IN Variant Measures And Markov Chains
With Random Transition Probabilities

By
Guoqi Lu AND Arunava Mukherjea (Tampa, Florida)

Abstract. In this paper, sufficient conditions for the existence of
(σ-finite) invariant measures for a class of Markov chains with random
transition probabilities are given. A special class of Markov chains
with random transition probabilities is also studied here to show the
relevance of attractors for certain iterated function systems to the in-
variant measures for these chains, and some of these results are illus-
trated with computer-generated pictures.

1. Introduction. One of the aims of this paper is to generalize the model of
Markov chains with random transition probabilities, extensively studied by
Cogburn in a series of papers (here we mention only [5]) and later by Orey
[16], to a locally compact Hausdorff state space. We provide a sufficient con-
dition for the existence of a non-trivial σ-finite and locally finite (i.e. finite on
compact subsets) invariant measure for the above-mentioned (skew) chains. In
Section 3, where we present this result, the method and other details are de-
scribed.

In Section 4, we consider a special class of the Markov chains we studied
in Section 3. These chains are induced by a number of affine maps on \( \mathbb{R}^d \). The
Markov random walks (with values in these affine maps) from these Markov
systems give rise to what has been called “attractors” in the literature. Such
attractors in the context of iterated function systems, in various situations, have
been studied extensively in recent years by M. Barnsley and his co-workers in
numerous papers. Also, recently in [13], in the i.i.d. context, such attractors
were studied in connection with invariant measures. The main purpose of this
section is to establish a connection between these attractors and the invariant
measures for the class of Markov chains that we study here. Our results in
Section 4 are then illustrated by appropriate computer-generated pictures of
attractors.

In Section 2, we present a result necessary and sufficient for the existence
of invariant probability measures for general Markov chains that have the Fel-
ler property (i.e. takes bounded continuous functions into bounded continuous functions) and also the property of taking continuous functions vanishing at infinity into continuous functions vanishing at infinity.

2. Invariant measure for general Markov chains. In this section, we present a general theorem on the existence of an invariant probability measure for a Markov chain \((X_n)\) with locally compact second countable state space \(X\).

Let \(P(x, A), x \in X\) and \(A \subset X\), be a Markov transition probability function so that, for each Borel subset \(A \subset X\), \(P(\cdot, A)\) is a Borel measurable function of \(x\) and, for each \(x \in X\), \(P(x, \cdot)\) is a probability measure on the Borel subsets of \(X\).

In the rest of this section, we make the following two assumptions:

(i) If \(A\) is a compact subset of \(X\), then the function \(x \to P(x, A)\) vanishes at infinity.

(ii) \(P\) has the Feller property, that is, for each bounded continuous function \(f\) on \(X\), the function \(Pf(x) = \int f(y) P(x, dy)\) is continuous.

Let us remark that in the context of random matrices we are often in situations where the conditions (i) and (ii) hold. For instance, if \(\mu\) is a (Borel) probability measure on \(X\), the set of \(d \times d\) non-negative matrices (with usual topology) which have no zero rows or zero columns, then the Markov kernel \(P(x, A)\), defined by

\[
P(x, A) = \mu \{y \in X: y \cdot x \in A\},
\]

satisfies both (i) and (ii).

**Theorem 2.1.** Suppose that conditions (i) and (ii) hold. Then the following are equivalent:

(a) There is an invariant probability measure \(\lambda\) for \(P\), that is,

\[
\lambda(A) = \int P(x, A) \lambda(dx)
\]

for every Borel subset \(A \subset X\).

(b) There is a compact subset \(K \subset X\) such that \(\sup_n n^{-1} \sum_{k=1}^nP^k(x, K)\) does not converge to 0 as \(n \to \infty\).

(c) \(\sup_{A: \text{compact}} \limsup_n n^{-1} \sum_{k=1}^nP^k(x, A) = 1\).

For the equivalence of (b) and (c), we do not need condition (ii).

**Remark.** Note that once we have established that (a) and (b) are equivalent, (c) follows trivially from (a). However, in the absence of condition (ii), the equivalence of (b) and (c) does not seem to be trivial. Let us also mention that our proof will show that the quantity on the left in (c) with "lim sup" there replaced by "lim inf" is also either 0 or 1, in the presence of condition (i).
Proof of Theorem 2.1. (a) ⇒ (b). This is easy since if λ is an invariant probability measure for P, then

\[ \lambda(A) = \int P(x, A) \lambda(dx) = \int P^n(x, A) \lambda(dx) = \frac{1}{n} \sum_{k=1}^{n} P^k(x, A) \lambda(dx). \]

(b) ⇒ (a). Note that (b) implies that there is a compact set K such that, for some subsequence \((n_i)\), there exist elements \(x_{n_i} \in X\) such that

\[ \frac{1}{n_i} \sum_{k=1}^{n_i} P^k(x_{n_i}, K) > \delta > 0 \quad \text{for all } i \geq 1. \]

Define the probability measures \((v_{n_i})\) by

\[ v_{n_i}(\cdot) = \frac{1}{n_i} \sum_{k=1}^{n_i} P^k(x_{n_i}, \cdot). \]

Then there exists a subsequence \((n_{i_j})\) such that \((v_{n_{i_j}})\) converges (weak) to some measure \(v\), \(v(X) \leq 1\). It follows that \(v(K) \geq \delta\).

Now we observe that, for any Borel set \(A\),

\[ \left| \int P(y, A) v_{n_i}(dy) - v_{n_i}(A) \right| = \left| \frac{1}{n_i} \sum_{k=1}^{n_i} P^{k+1}(x_{n_i}, A) - \frac{1}{n_i} \sum_{k=1}^{n_i} P^k(x_{n_i}, A) \right| \to 0 \quad \text{as } n_i \to \infty. \]

Now, writing \(\lambda_{n_i}(A) = \int P(y, A) v_{n_i}(dy)\), we have: for any bounded continuous function \(f\) vanishing at infinity,

\[ \int f d\lambda_{n_i} = \int \left[ \int f(z) P(y, dz) \right] v_{n_i}(dy) \to \int \left[ \int f(z) P(y, dz) \right] v(dy), \]

since the function \(Pf(y)\) also vanishes at infinity. Thus, since

\[ \left| \int f d\lambda_{n_i} - \int f d\nu \right| \to 0 \quad \text{as } i \to \infty, \]

we have

\[ \int \left[ \int f(z) P(y, dz) \right] v(dy) = \int f dv. \]

Now, given any compact set \(B \subset X\), the set \(B\) being a \(G_\delta\)-set, there exists \(f_n\) (continuous with compact support) such that \(f_n \searrow I_B\) so that \(v(B) = \int P(y, B) v(dy)\). Thus, (b) ⇒ (a).

(b) ⇒ (c). Suppose that \(a > 0\), where

\[ (2.1) \quad \sup_{A: \text{compact}} \limsup_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} P^k(x, A) = a. \]

If possible, let \(a < 1\). Choose \(\varepsilon > 0\), \(a < c < 1\), such that \(\varepsilon < c/4\), \(0 < \varepsilon < c - a\), \(0 < a < c - \varepsilon < c < 1\), \(c(1 + c)/2 < a\). Let \(A\) be a given compact set and \(x \in X\).
There exists a positive integer $m_0$ such that, for $m \geq m_0$,

$$
(2.2) \quad \frac{1}{m} \sum_{y}^{m} P^k(y, A) < c - \varepsilon.
$$

Using the condition (i), we can now find a compact set $E$ such that

$$
(2.3) \quad P^s(y, A) < \varepsilon, \quad 1 \leq s \leq m_0,
$$

whenever $y \notin E$. Now we have from (2.1): There exists $r_0$ such that, for $r \geq r_0$,

$$
(2.4) \quad \frac{1}{m_0} \sum_{k=m_0+1}^{m_0+2} P^k(x, E) < c - \varepsilon \quad \text{and} \quad \frac{c(c+1)}{2} + \frac{3}{r_0} < a.
$$

There exists a positive integer $s_0$, $1 \leq s_0 \leq m_0$, depending on $r$ such that, for $r \geq r_0$,

$$
(2.5) \quad \frac{1}{r} \sum_{k=1}^{r} P^{m_0+s_0}(x, E) < c - \varepsilon.
$$

To see this, just write

$$
\frac{1}{m_0} \sum_{k=m_0+1}^{m_0+2} P^k(x, E) = \frac{1}{m_0} \{ P^{m_0+1}(x, E) + P^{2m_0+1}(x, E) + \ldots + P^{2^{m_0}+1}(x, E) \\
+ P^{m_0+2}(x, E) + P^{2m_0+2}(x, E) + \ldots + P^{2^{m_0}+2}(x, E) + \ldots \\
+ P^{m_0+s_0}(x, E) + P^{2m_0+s_0}(x, E) + \ldots + P^{2^{m_0}+s_0}(x, E) \}.
$$

Now, let $n \geq m_0(r_0+1)$ and write $n = m_0(r_0+1) + t$, $0 \leq t < m_0$, where $r \geq r_0$. Let $s_0$ be such that (2.5) holds for this $s_0$.

We will now use (2.2), (2.3) and (2.5). Now we have: for $A$ in (2.2),

$$
\frac{1}{n} \sum_{k=1}^{n} P^k(x, A)
$$

$$
= \frac{1}{n} \left[ P(x, A) + \ldots + P^{m_0+s_0}(x, A) \\
+ \left\{ \sum_{E \subseteq k=1}^{m_0} P^k(y, A) P^{m_0+s_0}(x, dy) + \sum_{E \subseteq k=1}^{m_0} P^k(y, A) P^{m_0+s_0}(x, dy) \right\} \\
+ \left\{ \sum_{E \subseteq k=1}^{m_0} P^k(y, A) P^{2m_0+s_0}(x, dy) + \sum_{E \subseteq k=1}^{m_0} P^k(y, A) P^{2m_0+s_0}(x, dy) \right\} + \ldots \\
+ \left\{ \sum_{E \subseteq k=1}^{m_0} P^k(y, A) P^{(r-1)m_0+s_0}(x, dy) + \sum_{E \subseteq k=1}^{m_0} P^k(y, A) P^{(r-1)m_0+s_0}(x, dy) \right\} + \ldots \\
+ \sum_{k=1}^{t+m_0} P^{m_0+k}(x, A) \right].
$$
which is a contradiction. This proves that (b) \implies (c).

Notice that (c) \implies (b) trivially. The proof of the theorem is now complete. \quad \blacksquare

3. Invariant measures for Markov chains with random transition probabilities. In this section, we generalize the model of Markov chains with random transition probabilities initially and extensively investigated by Cogburn [5] and later by Orey [16] to a Hausdorff topological state space. Cogburn points out in [5] that "the existence of a \sigma-finite invariant measure \nu \ll \mu (\mu = \kappa \times \pi a given distribution) is a more difficult question" and leaves this problem open. Orey also mentions this problem (see Problem 1.3.1 on p. 916 in [16]) as one of his open problems. We present here a sufficient condition for the existence of a non-trivial, \sigma-finite, and locally finite invariant measure for Cogburn's Markov chains. Under appropriate conditions (mentioned at the end of this section), this invariant measure will be absolutely continuous with respect to a given \beta.

We exploit here the method given by Skorokhod in [18]. Also, in the final step of the proof of the theorem in this section, the formula for the invariant measure \nu in terms of the invariant measure \lambda for \( P_\theta \) that we use here was given in [18], and thus perhaps could have been omitted at the expense of creating, however, a hard-to-follow proof. Skorokhod's concern in his paper was to establish the uniqueness of the invariant measure, whereas our concern here is simply to establish the existence of a non-trivial, \sigma-finite, and locally finite invariant measure. Our idea is simply to find an appropriate function \( g \) so that \( \{ P_\theta^n(x, \bar{\theta}) \} \) is a tight sequence for a given distribution \( \beta \), thus yielding an invariant probability measure for \( P_\theta \), and then finding an invariant measure for \( P \) through Skorokhod's formula. We feel that the condition for the existence of a \sigma-finite invariant measure that we present here is new. Final-
ly, even though we say our approach is as given in [18], Skorokhod’s approach is essentially the same as Foguel’s in [7] and Skorokhod’s formula (that we use in Step V) is essentially the same as Harris’ (see [7], equation (3.10)).

We now present our theorem. A reader, familiar with [18], can skip easily Steps I, II, and V in the proof of Theorem 3.2.

Let \((X, \mathcal{A})\) be a locally compact Hausdorff second countable topological space and \((\Theta, \mathcal{B})\) be a complete metric separable space. We will write

\[
\Omega = X \times \Theta, \quad \mathcal{F} = \mathcal{A} \times \mathcal{B},
\]

where \(\Theta = \Theta^2\) and \(\mathcal{B} = \prod_{i \in \mathbb{Z}} \mathcal{B}_i, \mathcal{B}_i = \mathcal{B}\) with product topology in \(\Theta\) and \(\Omega\).

Let \(\{P_\theta(\cdot, \cdot): \theta \in \Theta\}\) be a given family of transition probabilities on \((X, \mathcal{A})\), and \(Y_0 = (X_0, \xi_0)\) be the Markov chain with initial distribution \(\beta\) and transition probability \(P((x, \tilde{\theta}), F)\) given by

\[
P((x, \tilde{\theta}), F) = P_{\theta_0}(x, (F)_{T\tilde{\theta}}),
\]

where \((F)_{T\tilde{\theta}}\) is defined by

\[
(F)_{T\tilde{\theta}} = \{y \in X: (y, T\tilde{\theta}) \in F\},
\]

\(T\) being the left shift on \(\Theta\).

Notice that if \(X\) is discrete, then

\[
P((x, \tilde{\theta}), y \times B) = P_{\theta_0}(x, y) I_B(T\tilde{\theta}),
\]

and

\[
P((x, \tilde{\theta}), F) = \sum_{y \in X} P_{\theta_0}(x, y) I_{(F)_{T\tilde{\theta}}}(T\tilde{\theta}).
\]

This is exactly the Markov chain investigated by Cogburn [5] and later by Orey [16].

**Proposition 3.1.**

(i) \(P\{\xi_n = T^n\tilde{\theta} \mid Y_0 = (x, \tilde{\theta})\} = 1\).

(ii) \(P^n((x, \tilde{\theta}), F) = P_{\theta_0} \cdots P_{\theta_{n-1}}(x, (F)_{T^{n-1}\tilde{\theta}})\).

**Proof.** We have

\[
P\{\xi_1 = T\tilde{\theta} \mid Y_0 = (x, \tilde{\theta})\} = P((x, \tilde{\theta}), X \times \{T\tilde{\theta}\}) = P_{\theta_0}(x, X) = 1.
\]

First, we use induction to prove (ii). For \(n = 1\), by definition, we have

\[
P((x, \tilde{\theta}), F) = P_{\theta_0}(x, (F)_{T\tilde{\theta}}).
\]

Suppose that, for \(k > 1\),

\[
P^k((x, \tilde{\theta}), F) = P_{\theta_0} \cdots P_{\theta_{k-1}}(x, (F)_{T^{k-1}\tilde{\theta}});
\]

then for \(k + 1\) we have

\[
P^{k+1}((x, \tilde{\theta})), F) = \int P^k((y, \tilde{\theta}), F) P((x, \tilde{\theta}), d(y, \tilde{\theta}))
\]

\[
= \int P^k((y, T\tilde{\theta}), F) P((x, \tilde{\theta}), dy \times \{T\tilde{\theta}\})
\]

\[
= \int P_{\theta_1} \cdots P_{\theta_k}(y, (F)_{T^{k+1}\tilde{\theta}}) P_{\theta_0}(x, dy) = P_{\theta_0} \cdots P_{\theta_k}(x, (F)_{T^{k+1}\tilde{\theta}}).
\]
Therefore (ii) is true. Now, we have
\[ P \{ x_n = T^n \vartheta \mid Y_0 = (x, \vartheta) \} = P \{ Y_n \in X \times \{ T^n \vartheta \} \mid Y_0 = (x, \vartheta) \} = P^n((x, \vartheta), X \times \{ T^n \vartheta \}) = P_{\theta_0} \cdots P_{\theta_{n-1}}(x, X) = 1, \]
which is (i). \( \blacksquare \)

Let \( \beta = m \times \pi \), where \( m \) is a probability measure on \( X \) and \( \pi \) is a probability measure on \( \vartheta \). Let \( C_0(\vartheta) \) be the class of all bounded real continuous functions on \( \vartheta \) that vanish at infinity and \( g \in C_0(\vartheta) \), \( 0 \leq g \leq 1 \). In what follows, we define the function \( P_{g}(\cdot, \cdot) \); as in [18], on \( (\vartheta, \mathcal{F}) \) as follows:

(3.1) \[ P_{g}((x, \vartheta), F) = E_{(x, \vartheta)}[g(Y_1)I_F(Y_1) + \sum_{n=2}^{\infty} (1-g(Y_1)) \cdots (1-g(Y_{n-1}))g(Y_n)I_F(Y_n)]. \]

Note that for any sequence \( a_i \), \( 0 \leq a_i < 1 \),
\[ a_1 + (1-a_1)a_2 + (1-a_1)(1-a_2)a_3 + \ldots = 1 - \prod_{i=1}^{\infty} (1-a_i) = 1 \quad \text{iff} \quad \sum_{i=1}^{\infty} a_i = \infty. \]
Thus, it is clear that \( P_{g} \) is a transition probability on \( (\vartheta, \mathcal{F}) \) if

(3.2) \[ P_{g}((x, \vartheta), \{ \sum_{n=1}^{\infty} g(Y_n) = \infty \}) = 1 \quad \text{for all } (x, \vartheta) \in \vartheta. \]

We can now state our results.

**Theorem 3.2.** Suppose that there exists a strictly positive function \( g \in C_0(\vartheta) \), \( 0 \leq g \leq 1 \), such that

(i) \( P_{(x, \vartheta)}(\{ \sum_{n=1}^{\infty} g(Y_n) = \infty \}) = 1 \) for all \( (x, \vartheta) \);
(ii) for \( \beta \) almost \( (x, \vartheta) \), there exists a Borel subset \( B((x, \vartheta)) \) of \( \vartheta \), containing \( (x, \vartheta) \), such that
\[ P_{(x, \vartheta)}(Y_n \in B((x, \vartheta))) \text{ i.o.)} = 0; \]
(iii) \( \lim_{n \to \infty} \frac{1}{n} \prod_{i=1}^{n} (1-g(y_i, T^i \vartheta)) P_{\theta_0}(x, dy_1) \cdots P_{\theta_{n-1}}(y_{n-1}, dy_n) = 0 \)
uniformly for \( (x', \vartheta') \) in each \( B((x, \vartheta)) \);
(iv) for any bounded continuous function \( f \) on \( \vartheta \),
\[ Pf(x, \vartheta) = \int f(y, \vartheta) P((x, \vartheta), dy, \vartheta') \]
is continuous.

Then there exists a non-trivial \( \sigma \)-finite and locally finite (that is, finite on compact subsets) invariant measure \( \mu \) on \( (\vartheta, \mathcal{F}) \) for the transition kernel \( P \).

**Remark.** Let us remark that in [7] Foguel proved the above theorem for a Markov process (with non-random transition probabilities) under the Feller condition (our assumption (ii) above) and an assumption (assumption (2.1))
in [7]) which implies our conditions (i), (ii) and (iii) above for some bounded continuous function with compact support with \( B((x, \theta)) = \Omega \) for each \((x, \theta)\). [His function \( \beta \) in his Section 3 (see his limit statement (3.3)) happens to be this continuous function with compact support.] Note that our Theorem 3.2 under assumptions (i)–(iv) for some continuous \( g \) with compact support (when \( \Theta \) is a singleton set) follows easily if we observe that the transition probability \( P_g \), in this case, acts on a compact space and, as such, has an invariant probability measure, which, in turn, gives a \( \sigma \)-finite invariant measure for \( P \).

Proof of Theorem 3.2. Notice that condition (i) implies that \( P_g \) is a transition probability on \( \Omega \). Now we separate the proof into several steps.

Step I. Let \( (\overline{Y}_n) \) be a copy of the Markov Chain \( (Y_n) \) with the same transition probability \( P \). Then we have

\[
E_{(x, \theta)} \left[ f(Y_1, \ldots, Y_n) E_{Y_n}(h(\overline{Y}_1, \ldots, \overline{Y}_m)) \right]
= E_{(x, \theta)} \left[ f(Y_1, \ldots, Y_n) h(Y_{n+1}, \ldots, Y_{n+m}) \right],
\]

where \( f \) and \( h \) are bounded Borel measurable functions on \( \Omega \). To prove this notice that

\[
E_{(x, \theta)} \left[ f(Y_1, \ldots, Y_n) E_{Y_n}(h(\overline{Y}_1, \ldots, \overline{Y}_m)) \right]
= \int \cdots \int \left[ f((y_1, T\theta), \ldots, (y_n, T^n \theta)) \right]
\times E_{(y_n, T^n \theta)}(h(\overline{Y}_1, \ldots, \overline{Y}_m)) P_{\theta_0}(x, dy_1) \cdots P_{\theta_{n-1}}(y_{n-1}, dy_n)
\times P_{\theta_n}(y_n, dy_{n+1}) \cdots P_{\theta_{n+m-1}}(y_{n+m-1}, dy_{n+m}) \cdots P_{\theta_{n+m}}(y_{n+m}, dy_{n+m})
\times P_{\theta_n}(y_{n-1}, dy_n) = E_{(x, \theta)} \left[ f(Y_1, \ldots, Y_n) h(Y_{n+1}, \ldots, Y_{n+m}) \right].
\]

This establishes Step I.

Step II. In this step, we show that for \( m \geq 1 \)

\[
P_{\theta}^m ((x, \theta), F) = E_{(x, \theta)} \sum_{n_1, \ldots, n_m = 1}^{\infty} \left[ \prod_{k=1}^{n_1-1} (1 - g(Y_k)) g(Y_{n_1}) \ldots \prod_{k=1}^{n_m-1} (1 - g(Y_{n_1+\ldots+n_{m-1}+k})) g(Y_{n_1+\ldots+n_m}) I_F(Y_{n_1+\ldots+n_m}) \right].
\]

To prove this, let \( (\overline{Y}_n) \) be a copy of the Markov chain \( (Y_n) \) so that \( \overline{Y}_n \) has also \( P \) as its transition probability. Now we use induction to establish the expression for \( P_{\theta}^m \). Using induction, then we have
Invariant measures and Markov chains

\[ P^m_{I_g}(x, \theta), F) = P^m_{I_g}(P_g I_F)(x, \theta) \]

\[ = E_{(x, \theta)} \sum_{n_1, \ldots, n_m}^{n_1-1} \prod_{k=1}^{n_m-1} (1-g(Y_k)) g(Y_n) \ldots \]

\[ \ldots (1-g(Y_{n_1+\ldots+n_m+k})) g(Y_{n_1+\ldots+n_m}) P_g I_F(Y_{n_1+\ldots+n_m}) \]

\[ = E_{(x, \theta)} \sum_{n_1, \ldots, n_m}^{n_1-1} \prod_{k=1}^{n_m-1} (1-g(Y_k)) g(Y_n) \ldots \]

\[ \prod_{k=1}^{n_m-1} (1-g(Y_{n_1+\ldots+n_m+k})) g(Y_{n_1+\ldots+n_m}) \]

\[ \times \sum_{n_m+1}^{n_m+1-1} \prod_{k=1}^{n_m+1-1} (1-g(Y_k)) g(Y_{n_m+1}) I_F(Y_{n_m+1}) \]

which, by Step I, is equal to

\[ E_{(x, \theta)} \sum_{n_1, \ldots, n_m+1}^{n_1-1} \prod_{k=1}^{n_m+1-1} (1-g(Y_k)) g(Y_n) \ldots \]

\[ \ldots (1-g(Y_{n_1+\ldots+n_m+k})) g(Y_{n_1+\ldots+n_m+1}) I_F(Y_{n_1+\ldots+n_m+1}) \]

This establishes Step II.

Step III. In this step, we establish that the sequence of probability measures

\[ \{ \int P^m_{I_g}(x, \theta), F) \beta(d(x, \theta)): m \geq 1 \} \]

is tight, that is, for \( \varepsilon > 0 \) there is a compact set \( F \subset \Omega \) such that for all \( m \geq 1 \)

\[ \int P^m_{I_g}(x, \theta), F) \beta(d(x, \theta)) > 1 - \varepsilon. \]

To prove this, notice that for a sequence \( \{a_i\}, 0 \leq a_i \leq 1, \)

\[ \prod_{i=1}^{n} (1-a_i) = \sum_{j=1}^{\infty} (1-a_i) \cdot a_{i+1}. \]

Therefore, with probability one, by assumption (i),

\[ 3.3 \quad g(Y_1) + \sum_{n=2}^{\infty} \prod_{k=1}^{n-1} (1-g(Y_k)) g(Y_n) = 1 - \prod_{k=1}^{\infty} (1-g(Y_k)) = 1. \]

Let \( \varepsilon > 0 \). For each positive integer \( N \), define the set \( \Omega_N(\varepsilon) \) by

\[ \Omega_N(\varepsilon) = \left\{(x, \theta) \mid P_{(x, \theta)}(y_n \in B((x, \theta))) \text{ for all } n \geq N > 1 - \varepsilon \right\}, \]

and \[ \int \prod_{i=1}^{N} (1-g(y_i, T^i \theta)) P_{\theta_0}(x', d y_1) \ldots P_{\theta_{N-1}}(y_{N-1}, d y_N) < \frac{\varepsilon}{2} \]

for every \( (x', \theta) \in B((x, \theta)) \).
It is clear that $\Omega_N(\varepsilon) \subset \Omega_{N+1}(\varepsilon)$ and that

$$\lim_{N \to \infty} \beta(\Omega_N(\varepsilon)) = 1$$

by condition (ii), where $\overline{\Omega_N(\varepsilon)}$ is the closure of $\Omega_N(\varepsilon)$. This means that there exists $N_0$ such that

$$\beta(\overline{\Omega_{N_0}(\varepsilon)}) > 1 - \varepsilon.$$ 

Now, by condition (iii), for $(x, \bar{\theta}) \in \Omega_N(\varepsilon)$, whenever $(x', \bar{\theta}') \in B((x, \bar{\theta}))$, we have

$$E_{(x', \bar{\theta})}[\sum_{n=1}^{\infty} \prod_{k=1}^{n-1} (1 - g(Y_k)) g(Y_n)] < \frac{\varepsilon}{2}.$$ 

Choose $\delta < \varepsilon/[2(N_0 + 1)]$; since $g \in C_0(\Omega)$, there is a compact subset $F \subset \Omega$ such that $g(x, \bar{\theta}) < \delta$ for $(x, \bar{\theta}) \notin F$. Note that for $(x, \bar{\theta}) \in \Omega_{N_0}(\varepsilon)$ and all $(x', \bar{\theta}') \in B((x, \bar{\theta}))$, we have

$$E_{(x', \bar{\theta})}[\sum_{n=1}^{\infty} \prod_{k=1}^{n-1} (1 - g(Y_k)) g(Y_n)] < \varepsilon.$$

Let $m \geq N_0 + 1$. We have, for $(x, \bar{\theta}) \in \Omega_{N_0}(\varepsilon)$,

$$P^m_g((x, \bar{\theta}) F^c) = E_{(x, \bar{\theta})} \sum_{n_1, \ldots, n_m-1} \prod_{k=1}^{n_m-1} (1 - g(Y_{n_1} + \ldots + n_{m-2} + k)) g(Y_{n_1} + \ldots + n_{m-1})$$

$$\times E_{Y_{n_1} + \ldots + n_{m-1}} \prod_{k=1}^{n_m-1} (1 - g(Y_{n_k})) g(Y_{n_m})$$

$$\leq E_{(x, \bar{\theta})} \sum_{n_1, \ldots, n_m-1} \prod_{k=1}^{n_m-1} (1 - g(Y_k)) g(Y_{n_1})$$

$$\times [I_{B((x, \bar{\theta}))}(Y_{n_1} + \ldots + n_{m-1}) + I_{B^c((x, \bar{\theta}))}(Y_{n_1} + \ldots + n_{m-1})]$$

$$\times E_{Y_{n_1} + \ldots + n_{m-1}} \prod_{k=1}^{n_m-1} (1 - g(Y_{n_k})) g(Y_{n_m})$$

$$\leq \varepsilon + P^m_{g^{-1}} I_{B^c((x, \bar{\theta}))}((x, \bar{\theta})) \quad \text{(by (3.4))}.$$

Let

$$D = \{\omega: Y_\omega((x, \bar{\theta})) \in B^c((x, \bar{\theta})) \text{ for some } n \geq N_0\}.$$
By (3.3), we obtain
\[ P_\alpha^{-1} I_{B^{\epsilon}(x, \bar{\theta})}(x, \bar{\theta}) \leq E_{(x, \bar{\theta})}(I_D) \]
since \((x, \bar{\theta}) \in \Omega\). Therefore,
\[ P_\alpha^m ((x, \bar{\theta}), F^c) \leq \epsilon + P_{(x, \bar{\theta})}(D) \leq 2\epsilon. \]
Now we show that, for \(m \geq 1\) and any \((x, \bar{\theta}) \in \Omega\),
\[ P_\alpha^m ((x, \bar{\theta}), F) \geq 1 - 3\epsilon. \]
To this end, we need the Feller property for the transition kernel \(P_\alpha\) that we will establish in Step IV. Notice that the compact subset \(F\) above has the property that \(F\) is a \(G_\alpha\)-set and that, for any \((x, \bar{\theta}) \in \Omega\),
\[ P_\alpha^m ((x, \bar{\theta}), F) \geq 1 - 2\epsilon. \]
Let us suppose that there is some \((x_0, \bar{\theta}_0) \in \Omega\) such that, for a positive integer \(m > 1\),
\[ P_\alpha^m ((x_0, \bar{\theta}_0), F) < 1 - 3\epsilon. \]
This means that there is an open set \(V, F \subset V\) such that
\[ P_\alpha^m ((x_0, \bar{\theta}_0), V) < 1 - 3\epsilon. \]
Let \(f\) be a continuous function on \(\Omega\) such that \(0 \leq f \leq 1, f = 1\) on \(F\) and \(0\) on \(V^c\). Then we have
\[ \int f(y, \bar{\theta}) P_\alpha^m ((x_0, \bar{\theta}_0), d(y, \bar{\theta})) < 1 - 3\epsilon. \]
Since the transition kernel \(P_\alpha^m\) has also the Feller property, this means that there must exist some \((x, \bar{\theta}) \in \Omega\) such that
\[ \int f(y, \bar{\theta}) P_\alpha^m ((x, \bar{\theta}), d(y, \bar{\theta})) < 1 - 3\epsilon, \]
which implies that
\[ P_\alpha^m ((x, \bar{\theta}), F) < 1 - 3\epsilon. \]
This is a contradiction.

The rest of the proof of Step III now follows easily.

Step IV. In this step, we show that there exists an invariant probability measure for \(P_\alpha\).

First, notice that for any bounded continuous function \(f\) on \(\Omega\), \(f \geq 0,
\[ P_\alpha f ((x, \bar{\theta})) = E_{(x, \bar{\theta})}[ \sum_{n=1}^{\infty} \prod_{k=1}^{n-1} (1-g(Y_k))g(Y_n)f(Y_n) ], \]
which is an infinite sum of non-negative bounded continuous functions (because of condition (iv)) and, as such, is lower semicontinuous. Also, notice that if
f above is bounded by $M$, then
\[ P_{\theta} [M - f] (x, \vartheta) = M - P_{\theta} f ((x, \vartheta)), \]
which is also lower semicontinuous so that $P_{\theta} f$ is also upper semicontinuous. It follows that, for any bounded continuous $f$, $P_{\theta} f$ is also continuous. Now we define, for any probability measure $\lambda$ on $\Omega$, the probability measure $P_{\theta} \lambda$ by
\[ P_{\theta} \lambda (B) = \int P_{\theta} ((x, \vartheta), B) \lambda (d (x, \vartheta)). \]
Notice that if $\lambda_n \to \lambda$ weakly in $P(\Omega)$, then for any bounded continuous function $f$ on $\Omega$ we have
\[ \int fd (P_{\theta} \lambda_n) = \int \left[ \int f ((y, \vartheta)) P_{\theta} ((x, \vartheta), d (y, \vartheta)) \right] \lambda_n (d (x, \vartheta)) \]
\[ \to \int \left[ \int f ((y, \vartheta)) P_{\theta} ((x, \vartheta), d (y, \vartheta)) \right] \lambda (d (x, \vartheta)) = \int fd (P_{\theta} \lambda), \]
so that $P_{\theta} \lambda_n \to P_{\theta} \lambda$ weakly as $n \to \infty$.

Let us now write, for any Borel subset $F \subset \Omega$,
\[ v_n (F) = \frac{1}{n} \sum_{k=1}^{n} \int P_{\theta}^k ((x, \vartheta), F) \beta (d (x, \vartheta)). \]
Notice that
\[ \int fdv_{n_k} = \frac{1}{n_k} \sum_{k=1}^{n_k} \int f (y, \vartheta) P_{\theta}^k ((x, \vartheta), d (y, \vartheta)) \beta (d (x, \vartheta)) \to \int fdv_0; \]
also,
\[ \int fd (P_{\theta} v_{n_k}) = \frac{1}{n_k} \sum_{k=1}^{n_k} \int f (y, \vartheta) P_{\theta}^{i+1} ((x, \vartheta), d (y, \vartheta)) \beta (d (x, \vartheta)) \to \int fd (P_{\theta} v_0). \]
It is now clear that, for any bounded continuous function $f$ on $\Omega$,
\[ \int fdv_0 = \int fd (P_{\theta} v_0). \]
Since $\Omega$ is a metric space, given any compact subset $F \subset \Omega$, $F$ is also a $G_{\delta}$-set so that there is a sequence of continuous functions uniformly bounded by 1, converging pointwise to $I_F$. This means that, for every compact set $F \in \Omega$,
\[ v_0 (F) = P_{\theta} v_0 (F). \]
By regularity of the measures, it follows that $v_0 = P_{\theta} v_0$. This completes the proof of Step IV.

Step V. Let $\lambda$ be the invariant probability measure for $P_{\theta}$. Define $v$ on $(\Omega, \mathcal{F})$ by
\[ v (F) = \int E_{(x, \vartheta)} [I_F (Y_1) + \sum_{n=2}^{\infty} \prod_{k=1}^{n-1} (1 - g (Y_k)) I_F (Y_n)] \lambda (d (x, \vartheta)). \]
The formula for this $v$ is actually a formula that probably was first considered by Harris (see [7]). We have taken it from [18]. Notice that $\Omega$ is $\sigma$-compact and
\[ v (\Omega) \geq \int P_{\theta} ((x, \vartheta), \Omega) \lambda (d (x, \vartheta)) = 1; \]
also, if $F$ is a compact subset of $\Omega$, there is a positive number $M$ such that

$$I_F(Y_n) \leq Mg(Y_n),$$

since $g$ is strictly positive and continuous on $\Omega$, so that

$$v(F) \leq M \int P_g(\{(x, \emptyset), \Omega\} \lambda (d(x, \emptyset)) \leq M.$$ 

This means that $v$ is a $\sigma$-finite, non-trivial and locally finite measure on $\Omega$.

Let us now establish that $v$ is invariant for $P$. (Our proof here is quite different from that given in Lemma 6 of [18].) Observe that

$$v(F) = \int E_{(x, \emptyset)} I_F(Y_n) d\lambda + \int E_{(x, \emptyset)} \left( \sum_{n=2}^{\infty} \prod_{k=1}^{n-2} (1-g(Y_k)) I_F(Y_n) \right) d\lambda$$

$$= \int P((x, \emptyset), F) d\lambda + \int E_{(x, \emptyset)} \left[ \sum_{n=2}^{\infty} \prod_{k=1}^{n-2} (1-g(Y_k)) I_F(Y_n) \right] d\lambda - \int \prod_{k=1}^{n-2} (1-g(Y_k)) g(Y_{n-1}) I_F(Y_n) d\lambda$$

$$= \int P((x, \emptyset), F) d\lambda + \int E_{(x, \emptyset)} \left[ \sum_{n=1}^{\infty} \prod_{k=1}^{n-1} (1-g(Y_k)) I_F(Y_n) \right] d\lambda - \int E_{(x, \emptyset)} \left[ \sum_{n=1}^{\infty} \prod_{k=1}^{n-1} (1-g(Y_k)) g(Y_n) I_F(Y_n) \right] d\lambda$$

This completes Step V and the proof of the theorem. \hfill \Box
Let us remark that we would often like to have the invariant measure \( \nu \) (for \( P \)) to be absolutely continuous with respect to the initial distribution \( \mu \) for the Markov chain \( (Y_n) \). This, of course, will necessitate imposing some conditions on the transition kernels \( P_\theta(x, \cdot) \), \( \theta \in \Theta \). One easy condition is, of course, that

\[
\beta(F) = 0 \implies P_\theta(x, F_{TF}) = 0 \quad \text{for all } x \in X \text{ and } \theta \in \Theta,
\]

where \( F_{TF} = \{ y \in X \mid (y, T\theta) \in F \} \). If this condition holds, it is obvious that the invariant measure \( \nu \) obtained in Step V is absolutely continuous with respect to \( \beta \). In the remainder of this section, we present two results relevant to the problem of the existence of a \( \sigma \)-finite invariant measure for \( P \) (in the context of Theorems A, C and D in \cite{g}, pp. 65-70). These results, even though likely known to experts, are recorded here (as we have not found them in the literature in the form we present them here).

Let us first note the following definition. A set \( F \in \mathcal{F} \) is closed if \( \beta(F) > 0 \) and \( I_F \leq PI_F \) \( \beta \) a.s., where \( \beta \) is a given measure (not necessarily finite) on \( (\Omega, \mathcal{F}) \) and

\[
PI_F(x, \theta) = P((x, \theta), F).
\]

The conservative set \( C \) is defined by

\[
C = \{(x, \theta) \mid \sum_{k=0}^{\infty} uP^k(x, \theta) = \infty\},
\]

where \( u \in L_1(\beta) \), \( u \) strictly positive and \( \langle uP, f \rangle = \langle u, Pf \rangle \) for \( f \in L_\infty(\beta) \). The set \( C \) is independent of \( u \), see \cite{g}. In what follows, we assume, like in \cite{g}, p. 2, for each \( (x, \theta) \in \Omega \), \( P((x, \theta), \cdot) \ll \beta \).

**Proposition 3.3.** Suppose that \( \Omega = C \). Then the following are equivalent:

1. \( P \) is \( \beta \)-irreducible, that is, for any \( F \) with \( \beta(F) > 0 \) \( \beta \)-a.s.,

\[
P_{(x, \theta)}(Y_n \in F \text{ for some } n \geq 1) > 0.
\]

2. \( \mathcal{F}_1 = \{ F \in \mathcal{F} \mid I_F = PI_F \ \beta \text{ a.s.} \} \) is trivial.

3. \( P \) is \( \beta \)-recurrent, that is, for any \( F \) with \( \beta(F) > 0 \) \( \beta \)-a.s.,

\[
P_{(x, \theta)}(Y_n \in F \text{ i.o.}) = 1.
\]

4. \( \Omega \) is minimal closed, that is, if \( F \subset \Omega \) and \( \beta(\Omega - F) > 0 \), then \( F \) is not closed.

5. If \( Pf = f, f \in L_\infty(\beta), \) then \( f \) is a constant \( \beta \)-a.s.

6. If \( f \in L_+^\infty(\beta), f \neq 0, \) then \( \sum_{k=1}^{\infty} P^k f = \infty \) \( \beta \)-a.s.

We omit the proof which is not difficult. Note that when \( \Omega = C \) and one of the conditions in Proposition 3.2 holds, then there exists a \( \sigma \)-finite \( P \)-invariant measure on \( \Omega \) equivalent to \( \beta \) provided the condition of Theorem D (on p. 70 of \cite{g}) holds. This then leads to the well-known fact that if \( P \) satisfies the Harris
condition, then there exists a $\sigma$-finite $P$-invariant measure equivalent to $\beta$ (see Theorem E on p. 73 in [8]). We close this section with the following result for the case when $X$ is countable and $\beta = \kappa \times \pi$ (like in [5], $\kappa$ being the counting measure on $X$), $\pi$ being a given stationary probability measure on $(\mathcal{E}, \mathcal{B})$.

**Proposition 3.4.** $P$ is $\beta$-irreducible iff for any $y \in X$ and $B \in \mathcal{B}$, $\pi(B) > 0$,

$$\sum_{k=1}^{\infty} P(\theta_0 \ldots \theta_{k-1}; x, y) I_B(T^k \overline{\theta}) > 0 \quad \text{for } (\beta-) \text{ almost all } (x, \overline{\theta}),$$

where $P(\theta_0 \ldots \theta_{k-1}; x, y)$ means the entry on the $x$-th row and the $y$-th column of the product of the $k$ stochastic matrices $P_{\theta_0}, P_{\theta_1}, \ldots, P_{\theta_{k-1}}$.

4. **Markov random walks, attractors, and invariant measures.** In this section, we consider a special class of Markov chains with random transition probabilities involving transformations from $\mathbb{R}^d$ into $\mathbb{R}^d$ of the form

$$T(x) = A(x) + x_0, \quad x \in \mathbb{R}^d,$$

where $x_0$ is a fixed element in $\mathbb{R}^d$ and $A$ is linear. The reason for including this section is to provide a concrete example of Cogburn chains and at the same time to show connections between the invariant measures, left and right attractors for the Markov random walks that we get from these chains. Similar results as presented here appeared earlier in [3]; but the discussion involved only a finite number of transformations and the proofs there are not valid for an arbitrary family of such transformations.

Let $\Theta$ be a compact metric space and $\{W_\theta; \theta \in \Theta\}$ be a family of transformations of the form (4.1) such that

$$W_\theta(x) = A_\theta(x) + B_\theta.$$

Then, if $\theta_i \in \Theta$, $0 \leq i \leq n$, for $x \in \mathbb{R}^d$ we have

$$\begin{cases} 
W_{\theta_n} W_{\theta_{n-1}} \ldots W_{\theta_0}(x) = A_{\theta_n} A_{\theta_{n-1}} \ldots A_{\theta_0}(x) \\
+ A_{\theta_n} A_{\theta_1} B_{\theta_0} + A_{\theta_n} A_{\theta_2} B_{\theta_1} + \ldots + A_{\theta_n} B_{\theta_{n-1}} + B_{\theta_n}, \\
W_{\theta_0} W_{\theta_1} \ldots W_{\theta_n}(x) = A_{\theta_0} A_{\theta_1} \ldots A_{\theta_n}(x) \\
+ A_{\theta_0} A_{\theta_1} \ldots A_{\theta_{n-1}} B_{\theta_n} + \ldots + A_{\theta_0} B_{\theta_1} + B_{\theta_0}. 
\end{cases}$$

Let $(\xi_n)$ be an indecomposable Markov chain with state space $\Theta$ and transition function $P$ such that the (initial) distribution $\pi_0$ of $\xi_0$ is invariant with respect to $P$. Then the stationary process gotten by taking $\pi_0$ as the distribution of $\xi_0$ is ergodic (see [4], Theorem 7.16, p. 136).

Let $\overline{\Theta} = \Theta^\mathbb{Z}^+$, $\mathcal{B} = \bigcap_{i \in \mathbb{Z}^+} \mathcal{B}_i$, $\mathcal{B}_i \equiv \mathcal{B}$, where $\mathcal{B}$ is the class of Borel subsets of $\Theta$. Let $\pi$ be the probability measure on $(\overline{\Theta}, \mathcal{B})$ induced by the stationary ergodic process $(\xi_n)$, so that it is stationary and ergodic with respect to the
left shift $T$ on $\mathcal{O}$ given by

$$T\mathcal{O} = (\theta_1, \theta_2, \ldots), \quad \text{where } \mathcal{O} = (\theta_0, \theta_1, \theta_2, \ldots).$$

Let $\pi^*$ be the unique probability measure on $(\mathcal{O}, \mathcal{F})$ such that, for any finite-dimensional rectangle of the form $A^{(n)} = \{\mathcal{O}: \theta_i \in A_i, 0 \leq i \leq n\}$,

$$\pi^* (A^{(n)}) = \pi \{\mathcal{O}: \theta_i \in A_{n-i}, 0 \leq i \leq n\}.$$

Let us assume that

(4.4) $\pi^*$ is stationary and ergodic.

[Let us remark that if $\Theta$ is finite and $P$ is a strongly ergodic stochastic matrix with a stationary distribution $\pi_0$ such that $\pi_0(\theta) > 0$ for each $\theta \in \Theta$, then $\pi$ induced by $\pi_0$ and $P$, as well as the corresponding $\pi^*$, as defined above, is stationary and ergodic.]

Let us write $E \equiv \mathbb{R}^d$ and $\Omega \equiv E \times \mathcal{O}$. For $(x, \mathcal{O}) \in \Omega$, let

(4.5) $\Phi(x, \mathcal{O}) = (W_{\theta_0}(x), T\mathcal{O}).$

Let $\mathcal{A}$ be the Borel subsets of $E$ and $\mathcal{F} = \mathcal{A} \times \mathcal{O}$, and for $F \in \mathcal{F}$ let

$$K((x, \mathcal{O}), F) = I_F(\Phi(x, \mathcal{O})).$$

Consider the Markov chain $(Y_n)$ on $\Omega$ with transition function $K$ so that $Y_n = \Phi^n(Y_0)$. Let us now make the following assumptions:

(i) For some $M_1, M_2 > 0$ and all $\theta \in \Theta$,

$$\|A_\theta\| \leq M_1, \quad \|B_\theta\| \leq M_2.$$

(ii) For $n \geq 0, x \in E$, and $r > 0$, the sets

(4.6) \{\mathcal{O} \in \mathcal{O}: \|W_{\theta_n} \cdots W_{\theta_0}(x)\| \leq r\} \quad \text{and} \quad \{\mathcal{O} \in \mathcal{O}: \|W_{\theta_n} \cdots W_{\theta_0}(x)\| \leq r\}

are in $\mathcal{F}$.

(iii) There are real numbers $b < a < 0$ such that

$$b < \int \log \|A_\theta\| \, \pi_0(d\theta) \leq a < 0.$$

Taking $f(\mathcal{O}) = \log \|A_\theta\|$ and using the ergodic theorem (and (4.6)), we infer that

(4.7) the limit $\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} \log \|A_{\theta_k}\|$ exists and is not greater than $a < 0$

for all $\mathcal{O} \in D \subset \mathcal{O}$, where $\pi(D) = \pi^*(D) = 1$; moreover, for $\mathcal{O} \in D$, $x \in E$,

(4.8) \sup \{\|W_{\theta_n} W_{\theta_1} \cdots W_{\theta_0}(x)\|: n \geq 0\} < \infty.

Thus, it follows from (4.3) and (4.8) that, for $\mathcal{O} \in D$ and $x \in E$,

(4.9) $Z(\mathcal{O}) = \lim_{n \to \infty} W_{\theta_0} W_{\theta_1} \cdots W_{\theta_n}(x)$ exists,

and this limit is independent of $x$ because of (4.3) and (4.7).
Now suppose that $Y_0$ has distribution $\beta$ such that
\[ \beta(E \times \bar{B}) = \pi(\bar{B}), \quad \bar{B} \subset \bar{\Omega}. \]
Given $\varepsilon > 0$, let $A$ be a compact subset of $E$ and $r > 0$ such that
\[ \beta(A \times \bar{\Omega}) > 1 - \varepsilon, \]
and $\pi^* \{ \bar{\theta} \in \bar{\Omega} : \| Z(\bar{\theta}) \| \leq r \} > 1 - \varepsilon$.
Given $x_0 \in K$, there exists $N > 0$ such that, for $n \geq N$,
\[ \pi^* \{ \bar{\theta} \in \bar{\Omega} : \| W_{\theta_n} W_{\theta_{n-1}} \ldots W_{\theta_0}(x_0) \| \leq 2r \} > 1 - \varepsilon. \]
It follows from the definition of $\pi^*$ that
\[ \pi \{ \bar{\theta} \in \bar{\Omega} : \| W_{\theta_n} W_{\theta_{n-1}} \ldots W_{\theta_0}(x_0) \| \leq 2r \} > 1 - \varepsilon. \]
For $n$ sufficiently large we obtain
\[ \beta \Phi^{-n}(\{ y \in E : \| y \| \leq 3r \} \times \bar{\Omega}) > 1 - 3\varepsilon. \]
This means that the sequence $\beta_n$, given by
\[ \beta_n = \frac{1}{n} \sum_{k=0}^{n-1} \beta \Phi^{-k}, \]
is tight, and converges weakly to some probability measure $\beta_0 \in P(\Omega)$ such that $\beta_0 = \beta_0 \cdot \Phi^{-1}$ and
\[ \beta_0(E \times \bar{B}) = \pi(\bar{B}), \quad \bar{B} \subset \bar{\Omega}. \]
(4.10)

Now notice that if $f_0$ is any continuous function on $E$ with compact support and $f(x, \bar{\theta}) = f_0(x)$ for $(x, \bar{\theta}) \in \Omega$, then
\[ \int f \, d\beta_0(x, \bar{\theta}) = \int f \, d\beta_0 \Phi^{-n}(x, \bar{\theta}) \]
\[ = \lim_{n \to \infty} \int f(W_{\theta_n} W_{\theta_{n-1}} \ldots W_{\theta_0}(x_0)) \, d\beta_0(x, \bar{\theta}) \]
\[ = \lim_{n \to \infty} \int f(W_{\theta_n} W_{\theta_{n-1}} \ldots W_{\theta_0}(x_0)) \pi(\, d\bar{\theta}) \]
\[ = \int f(Z(\bar{\theta})) \pi^*(\, d\bar{\theta}). \]
It follows that, for $A \subset E$,
\[ \beta_0(A \times \bar{\Omega}) = \pi^* \{ \bar{\theta} \in \bar{\Omega} : Z(\bar{\theta}) \in A \}. \]
(4.11)

Now, if we give $Y_0$ the initial distribution $\beta_0$, then the process $Y_n = \Phi^n(Y_0)$ becomes stationary. Let $f$ be any continuous function with compact support
on $E$. Using the ergodic theorem on $(\Omega, \mathscr{F}, \beta_0)$ and $\Phi$, we infer that the limit

$$
\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(W_{\theta_0} W_{\theta_{n-1}} \ldots W_{\theta_0}(x)) = g_f(x, \bar{\theta})
$$

exists for $\beta_0$-almost all $(x, \bar{\theta})$. It is easy to see that $g_f(x, \bar{\theta})$ is independent of $x$, and then since $\pi$ is ergodic, it is a constant, which is easily seen to be

$$
\int f(Z(\bar{\theta})) \pi^*(d\bar{\theta}).
$$

Since we have assumed that $\pi$ and $\pi^*$ are both stationary and ergodic with respect to the left shift on $\bar{\Theta}$, it is known (see [4], Corollary 6.24, p. 116) that either $\pi = \pi^*$ or $\pi$ is orthogonal to $\pi^*$, that is, there is a shift invariant subset $\bar{B} \subset \bar{\Theta}$ such that $\pi(\bar{B}) = 1$ and $\pi^*(\bar{B}) = 0$. In the case when $\Theta$ is finite and the transition matrix $P$ is symmetric or has identical rows, it is clear that $\pi = \pi^*$.

Let us now define the sets $A_\pi$ and $A_\pi(\bar{\theta})$ as follows:

1. $A_\pi \subset E$ and $A_\pi$ is the support of the probability measure $\beta_0(\cdot \times \bar{\theta})$, where $\beta_0$ is the unique $\Phi$-invariant measure on $(\Omega, \mathscr{F})$, satisfying (4.10). We will call $A_\pi$ the right attractor.

2. $A_\pi(\bar{\theta}) \subset E$ and it is given by $\{ y \in E \mid \text{given any open subset } N(y) \text{ containing } y, \text{ for any } x \in E \text{ there are infinitely many } n \text{ such that } W_{\theta_0} W_{\theta_{n-1}} \ldots W_{\theta_0}(x) \in N(y) \}$. We will call $A_\pi(\bar{\theta})$ the left attractor.

Note that when $\pi = \pi^*$, it follows from (4.11) that (since $E$ is second countable) there is a set $D_0 \subset \bar{\Theta}$ with $\pi(D_0) = 1$ such that, for each $\bar{\theta} \in D_0$, $A_\pi \subset A_\pi(\bar{\theta})$. Also, when $\pi = \pi^*$, if $y \notin A_\pi$, then there is an open set $N(y)$ containing $y$ such that

$$
\beta_0(N(y) \times \bar{\theta}) = \beta_0 \Phi^{-n}(N(y) \times \bar{\theta}) = 0 \quad \text{for every } n
$$

so that, for any $x \in E$,

$$
\pi \{ \bar{\theta} \mid W_{\theta_0} W_{\theta_{n-1}} \ldots W_{\theta_0}(x) \in N(y) \text{ i.o.} \} = 0,
$$

and, consequently, $y \notin A_\pi(\bar{\theta})$ $\pi$-a.e. Thus, we have proved the following:

**Theorem 4.1.** If $\pi$ and $\pi^*$ are both stationary and ergodic on $\bar{\Theta}$ with respect to the left shift $T$, then under assumption (4.6) there is a unique $\Phi$-invariant probability measure $\beta_0$ on $\Omega$, whose $\bar{\Theta}$-marginal is $\pi$. Furthermore, when $\pi = \pi^*$, there is a set $D_0 \subset \bar{\Theta}$ with $\pi(D_0) = 1$ such that, for any $\bar{\theta} \in D_0$, $A_\pi(\bar{\theta}) = A_\pi$, $A_\pi$ being the support of the $E$-marginal of $\beta_0$.

Note that when $\Theta$ is finite, it is easy to find conditions for $\pi = \pi^*$, as we mentioned earlier. When $\Theta$ is not finite, it is not so clear when (4.4) holds or
when $\pi = \pi^*$. Instead of (4.4), let us now assume that

\begin{equation}
\pi(D^*) = 1, \text{ where } D^* \text{ is given by}
\end{equation}

\[D^* = \{\bar{\theta} \in \Theta : \sup_{n \geq 0} \|W_{n_1} \cdots W_{n_k}(x)\| < \infty \text{ for } x \in E\}.\]

[Note that (4.13) holds if for each $W_\theta$, $\theta \in \Theta$, the corresponding $A_\theta$ has its norm less than a number $p < 1$.]

We can now state and prove the following theorem:

**Theorem 4.2.** Assume (4.6) and (4.13). Suppose that every basic open set (that is, every open rectangle) in $\Theta$ has $\pi$-measure positive. Suppose also that, for $\theta \in \Theta$, each entry in $W_\theta$ (as a matrix with respect to a fixed basis) is a real continuous function of $\theta$. Then there exists $D \subset \Theta$ such that $\pi(D) = 1$ and, for any $\theta, \bar{\theta} \in D$, $A_\theta(\bar{\theta}) = A_\theta(\bar{\theta})$; moreover, the left attractor $A_\theta(\bar{\theta})$, $\bar{\theta} \in D$, is the support of the $E$-marginal of $\beta_0$, where $\beta_0 \in P(Q)$, $\beta_0 = \beta_0 \Phi^{-1}$ such that its $\Theta$-marginal is $\pi$.

**Proof.** We separate the proof into several steps.

**Step I.** Let $A_1, A_2, \ldots, A_n$ be open subsets of $\Theta$ such that

\[\alpha = \pi(A_1 \times A_2 \times \cdots \times A_n \times \Theta \times \Theta \times \cdots) > 0.\]

Then there exists $D_1 \subset \Theta$, depending upon the sets $A_1, A_2, \ldots, A_n$, such that $\pi(D_1) = 1$ and, for $\bar{\theta} \in D_1$,

\[\theta_{m+i} \in A_i, \quad 1 \leq i \leq n,\]

for infinitely many $m$.

To see this, define the sets $E_m \subset \Theta$ by

\[E_m = \{\bar{\theta} \in \Theta : \theta_{m+i} \in A_i, \quad 1 \leq i \leq n\},\]

and let

\[g(\bar{\theta}) = \limsup_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} I_{E_m}(\bar{\theta}).\]

Then we have

\[g(T\bar{\theta}) = g(\bar{\theta});\]

moreover,

\[\int g d\pi \geq \limsup_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} \pi(E_m) = \alpha > 0.\]

Since $\pi$ is ergodic, $g$ is a positive constant for $\pi$-almost all $\bar{\theta}$. This establishes Step I.
Step II. There exists $D \subset \mathfrak{D}$ with $\pi(D) = 1$ such that $D \subset D^*$, $D^*$ as in (4.13), and for $\vartheta \in D$, (4.8) holds, and also, given any basic open rectangle in $\mathfrak{D}$ of the form $A_1 \times A_2 \times \ldots \times A_n \times \Theta \times \Theta \times \ldots$,

$$\theta_{m+1} \in A_i, \quad 1 \leq i \leq n,$$

for infinitely many $m$.

The proof of this step follows from that of Step I if we observe that $\mathfrak{D}$ is a compact metric space and, as such, it is second countable.

Step III. We show in this step that for $\vartheta, \vartheta' \in D$, $D$ as in Step II, $A_{\vartheta'}(\vartheta) = A_{\vartheta}(\vartheta')$.

With no loss of generality, we can assume that, for all $\vartheta \in D$,

$$\prod_{i=0}^{n} \| A_{\theta_i} \| \to 0 \quad \text{as } n \to \infty.$$ Let $h \in A_{\vartheta}(\vartheta')$. Then there is a subsequence $(n_k)$ of positive integers such that, for $x \in E$,

$$W_{\theta_n} \ldots W_{\theta_0}(x) \to h \quad \text{as } k \to \infty.$$ Let $\vartheta' \in D$ and $\delta > 0$. Fix an $x \in E$. Let $p$ be so large that $\|x\| < p$ and

$$\sup \{ W_{\theta_n} \ldots W_{\theta_0}(x) : n \geq 0 \} \leq p.$$ Let $U_\delta$ be the open set $\{ h' \in E : \| h' - h \| < \delta \}$. Choose $k$ so large that

$$W_{\theta_{n_k}} \ldots W_{\theta_0}(x) \in U_\delta \quad \text{and} \quad \prod_{i=0}^{n_k} \| A_{\theta_i} \| < \delta/(2p).$$ Let this $k$ be now fixed. Because of the continuity assumption in the proposition, there is a basic open set

$$V = B_0 \times \ldots \times B_{n_k} \times \Theta \times \Theta \times \ldots$$ such that for any $\vartheta'' \in D$ such that $\vartheta'' \in B_i$, $0 \leq i \leq n_k$,

$$W_{\theta_{n_k}} \ldots W_{\theta_0}(x) \in U_\delta \quad \text{and} \quad \prod_{i=0}^{n_k} \| A_{\theta_i} \| < \delta/(2p).$$ By Step I, $\vartheta_{m+i} \in B_i$, $0 \leq i \leq n_k$, for infinitely many $m$. This means that there exists a subsequence $(m_s)$ such that, for each $s \geq 1$,

$$W_{\theta_{m_s+n_k}} \ldots W_{\theta_0}(x) \in U_\delta \quad \text{and} \quad \prod_{i=m_s}^{m_s+n_k} \| A_{\theta_i} \| < \delta/(2p).$$ Notice that, for all $s \geq 1$, if $y_s = W_{\theta_{m_s-1}} \ldots W_{\theta_0}(x)$, then $\|y_s\| < p$ and

$$\| W_{\theta_{m_s+n_k}} \ldots W_{\theta_0}(x) - W_{\theta_{m_s+n_k}} \ldots W_{\theta_0}(x) \|$$

$$\leq \| A_{\theta_{m_s+n_k}} \ldots A_{\theta_0}(y_s) \| + \| A_{\theta_{m_s+n_k}} \ldots A_{\theta_0}(x) \| < \frac{\delta}{2p} \cdot p + \frac{\delta}{2p} \cdot p = \delta.$$
This means that, for $s \geq 1$,

$$W_{\theta_{n_1} \cdots \theta_{n_s}}(x) \in U_{2s},$$

so that $A_1(\theta) \cap U_{2s} \neq \emptyset$. Consequently, we have $h \in A_1(\theta)$. Thus, for $\theta, \theta' \in D$, $A_1(\theta) \subseteq A_1(\theta')$.

Step IV. Let $\beta_0 \in P(\varnothing)$ such that $\beta_0 \Phi^{-1} = \beta_0$ and $\beta_0(E \times \Theta) = \pi(\Theta)$ for $\Theta \subseteq \Theta$. (Such a $\beta_0$ exists by the same arguments used earlier.) Let $\mu(A) = \beta_0(A \times \varnothing)$. Then, for $\theta \in D$, $A_1(\theta) = \text{supp}(\mu)$.

To prove this let $y \notin A_1(\theta)$. Then there is an open ball $N_y$ containing $y$ such that $N_y \cap A_1(\theta) = \emptyset$, and $\pi(C) = 0$, where for $x_0 \in E$

$$C = \{\theta: W_{\theta_0} \cdots W_{\theta_0}(x_0) \in N_y \text{ i.o.}\}.$$

Then $\mu(N_y) = \beta_0(N_y \times \varnothing) = \beta_0 \Phi^{-n}(N_y \times \Theta)$ for $n \geq 1$; also

$$E \times C = \{(x, \theta): W_{\theta_0} \cdots W_{\theta_0}(x_0) \in N_y \text{ i.o.}\} = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \Phi^{-n}(N_y \times \Theta),$$

where $N_y$ is an open ball containing $y$, but with its diameter half that of $N_y$.

This means that

$$\mu(N_y) > 0 \Rightarrow \beta_0(E \times C) = \pi(C) > 0,$$

which is a contradiction. Thus, $y \notin \text{supp}(\mu)$. Thus, the support of $\mu$ is contained in $A_1(\theta)$ for $\theta \in D$. The converse is also clear by similar arguments. 

We conclude this section with a few computer-generated pictures of left and right attractors (for a finite $\Theta$).

Example 1. Here we take three maps $W_1$, $W_2$ and $W_3$ from $R^2$ into $R^2$, where $W_i(x) = A_i(x) + B_i$, $x \in R^2$, given by

$$A_1 = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.2 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0.5 & 1 \\ 0 & 0.5 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 0.5 & 0 \\ 0.5 & 0.5 \end{pmatrix};$$

$$B_1 = \begin{pmatrix} 50 \\ 80 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 50 \\ 200 \end{pmatrix}, \quad B_3 = \begin{pmatrix} 170 \\ 200 \end{pmatrix}.$$ 

Here we take three different transition matrices $P$, one with identical rows (and symmetric) and the other two non-symmetric. In the picture, instead of giving $P$, we take the matrix $Z$, where $Z_{ij} = 0$ or 1 according as $P_{ij} = 0$ or $> 0$. Note that if

$$M_{ij} = Z_{ij}/(\sum_{k=1}^{3} Z_{ik}),$$

then the left attractors corresponding to $P$ and $M$, but using the same maps $W_1$, $W_2$ and $W_3$, are the same, and the same is true for the right attractors.
[Let us remark that when $\Theta$ is finite, then if we define $P^\ast$ by
\[ P^\ast_{ij} = P_{ii}(\pi_0)_{ij}/(\pi_0), \]
where $\pi_0 P = \pi_0$ and $(\pi_0)_i > 0$ for all $i$, then $A_i(P) = A_r(P^\ast)$. Also, if $P_1 \approx P_2$ (that is, $(P_1)_{ij} > 0$ iff $(P_2)_{ij} > 0$), then $A_r(P_1) = A_i(P_2)$ and $A_r(P_1) = A_r(P_2)$.

**EXAMPLE 2.** Here we take 12 maps $W_i$, $1 \leq i \leq 12$, such that the corresponding $A_i$'s and $B_i$'s are given by

\[
A_i = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}, \quad 1 \leq i \leq 12;
\]

\[
B_1 = \begin{pmatrix} -100 \\ 0 \end{pmatrix}, \quad B_2 = \begin{pmatrix} -87 \\ 50 \end{pmatrix}, \quad B_3 = \begin{pmatrix} -50 \\ 87 \end{pmatrix}, \quad B_4 = \begin{pmatrix} 0 \\ 100 \end{pmatrix},
\]

\[
B_5 = \begin{pmatrix} 50 \\ 87 \end{pmatrix}, \quad B_6 = \begin{pmatrix} 87 \\ 50 \end{pmatrix}, \quad B_7 = \begin{pmatrix} 100 \\ 0 \end{pmatrix}, \quad B_8 = \begin{pmatrix} 87 \\ -50 \end{pmatrix},
\]

\[
B_9 = \begin{pmatrix} 50 \\ -87 \end{pmatrix}, \quad B_{10} = \begin{pmatrix} 0 \\ -100 \end{pmatrix}, \quad B_{11} = \begin{pmatrix} -50 \\ -87 \end{pmatrix}, \quad B_{12} = \begin{pmatrix} -87 \\ -50 \end{pmatrix}.
\]
Invariant measures and Markov chains

REFERENCES


Department of Mathematics  
University of South Florida  
Tampa, Florida 33620-5700, U.S.A.

Received on 10.3.1996