On non-regular $g$-measures

Sandro Gallo$^1$ and Frédéric Paccaut$^2$

$^1$ Departamento de Métodos Estatísticos, Instituto de Matemática, Universidade Federal do Rio de Janeiro, Caixa Postal 68530, CEP 21941-909, Rio de Janeiro, Brazil
$^2$ Laboratoire Amiénois de Mathématiques Fondamentales et Appliquées CNRS UMR 7352, Université de Picardie Jules Verne, 33 rue Saint Leu, 80039 AMIENS Cedex 1, France

E-mail: sandro@im.ufrj.br and frederic.paccaut@u-picardie.fr

Received 17 September 2012, in final form 18 December 2012
Published 7 February 2013
Online at stacks.iop.org/Non/26/763

Recommended by J Marklof

Abstract

We prove that $g$-functions whose set of discontinuity points has strictly negative topological pressure and which satisfy an assumption that is weaker than non-nullness, have at least one stationary $g$-measure. We also obtain uniqueness by adding conditions on the set of continuity points.

Mathematics Subject Classification: 60J05, 37E05

1. Introduction

The $g$-measures on $A^Z$ ($A$ discrete) are the measures for which the conditional probability of one state at any time, given the past, is specified by a function $g$, called the $g$-function. In this paper, $g$-measures will always refer to stationary measures. The main question we answer in the present paper is the following: what conditions on $g$-functions $g$ will ensure the existence of a (stationary) $g$-measure?

It is well-known that the continuity of $g$ implies existence if the alphabet $A$ is finite. Here we extend this result to discontinuous $g$-functions by proving that existence holds whenever the topological pressure of the set of discontinuities of $g$ is strictly negative, even when $g$ is not necessarily non-null.

The name $g$-measure was introduced by Keane (1972) in ergodic theory to refer to an extension of the Markov measures, in the sense that the function $g$ may depend on an unbounded part of the past. In the literature of stochastic processes, these objects already existed under the names ‘Chaînes à liaisons complètes’ or ‘chains of infinite order’, respectively coined by Doeblin and Fortet (1937) and Harris (1955). The function $g$ is also called a set of transition probabilities, or probability kernel. Given a function $g$ (or probability kernel), the most basic questions are the following: does it specify a $g$-measure (or stationary stochastic process)? If yes, is it unique? To answer these questions, the literature has mainly focused on the continuity
assumption for $g$ (see Onicescu and Mihoc (1935), Doeblin and Fortet (1937), Harris (1955), Keane (1972), Ledrappier (1974), Johansson and Oberg (2003), Fernández and Maillard (2005) and many others). This assumption gives ‘for free’ the existence of the $g$-measure. For this reason, uniqueness and the study of the statistical properties of the resulting unique measure have been the centre of attention from the beginning of the literature. Only recently, Gallo (2011), Cénac et al (2012), De Santis and Piccioni (2012) studied $g$-measures with functions $g$ that were not necessarily continuous. However, no general criteria have been given regarding the existence issue, either because these works are example-based, or because the obtained conditions are restrictive, implying both existence and uniqueness. This gives rise to a natural motivation to find a general criteria for the existence of $g$-measures.

A second motivation is the analogy with one-dimensional Gibbs measures. In statistical mechanics, the function specifying the conditional probabilities with respect to both past and future is called a specification. The theorem of Kozlov (1974) states that Gibbs measures have continuous and strictly positive specifications. Stationary measures with support on the set of points where the specification is continuous are called almost-Gibbsian (Meas et al 1999). Clearly, Gibbsian measures are almost-Gibbsian. Fernández et al (2011) proved that regular $g$-measures (associated to continuous and strictly positive function $g$) might not be Gibbs measures, still they are always almost-Gibbsian. Thus, although the nomenclature of Gibbsianity cannot be imported directly to the case of $g$-measures, it is tempting to try to find ‘almost-regular’ $g$-measures.

Going further in the analogy between $g$-measures and (almost-)Gibbs measures, a natural idea is to look for a $g$-measure with support inside the set of continuity points of $g$. Of course, it is not an easy task to control the support of a measure before knowing its existence. The idea is then to put a topological assumption on the set of discontinuity points of $g$, ensuring that this set will have $\mu$-measure 0, whenever the $g$-measure $\mu$ exists. In the vein of Buzzi et al (2001), this is done in the present paper by using the topological pressure of the set of discontinuity points of $g$. Theorem 1 states that there exist $g$-measures when the function $g$ has a set of discontinuity points with negative topological pressure, even without assuming non-nullness. As a corollary (corollary 1), a simple condition on the set of discontinuity points of a function $g$ is given, which may appear more intuitive to the reader not familiar with the concept of topological pressure. The set of discontinuity points of $g$ can be seen as a tree where each branch is $A^{-\mathbb{N}}$. The new condition is that the upper exponential growth rate of this tree is smaller than a constant that depends on $\inf_X g$ (or, if non-nullness is not assumed, on some parameter explicitly computable on $g$). Our last result (theorem 2), based on the work of Johansson and Oberg (2003), gives explicit sufficient conditions on the set of continuity points of discontinuous kernels $g$ (satisfying our conditions of existence) ensuring uniqueness.

2. Notations, definitions and main results

Let $(A, \mathcal{A})$ be a measurable space, where $A$ is a finite set (the alphabet) and $\mathcal{A}$ is the associated discrete $\sigma$-algebra. We will denote by $|A|$ the cardinal of $A$. Define $X = A^{-\mathbb{N}}$ (we use the convention that $\mathbb{N} = \{0, 1, 2, \ldots\}$), endowed with the product of discrete topologies and with the $\sigma$-algebra $\mathcal{F}$ generated by the coordinate applications. For any $x \in X$, we will use the notation $x = (x_i)_{i \in \mathbb{N}} = x_0^\infty = \ldots x_{-1}x_0$. For any $x \in X$ and $z \in X$, we denote, for any $k \geq 0$, $z x_{-k}^0 = \ldots z_{-2}z_{-1}z_0x_{-k}^0 \ldots x_0$, the concatenation between $x_{-k}^0$ and $z$. In other words, $z x_{-k}^0$ denotes a new sequence $y \in X$ defined by $y_i = z_{i+k}$ for any $i \leq -k - 1$ and $y_i = x_i$ for any $-k \leq i \leq 0$. Finally, the length of any finite string $v$ of elements of $A$, that is, the number of letters composing the string $v$, will be written $|v|$.
Define the shift mapping $T$ as follows:

$$T : X \to X$$

$$(x_n)_{n \leq 0} \mapsto (x_{n-1})_{n \leq 0}.$$  

The mapping $T$ is continuous and has $|A|$ continuous branches called $T_a^{-1}$, $a \in A$. Denote by $\mathcal{M}$ the set of Borelian probability measures on $X$, by $\mathcal{B}$ the set of bounded functions and by $\mathcal{C}$ the set of continuous functions. The characteristic functions will be written $I$.

A $g$-function is a $\mathcal{F}$-measurable function $g : X \to [0, 1]$ such that

$$\forall x \in X, \quad \sum_{y : T(y) = x} g(y) = \sum_{a \in A} g(a x) = 1.$$  

(2.1)

**Example 1.** Matrix transitions of $k$-steps Markov chains, $k \geq 1$, are the simplest example of $g$-functions. They satisfy $g(xa) = g(ya)$ whenever $x_{-k+1}^0 = y_{-k+1}^0$, $\forall a$.

**Example 2.** Let us introduce one of the simplest examples of non-Markovian $g$-function on $A = [0, 1]$. Let $(q_n)_{n \in \mathbb{N} \cup \{\infty\}}$ be a sequence of $[0, 1]$-valued real numbers. Set $\tilde{g}(x 1) = q_\ell(x)$ where $\ell(x) := \inf\{k \geq 0 : x_{-k} = 1\}$ for any $x \in A^{-\mathbb{N}}$ (with the convention that $\ell(x) = \infty$ whenever $x_i = 0$ for all $i \leq 0$). Notice that the value of $\tilde{g}(x)$ depends on the distance to occurrence of a symbol 1 in the sequence $\ldots x_{-1} x_0$. Therefore, for any $k \geq 1$ the property that $g(xa) = g(ya)$ whenever $x_{-k+1}^0 = y_{-k+1}^0 = 0_{k+1}^0$ does not hold. This is not the transition matrix of a Markov chain. We will come back to this motivating example several times throughout this paper.

**Definition 1.** An $A$-valued stochastic process $(\xi_n)$ defined on a probability space $(\Omega, \mathcal{G}, \mathbb{P})$ is specified by a given $g$-function $g$ if

$$\mathbb{P}(\xi_0 = a| (\xi_k)_{k < 0}) = g(\ldots \xi_{-2} \xi_{-1} a) \quad \mathbb{P} \text{ almost surely}.$$  

The distribution of a stationary process $(\xi_n)$ of this form is called a $g$-measure.

Here is a more ergodic oriented, equivalent definition:

**Definition 2.** Let $g$ be a $g$-function. A probability measure $\mu \in \mathcal{M}$ is called a $g$-measure if $\mu$ is $T$-invariant and for $\mu$ almost every $x \in X$ and for every $a \in A$:

$$\mathbb{E}_\mu(\mathbb{1}_{[x_0 = a]}|\mathcal{F}_1)(x) = g(T(x) a).$$

with $\mathcal{F}_1 = T^{-1}\mathcal{F}$.

Given a $g$-function, the existence of a corresponding $g$-measure is not always guaranteed. For instance, coming back to example 2, Cénac et al (2012) proved that if $\prod_{k \geq 1} \sum_{i=0}^{k-1} (1 - q_i) = \infty$ and $q_\infty > 0$, then there does not exist any $g$-measures for $\tilde{g}$. Another simple example is given by Keane (1972) on the torus. In general, a sufficient condition for the existence of a $g$-measure corresponding to some fixed $g$-function is to assume that $g$ is continuous in every point (see Keane (1972) for instance). Continuity here is understood with respect to the discrete topology, that is, $g$ is continuous at the point $x$ if for any $\varepsilon$, we have

$$g(z x_{-k}^0) \xrightarrow{k \to \infty} g(x).$$

Continuity is nevertheless not necessary for existence, as shown, one more time, by the $g$-function $\tilde{g}$ of example 2. For instance, let $q_i = \varepsilon < 1/2$ when $i$ is odd and $q_i = 1 - \varepsilon$ when $i$ is even, and put $q_\infty > 0$. Observe that in this case $\tilde{g}$ has a discontinuity at $0_{-\infty}^0$, since $\tilde{g}(1_{-\infty}^0 0_{-\infty}^1)$ oscillates between $\varepsilon$ and $1 - \varepsilon$ when $k$ increases. But it is well-known that $\tilde{g}$ has a $g$-measure (see Cénac et al (2012) or Gallo (2011) for instance).
The preceding observations pertain our first issue, which is to give a general condition on
the set of discontinuities of \( g \), under which there still exists a \( g \)-measure. This is the content
of theorem 1, which we will state after introducing some further definitions.

The cylinders are defined in the usual way by
\[
C_n(x) = \{ w \in X, w_{-n+1}^0 = x_{-n+1}^0 \}, \quad \forall x \in X,
\]
and the set of \( n \)-cylinders is
\[
C_n = \{ C_n(x), x \in X \}.
\]
Define, for \( x \in X \) and \( n \in \mathbb{N}, n \geq 1 \)
\[
g_n(x) = \prod_{i=0}^{n-1} g(T^i(x)).
\]

The topological pressure of a measurable set \( S \subset X \) is defined by
\[
P_g(S) = \limsup_{n \to +\infty} \frac{1}{n} \log \sum_{B \in C_n} \sup B g_n.
\]

Let \( D \) be the set of discontinuity points of \( g \). Let \( C_n(D) \) be the union of \( n \)-cylinders that intersect \( D \):
\[
C_n(D) = \bigcup_{x \in D} C_n(x).
\]
For \( n \in \mathbb{N} \), set \( \mathcal{E}_n = T^{-1} T^0 C_{n+1}(D) \) (notice that \( \mathcal{E}_0 = X \) and \( \mathcal{E}_{n+1} \subset \mathcal{E}_n \)). \( \mathcal{E}_n \) is the set of points that may be written \( yx_{-n}^0 a \), with \( a \in A, x_{-\infty}^0 \in D \) and \( y \in X \).

**Theorem 1.** Let \( g \) be a \( g \)-function with discontinuity set \( \mathcal{D} \). Assume
\[
\textbf{(H1)} \quad \exists N \in \mathbb{N}, \inf_{\mathcal{E}_N} g > 0,
\]
\[
\textbf{(H2)} \quad P_g(D) < 0,
\]
then there exists at least one \( g \)-measure. Moreover, the support of any \( g \)-measure is contained
in \( X \setminus \mathcal{D} \) as soon as \( \textbf{(H2)} \) is fulfilled.

**Remark 1.** Hypothesis \( \textbf{(H1)} \) is strictly weaker than the ‘strong non-nullness’ assumption
\( \inf g > 0 \), since the later corresponds to \( \textbf{(H1)} \) being satisfied for \( N = 0 \) and example 4 satisfies \( \textbf{(H1)} \) and is not strongly non-null.

**Remark 2.** Notice that \( \textbf{(H2)} \) is fulfilled for instance when \( \mathcal{D} \) is a finite set and \( \inf g > 0 \) (i.e. \( \textbf{(H1)} \) is fulfilled with \( N = 0 \)). This is, in particular, the case of our simplest example 2 when the \( q_i \) are oscillating between \( \varepsilon \) and \( 1 - \varepsilon \). Notice also that \( \textbf{(H2)} \) is fulfilled as well as soon as \( \mathcal{D} \) is finite, \( \textbf{(H1)} \) is fulfilled with \( N > 1 \) and \( T \mathcal{D} \subset \mathcal{D} \). This will be an easy consequence of corollary 1.

**Remark 3.** Notice also that \( \textbf{(H2)} \) implies that \( g \) cannot be everywhere discontinuous. Namely, the property 2.1 of a \( g \)-function entails:
\[
\forall n \in \mathbb{N}^*, \forall y \in X, \sum_{x_{-n+1}^0 \in A^y} g_n(yx_{-n+1}^0) = 1
\]
and therefore \( \sum_{B \in \mathcal{E}_n} \sup_B g_n \geq 1 \) which in turn implies that \( P_g(X) \geq 0 \).
Example 3. This example was presented in De Santis and Piccioni (2012) (see example 2 therein) on $[-1, +1]$. Here we adapt it on the alphabet $A = \{0, 1\}$. As $\tilde{g}$, the $g$-function we introduce here has a unique discontinuity point along $0^{0}_{-\infty}$, but the dependence on the past does not stop at the last occurrence of a 1. Recall that $\ell(x) := \inf\{k \geq 0 : x_{-k} = 1\}$. Let $g(0^{0}_{-\infty}) = \epsilon > 0$, and for any $x \neq 0^{0}_{-\infty}$ and any $a \in \{0, 1\}$ let

$$g(xa) = \epsilon + (1 - 2\epsilon) \sum_{n \geq 1} 1_{\{x_{-\ell(x)-n} = a\}} q_n^{(\ell(x))} ,$$

where, for any $l \geq 0$, $(q_n^l)_{n \geq 1}$ is a probability distribution on the integers. This kernel has a discontinuity at $0^{0}_{-\infty}$ since for each $k \in \mathbb{N}$,

$$g(\ldots 1110^{0}_{-k+1}1) = \epsilon + (1 - 2\epsilon) \sum_{n \geq 1} q_n^{k+1} = 1 - \epsilon \neq \epsilon,$$

but it is continuous at any other point, since for any $x$ such that $\ell(x) = l < +\infty$, for any $z$ and $k > l$

$$g(\ldots z_{-l}^{0}_{-k+1}1) = \epsilon + (1 - 2\epsilon) \sum_{j \geq 1} 1_{\{x_{-l-j} = 1\}} q_j^l + \sum_{j \geq k-l+1} 1_{\{z_{k-l+1-j} = 1\}} q_j^l$$

which converges to $g(x1) = \epsilon + (1 - 2\epsilon) \sum_{j \geq 1} 1_{\{1 = x_{-l-j}\}} q_j^l$. Under some assumptions on the set of distributions $(\{q_n^l\}_{n \geq 1})_{l \geq 0}$, De Santis and Piccioni (2012) proved existence, uniqueness and perfect formula, while our theorem 1 guarantees existence of a $g$-measure, without any further assumptions on this sequence of distributions.

Theorem 1 involves the notion of topological pressure, which is not always easy to extract from the set of discontinuities. We now introduce two simple criteria on the set $D$ of a $g$-function, that will imply existence.

**Definition 3.** For any $n \geq 0$, let us denote $D^n := \{x_{-n+1}^{0}_{-n+1}\}_{x \in D}$. The upper exponential growth rate of $D$ is

$$\bar{gr}(D) := \limsup_n |D^n|^{1/n} .$$

(2.2)

Although this nomenclature is generally reserved for trees, we use it here as there exists a natural way to represent the set $D$ as a rooted tree (a subtree of $A^{-\mathbb{N}}$) with the property that each branch, representing an element of $D$, is infinite, and each node has between 1 and $|A|$ sons. For instance, in the particular case of $\tilde{g}$ (example 2), the tree is the single branch $0^{0}_{-\infty}$ and $D^n = 0^{0}_{-n+1}$.

**Corollary 1.** Let $g$ be a $g$-function with discontinuity set $D$. Assume either,

(H1') \hspace{1cm} \exists \epsilon > 0, \inf_x g = \epsilon , \hspace{1cm} \text{or} \hspace{1cm} \bar{gr}(D) < [1 - (|A| - 1)\epsilon]^{-1} ,

or

(H1) \hspace{1cm} \exists N \in \mathbb{N}, \exists \epsilon > 0, \inf_x g = \epsilon , \hspace{1cm} \bar{gr}(D) < [1 - (|A| - 1)\epsilon]^{-1} , \hspace{1cm} \text{or} \hspace{1cm} TD \subset D ,

then there exists at least a $g$-measure and its support is contained in $X \setminus D$. 

Intuitively, corollary 1 states that if $\varepsilon$ (which plays the role of a 'non-nullness parameter' for $g$) is sufficiently large, it may compensate the set of discontinuities of $g$, allowing $g$-measures to exist, with support on the continuity points. Notice that this assumption allows $D$ to be uncountable, as shown in the following example.

**Example 4.** Let $A = \{0, 1, 2\}$, and consider the function $\ell$ defined as in examples 3 and 2. Let also $N_0, N_1$ and $N_2$ be three disjoint finite subsets of $\mathbb{N}$. The $g$-function is defined as follows: for $x \in \{0, 2\}^{-\mathbb{N}}$, put

$$g(x) = g(x) = g(0) = 0.3, \text{ for } x \text{ such that } \ell(x) \in N_0 \cup N_1 \cup N_2,$$

and for any $x$ such that $\ell(x) \in \mathbb{N} \setminus \{N_0 \cup N_1 \cup N_2\}$, put

$$g(x) = g(x) = g(0) = 0.26 + \sum_{k \geq 1} \theta_k x_{-\ell(x)-k},$$

where $(\theta_k)_{k \geq 1}$ satisfies $\theta_k \geq 0$ and $\sum_{k \geq 1} \theta_k < 0.03$. Observe that, for any $x \in \{0, 2\}^{-\mathbb{N}}$, $g(x)_{\{111x_0^0 1\}} < 0.29$ for any sufficiently large $k$, and therefore does not converge to 0.3. So $\{0, 2\}^{-\mathbb{N}} \subset D$. On the other hand, any point $x$ satisfying $\ell(x) \in N_0 \cup N_1 \cup N_2$ is trivially continuous, and any point $x$ satisfying $\ell(x) \in \mathbb{N} \setminus \{N_0 \cup N_1 \cup N_2\}$ is continuous since for any $k > l$ and any $y \in \{0, 1, 2\}^{-\mathbb{N}},$

$$g(\ldots y_{-2} y_{-1} y_0 x_{-k} 0 1) = 0.26 + \sum_{i=1}^{k-l} \theta_i x_{-\ell(y)-i} + \sum_{i \geq k-l+1} \theta_i y_{i-k-l-1},$$

which converges to $0.26 + \sum_{i \geq 1} \theta_i y_{-i}$. So $D = \{0, 2\}^{-\mathbb{N}}$ (which is uncountable), $|D'| = 2^\mathbb{N}$ and consequently $g(x) \in D$. Observe on the other hand that $\inf_x g = 0$, but there exists $N$ such that $\inf_{x \in N} g \geq 0.26$ (any $N > \max(N_0 \cup N_1 \cup N_2)$ will do the job). Thus, the hypothesis of corollary 1 are fulfilled since $1 - (|A| - 1)\varepsilon = 0.48 < 1/2$, and existence holds.

So far, we have focused on the existence issue. However, Bramson and Kalikow (1993) proved that even regular $g$-functions (continuous $g$-functions satisfying (H1')) might have several $g$-measures. In view of a result on uniqueness for non-regular $g$-measures, we now give a condition on the set of continuous pasts $X \setminus D$. To do so, we use the notion of context tree defined below.

**Definition 4.** A context tree $\tau$ on $A$ is a subset of $\bigcup_{k \geq 0} A^{+[k, \ldots, 0]} \cup X$ such that for any $x \in X$, there exists a unique element $v \in \tau$ satisfying $x_{[0]} = v \in [0, 1, 2]$. For any $g$-function $g$, we denote by $\tau^g$ the smallest context tree containing $D$, called the skeleton of $g$. For instance, coming back to example 2, $\tau^g = \bigcup_{j \geq 0} \{10^j\} \cup \{0^\infty, 1\}$ and is represented on figure 1. It is also the skeleton of any $g$-function having only $0^\infty_{\infty}$ as discontinuity point, such as the $g$-function introduced in example 3. Pictorially, any $g$-function can be represented as a set of transition probabilities associated to each leaf of the complete tree $A^{-\mathbb{N}}$ and $\tau^g$ is the smallest subtree of $A^{-\mathbb{N}}$ which contains $D$, such that every node has either $|L|$ or 0 sons. Figure 2 shows the (upper part of) the complete tree corresponding to some function $g$ having complicated sets $D$ and $\tau^g$.

Let us introduce the $n$-variation of a point $x \in X$ that quantifies the rate of continuity of $g$ as

$$\text{var}_n(x) := \sup_{y, y' \in [x]} |g(y) - g(x)|.$$
Notice that $\text{var}_n(x)$ converges to 0 if and only if $x$ is a continuity point for $g$. As $\text{var}_n(x)$ actually only depends on $x_{-n}^0$, the notation $\text{var}_n(x_{-n}^0)$ will sometimes be used. Now, observe that the set of continuous pasts of a given $g$-function $g$ is the set of pasts $x_{-\infty}^0$ such that there exists $v \in \tau^g$, $|v| < +\infty$ with $x_{-|v|+1}^0 = v_{|v|+1}^0$. In particular, for any $v \in \tau^g$ with $|v| < +\infty$,

$$\text{var}_n^v := \sup_{x,x_{-|v|+1}^0 = v} \text{var}_n(x) \xrightarrow{n \to +\infty} 0.$$  

For any $v \in \tau^g$, $|v| < +\infty$, let $R_v := \sum_{n \geq |v|} |\text{var}_n^v|^2$. Our assumption on the set of continuous pasts $X \setminus D$ is

$$(H4) \quad \sum_{v \in \tau^g, |v| < +\infty} \mu(v) R_v < +\infty,$$

where $\mu$ is any $g$-measure given by theorem 1. Observe that (H4) implies that $R_v < +\infty$ for any $v \in \tau$.

**Theorem 2.** Suppose that we are given a $g$-function $g$ satisfying (H1), (H2) and (H4) for some $g$-measure $\mu$, then this $g$-measure is unique.

**Remark 4.** In this theorem, hypothesis (H1) and (H2) are mainly used to get the existence of a $g$-measure. Therefore, thanks to corollary 1, the same conclusion holds either assuming (H1'), (H2') and (H4) or (H1), (H2'), (H3) and (H4).
This result is to be compared to the results of Johansson and Öberg (2003), which state, in particular, that uniqueness holds when \( \text{var}_n := \sup_{x_0} \text{var}_n(x_0) \) is in \( \ell^2 \). In fact, this is mainly what is assumed here, but only on the set of continuous pasts, which has full \( \mu \)-measure. This is formalized through the more complex hypothesis (H4). We now come back to examples 3 and 4 in order to illustrate theorem 2.

**Example 3 (Continued).** In this example, we have as skeleton \( \tau^g = (-\infty, 0^n) \cup \bigcup_{i \geq 0} \{10^i\} \), so that any \( v \in \tau^g \) with \( |v| = k < \infty \) writes \( v = 10^{n-1} \) and simple calculations yield, for any \( n \geq k \)

\[
\text{var}_n = (1 - 2\epsilon) \sum_{i \geq n-k+1} q_i.
\]

Hypothesis (H4) is satisfied as soon as

\[
\sum_{k \geq 1} (1 - \epsilon)^k \sum_{n \geq k+1} \left[ \sum_{l \geq n-k+1} q_l \right]^2 < +\infty.
\]

For instance, if for any \( k \geq 1 \), \( (q_i^k)_{i \geq 1} \) is the geometric distribution with parameter \( \alpha^k \), where \( 1 - \epsilon < \alpha < 1 \), then

\[
\sum_{k \geq 1} (1 - \epsilon)^k \sum_{n \geq 1} \left[ \sum_{l \geq n+1} q_l \right]^2 \leq \sum_{k \geq 1} [(1 - \epsilon) \alpha^{-1}]^k,
\]

which is summable. So we have uniqueness for this kernel.

**Example 4 (Continued).** The skeleton of \( g \) is

\[\tau^g = [0, 2]^\mathbb{N} \cup \{1\} \cup \bigcup_{i \geq 0} \{ i0^i \} \cup \bigcup_{i \geq 0} \{1x^0_{-i}\}\]

and for any \( v \in \tau^g \), \( |v| < \infty \),

\[
\text{var}_n \leq 2 \sum_{i \geq n - |v|} \theta_i, \quad \forall n > |v|.
\]

Since this upper bound does not depend on the length of the string \( |v| \), it follows that hypothesis (H4) is satisfied if \( \sum_{n \geq 1} \sum_{i \geq n} \theta_i^2 < +\infty \).

### 3. Proof of theorem 1

Let us define the Perron Frobenius operator \( L \) acting on measurable functions \( f \) as follows:

\[
L f (x) = \sum_{a \in A} g(xa) f(xa) = \sum_{x \in T(y)} g(y) f(y).
\]

For \( \mu \in \mathcal{M} \), let \( \mu^* \) denote the dual operator, that is

\[
\mu^*(f) = \mu(Lf)
\]

for any \( f \in B \). The relation between \( L^* \) and the \( g \)-measures is illuminated by the following result.

**Proposition 1 (Ledrappier (1974)).** \( \mu \) is a \( g \)-measure if and only if \( \mu \) is a probability measure and \( L^* \mu = \mu \).
We start with the proof of the second part of the theorem. Assume that (H2) is fulfilled and that \( \mu \) is a \( g \)-measure. Then, a consequence of proposition 1 and of the definition of pressure is that, for any \( \delta > 0 \), there exists \( N(\delta) \) such that, for any \( n > N(\delta) \),

\[
\mu(D) = \mu(L^n \1_D) \leq \sum_{C \in \mathcal{C}(D)} \sup_{C} g_{n} \leq (e^{P_{g}(D)+\delta})^{n}.
\]

Taking \( \delta = -P_{g}(D)/2 \) and letting \( n \) tend to infinity implies \( \mu(D) = 0 \).

As for the existence, in view of proposition 1, the strategy of the proof will be to find a fixed point for \( L^{*} \). When \( g \) is a continuous function, the operator \( L \) acts on \( \mathcal{C} \) and \( L^{*} \) acts on \( \mathcal{M} \), the existence of a \( g \)-measure \( \mu \) is then a straightforward consequence of the Schauder–Tychonoff theorem.

If \( g \) is not continuous, \( L \) does not preserve the set of continuous functions. More precisely, if \( D \) is the set of discontinuities of \( g \) and \( f \) is continuous, then the set of discontinuities of \( Lf \) is \( T^{D} \). Still, as \( g \) is bounded, \( L \) acts on the space \( \mathcal{B} \) of bounded functions. More precisely \( \|Lf\| \leq \|f\| \), where \( \| \cdot \| \) is the uniform norm. Therefore \( L^{*} \) acts on \( \mathcal{B}' \), the topological dual space of \( \mathcal{B} \) i.e.

\[
L^{*}\alpha(f) = \alpha(Lf)
\]

for all \( \alpha \in \mathcal{B}' \) and \( f \in \mathcal{B} \).

Firstly, the existence of a fixed point \( \Lambda \in \mathcal{B}' \) for \( L^{*} \) will be proved. Then the hypothesis (H1) and (H2) will be shown to imply \( \mu(D) = 0 \) and \( \mu(TD) = 0 \), where \( \mu \) is the restriction of \( \Lambda \) to the continuous functions. Finally, we will use these two equalities to prove that \( \mu \) is indeed a \( g \)-measure.

**Proposition 2.** There exists a positive functional \( \Lambda \in \mathcal{B}' \) with \( \Lambda(\1) = 1 \) such that \( L^{*}\Lambda = \Lambda \).

**Proof.** Consider the following subset \( C \) of \( \mathcal{B}' \)

\[
C = \{ \alpha \in \mathcal{B}', \alpha(\1) = 1 \text{ and } \alpha(f) \geq 0 \text{ for all } f \geq 0 \}.
\]

We consider the weak star topology on \( \mathcal{B}' \) and \( C \). In order to apply the Schauder–Tychonoff theorem (Dunford and Schwartz (1988) V.10.5), it is necessary that \( L^{*} \) is well defined and continuous for the weak star topology, that \( C \) is compact for this topology, non-empty and convex (the two last properties are straightforward). The continuity of \( L^{*} \) is given by a simplification of the proof in Buzzi et al (2001). The compactness of \( C \) follows from the Banach–Alaoglu theorem (Dunford and Schwartz (1988) V.4.2), as \( C \) is a closed subset of the unit ball of \( \mathcal{B}' \).

Since \( \Lambda_{C} \) is a positive linear form on \( \mathcal{C} \), the Riesz representation theorem implies that there exists \( \mu \), a positive Borel measure, such that:

\[
\forall f \in \mathcal{C} : \Lambda(f) = \mu(f).
\]

In particular, \( \mu(\1) = \Lambda(\1) = 1 \) so that \( \mu \) is a probability measure.

For all \( f \in \mathcal{C} \), \( \Lambda(Lf) = \Lambda(f) = \mu(f) \). But \( Lf \) is not necessarily continuous at points of \( TD \). Notice though that if \( f \in \mathcal{C} \) and \( Lf \in \mathcal{C} \) then \( \mu(f) = \mu(Lf) \). What remains to prove is that this is true for any \( f \in \mathcal{C} \).

Two more lemmas are needed to go on further in the proof.

**Lemma 1.**

\[
P_{g}(TD) \leq P_{g}(D)
\]
Proof. By definition:
\[ P_g(TD) = \limsup_{n \to \infty} \frac{1}{n} \log \sum_{B \in C_n} \sup_B g_n. \]
Let \( B \in C_n \) such that \( B \cap TD \neq \emptyset \), there exists \( C \in C_{n+1} \) such that \( C \cap D \neq \emptyset \). More precisely, there exists \( a \in A \) such that \( C = C_1(a) \cap T_{a}^{-1}(B) \). Moreover, let \( x \in B \), then \( T_{a}^{-1}(x) \in E_n \) and
\[ g_n(x) \leq \frac{g_{n+1}(T_{a}^{-1}(x))}{g(T_{a}^{-1}(x))} \leq \frac{1}{\inf_{E_n} g} \sup_C g_{n+1}. \]
Since \( E_{n+1} \subset E_n \), \( \sup_B g_n \leq \frac{1}{\inf_{E_n} g} \sup_C g_{n+1} \) for \( n \geq N \). Recall that \( \inf_{E_n} g > 0 \) by hypothesis (H1). It comes, for \( n \geq N \):
\[ \sum_{B \in C_n} \sup_B g_n \leq \frac{1}{\inf_{E_n} g} \sum_{C \in C_{n+1}} \sup_C g_{n+1} \]
and thus:
\[ \limsup_{n \to \infty} \frac{1}{n} \log \sum_{B \in C_n} \sup_B g_n \leq \limsup_{n \to \infty} \frac{1}{n} \log \frac{1}{\inf_{E_n} g} + \limsup_{n \to \infty} \frac{1}{n+1} \log \sum_{C \in C_{n+1}} \sup_C g_{n+1}. \]
\[ \square \]

Lemma 2. For all Borel sets \( B \),
\[ \mu(B) \leq \inf \{ \Lambda(O), \ O \text{ open}, \ O \supset B \} \]
Proof. Since \( \mu \) is a regular measure (as a Borel measure on a compact set):
\[ \mu(B) = \inf \{ \mu(O), \ O \text{ open}, \ O \supset B \}. \]
Let us fix an open set \( O \) and show that: \( \mu(O) \leq \Lambda(O) \), this will prove the lemma. Take \( \varepsilon > 0 \). Using again the regularity of \( \mu \), there exists \( K_{\varepsilon} \), a compact subset of \( O \), such that:
\[ \mu(O) < \mu(K_{\varepsilon}) + \varepsilon. \]
Let \( f_{\varepsilon} : X \to [0, 1] \) be continuous and such that:
\[
\begin{cases}
  f_{\varepsilon} = 1 & \text{in } K_{\varepsilon} \\
  f_{\varepsilon} = 0 & \text{in } O^c \\
  f_{\varepsilon} \leq 1 & \text{in } O \setminus K_{\varepsilon}.
\end{cases}
\]
On one hand, \( f_{\varepsilon} \leq 1_{O} \) so that
\[ \mu(f_{\varepsilon}) = \Lambda(f_{\varepsilon}) \leq \Lambda(O) \text{ and } \sup_{r > 0} \mu(f_{r}) \leq \Lambda(O). \]
On the other hand, \( \mu(f_{\varepsilon}) \geq \mu(K_{\varepsilon}) > \mu(O) - \varepsilon \) so that:
\[ \mu(O) < \mu(K_{\varepsilon}) + \varepsilon \leq \mu(f_{\varepsilon}) + \varepsilon \]
and \( \mu(O) \leq \sup_{r > 0} \mu(f_{r}) \leq \Lambda(O). \)
\[ \square \]
Now, we claim the following:

Lemma 3.
\[ \mu(D) = 0 \quad \text{and } \mu(TD) = 0. \]
Proof. The claim will follow from lemma 2 if we can find open neighbourhoods $V$ of $D$ and $W$ of $TD$ with $\Lambda(V)$ and $\Lambda(W)$ arbitrarily small. Let us write the proof for $D$. The same scheme will work for $TD$.

Recall that $C_n(D) = \cup \{ C \in C_n, C \cap D \neq \emptyset \}$. Using the fixed point property of $\Lambda$ and the definition of pressure, we get, for any $\delta > 0$, $N(\delta)$ such that, for all $n > N(\delta)$:

$$\Lambda(C_n(D)) = \Lambda(L^n \mathbb{1}_{C_n(D)}) \leq \sum_{C \in C_n(D)} \sup C g_n \leq (e^{P^1(D) + \delta})^n.$$ 

Taking $\delta = -P^1(D)/2$, which is positive by the main hypothesis (H2), we get $\lim_{n \to \infty} \Lambda(C_n(D)) = 0$ and for every $n$, $C_n(D)$ is an open neighbourhood of $D$. □

Finally, the proof of the main theorem writes as follows:

Proof. Fix $f \in C(X)$ non-negative.

Since $\mu$ is regular (as a Borel measure on a compact set) and as $\mu(D) = \mu(TD) = 0$ (lemma 3), for each $\varepsilon > 0$, there exist $U_\varepsilon$ open neighbourhood of $D$ and $V_\varepsilon$ open neighbourhood of $TD$ such that $\mu(U_\varepsilon) < \varepsilon$ and $\mu(V_\varepsilon) < \varepsilon$. Let $W_\varepsilon = U_\varepsilon \cap T^{-1}V_\varepsilon$. This is also a neighbourhood of $D$ such that $\mu(W_\varepsilon) < \varepsilon$. Moreover, as $TW_\varepsilon \subset V_\varepsilon$, it comes $\mu(TW_\varepsilon) < \varepsilon$.

Consider now $f_\varepsilon$ with compact support in $X \setminus D$ such that:

$$\begin{cases}
    f_\varepsilon = f & \text{in } X \setminus W_\varepsilon \\
    f_\varepsilon \leq f & \text{in } W_\varepsilon.
\end{cases}$$

First, $L f_\varepsilon$ is continuous on $X$. Namely, $f_\varepsilon$ is continuous on $X$ so $L f_\varepsilon$ is continuous on $X \setminus TD$. Now, if $x \in TD$, it may be easily checked that the potentially discontinuous part of $L f_\varepsilon$ actually vanishes. This continuity implies $\mu(L f_\varepsilon) = \mu(f)$ and

$$|\mu(L f) - \mu(f)| = |\mu(L f_\varepsilon) + \mu(L(f - f_\varepsilon)) - \mu(f)|
\leq |\mu(f_\varepsilon - f) + \mu(L(f - f_\varepsilon))|
\leq 2 \| f \| \mu(W_\varepsilon) + \mu(L(f - f_\varepsilon)).$$

We need to show that $\mu(L(f - f_\varepsilon))$ is small. By definition of $f_\varepsilon$,

$$L(f - f_\varepsilon)(x) = \sum_{a \in A} (f - f_\varepsilon)(ax) g(ax) \mathbb{1}_{W_\varepsilon}(ax)$$

therefore

$$\mu(L(f - f_\varepsilon)) \leq \| g \| \| f - f_\varepsilon \| \sum_{a \in A} \mu(\mathbb{1}_{W_\varepsilon} \circ T_a^{-1})
\leq 2 \| g \| \| f \| \sum_{a \in A} \mu(T_a(W_\varepsilon))
\leq (2 \| g \| \| f \| |A|) \mu(TW_\varepsilon).$$

Letting $\varepsilon$ go to zero gives $\mu(L f) = \mu(f)$. □

4. Proofs of corollary 1 and theorem 2

Proof of corollary 1 using (H1'), (H2'). In view of theorem 1, it is enough to show that hypothesis (H1') and (H2') imply (H2). Under hypothesis (H1')

$$g_n(x) \leq (1 - (|A| - 1)\varepsilon)^n \quad \text{for any } x.$$ (4.1)
It follows that
\[ P_g(D) \leq \limsup_{n \to +\infty} \frac{1}{n} \log |D^n|(1 - |A| - 1)\epsilon)^n. \]

Now, under (H2'), there exists \( \alpha \in (0, 1) \) such that \( |D_n| \leq |1 - (|A| - 1)\epsilon|^{n(1-\alpha)} \) for any sufficiently large \( n \). Thus,
\[
P_g(D) \leq \limsup_{n \to +\infty} \frac{1}{n} \log(1 - (|A| - 1)\epsilon)^{-n(1-\alpha)}(1 - (|A| - 1)\epsilon)^n
= \lim_{n \to +\infty} \frac{1}{n} \log(1 - (|A| - 1)\epsilon)^{\alpha}
= \alpha \log(1 - (|A| - 1)\epsilon) < 0.
\]

**Proof of corollary 1 using \{(H1), (H2'), (H3)\}**. In view of theorem 1, it is enough to show that hypothesis (H1), (H2') and (H3) imply (H2). Under (H1)
\[
\forall n \geq N + 1, \quad \forall x \in E_n, \quad g(x) \leq 1 - (|A| - 1)\epsilon.
\]
Take \( B \in C_n(D) \) and \( x \in B \). Hypothesis (H3) implies that \( T^i x \in E_{n-i} \subseteq E_N \) for all \( i \in \{1, \ldots, n - N - 1\} \). Therefore the identity \( \varphi_n(x) = g_{n-N}(x)g_N(T^{n-N}x) \) entails for \( n \geq N + 1 \)
\[
\forall B \in C_n(D), \quad \forall x \in B, \quad \varphi_n(x) \leq (1 - (|A| - 1)\epsilon)^{n-N}.
\]

It follows that
\[
P_g(D) \leq \limsup_{n \to +\infty} \frac{1}{n} \log |D^n|(1 - (|A| - 1)\epsilon)^{n-N}.
\]

The rest of the proof runs as before, using hypothesis (H2').

**Proof of theorem 2**. We already know that existence holds, thanks to hypothesis (H1) and (H2). Remark 2 in Johansson and Öberg (2003) states that, if for some stationary \( \mu \) we have
\[
\int_X \mu(dx) \sum_n [\text{var}_n(x)]^2 < +\infty,
\]
then \( \mu \) is unique. Notice that although Johansson and Öberg (2003) deal with continuous \( g \)-functions throughout the paper, their uniqueness result only requires existence of a \( g \)-measure, which is what we have here.

For any point \( x \in X \), the sequence \( \{\sum_{n=0}^N [\text{var}_n(x)]^2\}_{N \geq 0} \) is monotonically increasing and positive, therefore
\[
\int_X \mu(dx) \sum_n [\text{var}_n(x)]^2 = \lim_{N} \int_X \mu(dx) \sum_{n=0}^N [\text{var}_n(x)]^2 \\
= \lim_{N} \sum_{n=0}^N \int_X \mu(dx) [\text{var}_n(x)]^2 \\
= \sum_n \int_X \mu(dx) [\text{var}_n(x)]^2 \\
= \sum_n \sum_{x_{-n} \in A_{-n}} \mu(x_{-n}) [\text{var}_n(x_{-n})]^2.
\]
On non-regular $g$-measures

where we used the Beppo–Levi theorem in the first line, and the fact that $\text{var}_n(x)$ only depends on $x_{-n}$ in the last line. We now divide into two parts as follows:

$$
\int_X \mu(dx) \sum_n \left[ \sum_{x_{-n} \in D_n^{n+1}} \mu(x_{-n}) [\text{var}_n(x_{-n})]^2 \right] = \sum_n \sum_{x_{-n} \in A^{n+1} \setminus D_n^{n+1}} \mu(x_{-n}) [\text{var}_n(x_{-n})]^2.
$$

For the first term of the right-hand side of the equality, we majorate $\text{var}_n(x_{-n})$ by 1 and we use the fixed point property of the $g$-measure $\mu$ to obtain, for any $\delta > 0$, the existence of $N(\delta)$ such that for all $n > N(\delta)$:

$$
\sum_n \sum_{x_{-n} \in D_n^{n+1}} \mu(x_{-n}) = \mu(C_{n+1}(D)) = \mu(L^{n+1} \mathbb{1}_{C_{n+1}(D)}) \\
\leq \sum_{C \in C_{n+1}(D)} \sup \ g_n \leq (e^{P_g(D) + \delta} + 1).
$$

Taking $\delta = -\frac{P_g(D)}{2}$ (which is strictly positive by hypothesis (H2)) proves

$$
\sum_n \sum_{x_{-n} \in D_n^{n+1}} \mu(x_{-n}) < \infty.
$$

It remains to consider the second term. Recall that if $x_{-n} \in A_n^{n+1} \setminus D_n^{n+1}$ then there exists $v \in \pi^n$ with $|v| \leq n$ such that $v$ is a prefix of $x_{-n}$ (denoted by $x_{-n} \geq v$). It becomes, using (H4),

$$
\sum_n \sum_{x_{-n} \in A_n^{n+1} \setminus D_n^{n+1}} \mu(x_{-n}) [\text{var}_n(x_{-n})]^2 = \sum_n \sum_{v \in \pi^n : |v| \leq n+1} \sum_{x_{-n} \geq v} \mu(x_{-n}) [\text{var}_n(x_{-n})]^2 \\
= \sum_n \sum_{v \in \pi^n : |v| \leq n+1} \mu(v) [\text{var}_n(v)]^2 \\
= \sum_n \mu(v) R_v \\
= \sum_{v \in \pi^n : |v| < +\infty} \mu(v) R_v < +\infty. \quad \square
$$

5. Questions and perspectives

Notice that existence is ensured by an assumption on the set of discontinuous pasts, whereas uniqueness involves a condition on the set of continuous pasts. For continuous chains, Johansson and Oberg (2003) obtained conditions on the continuity rate of the kernel, ensuring uniqueness. Making the necessary changes in the hypothesis, theorem 2 states that for discontinuous kernel, the same kind of conditions can be used but restricted to the set of continuous pasts, when the measure does not charge the discontinuous pasts.

Concerning mixing properties, it is known (using the results of Comets et al (2002) for example) that chains having summable continuity rate enjoy summable $\phi$-mixing rate. It is natural to expect that, as for the problem of uniqueness, the chains we consider will enjoy the same mixing properties under the same assumption, restricted to the set of continuous pasts.

Finally, it is worth mentioning an interesting parallel with the literature of non-Gibbs state. In this literature, there are examples of stationary measures that are not *almost-Gibbs*, meaning
that there exist stationary measures that give positive weight to the set of discontinuities with respect to both past and future. We do not enter into details and refer to Maes et al (1999) for the definition of this notion. As far as we know, no such example exist in the world of $g$-measures. More precisely, an interesting question is whether there exist examples of stationary $g$-measures that are not almost-regular, or if, on the contrary, $\mu(D) = 0$ is valid for every stationary $g$-measure.

Acknowledgments

The authors are grateful to Xavier Bressaud for interesting discussions during the Workshop Jorma’s Razor II. They also thank Antonio Galves and NUMEC for their warm hospitality. Both authors were partially supported by CAPES grant AUXPE-PAE-598/2011 and by CNPq (grant 474233/2012-0). The first author was partially supported by FAPESP grant 2009/09809-1.

References

Bramson M and Kalikow S 1993 Nonuniqueness in $g$-functions Israel J. Math. 84 153–60
Buzzi J, Paccaut F and Schmitt B 2001 Conformal measures for multidimensional piecewise invertible maps Ergod. Theory Dyn. Syst. 21 1035–49
Cénaç P, Chauvin B, Paccaut F and Pouyanne N 2012 Variable length Markov chains and dynamical sources Sémin. Probabilités XLIV, Lecture Notes in Math. 2046 1–39
Comets F, Fernández R and Ferrari P A 2002 Processes with long memory: regenerative construction and perfect simulation Ann. Appl. Probab. 12 921–43
De Santis E and Piccioni M 2012 Backward coalescence times for perfect simulation of chains with infinite memory J. Appl. Probab. 49 319–37
Doeblin W and Fortet R 1937 Sur des chaînes à liaisons complètes Bull. Soc. Math. France 65 132–48
Dunford N and Schwartz J 1988 Linear Operators. Part I. General Theory (New York: Wiley)
Fernández R and Maillard G 2005 Chains with complete connections: general theory, uniqueness, loss of memory and mixing properties J. Stat. Phys. 118 555–88
Fernández R, Gallo S and Maillard G 2011 Regular $g$-measures are not always gibbsian Electron. J. Prob. 16 732–40
Gallo S 2011 Chains with unbounded variable length memory: perfect simulation and a visible regeneration scheme Adv. Appl. Probab. 43 735–59
Harris T E 1955 On chains of infinite order Pacific J. Math. 5 707–24
Johansson A and Öberg A 2003 Square summability of variations of $g$-functions and uniqueness of $g$-measures Math. Res. Lett. 10 587–601
Keane M 1972 Strongly mixing $g$-measures Invent. Math. 16 309–24
Kozlov O K 1974 A Gibbs description of a system of random variables Problemy Peredači Informacii 10 94–103
Ledrappier F 1974 Principe variationnel et systèmes dynamiques symboliques Z. Wahrscheinlichkeitsber. Verwandte Geb. 30 185–202
Maes C, Redig F, Van Moffaert A and Leuven K U 1999 Almost Gibbsian versus weakly Gibbsian measures Stochastic Process. Appl. 79 1–15
Onicescu O and Mihoc G 1935 Sur les chaînes de variables statistiques Bull. Sci. Math. 59 174–92