# ERGODIC THEOREMS ARISING IN CORRELATION DIMENSION ESTIMATION

by

Regis J. Serinko
Purdue University

Technical Report #95-34

Department of Statistics
Purdue University

July 1995

◇

# Ergodic Theorems Arising in Correlation Dimension Estimation

Regis J. Serinko
Department of Statistics
Purdue University
West Lafayette, IN 47907
e-mail *rege@stat.purdue.edu*

July 12, 1995

## Abstract

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 space that 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 that 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.



**Keywords:** *Glivenko-Cantelli theorem, fractal, almost sure convergence, moment estimators, least square estimators, dynamical systems, chaos.*
**Short Title:** *Ergodic Theorems*

# 1 Introduction

Let $\mu$ be a probability measure on the Borel sets $\mathcal{B}$ of a metric space $(X, \rho)$. Set $S_r = \{(x, x') \in X \times X : \rho(x, x') \leq r\}$. The *Grassberger-Procaccia (GP) spatial correlation integral* $C(r)$ of $\mu$ [13] is defined to be

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

where the measureability of $S_r$ relative to the product $\sigma$-field follows from the continuity of $\rho$. Clearly, $C(r)$ is the probability that two points sampled independently from $\mu$ are no more than a distance $r$ apart. Let $T$ be a measure preserving transformation with respect to $\mu$. Put $x_n = T(x_{n-1}) = T^{(n)}(x_0)$ for some $x_0 \in X$, where $T^{(n)}$ is the $n^{th}$-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 $\partial_x$ is the unit point mass at $x$. The *GP empirical spatial correlation integral* $C_n(r; x_0)$ [13] is given by

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

$C_n(r; x_0)$ 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. [1] The main result of this paper is the following Glivenko-Cantelli theorem.

**Theorem 1** *If $X = \Re^d$ and $\rho$ 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. $\mu$.*

The 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

[1] Some times the GP empirical correlation integral is defined to be the probability that two points selected with out 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.

characterize the system. A popular invariant to estimate is the *correlation dimension* $\nu$ [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 an up to date review ofdimension estimation, see Cutler [6].) The above result is used to obtain erogdic theorems for the moment estimators and standard least square estimators of correlation dimension. It is shown that these estimators converge almost surely to $\nu$, under the conditions of Theorem 1, if and only if there exists positive constants $c$ and $r_0$ such that,

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

This property is called *exact scaling.* In Serinko [19], 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 $\nu$.

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 fast algorithm [?] 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 $(\Omega, \mathcal{F}, m, S)$ be a dynamical system and let $(Y, \tau)$ be a metric space. Take $h$ measureable from $\Omega$ 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 $S'_r = \{(y, y') \in Y \times Y : \tau(y, y') \leq r\}$. 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 $\omega_0$ 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 $Y = \Re^d$ and $\tau$ is the 'max' metric, then ergodicity of $(\Omega, \mathcal{F}, m, S)$ implies,*

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

*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** *If $(X, \rho)$ is separable, then ergodicity implies*

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

*a.s. $\mu$ 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 *a.s.* $\mu$ to $C(r)$ at continuity points of $C(r)$ when $X = \Re^1$. In fact, as shown here, this argument yields a simple proof of Theorem 3. The question of uniform convergence when $(X, \rho)$ is separable is unanswered. Finally, it should be noted that Denker and Keller [8] have shown for certain weak Bernoulli dynamical systems in $\Re^d$ that

$$\sqrt{n}\,[C_n(r; x_0) - C(r)] \tag{12}$$

converges weakly to a normal distribution for each $r$.

The paper is organized as follows. The ergodic theorems for the estimtors of correlation dimension are stated and proven in the next section. The proofs of Theorems 1 through 3 are given in section 3. Finally, the proof of a lemma used in the proof of the Theorem 1 is presented for completeness in an appendix.

## 2 Ergodic Theorems for the Estimators

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 $r'$ $F(r|r')$ by

$$F(r|r') = \begin{cases} \frac{F(r)}{F(r')} & \text{if } r \leq r' \\ 1 & \text{if } r > r' \end{cases}, \tag{13}$$

and the quantile function $F^{-1}$ by

$$F^{-1}(u) = \inf\{x \in \Re : F(x) \geq u\} \qquad 0 < u \leq 1, \tag{14}$$
$$F^{-1}(0) = \lim_{\epsilon \to 0+} F^{-1}(\epsilon). \tag{15}$$

The following corollaries, which are standard results in the empirical process literature, will be used in the examples 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.* $\mu$. Therfore one has the following.

**Corollary 1** *Under the assumptions of Theorem 1,*

$$\lim_{n \to \infty} \int_0^\infty f(r)\, dC_n(r; x_0) = \int_0^\infty f(r)\, dC(r) \tag{16}$$

*a.s.* $\mu$ *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,*

$$\lim_{n \to \infty} \sup_{u_1 \leq u \leq u_2} |C_n^{-1}(u; x_0) - C^{-1}(u)| = 0, \tag{17}$$

*a.s.* $\mu$, *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 equation (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 $[u_1, u_2]$.

*Example 1.* MOMENT ESTIMATORS [20] [21]. Suppose that the GP spatial correlation integral satisfies

$$C(r) = a(r)r^{\nu}, \qquad 0 \le r \le r_0, \tag{18}$$

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

$$M(p|r') = \begin{cases} \int_0^{r'} \left(\frac{r}{r'}\right)^p dC(r|r') & \text{if } p > 0, \\ \int_0^{r'} \log\left(\frac{r}{r'}\right) dC(r|r') & \text{if } p = 0. \end{cases} \tag{19}$$

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

$$M(p) = \lim_{r \to 0^+} M(p|r'), \tag{20}$$

$p \ge 0$. (See Theorem 1 of Feller [11, p. 281].) Under the assumption of slow variation, it can be shown [21, 20] that

$$\nu = \begin{cases} pM(p)/[1 - M(p)] & \text{if } p > 0, \\ -1/M(p) & \text{if } p = 0. \end{cases} \tag{21}$$

The first step in this estimation procedure is to approximate $\nu$ by

$$\beta(p|r') = \begin{cases} pM(p|r')/[1 - M(p|r')] & \text{if } p > 0, \\ -1/M(p|r') & \text{if } p = 0, \end{cases} \tag{22}$$

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

$$\beta_n(p; x_0|r') = \begin{cases} pM_n(p; x_0|r')/[1 - M_n(p; x_0|r')] & \text{if } p > 0, \\ -1/M_n(p; x_0|r') & \text{if } p = 0, \end{cases} \tag{23}$$

where

$$M_n(p; x_0|r') = \begin{cases} \int_0^{r'} \left(\frac{r}{r'}\right)^p dC_n(r; x_0|r') & \text{if } p > 0, \\ \int_0^{r'} \log\left(\frac{r}{r'}\right) dC_n(r; x_0|r') & \text{if } p = 0. \end{cases} \tag{24}$$

Takens [20] was the first to propose $\beta_n(p; x_0|r')$ with $p = 0$ as an estimator of $\nu$. However, he only considered the special case of constant $a(r)$ and he did not consider its almost sure limit under realistic assumptions on the dynamics. Wells *et. al.*[21] 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*

$$\lim_{n\to\infty} \beta_n(p; x_0|r') = \beta(p|r'), \tag{25}$$

*a.s. $\mu$. Further, $\beta_n(p; x_0|r')$ converges to $\nu$ a.s. $\mu$ if and only if $C(r)$ satisfies exact scaling and $r \le r_0$.*

**Proof:** If $p > 0$, it follows immediately from Corollary 1 that

$$\lim_{n\to\infty} M_n(p; x_0|r') = M(p|r') \tag{26}$$

*a.s.* $\mu$. Therefore

$$\lim_{n\to\infty} \beta_n(p; x_0|r') = \beta(p|r') \tag{27}$$

*a.s.* $\mu$.

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 $M_n(p; x_0|r')$. Instead, Corollary 2 will be used to show that,

$$\lim_{n\to\infty} \int_0^{r'} \log r \, dC_n(r; x_0|r') = \int_0^{r'} \log r \, dC(r|r') \tag{28}$$

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

$$\begin{aligned} \int_0^{r'} \log r \, dC_n(r; x_0|r') - \int_0^{r'} \log r \, dC(r|r') &= \\ \int_0^1 \left[\log C_n^{-1}(u; x_0|r') - \log C^{-1}(u|r')\right] du &\le \\ \log\left[\int_0^1 \frac{C_n^{-1}(u; x_0|r')}{C^{-1}(u|r')} \, du\right] \end{aligned} \tag{29}$$

It follows from Theorem 1 that $C_n(r; x_0|r')$ converges uniformly to $C(r|r')$ *a.s.* $\mu$. This together with the assumed continuity of $C(r)$ near the origin implies that $C_n^{-1}(u;x_0|r')$ and $C^{-1}(u|r')$ obey the conclusion of Corollary 2 over $[0,1]$. Further, $\nu>1$ implies that $\int_0^1 \frac{1}{C^{-1}(u|r')}du < \infty$. Therefore one has

$$\lim_{n\to\infty} |\int_0^1 \frac{C_n^{-1}(u;x_0|r')}{C^{-1}(u|r')}du - 1| =$$
$$\lim_{n\to\infty} |\int_0^1 \frac{C_n^{-1}(u;x_0|r') - C^{-1}(u|r')}{C^{-1}(u|r')}du| \leq$$
$$\lim_{n\to\infty} \sup_u |C_n^{-1}(u;x_0|r') - C^{-1}(u|r')| \times |\int_0^1 \frac{1}{C^{-1}(u|r')}du| = 0 \quad (30)$$

*a.s.* $\mu$. This completes the proof of the first part of the lemma.

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

$$C(r')\,r'^p = (\nu+p)\int_0^{r'} r^{p-1}C(r)\,dr, \quad (31)$$

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

$$\frac{dC(r')}{dr'} = \nu C(r'), \quad (32)$$

$r'\leq r_0$. This equation has the solution,

$$C(r) = cr^\nu, \quad (33)$$

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

*Example 2.* STANDARD LEAST SQUARE ESTIMATORS. Suppose that the GP spatial correlation integral satisfies

$$C(r) = a(r)r^\nu, \quad (34)$$

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

$$\hat{\nu}_n(\mathbf{r};x_0) = \nu + \mathbf{d}(\mathbf{r}) + \epsilon_n(\mathbf{r};x_0), \quad (35)$$

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

$$\mathbf{d}(\mathbf{r}) = \sum_{i=1}^{m} v(r_i)(x_i - \overline{x})/S_{xx}, \tag{36}$$

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

$$\epsilon_n(\mathbf{r}; x_0) = \sum_{i=1}^{m}(\log C_n(r_i; x_0) - \log C(r_i))(x_i - \overline{x})/S_{xx}, \tag{37}$$

and

$$\mathbf{r} \in \mathcal{D}^m, \tag{38}$$
$$x_i = \log r_i, \tag{39}$$
$$\overline{x} = \frac{1}{m}\sum_{i=1}^{m} x_i, \tag{40}$$
$$v(r) = \log a(r), \tag{41}$$
$$S_{xx} = \frac{1}{m}\sum_{i=1}^{m}(x_i - \overline{x})^2, \tag{42}$$

with

$$\mathcal{D}^m = \{(r_1, r_2, \ldots, r_m) \in (0,\infty)^m : r_i \neq r_j \text{for some } i \neq j, i, j = 1, 2, \ldots m\}. \tag{43}$$

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

$$(\log r_i, \log C_n(r_i; x_0)), \qquad i = 1, 2, \ldots m. \tag{44}$$

As is seen in the next two results, the choise 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* $\mathbf{r} \in \mathcal{D}^m$ *with* $\max_{1 \leq i \leq m} r_i \leq r_0$*,* $m = 2, 3, \ldots$ *if and only if for some positive constant c*

$$C(r) = cr^{\nu} \qquad \text{if } r \leq r_0. \tag{45}$$

*2.* $\lim_{\lambda \to 0^+} \mathbf{d}(\lambda \mathbf{r}) = 0$ *for all* $\mathbf{r} \in \mathcal{D}^m$ *,* $m = 2, 3, \ldots$ *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*

$$\lim_{m\to\infty}\mathbf{d}\left(\mathbf{r}^{(m)}\right)=0 \tag{46}$$

**Proof:**

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

$$\begin{aligned}\overline{x}_k &= \sum_{i=1}^{k} x_i/k & (47)\\ \Delta_{i,k} &= x_i-\overline{x}_k, & (48)\end{aligned}$$

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

$$\mathbf{d}(\mathbf{r})=\sum_{i=1}^{m} v(r_i)\Delta_{i,m}=0. \tag{49}$$

Note that

$$\Delta_{i,m}=\begin{cases}\Delta_{i,m-1}+[\overline{x}_{m-1}-x_m]/m & \text{if } i=1,2\ldots m-1\\ (m-1)[x_m-\overline{x}_{m-1}]/m & \text{if } i=m.\end{cases} \tag{50}$$

Substitution of equation (50) into equation (49) gives,

$$\begin{aligned}\sum_{i=1}^{m} v(r_i)\Delta_{i,m} &= \sum_{i=1}^{m-1} v(r_i)\Delta_{i,m-1}\\ &+ \left\{\sum_{i=1}^{m-1} v(r_i)[\overline{x}_{m-1}-x_m]+v(r_m)(m-1)[x_m-\overline{x}_{m-1}]\right\}/m\\ &= \left\{\sum_{i=1}^{m-1} v(r_i)[\overline{x}_{m-1}-x_m]+v(r_m)(m-1)[x_m-\overline{x}_{m-1}]\right\}/m\\ &= 0. & (51)\end{aligned}$$

It immediatly follows that

$$v(r_m)=\sum_{i=1}^{m-1} v(r_i)/(m-1). \tag{52}$$

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

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

$$\begin{aligned}
\lim_{\lambda\to 0^+} \mathbf{d}(\lambda\mathbf{r}) &= \lim_{\lambda\to 0^+} \sum_{i=1}^{m} v(\lambda r_i)\Delta_{i,m} \\
&= \sum_{i=1}^{m} \lim_{\lambda\to 0^+} [v(\lambda r_i) - v(\lambda r')]\,\Delta_{i,m} \\
&= \sum_{i=1}^{m} \lim_{\lambda\to 0^+} \log\left(a(\lambda r_i)/a(\lambda r')\right)\Delta_{i,m} \\
&= 0. \qquad (53)
\end{aligned}$$

Next suppose that $\lim_{\lambda\to 0^+} \mathbf{d}(\lambda\mathbf{r}) = 0$ for any $\mathbf{r} \in \mathcal{D}^m$, $m = 2, 3\ldots.$ Then considerations similar to those leading to equation (52) give

$$\lim_{\lambda\to 0^+} \sum_{i=1}^{m-1} \log\left(a(\lambda r_i)/a(\lambda r_m)\right) = 0. \qquad (54)$$

This is equivalent to

$$\lim_{\lambda\to 0^+} \frac{a(\lambda r_1)}{a(\lambda r_m)}\frac{a(\lambda r_2)}{a(\lambda r_m)}\cdots\frac{a(\lambda r_{m-1})}{a(\lambda r_m)} = 1. \qquad (55)$$

This equation is invariant under the interchange of indices, therefore

$$\begin{aligned}
&\lim_{\lambda\to 0^+} \frac{a(\lambda r_1)}{a(\lambda r_{m-1})}\frac{a(\lambda r_2)}{a(\lambda r_{m-1})}\cdots\frac{a(\lambda r_{m-2})}{a(\lambda r_{m-1})}\frac{a(\lambda r_m)}{a(\lambda r_{m-1})} \\
&= \lim_{\lambda\to 0^+} \frac{a(\lambda r_1)}{a(\lambda r_m)}\frac{a(\lambda r_2)}{a(\lambda r_m)}\cdots\frac{a(\lambda r_{m-1})}{a(\lambda r_m)}\left(\frac{a(\lambda r_m)}{a(\lambda r_{m-1})}\right)^m \\
&= \lim_{\lambda\to 0^+} \left(\frac{a(\lambda r_m)}{a(\lambda r_{m-1})}\right)^m \\
&= 1. \qquad (56)
\end{aligned}$$

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*

$$\lim_{n\to\infty} \hat{\nu}_n(\mathbf{r}; x_0) = \mathbf{d}(\mathbf{r}) + \nu, \tag{57}$$

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

**Proof:**
One has,

$$\begin{aligned}\lim_{n\to\infty} \epsilon_n(\mathbf{r}; x_0) &= \lim_{n\to\infty} \sum_{i=1}^{m} (\log C_n(r_i; x_0) - \log C(r_i))(x_i - \overline{x})/S_{xx} \quad (58)\\ &= \sum_{i=1}^{m} \lim_{n\to\infty} (\log C_n(r_i; x_0) - \log C(r_i))(x_i - \overline{x})/S_{xx}\end{aligned}$$

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

$$\lim_{n\to\infty} (\log C_n(r_i; x_0) - \log C(r_i)) = 0, \tag{59}$$

*a.s.* $\mu$. Therefore

$$\begin{aligned}\lim_{n\to\infty} \hat{\nu}_n(\mathbf{r}; x_0) &= \mathbf{d}(\mathbf{r}) + \nu + \lim_{n\to\infty} \epsilon_n(\mathbf{r}; x_0) \quad (60)\\ &= \mathbf{d}(\mathbf{r}) + \nu,\end{aligned}$$

*a.s.* $\mu$. 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,*

$$\begin{aligned}\lim_{m\to\infty}\lim_{n\to\infty} \hat{\nu}_n\left(\mathbf{r}^{(m)}; x_0\right) &= \nu + \lim_{m\to\infty} \mathbf{d}\left(\mathbf{r}^{(m)}\right)\\ &+ \lim_{m\to\infty}\lim_{n\to\infty} \epsilon_n(\mathbf{r}^{(m)}; x_0) \quad (61)\\ &= \nu,\end{aligned}$$

*In Serinko [19], it is shown how the limits in (61) maybe taken simulateously to yield a consistent estimator of $\nu$ without additional assumptions on $C(r)$ beyond the existence of $\nu$. 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 $C_n(r; x_0)$ and $C(r)$ are distribution functions. Therefore it suffices to show that $C_n(r; x_0)$ converges to $C(r)$ *a.s.* $\mu$ for each $r$ and that $C_n(r^-, x_0) = \lim_{\epsilon \to 0^+} C_n(r - \epsilon, x_0)$ converges to $C(r^-) = \lim_{\epsilon \to 0^+} C(r - \epsilon)$ *a.s.* $\mu$ for each $r$. The uniformity will follow from the Glivenko-Cantelli theorem for distribution functions [4, pp. 275-276].

In what follows fix $r$. It will be convienent to write

$$C_n(r; x_0) - C(r) = \int_{S_r} \mu_n^{x_0} \times \mu_n^{x_0}(dz) - \int_{S_r} \mu \times \mu(dz). \tag{62}$$

A vertical section of $S_r$, $S_r^x = \overline{B}_r(x)$, where $\overline{B}_r(x) = \{x' \in \Re^d : \rho(x, x') \leq r\}$ is the closed ball in $\Re^d$ of radius $r$ centered at $x$. Therefore the measureability of $S_r$ with respect to the product $\sigma$-field along with Fubini's theorem [4, p. 240] and the addition and subtraction of terms gives,

$$\begin{aligned} C_n(r; x_0) - C(r) &= \int \mu_n^{x_0}(S_r^x)\, \mu_n^{x_0}(dx) - \int \mu(S_r^x)\, \mu(dx) \\ &= \int \mu_n^{x_0}\left(\overline{B}_r(x)\right) \mu_n^{x_0}(dx) - \int \mu\left(\overline{B}_r(x)\right) \mu(dx) \\ &= \int \left[\mu_n^{x_0}\left(\overline{B}_r(x)\right) - \mu\left(\overline{B}_r(x)\right)\right] \mu_n^{x_0}(dx) \qquad (63) \\ &+ \int \mu\left(\overline{B}_r(x)\right) \mu_n^{x_0}(dx) - \int \mu\left(\overline{B}_r(x)\right) \mu(dx). \qquad (64) \end{aligned}$$

Fubini's theorem also yields $\mu\left(\overline{B}_r(x)\right) \in L^1(\mu)$. Therefore the pointwise ergodic theorem [2, p. 13] implies that the term in (64) goes to zero as $n \to \infty$ *a.s.* $\mu$.

One has for the term in (63),

$$\int \left|\mu_n^{x_0}\left(\overline{B}_r(x)\right) - \mu\left(\overline{B}_r(x)\right)\right| \mu_n^{x_0}(dx) \leq \sup_x \left|\mu_n^{x_0}\left(\overline{B}_r(x)\right) - \mu\left(\overline{B}_r(x)\right)\right|.$$

The closed balls in $\Re^d$ with respect to the 'max' metric take the form

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

where $x = (x_1, x_2, \ldots x_d)$. Krickeberg[14] has proven the uniform convergence in $\Re^d$ of Cartesian products of connected real sets, of which these balls are a sub-family. 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

$$\lim_{n\to\infty} \sup_x \left|\mu_n^{x_0}\left(\overline{B}_r(x)\right) - \mu\left(\overline{B}_r(x)\right)\right| = 0, \tag{65}$$

*a.s.* $\mu$. For completeness a direct proof of (65) is contained in the appendix.

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

**Proof of Theorem 2:**

It suffices to note that the pointwise ergodic theorem [2, p. 13] along with the measureability of $h$ imply that

$$\lim_{n\to\infty} m_n^{\omega_0} h^{-1}\left(\overline{B}'_r(y)\right) = m h^{-1}\left(\overline{B}'_r(y)\right), \tag{66}$$

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

**Proof of Theorem 3:**

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

$$\lim_{n\to\infty} \left[ \int f(z)\, \mu_n^{x_0} \times \mu_n^{x_0}\,(dz) - \int f(z)\, \mu \times \mu\,(dz) \right] = 0 \tag{67}$$

*a.s.* $\mu$. 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 $C\,(r)$.

# 4 Appendix

**Lemma 4** *If $X = \Re^d$ and $\rho$ is the 'max' metric, then ergodicity implies*

$$\lim_{n\to\infty} \sup_{x\in X} \left| \mu_n^{x_0} \left( \overline{B}_r\,(x) \right) - \mu \left( \overline{B}_r\,(x) \right) \right| = 0 \tag{68}$$

*a.s. $\mu$.*

The proof of this theorem uses the following result, the proof of which can be found in Pollard [18, p. 8].

**Theorem 4** *Suppose that for each $\epsilon > 0$ there exists a finite class of functions $\mathcal{G}_\epsilon$ containing lower and upper approximations to each $g \in \mathcal{G}$, for which*

$$g_{\epsilon,L} \le g \le g_{\epsilon,U}, \tag{69}$$

*and*

$$\int_X (g_{\epsilon,U}(x) - g_{\epsilon,L}(x))\, \mu(dx) < \epsilon. \tag{70}$$

*Then ergodicity implies,*

$$\lim_{n\to\infty} \sup_{\mathcal{G}} \left| \int_X g\,(x)\ \mu_n^{x_0}\,(dx) - \int_X g(x)\, \mu\,(dx) \right| = 0 \tag{71}$$

*a.s. $\mu$.*

**Proof of Lemma 4:** The first step is to use the tightness of probability measures defined on the Borel sets of $\Re^d$ to "trim-off" infinity. By tightness, given an $\epsilon > 0$ there exists a compact set $K_0 \subseteq X$ such that

$$\mu(K_0^c) < \frac{\epsilon}{6}. \tag{72}$$

Let

$$K = \{x \in X : \inf_{x' \in K_0} \rho(x, x') \leq r\}. \tag{73}$$

Note that $K$ is closed and bounded and therefore by the Heine-Borel Theorem it too is compact. One has

$$\begin{aligned} &\sup_{x \in X} \left|\mu_n^{x_0}\left(\overline{B}_r(x)\right) - \mu\left(\overline{B}_r(x)\right)\right| \qquad (74) \\ &= \sup_{x \in X} \left|\textstyle\int_X I_{\overline{B}_r(x)}(x')\, \mu_n^{x_0}(dx') - \int_X I_{\overline{B}_r(x)}(x')\, \mu(dx')\right| \\ &\leq \sup_{x \in K} \left|\textstyle\int_X I_{\overline{B}_r(x)}(x')\, \mu_n^{x_0}(dx') - \int_X I_{\overline{B}_r(x)}(x')\, \mu(dx')\right| \\ &+ \sup_{x \in K^c} \left|\textstyle\int_X I_{\overline{B}_r(x)}(x')\, \mu_n^{x_0}(dx') - \int_X I_{\overline{B}_r(x)}(x')\, \mu(dx')\right| \end{aligned}$$

Consider the second term on the righthandside of the inequality. The triangle inequality gives

$$\begin{aligned} \sup_{x \in K^c} \left|\mu_n^{x_0}\left(\overline{B}_r(x)\right) - \mu\left(\overline{B}_r(x)\right)\right| &\leq \sup_{x \in K^c} \left|\mu_n^{x_0} \overline{B}_r(x)\right| \\ &+ \sup_{x \in K^c} \left|\mu\left(\overline{B}_r(x)\right)\right|. \end{aligned} \tag{75}$$

By construction, $\overline{B}_r(x) \subseteq K_0^c$, if $x \in K^c$. Therefore

$$\sup_{x \in K^c} \left|\mu\left(\overline{B}_r(x)\right)\right| \leq \mu(K_0^c) < \frac{\epsilon}{6} \tag{76}$$

and

$$\sup_{x \in K^c} \left|\mu_n^{x_0}\left(\overline{B}_r(x)\right)\right| \leq \mu_n^{x_0}(K_0^c). \tag{77}$$

The pointwise ergodic theorem gives for $n$ sufficiently large,

$$\begin{aligned} \left|\mu_n^{x_0}(K_0^c)\right| &\leq \mu(K_0^c) \\ &+ \left|\mu_n^{x_0}(K_0^c) - \mu(K_0^c)\right| \\ &\leq \frac{\epsilon}{6} + \frac{\epsilon}{6} = \frac{\epsilon}{3}, \end{aligned} \tag{78}$$

$a.s.\ \mu$. These consideration lead to

$$\sup_{x \in K^c} \left| \int_X I_{\overline{B}_r(x)}(x')\, \mu_n^{x_0}(dx') - I_{\overline{B}_r(x)}(x')\, \mu(dx') \right| \leq \frac{\epsilon}{2}. \tag{79}$$

a.s. $\mu$.

The first term on the righthand side of (74) is considered. The problem is to exhibit finite class $\mathcal{G}_{\frac{\epsilon}{2}}$, corresponding to the uncountably infinite class

$$\mathcal{G} = \{g(\cdot) = I_{\overline{B}_r(x)}(\cdot) : x \in K\}, \tag{80}$$

which satisfies the condition of Theorem 4. Because of the product structure of the closed balls,

$$I_{\overline{B}_r(x)}(y) = \prod_{j=1}^{d} I_{\overline{B}_r^{(j)}(x_j)} \circ \pi_j(y) \tag{81}$$

where $\overline{B}_r^{(j)}(x_j) = [x_j - r, x_j + r]$ and $\pi_j$ is the projection onto the $j^{th}$ coordinate, $j = 1, 2, \ldots, d$. Take $m$ an integer satisfying $\frac{1}{m} < \frac{\epsilon}{4d}$. A finite collection of envelope functions $\mathcal{G}_j^{(m)}$satisfying the conditions of Theorem 4 relative to the uncountable collection of functions

$$\mathcal{G}_j = \{g(\cdot) = I_{\overline{B}_r^{(j)}(x_j)} \circ \pi_j(\cdot); x_j \in \pi_j K\}. \tag{82}$$

$j = 1, 2, \ldots, d$ is constructed, such that

$$\mathcal{G}_{\frac{\epsilon}{2}} = \{g = \prod_{j=1}^{d} g_j : g_j \in \mathcal{G}_j^{(m)}\}. \tag{83}$$

For any distribution function $G$ its left-continuous inverse $G^{-1}$ is given by

$$G^{-1}(u) = \inf\{x : G(x) \geq u\} \tag{84}$$

for $0 < u < 1$. Also let

$$G(x^-) = \lim_{\delta \downarrow 0} G(x - \delta). \tag{85}$$

As a consequence of $K$ being closed and bounded, there exist $a_j \leq b_j$ such that $\pi_j K = [a_j, b_j]$, $j = 1, 2, \ldots, d$. Let $\mu_j = \mu \pi_j^{-1}$ and define,

$$\overline{F}_j(x) = \mu_j((-\infty, x + r]), \tag{86}$$

and

$$\underline{F}_j(x) = \mu_j((-\infty, x-r]), \tag{87}$$

$j = 1, 2, \ldots, d$,where $x, r \in \Re$. Further, let

$$\overline{x}_j^{(k,m)} = \overline{F}_j^{-1}\left(\frac{k}{m}\right), \tag{88}$$

and

$$\underline{x}_j^{(k,m)} = \underline{F}_j^{-1}\left(\frac{k}{m}\right), \tag{89}$$

$k = 1, 2, \ldots, m-1,\ j = 1, 2, \ldots, d$. The sequences $\overline{x}_j^{(k,m)}$ and $\underline{x}_j^{(k,m)}$, $k = 1, 2, \ldots, m-1;\ j = 1, 2, \ldots, d;$ are merged into a single sequence. The elements $a_j$ and $b_j + \delta$, for some $\delta > 0$ are added to the merged sequence, $j = 1, 2, \ldots, d$. Any elements less than $a_j$ or greater than $b_j + \delta$ are discarded from the merged sequence, $j = 1, 2, \ldots, d$. If two or more elements of the merged sequence are equal only one is retained. If the spacing between any two nearest neighbors in the remaining elements of the merged sequence is greater or equal than $2r$, intervening elements are added such that the new spacings are less than $2r$. Let $a_j = x_j^{(1,m)} < x_j^{(2,m)} < \ldots < x_j^{(s_j,m)} = b_j + \delta$ denote the new sequence; $j = 1, 2, \ldots, d$; so constructed.

Let

$$J_j^{(k,m)} = (x_j^{(k,m)}, x_j^{(k+1,m)}), \tag{90}$$

and

$$J_j^{(k+s_j-1,m)} = \{x_j^{(k,m)}\} \tag{91}$$

$k = 1, 2, \ldots, s_j - 1,\ j = 1, 2, \ldots, d$. If $x \in J_j^{(k,m)}$, then

$$I_{(-\infty, x_j^{(k,m)}+r]} \leq I_{(-\infty, x+r]} \leq I_{(-\infty, x_j^{(k+1,m)}+r)}, \tag{92}$$

and

$$I_{(-\infty, x_j^{(k,m)}-r]} \leq I_{(-\infty, x-r)} \leq I_{(-\infty, x_j^{(k+1,m)}-r)}, \tag{93}$$

$k = 1, 2, \ldots, s_j - 1,\ j = 1, 2, \ldots, d$. By construction,

$$x_j^{(k,m)} - r < x_j^{(k+1,m)} - r < x_j^{(k,m)} + r < x_j^{(k+1,m)} + r, \tag{94}$$

$k = 1, 2, \ldots, s_j - 1;\ j = 1, 2, \ldots, d$, therefore subtraction of (93) from (92) gives

$$\begin{aligned} g_{j,L}^{(k,m)}(\cdot) &= I_{[x_j^{(k+1,m)}-r, x_j^{(k,m)}+r]}(\cdot) \\ &\leq I_{\overline{B}_r^{(j)}(x)}(\cdot) \\ &\leq I_{(x_j^{(k,m)}-r, x_j^{(k+1,m)}+r)}(\cdot) \\ &= g_{j,U}^{(k,m)}(\cdot), \end{aligned} \tag{95}$$

for $x \in J_j^{(k,m)}$, $k = 1, 2, \ldots, s_j - 1;\ j = 1, 2, \ldots, d$. On the otherhand if $x \in \cup_{k=1}^{s_j-1} J_j^{(k+s_j-1,m)}$,[2] then let

$$g_{j,L}^{(k,m)}(\cdot) = I_{\overline{B}_r^{(j)}(x)}(\cdot) \leq I_{\overline{B}_r^{(j)}(x)}(\cdot) \leq I_{\overline{B}_r^{(j)}(x)}(\cdot) = g_{j,U}^{(k,m)}(\cdot), \tag{96}$$

$j = 1, 2, \ldots, d$.

Let $\mathcal{G}_j^{(m)} = \{g_{j,L}^{(k,m)}(\cdot), g_{j,U}^{(k,m)}(\cdot) :\ k = 1, 2, \ldots, 2s_j - 2\}$, $j = 1, 2, \ldots, d$. It follows from (95) and (96) that $\mathcal{G}_j^{(m)}$ satisfies (69) , since $[a_j, b_j] \subset \cup_{k=1}^{2s_j-2} J_j^{(k,m)}$ implies each $x \in [a_j, b_j]$ is in some $J_j^{(k,m)}$, $k = 1, 2, \ldots, 2s_j - 2;\ j = 1, 2, \ldots, d$.

It remains to show that $\mathcal{G}_j^{(m)}$ satisfies (70). Clearly, it is satisfied for $k = s_j, s_j+1, \ldots, 2s_j-2$; since $g_{j,U}^{(k,m)}(\cdot) - g_{j,L}^{(k,m)}(\cdot) = 0$, if $k = s_j, s_j+1, \ldots, 2s_j-2$. Next, suppose $k = 1, 2, \ldots, s_j - 1$, then

$$\begin{aligned} &\int_{\Re} \left( g_{j,U}^{(k,m)}(x) - g_{j,L}^{(k,m)}(x) \right) \mu_j(dx) = \\ &\int_{\Re} \left( I_{(x_j^{(k,m)}-r, x_j^{(k+1,m)}+r)}(x) - I_{[x_j^{(k+1,m)}-r, x_j^{(k,m)}+r]}(x) \right) \mu_j(dx) = \\ &\mu_j \left( (x_j^{(k,m)} + r, x_j^{(k+1,m)} + r) \right) + \mu_j \left( (x_j^{(k,m)} - r, x_j^{(k+1,m)} - r) \right) = \\ &\left[ \overline{F}_j \left( x_j^{(k+1,m)-} \right) - \overline{F}_j \left( x_j^{(k,m)} \right) \right] + \left[ \underline{F}_j \left( x_j^{(k+1,m)-} \right) - \underline{F}_j \left( x_j^{(k,m)} \right) \right] \end{aligned} \tag{97}$$

$j = 1, 2, \ldots, d$.

---

[2]Hyperplanes parallel to coordinate axes of positve $\mu$ mass cause discontinuities in $\overline{F}_j$ and $\underline{F}_j$; $j = 1, 2, \ldots, d$. The way in which the sequences $\overline{x}_j^{(k,m)}$ and $\underline{x}_j^{(k,m)}$ $k = 1, 2, \ldots, m-1$; $j = 1, 2, \ldots, d$are constructed implies that all discontinuities of $\overline{F}_j$ and $\underline{F}_j$ greater than $\frac{1}{m}$ occur at points of $\cup_{k=1}^{s_j-1} J_j^{(k+s_j-1,m)}$; $j = 1, 2, \ldots, d$. It is for this reason that the function itself is used as its own envelope functions for points in this finite set.

It follows directly from the definition of $\overline{x}_j^{(k,m)}$, that

$$\overline{F}_j\left(\overline{x}_j^{(k,m)-}\right) - \overline{F}_j\left(\overline{x}_j^{(k-1,m)}\right) \leq \frac{1}{m}, \tag{98}$$

$k = 2, 3, \ldots, m-1;\ j = 1, 2, \ldots, d;$

$$\overline{F}_j\left(\overline{x}_j^{(1,m)-}\right) \leq \frac{1}{m}, \tag{99}$$

and

$$1 - \overline{F}_j\left(\overline{x}_j^{(m-1,m)}\right) \leq \frac{1}{m}, \tag{100}$$

The spacing in the sequence $x_j^{(k,m)}$, $k = 1, 2, \ldots, s_j - 1$; is no larger than the spacing between distinct elements of the sequence $\overline{x}_j^{(k,m)}$; $k = 2, 3, \ldots, m - 1$and it contains all distinct elements of $\overline{x}_j^{(k,m)}$; $k = 2, 3, \ldots, m - 1$between $a_j$ and $b_j + \delta$; $j = 1, 2, \ldots, d$. Hence,

$$\overline{F}_j\left(x_j^{(k,m)-}\right) - \overline{F}_j\left(x_j^{(k-1,m)}\right) \leq \frac{1}{m}, \tag{101}$$

$k = 1, 2, \ldots, s_j - 1;\ j = 1, 2, \ldots, d$. Likewise,

$$\underline{F}_j\left(x_j^{(k,m)-}\right) - \underline{F}_j\left(x_j^{(k-1,m)}\right) \leq \frac{1}{m}, \tag{102}$$

$k = 1, 2, \ldots, s_j - 1\, j = 1, 2, \ldots, d$. Inequalities (101) and (102) along with (97) give,

$$\int_{\Re}\left(g_{j,U}^{(k,m)}(x) - g_{j,L}^{(k,m)}(x)\right)\mu_j(dx) \leq \frac{2}{m} \tag{103}$$

Now consider the original problem. Let $\mathbf{k} = (k_1, k_2, \ldots, k_d)$ where $k_j = 1, 2, \ldots, 2s_j - 2$; $j = 1, 2, \ldots, d$. Inequalities (95) and (96) imply that

$$g_L^{(\mathbf{k},m)}(\cdot) = \prod_{j=1}^{d} g_{j,L}^{(k_j,m)} \circ \pi_j(\cdot) \leq I_{\overline{B}_r(x)}(\cdot) \leq \prod_{j=1}^{d} g_{j,U}^{(k_j,m)} \circ \pi_j(\cdot) = g_U^{(\mathbf{k},m)}(\cdot), \tag{104}$$

if $x \in J^{(\mathbf{k},m)} = J_1^{(k_1,m)} \times J_2^{(k_2,m)} \times \ldots \times J_d^{(k_d,m)}$, $k_j = 1, 2, \ldots, 2s_j - 2$; $j = 1, 2, \ldots, d$. $K \subset \cup_{\mathbf{k}} J^{(\mathbf{k},m)}$, therefore (69) is satisfied.

It remains to show that (70) is satisfied. For each $\mathbf{k}$ one has,

$$\prod_{j=1}^{d}\left[g_{j,U}^{(k_j,m)}\circ\pi_j(\cdot)-g_{j,L}^{(k_j,m)}\circ\pi_j(\cdot)\right]\leq\sum_{j=1}^{d}\left(g_{j,U}^{(k_j,m)}\circ\pi_j(\cdot)-g_{j,L}^{(k_j,m)}\circ\pi_j(\cdot)\right). \tag{105}$$

The proof of this is by induction. The above inequality and (103) give

$$\int_{\Re^d}\left(g_U^{(\mathbf{k},m)}(x)-g_L^{(\mathbf{k},m)}(x)\right)\mu(dx)= \tag{106}$$

$$\int_{\Re^d}\prod_{j=1}^{d}\left(g_{j,U}^{(k_j,m)}\circ\pi_j(x)-g_{j,L}^{(k_j,m)}\circ\pi_j(x)\right)\mu(dx)\leq \tag{107}$$

$$\sum_{j=1}^{d}\int_{\Re^d}\left(g_{j,U}^{(k_j,m)}\circ\pi_j(x)-g_{j,L}^{(k_j,m)}\circ\pi_j(x)\right)\mu(dx)=$$

$$\sum_{j=1}^{d}\int_{\Re}\left(g_{j,U}^{(k_j,m)}(x)-g_{j,L}^{(k_j,m)}(x)\right)\mu_j(dx)\leq \tag{108}$$

$$\sum_{j=1}^{d}\frac{2}{m}=\frac{2d}{m}\leq\frac{\epsilon}{2}.$$

This completes the proof.

Finally, it is noted that with minor modification the same proof holds for open balls.

# References


[1] Aaronson, J., Burton, R., Dehling, H., Gilat, D., Hill, T., Weiss, B.(1993). Law of large numbers for U-statistics of dependent observation. *Preprint.*

[2] Billingsley, P.(1965). *Ergodic Theory and Information.* Wiley, New York.

[3] Billingsley, P.(1968).*Convergence of Probability Measure.* Wiley, New York.

[4] Billingsley, P.(1986).*Probability and Measure,*$2^{nd}$ *ed.*Wiley, New York.

[5] Cutler,C.(1991). Some results on the behavior and estimation of fractal dimensions of distributions on attractors. *J. Stat. Phys.* **62,** 651-708.

[6] Cutler,C.(1993). A review of the theory and estimation of fractal dimension. In *Nonlinear Time Series and Chaos, Vol 1: Dimension Estimation and Models,* ed. H. Tong, World Scientific, Singapore, 1-107.

[7] Cutler,C.(1994). A theory of correlation dimension for stationary time series. Technical Report STAT-94-17, Dept. of Statistics, University of Waterloo.

[8] Denker, M and Keller, G. (1986). Rigorous statistical procedures for data from dynamical systems. *J. Stat. Phys.* **44** 67-93.

[9] Eckmann, J.-P.and Ruelle, D.(1985). Ergodic theory of Chaos and strange attractors. *Rev. Mod. Phys.* **57,** 617-656.

[10] Eckmann, J.-P., Kamphorst, S. O., Ruelle, D., Ciliberto, S. (1986). Liapunov exponents from time series. *Phys. Rev.* **A34**, 4971-4979.

[11] Feller, W.(1971). *An Introduction to Probability Theory and Its Applications* **Vol. 2**. Wiley,New York.

[12] Gaenssler, P. B. (1974). Around the Glivenko-Cantelli theorem.*Colloquia Math. Societatis Janos Bolyai***11**, 93-103.

[13] Grassberger, P., Procaccia, I. (1983). Measuring the strangeness of strange attractors. *Physica* **D9**, 189-208.

[14] Krickeberg, K. (1976). An alternative approach to Glivenko-Cantelli theorems. In: *Lecture Notes in Mathematics* **566**, 57-67. Springer-Verlag, New York.

[15] Osborne, A. R. and Provenzale, A. (1989). Finite correlation dimension for stochastic systems with power-law spectra. *Physica* **D35** 357-381.

[16] Parthasarathy, K. R. (1967). *Probability Measures on Metric Spaces.* Academic, New York.

[17] Pesin, Ya. B. (1993). On rigorous mathematical definitions of correlation dimension and generalized spectrum for dimension. *J. Stat. Phys.* **71** 529-547.

[18] Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer-Verlag, New York.

[19] Serinko, R. (1994). A consistent approach to least squares estimation of correlation dimension in weak Bernoulli dynamical systems. *Ann. Appl. Probab.* **4** 1234-1254.

[20] Takens, F.(1984). On the numerical determination of the dimension of an attractor. In: *Lecture Notes in Mathematics* **1125**,99-106. Springer-Verlag, Berlin.

[21] Wells, R., Shire, H. N., Fosmire,C., Doran, J.A.(1993) Improved algorithms for estimating the correlation dimension and the associated probable errors. Penn State Dept. of Math. Tech. Report#**AM114**.

```
From seele Thu Jul 13 10:56:11 1995
To: rege@stat.psu.edu
Subject: Tech Report


Hi Rege,

Hope your summer is going well and you are staying cool!!

I'm working on the tech report you sent Norma and I have
some questions before I go ahead and make copies.

1.  On my page 12-13 it looks like an equation broke bad.
    It's in Section 3 The Proofs.

    One has for the term in (63)

      \int |\mu^{x_0}_n (.....  \leq

    it breaks here and then on the next page to the
    right margin it continues

      \sup_x |\mu^{x_0}_n.....

    Is this okay?

2.  The references..... in reference 19 .... Serinko, R. (1994).
    {\it A consistent approach to least squares estimation of
    correlation dimension in weak Bernoulli dynamical systems.\/}
    Ann. Appl. Probab. {\bf 4} {\it 1234--1254}.  <--- This is
    how it printed up.  The journal is in roman print and the
    other stuff came out in italics and bold as I did it here.

    Then references 20 and 21 are all in italics except for
    what you put in bold.  Is this the way you want it?

In case you would like to know, I assigned #95-34 to the report.

Let me know if all this is the way you want it and then I'll
go ahead and copy it.  Do you want any extra copies made for
yourself?  As usual, 15 copies go in 519.

Thanks and see you in August....

Teena
```

From sys Thu Jul 20 18:48:07 1995
Received: from student.stat.psu.edu (root@student.stat.psu.edu [128.118.181.11]) by a.sta
Received: from bayes.stat.psu.edu. by student.stat.psu.edu (4.1/SMI-4.1/stat.3)
id AA20112; Thu, 20 Jul 95 14:47:55 EDT
Date: Thu, 20 Jul 95 14:47:55 EDT
From: Regis J Serinko <rege@stat.psu.edu>
Message-Id: <9507201847.AA20112@student.stat.psu.edu>
To: seele@stat.purdue.edu
Subject: Re: Tech Report
In-Reply-To: Mail from 'Teena Seele <seele@stat.purdue.edu>'
dated: Thu, 13 Jul 1995 10:56:11 -0500
Status: R

Hi Teena, Sorry I didn't get back to you sooner. I was away in
Montreal for a workshop. Below are the answers to your questions.

> Message 18:
> From seele@stat.purdue.edu Thu Jul 13 11:56:17 1995
> To: rege@stat.psu.edu
> Subject: Tech Report
>
>
> Hi Rege,
>
> Hope your summer is going well and you are staying cool!!
>
> I'm working on the tech report you sent Norma and I have
> some questions before I go ahead and make copies.
>
> 1. On my page 12-13 it looks like an equation broke bad.
> It's in Section 3 The Proofs.
>
> One has for the term in (63)
>
> \int |\mu^{x_0}_n (..... \leq
>
> it breaks here and then on the next page to the
> right margin it continues
>
> \sup_x |\mu^{x_0}_n.....
>
> Is this okay?

No the first line should be pushed to the next page.

>
> 2. The references..... in reference 19 .... Serinko, R. (1994).
> {\it A consistent approach to least squares estimation of
> correlation dimension in weak Bernoulli dynamical systems.\/}
> Ann. Appl. Probab. {\bf 4} {\it 1234--1254}. <--- This is
> how it printed up. The journal is in roman print and the
> other stuff came out in italics and bold as I did it here.
>
> Then references 20 and 21 are all in italics except for
> what you put in bold. Is this the way you want it?
>

No, the italics should only be on the Journal name.

> In case you would like to know, I assigned #95-34 to the report.
>
> Let me know if all this is the way you want it and then I'll
> go ahead and copy it. Do you want any extra copies made for
> yourself? As usual, 15 copies go in 519.

1/2 dozen copies for me should be enough.

>
> Thanks and see you in August....
>
> Teena
>

I'll send you a corrected postscript file in the next e-mail.
Rege