NON-ERGODICITY FOR C^1 EXPANDING MAPS AND g-MEASURES

Anthony N. Quas

Department of Pure Mathematics and Mathematical Statistics, University of Cambridge

May 1994

Abstract. We introduce a procedure for finding C^1 Lebesgue measure-preserving maps of the circle isomorphic to one-sided shifts equipped with certain invariant probability measures. We use this to construct a C^1 expanding map of the circle which preserves Lebesgue measure, but for which Lebesgue measure is non-ergodic (that is there is more than one absolutely continuous invariant measure). This is in contrast with results for $C^{1+\epsilon}$ maps. We also show that this example answers in the negative a question of Keane's on uniqueness of g-measures, which in turn is based on a question raised by an incomplete proof of Karlin's dating back to 1953.

1. Introduction

In this paper, we will be considering differentiable maps of the circle. Such a map T will be called *expanding* if there is a $C > 1$ such that $|T'(x)| \geqslant C$ for all $x \in S^1$. An expanding map of the circle is a degree r map for some $|r| > 1$. This means the map is an $|r|$-fold cover of the circle by itself. We will only be considering orientation-preserving maps (that is those with $r > 1$), and throughout, we will consider the circle to be the interval $[0, 1)$ (mod 1). We shall be interested in the existence and number of absolutely continuous invariant measures (abbreviated to ACIMs) for these maps. We will also be working with

1991 *Mathematics Subject Classification.* 28D05, 58F11, 58F03.

Key words and phrases. absolutely continuous invariant measures, expanding maps, g-measures, spontaneous symmetry breaking, phase transition.

The author was financially supported by King's College, Cambridge

symbol spaces. The full shift on r symbols is defined to be the set $\Sigma_r = \{0, \dots, r-1\}^{\mathbb{Z}^+}$, with the metric

$$d(x,y) = \begin{cases} 2^{-n} & \text{if } x \text{ and } y \text{ differ in the } n\text{th place and not before} \\ 0 & \text{if } x = y. \end{cases}$$

The shift map on Σ_r is denoted by σ. If $x \in \Sigma_r$ and $a \in \{0, \dots, r-1\}^k$, then we will write ax to be the sequence in Σ_r defined by

$$(ax)_i = \begin{cases} a_i & \text{if } i < k \\ x_{i-k} & \text{if } i \geqslant k \end{cases}$$

We will also write $[x]^n$ for $\{y : d(x,y) < 2^{-n}\}$ (that is those y whose symbols agree with x from the 0th to the nth). This will be called the *n-cylinder* about x. Define

$$\mathrm{var}_n f = \sup_{\{x,y : d(x,y) < 2^{-n}\}} |f(x) - f(y)|.$$

There is considerable literature about expanding maps. Krzyżewski and Szlenk showed that for C^2 expanding maps of compact manifolds (that is C^2 maps whose Jacobian is everywhere bounded below by some $C > 1$), there is a unique ACIM (see [16] and [14]). Lasota and Yorke showed in [17] that any piecewise continuous and C^2 expanding map of the unit interval has an ACIM. This result clearly applies also to maps of the circle. Kowalski ([13]) improved this by showing that the same conclusion holds if the map is piecewise C^{1+1} (that is the map has Lipschitz derivative). Mañé's book ([20]) gives a refined proof showing that this remains true if the map is piecewise $C^{1+\epsilon}$ (that is the map has Hölder continuous derivative with exponent $\epsilon < 1$). Wong([23]) found that the conclusion holds when the assumption is altered to assuming that the map is piecewise C^1 with the reciprocal of the derivative, $1/T'$, of bounded variation.

Krzyżewski in [15] managed to show that the same conclusions do not in general hold for C^1 maps by showing that for any manifold M, there exist C^1 expanding maps of M which do not preserve any ACIM. His proof however was not constructive, so there was still some interest in constructing an explicit example of such a C^1 map in (for example) the simple case of the circle. This was done by Góra and Schmitt (see [7]).

Various authors then turned their attentions to the number of ergodic ACIMs in the piecewise C^2 case (where ACIMs are known to exist). Papers on this include [18], [4] and [3]. Note that these papers consider a more general situation than that considered in [16] and [14]. All of these in particular imply that if T is a C^2 expanding map of the circle then there is a unique ACIM for T. Such an ACIM would therefore necessarily be ergodic.

A natural question which remains is the one which we shall address in this paper: Does there exist a C^1 expanding map of the circle which preserves more than one ACIM? We will show that such a map does exist. We note that it is relatively straightforward to construct Lipschitz maps with this property (with an appropriate definition of expanding for Lipschitz maps). This was first done by Bose in [2] using generalized baker's transformations and was later done using the techniques of §2 of this paper by the author in [21].

THEOREM 1. *There exists a C^1 expanding map preserving Lebesgue measure, for which Lebesgue measure is not ergodic.*

In this paper, our main tool and an important point of subsidiary interest is that of a g-measure. These were introduced by Keane in [12], and are invariant measures with certain prescribed conditional probabilities. The concept of a g-measure is equivalent to the concept of a chain with complete connections introduced by Doeblin and Fortet ([6]) and also to that of a bounded uniform martingale introduced by Kalikow ([10]). To describe the concept of a g-measure, let T be an r-to-1 local homeomorphism of a compact metric space X to itself satisfying certain expansiveness conditions - for details, see [12]. We will only consider the case where T is either the degree r expanding map of the circle, $x \mapsto rx \bmod 1$, or the shift on Σ_r. Let $\mathcal{G}(X)$ be the class of those continuous functions $g : X \to (0,1)$ on X such that for all $x \in X$, $\sum_{y \in T^{-1}(x)} g(y) = 1$. We will write $\mathcal{G}$ where no confusion arises. Write $|A|$ for the diameter of a set A. A *g-measure* ν on X is then a Borel probability measure satisfying

$$\lim_{|A| \to 0, x \in A} \frac{\nu(A)}{\nu(T(A))} = g(x),$$

for all $x \in X$. In particular, if $X = \Sigma_r$, ν is a g-measure if and only if for each $i \in \{0, \dots, r-1\}$ and x,

$$\lim_{n \to \infty} \frac{\nu([ix]^{n+1})}{\nu([x]^n)} = g(ix). \tag{1}$$

Note that by compactness, the g-functions which we are considering are bounded away from 0. It is straightforward to show that for each $g \in \mathcal{G}$, there is at least one g-measure. Such a g-measure is automatically T-invariant and fully supported on X with no atoms (for proofs of these facts, see [12] and [22]). The set of g-measures for a given g forms a compact affine subset of the set of measures on X (in the weak*-topology), and the extreme points of this set are precisely the ergodic g-measures. Further, by the Krein-Mil'man theorem, the set of g-measures is the closed convex hull of the set of ergodic g-measures. It follows that if there is a g with a non-ergodic g-measure then there are at least two g-measures for that g.

We will now describe a more probabilistic interpretation of g-measures on Σ_r. We consider sequences $(X_n)_{n \in \mathbb{Z}}$ of random variables taking values in the set $\{0, \dots, r-1\}$, often regarding their values as outcomes of a sequence of experiments, one performed at each integer time. Strictly, one should consider the X_n as maps from some probability space Ω to $\{0,\dots,r-1\}$, and write $X_n(\omega)$ for X_n, but as we will be using the same probability space throughout, we often prefer to simply write X_n.

We will look at the evolution of the random variables by specifying the probabilities of the various outcomes of the 'present' experiment (that is X_0) conditional on the 'past' (that is $(X_n)_{n<0}$). The simplest non-trivial examples of this are given by Markov chains with stationary transition probabilities, where the probabilities of the outcomes of the present experiment are completely determined by the outcome of the previous one (that is $\mathbb{P}(X_n = i | X_{n-1} = j_1,\ X_{n-2} = j_2,\ \dots)$ is independent of $j_2,\ j_3,\ \dots$ and n). One can similarly consider the so-called 'finite range' processes or *k-step Markov chains*, where the probabilities are determined by the outcomes of the previous k experiments.

We will look at a generalization of these to 'infinite range' processes. Let $(X_n)_{n\in\mathbb{Z}}$ be a sequence of random variables taking values in $\{0,\dots,r-1\}$. Suppose the sequence satisfies

$$\mathbb{P}(X_n = i | X_{n-1} = a_1,\ X_{n-2} = a_2,\ \dots) = g(i, a_1, a_2, \dots), \tag{2}$$

where $g \in \mathcal{G}$. If we now fix an n, then we get a natural map $\rho_n : \Omega \to \Sigma_r$ given by $\rho_n(\omega)_i = X_{n-i}(\omega)$. This defines a natural map ρ_n^* from the measures on Ω to the measures on Σ_r by $\rho_n^*(\mathbb{P})(A) = \mathbb{P}(\{\omega : \rho_n(\omega) \in A\})$, this measure on Σ_r being called the push-forward of $\mathbb{P}$ under ρ_n. If $\mathbb{P}$ is a stationary probability distribution on Ω, satisfying (2), then by stationarity, we have $\rho_n^*(\mathbb{P})$, the push-forward of the distribution on those symbols before the nth, is independent of n. Call this measure ν, say. By integrating, we see

$$\nu([ix]^{n+1}) = \int_{[x]^n} g(iy)\ d\nu(y).$$

Using the continuity of g, it follows that ν satisfies (1), and hence is a g-measure. Clearly this also works in reverse. We will also be interested in the operator $\mathcal{L} : C(\Sigma_r) \to C(\Sigma_r)$ defined by $\mathcal{L}f(x) = \sum_{y\in T^{-1}(x)} g(y)f(y)$. We will sometimes write this as $\mathcal{L}_g$ to emphasise the dependence on g. A quick calculation shows that $\mathcal{L}^m f(x) = \sum_{y\in T^{-m}(x)} g^{(m)}(y)f(y)$, where $g^{(m)}(x) = g(x)g(\sigma(x))\cdots g(\sigma^{m-1}(x))$. The interpretation of these operators is that $\mathcal{L}^m f(x)$ is the expectation of $f(X_m,X_{m-1},\dots)$ conditioned on $X_{-i} = x_i$, for all $i \geqslant 0$, given that the (X_n) satisfy (2).

Keane ([12]) demonstrates that if g is Lipschitz then there is a unique g-measure, when $X = S^1$. Walters shows in [22] that the same conclusion holds when $X = \Sigma_r$ if g has summable variation (that is $\sum_{i=0}^{\infty} \mathrm{var}_i g < \infty$), or in particular if g is Hölder continuous. He shows further that in this situation, if the unique g-measure is ν then for any continuous function f, we have $\mathcal{L}^m f(x)$ converges uniformly to $\int f\, d\nu$ as $m \to \infty$. More recently, Berbee ([1]), also working on symbol spaces, took up the question, providing weaker conditions on g which also guarantee uniqueness of g-measures. Kaijser ([9]) worked on a situation which can be regarded as a generalization of the case $X = S^1$, where he managed to show uniqueness of g-measures under the very weak 'Dini condition'. Hulse ([8]) applied some ideas of statistical mechanics to find a new class of g which have unique g-measures. The paper was interesting as the result followed from general statistical mechanical restrictions on g, rather than stronger continuity conditions. He worked on Σ_r and introduced a partial order $\preceq$ on it:

$$x \preceq y \text{ if } x_i \leqslant y_i,\ \forall i \geqslant 0.$$

A g-function is then called *attractive* or *monotonic* if

$$\sum_{i\geqslant j} g(ix) \leqslant \sum_{i\geqslant j} g(iy), \text{ whenever } x \preceq y$$

Hulse showed that attractiveness greatly facilitated calculations with g-measures and we make considerable use of a similar property in what follows. We note that in the probabilistic interpretation, this says the bigger the past has been, the bigger the values we get at the present experiment.

All of the above papers gave conditions on g under which there exists a unique g-measure, but the question still remained, does there exist some $g \in \mathcal{G}$ for which there is more than one g-measure? This question was finally answered by Bramson and Kalikow in the ingenious paper [5], for the case $X = \Sigma_2$, in which they construct an example of a g-function, for which there are at least two g-measures. The methods used below are heavily based on those in [5], which may be regarded as inspiration for the proofs in this paper. The proof of Bramson and Kalikow made extensive use of 'coupling' arguments (also known as joinings). For a thorough description of these methods, the reader is referred to Lindvall's book, [19]. Briefly, suppose (X_n) and (Y_n) are sequences of random variables with distributions $\mathbb{P}_1$ and $\mathbb{P}_2$ respectively, then a *coupling* is a joint probability distribution $\mathbb{P}$ whose marginal distribution on the (X_n) is $\mathbb{P}_1$ and on the (Y_n) is $\mathbb{P}_2$.

The work of Bramson and Kalikow still leaves open the case $X = S^1$, which was the question originally asked by Keane in [12]. In fact, Keane points out that much earlier, Karlin had claimed to answer this question, but his results depended on strong continuity assumptions, similar to those in other papers discussed above. Keane also points out that his question had also been raised by Doeblin and Fortet in [6].

In this paper, we show that the non-uniqueness of g-measures when $X = S^1$ is equivalent to Theorem 1. To prove Theorem 1, we note that there is a third equivalent form of the question which relates to g-measures on symbol spaces, and it is in this formulation that we will be most easily able to prove the result. The proof of these equivalences is in fact little more than changes of coordinate. This is carried out in the next section. The remainder of the paper is concerned with proving the equivalent version of Theorem 1, which relates to g-measures on symbol spaces as described in §2. The technique employed there is based on one used by Bramson and Kalikow in [5], and in physics terms could be described as finding an example of a g-function with symmetry which exhibits spontaneous symmetry breaking in that the g-measures do not respect the symmetry of the g-function. The situation where there are two distinct g-measures is known in the physics literature as phase transition.

2. Equivalent formulations of Theorem 1

To state the result of this section, we need some further definitions. If $X = \Sigma_r$, let $u = 000\ldots$ and $v = r-1, r-1, r-1, \ldots$. We define an equivalence relation $\sim$ on X generated by

$$u \sim v$$

$$aju \sim aiv \text{ for } i < r-1,\ j = i+1 \text{ and any finite word } a.$$

Note that the cardinality of any $\sim$-equivalence class is either 1 or 2. The equivalence classes consist precisely of elements of Σ_r, which describe the same number (mod 1) when considered as a base r expansion (with a point at the front).

If $g \in \mathcal{G}$ is a g-function on X, then we say that g is *compatible* if $x \sim y$ implies $g(x) = g(y)$. Let $\mathcal{G}^{\text{comp}}$ denote the set of those $g \in \mathcal{G}$ which are compatible.

THEOREM 2. *The following are equivalent:*

(1) *There exists an orientation-preserving expanding C^1 map of S^1 preserving Lebesgue measure for which Lebesgue measure is not ergodic;*

(2) *There exists a $g \in \mathcal{G}^{\mathrm{comp}}$ which has more than one g-measure;*

(3) *There exists a $g \in \mathcal{G}(S^1)$, which has more than one g-measure.*

Proof. Suppose (1) holds. Then suppose the expanding map in question is T, a degree r map. By a change of coordinates of the form $x \mapsto x + \alpha$, we may assume that $T(0) = 0$. Then let the preimages of 0 be $0, a_1, \dots, a_{r-1}$. Then set $I_0 = [0, a_1]$, $I_1 = [a_1, a_2]$, $\dots$, $I_{r-1} = [a_{r-1}, 1]$. These sets form a Markov partition of the circle. By a standard argument, there is a semi-conjugacy π_1 from (σ, Σ_r) to (T, S^1). By the expanding condition, we can check that each point of the circle has a unique preimage under π_1, except for the countable number of points which are preimages under T of 0. In particular, $\pi_1(x) = \pi_1(y)$ if and only if $x \sim y$. We can therefore define a measure on Σ_r by defining $\nu(A) = \lambda(\pi_1(A))$ where λ denotes Lebesgue measure (as it will throughout this paper), since λ has no atoms. Now fix a point $x \in \Sigma_r$ and consider sets A containing x. We have

$$\frac{\nu(A)}{\nu(\sigma(A))} = \frac{\lambda(\pi_1(A))}{\lambda(T(\pi_1(A)))}.$$

As the diameter of A becomes smaller, the ratio on the right hand side approaches the limit $1/T'(\pi_1(x))$, so we have ν is a g-measure, where $g(x) = 1/T'(\pi_1(x))$. But π_1 is a measure-theoretic isomorphism $(\sigma, \Sigma_r, \nu) \to (T, S^1, \lambda)$, so since λ is not ergodic, we must have that ν is non-ergodic, hence there is more than one g-measure. This g-function is clearly compatible, so we have shown (2) holds.

Next, suppose (2) holds for $g \in \mathcal{G}^{\mathrm{comp}}(\Sigma_r)$. Then define $\pi_2 : \Sigma_r \to S^1$, $x \mapsto \Sigma_{i=0}^{\infty} x_i r^{-(i+1)}$.▮

This is a semiconjugacy from (σ, Σ_r) to (T_r, S^1), with the property that $x \sim y$ if and only if $\pi_2(x) = \pi_2(y)$. Then let ν be a non-ergodic g-measure. Define a measure μ on S^1 by pushing ν forward: $\mu(A) = \nu(\pi_2^{-1}(A))$. Given a point $x \in S^1$, consider sets A containing x. We have

$$\frac{\mu(A)}{\mu(T_r(A))} = \frac{\nu(\pi_2^{-1}(A))}{\nu(\sigma(\pi_2^{-1}(A)))}.$$

If x is not a preimage under T_r of 0, then this expression clearly converges to $g(\pi_2^{-1}(x))$ as $|A| \to 0$. But if x is a preimage of 0 under T_r, the same conclusion holds by the compatibility of g. It follows that μ is an h-measure, where $h(x) = g(\pi_2^{-1}(x))$. This can be seen to be well-defined and continuous because of the compatibility of g. Again, π_2 is a measure-theoretic isomorphism $(\sigma, \Sigma_r, \nu) \to (T_r, S^1, \mu)$, so it follows that μ is non-ergodic, and hence there is more than one h-measure. This proves (3).

Finally, suppose (3) holds. Let $g \in \mathcal{G}(S^1)$ be a g-function which has more than one g-measure. Then let ν be a non-ergodic g-measure. Define $h(x) = \nu[0, x]$. This is an orientation-preserving homeomorphism of the circle (since g-measures are fully supported and have no atoms). Let μ be the measure pushed forward by h. Then we have

$$\mu[0, x] = \nu(h^{-1}[0, x]) = \nu([0, h^{-1}(x)]) = h(h^{-1}(x)) = x.$$

It follows that μ is Lebesgue measure. Let $T = h \circ T_r \circ h^{-1}$. Then we have (T_r, S^1, ν) is measure-theoretically isomorphic to (T, S^1, λ) by h, so we see that λ is non-ergodic for T. Now suppose $x \in S^1$ is fixed and $y < x < z$. Then we have

$$\frac{T(z) - T(y)}{z - y} = \frac{\lambda(T[y, z])}{\lambda([y, z])} = \frac{\nu(T_r[h^{-1}(y), h^{-1}(z)])}{\nu([h^{-1}(y), h^{-1}(z)])}.$$

Taking the limit as y increases to x and z decreases to x, we get convergence to $1/g(h^{-1}(x))$. It follows that T is differentiable and expanding, preserving Lebesgue measure (which is non-ergodic for T), so we have shown (1) holds and thus completed the proof of the equivalences. $\square$

From the above, we know that to prove Theorem 1, it is sufficient to find an example satisfying condition (3) of Theorem 2.

3 Preliminaries for the Proof of Theorem 1

There are difficulties reconciling attractiveness with compatibility in general (see [21]), and for this reason, we will need to use a weaker order than that used in the definition of attractiveness given above. We will work for the rest of the paper with $r = 10$, so as to produce a ten-branched map (this started as an arbitrary choice as 2 was too small, but it turns out that this construction uses the fact that there are 10 branches in a non-trivial way).

We define a partial order on $\{0, 1, \ldots, 9\}$ by $3 \preceq i \preceq 6$, for all i. We then define a partial order on Σ_{10} by $x \preceq y$ if $x_i \preceq y_i$, for all $i \in \mathbb{Z}^+$. Write π for the map $\Sigma_{10} \to I$, defined by $x \mapsto \sum_{i=0}^{\infty} x_i 10^{-(i+1)}$. We will identify Σ_{10} with I and often omit reference to π, when applying functions on I to arguments in Σ_{10}.

We attempt to construct Hölder continuous functions $h_{\nearrow}$ and $h_{\searrow}$ on I such that $h_{\nearrow}(0) = 0,\ h_{\nearrow}(1) = 1,\ h_{\searrow}(0) = 1,\ h_{\searrow}(1) = 0$ and $x \preceq y \Rightarrow h_{\nearrow}(x) \leqslant h_{\nearrow}(y), h_{\searrow}(x) \leqslant h_{\searrow}(y)$. These will be used to write down the g-function later.

Define an operator $\Phi_{\nearrow} : C_{\nearrow}(I) \to C_{\nearrow}(I)$ as follows:

$$\Phi_{\nearrow} f(x) = \begin{cases} 0 & x < 0.4 \\ 1 & x \geqslant 0.6 \\ \frac{1}{2} f(\sigma(x)) & 0.4 \leqslant x < 0.5 \\ \frac{1}{2}(f(\sigma(x)) + 1) & 0.5 \leqslant x < 0.6, \end{cases}$$

where $C_{\nearrow}(I) = \{f : I \to I,\ f(0) = 0,\ f(1) = 1,\ f \text{ continuous}\}$. It is clear that this operator maps $C_{\nearrow}$ into itself, and it can also be seen to be a contraction map (with respect to the uniform norm). It follows that there is a unique fixed point $h_{\nearrow}$. It remains to show that this function respects the order and is Hölder continuous. To this end, suppose x and y lie in the same n-cylinder of Σ_{10} (where $n \geqslant 0$). Then since $h_{\nearrow}$ is a fixed point of $\Phi_{\nearrow}$, it follows that $|h_{\nearrow}(y) - h_{\nearrow}(x)| \leqslant \frac{1}{2}|h_{\nearrow}(\sigma(y)) - h_{\nearrow}(\sigma(x))|$. By induction, it follows that $|h_{\nearrow}(y) - h_{\nearrow}(x)| \leqslant 2^{-n}$. It follows that $h_{\nearrow}$ is Hölder continuous.

PROPOSITION 3. *Suppose $x \prec y$. Then $h_{\nearrow}(x) \leqslant h_{\nearrow}(y)$.*

Proof. By induction on the first digit where x and y differ. Note that by definition of $\preceq$, if x and y differ in the 0th place then either x starts with a 3 (in which case $h_{\nearrow}(x) = 0 \leqslant h_{\nearrow}(y)$) or y starts with a 6 (in which case $h_{\nearrow}(x) \leqslant 1 = h_{\nearrow}(y)$). Clearly, in either case the statement holds. Suppose the proposition holds for all x and y which differ first in the nth place (with $n \geqslant 0$), but suppose we are given $x \prec y$ with $x_i = y_i$, for all $i \leqslant n$ and $x_{n+1} \prec y_{n+1}$. If $x_0 < 4$ or $x_0 > 6$, then $h_{\nearrow}(x) = h_{\nearrow}(y)$. Otherwise, $h_{\nearrow}(y) - h_{\nearrow}(x) = \frac{1}{2}(h_{\nearrow}(\sigma(y)) - h_{\nearrow}(\sigma(x)))$. But $\sigma(x) \prec \sigma(y)$ and they differ first on the nth digit, so by the induction hypothesis, we see $h_{\nearrow}(y) - h_{\nearrow}(x) \geqslant 0$ and the proposition holds. □

Next, we construct $h_{\searrow}$. Let $C_{\searrow}$ be given by $\{f : I \to I;\ f(0) = 1,\ f(1) = 0,\ f \text{ continuous}\}$. Then define $\Phi_{\searrow} : C_{\searrow} \to C_{\searrow}$ by

$$\Phi_{\searrow} f(x) = \begin{cases} \frac{1}{2}(f(\sigma(x)) + 1) & x < 0.1 \\ \frac{1}{2} f(\sigma(x)) & 0.1 \leqslant x < 0.2 \\ h_{\nearrow}(x) & 0.2 \leqslant x < 0.8 \\ \frac{1}{2}(1 + f(\sigma(x))) & 0.8 \leqslant x < 0.9 \\ \frac{1}{2} f(\sigma(x)) & x \geqslant 0.9 \end{cases}$$

Then $\Phi_\searrow$ is also a contraction. Similar arguments to those given above show that the unique fixed point $h_\searrow$ also respects the partial order and is Hölder continuous (with the same constant).

We note that $h_\searrow(1-x) = 1 - h_\searrow(x)$ and $h_\nearrow(1-x) = 1 - h_\nearrow(x)$. These two functions reverse the partial order and will also be used in the construction of the g-function.

Given $m \in \mathbb{N}$, let S_m denote $\{0, \ldots, 9\}^m$, the set of words of length m. Given $a \in S_m$, define $a-1 \in S_m$ and $a+1 \in S_m$ in the obvious way (that is such that $13246+1 = 13247$ and $00000 - 1 = 99999$ for example). Define $\psi^i : S_m \to \{0, \ldots, m\}$ by setting $\psi^i(a)$ to be the number of occurrences of the symbol i in a. We will then be considering $\delta(a)$ given by $\delta(a) = \psi^6(a) - \psi^3(a)$.

Proposition 4. *We have* $|\delta(a) - \delta(a+1)| \leqslant 1$ *for all* $a \in S_m$.

The proof of this is straightforward and is omitted.

Write $[a]$ for the m-cylinder consisting of those elements of Σ_{10} whose first m terms are the block a. We then define $V^6_{m,n} : S^1 \to [0,1]$ cylinder by cylinder:

$$V^6_{m,n}\big|_{[a]}(x) = \begin{cases} 1 & \text{if } \delta(a) > n \\ 0 & \text{if } \delta(a) < n \\ 0 & \text{if } \delta(a) = n,\ \delta(a-1) \leqslant n,\ \delta(a+1) \leqslant n \\ h_\nearrow(\sigma^m(x)) & \text{if } \delta(a) = n,\ \delta(a-1) \leqslant n,\ \delta(a+1) > n \\ h_\searrow(\sigma^m(x)) & \text{if } \delta(a) = n,\ \delta(a-1) > n,\ \delta(a+1) \leqslant n \\ 1 & \text{if } \delta(a) = n,\ \delta(a-1) > n,\ \delta(a+1) > n \end{cases}$$

It is then immediate to check that $V^6_{m,n}$ is Hölder continuous and respects the partial order.

We then define $V^3_{m,n}(x) = V^6_{m,n}(1-x)$. By the earlier observations, we see that this is order-reversing and in fact satisfies $V^3_{m,n}(x) = 1 - V^6_{m,-n}(x)$. Now set

$$W^6_{m,n}(x) = \tfrac{1}{10} + \tfrac{1}{2}V^6_{m,n}(x)$$

$$W^3_{m,n}(x) = \tfrac{1}{10} + \tfrac{1}{2}V^3_{m,n}(x)$$

$$W^i_{m,n}(x) = \tfrac{1}{10} - \tfrac{1}{16}(V^6_{m,n}(x) + V^3_{m,n}(x)) \text{ for } i \neq 3, 6.$$

Notice that $\sum_i W^i_{m,n}(x) \equiv 1$. We will always be considering the case where $n > 0$ and in this case, it is easy to see that for each x, one of $V^6_{m,n}(x)$ and $V^3_{m,n}(x)$ is 0. It follows that $\frac{1}{10} \leqslant W^i_{m,n}(x) \leqslant \frac{6}{10}$ for $i = 3, 6$ and $\frac{3}{80} \leqslant W^i_{m,n}(x) \leqslant \frac{1}{10}$ otherwise. Our g-function will then be constructed as an infinite affine combination of the $W_{m,n}$.

4. Construction of the Example

From here, the proof resembles that given in [5], but the example is more complicated so some details differ. Let $q_j = \frac{1}{2}(\frac{2}{3})^j$, so $\sum_{j=1}^{\infty} q_j = 1$. We will choose n_j and m_j such that taking $g(ix) = \sum_{j=1}^{\infty} q_j W^i_{m_j,n_j}(x)$ will give a compatible continuous g with more than one g-measure. The choice of n_j and m_j will be made inductively, by considering certain Hölder continuous truncations of the final g-function. Suppose $n_1, \ldots, n_{k-1}$ and $m_1, \ldots, m_{k-1}$ are chosen. Then define vectors as follows:

$$u = (\tfrac{3}{80}, \tfrac{3}{80}, \tfrac{3}{80}, \tfrac{3}{5}, \tfrac{3}{80}, \tfrac{3}{80}, \tfrac{1}{10}, \tfrac{3}{80}, \tfrac{3}{80}, \tfrac{3}{80})$$

$$v = (\tfrac{3}{80}, \tfrac{3}{80}, \tfrac{3}{80}, \tfrac{1}{10}, \tfrac{3}{80}, \tfrac{3}{80}, \tfrac{3}{5}, \tfrac{3}{80}, \tfrac{3}{80}, \tfrac{3}{80})$$

$$z = (\tfrac{3}{80}, \tfrac{3}{80}, \tfrac{3}{80}, \tfrac{7}{20}, \tfrac{3}{80}, \tfrac{3}{80}, \tfrac{7}{20}, \tfrac{3}{80}, \tfrac{3}{80}, \tfrac{3}{80}),$$

with indices running from 0 to 9. Now define

$$g_k^1(ix) = \sum_{j=1}^{k-1} q_j W^i_{m_j,n_j}(x) + \sum_{j=k}^{\infty} q_j z_i$$
$$g_k^2(ix) = \sum_{j=1}^{k-1} q_j W^i_{m_j,n_j}(x) + q_k u_i + \sum_{j=k+1}^{\infty} q_j v_i$$
$$g_k^3(ix) = \sum_{j=1}^{k-1} q_j W^i_{m_j,n_j}(x) + q_k W^i_{M,N}(x) + \sum_{j=k+1}^{\infty} q_j v_i,$$

where $M > N > 0$. These are all Hölder continuous g-functions and as such have unique g-measures, which we call μ_k^e where $e = 1, 2, 3$. First note that g_k^1 is symmetric: $g_k^1(1-x) = 1-g_k^1(x)$. This means the unique invariant measure must be preserved under the involution $x \mapsto 1-x$. It follows that $\mu_k^1[6] = \mu_k^1[3]$. We will use the order-preserving properties of g to show that $\mu_k^3([6]) \geqslant \mu_k^2([6]) > \mu_k^1([6])$ and $\mu_k^3([3]) \leqslant \mu_k^2([3]) < \mu_k^1([3])$. Let $\alpha_k = \frac{1}{16}(\frac{2}{3})^k$.

Lemma 5. *We have $\mu_k^2([6]) \geqslant \mu_k^1([6]) + 2\alpha_k$ and $\mu_k^2([3]) \leqslant \mu_k^1([3]) - 2\alpha_k$. Further, suppose we are given $x \in \Sigma_{10}$. Then there is a coupling of the two processes (Y_n) and (Z_n) evolving under g_k^2 and g_k^3 respectively, conditioned on $Y_i = Z_i = x_{-i}$, for all $i \leqslant 0$ such that $Y_n \preceq Z_n$ with probability 1 for all n.*

Proof. The proof works by finding couplings of two processes evolving under different g-functions, which make it obvious that the required inequalities hold.

It is easy to check that $g_k^2(6x) - g_k^1(6x) = 2\alpha_k$ and $g_k^2(3x) - g_k^1(3x) = -2\alpha_k$, while $g_k^2(ix) = g_k^1(ix)$ for all $i \neq 3, 6$ and all x.

We use this to give an explicit coupling of two random processes (X_n) and (Y_n) evolving under g_k^1 and g_k^2 respectively as in (2). We write $P(ix, jy)$ for the probability that i is added to x and j is added to y. The transition probability will only be defined when $x \preceq y$, and it must therefore have $\mathbb{P}(ix \preceq jy) = 1$ in order that it can be applied repeatedly. Suppose

$x \preceq y$. Then define

$$P(ix,jy) = \begin{cases} g_k^1(6x) & \text{if } i=j=6 \\ \max\big(0, g_k^1(ix) - g_k^2(iy)\big) & \text{if } i \neq 3,6 \text{ and } j=6 \\ \min\big(g_k^1(ix), g_k^2(iy)\big) & \text{if } i=j\neq 3,6 \\ \max\big(0, g_k^2(jy) - g_k^1(jx)\big) & \text{if } i=3 \text{ and } j\neq 3,6 \\ \min\big(g_k^2(6y) - g_k^1(6x), g_k^1(3x) - g_k^2(3y)\big) & \text{if } i=3 \text{ and } j=6 \\ g_k^2(3y) & \text{if } i=j=3 \end{cases}$$

We note that all the transition probabilities are non-negative, and we must just check that the marginals of this coupling are as claimed. We compute one example as an illustration. We will show that under P, the probability that x goes to $3x$ is $g_k^1(3x)$ as required. By observation, we see that the probability that x goes to $3x$ is

$$\begin{aligned} &g_k^2(3y) + \min\big(g_k^2(6y) - g_k^1(6x), g_k^1(3x) - g_k^2(3y)\big) + \sum_{j\neq 3,6} \max(0, g_k^2(jy) - g_k^1(jx)) \\ =&g_k^1(3x) + \min\big(g_k^2(6y) - g_k^1(6x) - g_k^1(3x) + g_k^2(3y), 0\big) + 8\max\big(0, g_k^2(jy) - g_k^1(jx)\big) \\ =&g_k^1(3x) + \min\left(\sum_{i\neq 3,6} g_k^1(ix) - g_k^2(ix), 0\right) + 8\max\big(0, g_k^2(jy) - g_k^1(jx)\big) \\ =&g_k^1(3x) \end{aligned}$$

as required, where we are using the fact that $g(ix) = g(jx)$, for all $i, j \neq 3,6$. This shows that given that $x \preceq y$, we can choose i and j such that y evolves according to g_k^2 and x according to g_k^1 such that with probability 1, $ix \preceq jy$. Looking further at the coupling, we see that the probability that y is preceded by a 6 and x is not preceded by a 6 given that $x \preceq y$ is $g_k^2(6y) - g_k^1(6x)$, but $g_k^2(6y) - g_k^1(6y) = 2\alpha_k$ and $g_k^1(6y) \geqslant g_k^1(6x)$, so it follows that with (x,y) goes to $(ix, 6y)$ for some $i \neq 6$ with probability at least $2\alpha_k$. It follows that $\mu_k^2([6]) \geqslant \mu_k^1([6]) + 2\alpha_k$. A similar argument shows that $\mu_k^2([3]) \leqslant \mu_k^1([3]) - 2\alpha_k$.

To prove the remaining parts of the Lemma, it is necessary to consider a coupling of processes (Y_n) evolving under g_k^2 and (Z_n) evolving under g_k^3. This is done by a coupling exactly similar to the coupling above, with g_k^2 replacing g_k^1 and g_k^3 replacing g_k^2. The conclusion then is that given that $y \preceq z$, then y can be allowed to evolve under g_k^2 and z under g_k^3 in such a way that the ordering is preserved. For a more formal and general discussion of couplings, the reader is referred to Lindvall's book ([19]). □

We now describe the inductive choice of m_k and n_k. In each case, n_k is given by $\lfloor \alpha_k m_k \rfloor$. Suppose we have chosen $m_1, m_2, \dots m_{k-1}$ and hence $n_1, n_2, \dots n_{k-1}$. Let $\eta(x) = \chi_{[6]}(x) - \chi_{[3]}(x)$ and $\Delta_j(x) = \sum_{i=0}^{m_j-1} \eta(\sigma^i(x))$ Let G_j denote $\{x : \Delta_j(x) \geqslant n_j\}$ and H_j denote $\{x : \Delta_j(x) \geqslant 3n_j\}$. Then note that $\sigma^{-n_j}(H_j) \subseteq G_j$. Note also that if $x \in H_j$ and $y \succeq x$, then $y \in H_j$. Assume that there are $t_1, t_2, \dots t_{k-1}$ such that

$$\mathbb{P}\big((X_{t_j}, X_{t_j-1}, \dots) \in H_j | X_{-i} = x_i,\ \forall i \geqslant 0\big) \geqslant 1 - 4^{-j}, \tag{3}$$

for all $j < k$ and $x \in \Sigma_{10}$, where the X_n evolve according to g_j^3.

Let $A_m = \{x : \Delta_m(x) > 3\alpha_k m\}$. We know $\int \eta(x)\ d\mu_k^2(x) \geqslant 4\alpha_k$ and we will use this to show $\mu_k^2(A_m) \to 1$ as $m \to \infty$.

Lemma 6. *We have $\mu_k^2(A_m) \to 1$ as $m \to \infty$.*

Proof. Suppose the claim does not hold. Since we have $\mu_k^2(A_m) \leqslant 1$ for all m, the only way the claim can fail is if there exists an $\epsilon > 0$ and a sequence M_i such that $\mu_k^2(A_{M_i}) < 1 - \epsilon$ for all i. In this case, we have

$$\mu_k^2\left(\bigcup_{i>j} A_i{}^c\right) > \epsilon \text{ for all } j, \text{ so}$$

$$\mu_k^2\left(\bigcap_j \bigcup_{i>j} A_i{}^c\right) \geqslant \epsilon.$$

Let S be $\bigcap_j \bigcup_{i>j} A_i{}^c$. If $x \in S$ then

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \eta(\sigma^i(x)) \leqslant 3\alpha_k.$$

We have however that μ_k^2 is ergodic, so for almost all x (with respect to μ_k^2), we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \eta(\sigma^i(x)) \geqslant 4\alpha_k.$$

This is a contradiction. □

Next, pick m_k such that $\mu_k^2(A_{m_k}) > 1 - 4^{-k}$ and $\alpha_k m_k > t_{k-1}$. Now $H_k = A_{m_k}$. Since g_k^2 is Hölder continuous, we can apply Walters' theorem ([22]) to get that ${\mathcal{L}_{g_k^2}}^n \chi_{H_k}(x)$ converges uniformly to $\mu_k^2(H_k)$, which is greater than $1 - 4^{-k}$. It follows that there exists a t_k such that ${\mathcal{L}_{g_k^2}}^{t_k} \chi_{H_k}(x) \geqslant 1 - 4^{-k}$, for all $x \in \Sigma_{10}$. This says that for all $x \in \Sigma_{10}$,

$$\mathbb{P}\big((X_{t_k}, X_{t_k-1}, \dots) \in H_k | X_{-i} = x_i,\ \forall i \geqslant 0\big) \geqslant 1 - 4^{-k},$$

where the X_n evolve according to g_k^2, but by the second statement of Lemma 5, it follows that the same equation holds when the evolution is according to g_k^3. This is precisely the statement of (3) when we take j to be k. This completes the inductive step.

To complete the inductive construction of the example, it remains only to specify an initial case for the induction. For $k = 1$, however, the induction hypothesis, (3), holds vacuously. Taking $t_0 > 1$, applying the above induction step produces m_1, n_1 and t_1 which can be used as a starting point for the induction.

5. Completion of the Proof

Proof of Theorem 1. In the above section, the m_i and n_i were inductively constructed, so the g-function is now given by

$$g(ix) = \sum_{j=1}^{\infty} q_j W^i_{m_j,n_j}(x).$$

Note that to check compatibility, we must just check $g(000\dots) = g(999\dots)$ and $g(i999\dots) = g(j000\dots)$ for $i < 9$ and $j = i+1$. This is however straightforward as all of these quantities are equal to $\frac{1}{10}$. Continuity of this g is also clear as it is the uniform limit of continuous functions. It therefore remains to show that there are two distinct g-measures for this g.

We consider the events E_k^t that $(X_t, X_{t-1}, \dots) \in H_k$. Write $\mathbb{P}_6$ for the probability distribution of the X_n conditioned on $X_i = 6$ for all $i < 0$ with subsequent evolution under g. Informally, E_k^t is the event that the process has a 'large majority of 6s over 3s at the m_k scale at time t'. We then consider letting the process evolve from an initial condition of all 6s (so $\mathbb{P}_6(E_k^0) = 1,\ \forall k$). We show inductively that the events E_k^t have a high probability by induction on t, using the result of §4, which says that if the process has a large majority of 6s on scales $m_{k+1}, m_{k+2}, \dots$ at time $t - t_k$, then with high probability, the process will have a majority of 6s on scale m_k at time t.

Lemma 7. *We have*

$$\mathbb{P}_6(E_k^t) \geqslant 1 - \zeta_k, \tag{4}$$

for all $t \in \mathbb{Z}$ and $k \in \mathbb{N}$, where $\zeta_k = \frac{3}{2}4^{-k}$.

Proof. The proof is by induction on t. Note that the hypothesis is automatically true for all k if $t < 0$, so we need only prove the inductive step. Suppose (4) holds for all $t < s$ then pick $k \in \mathbb{N}$. Let $S = \bigcap_{j>k} E_j^{s-t_k}$. Then by the induction hypothesis, $\mathbb{P}_6(S^c) \leqslant \sum_{j>k} \zeta_j = \frac{1}{2}4^{-k}$. Now we decompose $E_k^{s\,c}$ as $(E_k^{s\,c} \cap S) \cup (E_k^{s\,c} \cap S^c)$. We then have

$$\mathbb{P}_6(E_k^{s\,c}) \leqslant \mathbb{P}_6(E_k^{s\,c} \cap S) + \mathbb{P}_6(S^c) \leqslant \mathbb{P}_6(E_k^{s\,c}|S) + \tfrac{1}{2}4^{-k}.$$

But now suppose $\omega \in S$. Then let $x = (X_{s-t_k}, X_{s-t_k-1}, \dots)$ and $z = (X_s, X_{s-1}, \dots)$. Then $x \in \bigcap_{j>k} H_j$. It follows that if $y \in \sigma^{-t}(x)$, for some $t \leqslant t_k$ then $y \in \bigcap_{j>k} G_j$. In particular, $g(y) = g_k^3(y)$, where the M and N in g_k^3 are taken to be m_k and n_k. It follows that the evolution of x for t_k steps takes place under g_k^3, but by (3), the probability that $z \in E_k^{sc}$ is no more than 4^{-k}. In particular, we have shown that $\mathbb{P}_6(E_k^{sc}) \leqslant \zeta_k$ as required. This completes the proof of the inductive step and hence of the lemma. □

We apply this by calculating $\mathbb{P}_6(X_n = 6)$. Using the Lemma above, this is bounded below by $\sum_{j\geqslant 1} q_j\left((1-\zeta_j)\frac{3}{5} + \zeta_j\frac{1}{10}\right)$. This turns out to be equal to $\frac{21}{40}$. Let $\mu_n = \rho_n^*(\mathbb{P}_6)$, as defined in §1. Then we have

$$\mu_{n+1}([ix]^{m+1}) = \int_{[x]^m} g(iy)\ d\mu_n(y).$$

Now let $\nu_n = \frac{1}{n}\sum_{j=0}^{n-1}\mu_j$. Then we see

$$\left|\nu_n([ix]^{m+1}) - \int_{[x]^m} g(iy)\ d\nu_n(y)\right| \leqslant \frac{2}{n}.$$

Taking a weak*-convergent subsequence $\nu_{n_k} \to \nu$ of the ν_n, we find

$$\nu([ix]^{m+1}) = \int_{[x]^m} g(iy)\ d\nu(y).$$

As noted in §1, this implies that ν is a g-measure. However $\mu_n([6]) \geqslant \frac{21}{40}$, for all n, so it follows that $\nu([6]) \geqslant \frac{21}{40}$. Since the g-function which we constructed was symmetric under $x \mapsto 1-x$, there is a second g-measure giving weight at least $\frac{21}{40}$ to the symbol 3. Since these are probability measures, they cannot be equal. It follows that there are two distinct g-measures for this g. This completes the proof of Theorem 1. □

It is envisaged that these techniques might find application in constructing other C^1 maps with varying degrees of ergodic properties. The author has already used them to find a C^1 expanding map of the circle which is ergodic, but not weak-mixing. I hope that this result will appear in another paper. It would be interesting to see if this method could be used to construct C^1 maps which are (for instance) exact, but not Bernoulli. It would also be of interest to construct a map with 2 branches for which Lebesgue measure is not ergodic, rather than the 10 branches in this paper, as it was in this form that the erroneous proof of Karlin's theorem was stated and also Keane's question based on Karlin's theorem.

I would like to express my thanks to Chris Bose for pointing out an error in an earlier version of this paper. Thanks too to Peter Walters, Yurii Suhov, Mike Keane, Thomas Kaijser, Jeff Steif, Bill Parry and Saul Jacka for helpful conversations.

References

1. H. Berbee. Chains with infinite connections: uniqueness and Markov representation'. *Probab. Th. Rel. Fields* **76** (1987), 243–253.
2. C. J. Bose. Generalized baker's transformations. *Ergodic Theory Dynamical Systems* **9** (1989), 1–17.
3. A. Boyarsky and W. Byers. A graph-theoretic bound on the number of independent absolutely continuous invariant measures. *J. Math. Anal. Appl.* **139** (1989), 139–151.
4. A. Boyarsky and M. Scarowsky. On a class of transformations which have unique absolutely continuous invariant measure. *Trans. Amer. Math. Soc.* **255** (1979), 243–262.
5. M. Bramson and S. A. Kalikow. Nonuniqueness in g-functions. *Israel J. Math.* **84** (1993), 153–160.
6. W. Doeblin and R. Fortet. Sur les chaînes à liaisons complètes. *Bull. Soc. Math. France* **65** (1937), 132–148.
7. P. Góra and B. Schmitt. Un exemple de transformation dilatante et C^1 par morceaux de l'intervalle, sans probabilité absolument continue invariante. *Ergodic Theory Dynamical Systems* **9** (1989), 101-113.
8. P. Hulse. Uniqueness and ergodic properties of ergodic g-measures. *Ergodic Theory Dynamical Systems* **11** (1991), 65–77.
9. T. Kaijser. On a Theorem of Karlin. *Preprint: Department of Electrical Engineering, Linköping University, Sweden* (1993).
10. S. A. Kalikow. Random Markov processes and uniform martingales. *Israel J. Math.* **71** (1990), 33–54.
11. S. Karlin. Some random walks arising in learning models, I. *Pac. J. Math.* **3** (1953), 725–756.
12. M. Keane. Strongly mixing g-measures. *Invent. Math* **16** (1972), 309–324.
13. Z. S. Kowalski. Invariant measures for piecewise monotonic transformations. *Proc. Winter School of Probability, Karpacz 1975. Springer Lecture Notes in Mathematics* **472** (1975), 77–94.
14. K. Krzyżewski. On expanding mappings. *Bull. Acad. Polon. Sci. Sér. Sci. Math.* **XIX** (1971), 23–24.
15. K. Krzyżewski. A remark on expanding mappings. *Colloq. Math.* **XLI** (1979), 291–295.
16. K. Krzyżewski and W. Szlenk. On invariant measures for expanding differentiable mappings. *Studia Math.* **XXXIII** (1969), 83–92.
17. A. Lasota and J. A. Yorke. On the existence of invariant measures for piecewise monotonic transformations. *Trans. Amer. Math. Soc.* **186** (1973), 481–488.
18. T. Y. Li and J. A. Yorke. Ergodic transformations from an interval into itself. *Trans. Amer. Math. Soc.* **235** (1978), 183–192.
19. T. Lindvall *Lectures on the Coupling Method.* Wiley: New York, 1992.

20. R. Mañé *Ergodic Theory and Differentiable Dynamics*. Springer: New York, 1988.
21. A. N. Quas. University of Warwick Thesis: Some problems in ergodic theory (1993).
22. P. Walters. Ruelle's operator theorem and g-measures. *Trans. Amer. Math. Soc.* **214** (1975), 375–387.
23. S. Wong. Some metric properties of piecewise monotonic mappings of the unit interval. *Trans. Amer. Math. Soc.* **246** (1978), 493–500.

Statistical Laboratory, Department of Pure Mathematics and Mathematical Statistics, 16 Mill Lane, Cambridge, CB2 1SB, ENGLAND