The Annals of Applied Probability
1994, Vol. 4, No. 4, 1234–1254

A CONSISTENT APPROACH TO LEAST SQUARES ESTIMATION OF CORRELATION DIMENSION IN WEAK BERNOULLI DYNAMICAL SYSTEMS

BY REGIS J. SERINKO

Pennsylvania State University

A new approach to the least squares procedure for correlation dimension estimation is suggested. Consistency of the new estimator is established for a class of dynamical systems that includes the Cantor map and the logistic map with parameter value 4. Unlike the proofs of consistency for other estimation procedures, no assumptions are made about the Grassberger–Procaccia spatial correlation integral beyond the existence of the correlation dimension.

1. Introduction. Consider a dynamical system $(\Omega, \mathscr{F}, \mu, T)$, where $\Omega \subseteq \Re^d$ is closed, $\mathscr{F}$ is the completion of the Borel σ-field with respect to μ and the projections of μ onto the coordinate axes have finite first moments. Let ρ denote either the Euclidean or max metric on $\Re^d$. The *Grassberger–Procaccia* (GP) *spatial correlation integral* $C(r)$ [12], in terms of which the correlation dimension is defined, is given by

$$C(r) = \iint_{\Omega \times \Omega} I_{S_r}(\omega, \omega')\, d\mu(\omega)\, d\mu(\omega'), \tag{1}$$

where I_E is the indicator of the set E and $S_r = \{(\omega, \omega') \in \Omega \times \Omega\colon \rho(\omega, \omega') \le r\}$. The measurability of S_r with respect to the product σ-field follows from the continuity of ρ. Whenever the limit exists, the *correlation dimension* [12] ν is defined by

$$\nu = \lim_{r \to 0^+} \frac{\log C(r)}{\log r}; \tag{2}$$

it is undefined whenever the limit fails to exist. The correlation dimension may be used to gauge the smoothness of μ over its support [7, 12]. It is also used along with other quantities such as Hausdorff dimension, Liapunov exponents and entropy to characterize dynamical systems. For a detailed discussion of correlation dimension and its relation to other dimensions, see Cutler [7].

The estimation of correlation dimension from an observed segment of an orbit is a popular goal in experimental study of dynamical systems [2, 6, 11].

Received April 1993; revised April 1994.
AMS 1991 *subject classifications*. 60F99 62G99 53F11 28A80
Key words and phrases. Grassberger–Procaccia spatial correlation integral, U-statistic, fractal, convergence in measure, nonlinear dynamics, chaos, itinerate process, absolutely regular.

Although there are a number of different procedures [7, 10–12, 17–19, 22] for this estimation, all of them begin with the estimation of $C(r)$ from a segment of an orbit of a point $\omega_0 \in \Omega$, $\{\omega_k = T^{(k)}(\omega_0)\colon k = 0, 1, 2, \ldots\}$. The standard estimator of $C(r)$ is the *Grassberger-Procaccia empirical spatial correlation integral* [12] $C_n(r; \omega_0)$, which is given by

$$(3) \qquad C_n(r; \omega_0) = \frac{2}{n(n-1)} \sum_{i=0}^{n-1} \sum_{j=i+1}^{n-1} I_{S_r}(\omega_i, \omega_j).$$

If the dynamical system is ergodic and ρ is the max metric, then $C_n(r; \omega_0)$ converges to $C(r)$ uniformly in r a.s. μ [16]. If the dynamical system satisfies a uniform mixing condition known as the *weak Bernoulli property*, $n^{1/2}(C_n(r; \omega_0) - C(r))$ is asymptotically normal for each r [8–10].

The least squares procedure for estimating correlation dimension [7, 10], which formalizes the original and the most commonly used procedure [11, 12], is considered here. The first step in this estimation procedure is to plot $\log C_n(r; \omega_0)$ versus $\log r$ for a variety of values of r. The second step is to identify a linear region in the plot and to fit a least squares line to the points $(\log r_k, \log C_n(r_k; \omega_0))$, $k = 1, 2, \ldots, m$, in this region. The slope of this line $\hat{\nu}_n(r_1, r_2, \ldots, r_m; \omega_0)$ is the least squares estimator of ν.

The motivation for this procedure comes from the observation that if $C(r)$ satisfies *exact scaling*, that is, for r sufficiently small

$$(4) \qquad C(r) = ar^{\nu},$$

where a is a positive constant, then for r near the origin, $\log C(r)$ is a linear function of $\log r$ with slope ν. Under this condition, if the dynamical system is ergodic and ρ is the max metric, the results of [7, 16] imply that $\hat{\nu}_n(r_1, r_2, \ldots, r_m; \omega_0)$ converges to ν a.s. μ, whenever the points $r_1, r_2, \ldots, r_m$ are sufficiently small.

However, dynamical systems that satisfy exact scaling are thought to be atypical [7]. In fact, two often studied dynamical systems—the Cantor map and the logistic map with parameter value 4—satisfy (2), but not (4). In the case of the Cantor map, the GP spatial correlation integral may be written as

$$(5) \qquad C(r) = b(r)r^{\nu},$$

where for r sufficiently small, $b(r)$ is a bounded oscillating function which fails to converge as $r \to 0$ [7, 20]. In the case of the logistic map with parameter value 4, the GP spatial correlation integral may be written as

$$(6) \qquad C(r) = c(r)r^{\nu},$$

where $c(r)$ diverges to infinity as $r \to 0$ and $\log c(r) = o(\log r)$ [7, 12].

Whenever exact scaling is not satisfied, the functional dependence of $\log C(r)$ on $\log r$ is nonlinear even for small r. In this case, if the dynamical system is ergodic, then the results of [7, 16] imply that $\hat{\nu}_n(r_1, r_2, \ldots, r_m; \omega_0)$ converges to $\nu(r_1, r_2, \ldots, r_m)$ a.s. μ, which depends on the points $r_1, r_2, \ldots, r_m$ to which the line is fitted. That is, the procedure is asymptotically biased. The other procedures [17–19, 22] for estimating ν are no better in this regard.

(See [20], [7] and [17, 18] for additional discussions of this problem.)

On the other hand, Cutler [7] has shown that if $r_k^{(m)} = r_0^{2m+1-k}$, $k = 1, 2, \dots, m$, for some $0 < r_0 < 1$, then the asymptotic bias

$$\nu\left(r_1^{(m)}, r_2^{(m)}, \dots, r_m^{(m)}\right) - \nu$$

vanishes as $m \to \infty$. Hence, if the dynamical system is ergodic,

$$\lim_{m\to\infty} \lim_{n\to\infty} \hat{\nu}_n\left(r_1^{(m)}, r_2^{(m)}, \dots, r_m^{(m)}; \omega_0\right) = \nu \quad \text{a.s. } \mu.$$

This suggests that the least squares method can be made consistent by taking the limits as n and m go to infinity simultaneously. However, the least squares estimator is a functional of $\log C_n(r_k^{(m)}; \omega_0)$, $k = 1, 2, \dots, m$, and the convergence of $\log C_n(r; \omega_0)$ to $\log C(r)$ is not, in general, uniform in r. As a result, the double limits cannot be taken along an arbitrary path in the mn plane.

Paths along which the double limit can be taken are exhibited here. First, the problem of the double limits is considered outside the context of least squares estimation. It is shown (Theorem 1), under a set of regularity conditions (see Section 3) on the dynamical system, which include the uniform mixing condition known as the weak Bernoulli property, that there exists a sequence of reals $\{b_n\}_{n=0}^{\infty}$ [see (59)] that vanish as n goes to infinity, such that for any other sequence of real $\{r_n\}_{n=0}^{\infty}$ that vanish no faster than the first,

$$\frac{C_n(r_n; \omega_0) - C(r_n)}{C(r_n)} \xrightarrow{\mu} 0 \tag{7}$$

as $n \to \infty$, whenever the correlation dimension exists. The rate at which $\{b_n\}_{n=0}^{\infty}$ goes to zero depends on the correlation dimension and the mixing rate. It is a nondecreasing function of the mixing rate and a decreasing function of the correlation dimension.

Second (Theorem 3), it is shown, under the same conditions that if r_n goes to zero no faster than b_n, $0 < s < 1$, and $m_n = \frac{1}{2}(1 + \log r_n / \log s)$, then

$$\hat{\nu}_n\left(r_1^{(m_n)}, r_2^{(m_n)}, \dots, r_{[m_n]}^{(m_n)}\right) - \nu\left(r_1^{(m_n)}, r_2^{(m_n)}, \dots, r_{[m_n]}^{(m_n)}\right) \xrightarrow{\mu} 0 \tag{8}$$

as $n \to \infty$, where $[\cdot]$ is the integer part of $\cdot$. Further, the rate of convergence is at least $r_n^{-\delta}(\log r_n)^{1/2}$, where $0 < \delta < 1$. The above result on the asymptotic bias [7] along with (8) leads to a consistent approach to the least squares estimation of correlation dimension (Corollary 1).

The proofs of these results use Chebyshev's inequality and they require an $\mathbf{L}^2$ bound on $C_n(r; \omega_0) - C(r)$. This bound builds on the work of Denker and Keller [9, 10] on the asymptotics of U-statistics for weak Bernoulli dynamical systems.

Both the Cantor map and the logistic map with parameter value 4 are shown to satisfy the conditions of Theorems 1 and 3.

This paper is organized as follows. Background from ergodic theory is presented in Section 2. Section 3 contains the assumptions and preliminary results. The main results, Theorems 1 and 3, are established in Section 4. Examples are presented in the final section.

2. Background from ergodic theory.

2.1. *Introduction.* In this section ideas and results from ergodic theory that are used in the sequel are gathered together. This will set notation and provide background for the uninitiated.

2.2. *Partitions and the weak Bernoulli property.* Throughout this paper it will be assumed that T is an endomorphism. The results extend easily to the case of an automorphism. Let $\alpha = \{A_1, A_2, \dots, A_m\}$ and $\beta = \{B_1, B_2, \dots, B_n\}$ be finite measurable partitions of Ω. From these one may construct the following partitions:

1. $\alpha \vee \beta = \{A \cap B\colon A \in \alpha,\ B \in \beta\}$.
2. $T^{-1}\alpha = \{T^{-1}A\colon A \in \alpha\}$.
3. $\alpha_r^s = T^{(-r)}\alpha \vee T^{(-r-1)}\alpha \vee \cdots \vee T^{(-s+1)}\alpha \vee T^{(-s)}\alpha$; $r < s$, $r, s \in \mathbf{N}$.

Let $\mathscr{F}_r^s$ denote the σ-algebra generated by α_r^s, $r < s$, $r, s \in \mathbf{N}$, and let $\mathscr{F}_0^\infty$ denote the smallest σ-algebra which contains all of the σ-algebras $\mathscr{F}_r^s$, $r < s$, $r, s \in \mathbf{N}$. (This is sometimes denoted by α_0^∞.) A partition is called a *generator* if the completion of $\mathscr{F}_0^\infty$ with respect of μ is $\mathscr{F}$.

A measureable partition α is said to be *weak Bernoulli* for a dynamical system $(\Omega, \mathscr{F}, \mu, T)$ if

$$\beta_k = \sup_{r, s \in N} \sum_{A \in \alpha_0^r} \sum_{B \in \alpha_{r+k}^{r+s+k}} |\mu(A \cap B) - \mu(A)\mu(B)| \tag{9}$$

goes to zero as $k \to \infty$. The β_k's are called the *mixing coefficients.* A dynamical system that has a weak Bernoulli partition will be called weak Bernoulli. The weak Bernoulli property is a uniform mixing condition, which is known in the stochastic processes literature as *absolute regularity.*

2.3. *A metric isomorphism.* Let $(\Omega_i, \mathscr{F}_i, \mu_i, T_i)$, $i = 1, 2$, be two dynamical systems. These systems are said to be *metrically isomorphic* or *equivalent* if there exist sets $\Omega_i' \subseteq \Omega_i$ of full μ_i measure, $i = 1, 2$ and an invertible map $\phi | \Omega_1' \to \Omega_2'$ such that (1) $\phi \circ T_1 = T_2 \circ \phi$ on Ω_1' and (2) $\mu_1 \phi^{-1}(E) = \mu_2(E)$ for all measurable $E \subseteq \Omega_2'$. (See, e.g. [15], page 4].

A condition under which a given dynamical system is equivalent to a shift dynamical system is considered. Let $\Sigma = \{1, 2, \dots, m\}^{\mathbf{N}}$. A point in Σ is a one-sided sequence of the first m integers, that is, if $a \in \Sigma$,

$$a = a_0 a_1 a_2 \cdots, \qquad a_k \in \{1, 2, \dots, m\}, \qquad k \in \mathbf{N}. \tag{10}$$

Let $\phi_\alpha | \Omega \to \Sigma$ be defined by

$$[\phi_\alpha(\omega)]_k = j \quad \text{if } T^{(k)}(\omega) \in A_j, \tag{11}$$

$j = 1, 2, \ldots, m$, $k \in \mathbf{N}$. The left-shift σ is an endomorphism of Σ defined by

$$[\sigma(a)]_k = a_{k+1}, \qquad k \in \mathbf{N}. \tag{12}$$

It is an immediate consequence of (11) and (12) that

$$\phi_\alpha \circ T = \sigma \circ \phi_\alpha. \tag{13}$$

This map induces a probability structure on Σ as follows. A cylinder set $\mathscr{C}(s; j_0, j_1, \ldots, j_r)$ on Σ is given by

$$\mathscr{C}(s; j_0, j_1, \ldots, j_r) = \{a \in \Sigma : a_{s+k} = j_k;\ k = 0, 1, 2, \ldots, r\}, \tag{14}$$

$j_{k'} \in \{1, 2, \ldots, m\}$, $k' = 0, 1, 2, \ldots, r$, and $r, s \in \mathbf{N}$. We denote by $\mathscr{C}_s^{r+s}$ the σ-algebra generated by all cylinders of the above form, $r, s \in \mathbf{N}$, whereas $\mathscr{C}$ will denote the smallest σ-algebra that contains all of these σ-algebras. A measure ν is defined on $\mathscr{C}$ by

$$\nu(\mathscr{C}(s; j_0, j_1, \ldots, j_r)) = \mu\big(T^{(-s)}A_{j_0} \cap T^{(-s-1)}A_{j_1} \cap \cdots \cap T^{(-s-r)}A_{j_r}\big). \tag{15}$$

(This measure does not appear outside of this section. Hence no confusion with correlation dimension should occur.) Two immediate consequences of these definitions are

$$\phi_\alpha^{-1}\mathscr{C} = \mathscr{F}_0^\infty \subseteq \mathscr{F} \tag{16}$$

and

$$\nu = \mu\phi_\alpha^{-1}. \tag{17}$$

The following proposition gives a necessary and sufficient condition for ϕ_α to be a metric isomorphism. Both of the examples considered here satisfy its conditions.

PROPOSITION 1. *For a dynamical system $(\Omega, \mathscr{F}, \mu, T)$, where Ω is a complete separable metric space and $\mathscr{F}$ is the completion of the Borel σ-field with respect to μ, ϕ_α is a metric isomorphism if and only if α is a generator.*

The proof of this result, which is omitted, can be found in [15], pages 16–17, 274.

2.4. *The itinerate process.* The considerations of the last subsection allow one to connect the dynamics of $(\Omega, \mathscr{F}, \mu, T)$ to the time evolution of a stationary stochastic process. Suppose that α is a weak Bernoulli generator for $(\Omega, \mathscr{F}, \mu, T)$. Let

$$\zeta_k(\omega) = [\phi_\alpha(\omega)]_k, \tag{18}$$

$k = 0, 1, 2, \ldots$, $\omega \in \Omega$. The sequence $\zeta_0, \zeta_1, \zeta_2, \ldots$ is a stationary stochastic process, called variously the *itinerate process* or the *label process*, defined on the probability space $(\Omega, \mathscr{F}, \mu)$ with state space $\{1, 2, 3, \ldots, m\}$ and marginals given in (15). A sample path ζ of the process is a random element of Σ given by

$$\zeta(\omega) = \zeta_0(\omega)\zeta_1(\omega)\zeta_2(\omega)\cdots. \tag{19}$$

As a consequence of the definition of $\mathscr{C}_r^s$, $\phi_\alpha^{-1}\mathscr{C}_r^s = \mathscr{F}_r^s$. Therefore, $\mathscr{F}_r^s$ is the σ-field generated by $\zeta_r, \zeta_{r+1}, \dots, \zeta_s$, $r < s$.

Some additional notation that will make later manipulations clearer is introduced. Because ϕ_α is a metric isomorphism, ϕ_α^{-1} exists a.s. ν. Let

$$X_0(\zeta) = \phi_\alpha^{-1}(\zeta) \quad \text{a.s. } \nu. \tag{20}$$

One has $X_0(\zeta(\omega)) = \omega$ a.s. μ. Next let

$$X_j(\zeta) = T^{(j)}(X_0(\zeta)) = X_0(\sigma^{(j)}(\zeta)) \quad \text{a.s. } \nu, \tag{21}$$

$j = 1, 2, \dots$. The second equality is a consequence of (13). This notation emphasizes the fact that points of Ω are functionals of the sample paths of a stationary stochastic process.

Because α is weak Bernoulli for the dynamical system, the process ζ_0, $\zeta_1, \dots$ is absolutely regular. However, the process $X_0(\zeta), X_1(\zeta), \dots$ is not absolutely regular because it is a functional of the entire sample path. An approximation to this latter sequence that is absolutely regular will be needed below. It is given by

$$X_j^{(l)}(\zeta_j, \zeta_{j+1}, \dots, \zeta_{j+l}) = \mathbf{E}\left[X_j(\zeta) | \alpha_j^{j+l}\right], \tag{22}$$

where for any $f \in \mathbf{L}^1(\mu)$,

$$\mathbf{E}\left[f | \alpha_j^{j+l}\right] = \sum_{A \in \alpha_j^{j+l}} \left[\frac{1}{\mu(A)} \int_A f \, d\mu\right] I_A. \tag{23}$$

The expectation on the right-hand side of (22) is finite because μ has finite first moment. The sequence $X_0^{(l)}, X_1^{(l)}, \dots$ is absolutely regular because each term in the sequence is a functional of a finite segment of the sample path. For a complete discussion of this type of approximation, see [4], Section 20.

3. Preliminary results. The main theorem requires a bound on $\|C_n(r) - C(r)\|_2^2$, where $\|\cdot\|_2$ is the $\mathbf{L}^2(\mu)$-norm. It will be slightly easier to work with

$$C_n'(r; \omega_0) = \frac{1}{n^2} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} I_{S_r}(\omega_i, \omega_j) \tag{24}$$

in place of $C_n(r; \omega_0)$. This is possible because

$$[C_n(r; \omega_0) - C_n'(r; \omega_0)] = O(n^{-1}). \tag{25}$$

Henceforth, the prime will be dropped; no confusion should result. The first step in obtaining this $\mathbf{L}^2$ bound is the following decomposition from the theory of U-statistics [14], first introduced in this context by Denker and Keller [10]. One writes,

$$\begin{aligned} C_n(r; \omega_0) &= \frac{1}{n^2} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} I_{S_r}(\omega_i, \omega_j) \\ &= C(r) + \frac{2}{n} \sum_{i=0}^{n-1} \left[\mu(\overline{B}_r(\omega_i)) - C(r)\right] + R_n(r; \omega_0), \end{aligned} \tag{26}$$

where

$$R_n(r;\omega_0) \tag{27}$$
$$= \frac{1}{n^2} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \left[C(r) - \mu\big(\overline{B}_r(\omega_i)\big) - \mu\big(\overline{B}_r(\omega_j)\big) + I_{S_r}(\omega_i, \omega_j) \right]$$

and $\overline{B}_r(\omega) = \{\omega' \in \Omega: \rho(\omega', \omega) \le r\}$. As a consequence of the measurability of S_r and Fubini's theorem ([5], page 234), $\mu(\overline{B}_r(\cdot))$ is a measurable function. Bounds are obtained separately for the main term and the remainder.

The assumptions used in establishing these bounds are collected together. Let

$$\eta^{(l)}(r) = \left\| \mu\big(\overline{B}_r(X_j)\big) - \mu\big(\overline{B}_r\big(X_j^{(l)}\big)\big) \right\|_2^2, \tag{28}$$

$$\psi_k^{(l)}(r) = \left\| I_{S_r}(X_i, X_j) - I_{S_r}\big(X_i^{(l)}, X_j^{(l)}\big) \right\|_2^2, \tag{29}$$

$i, j = 0, 1, 2, \dots,$ $k = |i - j|$. That $\eta^{(l)}$ does not depend on j and $\psi_k^{(l)}$ depends on $|i - j|$ are consequences of stationarity. Let $\overline{\eta}^{(l)} = \sup_r \eta^{(l)}(r)$ and $\overline{\psi}_k^{(l)} = \sup_r \psi_k^{(l)}(r)$.

ASSUMPTION 1. $(\Omega, \mathscr{F}, \mu, T)$ has a weak Bernoulli generator α.

ASSUMPTION 2. $(\Omega, \mathscr{F}, \mu, T)$ is such that the mixing coefficients satisfy $\beta_k^{\delta/(2+\delta)} = O(k^{-(1+\epsilon)})$ for some $\delta > 0$ and $0 < \epsilon < 1$.

ASSUMPTION 3. $(\Omega, \mathscr{F}, \mu, T)$ is such that $\overline{\eta}^{(l)^{1/2}} = o(l^{-(1+\gamma)})$ for some $\gamma > 0$ such that $\gamma/(1 + \gamma) > \epsilon$.

ASSUMPTION 4. $(\Omega, \mathscr{F}, \mu, T)$ is such that for any sequence of reals $\{c_n\}_{n=0}^{\infty}$ satisfying $\lim_{n \to \infty} c_n = \infty$ and $c_n = o(n^{1/2})$, one has

$$\sum_{k=0}^{n-1} \overline{\psi}_k^{(c_n)^{1/2}} = o(n^{1/2}).$$

The first result is an adaptation of a result due to Denker and Keller [10].

PROPOSITION 2 (Denker and Keller [10]). *Suppose that* $(\Omega, \mathscr{F}, \mu, T)$ *satisfies Assumptions* 1 *through* 4. *Then*

$$\|R_n(r;\cdot)\|_2 = o(n^{-1/2}) \tag{30}$$

uniformly in r.

PROOF. Let $R_n^{(l)}$ denote the functional obtained from R_n by replacing ω_i with $X_i^{(l)}$, $i = 0, 1, 2 \dots$. The triangle inequality for $\mathbf{L}^2$-norms gives

$$\|R_n(r;\cdot)\|_2 \le \|R_n^{(l)}(r;\cdot)\|_2 + 2\eta^{(l)^{1/2}}(r) + \frac{2}{n} \sum_{k=0}^{n-1} \psi_k^{(l)^{1/2}}(r) \tag{31}$$

for any $l = 1, 2, \ldots$. Assumptions 1 and 2 along with Proposition 2 of [9] as modified in [10] give

$$\|R_n^{(l)}(r;\cdot)\|_2 \leq \Gamma l n^{-1+\epsilon/2}, \tag{32}$$

where Γ is a positive constant that does not depend on r. Next let $l = c_n = [n^{1/2-\lambda}]$, where $\lambda > \epsilon/2$ and $[\cdot]$ is the integer part. Then

$$\|R_n^{(c_n)}(r;\cdot)\|_2 = o(n^{-1/2}) \tag{33}$$

uniformly in r. On the other hand, Assumption 4 gives

$$\frac{2}{n}\sum_{k=0}^{n-1} \psi_k^{(c_n)^{1/2}}(r) \leq \frac{2}{n}\sum_{k=0}^{n-1} \overline{\psi}_k^{(c_n)^{1/2}} = o(n^{-1/2}), \tag{34}$$

while Assumption 3 yields

$$\eta^{(c_n)^{1/2}}(r) \leq \overline{\eta}^{(c_n)^{1/2}} = o(n^{-(1+\gamma)(1/2-\lambda)}). \tag{35}$$

If $\lambda < \gamma/(2(1+\gamma))$, then

$$\eta^{(c_n)^{1/2}}(r) = o(n^{-1/2}) \tag{36}$$

uniformly in r. The bounds (31), (33), (34), and (36) lead to (30). This completes the proof. □

An $\mathbf{L}^2$ bound for the main term in (26) is obtained through the following two lemmas. The first is a modification of Lemma 1 in [4], page 170, Section 20 which holds for ϕ-mixing processes. Alternatives to the bound given in Lemma 1 may be obtained by exploiting the fact that an absolutely regular stochastic process is strong mixing. (This use of strong mixing should not be confused with its use in ergodic theory.) See [3] and references therein for these alternatives.

LEMMA 1. *Suppose that* $(\Omega, \mathscr{F}, \mu, T)$ *satisfies Assumption* 1, ξ *is measurable* $\mathscr{F}_0^l$ *and* η *is measurable* $\mathscr{F}_{k+l}^\infty$. *Then* $\sup|\xi| \leq 1$ *a.s.* $\mathbf{E}^{1/r}[|\eta|^r] < \infty$, $1 < r < \infty$ *and* $s = r/(r-1)$ *implies*

$$|\mathbf{E}[\xi\eta] - \mathbf{E}[\xi]\mathbf{E}[\eta]| \leq 2^{1/r}\beta_k^{1/s}\mathbf{E}^{1/r}[|\eta|^r]. \tag{37}$$

PROOF. Simple functions are dense in $\mathbf{L}^q$, $1 \leq q < \infty$; see, for example, [1], page 88. Hence without loss of generality one may assume that

$$\eta = \sum_i u_i I_{A_i}, \tag{38}$$

$$\xi = \sum_j v_j I_{B_j}, \tag{39}$$

where $\{A_i\}$ and $\{B_j\}$ are finite partitions of Ω that are measurable $\mathscr{F}_0^l$ and $\mathscr{F}_{k+l}^\infty$, respectively. Under this assumption Hölder's inequality gives

$$\begin{aligned}|\mathbf{E}[\xi\eta] - \mathbf{E}[\xi]\mathbf{E}[\eta]| &= \left|\sum_i u_i\, \mu(A_i)^{1/r}\left[\mu(A_i)^{1/s}\sum_j v_j[\mu(B_j|A_i) - \mu(B_j)]\right]\right| \\ &\le \mathbf{E}^{1/r}[|\eta|^r]\left[\sum_i \mu(A_i)\left|\sum_j v_j[\mu(B_j|A_i) - \mu(B_j)]\right|^s\right]^{1/s}.\end{aligned} \tag{40}$$

It will suffice to show

$$\sum_i \mu(A_i)\left|\sum_j v_j[\mu(B_j|A_i) - \mu(B_j)]\right|^s \le 2^{s/r}\beta_k. \tag{41}$$

For each i, Hölder's inequality gives

$$\begin{aligned}&\left|\sum_j v_j[\mu(B_j|A_i) - \mu(B_j)]\right|^s \\ &\quad\le \left[\sum_j |v_j|^s |\mu(B_j|A_i) - \mu(B_j)|\right]\left[\sum_j |\mu(B_j|A_i) - \mu(B_j)|\right]^{s/r}\end{aligned} \tag{42}$$

Further,

$$\sum_j |v_j|^s |\mu(B_j|A_i) - \mu(B_j)| \le \sum_j |\mu(B_j|A_i) - \mu(B_j)| \tag{43}$$

and

$$\left[\sum_j |\mu(B_j|A_i) - \mu(B_j)|\right]^{s/r} \le 2^{s/r}. \tag{44}$$

Therefore,

$$\begin{aligned}&\sum_i \mu(A_i)\left|\sum_j v_j[\mu(B_j|A_i) - \mu(B_j)]\right|^s \\ &\quad\le 2^{s/r}\sum_i \mu(A_i)\sum_j |\mu(B_j|A_i) - \mu(B_j)| \\ &\quad\le 2^{s/r}\beta_k,\end{aligned} \tag{45}$$

where the last inequality comes from the fact that the dynamical system is weak Bernoulli. This completes the proof. □

Let

$$m_j(r;\cdot) = \mu(\overline{B}_r(X_j(\zeta))), \tag{46}$$

$$m_j^{(l)}(r;\cdot) = \mathbf{E}[m_j(r;\cdot)|\alpha_j^{j+l}], \tag{47}$$

$$\Delta^{(l)}(r) = \|m_j(r;\cdot) - m_j^{(l)}(r;\cdot)\|_2^2 \tag{48}$$

and

$$\theta_k = \sqrt{2}\,\beta_{k-[k/3]}^{1/2} + 4\bar{\Delta}^{([k/3])^{1/2}}, \tag{49}$$

where $\bar{\Delta}^{(l)} = \sup_r \Delta^{(l)}(r)$, $j, k = 0, 1, 2, \dots$. $\Delta^{(l)}(r)$ does not depend on j as a consequence of stationarity.

LEMMA 2. *Suppose that* $(\Omega, \mathscr{F}, \mu, T)$ *satisfies Assumption* 1. *Then*

$$|\mathbf{E}[m_0(r;\cdot)m_k(r;\cdot)] - \mathbf{E}[m_0(r;\cdot)]\mathbf{E}[m_k(r;\cdot)]| \le \theta_k C^{1/2}(r), \tag{50}$$

$k = 0, 1, 2, \dots$.

PROOF. The triangle inequality gives

$$\begin{aligned} &|\mathbf{E}[m_0(r;\cdot)m_k(r;\cdot)] - \mathbf{E}[m_0(r;\cdot)]\mathbf{E}[m_k(r;\cdot)]| \\ &\quad \le \left|\mathbf{E}\left[m_0^{(l)}(r;\cdot)m_k^{(l)}(r;\cdot)\right] - \mathbf{E}\left[m_0^{(l)}(r;\cdot)\right]\mathbf{E}\left[m_k^{(l)}(r;\cdot)\right]\right| \\ &\qquad + \left|\mathbf{E}\left[m_0^{(l)}(r;\cdot)\left(m_k(r;\cdot) - m_k^{(l)}(r;\cdot)\right)\right]\right| \\ &\qquad + \left|\mathbf{E}\left[m_k^{(l)}(r;\cdot)\left(m_0(r;\cdot) - m_0^{(l)}(r;\cdot)\right)\right]\right| \\ &\qquad + \left|\mathbf{E}\left[\left(m_0(r;\cdot) - m_0^{(l)}(r;\cdot)\right)\left(m_k(r;\cdot) - m_k^{(l)}(r;\cdot)\right)\right]\right|, \end{aligned} \tag{51}$$

where $\mathbf{E}[m_k^{(l)}(r;\cdot)] = \mathbf{E}[m_k(r;\cdot)]$ has been used, $k = 0, 1, 2, \dots$, $l = 1, 2, 3, \dots$. Several applications of Hölder's inequality give

$$\begin{aligned} &|\mathbf{E}[m_0(r;\cdot)m_k(r;\cdot)] - \mathbf{E}[m_0(r;\cdot)]\mathbf{E}[m_k(r;\cdot)]| \\ &\quad \le \left|\mathbf{E}\left[m_0^{(l)}(r;\cdot)m_k^{(l)}(r;\cdot)\right] - \mathbf{E}\left[m_0^{(l)}(r;\cdot)\right]\mathbf{E}\left[m_k^{(l)}(r;\cdot)\right]\right| \\ &\qquad + 2\mathbf{E}^{1/2}\left[\left(m_0^{(l)}(r;\cdot)\right)^2\right]\Delta^{(l)^{1/2}}(r) + \Delta^{(l)}(r). \end{aligned} \tag{52}$$

Note that $0 < m_0^{(l)}(r;\cdot) \le 1$ implies $m_0^{(l)}(r;\cdot)^2 \le m_0^{(l)}(r;\cdot)$. Therefore,

$$\begin{aligned} \mathbf{E}^{1/2}\left[\left(m_0^{(l)}(r;\cdot)\right)^2\right] &\le \mathbf{E}^{1/2}\left[m_0^{(l)}(r;\cdot)\right] \\ &= \mathbf{E}^{1/2}\left[\mathbf{E}\left[m_0(r;\cdot)|\alpha_0^l\right]\right] \\ &= \mathbf{E}^{1/2}\left[m_0(r;\cdot)\right] \\ &= C^{1/2}(r). \end{aligned} \tag{53}$$

Further, the triangle inequality for $\mathbf{L}^2$-norms leads to

$$\Delta^{(l)^{1/2}}(r) = \mathbf{E}^{1/2}\left[\left(m_0(r;\cdot) - m_0^{(l)}(r;\cdot)\right)^2\right] \le 2\mathbf{E}^{1/2}\left[\left(m_0(r;\cdot)\right)^2\right]. \tag{54}$$

Therefore,

$$\begin{aligned} &|\mathbf{E}[m_0(r;\cdot)m_k(r;\cdot)] - \mathbf{E}[m_0(r;\cdot)]\mathbf{E}[m_k(r;\cdot)]| \\ &\quad \le \left|\mathbf{E}\left[m_0^{(l)}(r;\cdot)m_k^{(l)}(r;\cdot)\right] - \mathbf{E}\left[m_0^{(l)}(r;\cdot)\right]\mathbf{E}\left[m_k^{(l)}(r;\cdot)\right]\right| \\ &\qquad + 4\Delta^{(l)^{1/2}}(r)C^{1/2}(r). \end{aligned} \tag{55}$$

If $l = [k/3]$, then Lemma 1 with $r = s = 2$ gives

$$\begin{aligned}(56)\quad &\left|\mathbf{E}\left[m_0^{([k/3])}(r;\cdot)m_k^{([k/3])}(r;\cdot)\right] - \mathbf{E}\left[m_0^{([k/3])}(r;\cdot)\right]\mathbf{E}\left[m_k^{([k/3])}(r;\cdot)\right]\right| \\ &\quad \le \sqrt{2}\,\beta_{k-[k/3]}^{1/2}C^{1/2}(r).\end{aligned}$$

Equation (49) and bounds (55) and (56) lead to

$$(57)\quad \left|\mathbf{E}[m_0(r;\cdot)m_k(r;\cdot)] - \mathbf{E}[m_0(r;\cdot)]\mathbf{E}[m_k(r;\cdot)]\right| \le \theta_k C^{1/2}(r).$$

This completes the proof. □

4. The main results. To state the main theorems, let

$$(58)\qquad \lambda_n = \sum_{k=1}^{n-1} \theta_k,$$

where θ_k is defined in (49) and

$$(59)\qquad b_n = \max\left\{\left(\frac{\lambda_n}{n}\right)^{2/(3(\nu+\varepsilon_0))}, \left(\frac{1}{n}\right)^{1/(2(\nu+\varepsilon_1))}\right\},$$

where $\varepsilon_0, \varepsilon_1 > 0$.

THEOREM 1. *If $(\Omega, \mathscr{F}, \mu, T)$ satisfies Assumptions 1 through 4, then whenever ν exists,*

$$(60)\qquad \lim_{n\to\infty} r_n = 0$$

and

$$(61)\qquad \limsup_{n\to\infty} \frac{b_n}{r_n} < \infty$$

imply

$$(62)\qquad \frac{C_n(r_n;\omega_0) - C(r_n)}{C(r_n)} \xrightarrow{\mu} 0$$

as $n \to \infty$.

PROOF. It must be shown that for any $\varepsilon > 0$,

$$(63)\qquad \lim_{n\to\infty} \mu\left[\left\{\omega: \left|\frac{C_n(r_n;\omega_0) - C(r_n)}{C(r_n)}\right| \ge \varepsilon\right\}\right] = 0$$

whenever the sequence $\{r_n\}$ satisfies (60) and (61). For n sufficiently large, Chebyshev's inequality gives the bound

$$(64)\qquad \mu\left[\left\{\omega: \left|\frac{C_n(r_n;\omega_0) - C(r_n)}{C(r_n)}\right| \ge \varepsilon\right\}\right] \le \frac{\|C_n(r_n;\cdot) - C(r_n)\|_2^2}{(\varepsilon C(r_n))^2}$$

It suffices to show that the right-hand side of the above inequality vanishes as $n \to \infty$ under the conditions of this theorem.

To this end, note that (26) and the definition of $m_j(r;\cdot)$ give

$$\|C_n(r_n;\cdot) - C(r_n)\|_2^2 = \left\| \frac{2}{n} \sum_{i=0}^{n-1} [m_i(r_n;\cdot) - C(r_n)] + R_n(r_n;\cdot) \right\|_2^2$$

$$\leq \left\| \frac{2}{n} \sum_{i=0}^{n-1} [m_i(r_n;\cdot) - C(r_n)] \right\|_2^2 + \|R_n(r_n;\cdot)\|_2^2 \tag{65}$$

$$+ 2\left\| \frac{2}{n} \sum_{i=0}^{n-1} [m_i(r_n;\cdot) - C(r_n)] \right\|_2 \|R_n(r_n;\cdot)\|_2,$$

where the inequality comes from expanding the square in the norm and an application of Hölder's inequality. Under Assumptions 1 through 4, Proposition 2 holds. Therefore,

$$\|R_n(r_n;\cdot)\|_2^2 = o(n^{-1}). \tag{66}$$

Note that $C(r) = \mathbf{E}[m_j(r;\cdot)]$. Hence,

$$\left\| \frac{2}{n} \sum_{i=0}^{n-1} [m_i(r_n;\cdot) - C(r_n)] \right\|_2^2 \tag{67}$$

is the variance of a sum of a stationary process and one may write

$$\left\| \frac{2}{n} \sum_{i=0}^{n-1} [m_i(r_n;\cdot) - C(r_n)] \right\|_2^2$$

$$\leq \frac{4}{n} \|m_0(r_n;\cdot) - C(r_n)\|_2^2 \tag{68}$$

$$+ \frac{4}{n} \sum_{k=1}^{n-1} [\mathbf{E}[m_0(r_n;\cdot) m_k(r_n;\cdot)]$$

$$- \mathbf{E}[m_0(r_n;\cdot)]\mathbf{E}[m_k(r_n;\cdot)]].$$

It follows from an expansion of the square in the norm and inequality (53) that

$$\|m_0(r_n;\cdot) - C(r_n)\|_2^2 \leq C(r_n)(1 - C(r_n)). \tag{69}$$

Lemma 2, which holds under Assumption 1, gives

$$\left| \frac{1}{n} \sum_{k=1}^{n-1} [\mathbf{E}[m_0(r_n;\cdot) m_k(r_n;\cdot)] - \mathbf{E}[m_0(r_n;\cdot)]\mathbf{E}[m_k(r_n;\cdot)]] \right|$$

$$\leq \frac{1}{n} \sum_{k=1}^{n-1} \theta_k C^{1/2}(r_n) = \frac{\lambda_n}{n} C^{1/2}(r_n). \tag{70}$$

It follows from this series of inequalities that

$$\|C_n(r_n;\cdot) - C(r_n)\|_2^2 \leq \frac{4\lambda_n}{n} C^{1/2}(r_n)(1 + o(1)) + o(n^{-1})$$

$$+ \left(\frac{4\lambda_n}{n} C^{1/2}(r_n)(1 + o(1)) o(n^{-1}) \right)^{1/2}. \tag{71}$$

It suffices to show

$$\lim_{n\to\infty} \frac{\lambda_n}{n} C^{-3/2}(r_n) = 0 \tag{72}$$

and

$$\lim_{n\to\infty} n^{-1} C^{-2}(r_n) = 0. \tag{73}$$

Given $\epsilon > 0$, the definition of correlation dimension, (2), implies $C(r) = o(r^{\nu-\epsilon})$ and $r^{\nu+\epsilon} = o(C(r))$. Therefore, (72) and (73) are satisfied if

$$\limsup_{n\to\infty} \frac{\lambda_n}{n} \left(r_n^{\nu+\varepsilon_0}\right)^{-3/2} < \infty \tag{74}$$

and

$$\limsup_{n\to\infty} n^{-1} \left(r_n^{\nu+\varepsilon_1}\right)^{-2} < \infty, \tag{75}$$

for some $\varepsilon_0 > 0$ and $\varepsilon_1 > 0$. These are an immediate consequence of (59) and (61). This completes the proof. □

The modification of the least squares method of estimating the correlation dimension is now considered. It can be shown [7] that this estimator is given by

$$\hat{\nu}_n(\mathbf{r}; \omega_0) = \nu + \mathbf{d}(\mathbf{r}) + \epsilon_n(\mathbf{r}; \omega_0), \tag{76}$$

where $\mathbf{d}$ is the asymptotic bias, which is given by

$$\mathbf{d}(\mathbf{r}) = \sum_{i=1}^{m} (v(r_i) - \bar{v})(x_i - \bar{x})/S_{xx}, \tag{77}$$

and $\epsilon_n(\mathbf{r}; \omega_0)$ is the random error, which is given by

$$\epsilon_n(\mathbf{r}; \omega_0) = \sum_{i=1}^{m} (\log C_n(r_i; \omega_0) - \log C(r_i))(x_i - \bar{x})/S_{xx}, \tag{78}$$

and

$$\mathbf{r} = (r_1, r_2, r_3, \dots, r_m), \tag{79}$$

$$x_i = \log r_i, \tag{80}$$

$$\bar{x} = \frac{1}{m} \sum_{i=1}^{m} x_i, \tag{81}$$

$$v(r) = \log C(r) - \nu \log r, \tag{82}$$

$$\bar{v} = \frac{1}{m} \sum_{i=1}^{m} v(r_i), \tag{83}$$

$$S_{xx} = \sum_{i=1}^{m} (x_i - \bar{x})^2. \tag{84}$$

The components of $\mathbf{r}$ give the points $(\log r_k, \log C_n(r_k, \omega_0))$, $k = 1, 2, \dots, m$, to which the least squares line is fitted.

Note that exact scaling implies that $v(r)$ is a constant. Therefore, under exact scaling, if the components of $\mathbf{r}$ are sufficiently small, there is no asymptotic bias, that is, $\mathbf{d}(\mathbf{r}) = 0$. If exact scaling is not satisfied, then, in general, there will be an asymptotic bias. The next result shows that this bias can be made to vanish, if the entire orbit of the system is available.

THEOREM 2 (Cutler [7]). *For any* $0 < s < 1$, $r_0 > 0$ *and* $m \geq 1$,

$$\mathbf{r}^{(m)} = \left(s^{2m-[m]}r_0, s^{2m-[m]+1}r_0, s^{2m-[m]+2}r_0, \dots, s^{2m-1}r_0\right) \in \Re^{[m]}, \tag{85}$$

where $[\cdot]$ *is the integer part of* $\cdot$, *implies*

$$\lim_{m \to \infty} \mathbf{d}(\mathbf{r}^{(m)}) = 0. \tag{86}$$

In fact, this is a slight generalization of Cutler's [7] result. There the case of integer m is considered. However, the proof with noninteger m is nearly identical to the proof with integer m. Therefore, it is omitted.

For any $0 < s < 1$ and $r_0 > 0$, let

$$m_n = \frac{1}{2}\left(1 + \frac{\log r_n}{\log s}\right). \tag{87}$$

THEOREM 3. *If* $(\Omega, \mathscr{F}, \mu, T)$ *satisfies Assumptions* 1 *through* 4, *then whenever* ν *exists*, $\delta < \min\{\frac{3}{4}\varepsilon_0, \varepsilon_1\}$,

$$\lim_{n \to \infty} r_n = 0 \tag{88}$$

and

$$\limsup_{n \to \infty} \frac{b_n}{r_n} < \infty \tag{89}$$

imply

$$r_n^{-\delta}(\log r_n)^{1/2}\epsilon_n\left(\mathbf{r}^{(m_n)}; \omega_0\right) \xrightarrow{\mu} 0 \tag{90}$$

as $n \to \infty$.

PROOF. Define

$$S_{x_n, x_n} = \sum_{j=1}^{[m_n]} \left(x_{n,j} - \bar{x}_n\right)^2, \tag{91}$$

$$\bar{x}_n = \frac{1}{[m_n]} \sum_{j=1}^{[m_n]} x_{n,j}, \tag{92}$$

$$r_k^{(m_n)} = s^{2m_n - k} r_0, \tag{93}$$

$$x_{n,k} = \log r_k^{(m_n)}, \qquad k = 1, 2, \dots, [m_n]. \tag{94}$$

Then

$$(95)\qquad \epsilon_n(\mathbf{r}^{(m_n)}; \omega_0) = \sum_{j=1}^{[m_n]} \left(\log C_n\left(r_j^{(m_n)}; \omega_0\right) - \log C\left(r_j^{(m_n)}\right)\right) \times (x_{n,j} - \bar{x}_n)/S_{x_n, x_n}.$$

The Cauchy-Schwarz inequality gives

$$(96)\qquad |\epsilon_n(\mathbf{r}^{(m_n)}; \omega_0)| \le \left(\sum_{j=1}^{[m_n]} \left(\log C_n\left(r_j^{(m_n)}; \omega_0\right) - \log C\left(r_j^{(m_n)}\right)\right)^2\right)^{1/2} S_{x_n, x_n}^{-1/2}.$$

However,

$$(97)\qquad \begin{aligned} S_{x_n, x_n} &= (\log s)^2 \sum_{j=1}^{[m_n]} \left(j - \frac{[m_n]+1}{2}\right)^2 \\ &= \frac{1}{12}(\log s)^2([m_n]([m_n]+1)([m_n]-1)). \end{aligned}$$

Therefore,

$$(98)\qquad |\epsilon(\mathbf{r}^{(m_n)}; \omega_0)| \le \Delta_n(\omega_0) O(m_n^{-1/2}),$$

where

$$(99)\qquad \Delta_n(\omega_0) = \max_{1 \le j \le [m_n]} \left|\log C_n\left(r_j^{(m_n)}; \omega_0\right) - \log C\left(r_j^{(m_n)}\right)\right|.$$

As $m_n = O(\log r_n)$, it suffices to show, for any $\varepsilon > 0$, that

$$(100)\qquad \lim_{n \to \infty} \mu(\{\omega_0 : \Delta_n(\omega_0) > r_n^{\delta} \varepsilon\}) = 0$$

as $n \to \infty$. To this end, note that subadditivity gives

$$(101)\qquad \mu(\{\omega_0 : \Delta_n(\omega_0) > r_n^{\delta}\varepsilon\}) \le \sum_{j=1}^{[m_n]} \mu(\{\omega_0 : \Delta_{n,j}(\omega_0) > r_n^{\delta}\varepsilon\}),$$

where

$$(102)\qquad \Delta_{n,j}(\omega_0) = \left|\log \frac{C_n\left(r_j^{(m_n)}; \omega_0\right)}{C\left(r_j^{(m_n)}\right)}\right|.$$

Chebyshev's inequality leads to

$$(103)\qquad \mu(\{\omega_0 : \Delta_{n,j}(\omega_0) > r_n^{\delta}\varepsilon\}) \le \frac{\|C_n\left(r_j^{(m_n)}; \cdot\right) - C\left(r_j^{(m_n)}\right)\|_2^2}{C\left(r_j^{(m_n)}\right)^2 \tilde{\varepsilon}_n},$$

where $\tilde{\varepsilon}_n = 1 - 1/\cosh r_n^\delta \varepsilon$. This yields

$$(104)\quad \begin{aligned} \mu(\{\omega_0 : \Delta_n(\omega_0) > r_n^\delta \varepsilon\}) &\le \sum_{j=1}^{[m_n]} \mu(\{\omega_0 : \Delta_{n,j}(\omega_0) > r_n^\delta \varepsilon\}) \\ &\le \sum_{j=1}^{[m_n]} \frac{\|C_n(r_j^{(m_n)};\cdot) - C(r_j^{(m_n)})\|_2^2}{C(r_j^{(m_n)})^2 \tilde{\varepsilon}_n} \\ &\le [m_n] \max_{1 \le j \le [m_n]} \frac{\|C_n(r_j^{(m_n)};\cdot) - C(r_j^{(m_n)})\|_2^2}{C(r_j^{(m_n)})^2 \tilde{\varepsilon}_n}. \end{aligned}$$

It follows from inequality (71), $\tilde{\varepsilon}_n = r_n^{2\delta}\varepsilon^2(1 + o(1))$ and the definition of $r_k^{(m_n)}$, $k = 1, 2, \dots, [m_n]$, that

$$(105)\quad \begin{aligned} &[m_n] \max_{1 \le j \le [m_n]} \frac{\|C_n(r_j^{(m_n)};\cdot) - C(r_j^{(m_n)})\|_2^2}{C(r_j^{(m_n)})^2 \tilde{\varepsilon}_n} \\ &\quad \le [m_n] r_n^{-2\delta} \varepsilon^{-2}(1 + o(1)) \\ &\quad \quad \times \left\{ \frac{4\lambda_n}{n} C(r_1^{(m_n)})^{-3/2}(1 + o(1)) + C(r_1^{(m_n)})^{-2} o(n^{-1}) \right. \\ &\quad \quad \left. + \left(\frac{4\lambda_n}{n} C(r_1^{(m_n)})^{-3/2}(1 + o(1)) C(r_1^{(m_n)})^{-2} o(n^{-1}) \right)^{1/2} \right\}. \end{aligned}$$

The definition of m_n (87) gives $r_1^{(m_n)} = r_n$ and $[m_n] = O(\log r_n)$. Therefore, it suffices to show that

$$(106)\quad \lim_{n \to \infty} \frac{r_n^{-2\delta} \log r_n \lambda_n}{n} C(r_n)^{-3/2} = 0$$

and

$$(107)\quad \lim_{n \to \infty} \frac{r_n^{-2\delta} \log r_n}{n} C(r_n)^{-2} = 0.$$

The definition of correlation dimension (2) implies that for any $\varepsilon' > 0$, $\lim_{r \to 0^+}(r^{\nu + \varepsilon'}/C(r)) = 0$. Take $\varepsilon' < \min\{\varepsilon_0 - \frac{4}{3}\delta, \varepsilon_1 - \delta\}$. Therefore, in light of (74) and (75), (106) and (107) are satisfied because $\delta < \min\{\frac{3}{4}\varepsilon_0, \varepsilon_1\}$ implies

$$(108)\quad \liminf_{n \to \infty} (\log r_n) r_n^{(3/2)(\varepsilon_0 - (4/3)\delta - \varepsilon')} = 0$$

and

$$(109)\quad \liminf_{n \to \infty} (\log r_n) r_n^{2(\varepsilon_1 - \delta - \varepsilon')} = 0.$$

This completes the proof. □

Under the conditions of Theorem 3,

$$\hat{\nu}_n\big(\mathbf{r}^{(m_n)};\,\omega_0\big) - \nu = \mathbf{d}(\mathbf{r}^{(m_n)}) + \epsilon_n\big(\mathbf{r}^{(m_n)};\,\omega_0\big) \overset{\mu}{\to} 0$$

as $n \to \infty$. To make practical use of this, an explicit form for r_n is needed. One choice is $r_n = b_n$. However b_n is a function of ν and the, in general, unknown λ_n. Consequently, this choice does not result in a statistic. The next corollary addresses this problem.

COROLLARY 1. *If* $(\Omega, \mathscr{F}, \mu, T)$ *satisfies Assumptions* 1 *through* 4, $\Omega \subset \Re^d$, $\lambda_n = O(1)$ *and* $r_n = n^{-1/2(d+\epsilon_0)}$ *for some* $\epsilon_0 > 0$, *then* $\hat{\nu}_n(\mathbf{r}^{(m_n)};\,\omega_0)$ *is consistent for* ν.

PROOF. It suffices to show that if

$$r_n = n^{-1/2(d+\epsilon_0)}, \tag{110}$$

then

$$\limsup_{n\to\infty} \frac{b_n}{r_n} < \infty. \tag{111}$$

The definition of b_n (59) and $\lambda_n = O(1)$ give

$$b_n = n^{-1/2(\nu+\varepsilon_0)}. \tag{112}$$

The assumption $\Omega \subseteq \Re^d$ implies $\nu \le d$ [7]. Therefore, $b_n \le r_n$. This completes the proof. □

5. Examples. It is shown that both the Cantor map and the logistic map with parameter value 4 satisfy Assumptions 1 through 4. Further, in both of these examples, $\Omega = [0,1] \subseteq \Re$ and $\lambda_n = O(1)$. Therefore, Corollary 1 gives a consistent estimator in these systems. Other specific examples that satisfy these assumptions may be found in [10], whereas a class of dynamical systems, which satisfy the assumptions, are given in [13].

EXAMPLE 1 (The Cantor map). Consider the dynamical system $(\Omega, \mathscr{F}, \mu, T)$, where $\Omega = [0,1]$, $\mathscr{F}$ is the completion of the Borel σ-field with respect to μ, the Cantor distribution (see [1], page 77), and $T(\omega) = 3\omega \bmod 1$. The correlation dimension of μ is $\log 2/\log 3$ [7].

Assumptions 1 *and* 2. The partition $\alpha = \{(0, \frac{1}{3}], (\frac{1}{3}, \frac{2}{3}], (\frac{2}{3}, 1]\}$ is a generator, because $\mathscr{F}_0^\infty$ is the σ-field generated by the triadic rational intervals, which is just a Borel σ-field. The definition of μ gives

$$\mu(A \cap B) - \mu(A)\mu(B) = 0 \tag{113}$$

for $A \in \alpha_0^r$ and $B \in \alpha_{r+k}^{s+k}$, $r < s$, $k = 1, 2, \dots$. Hence α is weak Bernoulli for the dynamical system with $\beta_k = 0$ if $k \ge 1$. In fact, this is a stronger property known as the Bernoulli property.

Assumption 3. Jensen's inequality gives

$$\eta^{(l)}(r) = \left\| \mu(\overline{B}_r(X_0)) - \mu(\overline{B}_r(X_0^{(l)})) \right\|_2^2$$

$$(114) \qquad \le \iint_{\Omega\times\Omega} \left(I_{S_r}(\omega', X_0(\omega)) - I_{S_r}(\omega', X_0^{(l)}(\omega)) \right)^2 d\mu(\omega')\, d\mu(\omega)$$

$$\le \sum_{A\in\alpha_0^l} \mu(A) \int_\Omega \sup_{\omega,\, \omega''\in A} |I_{S_r}(\omega', \omega) - I_{S_r}(\omega', \omega'')|\, d\mu(\omega').$$

The fact that the square quantity in the second line is either 0 or 1 has been used. $A \in \alpha_0^l$ is a subinterval of $[0, 1]$. Therefore,

$$(115) \qquad \sum_{A\in\alpha_0^l} \sup_{\omega,\, \omega''\in A} |I_{S_r}(\omega', \omega) - I_{S_r}(\omega', \omega'')| \le \bigvee_0^1 I_{\overline{B}_r(\omega')},$$

where $\bigvee_a^b f$, $f|[a, b] \to \Re$, denotes the total variation of f. It is easily verified that $\bigvee_0^1 I_{\overline{B}_r(\omega')} \le 2$ and $\mu(A) = 1/2^l$, $A \in \alpha_0^l$. Therefore,

$$(116) \qquad \eta^{(l)}(r) \le 2^{1-l}.$$

Hence,

$$(117) \qquad \overline{\eta^{(l)}}^{1/2} = o(l^{-(1+\gamma)})$$

for any $\gamma > 0$.

Assumption 4. A bound is found for $\psi_k^{(l)}(r)$. Recall that

$$(118) \qquad \psi_k^{(l)}(r) = \int_\Omega \left(I_{S_r}(X_0(\omega), X_k(\omega)) - I_{S_r}(X_0^{(l)}(\omega), X_k^{(l)}(\omega)) \right)^2 d\mu(\omega).$$

Therefore,

$$(119) \qquad \begin{aligned} \psi_k^{(l)}(r) \le & \sum_{A,\, B\in\alpha_0^l} \mathbf{V}_{AB} |\mu(A\cap T^{-k}B) - \mu(A)\mu(B)| \\ & + \sum_{A,\, B\in\alpha_0^l} \mathbf{V}_{AB}\, \mu(A)\mu(B), \end{aligned}$$

where

$$(120) \qquad \mathbf{V}_{AB} = \sup_{z,\, z'\in A\times B} |I_{S_r}(z) - I_{S_r}(z')|.$$

One may easily verify that

$$(121) \qquad \mathbf{V}_{AB} = \begin{cases} 1, & \text{if } A\cap B + r \ne \varnothing \text{ or } A\cap B - r \ne \varnothing, \\ 0, & \text{otherwise,} \end{cases}$$

where $B \pm r = \{y \in \Re\colon y = x \pm r;\ x \in B\}$. The number of pairs of sets in the partition α_0^l for which $\mathbf{V}_{AB}$ is nonzero is no more than four times the number of sets in the partition. To see this, consider the sets

$$(122) \qquad L_\pm = \{(x, y) \in \Omega\times\Omega\colon y = x \pm r\},$$

$$(123) \qquad V_A = \{(x, y) \in \Omega\times\Omega\colon x = \sup A\},$$

$$(124) \qquad H_A = \{(x, y) \in \Omega\times\Omega\colon y = \sup A\}.$$

The quantity $\mathbf{V}_{AB} = 1$ if and only if one of the following hold:

$$[L_+ \cap V_A] \cap A \times B \neq \varnothing,$$
$$[L_+ \cap H_B] \cap A \times B \neq \varnothing,$$
$$[L_- \cap V_A] \cap A \times B \neq \varnothing,$$
$$[L_- \cap H_B] \cap A \times B \neq \varnothing.$$

However, each of the intersections in brackets contains at most one point. Therefore, the number of pairs of sets such that $\mathbf{V}_{AB} = 1$ is no more than four times the number of sets in the partition. The number of sets of nonzero measure in the partition is 2^l. Hence (119) gives

$$\psi_k^{(l)}(r) \leq \beta_k + 4 \times 2^{-l} = 2^{2-l}, \tag{125}$$

$k > l$. Therefore, for any sequence of reals $\{c_n\}_{n=0}^{\infty}$ satisfying $\lim_{n\to\infty} c_n = \infty$ and $c_n = o(n^{1/2})$, one has

$$\begin{aligned} \sum_{k=0}^{n-1} \overline{\psi}_k^{(c_n)^{1/2}} &= \sum_{k=0}^{[c_n]} \overline{\psi}_k^{(c_n)^{1/2}} + \sum_{k=[c_n]+1}^{n-1} \overline{\psi}_k^{(c_n)^{1/2}} \\ &\leq 2([c_n]+1) + ((n-1) - ([c_n]+1))(2)^{1/2(1-[c_n])} \\ &= o(n^{1/2}). \end{aligned} \tag{126}$$

Consistency. It is shown that $\lambda_n = \sum_{k=1}^{n-1}\theta_k = O(1)$ and, therefore, the corollary to Theorem 3 applies. The coefficient θ_k is considered. Because $\beta_k = 0$ if $k \geq 1$,

$$\theta_k = 4\overline{\Delta}^{([k/2])^{1/2}}. \tag{127}$$

Further,

$$\begin{aligned} \Delta^{(l)}(r) &= \left\| \mu(\overline{B}_r(X)) - \mathbf{E}\left[\mu(\overline{B}_r(X)) | \alpha_0^l \right] \right\|_2^2 \\ &\leq \sum_{A \in \alpha_0^l} \mu(A) \Big(\operatorname*{ess\,sup}_A \mu(\overline{B}_r(X)) - \operatorname*{ess\,inf}_A \mu(\overline{B}_r(X)) \Big)^2 \\ &\leq \sum_{A \in \alpha_0^l} \mu(A) \int_\Omega \sup_{\omega', \omega'' \in A} |I_{S_r}(\omega, \omega') - I_{S_r}(\omega, \omega'')| \, d\mu(\omega) \\ &\leq 2^{1-l}, \end{aligned} \tag{128}$$

where the last inequality follows from the same reasoning that led to (116). Consequently, Corollary 1 gives that the least squares estimator with $r_n = n^{-1/2(1+\varepsilon_0)}$ is consistent for ν.

EXAMPLE 2 (The logistic). Consider the dynamical system $(\Omega, \mathscr{F}, \mu, T)$, where $\Omega = [0,1]$. $\mathscr{F}$ is the completion of the Borel σ-field with respect to μ, which is given by

$$\mu(E) = \frac{1}{\pi}\int_E \frac{d\omega}{\sqrt{\omega(1-\omega)}}, \tag{129}$$

$E \in \mathscr{F}$, and

$$T(\omega) = 4\omega(1-\omega) \tag{130}$$

is the logistic with parameter 4. μ has correlation dimension 1 [12, 7].

Assumptions 1 *and* 2. The map $\Phi|\Omega \to \Omega$, given by

$$\Phi(\omega) = \frac{2}{\pi}\sin^{-1}(\sqrt{\omega}), \tag{131}$$

first introduced by Ulam and von Neumann [21], is a metric isomorphism between the original dynamical system and $(\Omega', \mathscr{F}', \lambda, S)$, where $\Omega' = (0, 1]$, $\mathscr{F}'$ is the completion of the Borel σ-field with respect to the Lebesgue measure λ and $S(\omega) = 2\omega \bmod 1$. S is the *dyadic map* of the interval. It is discussed at length in [5]. The partition $\kappa = \{(0, \frac{1}{2}], (\frac{1}{2}, 1]\}$ is a generator for the second dynamical system for which

$$|\lambda(A \cap B) - \lambda(A)\lambda(B)| = 0 \tag{132}$$

if $A \in \kappa_0^r$ and $B \in \kappa_{r+k}^{s+k}$, $r < s$, $k \geq 1$. Therefore, the partition α is a weak Bernoulli generator, with $\beta_k = 0$ if $k \geq 1$, for the dyadic map. Because Φ is a metric isomorphism, the partition $\alpha = \Phi^{-1}\kappa = \{(0, \frac{1}{2}], (\frac{1}{2}, 1]\}$ is a weak Bernoulli generator for the original dynamical system, with $\beta_k = 0$ if $k \geq 1$. Therefore, Assumptions 1 and 2 are satisfied.

Assumptions 3 *and* 4. Note that

$$\mu(A) = \mu\Phi^{-1}(B) = \lambda(B) = 2^{-l}, \tag{133}$$

where $A \in \alpha_0^l$ and $B = \Phi A \in \kappa_0^l$, and that the number of partitions of nonzero measure in α_0^l is the same as in κ_0^l, which is 2^l. This observation and the arguments in the previous example show that this dynamical system satisfies Assumptions 3 and 4.

Consistency. The same arguments used in Example 1 yield $\lambda_n = O(1)$. Therefore, the least squares estimator with $r_n = n^{-1/2(1+\varepsilon_0)}$ is consistent for ν.

Acknowledgments. Partial support for this work was provided by a postdoctoral fellowship from the Smeal College of Business, Pennsylvania State University. I have benefited during the early stages of this research from conversations with Arkady Templeman and Susan Murphy. I have also profited from discussions with Colleen Cutler.

REFERENCES

[1] ASH, R. B. (1972). *Real Analysis and Probability*. Academic Press, New York.

[2] ATTEN, P. and MALRAISON, B. (1988). Characterisation du chaos deterministe recontre dans les experiences. In *Le Chaos* (P. Berge, ed.) 283–325. Extrolles, Paris.

[3] BABU, G. J. and SINGH, K. (1978). Probabilities of moderate deviations for some stationary strong mixing process. *Sankhyā* **40** 38–43.

[4] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
[5] BILLINGSLEY, P. (1986). *Probability and Measure*. Wiley, New York.
[6] CASDAGLI, M. (1992). Chaos and deterministic versus stochastic nonlinear modeling. *J. Roy. Statist. Soc. B* **54** 303–328.
[7] CUTLER, C. D. (1991). Some results on the behavior and estimation of fractal dimension of distributions on attractors. *J. Statist. Phys.* **62** 651–708.
[8] DENKER, M. (1982). Statistical decision procedures and ergodic theory. *Math. Res.* **12** 35–47.
[9] DENKER, M. and KELLER, G. (1983). On *U*-statistics and v.Mises' statistics for weakly dependent processes. *Z. Wahrsch. Verw. Gebiete* **64** 505–522.
[10] DENKER, M. and KELLER, G. (1986). Rgorous statistical procedures for data from dynamical systems. *J. Statist. Phys.* **49** 67–93.
[11] ECKMANN, J.-P. and RUELLE, D. (1985). Ergodic theory of strange attractors. *Rev. Mod. Phys.* **57** 617–656.
[12] GRASSBERGER, P. and PROCACCIA, I. (1983). Measuring the strangeness of strange attractors. *Phys. D* **9** 189–208.
[13] HOFBAUER, F. and KELLER, G. (1982). Ergodic properties of invariant measures for piecewise monotonic transformations. *Math. Z.* **180** 119–140.
[14] HOEFFDING, W. (1948). A class of statistics with asymptotically normal distributions. *Ann. Math. Statist.* **19** 293–325.
[15] PETERSEN, K. (1983). *Ergodic Theory*. Cambridge Univ. Press.
[16] SERINKO, R. J. (1993). A Glivenko–Cantelli theorem for the Grassberger–Procaccia spatial correlation integral for ergodic dynamical systems. Technical Report 93-3, Dept. Statistics, Pennsylvania State Univ.
[17] SMITH, R. L. (1991). Optimal estimation of fractal dimension. In *Non linear Modeling and Chaos* (M. Casdagli and E. Eubank, eds.) **12**. Addison-Wesley, Reading, MA.
[18] SMITH, R. L. (1992). Estimating dimension in noisy chaotic time series. *J. Roy. Statist. Soc. B* **54** 329–351.
[19] TAKENS, F. (1985). On the numerical determination of dimension of a strange attractor. *Dynamical Systems and Bifurcations. Lecture Notes in Math.* **1125** 99–106. Springer, Berlin.
[20] THEILER, J. (1988). Lacunarity in a best estimator of correlation dimension. *Phys. Lett. A* **133** 195–200.
[21] ULAM, S. M. and VON NEUMANN, J. (1947). On combinations of stochastic and deterministic processes. *Bull. Am. Math. Soc.* **53** 1120.
[22] WELLS, R., SHIRE, H. N., FOSMIRE, C. and DORAN, J. A. (1993). Improved algorithms for estimating the correlation dimension and the associated probable errors. Technical Report AM114, Dept. Mathematics, Pennsylvania State Univ.

DEPARTMENT OF STATISTICS
PENNSYLVANIA STATE UNIVERSITY
UNIVERSITY PARK, PENNSYLVANIA 16802