Mathematical Statistics
Stockholm University

Central limit theorems for contractive Markov chains

Andreas Nordvall Lagerås and Örjan Stenflo

Research Report 2004:8

ISSN 1650-0377

Postal address:
Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:
http://www.math.su.se/matstat

Mathematical Statistics
Stockholm University
Research Report **2004:8**,
http://www.math.su.se/matstat

Central limit theorems for contractive Markov chains

Andreas Nordvall Lagerås and Örjan Stenflo*

Augusti 2004

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.

KEY WORDS: Markov chains, Iterated Function Systems, Central limit theorems, Stationary Measures, g-measures, Coupling.
MSC2000: 60F05, 60J05, 60B10, 37H99

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Stockholm, Sweden. E-mail: andreas@math.su.se, stenflo@math.su.se

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 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 Barnsley and Demko [1]. Hutchinson [11] 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 $K = \cup_{i=1}^N w_i(K)$. The set K is called the fractal set, or attractor, associated with the IFS. If the maps w_i are non-degenerated 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 the fact that fractals are totally deterministic objects, the most simple way to draw 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, μ, supported on K. Since

$$\frac{\sum_{j=0}^{n-1} f(Z_j(x))}{n} \to \int f d\mu \quad a.s., \tag{1}$$

as $n \to \infty$, 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 a way making the stationary distribution as uniform as possible on K and the stationary state is reached sufficiently fast. The choice of p_i's 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 (CLT), where

$$\frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} \left(f(Z_j(x)) - \int f d\mu \right) \tag{2}$$

converges to normal for f belonging to some suitably rich class of real-valued functions on X.

The purpose of this paper is to study Markov chains generated by iterated function systems with place-dependent probabilities. (Such Markov chains have also been studied under the name "Random systems with complete connections", see [12]). We are given a set of contraction maps $\{w_i\}$, with associated continuous functions $p_i = p_i(x)$, where p_i :

$X \to (0,1)$, 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 posses a unique stationary distribution, see [5], [18], [23], [24], and [3], but with some additional regularity conditions, uniqueness holds, see e.g. [10], [24], [13], and [25].

The operator T, (without the normalizing condition $\sum_i p_i(x) = 1$) is known as the Ruelle-Perron-Frobenius operator. Fan and Lau [9] proved a limit theorem for iterates of the Ruelle-Perron-Frobenius operator under the Dini-continuity assumptions on the p_i's, by lifting a similar result by Walters [26] on symbolic spaces. (Recall that p_i is Dini-continuous if $\int_0^1 \frac{\Delta_{p_i}(t)}{t} dt < \infty$, or equivalently $\sum_{n=1}^{\infty} \Delta_{p_i}(c^n) < \infty$, 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).

The Dini-condition is a condition that is somewhat stronger than the weakest known conditions for uniqueness in stationary probability distributions (in the normalized cases), 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 Fan and Lau [9]. This property is of importance when analyzing 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 central limit theorem, 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.

Central limit theorems for iterated random functions under conditions that imply exponential (or other rapid) rates of convergence have previously been proved in e.g. [20], [14], [15], [2], [28], and [27]. We discuss the connection between some of these results and our results in remarks 3 and 5 below.

2 Preliminaries

Let $\mathcal{B}$ denote the Borel σ-field generated by the metric d, and let $\mathbf{P} : X \times \mathcal{B} \to [0,1]$ 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 μ is stationary for $\mathbf{P}$ if $\mu(\cdot) = \int_X \mathbf{P}(x,\cdot)d\mu(x)$.

There are several ways to represent a Markov chain with a given transfer operator. One common way is to find a measurable function $w : X \times [0,1] \to X$, 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

$$Z_n(x) := w_{I_n} \circ w_{I_{n-1}} \circ \cdots \circ w_{I_1}(x),\ n \geq 1, \quad Z_0(x) := x, \tag{3}$$

for any $x \in X$, where

$$w_s(x) = w(x,s). \tag{4}$$

If (X,d) is a complete, separable, metric space then 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 e.g. [17].

For two fixed points $x, y \in X$, and $\mathbf{x} = (x,y)$ we can consider the Markov chain $\{\mathbf{Z}_n(\mathbf{x})\}$, on X^2, where $\mathbf{Z}_n(\mathbf{x}) := (Z_n(x), Z_n(y))$. When proving theorems based on contraction conditions we are typically interested in representations that minimizes $d(Z_n(x), Z_n(y))$ (in some average sense). (Note that the random map $\mathbf{Z}_n$ does not depend on $\mathbf{x}$ here.)

More generally, if $W : X^2 \times [0,1] \to X^2$, 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

$$\mathbf{Z}_n(\mathbf{x}) := W_{I_n} \circ W_{I_{n-1}} \circ \cdots \circ W_{I_1}(\mathbf{x}),\ n \geq 1, \quad \mathbf{Z}_0(\mathbf{x}) := \mathbf{x}, \tag{5}$$

where $W_s(\mathbf{x}) = W(\mathbf{x}, s)$, such that, for any $\mathbf{x} = (x,y) \in X^2$, the Markov chain $\mathbf{Z}_n(\mathbf{x}) =: (Z_n^{(x,y)}(x), Z_n^{(x,y)}(y))$ 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 $d_w(\pi,\nu) = \sup(\int f d(\pi - \nu); f : X \to \mathbb{R}, |f(x) - f(y)| \leq d(x,y)\, \forall x, y)$, for probability measures π and ν on X. The Monge-Kantorovich metric metrizes the topology of weak convergence on the set of probability measures on X, see e.g. Dudley [8]. It follows from the definitions, that for any stationary probability measure μ, we have

$$d_w(\mathbf{P}^n(x,\cdot), \mu(\cdot)) \leq \sup_{x,y \in X} Ed(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y)). \tag{6}$$

Therefore if $\sup_{x,y} Ed(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y)) \to 0$ as $n \to \infty$, 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 central limit theorem.

Proposition 1. *Suppose there exists a unique stationary distribution μ 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 $\delta > 0$,

$$\lim_{n\to\infty} n^{-1/2}(\log n)^{1+\delta} \sup_{x,y\in X} E\Big|\sum_{k=0}^{n-1}\Big(f(Z_k^{(x,y)}(x)) - f(Z_k^{(x,y)}(y))\Big)\Big| = 0. \tag{7}$$

Then the limit

$$\sigma^2 = \sigma^2(f) := \lim_{n\to\infty} \frac{1}{n} E\Big(\sum_{k=0}^{n-1}(f(Z_k(Z)) - \int f d\mu)\Big)^2 \tag{8}$$

exists and is finite, where Z is a μ-distributed random variable, independent of $\{I_j\}_{j=1}^{\infty}$, and

$$\lim_{n\to\infty} P\Big(\frac{\sum_{k=0}^{n-1}(f(Z_k(x)) - Ef(Z_k(x)))}{\sqrt{n}} \leq t\Big) = \Phi(\sigma t), \tag{9}$$

and

$$\lim_{n\to\infty} P\Big(\frac{\sum_{k=0}^{n-1}(f(Z_k(x)) - \int f d\mu)}{\sqrt{n}} \leq t\Big) = \Phi(\sigma t), \tag{10}$$

hold for any $x \in X$, where

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-s^2/2} ds$$

is the distribution function for a standardized normal random variable.

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

Remark 2. *General central limit theorems for Markov chains started at a point has been proved by Derriennic and Lin [6]. Proposition 1 above complements their result in cases of "uniform" ergodicity. The proof of Proposition 1 below relies on a slightly stronger result by Maxwell and Woodroofe [19] for Markov chains starting according to the unique stationary probability distribution.*

Remark 3. *The conditions in the central limit theorem (Theorem 2) of Wu and Woodroofe [28] imply (7) 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 centered 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 apriori information about the (possibly non-unique) stationary measures. The following corollary circumvents these problems and might therefore be more directly applicable in cases when (X,d) is compact;

Corollary 1. *If*

$$\lim_{n\to\infty} \sup_{x,y\in X} Ed(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y)) = 0, \tag{11}$$

then there exists a unique stationary distribution μ for $\mathbf{P}$.

Let f be a real-valued continuous function on X. Suppose $\Delta_f : \mathbb{R}^+ \to \mathbb{R}^+$ 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 (11), that for some $\delta > 0$,

$$\lim_{n\to\infty} \sqrt{n}(\log n)^{1+\delta} \Delta_f\Big(\sup_{x,y\in X} Ed(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y))\Big) = 0, \tag{12}$$

also holds, then then the conclusions of Proposition 1 hold for f, i.e. the limit (8) exists for f and is finite, and (9) and (10) hold.

Remark 4. *The function Δ_f may thus be chosen to be the modulus of uniform continuity of f in cases when this function is concave.*

Remark 5. *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 for average-contractive IFSs with place-independent probabilities, then it follows from Corollary 1 that the central limit theorem holds with respect to any f of modulus of uniform continuity Δ_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 Benda [2] (who considered Lipschitz-continuous f's). Wu and Shao [27] considered functions f that are stochastically Dini-continuous with respect to the stationary distribution. (It should be noted that [2] and [27] treated 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 (7).

Since

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

it follows from Corollary 1 of Maxwell and Woodroofe [19] that the limit

$$\sigma^2=\sigma^2(f):=\lim_{n\to\infty}\frac{1}{n}E\Big(S_nf(Z)-n\int f d\mu\Big)^2 \tag{13}$$

exists and is finite, where $S_nf(x):=\sum_{k=0}^{n-1}f(Z_k(x))$, and Z is a μ-distributed random variable, independent of $\{I_j\}_{j=1}^{\infty}$, and

$$\lim_{n\to\infty}P\Big(\frac{S_nf(Z)-n\int f d\mu}{\sqrt{n}}\le t\Big)=\Phi(\sigma t). \tag{14}$$

By Chebyshevs inequality

$$P\Big(\frac{|S_nf(x)-S_nf(Z)|}{\sqrt{n}}\ge\epsilon\Big)\ \le\ \frac{E|S_nf(x)-S_nf(Z)|}{\epsilon\sqrt{n}}, \tag{15}$$

for any $\epsilon>0$, and since

$$E|S_nf(x)-S_nf(Z)|\ \le\ \sup_{x,y\in X}E|S_nf(x)-S_nf(y)| \tag{16}$$

it follows from (15) and (16) and assumption (7), that

$$P\Big(\frac{|S_nf(x)-S_nf(Z)|}{\sqrt{n}}\ge\epsilon\Big)\le\frac{\sup_{x,y}E|\sum_{k=0}^{n-1}(f(Z_k^{(x,y)}(x))-f(Z_k^{(x,y)}(y)))|}{\epsilon\sqrt{n}}\to 0, \tag{17}$$

as $n \to \infty$. By Slutsky's theorem, see e.g. [4], it therefore follows from (14) and (17) that

$$\lim_{n\to\infty} P\Big(\frac{S_n f(x) - n\int f d\mu}{\sqrt{n}} \le t\Big) = \Phi(\sigma t), \tag{18}$$

for any $x \in X$, and since from the invariance of μ and assumptions,

$$\begin{aligned}\left|\frac{ES_n f(x) - n\int f d\mu}{\sqrt{n}}\right| &= \left|\frac{E\sum_{k=0}^{n-1}\Big(f(Z_k(x)) - \int f(Z_k(y))d\mu(y)\Big)}{\sqrt{n}}\right| \\ &\le \frac{\sup_{x,y} E|\sum_{k=0}^{n-1}(f(Z_k^{(x,y)}(x)) - f(Z_k^{(x,y)}(y)))|}{\sqrt{n}} \to 0,\end{aligned}$$

as $n \to \infty$, it follows that also

$$\lim_{n\to\infty} P\Big(\frac{S_n f(x) - ES_n f(x)}{\sqrt{n}} \le t\Big) = \Phi(\sigma t), \tag{19}$$

for any $x \in X$. Since (13), (19), and (18) are just reformulations of (8), (9), and (10) this completes the proof of Proposition 1.

□

Proof. (Corollary 1) The first part of the corollary follows from (6) above.

For the proof of the second part of Corollary 1, first note that by assumption (12)

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

implying that,

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

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

Thus,

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

Since by the definition of Δ_f and Jensen's inequality,

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

and

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

we see that an application of Proposition 1 completes the proof of the second part of Corollary 1.

□

3 Main results

Theorem 1. *Let $W : X^2 \times [0,1] \to X^2$, 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 $P(W^{(x,y)}(x) \in \cdot) = \mathbf{P}(x,\cdot)$ and $P(W^{(x,y)}(y) \in \cdot) = \mathbf{P}(y,\cdot)$, where P denotes the uniform probability measure on the Borel subsets of $[0,1]$.*

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

$$P(d(W^{(x,y)}(x), W^{(x,y)}(y)) \leq cd(x,y)) \geq 1 - \Delta(d(x,y)), \tag{20}$$

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

Then

(i) (Distributional stability theorem)

$$d_w(\mathbf{P}^n(x,\cdot), \mu(\cdot)) \leq \sup_{x,y\in X} Ed(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y)) \leq ED_n, \tag{21}$$

where D_n is a homogeneous Markov chain with $D_0 = diam(X) := \sup_{x,y} d(x,y)$,

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

and

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

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

If

$$\sum_{n=1}^{\infty} \prod_{k=1}^{n} (1 - \Delta(c^k)) = \infty, \tag{22}$$

then $ED_n \to 0$ and thus by Corollary 1 there is a unique stationary distribution, μ.

(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 $\alpha > 1/2$.

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

$$\begin{aligned} Z_n^{(x,y)}(x) &= W^{(Z_{n-1}^{(x,y)}(x), Z_{n-1}^{(x,y)}(y))}(Z_{n-1}^{(x,y)}(x)) \text{ and} \\ Z_n^{(x,y)}(y) &= W^{(Z_{n-1}^{(x,y)}(x), Z_{n-1}^{(x,y)}(y))}(Z_{n-1}^{(x,y)}(y)), \end{aligned}$$

as in (5). 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 (20) that

$$P(d(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y)) \leq ct \mid d(Z_{n-1}^{(x,y)}(x), Z_{n-1}^{(x,y)}(y)) \leq t)$$

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

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

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 (21) follows from (6).

Since $\{D_n\}$ is a non-ergodic Markov chain under condition (22) (see e.g. [21], p. 80), it follows that $ED_n \to 0$ as $n \to \infty$, if (22) 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 $\sum_{k=1}^{\infty} \Delta(c^k) < \infty$, implies that D_n is transient (see e.g. [21], p. 80). Therefore, see e.g. Shiryaev [22] p.575, $\sum_{k=0}^{\infty} P(D_k = diam(X)) < \infty$, and it follows that

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

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 Δ_f satisfying $\Delta_f(t) \leq c_1 t^\alpha$, for some constants c_1 and $\alpha > 1/2$, and any $t \geq 0$, it follows that for any $\delta > 0$,

$$\lim_{n\to\infty} \sqrt{n}(\log n)^{1+\delta}\Delta_f\Big(\sup_{x,y\in X} Ed(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y))\Big)$$
$$\leq \lim_{n\to\infty} \sqrt{n}(\log n)^{1+\delta} c_1 (c_0/n)^\alpha = 0.$$

An application of Corollary 1 now completes the proof of Theorem 1 (ii). □

4 Iterated function systems with place-dependent probabilities

Let $\{w_i\}_{i=1}^\infty$ be a set of uniformly 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
$$\Delta(t) = \frac{1}{2}\sup_{d(x,y)\leq t}\sum_{i=1}^\infty |p_i(x) - p_i(y)| = 1 - \inf_{d(x,y)\leq t}\sum_{i=1}^\infty \min(p_i(x), p_i(y)),$$
and let for any two points $x, y \in X$, $W^{(x,y)}(x)$ and $W^{(x,y)}(y)$ be random variables defined by
$$P(W^{(x,y)}(x) = w_i(x), W^{(x,y)}(y) = w_i(y)) = \min(p_i(x), p_i(y)) \tag{23}$$
and

$$P(W^{(x,y)}(x) = w_i(x), W^{(x,y)}(y) = w_j(y))$$
$$= \frac{(p_i(x) - \min(p_i(x), p_i(y)))(p_j(y) - \min(p_j(x), p_j(y)))}{1 - \sum_{k=1}^\infty \min(p_k(x), p_k(y))}, \tag{24}$$

when $i \neq j$. (If $p_i(x) = p_i(y)$, $\forall\ i$, then we understand the expression in (24) as zero.)

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

It follows from (23) 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:

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

Remark 6. *If $X = \{1, \ldots, N\}^{\mathbb{N}}$, and we for two elements $x = x_0x_1\ldots$ and $y = y_0y_1\ldots$ in X, 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 $\sum_{x_0=1}^{N} g(x_0x_1\ldots) = 1$ for all $x_1x_2\cdots \in X$. g describes an iterated function system with place-dependent probabilities: $\{(X,d), w_i(x), p_i(x), i \in \{1,\ldots,N\}\}$, where $w_i(x) = ix$ and $p_i(x) = g(ix)$, and Corollary 2 applies. This generalizes Theorem 1 in [25] and also implies a central limit theorem for the associated Markov chains under the "summable variations" condition used in [7] or [26]. Stationary probability measures for such Markov chains are sometimes called g-measures, a concept coined by Keane [16].*

References

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

[2] Benda, M. (1998). A central limit theorem for contractive stochastic dynamical systems. *J. Appl. Probab.* **35,** 200–205.

[3] Berger, N., Hoffman, C. and Sidoravicius, V. (2003). Nonuniqueness for specifications in $\ell^{2+\epsilon}$. *arXiv:math.PR/0312344*

[4] Billingsley, P. (1968). *Convergence of probability measures.* Wiley, New York.

[5] Bramson, M. and Kalikow, S. (1993). Non-uniqueness in g-functions. *Israel J. Math.* **84,** 153–160.

[6] Derriennic, Y. and Lin, M. (2003). The central limit theorem for Markov chains started at a point. *Probab. Theory Related Fields* **125,** 73–76.

[7] Doeblin, W. and Fortet, R. (1937). Sur les chaines à liaisons complètes. *Bull. Soc. Math. France* **65,** 132–148.

[8] Dudley, R. M. (1989). *Real analysis and probability.* Wadsworth & Brooks/Cole, Pacific Grove, CA.

[9] Fan, A. H. and Lau, K.-S. (1999). Iterated function system and Ruelle operator. *J. Math. Anal. Appl.* **231,** 319–344.

[10] Harris, T. E. (1955). On chains of infinite order. *Pacific J. Math.* **5,** 707–724.

[11] Hutchinson, J. E. (1981). Fractals and self-similarity. *Indiana Univ. Math. J.* **30,** 713–747.

[12] Iosifescu, M. and Grigorescu, S. (1990). *Dependence with complete connections and its applications.* Cambridge University Press.

[13] Johansson, A. and Öberg, A. (2003). Square summability of variations of g-functions and uniqueness of g-measures. *Math. Res. Lett.* **10,** 587-601.

[14] Kaijser, T. (1973). *Some limit theorems for Markov chains with applications to learning models and products of random matrices.* Doctoral Thesis, Institut Mittag-Leffler, Djursholm, Sweden.

[15] Kaijser, T. (1979). Another central limit theorem for random systems with complete connections. *Rev. Roumaine Math. Pures Appl.* **24,** 383–412.

[16] Keane, M. (1972). Strongly mixing g-measures. *Invent. Math.* **16,** 309–324.

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

[18] Lacroix, Y. (2000). A note on weak-$\star$ perturbations of g-measures. *Sankhyā Ser. A* **62,** 331–338.

[19] Maxwell, M. and Woodroofe, M. (2000). Central limit theorems for additive functionals of Markov chains. *Ann. Probab.* **28,** 713–724.

[20] Norman, M. F. (1972). *Markov processes and learning models.* Academic Press, New York-London.

[21] Prabhu, N. U. (1965). *Stochastic processes. Basic theory and its applications.* Macmillan, New York.

[22] Shiryaev, A. N. (1996). *Probability, 2nd ed.* Springer, New York.

[23] Stenflo, Ö. (2001). A note on a theorem of Karlin. *Statist. Probab. Lett.* **54,** 183–187.

[24] Stenflo, Ö. (2002). Uniqueness of invariant measures for place-dependent random iterations of functions. *IMA Vol. Math. Appl.* **132,** 13–32.

[25] Stenflo, Ö. (2003). Uniqueness in g-measures. *Nonlinearity* **16,** 403–410.

[26] Walters, P. (1975). Ruelle's operator theorem and g-measures. *Trans. Amer. Math. Soc.* **214,** 375–387.

[27] Wu, W. B. and Shao, X. (2004). Limit theorems for iterated random functions. *J. Appl. Probab.* **41,** 425–436.

[28] Wu, W. B. and Woodroofe, M. (2000). A central limit theorem for iterated random functions. *J. Appl. Probab.* **37,** 748–755.