Ergodic Theorems for Markov Chains represented by Iterated Function Systems

Örjan Stenflo

Abstract. 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 give 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 $w : X \times S \to X$. 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

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

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

1991 Mathematics Subject Classification: Primary 60B10, 60F05, 60J05; Secondary 58F11.

We call this particular type of stochastic dynamical system an IFS controlled by $\{I_n\}$. (Some authors use the name stochastically recursive sequence with driver $\{I_n\}$. See e.g. Borovkov and Foss (1994).)

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 chains 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. Kifer (1986)). 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.

A main ingredient in the proof of this theorem is the method of reversing time. This method was introduced by Letac (1986) and has been used in e.g. Burton and Rösler (1995), Łoskot and Rudnicki (1995), Ambroladze (1997) and Silvestrov and Stenflo (1998) in order to prove ergodic theorems.

In Section 3 we use the result derived in Section 2 to give estimates of distances between IFS generated invariant probability measures. A related result concerning continuity of the invariant measures for iterated function systems is given in Centore and Vrscay (1994).

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 "splitable" transition kernels.

2. Ergodic theorems for IFS controlled by i.i.d. sequences

Let BL denote the class of bounded continuous functions, $f : X \to \mathbb{R}$ (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 ν_1 and ν_2 we define the metric

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

It is well known see e.g. Shiryaev (1996) that this metric metrizes the topology of weak convergence of probability measures (on separable metric spaces).

Consider now the Kantorovich distance defined by

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

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 2.1. *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 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

$$\sup_{x \in K} d_k(\mu_n^x, \mu) \leq \frac{\gamma_K}{1 - c} c^n, \quad n \geq 0. \tag{2.1}$$

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

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

Remark 2. The limiting probability measure μ 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 (2.2), and (2.6) below.]

Remark 3. Obviously, we may replace the d_k-distance in (2.1) 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).

Proof. For $x \in X$, define the reversed iterates

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

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 μ by $\mu(\cdot) = P(\hat{Z} \in \cdot)$, we have the following sequence of inequalities:

$$\begin{aligned} d_k(\mu_n^x, \mu) &= \sup\{|\int_X f d(\mu_n^x - \mu)| : \|f\|_L \leq 1\} \\ &= \sup\{|E(f(\hat{Z}_n(x)) - f(\hat{Z}))| : \|f\|_L \leq 1\} \\ &\leq \sup\{E|f(\hat{Z}_n(x)) - f(\hat{Z})| : \|f\|_L \leq 1\} \\ &\leq Ed(\hat{Z}_n(x), \hat{Z}) = E \lim_{m \to \infty} d(\hat{Z}_n(x), \hat{Z}_m(x)) \\ &\leq E \lim_{m \to \infty} \sum_{k=n}^{m-1} d(\hat{Z}_k(x), \hat{Z}_{k+1}(x)) \\ &= E \sum_{k=n}^{\infty} d(\hat{Z}_k(x), \hat{Z}_{k+1}(x)) \end{aligned} \tag{2.2}$$

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,

$$d(\hat{Z}_n(x), \hat{Z}_m(x)) \leq \sum_{i=N}^{\infty} d(\hat{Z}_i(x), \hat{Z}_{i+1}(x)). \tag{2.3}$$

Thus if we prove that

$$E \sum_{i=N}^{\infty} d(\hat{Z}_i(x), \hat{Z}_{i+1}(x)) \to 0, \qquad as\ N \to \infty, \tag{2.4}$$

then

$$\sum_{i=N}^{\infty} d(\hat{Z}_i(x), \hat{Z}_{i+1}(x)) \stackrel{a.s.}{\to} 0 \tag{2.5}$$

and from (2.3) and (2.5) we conclude that $\{\hat{Z}_n(x)\}$ *a.s.* forms a Cauchy sequence.

Now, by recursively using condition (A) we obtain that,

$$\begin{aligned}
E\sum_{i=N}^{\infty} d(\hat{Z}_i(x), \hat{Z}_{i+1}(x)) &= \sum_{i=N}^{\infty} Ed(\hat{Z}_i(x), \hat{Z}_{i+1}(x)) \\
&= \sum_{i=N}^{\infty} E(E(d(\hat{Z}_i(x), \hat{Z}_{i+1}(x))|w_{I_1}, \dots, w_{I_i})) \\
&= \sum_{i=N}^{\infty} \Big(E(E(d(w_{I_0}(w_{I_1} \circ \cdots \circ w_{I_{i-1}}(x)), \\
&\qquad w_{I_0}(w_{I_1} \circ \cdots \circ w_{I_i}(x)))|w_{I_1}, \dots, w_{I_i}))\Big) \\
&\leq \sum_{i=N}^{\infty} cEd(w_{I_1} \circ \cdots \circ w_{I_{i-1}}(x), w_{I_1} \circ \cdots \circ w_{I_i}(x)) \\
&\leq \sum_{i=N}^{\infty} c^i Ed(x, w_{I_i}(x)) = \frac{c^N}{1-c} Ed(x, w_{I_0}(x)).
\end{aligned} \tag{2.6}$$

Since using condition (A) and (B),

$$\begin{aligned}
Ed(x, w_{I_0}(x)) &\leq d(x, x_0) + Ed(x_0, w_{I_0}(x_0)) + Ed(w_{I_0}(x_0), w_{I_0}(x)) \\
&\leq Ed(x_0, w_{I_0}(x_0)) + (c+1)d(x, x_0) < \infty,
\end{aligned} \tag{2.7}$$

it follows that (2.4) 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 $\epsilon > 0$,

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

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. Shiryaev (1996)) that

$$
d(\hat{Z}_n(x), \hat{Z}_n(y)) \to 0 \quad a.s. \tag{2.8}
$$

Define $\hat{Z} = \hat{Z}(x_0)$. From (2.8), the triangle inequality, and the fact of almost sure convergence of $\hat{Z}_n(x_0)$ to $\hat{Z}$, it follows that for any $x \in X$, $d(\hat{Z}_n(x), \hat{Z}) \stackrel{a.s.}{\to} 0$ as $n \to \infty$ establishing the a.s. independence of x.

Combining (2.2) and (2.6) we see that

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

Thus,

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

where $\gamma_K := \sup_{x \in K} Ed(x, w_{I_0}(x))$ is a finite constant, since, by taking supremums in (2.7)

$$
\gamma_K \leq Ed(x_0, w_{I_0}(x_0)) + (c+1) \sup_{x \in K} d(x, x_0) < \infty. \tag{2.9}
$$

It remains to prove that the probability measure μ 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 $g : X \to \mathbb{R}$ being a bounded and continuous function implies that the mapping

$$
x \mapsto Eg(w_{I_0}(x)) \tag{2.10}
$$

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

$$Eg(Z_n(x)) = \int_X Eg(w_{I_0}(y))P(Z_{n-1}(x) \in dy), \tag{2.11}$$

the invariance equation

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

will then follow by taking limits in (2.11) justified by using the continuity in (2.10).

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

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

as $n \to \infty$, 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

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

and the Feller property is established. Thus μ is invariant.

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

$$\int_X g d\nu = \int_X Eg(Z_1(x)) d\nu(x) = \ldots = \int_X Eg(Z_n(x)) d\nu(x)$$

$$\to \int_X \left(\int_X g d\mu \right) d\nu = \int_X g d\mu,$$

for bounded and continuous functions g, and thus μ and ν coincides. (See e.g. Billingsley (1968).) This completes the proof of Theorem 2.1.

□

3. Estimation of distances between IFS generated probability measures

We will now turn to a 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 ν_1 and ν_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 3.1. *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* $x \in X$*,*

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

Suppose

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

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

(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 $\Delta :\ [0, t] \to \mathbb{R}_+$ *such that,*

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

(F): There exists a function $T :\ [0, t] \to \mathbb{R}_+$ *bounded by* $\frac{(-\ln c)\gamma_{x_0}}{1-c}$*, such that,*

$$d_{TV}(P_\epsilon, P_0) \leq T(\epsilon),\ \ for\ all\ \epsilon \in [0, t],$$

where $P_\epsilon(\cdot) := P(I_0^\epsilon \in \cdot)$*, and where* γ_{x_0} *and* c *are the constants defined by conditions (C) and (D).*

Let μ^ϵ *denote the limiting invariant probability distribution for the Markov chain generated by the IFS and controlling sequence indexed by* ϵ *(these measures exist due to Theorem 2.1). Then there exist constants* α_0*,* α_1 *and* α_2 *such that*

$$d_w(\mu^\epsilon, \mu^0) \leq \alpha_0 \Delta(\epsilon) + \alpha_1 T(\epsilon) \ln T(\epsilon) + \alpha_2 T(\epsilon), \tag{3.1}$$

for all $\epsilon \in [0, t]$, *where*

$$\alpha_0 = \frac{1}{1-c}, \ \alpha_1 = \frac{2}{\ln c}, \ \textit{and} \ \alpha_2 = \frac{2}{\ln c}\left(\ln\left(\frac{-c(1-c)}{2\gamma_{x_0}\ln c}\right) + \ln 2 - 1\right).$$

Remark 5. If $T(\epsilon) = 0$ we interpret (3.1) as $d_w(\mu^\epsilon, \mu^0) \leq \alpha_0 \Delta(\epsilon)$.

Before we turn to the proof of Theorem 3.1, we illustrate the theorem with an example.

Example 3.2. 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 ϵ) for $i = 1, 2$. Applying the above theorem, with $c = 1/2$, $\Delta(\epsilon) = \epsilon$, and $T(\epsilon) = 0$, shows that $d_w(\mu^\epsilon, \mu^0) \leq 2\epsilon$ and thus $\mu^\epsilon \to \mu^0$ as $\epsilon \to 0$ weakly which at first glance may be somewhat conspicuous since we know that the supports of μ^ϵ, for $0 < \epsilon < 1/2$, are sets of Cantor type, while μ^0 is Lebesgue measure on $[0, 1]$.

We now turn to the proof of Theorem 3.1.

Proof. 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

$$d_w(\mu^\epsilon, \mu^0) \leq d_w(\mu^\epsilon, \mu_n^{\epsilon,x_0}) + d_w(\mu_n^{\epsilon,x_0}, \mu_n^{0,x_0}) + d_w(\mu_n^{0,x_0}, \mu^0). \tag{3.2}$$

From Theorem 2.1 using conditions (C) and (D) it follows that

$$d_w(\mu_n^{\epsilon,x_0}, \mu^\epsilon) \leq \frac{\gamma_{x_0}}{1-c} c^n, \ \textit{for all} \ n \geq 0, \ \textit{and} \ \epsilon \in [0, t]. \tag{3.3}$$

Define, for $x \in X$,

$$\tilde{Z}_n^\epsilon(x) := w^\epsilon_{I^0_{n-1}} \circ w^\epsilon_{I^0_{n-2}} \circ \cdots \circ w^\epsilon_{I^0_0}(x), \ n \geq 1, \qquad \tilde{Z}_0^\epsilon(x) = x.$$

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

We have the following inequalities,

$$
\begin{aligned}
d_w(\mu_n^{\epsilon,x_0},\mu_n^{0,x_0}) &= \sup_{f\in BL}\{|\int_X fd(\mu_n^{\epsilon,x_0}-\mu_n^{0,x_0})| : \|f\|_{BL}\le 1\} \\
&= \sup_{f\in BL}\{|Ef(Z_n^{\epsilon}(x_0))-Ef(Z_n^0(x_0))| : \|f\|_{BL}\le 1\} \\
&\le \sup_{f\in BL}\{|Ef(Z_n^{\epsilon}(x_0))-Ef(\tilde{Z}_n^{\epsilon}(x_0))| : \|f\|_{BL}\le 1\} \\
&\quad + \sup_{f\in BL}\{|Ef(\tilde{Z}_n^{\epsilon}(x_0))-Ef(Z_n^0(x_0))| : \|f\|_{BL}\le 1\} \\
&\le 2d_{TV}(P_\epsilon^{(n)},P_0^{(n)})+Ed(\tilde{Z}_n^{\epsilon}(x_0),Z_n^0(x_0)). \qquad (3.4)
\end{aligned}
$$

We may, (see Dobrushin (1970)), for any fixed ϵ and n assume that I_n^ϵ 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

$$
\begin{aligned}
d_{TV}(P_\epsilon^{(n)},P_0^{(n)}) &\le P((I_0^\epsilon,\dots,I_{n-1}^\epsilon)\neq(I_0^0,\dots,I_{n-1}^0)) \\
&= P(\cup_{i=0}^{n-1}\{I_i^\epsilon\neq I_i^0\})\le\sum_{i=0}^{n-1}P(I_i^\epsilon\neq I_i^0) \\
&= \sum_{i=0}^{n-1}d_{TV}(P_\epsilon,P_0)\le nT(\epsilon). \qquad (3.5)
\end{aligned}
$$

Studying the other term appearing in (3.4), we obtain that,

$$
Ed(\tilde{Z}_n^{\epsilon}(x_0),Z_n^0(x_0))\le Ed(w^\epsilon_{I^0_{n-1}}(\tilde{Z}^\epsilon_{n-1}(x_0)),w^0_{I^0_{n-1}}(Z^0_{n-1}(x_0)))
$$

$$
\begin{aligned}
&\le Ed(w^\epsilon_{I^0_{n-1}}(\tilde{Z}^\epsilon_{n-1}(x_0)),w^0_{I^0_{n-1}}(\tilde{Z}^\epsilon_{n-1}(x_0))) \\
&\quad +Ed(w^0_{I^0_{n-1}}(\tilde{Z}^\epsilon_{n-1}(x_0)),w^0_{I^0_{n-1}}(Z^0_{n-1}(x_0))) \\
&\le \sup_{s\in S}\sup_{x\in X}d(w_s^\epsilon(x),w_s^0(x)) \\
&\quad +E(E(d(w^0_{I^0_{n-1}}(\tilde{Z}^\epsilon_{n-1}(x_0)),w^0_{I^0_{n-1}}(Z^0_{n-1}(x_0)))|\tilde{Z}^\epsilon_{n-1}(x_0),Z^0_{n-1}(x_0))) \\
&\le \Delta(\epsilon)+cEd(\tilde{Z}^\epsilon_{n-1}(x_0),Z^0_{n-1}(x_0)), \qquad (3.6)
\end{aligned}
$$

and thus by a recursive use of (3.6), we see that

$$Ed(\tilde{Z}_n^\epsilon(x_0), Z_n^0(x_0)) \leq \Delta(\epsilon)\sum_{i=0}^{n-1} c^i \leq \Delta(\epsilon)\frac{1}{1-c}. \tag{3.7}$$

By inserting (3.5) and (3.7) in (3.4) we obtain that

$$d_w(\mu_n^{\epsilon,x_0}, \mu_n^{0,x_0}) \leq 2nT(\epsilon) + \Delta(\epsilon)\frac{1}{1-c}, \tag{3.8}$$

and thus from (3.2), (3.3), and (3.8) we see that

$$d_w(\mu^\epsilon, \mu^0) \leq \min_{n\geq 0}\left(\frac{2\gamma_{x_0}}{1-c}c^n + 2nT(\epsilon)\right) + \Delta(\epsilon)\frac{1}{1-c}. \tag{3.9}$$

In order to give a more explicit expression for the right hand side of (3.9) we investigate the function $f(x) = ax + bc^x$, $x > 0$, where $a, b > 0$, $0 < c < 1$. Suppose that $a < -b\ln c$. Then f attains its minimum $\frac{a}{\ln c}(\ln a + \ln(\frac{-1}{b\ln c}) - 1)$ at $x = \frac{1}{\ln c}\left(\ln a + \ln(\frac{-1}{b\ln c})\right) > 0$.

Let b_0 denote the smallest real number with $b_0 \geq \frac{1}{\ln c}\ln(\frac{-1}{b\ln c})$ such that $\frac{\ln a}{\ln c} + b_0$ is an integer. We have that

$$\begin{aligned} f\left(\frac{\ln a}{\ln c} + b_0\right) &\leq a\frac{\ln a}{\ln c} + a\left(\frac{1}{\ln c}\ln(\frac{-1}{b\ln c}) + 1\right) + bc^{\frac{1}{\ln c}\left(\ln a + \ln(\frac{-1}{b\ln c})\right)} \\ &= \frac{1}{\ln c}a\ln a + \frac{a}{\ln c}\left(\ln(\frac{-c}{b\ln c}) - 1\right). \end{aligned}$$

Thus, using this, with $a = 2T(\epsilon)$ and $b = \frac{2\gamma_{x_0}}{1-c}$, in (3.9), we see that

$$d_w(\mu^\epsilon, \mu^0) \leq \alpha_0\Delta(\epsilon) + \alpha_1 T(\epsilon)\ln T(\epsilon) + \alpha_2 T(\epsilon), \tag{3.10}$$

where $\alpha_0 = \frac{1}{1-c}$, $\alpha_1 = \frac{2}{\ln c}$, and

$$\alpha_2 = \frac{2}{\ln c}\left(\ln\left(\frac{-c(1-c)}{2\gamma_{x_0}\ln c}\right) + \ln 2 - 1\right).$$

This completes the proof of the theorem. □

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 4.1. 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. Kifer (1986)), suppose there exists a one-to-one Borel map $\phi : X \to \mathbb{R}$ such that $M = \phi(X)$ is a Borel subset of $\mathbb{R}$ and $\phi^{-1} : M \to X$ is also Borel measurable. Suppose that $\psi : \mathbb{R} \to X$ equals ϕ^{-1} on M and maps $\mathbb{R}\backslash M$ on some point $x_f \in X$. For each $x \in X$ define a probability measure on the Borel σ-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 (1972). 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 Quas (1991) for further results.

4.1. Recurrent Markov chains. We shall now give a corollary of Theorem 2.1. The result is well known, see e.g. Loève (1978), but the proof is non-standard and shows how classical ergodic theorems for exponentially ergodic Markov chains can be embedded within the theory of iterated function systems.

Corollary 4.2. *Suppose $\{Z_n\}$ is a homogeneous Markov chain with Polish state space X and with a transition kernel P which has the splitting* *property that*

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

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

$$\sup_{x\in X}\sup_{A\in B(X)} |P(Z_n(x)\in A) - \mu(A)| \leq \frac{\beta^{[n/n_0]}}{1-\beta}, \qquad (4.1)$$

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

Proof. It is sufficient to prove this theorem for the case $n_0 = 1$, since we can consider subsequences $\{Z_{n_0 n}\}_{n=0}^{\infty}$, and if the Markov chain $\{Z_{n_0 n}\}_{n=0}^{\infty}$ satisfies (4.1) then for $k = 0, 1, 2, \dots, n_0 - 1$ we observe that

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

and the conclusion of the theorem will then hold.

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) ν (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 that

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

where χ 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

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

$$\leq \beta Ed(f_{I_0'}(x), f_{I_0'}(y)) + (1-\beta)Ed(g_{I_0'}(x), g_{I_0'}(y)) \leq \beta d(x,y)$$

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 2.1 can be used. In fact, if $\mathcal{C} = \{f : X \to \mathbb{R};\ f = \chi$ for some indicator function $\chi\}$ then from Theorem 2.1, [using that $\gamma_K \leq 1$],

$$\sup_{A\in B(X)} |P(Z_n(x)\in A) - \mu(A)| = \sup_{f\in\mathcal{C}}\{|\int_X f d\mu_n^x - \int_X f d\mu|\}$$

$$\leq \sup\{|\int_X f d\mu_n^x - \int_X f d\mu| : \|f\|_L \leq 1\} \leq \frac{\beta^n}{1-\beta},$$

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

□

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 $p_i : X \to [0,1]$, $p_i(x) \geq 0,\ i \in S$ 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 considering this model are Barnsley *et al.* (1988), Kaijser (1994), and Lasota and Yorke (1994).

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 4.3. 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

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

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 2.1 we obtain:

Corollary 4.4. *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 μ for the Markov chain $\{Z_n(x)\}$ such that, for any bounded set $K \subseteq X$ there exists a positive constant γ such that

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

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 4.3 above. We are going to use Theorem 2.1. 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

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

The conditions in Theorem 2.1 are satisfied and thus there exists a probability measure μ such that, for any bounded set $K \subseteq X$ there exists a positive constant γ such that

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

□

Corollary 4.5. *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. $p_1(x) = p_1 + c_3 x$, 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 μ for the Markov chain $\{Z_n(x)\}$ and a positive constant γ 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 4.3 above. We are going to use Theorem 2.1. 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 2.1 are satisfied and thus the conclusion follows.

□

References

[1] Ambroladze, A. (1997) Ergodic properties of random iterations of analytic functions, *Research reports No 1*, Dept. of Mathematics, Umeå University. (To appear in *Ergodic Theory Dynam. Systems.*)

[2] Barnsley, M. F., Demko, S. G., Elton, J. H. and Geronimo, J. S. (1988) Invariant measures for Markov processes arising from iterated function systems with place-dependent probabilities, *Ann. Inst. H. Poincaré Probab. Statist.*, **24**, 367–394.

[3] Billingsley, P. (1968) *Convergence of Probability Measures*, Wiley, New York.

[4] Blumenthal, R. M. and Corson, H. H. (1972) On continuous collections of measures, *Proc. 6th Berkely Sympos. on Math. Stat. and Prob.*, **2**, 33–40.

[5] Borovkov, A. and Foss, S. G. (1994) Two ergodicity criteria for stochastically recursive sequences, *Acta Appl. Math.*, **34**, 125–134.

[6] Burton, R. M. and Rösler, U. (1995) An L_2 convergence theorem for random affine mappings, *J. Appl. Probab.*, **32**, 183–192.

[7] Centore, P. M. and Vrscay, E. R. (1994) Continuity of attractors and invariant measures for iterated function systems, *Canad. Math. Bull.*, **37**, 315–329.

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

[9] Kaijser, T. (1994) On a theorem of Karlin, *Acta Appl. Math.*, **34**, 51–69.

[10] Kifer, Y. (1986) *Ergodic theory of random transformations*, Birkhäuser, Boston.

[11] Lasota, A. and Yorke, J. A. (1994) Lower bound technique for Markov operators and iterated function systems, *Random Comput. Dynam.*, **2**, 41–77.

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

[13] Loève, M. (1978) *Probability Theory II*, 4th edition, Springer-Verlag, New York.

[14] Łoskot, K. and Rudnicki, R. (1995) Limit theorems for stochastically perturbed dynamical systems, *J. Appl. Probab.*, **32**, 459–469.

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

[16] Silvestrov, D. S. and Stenflo, Ö. (1996) Ergodic theorems for iterated function systems controlled by regenerative Sequences, *Research reports No 11*, Dept. of Mathematics, Umeå University. [To appear in *J. Theoret. Probab.*, **11**, (1998).]

[17] Quas, A. N. (1991) On representations of Markov chains by random smooth maps, *Bull. London Math. Soc.*, **23**, 487–492.

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