

# Central limit theorems for contractive Markov chains

Andreas Nordvall Lagerås and Örjan Stenflo

Department of Mathematics, Stockholm University, SE-10691 Stockholm, Sweden

E-mail: [andreas@math.su.se](mailto:andreas@math.su.se) and [stenflo@math.su.se](mailto:stenflo@math.su.se)

Received 9 February 2005, in final form 3 May 2005

Published 6 June 2005

Online at [stacks.iop.org/Non/18/1955](http://stacks.iop.org/Non/18/1955)

Recommended by C P Dettmann

## Abstract

We prove limit theorems for Markov chains under (local) contraction conditions. As a corollary we obtain a central limit theorem for Markov chains associated with iterated function systems with contractive maps and place-dependent Dini-continuous probabilities.

Mathematics Subject Classification: 60F05, 60J05, 60B10, 37H99

## 1. Introduction

Let  $(X, d)$  be a compact metric space, typically a closed and bounded subset of  $\mathbb{R}$  or  $\mathbb{R}^2$  with the Euclidean metric and let  $\{w_i\}_{i=1}^N$  be a family of (strict) contraction maps on  $X$ , i.e. there exists a constant  $c < 1$  such that  $d(w_i(x), w_i(y)) \leq cd(x, y)$ , for any  $x, y \in X$  and integer  $1 \leq i \leq N$ . Such a system is called an iterated function system (IFS) (see [1]). Hutchinson [12] and Barnsley and Demko [1] introduced these objects in order to describe fractals. It is easy to see that there exists a unique compact set  $K$  that is invariant for the IFS in the sense that  $K = \bigcup_{i=1}^N w_i(K)$ . The set  $K$  is called the fractal set, or attractor, associated with the IFS. If the maps  $w_i$  are non-degenerate and affine and the sets  $w_i(K)$ ,  $1 \leq i \leq N$ , are ‘essentially’ disjoint, then  $K$  will have the characteristic ‘self-similar’ property of a fractal. The huge class of examples of fractals that can be described in this way includes the Sierpinski gasket, Barnsley’s fern, the Cantor set and many, many others. Despite fractals being totally deterministic objects, the simplest way of drawing pictures of fractals is often via Barnsley’s ‘random iteration algorithm’: attach probabilities,  $p_i$ , to each map  $w_i$  ( $\sum_i p_i = 1$ ). Choose a starting point  $Z_0(x) := x \in X$ . Choose a function,  $w_{I_1}$ , at random from the IFS, with  $P(w_{I_1} = w_k) = p_k$ . Let  $Z_1(x) = w_{I_1}(x)$ . Next, independently, choose a function,  $w_{I_2}$ , in the same manner and let  $Z_2(x) = w_{I_2}(Z_1(x)) = w_{I_2} \circ w_{I_1}(x)$ . Repeat this ‘random iteration’ procedure inductively and define  $Z_n(x) = w_{I_n} \circ w_{I_{n-1}} \circ \cdots \circ w_{I_1}(x)$ . The random sequence  $\{Z_n(x)\}$  forms a Markov chain with a unique stationary probability distribution,  $\mu$ , supported on  $K$ . Since

$$\frac{\sum_{k=0}^{n-1} f(Z_k(x))}{n} \rightarrow \int f d\mu \quad \text{a.s.},$$

as  $n \rightarrow \infty$ , for any real-valued continuous function  $f$  on  $X$ , by Birkhoff's ergodic theorem (note that  $x$  can be chosen to be *any* fixed point by the contraction assumption), we will 'draw a picture of the attractor  $K$ ' by 'plotting' the orbit  $\{Z_n(x)\}$ , possibly ignoring some of the first points in order to reach the stationary regime. This algorithm will be an efficient way of 'drawing a picture of  $K$ ' provided the probabilities are chosen in such a way as to make the stationary distribution as uniform as possible on  $K$  and the stationary state is reached sufficiently fast. The choice of  $p_i$  can sometimes be made by inspection, by searching for a stationary distribution with the same dimension as  $K$  itself. The convergence rates towards the stationary state are 'heuristically justified' by central limit theorems (CLTs), where

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

converges in distribution to the normal distribution for  $f$  belonging to some suitably rich class of real-valued functions on  $X$ , or by stronger forms of CLTs, the so-called invariance principles or functional CLTs, where the stochastic process

$$\frac{1}{\sqrt{n}} \sum_{k=0}^{\lfloor nt \rfloor - 1} \left( f(Z_k(x)) - \int f \, d\mu \right), \quad 0 \leq t \leq 1 \quad (1)$$

converges in distribution to a Brownian motion. (Here  $[x]$  denotes the integer part of  $x$ .) Note that expression (1) above is a function-valued random element. See [5] for details about the concept of convergence in distribution for function-valued random elements.

The purpose of this paper is to study Markov chains generated by IFSs with place-dependent probabilities. (Such Markov chains have also been studied under the name 'random systems with complete connections', see [13].) We are given a set of contraction maps  $\{w_i\}$ , with associated continuous functions  $p_i = p_i(x)$ , where  $p_i : X \rightarrow (0, 1)$ , with  $\sum_i p_i(x) = 1$ , for any  $x \in X$ . The Markov chains are characterized by the transfer operator  $T$  defined for real-valued measurable functions  $f$  on  $X$  by  $Tf(x) = \sum_i p_i(x) f(w_i(x))$ . Intuitively, the Markov chains considered are generated by fixing a starting point  $x$  and letting  $Z_0(x) := x$ , and inductively letting  $Z_{n+1}(x) := w_i(Z_n(x))$  with probability  $p_i(Z_n(x))$  for  $n \geq 0$ .

One motivation for studying such chains is that it gives more freedom when trying to generate a 'uniform' stationary probability distribution on  $K$ . Such Markov chains also arise naturally in the thermodynamic formalism of statistical mechanics. It is well known that they do not necessarily possess a unique stationary distribution (see [4, 6, 20, 26, 27]), but with some additional regularity conditions, uniqueness holds (see [11, 14, 27, 28]).

The operator  $T$  (without the normalizing condition  $\sum_i p_i(x) = 1$ ) is known as the Ruelle–Perron–Frobenius operator. Fan and Lau [10] proved a limit theorem for iterates of the Ruelle–Perron–Frobenius operator under the Dini-continuity assumptions on the  $p_i$ , by lifting a similar result from Walters [29] on symbolic spaces. (Recall that  $p_i$  is Dini-continuous if  $\int_0^1 (\Delta_{p_i}(t)/t) \, dt < \infty$ , or equivalently  $\sum_{n=0}^{\infty} \Delta_{p_i}(c^n) < \infty$ , for some (and thus all)  $0 < c < 1$ , where  $\Delta_{p_i}(t) := \sup_{d(x,y) \leq t} |p_i(x) - p_i(y)|$  is the modulus of uniform continuity of  $p_i$ .) Uniqueness in stationary distributions still holds (in the normalized cases) if the contraction assumptions of the  $w_i$  are relaxed to 'average contraction' under the Dini-continuity assumption (see [2, 17]) but information about rates of convergence in these 'average contractive' cases seems to be unknown.

The Dini-condition is somewhat stronger than the weakest known conditions for uniqueness in stationary probability distributions (in the normalized cases with strict contraction maps), but weaker than, e.g., Hölder-continuity.

In the Dini-continuous cases it follows that the unique equilibrium measure will have the Gibbs (approximation) property (see [10]). This property is of importance when analysing the multidimensional spectra of measures.

In this paper we will prove the perhaps initially surprising fact (corollary 2) that Markov chains generated by IFSs with Dini-continuous probabilities obey a CLT, despite the well-known fact that such Markov chains do not typically converge with an exponential rate. Our main result, theorem 1, expresses this in a natural generality.

CLTs/functional CLTs for iterated random functions under conditions that imply exponential (or other rapid) rates of convergence have previously been proved in, e.g., [3, 15, 16, 22, 30, 31]. We discuss the connection between some of these results and our results in remarks 4 and 6.

## 2. Preliminaries

Let  $\mathcal{B}$  denote the Borel  $\sigma$ -field generated by the metric  $d$ , and let  $\mathbf{P} : X \times \mathcal{B} \rightarrow [0, 1]$  be a transition probability. That is, for each  $x \in X$ ,  $\mathbf{P}(x, \cdot)$  is a probability measure on  $(X, \mathcal{B})$  and for each  $A \in \mathcal{B}$ ,  $\mathbf{P}(\cdot, A)$  is  $\mathcal{B}$ -measurable. The transition probability generates a Markov chain with transfer operator defined by  $Tf(x) = \int_X f(y)\mathbf{P}(x, dy)$  for real-valued measurable functions  $f$  on  $X$ . A probability measure  $\mu$  is stationary for  $\mathbf{P}$  if  $\mu(\cdot) = \int_X \mathbf{P}(x, \cdot) d\mu(x)$ .

There are several ways of representing a Markov chain with a given transfer operator. One common way is to find a measurable function  $w : X \times [0, 1] \rightarrow X$ , let  $\{I_j\}_{j=1}^\infty$  be a sequence of independent random variables uniformly distributed in  $[0, 1]$ , and consider the random dynamical system defined by

$$Z_n(x) := w_{I_n} \circ w_{I_{n-1}} \circ \dots \circ w_{I_1}(x), \quad n \geq 1, \quad Z_0(x) := x,$$

for any  $x \in X$ , where

$$w_s(x) = w(x, s).$$

It is always possible to find such a representation,  $w$ , such that the transition probability generated by  $\{Z_n\}$  is  $\mathbf{P}$ , i.e.  $Ef(Z_n(x)) = T^n f(x)$ , for any  $x, n$  and  $f$  (see [19]).

For two fixed points  $x, y \in X$  and  $\mathbf{x} = (x, y)$  we can consider the Markov chain  $\{Z_n(\mathbf{x})\}$ , on  $X^2$ , where  $Z_n(\mathbf{x}) := (Z_n(x), Z_n(y))$ . When proving theorems based on contraction conditions we are typically interested in representations that minimize  $d(Z_n(x), Z_n(y))$  (in some average sense).

More generally, if  $W : X^2 \times [0, 1] \rightarrow X^2$ , is a measurable map and  $\{I_j\}_{j=1}^\infty$  is a sequence of independent random variables uniformly distributed in  $[0, 1]$ , we will consider the random dynamical system defined by

$$Z_n(\mathbf{x}) := W_{I_n} \circ W_{I_{n-1}} \circ \dots \circ W_{I_1}(\mathbf{x}), \quad n \geq 1, \quad Z_0(\mathbf{x}) := \mathbf{x}, \quad (2)$$

where  $W_s(\mathbf{x}) = W(\mathbf{x}, s)$ , such that, for any  $\mathbf{x} = (x, y) \in X^2$ , the Markov chain  $Z_n(\mathbf{x}) := (Z_n^{(x,y)}(x), Z_n^{(x,y)}(y))$  on  $X^2$  has marginals  $\mathbf{P}^n(x, \cdot) = P(Z_n^{(x,y)}(x) \in \cdot)$ , and  $\mathbf{P}^n(y, \cdot) = P(Z_n^{(x,y)}(y) \in \cdot)$ , for any  $n$ .

Thus  $\{Z_n^{(x,y)}(x)\}$  and  $\{Z_n^{(x,y)}(y)\}$  denote two Markov chains on  $X$ , defined on the same probability space, with the former starting at  $x \in X$  and the latter starting at  $y \in X$ , both with transition probability  $\mathbf{P}$ .

Let  $d_w$  be the Monge–Kantorovich metric defined by  $d_w(\pi, \nu) = \sup(\int f d(\pi - \nu))$ ;  $f : X \rightarrow \mathbb{R}$ ,  $|f(x) - f(y)| \leq d(x, y) \forall x, y$ , for probability measures  $\pi$  and  $\nu$  on  $X$ . The Monge–Kantorovich metric metrizes the topology of weak convergence on the set of probability

measures on  $X$  (see [9]). It follows from the definitions that for any stationary probability measure  $\mu$ , we have

$$d_w(\mathbf{P}^n(x, \cdot), \mu(\cdot)) \leq \sup_{x, y \in X} Ed(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y)). \quad (3)$$

Therefore if  $\sup_{x, y} Ed(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y)) \rightarrow 0$  as  $n \rightarrow \infty$ , then there is a unique stationary distribution for  $\mathbf{P}$ .

We will sometimes drop the upper index, i.e. write  $Z_n(x)$  instead of  $Z_n^{(x,y)}(x)$  etc, when we are not interested in the joint distribution of the pair  $(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y))$ .

The following proposition gives sufficient conditions for the existence of a CLT.

**Proposition 1.** *Suppose there exists a unique stationary distribution  $\mu$  for  $\mathbf{P}$ , and let  $f$  be a real-valued measurable function on  $X$  with  $\|f\|_{L^2}^2 = \int f^2 d\mu < \infty$ . Suppose that for some  $\delta > 0$ ,*

$$\lim_{n \rightarrow \infty} n^{-1/2} (\log n)^{1+\delta} \sup_{x, y \in X} E \sum_{k=0}^{n-1} |f(Z_k^{(x,y)}(x)) - f(Z_k^{(x,y)}(y))| = 0. \quad (4)$$

Let

$$S_n^x = \sum_{k=0}^{n-1} (f(Z_k(x)) - Ef(Z_k(x))),$$

$$S_n^{x,\mu} = \sum_{k=0}^{n-1} \left( f(Z_k(x)) - \int f d\mu \right)$$

and

$$B_n^x(t) = \frac{S_{[nt]}^x}{\sqrt{n}}, \quad 0 \leq t \leq 1,$$

$$B_n^{x,\mu}(t) = \frac{S_{[nt]}^{x,\mu}}{\sqrt{n}}, \quad 0 \leq t \leq 1.$$

Then the limit

$$\sigma^2 = \sigma^2(f) := \lim_{n \rightarrow \infty} \frac{1}{n} E[(S_n^Z)^2] \quad (5)$$

exists and is finite, where  $Z$  is a  $\mu$ -distributed random variable, independent of  $\{I_j\}_{j=1}^\infty$ . Furthermore, if  $B = \{B(t) : 0 \leq t \leq 1\}$  denotes the standard Brownian motion, then

$$B_n^x \xrightarrow{d} \sigma B \quad (6)$$

and

$$B_n^{x,\mu} \xrightarrow{d} \sigma B, \quad (7)$$

as  $n \rightarrow \infty$ , for any  $x \in X$ , where  $\xrightarrow{d}$  denotes convergence in distribution for random elements taking values in the space of right-continuous functions on  $[0, 1]$  with left-hand limits equipped with the Skorokhod topology.

**Remark 1.** Proposition 1 above is valid when  $(X, \mathcal{B})$  is a general measurable space.

**Remark 2.** General CLTs for Markov chains started at a point have been proved by Derriennic and Lin [7]. Proposition 1 complements their result in cases of ‘uniform’ ergodicity. The proof of proposition 1, given later, relies on a slightly stronger result by Peligrad and Utev [23] for

Markov chains starting according to the unique stationary probability distribution. Theorems about convergence, allowing Markov chains to start at a point, are important in the theory for Markov chain–Monte Carlo methods.

**Remark 3.** In an earlier draft of this paper we proved a weaker (non-functional) form of the CLT in proposition 1, where our result was based on a CLT by Maxwell and Woodroffe [21]. The recent paper by Peligrad and Utev [23], which was helpfully pointed out to us by a referee, enabled us to state our CLT in the current functional CLT form.

**Remark 4.** Wu and Woodroffe considered general state spaces in [31]. The conditions in their CLT (theorem 2) imply (4) in the case of a compact  $X$ . This can be seen as follows: their proof of this theorem amounts to showing that  $\sum_{n=0}^{\infty} \|T^n f\|_{L^2} < \infty$ , for centred functions  $f$ . Restricting  $X$  to be compact allows a strengthening of their lemma 3, so that its result holds even when starting  $\{Z_k(x)\}$  from a point. With some minor modifications to the proof, it is possible to show that  $\sum_{n=0}^{\infty} \sup_{x,y} E|f(Z_n^{(x,y)}(x)) - f(Z_n^{(x,y)}(y))| < \infty$ . Thus the conditions of our proposition 1 hold.

Checking the  $L^2$  boundedness condition could be difficult if we have no *a priori* information about the (possibly non-unique) stationary measures. The following corollary circumvents these problems and might therefore be more directly applicable in our case when  $(X, d)$  is compact.

**Corollary 1.** *If*

$$\lim_{n \rightarrow \infty} \sup_{x,y \in X} Ed(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y)) = 0, \tag{8}$$

*then there exists a unique stationary distribution  $\mu$  for  $P$ .*

*Let  $f$  be a real-valued continuous function on  $X$ . Suppose  $\Delta_f : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  is an increasing concave function with  $\Delta_f(t) \geq \sup_{d(x,y) \leq t} |f(x) - f(y)|$ , for any  $t \geq 0$  and suppose, in addition to (8), that for some  $\delta > 0$ ,*

$$\lim_{n \rightarrow \infty} \sqrt{n}(\log n)^{1+\delta} \Delta_f \left( \sup_{x,y \in X} Ed(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y)) \right) = 0 \tag{9}$$

*also holds, then the conclusions of proposition 1 hold for  $f$ , i.e. the limit (5) exists for  $f$  and is finite and (6) and (7) hold.*

**Remark 5.** The function  $\Delta_f$  may thus be chosen to be the modulus of uniform continuity of  $f$  in cases when this function is concave.

**Remark 6.** If  $\sup_{x,y \in X} Ed(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y)) \sim O(c^n)$ , for some constant  $c < 1$ , satisfied for instance the average-contractive IFSs with place-independent probabilities, then it follows from corollary 1 that the CLT holds with respect to any  $f$  of modulus of uniform continuity  $\Delta_f$ , of order  $\Delta_f(c^n) \sim o(1/\sqrt{n}(\log n)^{1+\delta})$ . This condition is satisfied by, e.g., Dini-continuous functions  $f$ . Corollary 1 thus strengthens theorem 2.4. of [3] (who considered Lipschitz-continuous  $f$ ). Wu and Shao [30] considered functions  $f$  that are stochastically Dini-continuous with respect to the stationary distribution. (It should be noted that [3] and [30] treated average contractive IFSs on more general metric spaces.)

**Proof (proposition 1).** Let  $f \in L^2(\mu)$  be a real-valued measurable function on  $X$  satisfying assumption (4).

Since

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{1}{n^{3/2}} \left\| \sum_{k=0}^{n-1} T^k \left( f - \int f \, d\mu \right) \right\|_{L^2} &= \sum_{n=1}^{\infty} \frac{1}{n^{3/2}} \left\| \sum_{k=0}^{n-1} \left( T^k f - \int f \, d\mu \right) \right\|_{L^2} \\
&\leq \sum_{n=1}^{\infty} \frac{1}{n^{3/2}} \sup_{x \in X} \left| \sum_{k=0}^{n-1} \left( T^k f(x) - \int f \, d\mu \right) \right| \\
&= \sum_{n=1}^{\infty} \frac{1}{n^{3/2}} \sup_{x \in X} \left| \sum_{k=0}^{n-1} \left( T^k f(x) - \int T^k f \, d\mu \right) \right| \\
&\leq \sum_{n=1}^{\infty} \frac{1}{n^{3/2}} \sup_{x, y \in X} \left| \sum_{k=0}^{n-1} \left( T^k f(x) - T^k f(y) \right) \right| \\
&= \sum_{n=1}^{\infty} \frac{1}{n^{3/2}} \sup_{x, y \in X} \left| E \sum_{k=0}^{n-1} \left( f(Z_k^{(x,y)}(x)) - f(Z_k^{(x,y)}(y)) \right) \right| \\
&\leq \sum_{n=1}^{\infty} \frac{1}{n^{3/2}} \sup_{x, y \in X} E \sum_{k=0}^{n-1} \left| f(Z_k^{(x,y)}(x)) - f(Z_k^{(x,y)}(y)) \right| < \infty,
\end{aligned}$$

it follows from theorem 1.1 of [23] that  $\sigma^2 = \lim_{n \rightarrow \infty} (1/n) E[(S_n^Z)^2]$  exists and is finite, and  $B_n^{Z, \mu} \xrightarrow{d} \sigma B$ , where  $Z$  is a  $\mu$ -distributed random variable, independent of  $\{I_j\}_{j=1}^{\infty}$ . By Chebyshev's inequality,

$$\begin{aligned}
&P \left( \sup_{0 \leq t \leq 1} |B_n^{x, \mu}(t) - B_n^{Z, \mu}(t)| \geq \epsilon \right) \\
&= P \left( \max_{0 \leq m \leq n} \frac{1}{\sqrt{n}} \left| \sum_{k=0}^{m-1} \left( f(Z_k^{(x,Z)}(x)) - f(Z_k^{(x,Z)}(Z)) \right) \right| \geq \epsilon \right) \\
&\leq P \left( \frac{1}{\sqrt{n}} \max_{0 \leq m \leq n} \sum_{k=0}^{m-1} \left| f(Z_k^{(x,Z)}(x)) - f(Z_k^{(x,Z)}(Z)) \right| \geq \epsilon \right) \\
&\leq P \left( \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} \left| f(Z_k^{(x,Z)}(x)) - f(Z_k^{(x,Z)}(Z)) \right| \geq \epsilon \right) \\
&\leq \frac{1}{\epsilon \sqrt{n}} E \sum_{k=0}^{n-1} \left| f(Z_k^{(x,Z)}(x)) - f(Z_k^{(x,Z)}(Z)) \right| \\
&\leq \frac{1}{\epsilon \sqrt{n}} \sup_{x, y \in X} E \sum_{k=0}^{n-1} \left| f(Z_k^{(x,y)}(x)) - f(Z_k^{(x,y)}(y)) \right| \rightarrow 0,
\end{aligned}$$

as  $n \rightarrow \infty$ . By theorem 4.1 in [5],  $B_n^{x, \mu} \xrightarrow{d} \sigma B$ .

The difference between  $S_n^x$  and  $S_n^{x,\mu}$  lies in how the summands are centred. The difference is negligible in the limit:

$$\begin{aligned} \sup_{0 \leq t \leq 1} |B_n^{x,\mu}(t) - B_n^x(t)| &= \max_{0 \leq m \leq n} \frac{1}{\sqrt{n}} \left| \sum_{k=0}^{m-1} \left( Ef(Z_k(x)) - \int f \, d\mu \right) \right| \\ &\leq \frac{1}{\sqrt{n}} \max_{0 \leq m \leq n} E \left| \sum_{k=0}^{m-1} \left( f(Z_k(x)) - \int f(Z_k(y)) \, d\mu(y) \right) \right| \\ &\leq \frac{1}{\sqrt{n}} E \sum_{k=0}^{n-1} \left| f(Z_k(x)) - \int f(Z_k(y)) \, d\mu(y) \right| \\ &\leq \frac{1}{\sqrt{n}} \sup_{x,y \in X} E \sum_{k=0}^{n-1} \left| f(Z_k^{(x,y)}(x)) - f(Z_k^{(x,y)}(y)) \right| \rightarrow 0, \end{aligned}$$

as  $n \rightarrow \infty$ . Thus also  $B_n^x \xrightarrow{d} \sigma B$ . □

**Proof (corollary 1).** The first part of the corollary follows from (3) above.

For the proof of the second part of corollary 1, first note that by assumption (9),

$$\Delta_f \left( \sup_{x,y \in X} Ed(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y)) \right) \sim o \left( \frac{1}{\sqrt{n}(\log n)^{1+\delta}} \right),$$

implying that

$$\sum_{k=0}^{n-1} \Delta_f \left( \sup_{x,y \in X} Ed(Z_k^{(x,y)}(x), Z_k^{(x,y)}(y)) \right) \sim o \left( \frac{\sqrt{n}}{(\log n)^{1+\delta}} \right).$$

(To see this, note that the derivative  $F'(t)$  of  $F(t) = \sqrt{t}/(\log t)^{1+\delta}$  satisfies  $F'(t) \geq 1/(3\sqrt{t}(\log t)^{1+\delta})$ , for large  $t$ .)

Thus,

$$\lim_{n \rightarrow \infty} n^{-1/2} (\log n)^{1+\delta} \sum_{k=0}^{n-1} \Delta_f \left( \sup_{x,y \in X} Ed(Z_k^{(x,y)}(x), Z_k^{(x,y)}(y)) \right) = 0.$$

Since by the definition of  $\Delta_f$  and Jensen's inequality,

$$\begin{aligned} \Delta_f \left( \sup_{x,y \in X} Ed(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y)) \right) &\geq \sup_{x,y \in X} \Delta_f(Ed(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y))) \\ &\geq \sup_{x,y \in X} E \Delta_f(d(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y))) \\ &\geq \sup_{x,y \in X} E |f(Z_n^{(x,y)}(x)) - f(Z_n^{(x,y)}(y))| \end{aligned}$$

and

$$\sum_{k=0}^{n-1} \sup_{x,y \in X} E |f(Z_k^{(x,y)}(x)) - f(Z_k^{(x,y)}(y))| \geq \sup_{x,y \in X} E \sum_{k=0}^{n-1} |f(Z_k^{(x,y)}(x)) - f(Z_k^{(x,y)}(y))|,$$

we see that an application of proposition 1 completes the proof of the second part of corollary 1.  $\square$

### 3. Main results

**Theorem 1.** Let  $W : X^2 \times [0, 1] \rightarrow X^2$  be a measurable map such that for any fixed  $(x, y) \in X^2$  the map  $W(x, y, \cdot) := (W^{(x,y)}(x), W^{(x,y)}(y))(\cdot)$  defines random variables with  $P(W^{(x,y)}(x) \in \cdot) = P(x, \cdot)$  and  $P(W^{(x,y)}(y) \in \cdot) = P(y, \cdot)$ , where  $P$  denotes the uniform probability measure on the Borel subsets of  $[0, 1]$ .

Let  $\Delta : [0, \infty) \rightarrow [0, 1]$ , be an increasing function with  $\Delta(0) = 0$ . Suppose there exists a constant  $c < 1$ , such that

$$P(d(W^{(x,y)}(x), W^{(x,y)}(y)) \leq cd(x, y)) \geq 1 - \Delta(d(x, y)), \quad (10)$$

for any two points  $x, y \in X$ .

Then

(i) (Distributional stability theorem)

$$d_w(\mathbf{P}^n(x, \cdot), \mu(\cdot)) \leq \sup_{x, y \in X} Ed(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y)) \leq ED_n, \quad (11)$$

for any stationary probability distribution  $\mu$ , where  $D_n$  is a homogeneous Markov chain with  $D_0 = \text{diam}(X) := \sup_{x, y} d(x, y)$ ,

$$P(D_{n+1} = ct \mid D_n = t) = 1 - \Delta(t)$$

and

$$P(D_{n+1} = \text{diam}(X) \mid D_n = t) = \Delta(t),$$

for any  $0 \leq t \leq \text{diam}(X)$ .

If

$$\sum_{n=1}^{\infty} \prod_{k=1}^n (1 - \Delta(c^k)) = \infty, \quad (12)$$

then  $ED_n \rightarrow 0$  and thus by corollary 1 there is a unique stationary distribution,  $\mu$ .

(ii) (Central limit theorem)

If  $\sum_{k=0}^{\infty} \Delta(c^k) < \infty$ , then the conclusions of proposition 1 hold for any Hölder-continuous function  $f$  with exponent  $\alpha > \frac{1}{2}$ .

**Proof (theorem 1(i)).** Fix two points  $x$  and  $y$  in  $X$ . Define  $Z_0^{(x,y)}(x) = x$ ,  $Z_0^{(x,y)}(y) = y$  and inductively

$$Z_n^{(x,y)}(x) = W^{(Z_{n-1}^{(x,y)}(x), Z_{n-1}^{(x,y)}(y))}(Z_{n-1}^{(x,y)}(x))$$

and

$$Z_n^{(x,y)}(y) = W^{(Z_{n-1}^{(x,y)}(x), Z_{n-1}^{(x,y)}(y))}(Z_{n-1}^{(x,y)}(y)),$$

as in (2). Then  $Z_n^{(x,y)}(x)$  and  $Z_n^{(x,y)}(y)$  are random variables such that  $Ef(Z_n^{(x,y)}(x)) = T^n f(x)$  and  $Ef(Z_n^{(x,y)}(y)) = T^n f(y)$ , for any  $n$ .

We have from assumption (10) that

$$\begin{aligned} P(d(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y)) \leq ct \mid d(Z_{n-1}^{(x,y)}(x), Z_{n-1}^{(x,y)}(y)) \leq t) \\ \geq 1 - \Delta(t) = P(D_n = ct \mid D_{n-1} = t), \end{aligned}$$



for any  $t \in \{c^k \text{diam}(X)\}_{k=0}^\infty$ . (Note that  $D_n$  takes values in the discrete state space  $\{c^k \text{diam}(X)\}_{k=0}^\infty$ .)

$D_n$  is therefore stochastically larger than  $d(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y))$ , and consequently  $ED_n \geq Ed(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y))$ , for any  $x, y \in X$ . The other inequality of (11) follows from (3).

Since  $\{D_n\}$  is a non-ergodic Markov chain under condition (12) (see [24], p 80), it follows that  $ED_n \rightarrow 0$  as  $n \rightarrow \infty$ , if (12) holds, and we have thus proved theorem 1(i).

In order to prove theorem 1(ii), we first observe that it is well known that  $\sum_{k=0}^\infty \Delta(c^k) < \infty$  implies that  $D_n$  is transient (see [24], p 80). Therefore (see [25], p 575),  $\sum_{k=0}^\infty P(D_k = \text{diam}(X)) < \infty$  and it follows that

$$\begin{aligned} \sum_{k=0}^\infty ED_k &= \sum_{k=0}^\infty \sum_{j=0}^k c^j \text{diam}(X) P(D_k = c^j \text{diam}(X)) \\ &\leq \text{diam}(X) \sum_{k=0}^\infty \sum_{j=0}^k c^j P(D_{k-j} = \text{diam}(X)) \\ &= \frac{\text{diam}(X)}{1-c} \sum_{k=0}^\infty P(D_k = \text{diam}(X)) < \infty. \end{aligned}$$

By stochastic monotonicity  $ED_k$  is decreasing, and thus  $\sum_{k=1}^n ED_k \geq nED_n$ , for any  $n$ . This implies that  $ED_n \leq c_0/n$ , for  $c_0 := \sum_{k=0}^\infty ED_k$ .

Thus  $\sup_{x,y} Ed(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y)) \leq c_0/n$ , for any  $n \geq 1$ . If  $f$  is a Hölder-continuous function on  $X$ , with modulus of uniform continuity  $\Delta_f$  satisfying  $\Delta_f(t) \leq c_1 t^\alpha$ , for some constants  $c_1$  and  $\alpha > \frac{1}{2}$ , and any  $t \geq 0$ , it follows that for any  $\delta > 0$ ,

$$\begin{aligned} \lim_{n \rightarrow \infty} \sqrt{n}(\log n)^{1+\delta} \Delta_f \left( \sup_{x,y \in X} Ed(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y)) \right) \\ \leq \lim_{n \rightarrow \infty} \sqrt{n}(\log n)^{1+\delta} c_1 \left( \frac{c_0}{n} \right)^\alpha = 0. \end{aligned}$$

An application of corollary 1 now completes the proof of theorem 1(ii). □

#### 4. IFSS with place-dependent probabilities

Let  $\{w_i\}_{i=1}^\infty$  be a set of strictly contracting maps, i.e. there exist a constant  $c < 1$  such that  $d(w_i(x), w_i(y)) \leq cd(x, y)$ , for any  $x, y \in X$  and any integer  $i$ . Let  $\{p_i(x)\}_{i=1}^\infty$  be associated place-dependent probabilities, i.e. non-negative continuous functions, with  $\sum_i p_i(x) = 1$ , for any  $x \in X$ . This system defines a Markov chain with transfer operator defined by  $Tf(x) = \sum_{i=1}^\infty p_i(x) f(w_i(x))$ , for real-valued measurable functions  $f$  on  $X$ .

Let

$$\Delta(t) = \frac{1}{2} \sup_{d(x,y) \leq t} \sum_{i=1}^\infty |p_i(x) - p_i(y)| = 1 - \inf_{d(x,y) \leq t} \sum_{i=1}^\infty \min(p_i(x), p_i(y)) \tag{13}$$

and let for any two points  $x, y \in X$ ,  $W^{(x,y)}(x)$  and  $W^{(x,y)}(y)$  be random variables defined by

$$P(W^{(x,y)}(x) = w_i(x), W^{(x,y)}(y) = w_i(y)) = \min(p_i(x), p_i(y)) \tag{14}$$

and

$$\begin{aligned} P(W^{(x,y)}(x) = w_i(x), W^{(x,y)}(y) = w_j(y)) \\ = \frac{(p_i(x) - \min(p_i(x), p_i(y)))(p_j(y) - \min(p_j(x), p_j(y)))}{1 - \sum_{k=1}^{\infty} \min(p_k(x), p_k(y))}, \end{aligned} \quad (15)$$

when  $i \neq j$ . (If  $p_i(x) = p_i(y), \forall i$ , then we understand the expression in (15) as zero.)

It is straightforward to check that by construction  $P(W^{(x,y)}(x) = w_i(x)) = p_i(x)$ , and  $P(W^{(x,y)}(y) = w_j(y)) = p_j(y)$  for any  $i$  and  $j$ .

It follows from (14) that

$$P(d(W^{(x,y)}(x), W^{(x,y)}(y)) \leq cd(x, y)) \geq \sum_{i=1}^{\infty} \min(p_i(x), p_i(y)) \geq 1 - \Delta(d(x, y)),$$

and we may therefore apply theorem 1 to obtain the following.

**Corollary 2.** *Let  $\{w_i\}_{i=1}^{\infty}$  be an IFS with strictly contractive maps, and let  $\{p_i(x)\}$  be associated place-dependent probabilities. Then the conclusions of theorem 1 hold with  $\Delta$  defined as in (13) above.*

Let us illustrate the above corollary with an example.

**Example 1.** Let  $w_1$  and  $w_2$  be two maps from  $[0, 1]$  into itself defined by

$$w_1(x) = \beta x \quad \text{and} \quad w_2(x) = \beta x + (1 - \beta),$$

where  $0 < \beta < 1$  is a constant parameter. Consider the Markov chain with transfer operator  $T : C([0, 1]) \rightarrow C([0, 1])$  defined by

$$Tf(x) = p(x)f(w_1(x)) + (1 - p(x))f(w_2(x)), \quad f \in C([0, 1]),$$

where  $p : [0, 1] \rightarrow (0, 1)$ , is a continuous function with modulus of uniform continuity  $\Delta = \Delta_p$ .

The case when  $p(x) \equiv \frac{1}{2}$  and  $\beta = \frac{1}{2}$ , where the uniform distribution on  $[0, 1]$  is the unique stationary distribution, and the case when  $p(x) \equiv \frac{1}{2}$  and  $\beta = \frac{1}{3}$ , where the uniform distribution on the (middle third) Cantor set is the unique stationary distribution, are two important particular cases of this model.

For general  $p$ , Markov chains of this form always possess a stationary probability distribution, but they may possess more than one stationary probability distribution (see [26]).

From theorem 1 it follows that the distribution will be unique (for any fixed value of the parameter  $\beta$ ) provided (12) holds, and this theorem also enables us to quantify the rate of convergence as a function of the modulus of uniform continuity of  $p$ . It also follows that this Markov chain will obey the functional CLT (6) and (7) for Hölder-continuous functions  $f$  with exponent  $\alpha > \frac{1}{2}$  provided  $p$  is Dini-continuous. Observe that our conditions are only sufficient. It is an interesting open problem to try to find critical smoothness properties of  $p$  to ensure a unique stationary measure and a CLT.

**Remark 7.** If  $X = \{1, \dots, N\}^{\mathbb{N}}$  and for two elements  $x = x_0x_1\dots$  and  $y = y_0y_1\dots$  in  $X$ , we define  $d(x, y) := 2^{-\min(k \geq 0; x_k \neq y_k)}$  if  $x \neq y$ , and  $d(x, y) := 0$  if  $x = y$ , then  $(X, d)$  is a compact metric space. Let  $g$  be a continuous function from  $X$  to  $(0, 1]$ , such that  $\sum_{x_0=1}^N g(x_0x_1\dots) = 1$  for all  $x_1x_2\dots \in X$ .  $g$  describes an IFS with place-dependent probabilities:  $\{(X, d), w_i(x), p_i(x), i \in \{1, \dots, N\}\}$ , where  $w_i(x) = ix$  and  $p_i(x) = g(ix)$ , and corollary 2 applies. This generalizes theorem 1 in [28] and also implies a CLT for the associated Markov chains under the ‘summable variations’ condition used in [8] or [29]. Stationary probability measures for such Markov chains are sometimes called  $g$ -measures, a concept coined by Keane [18].

## Acknowledgments

We are grateful to the referees for pointing out useful references and for providing us with comments for improving this paper.

## References

- [1] Barnsley M F and Demko S 1985 Iterated function systems and the global construction of fractals *Proc. R. Soc. Lond. A* **399** 243–75
- [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. Stat.* **24** 367–94  
Barnsley M F, Demko S G, Elton J H and Geronimo J S 1989 *Ann. Inst. H Poincaré Probab. Stat.* **25** 589–90 (erratum)
- [3] Benda M 1998 A central limit theorem for contractive stochastic dynamical systems *J. Appl. Probab.* **35** 200–5
- [4] Berger N, Hoffman C and Sidoravicius V 2003 Nonuniqueness for specifications in  $\ell^{2+\epsilon}$  *Preprint math PR/0312344*
- [5] Billingsley P 1968 *Convergence of Probability Measures* (New York: Wiley)
- [6] Bramson M and Kalikow S 1993 Non-uniqueness in  $g$ -functions *Israel J. Math.* **84** 153–60
- [7] Derriennic Y and Lin M 2003 The central limit theorem for Markov chains started at a point *Probab. Theory Rel. Fields* **125** 73–6
- [8] Doebelin W and Fortet R 1937 Sur les chaînes à liaisons complètes *Bull. Soc. Math. France* **65** 132–48
- [9] Dudley R M 1989 *Real Analysis and Probability* (Pacific Grove, CA: Wadsworth and Brooks/Cole)
- [10] Fan A H and Lau K-S 1999 Iterated function system and Ruelle operator *J. Math. Anal. Appl.* **231** 319–44
- [11] Harris T E 1955 On chains of infinite order *Pac. J. Math.* **5** 707–24
- [12] Hutchinson J E 1981 Fractals and self-similarity *Indiana Univ. Math. J.* **30** 713–47
- [13] Iosifescu M and Grigorescu S 1990 *Dependence with Complete Connections and its Applications* (Cambridge: Cambridge University Press)
- [14] Johansson A and Öberg A 2003 Square summability of variations of  $g$ -functions and uniqueness of  $g$ -measures *Math. Res. Lett.* **10** 587–601
- [15] Kaijser T 1973 Some limit theorems for Markov chains with applications to learning models and products of random matrices *PhD Thesis* Institut Mittag-Leffler, Djursholm, Sweden
- [16] Kaijser T 1979 Another central limit theorem for random systems with complete connections *Rev. Roum. Math. Pures Appl.* **24** 383–412
- [17] Kaijser T 1981 On a new contraction condition for random systems with complete connections *Rev. Roum. Math. Pures Appl.* **26** 1075–117
- [18] Keane M 1972 Strongly mixing  $g$ -measures *Invent. Math.* **16** 309–24
- [19] Kifer Y 1986 *Ergodic Theory of Random Transformations* (Boston, MA: Birkhäuser)
- [20] Lacroix Y 2000 A note on weak- $\star$  perturbations of  $g$ -measures *Sankhyā A* **62** 331–8
- [21] Maxwell M and Woodroffe M 2000 Central limit theorems for additive functionals of Markov chains *Ann. Probab.* **28** 713–24
- [22] Norman M F 1972 *Markov Processes and Learning Models* (New York: Academic)
- [23] Peligrad M and Utev S 2005 A new maximal inequality and invariance principle for stationary sequences *Ann. Probab.* **33** 798–815
- [24] Prabhu N U 1965 *Stochastic Processes. Basic Theory and its Applications* (New York: Macmillan)
- [25] Shiryaev A N 1996 *Probability* 2nd edn (New York: Springer)
- [26] Stenflo Ö 2001 A note on a theorem of Karlin *Stat. Probab. Lett.* **54** 183–7
- [27] Stenflo Ö 2002 Uniqueness of invariant measures for place-dependent random iterations of functions *IMA Vol. Math. Appl.* **132** 13–32
- [28] Stenflo Ö 2003 Uniqueness in  $g$ -measures *Nonlinearity* **16** 403–10
- [29] Walters P 1975 Ruelle’s operator theorem and  $g$ -measures *Trans. Am. Math. Soc.* **214** 375–87
- [30] Wu W B and Shao X 2004 Limit theorems for iterated random functions *J. Appl. Probab.* **41** 425–36
- [31] Wu W B and Woodroffe M 2000 A central limit theorem for iterated random functions *J. Appl. Probab.* **37** 748–55