Estimation and prediction for nonlinear time series

Borovkova, Svetlana Alfredovna

3 Consistency of the Takens estimator for the correlation dimension

3.1 Examples

Recall that the Takens estimator for the correlation dimension is given by

$$\hat{\alpha}_n^T = - \left(\frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \log \frac{\| X_i - X_j \|}{r_0} \right)^{-1}. \qquad (3.1)$$

To establish its consistency for general sequences $X_1, ..., X_n$, we have to study the average

$$\frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \log \| X_i - X_j \|, \qquad (3.2)$$

which has the form of a U-statistic. The question of consistency of the Takens estimator cannot be answered just by applying Theorem A, since the kernel function $h(X_i, X_j) = \log \| X_i - X_j \|$ is not bounded for observations (X_i, X_j) which are close to each other. It turns out that the Takens estimator is in general *not* consistent for stationary and ergodic sequences, in fact it is not even consistent for absolutely regular sequences. This is confirmed by the following counterexample by Aaronson *et.al.* ([1]):

Example 3.1   Let $W_1, W_2, ...$ be i.i.d. random variables with a continuous distribution F, satisfying

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

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

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

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 (3.2) diverges to $-\infty$ almost surely and hence, the Takens estimator (3.1) is not consistent.

    It is interesting to see what went wrong in this example. This has something to do with the drastic difference between the product distribution $F \times F$ and the 2-dimensional joint distributions $\mathbf{P}_{ij}$ induced by the pairs (X_i, X_j). Note that for any n, $X_n = W_i$ for some $i = 1, 2, ..., n$. So the condition above implies that $\mathbf{E}_{F \times F}(|\log|X_i - X_j||) < \infty$ still holds for all i, j. However, there is positive probability that $X_i = X_j$. Thus, the same expectation, but taken with respect to $\mathbf{P}_{ij}$, will be infinite, i.e.

$$\sup_{i,j} \mathbf{E}_{\mathbf{P}_{ij}}(|\log|X_i - X_j||) = \infty. \tag{3.3}$$

The previous example is quite crude, because there $|X_i - X_j| = 0$ with positive probability. In our next example we consider a stationary sequence $\{X_n\}_{n \in \mathbf{Z}}$ for which pairs (X_i, X_j) have a density with respect to Lebesgue measure and $\mathbf{E}|\log|X_i - X_j|| < \infty$, but

$$\lim_{n \longrightarrow \infty} \mathbf{E}(|\log|X_n - X_1||) = \infty,$$

and the Takens estimator is not consistent.

Example 3.2   Let $\{\tilde{X}_n\}_{n \in \mathbf{Z}}$ be i.i.d. random variables, uniformly distributed on $(0, 1)$, and $\{Y_n\}_{n \in \mathbf{Z}}$ be i.i.d. Bernoulli random variables with $\mathbf{P}[Y_n = 0] = \mathbf{P}[Y_n = 1] = 1/2$. Suppose, moreover, that $\{\tilde{X}_n\}$ and $\{Y_n\}$ are independent of each other. Define a new sequence $\{X_n\}$ by

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

where k is the maximal index such that $Y_{n-1} = ... = Y_{n-k} = 0$, and choose $\phi(k) = \exp(2^{2k})$, $\{Z_n\}$ i.i.d. sequence of uniformly $[0, 1]$-distributed random variables, independent of $\{\tilde{X}_n\}, \{Y_n\}$.

Note that $\{X_n\}_{n\in\mathbf{Z}}$ is a stationary, absolutely regular sequence. For this sequence

$$\sup_r \mathbf{E}|\log|X_0 - X_r|| = \infty,$$

since, due to the fact that $|X_0 - X_r| \le 1$ and so, $-\log|X_0 - X_r| \ge 0$,

$$
\begin{aligned}
\mathbf{E}(-\log|X_0 - X_r|) &= \mathbf{E}\left[\mathbf{E}(-\log|X_0 - X_r| \,|\{Y_n\})\right] \qquad (3.4)\\
&\ge \mathbf{E}\left(-\log|X_0 - X_r| \,|Y_r = 1,\, Y_{r-1} = ... = Y_0 = 0,\, Y_{-1} = 1\right) \cdot 2^{-r-2}\\
&= \left[\mathbf{E}(-\log Z_r) + \log\phi(r)\right] \cdot 2^{-r-2}
\end{aligned}
$$

and the supremum over r of (3.4) diverges to infinity, because

$$\sup_r \log\phi(r) \cdot 2^{-r-2} = \sup_r 2^{r-2} = \infty$$

for our choice of $\phi(r)$.

The U-statistic (3.2) for this sequence is divergent a.s. To show that we shall make use of the following lemma.

Lemma 3.1 *If* $Y_1, Y_2, ...$ *are non-negative i.i.d. random variables with* $\mathbf{E}[Y_1]^{1/2} = \infty$, *then a.s.*

$$\limsup_{n\longrightarrow\infty} \frac{1}{n^2} \sum_{i=1}^{n} Y_i = \infty.$$

Proof Since all Y_i's are positive,

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

and the fact that the r.h.s. of this inequality is infinite is the consequence of the following line of equivalent statements:

$$\mathbf{E}[Y_1]^{1/2} = \infty \Longleftrightarrow \sum_{n=1}^{\infty} \mathbf{P}(Y_n > \epsilon n^2) = \infty \Longleftrightarrow \mathbf{P}\{Y_n > \epsilon n^2 \text{ i.o.}\} = 1 \quad \forall\epsilon > 0,$$

which follow from the Borel-Cantelli lemma. $\square$.
Note that

$$- \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \log|X_i - X_j| \qquad (3.5)$$

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

where $R_1, ..., R_M$ are lengths of full zero-blocks of Y_i contained in the sample of size n. An application of the ergodic theorem yields that $\frac{M}{n} \longrightarrow \frac{1}{4}$ as $n \longrightarrow \infty$. The last term in (3.5) is divergent a.s., according to Lemma 3.1, 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

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

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

In the last section of this chapter we shall give a numerical example which illustrates the divergence of the Takens estimator in the case of infinite expectation in (3.3), while the Grassberger-Procaccia approach gives a reasonable estimate of the correlation dimension. However, in general the Takens estimator has its advantages over the Grassberger-Procaccia method, such as computational efficiency. And it turns out that additional conditions on the expectations in (3.3) lead 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 with unbounded kernel, which we present in the next section.

3.2   Weak consistency of U-statistics

This section contains the main theoretical results of the present chapter. We prove two consistency results for U-statistics of stationary ergodic, respec-

tively absolutely regular sequences. Compared with the results of Aaronson et al. [1], 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, i.e. $k = 1$. We note however that the results continue to hold for general m and k.

Theorem 3.1 *Let $\{X_n\}_{n \geq 1}$ be a stationary ergodic process with marginal distribution F, and let $h : \mathbf{R} \times \mathbf{R} \longrightarrow \mathbf{R}$ 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 $n \to \infty$,*

$$U_n \longrightarrow \theta(F) \ \text{in probability}. \tag{3.6}$$

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

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

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

We now define the truncated kernel $h_K(x, y) = h(x, y)\mathbf{1}\{|h(x, y)| \leq K\}$, where K is such that $(F \times F)\{(x, y) : |h(x, y)| = K\} = 0$. As $h_K(x, y)$ is bounded and $F \times F$-a.e. continuous, we get

$$\int |h_K(x, y)| d(\frac{1}{n} \sum_{k=1}^{n} \mu_k)(x, y) \to \int |h_K(x, y)) dF(x) dF(y)$$

and thus

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

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

Moreover, $h_K(x, y)$ satisfies all the conditions of Theorem A1 of Aaronson et al., and hence

$$\frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} h_K(X_i, X_j) \xrightarrow{a.s.} \int \int h_K(x, y) dF(x) dF(y) \quad as \ \ n \longrightarrow \infty. \tag{3.7}$$

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

$$| \int \int h(x, y) dF(x) dF(y) - \int \int h_K(x, y) dF(x) dF(y)| \to 0 \tag{3.8}$$

as $K \longrightarrow \infty$. Uniform integrability of $\{h(X_i, X_j), i, j \geq 1\}$ 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 $K \longrightarrow \infty$. This implies that

$$\mathbf{E} \left| \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} h_K(X_i, X_j) - \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} h(X_i, X_j) \right| \longrightarrow 0, \tag{3.9}$$

as $K \longrightarrow \infty$. Combining now (3.7), (3.8) and (3.9), the statement of the theorem follows. $\qquad\square$

In case of an absolutely regular process we can drop the continuity condition on the kernel, as the next theorem shows. The 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 [58].

Lemma 3.2 (Theorem 3.4 in Philipp [58]) *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* $\xi_1', \xi_2', \dots$, *such that for all* $k = 1, 2, \dots$

$$\mathbf{P}(\xi_k = \xi_k') = 1 - \beta_m, \tag{3.10}$$

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

Theorem 3.2 *Let* $\{X_n\}_{n \in \mathbf{N}}$ *be a stationary and absolutely regular process with marginal distribution F, and let* $h : \mathbf{R}^2 \longrightarrow \mathbf{R}$ *be measurable. Suppose moreover that the family of random variables* $\{h(X_i, X_j) : i, j \geq 1\}$ *is uniformly integrable. Then*

$$U_n \longrightarrow \theta(F) \ in \ L_1,$$

and hence also in probability.

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

$$E|h(X_i, X_j)|1_B \leq \epsilon \tag{3.11}$$

holds for all measurable sets B with $P(B) < \delta$. Then choose m, N so big that $2\beta_m < \delta$ and $\frac{m}{N} < \epsilon$. 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 $p = [\frac{n}{N+m}]$. By Lemma 3.2 there exists a sequence of independent N-dimensional vectors $\xi_1', \xi_2', \ldots$ with the same marginal distribution as (ξ_k) such that (3.10) 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 $H : \mathbf{R}^N \times \mathbf{R}^N \to \mathbf{R}$ by

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

where $\xi = (x_1, \ldots, x_N)$ and $\eta = (y_1, \ldots, y_N)$. From (3.11) we can infer that

$$E|H(\xi_k, \xi_l)|1_B \leq \epsilon \tag{3.12}$$

for all sets B with $P(B) < \delta$. The same holds if $(\xi_k)_k$ is replaced by (ξ_k'), by $F \times F$-integrability of H.

Independence of ξ_k' and ξ_l' implies that $EH(\xi_k', \xi_l') = \int \int h(x,y) dF(x) dF(y) = \theta(F)$ for all $k \neq l$. Thus by the U-statistics law of large numbers for independent observations

$$\frac{1}{p(p-1)} \sum_{1 \leq k \neq l \leq p} H(\xi_k', \xi_l') \to \theta(F) \qquad (3.13)$$

almost surely and in L_1. By construction of (ξ_k'), we have $P(\xi_k \neq \xi_k'$ or $\xi_l \neq \xi_l') \leq 2\beta_m < \delta$ and thus by (3.12)

$$E|H(\xi_k, \xi_l) - H(\xi_k', \xi_l')| = E|H(\xi_k, \xi_l) - H(\xi_k', \xi_l')| 1_{\{\xi_k \neq \xi_k' \text{ or } \xi_l \neq \xi_l'\}} \leq 2\epsilon$$

Hence

$$E \left| \frac{1}{p(p-1)} \sum_{1 \leq k \neq l \leq p} H(\xi_k, \xi_l) - \frac{1}{p(p-1)} \sum_{1 \leq k \neq l \leq p} H(\xi_k', \xi_l') \right| \leq 2\epsilon. \qquad (3.14)$$

Moreover $\left| \frac{1}{p(p-1)} - \frac{N^2}{n(n-1)} \right| \leq \frac{2\epsilon}{p(p-1)}$ for p large enough, and thus

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

where $C_0 = \sup_{i,j} E|h(\xi_i, \xi_j)|$. This last estimate together with (3.13) and (3.14) show that for n large enough

$$E \left| \frac{N^2}{n(n-1)} \sum_{1 \leq k \neq l \leq p} H(\xi_k, \xi_l) - \theta(F) \right| \leq C\epsilon \qquad (3.15)$$

Now, decompose the original U-statistics as follows:

$$\sum_{1 \leq i \neq j \leq n} h(X_i, X_j) = \sum_{1 \leq k \neq l \leq p} \sum_{i=n_k+1}^{n_k+N} \sum_{j=n_l+1}^{n_l+N} h(X_i, X_j)$$

$$+ \sum_{k=1}^{p} \sum_{n_k+1 \leq i \neq j \leq n_k+N} h(X_i, X_j)$$

$$+ 2 \sum_{1 \leq k, l \leq p} \sum_{i=n_k+N+1}^{n_{k+1}} \sum_{j=n_l+1}^{n_l+N} h(X_i, X_j)$$

$$+ \sum_{1 \le k \ne l \le p} \sum_{i=n_k+N+1}^{n_{k+1}} \sum_{j=n_l+1}^{n_l+N} h(X_i, X_j)$$

$$+ \sum_{k=1}^{p} \sum_{n_k+N+1 \le i \ne j \le n_{k+1}} h(X_i, X_j)$$

$$+ \sum_{i=n_p+N+1}^{n} \sum_{j=1}^{n} h(X_i, X_j)$$

$$+ \sum_{i=1}^{n_p+N} \sum_{j=n_p+N+1}^{n} h(X_i, X_j)$$

A careful study of the index set shows that

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

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

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

for n large enough. This, together with (3.15), proves the theorem. $\qquad \square$

3.3 Application to the Takens estimator

From Theorem 3.2 the consistency of the Takens estimator for absolutely regular sequence $X_1, X_2, ..., X_i \in \mathbf{R}^k$, follows if for some $\delta > 0$

$$\sup_{i,j} \mathbf{E} | \log \| X_i - X_j \| |^{1+\delta} < \infty. \qquad (3.16)$$

This condition can be translated into the distributional properties of the sequence.

Let $X_i \in A$ for all i, where A is some bounded subset of $\mathbf{R}^k$. Note that, if the joint distribution of (X_i, X_j) has a density f_{ij} which is *bounded* (by

C, say), then

$$\mathbf{E} |\log \| X_i - X_j \| |^{1+\delta} = \int_{A^2} \int |\log \| x - y \| |^{1+\delta} f_{ij}(x, y) dx dy$$
$$\leq C \int_{A^2} \int |\log \| x - y \| |^{1+\delta} dx dy < \infty,$$

and the condition (3.16) is satisfied.

On the other hand, if the distribution of the distances $\rho_{i,j} = \| X_i - X_j \|$ has a density $p(x)$ then the expectation in (3.16) can also be expressed as

$$\mathbf{E} |\log \| X_i - X_j \| | = \int_0^{r_0} |\log r|^{1+\delta} p(r) dr.$$

This integral is finite if $p(x)$ is of the order $O(x^{-\delta})$ for some $\delta < 1$ in the neighbourhood of zero, and the condition (3.16) is fulfilled in this case as well.

3.4 Numerical example

In this section we apply the Takens estimator and the Grassberger-Procaccica method to the stationary ergodic process $\{X_n\}_{n \in \mathbf{N}}$, defined by

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

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 Fig.1 the delay map X_{n+1} vs. X_n is shown.

Note that for this process

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

and, according to the results above, we expect the Takens estimator to diverge. And, indeed, computing $\hat{\alpha}^T$, as in (3.1), gives us extremely low values of the estimate, such as

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

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

On the other hand, this is no danger for the Grassberger-Procaccia estimator (2.4). In Fig.2 we plotted $\log C_n(r)$ vs. $\log r$ 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:
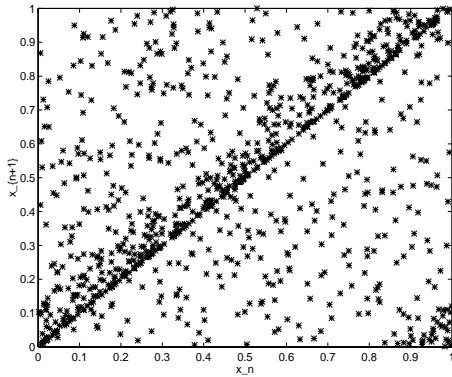
$$\hat{\alpha}^{GP} = 0.89.$$
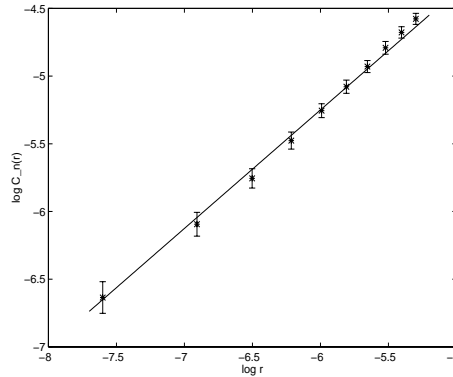

Figure 3.1: Delay map x_{n+1} vs. x_n Figure 3.2: Linear regression $\log C_n(r)$ vs. $\log r$

The problem of small distances in the Takens estimator can be attacked by introducing not only an upper (r_0), but also a lower cut-off 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. (Note that this is the Ellner estimator, considered in Chapter 2). For our numerical example it gives the values of the estimate (when the lower cut-off 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 Grassberger-Proccacia estimate. Moreover, this "cut-off"-Takens estimator has the same advantage over least-squares as the original Takens estimator, i.e. it is computationally more efficient.