

## Ergodic Theorems for Markov Chains Represented by Iterated Function Systems

by

Örjan STENFLO

*Presented by Andrzej LASOTA on October 21, 1999<sup>(\*)</sup>*

**Summary.** We consider Markov chains represented in the form  $X_{n+1} = f(X_n, I_n)$ , where  $\{I_n\}$  is a sequence of independent, identically distributed (i.i.d.) random variables, and where  $f$  is a measurable function. Any Markov chain  $\{X_n\}$  on a Polish state space may be represented in this form i.e. can be considered as arising from an iterated function system (IFS).

A distributional ergodic theorem, including rates of convergence in the Kantorovich distance is proved for Markov chains under the condition that an IFS representation is "stochastically contractive" and "stochastically bounded".

We apply this result to prove our main theorem giving upper bounds for distances between invariant probability measures for iterated function systems.

We also give some examples indicating how ergodic theorems for Markov chains may be proved by finding contractive IFS representations. These ideas are applied to some Markov chains arising from iterated function systems with place dependent probabilities.

**1. Introduction.** Let  $(X, d)$  be a complete metric space, and let  $S$  be a measurable space. Consider a measurable function  $w : X \times S \rightarrow X$ . For each fixed  $s \in S$ , we write  $w_s(x) := w(x, s)$ . We call the set  $\{X; w_s, s \in S\}$  an iterated function system (IFS). Let  $\{I_n\}_{n=0}^\infty$  be a stochastic sequence with state space  $S$ . 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

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

We call this particular type of stochastic dynamical system an IFS controlled by  $\{I_n\}$ . (Some authors use the name stochastically recursive sequence with

---

1991 MS Classification: primary 60B10, 60F05, 60J05; secondary 58F11.

Key words: ergodic theorems, iterated function systems (IFS), convergence rates, Markov chains.

<sup>(\*)</sup>Revised version received on March 29, 2000.

driver  $\{I_n\}$ . See e.g. Borovkov and Foss [6].) Surveys of IFS related results can be found e.g. in the introduction paper in Stenflo [27] and in Diaconis and Freedman [8].

In this paper, we shall consider the model when the controlling sequence  $\{I_n\}$  is a sequence of independent, identically distributed (i.i.d.) random variables. Any homogeneous Markov chain on a complete, separable metric space can be represented in this form with the i.i.d. random variables,  $\{I_n\}$ , being uniformly distributed in  $(0, 1)$  (see e.g. [14]). A representation, however, is not in general unique. In Section 4 we will describe this in more detail.

In Section 2 we are going to prove a weak ergodic theorem including rate of convergence for Markov chains under a stochastic boundedness condition and an average contraction condition posed on a representing IFS. Similar results of this type can be found in e.g. [11], [21], [5], [26] and [8]. Our generality, also including situations with representations containing an arbitrary number of discontinuous maps on a non-compact state space, is however of importance since in general it cannot be expected that a representation with continuous maps exists.

A main ingredient in the proof of this theorem is the technique of reversing time. This technique has flourished around the last 30–40 years and was formulated as a principle by Letac [17]. In the paper by Propp and Wilson [22] this method was used as a basis for their nowadays widely known algorithm for exact simulation of invariant probability measures of Markov chains. See [8] for a nice overview. Some additional recent papers where this method has been used in order to prove ergodic theorems are e.g. [19], [1], [25] and [26].

In Section 3 we use the result derived in Section 2 to prove our main theorem giving estimates of distances between IFS generated invariant probability measures. Three related results concerning continuity of the invariant measures for iterated function systems can be found in [10], [7], and [20]. See also e.g. [15] and [12] for corresponding results about the invariant sets.

The escape from using a continuity condition in our theorems enables us to give a new approach towards Markov chains generated by iterated function systems with place dependent probabilities. This is done by representing the system by another IFS with place independent probabilities, i.e. an IFS controlled by a sequence of i.i.d. random variables, and use the theorem derived in Section 2. An example of this can be found in Section 4 as well as a new proof for the classical ergodic theorem for Markov chains with “splittable” transition kernels.

**2. Ergodic theorems for IFS controlled by i.i.d. sequences.** Let  $BL$  denote the class of bounded continuous functions,  $f : X \rightarrow \mathbb{R}$  (with

$\|f\|_\infty = \sup_{x \in X} |f(x)| < \infty$ ) that also satisfy the Lipschitz condition

$$\|f\|_L = \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x, y)} < \infty.$$

We set  $\|f\|_{BL} = \|f\|_\infty + \|f\|_L$ . For Borel probability measures  $\nu_1$  and  $\nu_2$  we define the metric

$$d_w(\nu_1, \nu_2) = \sup_{f \in BL} \left\{ \left| \int_X f d(\nu_1 - \nu_2) \right| : \|f\|_{BL} \leq 1 \right\}$$

It is well known, see e.g. [24], that this metric metrizes the topology of weak convergence of probability measures (on separable metric spaces).

Consider now the Kantorovich distance defined by

$$d_k(\nu_1, \nu_2) = \sup \left\{ \left| \int_X f d(\nu_1 - \nu_2) \right| : \|f\|_L \leq 1 \right\}.$$

(See [23] for more on probability metrics.) It is evident that  $d_w(\nu_1, \nu_2) \leq d_k(\nu_1, \nu_2)$ . Denote by  $\mu_n^x(\cdot) = P(Z_n(x) \in \cdot)$ .

We have the following theorem:

**THEOREM 1.** *Suppose*

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

$$Ed(w_{I_0}(x), w_{I_0}(y)) \leq cd(x, y),$$

*for all  $x, y \in X$ .*

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

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

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

*Remark 1.* An explicit expression and upper bound for  $\gamma_K$  is given by

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

*Remark 2.* The limiting probability measure  $\mu$  is concentrated in the sense of a bounded first moment, or to be more precise

$$\int_X d(x_0, x) d\mu(x) \leq \frac{Ed(x_0, w_{I_0}(x_0))}{1-c},$$

where  $c$  and  $x_0$  are defined by the conditions (A) and (B). [See the inequalities in (2) and (6) below.]

*Remark 3.* Obviously, we may replace the  $d_k$ -distance in (1) by the  $d_w$ -distance and thus in particular, we have weak convergence with exponential rate of convergence.

*Remark 4.* Note that the functions  $w_s$  are not assumed to be continuous for any  $s \in S$ . [See the example concerning IFS with place dependent probabilities given in Section 4.2.] In the case when  $S$  is countable, however, continuity is a consequence of condition (A).

**P r o o f.** For  $x \in X$ , define the reversed iterates

$$\hat{Z}_n(x) := w_{I_0} \circ w_{I_1} \circ \cdots \circ w_{I_{n-1}}(x), \quad n \geq 1, \quad \hat{Z}_0(x) = x.$$

The random variables  $\hat{Z}_n(x)$  and  $Z_n(x)$  are identically distributed. We are first going to prove that there exists a random variable  $\hat{Z}$ , such that  $\hat{Z}_n(x)$  converges almost surely (a.s.) to  $\hat{Z}$ . If we then define  $\mu$  by  $\mu(\cdot) = P(\hat{Z} \in \cdot)$ , we have the following sequence of inequalities:

$$\begin{aligned} (2) \quad d_k(\mu_n^x, \mu) &= \sup \left\{ \left| \int_X f d(\mu_n^x - \mu) \right| : \|f\|_L \leq 1 \right\} \\ &= \sup \{ |E(f(\hat{Z}_n(x)) - f(\hat{Z}))| : \|f\|_L \leq 1 \} \\ &\leq \sup \{ E|f(\hat{Z}_n(x)) - f(\hat{Z})| : \|f\|_L \leq 1 \} \\ &\leq E d(\hat{Z}_n(x), \hat{Z}) = E \lim_{m \rightarrow \infty} d(\hat{Z}_n(x), \hat{Z}_m(x)) \\ &\leq E \lim_{m \rightarrow \infty} \sum_{k=n}^{m-1} d(\hat{Z}_k(x), \hat{Z}_{k+1}(x)) \\ &= E \sum_{k=n}^{\infty} d(\hat{Z}_k(x), \hat{Z}_{k+1}(x)) \end{aligned}$$

We shall prove the existence of  $\hat{Z}$  by first proving that  $\{\hat{Z}_n(x)\}$  is a.s. a Cauchy sequence, which converges since  $X$  is complete, and then prove that the limit is independent of  $x$ .

For  $N \leq n \leq m$  we have

$$(3) \quad d(\hat{Z}_n(x), \hat{Z}_m(x)) \leq \sum_{i=N}^{\infty} d(\hat{Z}_i(x), \hat{Z}_{i+1}(x)).$$

Thus if we prove that

$$(4) \quad E \sum_{i=N}^{\infty} d(\hat{Z}_i(x), \hat{Z}_{i+1}(x)) \rightarrow 0 \quad \text{as } N \rightarrow \infty,$$

then

$$(5) \quad \sum_{i=N}^{\infty} d(\hat{Z}_i(x), \hat{Z}_{i+1}(x)) \xrightarrow{\text{a.s.}} 0$$

and from (3) and (5) we conclude that  $\{\hat{Z}_n(x)\}$  a.s. forms a Cauchy sequence.

Now, by recursively using condition (A) we obtain

$$\begin{aligned}
 (6) \quad E \sum_{i=N}^{\infty} d(\hat{Z}_i(x), \hat{Z}_{i+1}(x)) &= \sum_{i=N}^{\infty} E d(\hat{Z}_i(x), \hat{Z}_{i+1}(x)) \\
 &= \sum_{i=N}^{\infty} E(E(d(\hat{Z}_i(x), \hat{Z}_{i+1}(x)) | w_{I_1}, \dots, w_{I_i})) \\
 &= \sum_{i=N}^{\infty} (E(E(d(w_{I_0}(w_{I_1} \circ \dots \circ w_{I_{i-1}}(x)), \\
 &\quad w_{I_0}(w_{I_1} \circ \dots \circ w_{I_i}(x))) | w_{I_1}, \dots, w_{I_i}))) \\
 &\leq \sum_{i=N}^{\infty} c E d(w_{I_1} \circ \dots \circ w_{I_{i-1}}(x), w_{I_1} \circ \dots \circ w_{I_i}(x)) \\
 &\leq \sum_{i=N}^{\infty} c^i E d(x, w_{I_i}(x)) = \frac{c^N}{1-c} E d(x, w_{I_0}(x)).
 \end{aligned}$$

Since using condition (A) and (B),

$$\begin{aligned}
 (7) \quad E d(x, w_{I_0}(x)) &\leq d(x, x_0) + E d(x_0, w_{I_0}(x_0)) + E d(w_{I_0}(x_0), w_{I_0}(x)) \\
 &\leq E d(x_0, w_{I_0}(x_0)) + (c+1)d(x, x_0) < \infty,
 \end{aligned}$$

it follows that (4) holds and thus  $\{\hat{Z}_n(x)\}$  converges a.s. to some random element  $\hat{Z}(x)$  for each  $x \in X$ . It remains to prove that the limit is independent of  $x$ .

By the Chebyshev inequality, and by a recursive use of condition (A), for any two points  $x$  and  $y$  in  $X$ , and for any  $\epsilon > 0$ ,

$$\begin{aligned}
 P(d(\hat{Z}_n(x), \hat{Z}_n(y)) > \epsilon) &\leq \frac{E d(\hat{Z}_n(x), \hat{Z}_n(y))}{\epsilon} \\
 &\leq \frac{1}{\epsilon} E(E(d(\hat{Z}_n(x), \hat{Z}_n(y)) | w_{I_1}, \dots, w_{I_{n-1}}))
 \end{aligned}$$

$$\begin{aligned} &\leq \frac{c}{\epsilon} Ed(w_{I_1} \circ \dots \circ w_{I_{n-1}}(x), w_{I_1} \circ \dots \circ w_{I_{n-1}}(y)) \\ &\leq \dots \leq \frac{c^n}{\epsilon} d(x, y). \end{aligned}$$

Thus

$$\sum_{n=0}^{\infty} P(d(\hat{Z}_n(x), \hat{Z}_n(y)) > \epsilon) \leq \sum_{n=0}^{\infty} \frac{c^n}{\epsilon} d(x, y) < \infty,$$

and it follows (see e.g. [24]) that

$$(8) \quad d(\hat{Z}_n(x), \hat{Z}_n(y)) \rightarrow 0 \quad \text{a.s.}$$

Define  $\hat{Z} = \hat{Z}(x_0)$ . From (8), the triangle inequality, and the fact of almost sure convergence of  $\hat{Z}_n(x_0)$  to  $\hat{Z}$ , it follows that for any  $x \in X$ ,  $d(\hat{Z}_n(x), \hat{Z}) \xrightarrow{\text{a.s.}} 0$  as  $n \rightarrow \infty$  establishing the a.s. independence of  $x$ .

Combining (2) and (6) we see that

$$d_k(\mu_n^x, \mu) \leq \frac{c^n}{1-c} Ed(x, w_{I_0}(x)), \quad n \geq 0.$$

Thus,

$$\sup_{x \in K} d_k(\mu_n^x, \mu) \leq \frac{c^n}{1-c} \gamma_K, \quad n \geq 0,$$

where  $\gamma_K := \sup_{x \in K} Ed(x, w_{I_0}(x))$  is a finite constant, since, by taking suprema in (7)

$$(9) \quad \gamma_K \leq Ed(x_0, w_{I_0}(x_0)) + (c+1) \sup_{x \in K} d(x, x_0) < \infty.$$

It remains to prove that the probability measure  $\mu$  is invariant and unique with this property. To do this, we prove that the Markov chain  $\{Z_n(x)\}$  has the (weak) Feller property which in our terminology means that  $g : X \rightarrow \mathbb{R}$  being a bounded and continuous function implies that the mapping

$$(10) \quad x \mapsto Eg(w_{I_0}(x))$$

is continuous. It is well known that the limiting probability measure of an ergodic Markov chain with the Feller property is invariant. To be self-contained, we explain why before proving that our Markov chain has this property.

Since

$$(11) \quad Eg(Z_n(x)) = \int_X Eg(w_{I_0}(y)) P(Z_{n-1}(x) \in dy),$$

the invariance equation

$$\int_X g d\mu = \int_X Eg(w_{I_0}(y)) d\mu(y)$$

will then follow by taking limits in (11) justified by using the continuity in (10).

We shall now prove that our Markov chain has the Feller property. Let  $\{y_n\}$  be a sequence in  $X$  with  $\lim_{n \rightarrow \infty} y_n = y$ . Since, for fixed  $\epsilon > 0$ , by the Chebyshev inequality, and from condition (A),

$$P(d(w_{I_0}(y_n), w_{I_0}(y)) > \epsilon) \leq \frac{Ed(w_{I_0}(y_n), w_{I_0}(y))}{\epsilon} \leq \frac{cd(y_n, y)}{\epsilon} \rightarrow 0$$

as  $n \rightarrow \infty$ , we have proved that  $w_{I_0}(y_n)$  converges in probability to  $w_{I_0}(y)$ . Thus for any bounded and continuous function  $g$

$$\lim_{n \rightarrow \infty} Eg(w_{I_0}(y_n)) = Eg(w_{I_0}(y)),$$

and the Feller property is established. Thus  $\mu$  is invariant.

The uniqueness follows since if  $\nu$  is another invariant measure, we obtain by using Lebesgue's convergence theorem, that

$$\begin{aligned} \int_X g d\nu &= \int_X Eg(Z_1(x)) d\nu(x) = \dots = \int_X Eg(Z_n(x)) d\nu(x) \\ &\rightarrow \int_X \left( \int_X g d\mu \right) d\nu = \int_X g d\mu, \end{aligned}$$

for bounded and continuous functions  $g$ , and thus  $\mu$  and  $\nu$  coincides (see e.g. [3]). This completes the proof of Theorem 1.  $\square$

**3. Estimation of distances between IFS generated probability measures.** We will now turn to our main theorem proving that under uniform contractivity and stochastic boundedness assumptions (condition (C) and (D) below), we can give upper bounds for  $d_w$ -distances between IFS generated probability measures and in particular prove that the limiting probability measure depends "continuously" on the parameters in the system.

For probability measures  $\nu_1$  and  $\nu_2$  defined on the same measurable space,  $(M, M)$ , let  $d_{TV}$  denote the total variation distance defined by

$$d_{TV}(\nu_1, \nu_2) = \sup_{A \in M} |\nu_1(A) - \nu_2(A)|.$$

**THEOREM 2.** Let  $\mathcal{F}^\epsilon = \{X; w_s^\epsilon, s \in S\}$ , with  $\epsilon \in [0, t]$  for some  $t > 0$ , be an indexed family of iterated function systems, and let  $\{I_n^\epsilon\}$ , respectively, be associated i.i.d. controlling sequences. Define for  $x \in X$ ,

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

Suppose

(C) There exists a constant  $0 < c < 1$  such that

$$Ed(Z_1^\epsilon(x), Z_1^\epsilon(y)) \leq cd(x, y), \quad \text{for all } x, y \in X, \quad \text{and all } \epsilon \in [0, t].$$

- (D)  $\gamma_{x_0} := \sup_{\epsilon \in [0, t]} Ed(x_0, w_{I_0^\epsilon}(x_0)) < \infty$ , for some  $x_0 \in X$ .  
 (E) There exists a function  $\Delta : [0, t] \rightarrow \mathbb{R}_+$  such that,

$$\sup_{s \in S} \sup_{x \in X} d(w_s^\epsilon(x), w_s^0(x)) \leq \Delta(\epsilon) \quad \text{for all } \epsilon \in [0, t].$$

- (F) There exists a function  $S : [0, t] \rightarrow \mathbb{R}_+$  such that

$$d_{TV}(P_\epsilon, P_0) \leq \gamma_{x_0} S(\epsilon), \quad \text{for all } \epsilon \in [0, t],$$

where  $P_\epsilon(\cdot) := P(I_0^\epsilon \in \cdot)$ , and where  $\gamma_{x_0}$  is the constant defined by condition (D).

Let  $\mu^\epsilon$  denote the limiting invariant probability distribution for the Markov chain generated by the IFS and controlling sequence indexed by  $\epsilon$  (these measures exist due to Theorem 1). Then for any  $\epsilon \in [0, t]$ , we have

$$(12) \quad d_w(\mu^\epsilon, \mu^0) \leq \begin{cases} \frac{\Delta(\epsilon) + 2\gamma_{x_0} S(\epsilon)}{1 - c} + \frac{2\gamma_{x_0} S(\epsilon) \ln S(\epsilon)}{\ln c}, & S(\epsilon) < 1 \\ \frac{\Delta(\epsilon) + 2\gamma_{x_0}}{1 - c}, & S(\epsilon) \geq 1. \end{cases}$$

*Remark 5.* If  $S(\epsilon) = 0$  we interpret (12) as  $d_w(\mu^\epsilon, \mu^0) \leq \frac{\Delta(\epsilon)}{1 - c}$ .

Before we turn to the proof of Theorem 2, we illustrate the theorem with an example.

*Example 1.* Consider the family of iterated function systems  $\mathcal{F}^\epsilon = \{[0, 1], w_1^\epsilon(x) = (1/2 - \epsilon)x, w_2^\epsilon(x) = (1/2 - \epsilon)x + 1/2 + \epsilon\}$ ,  $0 \leq \epsilon \leq 1/2$ , with  $P(I_0^\epsilon = i) = 1/2$ , (i.e. independent of  $\epsilon$ ) for  $i = 1, 2$ . Applying the above theorem, with  $c = 1/2$ ,  $\Delta(\epsilon) = \epsilon$ , and  $S(\epsilon) = 0$ , shows that  $d_w(\mu^\epsilon, \mu^0) \leq 2\epsilon$  and thus  $\mu^\epsilon \rightarrow \mu^0$  as  $\epsilon \rightarrow 0$  weakly which at first glance may be somewhat conspicuous since we know that the supports of  $\mu^\epsilon$ , for  $0 < \epsilon < 1/2$ , are sets of Cantor type, while  $\mu^0$  is Lebesgue measure on  $[0, 1]$ .

We now turn to the proof of Theorem 2.

**P r o o f.** Define  $\mu_n^{\epsilon, x}(\cdot) := P(Z_n^\epsilon(x) \in \cdot)$ , for  $x \in X$ , and let  $x_0$  be the point defined in condition (D). By the triangle inequality we have that

$$(13) \quad d_w(\mu^\epsilon, \mu^0) \leq d_w(\mu^\epsilon, \mu_n^{\epsilon, x_0}) + d_w(\mu_n^{\epsilon, x_0}, \mu_n^{0, x_0}) + d_w(\mu_n^{0, x_0}, \mu^0).$$

Putting in the assertion of Theorem 1,  $K = \{x_0\}$  and applying conditions (C) and (D) we get

$$(14) \quad d_w(\mu_n^{\epsilon, x_0}, \mu^\epsilon) \leq \frac{\gamma_{x_0}}{1 - c} c^n, \quad \text{for all } n \geq 0, \quad \text{and } \epsilon \in [0, t].$$

Define, for  $x \in X$ ,

$$\tilde{Z}_n^\epsilon(x) := w_{I_{n-1}^0}^\epsilon \circ w_{I_{n-2}^0}^\epsilon \circ \cdots \circ w_{I_0^0}^\epsilon(x), \quad n \geq 1, \quad \tilde{Z}_0^\epsilon(x) = x.$$



Let for each fixed  $\epsilon$  and  $n$ ,  $P_\epsilon^{(n)}$  denote the probability distribution of the random vector  $\{I_0^\epsilon, I_1^\epsilon, \dots, I_{n-1}^\epsilon\}$ . ( $P_\epsilon^{(1)} = P_\epsilon$ )

We have the following inequalities,

$$\begin{aligned}
 (15) \quad d_w(\mu_n^{\epsilon, x_0}, \mu_n^{0, x_0}) &= \sup_{f \in BL} \left\{ \left| \int_X f d(\mu_n^{\epsilon, x_0} - \mu_n^{0, x_0}) \right| : \|f\|_{BL} \leq 1 \right\} \\
 &= \sup_{f \in BL} \{ |Ef(Z_n^\epsilon(x_0)) - Ef(Z_n^0(x_0))| : \|f\|_{BL} \leq 1 \} \\
 &\leq \sup_{f \in BL} \{ |Ef(Z_n^\epsilon(x_0)) - Ef(\tilde{Z}_n^\epsilon(x_0))| : \|f\|_{BL} \leq 1 \} \\
 &\quad + \sup_{f \in BL} \{ |Ef(\tilde{Z}_n^\epsilon(x_0)) - Ef(Z_n^0(x_0))| : \|f\|_{BL} \leq 1 \} \\
 &\leq 2d_{TV}(P_\epsilon^{(n)}, P_0^{(n)}) + Ed(\tilde{Z}_n^\epsilon(x_0), Z_n^0(x_0)).
 \end{aligned}$$

For any fixed  $\epsilon$  and  $n$  we may assume (see [9]) that  $I_n^\epsilon$  and  $I_n^0$  are defined on the same probability space with  $P(I_n^\epsilon \neq I_n^0) = d_{TV}(P_\epsilon, P_0)$ . It can also be assumed that  $\{(I_n^\epsilon, I_n^0)\}$  is a sequence of i.i.d. random variables. Thus

$$\begin{aligned}
 (16) \quad d_{TV}(P_\epsilon^{(n)}, P_0^{(n)}) &\leq P((I_0^\epsilon, \dots, I_{n-1}^\epsilon) \neq (I_0^0, \dots, I_{n-1}^0)) \\
 &= P\left(\bigcup_{i=0}^{n-1} \{I_i^\epsilon \neq I_i^0\}\right) \leq \sum_{i=0}^{n-1} P(I_i^\epsilon \neq I_i^0) \\
 &= \sum_{i=0}^{n-1} d_{TV}(P_\epsilon, P_0) \leq n\gamma_{x_0} S(\epsilon).
 \end{aligned}$$

Studying the second term appearing in (15), we obtain

$$\begin{aligned}
 (17) \quad Ed(\tilde{Z}_n^\epsilon(x_0), Z_n^0(x_0)) &\leq Ed(w_{I_{n-1}^0}^\epsilon(\tilde{Z}_{n-1}^\epsilon(x_0)), w_{I_{n-1}^0}^0(Z_{n-1}^0(x_0))) \\
 &\leq Ed(w_{I_{n-1}^0}^\epsilon(\tilde{Z}_{n-1}^\epsilon(x_0)), w_{I_{n-1}^0}^0(\tilde{Z}_{n-1}^\epsilon(x_0))) \\
 &\quad + Ed(w_{I_{n-1}^0}^0(\tilde{Z}_{n-1}^\epsilon(x_0)), w_{I_{n-1}^0}^0(Z_{n-1}^0(x_0))) \\
 &\leq \sup_{s \in S} \sup_{x \in X} d(w_s^\epsilon(x), w_s^0(x)) \\
 &\quad + E(E(d(w_{I_{n-1}^0}^0(\tilde{Z}_{n-1}^\epsilon(x_0)), w_{I_{n-1}^0}^0(Z_{n-1}^0(x_0))) | \tilde{Z}_{n-1}^\epsilon(x_0), Z_{n-1}^0(x_0))) \\
 &\leq \Delta(\epsilon) + cEd(\tilde{Z}_{n-1}^\epsilon(x_0), Z_{n-1}^0(x_0)),
 \end{aligned}$$

and thus by a recursive use of (17), we see that

$$(18) \quad Ed(\tilde{Z}_n^\epsilon(x_0), Z_n^0(x_0)) \leq \Delta(\epsilon) \sum_{i=0}^{n-1} c^i \leq \Delta(\epsilon) \frac{1}{1-c}.$$

By inserting (16) and (18) in (15) we obtain that

$$(19) \quad d_w(\mu_n^{\epsilon, x_0}, \mu_n^{0, x_0}) \leq 2n\gamma_{x_0}S(\epsilon) + \Delta(\epsilon)\frac{1}{1-c},$$

and thus from (13), (14), and (19) we see that

$$(20) \quad d_w(\mu^\epsilon, \mu^0) \leq \min_{n \geq 0} \left( \frac{2\gamma_{x_0}}{1-c}c^n + 2n\gamma_{x_0}S(\epsilon) \right) + \Delta(\epsilon)\frac{1}{1-c}.$$

In order to give a more explicit expression for the right hand side of (20) we investigate the function  $f(x) = ax + bc^x$ ,  $x > 0$ , where  $a, b > 0$ ,  $0 < c < 1$ . Suppose that  $a < b(1-c)$ . For  $x = (1/\ln c) \ln(a/(b(1-c))) > 0$ , we have that

$$f(x) = f(x+1) = \frac{a \ln(a/(b(1-c)))}{\ln c} + \frac{a}{1-c}.$$

Since  $f$  is convex it follows that for the unique integer  $n$  such that  $x \leq n < x+1$  we have that  $f(n) \leq f(x)$ .

Thus with  $a = 2\gamma_{x_0}S(\epsilon)$  and  $b = 2\gamma_{x_0}/(1-c)$ , we see that if  $S(\epsilon) < 1$  then

$$\min_{n \geq 0} \left( \frac{2\gamma_{x_0}}{1-c}c^n + 2n\gamma_{x_0}S(\epsilon) \right) \leq \frac{2\gamma_{x_0}S(\epsilon)}{1-c} + \frac{2\gamma_{x_0}S(\epsilon) \ln S(\epsilon)}{\ln c}.$$

(If  $S(\epsilon) \geq 1$  we use the trivial bound  $\min_{n \geq 0} ((2\gamma_{x_0}/(1-c))c^n + 2n\gamma_{x_0}S(\epsilon)) \leq 2\gamma_{x_0}/(1-c)$ .)

Using these estimates in (20) completes the proof of the theorem.  $\square$

**4. IFS representations of Markov chains.** In this section, we are going to describe how a Markov chain,  $\{X_n\}$  with state space  $X$ , may be represented as an IFS controlled by a sequence of i.i.d. random variables. (We shall call such a representation an IFS representation.)

Our aim is to prepare for Sections 4.1 and 4.2 where we apply the ergodic theorem for IFS controlled by i.i.d. sequences (developed in Section 2) to prove ergodic theorems for Markov chains by finding suitable IFS representations.

We shall start with an example where the state space  $X$  is as simple (non-trivial) as possible.

One purpose of this example is to illustrate the non-uniqueness of an IFS representation of a given Markov chain.

*Example 2.* Consider a time homogeneous Markov chain  $\{X_n\}$  with state space  $X = \{0, 1\}$  and transition matrix

$$P = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}.$$

To find all IFS representations of this Markov chain, we take the four possible functions  $w_1(x) = x$ ,  $w_2(x) = 0$ ,  $w_3(x) = 1$ ,  $w_4(x) = 1 - x$ , and let  $\{I_n\}$  be i.i.d. with  $P(I_n = i) = p_i$ , for  $i = 1, \dots, 4$ . We obtain the following system of linear equations:

$$\begin{cases} p_1 + p_2 = p_{00} \\ p_1 + p_3 = p_{11} \\ p_1 + p_2 + p_3 + p_4 = 1 \\ 0 \leq p_1, p_2, p_3, p_4 \leq 1 \end{cases}$$

Solving this system for  $p_i$ ,  $i = 1, \dots, 4$ , we finally get

$$\begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix} = \begin{pmatrix} 0 \\ p_{00} \\ p_{11} \\ 1 - (p_{00} + p_{11}) \end{pmatrix} + t \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix},$$

$$\max(0, p_{00} + p_{11} - 1) \leq t \leq \min(p_{00}, p_{11}).$$

In this example we made a total investigation finding all representing iterated function systems. In Section 2 we proved a theorem for iterated function systems based on contraction conditions of the involved functions. Therefore we see that the representation above with parameter  $t$  as small as possible is the best representation provided that we want to choose the contractive functions ( $w_2$  and  $w_3$  in the above example) with as high probability as possible.

When the state space is larger however, e.g.  $\mathbb{R}$ , we can no longer make this type of total investigation for representing iterated function systems. Therefore we will be pleased if we can find an algorithm constructing *one* "contractive" IFS from a given Markov chain.

If  $\{X_n\}$  is a Markov chain on (a subset of)  $\mathbb{R}$  with transition kernel  $P(x, A)$ , then we can define a representing IFS with,

$$w_s(x) = \inf\{y : P(x, (-\infty, y]) \geq s\}$$

and with  $\{I_n\}$  being an independent sequence of random variables with values uniformly distributed on  $(0, 1)$ . This representation corresponds to the "most contractive" IFS in the above example. Note however that this is not a general statement for larger state spaces since the above construction depends on the natural ordering of  $\mathbb{R}$ .

The above algorithm for creating a representing IFS can be generalized if the state space,  $X$ , is Borel measurably isomorphic to a Borel subset of the real line,  $\mathbb{R}$ , satisfied for instance if  $X$  is a Polish space.

In fact (see e.g. [14]), suppose there exists a one-to-one Borel map  $\phi : X \rightarrow \mathbb{R}$  such that  $M = \phi(X)$  is a Borel subset of  $\mathbb{R}$  and  $\phi^{-1} : M \rightarrow X$  is also Borel measurable. Suppose that  $\psi : \mathbb{R} \rightarrow X$  equals  $\phi^{-1}$  on  $M$  and maps  $\mathbb{R} \setminus M$  on some point  $x_f \in X$ . For each  $x \in X$  define a probability measure on the Borel  $\sigma$ -field  $B(\mathbb{R})$  of  $\mathbb{R}$  by  $\tilde{P}(x, B) = P(x, \phi^{-1}(B \cap M))$ , for  $B \in B(\mathbb{R})$ ,

and for each  $x \in X$  and  $s \in (0, 1)$  let  $q_s(x) = \inf\{y : \tilde{P}(x, (-\infty, y]) \geq s\}$ . If we for each  $s \in (0, 1)$  let  $w_s = \psi \circ q_s$  then the construction is completed.

To know whether or not there exists a representing IFS with continuous functions, one sufficient condition is given in Blumenthal and Corson [4]. Their condition is that the state space  $X$  is a connected and locally connected compact metric space, the transition kernel  $P$  has the (weak) Feller property, (i.e. the operator  $T$  defined by  $Tf(x) := \int_X P(x, dy) f(y)$ , for all  $x$ , maps bounded and continuous functions to itself), and for each fixed  $x \in X$ , the support of  $P(x, \cdot)$  is all of  $X$ . See also [28] for further results.

**4.1. Recurrent Markov chains.** We shall now as a corollary of Theorem 1 present a proof of ergodicity for Markov chains satisfying the classical Doeblin "splitting" condition. Although the result is more than fifty years old and many different proofs exist in the literature, see e.g. [18], the proof given here is non-standard and may thus be of complementary interest. The proof also awakes the general open question on how to embed the theory of exponentially ergodic Markov chains within the theory of iterated function systems.

**COROLLARY 1.** *Suppose  $\{Z_n\}$  is a homogeneous Markov chain with Polish state space  $X$  and with a transition kernel  $P$  which has the splitting property:*

$$P^{n_0}(x, \cdot) = \beta Q(x, \cdot) + (1 - \beta)\nu(\cdot),$$

*for some  $n_0 \geq 1$ , transition kernel  $Q$  and probability measure  $\nu$ , and for every  $x \in X$ , where  $0 \leq \beta < 1$ . Then there exists a unique invariant probability measure  $\mu$  for  $\{Z_n\}$ , such that*

$$(21) \quad \sup_{x \in X} \sup_{A \in B(X)} |P(Z_n(x) \in A) - \mu(A)| \leq \frac{\beta^{[n/n_0]}}{1 - \beta},$$

*where  $B(X)$  denotes the class of Borel sets in  $X$ , and  $Z_n(x)$  denotes the Markov chain with initial distribution concentrated at  $x \in X$ .*

**P r o o f.** It is sufficient to prove this theorem for the case  $n_0 = 1$ , since we can consider subsequences  $\{Z_{nn_0}\}_{n=0}^\infty$ , and if the Markov chain  $\{Z_{nn_0}\}_{n=0}^\infty$  satisfies (21) then for  $k = 0, 1, 2, \dots, n_0 - 1$  we observe that

$$\begin{aligned} & \sup_{x \in X} \sup_{A \in B(X)} |P(Z_{nn_0+k}(x) \in A) - \mu(A)| \\ & \leq \sup_{x \in X} \sup_{A \in B(X)} \left| \int_X (P(Z_{nn_0}(y) \in A) - \mu(A)) P^k(x, dy) \right| \\ & \leq \sup_{x \in X} \int_X \sup_{A \in B(X)} |P(Z_{nn_0}(y) \in A) - \mu(A)| P^k(x, dy) \leq \frac{\beta^n}{1 - \beta}, \end{aligned}$$

and the conclusion of the theorem holds.

Using the algorithm described in Section 4, let  $f_s, s \in (0, 1)$  be a set of functions representing a Markov chain with transition kernel  $Q$  together with a sequence  $\{I'_n\}$  of i.i.d. random variables with values in  $(0, 1)$ . Furthermore, let  $g_s, s \in (0, 1)$  be a set of functions representing a Markov chain with transition kernel (measure)  $\nu$  (together with  $\{I'_n\}$ ). Let  $\{I''_n\}$  be another (independent) such i.i.d. sequence.

Then  $\{I_n\}$ , with  $I_n = (I'_n, I''_n)$  forms an independent sequence uniformly distributed in  $(0, 1) \times (0, 1)$ . If we define  $w_{s,t} = f_s$  for  $0 < t \leq \beta$  and  $g_s$  otherwise, we obtain

$$w_{I_n} = \chi(I''_n \leq \beta) f_{I'_n} + \chi(I''_n > \beta) g_{I'_n},$$

where  $\chi$  denotes the indicator function.

Let  $d$  denote the discrete metric. The space  $(X, d)$  then constitutes a complete metric space.

Since  $g_s, s \in (0, 1)$  will all be constant maps, it follows that

$$\begin{aligned} Ed(w_{I_0}(x), w_{I_0}(y)) \\ \leq \beta Ed(f_{I'_0}(x), f_{I'_0}(y)) + (1 - \beta) Ed(g_{I'_0}(x), g_{I'_0}(y)) \leq \beta d(x, y). \end{aligned}$$

Thus  $\{(X, d), w_{s,t}, (s, t) \in (0, 1) \times (0, 1)\}$  together with  $\{I_n\}$  forms a contractive IFS representing the transition kernel  $P$  and Theorem 1 can be used. In fact, if  $\mathcal{C} = \{f : X \rightarrow \mathbb{R}; f = \chi \text{ for some indicator function } \chi\}$  then from Theorem 1, using  $\gamma_K \leq 1$ ,

$$\begin{aligned} \sup_{A \in \mathcal{B}(X)} |P(Z_n(x) \in A) - \mu(A)| \\ = \sup_{f \in \mathcal{C}} \left\{ \left| \int_X f d\mu_n^x - \int_X f d\mu \right| \right\} \\ \leq \sup \left\{ \left| \int_X f d\mu_n^x - \int_X f d\mu \right| : \|f\|_L \leq 1 \right\} \leq \frac{\beta^n}{1 - \beta}, \end{aligned}$$

and the above inequality holds uniformly for all  $x \in X$ . This completes the proof of the theorem.  $\square$

**4.2. Iterated function systems with place dependent probabilities.** If we have an IFS with  $S = \{1, \dots, N\}$ , for some  $N \geq 1$ , and to each  $i \in S$  we have associated probability weights  $p_i : X \rightarrow [0, 1]$ ,  $p_i(x) \geq 0$ ,  $i \in S$  and  $\sum_{i=1}^N p_i(x) = 1$ , for each  $x \in X$ , we call the set  $\{X; w_i, p_i(x), i \in S\}$  an IFS with place dependent probabilities.

Specify a point  $x \in X$ . Using this system we can construct a Markov chain  $\{Z_n(x)\}$  in the following way: Put  $Z_0(x) := x$ , and let  $Z_n(x) := w_i(Z_{n-1}(x))$  with probability  $p_i(Z_{n-1}(x))$ , for each  $n \geq 1$ . Some papers where this model is considered are [2], [13], and [16].

For any IFS with place dependent probabilities, there is an IFS with place *independent* probability weights (i.e. an IFS controlled by an i.i.d. sequence) generating the same Markov chain. (We call iterated function systems generating the same Markov chain equivalent.)

We illustrate this with the following example.

*Example 3.* Consider the IFS with place dependent probabilities  $\{X; w_i(x), p_i(x), i \in \{1, 2\}\}$ , with  $w_i, i = 1, 2$ , being continuous.

The IFS  $\{X; f_s, s \in (0, 1)\}$  with

$$f_s(x) = \begin{cases} w_1(x) & \text{if } p_1(x) \geq s, x \in X, \\ w_2(x) & \text{otherwise,} \end{cases}$$

controlled by a sequence of independent random variables, uniformly distributed in  $(0, 1)$ , is equivalent with the above system. It is more well behaved in the sense that it has place independent probabilities but the loss is that it generally has a denumerable set of discontinuous functions.

As corollaries of Theorem 1 we obtain:

**COROLLARY 2.** *Let  $w_1$  and  $w_2$  be bounded contractions i.e. functions satisfying the Lipschitz conditions  $d(w_i(x), w_i(y)) \leq cd(x, y)$ ,  $c < 1$ , for all  $x, y \in X$ , and  $i = 1, 2$ , with  $b := \sup_{x, y \in X} d(w_1(x), w_2(y)) < \infty$ .*

*Suppose, for some  $r > 1$  and all  $x, y \in X$ ,*

$$|p_1(x) - p_1(y)| \leq \frac{1-c}{rb} d(x, y).$$

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

$$\sup_{x \in K} d_w(\mu_n^x, \mu) \leq \gamma \left( c + \frac{1-c}{r} \right)^n$$

*where  $\mu_n^x(\cdot) := P(Z_n(x) \in \cdot)$ .*

**Proof.** Take the representing IFS  $\{X; f_s, s \in (0, 1)\}$  constructed as in Example 3 above. We are going to use Theorem 1. We thus have to check condition (A). (Condition (B) trivially holds.)

Now, for  $x, y \in X$  we may suppose that  $p_1(x) \leq p_1(y)$  and thus

$$\begin{aligned} Ed(f_{I_0}(x), f_{I_0}(y)) &\leq p_1(x)d(w_1(x), w_1(y)) \\ &\quad + (p_1(y) - p_1(x))d(w_2(x), w_1(y)) \\ &\quad + (1 - p_1(y))d(w_2(x), w_2(y)) \\ &\leq d(x, y)(cp_1(x) + \frac{1-c}{rb}b + c(1 - p_1(y))) \end{aligned}$$

$$\begin{aligned} &\leq d(x, y) \left( c(p_1(x) + 1 - p_1(y)) + \frac{1-c}{r} \right) \\ &\leq \left( c + \frac{1-c}{r} \right) d(x, y). \end{aligned}$$

The conditions in Theorem 1 are satisfied and thus there exists a probability measure  $\mu$  such that, for any bounded set  $K \subseteq X$  there exists a positive constant  $\gamma$  such that

$$\sup_{x \in K} d_w(\mu_n^x, \mu) \leq \gamma \left( c + \frac{1-c}{r} \right)^n \quad \square$$

**COROLLARY 3.** *Consider the functions  $w_1(x) = cx + c_1$  and  $w_2(x) = cx + c_2$ , with  $c < 1$ , on a compact subset  $K$  of  $\mathbb{R}$  (together with the Euclidean metric).*

*Suppose, the probability weights are affine i.e.  $p_1(x) = p_1 + c_3x$ , for some constants  $p_1$  and  $c_3$ . Denote by  $L := c + |c_3||c_1 - c_2|$ . If  $L < 1$  then there exists a unique invariant probability measure  $\mu$  for the Markov chain  $\{Z_n(x)\}$  and a positive constant  $\gamma$  such that*

$$\sup_{x \in K} d_w(\mu_n^x, \mu) \leq \gamma L^n$$

where  $\mu_n^x(\cdot) := P(Z_n(x) \in \cdot)$ .

**PROOF.** Take the representing IFS  $\{X; f_s, s \in (0, 1)\}$  constructed as in Example 3 above. We are going to use Theorem 1. We thus have to check condition (A). (Condition (B) trivially holds.)

Now, (for  $p_1(x) \leq p_1(y)$ )

$$\begin{aligned} E|f_{I_0}(x) - f_{I_0}(y)| &\leq p_1(x)c|x - y| + (p_1(y) - p_1(x))(cx + c_2 - (cy + c_1)) \\ &\quad + (1 - p_1(y))c|x - y| \\ &\leq c|x - y|(1 - (p_1(y) - p_1(x))) \\ &\quad + (p_1(y) - p_1(x))(c|x - y| + |c_1 - c_2|) \\ &= c|x - y| + (p_1(y) - p_1(x))|c_1 - c_2| \\ &\leq |x - y|(c + |c_3||c_1 - c_2|) = L|x - y|. \end{aligned}$$

The conditions in Theorem 1 are satisfied and thus the conclusion follows.  $\square$

**Acknowledgements.** A version of this paper was included as a part of the authors Ph.D. thesis. The author is grateful to his supervisors Dmitrii S. Silvestrov and Hans Wallin for their support and useful comments.

## REFERENCES

- [1] A. Ambroladze, *Ergodic properties of random iterations of analytic functions*, Ergodic Theory Dynam. Systems, **19** (1999) 1379–1388.
- [2] M. F. Barnsley, S. G. Demko, J. H. Elton, J. S. Geronimo, *Invariant measures for Markov processes arising from iterated function systems with place-dependent probabilities*, Ann. Inst. H. Poincaré Probab. Statist., **24** (1988) 367–394.
- [3] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York 1968.
- [4] R. M. Blumenthal, H. H. Corson, *On continuous collections of measures*, Proc. 6th Berkely Sympos. on Math. Stat. and Prob., **2** (1972) 33–40.
- [5] A. A. Borovkov, *Ergodicity and stability of stochastic processes*, Wiley, Chichester 1998.
- [6] A. A. Borovkov, S. G. Foss, *Two ergodicity criteria for stochastically recursive sequences*, Acta Appl. Math., **34** (1994) 125–134.
- [7] P. M. Centore, E. R. Vrscaj, *Continuity of attractors and invariant measures for iterated function systems*, Canad. Math. Bull., **37** (1994) 315–329.
- [8] P. Diaconis, D. Freedman, *Iterated random functions*, SIAM Rev., **41** (1999) 45–76.
- [9] R. L. Dobrushin, *Prescribing a system of random variables by conditional distributions*, Theory Probab. Appl., **15** (1970) 458–486.
- [10] L. Dubins, D. Freedman, *Invariant probabilities for certain Markov processes*, Ann. Math. Statist., **37** (1966) 837–844.
- [11] M. Iosifescu, *A basic tool in mathematical chaos theory: Doeblin and Fortet's ergodic theorem and Ionescu Tulcea and Marinescu's generalization. Doeblin and modern probability (Blaubeuren, 1991)*, Contemp. Math., **149** (1993) 111–124.
- [12] J. Jachymski, *Continuous dependence of attractors of iterated function systems*, J. Math. Anal. Appl., **198** (1996) 221–226.
- [13] T. Kaijser, *On a theorem of Karlin*, Acta Appl. Math., **34** (1994) 51–69.
- [14] Y. Kifer, *Ergodic theory of random transformations*, Birkhäuser, Boston 1986.
- [15] M. Kwieciński, *A note on continuity of fixed points*, Univ. Iagel. Acta Math., **29** (1992) 19–24.
- [16] A. Lasota, J. A. Yorke, *Lower bound technique for Markov operators and iterated function systems*, Random Comput. Dynam., **2** (1994) 41–77.
- [17] G. Letac, *A contraction principle for certain Markov chains and its applications*, Contemp. Math., **50** (1986) 263–273.
- [18] M. Loève, *Probability Theory II*, 4th Edition, Springer-Verlag, New York 1978.
- [19] K. Łoskot, R. Rudnicki, *Limit theorems for stochastically perturbed dynamical systems*, J. Appl. Probab., **32** (1995) 459–469.
- [20] F. Mendivil, *A generalization of ifs with probabilities to infinitely many maps*, Rocky Mountain J. Math., **28** (1998) 1043–1051.
- [21] S. P. Meyn, R. L. Tweedie, *Markov chains and stochastic stability*, Springer-Verlag, London 1993.
- [22] J. G. Propp, D. B. Wilson, *Exact sampling with coupled Markov chains and applications to statistical mechanics*, Random Structures Algorithms, **9** (1996) 223–252.
- [23] S. T. Rachev, *Probability metrics*, Wiley, Chichester 1991.
- [24] A. N. Shiryaev, *Probability*, Second Edition, Springer-Verlag, New York 1996.
- [25] D. S. Silvestrov, Ö. Stenflo, *Ergodic theorems for iterated function systems controlled by regenerative sequences*, J. Theoret. Probab., **11** (1998) 589–608.



- [26] D. Steinsaltz, *Locally contractive iterated function systems*, Ann. Probab., **27** (1999) 1952–1979.
- [27] Ö. Stenflo, *Ergodic theorems for iterated function systems controlled by stochastic sequences*, Ph.D. Thesis No 14, Department of Mathematics, Umeå University, 1998. (Available at <http://www.math.umu.se/~stenflo/avhandling.html>)
- [28] A. N. Quas, *On representations of Markov chains by random smooth maps*, Bull. London Math. Soc., **23** (1991) 487–492.