Remarks on the linear regression approach to dimension estimation

Gerhard Keller and Ralph Sporer

Abstract

We discuss some statistical theory of the simultaneous estimation of correlation integrals from dynamical data with varying radii and embedding dimensions. Thereby we focus on the estimation of the covariance matrix of these estimators taking into account the finite sample size and the correlation time effects observed by Theiler [23]. As applications we discuss linear model statistics like linear regression estimates of correlation dimension and entropy and the detection of noise.

1 Introduction

Let $X_1, X_2, X_3, \ldots$ be a real-valued stationary time series that is mixing in a sense to be made precise later. Typical examples would be a) independent identically distributed (i.i.d.) random observations, b) observation on a "chaotic" dynamical system, c) observations on a noisy system. In particular there may be some interesting dependence between consecutive observations that can be studied by looking at the distribution μ_ℓ of blocks

$$Y_i^\ell := (X_i, \ldots, X_{i+\ell-1}) \in \mathbb{R}^\ell \ .$$

The length ℓ of the blocks is called the *embedding dimension*, and for fixed ℓ the sequence $Y_1^\ell, Y_2^\ell, Y_3^\ell, \ldots$ is again stationary and mixing.

Some aspects of the geometry of the distribution μ_ℓ can be described by means of the *correlation integrals*

$$C(r,\ell) := \int\int 1_{\{\|y-y'\|<r\}}\, d\mu_\ell(y)\, d\mu_\ell(y')$$

where $\|y - y'\|$ denotes the euclidean (or any other suitable) distance of y and y'. Grassberger and Procaccia [14] used the functional dependence of $\log C(r,\ell)$ on r and ℓ to describe quantitative features of *deterministic* chaotic systems. They observed that in many cases there are real numbers $\nu > 0$ and $h > 0$ such that

$$\log C(r,\ell) = \begin{cases} \nu \cdot \log r + o(\log r) & \text{as } r \to 0 \text{ when } \ell \text{ is large,} \\ -\ell \cdot h + o(\ell) & \text{as } \ell \to \infty \text{ when } r \text{ is small.} \end{cases} \tag{1}$$

ν is called the *correlation dimension* and h is an entropy like quantity. Cutler [10] gives a rather comprehensive review of much of the underlying theory.

If, on the other hand, the X_i are i.i.d. observations and if $\|y - y'\|$ denotes the maximum norm of $y - y'$, then $C(r,\ell) = (C(r,1))^\ell$ such that the following model assumption makes sense

$$\log C(r,\ell) = \ell \cdot \nu_1 \cdot (\log r + o(\log r)) \quad \text{as } r \to 0, \tag{2}$$

where ν_1 denotes the correlation dimension of the distribution μ_1 of the X_i, see [6]. The two model assumptions (1) and (2) differ drastically in the sense that (1) describes a finite entropy situation whereas (2) reflects the in general infinite entropy of true random observations.

The above considerations motivate the following decomposition of $\log C(r,\ell)$ into a linear part with constant term and a (hopefully small) nonlinear remainder $\delta(r,\ell)$:

$$\log C(r,\ell) = \nu \cdot \log r - h \cdot \ell + \nu_1 \cdot \ell \cdot \log r + C + \delta(r,\ell) \ . \tag{3}$$

This model, although appropriate if $\|y - y'\|$ denotes the maximum norm, can be improved if the euclidean distance is used by replacing $\log r$ with $\log \frac{r}{\sqrt{\ell}}$, *i.e* by using dimension scaled distances, see [13].

In practice the unknown coefficients ν, h, ν_1, C must be estimated from a finite number of observations $X_1, \ldots, X_N$. Therefore (3) (or any other linear or nonlinear model describing $\log C(r,\ell)$) can be fitted only on finitely many parameter pairs $(r_1,\ell_1), \ldots, (r_p,\ell_p)$ the appropriate choice of which depends on the sample size and the nature of the observed data. In particular we make no attempt to calculate limits as $r \to 0$ or $\ell \to \infty$. Using the notation

$$(\underline{r},\underline{\ell}) = \begin{pmatrix} r_1 & \ell_1 \\ \vdots & \vdots \\ r_p & \ell_p \end{pmatrix} , \quad Z(\underline{r},\underline{\ell}) = \begin{pmatrix} \log C(r_1,\ell_1) \\ \vdots \\ \log C(r_p,\ell_p) \end{pmatrix} , \quad \delta = \begin{pmatrix} \delta(r_1,\ell_1) \\ \vdots \\ \delta(r_p,\ell_p) \end{pmatrix}$$

we write our model (3) as

$$Z(\underline{r},\underline{\ell}) = M\beta + \delta \tag{4}$$

where M is a matrix involving only the controlled parameters $\log \underline{r}$ and $\underline{\ell}$ and where $\beta = (\nu, h, \nu_1, C)^t$ *is that coefficient vector that yields the best fit in (4) in the least squares sense,* i.e.

$$\beta = B \cdot Z(\underline{r},\underline{\ell}) \quad \text{where} \quad B = (MM^t)^{-1}M^t \ .$$

Our aim is to estimate this coefficient vector β from observed data.

Suppose now that $\hat{Z} = \hat{Z}(X_1, \ldots, X_N) \in \mathbb{R}^p$ is a "suitable" estimator for the p-vector $Z(\underline{r},\underline{\ell})$ of logarithms of correlation integrals, suitable in the following sense:

$$\hat{Z} = Z(\underline{r},\underline{\ell}) + \epsilon = M\beta + \delta + \epsilon$$

where ϵ is approximately $\mathcal{N}(\underline{0}, V)$–distributed (approximately in the sense of a central limit theorem), and there is a reliable (*i.e.* "consistent" in statistical terms) estimate $\hat{V} = \hat{V}(X_1, \ldots, X_N)$ of V. Then $\hat{\beta} = B\hat{Z}$ is the least squares estimator for β and $\hat{\epsilon} = \hat{Z} - M\hat{\beta} = (\mathbf{1} - MB)\hat{Z}$ are the corresponding residues. The distributions of $\hat{\beta}$ and $\hat{\epsilon}$ are known, namely $\hat{\beta} \sim \mathcal{N}(\beta, BVB^t)$ and $\hat{\epsilon} \sim \mathcal{N}(\delta, SVS^t)$ where $S = \mathbf{1} - MB$. All this is well known, see *e.g.* [24]. Replacing V by the consistent estimate $\hat{V}$ we can *e.g.* calculate confidence bounds for β, test the hypothesis "$\nu_1 = 0$", *i.e.* the absence of true randomness on the scale of radii $r_1, \ldots, r_p$, or discriminate between the systematic error δ of our linear model (3) and the statistical error ϵ. In section 3 we illustrate this by the results of some numerical simulations. Beforehand we turn to the problem of how to obtain good estimates $\hat{Z}$ of $Z(\underline{r},\underline{\ell})$ and $\hat{V}$ of V.

2 Covariance estimates

2.1 The covariance matrix of the correlation integrals

Recall that $C(r,l) = \int\int h(y,y')\, d\mu_\ell(y)\, d\mu_\ell(y')$ where $h(y,y') = 1_{\{\|y-y'\|<r\}}$. Quantities represented by integrals in this way are usually estimated by a so called U-statistic

$$\frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j\neq i}}^{N} h(Y_i, Y_j) \ .$$

In the case of i.i.d. observations Y_i the U-statistic is an asymptotically normal unbiased minimal variance estimator of $C(r,l)$, see [15]. These properties essentially persist if the observations are mixing with exponentially decaying correlations. The concept of mixing which is most useful in this context is that of ***absolute regularity***, see [25] for a purely probabilistic treatment and [12] for its applicability to chaotic dynamical systems. Although strong mixing properties like this one are difficult to verify on a theoretically level in the case of non-uniformly hyperbolic systems, one can expect that many chaotic systems show the same central limit behaviour. In the last section of this note we discuss this aspect in some more detail.

For computational purposes we assume that the observations Y_i^ℓ are t-dependent, *i.e.* Y_i^ℓ and Y_j^ℓ are independent, if $|i-j| > t$. For a given data set t should be carefully chosen. In our Hénon example in the next section $t = 10$ seems to be a good choice. We modify the U-statistic estimator of $C(r,l)$ accordingly:

$$U_N(r,l) = \frac{1}{\pi(N,t)} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ |j-i|>2t}}^{N} h(Y_i^\ell, Y_j^\ell) \ . \tag{5}$$

Here $\pi(N,t) = (N-2t)(N-2t-1)$ is the number of pairs in the sum. The minimal index distance $2t$ (instead of t) will help to avoid some dependencies when we calculate the variance of $U_N(r,l)$. As we are going to estimate $C(r_1,\ell_1),\ldots,C(r_p,\ell_p)$ simultaneously, we have indeed to calculate the covariance matrix of the $\mathbb{R}^p$–valued estimator

$$U_N(\underline{r},\underline{\ell}) = \begin{pmatrix} U_N(r_1,\ell_1) \\ \vdots \\ U_N(r_p,\ell_p) \end{pmatrix} \ .$$

A tedious, though elementary calculation yields the following expression for the covariance matrix K of $U_N(\underline{r},\underline{\ell})$:

$$K = 4 \cdot \left(N^{-1}P + N^{-2}(\frac{1}{2}Q - R - (1+2t)P) + O(N^{-3}) \right) \tag{6}$$

where P, Q and R are $p\times p$–matrices, that can be described as follows: For $u = 1,\ldots,p$ let

$$h^{(u)}(y,y') \quad = \quad 1_{\{\|y-y'\|<r_u\}}$$

$$
\begin{aligned}
h_0^{(u)} &= \int\int h^{(u)}(y,y')\,d\mu_{\ell_u}(y)\,d\mu_{\ell_u}(y') \quad (= C(r_u,\ell_u)) \\
h_1^{(u)}(y) &= \int h^{(u)}(y,y')\,d\mu_{\ell_u}(y') \\
h_2^{(u)}(y,y') &= h^{(u)}(y,y') - h_1^{(u)}(y) - h_1^{(u)}(y') + h_0^{(u)} \ .
\end{aligned}
$$

Then for $u,v = 1,\ldots,p$

$$
\begin{aligned}
P_{u,v} &= \sum_{k=-t}^{t} \left(E\left[h_1^{(u)}(Y_i^{\ell_u}) \cdot h_1^{(v)}(Y_{i+k}^{\ell_v}) \right] - h_0^{(u)} \cdot h_0^{(v)} \right) \quad (i \text{ arbitrary}) \\
Q_{u,v} &= \sum_{p=-t}^{t} \sum_{q=-t}^{t} \left(E\left[h^{(u)}(Y_i^{\ell_u}, Y_j^{\ell_u}) \cdot h^{(v)}(Y_{i+p}^{\ell_v}, Y_{j+q}^{\ell_v}) \right] - h_0^{(u)} \cdot h_0^{(v)} \right) \\
&\qquad (\text{where } |i-j| > 3t) \\
R_{u,v} &= \sum_{k=-t}^{t} |k| \cdot \left(E\left[h_1^{(u)}(Y_i^{\ell_u}) \cdot h_1^{(v)}(Y_{i+k}^{\ell_v}) \right] - h_0^{(u)} \cdot h_0^{(v)} \right) \quad (i \text{ arbitrary})
\end{aligned}
$$

Remarks:

(i) In the case of a single radius and embedding dimension the formula for K reduces to that of Theiler [23] if $t = 0$, *i.e.* if the Y_i are independent, namely $K = 4(N^{-1}P + N^{-2}(\frac{1}{2}Q - P))$. [1]

(ii) As we assume Y_i and Y_{i+k} to be independent for $|k| > t$, we have the following relation between $P_{u,u}$ and Theiler's *correlation time* τ: $P_{u,u} = \tau \cdot E\left[(h_1^{(u)}(Y_i^{\ell_u}))^2 \right]$

(iii) Asymptotically (as $N \to \infty$) the N^{-1}–term in the decomposition of K dominates. However, for small radii r_u, $P_{u,u}$ tends to be much smaller than $Q_{u,u}$ such that the N^{-2}–term may be comparable in size to the N^{-1}–term even for $N = 10000$. In the case of independent observations the statistic $U_N(r_u,\ell_u) - h_0^{(u)}$ can be decomposed into a sum of two random variables in analogy to the decomposition of its variance into an N^{-1}– and N^{-2}–term. The first part is asymptotically normal, the second one is a weighted sum of squares of normal random variables, see [11, Example 2.2.7]. Therefore, if the N^{-2}–term is not neglectable, $U_N(r_u,\ell_u) - h_0^{(u)}$ is not close to an exact normal but to a slightly skewed normal distribution. It seems impossible, however, to estimate the size of this effect from the data.

As $U_N(\underline{r},\underline{\ell}) - C(\underline{r},\underline{\ell})$ is approximately $\mathcal{N}(0,K)$–distributed, standard results from probability theory guarantee that $\log U_N(\underline{r},\underline{\ell}) - \log C(\underline{r},\underline{\ell})$ is approximately $\mathcal{N}(0,V)$–distributed where $V_{u,v} = K_{u,v}/(C(r_u,\ell_u)C(r_v,\ell_v))$, $u,v = 1,\ldots,p$. Higher order correction terms which occur also in this approximation are small compared to the leading term and can be neglected.

[1] Theiler gives $(\frac{1}{2}Q + \frac{1}{2}P)$ instead of $(\frac{1}{2}Q - P)$, but as P occurs also at the order of N^{-1}, this is practically meaningless.

2.2 An estimator for the covariance matrix

An unbiased and consistent estimator for $P_{u,v}$ is

$$\sum_{k=-t}^{t} \frac{1}{N-k} \sum_{i=1}^{N-k} \left(h_1^{(u)}(Y_i^{\ell_u}) \cdot h_1^{(v)}(Y_{i+k}^{\ell_v}) - h_0^{(u)} \cdot h_0^{(v)} \right) .$$

As the $h_0^{(u)}$ and $h_1^{(u)}(Y_i^{\ell_u}) = \int h^{(u)}(Y_i^{\ell_u}, y')\, d\mu_{\ell_u}(y')$ are not explicitly known (they are defined in terms of the unknown distribution μ_{ℓ_u}), we replace them through estimators

$$H_0^{(u)} := U_N(r_u, \ell_u) \quad \text{and} \quad H_{1,i}^{(u)} := \frac{1}{N-2t-1} \sum_{\substack{j=1 \\ |j-i|>t}}^{N} h^{(u)}(Y_i^{\ell_u}, Y_j^{\ell_u}) \tag{7}$$

and denote

$$\hat{P}_{u,v} := \sum_{k=-t}^{t} \frac{1}{N-k} \sum_{i=\max\{1,1-k\}}^{\min\{N,N-k\}} \left(H_{1,i}^{(u)} \cdot H_{1,i+k}^{(v)} - H_0^{(u)} \cdot H_0^{(v)} \right) . \tag{8}$$

Another tedious but elementary calculation shows that

$$\begin{aligned} E[4N^{-1}\hat{P}] &= 4 \cdot \left(N^{-1}P + N^{-2}(Q - R - 5(1+2t)P) + O(N^{-3}) \right) \\ &= K + 2N^{-2}Q - 16(1+2t)N^{-2}P + O(N^{-3}) , \end{aligned}$$

i.e. $4N^{-1}\hat{P}$ is not an unbiased estimator for P. As a matter of fact, the correction term can be both, negative or positive, depending on the particular situation. An unbiased estimator for Q is easily found:

$$\hat{Q}_{u,v} := \frac{1}{\tilde{\pi}(N,t)} \sum_{\substack{i,j=1 \\ |j-i|>3t}}^{N} \sum_{p,q=-t}^{t} \left(h^{(u)}(Y_i^{\ell_u}, Y_j^{\ell_u}) \cdot h^{(v)}(Y_{i+p}^{\ell_v}, Y_{j+q}^{\ell_v}) - H_0^{(u)} \cdot H_0^{(v)} \right) \tag{9}$$

where $\tilde{\pi}(N,t) = (N-3t)(N-3t-1)$. The importance of the "$p=q$"-terms in $\hat{Q}_{u,v}$ for deterministic data is obvious: If $Y_j^{\ell_u}$ is close to $Y_i^{\ell_u}$ (because the system returned close to a previously attained state), also $Y_{j+p}^{\ell_v}$ and $Y_{i+p}^{\ell_v}$ are likely to be close to each other for small p. The "$p \neq q$"-terms contribute to $\hat{Q}_{u,v}$ if the system spends some time near a fix point or a periodic orbit of small period.

To summarize: $\hat{K} := 4N^{-1}(1 + 4(1+2t)N^{-1})\hat{P} - 2N^{-2}\hat{Q}$ is an estimator for the covariance matrix K of $U_N(\underline{r}, \underline{\ell})$ which is unbiased up to terms of order $O(N^{-3})$. It is defined by (8) and (9).

As the computation of $\hat{Q}$ involves about $p^2(2t+1)^2N^2$ terms, we determine only the diagonal terms $\hat{Q}_{u,u}$ and use the following modified estimator $\hat{K}'$ for K: Decompose $4N^{-1}\hat{P} = D^{1/2}CD^{1/2}$ where D is the diagonal matrix made up from the diagonal

elements of $4N^{-1}\hat{P}$. Thus C is the correlation matrix corresponding to $4N^{-1}\hat{P}$. We modify D by subtracting the diagonal elements of $2N^{-2}\hat{Q}$: $D'_{u,u} := D_{u,u} - 2N^{-2}\hat{Q}_{u,u}$, and obtain the estimator $\hat{K}' := D'^{1/2} C D'^{1/2}$ for K. The diagonal of $\hat{K}'$ coincides with that of $\hat{K}$ whereas its correlation structure is that of the leading term $4N^{-1}\hat{P}$ of $\hat{K}$. This leads us finally to the following estimator $\hat{V}$ for the covariance matrix of $\log C(\underline{r}, \underline{\ell})$:

$$\hat{V}_{u,v} := \frac{\hat{K}'_{u,v}}{U_N(r_u, \ell_u)\, U_N(r_v, \ell_v)} \,.$$

3 Simulations

We tested the reliability of the statistical procedures described in the previous sections on Hénon time series. More exactly: $X_1, \ldots\ldots, X_N$ are the x-coordinates of N consecutive iterates of a Hénon system with parameters $a = 1.4$ and $b = 0.3$. (In some cases noise was added, see below.) We always used a set of radii $r_0, \ldots, r_{20}$ ranging from $r_0 = 0.01$ to $r_{20} = 0.41$. Instead of working with the maximum-norm distance we used dimension scaled euclidean distances, see [13]. Based on the observed decay of correlations for the random variables $H^{(u)}_{1,i}$ from (7) we assumed that X_i and X_j are independent if $|i - j| > t = 10$.

(i) *Fixed embedding dimension $\ell = 4$; no noise; $N = 1000$ and $N = 10000$:*
We performed 100 independent runs recording for the j-th run the estimated logarithms of the correlation integrals $Z_j(r_i) = \log C_j(r_i)$ and their estimated standard deviations $\sigma_j(r_i)$. The averages of these quantities over all runs are denoted $\bar{Z}(r_i)$ and $\bar{\sigma}(r_i)$, respectively. The sample standard deviation of the $Z_j(r_i)$ is denoted by S_i. The comparison of $\bar{\sigma}(r_i)$ and S_i in Figure 1, where we give both, the values of $\bar{\sigma}_i$ we obtained using the uncorrected variance estimator $\hat{P}_{u,u}$ and those using the corrected estimator $\hat{Q}_{u,u}$, shows that, at least for small radii the corrected estimates for the standard deviation are on the average closer to the sample standard deviation than the uncorrected values. As to be expected, this effect is much stronger for $N = 1000$ than for $N = 10000$. We remark that for $N = 1000$ there were 4 runs where the corrected estimator for $\sigma_j^2(r_i)$ took negative values for some radii. As this estimator is the difference of two values, it is not surprising that this happened a few times. For our further calculations we set these values to 0.

From the same data we calculated 100 estimates ν_j for the correlation dimension and their estimated standard deviations σ_j and denoted their 95%–confidence intervals by $J_j = [\nu_j - 1.96\,\sigma_j, \nu_j + 1.96\,\sigma_j]$. As the true value of ν is not known we could not simply count the number of runs where the true value is inside the confidence interval. Instead we counted the number of runs where $\bar{\nu} = \frac{1}{100}\sum_{j=1}^{100} \nu_j$ is in this interval [2]. As long as the distribution of the σ_j^2 is close

[2] The average correlation dimension was $\bar{\nu} = 1.19330 \pm 0.00111$ for r_0–r_{10} and $\bar{\nu} = 1.20881 \pm 0.00078$ for r_{10}–r_{20}

to normal one can approximately calculate the expected number of successes. The observed success probabilities with the expected ones in brackets are:

	$N = 1000$		$N = 10000$	
	r_0–r_{10}	r_{10}–r_{20}	r_0–r_{10}	r_{10}–r_{20}
uncorrected	0.99 (0.934)	0.95 (0.930)	0.98 (0.945)	0.98 (0.945)
corrected	0.93	0.93	0.96 (0.941)	0.96 (0.945)

(The corrected σ_j^2 for $N = 1000$ were so non-normal that no useful expected value of the success probabilities could be calculated.) So, except for large radii and small numbers of observations, the uncorrected values seem to give too pessimistic confidence intervals whereas the corrected ones work very well for $N = 10000$ but tend to give slightly too small confidence intervals for $N = 1000$. Compared to other approaches that provide confidence intervals for correlation integrals or the correlation dimension (*e.g.* [21, 17]) we neither need parametric assumptions on the dynamical system producing the data nor very large sample sizes in order to produce reasonable estimates.

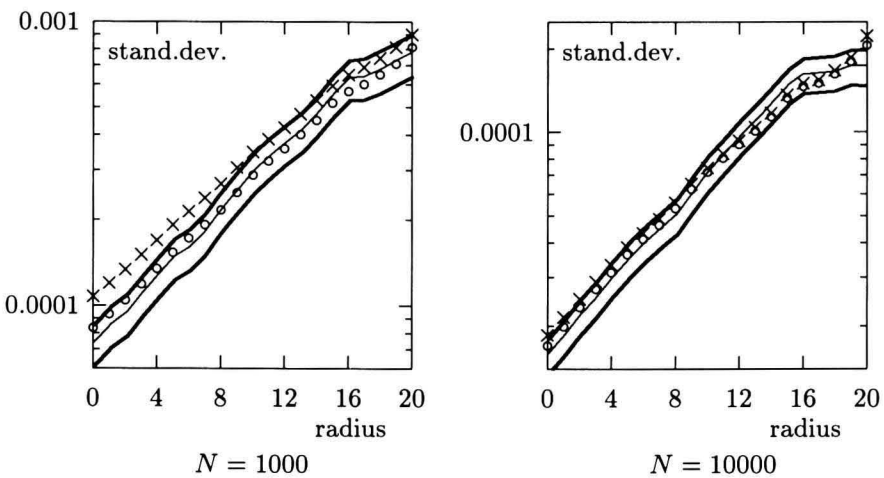

Figure 1: Comparison of the average estimated standard deviation $\bar{\sigma}(r_i)$ to the sample standard deviation S_i for time series of length 1000 and 10000. Logarithmic plot. Lines: S_i with 95%–confidence intervals. Crosses: uncorrected variance estimator. Circles: corrected variance estimator.

(ii) *Simultaneous estimation of ν, h, and ν_1; $\ell = 4, \ldots, 7$; $N = 10000$:*
We performed single runs on estimating ν, h, and ν_1 from the linear model (3) using the radii $r_0 = 0.01, \ldots, r_{10} = 0.064$. In all cases we calculated 95%–confidence intervals:

(a) *No noise:* As ν_1 was not significantly different from 0 (interpretation: no noise!) we excluded the variable ν_1 from the model and estimated ν and h again:

	ν	h	ν_1
ν_1 included	$[1.035, 1.172]$	$[0.241, 0.387]$	$[-0.013, 0.025]$
ν_1 excluded	$[1.077, 1.188]$	$[0.277, 0.313]$	–

(b) 1% *noise in the system:* As ν was not significantly different from 0 we excluded the variable ν from the model and estimated h and ν_1 again:

	ν	h	ν_1
ν included	$[-0.050, 0.441]$	$[2.471, 2.992]$	$[0.445, 0.560]$
ν excluded	–	$[2.789, 2.963]$	$[0.520, 0.553]$

(c) 1% *measurement noise:* As ν was not significantly different from 0 we excluded the variable ν from the model and estimated h and ν_1 again:

	ν	h	ν_1
ν included	$[-0.216, 0.224]$	$[2.717, 3.167]$	$[0.531, 0.631]$
ν excluded	–	$[2.868, 3.022]$	$[0.567, 0.596]$

(iii) *Discrimination between systematic and statistical errors; fixed embedding dimension $\ell = 4$; $r_0, \ldots, r_{10}$; $N = 10000$:*
Recall from the introduction that the residues $\hat{\epsilon} = \hat{Z} - M\hat{\beta}$ are normally distributed with mean δ (= the systematic error) and a degenerate covariance matrix SVS^t (representing the statistical error) that can be estimated from the data. In particular, the euclidean length of the difference vector $\hat{\epsilon} - \delta$ has a distribution that can approximately be determined by simulations with normal random numbers. This distribution can be compared to the actually observed length $|\hat{\epsilon}|$ of $\hat{\epsilon}$. In our numerical example we found $|\hat{\epsilon}| = 0.0533$. At the same time a 95%–confidence interval for $|\hat{\epsilon} - \delta|$ was $[0, 0.0129]$. Indeed, the largest among 500 simulated values of $|\hat{\epsilon} - \delta|$ was 0.0210. So more than 2/3 of the length of the observed residue vector is due to the nonlinearity of the function $\log r \mapsto \log C(r)$ and only a smaller part of it can be explained by statistical fluctuations.

4 Remarks on mixing in chaotic systems

In [12, 9, 22] the following framework for statistics on data from dynamical system was considered: Let $T : M \to M$ describe a time-discrete, deterministic dynamical system on a metric space (M, d). Suppose there is an ergodic, T-invariant Borel probability measure P on M, and fix a finite partition $\mathcal{Z} = (Z_1, \ldots, Z_\ell)$ of M. Then

$$\begin{aligned} X_n(\omega) &:= T^n(\omega) \quad \text{and} \\ \xi_n(\omega) &:= j \quad \text{if} \quad X_n(\omega) \in Z_j \quad (\omega \in M) \end{aligned}$$

define ergodic stationary processes on the probability space (M, P). Here the index n ranges over $I = \mathbb{Z}$ if T is invertible and over $I = \mathbb{N}$ otherwise.

Sometimes it is possible to recover the process $(X_n)_{n\in I}$ from the label process $(\xi_n)_{n\in I}$ via a functional $\Phi : \{1, \dots, \ell\}^I \to M$ for which there are real constants $C > 0$ and $\alpha \in (0, 1)$ such that for P-a.e. ω

$$d\left(\Phi((k_i)_{i\in I}) - X_n(\omega)\right) \leq C \cdot \alpha^m, \tag{10}$$

whenever $k_i = \xi_{n+i}(\omega)$ for all $i \in I$ with $|i| \leq m$. In particular, $\Phi((\xi_{n+i}(\omega))_{i\in I}) = X_n(\omega)$.

In such a situation good mixing properties of the process $(\xi_n)_{n\in I}$ guarantee the asymptotic normality of $U_N(\underline{r}, \underline{\ell})$. In particular it suffices that the process $(\xi_n)_{n\in I}$ is absolutely regular with mixing coefficients β_n decreasing at a rate of $n^{-(2+\delta)}$ or faster. In the language of ergodic theory this means that $\mathcal{Z}$ is a weak Bernoulli partition for (T, P) with mixing rate β_n.

The asymptotic normality of the $\mathbb{R}^p$-valued process $U_N(\underline{r}, \underline{\ell})$ and the structure and estimability of its limiting covariance matrix are discussed in [12, 9]. Indeed, besides the absolute regularity and property (10) a mild regularity assumption on the measure P is needed, see [12, Theorem 1]. For our particular U-statistic $U_N(\underline{r}, \underline{\ell})$ it is not hard to show that condition (3.6) of [12] is satisfied if the functions $r \mapsto C(r, \ell)$ are Hölder continuous, a very reasonable assumption in the context of dimension estimation.

Absolute regularity (with even exponentially decreasing β_n) and property (10) are known since long for many uniformly hyperbolic or uniformly expanding systems such as mixing torus automorphisms, Axiom-A-diffeomorphisms [5], piecewise expanding interval maps [16] and others. For non-uniformly hyperbolic dynamical systems the situation is much more complicated, but results for two prototype systems indicate that also for such systems the statistical approach to dimension estimation via U-statistics is justified: For Collet-Eckmann maps (*i.e.* quadratic interval maps where the critical point has a positive Lyapunov exponent) it was essentially proved in [8] that there exists an absolutely continuous f-invariant probability measure μ on $[0, 1]$. Without essential loss of generality one may assume that (f, μ) is mixing, cf. [4]. The exponential weak Bernoulli property of the partition $([0, \frac{1}{2}], (\frac{1}{2}, 1])$ for the system (f, μ) is proved in [9] (building on results from [19]), and the approximation property (10) is an immediate consequence of [20]. In view of the work of Benedicks and Carleson [1] one might hope that the Collet-Eckmann property is in a sense typical for unimodal maps which have no stable attractor. In the case of the Hénon family there is not yet a complete theoretical justification for the statistical approach, but the relevant results obtained during the last years are nevertheless impressive: For a set of parameters of positive Lebesgue measure in the Hénon family Benedicks and Young [3] proved (building on results from [2]) the existence of a SBR-measure (that is a physically observable invariant measure), and more recently they proved exponential decay of correlations and a Central Limit Theorem for Hölder continuous observables of these systems [3]. Finally we mention the work of Chernov [7] who offers a kind of tool-box to investigate mixing properties of a broad class of dynamical systems.

[3]Reported at the Banach Center Symposium on Ergodic Theory and Dynamical Systems at Warsaw, June 1995

References

[1] Benedicks, M. and L. Carleson, On iterations of $1 - ax^2$ on $(-1, 1)$, *Annals Math.* **122** (1985), 1–25.

[2] Benedicks, M. and L. Carleson, The dynamics of the Hénon map, *Annals Math.* **133** (1991), 73–169.

[3] Benedicks, M. and L.S. Young, Sinai-Bowen-Ruelle measures for certain Hénon maps, *Inventiones Math.* **112** (1993), 541–576.

[4] Blokh, A.M. and M.Yu. Lyubich, Measurable dynamics of S-unimodal maps of the interval, *Ann. Sc. E.N.S. 4e Série* **24** (1991), 545–573.

[5] Bowen, R. Equilibrium states and the ergodic theory of Anosov diffeomorphisms, Springer Lecture Notes in Math. **470** (1975).

[6] Brock, W.A., W.D. Dechert and J.A. Scheinkman, A test for independence based on the correlation dimension, Preprint SSRI W.P. 8702 (1987).

[7] Chernov, N., Limit theorems and Markov approximations for chaotic dynamical systems, *Prob. Th. Rel. Fields* **101** (1995), 321–362.

[8] Collet, P. and J.-P. Eckmann, Positive Liapounov exponents and absolute continuity for maps of the interval, *Ergod. Th. & Dynam. Sys.* **3** (1983), 13–46.

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

[10] Cutler, C.D., A review of the theory and estimation of fractal dimension, in: *Nonlinear Times Series and Chaos, Vol. I: Dimension Estimation and Models,* pp. 1–107, ed. H. Tong, World Scientific, Singapore, 1993.

[11] Denker, M., *Asymptotic Distribution Theory in Nonparametric Statistics,* Vieweg, Braunschweig/Wiesbaden, 1985.

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

[13] Frank, M. et al., Improvement of K_2-entropy calculations by means of dimension scaled distances, *Physica D* **65** (1993), 359–364.

[14] Grassberger, P. and I. Procaccia, Measuring the strangeness of attractors, *Physica D* **9** (1983), 189–208.

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

[16] Hofbauer, F. and G. Keller, Ergodic properties of invariant measures for piecewise monotonic transformations, *Math. Zeitschrift* **180** (1982), 119–140.

[17] Judd, K., An improved estimator of dimension and some comments on providing confidence intervals, *Physica D* **56** (1992), 216–228.

[18] Keller, G., Exponential weak Bernoulli mixing for Collet-Eckmann maps, *Israel J. Math.* **86** (1994), 301–310.

[19] Keller, G. and T. Nowicki, Spectral theory, zeta functions and the distribution of periodic orbits for Collet-Eckmann maps, *Commun. Math. Phys.* **149** (1992), 31–69.

[20] Nowicki, T. Symmetric S-unimodal mappings and positive Liapunov exponents, *Ergod. Th. & Dynam. Sys.* **5** (1985), 611–616.

[21] Ramsey, J.B. and H.-J. Yuan, (a) Bias and error bars in dimension calculations and their evaluation in some simple models, *Physics Letters A* **134(5)** (1989), 287-297. (b) The statistical properties of dimension calculations using small data sets, *Nonlinearity* **3** (1990), 155–176.

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

[23] Theiler, J., Statistical precision of dimension estimators, *Physical Review A* **41** (1990), 3038–3051.

[24] Weisberg, S., *Applied Linear Regression,* John Wiley & Sons, New York, 1985.

[25] Yoshihara, K., Limiting behaviour of U-statistics for stationary, absolutely regular processes, *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **35** (1976), 237–252.

G.Keller, Math. Institut, Universität Erlangen, Germany
R. Sporer, Phys. Institut, Universität Erlangen, Germany