Limit Theorems for Iterated Random Functions

WEI BIAO WU AND XIAOFENG SHAO [1]

November 19, 2003

Department of Statistics, The University of Chicago, Chicago, IL 60637, USA

Abstract

We study geometric-moment contracting properties of nonlinear time series that are expressed in terms of iterated random functions. Under Dini-continuity condition, a central limit theorem for additive functionals of such systems is established. The empirical processes of sample paths are shown to converge to Gaussian processes in the Skorokhod space. An exponential inequality is established. We present a bound for joint cumulants, which ensures the applicability of several asymptotic results in spectral analysis of time series. Our results provide a vehicle for statistical inferences for fractals and many nonlinear time series models.

Keywords: Stationarity; Iterated random function; Central limit theorem; Dini Continuity; Exponential inequality; Martingale; Markov chains; Fractal; Nonlinear time series; Cumulants

1 Introduction

Let $(\mathcal{X}, \rho)$ be a complete, separable metric space with Borel sets $\mathbb{X}$. An iterated random function system on the state space $\mathcal{X}$ is defined as

$$X_n = F_{\theta_n}(X_{n-1}),\ n \in \mathbb{N}, \tag{1}$$

where $\theta, \theta_n, n \in \mathbb{N}$, take values in a second measurable space Θ, and are independent with identical marginal distribution H. Here, $F_\theta(\cdot) = F(\cdot, \theta)$ is the θ-section of a jointly measurable function $F : \mathcal{X} \times \Theta \mapsto \mathcal{X}$ and X_0 is independent of $(\theta_n)_{n\geq 1}$. The simple iteration (1) unifies many interesting branches in probability theory, such as Markov chains, nonlinear time series, queuing etc. The problem of the existence of stationary distributions

[1] *E-mail address*: wbwu@galton.uchicago.edu, shao@galton.uchicago.edu

and related convergence issues has received considerable attention recently; see for example, Barnsley and Elton (1988), Elton (1990), Arnold (1998), Stenflo (1998), Diaconis and Freedman (1999), Steinsaltz (1999), Alsmeyer and Fuh (2001), Jarner and Tweedie (2001) among others. Various sufficient conditions are presented in those works to ensure the weak convergence $X_n \Rightarrow \pi$, where π is the stationary distribution.

In this paper, we shall establish the convergence of X_n to π in the sense of *geometric-moment contraction* (to be defined below), and obtain limit theorems for additive functionals and empirical processes for X_n. Unlike strong mixing conditions, the geometric moment contraction seems easily verifiable and sufficiently mild, and it provides a natural base from which the limit theorems related to X_n can be systematically derived.

To define geometric moment contraction, let $X_0' \sim \pi$ be independent of $X_0 \sim \pi$ and $(\theta_k)_{k\geq 1}$ and define $X_n(x) = F_{\theta_n} \circ F_{\theta_{n-1}} \circ \ldots \circ F_{\theta_1}(x)$. Thus $X_n(X_0')$ can be viewed as a coupled version of $X_n(X_0)$. We say that X_n is *geometric-moment contracting* if there exist $\alpha > 0$, $C = C(\alpha) > 0$ and $0 < r = r(\alpha) < 1$ such that for all $n \in \mathbb{N}$,

$$E\{\rho^\alpha[X_n(X_0'),\ X_n(X_0)]\} \leq Cr^n. \tag{2}$$

The inequality (2) implies that, starting from two independent initial points X_0 and X_0', the orbits $X_n(X_0')$ and $X_n(X_0)$ will be close to each other at an exponential rate. Steinsaltz (1999) considered rate of convergence with $\alpha = 1$.

The rest of the paper is organized as follows. Geometric moment contraction is discussed in Section 2. In Section 3 we present a central limit theorem for $S_{n,l}(g) = \sum_{i=1}^n g(Y_i)$, where the functional g is stochastic Dini-continuous and $Y_i = (X_{i-l+1}, X_{i-l+2}, \ldots, X_i)$ (see Remark 2 for the definition of X_k with negative subscripts k). The convergence of empirical processes towards Gaussian processes is also studied. A bound on joint cumulants is obtained in Section 3.3.

2 Geometric-moment contraction

We start by imposing regularity conditions on the underlying evolution mechanism $F_\theta(\cdot)$. Our main result regarding stationarity is Theorem 2 which asserts the existence of the stationary distribution together with a geometric convergence rate in the sense of (2).

Condition 1. *There exist $y_0 \in \mathcal{X}$ and $\alpha > 0$ such that*

$$I(\alpha, y_0) := E\{\rho^\alpha[y_0, F_\theta(y_0)]\} = \int_\Theta \rho^\alpha[y_0, F_\theta(y_0)]H\{d\theta\} < \infty. \tag{3}$$

Condition 2. *There exist $x_0 \in \mathcal{X}, \alpha > 0, r(\alpha) \in (0,1)$ and $C(\alpha) > 0$ such that*

$$E\{\rho^\alpha[X_n(x), X_n(x_0)]\} \le C(\alpha)r^n(\alpha)\rho^\alpha(x, x_0) \tag{4}$$

holds for all $x \in \mathcal{X}, n \in \mathbb{N}$.

Condition 1 provides a bound on the intercept of the random transform F; condition 2 is of Lyapunov type ensuring that the forward iteration X_n is contracting on average. Unless otherwise specified, we will assume hereafter that $0 < \alpha \le 1$ in Conditions 1 and 2 since if (3) and (4) are satisfied for some $\alpha > 1$, then they are valid for all $\alpha \le 1$ by Hölder's inequality. Actually, for any $\beta \in (0, \alpha)$, let $C(\beta) = C(\alpha)^{\beta/\alpha}$ and $r(\beta) = r(\alpha)^{\beta/\alpha} \in (0,1)$. Then

$$\begin{aligned} E\{\rho^\beta[X_n(x), X_n(x_0)]\} &\le (E\{\rho^\alpha[X_n(x), X_n(x_0)]\})^{\frac{\beta}{\alpha}} \\ &\le C(\alpha)^{\frac{\beta}{\alpha}}[r(\alpha)^{\frac{\beta}{\alpha}}]^n\rho^\beta(x, x_0). \end{aligned}$$

Introduce the backward iteration process $Z_n(x) = F_{\theta_1} \circ F_{\theta_2} \circ \ldots \circ F_{\theta_n}(x)$. Notice that for all $x \in \mathcal{X}$, $Z_n(x) \overset{\mathcal{D}}{=} X_n(x)$. If $Z_n(x)$ converges a.s. to a proper random variable, then $X_n(x)$ converges in distribution. Clearly, $X_n(x) = F_{\theta_n} \circ X_{n-1}(x)$ and $Z_n(x) = Z_{n-1} \circ F_{\theta_n}(x)$. A typical result for the existence of stationarity of (1) is given in Diaconis and Freedman (cf Theorem 1). A random variable Y is said to have an *algebraic tail* if there exist $A, B > 0$ such that $P(|Y| > y) < A/y^B$ for all $y > 0$. Equivalently, $E(|Y|^\alpha) < \infty$ for some $\alpha > 0$.

Theorem 1. (Diaconis and Freedman, 1999) *Assume (3),*

$$E(\log K_\theta) = \int_\Theta \log K_\theta H\{d\theta\} < 0, \quad \textit{where } K_\theta = \sup_{x' \ne x} \frac{\rho[F_\theta(x'), F_\theta(x)]}{\rho(x', x)}, \tag{5}$$

and that K_θ has an algebraic tail. Then there exists a unique stationary distribution π for (1) and $Z_n(x) \to Z_\infty \sim \pi$ at a geometric rate. The limit Z_∞ does not depend on x.

Theorem 2. *Suppose that Conditions 1 and 2 hold. Then there exists a random variable Z_∞ such that for all $x \in \mathcal{X}$, $Z_n(x) \to Z_\infty$ almost surely. The limit Z_∞ is $\sigma(\theta_1, \theta_2, \ldots)$-measurable and does not depend on x. Moreover, for every $n \in \mathbb{N}$,*

$$E\{\rho^\alpha[Z_n(x), Z_\infty]\} \le Cr^n(\alpha) \tag{6}$$

where $C > 0$ depends solely on x, x_0, y_0 and α, and $0 < r(\alpha) < 1$. In addition, (2) holds.

Remark 1. Condition 2 is slightly weaker than (5). A simple but useful observation pointed out in Wu and Woodroofe (2000, WW hereafter) is that if K_θ has an algebraic tail, then (5) implies that $E(K_\theta^\alpha) < 1$ for sufficiently small $\alpha > 0$. Hence (4) holds with $C(\alpha) = 1$ and $r(\alpha) = E(K_\theta^\alpha)$ by Fatou's lemma:

$$1 > E(K_\theta^\alpha) = \int_\Theta \sup_{x' \neq x} \frac{\rho^\alpha[F_\theta(x'), F_\theta(x)]}{\rho^\alpha(x', x)} H\{d\theta\} \geq \sup_{x' \neq x} \int_\Theta \frac{\rho^\alpha[F_\theta(x'), F_\theta(x)]}{\rho^\alpha(x', x)} H\{d\theta\}. \tag{7}$$

Actually, (7) implies $E\{\rho^\alpha[X_1(x'), X_1(x)]\} \leq r(\alpha)\rho^\alpha(x', x)$ and consequently (4) by a simple induction.

The proof of Theorem 2 seems simpler than the one in Diaconis and Freedman (1999). On the other hand, the geometric-moment contraction (2) asserted by Theorem 2 plays a key role for central limit theorems and concentration inequalities (cf Section 3.)

Proof of Theorem 2. Let $0 < \alpha \leq 1$ satisfy both Conditions 1 and 2. By (4) and the triangle inequality, $I(\alpha, x_0) \leq \rho^\alpha(x_0, y_0) + I(\alpha, y_0) + E\{\rho^\alpha[F_\theta(x_0), F_\theta(y_0)]\} < \infty$. By (4),

$$\begin{aligned} E\{\rho^\alpha[Z_{n+1}(x_0), Z_n(x_0)]\} &= E\left[E\{\rho^\alpha[Z_n \circ F_{\theta_{n+1}}(x_0), Z_n(x_0)]|\theta_{n+1}\}\right] \\ &\leq C(\alpha)r^n(\alpha)E\{\rho^\alpha[F_{\theta_{n+1}}(x_0), x_0]\} = C(\alpha)r^n(\alpha)I(\alpha, x_0) =: \delta_n. \end{aligned}$$

Then $P(\rho[Z_{n+1}(x_0), Z_n(x_0)] \geq \delta_n^{\frac{1}{2\alpha}}) \leq \delta_n^{\frac{1}{2}}$, which by the Borel–Cantelli lemma yields $P(\rho[Z_{n+1}(x_0), Z_n(x_0)] \geq \delta_n^{\frac{1}{2\alpha}}$ infinitely often$) = 0$. Since $\delta_n^{\frac{1}{2\alpha}}$ is summable, $Z_n(x_0) \to Z_\infty$ a.s. due to the completeness of $\mathcal{X}$. Clearly Z_∞ is $\sigma(\theta_1, \theta_2, \ldots)$-measurable. Again by the triangle inequality,

$$\begin{aligned} E\{\rho^\alpha[Z_n(x_0), Z_\infty]\} &\leq E\left\{\sum_{j=0}^\infty \rho[Z_{n+1+j}(x_0), Z_{n+j}(x_0)]\right\}^\alpha \\ &\leq \sum_{j=0}^\infty E\{\rho^\alpha[Z_{n+1+j}(x_0), Z_{n+j}(x_0)]\} \leq \frac{\delta_n}{1 - r(\alpha)}. \end{aligned}$$

Let $C = C(\alpha)[I(\alpha, x_0)/(1 - r(\alpha)) + \rho^\alpha(x, x_0)]$. Then (6) follows from (4) and

$$\begin{aligned} E\{\rho^\alpha[Z_n(x), Z_\infty]\} &\leq E\{\rho^\alpha[Z_n(x_0), Z_\infty]\} + E\{\rho^\alpha[Z_n(x), Z_n(x_0)]\} \\ &\leq \frac{\delta_n}{1 - r(\alpha)} + C(\alpha)r^n(\alpha)\rho^\alpha(x, x_0) = Cr^n(\alpha). \end{aligned}$$

So $Z_n(x) \to Z_\infty$ almost surely. Hence for any x, the limit $V_n = \lim_{m\to\infty} F_{\theta_{1+n}} \circ F_{\theta_{2+n}} \circ \ldots \circ F_{\theta_{n+m}}(x)$ exists almost surely. Observe that V_n is identically distributed as $Z_\infty = Z_n(V_n) \sim$

π and V_n is independent of $(\theta_i)_{1\le i\le n}$. Hence we have

$$\begin{aligned} E\{\rho^\alpha[X_n(X_0'),X_n(X_0)]\} &\le E\{\rho^\alpha[X_n(X_0'),X_n(x_0)]\}+E\{\rho^\alpha[X_n(x_0),X_n(X_0)]\} \\ &= 2E\{\rho^\alpha[Z_n(V_n),Z_n(x_0)]\}=2E\{\rho^\alpha[Z_\infty,Z_n(x_0)]\}\le \frac{2\delta_n}{1-r(\alpha)}, \end{aligned}$$

which entails (2). $\diamondsuit$

Remark 2. Theorem 2 suggests a simple way to define X_i with negative subscripts $i\le 0$ such that the relation $X_i=F_{\theta_i}(X_{i-1})$ holds for $i\le 0$ as well. Let $(\theta_i)_{i\in\mathbb{Z}}$ be iid random variables. Then for all $x\in\mathcal{X}$, the limit

$$\lim_{m\to\infty} F_{\theta_i}\circ F_{\theta_{i-1}}\circ\ldots\circ F_{\theta_{i-m}}(x)$$

exists almost surely and does not depend on x. Denote the limit by $X_i=M(\ldots,\theta_{i-1},\theta_i)$, where M is a measurable function. Then $X_i=F_{\theta_i}(X_{i-1})$ holds for all $i\in\mathbb{Z}$.

The following Lemma 1 shows an interesting equivalence between geometric-moment contraction inequalities.

Lemma 1. *Assume $E[\rho^p(X_0,x)]<\infty$ for some $p>0$ and $x\in\mathcal{X}$. If (2) holds for an $\alpha\in(0,p)$, then (2) holds for all $\alpha\in(0,p)$.*

Proof of Lemma 1. It suffice to show that (2) holds for $a\in(\alpha,p)$. Let $q=1/(1-a/p)$, $\delta_n=r^{n/(2\alpha)}$ and $T_n=\rho[X_n(X_0'),X_n(X_0)]$. Then

$$\begin{aligned} E(T_n^a) &= E(T_n^a\times \mathbf{1}_{T_n<\delta_n}+T_n^a\times\mathbf{1}_{T_n\ge\delta_n}) \\ &\le \delta_n^a+2^{1+a}E(\{\rho^a[X_n(X_0'),x]+\rho^a[x,X_n(X_0)]\}\times\mathbf{1}_{T_n\ge\delta_n}) \\ &= \delta_n^a+2^{2+a}E\{\rho^a[X_n(X_0'),x]\times\mathbf{1}_{T_n\ge\delta_n}\} \end{aligned}$$

By Hölder's and Markov's inequalities,

$$\begin{aligned} E\{\rho^a[X_n(X_0'),x]\times\mathbf{1}_{T_n\ge\delta_n}\} &\le \{E\rho^p[X_n(X_0'),x]\}^{a/p}\times\{E(\mathbf{1}_{T_n\ge\delta_n})\}^{1/q} \\ &\le \{E\rho^p(X_0,x)\}^{a/p}\times\{\delta_n^{-\alpha}E(T_n^\alpha)\}^{1/q} \\ &= \mathcal{O}[(\delta_n^{-\alpha}r^n)^{1/q}]=\mathcal{O}[r^{n/(2q)}]. \end{aligned}$$

Therefore (2) holds with $r(a)=\max[r^{a/(2\alpha)},r^{1/(2q)}]$. $\diamondsuit$

3 Central limit problems

Many nonlinear time series adopt the form $X_n = F(X_{n-1}, \theta_n; \xi)$, where the parameter $\xi \in \Xi \subset \mathbb{R}^d$. For example, the threshold AR(1) (TAR) model is given by $X_n = \xi_1 X_{n-1}^+ + \xi_2 X_{n-1}^- + \theta_n$ (see Tong, 1990). The AR with conditional heteroscedasticity (ARCH, Engle, 1982) model has recursion $X_n = \theta_n \sqrt{\xi_1^2 + \xi^2 X_{n-1}^2}$. Random coefficient model assumes $X_n = (\xi_1 + \xi_2 \theta_{n,1}) X_{n-1} + \xi_3 \theta_{n,2}$ (Nicholls and Quinn, 1982).

The estimation for unknown parameter ξ often involves additive functionals $S_{n,l}(g) = \sum_{i=1}^n g(X_{i-l+1}, X_{i-l+2}, \ldots, X_i)$. For example, the least square estimators of ξ_1 and ξ_2 in the TAR model are given by $\hat{\xi}_{1n} = \sum_{i=1}^n X_i X_{i-1}^+ / \sum_{i=1}^n (X_{i-1}^+)^2$ and $\hat{\xi}_{2n} = \sum_{i=1}^n X_i X_{i-1}^- / \sum_{i=1}^n (X_{i-1}^-)^2$ respectively. Let θ_n have mean 0 and variance 1 in an ARCH model $X_n = \theta_n \sqrt{\xi_1^2 + \xi_2^2 X_{n-1}^2}$. Then $EX_n^2 = \xi_1^2 + \xi_2^2 EX_{n-1}^2$ and $E(X_n^2 X_{n-1}^2) = \xi_1^2 EX_{n-1}^2 + \xi_2^2 EX_{n-1}^4$. These identities yield estimators for ξ_1^2 and ξ_2^2 from the estimated moments $\widehat{E}X_n^2 = \sum_{i=1}^n X_i^2/n$, $\widehat{E}X_n^4 = \sum_{i=1}^n X_i^4/n$ and $\widehat{E}(X_{n-1}^2 X_n^2) = \sum_{i=1}^n X_{i-1}^2 X_i^2/n$. The limiting behavior of $S_{n,l}(g)$ is needed for statistical inference based on estimation equations.

Theorem 3 aims at establishing central limit theorems for $S_{n,l}(g)$ under mild conditions, and thus provides an inferential base for nonlinear time series. Some special models have been discussed earlier; see for example, Petruccelli and Woolford (1984), Nicholls and Quinn (1982). See WW (2000) and Herkenrath et al (2003) for some recent work. Let the l-dimensional vector $Y_i = (X_{i-l+1}, X_{i-l+2}, \ldots, X_i)$. For a random variable Z let $\|Z\|_r = [\mathbb{E}(|Z|^r)]^{1/r}$ and $\|Z\| = \|Z\|_2$. If $l > 1$, then g is said to be *non-instantaneous*. For $\delta > 0$ define

$$\Delta_g(\delta) = \sup\left\{ \|[g(Y) - g(Y_1)] \times \mathbf{1}_{[\rho(Y, Y_1) \le \delta]}\| : Y \text{ and } Y_1 \text{ are identically distributed} \right\}, \quad (8)$$

where $\rho(\cdot, \cdot)$ is the product metric: $\rho(z, z') = \sqrt{\sum_{i=1}^l \rho^2(z_i, z_i')}$ for $z = (z_1, \ldots, z_l), z' = (z_1', \ldots, z_l') \in \mathcal{X}^l$.

Theorem 3. *Assume (2), $X_1 \sim \pi$, $E[g(Y_1)] = 0$, $E[|g(Y_1)|^p] < \infty$ for some $p > 2$, and*

$$\int_0^1 \frac{\Delta_g(t)}{t} dt < \infty. \quad (9)$$

Then there exists $\sigma_g \ge 0$ such that for almost all x (π), $\{S_{\lfloor nu \rfloor, l}(g)/\sqrt{n}, 0 \le u \le 1\}$ given $X_0 = x$ converges to $\sigma_g \mathbb{B}$, where $\mathbb{B}$ is a standard Brownian motion and $\lfloor v \rfloor = \max\{k \in \mathbb{Z} : k \le v\}$.

Proof. We adopt the argument in Gordin and Lifsic (1978). Suppose the probability space is rich enough to carry iid random variables θ_k, $k \in \mathbb{Z}$. Let $\Theta_n = (\ldots, \theta_{n-1}, \theta_n)$, $n \in \mathbb{Z}$ be the shift process. Clearly Θ_n is Markovian. Let X_0', an independent copy of X_0, be independent of θ_k, $k \in \mathbb{Z}$; let $X_n' = F_{\theta_n} \circ \ldots F_{\theta_1}(X_0')$ and $Y_n' = (X_{n-l+1}', \ldots, X_n')$. By (2), $E[\rho^\alpha(Y_n, Y_n')] \leq Cr^n$ for some $C > 0$, $0 < r < 1$. Set $\phi = r^{\frac{1}{2\alpha}}$. For $n > l$, since $E[g(Y_n')|X_0] = 0$, we have by Cauchy's inequality,

$$\begin{aligned}
\|E[g(Y_n)|\Theta_0]\| &\leq \|g(Y_n) - g(Y_n')\| \\
&\leq \|[g(Y_n) - g(Y_n')]\mathbf{1}_{\rho(Y_n, Y_n') \leq \phi^n}\| + \|[g(Y_n) - g(Y_n')]\mathbf{1}_{\rho(Y_n, Y_n') > \phi^n}\| \\
&\leq \Delta(\phi^n) + \{\|[g(Y_n) - g(Y_n')]^2\|_{q'} \times \|\mathbf{1}_{\rho(Y_n, Y_n') > \phi^n}\|_q\}^{1/2} \\
&\leq \Delta(\phi^n) + \mathcal{O}\{E[\rho^\alpha(Y_n, Y_n')]/\phi^{n\alpha}\}^{\frac{1}{2q}} \leq \Delta(\phi^n) + \mathcal{O}(\phi^{\frac{\alpha n}{2q}}),
\end{aligned}$$

where we have applied Hölder's inequality with $q' = p/2 > 1$ and $q = q'/(q' - 1)$ and Markov's inequality $P(|Z| > z) \leq E(|Z|^\alpha)/z^\alpha$ with $z = \phi^n$. Since

$$\sum_{n=1}^{\infty} \Delta(\phi^n) \leq \frac{1}{1 - \phi} \int_0^1 \frac{\Delta_g(t)}{t} dt$$

and $\sum_{n=1}^{\infty} \phi^{\frac{\alpha n}{2q}} < \infty$, $h(\Theta_0) = \sum_{k=0}^{\infty} E[g(Y_k)|\Theta_0]$ converges in L^2 in view of (9). Observe that $g(Y_0) = h(\Theta_0) - E[h(\Theta_1)|\Theta_0]$, we have

$$\sum_{k=1}^{n} g(Y_k) = \sum_{k=1}^{n} D_k + R_n, \tag{10}$$

where $D_k = h(\Theta_k) - E[h(\Theta_k)|\Theta_{k-1}]$ and $R_n = E[h(\Theta_1)|\Theta_0] - E[h(\Theta_{n+1})|\Theta_n] = \mathcal{O}_p(1)$. Thus $S_n(g)/\sqrt{n} \Rightarrow N(0, \|D_1\|^2)$ by applying the Martingale central limit theorem to the stationary and ergodic martingale differences D_k, $k \in \mathbb{Z}$. The Martingale central limit theorem also asserts that for almost all x (π), the partial sum process $\{S_{\lfloor nu \rfloor, l}(g)/\sqrt{n}, 0 \leq u \leq 1\}$ given $X_0 = x$ converges to Brownian motion (cf Corollary 2 in WW). $\diamondsuit$

A function f is Dini-continuous if $\int_0^1 \Delta_f(x)/x dx < \infty$, where $\Delta_f(x) = \sup\{|f(y) - f(y')| \times \mathbf{1}_{|y-y'| \leq x}\}$. Thus it is natural to say that g is *stochastic Dini-continuous* with respect to the distribution of Y_1 if (9) holds. Clearly if g is Dini-continuous, then it is necessarily stochastic Dini-continuous. However the reverse is not true (cf. Corollary 1 by noticing that the indicator function $f_\lambda(x) = \mathbf{1}_{x \leq \lambda}$ is not Dini-continuous).

Theorem 3 goes beyond the earlier work by WW in several aspects. In the latter paper a central limit theorem is derived for instantaneous filters g, namely $l = 1$. The non-instantaneous transformation in Theorem 3 facilitates statistical inference for non-linear time series. Even though the vector process Y_i can be viewed as a new iterated function system defined by $Y_n = G(Y_{n-1}, \theta_n)$, where $G(y, \theta) = (y^{(2)}, \ldots, y^{(l)}, F(y^{(l)}, \theta))$ for $y = (y^{(1)}, \ldots, y^{(l)})$, the result in WW is not directly applicable here. To see this, let $L(\theta)$ be the Lipschitz constant for $F(\cdot, \theta)$. Then under the Euclidean distance, G has a non-contracting Lipschitz constant $\max[1, L(\theta)]$.

Conditions on g in WW are also stronger than the stochastic Dini-continuity. Let $l = 1$, π be the uniform(0,1) distribution and $g(x) = x^{-1/3} - EX_1^{-1/3} = x^{-1/3} - 3/2$. Then it is easily verified that $K(g, \psi; x)$ in WW is ∞ for all $x \in (0, 1)$. Hence the conditions on g in the former paper are violated. However (9) is satisfied since $\Delta_g(t) = \mathcal{O}(t^{1/32})$ as $t \downarrow 0$. To see this, let X, Y be uniform(0,1) distributed random variables. Then

$$\begin{aligned}\|[g(X) - g(Y)]\mathbf{1}_{[|X-Y|\le\delta]}\| &\le \|[g(X) - g(Y)]\mathbf{1}_{[|X-Y|\le\delta]} \times \mathbf{1}_{[|X-Y|\le Y^2\sqrt{\delta}]}\| \\ &+ \|[g(X) - g(Y)]\mathbf{1}_{[|X-Y|\le\delta]} \times \mathbf{1}_{[|X-Y|> Y^2\sqrt{\delta}]}\| := A + B.\end{aligned}$$

For the term B observe that necessarily $Y^2\sqrt{\delta} \le \delta$, and hence by Hölder's inequality $B^2 \le \|[g(X) - g(Y)]^2\|_{4/3} \times \|\mathbf{1}_{[Y^2\sqrt{\delta}\le\delta]}\|_4 = \mathcal{O}(\delta^{1/16})$. On the other hand, by the mean-value theorem, there exists $\xi \in (-1, 1)$ such that under $|X - Y| \le Y^2\sqrt{\delta}$, $|g(X) - g(Y)| \le Y^2\sqrt{\delta}|g'(Y + \xi Y^2\sqrt{\delta})|$. Thus $A^2 = \mathcal{O}(\delta)$.

3.1 Empirical processes

Let $\mathcal{X} = \mathbb{R}$ and $\rho(\cdot, \cdot)$ be the Euclidean distance; let $G(x) = P[X_1 \le x]$ and $G_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{X_i\le x}$ be the distribution and empirical distribution functions of X_1 and $P_n(x) = \sqrt{n}[G_n(x) - G(x)]$. Empirical processes play a paramount role in statistics. Corollary 1 asserts the asymptotic normality for $P_n(x)$ for a fixed x and Theorem 4 states a functional central limit theorem.

Corollary 1. *Let $\tau(x; t) = \min[G(x + t) - G(x),\ G(x) - G(x - t)]$. Assume (2) and*

$$\int_0^1 \frac{\sqrt{\tau(x; t)}}{t} dt < \infty. \tag{11}$$

Then there exists a $\sigma(x) < \infty$ *such that*

$$P_n(x) \Rightarrow N[0, \sigma^2(x)]. \tag{12}$$

Proof. Let $g_y(u) = \mathbf{1}_{u \le y} - G(y)$ and $X, X_1 \sim \pi$. Then

$$\begin{aligned} \|[g_y(X_1) - g_y(X)]\mathbf{1}_{|X-X_1|\le\delta}\|^2 &= P(X \le y, X_1 > y, |X - X_1| \le \delta) \\ &+ P(X_1 \le y, X > y, |X - X_1| \le \delta) \le 2P(y < X \le y + \delta). \end{aligned}$$

So $\Delta_g^2(\delta) \le 2[G(y+\delta) - G(y)]$. Similarly, $\Delta_g^2(\delta) \le 2[G(y) - G(y-\delta)]$. So (12) follows from Theorem 3 in view of (11). $\diamondsuit$

Theorem 4. *Assume (2) and there exists* $\kappa > 5/2$ *and* $C > 0$ *such that for all* $0 < \delta \le 1/2$,

$$\sup_{x\in\mathbb{R}} |G(x+\delta) - G(x)| \le C \log^{-\kappa}(\delta^{-1}). \tag{13}$$

Then $\{P_n(y),\ y \in \mathbb{R}\}$ *converges in* $\mathcal{D}(\mathbb{R})$ *to a Gaussian process* W *with mean zero and covariance function*

$$E[W(x)W(y)] = \sum_{k\in\mathbb{Z}} \mathrm{Cov}[\mathbf{1}_{X_0\le x},\ \mathbf{1}_{X_k\le y}].$$

Proof. Corollary 1 implies the finite dimensional covergence in view of (11) and (13). By Proposition 2 in Doukhan and Louhichi (1999), for the tightness it suffices to verify that

$$d_n := \sup |\mathrm{Cov}\,[g(X_{-n_1})g(X_0),\ g(X_n)g(X_{n+n_2})]| = \mathcal{O}(n^{-(\kappa+5/2)/2}), \tag{14}$$

where the sup is taken over all $n_1, n_2 \ge 0$ and $g \in \mathcal{G} = \{x \mapsto \mathbf{1}_{s<x\le t} : s, t \in \mathbb{R}\}$. To this end, we shall apply the idea of coupling by letting $X'_k = X_k(X'_0)$, $k \in \mathbb{N}$. Then

$$\begin{aligned} &|\mathrm{Cov}\,[g(X_{-n_1})g(X_0),\ g(X_n)g(X_{n+n_2})]\,| \\ &= |E\{g(X_{-n_1})g(X_0)[g(X_n)g(X_{n+n_2}) - g(X'_n)g(X'_{n+n_2})]\}| \\ &\le |E\{\xi_1[g(X_n) - g(X'_n)]\}| + |E\{\xi_2[g(X_{n+n_2}) - g(X'_{n+n_2})]\}| \\ &\le \|\xi_1\|_p\|g(X_n) - g(X'_n)\|_q + \|\xi_2\|_p\|g(X_{n+n_2}) - g(X'_{n+n_2})\|_q \end{aligned}$$

by Hölder's inequality, where $\xi_1 = g(X_{-n_1})g(X_0)g(X_{n+n_2})$, $\xi_2 = g(X_{-n_1})g(X_0)g(X'_n)$, $q = (3\kappa + 5/2)/(2\kappa + 5) > 1$ and $p = q/(q-1)$. Let $\beta = r^{1/(2\alpha)}$. Then by (2),

$$\begin{aligned} E|\mathbf{1}_{X_n\le s} - \mathbf{1}_{X'_n\le s}| &\le P(|X_n - X'_n| \ge \beta^n) + 2P(X_n \le s, X'_n > s, |X_n - X'_n| < \beta^n) \\ &\le Cr^n/\beta^{n\alpha} + 2C\log^{-\kappa}(\beta^{-n}) = \mathcal{O}(n^{-\kappa}), \end{aligned}$$

which implies that $d_n = \mathcal{O}(n^{-\kappa/q})$ since $|\xi_1|$, $|\xi_2| \le 1$. Thus (14) follows. $\diamondsuit$

Example 1. Consider the $AR(1)$ model $X_n = aX_{n-1} + (1-a)\theta_n$, where θ_n are iid Bernoulli random variables with success probability $1/2$. Then X_n is a Markov chain which is neither strong mixing nor irreducible (hence it cannot be Harris ergodic although it has stationary distribution).

In the case $a = 1/2$ it is a Bernoulli shift model which takes uniform(0,1) as invariant distribution. Clearly (13) is satisfied for any $0 < x < 1$ since $\pi(x) = x$, and hence $P_n(x) \Rightarrow W(x)$.

Solomyak (1995) showed that for almost all $a \in (1/2, 1)$ (Lebesgue), the invariant measure π is absolutely continuous. Therefore for those a, (13) trivially holds for all $x \in (0, 1)$.

Now consider the case $0 < a < 1/2$. Then the invariant distribution π has a compact, fractal-set support (Hutchinson, 1981), and π is singular with respect to the Lebesgue measure. If $a = 1/3$, support(π) is the well-known Cantor set. Consider the 2^k points $x_1 < x_2 < \ldots < x_{2^k}$ in the set $\{\sum_{i=1}^k a^{i-1}(1-a)z_i : \ z_i = 0 \text{ or } 1\}$. It is easily seen that $x_j - x_{j-1} \geq a^{k-1}(1-a)$ for $j = 2, \ldots, 2^k$ and $P(x_j \leq X_0 \leq x_j + a^k) = 2^{-k}$. Let $t_k = a^{k-1}(1-a) - a^k$. Notice that support(π) is a subset of $\cup_{j=1}^{2^k}[x_j, x_j + a^k]$. For any x, the interval $(x, x + t_k]$ intersects at most one of the 2^k intervals. For $\delta \in (0, 1-2a)$ let $k = k(\delta)$ satisfy $t_k > \delta \geq t_{k+1}$. Then

$$\sup_x |G(x+\delta) - G(x)| \leq \sup_x |G(x+t_k) - G(x)| = \sup_x P(x < X_0 \leq x + t_k) \leq 2^{-k}.$$

Then $\lim_{\delta \to 0} k^{-1}(\log \delta) = \log a$ and (13) holds in view of $2^{-k} = \mathcal{O}[\delta^{-(\log 2)/\log a}]$.

3.2 An Exponential Inequality

Recall $Y_i = (X_{i-l+1}, X_{i-l+2}, \ldots, X_i)$ and $S_n(g) = \sum_{i=1}^n g(Y_i)$. Exponential inequalities play important roles in stochastic processes; see Chapter 1 in Bosq (1996) for an extensive treatment, where applications to nonparametric inference are discussed. However, rigid strong mixing conditions are needed in the latter book, which may fail for many interesting applications. Here we provide an exponential inequality without strong mixing conditions. It is unclear whether similar inequality exists without the restriction (15).

Proposition 1. *Let g be a bounded function, $E[g(Y_n)] = 0$ and*

$$C := \sup_{x\in\mathcal{X}} \sum_{n=0}^{\infty} |E[g(Y_n)|X_0 = x]| < \infty. \tag{15}$$

Then there exists $c_1, c_2 > 0$ which only depend on $\{Y_n\}$ and g such that for all $\lambda > 0$,

$$P(|S_n(g)| \geq n\lambda) \leq c_1 e^{-n\lambda^2 c_2}. \tag{16}$$

Proof. Under (15), we have the decomposition (10) and $h(\Theta_0) = \sum_{k=0}^{\infty} E[g(Y_k)|\Theta_0]$ exists and is bounded. Thus R_n and D_n are also bounded. Let $|R_n| \leq r$ and $|D_n| \leq d$; let $I(y) = e^y - 1 - y$. Applying the exponential inequality for bounded martingale differences (for example Freedman (1975)), we have

$$E\{\exp[\beta(\sum_{i=1}^{n} D_i + R_n)]\} \leq e^{r+nI(\beta d)}$$

for $\beta \geq 0$ and similarly $E\{\exp[-\beta S_n(g)]\} \leq e^{r+nI(-\beta d)}$. So (16) easily follows. $\diamondsuit$

Example 1 (continued). Let $X_n = (X_{n-1} + \theta_n)/2$, where θ_n are iid Bernoulli random variables with success probability 1/2 and g has bounded variation on $[0,1]$. Then (15) is satisfied. To see this, assume $|g| \leq 1$ and let $L = \sup\{\sum_{i=0}^{I} |g(t_i) - g(t_{i-1})|,\ 0 \leq t_0 < \ldots < t_I \leq 1\} < \infty$ be the total variation of g over $[0,1]$. For $x \in (0,1)$, since $E[g(X_n)] = \int_0^1 g(u)du = 0$, (15) follows from

$$|E[g(X_n)|X_0 = x]| = 2^{-n} \left| \sum_{i=0}^{2^n-1} g(\frac{x+i}{2^n}) \right| \leq \int_0^{\frac{1}{2^n}} \sum_{i=0}^{2^n-1} \left| g(\frac{i}{2^n} + u) - g(\frac{x+i}{2^n}) \right| du \leq \frac{L}{2^n}.$$

3.3 Joint Cumulants

Let $(U_1, \ldots, U_k)$ be a random vector. Then the joint cumulant is defined as

$$\mathrm{Cum}(U_1, \ldots, U_k) = \sum (-1)^p (p-1)! E\left[\prod_{j\in V_1} U_j\right] \ldots E\left[\prod_{j\in V_p} U_j\right], \tag{17}$$

where $V_1, \ldots, V_p$ is a partition of the set $\{1, 2, \ldots, k\}$ and the sum is taken over all such partitions. For example, $\mathrm{Cum}(U_1, U_2) = E(U_1U_2) - E(U_1)E(U_2) = \mathrm{Cov}(U_1, U_2)$ and $\mathrm{Cum}(U_1, U_1, U_1) = E[U_1 - E(U_1)]^3$. It is easily seen in view of Hölder's inequality that, if $E(|U_i|^k) < \infty$ for all $i = 1, \ldots, k$, then $\mathrm{Cum}(U_1, \ldots, U_k)$ is well-defined. Cumulants are closely related to joint characteristic functions; see Rosenblatt [1984, 1985 (p. 138)] for more details. Many important asymptotic results in the spectral analysis of time series require certain summability conditions on joint cumulants. For example, Rosenblatt (1985, p 138) established a central limit theorem for the spectral density estimator of the strongly mixing stationary process $(X_k)_{k\in\mathbb{Z}}$ under the condition

$$\sum_{s_1,\ldots,s_7\in\mathbb{Z}} |\mathrm{Cum}(X_0, X_{s_1}, \ldots, X_{s_7})| < \infty. \tag{18}$$

Conditions of similar nature can be found in Brillinger (1981). To ensure the applicability of such results, it is critical to have a bound for $|\mathrm{Cum}(X_0, X_{s_1}, \ldots, X_{s_k})|$. In this section we show that the geometric-moment contraction (2) does imply an exponential decay rate of joint cumulants, which consequently guarantees such summability conditions (cf Proposition 2 and Remark 3).

We formulate our result in a framework slightly more general than (1). Recall the shift process $\Theta_n = (\ldots, \theta_{n-1}, \theta_n)$. Let M be a measurable function such that $X_n = M(\Theta_n)$ is a well-defined random variable (cf Remark 2). Then $(X_n)_{n\in\mathbb{Z}}$ is a stationary and ergodic process. Let $(\theta_n^*)_{n\in\mathbb{Z}}$ be an iid copy of $(\theta_n)_{n\in\mathbb{Z}}$, $\Theta_n^* = (\ldots, \theta_{n-1}^*, \theta_n^*)$ and, for $m \geq 0$, $X'_m = M(\Theta_0^*, \theta_1, \ldots, \theta_m)$. Namely X'_m is a coupled version of X_m with the past Θ_0 replaced by the iid copy Θ_0^*.

Proposition 2. *Assume that there exist $C_1 > 0$, $0 < r_1 < 1$ and integer $k \geq 2$ such that $E(|X_0|^k) < \infty$ and $E(|X_n - X'_n|^k) \leq C_1 r_1^n$ for all $n \geq 0$. Then for all $0 \leq m_1 \leq \ldots \leq m_{k-1}$,*

$$|\mathrm{Cum}(X_0, X_{m_1}, \ldots, X_{m_{k-1}})| \leq C r_1^{m_{k-1}/[k(k-1)]}, \tag{19}$$

where the constant $C > 0$ is independent of $m_1, \ldots, m_{k-1}$.

Proof of Proposition 2. Let $C > 0$ be a generic constant which is independent of m_1, $\ldots, m_{k-1}$. In the proof C may vary from line to line and it only depends on C_1, r_1 and the moments $E(|X_0|^i)$, $1 \leq i \leq k$. Let $J = \mathrm{Cum}(X_0, X_{m_1}, \ldots, X_{m_{k-1}})$, $m_0 = 0$ and $n_l = m_l - m_{l-1}$, $1 \leq l \leq k-1$; let the random vector $Y_0 = Y_{0,l} = (X_{m_0-m_{l-1}}, \ldots, X_{m_{l-2}-m_{l-1}}, X_0)$. By the stationarity and the additive property of cumulants,

$$J = \mathrm{Cum}(Y_0, X_{m_l-m_{l-1}}, X_{m_{l+1}-m_{l-1}}, \ldots, X_{m_{k-1}-m_{l-1}})$$

$$
\begin{aligned}
&= \mathrm{Cum}(Y_0, X_{m_l-m_{l-1}} - X'_{m_l-m_{l-1}}, X_{m_{l+1}-m_{l-1}}, \ldots, X_{m_{k-1}-m_{l-1}}) \\
&\quad + \sum_{j=1}^{k-l-1} \mathrm{Cum}(Y_0, X'_{m_l-m_{l-1}}, \cdots, X'_{m_{l+j-1}-m_{l-1}}, \\
&\qquad\qquad X_{m_{l+j}-m_{l-1}} - X'_{m_{l+j}-m_{l-1}}, X_{m_{l+j+1}-m_{l-1}}, \ldots, X_{m_{k-1}-m_{l-1}}) \\
&\quad + \mathrm{Cum}(Y_0, X'_{m_l-m_{l-1}}, \ldots, X'_{m_{k-1}-m_{l-1}}) =: A_0 + \sum_{j=1}^{k-l-1} A_j + B.
\end{aligned}
$$

Since Y_0 and the random vector $(X'_{m_l-m_{l-1}}, \cdots, X'_{m_{k-1}-m_{l-1}})$ are independent, we have $B = 0$ [cf Property (ii), Rosenblatt (1985, p 35)]. We shall now use the definition (17) and show that $|A_0| \le C r_1^{n_l/k}$. To this end, let $U_j = X_{m_j - m_{l-1}}$ for $0 \le j \le k-1$, $j \ne l$ and $U_l = X_{n_l} - X'_{n_l}$. Let $|V|$ be the cardinality of the set V. For any subset $V \subset \{0, 1, \ldots, k-1\}$ such that $l \notin V$, by Hölder's and Jensen's inequalities, we have $|E(\prod_{j\in V} U_j)| \le E(|X_0|^{|V|})$ and

$$
\begin{aligned}
\left| E\left[U_l \prod_{j\in V} U_j \right] \right| &\le \|U_l\|_{1+|V|} \left[E \prod_{j\in V} |U_j|^{(|V|+1)/|V|} \right]^{|V|/(1+|V|)} \\
&\le \|U_l\|_k (E|X_0|^{|V|+1})^{|V|/(1+|V|)} \le (C_1 r_1^{n_l})^{1/k} C'
\end{aligned}
$$

by letting $C' = \sum_{i=0}^{k-1} (E|X_0|^{i+1})^{i/(1+i)}$. By (17), $|A_0| \le C r_1^{n_l/k}$ for some constant C. Similarly, for $1 \le j \le k-l-1$, $|A_j| \le C r_1^{(m_{l+j}-m_{l-1})/k} \le C r_1^{n_l/k}$. Hence $|J| \le C r_1^{n_l/k}$, which implies (19) in view of $|J| \le C \min_{1\le l\le k-1} r_1^{n_l/k}$ and $m_{k-1} = \sum_{l=1}^{k-1} n_l \le (k-1) \max_{1\le l\le k-1} n_l$. $\diamondsuit$

Proposition 2 requires the geometric-moment contraction (2) with $\alpha = k$. If $E(|X_0|^p) < \infty$ for some $p > k$, then by Lemma 1, it suffices to assume (2) with some $\alpha > 0$.

Remark 3. The inequality (19) implies (18) in view of

$$
\begin{aligned}
\sum_{s_1,\ldots,s_7\in\mathbb{Z}} |\mathrm{Cum}(X_0, X_{s_1}, \ldots, X_{s_7})| &\le 2\sum_{s=0}^{\infty} \sum_{(s_1,\ldots,s_7)\in L(s)} |\mathrm{Cum}(X_0, X_{s_1}, \ldots, X_{s_7})| \\
&= \sum_{s=0}^{\infty} \mathcal{O}(s^6 r^s) < \infty,
\end{aligned}
$$

where $r = r_1^{1/[8(8-1)]} = r_1^{1/56}$ and $L(s) = \{(s_1, \ldots, s_7) \in \mathbb{Z}^7 : \ \max_{1\le i\le 7} |s_i| = s\}$.

Acknowledgement

The authors gratefully acknowledge the helpful and constructive comments provided by the anonymous referee.

References

Alsmeyer, G., Fuh, C. D., 2001. Limit theorems for iterated random functions by regenerative methods, Stoch. Proc. Appl. 96, 123–142.

Anderson, T. W. 1971. The statistical analysis of time series. Wiley, New York.

Arnold, L., 1998. Random dynamical systems. Springer, Berlin.

Barnsley, M. F., Elton, J. H., 1988. A new class of Markov processes for image encoding. Adv. Appl. Probab. 20. 14–32

Bosq, D. (1996). *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction. Lecture Notes in Statist.* **110**. Springer, New York.

Brillinger, D. R. 1981. Time series: data analysis and theory 2nd edition, San Francisco: Holden-Day.

Chan, K. S., Tong, H., 2001. Chaos: a statistical perspective. Springer, New York.

Diaconis, P., Freedman. D., 1999. Iterated random functions. SIAM Rev. 41, 41–76.

Paul Doukhan and Sana Louhichi (1999). A new weak dependence condition and applications to moment inequalities, *Stoc. Proc. Appl.* **84**, 313-342.

Elton, J. H., 1990. A multiplicative ergodic theorem for Lipschitz maps. Stoc. Proc. Appl. 34, 39–47.

Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica. 50, 987–1007.

Freedman, David A. (1975). On tail probabilities for martingales *Ann. Probab.* **3** 100–118

Gordin, M.I. and Lifsic, B. (1978). The central limit theorem for stationary Markov processes. *Doklady*, **19** 392–394.

Herkenrath, U., Iosifescu, M. and Rudolph, A. (2003) A note on invariance principles for iterated random functions. J. Appl. Probab. 40, 834-837

Hutchinson, J., 1981. Fractals and self similarity. Indiana University Math. J., 30, 713–747.

Jarner, S., Tweedie R., 2001. Locally contracting iterated random functions and stability of Markov chains. J. Appl. Probab. 38, 494–507.

Nicholls, D. F., Quinn, B. G., 1982. Random coefficient autoregressive models: an introduction. Springer, New York.

Petruccelli, J. D., Woolford, S. W., 1984. A threshold AR(1) model. J. Appl. Probab. 21, 270–286

Rosenblatt, M, 1984. Asymptotic normality, strong mixing, and spectral density estimates. Ann. Probab. 12. 1167-1180.

Rosenblatt,M, 1985. Stationary sequences and random fields. Boston

Solomyak, B. M., 1995. On the random series $\sum \pm\lambda^n$ (an Erdös problem). Ann. of Math. 142, 611–625.

Steinsaltz, D., 1999. Locally contractive iterated function systems. Ann. Probab. 27, 1952–1979.

Stenflo, Ö., 1998. Ergodic theorems for iterated function systems controlled by stochastic sequences. Doctoral thesis No 14, Umeå University, Sweden.

Tong, H., 1990. Non-linear time series: a dynamical system approach. Oxford University Press.

Wu, W. B., Woodroofe, M., 2000. A central limit theorem for iterated random functions. J. Appl. Probab. 37, 748–755.