Probability Theory and Related Fields
https://doi.org/10.1007/s00440-024-01356-3

Uniqueness and mixing properties of Doeblin measures

Noam Berger[1] · **Diana Conache**[1] · **Anders Johansson**[2] · **Anders Öberg**[3]

Received: 8 February 2024 / Revised: 12 December 2024 / Accepted: 19 December 2024
© The Author(s) 2025

Abstract
In this paper we solve two open problems in ergodic theory. We prove first that if a Doeblin function g (a g-function) satisfies

$$\limsup_{n\to\infty} \frac{\mathrm{var}_n \log g}{n^{-1/2}} < 2,$$

then we have a unique Doeblin measure (g-measure). This result indicates a possible phase transition in analogy with the long-range Ising model. Secondly, we provide an example of a Doeblin function with a unique Doeblin measure that is not weakly mixing, which implies that the sequence of iterates of the transfer operator does not converge, solving a well-known folklore problem in ergodic theory. Previously it was only known that uniqueness does not imply the Bernoulli property.

Keywords Doeblin measure · Ergodic theory · g-measure · Chains with complete connections · Transfer operator · Mixing · Phase transition

Mathematics Subject Classification Primary 37A05 · 37A25 · 37A50 · 60G10

✉ Noam Berger
noam.berger@tum.de

✉ Anders Johansson
ajj@hig.se

Diana Conache
diana.conache@tum.de

Anders Öberg
anders@math.uu.se

1 School of Computation, Information and Technology, Technische Universität München, Boltzmannstrasse 3, 85748 Garching bei München, Germany

2 Department of Mathematics, University of Gävle, 801 76 Gävle, Sweden

3 Department of Mathematics, Uppsala University, P.O. Box 480, 751 06 Uppsala, Sweden

Published online: 17 February 2025

1 Introduction

In this paper we provide, in terms of the variations of a g-function, the so far best known (and possibly sharp) uniqueness condition for what is usually called a g-measure. Keane [23] introduced the terminology of g-functions and g-measures in a groundbreaking paper and this terminology became standard after further important papers by Ledrappier [25] and Walters [28]. These measures had already appeared in the pioneering probability theory work by Doeblin and Fortet [9], where they proved uniqueness of stationary distributions for what they called chains of complete connections. This is possibly the first time coupling was used in probability theory.

We propose to change the name of g-measures to Doeblin measures, and also g-functions to Doeblin functions. We think the name g-measure is rather confusing, since it just refers to some function g. Also, many authors have used other names than g-measures, such as (stationary distributions of) stochastic chains with unbounded memory ([13]).

Moreover, we would also like to pay tribute to Wolfgang Doeblin, a French Mathematician of German–Jewish origin, who died early in the Second World War, leaving some great contributions to probability theory. We refer to Lindvall's biographical paper [26] for more on Doeblin's life and work.

In Sect. 2, we show that we have a unique Doeblin measure if the Doeblin function $g > 0$ satisfies

$$\limsup_{n\to\infty} \frac{\mathrm{var}_n \log g}{n^{-1/2}} < 2. \tag{1}$$

We were motivated by the earlier result by Johansson et al. [19], where they proved that uniqueness follows if

$$\mathrm{var}_n \log g = o(n^{-1/2}),$$

as $n \to \infty$. We conjecture that the condition (1) is sharp, i.e. that there are systems where a phase transition occurs in analogy with the Ising models with long-range interactions (see e.g. [1]). The best known bound in this direction is that provided in Berger et al. [4] showing that the power $-1/2$ in the denominator of (1) is sharp. To show that there exists a Doeblin function g possessing multiple Doeblin measures for which the left hand side of (1) is finite remains however an open challenge.

In Sect. 3 we treat the well-known problem whether a unique Doeblin measure is mixing, in the sense that uniqueness of a Doeblin measures implies that the sequence of the iterates of the transfer operator converges. We answer this negatively by means of a counterexample, and this also provides an example of a unique Doeblin measure for which even weak mixing fails. Kalikow [22] gave an example of a system that is K (implying uniqueness of a Doeblin measure) but not Bernoulli, that is, that the system is (not) measurably isomorphic to a Bernoulli shift.

For our counterexample in Sect. 3 we use the result of [4], which gives examples of functions $g > 0$ possessing multiple Doeblin measures. We have supplied this paper with an appendix in order to clarify that the construction in [4] gives *precisely* two

extremal (ergodic) measures, since this is important for our construction. This is not known for the construction of more than one Doeblin measure in the paper by Bramson and Kalikow [6].

1.1 Preliminaries

1.1.1 Notation and basic definitions

For a Borel space $(X, \mathcal{B})$ we let $\mathcal{C}(X)$ denote the space of continuous functions on X and $\mathcal{M}(X)$ denote the space of probability measures on X. We write $\delta_x \in \mathcal{M}$ for the Dirac measure at $x \in X$. Assuming the variable x for values in X and given a measure $\mu \in \mathcal{M}(X)$, we write $x \sim \mu$ to (somewhat informally) state that a variable x has (or is sampled with) distribution μ. The expression $\mu(f(x))$ refers to the distribution $\mu \circ f^{-1}$ of the function $f(x)$ defined on X. Similarly, the expression $\mu(f(x)|g(x))$ signifies the *conditional distribution* of $f(x)$ with respect to the sigma-algebra generated by g. We always assume the existence of an underlying probability space Ω with probability measure $A \mapsto \mathsf{P}(A)$ and expectation operator $f \mapsto \mathsf{E}(f)$. We write discrete time stochastic processes defined on Ω as $u^{(n)}$, i.e. with the time index $n \in \mathbb{Z}$ as a superscript in parentheses.

1.1.2 The state space *X*

Let A be a finite set and let $X = A^{\mathbb{N}}$ be the space of one-way infinite strings written left-to right as $x = x_0 x_1 \ldots$. We endow X with the product topology and the corresponding Borel sigma-algebra $\mathcal{B}$. We consider the symbolic shift system (X, T) where $T : X \to X$, $x_0 x_1 \ldots \mapsto Tx = x_1 x_2 \ldots$, is the destructive left shift on X. The fiber at x is thus the set $T^{-1}x = \{ax : a \in A\}$, where ax denotes the string x with prefix $a \in A$ prepended.

We will occasionally use the following *natural extension* $(\bar{X}, \bar{T})$ of (X, T) with projection $\pi : \bar{X} \to X$ such that $\pi \circ \bar{T} = T \circ \pi$. We take $\bar{X} = A^{\mathbb{Z}}$ with the projection $\pi : \bar{X} \to X$ given by

$$\bar{x} = (\bar{x}_i)_{i \in \mathbb{Z}} \mapsto \pi(\bar{x}) = \bar{x}_0 \bar{x}_{-1} \bar{x}_{-2} \cdots$$

and the non-destructive right shift $\bar{T} : \bar{X} \to \bar{X}$ defined by $(\bar{T}\bar{x})_i = \bar{x}_{i-1}$.

For $x \in X$, let $\iota_n(x) \in A^n$ denote the initial prefix of x of length n, i.e. $\iota_n(x) = x_0 x_1 \ldots x_{n-1}$ and let $[x]_n = \iota_n^{-1}(\iota_n(x))$ be the initial cylinder set of length n. Denote by $\mathcal{F}_n$ the sigma-algebra generated by the initial cylinders. For $x, y \in X$, we denote by $\kappa(x, y) \geq 0$ the maximum length of an initial common cylinder, i.e.

$$\kappa(x, y) = \sup\{n \geq 0 : \iota_n(x) = \iota_n(y)\}. \tag{2}$$

The *variation of order* n, $\mathrm{var}_n f$, $n \geq 0$, of a function $f : X \to \mathbb{R}$ is

$$\mathrm{var}_n f = \sup_{x,y}\{|f(x) - f(y)| : \kappa(x, y) \geq n\}.$$

A function $f(x)$ is *continuous* with respect to the product topology precisely if $\mathrm{var}_n\, f \to 0$ as $n \to \infty$.

1.1.3 Doeblin chains and Doeblin measures

A chain with complete (or infinite) connections is a generalization of Markov chains to having a continuous dependence on an infinite past, instead of a finite past. Our study focuses on discrete time stochastic processes $x^{(n)} \in X$ with state space X, such that, for all times n, the previous state $x^{(n-1)}$ is the shift $Tx^{(n)}$ of the current state $x^{(n)}$. Hence, the process is deterministic in the direction of the past and has a stochastic *transition* in the direction of the future: Schematically,

$$x^{(n-1)} \xleftarrow{T} x^{(n)} \rightsquigarrow x^{(n+1)} \tag{3}$$

where $x^{(n+1)} = ax^{(n)}$ is obtained by prepending a random prefix $a \in A$ to $x^{(n)}$. The process $x^{(n)}$ trivially has the Markov property since the current state determines all past states: We can capture the full realisation $\{x^{(n)}\}$ from an element $\bar{x} \in \bar{X}$ by

$$x^{(n)}(\bar{x}) = \pi(\bar{T}^{-n}\,\bar{x}) = \bar{x}_n \bar{x}_{n-1} \ldots .$$

At each transition $x^{(n)} \rightsquigarrow x^{(n+1)}$, we thus *expose* the coordinate $\bar{x}_{n+1}$ of $\bar{x}$.
We also require that, given the current state $x^{(n)}$, the transition should sample the next state $x^{(n+1)}$ in a time invariant manner and have a *continuous* dependence on the past. This means that

$$\mathsf{P}\left(x^{(n+1)} = ax \mid x^{(n)} = x\right) = g(ax) \tag{4}$$

where $g : X \to [0, 1]$ is a continuous function such that

$$\sum_{a \in A} g(ax) = 1 \quad \text{for all } x \in X. \tag{5}$$

Definition 1 We say that $g \in \mathcal{C}(X)$ satisfying (5) is a *Doeblin function* (a.k.a. a g-function) and if $g > 0$ we say that g is *regular*. For a given Doeblin function $g(x)$, we refer to processes $\{x^{(n)}\}$ satisfying (4) as *g-chains*.

Note that (3) is a consequence of (4). Without reference to a fixed Doeblin function g, we speak of *Doeblin chains*. We can specify a Doeblin chain by specifying the *system* (X, T, g) together with an initial distribution $\nu \in \mathcal{M}(X)$ for the state $x^{(0)}$ at time zero. An alternative way to describe g-chains uses the associated *transfer operator*, $L = L_g$, which acts on continuous functions $f \in \mathcal{C}(X)$ according to

$$f(x) \mapsto Lf(x) = \sum_{y \in T^{-1}(x)} g(y) f(y).$$

Then L is a transition operator for the process $x^{(n)}$, i.e. $Lf(x) = \mathsf{E}\left(f(x^{(n+1)}) \mid x^{(n)} = x\right)$ for all n and all functions $f \in \mathcal{C}(X)$. For the dual operator L^* acting on $\mathcal{M}(X)$, it holds that the conditional distribution $\mathsf{P}(x^{(n+1)} \mid x^{(n)}) = L^* \delta_{x^{(n)}}$. A measure $\mu \in \mathcal{M}(X)$ is the distribution of the n-th term in a g-chain governed by (4) if and only if $\mu = L^{*n}(\mu \circ T^{-n})$. We refer to such μ as g-chains of *age* n. Denote by $\mathcal{K}_g^{(n)} \subset \mathcal{M}$ the set of g-chains of age n and $\mathcal{K}_g = \cap_{n=1}^{\infty} \mathcal{K}_g^{(n)}$ the space $\mathcal{K}_g \subset \mathcal{M}(X)$ of g-chains of infinite age.

Definition 2 A *Doeblin measure* is the invariant distribution of a *stationary* g-chain.

Equivalently, a Doeblin measure $\mu \in \mathcal{M}(X)$ is a fixed point to the dual transfer operator L^*, i.e. $\mu = L^*\mu$. We denote by $\mathcal{M}_g$ the space of Doeblin measures corresponding to g. A Doeblin measure μ is translation invariant, i.e. $\mathcal{M}_g \subset \mathcal{M}_T$, where $\mathcal{M}_T := \{\mu : \mu = \mu \circ T^{-1}\}$. Note that a translation invariant measure μ is a g-chain of infinite age and that $\mathcal{M}_g = \mathcal{M}_T \cap \mathcal{K}_g^{(1)}$.

Our first result, Theorem 1, concerns regularity conditions on g ensuring that $\mathcal{M}_g = \{\mu\}$, i.e. conditions implying the uniqueness of the Doeblin measure. In our second result, Theorem 5, we construct an example of a system $(X, T, \tilde{g})$, where we have multiple Doeblin chains but a unique Doeblin measure, i.e. $\mathcal{K}_{\tilde{g}} \supsetneq \mathcal{M}_{\tilde{g}} = \{\mu\}$.

1.2 Previous results on uniqueness and non-uniqueness

We say that $g(x)$ is *local* if g is $\mathcal{F}_N$-measurable for some N, i.e. determined by a finite set of coordinates. For a local Doeblin function $g(x)$ the corresponding g-chains $x^{(n)}$ reduce to finite-state Markov chains $\iota_N(x^{(n)})$ with state space A^N and g is therefore uniquely ergodic if we furthermore assume that it is *regular*, i.e. that $g > 0$.
For a system (X, T, g), the existence of a Doeblin measure follows from the Schauder-Tychonoff fixed-point theorem, since the compact convex set $\mathcal{M}_T$ of translation invariant measures maps to itself under the action of L^*. Note that the requirement that the Doeblin function g is continuous is essential. Consider for example the "discontinuous Doeblin function" $g(x)$, $x \in \{-1, 1\}^{\mathbb{N}}$, given by

$$g(x) = 0.5 - 0.25 \cdot x_0 \cdot \mathbf{1}_{M(Tx) \geq 0},$$

where $M(Tx) = \limsup_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} x_k$. For this g the space $\mathcal{M}_g$ is empty, since, for an ergodic $\mu \in \mathcal{M}_g$, we have that $M(x) = M(Tx)$ and is equal to the constant $\mu(x_0) = \mu(\mu(x_0|Tx))$, μ-almost surely, which gives the contradiction

$$M(x) = \begin{cases} 0.25(+1) + 0.75(-1) = -0.5 & \text{if } M(x) \geq 0 \\ 0.5(+1) + 0.5(-1) = 0 & \text{if } M(x) < 0 \end{cases}.$$

Doeblin and Fortet [9] proved that the Doeblin measure is unique whenever the regular Doeblin function g satisfies "summability of variations"

$$\sum_{n=1}^{\infty} \operatorname{var}_n g < \infty.$$

In [16], Johansson and Öberg showed that the condition of "square summability" (or ℓ^2-summability)

$$\sum_{n=1}^{\infty} (\mathrm{var}_n\, g)^2 < \infty, \tag{6}$$

implies uniqueness. In [18] Johansson, Öberg and Pollicott extended this result to countable state shifts.

The condition of square summable variations was further explored by Gallesco, Gallo and Takahashi [13]. They also made a distinction between equilibrium and dynamic uniqueness (similar to the distinction between weak and strong uniqueness of Gibbs measures of disordered spin systems in rigorous statistical mechanics [11]). Equilibrium uniqueness is what we call uniqueness of a Doeblin measure, and dynamic uniqueness is a stronger notion. They show that a.e. square summable variations (first introduced in [16]) is equivalent to dynamic uniqueness and also to the weak Bernoulli property.

Berbee's results from the late 1980s [2, 3] are also intriguing in this context. He proved that there is a unique Doeblin measure if

$$\sum_{n=1}^{\infty} e^{-r_1-\cdots-r_n} = \infty, \tag{7}$$

where $r_n = \mathrm{var}_n \log g$ denotes the (logarithmic) variations of the Doeblin function. This condition for uniqueness is not comparable with that of (6), i.e., there are cases where one is valid but not the other, although it is worth noting that Berbee's condition is valid for $r_n = 1/n$, but not for $r_n = 1/n^\alpha$, if $\alpha < 1$, whereas (6) is valid if $\alpha > 1/2$. Also, Berbee's condition is not valid if $r_n = c/n$ and if $c > 1$.

In [19], the authors found a condition for uniqueness that subsumes Berbee's condition (7) for uniqueness with square summable variations (6). As a corollary, uniqueness was obtained under the condition

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

Fernandez and Maillard [12] proved that uniqueness and mixing properties for a Doeblin measure (and some generalisations) follow if the Doeblin function satisfies the classical Dobrushin condition, which is incomparable with variational conditions. In statistical mechanics, Bissacot et al. [5] show that the "g-measure property" is violated for certain Gibbs measures in low temperatures. Other papers that draw inspiration from statistical mechanics are Verbitskiy [27], Johansson et al. [20], and the forthcoming [21]. In the last two papers the difference between conditions for uniqueness of Gibbs measures for general potentials and the uniqueness of Doeblin measures are further explored and to some extent clarified.

Concerning non-uniqueness of Doeblin measures, we refer to the breakthrough of Bramson and Kalikow [6], who provided the first counterexample to uniqueness of

Doeblin measures consistent with a positive Doeblin function, although the Doeblin function is continuous. The paper by Berger et al. [4] was a direct response to [16] and provided non-unique systems where the variations of the Doeblin functions are in $\ell^{2+\epsilon}$ for $\epsilon > 0$, proving that the square summable variations condition is sharp for uniqueness in this sense. We would also like to mention the paper of Hulse [14], where he demonstrates, using an example, that the Dobrushin type condition in [12] cannot be essentially improved.

1.3 Our results

We now present the main two results of this paper.

Theorem 1 *Let g be a regular Doeblin function that satisfies (1), namely*

$$\limsup_{n\to\infty} \frac{r_n}{\sqrt{n}^{-1}} < 2, \text{ with } r_n := \operatorname{var}_n \log g.$$

Then g possesses a unique Doeblin measure.

Theorem 2 *There exists a Doeblin function with a unique Doeblin measure which is not weakly mixing.*

This implies that the sequence of the iterates of the transfer operator does not converge.

Remark 1 It is natural to ask whether there exists a positive constant c^* such that if

$$\operatorname{var}_n \log g \le \frac{c^*}{\sqrt{n}}, \text{ for all } n \text{ large},$$

then uniqueness holds, but for all larger constants one can find an example of non-uniqueness, i.e. whether a phase transition occurs when the variations of $\log g$ are of order $1/\sqrt{n}$. We conjecture that this is indeed the case. In Theorem 1, we provide 2 as a lower bound for c^*. If a phase transition is established, we might ask whether 2 is indeed the critical value.

Remark 2 Sometimes authors use the variations of g, $\operatorname{var}_n g$ rather than $\operatorname{var}_n \log g$, in their conditions for uniqueness (e.g. in Doeblin and Fortet [9] and in Johansson and Öberg [17]) or for counterexamples (Berger et al. [4]), but these conditions are many times equivalent, owing to the compactness of X and that $g > 0$. Using $\operatorname{var}_n \log g$ conforms more to the thermodynamic formalism of ergodic theory, since for various conditions one considers $\operatorname{var}_n \phi$, where the function e^{ϕ} is the weight function of the transfer operator. The choice between using $\operatorname{var}_n g$ or $\operatorname{var}_n \log g$ is usually just a matter of taste or context, and even in Johansson et al. [17] one can replace the uniqueness condition $\operatorname{var}_n \log g = o(1/\sqrt{n})$, $n \to \infty$, with the equivalent condition $\operatorname{var}_n g = o(1/\sqrt{n})$, $n \to \infty$. However, in our context, the two conditions are not equivalent, and we ask whether Theorem 1 should hold with $\operatorname{var}_n g$ instead of $\operatorname{var}_n \log g$ and with possibly a different constant.

Remark 3 In Johansson et al. [19] the Bernoulli property follows when $\mathrm{var}_n \log g = o(1/\sqrt{n})$, $n \to \infty$. This consequence does not follow from the method of proof in the present paper. As was clear already from Kalikow's paper [22], the Bernoulli property is not automatic for a unique Doeblin measure. We ask under what conditions uniqueness implies the Bernoulli property (very weak Bernoulli property).

2 Proof of Theorem 1

We start by constructing our main tool, namely the block-coupling.

2.1 The block-coupling

The block coupling that we use here was first developed in [19], where the authors used it to estimate $\bar{d}$-metrics between various chains. In this paper we use the same block coupling, but for a different purpose, and in particular we do not use (and thus also do not define) $\bar{d}$-metrics. We start here by defining the block coupling, then, in Sect. 2.2, we estimate its success probability, and in subsequent subsections we use it to prove Theorem 1.

2.1.1 The block extension ξ_x^b and the block-coupling $\zeta_{x,\tilde{x}}^b$

Let $b \geq 1$ be fixed. For a g-chain $x^{(n)}$, conditioned on the current state $x^{(n)} = x$, the distribution of the state $y = x^{(n+b)}$, at b time steps in the future, is $L^{*b}\delta_x$. The transition from x to y means prepending a randomly sampled *prefix* $p = \iota_b(y) \in A^b$ with distribution $p \sim \xi_x^b = L^{*b}\delta_x \circ \iota_b^{-1}$. This means that we sample the prefix $p \in A^b$ with probability

$$\xi_x^b(p) = \prod_{k=0}^{b-1} g(T^k(px)). \tag{8}$$

The intermediate states are then sampled with the right distribution, i.e. $x^{(n+i)} \sim L^{*i}\delta_x$, for $1 \leq i \leq b$. We refer to the sampling of $y \in \{px : p \in A^b\}$ as a *block extension*. In the picture of the natural extension $\bar{x}$, it amounts to exposing a *block* $\bar{x}_{n+1}\bar{x}_{n+2}\cdots\bar{x}_{n+b}$ of b coordinates of the g-chain $\bar{x} \in \bar{X}$ in one step.
Now consider a pair of g-chains $(x^{(n)}, \tilde{x}^{(n)})$ and two present states $x = x^{(n)}$ and $\tilde{x} = \tilde{x}^{(n)}$ in X. A *block-coupling* of length $b \geq 1$ at $(x, \tilde{x})$ is a sampling of $(y, \tilde{y}) = (x^{(n+b)}, \tilde{x}^{(n+b)})$ such that the marginal distributions of $(y, \tilde{y})$ are block extensions. That is, a distribution $\zeta(y, \tilde{y}) \in \mathcal{M}(T^{-b}x \times T^{-b}\tilde{x})$, such that

$$\zeta(y) = L^{*b}\delta_x \text{ and } \zeta(\tilde{y}) = L^{*b}\delta_{\tilde{x}}.$$

The event of *success* means that

$$(y, \tilde{y}) \in S := \{(y, \tilde{y}) \mid \iota_b(y) = \iota_b(\tilde{y})\} = (\iota_b \times \iota_b)^{-1}(\{(p, p) | p \in A^b\})$$

means that the prefixes, $p = \iota_b(y)$ and $\tilde{p} = \iota_b(\tilde{y})$, prepended to x and $\tilde{x}$ are equal. It is a standard result that we can find a coupling $\zeta^b_{x,\tilde{x}}$ such that

$$\zeta^b_{x,\tilde{x}}(S) = 1 - d_{\mathrm{TV}}(\xi^b_x, \xi^b_{\tilde{x}}), \tag{9}$$

where d_{TV} is the total variation distance and ξ^b_x and $\xi^b_{\tilde{x}}$ are as in (8). Moreover, the event S of success means that the agreement length κ from (2) increases, i.e. for $0 \le i \le b$ we have

$$(y, \tilde{y}) \in S \implies \kappa(T^i y, T^i \tilde{y}) = \kappa(x, \tilde{x}) + b - i. \tag{10}$$

An important observation from [19] is that, for longer blocks, the total variation distance $d_{\mathrm{TV}}(\xi, \tilde{\xi})$ is bounded by a function of the sum of the squares of the variations of $\log g$ rather than, say, a function of the sum of the variations. This observation will be made precise and will be used in Sect. 2.2 in order utilize our variational condition for the purpose of controlling the coupling.

2.1.2 The associated Markov chain

The way in which we use the block coupling is to construct a pair of coupled g-chains $(x^{(n)}, \tilde{x}^{(n)})$, $n \ge 0$, that agree in the coordinates to a large extent. Similarly to [16], we associate the pair $(x^{(n)}, \tilde{x}^{(n)})$ of coupled g-chains with a Markov chain $Y^{(n)}$ with state space $\mathbb{Z}$. The chain $Y^{(n)}$ dominates the process $\kappa(x^{(n)}, \tilde{x}^{(n)})$ such that for all $n \ge 0$

$$\kappa(x^{(n)}, \tilde{x}^{(n)}) \ge Y^{(n)}. \tag{11}$$

Let $\mathcal{B} = \{B_\ell : \ell \ge 0\} \subseteq \mathbb{N}_0$ be a set of states, which we call "breaking states", ordered so that $0 = B_0 < B_1 < B_2 < \dots$. If the Markov chain $Y^{(n)}$ is in a state outside of $\mathcal{B}$, then it gets incremented by 1 with probability 1. If the Markov chain is in a state in $\mathcal{B}$, then the transition is more complex. Let $b_\ell = B_{\ell+1} - B_\ell$ be the distance between two consecutive breaks. Let also

$$p_\ell = \min\left\{1/2, \inf\{1 - d_{\mathrm{TV}}(\xi^{b_\ell}_x, \xi^{b_\ell}_{\tilde{x}}) \mid \kappa(x, \tilde{x}) \ge B_\ell\}\right\}. \tag{12}$$

We start the chain at $Y^{(0)} = 0$ and, for $n \ge 0$, we let

$$Y^{(n+1)} = \begin{cases} Y^{(n)} + 1 & \text{w. prob. 1 if } Y^{(n)} \notin \mathcal{B} \\ Y^{(n)} + 1 & \text{w. prob. } p_\ell \text{if } Y^{(n)} = B_\ell \in \mathcal{B} \\ -b_\ell + 1 & \text{w. prob. } 1 - p_\ell \text{ if } Y^{(n)} = B_\ell \in \mathcal{B}. \end{cases} \tag{13}$$

Thus, the chain $Y^{(n)}$ moves deterministically upwards at unit speed, i.e. $Y^{(n+1)} = Y^{(n)} + 1$, whenever $Y^{(n)} \notin \mathcal{B}$. In case it reaches a breaking state $Y^{(n)} = B_\ell$, it continues upwards with probability p_ℓ and with probability $1 - p_\ell$, it jumps to the state $-b_\ell + 1$ below zero.

If $Y^{(n)} = B_\ell$ then (11) implies $\kappa(x^{(n)}, \tilde{x}^{(n)}) \geq B_\ell$. Hence, the definition (12) implies that

$$p_\ell \leq 1 - d_{\mathrm{TV}}(\xi^{b_\ell}_{x^{(n)}}, \xi^{b_\ell}_{\tilde{x}^{(n)}}).$$

Hence, we may couple the event $Y^{(n+1)} = Y^{(n)} + 1$ in state B_ℓ with a *successful* block extension $(x^{(n+b_\ell)}, \tilde{x}^{(n+b_\ell)})$ of the chains. In this case, the chain $Y^{(m)}$ moves deterministically upwards until it reaches the next state $B_{\ell+1} = B_\ell + b_\ell$ at time $n + b_\ell$ and relation (10) implies that (11) remains valid for times m where $n \leq m \leq n + b_\ell$. In case $Y^{(n+1)} = -b_\ell + 1$, the chain moves deterministically upwards until it hits state zero at time $n + b_\ell$. In this case (11) remains trivially true for $n \leq m \leq n + b_\ell$. Lemma 3 below shows that there exists B_1, such that the definition $B_0 = 0$ and for $\ell = 1, 2, \ldots$

$$B_\ell := B_1 \cdot K^{\ell-1}, \quad \text{i.e. } b_\ell = (K-1)B_\ell. \tag{14}$$

implies

$$p_\ell > \frac{1}{K} \tag{15}$$

for any $K \geq 2$. Thus, we define the breaking states $\mathcal{B}$ according to (14) above.
In Sect. 2.2 we state and prove Lemma 3 which we referred to before. In Sect. 2.3, we use renewal theory and a theorem by Kesten [24] to further study the chain $Y^{(n)}$. Then in Sect. 2.4 we use this construction to prove Theorem 1.

2.2 Bounding the total variation distance

In order to provide the constant 2 in our theorem, we use some familiar estimates of the total variation metric based on the Hellinger integral (cf. [15, 19]). The Hellinger integral is useful for us because of its use to bound the total variation distance. Indeed, let $\xi = \xi^b_x$ and $\tilde{\xi} = \xi^b_{\tilde{x}}$ be as in (9) so that the probability of a successful coupling is at least $1 - d_{\mathrm{TV}}(\xi, \tilde{\xi})$. Using Cauchy-Schwarz, it is easy to verify (as in e.g. [15]) that

$$1 - d_{\mathrm{TV}}(\xi, \tilde{\xi}) \geq 1 - \sqrt{1 - H^2} \geq \frac{H^2}{2}, \tag{16}$$

where, for the discrete measures ξ and $\tilde{\xi}$, the Hellinger integral has the form

$$H = H(\xi, \tilde{\xi}) = \sum_{p \in A^b} \prod_{j=0}^{b-1} \sqrt{g(T^j y) g(T^j \tilde{y})}. \tag{17}$$

Lemma 3 *Assume (1) holds. It holds for integers B and K sufficiently large and $b = (K-1)B$, that*

$$\inf\left\{1 - d_{TV}(\xi_x^b, \xi_{\tilde{x}}^b) : \kappa(x,\tilde{x}) \geq B\right\} > \frac{1}{K}. \tag{18}$$

Proof We have for all $j \in \{0,\ldots,b-1\}$

$$\sqrt{g(T^j y)\, g(T^j \tilde{y})} = \frac{g(T^j y) + g(T^j \tilde{y})}{2} \cdot \frac{1}{\cosh(\delta_j)}, \tag{19}$$

where

$$\delta_j(u) = \left|\frac{\log g(T^j y) - \log g(T^j \tilde{y})}{2}\right| \leq \frac{r_{B+b-j}}{2}, \tag{20}$$

since, by (10), $\kappa(T^j y, T^j \tilde{y}) \geq B + b - j$ when $\kappa(x,\tilde{x}) \geq B$. Since, for all $w, \tilde{w} \in X$, we have $\sum_{a\in A}(g(aw) + g(a\tilde{w}))/2 = 1$, it follows that

$$\sum_{p\in A^b} \prod_{j=0}^{b-1} \frac{g(T^j y) + g(T^j \tilde{y})}{2} = 1. \tag{21}$$

Since $B + b = KB$, we deduce from (17)–(21) that

$$H \geq \exp\left(-\sum_{k=B+1}^{KB} \log\cosh(r_k/2)\right).$$

Since $\log\cosh(u/2) = \frac{1}{8}u^2 + O\left(u^4\right)$, we have for any $c > \limsup r_n\sqrt{n}$ and a constant $c_1 \geq 0$ that

$$\begin{aligned}\log H &\geq -\frac{1}{8}\sum_{k=B+1}^{KB} r_k^2 - c_1 \sum_{k=B+1}^{KB} r_k^4 \\ &\geq -\frac{c^2}{8}\cdot\sum_{k=B+1}^{KB}\frac{1}{k} - c^4\cdot c_1 \sum_{k=B+1}^{KB}\frac{1}{k^2},\end{aligned}$$

where the second inequality follows from choosing B large enough.
Using the formula

$$\sum_{k=1}^{n}\frac{1}{k} = \log n + \gamma + \frac{1}{2n} + O\left(\frac{1}{n^2}\right)$$

for the partial sums of the harmonic series and the integral estimate

$$\sum_{k=B+1}^{KB} \frac{1}{k^2} \leq \frac{1}{B+1} = O\left(1/B\right),$$

we obtain $\log H \geq -\frac{c^2}{8} \log K - \frac{c_2}{B}$, for some constant $c_2 \geq 0$.
Thus

$$H^2 \geq \frac{2}{K} \cdot \exp\left(\epsilon \log K - \log 2 - \frac{2c_2}{B}\right), \tag{22}$$

where $\epsilon := 1 - c^2/4$, which is positive by assumption (1). We can choose $K \geq 2$ large enough so that $\epsilon \log K - \log 2 - \frac{2c_2}{B} > \epsilon/2$. Using (16) with the bound (22) we obtain

$$1 - d_{\mathrm{TV}} \geq \frac{1}{K} \cdot e^{\epsilon/2} > \frac{1}{K},$$

which concludes the proof of Lemma 3. □

2.3 A property of the associated Markov chain $Y^{(n)}$

Proposition 4 *Under the assumptions of Lemma 3*

$$\limsup_{n\to\infty} \frac{Y^{(n)}}{n} = 1 \ a.s. \tag{23}$$

Proof of Proposition 4 In order to establish (23), we let T_k, $k \geq 0$, denote the renewal process of return times to state 0 of the Markov chain $Y^{(n)}$, i.e. $T_0 = 0$ and $T_k = \inf\{n > T_{k-1} : Y_n = 0\}$. The differences $M_k = T_k - T_{k-1}$, $k = 1, 2, \ldots$, constitute an iid sequence, such that $M_k < \infty$ a.s.
Note that, $M_k = B_{L_k+1}$ where B_{L_k} is the last visited breaking state by $Y^{(n)}$ in excursion number k, i.e. $L_k := \sup\{\ell \in \mathbb{Z}_{\geq 0} : Y^{(n)} = B_\ell, n \in \left[T_{k-1}, T_k\right)\}$. If $L_k = 0$ then $M_k = B_1$ and, otherwise, if $L_k \geq 1$, we have

$$\sup\{Y^{(n)} : T_{k-1} < n \leq T_k\} = B_{L_k} = M_k/K. \tag{24}$$

Since, a.s., $L_k \geq 1$ infinitely often, we have, almost surely,

$$\begin{aligned}\limsup_{n\to\infty} \frac{Y^{(n)}}{n} &\geq \limsup_k \frac{M_k/K}{T_{k-1} + M_k/K} \\ &\geq 1 - \liminf \frac{K \cdot T_{k-1}}{M_k},\end{aligned}$$

which equals 1 almost surely if and only if

$$\limsup \frac{M_k}{T_{k-1}} = \limsup \frac{M_k}{M_1 + M_2 + \cdots + M_{k-1}} = \infty, \quad \text{a.s.} \tag{25}$$

A result by Kesten ([24], 1971) says that (25) holds as soon as

$$\mathsf{E}\,(M_k) = \infty, \tag{26}$$

which is true in our case. We have

$$\mathsf{E}(M_k) = \mathsf{E}(K^{L_k} B_1) = \sum_{\ell=0}^{\infty} \mathsf{P}(L_k = \ell) K^{\ell} B_1 = \infty,$$

since, for $\ell \geq 1$,

$$\mathsf{P}(L_k = \ell) = \left(\prod_{i=0}^{\ell-1} p_i \right) \cdot (1 - p_\ell) > p_0 \Big(\frac{1}{K} \Big)^{\ell-1} \cdot \frac{1}{2},$$

where last inequality follow from Lemma 3 and $p_\ell \leq \frac{1}{2}$ by (12). This establishes that the construction satisfies (23). □

2.4 Finishing the proof of Theorem 1

Proof of Theorem 1 Assume for contradiction that there are two distinct Doeblin measures μ and $\tilde{\mu}$ with respect to g. Since the space $\mathcal{M}_g$ of Doeblin measures with respect to g is convex and compact, we may assume that μ and $\tilde{\mu}$ are extremal, i.e. *ergodic*.

Now, we choose the initial states $x^{(0)}$ resp. $\tilde{x}^{(0)}$ independently according to the distributions μ resp. $\tilde{\mu}$, and then complete them to chains $x^{(n)}, \tilde{x}^{(n)} \in X$, $n \geq 1$, according to the coupling from Sect. 2.1. Since we sample the initial states using stationary distributions, $x^{(n)}$ and $\tilde{x}^{(n)}$ have distribution μ and $\tilde{\mu}$, for all $n \in \mathbb{Z}$.

Since $\mu \neq \tilde{\mu}$, we have some local function f such that $\int f \, d\mu > \int f \, d\tilde{\mu}$. Let N be s.t. f is in $\mathcal{F}_N$. Then, by the pointwise Birkhoff Ergodic Theorem, almost surely, the following limits exist

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(x^{(i)}) = \int f \, d\mu > \int f \, d\tilde{\mu} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(\tilde{x}^{(i)}).$$

In particular, it must hold almost surely that

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \big(f(x^{(i)}) - f(\tilde{x}^{(i)}) \big) > 0. \tag{27}$$

The fact that f is $\mathcal{F}_N$-measurable means that $f(x^{(i)}) - f(\tilde{x}^{(i)}) = 0$ if $\kappa(x^{(i)}, \tilde{x}^{(i)}) \geq N$. Thus, by (11), we have

$$\frac{1}{n}\sum_{i=1}^{n}\left(f(x^{(i)}) - f(\tilde{x}^{(i)})\right) \leq \frac{1}{n}\sum_{i=1}^{n} 2\|f\|_\infty \cdot 1_{Y^{(i)}\leq N} \leq 2\|f\|_\infty \frac{n - Y^{(n)} + N}{n}.$$

However, (23) implies that the right hand side above is arbitrarily close to zero for infinitely many n, in contradiction to (27). We conclude that there can only be one Doeblin measure corresponding to g. □

Remark 4 Here we use ergodicity in the sense of stochastic processes, i.e. that

$$\lim \frac{1}{n}\sum_{k=1}^{n} f(x^{(k)}) = \int f\, d\mu \quad \mu\text{ -a.e. },$$

i.e. in terms of iterations over "future" values. In ergodic theory, Birkhoff's theorem is usually stated as

$$\lim \frac{1}{n}\sum_{k=1}^{n} f(T^k x) = \int f\, d\mu \quad \mu\text{ -a.e. },$$

i.e. in terms of iterations over "past" values. But, for Doeblin measures, we have equivalence between these two statements, since

$$(X, T, \mu) \text{ ergodic} \iff (\bar{X}, \bar{T}, \bar{\mu}) \text{ ergodic} \iff (\bar{X}, \bar{T}^{-1}, \bar{\mu}) \text{ ergodic},$$

where $(\bar{X}, \bar{T}, \bar{\mu})$, $\pi : \bar{X} \to X$, is the natural extension of (X, T, μ). But then,

$$\lim \frac{1}{n}\sum_{k=1}^{n} f(x^{(k)}) = \lim \frac{1}{n}\sum_{k=1}^{n} f(\pi(\bar{T}^{-k}\bar{x})) \underset{\text{a.e.}}{=} \int f(\pi(\bar{x}))d\,\bar{\mu}(\bar{x}) = \int f\, d\mu.$$

3 A non-mixing unique Doeblin measure

In this section, we answer a well-known folklore question that we in particular attribute to Peter Walters, namely whether uniqueness of a Doeblin measure (in the case of finitely many symbols and $g > 0$) implies that the iterates of the transfer operator L_g converge. In fact we show by means of an example that this is not necessary and that the unique Doeblin measure in our construction is not even weakly mixing.
Recall that the transfer operator here is defined for continuous functions f by

$$L_g f(x) = \sum_{y \in T^{-1}x} g(y) f(y).$$

We know from Breiman [7] that a unique Doeblin measure μ implies that for any continuous function $f : X \to \mathbb{R}$, we have

$$\lim_{n\to\infty} \sup_x \left| \frac{1}{n} \sum_{k=1}^{n} L_g^k f(x) - \int f \, d\mu \right| = 0. \tag{28}$$

The full uniform convergence of the iterates of the transfer operator implies *strong mixing* for any Doeblin measure μ. That is if

$$\lim_{n\to\infty} \sup_x \left| L^n f(x) - \int f \, d\mu \right| = 0,$$

then we also have the strong mixing of μ:

$$\mu(T^{-n} A \cap B) \to \mu(A)\mu(B),$$

as $n \to \infty$, for any Borel subsets A and B of X.

We construct a unique Doeblin measure μ that is not even *weakly mixing*, hence not strongly mixing, and then the iterates of the transfer operator do not necessarily converge (uniformly or pointwise). Weak mixing for a measure μ means that

$$\frac{1}{n} \sum_{k=1}^{n} |\mu(T^{-n} A \cap B) - \mu(A)\mu(B)| = 0,$$

for any Borel subsets A and B of X. Observe that this does not follow from Breiman's property (28).

3.1 The construction

Let $\mathcal{K}_g$ denote the set of g-chains (of infinite age) on the state space X. The members of $\mathcal{K}_g$ have also been studied under the name G-measures, see e.g. [8, 10]. The g-chain condition takes the form

$$\mu(x \mid T^n x) = g(x) \cdot g(Tx) \cdots g(T^{n-1} x) \quad \text{for all } n \geq 0. \tag{29}$$

The space of Doeblin measures $\mathcal{M}_g(X)$ is the set of *stationary* g-chains. Our aim is to demonstrate that there is a system $(X, T, \tilde{g})$ where the Doeblin measure is unique but there are multiple g-chains: That is,

$$\mathcal{M}_g = \{\mu\} \text{ and } \mathcal{M}_g \subsetneq \mathcal{K}_g. \tag{30}$$

Our aim is to show the following theorem.

Theorem 5 *There exists a continuous Doeblin function $g > 0$ on $X = A^{\mathbb{N}}$, $|A| < \infty$, for which (30) holds.*

Since T is the full shift on X, we know from Walters [29], p. 334, p. 340) that we have a unique measure in $\mathcal{K}_g$ if and only if we have uniform (or equivalently: pointwise) convergence of the sequence $\{L^n f(x)\}$. Hence we have the following corollary.

Corollary 6 *There exists a Doeblin function g with a unique Doeblin measure such that we do not have pointwise convergence of the sequence of iterates of the associated transfer operator L_g.*

Our result also implies that there exists a system with a unique Doeblin measure which is not Bernoulli, but that was already proved by Kalikow [22].

3.2 The proof of Theorem 5

Recall that an involution on X is a bijection $F : X \to X$ such that $F \neq \mathsf{I}$ and $F^2 = \mathsf{I}$, where I is the identity map. If a Doeblin function $g(x)$ is invariant under an involution M commuting with T then both $\mathcal{M}_g$ and $\mathcal{K}_g$ are invariant under the direct image $M_* : \mathcal{M}(X) \to \mathcal{M}(X), \nu \mapsto \nu \circ M^{-1}$. We look for a triple (X, T, g) with two distinct ergodic measures ν_+ and ν_- conjugated by an involution symmetry M. That is, we have

$$g(Mx) = g(x), \quad MT = TM, \quad M^2 = \mathsf{I} \quad \text{and} \quad \nu_- = \nu_+ \circ M^{-1}. \tag{31}$$

We also assume that there is a second involution $F : X \to X$, $F \neq M$, with the commutation rules

$$MF = FM \quad \text{and} \quad FT = MTF. \tag{32}$$

In addition, we also assume that ν_+ and ν_- are the *only* extremal g-chains, so that $\mathcal{M}_g$ and $\mathcal{K}_g$ are the convex hull

$$\mathcal{K}_g = \mathcal{M}_g = \text{conv}\{\nu_+, \nu_-\}. \tag{33}$$

A candidate for a system that satisfies (31), (32) and (33), would be the Doeblin function constructed by Bramson and Kalikow [6] on $X = \{+1, -1\}^{\mathbb{N}}$ with T the usual shift. With $M[(x_i)] = (-x_i)$ and $F[(x_i)] = \left((-1)^i x_i\right)$ it is easy to verify (31) and (32). However, the condition (33) seems hard to prove. Instead, we use below a system on four symbols constructed in [4], originally for the purpose of providing a system with multiple Doeblin measures, but with $\ell^{2+\epsilon}$-summable variations of g.

Lemma 7 *Assume a system (X, T, g) satisfies (31), (32) and (33). Let $\tilde{g}(x) = g(Fx)$ and $\tilde{\nu}_\pm = \nu_\pm \circ F^{-1}$. Then, for the system $(X, T, \tilde{g})$, we have*

$$\mathcal{M}_{\tilde{g}} = \{\frac{1}{2}(\tilde{\nu}_+ + \tilde{\nu}_-)\} \quad \textit{and} \quad \mathcal{K}_{\tilde{g}} = \text{conv}\{\tilde{\nu}_+, \tilde{\nu}_-\}.$$

Proof We claim that the map $\nu \mapsto \tilde{\nu} = \nu \circ F^{-1}$ takes $\mathcal{K}_g$ bijectively onto $\mathcal{K}_{\tilde{g}}$. To see this, let $\mu \in \mathcal{K}_g$ be a g-chain and note that

$$\tilde{\mu}(x \mid T^k x) = \mu(Fx|T^k Fx)$$

Then it follows that

$$\begin{aligned}\tilde{\mu}(x \mid T^n x) &= g(Fx) \cdot g(TFx) \cdot g(T^2 Fx) \cdots g(T^{n-1} Fx) \\ &= g(Fx) \cdot g(MFTx) \cdot g(M^2 FT^2 x) \cdots g(M^{n-1} FT^{n-1} x) \\ &= \tilde{g}(x) \cdot \tilde{g}(Tx) \cdot \tilde{g}(T^2 x) \cdots \tilde{g}(T^{n-1} x).\end{aligned}$$

Since this holds for all $k \geq 0$, it shows that $\tilde{\mu}$ is a $\tilde{g}$-chain. The first equality is due to μ being a g-chain. The second equality follows from the rule (32). The final equality is due to $g(M^k x) = g(x)$ by the symmetry and that $g(Fx) = \tilde{g}(x)$ by definition. Since F is an involution, it is clear that the argument gives a bijection between $\mathcal{K}_g$ and $\mathcal{K}_{\tilde{g}}$. None of the measures $\tilde{\nu}_+$ and $\tilde{\nu}_-$ are translation invariant since

$$\tilde{\nu}_+ \circ T^{-1} = \nu_+ \circ F \circ T^{-1} = \nu_+ \circ M \circ F = \nu_- \circ F = \tilde{\nu}_-.$$

Thus chains with distribution $\tilde{\nu}_\pm$ are periodic in distribution with period 2. It is then clear that the only candidate for a stationary $\tilde{g}$-chain in $\mathcal{K}_{\tilde{g}} = \text{conv}\{\tilde{\nu}_+, \tilde{\nu}_-\}$ is the midpoint $\tilde{\mu} = (\tilde{\nu}_+ + \tilde{\nu}_-)/2$ between them. □

To prove Theorem 5, we use the following counterexample to uniqueness of a Doeblin measure based on the example provided in [4]. We consider an alphabet $A = \{0, 1\} \times \{-1, 1\}$ and the input to the Doeblin function g as two sequences: $(x_n)_{n\geq 0}$ with values in $\{-1, 1\}$ and $(y_n)_{n\geq 0}$ with values in $\{0, 1\}$.
We construct a Doeblin function g as a mechanism for sampling x_0 and y_0 given $(x_n)_{n\geq 1}$ and $(y_n)_{n\geq 1}$, using the following scheme:

(1) Sample y_0 as a $1/2$-Bernoulli variable independently of $(x_n, y_n)_{n\geq 1}$.
(2) Using $(y_n)_{n\geq 1}$ choose a finite set $S \subseteq \{k : k \geq 1\}$. Either $|S|$ is odd or $S = \emptyset$.
(3) Using $(y_n)_{n\geq 1}$ choose a number ξ between .5 and .75.
(4) If $S \neq \emptyset$, then take $x_0 = \text{sgn} \sum_{k\in S} x_k$ with probability ξ and $x_0 = -\text{sgn} \sum_{k\in S} x_k$ with probability $1 - \xi$. If $S = \emptyset$, choose $x_0 = 1$ with probability $1/2$ and $x_0 = -1$ with probability $1/2$.

The following theorem is the main result of [4] in combination with the appendix of the present paper, where we show, for the sake of clarification, that there are precisely two ergodic measures in this example.

Theorem 8 [4] *There exists a way to choose the set S and the parameter ξ such that the Doeblin function g has exactly two ergodic (extremal) Doeblin measures μ_+ and μ_-. Moreover, μ_+ and μ_- are also the only extremal Doeblin chains for g.*

Now, we define the action of M and F as in the Bramson-Kalikow example mentioned above, but restricted to the x-sequence. That is, let $M((x_i, y_i)_{i=1}^\infty) = (-x_i, y_i)_{i=0}^\infty$

and $F((x_i, y_i)_{i=1}^{\infty}) = ((-1)^i x_i, y_i)_{i=1}^{\infty}$. It is then straightforward to verify that the commutation rules (31) and (32) hold. From Theorem 8 we also demonstrate (33). An application of Lemma 7 thus concludes the proof. □

Remark 5 It is still an open question whether the two measures constructed by Bramson and Kalikow [6] are the only extremal measures in K_g. In the affirmative case, their construction can also be adapted to provide a similar counterexample.

3.3 Proof of Theorem 2

We conclude by giving the proof of Theorem 2 since it uses the construction in the proof of Theorem 5. We claim that the translation invariant measure $\tilde{\mu} = (\tilde{\nu}_+ + \tilde{\nu}_-)/2$ constructed above is not weakly mixing. Let A be a set on which $\tilde{\nu}_+$ is concentrated and such that $\tilde{\nu}_-(A) = 0$ ($\tilde{\nu}_+$ and $\tilde{\nu}_-$ are mutually singular). We obtain an oscillating sequence

$$\mu((T^{-n}A) \cap A) = \nu_+(T^{-n}A) = \begin{cases} 1/2 & n \text{ even} \\ 0 & n \text{ odd.} \end{cases}$$

Thus the terms in the Cesàro mean

$$\frac{1}{n}\sum_{k=1}^{n} |\mu(T^{-n}A \cap A) - \mu(A)\mu(A)|$$

are all equal to $1/4$ and hence the limit is $1/4$ and not zero, disproving weak mixing.

Remark 6 Note that Theorem 2 does not directly follow from Theorem 5 and Corollary 6, since strong mixing implies weak mixing. The existence of a non weakly mixing ergodic measure will hold whenever the map $\nu \to \nu \circ T^{-1}$ have periodic points in $\mathcal{K}_g \setminus \mathcal{M}_g$.

Remark 7 In our example, the Doeblin measure is ergodic, but not totally ergodic. We may ask whether a unique totally ergodic Doeblin measure has to be weakly mixing.

Appendix A: Exactly two extremal measures in Theorem 8

In this appendix we prove a fact that should have been proven in [4]. Both the fact and its proof were known to the authors of [4] when they wrote their paper. All throughout this appendix we use the terminology and the notation of [4], and we assume knowledge of that paper.

Let $A = \{-1, +1\}^2$, and let g be the Doeblin function ('specification' in the terminology of [4]) defined in [4]. In order to be consistent with the notations in [4], we think of g as a function of $(x_{-j}, y_{-j} : j \in \mathbb{N}_0)$.

The goal of this appendix is to prove that indeed there are exactly two Doeblin measures. In [4] it was proven that there are at least two, so here we only need to prove that there are no more than two.

Proof that there are exactly two extremal Doeblin measures in Theorem 8 Let K be as in section 2 of [4], and for every $k \geq K$, let O_k be the beginning and let C_k be the opening of the k block containing 0 (see Definitions 3 and 4 in [4], p. 1345 in the journal version). Note that by definition, k blocks are disjoint. Let

$$X_k = \operatorname{sign} \sum_{t \in C_k} x_t$$

be the signature of the block. By (the proof of) Lemma 7 of [4] (p. 1350), if μ is a Doeblin chain for g, then

$$X = \lim_{k \to \infty} X_k$$

exists μ-almost surely. Further, if μ is extremal, then the limit is an almost sure constant. By the definition of X_k and the fact that the openings are of odd sizes, we get that X has to be $+1$ or -1.

The proof will be complete once we have shown that if X has the same distribution under two Doeblin measures μ_1 and μ_2 then $\mu_1 = \mu_2$. Without loss of generality we assume that

$$\mu_1(X = 1) = \mu_2(X = 1) = 1. \tag{34}$$

In order to prove (34), for every $\epsilon > 0$ and $N \in \mathbb{N}$ we find a coupling P between μ_1 and μ_2 such that

$$P\left(\forall_{-N<j<N}\ x_j^{(1)} = x_j^{(2)} \text{ and } y_j^{(1)} = y_j^{(2)}\right) > 1 - \epsilon.$$

Fix now such an $\epsilon > 0$ and $N \in \mathbb{N}$. Since the sequence $(y_j)_j$ has the same marginal distribution in μ_1 and in μ_2 (and in fact in every Doeblin chain of g), we first sample $\left(y_j^{(1)}\right)_{j \in \mathbb{Z}}$ and $\left(y_j^{(2)}\right)_{j \in \mathbb{Z}}$ in a way that $y_j^{(1)} = y_j^{(2)}$ for all j. In particular, this implies that both sequences have exactly the same block structure.

For $k \geq K$ let

$$A(k) = A_N(k) = \{O_k \neq \emptyset\} \cap \{O_k \cap [-N, N] \neq \emptyset\}.$$

Now let k_0 be so large that

$$\mu_1\left(X_{k_0} \neq 1\right) + \mu_2\left(X_{k_0} \neq 1\right) + \mu\left(A_N(k_0)^c\right) < \epsilon. \tag{35}$$

where for the events related to (y_j) we simply write μ because, as already mentioned, all Doeblin measures agree on the marginal distribution of the sequence (y_j).

Condition on the event $A_N(k_0)$. On the complement event our coupling fails, but the probability of $A_N(k_0)^c$ is small enough.

Let $j_0 = \max(O_{k_0})$. We now sample $\left(x_j^{(1)}\right)_{j<j_0}$ according to μ_1 and $\left(x_j^{(2)}\right)_{j<j_0}$ according to μ_2, say independently. Note that $P\left(A(k_0)^c \cup \left\{X_{k_0}^{(1)} \neq X_{k_0}^{(1)}\right\}\right) < \epsilon$, so in what follows we condition on the complement event $A(k_0) \cap \left\{X_{k_0}^{(1)} = X_{k_0}^{(1)}\right\}$. We now continue to sample $\left(x_j^{(1)}, x_j^{(2)}\right)_{j=j_0+1}^{\infty}$ sequentially as follows: Once we have $\left(x_i^{(1)}, x_i^{(2)}\right)_{i<j}$, we sample $x_j^{(1)}$ and $x_j^{(2)}$ according to the probabilities given by g, coupled in a way that maximizes $P\left(x_j^{(1)} = x_j^{(2)}\right)$.

Clearly P is a coupling of μ_1 and μ_2, and we need the following fact. □

Lemma 9

$$P\left[\forall_{j\in[-N,N]}\, x_j^{(1)} = x_j^{(2)} \;\middle|\; A(k_0) \cap \left\{X_{k_0}^{(1)} = X_{k_0}^{(1)}\right\}\right] = 1.$$

Proof For every j write S_j for the set whose majority we take in deciding $x_j^{(1)}$ and $x_j^{(2)}$. By induction, for every $j > j_0$, we have that either $S_j = C_{k_0}$ or $S_j \subseteq [j_0+1, \infty] \cap \mathbb{Z}$. Therefore, again by induction, on the event $A(k_0) \cap \left\{X_{k_0}^{(1)} = X_{k_0}^{(1)}\right\}$ we have that $x_j^{(1)} = x_j^{(2)}$ for all $j > j_0$. □

Using Lemma 9 we see that the total variation between the restrictions of μ_1 and μ_2 on $[-N, N]$ is arbitrarily small. As N is arbitrarily large, we get that $\mu_1 = \mu_2$. □

Acknowledgements We would like to thank Mark Pollicott and Peter Walters for valuable discussions. The fourth author wishes to thank the Knut and Alice Wallenberg Foundation for financial support.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

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. Stat. Phys. **50**(1–2), 1–40 (1988)
2. Berbee, H.: Chains with infinite connections: uniqueness and Markov representation. Probab. Theory Relat. Fields **76**, 243–253 (1987)
3. Berbee, H.: Uniqueness of Gibbs measures and absorption probabilities. Ann. Probab. **17**(4), 1416–1431 (1989)
4. Berger, N., Hoffman, C., Sidoravicius, V.: Nonuniqueness for specifications in $l^{2+\epsilon}$. Ergod. Theory Dyn. Syst. **38**(4), 1342–1352 (2018)
5. Bissacot, R., Endo, E.O., van Enter, A.C.D., Le Ny, A.: Entropic Repulsion and Lack of the g-measure Property for Dyson Models. Commun. Math. Phys. **363**, 767–788 (2018)

6. Bramson, M., Kalikow, S.: Nonuniqueness in g-functions. Israel J. Math. **84**, 153–160 (1993)
7. Breiman, L.: The strong law of large numbers for a class of Markov chains. Ann. Math. Stat. **31**, 801–803 (1960)
8. Brown, G., Dooley, A.H.: Odometer actions on G-measures. Ergod. Theory Dyn. Syst. **11**, 279–307 (1991)
9. Doeblin, W., Fortet, R.: Sur des chaînes à liaisons complètes. Bull. Soc. Math. France **65**, 132–148 (1937)
10. Dooley, A.H., Rudolph, D.J.: Non-uniqueness in G-measures. Ergod. Theory Dyn. Syst. **32**(2), 575–586 (2012)
11. Endo, E.O., van Enter, A.C.D., Le Ny, A.: The Roles of Random Boundary Conditions in Spin Systems, *Progress in Probability* **77** (2020)
12. Fernandez, R., Maillard, G.: Chains with complete connections: general theory, uniqueness, loss of memory and mixing properties. J. Stat. Phys. **118**, 555–588 (2005)
13. Gallesco, C., Gallo, S., Takahashi, D.Y.: Dynamic uniqueness for stochastic chains with unbounded memory. Stoch. Process. Their Appl. **128**(2), 689–706 (2018)
14. Hulse, P.: An example of non-unique g-measures. Ergod. Theory Dyn. Syst. **26**(2), 439–445 (2006)
15. Jacod, J., Shiryaev, A.N.: Limit Theorems for Stochastic Processes, Grundlehren der mathematischen Wissenschaften, vol. 288, 2nd edn. Springer, Berlin (2003)
16. Johansson, A., Öberg, A.: Square summability of variations of g-functions and uniqueness of g-measures. Math. Res. Lett. **10**(5–6), 587–601 (2003)
17. Johansson, A., Öberg, A.: Square summability of variations and convergence of the transfer operator. Ergod. Theory Dyn. Syst. **28**(4), 1145–1151 (2008)
18. Johansson, A., Öberg, A., Pollicott, M.: Countable state shifts and uniqueness of g-measures. Am. J. Math. **129**(6), 1501–1511 (2007)
19. Johansson, A., Öberg, A., Pollicott, M.: Unique Bernoulli g-measures. J. Eur. Math. Soc. (JEMS) **14**, 1599–1615 (2012)
20. Johansson, A., Öberg, A., Pollicott, M.: Phase transitions in long-range Ising models and an optimal condition for factors of g-measures. Ergod. Theory Dyn. Syst. **39**(5), 1317–1330 (2019)
21. Johansson, A., Öberg, A., Pollicott, M.: Continuous eigenfunctions for the transfer operator of the Dyson model, preprint arXiv:2304.04202
22. Kalikow, S.: Random Markov processes and uniform martingales. Israel J. Math. **71**(1), 33–54 (1990)
23. Keane, M.: Strongly mixing g-measures. Invent. Math. **16**, 309–324 (1972)
24. Kesten, H.: Sums of Random Variables with Infinite Expectation, Solutions of Advanced Problems. Am. Math. Mon. **78**(3), 305–308 (1971)
25. Ledrappier, F.: Principe variationnel et systèmes dynamiques symboliques. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **30**, 185–202 (1974)
26. Lindvall, T., Doeblin, W.: 1915–1940. Ann. Probab. **19**(3), 929–934 (1991)
27. Verbitskiy, E.: On factors of g-measures. Indag. Math. (N.S.) **22**(3–4), 315–329 (2011)
28. Walters, P.: Ruelle's operator theorem and g-measures. Trans. Am. Math. Soc. **214**, 375–387 (1975)
29. Walters, P.: Convergence of the Ruelle operator for functions satisfying Bowen's condition. Trans. Am. Math. Soc. **353**(1), 327–347 (2001)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.