J. Eur. Math. Soc. 14, 1599–1615

DOI 10.4171/JEMS/342

Anders Johansson · Anders Öberg · Mark Pollicott

Unique Bernoulli g-measures

Received March 30, 2010

Abstract. We improve and subsume the conditions of Johansson and Öberg [18] and Berbee [2] for uniqueness of a g-measure, i.e., a stationary distribution for chains with complete connections. In addition, we prove that these unique g-measures have Bernoulli natural extensions. In particular, we obtain a unique g-measure that has the Bernoulli property for the full shift on finitely many states under any one of the following additional assumptions.

(1) $\sum_{n=1}^{\infty}(\mathrm{var}_n \log g)^2 < \infty$,
(2) For any fixed $\epsilon > 0$, $\sum_{n=1}^{\infty} e^{-(1/2+\epsilon)(\mathrm{var}_1 \log g+\cdots+\mathrm{var}_n \log g)} = \infty$,
(3) $\mathrm{var}_n \log g = o(1/\sqrt{n})$ as $n \to \infty$.

That the measure is Bernoulli in the case of (1) is new. In (2) we have an improved version of Berbee's [2] condition (concerning uniqueness and Bernoullicity), allowing the variations of $\log g$ to be essentially twice as large. Finally, (3) is an example that our main result is new both for uniqueness and for the Bernoulli property.

We also conclude that we have convergence in the Wasserstein metric of the iterates of the adjoint transfer operator to the g-measure.

Keywords. Bernoulli measure, g-measure, chains with complete connections

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 $X_- = S^{\mathbb{Z}\setminus\mathbb{Z}_+}$. For each $n \in \mathbb{Z}$, a bi-infinite sequence $x \in X$ 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.

A. Johansson: Division of Mathematics and Statistics, University of Gävle,
SE-801 76 Gävle, Sweden and Department of Mathematics, Uppsala University,
P.O. Box 480, SE-751 06 Uppsala, Sweden; e-mail: ajj@hig.se

A. Öberg: Department of Mathematics, Uppsala University, P.O. Box 480,
SE-751 06 Uppsala, Sweden; e-mail: anders@math.uu.se

M. Pollicott: Mathematical Institute, University of Warwick, Coventry CV4 7AL, UK;
e-mail: mpollic@maths.warwick.ac.uk

Mathematics Subject Classification (2010): Primary 37A05, 37A35, 60G10

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

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

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

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

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 complete*, or *infinite*, *connections*.

In this paper, we will restrict our attention to the case when S is a finite set and $g > 0$. A stationary measure for our process is sometimes called a g-*measure* (see Keane [22], who introduced this notion in ergodic theory). Important contributions were also provided by Ledrappier [23], who in particular showed that g-measures are equilibrium states, and Walters [26], who connected the theory of g-measures with the transfer operator theory for general potentials. The theory has also had a long, but slightly different appearance in the probability theory of chains with complete connections: see e.g. Doeblin and Fortet (1937) [9], where it was proved that uniqueness of g-measures follows from summable variations, and the works by Iosifescu and co-authors, for instance that with Theoderescu [16] and with Grigorescu [15]. The theory is also connected to that of iterated function systems, or iterated random functions; see Diaconis and Freedman [8] and the references therein. A recent contribution by Iosifescu is [14]. We have not attempted to give a complete survey of the literature, but rather to point the reader in some important main directions of the different appearances of the problems we are considering here.

If T is the left shift map on X_+, then a g-measure can alternatively be viewed as a T-invariant probability measure $\mu \in \mathcal{M}(X_+)$, with the property that $g = d\mu/d(\mu \circ T)$. Since X_+ is compact due to the finiteness of S, it follows that there always exists a g-measure. 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 [13].

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 we 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, $\phi = \log g$, where g satisfies (1.1). In [18], 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 [10] 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 dichotomy in terms of summability of powers of variations, Berbee's two results from the late 1980s are intriguing. 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 $r_n = 1/n$. 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 covers sequences $r_n = 1/n^{1/2+\epsilon}$, $\epsilon > 0$.

Since it was shown in [18] 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 contain both square summability of variations and Berbee's condition for a unique g-measure.

Our method of proof also 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.,

$$x^{(n)} = (x_{-n}, x_{-n+1}, \dots), \qquad n \geq 0,$$

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 [19]). 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 $\epsilon > 0$, $\sum_{n=1}^{\infty} e^{-(1/2+\epsilon)(r_1+\cdots+r_n)} = \infty$;
(3) $\mathrm{var}_n \log g = o(1/\sqrt{n})$ as $n \to \infty$.

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 [28].

It would be interesting to investigate whether there is a sharp constant so that we have uniqueness and perhaps the Bernoulli property for $\mathrm{var}_n \log g \leq c/\sqrt{n}$. 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 [17] to calculate the probability of not having a conflict (that is, different entries in a corresponding coordinate) 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 maximal probability of not having 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 an increasing sequence of natural numbers $B_l = B_{l-1} + b_l$, $l \geq 1$, $B_0 = 0$, and define for any $\epsilon > 0$

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

then we obtain $\rho_l \leq r_l$ for

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

In the special cases (1) and (3) above, we have 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, our method allows one to improve Berbee's result by a constant; essentially, the variations are allowed to be twice as big.

We can now state one version of our main result.

Theorem 1.1. *We obtain a unique g-measure which is Bernoulli if there is a sequence $\{b_l\}_{l=1}^{\infty}$ of positive integers such that, with $\{r_l\}$ defined from $\{b_l\}$ as above,* $\limsup r_l = 0$ *and*

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

2. Preliminaries

2.1. The Bernoulli property and the $\bar{d}$-metric

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 $\mathcal{M}_T^g(X)$ 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 $\mathcal{B}$.

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 $(x, \hat{x}) \in X \times \hat{X}$ such that the marginals are given by $x \sim \mu$ and $\hat{x} \sim \hat{\mu}$. For a pair $(\mu, \hat{\mu})$ of probability measures 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 corresponds to the $\bar{d}$-metric introduced by Ornstein (for a reference, see e.g. [7] or [25]), 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 $X \times X$. However, 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' denotes 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 [25] 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, this is the case 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. [25] or [7]. We also remind the reader of the results of Walters [28].

It is easy to see that any given g-function g with $\mathrm{var}_N \log g \to 0$ 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

$$\|\log \hat{g}_N - \log g\|_\infty \le \mathrm{var}_N \log g.$$

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 $s = \|\log g - \log \hat{g}\|_\infty$ and that fixes zero.

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

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

where for simplicity we have adopted the convention that $e^{-r_1 - \cdots - r_{l-1}} = 1$ for $l = 1$. A *block-variation function* r associates a positive real number $r(B, b)$ to integers $B \ge 0$ and $b > 0$. Given a block-structure $\{b_l\}$ and a block-variation function r, we define the corresponding sequence $\{r_l\}$ by setting

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

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

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 $(\{r_l\}, \{b_l\})$ satisfying*

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

we have

$$\bar{d}(\mu, \hat{\mu}) \le \bar{\delta}(\{r_l + s b_l\}, \{b_l\}) \tag{2.4}$$

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

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

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

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

To see how we can deduce 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

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

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 [25], 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 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 blocks 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 $0, 1, \ldots, l-1$ of a total length

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

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

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

The event of success (or "no conflict") means 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) 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 $(\{r_l\}, \{b_l\})$, 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 $Y_n = B_{l-1}$ for some $l = 1, \ldots, M$, then

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

If $Y_n = B_M$ we set $Y_{n+1} = 0$, because we want to avoid 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 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).

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 true up 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 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 r_l being valid implies that $r_l + sb_l$ 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 complete the proof of Lemma 2.1 is to show that

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

and to give an explicit 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. [24]). The success probability is given by

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

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

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

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

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

and the right hand side equals $e^{-bs} \cdot d\tilde{\eta}/d\eta$, where

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

We then deduce from (2.9) that

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

where

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

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

2.3. *Estimates using Hellinger integrals*

In order to arrive at verifiable conditions that ensure $\inf \bar{\delta}(r, \{b_l\}) = 0$, 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 [20]. Define the Hellinger block-variation $h(B, b) = h^g(B, b)$ by

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

where

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

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

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:*

$$\rho^g \leq -\log\bigl(1-\sqrt{1-\exp(-2h^g)}\bigr), \quad \textit{and, in particular,} \tag{2.14}$$

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

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

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 $w \to 0$,

$$\rho^g(B,b) \leq (1+O(w))\frac{1}{2}w \quad \textit{where} \quad w = \sqrt{\sum_{k=B}^{B+b}(\mathrm{var}_k \log g)^2}. \tag{2.18}$$

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

Theorem 2.5. *A sufficient condition for the conclusions of Theorem 2.2 to hold is that there is some infinite eventually valid block-variation pair* $(\{r_l\}_{l=1}^{\infty}, \{b_l\}_{l=1}^{\infty})$ *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 $1-e^{-r_l} \leq r_l$. 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 $r_l = (1/2+\epsilon)\,\mathrm{var}_l \log g$ eventually dominates ρ_l^g by (2.18), we can deduce the special case (2) in the Introduction. We now show the results under the hypotheses in 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 [18].

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

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

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

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

Define r_l by

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

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

$$r_l^2 \leq o(1) \sum_{n=B_{l-1}}^{B_l-1} \frac{1}{n} \leq o(1) \int_{c^{l-1}/(c-1)}^{c^l/(c-1)} \frac{1}{x}\, dx = o(\log c) = o((\log c)^2).$$

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

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 [18] also gives the Bernoulli property.

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

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

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

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

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

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

we have $r_{l+1} \leq O(\sqrt{L/2^l})$ and $\{r_l\}$ is clearly a 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 (2.14) *and* (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 [17, p. 311]) on the total variaton metric:

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

This relation immediately gives (2.14) by rewriting the relations in terms of ρ^g and h^g. From this, we obtain (2.15) as a useful approximation by 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. Slightly 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 $(x,\tilde{x})\in X_+\times X_+$ satisfies $\kappa(x,\tilde{x})\ge B+b$. Taking the conditional η-expectation of $\sqrt{g(\tilde{x})/g(x)}$ 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

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

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

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

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

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

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

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 $(y, \tilde{y}) = (\alpha T x, \alpha T \tilde{x})$, $\alpha \in S$, gives

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

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). □

3.2. *Proof of Lemma 2.3*

We now use renewal theory to show Lemma 2.3. Our aim is to prove that

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

The Markov chain $\{Y_n\}$ will return to 0 at random times $\{S_0, S_1, S_2, \ldots\}$ 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 $Y_{S_{N_n}} \ldots Y_{S_{N_n+1}}$ 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

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

where we use our convention that $e^{r_1 - \cdots - r_{l-1}} = 1$ when $l = 1$, i.e., $q_1 = 1 - e^{-r_1}$. Note that

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

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

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

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 [11, 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 $T_1 \le B_M < \infty$ 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

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

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

$$\sum_{l=1}^{M-1} (e^{-r_1 - \cdots - r_{l-1}} - e^{-r_1 - \cdots - r_l})(b_1 + \cdots + b_l) + e^{-r_1 - \cdots - r_{M-1}} (B_{M-1} + b_M)$$
$$= \sum_{l=1}^{M} b_l e^{-r_1 - \cdots - r_{l-1}},$$

which is the denominator in (2.1). □

3.3. *Proof of the last statement in Theorem 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

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

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 the $\bar{d}$-distance between any pair of g-chains is zero. In other words

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

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

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

Since d_{W} metrizes the weak topology, (3.10) is equivalent to stating that g has a unique *attractive* g-measure, i.e. 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

$$\begin{aligned}\limsup_{n} d_{\mathrm{W}}(\mathcal{L}^{*n}\mu, \mathcal{L}^{*n}\tilde{\mu}) &\leq 2^{-N} + \limsup_{n} \inf_{\nu} E_{\nu}\Big[\sum_{n=0}^{N-1} \mathbf{1}_{\kappa=0}(T^k x^{(n)}, T^k \tilde{x}^{(n)})\Big] \\ &\leq 2^{-N} + N \cdot 0.\end{aligned}$$

Since N was arbitrary, this concludes the proof. □

Acknowledgments. We would like to thank Jean-Pierre Conze for valuable discussions. AJ acknowledges support from the Human Frontiers Science Program grant RGP51/2007. In addition, we would like to acknowledge the hospitality of the Institut Mittag-Leffler, where this paper was completed.

References

[1] Aizenman, M., Chayes, J. T., Chayes, L., Newman, C. M.: Discontinuity of the magnetization in one-dimensional $1/|x-y|^2$ Ising and Potts models. J. Statist. Phys. **50**, 1–40 (1988) Zbl 1084.82514 MR 0939480

[2] Berbee, H.: Chains with infinite connections: Uniqueness and Markov representation. Probab. Theory Related Fields **76**, 243–253 (1987) Zbl 0611.60059 MR 0906777

[3] Berbee, H.: Uniqueness of Gibbs measures and absorption probabilities. Ann. Probab. **17**, 1416–1431 (1989) Zbl 0692.60086 MR 1048934

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

[5] Bramson, M., Kalikow, S.: Nonuniqueness in g-functions. Israel J. Math. **84**, 153–160 (1993) Zbl 0786.60043 MR 1244665

[6] Bressaud, X., Fernández, R., Galves, A.: Decay of correlations for non-Hölderian dynamics. A coupling approach. Electron. J. Probab. **4**, no. 3, 19 pp. (1999) Zbl 0917.58017 MR 1675304

[7] Coelho, Z., Quas, A.: Criteria for $\bar{d}$-continuity. Trans. Amer. Math. Soc. **350**, 3257–3268 (1998) Zbl 0907.28013 MR 1422894

[8] Diaconis, P., Freedman, D.: Iterated random functions. SIAM Rev. **41**, 45–76 (1999) Zbl 0926.60056 MR 1669737

[9] Doeblin, W., Fortet, R.: Sur les chaînes à liaisons complètes. Bull. Soc. Math. France **65**, 132–148 (1937) JFM 63.1077.05 MR 1505076

[10] Dyson, F. J.: Non-existence of spontaneous magnetisation in a one-dimensional Ising ferromagnet. Comm. Math. Phys. **12**, 212–215 (1969)

[11] Feller, W.: An Introduction to Probability Theory and Its Applications. Vol. II, Wiley (1971) Zbl 0219.60003 MR 0270403

[12] Harris, T. E.: On chains of infinite order. Pacific J. Math. **5**, 707–724 (1955) Zbl 0066.11402 MR 0075482

[13] Hulse, P.: An example of non-unique g-measures. Ergodic Theory Dynam. Systems **26**, 439–445 (2006) Zbl 1087.37003 MR 2218769

[14] Iosifescu, M.: Iterated function systems. A critical survey. Math. Rep. (Bucur.) **11** (61), 181–229 (2009) Zbl 1212.60056 MR 2551080

[15] Iosifescu, M., Grigorescu, S.: Dependence with Complete Connections and its Applications. Cambridge Univ. Press (1990) Zbl 0749.60067 MR 1070097

[16] Iosifescu, M., Theodorescu, R.: Random Processes and Learning. Grundlehren Math. Wiss. 150, Springer, New York (1969) Zbl 0194.51101 MR 0293704

[17] Jacod, J., Shiryaev, A. N.: Limit Theorems for Stochastic Processes. 2nd ed., Grundlehren Math. Wiss. 288, Springer (2003) Zbl 1018.60002 MR 1943877

[18] Johansson, A., Öberg, A.: Square summability of variations of g-functions and uniqueness of g-measures. Math. Res. Lett. **10**, 587–601 (2003) Zbl 1051.28008 MR 2024717

[19] Johansson, A., Öberg, A.: Square summability of variations and convergence of the transfer operator. Ergodic Theory Dynam. Systems **28**, 1145–1151 (2008) Zbl 1177.37020 MR 2437224

[20] Johansson, A., Öberg, A., Pollicott, M.: Countable state shifts and uniqueness of g-measures. Amer. J. Math. **129**, 1501–1511 (2007) Zbl 1128.37002 MR 2369887

[21] Kabanov, Yu. M., Lipster, R. Sh., Shiryaev, A. N.: On the variation distance for probability measures defined on a filtered space. Probab. Theory Related Fields **71**, 19–35 (1986) Zbl 0554.60006 MR 0814659

[22] Keane, M.: Strongly mixing g-measures. Invent. Math. **16**, 309–324 (1972) Zbl 0241.28014 MR 0310193

[23] Ledrappier, F.: Principe variationnel et systèmes dynamiques symboliques. Z. Wahrsch. Verw. Gebiete **30**, 185–202 (1974) Zbl 0276.93004 MR 0404584

[24] Lindvall, T.: Lectures on the Coupling Method. Wiley, New York (1992) Zbl 0850.60019 MR 1180522

[25] Ornstein, D. S.: Ergodic Theory, Randomness, and Dynamical Systems. Yale Univ. Press (1974) Zbl 0296.28016 MR 0447525

[26] Walters, P.: Ruelle's operator theorem and g-measures. Trans. Amer. Math. Soc. **214**, 375–387 (1975) Zbl 0331.28013 MR 0412389

[27] Walters, P.: Invariant measures and equilibrium states for some mappings which expand distances. Trans. Amer. Math. Soc. **236**, 121–153 (1978) Zbl 0375.28009 MR 0466493

[28] Walters, P.: Regularity conditions and Bernoulli properties of equilibrium states and g-measures. J. London Math. Soc. **71**, 379–396 (2005) Zbl 1069.37021 MR 2122435