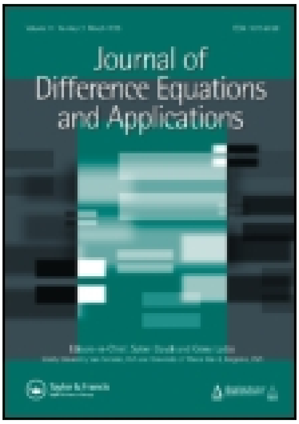
This article was downloaded by: [Uppsala universitetsbibliotek]
On: 16 August 2012, At: 00:29
Publisher: Taylor & Francis
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of Difference Equations and Applications
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gdea20

A survey of average contractive iterated function systems

Örjan Stenflo [a]

[a] Department of Mathematics, Uppsala University, 751 06, Uppsala, Sweden

Version of record first published: 09 Nov 2011

To cite this article: Örjan Stenflo (2012): A survey of average contractive iterated function systems, Journal of Difference Equations and Applications, 18:8, 1355-1380

To link to this article: http://dx.doi.org/10.1080/10236198.2011.610793

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Journal of Difference Equations and Applications
Vol. 18, No. 8, August 2012, 1355–1380

Taylor & Francis
Taylor & Francis Group

A survey of average contractive iterated function systems

Örjan Stenflo*

Department of Mathematics, Uppsala University, 751 06 Uppsala, Sweden

(*Received 22 November 2010; final version received 28 July 2011*)

Iterated function systems (IFSs) are useful for creating fractals, interesting probability distributions and enable a unifying framework for analysing stochastic processes with Markovian properties. In this paper, we present a survey of some basic results within the theory of random iterations of functions from an IFS based on average contraction conditions.

Keywords: Markov chains; iterated function systems; fractals; coupling from the past

2000 Mathematics Subject Classification: Primary: 60J05; Secondary: 28A80; 37H99; 60F05; 60J10; 65C05

Downloaded by [Uppsala universitetsbibliotek] at 00:29 16 August 2012

1. Introduction

Let (X, d) be a complete separable metric space. Any homogeneous Markov chain, $\{X_n\}$, on X can be represented in the form of a solution of a stochastic difference equation of the form

$$X_{n+1} = f(X_n, I_n), \tag{1}$$

where $f : X \times [0, 1] \rightarrow X$ is a measurable function and $\{I_n\}$ is a sequence of independent and identically distributed (i.i.d.) random variables. The randomness involved for a (homogeneous) Markov chain is thus nothing more complicated than that of a sequence of independent (and identically distributed) random variables.

Writing $f_s(x) := f(x, s)$, we may express (1) as $X_{n+1} = f_{I_n}(X_n)$ and thus regard $\{X_n\}$ as having been generated from random (i.i.d.) iterations of functions. The set of possible functions to iterate in each step, $\{f_s\}$, is called an iterated function system (IFS).

In this paper, we will present a survey of some variants of a basic convergence result in the theory of iterated random functions based on average contraction conditions and some applications exploiting the simple i.i.d. randomness of Markov chains.

The most well-known survey of iterated random functions with an extensive bibliography is the paper by Diaconis and Freedman [14]. See also Kaijser [31], the introduction paper in Stenflo [44] and Iosifescu [24] for other surveys. Limit theory for stochastic dynamical systems of the form (1) can also be found in many other sources including books by Borovkov [10], Bhattacharya and Majumdar [9] and Meyn and Tweedie [36].

This paper complements the above surveys in that it contains a more detailed discussion of different kinds of average contraction conditions appearing in the literature,

*Email: stenflo@math.uu.se

ISSN 1023-6198 print/ISSN 1563-5120 online
© 2012 Taylor & Francis
http://dx.doi.org/10.1080/10236198.2011.610793
http://www.tandfonline.com

Downloaded by [Uppsala universitetsbibliotek] at 00:29 16 August 2012

see Sections 2 and 3. Our basic distributional convergence theorems based on average contraction conditions (Theorems 1 and 2) are proved in Section 2.

The contraction conditions used in Section 2 are global. In Section 3, we discuss local average contraction conditions giving rise to equicontinuous Markov chains discussed in Section 4.

Many interesting probability measures supported, e.g. on fractal sets can be generated using random iterations, most commonly as stationary probabilities for Markov chains constructed according to (1), but also by other means. A general framework for describing fractals and probability measures supported on fractals is presented in Section 5.

Basic courses in finite state space Markov chains traditionally lack discussions of representations of the form (1) when proving convergence results with the unfortunate consequence that connections to basic time-series models, e.g. the autoregressive moving average model, appear vague. Hopefully, Sections 6 and 7 could serve as an introduction to the present unifying average- contractive- iterated-random-functions framework for readers missing this piece.

There are many fields of mathematics where IFSs can be used. See for example Stenflo [46] for a survey of place-dependent random iterations and connections to the Ruelle-Perron-Frobenius theorem of statistical mechanics, Roberts and Rosenthal [39] for a survey of Markov Chain Monte Carlo algorithms, Kaijser [30] for connections to the theory of products of random matrices, Iosifescu and Grigorescu [25] for connections to the theory of continued fractions, Barnsley [3] and Falconer [18] for more on connections to the theory to fractals, Jorgensen [29] for connections to representation theory and the theory of wavelets and Pesin [37] and Keller [32] for connections to the theory of (deterministic) dynamical systems. Naturally, the above 'classification' of topics is rough, and there is of course a substantial overlap and many areas are not mentioned.

1.1 Preliminaries: iterated function systems and Markov chains

Let (X, d) be a Polish (= complete separable metric) space, and let S be a measurable space. Consider a measurable function $f : X \times S \to X$. For each fixed $s \in S$, we write $f_s(x) := f(x, s)$. Following the terminology introduced by Barnsley and Demko [5][1], we call the set $\{(X, d); f_s, s \in S\}$ an IFS.

Let $\{I_n\}_{n=1}^{\infty}$ be a sequence of i.i.d. random variables with values in S. Let $P(\cdot) = P(I_1 \in \cdot)$ denote their common distribution. We call the set $\{(X, d); f_s, s \in S, P\}$ an IFS with probabilities.

Define for each fixed $x \in X$,

$$Z_n(x) := f_{I_n} \circ \cdots \circ f_{I_1}(x), \quad n \geq 1, \quad Z_0(x) = x.$$

The sequence $\{Z_n(x)\}$ is a Markov chain. Any Markov chain can be generated by an IFS with probabilities, in this way, see, e.g. Kifer [33]. An IFS representation of a Markov chain is typically not unique, see, e.g. Stenflo [44] for some simple examples.

Define for $g \in C(X)$, the bounded continuous functions $g : X \to \mathbb{R}$,

$$T^n g(x) = Eg(Z_n(x)), \quad n \geq 0.$$

Let T^* denote the adjoint of the operator T. T^* satisfies the equation $T^{*n}\nu(\cdot) = \int_X P(Z_n(x) \in \cdot) \mathrm{d}\nu(x)$, for any $n \geq 0$ and probability measure ν.

A probability measure, μ, satisfying the equation $T^*\mu = \mu$ is called an invariant probability measure. A Markov chain 'starting' according to an invariant probability

Downloaded by [Uppsala universitetsbibliotek] at 00:29 16 August 2012

measure is a stationary sequence. We therefore also refer to μ as a stationary distribution for the Markov chain.

Stationary distributions exist under mild conditions, see, e.g. Meyn and Tweedie [36] or Szarek [47].

In this paper, we will discuss average contraction conditions on $\{(X,d); f_s, s \in S, P\}$ ensuring the existence of a unique attractive stationary probability measure, μ in the sense that $T^{*n}\nu$ converges weakly to μ, for any probability measure ν.

We want the convergence to be as quick as possible. The rate of convergence can be measured in various ways for instance by metrics metrizing the topology of weak convergence, see Section 2 below.

An interesting object in our analysis is the reversed iterates

$$\hat{Z}_n(x) := f_{I_1} \circ \cdots \circ f_{I_n}(x), \quad n \geq 1, \quad \hat{Z}_0(x) = x. \tag{2}$$

The random variables $Z_n(x)$ and $\hat{Z}_n(x)$ are identically distributed for each fixed $n \geq 0$ and $x \in X$. We may, therefore, prove distributional limit theorems for $Z_n(x)$ by instead studying the pathwise more well-behaved $\hat{Z}_n(x)$.

Under typical conditions of ergodicity for a Markov chain, an IFS representation may be chosen such that the limit

$$\hat{Z} = \lim_{n\to\infty} \hat{Z}_n(x) \tag{3}$$

exists a.s., and the limiting random variable is independent of x.

In particular, if the Markov chain satisfies the Doeblin condition (which is a generalization of the standard conditions of ergodicity for finite-state-space Markov chains), then the Markov chain can be generated by an IFS with probabilities having the property that it actually exists a random positive integer T (finite with probability 1) such that $\{\hat{Z}_n(x)\}_{n\geq T}$ does not depend on x, see Athreya and Stenflo [2]. Propp and Wilson [38] discovered that this can be used to simulate 'exact' samples from the stationary distribution, μ, of the Markov chain since $\hat{Z}_T(x)$ is then a random variable with distribution $\mu(\cdot) = \lim_{n\to\infty} P(Z_n(x) \in \cdot)$. Details on such a construction in the special case when the state space is finite are presented in Section 6.

2. Global average contraction conditions for IFSs

In Stenflo [44] and Barnsley et al. [7], we proved the following theorem:

THEOREM 1. *Suppose*

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

$$\sup_{x\neq y} E\frac{d(Z_1(x), Z_1(y))}{d(x,y)} \leq c \quad \text{(global average contraction).} \tag{4}$$

(*B*) $Ed(x_0, Z_1(x_0)) < \infty$, *for some* $x_0 \in X$.

Then there exists a random variable $\hat{Z}$ *such that for any* $x \in X$ *there exists a constant* γ_x *such that*

$$Ed(\hat{Z}_n(x), \hat{Z}) \leq \gamma_x c^n, \quad n \geq 0. \tag{5}$$

Downloaded by [Uppsala universitetsbibliotek] at 00:29 16 August 2012

The constant γ_x satisfies

$$\gamma_x = Ed(x, \hat{Z}) \le d(x, x_0) + \frac{Ed(x_0, Z_1(x_0))}{1 - c}.$$

Theorem 2 below is a slightly more notationally involved theorem generalizing Theorem 1.

Suppose, for the moment, that $\{I_n\}_{n=-\infty}^{\infty}$ is (two-sided and) i.i.d. Define for $n \ge m$

$$Z_{n|m}(x) := f_{I_n} \circ \cdots \circ f_{I_{m+1}}(x), \quad n > m, \quad Z_{m|m}(x) = x.$$

Thus, $Z_{n|m}(x)$ represents the location of a Markov chain at time n 'starting' at $x \in X$ at time m.

THEOREM 2. *Suppose:*

($A*$) *There exists an integer $n_0 \ge 0$ and a constant $c < 1$ such that*

$$Ed(Z_{n_0+1|0}(x), Z_{n_0+1|0}(y)) \le c \;\; Ed(Z_{n_0|0}(x), Z_{n_0|0}(y)) < \infty, \;\; \text{for any} \;\; x, y \in X.$$

($B*$) $Ed(Z_{n_0|0}(x_0), Z_{n_0|-1}(x_0)) < \infty$, *for some $x_0 \in X$.*

Then the limits

$$Z_{n|-\infty} = \lim_{m \to \infty} Z_{n|-m}(x) \tag{6}$$

exist almost surely for any $n \in \mathbb{Z}$, and the a.s. limits are independent of x.

The rate of convergence in (6) *is exponential. More precisely, for any $x \in X$, there exists a constant γ_x such that*

$$Ed(Z_{n|n-m}(x), Z_{n|-\infty}) \le \gamma_x c^m, \tag{7}$$

for any $n \in \mathbb{Z}$ and any $m \ge n_0$.

If $X_n := Z_{n|-\infty}$, then $\{X_n\}_{n=-\infty}^{\infty}$ is stationary and ergodic.

Remark 1. Theorems 1 and 2 are essentially equivalent:

Theorem 1 corresponds to the special case when $n_0 = 0$ in Theorem 2. Note that $\{Z_{n|n-m}(x)\}_{m=0}^{\infty}$ has the same joint distributions as $\{\hat{Z}_m(x)\}_{m=0}^{\infty}$ for any fixed n and x.

In essence, we obtain Theorem 2 by changing distance in Theorem 1 to $d*(x, y) = Ed(Z_{n_0}(x), Z_{n_0}(y))$, waving hands slightly since this is not necessarily a metric. The method of changing metric will be discussed more in Section 3.

Proof. (Theorem 2)

The statement of the theorem was inspired by the 'changing distance in Theorem 1'–heuristics described in the remark above and the structure of the proof is the same as the one given in Stenflo [44] (Paper C) corresponding to Theorem 1. For completeness we give some details;

Let n be an arbitrary fixed integer. From condition ($A*$) and the i.i.d. assumption of $\{I_n\}$ we have

$$Ed(Z_{n|n-n_0-1}(x), Z_{n|n-n_0-1}(y)) \le cEd(Z_{n|n-n_0}(x), Z_{n|n-n_0}(y))$$

for any $x, y \in X$.

By induction

$$\begin{aligned} Ed(Z_{n|n-k}(x), Z_{n|n-k}(y)) &\leq cEd(Z_{n|n-k+1}(x), Z_{n|n-k+1}(y)) \\ &\leq \ldots (\text{recursion}) \ldots \\ &\leq c^{k-n_0} Ed(Z_{n|n-n_0}(x), Z_{n|n-n_0}(y)) \end{aligned} \tag{8}$$

for any $x, y \in X$ and any $k \geq n_0$.

Thus,

$$\begin{aligned} \sum_{k=m}^{\infty} Ed(Z_{n|n-k}(x), Z_{n|n-k-1}(x)) &= \sum_{k=m}^{\infty} Ed(Z_{n|n-k}(x), Z_{n|n-k}(f_{I_{n-k}}(x))) \\ &= \sum_{k=m}^{\infty} c^{k-n_0} Ed(Z_{n|n-n_0}(x), Z_{n|n-n_0}(f_{I_{n-k}}(x))) \\ &= \sum_{k=m}^{\infty} c^{k-n_0} Ed(Z_{n_0|0}(x), Z_{n_0|-1}(x)) \\ &= c^{m-n_0} \frac{Ed(Z_{n_0|0}(x), Z_{n_0|-1}(x))}{1-c}, \end{aligned} \tag{9}$$

for any $m \geq n_0$.

Therefore, if $\delta_m(x) := \sum_{k=m}^{\infty} d(Z_{n|n-k}(x), Z_{n|n-k-1}(x))$ then using condition $(B*)$ we see that $E\delta_m(x_0) \to 0$, which by monotonicity and the Chebychev inequality imply that $\delta_m(x_0) \to 0$ a.s. as $m \to \infty$.

Since X is complete it therefore follows that the limit

$$Z_{n|-\infty} := \lim_{m\to\infty} Z_{n|-m}(x_0)$$

exists for any n.

Since from (8) and condition $(A*)$ it follows that $d(Z_{n|-m}(x_0), Z_{n|-m}(x)) \to 0$, as $m \to \infty$, for any $x \in X$, a.s., it follows that $Z_{n|-\infty}$ must be independent of x_0.

The convergence rate is quantified by

$$\begin{aligned} Ed(Z_{n|n-m}(x), Z_{n|-\infty}) &= Ed(Z_{n|n-m}(x), Z_{n|n-m}(Z_{n-m|-\infty})) \\ &\leq c^{m-n_0} Ed(Z_{n|n-n_0}(x), Z_{n|n-n_0}(Z_{n-m|-\infty})) \\ &= c^{m-n_0} Ed(Z_{n|n-n_0}(x), Z_{n|n-n_0}(Z_{n-n_0|-\infty})) \\ &= c^{m-n_0} Ed(Z_{n|n-n_0}(x), Z_{n|-\infty}) \; = \; \gamma_x c^m \end{aligned}$$

for any $m \geq n_0$, where

$$\begin{aligned} \gamma_x &= c^{-n_0} Ed(Z_{n|n-n_0}(x), Z_{n|-\infty}) = c^{-n_0} Ed(Z_{n_0|0}(x), Z_{n_0|-\infty}) \\ &\leq c^{-n_0} \left(Ed(Z_{n_0|0}(x), Z_{n_0|0}(x_0)) + \frac{Ed(Z_{n_0|0}(x_0), Z_{n_0|-1}(x_0))}{1-c} \right) < \infty, \end{aligned}$$

where the last inequality follows from the triangle inequality and (9). If we define $X_n = Z_{n|-\infty}$, then X_n is a measurable function of $I_n, I_{n-1}, \ldots$ and since $\{I_n\}$ is a stationary and ergodic sequence it follows by standard results in ergodic theory that also $\{X_n\}$ is stationary and ergodic. This completes the proof of Theorem 2. □

Downloaded by [Uppsala universitetsbibliotek] at 00:29 16 August 2012

Remark 2. See Elton [17] for related results in the more general setting when $\{I_n\}$ is assumed to be 'two-sided' and stationary. Despite the generality of Theorem 2, we will restrict our attention to the case when $\{I_n\}$ is one-sided and i.i.d. for the rest of this paper, keeping Remark 1 in mind.

Remark 3. All contraction conditions discussed in this paper can be made weaker by replacing $Z_1(x)$ with $Z_{m_0}(x)$ for some m_0 (and similarly the corresponding quantities in the two-sided case). By considering subsequences $\{Z_{km_0}(x)\}_{k=0}^{\infty}$ we see that it is sufficient to consider the case $m_0 = 1$.

Probability metrics

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 $\|f\|_{BL} = \|f\|_\infty + \|f\|_L$, and let BL_1 denote the set of $f \in BL$, with $\|f\|_{BL} \leq 1$. Let $M(X)$ denote the set of Borel probability measures on X. For probability measures ν_1, $\nu_2 \in M(X)$ define the metric

$$d_w(\nu_1, \nu_2) = \sup_{f\in BL_1} \{|\int_X fd(\nu_1 - \nu_2)|\}.$$

It is well known, see, e.g. Shiryaev [41], that this metric metrizes the topology of weak convergence of probability measures (on separable metric spaces).

Another metric metrizing the topology of weak convergence of probability measures is the Prokhorov metric;

For a set $A \subset X$ and $\epsilon > 0$, let $A^\epsilon = \{y \in X : d(x, y) < \epsilon \text{ for some } x \in A\}$. The Prokhorov distance between two probability measures ν_1, $\nu_2 \in M(X)$ is defined as

$$\rho_{\text{Prokhorov}}(\nu_1, \nu_2) = \inf\{\epsilon > 0 \,|\, \nu_1(A) \leq \nu_2(A^\epsilon) + \epsilon, \ \text{ for all Borel sets } A \subseteq X\}.$$

Remark 4. If we let μ_n^x and μ denote the probability distributions of $Z_n(x)$ and $\hat{Z}$, respectively, then by definition of the d_w-metric

$$d_w\left(\mu_n^x, \mu\right) \leq Ed(\hat{Z}_n(x), \hat{Z}). \tag{10}$$

Since

$$\rho_{\text{Prokhorov}}\left(\mu_n^x, \mu\right) \leq \sqrt{\frac{3}{2} d_w(\mu_n^x, \mu)},$$

see Dudley [15], p. 398, Problem 5, it follows from (5) that we have exponential convergence rates in both the d_w and the Prokhorov metric under the conditions of Theorem 1.

It is common in the literature to assume a log-contraction condition instead of the more restrictive contraction condition of Theorem 1. This conspicuous improvement is, however, more of cosmetic nature, since cases when log-contraction conditions are fulfilled but average contraction conditions are not typically indicates that a change of

Downloaded by [Uppsala universitetsbibliotek] at 00:29 16 August 2012

Downloaded by [Uppsala universitetsbibliotek] at 00:29 16 August 2012

metric is appropriate. For example, if

$$\sup_{x \neq y} E \log \frac{d(Z_1(x), Z_1(y))}{d(x, y)} < 0, \tag{11}$$

then there exists a q with $0 < q \leq 1$ such that

$$\sup_{x \neq y} E \left(\frac{d(Z_1(x), Z_1(y))}{d(x, y)} \right)^q < 1, \tag{12}$$

provided the left-hand side of (12) is finite for some $q > 0$. To see this, note that the function

$$f(q) := \sup_{x \neq y} E \left(\frac{d(Z_1(x), Z_1(y))}{d(x, y)} \right)^q$$

satisfies $f(0) = 1$ and $f'(0) < 0$, by (11). Since d^q is a metric for $0 < q \leq 1$, (12) reduces to the ordinary average contraction condition in this new metric. By applying Theorem 1, we obtain the following.

Corollary 1. *Suppose:*

($\tilde{A}$)

$$\sup_{x \neq y} E \log \frac{d(Z_1(x), Z_1(y))}{d(x, y)} < 0 \quad (\text{global log} - \text{contraction}).$$

($\tilde{B}$) $Ed^{q_0}(x_0, Z_1(x_0)) < \infty$, *for some* $x_0 \in X$ *and* $q_0 > 0$, *and*

$$\sup_{x \neq y} E \left(\frac{d(Z_1(x), Z_1(y))}{d(x, y)} \right)^{q_0} < \infty.$$

Then for some $c < 1$ *and* $q > 0$, *there exists a random variable* $\hat{Z}$ *such that for any* $x \in X$,

$$Ed^q \left(\hat{Z}_n(x), \hat{Z} \right) \leq \gamma c^n, \quad n \geq 0, \tag{13}$$

for some positive constant γ *(depending on* x*). Thus, if* μ_n^x *and* μ *denote the probability distributions of* $Z_n(x)$ *and* $\hat{Z}$*, respectively, then* $\rho_{\text{Prokhorov}}\left(\mu_n^x, \mu\right)$ *converges to zero with an exponential rate as* $n \to \infty$.

Remark 5. A variant of Corollary 1, in the special case when

$$\sup_{x \neq y} E \log \frac{d(Z_1(x), Z_1(y))}{d(x, y)} < 0 \quad (\text{global log} - \text{contraction})$$

is replaced by the more restrictive condition

$$E \log \sup_{x \neq y} \frac{d(Z_1(x), Z_1(y))}{d(x, y)} < 0, \quad (\text{global log} - \text{contraction (strong sense)}), \tag{14}$$

was proved in the survey paper by Diaconis and Freedman [14].

(Note that condition (14) implies that almost all maps, f_s, $s \in S$, are Lipschitz continuous.)

The paper by Diaconis and Freedman [14] sparked attention on theorems of the same kind as Theorem 1 and its consequences on other limit theorems, see, e.g. Wu and Woodroofe [50], Alsmeyer and Fuh [1], Jarner and Tweedie [28], Iosifescu [23], Wu and Shao [49] and Herkenrath and Iosifescu [20].

See Hutchinson and Rüschendorf [22], Stenflo [45] and Barnsley et al. [7] for results corresponding to Theorem 1 for random IFSs with probabilities.

Remark 6. Note that the, somewhat ugly, condition ($\tilde{B}$) is fulfilled if the family f_s, $s \in S$, is a finite collection of Lipschitz continuous maps. This is a natural assumption in the theory of fractals, see, e.g. Sections 5.1 and 5.2 below.

2.1 Global stability conditions

If we know apriori that an invariant measure exists, then the following global stability condition is sufficient for uniqueness:

PROPOSITION 1. *Suppose a Markov chain with an invariant probability measure, μ, satisfies*

$$\lim_{n\to\infty} |T^n g(x) - T^n g(y)| = 0, \tag{15}$$

for any $x, y \in X$, and $g \in BL$. Then $d_w(T^{*n}\nu, \mu) \to 0$ as $n \to \infty$, for any probability measure ν.

Proof. By Theorem 11.3.3. on p.395 in Dudley [15], it is sufficient to prove that

$$\int_X g \mathrm{d}T^{*n}\nu \to \int_X g \mathrm{d}\mu,$$

as $n \to \infty$, for all $g \in BL$.

The proof of Proposition 1 is completed by observing that by the invariance of μ, (15), and Lebesgue's dominated convergence theorem

$$\begin{aligned}\left|\int_X g\mathrm{d}T^{*n}\nu - \int_X g\mathrm{d}\mu\right| &= \left|\int_X T^n g(x)\mathrm{d}\nu(x) - \int_X T^n g(y) d\mu(y)\right| \\ &\leq \int_X \int_X |T^n g(x) - T^n g(y)| \mathrm{d}\nu(x)\mathrm{d}\mu(y) \to 0,\end{aligned}$$

as $n \to \infty$. □

Remark 7. Since

$$\begin{aligned}|T^n g(x) - T^n g(y)| &= |Eg(Z_n(x)) - Eg(Z_n(y))| \leq E|g(Z_n(x)) - g(Z_n(y))| \\ &\leq \|g\|_L Ed(Z_n(x), Z_n(y))\end{aligned}$$

Downloaded by [Uppsala universitetsbibliotek] at 00:29 16 August 2012

a sufficient condition for (15) is that

$$\lim_{n\to\infty} Ed(Z_n(x), Z_n(y)) = 0,$$

for all $x, y \in X$.

3. Local average contraction conditions for IFSs

There are many ways of expressing the idea of locally contractive Markov chains. Globally average contractive IFSs are convenient in that they can be analysed by many different methods including coarse operator-theoretical methods. The analysis of locally contractive systems requires a more refined analysis, and probabilistic methods usually play a central role. In this section, we discuss natural local average contraction conditions for IFSs. Locally average contractive IFSs have been studied by Kaijser [30] and later independently by Steinsaltz [42], [43] and Carlsson [13]. See also Jarner and Tweedie [28] and Lagerås and Stenflo [34]. Such systems generate equicontinuous Markov chains, discussed in Section 4.

Definition 1. *Let $\epsilon > 0$ be a fixed constant. We say that an IFS is ϵ-local (average) contractive if*

$$\sup_{0<d(x,y)<\epsilon} E\frac{d(Z_1(x), Z_1(y))}{d(x, y)} \leq c, \tag{16}$$

for some constant $c < 1$.

Definition 2. *We say that an IFS is locally contractive in the weak sense, if*

$$\sup_{x\in X}\left(\limsup_{y\to x} E\frac{d(Z_1(x), Z_1(y))}{d(x, y)}\right) \leq c, \tag{17}$$

for some constant $c < 1$.

Among these conditions, local contraction in the weak sense is clearly a less restrictive condition.

Remark 8. We will not discuss local log-average contraction conditions in detail since we can typically reduce such conditions to the above conditions via a change of metric, c.f. the arguments preceding Corollary 1.

3.1 Spaces without differences between local and global contraction conditions

Local contraction conditions are of course weaker than the corresponding global contraction conditions. It is a natural question to ask if local contractivity is indeed equivalent to global contractivity. In some cases, there is no difference:

Definition 3. *We say that (X, d) is a geodesically convex metric space if for any two points x and y in X there exists a geodesic curve γ between x and y, i.e. a continuous function $\gamma : [0, 1] \to X$ with $\gamma(0) = x$ and $\gamma(1) = y$, such that*

$$d(x, y) = \sup\left\{\sum_{i=0}^{n-1} d(\gamma(t_i), \gamma(t_{i+1})) : n \in \mathbb{N} \text{ and } 0 = t_0 < t_1 < \cdots < t_n = 1\right\}.$$

Downloaded by [Uppsala universitetsbibliotek] at 00:29 16 August 2012

Downloaded by [Uppsala universitetsbibliotek] at 00:29 16 August 2012

Lemma 1. *Let (X, d) be a geodesically convex metric space. Then a locally contractive IFS in the weak sense* (17) *is also globally contractive.*

Proof. (Lemma 1) Fix a real number c^* with $c < c^* < 1$. Consider two arbitrary points x and y in X, and let $\gamma(t)$ be a geodesic curve between x and y parametrized by the unit interval with the property that $\gamma(0) = x$ and $\gamma(1) = y$. Let $\{t_i\}_{i=0}^n$ be an increasing sequence in $[0, 1]$, with $t_0 = 0$ and $t_n = 1$ such that $Ed(Z_1(\gamma(t_i)), Z_1(\gamma(t_{i+1}))/d(\gamma(t_i), \gamma(t_{i+1})) \leq c^*$, for all $i = 0, \ldots, n-1$. Such a sequence exists by the compactness of [0,1].

We have

$$
\begin{aligned}
Ed(Z_1(x), Z_1(y)) &\leq \sum_{i=0}^{n-1} E \frac{d(Z_1(\gamma(t_i)), Z_1(\gamma(t_{i+1}))}{d(\gamma(t_i), \gamma(t_{i+1}))} d(\gamma(t_i), \gamma(t_{i+1})) \\
&\leq c* \sum_{i=0}^{n-1} d(\gamma(t_i), \gamma(t_{i+1})) \leq c* d(x, y)
\end{aligned}
$$

and thus the IFS is globally average contractive. □

As a consequence of Lemma 1, we may apply Theorem 1 for locally contractive IFSs in the weak sense (17) on a geodesically convex metric space:

Theorem 3. *If a locally contractive IFS in the weak sense* (17) *on a geodesically convex metric space satisfies condition* (*B*)*, then the conclusion of Theorem* 1 *holds, and thus in particular, there is a unique stationary probability measure* μ*, and the n-th step probability distributions converge to* μ *with exponential rate in the* d_w*-metric.*

Remark 9. Theorem 3 is not true if we replace condition (17) by local log-contractivity. The following example was considered by Carlsson [13]. Consider the IFS with probabilities, $\{[0, 1], f_1, f_2, P\}$, where $f_1(x) = (x + x^2)/2$ and $f_2(x) = \log_3(1 + 2x)$, $P(1) = 1/2, P(2) = 1/2$. This system is easily checked to be ϵ-locally log-contractive, i.e.

$$
E \log \sup_{0<d(x,y)<\epsilon} \frac{d(Z_1(x), Z_1(y))}{d(x, y)} < 0,
$$

for some $\epsilon > 0$, but has more than one invariant probability measure (since $x = 0$ and $x = 1$ are fixed points for both f_1 and f_2).

The geodesic convexity assumption of Theorem 3 implies that (X, d) is connected. The following example illustrates the necessity of that property in Theorem 3:

Example 1. Consider the function $w : [0, 1] \cup [2, 3] \to [0, 1] \cup [2, 3]$ defined by $w(x) = 0$, if $x \leq 1$ and $w(x) = 3$ if $x \geq 2$. This trivial 'system' is clearly locally contractive. Observe that any probability measure of the form $\mu = \alpha\delta_0 + (1 - \alpha)\delta_3$, where $0 \leq \alpha \leq 1$ and where δ_x denotes the Dirac probability measure concentrated at $x \in \{0, 3\}$, is a stationary probability measure, and thus there is not necessarily a unique stationary probability measure for a locally contractive system.

In general, we need to add some 'blending' condition in order to prove uniqueness in invariant probability measures for locally contractive systems, see Section 4.

3.2 Making weighted local contraction conditions non-weighted via a change of metric

Steinsaltz [42] and Carlsson [13] considered weighted local-contraction conditions in cases when (X, d) is Euclidean. In this section, we will extend their results and show how we can reduce these conditions to non-weighted ones via a change of metric.

Suppose (X, d) is geodesically convex and for any two points x and y in X we may express $d(x, y)$ as

$$d(x,y) = \inf_{\gamma\in\Gamma}\int_0^1 |\gamma'(t)|\mathrm{d}t,$$

where $|\gamma'(t)| = \limsup_{\delta\to 0} d(\gamma(t+\delta), \gamma(t))/|\delta|$, and the infimum is taken over the set Γ of all curves $\gamma : [0,1] \to X$ between x and y.

If $\phi : X \to (0,\infty)$ is a continuous function with $\inf_{x\in X}\phi(x) > 0$, then we can define a new metric $\hat{d}$ by

$$\hat{d}(x,y) = \inf_{\gamma\in\Gamma}\int_0^1 \phi(\gamma(t))|\gamma'(t)|\mathrm{d}t. \tag{18}$$

For any curve γ between x and y, $\bar{\gamma}(t) := Z_1(\gamma(t))$ is a curve between $Z_1(x)$ and $Z_1(y)$ (provided Z_1 is continuous). Therefore, by (18),

$$\hat{d}(Z_1(x), Z_1(y)) \leq \int_0^1 \phi(\bar{\gamma}(t))|\bar{\gamma}'(t)|\mathrm{d}t, \tag{19}$$

where

$$|\bar{\gamma}'(t)| \leq \limsup_{\delta\to 0} \frac{d(Z_1(\gamma(t+\delta)), Z_1(\gamma(t)))}{d(\gamma(t+\delta), \gamma(t))}|\gamma'(t)|. \tag{20}$$

Thus, from (19) and (20), we see that

$$\begin{aligned}
E\hat{d}(Z_1(x),Z_1(y)) &\leq \inf_{\gamma}\int_0^1 |\gamma'(t)|E\left(\phi(Z_1(\gamma(t)))\limsup_{\delta\to 0}\frac{d(Z_1(\gamma(t+\delta)),Z_1(\gamma(t)))}{d(\gamma(t+\delta),\gamma(t))}\right)\mathrm{d}t \\
&\leq \inf_{\gamma}\sup_{t} E\left(\frac{\phi(Z_1(\gamma(t)))}{\phi(\gamma(t))}\limsup_{\delta\to 0}\frac{d(Z_1(\gamma(t+\delta)),Z_1(\gamma(t)))}{d(\gamma(t+\delta),\gamma(t))}\right)\int_0^1 \phi(\gamma(t))|\gamma'(t)|\mathrm{d}t \\
&\leq \sup_{\gamma}\sup_{t} E\left(\frac{\phi(Z_1(\gamma(t)))}{\phi(\gamma(t))}\limsup_{\delta\to 0}\frac{d(Z_1(\gamma(t+\delta)),Z_1(\gamma(t)))}{d(\gamma(t+\delta),\gamma(t))}\right)\hat{d}(x,y) \\
&\leq \sup_{x} E\left(\frac{\phi(Z_1(x))}{\phi(x)}\limsup_{y\to x}\frac{d(Z_1(x),Z_1(y))}{d(x,y)}\right)\hat{d}(x,y)
\end{aligned}$$

Therefore, in particular, if

$$\sup_{x} E\left(\frac{\phi(Z_1(x))}{\phi(x)}\limsup_{y\to x}\frac{d(Z_1(x),Z_1(y))}{d(x,y)}\right) < 1, \tag{21}$$

then $(X,\hat{d})$ is globally contractive (4), and we may apply Theorem 1.

Conversely, if $\Phi : X \to X$ is $1-1$ and (X, d) is a metric space, then we can define another metric $\hat{d}$ on X by $\hat{d}(x,y) := d(\Phi(x), \Phi(y))$.

Downloaded by [Uppsala universitetsbibliotek] at 00:29 16 August 2012

Downloaded by [Uppsala universitetsbibliotek] at 00:29 16 August 2012

Since

$$E\frac{\hat{d}(Z_1(x),Z_1(y))}{\hat{d}(x,y)} = E\frac{d(\Phi(Z_1(x)),\Phi(Z_1(y)))}{d(\Phi(x),\Phi(y))}$$
$$= E\frac{d(\Phi(Z_1(x)),\Phi(Z_1(y)))}{d(Z_1(x),Z_1(y))}\frac{d(Z_1(x),Z_1(y))}{d(x,y)}\frac{d(x,y)}{d(\Phi(x),\Phi(y))},$$

we can express global contraction in the $\hat{d}$-metric without explicitly referring to it as

$$\sup_{x\neq y} E\frac{d(\Phi(Z_1(x)),\Phi(Z_1(y)))}{d(Z_1(x),Z_1(y))}\frac{d(Z_1(x),Z_1(y))}{d(x,y)}\frac{d(x,y)}{d(\Phi(x),\Phi(y))} < 1. \tag{22}$$

Remark 10. Condition (21) is the local contraction condition corresponding to (22) we would obtain if $\phi(x) := \lim_{y\to x} d(\Phi(x),\Phi(y))/d(x,y)$ exists for each $x\in X$. The existence of this limit is in general a very strong assumption.

Remark 11. Suppose we have an IFS which is not average contractive with respect to the original metric after one step but where we believe we have exponential convergence towards an equilibrium. Suppose for simplicity that $X=\mathbb{R}$, d is the Euclidean metric, all IFS maps are smooth and $EZ'_{n+1}(x) < cEZ'_n(x)$, for some constant $c<1$ and all $x\in\mathbb{R}$, for some n. If we let $\phi(x)=EZ'_n(x)$, then it follows that (21) holds. Therefore, we may change the metric to $\hat{d}$ (defined like in (18)) and obtain average contraction w.r.t. $\hat{d}$.

Typically, it is hard to exactly compute $EZ'_n(x)$. Locally estimated Lyapunov-exponents might, however, give a good guess what a weight function $\phi(x)$ giving global average contractivity w.r.t. the metric $\hat{d}$ could be.

See Steinsaltz [42], [43] and Carlsson [13] for further details and interesting applications, e.g. in the analysis of random logistic maps.

4. Equicontinuous Markov chains

Equicontinuity is a local stability condition for Markov chains satisfied for locally contractive systems. The first systematic studies of equicontinuous Markov chains originate from the 60th and papers by Jamison, Rosenblatt and others, see, e.g. [26], [27] and [40]. A survey is given in the book by Meyn and Tweedie [36], where it is assumed that the state space (X,d) is a locally compact metric space.

The following definition is taken from [36].

Definition 4. *Let $C_c(X)$ denote the set of compactly supported continuous $g: X\to\mathbb{R}$. We say that the Markov chain $\{Z_n(x)\}$ is an e-chain if $\{T^n g\}$ is equicontinuous on compact sets for any $g\in C_c(X)$.*

Consider the average over time of the n-step transition probabilities

$$\bar{\mu}_n^x(\cdot) := (1/n)\sum_{j=0}^{n-1} P(Z_j(x)\in\cdot). \tag{23}$$

If an e-chain $\{Z_n(x)\}$ is bounded in probability on average, i.e. if for any $\epsilon>0$ and $x\in X$ there exists a compact set K, such that

$$\bar{\mu}_n^x(K) \geq 1-\epsilon, \quad \text{for all } n\geq 0, \tag{24}$$

Downloaded by [Uppsala universitetsbibliotek] at 00:29 16 August 2012

then it is proved in [36] that $\bar{\mu}_n^x(\cdot)$ converges weakly to a stationary distribution μ^x for any $x \in X$, and the map $\Lambda_g(x) = \int g\mathrm{d}\mu^x$ is continuous for any $g \in C_\mathrm{c}(X)$.

Basic tools in the proofs are the Ascoli theorem, the Prokhorov theorem and the separability of $C_\mathrm{c}(X)$ (which is a consequence of the local compactness assumption of (X, d) used in [36]).

In order to prove uniqueness in stationary distributions, we also need that the Markov chain satisfies some 'blending' assumptions. (Examples of Markov chains with non-unique stationary probabilities are easily constructed by, e.g. letting the Markov chain have two or more disjoint absorbing sets, see, e.g. Remark 9. Such Markov chains have the property that there exist non-communicating states.)

If we for any fixed $g \in C_\mathrm{c}(X)$ and $x \in X$ define

$$M_n = M_n(g,x) = \int_X g\mathrm{d}\mu^{Z_n(x)}, \tag{25}$$

then it is straightforward to check that $\{M_n\}$ is a bounded martingale w.r.t. $\{Z_n(x)\}$.

This property enables a powerful Martingale technique for proving uniqueness in stationary distributions originating from [27] illustrated in Theorem 4 below.

Definition 5. *We say that the Markov chain $\{Z_n(x)\}$ is asymptotically irreducible if for any $x, y \in X$, there exists a compact set $K \subset X$ such that*

$$\lim_{n\to\infty} \inf_{(j,k)\in J_n(K)} d(Z_j(x), Z_k(y)) = 0, \quad a.s., \tag{26}$$

where $J_n(K) = \{(j,k) : j \ge n, k \ge n, Z_j(x) \in K, Z_k(y) \in K\}$.

Theorem 4. *If an e-chain is bounded in probability on average and asymptotically irreducible, then there exists a unique stationary probability measure μ, and*

$$\lim_{n\to\infty} d_w\left(\bar{\mu}_n^x, \mu\right) = 0,$$

for any $x \in X$.

Proof. Fix $g \in C_c(X)$ and $x \in X$. Since $\{M_n\}$ defined as in (25) is a bounded martingale, it follows from the Martingale convergence theorem that there exists a random variable $M = M(g,x)$ such that

$$\lim_{n\to\infty} \int_X g\mathrm{d}\mu^{Z_n(x)} = M, \quad a.s.,$$

and $E(M) = E\int_X g\mathrm{d}\mu^{Z_n(x)} = \int_X g\mathrm{d}\mu^x$.

Since $\int_X g\mathrm{d}\mu^x$ is continuous as function of x for any $g \in C_c(X)$, it follows from assumption (26) that

$$\lim_{n\to\infty}\left(\inf_{j,k\ge n}\left|\int_X g\mathrm{d}\mu^{Z_j(x)} - \int_X g\mathrm{d}\mu^{Z_k(y)}\right|\right) = 0, \quad a.s.$$

Thus,

$$M(g,x) = M(g,y) \quad a.s.$$

It follows that

$$\int_X g\mathrm{d}\mu^x = E(M) = \int_X g\mathrm{d}\mu^y,$$

and thus $\mu^x = \mu^y$. If we fix $x_0 \in X$, it thus follows that $\mu := \mu^{x_0} = \mu^x$ for any $x \in X$, and μ is uniquely invariant, since if ν is an arbitrary invariant probability measure, and $g \in C_c(X)$, then

$$\int g\mathrm{d}\nu = \int \left(\frac{\sum_{j=0}^{n-1} T^j g}{n} \right) \mathrm{d}\nu \to \int\int g\mathrm{d}\mu\mathrm{d}\nu = \int g\mathrm{d}\mu,$$

and thus $\nu = \mu$. □

Remark 12. A sufficient condition for a Markov chain $\{Z_n(x)\}$ to be asymptotically irreducible is that it has a reachable recurrent point $x^* \in X$ in the sense that for any $x \in X$,

$$\liminf_{n\to\infty} d(Z_n(x), x^*) = 0 \quad a.s., \tag{27}$$

i.e. every open neighbourhood of x^* is visited infinitely often by $\{Z_n(x)\}$ a.s. for any $x \in X$.

Remark 13. Sufficient conditions for a Markov chain to be bounded in probability on average can be found in Meyn and Tweedie [36]. See also Szarek [47] who considers the general case when (X, d) is a Polish (= complete and separable metric) space. Under further aperiodicity conditions (in addition to the assumptions of Theorem 4), it can be proved that the non-averaged sequence $\mu_n^x(\cdot) := P(Z_n(x) \in \cdot)$ converges weakly to μ, see [36] for details.

Equicontinuity can be expressed w.r.t. many different classes of functions. The theory described in [36] demonstrates the suitability of the definition of e-chains above in the case when the state space (X, d) is a locally compact metric space.

There are many alternative useful definitions in the case when (X, d) is a Polish space, see, e.g. Szarek [47]. For our purpose of analysing locally average contractive IFSs, the following definition is convenient.

Definition 6. *We say that the Markov chain $\{Z_n(x)\}$ is (uniformly) equicontinuous if $\{T^n g\}$ is uniformly equicontinuous for any bounded Lipschitz-continuous function g.*

Proposition 2. *An ϵ-local contractive IFS,* (16), *generates an equicontinuous Markov chain.*

The proof relies on the following lemma telling that the probability of ever leaving the 'contractive zone' can be made arbitrarily close to zero provided we start sufficiently deep into the zone.

Lemma 2. *Suppose ϵ-local contractivity* (16). *Then for any positive $\epsilon' \leq \epsilon$ and $r < 1$,*

$$\sup_{x,y;\, d(x,y)\leq r\epsilon'} P(d(Z_n(x), Z_n(y)) > \epsilon', \ \text{for some}\, n) \leq \frac{c}{(1-c)} r. \tag{28}$$

Downloaded by [Uppsala universitetsbibliotek] at 00:29 16 August 2012

Proof. By ϵ-local contractivity, we obtain recursively

$$\begin{aligned} E(d(Z_n(x), Z_n(y)) \,|\, d(Z_k(x), Z_k(y)) < \epsilon, \;\; \text{for all} \;\; k \leq n-1) \\ \leq cE(d(Z_{n-1}(x), Z_{n-1}(y)) \,|\, d(Z_k(x), Z_k(y)) < \epsilon, \;\; \text{for all} \;\; k \leq n-1) \\ \leq cE(d(Z_{n-1}(x), Z_{n-1}(y)) \,|\, d(Z_k(x), Z_k(y)) < \epsilon, \;\; \text{for all} \;\; k \leq n-2) \\ \leq \cdots \leq c^n d(x, y). \end{aligned}$$

By the Chebyshev inequality, we therefore have

$$P(d(Z_n(x), Z_n(y)) > \epsilon' \,|\, d(Z_k(x), Z_k(y)) < \epsilon, \;\; \text{for all} \;\; k \leq n-1) \leq \frac{c^n d(x, y)}{\epsilon'}.$$

Thus, if $d(x, y) \leq r\epsilon'$,

$$\begin{aligned} &P(d(Z_n(x), Z_n(y)) > \epsilon', \;\; \text{for some}\, n) \\ &\leq \sum_{n=1}^{\infty} P(d(Z_n(x), Z_n(y)) > \epsilon' \,|\, d(Z_k(x), Z_k(y)) < \epsilon, \;\; \text{for all} \;\; k \leq n-1) \\ &\leq \sum_{n=1}^{\infty} rc^n = \frac{c}{1-c} r. \end{aligned}$$

□

Proof. (Proposition 2)

Let $g \in BL$ and $d(x, y) \leq r\epsilon'$, where $0 < r < 1$ and $0 < \epsilon' \leq \epsilon$. By Lemma 2, we obtain

$$\begin{aligned} |Eg(Z_n(x)) - Eg(Z_n(y))| &\leq 2\|g\|_\infty P(d(Z_n(x), Z_n(y)) > \epsilon', \;\; \text{for some}\, n) \\ &\quad + \epsilon'\|g\|_L P(d(Z_n(x), Z_n(y)) \leq \epsilon', \;\; \text{for all}\, n) \\ &\leq 2\|g\|_\infty \frac{c}{1-c} r + \epsilon'\|g\|_L. \end{aligned} \tag{29}$$

This completes the proof of Proposition 2 since the right-hand side of (29) can be made arbitrarily small by choosing r and ϵ' sufficiently small. □

For $\delta > 0$, define

$$N_\delta^{x,y} := \inf\{n : d(Z_n(x), Z_n(y)) < \delta\}.$$

Definition 7. *We say that an IFS is centre recurrent if for any $\delta > 0$,*

$$P\left(N_\delta^{x,y} < \infty\right) = 1,$$

for any $x, y \in X$.

We have the following uniqueness theorem for centre recurrent equicontinuous Markov chains.

Theorem 5. *If a centre recurrent (in sense of Definition 7 above) equicontinuous Markov chain on a Polish space (X, d) has a stationary distribution, μ, then this measure is attractive, i.e. ($P(Z_n(x) \in \cdot)$ converges weakly to $\mu(\cdot)$, for any $x \in X$), and thus μ is necessarily uniquely stationary.*

Downloaded by [Uppsala universitetsbibliotek] at 00:29 16 August 2012

Downloaded by [Uppsala universitetsbibliotek] at 00:29 16 August 2012

Since average contractive IFSs generate weak Feller Markov chains, i.e. the transfer operator T maps $C(X)$ into itself, see, e.g. [44], and weak Feller chains on compact metric spaces always possess stationary distributions, see, e.g. [36], Proposition 2 gives the following Corollary.

COROLLARY 2. *An ϵ-local contractive IFS,* (16)*, on a compact metric space that generates a centre recurrent Markov chain* (*in sense of Definition* 7 *above*) *has an attractive, and thus necessarily unique, stationary probability measure.*

Remark 14. See Kaijser [30] for a closely related result.

Proof. (Theorem 5) From the equicontinuity assumption, it follows that for any $g \in BL$ and any $\delta_1 > 0$, there exists a $\delta_2 > 0$, such that

$$|E(g(Z_n(x)) - g(Z_n(y)))| \leq \delta_1, \tag{30}$$

(uniformly in n) whenever $d(x, y) < \delta_2$.

Thus, for any $g \in BL$, using the Markov property, we obtain

$$\begin{aligned}
|E(g(Z_n(x)) - g(Z_n(y)))| &\leq \sum_{k=0}^{n} \left| E\Big(g(Z_n(x)) - g(Z_n(y)) \,|\, N_{\delta_2}^{x,y} = k\Big) \right| P\Big(N_{\delta_2}^{x,y} = k\Big) \\
&\quad + \left| E\Big(g(Z_n(x)) - g(Z_n(y)) \,|\, N_{\delta_2}^{x,y} > n\Big) \right| P\Big(N_{\delta_2}^{x,y}) > n\Big) \\
&\leq \delta_1 + 2\sup_{x \in X} |g(x)| \, P\Big(N_{\delta_2}^{x,y} > n\Big).
\end{aligned}$$

Since δ_1 can be chosen arbitrarily small, and $P(N_\delta^{x,y} > n) \to 0$, as $n \to \infty$, for any $\delta > 0$, it, therefore, follows that

$$\lim_{n \to \infty} |E(g(Z_n(x)) - g(Z_n(y)))| = 0,$$

for any $x, y \in X$ and $g \in BL$, and we may thus apply Proposition 1 (with μ being the, by assumption existing, invariant measure, and ν being Dirac measure concentrated in x) to obtain that $d_w\left(\mu_n^x, \mu\right) \to 0$, as $n \to \infty$. □

5. Generating probability distributions and fractals using IFSs

5.1 IFS attractors

The concept of fractals was coined by Mandelbrot in 1975 for certain types of 'fractured' sets possessing self-similarity properties. The word fractal was derived from the latin fractus meaning 'broken' or 'fractured'.

In [21], Hutchinson developed a theory for fractal sets and probability measures supported on these sets and showed that a big class of fractals can be described in terms of reversed iterations of functions as the attractor of an IFS.

Consider an IFS $\{(X, d); f_s, s \in S\}$, where S is a finite set and f_s are strict contractions, i.e. functions $f_s : X \to X$, satisfying $d(f_s(x), f_s(y)) \leq cd(x, y)$, for some constant $c < 1$.

From the contractivity assumption, it follows that the map

$$\hat{Z}(\mathbf{i}) = \lim_{n \to \infty} f_{i_1} \circ f_{i_2} \circ \cdots \circ f_{i_n}(x_0) \tag{31}$$

exists for *any* $\mathbf{i} = i_1 i_2 \ldots \in S^N$ and the limit is independent of $x_0 \in X$. The set of all limit points

$$F = \{\hat{Z}(\mathbf{i}) : \mathbf{i} \in S^{\mathbb{N}}\} \subseteq X \tag{32}$$

is called the (set-)attractor of the IFS.

Remark 15. In fractal geometry, X is most commonly assumed to be some Euclidean space. Properties such as self-similarity of F are inherited from the properties of $\{f_s\}$. Barnsley [3] called $\mathbf{i}$ an address of the point $\hat{Z}(\mathbf{i})$ on the attractor F. The set F will typically have a 'fractal' appearance in cases when the maps are affine and the address function has simple properties such as being $1 - 1$.

5.2 Probability measures on IFS attractors

A simple way to construct a probability measure on a set F being the attractor of an IFS $\{(X, d); f_s, s \in S\}$ is to use a probability measure P on S. The measure attractor of the IFS with probabilities $\{(X, d); f_s, s \in S, P\}$ is the unique stationary probability measure, μ, of the Markov chain, $\{Z_n(x)\}$, obtained from random (independent) iterations where the functions f_s are chosen at random according to the probability measure P in each step, i.e. the stationary distribution of the Markov chain defined as

$$Z_n(x) = f_{I_n} \circ f_{I_{n-1}} \circ \cdots \circ f_{I_1}(x), \quad Z_0(x) = x,$$

for any $x \in X$, where $\{I_n\}$ is a sequence of independent random variables with $P(I_n \in \cdot) = P(\cdot)$ for any n.

We can regard the function $\hat{Z} : S^{\mathbb{N}} \to X$ defined in (31) as a μ-distributed random variable if we equip $S^{\mathbb{N}}$ with the (Bernoulli) probability measure P_* generated by the probability measure P on S. Thus, μ can alternatively be defined as the push-forward of P_* via $\hat{Z}$, i.e. $\mu(\cdot) = P_*(\mathbf{i} \in S^{\mathbb{N}} : \hat{Z}(\mathbf{i}) \in \cdot)$.

Remark 16. One important question in fractal geometry is the problem of calculating various notions of dimensions of F and μ, see, e.g. Falconer [18].

When studying such local properties, it substantially simplifies if every point in F has a unique or special form of address and the maps are non-singular. It is common to assume that f_s belongs to some particular class of affine maps.

Remark 17. Note that an IFS with a finite number of strict contractions satisfies the conditions of Theorem 1 regardless of what probability measure P we take.

The strict contraction assumption guarantees the existence of the limit $\hat{Z}(\mathbf{i})$ in (31) for any $\mathbf{i}$. For average contractive systems, Theorem 1 only guarantees the existence of this limit for almost all $\mathbf{i}$, and the set attractor may, therefore, not in general be well defined for such systems. The measure attractor μ of an average contractive IFS with probabilities is, however, still uniquely defined as the unique stationary distribution of the associated Markov chain.

Remark 18. A simple way to approximate the unique stationary distribution, μ, is simply to run the Markov chain from an arbitrary fixed initial state, and the proportion of 'time' spent in a set, B, converges to $\mu(B)$ a.s., provided the boundary of B has zero μ-measure, see, e.g. Breiman [11], Elton [16] or Barnsley et al. [7].

Downloaded by [Uppsala universitetsbibliotek] at 00:29 16 August 2012

This theorem forms the theoretical bases of Barnsley's 'chaos game algorithm' saying that we can draw a picture of an IFS attractor by successively plotting a realization of a random trajectory of the associated Markov chain starting at some arbitrary given point on the set attractor.

Example 2. Sierpinski triangle.

Pick three points with coordinates A, B and C in the plane building the vertices of a triangle. Consider a Markov chain starting at $X_0 = A$ where, in each step, if $X_n = x$ we let

$$X_{n+1} = f_A(X_n) = \text{the point obtained by moving } x \text{ half the distance towards } A$$
$$= x + \frac{1}{2}(A - x) = \frac{x}{2} + \frac{A}{2},$$

with probability 1/3,

$$X_{n+1} = f_B(X_n) = \text{the point obtained by moving } x \text{ half the distance towards } B$$
$$= x + \frac{1}{2}(B - x) = \frac{x}{2} + \frac{B}{2},$$

with probability1/3,

$$X_{n+1} = f_C(X_n) = \text{the point obtained by moving } x \text{ half the distance towards } C$$
$$= x + \frac{1}{2}(C - x) = \frac{x}{2} + \frac{C}{2},$$

with probability 1/3.

If $\{I_n\}$ is a sequence of independent random variables uniformly distributed on $\{A, B, C\}$, then $X_{n+1} = f_{I_n}(X_n)$. The only reachable points are points on the *Sierpinski triangle*. A random trajectory will, with probability 1, 'draw' the Sierpinski triangle with vertices in A, B and C (Figure 1).

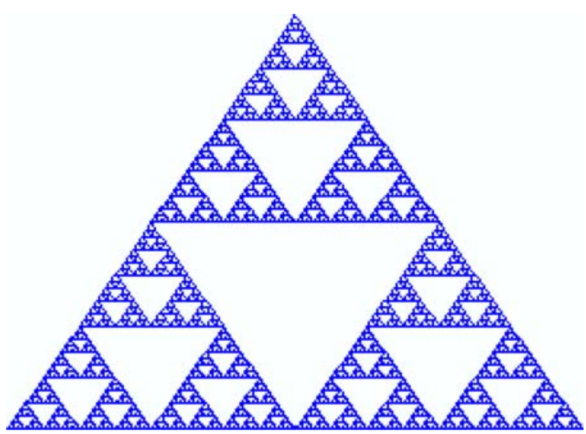


Figure 1. Sierpinski triangle.

Example 3. Cantor set: $X = [0, 1]$

The Markov chain $\{X_n\}$ generated by independent random iterations with the functions $f_1(x) = x/3$ and $f_2(x) = x/3 + 2/3$ chosen with equal probabilities (starting at, e.g.

Downloaded by [Uppsala universitetsbibliotek] at 00:29 16 August 2012

$X_0 = 0$) generates points on the middle-third *Cantor set*. A random trajectory will, with probability 1, 'draw' the Cantor set (Figure 2).

Figure 2. The first five sets approximating the limiting Cantor set. If $\{X_n\}$ is a Markov chain on $X = [0, 1]$ generated by i.i.d. iterations with the maps $f_1(x) = x/3$ and $f_2(x) = x/3 + 2/3$, then the value of X_n can be found in the $n : th$ approximating set of the Cantor set.

Remark 19. Hutchinson [21] studied fractals and probability measures on fractals via reversed iterations of functions.

The idea of regarding probability distributions on fractals as unique stationary distributions of Markov chains generated by random iterations seems to have been first explored by Barnsley and Demko [5].

In the theory of random processes, the historical order was reversed in the sense that 'reversed' random iterations were introduced as a tool to prove distributional theorems for 'forward' random iterations. This tool, which has been 'folklore' in the probabilistic literature the last 50 years, was formulated as a principle in [35].

The above formalism for describing certain types of sets and measures as attractors of an IFS or an IFS with probabilities can be generalized in various ways. The conditions of S being finite and the IFS maps being contractive are clearly far from being necessary.

We will keep these assumptions also in the next sub-section here in order not to blur the basic ideas and instead look at an interesting class of fractals that can be constructed using more than one IFS.

5.3 Code-tree fractals

Let $\mathbf{F} = \{(X, d); f_s^\lambda, s \in S\}_{\lambda \in \Lambda}$ be in indexed family of IFSs, where $S = \{1, 2, \ldots, M\}$, for some $M < \infty$, and Λ is a finite set, and suppose the maps f_s^λ are uniformly contractive. Consider a function $\omega : \cup_{k=0}^{\infty}\{1, \ldots, M\}^k \to \mathbf{F}$. We call ω a code tree corresponding to $\mathbf{F}$. A code tree can be identified with a labelled M-ary tree with nodes labelled with an index of an IFS.

Define

$$\hat{Z}^\omega(\mathbf{i}) = \lim_{k\to\infty} f_{i_1}^{\omega(\emptyset)} \circ f_{i_2}^{\omega(i_1)} \circ \ldots \circ f_{i_k}^{\omega(i_1 \ldots i_{k-1})}(x_0), \quad \text{for } \mathbf{i} = i_1 i_2 \ldots \in S^{\mathbb{N}},$$

and

$$F^\omega = \{\hat{Z}^\omega(\mathbf{i}); \mathbf{i} \in S^{\mathbb{N}}\},$$

for some fixed $x_0 \in X$. (It does not matter which x_0 we choose, since the limit is, as before, independent of x_0.)

We call F^ω the attractor or the code-tree fractal corresponding to the code tree ω and $\mathbf{F}$.

Downloaded by [Uppsala universitetsbibliotek] at 00:29 16 August 2012

5.3.1 V-variable fractals

The sub-code trees of a code tree ω corresponding to a node $i_1 \dots i_k$ is the code tree $\omega_{i_1 \dots i_k}$ defined by $\omega_{i_1 \dots i_k}(j_1 j_2 \dots j_n) := \omega(i_1 \dots i_k j_1 \dots j_n)$, for any $n \geq 0$ and $j_1 \dots j_n \in S^n$.

Let $V \geq 1$ be a positive integer. We call a code-tree V-variable if for any k the set of code trees $\left\{\omega_{i_1 \dots i_k}; i_1 \dots i_k \in S^k\right\}$ contains at most V distinct elements.

A code-tree fractal F^{ω} is said to be V-variable if ω is a V-variable code tree.

A V-variable fractal is intuitively a fractal having at most V distinct 'forms' or 'shapes' at any level of magnification.

All IFS attractors can be regarded as being 1-variable. See, e.g. Barnsley [4] and Barnsley et al. [6], [7] and [8] for more on the theory of V-variable fractals.

Downloaded by [Uppsala universitetsbibliotek] at 00:29 16 August 2012

5.4 Random (code-tree) fractals

If the code tree ω of a code-tree fractal is chosen at random, then the map $\omega \mapsto F^{\omega}$ will be a random fractal. A code tree can be identified with a labelled M-ary tree with nodes labelled with an IFS. We can thus generate random fractals if we choose labels of nodes at random. The case when nodes are chosen in an i.i.d. manner gives 'standard random fractals'. 'Random homogeneous fractals/or random 1-variable' are obtained if we choose labels of nodes at different levels in an i.i.d. manner, but let all nodes within a level be the same. The natural way of choosing random V-variable fractals for $V \geq 2$ requires a more involved (non-Markovian) dependence structure between the random choices of nodes, see, e.g. [8] for details (Figure 3).

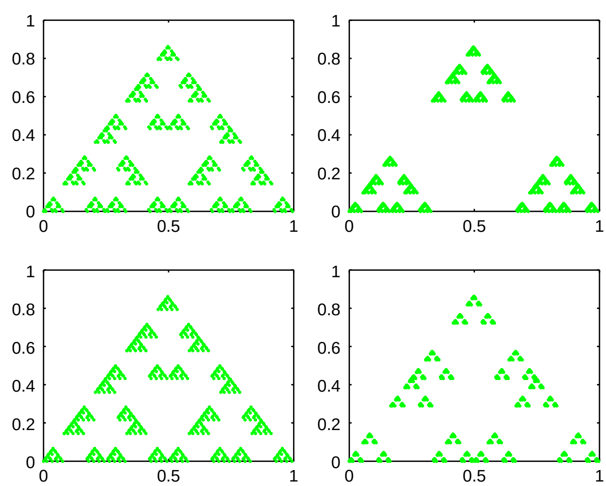


Figure 3. Four examples of realizations of random 1-variable Sierpinski-triangles generated by two IFSs. 1-variable fractals have only one shape at any level of magnification.

Remark 20. Generalizations of the above setting includes average contractive systems, different (possibly infinite) number of maps in the defining IFSs and more general types of 'transformations', see, e.g. [4], [7] and [8].

Downloaded by [Uppsala universitetsbibliotek] at 00:29 16 August 2012

6. The convergence theorem for finite Markov chains and the Propp–Wilson perfect sampling algorithm

In this section, we will present a proof of the Markov chain convergence theorem for finite Markov chains based on iterated random functions. One substantial advantage of this proof in comparison with other proofs not involving iterated random functions, in addition to being simple, is that it gives a method for sampling from the unique stationary distribution 'for free'. This method is known as the Propp–Wilson perfect sampling (or coupling from the past (CFTP)) method.

Consider a Markov chain with state space $X = \{1, 2, \ldots, n_X\}$ and transition matrix $P = (p_{ij})$, $1 \le i, j \le n_X$, where $n_X = |X| < \infty$. Let f be an IFS representation of P, i.e. a measurable function $f : X \times [0, 1] \to X$, such that if U is a random variable, uniformly distributed on the unit interval, then

$$P(f(i, U) = j) = p_{ij},$$

for any $1 \le i, j \le n_X$. If $\{I_n\}_{n=1}^{\infty}$ is a sequence of independent random variables uniformly distributed in the unit interval, and $f_s(x) := f(x, s)$, then it follows by induction that

$$P(f_{I_1} \circ f_{I_2} \circ \cdots \circ f_{I_n}(i) = j) = p_{ij}^{(n)},$$

where $p_{ij}^{(n)}$ denotes the element on row i and column j in the matrix P^n.

If $p_{ij} > 0$ for any i and j, then it is possible to choose f, such that

$$\delta = P(f(i, U) \text{ does not depend on } i) > 0,$$

i.e. a Markov chain with a strictly positive transition matrix can be generated by iterated random maps where constant maps are chosen with positive probability. (One simple representation is $f(i, s) = \inf\left\{k : \sum_{j=1}^{k} p_{ij} \ge s\right\}$, but we typically get a better representation (larger δ) if we base our representation on the fact that P can be expressed as $P = \delta\hat{P} + (1 - \delta)\tilde{P}$ for some $\delta > 0$, and some transition matrices $\hat{P}$ and $\tilde{P}$, where $\hat{P}$ has rank 1, see Athreya and Stenflo [2].)

If constant maps are chosen with positive probability, then it follows that the random maps

$$\hat{Z}_n(i) := f_{I_1} \circ f_{I_2} \circ \cdots \circ f_{I_n}(i),$$

will not depend on i (i.e. will be constant) for all $n \ge T$, where

$$T = \inf\{n : \hat{Z}_n(i) \text{ does not depend on } i\} \tag{33}$$

is the random first time when the mapping $\hat{Z}_n(i)$ does not depend on i. Note, by construction, that T is dominated by a geometrically distributed random variable.

Thus, $\hat{Z} := f_{I_1} \circ f_{I_2} \circ \cdots \circ f_{I_T}(i)$ is a random variable with distribution $\mu = (\mu_i)_{i=1}^{n_X}$ being the unique stationary distribution of the Markov chain described by the common row vector of $\lim_{n\to\infty} P^n$. An upper bound for the convergence rate (in total variation norm) follows from the inequality

$$\sum_{j=1}^{n_X}\left|p_{ij}^{(n)}-\mu_j\right|=\sum_{j=1}^{n_X}\left|P\left(\hat{Z}_n(i)=j\right)-P(\hat{Z}=j)\right|\leq 2P(T>n)\leq 2(1-\delta)^n.$$

Since $\hat{Z}=\hat{Z}_T$ is μ-distributed, we have derived the Propp–Wilson [38] algorithm for simulating from μ: generate realizations $i_1, i_2, \ldots i_T$ of uniformly distributed random variables $I_1, I_2, \ldots, I_T$ until the corresponding realization of $\hat{Z}_T$ is a constant. Then that realization is a realization of a μ-distributed random variable.

The result described above generalizes in a straightforward manner to Markov chains on general state spaces satisfying the so-called 'Doeblin condition', see Athreya and Stenflo [2] for details.

Remark 21. A standard result within the theory of finite Markov chains states that an irreducible and aperiodic Markov chain with transition matrix P has the property that P^{m_0} has only positive elements for some $m_0>0$, see, e.g. Häggström [19]. The above result could, therefore, be applied to prove the convergence theorem for Markov chains under the standard conditions of irreducibility and aperiodicity, c.f. Remark 3.

Remark 22. The algorithm by Propp and Wilson [38] for simulating non-biased (perfect) samples from π is also referred to as the Propp–Wilson coupling from the past (CFTP) algorithm. See Wilson [48] for related literature.

An explanation of the name CFTP requires a notation with $\{I_n\}$ being two-sided, as in Theorem 2. c.f. Remark 1.

It is not important to stop simulating exactly at time T defined as in (33) since $\hat{Z}_n=\hat{Z}_T$, for any $n\geq T$. In practice, it is common to take some fixed increasing sequence of times $t_1<t_2<t_3<\ldots$, and stop simulating at time t_k when the simulated realization of $\hat{Z}_{t_k}$ is constant.

It can be hard to check if a realization of the random function $\hat{Z}_n(x)$ is constant if the state space is large. Substantial computational simplifications are obtained in case the maps f_s are monotone.

7. Time series

The goal of time series analysis is to draw conclusions about the data observed at different times. In a time series model, we interpret the data as observations from a stochastic process. As always in stochastic modelling a useful model should be simple, have a well developed theory and describe the observed data well. For a model of the form (1), this could mean that the function f can be described with few parameters, that there is a well-developed theory for estimating these parameters and that the process $\{X_n\}$ has a well behaved long-run dynamics.

A classical simple class of time series models is the $AR(p)$-processes (autoregressive process of order p) being stationary solutions $\{U_n\}$ to equations of the form

$$U_n-\phi_1U_{n-1}-\ldots-\phi_pU_{n-p}=W_n, \tag{34}$$

where $\phi_1,\ldots,\phi_p$ are real valued constants and $\{W_n\}$ are i.i.d. random variables with mean zero and finite variance.

The $AR(p)$ process, $\{U_n\}$, is said to be causal (or future-independent) if it can be expressed in the form $U_n=\sum_{i=0}^{\infty}\psi_iW_{n-i}$, for some constants ψ_i with $\sum_{i=0}^{\infty}|\psi_i|<\infty$. The

Downloaded by [Uppsala universitetsbibliotek] at 00:29 16 August 2012

condition

$$\phi(z) = 1 - \phi_1 z - \dots \phi_p z^p \neq 0, \quad \text{for all} \ |z| \leq 1, \tag{35}$$

is a well-known necessary and sufficient condition for causality for an $AR(p)$-process, see Brockwell and Davis [12].

If we define the $p \times 1$-vectors $X_n = (U_{n-p+1}, U_{n-p+2}, \dots, U_n)'$, then

$$X_{n+1} = \underbrace{\begin{pmatrix} 0 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 1 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 0 & 1 \\ \phi_p & \phi_{p-1} & \phi_{p-2} & \cdot & \cdot & \phi_1 \end{pmatrix}}_{=:A} X_n + \underbrace{\begin{pmatrix} 0 \\ 0 \\ \cdot \\ 0 \\ W_{n+1} \end{pmatrix}}_{=:I_n}, \tag{36}$$

and we may thus study $AR(p)$-processes via random iterations on $X = \mathbb{R}^p$ with affine maps of the form $f_s(x) = Ax + s$, where the translation vector s is chosen at random.

It can be proved that all eigenvalues of A have absolute values strictly less than 1 if (35) holds, see Brockwell and Davis [12], p.311, exercise 8.3 and therefore (e.g. by using Theorem 2) it follows that the limit

$$X_n = \lim_{m \to -\infty} f_{I_n} \circ \cdots \circ f_{I_{m+1}}(x) = \sum_{k=0}^{\infty} A^k I_{n-(k+1)} = I_{n-1} + AI_{n-2} + A^2 I_{n-3} + \dots,$$

is the unique solution to (36), if (35) holds.

More generally, consider an $ARMA(p, q)$-process, i.e. a stationary process $\{Y_n\}$ satisfying the (autoregressive moving-average) equation

$$Y_n - \phi_1 Y_{n-1} - \dots - \phi_p Y_{n-p} = W_n + \theta_1 W_{n-1} + \dots + \theta_q W_{n-q}, \tag{37}$$

where $\phi_1, \phi_2, \dots, \phi_p$ and $\theta_1, \theta_2, \dots, \theta_q$ are constants, and $\{W_n\}$ is i.i.d. with mean zero and finite variance. Suppose the AR-coefficients satisfy (35), and let $\{U_n\}$ be the causal $AR(p)$-process satisfying (34). Let $r = \max(p, q+1)$. We may regard $\{U_n\}$ as a causal $AR(r)$-process if we let $\phi_j = 0$, for $j > p$. Let $X_n = (U_{n-r+1}, U_{n-r+2}, \dots, U_n)'$. The Markov chain $\{X_n\}$ can (as before) be analysed via iterated random functions, but the $r \times 1$-state vector X_n is now typically 'unobservable'.

It is straightforward to check that $Y_n = U_n + \theta_1 U_{n-1} + \cdots + \theta_q U_{n-q}$ is a solution to (37).

Thus, if we define $\theta_j = 0$, for $j > q$, then the 'observations' Y_n satisfy the 'observation equation'

$$Y_n = BX_n, \tag{38}$$

where $B = [\theta_{r-1}, \theta_{r-2}, \dots, \theta_1, 1]$. The above expresses the $ARMA(p, q)$-process $\{Y_n\}$ in the form of a hidden Markov chain, where the 'observed state' Y_n is a function of the hidden 'true' state X_n, where $\{X_n\}$ is a Markov chain that can be expressed in the form of an average contractive IFS.

The above model with 'state equation' (36) and 'observation equation' (38) is a particular kind of a stable *linear* state space model. A useful tool in the analysis of such

Downloaded by [Uppsala universitetsbibliotek] at 00:29 16 August 2012

models is Kalman recursions, which can be used both for parameter estimation and for prediction in models that can be expressed in this form. See Brockwell and Davis [12] for further reading.

The name 'state space model' originates from control engineering where the main problem is to use the 'observation' sequence $\{Y_n\}$ to control/predict the hidden 'state' X_n.

The convergence results discussed in this paper could be used within the theory of *nonlinear* state space models.

Acknowledgement

I am grateful to John Hutchinson for many helpful discussions and for kind hospitality during my visits to the Australian National University in Canberra.

Note

1. Barnsley and Demko [5] used the word IFS for a finite family of contractive maps. This is the standard use of the term IFS in the theory of fractals since this condition ensures the existence of an attractor of the IFS, see Section 5.1.

References

[1] G. Alsmeyer and C.-D. Fuh, *Limit theorems for iterated random functions by regenerative methods*, Stochastic Process. Appl. 96(1) (2001), pp. 123–142; Corrigendum: Stochastic Process. Appl. 97(2): 341–345, 2002.

[2] K.B. Athreya and Ö. Stenflo, *Perfect sampling for Doeblin chains*, Sankhyā 65(4) (2003), pp. 763–777.

[3] M.F. Barnsley, *Fractals Everywhere*, 2nd ed., Academic Press, Boston, MA, 1993.

[4] M.F. Barnsley, *Superfractals*, Cambridge University Press, Cambridge, 2006.

[5] M.F. Barnsley and S. Demko, *Iterated function systems and the global construction of fractals*, Proc. Roy. Soc. London Ser. A 399(1817) (1985), pp. 243–275.

[6] M.F. Barnsley, J.E. Hutchinson, and Ö. Stenflo, *A fractal valued random iteration algorithm and fractal hierarchy*, Fractals 13 (2005), pp. 111–146.

[7] M.F. Barnsley, J.E. Hutchinson, and Ö. Stenflo, *V-Variable fractals: Fractals with partial self similarity*, Adv. Math. 218 (2008), pp. 2051–2088.

[8] M.F. Barnsley, J.E. Hutchinson, and Ö. Stenflo, *V-variable fractals: Dimension results*, Forum Math, DOI: 10.1515/FORM.2011.075.

[9] R. Bhattacharya and M. Majumdar, *Random Dynamical Systems: Theory and Applications*, Cambridge University Press, Cambridge, 2007.

[10] A.A. Borovkov, *Ergodicity and Stability of Stochastic Processes*, Wiley, Chichester, 1998.

[11] L. Breiman, *The strong law of large numbers for a class of Markov chains*, Ann. Math. Statist. 31 (1960), pp. 801–803.

[12] P.J. Brockwell and R.A. Davis, *Introduction to Time Series and Forecasting*, Springer-Verlag, New York, 2002.

[13] N. Carlsson, *A contractivity condition for iterated function systems*, J. Theoret. Probab. 15(3) (2002), pp. 613–630.

[14] P. Diaconis and D. Freedman, *Iterated random functions*, SIAM Rev. 41(1) (1999), pp. 45–76 (electronic).

[15] R.M. Dudley, *Real Analysis and Probability*, Revised reprint of the 1989 original. Cambridge Studies in Advanced Mathematics, 74, Cambridge University Press, Cambridge, 2002.

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

[17] J.H. Elton, *A multiplicative ergodic theorem for Lipschitz maps*, Stochastic Process. Appl. 34(1) (1990), pp. 39–47.

[18] K. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, 2nd ed., Wiley, Hoboken, NJ, 2003.

Downloaded by [Uppsala universitetsbibliotek] at 00:29 16 August 2012

[19] O. Häggström, *Finite Markov Chains and Algorithmic Applications*, Cambridge University Press, Cambridge, 2002.
[20] U. Herkenrath and M. Iosifescu, *On a contractibility condition for iterated random functions*, Rev. Roumaine Math. Pures Appl. 52(5) (2007), pp. 563–571.
[21] J.E. Hutchinson, *Fractals and self-similarity*, Indiana Univ. Math. J. 30(5) (1981), pp. 713–747.
[22] J.E. Hutchinson and L. Rüschendorf, *Random fractals and probability metrics*, Adv. Appl. Probab. 32(4) (2000), pp. 925–947.
[23] M. Iosifescu, *A simple proof of a basic theorem on iterated random functions*, Proc. Rom. Acad. Ser. A Math. Phys. Tech. Sci. Inf. Sci. 4(3) (2003), pp. 167–174.
[24] M. Iosifescu, *Iterated function systems: A critical survey*, Math. Rep. (Bucur.) 11(61) (2009), pp. 181–229.
[25] M. Iosifescu and S. Grigorescu, *Dependence with Complete Connections and its Applications*, Cambridge University Press, Cambridge, 1990.
[26] B. Jamison, *Asymptotic behavior of successive iterates of continuous functions under a Markov operator*, J. Math. Anal. Appl. 9 (1964), pp. 203–214.
[27] B. Jamison, *Ergodic decompositions induced by certain Markov operators*, Trans. Amer. Math. Soc. 117 (1965), pp. 451–468.
[28] S.F. Jarner and R.L. Tweedie, *Locally contracting iterated functions and stability of Markov chains*, J. Appl. Probab. 38 (2001), pp. 494–507.
[29] P.E.T. Jorgensen, *Analysis and Probability: Wavelets, Signals, Fractals*, Graduate Texts in Mathematics, 234, Springer, New York, 2006.
[30] T. Kaijser, *A limit theorem for Markov chains in compact metric spaces with applications to products of random matrices*, Duke Math. J. 45(2) (1978), pp. 311–349.
[31] T. Kaijser, *On a new contraction condition for random systems with complete connections*, Rev. Roumaine Math. Pures Appl. 26(8) (1981), pp. 1075–1117.
[32] G. Keller, *Equilibrium States in Ergodic Theory*, London Mathematical Society Student Texts, 42, Cambridge University Press, Cambridge, 1998.
[33] Y. Kifer, *Ergodic Theory of Random Transformations*, Birkhäuser Boston Inc., Boston, MA, 1986.
[34] A.N. Lagerås and Ö. Stenflo, *Central limit theorems for contractive Markov chains*, Nonlinearity 18(5) (2005), pp. 1955–1965.
[35] G. Letac, *A contraction principle for certain Markov chains and its applications*, in *Random Matrices and their Applications* (Brunswick, Maine, 1984), Amer. Math. Soc., Providence, RI, 1986, pp. 263–273.
[36] S.P. Meyn and R.L. Tweedie, *Markov Chains and Stochastic Stability*, 2nd ed., With a prologue by Peter W. Glynn, Cambridge University Press, Cambridge, 2009.
[37] Y.B. Pesin, *Dimension Theory in Dynamical Systems: Contemporary Views and Applications*, Chicago Lectures in Mathematics, University of Chicago Press, Chicago, IL, 1997.
[38] J.G. Propp and D.B. Wilson, *Exact sampling with coupled Markov chains and applications to statistical mechanics*, Random Structures Algorithms 9(1–2) (1996), pp. 223–252.
[39] G.O. Roberts and J.S. Rosenthal, *General state space Markov chains and MCMC algorithms*, Probab. Surv. 1 (2004), pp. 20–71.
[40] M. Rosenblatt, *Equicontinuous Markov operators*, Teor. Verojatnost. i Primenen. 9 (1964), pp. 205–222.
[41] A.N. Shiryaev, *Probability*, Springer-Verlag, New York, 1996.
[42] D. Steinsaltz, *Locally contractive iterated function systems*, Ann. Probab. 27(4) (1999), pp. 1952–1979.
[43] D. Steinsaltz, *Random logistic maps and Lyapunov exponents*, Indag. Math. (N.S.) 12(4) (2001), pp. 557–584.
[44] Ö. Stenflo, *Ergodic theorems for iterated function systems controlled by stochastic sequences*, Doctoral thesis No 14, Department of Mathematics, Umeå University, 1998.
[45] Ö. Stenflo, *Markov chains in random environments and random iterated function systems*, Trans. Amer. Math. Soc. 353(9) (2001), pp. 3547–3562.
[46] Ö. Stenflo, *Uniqueness of invariant measures for place-dependent random iterations of functions*, IMA Vol. Math. Appl. 132 (2002), pp. 13–32.
[47] T. Szarek, *Feller processes on non-locally compact spaces*, Ann. Probab. 34(5) (2006), pp. 1849–1863.

Downloaded by [Uppsala universitetsbibliotek] at 00:29 16 August 2012

[48] D.B. Wilson, *Web site for perfectly random sampling with Markov chains*, Available at http://www.dbwilson.com/exact/
[49] W.B. Wu and X. Shao, *Limit theorems for iterated random functions*, J. Appl. Probab. 41(2) (2004), pp. 425–436.
[50] W.B. Wu and M. Woodroofe, *A central limit theorem for iterated random functions*, J. Appl. Probab. 37(3) (2000), pp. 748–755.

Downloaded by [Uppsala universitetsbibliotek] at 00:29 16 August 2012