Ergodic Theorems for Iterated Function Systems with Time Dependent Probabilities

Örjan Stenflo

Abstract. Consider a discrete-time nonhomogeneous Markov chain on a compact state space obtained by random iteration of functions chosen independently in each iteration step from some countable family $\{w_i\}_{i=1}^{\infty}$ of functions. In iteration step n we choose map w_i with probability p_i^n (depending on "time" n). Suppose $p_i^n \to p_i$ for all $i \in \mathbb{N}$. We give some sufficient conditions in order for the nonhomogeneous chain to have similar limiting behavior as the corresponding homogeneous Markov chain (with function w_i chosen with probability p_i in each iteration step).

1. Introduction

In this paper, we will consider Markov chains on compact state spaces obtained by random iteration of functions chosen independently in each iteration step from some countable family of functions.

To be more formal, let (K, d) be a compact metric space and $\{w_i\}_{i=1}^{\infty}$ a family of measurable functions $w_i : K \to K$. The set $\{K;\ w_i,\ i \in \mathbb{N}\}$ is called an iterated function system (IFS). Let $\{I_n\}_{n=0}^{\infty}$ be a sequence of independent random variables with values in $\mathbb{N}$. Specify a starting point $x \in K$. The stochastic sequence $\{I_n\}$ then controls the stochastic dynamical system $\{Z_n(x)\}_{n=0}^{\infty}$, where

$$Z_n(x) := w_{I_{n-1}} \circ w_{I_{n-2}} \circ \cdots \circ w_{I_0}(x),\ n \geq 1, \quad Z_0(x) = x.$$

The sequence $\{Z_n(x)\}_{n=0}^{\infty}$ forms a (in general nonhomogeneous) Markov chain. This Markov chain may be characterized by the IFS $\{K;\ w_i,\ i \in \mathbb{N}\}$ together with the probabilities $\{p_i^n\}_{i=1,n=0}^{\infty}$ where $P(I_n = i) = p_i^n$. Therefore, naturally generalizing the terminology introduced by Barnsley

Key words: Ergodic theorems, Iterated Function Systems (IFS), IFS with time dependent probabilities, Markov chains.

1991 Mathematics Subject Classification: Primary 60J05, 60B10; Secondary 60F05, 60F15.

and Demko (1985), we call this model iterated function systems with time dependent probabilities.

Barnsley and Demko (1985) treat the model with time independent probabilities. (The Markov chain $\{Z_n(x)\}_{n=0}^{\infty}$ is homogeneous i.e. the transition probabilities do not depend on "time" n iff the controlling sequence $\{I_n\}$ consists of independent and identically distributed (i.i.d.) random variables.)

Other related papers concerning the model with i.i.d. controlling sequence are e.g. Kaijser (1978), Lasota and Mackey (1989), Elton and Piccioni (1992), Łoskot and Rudnicki (1995) and Öberg (1997).

Markov chains represented in dynamical form is a natural model in many applications and have been studied e.g. in connection with learning processes by Iosifescu and Theodorescu (1969), and Norman (1972).

We are going to approach nonhomogeneous Markov chains using coupling arguments and the method of reversing time as basic ingredients in our proofs.

The main results are given in Theorems 2.1 and 2.2 below.

2. Main Results

2.1. Statements. We shall here state the main results of this paper concerning IFS with time dependent probabilities. First, however, we need to introduce some definitions and concepts.

Let $\{Z_n(x)\}$ be a (time homogeneous) Markov chain arising from the IFS $\{K;\ w_i,\ i \in \mathbb{N}\}$ with probabilities $\{p_i\}_{i=1}^{\infty}$. Let $\{S_n(x)\}$ be a (time inhomogeneous) Markov chain arising from the same IFS but with "time" dependent probabilities $\{p_i^n\}_{i=1}^{\infty}$. That is,

$$Z_n(x) = w_{I'_{n-1}} \circ w_{I'_{n-2}} \circ \cdots \circ w_{I'_0},\ n \geq 1, \qquad Z_0(x) = x,$$

and

$$S_n(x) = w_{I''_{n-1}} \circ w_{I''_{n-2}} \circ \cdots \circ w_{I''_0},\ n \geq 1, \qquad S_0(x) = x,$$

with $P(I_n^{'} = i) = p_i$ and $P(I_n^{''} = i) = p_i^n$, for each n and i, with $\{I_n^{'}\}$ and $\{I_n^{''}\}$ being sequences of independent random variables.

Denote by $P(\cdot) = P(I_n^{'} \in \cdot)$, $P_n(\cdot) = P(I_n^{''} \in \cdot)$, and $\mu_n^x(\cdot)$ the probability distribution of $S_n(x)$.

Let, for each n,

$$\|P_n - P\| = \sum_{i=1}^{\infty} |p_i^n - p_i|$$

denote the total variation distance between P_n and P.

For Borel probability measures μ_1 and μ_2, let d_k denote the Kantorovich distance defined by

$$d_k(\mu_1, \mu_2) = \sup_{f \in Lip_1} \left| \int_K f d(\mu_1 - \mu_2) \right|,$$

where

$$Lip_1 = \{f : K \to \mathbb{R},\ |f(x) - f(y)| \leq d(x, y) \ for\ all\ x, y \in K\}.$$

Convergence in this metric (sometimes also referred to as the Hutchinson, or Wasserstein, metric) is equivalent to weak convergence i.e.

$$d_k(\mu_n, \mu) \to 0$$

$$\Leftrightarrow$$

$$\int_K f d\mu_n \to \int_K f d\mu,\ for\ all\ f \in C(K),$$

where $C(K)$ denotes the set of real-valued continuous functions on K. (See e.g. Dudley (1989).)

We can now state our main results.

Theorem 2.1. *Suppose*

A: $\sup_{x,y \in K} d(Z_n(x), Z_n(y)) \xrightarrow{P} 0$ *as* $n \to \infty$,
and
B: $\|P_n - P\| \to 0$ *as* $n \to \infty$.

Then there exists a probability measure μ *such that*

$$\sup_{x \in K} d_k(\mu_n^x, \mu) \to 0 \ as\ n \to \infty,$$

i.e. the distribution of $\{S_n(x)\}$ *converges weakly to* μ *uniformly with respect to initial point* $x \in K$.

Under the following stronger assumptions, we may sharpen our conclusion and also obtain convergence rates and a law of large numbers.

Theorem 2.2. *Suppose*

A′: There exists a constant $c < 1$ *such that*

$$Ed(Z_1(x), Z_1(y)) \leq c d(x, y) \qquad for\ all\ x, y \in K,$$

and

B': There exist positive constants c_0 and c_1 with $c_1 < 1$ such that

$$\|P_n - P\| \leq c_0 c_1^n, \ n = 0, 1, \ldots$$

Then:

(i) There exists a probability measure μ and a positive constant ρ such that for any $d > \max(c, c_1)$,

$$\sup_{x \in K} d_k(\mu_n^x, \mu) \leq \rho d^n.$$

(ii) For any $f \in C(K)$, and any $x \in K$, we have that,

$$\frac{\sum_{k=0}^{n-1} f(S_k(x))}{n} \overset{a.s.}{\to} \int f d\mu \quad \text{as } n \to \infty.$$

Remarks:

1. An explicit expression for ρ is given in (2.18) and (2.19) below.
2. If $c_1 \neq c$ we may choose $d = \max(c, c_1)$.
3. If the rate of convergence in B' is slower than geometrical, then the rate of convergence in (i), is determined by the rate in B' (see (2.10) – (2.13) below), and thus in particular we obtain the conclusion of Theorem 2.1 under conditions A' and B.
4. A sufficient condition for (ii) is that A' holds and that $\sum_{n=0}^{\infty} \|P_n - P\| < \infty$ (See (2.20ff.) in the proof below.)
5. Condition A' implies that the functions w_i, $i \in \mathbb{N}$ need to be continuous.
6. Theorem 2.2 (ii) may be generalized under the same assumptions to state: For any $f \in C(K)$, and any sequence of points $\{x_n\}$ in K, we have that,

$$\frac{\sum_{k=0}^{n-1} f(S_k(x_k))}{n} \overset{a.s.}{\to} \int f d\mu \quad \text{as } n \to \infty.$$

2.2. Proofs. We start with some preliminaries common for both the proof of Theorem 2.1 and that of Theorem 2.2. Intuitively, we are going to consider the two sequences $\{S_n(x)\}$ and $\{Z_n(x)\}$ as being defined on the same probability space and realize a coupling type construction.

Since condition A and A', respectively, implies the special cases of the theorems obtained by replacing $\{S_n(x)\}$ with $\{Z_n(x)\}$, (an explanation of this fact is given below), our proofs will be completed if we can do our coupling construction in such a way, that $d(S_n(x), Z_n(x))$ converges to 0, (in the "right" sense).

In the construction below we are, for fixed n, maximizing the conditional probability that $Z_{n+1}(x) = S_{n+1}(x)$ given that $Z_n(x) = S_n(x)$.

Let $\{I_n\}$ be a sequence of i.i.d. random variables uniformly distributed in $[0, 1)$.

Let μ_{Leb} denote the Lebesgue measure. We shall first construct functions g and $g_n : [0, 1) \to \mathbb{N}$, $n = 0, 1, \dots$, such that $\mu_{Leb}(s : g(s) = i) = p_i$, $\mu_{Leb}(s : g_n(s) = i) = p_i^n$ for each $i \in \mathbb{N}$, and such that $\mu_{Leb}(s : g(s) = g_n(s))$ is maximized.

Define $g : [0, 1) \to \mathbb{N}$ by $g(s) = k$ if $t_{k-1} \leq s < t_k$, where $t_q := \sum_1^q p_i$, $t_0 := 0$. Let $A_k^{(n)} = \{s \in [0, 1) : g(s) = k,\ t_{k-1} \leq s < t_{k-1} + p_k^n\}$. Let $A^{(n)} = \{i : p_i < p_i^n\}$. Denote by $i_k^{(n)}$ the k:th smallest element of $A^{(n)}$. Define $q_0^{(n)} = 0$ and for $m \geq 1$,

$$q_m^{(n)} = \inf\{s : \mu_{Leb}([q_{m-1}^{(n)}, s) \backslash (\cup_{k=1}^{\infty} A_k^{(n)})) \geq p^n_{i_m^{(n)}} - p_{i_m^{(n)}}\}.$$

Finally we define

$$g_n(s) = \begin{cases} g(s) & \text{if } \ s \in \cup_{k=1}^{\infty} A_k^{(n)} \\ i_m^{(n)} & \text{if } \ s \in [q_{m-1}^{(n)}, q_m^{(n)}) \backslash (\cup_{k=1}^{\infty} A_k^{(n)}),\ m = 1, 2, \dots \end{cases}$$

From this construction, we see that $I_n^{'}$ and $g(I_n)$ are identically distributed for each n as well as $I_n^{''}$ and $g_n(I_n)$. Since we are only interested in distributional questions, we may consider $Z_n(x)$ and $S_n(x)$ as defined by

$$Z_n(x) := w_{g(I_{n-1})} \circ \cdots \circ w_{g(I_0)}(x),\ n \geq 1, \qquad Z_0(x) = x,$$

and

$$S_n(x) := w_{g_{n-1}(I_{n-1})} \circ \cdots \circ w_{g_0(I_0)}(x),\ n \geq 1, \qquad S_0(x) = x.$$

Proof. (Theorem 2.1)

Define the reversed iterates,

$$\hat{Z}_n(x) := w_{g(I_0)} \circ \cdots \circ w_{g(I_{n-1})}(x),\ n \geq 1, \qquad \hat{Z}_0(x) = x,$$

and

$$\hat{S}_n(x) := w_{g_{n-1}(I_0)} \circ \cdots \circ w_{g_0(I_{n-1})},\ n \geq 1, \qquad \hat{S}_0(x) = x.$$

Note that it is only the i.i.d. sequence $\{I_n\}$ which is reversed, and thus $S_n(x)$ has the same probability distribution as $\hat{S}_n(x)$ for each n. The reason to introduce $\{\hat{Z}_n(x)\}$ and $\{\hat{S}_n(x)\}$ is that these new random sequences converge *a.s.* which in general does not hold for the original sequences.

In fact, we will (see below) prove that there exists a random variable $\hat{Z}$ such that

$$\sup_{x\in K} d(\hat{S}_n(x), \hat{Z}) \stackrel{a.s.}{\to} 0 \quad \text{as } n \to \infty, \tag{2.1}$$

from which the conclusion of the theorem will follow. To see this, note that (2.1) implies that $\hat{S}_n(x_n) \stackrel{a.s.}{\to} \hat{Z}$, for any sequence $\{x_n\}$. Consequently, since almost sure convergence implies weak convergence, $d(\mu_n^{x_n}, \mu) \to 0$, which implies that $\sup_{x\in K} d(\mu_n^x, \mu) \to 0$.

In order to prove (2.1), the following lemma telling that (2.1) holds if $\hat{S}_n(x)$ is replaced by $\hat{Z}_n(x)$, will serve as a starting point.

Lemma 2.3. *The following two conditions are equivalent:*

A: $\sup_{x,y\in K} d(Z_n(x), Z_n(y)) \stackrel{P}{\to} 0 \quad \text{as } n \to \infty,$
and
$\tilde{A}$*: There exists a random variable* $\hat{Z}$ *such that*

$$\sup_{x\in K} d(\hat{Z}_n(x), \hat{Z}) \stackrel{a.s.}{\to} 0 \quad \text{as } n \to \infty.$$

Remark. If all functions w_i, $i \in \mathbb{N}$ are continuous then $\mu(\cdot) = P(\hat{Z} \in \cdot)$ is invariant and unique with this property. Condition A was introduced by Öberg (1997) and a similar condition as Condition $\tilde{A}$ by Letac (1986).

Proof. (Lemma 2.3) Since the random variables $\sup_{x,y\in K} d(Z_n(x), Z_n(y))$ and $\sup_{x,y\in K} d(\hat{Z}_n(x), \hat{Z}_n(y))$ have the same distribution (for each fixed n), we have that

$$\begin{aligned} &\sup_{x,y\in K} d(Z_n(x), Z_n(y)) \stackrel{P}{\to} 0 \\ \Leftrightarrow &\sup_{x,y\in K} d(\hat{Z}_n(x), \hat{Z}_n(y)) \stackrel{P}{\to} 0. \end{aligned}$$

Since $\hat{Z}_n(K) := \{\hat{Z}_n(x);\ x \in K\}$ is a nested nonincreasing sequence of sets we see that

$$\begin{aligned} &\sup_{x,y\in K} d(\hat{Z}_n(x), \hat{Z}_n(y)) \stackrel{P}{\to} 0 \\ \Leftrightarrow &\sup_{x,y\in K} d(\hat{Z}_n(x), \hat{Z}_n(y)) \stackrel{a.s.}{\to} 0. \end{aligned}$$

Thus

$$\sup_{x,y\in K} d(Z_n(x), Z_n(y)) \stackrel{P}{\to} 0 \Leftrightarrow \sup_{x,y\in K} d(\hat{Z}_n(x), \hat{Z}_n(y)) \stackrel{a.s.}{\to} 0. \tag{2.2}$$

To prove that $A \Rightarrow \tilde{A}$, let y_0 be an arbitrary point in K. From the construction of $\hat{Z}_n$ we observe, for $m > n$, that

$$\begin{aligned} d(\hat{Z}_n(y_0), \hat{Z}_m(y_0)) &= d(\hat{Z}_n(y_0), \hat{Z}_n(w_{g(I_n)} \circ \cdots \circ w_{g(I_{m-1})}(y_0))) \\ &\leq \sup_{x,y \in K} d(\hat{Z}_n(x), \hat{Z}_n(y)). \end{aligned}$$

It follows from Condition A and (2.2) that

$$\sup_{x,y \in K} d(\hat{Z}_n(x), \hat{Z}_n(y)) \stackrel{a.s.}{\to} 0 \tag{2.3}$$

and thus $\{\hat{Z}_n(y_0)\}$ is $a.s.$ a Cauchy sequence which converges, to say $\hat{Z}$, since K is complete. Thus since

$$\sup_{x \in K} d(\hat{Z}_n(x), \hat{Z}) \leq \sup_{x \in K} d(\hat{Z}_n(x), \hat{Z}_n(y_0)) + d(\hat{Z}_n(y_0), \hat{Z}),$$

it follows using (2.3), that $\sup_{x \in K} d(\hat{Z}_n(x), \hat{Z}) \to 0$, and $A \Rightarrow \tilde{A}$ is proved.

The proof of $\tilde{A} \Rightarrow A$ follows immediately from (2.2) and the triangle inequality.

□

We now return to the proof of Theorem 2.1. From the assumptions together with Lemma 2.3 it follows that there exists a random variable $\hat{Z}$ such that

$$\sup_{x \in K} d(\hat{Z}_n(x), \hat{Z}) \stackrel{a.s.}{\to} 0 \quad as\ n \to \infty. \tag{2.4}$$

It remains to prove that (2.4) holds with $\hat{Z}_n$ replaced by $\hat{S}_n$. This we will do by using a comparison method.

For any $\epsilon > 0$ there exists a random integer N_1 (finite with probability one) such that

$$\sup_{x \in K} d(\hat{Z}_{N_1}(x), \hat{Z}) < \epsilon.$$

Define, for $n \geq m$ with $m \geq 1$ fixed,

$$\hat{S}_m^n(x) := w_{g_{n-1}(I_0)} \circ \cdots \circ w_{g_{n-m}(I_{m-1})}(x).$$

As a consequence of condition B, $g_n \to g$ $a.e.$ with respect to the Lebesgue measure in the discrete metric i.e.

$$\mu_{Leb}(\cap_{N=1}^{\infty} \cup_{n=N}^{\infty} \{s \in [0,1); g_n(s) \neq g(s)\}) = 0.$$

Thus there exists a random integer $N_2 = N_2(1)$ (finite with probability one) such that $\hat{S}_1^n = \hat{Z}_1$ if $n \geq N_2$, and by an inductive argument we can show that for all m, there exists a random integer $N_2 = N_2(m)$ (finite with probability one) such that $\hat{S}^n_m = \hat{Z}_m$ if $n \geq N_2$.

Let N_3 be a random integer ($> N_1$) such that $\hat{S}^n_{N_1} = \hat{Z}_{N_1}$ if $n \geq N_3$. (From the above we see that it is possible to choose N_3 to be finite with probability one.)

For $n \geq N_3$ we then have that

$$\begin{aligned}\sup_{x\in K} d(\hat{S}_n(x), \hat{Z}) &\leq \sup_{x\in K} d(\hat{S}^n_{N_1}(w_{g_{n-N_1-1}(I_{N_1})} \circ \cdots \circ w_{g_0(I_{n-1})}(x)), \hat{Z}) \\ &\leq \sup_{x\in K} d(\hat{Z}_{N_1}(w_{g_{n-N_1-1}(I_{N_1})} \circ \cdots \circ w_{g_0(I_{n-1})}(x)), \hat{Z}) \\ &\leq \sup_{x\in K} d(\hat{Z}_{N_1}(x), \hat{Z}) < \epsilon,\end{aligned}$$

and thus the proof of Theorem 2.1 is completed. □

Proof. (Theorem 2.2)

Let $D_K = \sup_{x,y\in K} d(x, y)$ denote the diameter of K. Define

$$\delta_n = D_K \|P_n - P\|. \tag{2.5}$$

We have the following inequality

$$\begin{aligned}&\sup_{x\in K} Ed(w_{g(I_0)}(x), w_{g_n(I_0)}(x)) \leq D_K \mu_{Leb}\{s : g_n(s) \neq g(s)\} \\ &\leq D_K(1 - \sum_{i=1}^{\infty} \min(p_i, p_i^n)) \leq D_K \sum_{i=1}^{\infty}(\max(p_i, p_i^n) - \min(p_i, p_i^n)) \\ &\leq D_K \sum_{i=1}^{\infty} |p_i^n - p_i| = \delta_n. \qquad (2.6)\end{aligned}$$

Let x be an arbitrary point in K. By using (2.6) and condition A' in the triangle inequality, we see that

$$\begin{aligned}Ed(S_n(x), Z_n(x)) &= Ed(w_{g_{n-1}(I_{n-1})}(S_{n-1}(x)), w_{g(I_{n-1})}(Z_{n-1}(x))) \\ &\leq Ed(w_{g_{n-1}(I_{n-1})}(S_{n-1}(x)), w_{g(I_{n-1})}(S_{n-1}(x))) \\ &\quad + Ed(w_{g(I_{n-1})}(S_{n-1}(x)), w_{g(I_{n-1})}(Z_{n-1}(x))) \\ &\leq \delta_{n-1} + cEd(S_{n-1}(x), Z_{n-1}(x)), \qquad (2.7)\end{aligned}$$

and using (2.7) recursively we obtain the inequality

$$Ed(S_n(x), Z_n(x)) \leq \sum_{i=0}^{n-1} c^{n-1-i} \delta_i. \tag{2.8}$$

Let ν_n^x denote the probability distribution of $Z_n(x)$. As was shown in Stenflo (1998), with the use of reversing time techniques, condition A' implies the existence of a probability measure μ (invariant for $\{Z_n(x)\}$) such that

$$\sup_{x \in K} d_k(\nu_n^x, \mu) \leq \frac{D_K}{1-c} c^n. \tag{2.9}$$

Using this and the triangle inequality, we first observe that

$$\begin{aligned} \sup_{x \in K} d_k(\mu_n^x, \mu) &\leq \sup_{x \in K} d_k(\mu_n^x, \nu_n^x) + \sup_{x \in K} d_k(\nu_n^x, \mu) \\ &\leq \sup_{x \in K} d_k(\mu_n^x, \nu_n^x) + \frac{D_K}{1-c} c^n. \end{aligned} \tag{2.10}$$

Now,

$$\begin{aligned} d_k(\mu_n^x, \nu_n^x) &= \sup_{f \in Lip_1} \left| \int f d(\mu_n^x - \nu_n^x) \right| = \sup_{f \in Lip_1} |E(f(S_n(x)) - f(Z_n(x)))| \\ &\leq \sup_{f \in Lip_1} E|f(S_n(x)) - f(Z_n(x))| \leq Ed(S_n(x), Z_n(x)). \end{aligned} \tag{2.11}$$

From (2.5), (2.8), and condition B' we obtain that

$$Ed(S_n(x), Z_n(x)) \leq \sum_{i=0}^{n-1} c^{n-1-i} \delta_i \leq \sum_{i=0}^{n-1} c^{n-1-i} D_K c_0 c_1^i. \tag{2.12}$$

By combining (2.11) and (2.12) we thus see that

$$d_k(\mu_n^x, \nu_n^x) \leq \sum_{i=0}^{n-1} c^{n-1-i} D_K c_0 c_1^i. \tag{2.13}$$

If $c > c_1$, we see from (2.13) that

$$d_k(\mu_n^x, \nu_n^x) \leq D_K c_0 c^{n-1} \sum_{i=0}^{n-1} \left(\frac{c_1}{c} \right)^i \leq D_K \frac{c_0}{c - c_1} c^n,$$

and similarly if $c < c_1$ (by interchanging c and c_1),

$$d_k(\mu_n^x, \nu_n^x) \leq D_K \frac{c_0}{c_1 - c} c_1^n.$$

Thus for $c \neq c_1$ we have that

$$d_k(\mu_n^x, \nu_n^x) \leq D_K \left| \frac{c_0}{c - c_1} \right| (\max(c, c_1))^n. \tag{2.14}$$

Using (2.10) and (2.14) we see that

$$\begin{aligned} \sup_{x \in K} d_k(\mu_n^x, \mu) &\leq D_K \left| \frac{c_0}{c - c_1} \right| (\max(c, c_1))^n + \frac{D_K}{1-c} c^n \\ &\leq D_K \left(\left| \frac{c_0}{c - c_1} \right| + \frac{1}{1-c} \right) (\max(c, c_1))^n. \end{aligned} \tag{2.15}$$

If $c = c_1$ we obtain from (2.13) that

$$d_k(\mu_n^x, \nu_n^x) \leq D_K c_0 n c^{n-1}, \tag{2.16}$$

and finally by inserting this in (2.10), we obtain that

$$\sup_{x \in K} d_k(\mu_n^x, \mu) \leq D_K (\frac{c}{1-c} + n c_0) c^{n-1}. \tag{2.17}$$

All together if $c \neq c_1$, let

$$\rho = D_K \left(\left| \frac{c_0}{c - c_1} \right| + \frac{1}{1-c} \right), \tag{2.18}$$

and if $c = c_1$ let

$$\rho = D_K \sup_{n \geq 0} ((\frac{c}{d(1-c)} + n \frac{c_0}{d})(\frac{c}{d})^{n-1}). \tag{2.19}$$

Then

$$\sup_{x \in K} d_k(\mu_n^x, \mu) \leq \rho d^n.$$

This completes the proof of Theorem 2.2 (i).

To prove Theorem 2.2 (ii), let x be an arbitrary point K. From (2.8) it follows that

$$\begin{aligned} \sum_{n=0}^{\infty} Ed(S_n(x), Z_n(x)) &\leq \sum_{n=1}^{\infty} \sum_{i=0}^{n-1} c^{n-1-i} \delta_i \leq \lim_{n \to \infty} \sum_{i=0}^{n} \delta_i \frac{1 - c^{n+1-i}}{1-c} \\ &\leq \frac{1}{1-c} \sum_{i=0}^{\infty} \delta_i, \end{aligned} \tag{2.20}$$

and thus from (2.5) and condition B', $\sum_{n=0}^{\infty} Ed(S_n(x), Z_n(x)) < \infty$.

Thus $d(S_n(x), Z_n(x)) \overset{\text{a.s.}}{\to} 0$ by the Chebyshev inequality and the Borel–Cantelli lemma, and consequently for any $f \in C(K)$, we have that

$$|f(S_n(x)) - f(Z_n(x))| \overset{\text{a.s.}}{\to} 0. \tag{2.21}$$

From condition A' it follows (see Stenflo (1998) for details) that $\{Z_n(x)\}$ has a unique invariant probability measure μ, and has the Feller property, i.e. if $f \in C(K)$ then also $Ef(Z_1) \in C(K)$. Since we have a Markov chain with a unique invariant probability measure on a compact metric space having the Feller property, the conditions in a theorem by Breiman (1960) are satisfied and we may use it to obtain

$$\left| \frac{\sum_{k=0}^{n-1} f(Z_k(x))}{n} - \int f d\mu \right| \overset{a.s.}{\to} 0,$$

for any $x \in K$. Using (2.21) and the fact that convergence implies convergence in the Cesaro sense, we see that, for any $x \in K$,

$$\left| \frac{\sum_{k=0}^{n-1} f(S_k(x))}{n} - \int f d\mu \right| \leq \frac{\sum_{k=0}^{n-1} |f(S_k(x)) - f(Z_k(x))|}{n}$$

$$+ \left| \frac{\sum_{k=0}^{n-1} f(Z_k(x))}{n} - \int f d\mu \right| \overset{a.s.}{\to} 0.$$

This completes the proof of Theorem 2.2.

□

References

[1] Barnsley, M. F. and Demko, S. (1985) Iterated function systems and the global construction of fractals, *Proc. Roy. Soc. London Ser. A*, **399**, 243–275.

[2] Breiman, L. (1960) The strong law of large numbers for a class of Markov chains, *Ann. Math. Statist.*, **31**, 801–803.

[3] Dudley, R. M. (1989) *Real Analysis and Probability*, Wadsworth, New York.

[4] Elton, J. H. and Piccioni, M. (1992) Iterated function systems arising from recursive estimation problems, *Probab. Theory Related Fields*, **91**, 103–114.

[5] Iosifescu, M. and Theodorescu, R. (1969) *Random Processes and Learning*, Springer-Verlag, New York.

[6] Kaijser, T. (1978) A limit theorem for Markov chains in compact metric spaces with applications to products of random matrices, *Duke Math. J.*, **45**, 311–349.

[7] Lasota, A. and Mackey, M. (1989) Stochastic perturbation of dynamical systems: the weak convergence of measures, *J. Math. Anal. Appl.*, **138**, 232–248.

[8] Letac, G. (1986) A contraction principle for certain Markov chains and its applications, *Contemp. Math.*, **50**, 263–273.

[9] Łoskot, K. and Rudnicki, R. (1995) Limit theorems for stochastically perturbed dynamical systems, *J. Appl. Probab.*, **32**, 459–469.
[10] Norman, M. F. (1972) *Markov Processes and Learning Models*, Academic Press, New York.
[11] Öberg, A. (1997) Approximation of invariant measures for random iterations, *Research reports No 7*, Dept. of Mathematics, Umeå University.
[12] Stenflo, Ö. (1998) Ergodic theorems for Markov chains represented by iterated function systems, *Research reports No 2*, Dept. of Mathematics, Umeå University.

E-mail address: `stenflo@math.umu.se`