $i \geq 0$, we have by construction of the coupling \mathbb{P} and Wald's equality

$$\mathbb{P}(X_0^k \neq X_0^{k+1}) = \mathbb{P}\left(\bigcup_{i=0}^{\eta_k} S_i^c\right) \leq \mathbb{E}\left(\sum_{i=0}^{\eta_k} \mathbf{1}_{\{(S_i)^c\}}\right) = (\mathbb{E}(\eta_k) + 1)\mathbb{P}(S_0^c). \quad (21)$$

Combining (20) and (21) we obtain

$$\bar{d}(\mu_k, \mu_{k+1}) \leq (\mathbb{E}(\eta_k) + 1)\mathbb{P}(S_0^c) \quad (22)$$

for all $k \geq 0$.

To obtain an upper bound for $\mathbb{E}(\eta_k)$ we use Lemma 1; for $\mathbb{P}(S_0^c)$ we proceed as follows. Let $r := r_{k+1}$ and $(Y_j^{r, k+1})_{j \in \mathbb{Z}}$ be the process compatible with $q_{r, k+1}$. Observe that for all $k \geq r \geq 0$ we have $q_{r, k+1} \in \mathcal{M}_r$. Also note that, for any $n \geq 1$, and integers l_1, \dots, l_n , we have by construction that

$$\mathbb{P}\left(\bigcup_{j=1}^n \{X_{l_j}^{k+1} < Y_{l_j}^{r, k+1}\}\right) = 0,$$

and therefore

$$\mathbb{P}\left(\sum_{j=1}^{m_{k+1}} X_j^{k+1} < 0\right) \leq \mathbb{P}\left(\sum_{j=1}^{m_{k+1}} Y_j^{r, k+1} < 0\right). \quad (23)$$

Furthermore, we have

$$\mathbb{P}\left(\sum_{j=1}^{m_{k+1}} Y_j^{r, k+1} < 0\right) \leq \mathbb{P}\left(\left|\frac{1}{m_{k+1}} \sum_{j=1}^{m_{k+1}} Y_j^{r, k+1} - \mathbb{E}(Y_0^{r, k+1})\right| \geq \frac{\mathbb{E}(Y_0^{r, k+1})}{2}\right), \quad (24)$$

and therefore we can upper bound $\mathbb{P}(S_0^c)$ using a concentration of measure inequality for a Markov chain of order $r < k + 1$.

Combining (22), (23), (24), and Lemmas 2 and 3, we deduce that, for all $k \geq 0$,

$$\bar{d}(\mu_k, \mu_{k+1}) \leq 2(\mathbb{E}(\eta_k) + 1) \exp\left(-\frac{m_{k+1} \left(\sum_{j \geq k+2} \lambda_j - \sum_{j=r+1}^{k+1} \lambda_j\right)^2 (1 - 2\epsilon)^2}{8 \left(1 + \mathbb{E}(\theta^{q_{r, k+1}})\right)^2}\right).$$

Let $\alpha > 0$ such that $\alpha < \frac{1}{2} - \epsilon$. Define

$$A_0 := 8(1 - 2\epsilon)^{-2} \ln(4\alpha^{-1})$$

and for all $k \geq 1$,

$$A_k := 8(1 - 2\epsilon)^{-2} (1 + m_r(2\epsilon)^{-m_r})^2 \ln\left(2^{k+2}(1 + m_k(2\epsilon)^{-m_k})\alpha^{-1}\right).$$

Then, for all $k \geq 0$ choose m_{k+1} as the first odd integer such that

$$m_{k+1} \geq \frac{A_k}{\left(\sum_{j \geq k+2} \lambda_j - \sum_{j=r+1}^{k+1} \lambda_j\right)^2}. \quad (25)$$

With these choices, using Lemma 1 we obtain

$$\bar{d}(\mu_k, \mu_{k+1}) \leq \frac{\alpha}{2^{k+1}}$$

for all $k \geq 0$. Since $\alpha < \frac{1}{2} - \epsilon$ we obtain (19), which proves the theorem. \square

Proof of Corollary 1.

If, for $j \geq 1$, we choose $\lambda_j = \frac{1}{2} \left(\frac{2}{3}\right)^j$, we have for $k \geq 1$, $\sum_{j \geq k+1} \lambda_j - \lambda_k \geq 0$, *i.e.*, we have a function r in (4) defined by $r_k = k - 1$. Let $\epsilon = 1/4$ and $\alpha = 1/8$, then $A_0 = 160 \ln 2$ and, by (25), m_1 must be chosen greater than $320 \left(\frac{3}{2}\right)^2 \ln 2 \approx 216, 74$. Let us take $m_1 = 217$. Now, from (4), we can see that in this case, the sequence $(m_k)_{k \geq 1}$ must satisfy $m_k \geq k + 1$ for all $k \geq 1$. Therefore, for $k \geq 1$, we have

$$\begin{aligned} \frac{A_k}{\left(\sum_{j \geq k+2} \lambda_j - \lambda_{k+1}\right)^2} &\leq 512 \left(\frac{9}{2}\right)^{k+1} (1 + m_k 2^{m_k})^3 \\ &\leq 512 \left(\frac{9}{2}\right)^{k+1} (64)^{m_k} \\ &\leq (577)^{m_k}. \end{aligned}$$

\square

Proof of Corollary 2.

Let $b_1 = 1$ and c a positive integer to be fixed afterwards. For $l \geq 2$, we define $b_l = \left\lceil 2^{(c \sum_{j=1}^{l-1} b_j)^2} \right\rceil$, where $\lceil \cdot \rceil$ is the ceiling function. Let $s = 3/4$, for $l \geq 1$ and $j \in \{\sum_{k=1}^{l-1} b_k + 1, \dots, \sum_{k=1}^l b_k\}$ we define

$$\lambda_j = \frac{s^{l-1} - s^l}{b_l}.$$

It is straightforward to verify that $\sum_{j \geq 1} \lambda_j = 1$. Let $r_k = \left\lfloor \sqrt{\log(k)/c} \right\rfloor$ where \log is base 2 logarithm and $\lfloor \cdot \rfloor$ is the floor function. We observe that by construction, for

$l \geq 1$, we have

$$\sum_{j \geq k+1} \lambda_j - \sum_{j=r(k)+1}^k \lambda_j \geq s^l - (s^{l-2} \wedge 1 - s^l) = \frac{1}{8} \left(\frac{3}{4} \right)^{l-2}.$$

We set $m_j = \lfloor 2^{cj^2} \rfloor$ if $\lfloor 2^{cj^2} \rfloor$ is odd, otherwise $m_j = \lfloor 2^{cj^2} \rfloor - 1$. We want to obtain a sequence $(m_j)_{j \geq 1}$ that satisfies (4) and (5). Let

$$B_k = 4(k+1)^2 2^{2(k+1)(\log(2\epsilon)^{-1})}.$$

We have the following upper bound for (5):

$$A_k \leq \frac{8(k+2) \ln \frac{2}{\alpha}}{(1-2\epsilon)^2} B_k + \frac{8 \ln \left(1 + 2^{ck^2} (2\epsilon)^{-2ck^2} \right)}{(1-2\epsilon)^2} B_k.$$

Now, taking $\epsilon = 1/4$ and $\alpha = 1/8$, we have,

$$\begin{aligned} A_k &\leq 128 \ln 2 B_k (k+2) + 32 \ln 2 B_k (ck^2 + 1) + 32 B_k 2^{ck^2} \\ &\leq 81 B_k 2^{ck^2}. \end{aligned}$$

Also, we observe that for $c \geq 8$, $B_k \leq 2^{2+2\log(k+1)+2(k+1)} \leq 2^{ck}$, and, therefore,

$$A_k \leq 81 \cdot 2^{c(k^2+k)}.$$

Also for $c \geq 2$, we have $\frac{1}{8} \left(\frac{3}{4} \right)^{l-2} \geq \left(\frac{1}{2} \right)^{l+1} \geq 2^{-ck}$. Finally, to satisfy the conditions in Theorem 2, it is enough that

$$2^{c(k+1)^2} \geq 81 \cdot 2^{c(k^2+2k)}.$$

The above inequality is satisfied if $c \geq 8$. □

5. PROOFS OF THEOREMS 2 AND 3

Proof of Theorem 2.

We proceed in three main steps. First, we prove the existence of a subsequence $(\mu_{v_j})_{j \geq 0}$ that converges in \bar{d} to a measure μ compatible with g . The same naturally holds for a subsequence $(\mu'_{u_j})_{j \geq 0}$ and a measure μ' compatible with g . Then we prove that under the conditions of the theorem μ and μ' are actually distinct. The statement about bernoullicity then follows directly from the well-known fact that \bar{d} -limit of regular Markov chains are Bernoulli.

For the first step, we will prove that there exists a subsequence $(\mu_{v_j})_{j \geq 0}$ converging weakly and in entropy to a measure μ compatible with g . Because regular Markov processes are finitely determined and, for this class of processes, the weak

convergence and the convergence of the entropy together imply convergence in \bar{d} , we conclude that μ_{v_j} converges to μ in \bar{d} -distance (see definition in p.221 and Theorem IV.2.9 of Shields (1996)).

We consider processes on finite alphabet, therefore the space of respective probability measures endowed with the weak topology is compact. Hence, for any sequence $(\mu_j)_{j \geq 0}$ there exists a convergent subsequence $(\mu_{v_j})_{j \geq 0}$. Let μ be its weak limit. From the weak convergence of μ_{v_j} to μ and the convergence of g_{v_j} to g in $\mathcal{C}(\mathcal{X})$, it is immediate that μ is a g -measure.

We observe that the entropy $H(\mu_{v_j})$ of an ergodic Markov process μ_{v_j} can be written as

$$H(\mu_{v_j}) = - \int_{\mathcal{X}} \log g_{v_j} d\mu_{v_j}.$$

A standard computation shows that the entropy $H(\mu)$ of μ is given by

$$H(\mu) = - \int_{\mathcal{X}} \log g d\mu.$$

We note that $g \in \mathcal{G}$ and therefore $\log g \in \mathcal{C}(\mathcal{X})$. Again, because $\mu_{v_j} \rightarrow \mu$ weakly and $\log g_{v_j}$ converges to $\log g$ in $\mathcal{C}(\mathcal{X})$ we have that

$$\int_{\mathcal{X}} \log g_{v_j} d\mu_{v_j} \rightarrow \int_{\mathcal{X}} \log g d\mu$$

Thus, we conclude that μ_{v_j} converges in \bar{d} to μ .

We now come to the second step, and prove that the limits μ and μ' are distinct. Taking $v_0 = k$, we have

$$\lim_{j \rightarrow \infty} \bar{d}(\mu_k, \mu_{v_j}) = \bar{d}(\mu_k, \mu).$$

We also have

$$\bar{d}(\mu_k, \mu_{v_j}) \leq \sum_{j=k}^{\infty} \bar{d}(\mu_j, \mu_{j+1})$$

and, therefore,

$$\bar{d}(\mu_k, \mu) \leq \sum_{j=k}^{\infty} \bar{d}(\mu_j, \mu_{j+1}).$$

Similarly,

$$\bar{d}(\mu'_k, \mu') \leq \sum_{j=k}^{\infty} \bar{d}(\mu'_j, \mu'_{j+1}).$$

Thus, if (6) is satisfied, we have

$$\bar{d}(\mu_k, \mu) + \bar{d}(\mu'_k, \mu') < \bar{d}(\mu_k, \mu'_k)$$

showing that there exist two distinct g -measures μ and μ' . □

Proof of Theorem 3.

The proof follows from the properties of attractive kernels $g \in \mathcal{G}$ and associated g -measures described in Hulse (2006, 1991).

For $f \in \mathcal{C}(\mathcal{X})$, $g \in \mathcal{G}$, and all $x \in \mathcal{X}$, we define the *Ruelle operator* L_g by

$$L_g f(x) = \sum_{a \in A} g(ax) f(ax).$$

Recall that $\underline{+1}, \underline{-1} \in \mathcal{X}$ are defined by $\underline{+1}_i = 1$ and $\underline{-1}_i = -1$ for $i \leq 0$. By Lemma 2.1 in Hulse (1991), if g is attractive, for $f \in \mathcal{C}(\mathcal{X})$, we have

$$\lim_{n \rightarrow \infty} L_g^n f(\underline{+1}) = \int_{\mathcal{X}} f d\mu^+, \quad (26)$$

$$\lim_{n \rightarrow \infty} L_g^n f(\underline{-1}) = \int_{\mathcal{X}} f d\mu^-,$$

where μ^+ and μ^- are extremal g -measures. If $\mu^+ = \mu^-$ we have a unique g -measure. Let h_1 and h_2 be elements of \mathcal{G} and $x, y \in \mathcal{X}$. We say that h_1 *dominates* h_2 if for all $x \geq y$ we have $h_1(1x) \geq h_2(1y)$. From Hulse (2006) (p.442) if h_1 dominates h_2 , for any increasing function $f \in \mathcal{C}(\mathcal{X})$, $x \geq y$, and $n \geq 1$ we have

$$L_{h_1}^n f(x) \geq L_{h_2}^n f(y).$$

We start with the following lemma.

Lemma 4. *Let h_1, h_2 be attractive and h_1 dominates h_2 . If ν_1^+ and ν_2^+ are the extremal h_1 and h_2 -measures defined respectively by iterating L_{h_1} and L_{h_2} as in (26), we have that*

$$\bar{d}(\nu_1^+, \nu_2^+) = \nu_1^+(\{x \in \mathcal{X} : x_0 = 1\}) - \nu_2^+(\{x \in \mathcal{X} : x_0 = 1\}).$$

Proof. As in Hulse (1991), for $x, y \in \mathcal{X}$, $a, b \in A$, and $h_1, h_2 \in \mathcal{C}(\mathcal{X})$, we define the kernel $P : \mathcal{X} \times \mathcal{X} \rightarrow [0, 1]$ by

$$P(ax, by) = \begin{cases} \min \{h_1(ax), h_2(by)\} & \text{if } a = b \\ \max \{h_1(ax) - h_2(by), 0\} & \text{otherwise,} \end{cases}$$

and $\sum_{a \in A} \sum_{b \in A} P(ax, by) = 1$. Let $f \in \mathcal{C}(\mathcal{X})$.

We define ν_1^+ and ν_2^+ by $\lim_{n \rightarrow \infty} L_{h_1}^n f(\underline{+1}) = \int_{\mathcal{X}} f d\nu_1^+$ and $\lim_{n \rightarrow \infty} L_{h_2}^n f(\underline{+1}) = \int_{\mathcal{X}} f d\nu_2^+$, respectively. If h_1 and h_2 are attractive and h_1 dominates h_2 , we can use P to define a coupling between ν_1^+ and ν_2^+ . To see this, let $f_1, f_2 \in \mathcal{C}(\mathcal{X})$ and $x, y \in \mathcal{X}$. We introduce the Ruelle operator L_P as

$$L_P(f_1 \otimes f_2)(x, y) = \sum_{a \in A} \sum_{b \in A} P(ax, by) f_1(ax) f_2(by).$$

For any functions $f_1, f_2 \in \mathcal{C}(\mathcal{X})$, we have that $\lim_{n \rightarrow \infty} L_P^n(f_1 \otimes f_2)(+\underline{1}, +\underline{1})$ exists and defines a coupling ν between ν_1^+ and ν_2^+ (Hulse, 1991). By construction and definition of P , this coupling has the property that $\nu(\{(x, y) \in \mathcal{X} \times \mathcal{X} : x_0 < y_0\}) = 0$. This implies that

$$\begin{aligned} \bar{d}(\nu_1, \nu_2) &\leq \nu(\{(x, y) \in \mathcal{X} \times \mathcal{X} : x_0 \neq y_0\}) \\ &= \nu_1(\{x \in \mathcal{X} : x_0 = 1\}) - \nu_2(\{x \in \mathcal{X} : x_0 = 1\}). \end{aligned}$$

Moreover, we also have by definition of \bar{d} -distance that

$$\bar{d}(\nu_1^+, \nu_2^+) \geq \nu_1^+(\{x \in \mathcal{X} : x_0 = 1\}) - \nu_2^+(\{x \in \mathcal{X} : x_0 = 1\}),$$

which implies that

$$\bar{d}(\nu_1^+, \nu_2^+) = \nu_1^+(\{x \in \mathcal{X} : x_0 = 1\}) - \nu_2^+(\{x \in \mathcal{X} : x_0 = 1\}).$$

□

We introduce a sequence of kernels $g_j, g'_j \in \mathcal{M}$ for each $j \geq 1$ and $x \in \mathcal{X}$ by

$$g_j(1x_{-\infty}^{-1}) = \sup_{y \in \mathcal{X}} g(1x_{-j}^{-1}y),$$

and

$$g'_j(1x_{-\infty}^{-1}) = \inf_{y \in \mathcal{X}} g(1x_{-j}^{-1}y).$$

For $j = 0$ we define $g_0(1x_{-\infty}^{-1}) = \sup_{y \in \mathcal{X}} g(1y)$ and $g'_0(1x_{-\infty}^{-1}) = \inf_{y \in \mathcal{X}} g(1y)$. Observe that if g is attractive, g_j and g'_j are also attractive. Moreover, for all $j \geq 0$, g_j dominates g and g dominates g'_j .

Let μ_j be the unique g -measure of g_j . From Lemma 4

$$\bar{d}(\mu_j, \mu^+) = \mu_j(\{x \in \mathcal{X} : x_0 = 1\}) - \mu^+(\{x \in \mathcal{X} : x_0 = 1\}).$$

Now, by definition of μ^+ , for any g -measure μ we have

$$\mu^+(\{x \in \mathcal{X} : x_0 = 1\}) - \mu(\{x \in \mathcal{X} : x_0 = 1\}) \geq 0. \quad (27)$$

Let μ_{v_j} be any subsequence converging weakly to some g -measure. Because, for all $j \geq 1$, g_{v_j} dominates g , we have

$$\lim_{j \rightarrow \infty} \mu_{v_j}(\{x \in \mathcal{X} : x_0 = 1\}) - \mu^+(\{x \in \mathcal{X} : x_0 = 1\}) \geq 0.$$

The above equation together with (27) implies that

$$\lim_{j \rightarrow \infty} \mu_{v_j}(\{x \in \mathcal{X} : x_0 = 1\}) = \mu^+(\{x \in \mathcal{X} : x_0 = 1\}).$$

As this holds for any subsequence, we have that

$$\lim_{j \rightarrow \infty} \mu_j(\{x \in \mathcal{X} : x_0 = 1\}) = \mu^+(\{x \in \mathcal{X} : x_0 = 1\}).$$

The above equation and Lemma 4 imply

$$\lim_{j \rightarrow \infty} \bar{d}(\mu_j, \mu^+) = 0,$$

and we conclude that

$$\sum_{j=k}^{\infty} \bar{d}(\mu_j, \mu_{j+1}) = \mu_k(\{x \in \mathcal{X} : x_0 = 1\}) - \mu^+(\{x \in \mathcal{X} : x_0 = 1\}). \quad (28)$$

Now, let μ'_j be the g'_j measures. Repeating again the above arguments, we have that

$$\sum_{j=k}^{\infty} \bar{d}(\mu'_j, \mu'_{j+1}) = \mu^-(\{x \in \mathcal{X} : x_0 = 1\}) - \mu'_k(\{x \in \mathcal{X} : x_0 = 1\}). \quad (29)$$

Using Lemma 4 again, we have

$$\bar{d}(\mu_k, \mu'_k) = \mu_k(\{x \in \mathcal{X} : x_0 = 1\}) - \mu'_k(\{x \in \mathcal{X} : x_0 = 1\}). \quad (30)$$

Combining (28), (29), and (30) we have that inequality

$$\sum_{j=k}^{\infty} \bar{d}(\mu_j, \mu_{j+1}) + \sum_{j=k}^{\infty} \bar{d}(\mu'_j, \mu'_{j+1}) < \bar{d}(\mu_k, \mu'_k)$$

is equivalent to

$$\mu^+(\{x \in \mathcal{X} : x_0 = 1\}) - \mu^-(\{x \in \mathcal{X} : x_0 = 1\}) > 0. \quad (31)$$

Finally, from Theorem 2.2 in (Hulse, 1991), we have that inequality (31) holds if and only if there are several g -measures. \square

REFERENCES

- BERGER, N., HOFFMAN, C. & SIDORAVICIUS, V. (2005). Nonuniqueness for specifications in $\ell^{2+\epsilon}$. *arXiv:math/0312344*.
- BRAMSON, M. & KALIKOW, S. (1993). Nonuniqueness in g -functions. *Israel J. Math.* **84**(1-2), 153–160.
- CHAZOTTES, J.-R., COLLET, P., KÜLSKE, C. & REDIG, F. (2007). Concentration inequalities for random fields via coupling. *Probab. Theory Related Fields* **137**(1-2), 201–225.
- COMETS, F., FERNÁNDEZ, R. & FERRARI, P. A. (2002). Processes with long memory: regenerative construction and perfect simulation. *Ann. Appl. Probab.* **12**(3), 921–943.
- DE SANTIS, E. & PICCIONI, M. (2012). Backward coalescence times for perfect simulation of chains with infinite memory. *J. Appl. Probab.* **49**(2), 319–337.
- DOEBLIN, W. & FORTET, R. (1937). Sur des chaînes à liaisons complètes. *Bull. Soc. Math. France* **65**, 132–148.

- FERNÁNDEZ, R., GALLO, S. & MAILLARD, G. (2011). Regular g -measures are not always Gibbsian. *Electron. Commun. Probab.* **16**, 732–740.
- FERNÁNDEZ, R. & MAILLARD, G. (2005). Chains with complete connections: general theory, uniqueness, loss of memory and mixing properties. *J. Stat. Phys.* **118**(3-4), 555–588.
- FRIEDLI, S. (2010). A note on the Bramson-Kalikow process. *Preprint* , <http://www.mat.ufmg.br/sacha/textos/BK/plateaux.pdf>.
- GALLO, S. & PACCAUT, F. (2013). On non-regular g -measures. *Nonlinearity* **26**(3), 763–776.
- GALLO, S. & TAKAHASHI, D. (2013). Attractive regular stochastic chains: perfect simulation and phase transition. *Ergodic Theory and Dynamical System* , FirstView Article, 1–20, <http://dx.doi.org/10.1017/etds.2013.7>.
- HARRIS, T. E. (1955). On chains of infinite order. *Pacific J. Math.* **5**, 707–724.
- HULSE, P. (1991). Uniqueness and ergodic properties of attractive g -measures. *Ergodic Theory Dynam. Systems* **11**(1), 65–77.
- HULSE, P. (2006). An example of non-unique g -measures. *Ergodic Theory Dynam. Systems* **26**(2), 439–445.
- JOHANSSON, A., ÖBERG, A. & POLLICOTT, M. (2012). Unique bernoulli g -measures. *Journal of the European Mathematical Society* **14**, 1599–1615.
- KEANE, M. (1972). Strongly mixing g -measures. *Invent. Math.* **16**, 309–324.
- LACROIX, Y. (2000). A note on weak- \star perturbations of g -measures. *Sankhyā Ser. A* **62**(3), 331–338.
- LALLEY, S. P. (1986). Regenerative representation for one-dimensional Gibbs states. *Ann. Probab.* **14**(4), 1262–1271.
- ONICESCU, O. & MIHOC, G. (1935). Sur les chaînes de variables statistiques. *Bull. Sci. Math* **59**(2), 174–192.
- PROPP, J. G. & WILSON, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. In: *Proceedings of the Seventh International Conference on Random Structures and Algorithms (Atlanta, GA, 1995)*, vol. 9.
- SHIELDS, P. C. (1996). *The ergodic theory of discrete sample paths*, vol. 13 of *Graduate Studies in Mathematics*. Providence, RI: American Mathematical Society.
- STENFLO, Ö. (2003). Uniqueness in g -measures. *Nonlinearity* **16**(2), 403–410.
- WALTERS, P. (1975). Ruelle’s operator theorem and g -measures. *Transactions of American Mathematical Society* **214**, 375–387.

¹DEPARTAMENTO DE ESTATÍSTICA, INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E CIÊNCIA DE COMPUTAÇÃO, UNIVERSIDADE DE CAMPINAS, BRASIL

E-mail address: gallesco@ime.unicamp.br

²DEPARTAMENTO DE MÉTODOS ESTATÍSTICOS, INSTITUTO DE MATEMÁTICA, UNIVERSIDADE FEDERAL DE RIO DE JANEIRO, BRASIL.

E-mail address: sandro@im.ufrj.br

³NEUROSCIENCE INSTITUTE AND PSYCHOLOGY DEPARTMENT, PRINCETON UNIVERSITY, USA.

E-mail address: takahashi@d@gmail.com