Consistency of the Takens estimator for the correlation dimension

Borovkova, S; Burton, R; Dehling, H

Published in:
Annals of applied probability

Publication date:
1999

The Annals of Applied Probability
1999, Vol. 9, No. 2, 376–390

CONSISTENCY OF THE TAKENS ESTIMATOR FOR THE CORRELATION DIMENSION[1]

BY S. BOROVKOVA, R. BURTON AND H. DEHLING

University of Groningen, Oregon State University and University of Groningen

Motivated by the problem of estimating the fractal dimension of a strange attractor, we prove weak consistency of U-statistics for stationary ergodic and mixing sequences when the kernel function is unbounded, extending by this earlier results of Aaronson, Burton, Dehling, Gilat, Hill and Weiss. We apply the obtained results to show consistency of the Takens estimator for the correlation dimension.

1. Introduction. Estimation of the fractal dimension of a strange attractor from a chaotic time series has attracted considerable attention in the past few years and has become one of the tools in the analysis of the underlying dynamics. Though there are various notions of noninteger dimensions, most attention has been devoted to the *correlation dimension*. This is mainly because this type of dimension is relatively easy to estimate, and it provides a good measure of the complexity of the dynamics, that is, the number of active degrees of freedom.

Suppose $(\mathscr{X}, \mathscr{F}, \mu, T)$ is a dynamical system, where $\mathscr{X} \subseteq \mathbf{R}^p$ and $T: \mathscr{X} \to \mathscr{X}$ is a measurable transformation with invariant probability measure μ. We define the *correlation integral*

$$C_\mu(r) = (\mu \times \mu)\{(x, y): \|x - y\| \leq r\}$$

for $r \geq 0$. In many examples it turns out that there exists a constant α such that

$$C_\mu(r) \approx \text{const} \cdot r^\alpha \quad \text{as } r \to 0. \tag{1.1}$$

Then the exponent α is called the *correlation dimension* of μ. More formally, one defines the correlation dimension by $\alpha := \lim_{r \to 0} \log C(r)/\log r$, provided this limit exists.

In most practical situations, the dynamical system and thus also the invariant measure μ are unknown and one has to rely on (partial) observations of a finite orbit $(T^k\omega)_{0 \leq k \leq n}$ of the system. Most models assume that the actual observations are functions of the state. More precisely, one postulates existence of a so-called read-out function $f: \mathscr{X} \to \mathbf{R}$ such that $y_n = f(T^n\omega)$ is observed at time n. Of course, one cannot hope to get much information about the state

Received October 1996; revised December 1997.

[1]Supported by Netherlands Organization for Scientific Research Grant NLS 61-277, NSF Grant DMS-96-26575 and NATO Collaborative Research Grant CRG 930819.

AMS 1991 *subject classifications*. Primary 60F05, 62M10; secondary 62G05, 60G10.

Key words and phrases. U-statistics, correlation dimension, Takens estimator, ergodic sequences, absolute regularity.

ω by just observing $f(\omega)$. This, however, changes completely if one replaces $f(\omega)$ by the vector of observations $f(T^i\omega)$ at q consecutive time points, that is,

$$\text{Rec}(\omega) := (f(\omega), f(T\omega), \ldots, f(T^{q-1}\omega)).$$

The Takens reconstruction theorem [Takens (1981)] then assures that in generic situations, Rec: $\mathscr{X} \to \mathbf{R}^q$ defines an embedding, provided $q \geq 2p + 1$. Consequently we can obtain information about the state space and the dynamics of T by studying the process of reconstruction vectors

$$X_n = \text{Rec}(T^n\omega), \qquad n \geq 0.$$

Among other things, one can show that the correlation dimension of the invariant measure μ coincides with that of the marginal distribution F of X_n, again provided we are in the generic situation and $q \geq 2p + 1$. For smaller values of q, the correlation dimension of F will equal the embedding dimension.

It is thus of interest to estimate the correlation dimension of the marginal distribution of a stationary stochastic process from a finite sample $X_1, \ldots, X_n$. Note that the correlation integral can be written as

$$C(r) = C_F(r) = P(\|X - X'\| \leq r),$$

where X and X' are independent copies of X_1.

A number of procedures for estimating the correlation dimension has been introduced in the literature. Here we concentrate our attention on the approach proposed by Takens (1985). Assume for a moment that in a neighborhood of $r = 0$ an exact scaling law holds for the correlation integral, that is,

$$C(r) = \text{const} \cdot r^\alpha, \qquad r \leq r_0 \tag{1.2}$$

for some $r_0 > 0$. Then Takens first considered estimating α from i.i.d. realizations $R_i = \|X_i - Y_i\|$ of the distance $\|X - Y\|$, where X_i and Y_i are independent with distribution μ. If (1.2) holds, the conditional distribution of $U_i := R_i/r_0$ given $R_i \leq r_0$ is given by

$$\mathbf{P}(U_i \leq t | U_i \leq 1) = t^\alpha \quad \text{for } t \in [0, 1].$$

Then the (conditional) distribution of $S_i = -\log U_i$ is exponential with parameter α, that is, S_i has density

$$g(s) = \alpha e^{-\alpha s} 1_{[0, \infty)}(s).$$

Given an i.i.d. sample $S_1, S_2, \ldots, S_N$ of exponentially distributed random variables, the maximum likelihood estimator (MLE) of the parameter α is given by

$$\hat{\alpha}_{\text{ML}} = \frac{N}{\sum_{i=1}^N S_i}. \tag{1.3}$$

It turns out that the ML estimator is biased but that the variant, $\hat{\alpha} = (N - 1)/\sum_{i=1}^N S_i$, is unbiased and actually the uniformly minimum variance unbiased (UMVU) estimator. This is a simple consequence of the Lehman–Scheffé lemma and the fact that $\sum_{i=1}^N S_i$ is a complete, sufficient statistic.

In general, independent realizations of the distances $\|X - Y\|$ will not be available and thus a modification of the estimator (1.3) becomes necessary. Given a finite segment $X_1, \ldots, X_n$ of an orbit, we can form $n(n-1)/2$ pairwise distances $\|X_i - X_j\|$. Motivated by the ML estimator (1.3), Takens proposed to use

$$\hat{\alpha}_T = -\left(\frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \log \frac{\|X_i - X_j\|}{r_0}\right)^{-1} \tag{1.4}$$

as estimator for the correlation dimension.

A completely different approach was suggested by Grassberger and Procaccia (1983), actually in the same paper where they introduced the notion of correlation dimension. The Grassberger–Procaccia procedure is motivated by the approximately linear relationship $\log C(r) \approx \beta + \alpha \log r$, for r small, obtained by taking logarithms on both sides of (1.1). Estimating now the correlation integral by its empirical analogue,

$$C_n(r) = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \mathbf{1}\{\|X_i - X_j\| \le r\}, \tag{1.5}$$

for a vector of distances $(r_1, \ldots, r_k)$, Grassberger and Procaccia use least-squares linear regression of $\log C_n(r_i)$ versus $\log r_i$ to estimate α.

In the analysis of both the Grassberger–Procaccia and the Takens estimator, the theory of U-statistics plays a central role. Let $h\colon (\mathbf{R}^q)^m \to \mathbf{R}$ be a measurable function that is symmetric in its arguments. Given a sample $X_1, \ldots, X_n$ from a distribution function F, define the associated U-statistic of degree m by

$$U_n = \binom{n}{m}^{-1} \sum_{1 \le i_1 < \cdots < i_m \le n} h(X_{i_1}, \ldots, X_{i_m}).$$

Note that $U_n(h)$ can be viewed as the empirical analogue of the parameter

$$\theta(F) = \int_{\mathbf{R}^{qm}} h(x_1, \ldots, x_m)\, dF(x_1) \cdots dF(x_m).$$

It turns out that $U_n(h)$ is an unbiased estimator of $\theta(F)$ and, in the case of an i.i.d. sample, even the UMVU estimator, provided the distribution is completely unknown.

In order to establish consistency of the Takens estimator, we have to study the U-statistic

$$\frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \log \|X_i - X_j\|. \tag{1.6}$$

For the Grassberger–Procaccia estimator, consistency of the U-statistic (1.5) has to be investigated.

Halmos (1946) and Hoeffding (1948) independently introduced U-statistics. In the case of i.i.d. observations $(X_n)_{n \ge 1}$, their asymptotic behavior is well understood. The law of large numbers and the central limit theorem were already established by Hoeffding (1948). Mainly motivated by the above-mentioned

applications to dimension estimation, there has been in recent years considerable interest in U-statistics of stationary dependent observations. Early contributions were Yoshihara's (1976) central limit theorem for absolutely regular processes and Denker and Keller's (1983, 1986) CLT for certain Lipschitz functionals of an absolutely regular process.

Surprisingly, the law of large numbers for U-statistics of stationary ergodic sequences has not been investigated until the recent paper by Aaronson, Burton, Dehling, Gilat, Hill and Weiss (1996). They first showed, by means of a simple counterexample, that the law of large numbers might fail unless extra conditions on either the kernel or the process (X_n) are imposed. One of the main results of Aaronson et al., which we state below, implies strong consistency of $U_n(h)$ for bounded continuous kernels.

THEOREM A [Aaronson et al. (1996)]. *Let $(X_n)_{n\in\mathbf{N}}$ be a stationary ergodic sequence with marginal distribution F, and let h: $\mathbf{R}^{qm} \to \mathbf{R}$ be a measurable, bounded and F^m-a.e. continuous function. Then*

$$U_n \overset{a.s.}{\to} \theta(F) \quad \text{as } n \to \infty.$$

Stronger results can be obtained if the underlying process $(X_n)_n$ satisfies some mixing conditions. In the context of U-statistics laws of large numbers, *absolute regularity* turns out to be an important concept. Denote by $\mathscr{M}_a^b$ the σ-field generated by $\{X_i,\ a \le i < b\}$, where $0 \le a \le b \le \infty$. We define the mixing coefficients $(\beta_k)_{k\ge 0}$ by

$$\beta_k = \sup_a \mathbf{E}\Big\{ \sup_{A\in\mathscr{M}_{a+k}^{\infty}} |\mathbf{P}(A|\mathscr{M}_1^a) - \mathbf{P}(A)| \Big\}.$$

The process $(X_n)_n$ is called *absolutely regular* if $\beta_k \to 0$ as $k \to \infty$.

For absolutely regular sequences, Aaronson et al. established a result similar to Theorem A under milder conditions on the kernel function. Along a different line, Yoshihara (1976) could show that U-statistics of absolutely regular sequences are asymptotically normal, provided certain extra assumptions on the moments of h and on the rates (β_k) hold.

For stationary and ergodic sequences strong consistency of the estimator (1.5) of the correlation integral follows from Theorem A if $(F \times F)\{(x, y): \|x - y\| = r\} = 0$, that is, if r is a continuity point of the correlation integral $C(r)$. If $C(r)$ is continuous, convergence of $C_n(r)$ to $C(r)$ is even uniform in $r \in [0, r_0]$, as can be shown using monotonicity properties of $C_n(r)$ and $C(r)$. However, none of the theorems of Aaronson et al. (1996) is directly applicable to obtain consistency of the Takens estimator, as $\log|x - y|$ is an unbounded kernel. In the next section we will provide several counterexamples showing that the Takens estimator can indeed be inconsistent. General consistency results will be obtained in Section 3 under extra moment assumptions.

2. Counterexamples. A first simple counterexample to consistency of the Takens estimator was already provided by Aaronson et al. (1996). We include it here for completeness and reference.

EXAMPLE 1. Let $W_1, W_2, \ldots$ be i.i.d. random variables with a continuous distribution function F, satisfying

$$\mathbf{E}\big(\big|\log|W_1 - W_2|\,\big|\big) < \infty. \tag{2.1}$$

Let $Y_1, Y_2, \ldots$ be i.i.d. Bernoulli random variables, independent of $(W_n)_{n\geq 1}$, with $\mathbf{P}(Y_i = 1) = p$ $(0 < p < 1)$. Define the process $(X_n)_{n\in\mathbf{N}}$ by $X_1 = W_1$ and

$$X_n = W_n(1 - Y_n) + X_{n-1}Y_n \quad \text{for } n > 1.$$

This process is stationary and absolutely regular with marginal distribution F. Observe that $X_n = X_{n-1}$ whenever $Y_n = 1$. As the latter occurs infinitely often, the U-statistics (1.6) diverges to $-\infty$, almost surely, showing that the Takens estimator (1.4) is not consistent.

It is instructive to analyze what went wrong in this example. The main problem lies in the drastic difference between the product distribution $F \times F$ and the two-dimensional joint distributions $\mathbf{P}_{ij}$ induced by the pairs (X_i, X_j). For any pair (i, j) there is positive probability that $X_i = X_j$ and hence

$$\mathbf{E}_{\mathbf{P}_{ij}}\big(\big|\log|X_i - X_j|\,\big|\big) = \infty. \tag{2.2}$$

This is in contrast to $\mathbf{E}_{F\times F}(|\log|X_i - X_j|\,|) < \infty$, which is a consequence of (2.1) and the fact that $(X_n)_{n\geq 1}$ has marginal distribution F.

The previous example is quite crude in the sense that $|\log|X_i - X_j|| = \infty$ with positive probability, and thus $E|\log|X_i - X_j|| = \infty$ for all pairs (i, j). In our next example we consider a stationary sequence $(X_n)_{n\in\mathbf{Z}}$ for which pairs (X_i, X_j) have a bounded density with respect to Lebesgue measure and $\mathbf{E}|\log|X_i - X_j|| < \infty$, but

$$\lim_{n\to\infty} \mathbf{E}\big|\log|X_n - X_1|\,\big| = \infty$$

and for which the Takens estimator is inconsistent.

EXAMPLE 2. Let $(\tilde{X}_n)_{n\in\mathbf{Z}}$, $(Y_n)_{n\in\mathbf{Z}}$ and $(Z_n)_{n\in\mathbf{Z}}$ be three mutually independent i.i.d. processes, with $\tilde{X}_n$ and Z_n uniformly distributed on $[0, 1]$ and Y_n symmetric Bernoulli variables, that is, $\mathbf{P}[Y_n = 0] = \mathbf{P}[Y_n = 1] = 1/2$. Let $\phi(k) := \exp(2^{2k})$ and define a new process $(X_n)_{n\geq 1}$ by

$$X_n = \begin{cases} \dfrac{1}{2}\tilde{X}_{n-k} + \dfrac{Z_n}{\phi(k)}, & \text{if } Y_n = 1,\ Y_{n-1} = \cdots = Y_{n-k} = 0,\ Y_{n-k-1} = 1, \\ \dfrac{1}{2}\tilde{X}_n, & \text{if } Y_n = 0. \end{cases}$$

By stationarity of the processes $(\tilde{X}_n)$, (Z_n) and (Y_n), also $(X_n)_{n\geq 1}$ is stationary. In addition, one can show quite easily that $(X_n)_{n\geq 1}$ is absolutely regular. Moreover, $|X_i - X_j|$ has a bounded density and hence $E|\log|X_i - X_j|| < \infty$ for all pairs (i, j).

Define the event $A_r := \{Y_r = 1,\ Y_{r-1} = \cdots = Y_0 = 0,\ Y_{-1} = 1\}$, and note that $P(A_r) = 2^{-(r+2)}$. On A_r we have $X_0 = \frac{1}{2}\tilde{X}_0$ and $X_r = \frac{1}{2}\tilde{X}_0 + Z_r/\phi(r)$ and thus

$$\begin{aligned}\mathbf{E}\big(\big|\log|X_0 - X_r|\big|\big) &\geq \mathbf{E}\big(\big|\log|X_0 - X_r|\big|\,|A_r\big)2^{-(r+2)}\\ &= 2^{-(r+2)}\big[\mathbf{E}(-\log Z_r) + \log\phi(r)\big]\\ &\geq 2^{-(r+2)}2^{2r} = 2^{r-2}.\end{aligned}$$

Hence $\sup_{i,j} E|\log|X_i - X_j|| = \infty$. This result already suggests that the corresponding U-statistic (1.6) might be divergent, and this is indeed the case as we shall show now. We will make use of the following lemma.

LEMMA 1. *Suppose that $(Y_n)_{n\geq 1}$ is an i.i.d. sequence of nonnegative random variables with $\mathbf{E}[Y_1]^{1/2} = \infty$. Then*

$$\limsup_{n\to\infty} \frac{1}{n^2}\sum_{i=1}^{n} Y_i = \infty \quad \textit{almost surely}.$$

PROOF. Since all Y_i's are positive,

$$\limsup_{n\to\infty} \frac{1}{n^2}\sum_{i=1}^{n} Y_i \geq \limsup_{n\to\infty} \frac{Y_n}{n^2},$$

and the fact that the r.h.s. of this inequality is infinite is a consequence of the following line of equivalent statements, valid for all $\varepsilon > 0$:

$$\mathbf{E}[Y_1]^{1/2} = \infty \iff \sum_{n=1}^{\infty} \mathbf{P}(Y_n > \varepsilon n^2) = \infty \iff \mathbf{P}\{Y_n > \varepsilon n^2 \text{ i.o.}\} = 1,$$

which follow from the Borel–Cantelli lemma. □

Note that

$$\begin{aligned}&-\frac{2}{n(n-1)}\sum_{1\leq i<j\leq n}\log|X_i - X_j|\\ &\quad\geq \frac{2}{n(n-1)}\Bigg[\sum_{X_i=\frac{1}{2}\tilde{X}_i,\,X_j=\frac{1}{2}\tilde{X}_j} -\log|X_i - X_j|\\ &\qquad\qquad + \sum_{i,\,j:\,X_i=\frac{1}{2}\tilde{X}_i,\,X_j=\frac{1}{2}\tilde{X}_i+Z_j/\phi(j-i)} -\log|X_i - X_j|\Bigg]\\ &\quad= \frac{2}{n(n-1)}\Bigg[\sum_{X_i=\frac{1}{2}\tilde{X}_i,\,X_j=\frac{1}{2}\tilde{X}_j} -\log|X_i - X_j|\\ &\qquad\qquad + \sum_{i,\,j:\,X_i=\frac{1}{2}\tilde{X}_i,\,X_j=\frac{1}{2}\tilde{X}_i+Z_j/\phi(j-i)} -\log Z_j\\ &\qquad\qquad\qquad + \sum_{m=1}^{M}\log\phi(R_m)\Bigg],\end{aligned} \tag{2.3}$$

where $R_1, \ldots, R_M$ are lengths of full zero-blocks of Y_i contained in the sample of size n. An application of the ergodic theorem yields that $M/n \to \frac{1}{4}$ as $n \to \infty$. The last term in (2.3) is divergent a.s. according to Lemma 1, because

$$\frac{1}{n^2} \sum_{m=1}^{M} \log \phi(R_m) = \left(\frac{M}{n}\right)^2 \frac{1}{M^2} \sum_{m=1}^{M} \log \phi(R_m)$$

and

$$\mathbf{E}[\log \phi(R_m)]^{1/2} = \sum_r [\log \phi(r)]^{1/2} 2^{-r} = \sum_r 1 = \infty.$$

So the Takens estimator is not consistent for this example as well.

In the last section we shall give a numerical example which illustrates the divergence of the Takens estimator in the case of infinite expectation in (2.2), while the Grassberger–Procaccia approach gives a reasonable estimate of the correlation dimension. However, in general, the Takens estimator has advantages, such as computational efficiency, over the Grassberger–Procaccia method. It also turns out that if one imposes some additional conditions on the expectations in (2.2), this leads to the weak consistency of the Takens estimator in the case of absolutely regular and stationary ergodic processes. This is the consequence of more general results on the weak consistency of U-statistics, which we present in the next section.

3. Weak consistency of *U*-statistics. This section contains the main theoretical results of the present paper. We prove two consistency results for U-statistics of stationary ergodic, respectively, absolutely regular sequences. Compared with the results of Aaronson et al. (1996), we replace their condition that the kernel $h(x, y)$ be bounded by a uniform integrability requirement on $h(X_i, X_j)$, $i, j \geq 1$. For simplicity we formulate and prove our theorems here only for U-statistics of degree $m = 2$ and with one-dimensional inputs X_i, that is, $q = 1$. Nonetheless the results continue to hold for general m and q.

THEOREM 1. *Let $(X_n)_{n\geq 1}$ be a stationary ergodic process with marginal distribution F, and let $h: \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ be measurable and $(F \times F)$ - a.e. continuous. Suppose moreover that the family of random variables $\{h(X_i, X_j): i, j \geq 1\}$ is uniformly integrable. Then, as $n \to \infty$,*

$$U_n \to \theta(F) \quad \textit{in probability}. \tag{3.1}$$

In particular this holds, if $\sup_{i,j} E|h(X_i, X_j)|^{1+\delta} < \infty$ for some $\delta > 0$.

PROOF. A well-known result in ergodic theory states that, given a stationary ergodic process $(X_n)_{n\geq 1}$ with marginal distribution F, one has for all measurable sets A, B that

$$\frac{1}{n} \sum_{k=1}^{n} P(X_1 \in A,\ X_k \in B) \to F(A)F(B)$$

as $n \to \infty$. Denoting by μ_k the joint distribution of (X_1, X_k), this implies that $(1/n)\sum_{k=1}^{n} \mu_k$ converges weakly to the product measure $F \times F$.

We now define the truncated kernel $h_K(x, y) = h(x, y)\mathbf{1}_{\{|h(x,y)|\le K\}}$, where K is such that $(F \times F)\{(x, y)\colon |h(x, y)| = K\} = 0$. As $h_K(x, y)$ is bounded and $F \times F$-a.e. continuous, we get $\int |h_K(x, y)|\, d((1/n)\sum_{k=1}^{n} \mu_k)(x, y) \to \int |h_K(x, y)|\, dF(x)\, dF(y)$ and thus

$$\begin{aligned}\int |h_K(x, y)|\, dF(x)\, dF(y) &= \lim_{n\to\infty} \int |h_K(x, y)|\, d\left(\frac{1}{n}\sum_{k=1}^{n} \mu_k\right)(x, y)\\ &= \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} E|h_K(X_1, X_k)|\\ &\le \limsup_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} E|h(X_1, X_k)|\\ &\le \sup_k E|h(X_1, X_k)|.\end{aligned}$$

By uniform integrability of $\{h(X_i, X_j)\colon i, j \ge 1\}$, the right-hand side is finite. Hence we may conclude that $\int |h(x, y)|\, dF(x)\, dF(y) < \infty$, that is h is $F \times F$-integrable.

Moreover, $h_K(x, y)$ satisfies all the conditions of Theorem A, and hence

$$\begin{aligned}&\frac{1}{n(n-1)} \sum_{1\le i\ne j\le n} h_K(X_i, X_j)\\ &\qquad \to \int\int h_K(x, y)\, dF(x)\, dF(y) \quad \text{a.s. as } n \to \infty.\end{aligned} \tag{3.2}$$

By $F \times F$-integrability of h we obtain

$$\left|\int\int h(x, y)\, dF(x)\, dF(y) - \int\int h_K(x, y)\, dF(x)\, dF(y)\right| \to 0 \tag{3.3}$$

as $K \to \infty$. Uniform integrability of $\{h(X_i, X_j), i, j \ge 1\}$ implies

$$\sup_{i,j} \mathbf{E}|h_K(X_i, X_j) - h(X_i, X_j)| = \sup_{i,j} \mathbf{E}|h(X_i, X_j)|1_{\{|h(X_i, X_j)|>K\}} \to 0$$

as $K \to \infty$. This implies that

$$\mathbf{E}\left|\frac{1}{n(n-1)} \sum_{1\le i\ne j\le n} h_K(X_i, X_j) - \frac{1}{n(n-1)} \sum_{1\le i\ne j\le n} h(X_i, X_j)\right| \to 0, \tag{3.4}$$

as $K \to \infty$. Combining now (3.2), (3.3) and (3.4), the statement of the theorem follows. □

In case of an absolutely regular process, we can drop the continuity condition on the kernel, as the next theorem shows. Absolute regularity of a process implies that the sequence of long segments of this process, separated by short ones, can be perfectly coupled with another sequence of long segments, which

are independent and have the same distribution as those of the original process. This is stated precisely in the following result of Philipp (1986).

LEMMA 2 [Theorem 3.4 in Philipp (1986)]. *If $(X_n)_{n\in\mathbf{N}}$ is stationary and absolutely regular with mixing coefficients β_k, then for every $m, N > 0$ there exists an i.i.d. sequence of N-dimensional random vectors $\xi_1', \xi_2', \ldots$, such that for all $k = 1, 2, \ldots$,*

$$\mathbf{P}(\xi_k = \xi_k') = 1 - \beta_m, \tag{3.5}$$

where $\xi_k = (X_{(k-1)(N+m)+1}, \ldots, X_{kN+(k-1)m})$, and the vectors ξ_k and ξ_k' have the same marginal distributions.

THEOREM 2. *Let $(X_n)_{n\in\mathbf{N}}$ be a stationary and absolutely regular process with marginal distribution F, and let h: $\mathbf{R}^2 \to \mathbf{R}$ be measurable. Suppose moreover that the family of random variables $\{h(X_i, X_j)\colon i, j \geq 1,\ i \neq j\}$ is uniformly integrable. Then, as $n \to \infty$,*

$$U_n \to \theta(F) \quad \text{in } L_1,$$

and hence also in probability.

PROOF. Let $\varepsilon > 0$ be given. By uniform integrability of $\{h(X_i, X_j)\colon i, j \geq 1, i \neq j\}$ there exists a $\delta > 0$ such that

$$E|h(X_i, X_j)|1_B \leq \varepsilon \tag{3.6}$$

holds for all measurable sets B with $P(B) < \delta$. The same holds if (X_i, X_j) is replaced by an independent pair (X_i', X_j'). To see this, note first that absolute regularity of the process $(X_i)_{i\geq 1}$ implies that the joint distribution of (X_1, X_n) converges in total variation norm to the product measure $F \times F$. Thus for the truncated kernels $h_K(x, y) = h(x, y)1_{\{|h(x,y)|\leq K\}}$ we get

$$\begin{aligned}\int\int |h_K(x, y)|\, dF(x)\, dF(y) &= \lim_{n\to\infty} E|h_K(X_1, X_n)| \\ &\leq \sup_n E|h(X_1, X_n)| < \infty.\end{aligned}$$

Letting $K \to \infty$, we find that h is $F \times F$-integrable, and hence we can find a $\delta > 0$ such that

$$E|h(X_i', X_j')|1_B \leq \varepsilon \tag{3.7}$$

holds for all measurable sets B with $P(B) < \delta$.

Then choose m, N so big that $2\beta_m < \delta$ and $m/N < \varepsilon$. Define integers $n_k = (k-1)(m+N)$ and consider the blocks

$$\xi_k = (X_{n_k+1}, \ldots, X_{n_k+N}).$$

Observe that given the sample size n, the index of the last block ξ_k fully contained in $(X_1, \ldots, X_n)$ is $p = [n/(N+m)]$. By Lemma 2 there exists

a sequence of independent N-dimensional vectors $\xi'_1, \xi'_2, \ldots$ with the same marginal distribution as (ξ_k) such that (3.5) holds.

In the rest of the proof we will show that the random variables in the small separating blocks of length m can be neglected and that the error introduced by replacing ξ_i by ξ'_i is negligible. The main term will then be a U-statistic with independent vector valued inputs (ξ'_i) that can be treated by Hoeffding's classical U-statistic law of large numbers. To this end we define a new kernel $H\colon \mathbf{R}^N \times \mathbf{R}^N \to \mathbf{R}$ by

$$H(\xi, \eta) := \frac{1}{N^2} \sum_{1 \le i,\, j \le N} h(x_i, y_j),$$

where $\xi = (x_1, \ldots, x_N)$ and $\eta = (y_1, \ldots, y_N)$. From (3.6) we can infer that, for $k \neq l$,

$$E|H(\xi_k, \xi_l)|1_B \le \varepsilon \tag{3.8}$$

for all sets B with $P(B) < \delta$. The same holds if $(\xi_k)_k$ is replaced by (ξ'_k), by (3.7).

Independence of ξ'_k and ξ'_l implies that

$$EH(\xi'_k, \xi'_l) = \int\int h(x, y)\, dF(x)\, dF(y) = \theta(F) \quad \text{for all } k \neq l.$$

Thus by the U-statistics law of large numbers for independent observations,

$$\frac{1}{p(p-1)} \sum_{1 \le k \neq l \le p} H(\xi'_k, \xi'_l) \to \theta(F) \tag{3.9}$$

almost surely and in L_1. By construction of (ξ'_k), we have $P(\xi_k \neq \xi'_k \text{ or } \xi_l \neq \xi'_l) \le 2\beta_m < \delta$ and thus by (3.8),

$$E|H(\xi_k, \xi_l) - H(\xi'_k, \xi'_l)| = E|H(\xi_k, \xi_l) - H(\xi'_k, \xi'_l)|1_{\{\xi_k \neq \xi'_k \text{ or } \xi_l \neq \xi'_l\}} \le 2\varepsilon.$$

Hence,

$$E\left|\frac{1}{p(p-1)} \sum_{1 \le k \neq l \le p} H(\xi_k, \xi_l) - \frac{1}{p(p-1)} \sum_{1 \le k \neq l \le p} H(\xi'_k, \xi'_l)\right| \le 2\varepsilon. \tag{3.10}$$

Moreover,

$$\left|\frac{1}{p(p-1)} - \frac{N^2}{n(n-1)}\right| \le \frac{2\varepsilon}{p(p-1)}$$

for p large enough, and thus

$$E\left|\frac{1}{p(p-1)} \sum_{1 \le k \neq l \le p} H(\xi_k, \xi_l) - \frac{N^2}{n(n-1)} \sum_{1 \le k \neq l \le p} H(\xi_k, \xi_l)\right| \le 2C_0\varepsilon,$$

where $C_0 = \sup_{i,\, j} E|h(\xi_i, \xi_j)|$. This last estimate, together with (3.9) and (3.10), shows that for n large enough,

$$E\left|\frac{N^2}{n(n-1)} \sum_{1 \le k \neq l \le p} H(\xi_k, \xi_l) - \theta(F)\right| \le C\varepsilon. \tag{3.11}$$

Now, decompose the original U-statistics as follows:

$$\begin{aligned}\sum_{1\le i\ne j\le n} h(X_i, X_j) = &\sum_{1\le k\ne l\le p} \sum_{i=n_k+1}^{n_k+N} \sum_{j=n_l+1}^{n_l+N} h(X_i, X_j)\\ &+ \sum_{k=1}^{p} \sum_{n_k+1\le i\ne j\le n_k+N} h(X_i, X_j)\\ &+ 2\sum_{1\le k,\, l\le p} \sum_{i=n_k+N+1}^{n_{k+1}} \sum_{j=n_l+1}^{n_l+N} h(X_i, X_j)\\ &+ \sum_{1\le k\ne l\le p} \sum_{i=n_k+N+1}^{n_{k+1}} \sum_{j=n_l+N+1}^{n_{l+1}} h(X_i, X_j)\\ &+ \sum_{k=1}^{p} \sum_{n_k+N+1\le i\ne j\le n_{k+1}} h(X_i, X_j)\\ &+ \sum_{i=n_p+N+1}^{n} \sum_{j=1}^{n} h(X_i, X_j)\\ &+ \sum_{i=1}^{n_p+N} \sum_{j=n_p+N+1}^{n} h(X_i, X_j).\end{aligned}$$

A careful study of the index set shows that

$$\begin{aligned}E\Bigg| \sum_{1\le i\ne j\le n} h(X_i, X_j) - \sum_{1\le k\ne l\le p} \sum_{i=n_k+1}^{n_k+N} \sum_{j=n_l+1}^{n_l+N} h(X_i, X_j)\Bigg|\\ \le C_0(pN^2 + 2p^2mN + p^2mN + p^2m^2 + 2n(m+N)),\end{aligned}$$

where $C_0 = \sup_{i,\,j} E|h(X_i, X_j)|$. As $p \le n/N$ and $m \le \varepsilon N$, the r.h.s. of the above inequality is bounded by $C(\varepsilon + N/n)n^2$ and hence,

$$E\left|\frac{1}{n(n-1)} \sum_{1\le i\ne j\le n} h(X_i, X_j) - \frac{1}{n(n-1)} \sum_{1\le k\ne l\le p} N^2 H(\xi_k, \xi_l)\right| \le C\varepsilon$$

for n large enough. This, together with (3.11), proves the theorem. □

4. Application to the Takens estimator. By Theorem 2, consistency of the Takens estimator for an absolutely regular sequence $X_1, X_2, \ldots$ of q-dimensional random vectors follows if, for some $\delta > 0$,

$$\sup_{i,\,j} \mathbf{E}\big|\log \|X_i - X_j\|\big|^{1+\delta} < \infty. \tag{4.1}$$

In this section we will briefly indicate some conditions on the distribution of the process (X_n) that imply (4.1).

Observe first that for any bounded domain $A \subset \mathbf{R}^q$ we have, for all $\delta > 0$,

$$\int_A \int_A \left|\log \|x - y\|\right|^{1+\delta} dx\, dy < \infty.$$

Now suppose that X_i takes values in A and that the joint density $f_{ij}(x, y)$ of (X_i, X_j) is bounded, say by C. Then

$$\begin{aligned}\mathbf{E}\left|\log \|X_i - X_j\|\right|^{1+\delta} &= \int\int \left|\log \|x - y\|\right|^{1+\delta} f_{ij}(x, y)\, dx\, dy \\ &\le C \int_A \int_A \left|\log \|x - y\|\right|^{1+\delta} dx\, dy < \infty\end{aligned}$$

and the condition (4.1) is satisfied.

On the other hand, if the distribution of the distances $R_{i,j} = \|X_i - X_j\|$ has a density $p_{ij}(x)$, then the expectation in (4.1) can also be expressed as

$$\mathbf{E}\left|\log \|X_i - X_j\|\right| = \int_0^{r_0} |\log r|^{1+\delta} p_{ij}(r)\, dr.$$

This integral is bounded for all $\delta > 0$, for example, if $\sup_{i,j} p_{ij}(x) = O(x^{-\alpha})$ as $x \to 0$ for some $\alpha < 1$. Thus in this case the condition (4.1) is fulfilled as well.

5. Numerical example. In this section we apply both the Takens estimator and the Grassberger–Procaccia method to the stationary ergodic process $\{X_n\}_{n\in\mathbf{N}}$, defined by

$$X_{n+1} = \begin{cases} (X_n + \exp(-1/Y_{n+1})) \bmod 1, & \text{if } Y_{n+1} < 1/2, \\ \tilde{X}_n, & \text{otherwise,} \end{cases}$$

where $\{Y_n\}_{n\in\mathbf{N}}$, $\{\tilde{X}_n\}_{n\in\mathbf{N}}$ are i.i.d. sequences of uniformly $[0, 1]$-distributed random variables and X_0 is uniform $[0, 1]$.

This process is absolutely regular and has the Lebesgue measure as its marginal distribution, so the correlation dimension in this case is $\alpha = 1$.

We generated a sample of the size 1000 of this process. In Figure 1 the delay map X_{n+1} versus X_n is shown.

Note that for this process,

$$\mathbf{E}\left|\log |X_n - X_{n+1}|\right| \ge \frac{1}{2} \int_0^{1/2} \frac{dy}{y} = \infty,$$

and, according to the results above, we expect the Takens estimator to diverge. And, indeed, computing $\hat{\alpha}^T$ as in (1.4) gives us extremely low values of the estimate, such as

$$\hat{\alpha}^T = 8 \cdot 10^{-3},$$

that is, the reciprocal of $\hat{\alpha}^T$ indeed diverges due to the pairs (X_n, X_{n+1}) which are close to the diagonal.

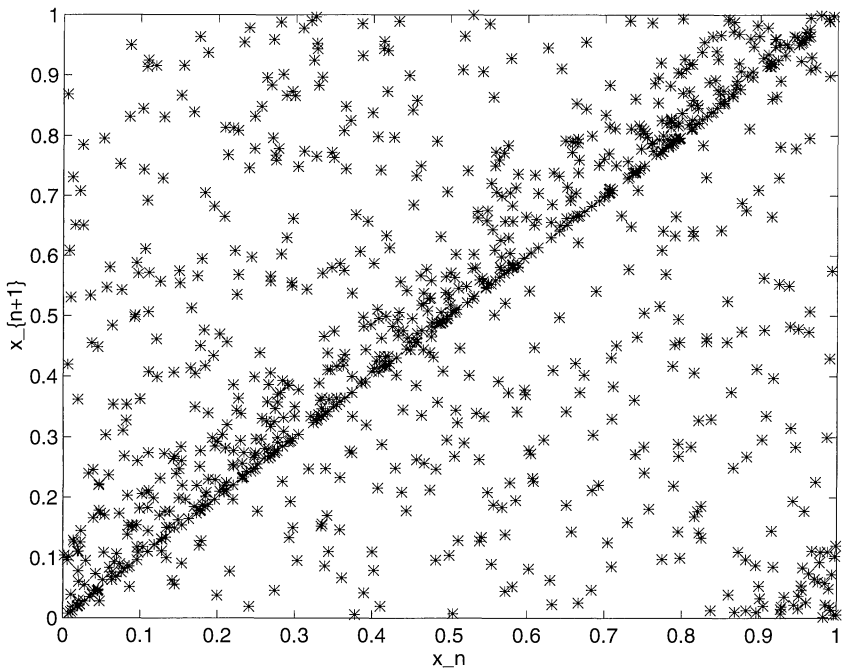

FIG. 1. *Delay map x_{n+1} versus x_n.*

On the other hand, this is no danger for the Grassberger–Procaccia estimator (1.5). In Figure 2 we plotted $\log C_n(r)$ versus $\log r$ for a number of small r, together with estimated confidence bounds for $\log C_n(r)$. The straight line fit is good and it gives the value of the estimate for the correlation dimension

$$\hat{\alpha}^{GP} = 0.89.$$

The problem of small distances in the Takens estimator can be attacked in the following way: introduce not only an upper (r_0), but also a lower cutoff distance $r_1 > 0$, which still can be very close to 0, and consider only those distances between points in the orbit which lie between r_1 and r_0. This certainly brings a bias into the estimate, but it keeps the estimator from diverging. [Such an estimator for the correlation dimension was first suggested by Ellner (1988).] For our numerical example it gives the values of the estimate (when the lower cutoff distances were taken $r_1^{(1)} = 10^{-3}$, $r_1^{(2)} = 10^{-4}$, $r_1^{(3)} = 10^{-5}$)

$$\hat{\alpha}^{T,\, r_1^{(1)}} = 0.98,$$

$$\hat{\alpha}^{T,\, r_1^{(2)}} = 0.96,$$

$$\hat{\alpha}^{T,\, r_1^{(3)}} = 0.95,$$

which is closer to the real value than the Grassberger–Proccacia estimate. Moreover, this "cutoff" Takens estimator has the same advantage over least-

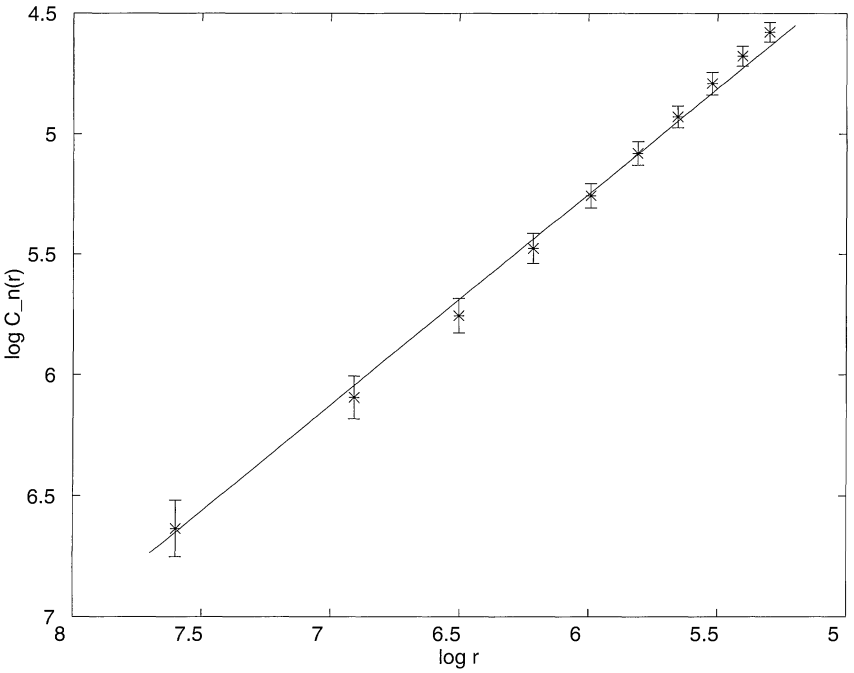

FIG. 2. *Linear regression,* $\log C_n(r)$ *versus* $\log r$.

squares as the original Takens estimator, that is, it is computationally more efficient.

Acknowledgment. The authors thank the referees for their very helpful comments that improved the presentation of the paper, as well as for spotting an error in an earlier version of the proof of Theorem 2.

REFERENCES

AARONSON, J., BURTON, R., DEHLING, H., GILAT, D., HILL, T. and WEISS, B. (1996). Strong laws for L- and U-statistics. *Trans. Amer. Math. Soc.* **348** 2845–2865.

DENKER, M. and KELLER, G. (1983). On U-statistics and von Mises's statistics for weakly dependent processes. *Z. Wahrsch. Verw. Gebiete* **64** 505–522.

DENKER, M. and KELLER, G. (1986). Rigorous statistical procedures for data from dynamical systems. *J. Statist. Phys.* **44** 67–93.

ELLNER, S. (1988). Estimating attractor dimensions from limited data: a new method, with error estimates. *Phys. Lett. A* **133** 128–133.

GRASSBERGER, P. and PROCACCIA, I. (1983). Characterization of strange attractors. *Phys. Rev. Lett.* **50** 346–349.

HALMOS, P. R. (1946). The theory of unbiased estimation. *Ann. Math. Statist.* **17** 34–43.

HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19** 293–325.

PHILIPP, W. (1986). Invariance principles for independent and weakly dependent random variables. In *Dependence in Probability and Statistics* (E. Eberlein and M. Taqqu, eds.) 225–268. Birkhäuser, Boston.

TAKENS, F. (1981). Detecting strange attractors in turbulence. *Dynamical Systems and Turbulence. Lecture Notes in Math.* **898** 366–381. Springer, Berlin.

TAKENS, F. (1985). On the numerical determination of the dimension of an attractor. *Dynamical Systems and Bifurcations. Lecture Notes in Math.* **1125** 99–106. Springer, Berlin.

YOSHIHARA, K. (1976). Limiting behaviour of U-statistics for stationary, absolutely regular processes. *Z. Wahrsch. Verw. Gebiete* **35** 237–252.

S. BOROVKOVA
SHELL RESEARCH AND TECHNOLOGY CENTRE
P.O. BOX 38000
1030 BN AMSTERDAM
THE NETHERLANDS
E-MAIL: Svetlana.A.Borovkova@opc.shell.com

R. BURTON
DEPARTMENT OF MATHEMATICS
OREGON STATE UNIVERSITY
KIDDER HALL 368
CORVALLIS, OREGON 97331
E-MAIL: burton@math.orst.edu

H. DEHLING
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF GRONINGEN
P.O. BOX 800
9700 AV GRONINGEN
THE NETHERLANDS
E-MAIL: dehling@math.rug.nl