J. Appl. Prob. **41,** 425–436 (2004)
Printed in Israel

LIMIT THEOREMS FOR ITERATED RANDOM FUNCTIONS

WEI BIAO WU [*] [**] AND
XIAOFENG SHAO,[*] [***] *University of Chicago*

Abstract

We study geometric moment contracting properties of nonlinear time series that are expressed in terms of iterated random functions. Under a 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 chain; fractal; nonlinear time series; cumulants

2000 Mathematics Subject Classification: Primary 60F05
Secondary 60F17; 60J10; 60G42

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

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

where θ and θ_n for $n \in \mathbb{N}$ take values in a second measurable space, Θ, and are independently distributed with identical marginal distribution H. Here, $F_\theta(\cdot) = F(\cdot, \theta)$ is the θ-section of a jointly measurable function $F : \mathcal{X} \times \Theta \to \mathcal{X}$ 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, queueing etc. The problem of the existence of stationary distributions 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, geometric moment contraction

Received 31 March 2003; revision received 1 December 2003.
[*] Postal address: Department of Statistics, University of Chicago, Chicago, IL 60637, USA.
[**] Email address: wbwu@galton.uchicago.edu
[***] Email address: shao@galton.uchicago.edu

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 $X_n(x) = F_{\theta_n} \circ F_{\theta_{n-1}} \circ \cdots \circ F_{\theta_1}(x)$. 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 an $\alpha > 0$, a $C = C(\alpha) > 0$ and an $r = r(\alpha) \in (0, 1)$ such that, for all $n \in \mathbb{N}$,

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

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 the 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 stochastically Dini continuous and $Y_i = (X_{i-l+1}, X_{i-l+2}, \ldots, X_i)$ (see Remark 2 below for the definition of X_k when $k < 0$). 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 a $y_0 \in \mathcal{X}$ and an $\alpha > 0$ such that*

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

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

$$\mathrm{E}\{\rho(X_n(x), X_n(x_0))^\alpha\} \leq C(\alpha) r(\alpha)^n \rho(x, x_0)^\alpha \tag{4}$$

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

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 $0 < \alpha \leq 1$ in Conditions 1 and 2 since, if (3) and (4) are satisfied for some $\alpha > 1$, then they are valid for all $\alpha \leq 1$ 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

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

Introduce the backward iteration process $Z_n(x) = F_{\theta_1} \circ F_{\theta_2} \circ \cdots \circ F_{\theta_n}(x)$. Notice that, for all $x \in \mathcal{X}$, $Z_n(x) \overset{\mathrm{D}}{=} X_n(x)$. If $Z_n(x)$ converges almost surely 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) was given by Diaconis and Freedman (1999) (see Theorem 1). A random variable Y is said to have an *algebraic tail* if there exist $A, B > 0$ such that $\mathrm{P}(|Y| > y) < A/y^B$ for all $y > 0$. Equivalently, $\mathrm{E}\{|Y|^\alpha\} < \infty$ for some $\alpha > 0$.

Theorem 1. (Diaconis and Freedman (1999).) *Assume that Condition 1 holds, that*

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

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 $n \in \mathbb{N}$,*

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

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 by Wu and Woodroofe (2000) is that, if K_θ has an algebraic tail, then (5) implies that $\mathrm{E}\{K_\theta^\alpha\} < 1$ for sufficiently small $\alpha > 0$. Hence, (4) holds with $C(\alpha) = 1$ and $r(\alpha) = \mathrm{E}\{K_\theta^\alpha\}$ by Fatou's lemma:

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

Actually, (7) implies that $\mathrm{E}\{\rho(X_1(x'), X_1(x))^\alpha\} \le r(\alpha)\rho(x', x)^\alpha$ and, consequently, that (4) holds by a simple induction.

Our proof of Theorem 2 seems simpler than that of 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 (see Section 3)

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

$$\begin{aligned}\mathrm{E}\{\rho(Z_{n+1}(x_0), Z_n(x_0))^\alpha\} &= \mathrm{E}\{\mathrm{E}\{\rho(Z_n \circ F_{\theta_{n+1}}(x_0), Z_n(x_0))^\alpha \mid \theta_{n+1}\}\} \\ &\le C(\alpha)r(\alpha)^n\, \mathrm{E}\{\rho(F_{\theta_{n+1}}(x_0), x_0)^\alpha\} = C(\alpha)r(\alpha)^n I(\alpha, x_0) =: \delta_n.\end{aligned}$$

Then $\mathrm{P}(\rho(Z_{n+1}(x_0), Z_n(x_0)) \ge \delta_n^{1/2\alpha}) \le \delta_n^{1/2}$, which by the Borel–Cantelli lemma yields that $\mathrm{P}(\rho(Z_{n+1}(x_0), Z_n(x_0)) \ge \delta_n^{1/2\alpha}$ infinitely often$) = 0$. Since $\delta_n^{1/2\alpha}$ is summable, $Z_n(x_0) \to Z_\infty$ almost surely due to the completeness of $\mathcal{X}$. Clearly, Z_∞ is $\sigma(\theta_1, \theta_2, \ldots)$-measurable. Again by the triangle inequality,

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

Let

$$C = C(\alpha)\left[\frac{I(\alpha, x_0)}{1 - r(\alpha)} + \rho(x, x_0)^\alpha\right].$$

Then (6) follows from (4) and

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

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

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

which entails (2).

Remark 2. Theorem 2 suggests a simple way to define X_i when $i \le 0$ in such a way that the relation $X_i = F_{\theta_i}(X_{i-1})$ still holds. Let $(\theta_i)_{i\in\mathbb{Z}}$ be independent and identically distributed (i.i.d.) random variables. Then, for all $x \in \mathcal{X}$, the limit

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

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 $X_i = F_{\theta_i}(X_{i-1})$ for all $i \in \mathbb{Z}$.

The following lemma shows an interesting equivalence between geometric moment contraction inequalities.

Lemma 1. *Assume that* $\mathrm{E}\{\rho(X_0, x)^p\} < \infty$ *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. 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

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

By the Hölder and Markov inequalities,

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

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

3. Central limit problems

Many nonlinear time series adopt the form $X_n = F(X_{n-1}, \theta_n; \xi)$, where the parameter ξ is in a set $\Xi \subset \mathbb{R}^d$. For example, the threshold AR(1) model (TAR) is given by

$$X_n = \xi_1 X_{n-1}^+ + \xi_2 X_{n-1}^- + \theta_n$$

(see Tong (1990)). The autoregressive with conditional heteroscedasticity model (ARCH; see Engle (1982)) has recursion

$$X_n = \theta_n \sqrt{\xi_1^2 + \xi^2 X_{n-1}^2}.$$

The random coefficient model assumes that

$$X_n = (\xi_1 + \xi_2 \theta_{n,1}) X_{n-1} + \xi_3 \theta_{n,2}$$

(see Nicholls and Quinn (1982)).

The estimation of the 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

$$\hat{\xi}_{1n} = \frac{\sum_{i=1}^n X_i X_{i-1}^+}{\sum_{i=1}^n (X_{i-1}^+)^2} \quad \text{and} \quad \hat{\xi}_{2n} = \frac{\sum_{i=1}^n X_i X_{i-1}^-}{\sum_{i=1}^n (X_{i-1}^-)^2}$$

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 $\mathrm{E}\{X_n^2\} = \xi_1^2 + \xi_2^2\,\mathrm{E}\{X_{n-1}^2\}$ and $\mathrm{E}\{X_n^2 X_{n-1}^2\} = \xi_1^2\,\mathrm{E}\{X_{n-1}^2\} + \xi_2^2\,\mathrm{E}\{X_{n-1}^4\}$. These identities yield estimators for ξ_1^2 and ξ_2^2 from the estimated moments

$$\hat{\mathrm{E}}\{X_n^2\} = \sum_{i=1}^n \frac{X_i^2}{n}, \qquad \hat{\mathrm{E}}\{X_n^4\} = \sum_{i=1}^n \frac{X_i^4}{n} \quad \text{and} \quad \hat{\mathrm{E}}\{X_{n-1}^2 X_n^2\} = \sum_{i=1}^n \frac{X_{i-1}^2 X_i^2}{n}.$$

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) and Nicholls and Quinn (1982). See Wu and Woodroofe (2000) and Herkenrath *et al.* (2003) for some recent work. Define the l-dimensional vector $Y_i = (X_{i-l+1}, X_{i-l+2}, \ldots, X_i)$. For a random variable Z, let $\|Z\|_r = \mathrm{E}\{|Z|^r\}^{1/r}$ and $\|Z\| = \|Z\|_2$. If $l > 1$, then g is said to be *noninstantaneous*. For $\delta > 0$, define

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

where $\rho(\cdot,\cdot)$ is the product metric:

$$\rho(z, z') = \sqrt{\sum_{i=1}^l \rho(z_i, z_i')^2} \quad \text{for } z = (z_1, \ldots, z_l),\, z' = (z_1', \ldots, z_l') \in \mathcal{X}^l.$$

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

$$\int_0^1 \frac{\Delta_g(t)}{t}\,\mathrm{d}t < \infty. \tag{9}$$

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

Proof. We adopt the argument of Gordin and Lifšic (1978). Suppose that the probability space is rich enough to carry i.i.d. random variables $\theta_k,\ k \in \mathbb{Z}$. Let $\Theta_n = (\ldots, \theta_{n-1}, \theta_n)$ for $n \in \mathbb{Z}$ be the shift process. Clearly Θ_n is Markovian. Let X'_0, an independent copy of X_0, be independent of θ_k for $k \in \mathbb{Z}$; let $X'_n = F_{\theta_n} \circ \cdots \circ F_{\theta_1}(X'_0)$ and $Y'_n = (X'_{n-l+1}, \ldots, X'_n)$. By (2), $\mathrm{E}\{\rho(Y_n, Y'_n)^\alpha\} \leq Cr^n$ for some $C > 0$ and $r \in (0, 1)$. Set $\phi = r^{1/2\alpha}$. For $n > l$, since $\mathrm{E}\{g(Y'_n) \mid X_0\} = 0$, we have by Cauchy's inequality that

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

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

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

and $\sum_{n=1}^{\infty} \phi^{\alpha n/2q} < \infty$, $h(\Theta_0) = \sum_{k=0}^{\infty} \mathrm{E}\{g(Y_k) \mid \Theta_0\}$ converges in L^2 in view of (9). Observing that $g(Y_0) = h(\Theta_0) - \mathrm{E}\{h(\Theta_1) \mid \Theta_0\}$, it follows that

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

where $D_k = h(\Theta_k) - \mathrm{E}\{h(\Theta_k) \mid \Theta_{k-1}\}$ and $R_n = \mathrm{E}\{h(\Theta_1) \mid \Theta_0\} - \mathrm{E}\{h(\Theta_{n+1}) \mid \Theta_n\} = \mathcal{O}_p[1]$. Thus $S_n(g)/\sqrt{n} \Rightarrow \mathrm{N}(0, \|D_1\|^2)$ 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 π-almost all x, the partial sum process $\{S_{\lfloor nu \rfloor, l}(g)/\sqrt{n},\ 0 \leq u \leq 1\}$ given that $X_0 = x$ converges to Brownian motion (see Corollary 2 in Wu and Woodroofe (2000)).

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

Theorem 3 goes beyond the earlier work of Wu and Woodroofe (2000) in several aspects. In that paper, a central limit theorem is derived for instantaneous filters g, namely when $l = 1$.

The noninstantaneous transformation in Theorem 3 facilitates statistical inference for nonlinear 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 of Wu and Woodroofe (2000) 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 noncontracting Lipschitz constant $\max[1, L(\theta)]$.

Conditions on g in Wu and Woodroofe (2000) are also stronger than the stochastic Dini continuity. Let $l = 1$, let π be the uniform(0, 1) distribution and let $g(x) = x^{-1/3} - \mathrm{E}\{X_1^{-1/3}\} = x^{-1/3} - \frac{3}{2}$. Then it is easily verified that the term $K(g, \psi; x)$ defined in Wu and Woodroofe (2000) 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

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

For the term B observe that, necessarily, $Y^2\sqrt{\delta} \le \delta$ and, hence, by Hölder's inequality, $B^2 \le \|[g(X) - g(Y)]^2\|_{4/3} \|\,\mathbf{1}_{(Y^2\sqrt{\delta} \le \delta)}\,\|_4 = \mathcal{O}[\delta^{1/16}]$. On the other hand, by the mean-value theorem, there exists a $\xi \in (-1, 1)$ 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 let $\rho(\cdot, \cdot)$ be the Euclidean distance; let $G(x) = \mathrm{P}[X_1 \le x]$ and $G_n(x) = (1/n) \sum_{i=1}^{n} \mathbf{1}_{(X_i \le x)}$ be, respectively, 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 that (2) holds and that*

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

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

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

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

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

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).

Theorem 4. *Assume that (2) holds and there exist a $\kappa > \frac{5}{2}$ and a $C > 0$ such that for all $\delta \in (0, \frac{1}{2}]$,*

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

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*

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

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

$$d_n := \sup_{n_1,n_2\ge 0} |\operatorname{cov}[g(X_{-n_1})g(X_0), g(X_n)g(X_{n+n_2})]| = \mathcal{O}[n^{-(\kappa+5/2)/2}], \tag{14}$$

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

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

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

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

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

Example 1. Consider the AR(1) model $X_n = aX_{n-1} + (1-a)\theta_n$, where the θ_n are i.i.d. Bernoulli random variables with success probability $\frac{1}{2}$. 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 $a = \frac{1}{2}$, it is a Bernoulli shift model which takes uniform(0, 1) as invariant distribution. Clearly (13) is satisfied for any $x \in (0,1)$ since $\pi(x) = x$, and hence $P_n(x) \Rightarrow W(x)$.

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

Now consider the case when $a \in (0, \frac{1}{2})$. Then the invariant distribution π has a compact, fractal-set support (Hutchinson (1981)), and π is singular with respect to the Lebesgue measure. If $a = \frac{1}{3}$, the support of π is the well-known Cantor set. Consider the 2^k points $x_1 < x_2 < \cdots < x_{2^k}$ in the set

$$\left\{\sum_{i=1}^{k} a^{i-1}(1-a)z_i : z_i \text{ is 0 or 1}\right\}.$$

It is easily seen that $x_j - x_{j-1} \ge a^{k-1}(1-a)$ for $j = 2, \ldots, 2^k$ and $\mathrm{P}(x_j \le X_0 \le x_j + a^k) = 2^{-k}$. Let $t_k = a^{k-1}(1-a) - a^k$. Notice that the support of π is a subset of $\bigcup_{j=1}^{2^k}[x_j, x_j + a^k]$.

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

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

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

3.2. An exponential inequality

Recall that $Y_i = (X_{i-l+1}, X_{i-l+2}, \ldots, X_i)$ and $S_n(g) = \sum_{i=1}^n g(Y_i)$. Exponential inequalities play important roles in stochastic processes; see Chapter 1 of 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 a similar inequality exists without the restriction (15).

Proposition 1. *Let g be a bounded function such that* $\mathrm{E}\{g(Y_n)\} = 0$ *and let*

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

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

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

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

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

for $\beta \geq 0$ and, similarly,

$$\mathrm{E}\{\exp[-\beta S_n(g)]\} \leq \mathrm{e}^{r + nI(-\beta d)}.$$

So (16) easily follows.

Example 1. (*Continued.*) Let $X_n = (X_{n-1} + \theta_n)/2$, where the θ_n are i.i.d. Bernoulli random variables with success probability $\frac{1}{2}$ and g has bounded variation on $[0, 1]$. Then (15) is satisfied. To see this, assume that $|g| \leq 1$ and let

$$L = \sup\left\{\sum_{i=0}^{I} |g(t_i) - g(t_{i-1})|,\ 0 \leq t_0 < \cdots < t_I \leq 1\right\} < \infty$$

be the total variation of g over [0, 1]. For $x \in (0, 1)$, since $\mathrm{E}\{g(X_n)\} = \int_0^1 g(u)\,\mathrm{d}u = 0$, (15) follows from the inequality

$$
\begin{aligned}
|\,\mathrm{E}\{g(X_n) \mid X_0 = x\}| &= 2^{-n}\left|\sum_{i=0}^{2^n-1} g\left(\frac{x+i}{2^n}\right)\right| \\
&\le \int_0^{1/2^n} \sum_{i=0}^{2^n-1}\left|g\left(\frac{i}{2^n}+u\right) - g\left(\frac{x+i}{2^n}\right)\right| \mathrm{d}u \\
&\le \frac{L}{2^n}.
\end{aligned}
$$

3.3. Joint cumulants

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

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

where $V_1, \dots, V_p$ is a partition of the set $\{1, 2, \dots, k\}$ and the sum is taken over all such partitions. For example,

$$
\mathrm{cum}(U_1, U_2) = \mathrm{E}\{U_1 U_2\} - \mathrm{E}\{U_1\}\,\mathrm{E}\{U_2\} = \mathrm{cov}(U_1, U_2)
$$

and

$$
\mathrm{cum}(U_1, U_1, U_1) = \mathrm{E}\{U_1 - \mathrm{E}\{U_1\}\}^3.
$$

It is easily seen in view of Hölder's inequality that, if $\mathrm{E}\{|U_i|^k\} < \infty$ for all $i = 1, \dots, k$, then $\mathrm{cum}(U_1, \dots, U_k)$ 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

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

Conditions of a similar nature can be found in Brillinger (1981). To ensure the applicability of such results, it is critical to have a bound for $|\mathrm{cum}(X_0, X_{s_1}, \dots, X_{s_k})|$. 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 (compare with Proposition 2 and Remark 3).

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

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

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

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

Proof. 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 $\mathrm{E}\{|X_0|^i\}$, $1 \leq i \leq k$. Let $J = \mathrm{cum}(X_0, X_{m_1}, \ldots, X_{m_{k-1}})$, $m_0 = 0$ and $n_l = m_l - m_{l-1}$, $1 \leq l \leq k-1$; define 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,

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

Since Y_0 and the random vector $(X'_{m_l - m_{l-1}}, \ldots, X'_{m_{k-1} - m_{l-1}})$ are independent, we have $B = 0$ (see Property (ii) of Rosenblatt (1985, p. 35)). We shall now use the definition (17) and show that $|A_0| \leq C r_1^{n_l/k}$. To this end, let $U_j = X_{m_j - m_{l-1}}$ for $j = 0, 1, \ldots, k-1$, $j \neq l$ 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 $|\mathrm{E}\{\prod_{j \in V} U_j\}| \leq \mathrm{E}\{|X_0|^{|V|}\}$ and

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

by letting $C' = \sum_{i=0}^{k-1} (\mathrm{E}\{|X_0|^{i+1}\})^{i/(1+i)}$. By (17), $|A_0| \leq C r_1^{n_l/k}$ for some constant C. Similarly, for $j = 1, \ldots, k-l-1$,

$$|A_j| \leq C r_1^{(m_{l+j} - m_{l-1})/k} \leq C r_1^{n_l/k}.$$

Hence, $|J| \leq C r_1^{n_l/k}$, which implies (19) since $|J| \leq C \min_{1 \leq l \leq k-1} r_1^{n_l/k}$ and $m_{k-1} = \sum_{l=1}^{k-1} n_l \leq (k-1) \max_{1 \leq l \leq k-1} n_l$.

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

Remark 3. The inequality (19) implies (18) since

$$\sum_{s_1,\ldots,s_7\in\mathbb{Z}} |\mathrm{cum}(X_0, X_{s_1}, \ldots, X_{s_7})| \le 2\sum_{s=0}^{\infty} \sum_{(s_1,\ldots,s_7)\in L(s)} |\mathrm{cum}(X_0, X_{s_1}, \ldots, X_{s_7})|$$
$$= \sum_{s=0}^{\infty} \mathcal{O}[s^6 r^s] < \infty,$$

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. and Fuh, C. D. (2001). Limit theorems for iterated random functions by regenerative methods. *Stoch. Process. Appl.* **96,** 123–142.

Anderson, T. W. (1971). *The Statistical Analysis of Time Series.* John Wiley, New York.

Arnold, L. (1998). *Random Dynamical Systems.* Springer, Berlin.

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

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

Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory,* 2nd edn. Holden-Day, San Francisco.

Chan, K. S. and Tong, H. (2001). *Chaos: A Statistical Perspective.* Springer, New York.

Diaconis, P. and Freedman. D. (1999). Iterated random functions. *SIAM Rev.* **41,** 41–76.

Doukhan, P. and Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stoch. Process. Appl.* **84,** 313–342.

Elton, J. H. (1990). A multiplicative ergodic theorem for Lipschitz maps. *Stoch. Process. 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, D. A. (1975). On tail probabilities for martingales. *Ann. Prob.* **3,** 100–118.

Gordin, M. I. and Lifšic, B. A. (1978). Central limit theorem for stationary Markov processes. *Dokl. Adad. Nauk SSSR* **239,** 766–767 (in Russian). English translation: *Soviet Math. Dokl.* **19** (1978), 392–394.

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

Hutchinson, J. (1981). Fractals and self similarity. *Indiana Univ. Math. J.* **30,** 713–747.

Jarner, S. F. and Tweedie R. L. (2001). Locally contracting iterated functions and stability of Markov chains. *J. Appl. Prob.* **38,** 494–507.

Nicholls, D. F. and Quinn, B. G. (1982). *Random Coefficient Autoregressive Models: An Introduction.* Springer, New York.

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

Rosenblatt, M. (1984). Asymptotic normality, strong mixing, and spectral density estimates. *Ann. Prob.* **12,** 1167–1180.

Rosenblatt, M. (1985). *Stationary Sequences and Random Fields.* Birkhäuser, Boston, MA.

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

Steinsaltz, D. (1999). Locally contractive iterated function systems. *Ann. Prob.* **27,** 1952–1979.

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

Tong, H. (1990). *Nonlinear Time Series. A Dynamical System Approach.* Oxford University Press.

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