Journal of Statistical Physics, Vol. 85, Nos. 1/2, 1996

Ergodic Theorems Arising in Correlation Dimension Estimation

Regis J. Serinko[1]

Received September 6, 1995; final November 15, 1995

The Grassberger–Procaccia (GP) empirical spatial correlation integral, which plays an important role in dimension estimation, is the proportion of pairs of points in a segment of an orbit of length n, of a dynamical system defined on a metric space, which are no more than a distance r apart. It is used as an estimator of the GP spatial correlation integral, which is the probability that two points sampled independently from an invariant measure of the system are no more than a distance r apart. It has recently been proven, for the case of an ergodic dynamical system defined on a separable metric spaceythat the GP empirical correlation integral converges a.s. to the GP correlation integral at continuity points of the latter as $n \to \infty$. It is shown here that for ergodic systems defined on $\Re^d$ with the "max" metric the convergence is uniform in r. Further, a simplified proof based on weak convergence arguments of the result in separable spaces is given. Finally, the Glivenko–Cantelli theorem is used to obtain ergodic theorems for both the moment estimators and least square estimators of correlation dimension.

KEY WORDS: Glivenko–Cantelli theorem; fractal; almost sure convergence; moment estimators; least square estimators; dynamical systems; chaos.

1. INTRODUCTION

Let μ be a probability measure on the Borel sets $\mathscr{B}$ of a metric space (X, ρ).

Set $S_r = \{(x, x') \in X \times X: \rho(x, x') \leqslant r\}$. The *Grassberger–Procaccia (GP) spatial correlation integral* $C(r)$ of μ is defined to be[13]

$$C(r) = \mu \times \mu(S_r) \tag{1}$$

[1] Department of Statistics, Purdue University, West Lafayette, Indiana 47907; e-mail: rege@stat.purdue.edu.

0022-4715/96/1000-0025$09.50/0

where the measureability of S_r relative to the product σ-field follows from the continuity of ρ. Clearly, $C(r)$ is the probability that two points sampled independently from μ are no more than a distance r apart. Let T be a measure preserving transformation with respect to μ. Put $x_n = T(x_{n-1}) = T^{(n)}(x_0)$ for some $x_0 \in X$, where $T^{(n)}$ is the nth-fold composition of T with itself and let

$$\mu_n^{x_0} = \frac{1}{n} \sum_{k=0}^{n-1} \partial_{x_k} \tag{2}$$

where ∂_x is the unit point mass at x. The GP *empirical spatial correlation integral* $C_n(r; x_0)$ is given by[13]

$$C_n(r; x_0) = \mu_n^{x_0} \times \mu_n^{x_0}(S_r) \tag{3}$$

$C_n(r; x_0)$ is the probability that two points selected with replacement from the first n points of the orbit of x_0 are no more than a distance r apart.[2] The main result of this paper is the following Glivenko–Cantelli theorem.

Theorem 1. If $X \subseteq \Re^d$ and ρ is the "max" metric, then ergodicity implies

$$\lim_{n \to \infty} \sup_r |C_n(r; x_0) - C(r)| = 0 \tag{4}$$

a.s. μ.

Correlation integrals arise in the empirical studies of dynamical systems. One objective in such studies is the estimation of invariants of a system from the observation of time series produced by it.[9] These, in turn, are used to characterize the system. A popular invariant to estimate is the *correlation dimension* ν,[13] which is defined by

$$\nu = \lim_{r \to 0^+} \frac{\log C(r)}{\log r} \tag{5}$$

whenever the limit exists; it is undefined otherwise. (For a recent review of dimension estimation, see Cutler.[6]) The above result is used to obtain ergodic theorems for the moment estimators and standard least square estimators of correlation dimension. Only the proof of the almost sure convergence of the moment estimator, with $p = 0$ (see Lemma 1), makes use of

[2] Sometimes the GP empirical correlation integral is defined to be the probability that two points selected without replacement from the first n points of the orbit are no more than a distance r apart. The difference in these two quantities is $O(n^{-1})$. Hence the conclusions of this paper apply equally to either.

the uniformity of the convergence of $C_n(r; x_0)$ to $C(r)$. It is shown that moment and standard least square estimators converge almost surely to ν, under the conditions of Theorem 1, if and only if there exist positive constants c and r_0 such that

$$C(r) = cr^{\nu} \qquad \text{if} \quad r \leqslant r_0 \tag{6}$$

This property is called *exact scaling*. The necessity of exact scaling is a weakness of these estimators. In ref. 18 it is shown how to modify the least square estimators to obtain an estimator of correlation dimension which is consistent under the assumption of the existence of ν.

The choice of metric space in Theorem 1 is that most often encountered in practice. The primary reason for using the "max" metric instead of the Euclidean metric is that a fast algorithm[10] exists for computing $C_n(r; x_0)$ in this case, while there is no price to pay for this convenience since the correlation dimension is the same in either case.

Often one is concerned with the image of a segment of an orbit under some function, rather than the orbit itself. Theorem 1 can easily be generalized to this setting. To do so, let $(\Omega, \mathscr{F}, m, S)$ be a dynamical system and let (Y, τ) be a metric space. Take h measurable from Ω to Y. The GP spatial correlation integral of mh^{-1} is given by

$$C(h)(r) = mh^{-1} \times mh^{-1}(S'_r) \tag{7}$$

where $S'_r = \{(y, y') \in Y \times Y: \tau(y, y') \leqslant r\}$. Put $\omega_k = S(k)(\omega_0)$ for some $\omega_0 \in \Omega$ and $y_k = h(\omega_k)$, $k = 1, 2, \ldots$. The sequence $\{y_k\}$ is the image of the orbit of ω_0 with respect to h. set

$$m_n^{\omega_0} h^{-1} = \frac{1}{n} \sum_{k=0}^{n-1} \partial_{\omega_k} h^{-1} = \frac{1}{n} \sum_{k=0}^{n-1} \partial_{y_k} \tag{8}$$

The empirical GP spatial correlation integral for the image of the orbit is given by

$$C_n^{(h)}(r; \omega_0) = m_n^{\omega_0} h^{-1} \times m_n^{\omega_0} h^{-1}(S'_r) \tag{9}$$

The modification of Theorem 1 is as follows.

Theorem 2. If $Y = \mathfrak{R}^d$ and τ is the "max" metric, then ergodicity of $(\Omega, \mathscr{F}, m, S)$ implies

$$\lim_{n \to \infty} \sup_r |C_n^{(h)}(r; \omega_0) - C^{(h)}(r)| = 0 \tag{10}$$

a.s. m.

Theorem 2 is of particular value in the understanding[6, 7] of the OP phenomena.[15] In that case, $\Omega = C[0, \infty)$, S is the left shift, and h is a finite dimensional projection.

A result [17] in the same direction as Theorem 1 is the following,

Theorem 3.[3] If (X, ρ) is separable, then ergodicity implies

$$\lim_{n \to \infty} C_n(r; x_0) = C(r) \tag{11}$$

a.s. μ at continuity points of $C(r)$.

In addition, Aaronson *et al.*[1] used a weak convergence argument to show that $C_n(r; x_0)$ converges to $C(r)$ a.s. μ at continuity points of $C(r)$ whenever $X \subseteq \Re^1$. In fact, as shown here, this argument yields a simple proof of Theorem 3 and ties this result to a large literature on almost sure weak convergence.

The question of uniform convergence when (X, ρ) is separable is unanswered unless $C(r)$ is a continuous function. However, $C(r)$ need not be continuous. There are two ways in which discontinuities can arise when the "max" metric is used. First, under the assumption of ergodicty, $C(r)$ will be discontinuous if μ is supported on a finite set. In this case, a simple combinatoric argument yields Theorem 1. The other way in which discontinuities can arise is if μ assigns positive mass to parallel hyperplanes. The following two examples show that such a measure can be an invariant measure of an ergodic system. In fact, the second example is Bernoulli.

Example 1. Let $X_1 = \{0, 1\}$,

$$T_1(x) = \begin{cases} 0 & \text{if } x = 1 \\ 1 & \text{if } x = 0 \end{cases} \tag{12}$$

$\mu_1 = 1/2(\partial_0 + \partial_1)$ and and let $([0, 1], \mathscr{B}, \mu_2, T_2)$ be a weak-mixing dynamical system. Set $X = X_1 \times X_2$, $T = T_1 \times T_2$, and $\mu = \mu_1 \times \mu_2$. The ergodicity of the product dynamical system follows from the weak mixing of the second factor and the ergodicity of the first factor. It is easy to show that $C(r)$ for the product system has a discontinuity at $r = 1$ with jump size equal to 1/2. This construction can be used to obtain ergodic dynamical systems in any dimension greater than 1 with correlation integrals with a finite number of discontinuities.

[3] The statement of this theorem conforms with the proof in ref. 17, not the statement given there.

Example 2. Let $Z=[0,1)$, $T_1(z)=2z \bmod 1$, and μ_1 be the Lebesgue measure on the Borel sets of Z. It is well known that the partition $\{[0,1/2),[1/2,1)\}$ is Bernoulli for this dynamical system. Let

$$\begin{aligned} Z_n &= [1-1/2^{n-1}, 1-1/2^n) \\ X_n &= [0, 1/2^n), \qquad n=1,2,... \end{aligned} \tag{13}$$

Define $X=\bigcup_{n=1}^{\infty}(X_n \times \{1/2^{n-1}\})$ and $\varphi \mid Z \to X$ by

$$\varphi(z)=(z-\varphi_0(z), 1-\varphi_0(z)) \tag{14}$$

where $\varphi_0(z)=1-\sum_n 1/2^{n-1} I_{Z_n}(z)$. Here $\varphi^{-1} \mid X \to Z$ is given by $\varphi^{-1}(x_1,x_2)=x_1+1-x_2$. Let $T \mid X \to X$ be given by $T=\varphi \circ T_1 \circ \varphi^{-1}$ and define μ on the Borel sets of $\Re^d$ restricted to X by $\mu=\mu_1\varphi^{-1}$. By construction, the second system is metrically isomorphic to the first; therefore it is Bernoulli. Further, the second system has a correlation integral with discontinuities at $r=2^{-n}$ of size 4^{-n}, $n=1,2,...$.

Finally, Denker and Keller[8] have shown for certain weak Bernoulli dynamical systems in $\Re^d$ that

$$\sqrt{n}\,[C_n(r;x_0)-C(r)] \tag{15}$$

converges weakly to a normal distribution for each r.

The paper is organized as follows. The ergodic theorems for the estimators of correlation dimension are stated and proven in the next section. The proofs of Theorems 1–3 are given in Section 3.

2. ERGODIC THEOREMS FOR THE ESTIMATORS

2.1. Background

The following definitions are needed below. For any distribution function $F(r)$ and real r', such that $F(r')>0$, define the truncated distribution function with truncation point r', $F(r \mid r')$, by

$$F(r \mid r')=\begin{cases} F(r)/F(r') & \text{if } r \leqslant r' \\ 1 & \text{if } r > r' \end{cases} \tag{16}$$

and the quantile function F^{-1} by

$$F^{-1}(u)=\inf\{x \in \Re = F(x) \geqslant u\}, \qquad 0<u\leqslant 1 \tag{17}$$

$$F^{-1}(0)=\lim_{\varepsilon \to 0^+} F^{-1}(\varepsilon) \tag{18}$$

The following corollaries, which are standard results in the empirical process literature, will be used in the proofs of this section. They are stated for completeness. Theorem 1 and the fact that $C_n(r; x_0)$ and $C(r)$ are distribution functions imply that $C_n(r; x_0)$ converges weakly to $C(r)$ a.s. μ. Therefore one has the following.

Corollary 1. Under the assumptions of Theorem 1,

$$\lim_{n\to\infty} \int_0^\infty f(r)\, dC_n(r; x_0) = \int_0^\infty f(r)\, dC(r) \tag{19}$$

a.s. μ for all f which are bounded and continuous on the support of $dC(r)$.

Corollary. 2 Suppose that $C(r)$ is strictly increasing for $s < r < t$. Then under the assumptions of Theorem 1,

$$\lim_{n\to\infty} \sup_{u_1 \leqslant u \leqslant u_2} |C_n^{-1}(u; x_0) - C^{-1}(u)| = 0 \tag{20}$$

a.s. μ, where $C(s) < u_1 < u_2 < C(t)$. Further, if $-\infty < s < t < \infty$; $C(s') = 0$, $s' < s$, and $C(t) = 1$, then Eq. (17) holds with $u_1 = 0$ and $u_2 = 1$.

This is a consequence of Theorem 1 and the uniform continuity of C^{-1} on $[u_1, u_2]$.

2.2. Moment Estimators

Suppose that the GP spatial correlation integral satisfies

$$C(r) = a(r)\, r^{v}, \qquad 0 \leqslant r \leqslant r_0 \tag{21}$$

for some $r_0 < 0$, where $a(r)$ is a slowly varying function, i.e., $\lim_{r\to 0^+} a(tr)/a(r) = 1$, $t > 0$. Set

$$M(p \mid r') = \begin{cases} \displaystyle\int_0^{r'} \left(\frac{r}{r'}\right)^p dC(r \mid r') & \text{if} \quad p > 0 \\[2ex] \displaystyle\int_0^{r'} \log\left(\frac{r}{r'}\right) dC(r \mid r') & \text{if} \quad p = 0 \end{cases} \tag{22}$$

The slow variation of $a(r)$ is equivalent to the existence of the following limit

$$M(p) = \lim_{r'\to 0^+} M(p \mid r'), \qquad p \geqslant 0 \tag{23}$$

(See Theorem 1 ref. 11, p. 281.) Under the assumption of slow variation, it can be shown[20, 19] that

$$\nu = \begin{cases} pM(p)/[1 - M(p)] & \text{if} \quad p > 0 \\ -1/M(p) & \text{if} \quad p = 0 \end{cases} \tag{24}$$

The first step in this estimation procedure is to approximate ν by

$$\beta(p \mid r') = \begin{cases} pM(p \mid r')/[1 - M(p \mid r')] & \text{if} \quad p > 0 \\ -1/M(p \mid r') & \text{if} \quad p = 0 \end{cases} \tag{25}$$

for some $0 < r' \leqslant r_0$. The second step is to estimate $\beta(p \mid r')$ by

$$\beta_n(p; x_0 \mid r') = \begin{cases} pM_n(p; x_0 \mid r')/[1 - M_n(p; x_0 \mid r')] & \text{if} \quad p > 0 \\ -1/M_n(p; x_0 \mid r') & \text{if} \quad p = 0 \end{cases} \tag{26}$$

where

$$M_n(p; x_0 \mid r') = \begin{cases} \displaystyle\int_0^{r'} \left(\frac{r}{r'}\right)^p dC_n(r; x_0 \mid r') & \text{if} \quad p > 0 \\[2ex] \displaystyle\int_0^{r'} \log\left(\frac{r}{r'}\right) dC_n(r; x_0 \mid r') & \text{if} \quad p = 0 \end{cases} \tag{27}$$

Takens[19] was the first to propose $\beta_n(p; x_0 \mid r')$ with $p = 0$ as an estimator of ν. However, he only considered the special case of exact scaling. Further, he did not consider its almost sure limit under reatistic assumptions on the dynamics. Wells *et al.*[20] were the first to consider the case of $p > 0$. They were able to find almost sure limits under strong mixing, for a slightly modified estimators, which are based on two independent orbits of the dynamical system. With Theorem 1 it is possible to obtain the almost sure limits under ergodicity.

Lemma 1. Under the conditions of Theorem 1, if $p > 0$, or $p = 0$, $\nu > 1$, and $C(r)$ is continuous in a neighborhood of the origin, then

$$\lim_{n \to \infty} \beta_n(p; x_0 \mid r') = \beta(p \mid r') \tag{28}$$

a.s. μ. Further, $\beta_n(p; x_0 \mid r')$ converges to ν a.s. μ if and only if $C(r)$ satisfes exact scaling and $r \leqslant r_0$.

The condition $\nu > 1$ only assures the continuity of $C(r)$ at the origin. Example 2 shows that there are correlation integrals which are continuous at the origin, but not continuous in any neighborhood of the origin. Hence, continuity in a neighborhood of the origin does not follow from ergodicity and $\nu > 1$.

Proof of Lemma 1. If $p > 0$, it follows immediately from Corollary 1 that

$$\lim_{n\to\infty} M_n(p; x_0 \mid r') = M(p \mid r') \tag{29}$$

a.s. μ. Therefore

$$\lim_{n\to\infty} \beta_n(p; x_0 \mid r') = \beta(p \mid r') \tag{30}$$

a.s. μ.

On the other hand, note that $\log r$ is not bounded on $(0, r_0)$; therefore, if $p = 0$, Corollary 1 cannot be used to obtain the almost sure limit of $M_n(p; x_0 \mid r')$. Instead, Corollary 2 will be used to show that

$$\lim_{n\to\infty} \int_0^{r'} \log r \, dC_n(r; x_0 \mid r') = \int_0^{r'} \log r \, dC(r \mid r') \tag{31}$$

a.s. μ. Two changes of variables and the concavity of $\log x$ give

$$\begin{aligned} &\int_0^{r'} \log r \, dC_n(r; x_0 \mid r') - \int_0^{r'} \log r \, dC(r \mid r') \\ &\quad = \int_0^1 [\log C_n^{-1}(u; x_0 \mid r') - \log C^{-1}(u \mid r')] \, du \\ &\quad \leqslant \log\left[\int_0^1 \frac{C_n^{-1}(u; x_0 \mid r')}{C^{-1}(u \mid r')} \, du\right] \end{aligned} \tag{32}$$

It follows from Theorem 1 that $C_n(r; x_0 \mid r')$ converges uniformly to $C(r \mid r')$ a.s. μ. This, together with the assumed continuity of $C(r)$ near the origin and the slow variation of $a(r)$, implies that $C_n^{-1}(u; x_0 \mid r')$ and $C^{-1}(u \mid r')$ obey the conclusion of Corollary 2 over $[0,1]$. Further, $\nu > 1$ implies that

$$\int_0^1 \frac{1}{C^{-1}(u \mid r')} \, du < \infty$$

Therefore one has

$$\lim_{n\to\infty}\left|\int_0^1 \frac{C^{-1}(u; x_0|r')}{C^{-1}(u|r')}\,du - 1\right|$$

$$= \lim_{n\to\infty}\left|\int_0^1 \frac{C_n^{-1}(u; x_0|r') - C^{-1}(u|r')}{C^{-1}(u|r')}\,du\right|$$

$$\leqslant \lim_{n\to\infty}\sup_u |C_n^{-1}(u; x_0|r') - C^{-1}(u|r')|$$

$$\times\left|\int_0^1 \frac{1}{C^{-1}(u|r')}\,du\right| = 0 \qquad (33)$$

a.s. μ. This completes the proof of the first part of the lemma.

Clearly, if $C(r)$ satisfies exact scaling, then $\beta(p|r') = \nu$, $p \geqslant 0$, if $r' \leqslant r_0$. Therefore the moment estimators converge to ν a.s. μ if $r' \leqslant r_0$. Next, suppose that $\beta_n(p; x_0|r')$ converges to ν a.s. μ if $r' \leqslant r_0$, that is, $\beta(p|r') = \nu$ if $r' \leqslant r_0$, $p \geqslant 0$. Then the definition of $\beta(p|r')$ yields, after some manipulation,

$$C(r')\,r'^p = (\nu + p)\int_0^{r'} r^{p-1} C(r)\,dr \qquad (34)$$

$r' \leqslant r_0$, $p \geqslant 0$. The right-hand side is differentiable, therefore one has

$$\frac{dC(r')}{dr'} = \nu C(r') \qquad (35)$$

$r' \leqslant r_0$. This equation has the solution

$$C(r) = cr^\nu \qquad (36)$$

if $r \leqslant r_0$, where c is a positive constant. This completes the proof.

2.3. Standard Least Square Estimators

Suppose that the GP spatial correlation integral satisfies

$$C(r) = a(r)\,r^\nu \qquad (37)$$

with $\lim_{r\to 0^+} \log a(r)/\log r = 0$. A standard least square estimator of ν is given by

$$\hat{\nu}_n(\mathbf{r}; x_0) = \nu + \mathbf{d}(\mathbf{r}) + \varepsilon_n(\mathbf{r}; x_0) \qquad (38)$$

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

$$\mathbf{d}(\mathbf{r}) = \sum_{i=1}^{m} v(r_i)(x_i - \bar{x})/S_{xx} \tag{39}$$

and $\varepsilon_n(\mathbf{r}; x_0)$ is the random error, which is given by

$$\varepsilon_n(\mathbf{r}; x_0) = \sum_{i=1}^{m} [\log C_n(r_i; x_0) - \log C(r_i)](x_i - \bar{x})/S_{xx} \tag{40}$$

and

$$\mathbf{r} \in \mathscr{D}^m \tag{41}$$

$$x_i = \log r_i \tag{42}$$

$$\bar{x} = \frac{1}{m} \sum_{i=1}^{m} x_i \tag{43}$$

$$v(r) = \log a(r) \tag{44}$$

$$S_{xx} = \frac{1}{m} \sum_{i=1}^{m} (x_i - \bar{x})^2 \tag{45}$$

with

$$\mathscr{D}^m = \{(r_1, r_2, \ldots, r_m) \in (0, \infty)^m: r_i \neq r_j \text{ for some } i \neq j,\ i, j = 1, 2, \ldots, m\} \tag{46}$$

The estimator is the slope of the least square line fit to the points

$$(\log r_i, \log C_n(r_i; x_0)), \qquad i = 1, 2, \ldots, m \tag{47}$$

As is seen in the next two results, the choice of points to which the line is fit will effect the asymptotic accuracy of the estimator, unless exact scaling is satisfied.

Lemma 2. 1. $\mathbf{d}(\mathbf{r}) = 0$ for all $\mathbf{r} \in \mathscr{D}^m$ with $\max_{1 \leqslant i \leqslant m} r_i \leqslant r_0$, $m = 2, 3, \ldots$, if and only if for some positive constant c

$$C(r) = cr^{\nu} \qquad \text{if} \quad r \leqslant r_0 \tag{48}$$

2. $\lim_{\lambda \to 0^+} \mathbf{d}(\lambda \mathbf{r}) = 0$ for all $\mathbf{r} \in \mathscr{D}^m$, $m = 2, 3, \ldots$, if and only if $a(r)$ is slowly varying.

3. Take $r>0$ and $0<s<1$. For each m, let $\mathbf{r}^{(m)}=(s^m r, s^{m+1}r,\ldots, s^{2m-1}r)$. Then

$$\lim_{m\to\infty} \mathbf{d}(\mathbf{r}^{(m)})=0 \tag{49}$$

Proof. 1. If $C(r)$ satisfies exact scaling and $\max_{1\leqslant i\leqslant m} r_i \leqslant r_0$, then it is easily shown that $\mathbf{d}(\mathbf{r})=0$. Next suppose that $\mathbf{d}(\mathbf{r})=0$ for any $\mathbf{r}\in\mathscr{D}^m$, with $\max_{1\leqslant i\leqslant m} r_i\leqslant r_0$. Let

$$\bar{x}_k=\sum_{i=1}^{k} x_i/k \tag{50}$$

$$\Delta_{i,k}=x_i-\bar{x}_k \tag{51}$$

$i=1, 2,\ldots, k$; $k=2, 3,\ldots, m$. In this notation,

$$\mathbf{d}(\mathbf{r})=\sum_{i=1}^{m} v(r_i)\,\Delta_{i,m}=0 \tag{52}$$

Note that

$$\Delta_{i,m}=\begin{cases}\Delta_{i,m-1}+[\bar{x}_{m-1}-x_m]/m & \text{if}\quad i=1, 2,\ldots, m-1\\ (m-1)[x_m-\bar{x}_{m-1}]/m & \text{if}\quad i=m\end{cases} \tag{53}$$

Substitution of Eq. (53) into Eq. (52) gives

$$\begin{aligned}\sum_{i=1}^{m} v(r_i)\,\Delta_{i,m}&=\sum_{i=1}^{m-1} v(r_i)\,\Delta_{i,m-1}\\ &\quad+\left\{\sum_{i=1}^{m-1} v(r_i)[\bar{x}_{m-1}-x_m]+v(r_m)(m-1)[x_m-\bar{x}_{m-1}]\right\}\Big/m\\ &=\left\{\sum_{i=1}^{m-1} v(r_i)[\bar{x}_{m-1}-x_m]+v(r_m)(m-1)[x_m-\bar{x}_{m-1}]\right\}\Big/m\\ &=0\end{aligned} \tag{54}$$

It immediately follows that

$$v(r_m)=\sum_{i=1}^{m-1} v(r_i)/(m-1) \tag{55}$$

The right-hand side does not depend on r_m if $\max_{1\leqslant i\leqslant m} r_i \leqslant r_0$. Therefore, the left-hand side is constant for $r_m \leqslant r_0$. Hence $C(r)$ satisfies exact scaling.

2. If $a(r)$ is slowly varying, then for any $r' > 0$,

$$\begin{aligned}\lim_{\lambda\to 0^+} \mathbf{d}(\lambda \mathbf{r}) &= \lim_{\lambda\to 0^+} \sum_{i=1}^{m} v(\lambda r_i)\, \Delta_{i,m}\\ &= \sum_{i=1}^{m} \lim_{\lambda\to 0^+} [v(\lambda r_i) - v(\lambda r')]\, \Delta_{i,m}\\ &= \sum_{i=1}^{m} \lim_{\lambda\to 0^+} \log[a(\lambda r_i)/a(\lambda r')]\, \Delta_{i,m}\\ &= 0\end{aligned} \tag{56}$$

Next suppose that $\lim_{\lambda\to 0^+} \mathbf{d}(\lambda \mathbf{r}) = 0$ for any $\mathbf{r} \in \mathscr{D}^m$, $m = 2, 3,\ldots$. Then considerations similar to those leading to Eq. (55) give

$$\lim_{\lambda\to 0^+} \sum_{i=1}^{m-1} \log[a(\lambda r_i)/a(\lambda r_m)] = 0 \tag{57}$$

This is equivalent to

$$\lim_{\lambda\to 0^+} \frac{a(\lambda r_1)}{a(\lambda r_m)} \frac{a(\lambda r_2)}{a(\lambda r_m)} \cdots \frac{a(\lambda r_{m-1})}{a(\lambda r_m)} = 1 \tag{58}$$

This equation is invariant under the interchange of indices; therefore

$$\begin{aligned}&\lim_{\lambda\to 0^+} \frac{a(\lambda r_1)}{a(\lambda r_{m-1})} \frac{a(\lambda r_2)}{a(\lambda r_{m-1})} \cdots \frac{a(\lambda r_{m-2})}{a(\lambda r_{m-1})} \frac{a(\lambda r_m)}{a(\lambda r_{m-1})}\\ &= \lim_{\lambda\to 0^+} \frac{a(\lambda r_1)}{a(\lambda r_m)} \frac{a(\lambda r_2)}{a(\lambda r_m)} \cdots \frac{a(\lambda r_{m-1})}{a(\lambda r_m)} \left(\frac{a(\lambda r_m)}{a(\lambda r_{m-1})}\right)^m\\ &= \lim_{\lambda\to 0^+} \left(\frac{a(\lambda r_m)}{a(\lambda r_{m-1})}\right)^m\\ &= 1\end{aligned} \tag{59}$$

Hence, $\lim_{\lambda\to 0^+} a(\lambda r_m)/a(\lambda r_{m-1}) = 1$ for any $r_j > 0$, $j = m-1, m$. Therefore $a(r)$ is slowly varying.

3. See Cutler.[5]

Lemma 3. Under the conditions of Theorem 1,

$$\lim_{n \to \infty} \hat{\nu}_n(\mathbf{r}; x_0) = \mathbf{d}(\mathbf{r}) + \nu \tag{60}$$

a.s. μ. Further, $\hat{\nu}_n(\mathbf{r}; x_0)$ converges to ν a.s. μ if and only if $C(r)$ satisfies exact scaling and $\max_{1 \leqslant i \leqslant m} r_i \leqslant r_0$.

Proof. One has

$$\begin{aligned} \lim_{n \to \infty} \varepsilon_n(\mathbf{r}; x_0) &= \lim_{n \to \infty} \sum_{i=1}^{m} [\log C_n(r_i; x_0) - \log C(r_i)](x_i - \bar{x})/S_{xx} \\ &= \sum_{i=1}^{m} \lim_{n \to \infty} [\log C_n(r_i; x_0) - \log C(r_i)](x_i - \bar{x})/S_{xx} \end{aligned} \tag{61}$$

By Theorem 1 and the continuity of $\log x$, one has

$$\lim_{n \to \infty} [\log C_n(r_i; x_0) - \log C(r_i)] = 0 \tag{62}$$

a.s. μ. Therefore

$$\begin{aligned} \lim_{n \to \infty} \hat{\nu}_n(\mathbf{r}; x_0) &= \mathbf{d}(\mathbf{r}) + \nu + \lim_{n \to \infty} \varepsilon_n(\mathbf{r}; x_0) \\ &= \mathbf{d}(\mathbf{r}) + \nu \end{aligned} \tag{63}$$

a.s. μ. This completes the proof of the first part of the lemma.

The second part of the lemma follows from the first part and Lemma 2, part 1. This completes the proof.

Remark 1. Exact scaling of $C(r)$ is not sufficient for either the moment estimators or the standard least square estimators to be strongly consistent. One must also know r_0. In practice, this is unlikely to be the case. Consequently, even in this, the best-behaved case, modifications to these estimators are needed to make them consistent.

Remark 2. Without any assumption on $a(r)$ one has under the conditions of Lemma 3,

$$\begin{aligned} \lim_{n \to \infty} \lim_{n \to \infty} \hat{\nu}_n(\mathbf{r}^{(m)}; x_0) &= \nu + \lim_{m \to \infty} \mathbf{d}(\mathbf{r}^{(m)}) \\ &\quad + \lim_{m \to \infty} \lim_{n \to \infty} \varepsilon_n(\mathbf{r}^{(m)}; x_0) \\ &= \nu \end{aligned} \tag{64}$$

In ref. 18 it is shown how the limits in (64) may be taken simultaneously to yield a consistent estimator of ν without additional assumptions on $C(r)$ beyond the existence of ν. However, the assumption of ergodicity is strengthen to the weak Bernoulli mixing and the almost sure limit is weakened to a limit in measure.

3. THE PROOFS

Proof of Theorem 1. First note that $C_n(\mathbf{r}; x_0)$ and $C(r)$ are distribution functions. Therefore it suffices to show that $C_n(r; x_0)$ converges to $C(r)$ a.s. μ for each r and that $C_n(r^-, x_0) = \lim_{\varepsilon \to 0^+} C_n(r; x_0)$ converges to $C(r^-) = \lim_{\varepsilon \to 0^+} + C(r-\varepsilon)$ a.s. μ for each r. The uniformity will follow from the Glivenko–Cantelli theorem for distribution functions (ref.4, pp.275–276).

In what follows fix r. It will be convenient to write

$$C_n(r; x_0) - C(r) = \int_{S_r} \mu_n^{x_0} \times \mu_n^{x_0}(dz) - \int_{S_r} \mu \times \mu(dz) \tag{65}$$

A vertical section of S_r, $S_r^x = \bar{B}_r(x)$, where $\bar{B}_r(x) = \{x' \in \Re^d \colon \rho(x, x') \leqslant r\}$ is the closed ball in $\Re^d$ of radius r centered at x. Therefore the measurability of S_r with respect to the product σ-field along with Fubini's theorem (ref. 4, p. 240) and the addition and subtraction of terms gives

$$\begin{aligned} C_n(r; x_0) - C(r) &= \int \mu_n^{x_0}(S_r^x)\, \mu_n^{x_0}(dx) - \int \mu(S_r^x)\, \mu(dx) \\ &= \int \mu_n^{x_0}(\bar{B}_r(x))\, \mu_n^{x_0}(dx) - \int \mu(\bar{B}_r(x))\, \mu(dx) \\ &= \int [\mu_n^{x_0}(\bar{B}_r(x)) - \mu(\bar{B}_r(x))]\, \mu_n^{x_0}(dx) \qquad (66) \\ &\quad + \int \mu(\bar{B}_r(x))\, \mu_n^{x_0}(dx) - \int \mu(\bar{B}_r(x))\, \mu(dx) \qquad (67) \end{aligned}$$

Fibini's theorem also yields $\mu(\bar{B}_r(x)) \in L^1(\mu)$. Therefore the pointwise ergodic theorem (ref. 2, p. 13] implies that the term in (67) goes to zero as $n \to \infty$ a.s. μ. one has for the term in (66), as

$$\begin{aligned} &\int |\mu_n^{x_0}(\bar{B}_r(x)) - \mu(\bar{B}_r(x))|\, \mu_n^{x_0}(dx) \\ &\quad \leqslant \sup_x |\mu_n^{x_0}(\bar{B}_r(x)) - \mu(\bar{B}(x))| \end{aligned}$$

The closed balls in $\mathfrak{R}^d$ with respect to the "max" metric take the form

$$\bar{B}_r(x) = [x_1 - r, x_1 + r] \times [x_2 - r, x_2 + r] \times \cdots \times [x_d - r, x_d + r]$$

where $x = (x_1, x_2, ..., x_d)$. Krickeberg[14] has proven the uniform convergence in $\mathfrak{R}^d$ of Cartesian products of connected real sets, of which these balls are a subfamily. His argument uses a theorem due to Gaenssler[12] which assumes i.i.d. observations, but only in order to use the strong law of large numbers. One may substitute the pointwise ergodic theorem in place of the strong law of large numbers; hence the conclusion of Krickeberg's result holds under the assumptions of this theorem. Therefore

$$\lim_{n \to \infty} \sup_x |\mu_n^{x_0}(\bar{B}_r(x)) - \mu(\bar{B}_r(x))| = 0 \tag{68}$$

a.s. μ.

It follows from the above argument that $C_n(r; x_0)$ converges to $C(r)$ as $n \to \infty$ a.s. μ for each r. This same argument works with open balls to give convergence of $C_n(r^-, x_0)$ to $C(r^-)$ a.s. μ. for each r. This completes the proof.

Proof of Theorem 2. It suffices to note that the pointwise ergodic theorem (ref. 2, p. 13) along with the measureability of h imply that

$$\lim_{n \to \infty} m_n^{\omega_0} h^{-1}(\bar{B}'_r(y)) = mh^{-1}(\bar{B}'_r(y)) \tag{69}$$

a.s. m, where $\bar{B}'_r(y) = \{y' \in \mathfrak{R}^d: \tau(y, y') \leq r\}$. Consequently, the proof of Theorem 1 carries over with μ and $\mu_n^{x_0}$ replaced by mh^{-1} and $m_n^{\omega_0}h^{-1}$, respectively.

Proof of Theorem 3. The pointwise ergodic theorem (ref. 2, p. 13) gives $\mu_n^{x_0}(A)$ converges to $\mu(A)$ as $n \to \infty$ a.s. μ for any Borel set A. In a separable metric space this implies that $\mu_n^{x_0}$ converges weakly to μ as $n \to \infty$ a.s. μ (ref. 16, p. 53). Again, in a separable metric space the weak convergence of $\mu_n^{x_0}$ to μ a.s. μ implies that $\mu_n^{x_0} x \mu_n^{x_0}$ converges weakly to $\mu \times \mu$ as $n \to \infty$ a.s. μ (ref. 3, p. 21). This, in turn, implies for any $f \mid X \times X \to \mathfrak{R}$ which is bounded and continuous almost everywhere $\mu \times \mu$ that

$$\lim_{n \to \infty} \left[\int f(z)\, \mu_n^{x_0} \times \mu_n^{x_0}(dz) - \int f(z)\, \mu \times \mu(dz) \right] = 0 \tag{70}$$

a.s. μ. The theorem follows immediately from the fact that the indicator of $S_r I_{S_r}$ is bounded and continuous almost everywhere, $\mu \times \mu$ if r is a continuity point of $C(r)$.

ACKNOWLEDGMENT

I would like to thank Steve Lalley for providing Example 2.

REFERENCES

1. J. Aaronson, R. Burton, H. Dehling, D. Gilat, T. Hili, and B. Weiss, Law of large numbers for U-statistics of dependent observation, *Preprint* (1993).
2. P. Billingsley, *Ergodic Theory and Information* (Wiley, New York, 1965).
3. P. Billingsley, *Convergence of Probability Measure* (Wiley, New York, 1968).
4 P. Billingsley, *Probability and Measure, 2nd ed.* (Wiley, New York, 1986).
5. C. Cutler, Some results on the behavior and estimation of fractal dimensions of distributions on attractors, *J. Stat. Phys.* **62**:651–708 (1991).
6 C. Cutler, A review of the theory and estimation of fractal dimension, in *Nonlinear Time Series and Chaos, Vol 1: Dimension Estimation and Models*, H. Tong, ed. (World Scientific, Singapore, 1993), pp. 1–107.
7. C. Cutler, A theory of correlation dimension for stationary time series, Technical Report STAT-94-17, Department of Statistics, University of Waterloo (1994).
8. M. Denker, and G. Keller, Rigorous statistical procedures for data from dynamical systems, *J. Stat. Phys.* **44**:67–93 (1986).
9. J. -P. Eckmann, and D. Ruelle, Ergodic theory of chaos and strange attractors, *Rev. Mod. Phys.* **57**:617–656 (1985).
10. J.-P. Eckmann, S. O. Kamphorst, D. Ruelle, and S. Ciliberto, Liapunov exponents from time series, *Phys. Rev. A* **34**:4971–4979 (1986).
11. W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 2 (Wiley, New York, 1971).
12. P. B. Gaenssler, Around the Glivenko–Cantelli theorem, *Coll. Math. Soc. Janos Bolyai* **11**:93–103 (1974).
13. P. Grassberger, and I. Procaccia, Measuring the strangeness of strange attractors, *Physica D* **9**:189–208 (1983).
14. K. Krickeberg, An alternative approach to Glivenko–Cantelli theorems, in *Lecture Notes in Mathematics* No. 566, (Springer-Verlag, New York. 1976), pp. 57–67.
15. A. R. Osborne, and A. Provenzale, Finite correlation dimension for stochastic systems with power-law spectra, *Physica D* **35**:357–381 (1989).
16. K. R. Parthasarathy, Probability measures on Metric Spaces (Academic Press, New York, (1967).
17. Ya. B. Pesin, On rigorous mathematical definitions of correlation dimension and generalized spectrum for dimension, *J. Stat. Phys.* **71**:529–547 (1993).
18. R. Serinko, A consistent approach to least squares estimation of correlation dimension in weak Bernoulli dynamical system, *Ann. Appl. Prob.* **4**:1234–1254 (1994).
19. F. Takens, On the numerical determination of the dimension of an attractor, in *Lecture Notes in Mathematics*, No. 1125 (Springer-Verlag, Berlin, 1984), pp. 99–106.
20. R. Wells, H. N. Shire, C. Fosmire, and J. A. Doran, Improved algorithms for estimating the correlation dimension and the associated probable errors, Penn State Department of Mathematics Technical Report, #AM114 (1993).