THE WEAK BERNOULLI PROPERTY AND (POSSIBLY) FACTORS OF g-MEASURES

ANDERS JOHANSSON AND ANDERS ÖBERG

Abstract.

1. Introduction

Let S be a countable set. Let $\mathbb{Z}_+ = \{0, 1, 2, \ldots\}$, $\mathbb{Z} = \{\ldots, -1, 0, 1, 2 \ldots\}$, $X = S^{\mathbb{Z}}$, $X_+ = S^{\mathbb{Z}_+}$ and $X_- = S^{\mathbb{Z} \setminus \mathbb{Z}_+}$. Any bi-infinite sequence $x \in X$ and $n \in \mathbb{Z}$, gives a one-sided infinite sequence $x^{(n)} = (x_{-n}, x_{-n+1}, \ldots)$ in X_+. Moreover, the stochastic process $\{x^{(n)}\}_{n \in \mathbb{Z}}$ has the Markov property for any distribution of x in $\mathcal{M}(X)$, where $\mathcal{M}(X)$ denotes the Borel probability measures on X, with respect to the product topology on X.

Let $g \geq 0$ be a continuous function on X_+ such that

$$\sum_{x_0 \in S} g(x_0 x) = 1, \; x \in X_+. \tag{1.1}$$

A distribution $\mu \in \mathcal{M}(X)$ of $x \in X$ is a g-*chain* if

$$\mu\left(x^{(n)} | x^{(n-1)}\right) = g\left(x^{(n)}\right) \tag{1.2}$$

for all $n \geq 0$. Thus, the process depends on the past according to the g-function. Note that the distribution of a g-chain is uniquely determined by the distribution $\mu \circ (x^{(0)})^{-1} \in \mathcal{M}(X_+)$ of its "initial" value $x^{(0)}$.

If g depends only on the choice of the new state then we have an *i.i.d.* process, and if g depends on the new state and the previous one, then we have a Markov chain on the countable set S. If we have dependence on the k previous states, before moving to the new state, we have a k-chain, and if there is no such restriction on the dependence, we have a chain with infinite connections.

A stationary measure for our process is sometimes called a g-measure (see Keane [17]). If T is the left shift map on X_+, then a g-measure can alternatively be

Date: February 3, 2012.

2000 *Mathematics Subject Classification.* Primary 37A05, 37A35, 60G10.

Key words and phrases. Bernoulli properties, g-measures, mixing, probability metrics, chains with infinite connections.

viewed as T-invariant probability measure $\mu \in \mathcal{M}(X_+)$, with the property that $d\mu/d(\mu \circ T) = g$, for a continuous function g. For a given g-function g, there always exists a g-measure, since X_+ is compact due to the finiteness of S. Uniqueness is however not automatic, as was clarified by Bramson and Kalikow in [5]. Examples of non-uniqueness have since then been provided in, e.g., [4] and [11].

A useful way of viewing a g-measure is as a fixed point of the dual $\mathcal{L}^*$ of the transfer operator $\mathcal{L}$, defined pointwise by

$$\mathcal{L}f(x) = \sum_{Ty=x} g(y)f(y),$$

where $\mathcal{L} : C(X_+) \to C(X_+)$. Hence, a g-measure can be viewed as a probability measure satisfying $\mathcal{L}^*\mu = \mu$.

If we do not impose the probability assumption (1.1), the eigen-measure of the dual of the transfer operator is not invariant in general, but may instead look for eigen-measure solutions ν of $\mathcal{L}^*\nu = \lambda\nu$, where $\lambda > 0$ is the greatest eigenvalue of the unrestricted transfer operator $\mathcal{L}$,

$$\mathcal{L}f(x) = \sum_{Ty=x} e^{\phi}(y)f(y),$$

where ϕ is the potential function, usually belonging to a function space with the same regularity conditions as the test functions f.

In this paper our results only concern the case of probabilistic weight functions, that is $\phi = \log g$, with $\sum_{Ty=x} g(y) = 1$, for all x. In [13], it was proved that there exists a unique g-measure if $g > 0$ and

$$\sum_{n=1}^{\infty} (\mathrm{var}_n \log g)^2 < \infty, \tag{1.3}$$

where the nth variation of a function f is defined as

$$\mathrm{var}_n f = \sup_{x \sim_n y} |f(x) - f(y)|,$$

where $x \sim_n y$ means that x and y coincide in the first n coordinates.

This condition of *square summability of variations* of the g-function for the g-chain is proven [4] to be sharp, in the sense that for all $\epsilon > 0$ there exists a g-function such that

$$\sum_{n=1}^{\infty} (\mathrm{var}_n \log g)^{2+\epsilon} < \infty,$$

with more than one g-measure. This should be compared to an older result of Dyson [8] for general potentials ϕ, identifying *summability of variations* as sharp, in the sense that we may have multiple eigen-measure solutions of $\mathcal{L}\nu = \lambda\nu$, when

$$\sum_{n=1}^{\infty} (\mathrm{var}_n\, \phi)^{1+\epsilon} < \infty.$$

In view of this case against uniqueness, Berbee's two results from the late 1980s are intrigueing. He proves uniqueness of a g-measure and of an eigen-measure in the general case, when

$$\sum_{n=1}^{\infty} e^{-r_1-\cdots-r_n} = \infty, \tag{1.4}$$

where $r_n = \mathrm{var}_n \log g$ or $r_n = \mathrm{var}_n\, \phi$, respectively. This allows for the non-summable sequence $r_n = \frac{1}{n}$. In the case of general potentials this is sharp, modulo a constant factor, see [1], but obviously not for g-measures, since square summability of variations cover sequences $r_n = \frac{1}{n^{1/2+\epsilon}}$, $\epsilon > 0$.

Since it was shown in [13] that there are sequences that satisfy Berbee's condition but not square summability, it becomes interesting in the case of proving uniqueness of a g-measure to ask if there is a condition that subsumes in a natural way these two uniqueness conditions. We provide conditions for uniqueness that contains both square summability of variations and Berbee's condition for a unique g-measure.

Our method of proof allows us to conclude that the unique g-measure is *Bernoulli*, meaning that if we look at the *natural extension* of the dynamical system, i.e.,

$$x^{(n)} = (x_{-n}, x_{-n+1}, \ldots),$$

$n \geq 0$, with the g-measure μ as initial distribution for $x^{(0)}$, then this stochastic process is isomorphic to an *i.i.d.* process.

The Bernoulli property was also proved by Berbee, but is new for square summability of variations (convergence for the iterates of the transfer operator is known from [14]). For instance we prove that we have a *unique g-measure* that is furthermore *Bernoulli* under the following three special conditions.

(1)
$$\sum_{n=1}^{\infty} (\mathrm{var}_n \log g)^2 < \infty,$$

(2) For any fixed $\epsilon > 0$,
$$\sum_{n=1}^{\infty} e^{-\left(\frac{1}{2}+\epsilon\right)(r_1+\cdots+r_n)} = \infty,$$

(3)
$$\mathrm{var}_n \log g = o\left(\frac{1}{\sqrt{n}}\right), \quad n \to \infty.$$

The last example is in a sense the weakest condition we have for a unique Bernoulli g-measure. The second is an improvement of Berbee's condition with a constant, owing to our method. For other results concerning the Bernoulli property for g-measures and equilibrium states for general potentials, see [?].

It would be interesting to investigate whether there is a sharp constant so that we have uniquness and perhaps the Bernoulli property for $\mathrm{var}_n \log g \leq \frac{c}{\sqrt{n}}$. Perhaps the $\leq$ should be replaced by a $<$ and perhaps the constants are different for uniqueness and for the Bernoulli property.

Our method of proof relies on two main ideas.

Firstly, we use a forward block coupling, including solving the renewal equation to obtain an estimate of the probability of having conflicts between two extensions of a g-chain, starting from two different distributions. This argument is then applied to a perturbation of one of the extensions to a sequence of g-functions corresponding to a sequence of Bernoulli measures that converges in the $\bar{d}$-metric to the unique g-measure under investigation.

Secondly, we use Hellinger integral estimates from [12] to calculate the probability of not having a conflict in the extensions of two initial distributions when we add a new block of positive integer length b_l (at a certain height $l \geq 1$ in the extension). We show that if these probabilities are $e^{-\rho_l}$, the minimal probability of a conflict, as defined through the total variations distance, then we can approximate ρ_l in such a way that it asymptotically includes a square sum of the variations, where the sums are taken over the increasing blocks. More precisely, if we define recursively $B_l = B_{l-1} + b_l$, $l \geq 1$, $B_0 = 0$, we get the estimate

$$\rho_l \leq (1 + o(1))s_l,$$

where

$$s_l := \sum_{k=B_{l-1}}^{B_l - 1} \frac{1}{8} (\mathrm{var}_k \log g)^2.$$

We also define

$$r_l = \sqrt{2s_l} + 2s_l.$$

In the special cases (1) and (3) above, we have only found examples of exponential increase of b_l in l. If $b_l = 1$ for all $l \geq 1$, we obtain Berbee's situation, in which case $\rho_l \leq r_l = \mathrm{var}_l \log g$. However our estimates show that although this is of the right order, one can improve Berbee's result by a constant; essentially, the variations are allowed to be twice as big.

We can now also state one version of our main result.

Theorem 1.1. *We obtain a unique g-measure which is Bernoulli, if there is a sequence of positive integers $\{b_l\}_{l=1}^{\infty}$ such that, with $\{r_l\}$ defined from $\{b_l\}$ as above,*

$\limsup r_l = 0$ *and*

$$\sum_{l=1}^{\infty} b_l \, e^{-r_1 - \cdots - r_l} = \infty.$$

Acknowledgement. We would like to thank Jean-Pierre Conze for valuable discussions.

2. Preliminaries

In this section we give an overview of the main ideas and the basic notation.

Let $\mathcal{M}^g(X) \subset \mathcal{M}(X)$ denote the set of g-chains corresponding to the g-function g, i.e. the set of μ such that

$$\mu \circ (x^{(n)})^{-1} = \mathcal{L}^{*n}[\mu \circ (x^{(0)})^{-1}].$$

Let $\mathcal{M}^g_T(X)$ denote the set of g-measures.

On $\mathcal{M}(X_+)$ we have the natural filtration $\{\mathcal{F}_n\}$ of the Borel σ-algebra, where $\mathcal{F}_n = \sigma(x_0, \ldots, x_{n-1})$. For a measure $\nu \in \mathcal{M}(X_+)$ and a sub σ-algebra $\mathcal{B} \subset \mathcal{F}$, we let $\nu|_{\mathcal{B}}$ denote the restriction to $\mathcal{B}$.

For $(y, \hat{y}) \in X_+ \times X_+$, define the concordance time as the non-negative integer

$$\kappa(y, \hat{y}) = \sup\{k \geq 0 : y \sim_k \hat{y}\}.$$

2.1. **The Bernoulli property and the $\bar{d}$-metric.** Recall that *coupling* (or *joining*) between two probability distributions $\mu \in \mathcal{M}(X, \mathcal{F})$ and $\hat{\mu} \in \mathcal{M}(\hat{X}, \hat{\mathcal{F}})$ is a probability distribution $\nu \in \mathcal{M}(X \times Y, \mathcal{F} \otimes \hat{\mathcal{F}})$ of a pair $(x, \hat{x}) \sim X \times \hat{X}$ such that the marginals are are given by $x \sim \mu$ and $\hat{x} \sim \hat{\mu}$. For a pair of probability measures $(\mu, \hat{\mu})$ on the measure space $\mathcal{M}(X, \mathcal{F})$, where $X = S^{\mathbb{Z}}$ and $\mathcal{F}$ denotes the corresponding product σ-algebra, let

$$\bar{d}(\mu, \hat{\mu}) := \inf_{\nu} \limsup_{n \to \infty} \nu\{x_{-n} \neq \hat{x}_{-n}\},$$

where the infimum is taken over all couplings ν between μ and $\hat{\mu}$. This make up the $\bar{d}$-metric introduced by Ornstein (for a reference, see e.g., [7] or [19]), if we take the restriction to the space $\mathcal{M}_T(X)$ of shift invariant measures; on $\mathcal{M}(X)$ it is a pseudo-metric. Notice that in our case, the definition of $\bar{d}$ uses couplings that are not necessarily translation invariant even if the marginals are. In [7], the authors define $\bar{d}$ on $\mathcal{M}_T(X)$ by taking the infimum over couplings that are invariant under the transformation $T \times T$ on $X \times X$. The original definition by Ornstein does not presuppose translation invariant couplings.

An invariant measure $\mu \in \mathcal{M}_T(X)$ is *Bernoulli* if it can be realised by an isomorphism with a Bernoulli shift. In other words, there is a bijectively measurable mapping $\phi : A^{\mathbb{Z}} \to X$ such that $\phi \circ T' = T \circ \phi$, where T' denote the shift on $A^{\mathbb{Z}}$ and such that $\mu = \mu' \circ \phi^{-1}$ where μ' is a Bernoulli shift, which means that, under μ', each symbol is chosen independently according to some fixed discrete probability on the finite set A. Ornstein proves in [19] that the set $\mathcal{B}$ of measures in $\mathcal{M}_T(X)$ having the Bernoulli property is closed in the topology induced by the $\bar{d}$-metric. Many classes of g-functions are well-known to give rise to unique g-measures with the Bernoulli property. In particular, if the g-function is determined by a finite number of coordinates, i.e., it is the transition probabilities for N-chains, for some finite N; see e.g. [19] or [7]. We also remind the reader of the results of Walters, see [**?**].

It easy to see that any given g-function g with $\operatorname{var}_N \log g \to 0$ as $N \to \infty$ can be arbitrarily well approximated by finitely determined g-functions, e.g. let $\hat{g}_N(x) = g(x_0, x_1, \ldots, x_N z)$, for a fixed $z \in X_+$, whence

$$\| \log \hat{g}_N - \log g \|_\infty \leq \operatorname{var}_N \log g.$$

Let μ and $\hat{\mu}$ denote g-chains corresponding to the g-functions g and $\hat{g}$, respectively. Our strategy — which is similar to that in used in [7] — for proving that the g-measure μ is Bernoulli, is first to show that, the $\bar{d}$-distance between μ and $\hat{\mu}$ can be bounded by a function which is continuous in $s = \| \log g - \log \hat{g} \|_\infty$.

A finite *block-structure* is a sequence $\{b_l\}_{l=1}^M$ of positive integer $b_l \geq 0$. We refer to the index l as *levels*. By a *block-variation pair*, we mean a block-structure $\{b_l\}$ in conjunction with a sequence $\{r_l\}$ of positive real numbers. For a block-variation pair $(\{b_l\}, \{r_l\}) = (\{r_l\}_{l=1}^M, \{b_l\}_{l=1}^M)$ we define

$$\bar{\delta}(\{r_l\}, \{b_l\}) := \frac{1 + \sum_{l=1}^M b_l e^{-r_1 - \cdots - r_{l-1}} \left(1 - e^{-r_l}\right)}{\sum_{l=1}^M b_l e^{-r_1 - \cdots - r_{l-1}}}. \tag{2.1}$$

A block-variation function r, is a real and positive $r(B, b)$ defined for integers $B \geq 0$ and $b > 0$. Given a block-structure $\{b_l\}$ and a block-variation function r, we define the corresponding sequence $\{r_l\}$ by setting

$$r_l := r(b_1 + b_2 + \cdots + b_{l-1}, b_l). \tag{2.2}$$

In this context, we will refer to the pair $(\{b_l\}, \{r_l\})$ by $(\{b_l\}, r)$.

Our first lemma establishes a bound on the $\bar{d}$-metric between g-chains which is continuous in the supremum norm.

Lemma 2.1. *Let g and μ be as above. There is a block-variation function $\rho^g(B, b)$, such that for any block-variation pair $(\{b_l\}, \{r_l\})$ satisfying*

$$\rho_l^g \leq r_l \tag{2.3}$$

we have

$$\bar{d}(\mu, \hat{\mu}) \leq \bar{\delta}\left(\{r_l + s \cdot b_l\}, \{b_l\}\right), \tag{2.4}$$

for all g-chains $\hat{\mu}$ corresponding to a g-functions $\hat{g}$ with

$$\|\log g - \log \hat{g}\|_\infty = s.$$

We say that pairs $(\{b_l\}, \{r_l\})$ satisfying (2.3) are *valid* for g. We prove this lemma in the next subsection. Note that, for a fixed finite pair $(\{b_l\}_{l=1}^M, \{r_l\}_{l=1}^M)$, the quantity $\bar{\delta}(\{r_l\}, \{b_l\})$ is clearly continuous in $\{r_l\}$ so that in particular

$$\lim_{s \to 0+} \bar{\delta}(\{r_l + s b_l\}, \{b_l\}) = \bar{\delta}(\{r_l\}, \{b_l\}).$$

To see how we can deduce the the Bernoulli property, notice that if

$$\inf_{\{r_l\},\{b_l\}} \bar{\delta}(\{r_l\}, \{b_l\}) = 0, \tag{2.5}$$

where the infimum is taken over all pairs $(\{r_l\}, \{b_l\})$ that are valid for g. Then, for every $\epsilon > 0$, we can find a block-structure $\{b_l^\epsilon\}_{l=1}^M$ with $\bar{\delta}(g, \{b_l^\epsilon\}) < \epsilon$. By the continuity of $\bar{\delta}(\cdot, \{b_l^\epsilon\})$ we can take a finitely determined (locally constant) g-function $\hat{g}$ with g-measure $\hat{\mu}$ such that

$$\bar{d}(\mu, \hat{\mu}) \leq \bar{\delta}\left(r + \|\log g - \log \hat{g}\|_\infty, \{b_l^\epsilon\}\right) < 2\epsilon,$$

say. It follows that the $\bar{d}$-distance between the g-measure μ of g and the set $\mathcal{B}$ of Bernoulli measures is zero and since $\mathcal{B}$ is closed with respect to the $\bar{d}$-distance, we conclude that $\mu \in \mathcal{B}$. Moreover, it is well-known and easy to see that this g-measure corresponding to g must be unique. We collect the conclusions in the following Theorem.

Theorem 2.2. *If* (2.5) *holds then we have a unique Bernoulli g-measure μ corresponding to g. Moreover, μ is attractive in the sense that $\mathcal{L}^{*n}\nu$ converges weakly to μ for any initial distribution $\nu \in \mathcal{M}(X_+)$.*

We prove the last statement separately in Section 3.

2.2. **The coupling argument and the proof of Lemma 2.1.**

In order to obtain the bound in (2.4), we will need to construct a coupling between a g-chain μ and a $\hat{g}$-chain $\hat{\mu}$, by defining the two chains $x \sim \mu$ and $\hat{x} \sim \hat{\mu}$ on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Assume that $s = \|\log g - \log \hat{g}\|_\infty$. The distributions of $x^{(0)}$ and $\hat{x}^{(0)}$ are arbitrary.

The coupling we construct uses a block-structure $\{b_l\}$, where we, at certain times n, extend the two g-chains with block of symbols of length b_l until we reach a conflict — i.e. a coordinate with different symbols — in the extension. Extending the two chains $x^{(n)}$ and $\hat{x}^{(n)}$ with a block of length b_l, means specifying a distribution of the pair $(x^{(n+b_l)}, \hat{x}^{(n+b_l)})$ such that $x^{(n+b_l)}$ has distribution $\mathcal{L}_g^{*b_l}\delta_{x^{(n)}}$ and $\hat{x}^{(n+b_l)}$ has distribution $\mathcal{L}_{\hat{g}}^{*b_l}\delta_{\hat{x}^{(n)}}$. We are at *level l* when we extend with a b_l-block and this presupposes, that previously, without conflict, we have extended with blocks at levels $0, 1, \dots, l-1$ of a total length

$$B_{l-1} = b_1 + b_2 + \cdots + b_{l-1}.$$

The event of success (or "no conflict") means that that

$$\kappa(x^{(n+b_l)}, \hat{x}^{(n+b_l)}) = \kappa(x^{(n)}, \hat{x}^{(n)}) + b_l.$$

We always use a *maximal coupling* between the chains, i.e., a coupling that makes the probability of success maximal.

We show (2.4) in Lemma 2.1, by defining on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ a Markov chain Y_n taking values in $\mathbb{Z}$. Given a block-variation pair $(\{b_l\}, \{r_l\})$, we define an associated Markov chain $Y_n = Y_n^{\{r_l\},\{b_l\}}$, $n \geq 0$, as follows: Let $Y_0 = 0$. If $Y_n \neq B_l$ for some l, simply let $Y_{n+1} = Y_n + 1$, but, if $Y_n = B_l$ for some level $l = 0, 1, \dots, M-1$ then

$$Y_{n+1} = \begin{cases} B_{l-1} + 1 & \text{w. prob. } e^{-r_l} \\ -b_l & \text{w. prob. } 1 - e^{-r_l} \end{cases}. \tag{2.6}$$

If $Y_n = B_M$ we set $Y_{n+1} = 0$, because we want to avoid to have infinite waiting time in mean when we later solve the renewal equation.

By using the Renewal Theorem, we show in section 3 the following.

Lemma 2.3. *Assume that the Markov chain Y_n is defined from parameters r and $\{b_l\}$ as in* (2.6)*. Then*

$$\limsup_{n\to\infty} \mathbb{P}\{Y_n \leq 0\} \leq \bar{\delta}(r, \{b_l\})$$

where $\bar{\delta}$ is defined in (2.1)*.*

We couple the Markov chain $Y_n = Y_n^{\{r_l+sb_l\},\{b_l\}}$ with the block-extensions such that, for all n,

$$\kappa(x^{(n)}, \hat{x}^{(n)}) \geq Y_n. \tag{2.7}$$

Since $x_{-n} \neq \hat{x}_{-n}$ precisely when $\kappa(x^{(n)}, \hat{x}^{(n)}) = 0$ it then becomes clear from Lemma 2.3 that

$$\bar{d}(\mu, \hat{\mu}) \leq \limsup \mathbb{P}\{Y_n \leq 0\} \leq \bar{\delta}(\{r_l\}, \{b_l\}), \tag{2.8}$$

which is (2.4) in Lemma 2.1.

We execute, at time n, a block-extension at level l, precisely when $Y_n = B_{l-1}$. In order to maintain (2.7), we should couple the transition of Y_n so that $Y_n = -b_l$ if the extension is unsuccessful; then (2.7) holds up true to time $n + b_l$ even if coordinates between $-n$ and $-n - b_l$ should disagree. A sufficient and necessary condition for the mechanism to work is therefore that the probability that Y_n of moves up one level, i.e. e^{-r_l}, is less than the probability that the block-extension is successful. We define $\rho^{g,\hat{g}}(B_{l-l}, b_l)$ as the infimum, over $(x^{(n)}, \hat{x}^{(n)})$, of the probability of success, conditioned on $(x^{(n)}, \hat{x}^{(n)})$, under the restriction that $\kappa(x^{(n)}, \hat{x}^{(n)}) \geq B_{l-1}$. More precisely, we need to show that, the condition that r_l is valid implies that $r_l + s \cdot b_l$ is less than $\rho^{g,\hat{g}}(B_{l-1}, b_l)$. As before, we assume that a maximal coupling is used. Notice that, if the extension is executed at level l, we have $\kappa(x^{(n)}, \hat{x}^{(n)}) \geq Y_n = B_{l-1}$, by (2.7).

What remains to do in the proof of Lemma 2.1 is to show that

$$\rho^{g,\hat{g}}(B, b) \geq \rho^{g,g}(B, b) + s \cdot b,$$

and to give an explicite expression for $\rho^g := \rho^{g,g}$.

It is well-known that the probability for a successful extension in a maximal coupling is given by the total variation metric between the marginals of the extension, see e.g. [18]. The success probability is given by

$$\int \left(\frac{d\hat{\eta}}{d\eta} \wedge 1 \right) d\eta = \left(1 - \frac{1}{2} \cdot d_{TV}\left(\eta, \hat{\eta}\right) \right).$$

In our situations we can identify the marginals η and $\hat{\eta}$ with the distributions on $\mathcal{F}_b$ given by

$$\eta = \mathcal{L}_g^{*b} \delta_{x^{(n)}}|_{\mathcal{F}_b}, \quad \hat{\eta} = \mathcal{L}_{\hat{g}}^{*b} \delta_{\hat{x}^{(n)}}|_{\mathcal{F}_b}.$$

for some $x^{(n)}$ and $\hat{x}^{(n)}$ that satisfy $\kappa(x^{(n)}, \hat{x}^{(n)}) \geq B$. Let $\mathcal{M}_{B,b}^{g,\hat{g}}$ denote the set of such pairs $(\eta, \hat{\eta})$.

We then define

$$\rho^{g,\hat{g}}(B, b) := \sup \left\{ - \log \int \left(\frac{d\hat{\eta}}{d\eta} \wedge 1 \right) d\eta : (\eta, \hat{\eta}) \in \mathcal{M}_{B,b}^{g,\hat{g}} \right\}. \tag{2.9}$$

Notice that, since $\hat{g}/g \geq e^{-s}$, we have

$$\frac{d\hat{\eta}}{d\eta} = \frac{\hat{g}(\hat{x})\hat{g}(T\hat{x}) \cdots \hat{g}(T^{b-1}\tilde{x})}{g(x)g(Tx) \cdots g(T^{b-1}x)} \geq e^{-bs} \cdot \frac{g(\tilde{x})g(T\tilde{x}) \cdots g(T^{b-1}\tilde{x})}{g(x)g(Tx) \cdots g(T^{b-1}x)} \tag{2.10}$$

which equals $e^{-bs} \cdot d\tilde{\eta}/d\eta$ where

$$\tilde{\eta} := \mathcal{L}_g^{*b} \delta_{\hat{x}^{(n)}}.$$

We then obtain from (2.9) that

$$\rho^{g,\hat{g}}(B, b) \leq \rho^{g,g}(B, b) + s \cdot b, \tag{2.11}$$

where

$$\rho^{g,g} = \sup \left\{ - \log \int \left(\frac{d\tilde{\eta}}{d\eta} \wedge 1 \right) d\eta : (\eta, \tilde{\eta}) \in \mathcal{M}_{B,b}^{g,g} \right\}. \tag{2.12}$$

Since $\rho^g = \rho^{g,g}$, this concludes the proof of Lemma 2.1. $\square$

2.3. **Estimates using Hellinger integrals.** In order to arrive at verifiable conditions that ensures that $\inf \bar{\delta}(r, \{b_l\}) = 0$, i.e. the assumption (2.5) in Theorem 2.2, we estimate the total variation metric using the Hellinger integral. This was done in some special cases also in our earlier paper [15]. Define the "Hellinger block-variation" $h(B, b) = h^g(B, b)$ by

$$h^g(B, b) = \sup\left\{ -\log H(\eta, \tilde{\eta})\, d\eta : (\eta, \tilde{\eta}) \in \mathcal{M}^{g,g}_{B,b} \right\} \tag{2.13}$$

where

$$H(\eta, \tilde{\eta}) = \int \left(\frac{d\tilde{\eta}}{d\eta} \right)^{\frac{1}{2}} d\eta$$

is the Hellinger integral of η and $\tilde{\eta}$. We always have $0 \le H \le 1$.

The relevant estimates we will need are collected in the following lemma.

Lemma 2.4. *We have the following relations between the block-variations defined above*

$$\rho^g \le -\log\left(1 - \sqrt{1 - \exp(-2h^g)}\right), \tag{2.14}$$

$$\rho^g \le \sqrt{2h^g} + 2h^g, \tag{2.15}$$

$$h^g(B, b) \le \sum_{k=B}^{B+b-1} h^g(k, 1), \tag{2.16}$$

As $k \to \infty$

$$h^g(k, 1) = (1 + o(1))\frac{1}{8}(\mathrm{var}_k \log g)^2, \tag{2.17}$$

and as $w \to 0$

$$\rho^g(B, b) \le (1 + O(w))\, \frac{1}{2} \cdot w \tag{2.18}$$

where

$$w = \sqrt{\sum_{k=B}^{B+b} (\mathrm{var}_k \log g)^2}.$$

A condition ensuring that condition (2.5) is satisfied is given in the following Theorem. We say that a block-variation $(\{b_l\}, \{r_l\})$ is *eventually valid* if for some l_0, we have $r_l \ge \rho^g_l$ for $l \ge l_0$.

Theorem 2.5. *A sufficent condition for the conclusions of Theorem 2.2 to hold is that there is some infinite eventually valid block variation pair* $(\{b_l\}_{l=1}^\infty, \{r_l\}_{l=1}^\infty)$ *such that* $\limsup r_l = 0$ *and*

$$\sum_{l=1}^{\infty} e^{-r_1 - \cdots - r_{l-1}}\, b_l = \infty. \tag{2.19}$$

Proof. We verify (2.5), that is, we show that

$$\inf_{\{r_l\}_{l=1}^M,\{b_l\}_{l=1}^M} \frac{1+\sum_{l=1}^M b_l e^{-r_1-\cdots-r_{l-1}}(1-e^{-r_l})}{\sum_{l=1}^M b_l e^{-r_1-\cdots-r_{l-1}}} = 0. \tag{2.20}$$

To see this, note that $(1-e^{-r_l}) \le r_l$. Hence, by the assumption (2.19) and since $r_l \to 0$, as $l \to \infty$, we have

$$\inf_{\{r_l\}_{l=1}^M,\{b_l\}_{l=1}^M} \frac{1+\sum_{l=1}^M b_l e^{-r_1-\cdots-r_{l-1}} r_l}{\sum_{l=1}^M b_l e^{-r_1-\cdots-r_{l-1}}} = 0,$$

and the conclusion follows. □

2.4. **Examples.** By setting $b_l = 1$ and noting that $r_l = (1/2+\epsilon)\operatorname{var}_l \log g$ eventually dominates ρ_l^g by (2.18), we can deduce the special case (2) in the Introduction. We now show that the special cases (1) and (3), by verifying that the conditions in Theorem 2.5 are satisfied.

Note that the following proposition gives a uniqueness result that is not covered by earlier results, for instance in [13].

Proposition 2.6. *We have a unique g-measure with the Bernoulli property if*

$$\operatorname{var}_n \log g = o\left(\frac{1}{\sqrt{n}}\right).$$

Proof. Take a real number $c > 1$. Let $B_0 = 0$ and let $B_l = \lceil c^l/(c-1)\rceil$ for $l \ge 1$, so that for $l \ge 2$ $b_l = B_l - B_{l-1}$ satisfies

$$b_l \ge \left\lfloor c^l/(c-1) - c^{l-1}/(c-1)\right\rfloor = \lfloor c^l \rfloor \ge 1.$$

Define r_l by

$$r_l^2 = \sum_{n=B_{l-1}}^{B_l-1} (\operatorname{var}_n \log g)^2.$$

For $l \ge 2$, we have by assumption that (as $l \to \infty$)

$$\begin{aligned} r_l^2 &\le o(1) \cdot \sum_{n=B_{l-1}}^{B_l-1} \frac{1}{\sqrt{n}}, \\ &\le o(1) \cdot \int_{c^{l-1}/(c-1)}^{c^l/(c-1)} \frac{1}{x}\,dx \\ &= o(\log c) = o\big((\log c)^2\big). \end{aligned}$$

The integral estimate of the partial sums of the harmonic series follows since $B_{l-1} \ge c^{l-1}/(c-1)$ and $B_l - 1 \le c^l/(c-1)$.

Since, by (2.18), $\rho_l^g \leq r_l$ eventually, we can apply Theorem 2.5. We already know that $r_l = o(\log c) \to 0$ as $l \to \infty$. Moreover, each term in the sum of (2.19) can be estimated as

$$b_l e^{-r_1-\cdots-r_l} \geq \exp\{l \log c - l \cdot o(\log c)\} \to \infty \tag{2.21}$$

which verifies (2.19). □

We now show that the uniqueness condition of [13] also gives the Bernoulli property.

Proposition 2.7. *We have a unique g-measure with the Bernoulli property if*

$$\sum_n (\mathrm{var}_n \log g)^2 < \infty.$$

Proof. First note that if $\{r_l\}$ is a block-variation relative blocks $\{b_l\}$ such that

$$r_1 + r_2 + \cdots < \infty,$$

then it is clear that the conditions in Theorem 2.5 hold for $\{r_l\}$ and $\{b_l\}$.

We define the blocks B_l such that $B_0 = 0$ and

$$B_l = \inf\left\{ B > B_{l-1} : \sum_{n=B}^{\infty} (\mathrm{var}_n \log g)^2 \leq L/2^l \right\}$$

where $L = \sum_{n=0}^{\infty} (\mathrm{var}_n \log g)^2$. Then with r_l defined by

$$r_l^2 = \sum_{n=B_{l-1}}^{B_l-1} (\mathrm{var}_n \log g)^2,$$

we have $r_{l+1} \leq O\left(\sqrt{L/2^l}\right)$ and $\{r_l\}$ is clearly a summable summable sequence since it decreases geometrically. Moreover, $\rho_l^g \leq r_l$ eventually by (2.18). □

3. Remaining proofs

3.1. **Proof of Lemma 2.4.** Note that (2.18) is easily deduced from (2.17) and (2.16).

Proof of (2.15). In order to relate the two variation functions ρ^g and h^g, we use the following bound (Proposition V.4.4 in [12, p. 311]) on the total variaton metric

$$d_{TV}(\eta, \tilde{\eta}) \leq 2\sqrt{1 - H(\eta, \tilde{\eta})^2} \tag{3.1}$$

and (2.14) and (2.15) follows from this using easy calculations. In the estimate (2.15), the first term $\sqrt{2} \cdot \sqrt{h^g}$ is sharp ($\sqrt{2}$ is the sharp number), but the second, $2 \cdot h^g$, is not (lower numbers than 2 are possible). □

Proof of (2.16). Let $(\eta, \tilde{\eta}) \in \mathcal{M}^{g,g}_{B,b}$. We can explicitly write

$$H(\eta, \tilde{\eta}) = \int \left(\frac{g(\tilde{x})g(T\tilde{x}) \cdots g(T^{K-1}\tilde{x})}{g(x)g(Tx) \cdots g(T^{K-1}x)} \right)^{1/2} d\eta(x), \tag{3.2}$$

where $(x, \tilde{x}) \in (X_+, X_+)$ satisfies $\kappa(x, \tilde{x}) \geq B + b$. Taking the conditional η-expectation of $\sqrt{\frac{g(\tilde{x})}{g(x)}}$ conditioned on Tx gives

$$H(\eta, \tilde{\eta}) = \int h(Tx, T\tilde{x}) \left(\frac{g(T\tilde{x}) \cdots g(T^{K-1}\tilde{x})}{g(Tx) \cdots g(T^{K-1}x)} \right)^{1/2} d\eta(x)$$

where we have

$$h(y, \tilde{y}) = \sum_{\alpha \in S} \sqrt{g(\alpha\tilde{y})}\sqrt{g(\alpha y)} \tag{3.3}$$

Since $-\log h(Tx, T\tilde{x})) \leq -h^g(B + b - 1, 1)$, we obtain the recursive expression

$$-\log H(\eta, \tilde{\eta}) \leq h^g(B + b - 1, 1) \cdot \left\{-\log H(\eta', \tilde{\eta}')\right\},$$

where $(\eta', \tilde{\eta}') \in \mathcal{M}^{g,g}_{B-1,b-1}$. This proves (2.16). $\square$

Proof of (2.17). The relation (2.17) follows from the Arithmetic–Geometric mean inequality: Fix $(x, \tilde{x}) \in X_+ \times X_0$, and assume that $g(\tilde{x}) = e^{\delta(x,\tilde{x})}g(x)$, say, where $|\delta(x, \tilde{x})| \leq \mathrm{var}_{\kappa(x,\tilde{x})} \log g$. Then

$$\sqrt{g(\tilde{x})}\sqrt{g(x)} = \frac{1}{2}\left(g(x) + g(\tilde{x})\right) - \delta^2 f(\delta)g(x), \tag{3.4}$$

where f is the continuous and strictly positive function

$$f(\delta) = \frac{1}{\delta^2}\left(\frac{1}{2}(1 + e^{\delta}) - e^{\delta/2}\right),$$

tending to 1/8 as $\delta \to 0$. Summing (3.4) over y and $\tilde{y}$ such that $(y, \tilde{y}) = (\alpha Tx, \alpha T\tilde{x})$, $\alpha \in S$, gives that

$$\begin{aligned} -\log h(Tx, T\tilde{x}) &= -\log(1 - \sum_y \delta^2(y, \tilde{y}) f(\delta(y, \tilde{y}))g(y)) \\ &= (1 + o(1))\delta^2 f(\delta). \end{aligned}$$

where h as in (3.3). Taking the infimum over $(Tx, T\tilde{x})$ such that $\kappa(Tx, T\tilde{x}) \geq k$ proves (2.17). $\square$

3.2. **Proof of Lemma 2.3.** We now use renewal theory to show Lemma 2.3. Our aim is to prove that

$$\mathbb{P}(Y_n \leq 0) \to 0 \text{ as } n \to \infty.$$

The Markov chain $\{Y_n\}$ will return to 0 at random times $\{S_0, S_1, S_2, \dots\}$ where $S_0 = 0$, since $Y_0 = 0$. For time n, define the number N_n of returns as

$$N_n = |\{k : 0 \leq k \leq n, Y_k = 0\}| = \sup\{k : S_k \leq n\}.$$

Define *the waiting times* $T_k = S_k - S_{k-1}$ which are independent and identically distributed waiting times due to the Markov property of Y_n. The waiting time T_{N_n} is the length of the "cycle" that Y_n currently completes and this cycle $Y_{S_{N_n}} \dots Y_{S_{N_n+1}}$ has length B_l for some level l. Let L_n denote this level, i.e. $B_{L_n} = T_{N_n}$.

We now use the renewal equation to analyse

$$A_n = \mathbb{P}(Y_n \leq 0). \tag{3.5}$$

The expansion

$$A_n = \mathbb{P}(Y_n \leq 0, N_n = 1) + \mathbb{P}(Y_n \leq 0, N_n > 1) \tag{3.6}$$

leads to the renewal equation

$$A_n = a_n + \sum_{j=1}^{\infty} A_{n-j} p_j, \tag{3.7}$$

where $a_n = \mathbb{P}(Y_n \leq 0, N_n = 1)$ and $p_j = \mathbb{P}(T_1 = j)$.

Let $q_l = \mathbb{P}\{L_n = l\}$. Then

$$q_l = \mathbb{P}\{L_n \geq l\} - \mathbb{P}\{L_n \geq l+1\} = e^{-r_1 - \dots - r_{l-1}}(1 - e^{-r_l}).$$

Note that

$$p_j = \begin{cases} q_l, & j = B_l,\ l = 1, 2, \dots, M-1 \\ 1 - \sum_{l=1}^{M-1} q_l, & j = B_M \\ 0, & \text{otherwise.} \end{cases}$$

Since, q_l is the probability that, in the first cycle, $Y_n \leq 0$ for $B_{l-1} < n \leq B_l = T_1$, we obtain

$$a_n = \begin{cases} 1, & n = 0 \\ q_l, & B_{l-1} < n \leq B_l,\ l = 1, 2, \dots, M \\ 0, & \text{otherwise.} \end{cases}$$

It is well known that the renewal equation (3.7) has the solution

$$A_n = \sum_{j=0}^{\infty} u_{n-j} a_j, \tag{3.8}$$

where $u_n = \mathbb{E}[N_n] - \mathbb{E}[N_{n-1}]$ and the theorem in [9, p. 362] states that

$$\lim_{n\to\infty} A_n = \frac{\sum_{j=0}^{\infty} a_j}{\mathbb{E}[T_1]},$$

provided $\sum_{j=0}^{\infty} |a_j| < \infty$. In our case we have $T_1 \le B_M < \infty$ and this condition is trivially satisfied.

The ratio $\sum_j a_j / \mathbb{E}[T_1]$ can be transformed to that in (2.1). We have

$$\sum_{j=0}^{\infty} a_j = 1 + \sum_{l=1}^{M} b_l q_l = 1 + \sum_{l=1}^{M} b_l e^{-r_1 - \cdots - r_{l-1}} (1 - e^{-r_l}),$$

and

$$\mathbb{E}[T_1] = \sum_{l=1}^{M} q_l B_l + (1 - \sum_{l=1}^{M-1} q_l) B_M,$$

where the last term is due to the fact that we let $Y_{n+1} = 0$ whenever $Y_n = B_M$. Since $B_l = b_1 + b_2 + \cdots + b_l$, the denominator $\mathbb{E}[T_1]$ equals

$$\sum_{l=1}^{M} q_l B_l + (1 - \sum_{l=1}^{M} q_l) B_M = \sum_{l=1} b_l \left(\sum_{k=l}^{M-1} q_l + e^{-r_1 - \cdots - r_M} \right),$$

and, by the definition of q_l, the sum

$$\sum_{k=l}^{M-1} q_l = e^{-r_1 - \cdots - r_{l-1}} - e^{-r_1 - \cdots - r_M}$$

and we obtain that $\mathbb{E}[T_1]$ equals the denominator in (2.1). □

3.3. **Proof of the last statement in Lemma 2.2.** We now prove the remaining statement in Theorem 2.2: That (2.5) implies that $\mathcal{L}^{*n}\mu'$ converges weakly to (the necessarily unique) g-measure in $\mathcal{M}_T^g$ for any initial distribution $\mu' \in \mathcal{M}(X_+)$.

In fact, we prove convergence in the Wasserstein metric. Given an underlying (pseudo-) metric d on the space Y, the corresponding Wasserstein (pseudo-) metric d_W between probability measures $\mu, \tilde{\mu} \in \mathcal{M}(Y)$ is defined as

$$d_W(\mu, \tilde{\mu}) := \inf_{\lambda} E_\lambda [d(x, \tilde{x})],$$

where the infimum is taken over all couplings $\lambda \in \mathcal{M}(Y \times Y)$ of μ and $\tilde{\mu}$. On the space X_+, we consider the underlying metric $d(x, \tilde{x}) = 2^{-\kappa(x,\tilde{x})}$ and the corresponding Wasserstein metric d_W.

We already know that the condition (2.5) in Lemma 2.2 implies that $\bar{d}$-distance between any pair of g-chains is zero. In other words

$$\inf_{\nu} \lim_{n\to\infty} E_\nu \left[\mathbf{1}_{\kappa=0}(x^{(n)}, \tilde{x}^{(n)}) \right] = 0 \tag{3.9}$$

where $\nu \in \mathcal{M}(X \times X)$ signifies couplings of the two arbitrary g-chains. We shall show that (3.9) implies that

$$\limsup_{n\to\infty} d_W(\mathcal{L}^{*n}\mu, \mathcal{L}^{*n}\tilde{\mu}) = 0. \tag{3.10}$$

Since d_W metrizes the weak topology, (3.10) is equivalent to stating that g has a unique *attractive* g-measure, i.e. is a for any μ, $\{\mathcal{L}^{*n}\mu\}$ converges weakly to a unique g-measure as $n \to \infty$.

The statement (3.10) follows readily from (3.9): Let $N \geq 0$ be fixed but arbitrary. A coupling $\nu \in \mathcal{M}(X \times X)$ of the g-chains with initial distributions μ and $\tilde{\mu}$ also gives a coupling $\lambda = \nu \circ (x^{(n)})^{-1} \otimes \nu \circ (\tilde{x}^{(n)})^{-1}$ of $\mathcal{L}^{*n}\mu$ and $\mathcal{L}^{*n}\tilde{\mu}$. Since

$$d(x, \tilde{x}) \leq 2^{-N} + \mathbf{1}_{\kappa \leq N}(x, \tilde{x}) \leq 2^{-N} + \sum_{n=0}^{N-1} \mathbf{1}_{\kappa=0}(T^n x, T^n \tilde{x})$$

it therefore follows from (3.9) that

$$\limsup_{n} d_W(\mathcal{L}^{*n}\mu, \mathcal{L}^{*n}\tilde{\mu}) \leq 2^{-N} + \limsup_{n} \inf_{\nu} E_\nu \left[\sum_{n=0}^{N-1} \mathbf{1}_{\kappa=0}(T^k x^{(n)}, T^k \tilde{x}^{(n)}) \right] \leq 2^{-N} + N \cdot 0.$$

Since N was arbitrary, this concludes the proof. □

References

[1] M. Aizenman, J.T. Chayes, L. Chayes, C.M. Newman, *Discontinuity of the magnetization in one-dimensional* $1/|x-y|^2$ *Ising and Potts models*, J. Statist. Phys. **50** (1988), no. 1-2, 1–40.

[2] H. Berbee, Chains with Infinite Connections: Uniqueness and Markov Representation, *Probab. Theory Related Fields* **76** (1987), 243–253.

[3] H. Berbee, Uniqueness of Gibbs measures and absorption probabilities, *Ann. Probab.* **17** (1989), no. 4, 1416–1431.

[4] N. Berger, C. Hoffman and Sidoravicius, Nonuniqueness for specifications in $l^{2+\epsilon}$. Preprint available on www.arxiv.org (PR/0312344).

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

[6] X. Bressaud, R. Fernández and A. Galves, Decay of correlations for non Hölderian dynamics. A coupling approach, *Electron. J. Proabab.* **4** (1999), no. 3, 1–19.

[7] Z. Coelho and A. Quas, Criteria for $\bar{d}$-continuity, *Trans. Amer. Math. Soc.* **350** (1998), no. 8, 3257–3268.

[8] F.J. Dyson, *Non-existence of spontaneous magnetisation in a one-dimensional Ising ferromagnet*, Commun. Math. Phys. **12** (1969), no. 3, 212–215.

[9] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. II, John Wiley & Sons, 1971.

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

[11] P. Hulse, An example of non-unique g-measures, *Ergodic Theory Dynam. Systems* **26** (2006), no. 2, 439–445.

[12] J. Jacod and A.N. Shiryaev, *Limit Theorems for Stochastic Processes*, 2nd ed., Grundlehren der mathematischen Wissenschaften **288**, Springer-Verlag 2003.

[13] A. Johansson and A. Öberg, Square summability of variations of g-functions and uniqueness of g-measures, *Math. Res. Lett.* **10** (2003), no. 5-6, 587–601.

[14] A. Johansson and A. Öberg, *Square summability of variations and convergence of the transfer operator*, *Ergodic Theory & Dynam. Systems* **28** (2008), no. 4, 1145–1151.

[15] A. Johansson, A. Öberg and M. Pollicott, *Countable state shifts and uniqueness of g-measures*, Amer. J. Math. **129** (2007), no. 6, 1501–1511.

[16] Yu.M. Kabanov, R.Sh. Lipster and A.N. Siryaev, On the variation distance for probability measures defined on a filtered space, *Probab. Theory Related Fields* **71** (1986), 19–35.

[17] M. Keane, Strongly Mixing g-Measures, *Invent. Math.* **16** (1972), 309–324.

[18] T. Lindvall, *Lectures on the coupling method*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1992.

[19] D.S. Ornstein, *Ergodic Theory, Randomness, and Dynamical Systems*, Yale University Press, 1974.

[20] P. Walters, Ruelle's operator theorem and g-measures, *Trans. Amer. Math. Soc.* **214** (1975), 375–387.

[21] P. Walters, Invariant measures and equilibrium states for some mappings which expand distances, *Trans. Amer. Math. Soc.* **236** (1978), 121–153.

[22] P. Walters, Regularity conditions and Bernoulli properties of equilibrium states and g-measures, *J. London Math. Soc.* **71** (2005), no. 2, 379–396.

Anders Johansson, Academy of technology and environmental sciences, University of Gävle, SE-801 76 Gävle, Sweden, Uppsala, Sweden

E-mail address: ajj@hig.se

Anders Öberg, Department of Mathematics, Uppsala University, P.O. Box 480 SE-751 06 Uppsala, Sweden

E-mail address: anders@math.uu.se