Ergodic Theorems for Time-Dependent Random Iteration of Functions

Örjan Stenflo

Abstract. Time-dependent iterated function systems with time dependent probabilities are introduced, generalizing the concept of iterated function systems by Barnsley and Demko (1985). A distributional ergodic theorem including rates of convergence and a law of large numbers are obtained under the assumptions that the system is asymptotically average-contractive.

1. Introduction

Let (X, d) be a complete metric space, S a measurable space, and P a probability measure defined on the measurable subsets of S. Consider a measurable function $w : X \times S \to X$. For each fixed $s \in S$, we write $w_s(x) := w(x, s)$. Following the terminology introduced by Barnsley and Demko (1985) we call the set $\{X;\ w_s,\ s \in S,\ P\}$ an iterated function system (IFS) with probabilities. Let $\{I_n\}_{n=0}^{\infty}$ be a sequence of independent identically distributed (i.i.d.) random variables with values in S. We assume that $P(\cdot) = P(I_n \in \cdot)$. Specify a starting point $x \in X$. The stochastic sequence $\{I_n\}$ then controls the stochastic dynamical system $\{Z_n(x)\}_{n=0}^{\infty}$, where

$$Z_n(x) := w_{I_{n-1}} \circ w_{I_{n-2}} \circ \cdots \circ w_{I_0}(x),\ n \geq 1, \qquad Z_0(x) = x. \tag{1.1}$$

We call this particular type of stochastic dynamical system an IFS controlled by $\{I_n\}$. The sequence $\{Z_n(x)\}_{n=0}^{\infty}$ forms a homogeneous Markov chain.

Key words: Ergodic theorems, Iterated Function Systems (IFS), Markov chains.
1991 Mathematics Subject Classification: Primary 60B10, 60J05, 60F05; Secondary 60F15.

Denote by $\nu_n^x(\cdot) = P(Z_n(x) \in \cdot)$. For Borel probability measures μ_1 and μ_2, let d_k denote the Kantorovich distance defined by

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

where

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

In Stenflo (1998) the following theorem is proved:

Theorem 1.1. *Assume that:*

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

$$Ed(w_{I_0}(x), w_{I_0}(y)) \leq cd(x, y) \quad for\ all\ x, y \in X,$$

and
(B): $Ed(x_0, w_{I_0}(x_0)) < \infty$, *for some $x_0 \in X$.*

Then there exists a unique invariant probability measure μ for the Markov chain $\{Z_n(x)\}$ such that for any bounded set $K \subseteq X$ there exists a positive constant γ_K such that

$$\sup_{x \in K} d_k(\nu_n^x, \mu) \leq \frac{\gamma_K}{1 - c} c^n,\ \ n \geq 0. \tag{1.2}$$

Remark 1. An explicit expression and upper bound for γ_K is given by

$$\gamma_K := \sup_{x \in K} Ed(x, w_{I_0}(x)) \leq Ed(x_0, w_{I_0}(x_0)) + (c + 1) \sup_{x \in K} d(x_0, x) < \infty.$$

The purpose of this paper is to generalize the above result to the class of nonhomogeneous Markov chains that are generated by time-dependent (asymptotically time-independent) iteration of functions. We are also going to prove a law of large numbers for such processes.

This result also generalizes Theorem 2.2 in Stenflo (1997) where time-dependent random iteration of functions chosen from a countable family, $\{w_i\}_{i \in \mathbb{N}}$, of functions on a compact state space, K, is considered.

2. Main theorems

Formalizing the introduction, we define a time-dependent IFS with time-dependent probabilities as a set of iterated function systems $\{X;\ w_s^{(n)}, s \in S,\ P_n\}_{n=0}^{\infty}$.

Let $\{S_n\}$ be a (time-inhomogeneous) Markov chain arising from independent iteration of functions, choosing a function to iterate in the $(n+1)$:th iteration step from the family $\{w_s^{(n)}, s \in S\}$ of functions according to the probability measure P_n.

That is,

$$S_n(x) = w_{I'_{n-1}}^{(n-1)} \circ w_{I'_{n-2}}^{(n-2)} \circ \cdots \circ w_{I'_0}^{(0)}(x),\ n \geq 1, \qquad S_0(x) = x, \tag{2.1}$$

where $\{I'_n\}$ is an S-valued sequence of independent random variables, with $P(I'_n \in \cdot) = P_n(\cdot)$, for each $n \geq 0$.

For each $n \geq 0$, let μ_n^x denote the probability distribution of $S_n(x)$, and let $\mathbf{M_n}$ denote the set of probability measures Q_n on the measurable subsets of $S \times S$ such that $Q_n(\cdot, S) = P(\cdot)$, and $Q_n(S, \cdot) = P_n(\cdot)$. We may, without loss of generality, consider $\{I_n\}$ and $\{I'_n\}$ as being defined on the same probability space with $P((I_n, I'_n) \in \cdot) = Q_n(\cdot)$ for some $Q_n \in \mathbf{M_n}$, with $\{(I_n, I'_n)\}$ being a sequence of independent random vectors.

Define

$$\delta_n = \inf_{Q_n \in \mathbf{M_n}} \sup_{x \in X} E_{Q_n} d(w_{I_n}(x), w_{I'_n}^{(n)}(x)),\ n \geq 0.$$

(In Section 3 we give some upper bounds for δ_n.)

We can now state the main results of this paper. (See Example 3.1 in Section 3 for an illustration of the theorem below.)

Theorem 2.1. *Suppose the conditions (A) and (B) hold. Let K be a bounded set in X, c be a constant defined by Condition (A), and let μ denote the unique invariant probability measure for the Markov chain $\{Z_n(x)\}$ (existing due to Theorem 1.1).*

(i): Assume that

$$\Delta_\delta := \sum_{n=0}^{\infty} \frac{\delta_n}{c^n} < \infty.$$

Then

$$\sup_{x \in K} d_k(\mu_n^x, \mu) \leq \alpha_\delta c^n,\ n \geq 0,$$

where $\alpha_\delta := \frac{1}{c}\Delta_\delta + \frac{\gamma_K}{1-c}$ is a finite constant.

(ii): Assume that there exists a log-convex sequence $\{q_n\}_{n=0}^{\infty}$ of positive real numbers with $q_n \to 0$, and $q_n \geq \delta_n$ for each $n \geq 0$, with the property

that

$$\Delta_q^{'} := \sum_{n=1}^{\infty} \frac{c^n}{q_n} < \infty.$$

Then

$$\sup_{x \in K} d_k(\mu_n^x, \mu) \leq \beta_q q_n, \ \ n \geq 0,$$

where $\beta_q = \frac{q_0}{c}\Delta_q^{'} + \frac{\gamma_K}{1-c}\sup_{n\geq 0}\frac{c^n}{q_n}$ *is a finite constant.*

Remark 2. If $c = 0$, and thus $\Delta_q^{'} = 0$, then β_q should be interpreted as $\beta_q = \frac{q_0}{q_1} + \frac{\gamma_K}{q_0}$, and if also $\delta_n = 0$, for each $n \geq 0$, we should interpret α_δ as $\alpha_\delta = \gamma_K$.

Remark 3. Note that Theorem 1.1 corresponds to the case when $\delta_n = 0$, for each $n \geq 0$.

Remark 4. We stress that if $\delta_n \to 0$ essentially faster than c^n we may use Theorem 2.1(i) and if $\delta_n \to 0$ with slower rates we use Theorem 2.1(ii).

Theorem 2.2. *Suppose conditions (A) and (B) hold and that*

$$\sum_{n=0}^{\infty} \delta_n < \infty.$$

Then for any uniformly continuous function $f : X \to \mathbb{R}$*, with* $\int |f| d\mu < \infty$*, and any* $x \in X$ *we have that,*

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

Proofs. Let $\{\epsilon_n\}$ be an arbitrary fixed sequence of positive real numbers. Since we are only interested in distributional questions, we may consider $\{I_n\}$ and $\{I_n^{'}\}$ as being defined on the same probability space such that $\{(I_n, I_n^{'})\}$ is an independent sequence with the property that,

$$\sup_{x \in X} Ed(w_{I_n}(x), w_{I_n^{'}}^{(n)}(x)) < \delta_n + \epsilon_n, \tag{2.2}$$

for all $n \geq 0$. (All expectations with which we will be concerned, will be with respect to the, by the Ionescu Tulcea's Theorem, unique probability measure Q generated on the space of infinite sequences of $S \times S$, by $\{(I_n, I_n^{'})\}$.)

Let x and y be arbitrary points in X. By using the triangle inequality, Condition (A) and (2.2) we see that

$$
\begin{aligned}
Ed(Z_n(x), S_n(y)) &= Ed(w_{I_{n-1}}(Z_{n-1}(x)), w^{(n-1)}_{I'_{n-1}}(S_{n-1}(y))) \\
&\leq Ed(w_{I_{n-1}}(Z_{n-1}(x)), w_{I_{n-1}}(S_{n-1}(y))) \\
&\quad + Ed(w_{I_{n-1}}(S_{n-1}(y)), w^{(n-1)}_{I'_{n-1}}(S_{n-1}(y))) \\
&\leq E(E(d(w_{I_{n-1}}(Z_{n-1}(x)), w_{I_{n-1}}(S_{n-1}(y)))|Z_{n-1}(x), S_{n-1}(y))) \\
&\quad + (\delta_{n-1} + \epsilon_{n-1}) \\
&\leq cEd(Z_{n-1}(x), S_{n-1}(y)) + (\delta_{n-1} + \epsilon_{n-1}), \qquad (2.3)
\end{aligned}
$$

and using (2.3) recursively we obtain the inequality

$$Ed(Z_n(x), S_n(y)) \leq c^n d(x, y) + \sum_{i=0}^{n-1} c^{n-1-i}(\delta_i + \epsilon_i). \qquad (2.4)$$

Now observe that

$$
\begin{aligned}
d_k(\nu_n^x, \mu_n^x) &= \sup_{f\in Lip_1} |\int f d(\nu_n^x - \mu_n^x)| = \sup_{f\in Lip_1} |E(f(Z_n(x)) - f(S_n(x)))| \\
&\leq \sup_{f\in Lip_1} E|f(Z_n(x)) - f(S_n(x))| \leq Ed(Z_n(x), S_n(x)), \qquad (2.5)
\end{aligned}
$$

and combining this with (2.4) (with $x = y$) we find that

$$d_k(\nu_n^x, \mu_n^x) \leq \sum_{i=0}^{n-1} c^{n-1-i}(\delta_i + \epsilon_i).$$

Since $\{\epsilon_n\}$ is arbitrary we obtain

$$d_k(\nu_n^x, \mu_n^x) \leq \sum_{i=0}^{n-1} c^{n-1-i}\delta_i. \qquad (2.6)$$

From Theorem 1.1, condition (A) and (B) implies the existence of a probability measure μ (invariant for $\{Z_n(x)\}$) such that

$$\sup_{x\in K} d_k(\nu_n^x, \mu) \leq \frac{\gamma_K}{1-c} c^n, \ n \geq 0. \qquad (2.7)$$

By the triangle inequality (2.6) and (2.7) we obtain that

$$
\begin{aligned}
\sup_{x\in K} d_k(\mu_n^x, \mu) &\leq \sup_{x\in K} d_k(\mu_n^x, \nu_n^x) + \sup_{x\in K} d_k(\nu_n^x, \mu) \\
&\leq \sum_{i=0}^{n-1} c^{n-1-i}\delta_i + \frac{\gamma_K}{1-c} c^n. \qquad (2.8)
\end{aligned}
$$

Consequently, if $\Delta_\delta = \sum_{i=0}^{\infty} \frac{\delta_i}{c^i} < \infty$, $(c > 0)$

$$\frac{\sup_{x\in K} d_k(\mu_n^x, \mu)}{c^n} \leq \frac{1}{c}\sum_{i=0}^{n-1} \frac{\delta_i}{c^i} + \frac{\gamma_K}{1-c} \leq \alpha_\delta,$$

where $\alpha_\delta := \frac{1}{c}\Delta_\delta + \frac{\gamma_K}{1-c}$ is (by assumption) a finite constant. (If $c = 0$ we see from (2.8) that $\sup_{x\in K} d_k(\mu_n^x, \mu) \leq \delta_{n-1}$, for $n \geq 1$, and $\sup_{x\in K} d_k(\mu_0^x, \mu) = \gamma_K$). This completes the proof of Theorem 2.1 (i).

To prove Theorem 2.1 (ii), let $\{q_n\}_{n=0}^{\infty}$ be the positive sequence with $q_n \geq \delta_n$ for each $n \geq 0$ existing by assumption. Using (2.8) we obtain that

$$\frac{\sup_{x\in K} d_k(\mu_n^x, \mu)}{q_n} \leq \sum_{i=0}^{n-1} \frac{c^{n-1-i} q_i}{q_n} + \frac{\gamma_K}{1-c}\frac{c^n}{q_n}. \tag{2.9}$$

Since q_n is log-convex, i.e. $q_{n+1}q_{n-1} - q_n^2 \geq 0$ for each $n \geq 1$, it follows that $\frac{q_i}{q_n} \leq \frac{q_0}{q_{n-i}}$ for each $n \geq i \geq 0$. Substituting this in (2.9) we obtain

$$\begin{aligned} \frac{\sup_{x\in K} d_k(\mu_n^x, \mu)}{q_n} &\leq \sum_{i=0}^{n-1} \frac{c^{n-1-i} q_0}{q_{n-i}} + \frac{\gamma_K}{1-c}\frac{c^n}{q_n} \\ &\leq \sum_{j=0}^{n-1} \frac{c^j q_0}{q_{j+1}} + \frac{\gamma_K}{1-c}\frac{c^n}{q_n}. \end{aligned} \tag{2.10}$$

Thus, if $c > 0$,

$$\frac{\sup_{x\in K} d_k(\mu_n^x, \mu)}{q_n} \leq \beta_q,$$

where $\beta_q = \frac{q_0}{c}\Delta_q^{'} + \frac{\gamma_K}{1-c}\sup_{n\geq 0}\frac{c^n}{q_n}$ is a finite constant since by assumption $\Delta_q^{'} = \sum_{n=1}^{\infty} \frac{c^n}{q_n} < \infty$. (If $c = 0$, we see from (2.10) that we may choose $\beta_q = \frac{q_0}{q_1} + \frac{\gamma_K}{q_0}$.) This completes the proof of Theorem 2.1 (ii).

In order to prove Theorem 2.2, let x and y be two arbitrary points in X. From (2.4) it follows that

$$\sum_{n=0}^{\infty} Ed(Z_n(x), S_n(y)) \leq \sum_{n=0}^{\infty}(c^n d(x,y) + \sum_{i=0}^{n-1} c^{n-1-i}(\delta_i + \epsilon_i))$$

$$\leq \frac{1}{1-c} d(x,y) + \lim_{n\to\infty} \sum_{i=0}^{n} (\delta_i + \epsilon_i) \frac{1-c^{n+1-i}}{1-c}$$

$$\leq \frac{1}{1-c} d(x,y) + \frac{1}{1-c} \sum_{i=0}^{\infty} (\delta_i + \epsilon_i) \tag{2.11}$$

and thus, since by assumption $\sum_{i=0}^{\infty} \delta_i < \infty$, and since $\{\epsilon_n\}$ is arbitrary, it follows that we may assume that

$$\sum_{n=0}^{\infty} Ed(Z_n(x), S_n(y)) < \infty.$$

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

$$|f(Z_n(x)) - f(S_n(y))| \overset{\text{a.s.}}{\to} 0. \tag{2.12}$$

From conditions (A) and (B) it follows by Theorem 1.1 (see Stenflo (1998) for details) that the Markov Chain $\{Z_n(x)\}$ has a unique invariant probability measure μ.

Thus, if Z_0 is chosen to have probability distribution μ and being independent of $\{I_n\}_{n=0}^{\infty}$, then $\{Z_n(Z_0)\}$ will form a stationary ergodic (see Elton (1987)) sequence and in particular, by Birkhoff's theorem, there exist a point $y_0 \in X$ such that

$$\left| \frac{\sum_{k=0}^{n-1} f(Z_k(y_0))}{n} - \int f d\mu \right| \overset{a.s.}{\to} 0.$$

Using (2.12) and the fact that convergence implies convergence in the Cesaro sense, we see that, for any $x \in X$,

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

$$\leq \frac{\sum_{k=0}^{n-1} |f(S_k(x)) - f(Z_k(y_0))|}{n} + \left| \frac{\sum_{k=0}^{n-1} f(Z_k(y_0))}{n} - \int f d\mu \right| \overset{a.s.}{\to} 0.$$

Thus, for any $x \in X$,

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

and this completes the proof of Theorem 2.2. □

3. Upper bounds for δ_n

In this section we are going to give some upper bounds for δ_n. Recall that

$$\delta_n = \inf_{Q_n \in \mathbf{M_n}} \sup_{x \in X} E_{Q_n} d(w_{I_n}(x), w^{(n)}_{I'_n}(x)),\ n \geq 0,$$

where $\mathbf{M_n}$ denotes the set of probability measures Q_n on the measurable subsets of $S \times S$ such that $Q_n(\cdot, S) = P(\cdot)$, and $Q_n(S, \cdot) = P_n(\cdot)$, and (I_n, I'_n) are independent Q_n distributed random variables. In order to separate distributional perturbations and perturbations in the functions, define, for $n \geq 0$,

$$\delta^1_n := \inf_{Q_n \in \mathbf{M_n}} \sup_{x \in X} E_{Q_n} d(w_{I_n}(x), w_{I'_n}(x)),$$

$$\delta^2_n := \sup_{x \in X} Ed(w_{I'_n}(x), w^{(n)}_{I'_n}(x)),$$

and

$$\delta^{(1)}_n := \inf_{Q_n \in \mathbf{M_n}} \sup_{x \in X} E_{Q_n} d(w^{(n)}_{I_n}(x), w^{(n)}_{I'_n}(x)),$$

$$\delta^{(2)}_n := \sup_{x \in X} Ed(w_{I_n}(x), w^{(n)}_{I_n}(x)).$$

By the triangle inequality, $\delta_n \leq \delta^1_n + \delta^2_n$, and $\delta_n \leq \delta^{(1)}_n + \delta^{(2)}_n$.

Note also that $\delta^1_n = \delta^{(1)}_n = 0$ if $P_n = P$. (This is obvious since in this case the infimum is attained for $I_n = I'_n$, for each $n \geq 0$.)

It is known (see Dobrushin (1970)) that for each $n \geq 0$, there exists Q_n-distributed random vectors (I_n, I'_n), with $Q_n \in \mathbf{M_n}$ such that $Q_n(I_n \neq I'_n) = \|P_n - P\|$, where $\|P_n - P\|$ denotes the total variation distance between P_n and P.

Suppose all maps have the boundedness property that

$$B_w := \sup_{x \in X} \sup_{s,t \in S} d(w_s(x), w_t(x)) < \infty.$$

Note that $B_w < \infty$ if X is bounded. From the result by Dobrushin (1970) we see that $\delta^1_n \leq B_w \|P_n - P\|$.

Define

$$d(w, w^{(n)}) := \sup_{s \in S} \sup_{x \in X} d(w_s(x), w^{(n)}_s(x)).$$

It is evident that $\delta^2_n \leq d(w, w^{(n)})$ and $\delta^{(2)}_n \leq d(w, w^{(n)})$.

We conclude this paper with an example illustrating the Theorems of the previous section.

Example 3.1. Consider the time-dependent IFS with time-dependent probabilities $\{\mathbb{R}_+;\ w_i^{(n)},\ i \in S = \{1,2\},\ P_n\}_{n=0}^{\infty}$, with

$$w_1^{(n-1)}(x) = \begin{cases} \frac{3x}{2} + \frac{1}{n^2} & \text{if } x \le 1 \\ \sqrt{x} + \frac{1}{2} + \frac{1}{n^2} & \text{if } x > 1 \end{cases},\quad w_2^{(n-1)}(x) = \begin{cases} \frac{x}{4} + \frac{3}{4} + \frac{1}{n^3} & \text{if } x \le 1 \\ \sqrt{x} + \frac{1}{n^3} & \text{if } x > 1 \end{cases},$$

$P_{n-1}(1) = 1/2 + (1/2)^n$, and $P_{n-1}(2) = 1/2 - (1/2)^n$, for $n \ge 1$. This system generates, see (2.1), a nonhomogeneous Markov chain $\{S_n(x)\}$.

Let μ denote the unique invariant probability measure for the homogeneous Markov chain $\{Z_n(x)\}$, defined in (1.1), generated by the IFS $\{\mathbb{R}_+;\ w_i,\ i \in S = \{1,2\},\ P\}$, with

$$w_1(x) = \begin{cases} \frac{3x}{2} & \text{if } x \le 1 \\ \sqrt{x} + \frac{1}{2} & \text{if } x > 1 \end{cases},\quad w_2(x) = \begin{cases} \frac{x}{4} + \frac{3}{4} & \text{if } x \le 1 \\ \sqrt{x} & \text{if } x > 1 \end{cases},$$

$P(1) = 1/2$, and $P(2) = 1/2$. (The existence of this probability measure follows from Theorem 1.1.)

It is evident that $B_w = \frac{3}{4}$ and $\|P_n - P\| = (1/2)^n$ and thus $\delta_n^1 \le \frac{3}{4}(1/2)^n$, and it also directly follows that $\delta_n^2 \le \frac{1}{n^2}$. Thus $\delta_n \le \frac{3}{4}(1/2)^n + \frac{1}{n^2}$, and it follows from Theorem 2.1 that the distribution μ_n^x, of $S_n(x)$, converges in the Kantorovich distance (and thus weakly, see Shiryaev (1996)) to μ uniformly on finite subsets of $\mathbb{R}_+$, with a rate of order $O(1/n^2)$. From Theorem 2.2 we also obtain (since $\sum \delta_n < \infty$) a law of large numbers for $\{S_n(x)\}$.

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] Dobrushin, R. L. (1970) Prescribing a system of random variables by conditional distributions, *Theory Probab. Appl.*, **15**, 458–486.

[3] Elton, J. H. (1987) An ergodic theorem for iterated maps, *Ergodic Theory Dynam. Systems*, **7**, 481–488.

[4] Shiryaev, A. N. (1996) *Probability*, second edition, Springer-Verlag, New York.

[5] Stenflo, Ö. (1997) Ergodic theorems for iterated function systems with time dependent probabilities, *Research reports No 18*, Dept. of Mathematics, Umeå University. [To appear in *Theory Stoch. Process.*]

[6] 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