On estimators for information dimension

Manfred Denker and Aleksey Min [1]

Georg-August-Universität Göttingen and Technische Universität München

Abstract

We introduce a new estimator for the information dimension. Keller (1997) posed the question whether this estimator is an alternative to standard procedures. We show consistency and asymptotic normality of this estimator in the case of i.i.d. random vectors. Simulations show that it performs well.

1 Introduction

At least since the pioneering work in Cutler (1993) the problem of estimation dimensions of probability measures received considerable interest. The main objective is to estimate their information dimension [see Cutler (1991) for a definition]. The interest in this question arose originally in the need of estimating dimensions of attractors in dynamical systems, like the correlation dimension of Grassberger and Procaccia (1983). This dimension can be estimated by a combined method using U-statistics as an empirical correlation integral and least square regression analysis [see Denker and Keller (1986)]. A new technique has been introduced by Cutler and Dawson (1989, 1990) using nearest neighbour analysis. This technique avoids the

[1]Acknowledgements: The research of the second author was supported by the Graduiertenkolleg *Strömungsinstabilitäten und Turbulenz* at the University of Göttingen.

AMS 2000 subject classifications. Primary 62E20; Secondary 60F05

Keywords and phrases. Information dimension, local dimension, central limit theorem, U-statistics

unpleasant problem of having enough observations to effectively estimating measures of balls. Keller (1997) overcame the problem of insufficient information from the data for estimating measures of small balls by truncation with score functions. He investigated a new least square approach to information dimension estimation of an invariant distribution of a dynamical system. His method is computationally similar to the Grassberger-Procaccia algorithm for estimating the correlation dimension. He mentioned another estimator for the information dimension, based on ergodicity and local dimensions, but did not prove any result for it. The goal of this note is to derive asymptotic properties of this estimator including some results from simulation studies.

Let $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ be i.i.d. random vectors sampled from a distribution μ. We shall consider statistics of the form

$$T_n = \begin{cases} \frac{1}{n}\sum_{j=1}^{n} \log\left(\frac{1}{n-1}\sum_{\substack{i=1\\ i\neq j}}^{n} h(\mathbf{X}_i, \mathbf{X}_j)\right) & \text{if } \sum_{\substack{i=1\\ i\neq j}}^{n} h(\mathbf{X}_i, \mathbf{X}_j) > 0 \text{ for } j = 1, \dots, n \\ -\infty & \text{otherwise} \end{cases} \tag{1}$$

where $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ denotes a symmetric function $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$. The estimator for the information dimension will be a special case when setting

$$h(\mathbf{x}, \mathbf{y}) = \mathbb{1}\{||\mathbf{x} - \mathbf{y}|| \leq \epsilon\}.$$

Heuristically this means that T_n is the estimator of the functional

$$T(\mu) = \int \log \mu(B(\mathbf{x}, \epsilon))\mu(d\mathbf{x})$$

when inserting the empirical distribution (disregarding that $i \neq j$ is required in the second summation in T_n).

Given a symmetric kernel function $h : \mathbb{R}^{2d} \to \mathbb{R}$ and a distribution μ on $\mathbb{R}^d$ with $\iint |h(\mathbf{x}, \mathbf{y})|\mu(d\mathbf{x})\mu(d\mathbf{y}) \leq \infty$ define

$$\widetilde{h}_1(\mathbf{x}) = \int h(\mathbf{x}, \mathbf{y})\mu(d\mathbf{y}),$$

i.e. $\widetilde{h}_1(\mathbf{x})$ is the conditional distribution $E(h(\mathbf{Y}, \mathbf{X})|\mathbf{X} = \mathbf{x})$ where $\mathbf{X}$ and $\mathbf{Y}$ are independent and μ-distributed. We also set

$$h_2(\mathbf{x}, \mathbf{y}) = h(\mathbf{x}, \mathbf{y}) - \widetilde{h}_1(\mathbf{x}) - \widetilde{h}_1(\mathbf{y}) + \iint h(\mathbf{x}, \mathbf{y})\mu(d\mathbf{x})\mu(d\mathbf{y})$$

and $\theta(\mu) := \int \log \widetilde{h}_1(\mathbf{x})\mu(d\mathbf{x})$. We shall prove the following results for a sequence of i.i.d. random vectors $(\mathbf{X}_n)_{n\in\mathbf{N}}$ with distribution μ.

Theorem 1.1. *If*

$$P\{\tilde{h}_1(\mathbf{X}_1) \geq A\} = 1 \text{ for some constant } A > 0 \text{ and} \tag{2}$$

$$Eh^4(\mathbf{X}_1, \mathbf{X}_2) < \infty, \tag{3}$$

then $T_n \to \theta(\mu)$ *in probability provided* $\liminf_{n\to\infty} T_n > -\infty$ *a.s.*

Theorem 1.2. *Let* $\liminf_{n\to\infty} T_n > -\infty$ *a.s. and*

$$P(\tilde{h}_1(\mathbf{X}_1) \geq A) = 1 \quad \text{for some} \quad A > 0; \tag{4}$$

$$Eh^4(\mathbf{X}_1, \mathbf{X}_2) < \infty; \tag{5}$$

$$\sigma^2 = \text{Var}\left[\log(\widetilde{h}_1(\mathbf{X})) + \int \left(\frac{\widetilde{h}_1(\mathbf{X})}{\widetilde{h}_1(\mathbf{y})}\right)\mu(d\mathbf{y}) + \int \frac{h_2(\mathbf{X}, \mathbf{y})}{\widetilde{h}_1(\mathbf{y})}\mu(d\mathbf{y})\right] > 0. \tag{6}$$

Then $\sqrt{n}\sigma^{-1}(T_n - \theta(\mu))$ *is asymptotically normal with mean 0 and variance 1.*

We just mention two examples which were carried out by the second named author Min (2004). Details will appear elsewhere.

Example 1: If we denote by C the Cantor set then the two-dimensional Cantor distribution is the uniform distribution on the Cartesian product of $C \times C$. It is known that the information dimension σ_μ of this distribution is approximately equal to 1.2619 [see Cutler (1991, 1993)].

We have produced 100 simulations. In each simulation a sample of a size 5000 was randomly drawn from the two-dimensional Cantor distribution. $\varepsilon_k = 0.0021 + 0.0001 \cdot k$, $k = 0, \ldots, 8$ were chosen as a radii for balls $B(X_i, \varepsilon)$, for $i = 1, \ldots, 5000$. Points which did not have any neighbor in their ε_1-neighborhood have been discarded for the analysis. Results of the simulations show good agreement with the theoretical value.

Example 2: Similarly, consider a generalization of the 3-dimensional Cantor distribution on the Cartesian product $C \times C \times C$. First we describe a construction of such a measure by the following procedure.

Consider the unit cube I^3 in $\mathbb{R}^3$. Divide it into 27 non-overlapping cubes $I_1, \ldots, I_{27}$ of equal size. We number them beginning at the lower left vertice of some face counting the cubes on this face from left to right and down to up, then the cubes adjacent to it in the same fashion finishing with the cubes of the opposite face. Now define a vector of probabilities $\mathbf{p} = (p_1, \ldots, p_{27})$ whose entries are given by $p_1 = 0.8^3$, $p_3 = p_7 = p_{19} = 0.8^2 \cdot 0.2$, $p_9 = p_{21} = p_{25} = 0.8 \cdot 0.2^2$, $p_{27} = 0.2^3$ and $p_i = 0$ otherwise. Furthermore, we assign the probability p_i to I_i.
Repeat this process for each I_i to obtain a (unique) selfsimilar probability measure $\mu_{\mathbf{p}}$ defined on the Borel sets of I^3 and cubes $I_{i_1 \ldots i_n}$, $(n \geq 1)$, such that $\mu_{\mathbf{p}}(I_{i_1 \ldots i_n}) = p_{i_1} \cdot \ldots \cdot p_{i_n}$. It is known that its information dimension is approximately equal to 1.3665 [see Cutler (1991)]. If the nonzero entries of a vector of probabilities $\mathbf{p}$ are given by $p_1 = p_3 = p_7 = p_9 = p_{19} = p_{21} = p_{25} = p_{27} = 0.5^3$ then we obtain the three-dimensional Cantor distribution.
We have produced 100 simulations. In each simulation sample of size 5000 was randomly drawn from the generalized three-dimensional Cantor distribution. $\varepsilon_k = 0.030 + 0.001 \cdot k$, $k = 0, \ldots, 8$ were chosen as a radii for balls $B(X_i, \varepsilon)$, for $i = 1, \ldots, 5000$. Points which did not have any neighbor in their ε_1-neighborhood were not considered. Results of the simulations are also in good agreement with the theoretical result.

2 Consistency

We begin with a simple observation and a lemma.
Setting $h_1(\mathbf{x}) = \widetilde{h}_1(\mathbf{x}) - \iint h(\mathbf{x}, \mathbf{y})\mu(d\mathbf{x})\mu(\mathbf{y})$ and

$$\eta_{k,n} = \frac{1}{n-1} \sum_{i=1: i \neq k}^{n} h_1(\mathbf{X}_i) + \frac{1}{n-1} \sum_{i=1: i \neq k}^{n} h_2(\mathbf{X}_i, \mathbf{X}_k) \tag{7}$$

we can rewrite (1) (when the first case holds) as

$$T_n = \frac{1}{n} \sum_{k=1}^{n} \log \left(\tilde{h}_1(\mathbf{X}_k) + \eta_{k,n} \right). \tag{8}$$

Lemma 2.1. *The random variables $\eta_{k,n}$ are identically distributed w.r. to k for every fixed n and, moreover, if $Eh^4(\mathbf{X}_1, \mathbf{X}_2) < \infty$, then for every fixed k*

$$E\eta_{k,n}^4 = O\left(\frac{1}{n^2}\right)$$

and consequently $\eta_{k,n} \to 0$ completely as $n \to \infty$.

Proof. The first statement is obvious. The last statement follows immediately from the second one.

Let us prove the second statement. By the c_r-inequality, it follows that

$$\begin{aligned} E\eta_{j,n}^4 &\leq 8E\left(\frac{1}{n-1}\sum_{i=1:i\neq j}^{n} h_1(\mathbf{X}_i)\right)^4 + 8E\left(\frac{1}{n-1}\sum_{i=1:i\neq j}^{n} h_2(\mathbf{X}_i, \mathbf{X}_j)\right)^4 \\ &=: 8W_{r1} + 8W_{r2}. \end{aligned}$$

Consider W_{r1}. Using degeneracy of h_1 and the assumption, we have

$$\begin{aligned} W_{r1} &= \frac{1}{(n-1)^4}\sum_{i=1:i\neq j}^{n} Eh_1^4(\mathbf{X}_i) + \frac{3}{(n-1)^4}\sum_{i=1:i\neq j}^{n}\sum_{k=1:k\neq j}^{n} Eh_1^2(\mathbf{X}_i)Eh_1^2(\mathbf{X}_k) \\ &= O\left(\frac{1}{n^2}\right) . \end{aligned}$$

A similar argument yields a bound for W_{r2}

$$\begin{aligned} W_{r2} &= \frac{1}{(n-1)^4}\sum_{i=1:i\neq j}^{n} Eh_2^4(\mathbf{X}_i, \mathbf{X}_j) \\ &\quad + \frac{3}{(n-1)^4}\sum_{i=1:i\neq j}^{n}\sum_{k=1:k\neq i:k\neq j}^{n} Eh_2^2(\mathbf{X}_i, \mathbf{X}_j)Eh_2^2(\mathbf{X}_k, \mathbf{X}_j) = O\left(\frac{1}{n^2}\right). \end{aligned}$$

The claim follows from these relations. □

Remark 2.1. *The complete convergence can be proved assuming $Eh^2(\mathbf{X}_1, \mathbf{X}_2) < \infty$. In this case one needs to use results of Hsu and Robbins* (1947) *and of Dehling* (1989).

Proof of Theorem 1.1. First note that $0 < E\tilde{h}_1(\mathbf{X}_1) \leq E|h(\mathbf{X}_1, \mathbf{X}_2)|$ and $|\log u| \leq \max\{\log A, u\}$ for $u \geq A > 0$ which shows that

$$E|\log \tilde{h}_1(\mathbf{X}_1)| \leq \infty \ . \tag{9}$$

Fix $\varepsilon > 0$.

The representation (8) of T_n and a simple argument now imply that

$$\begin{aligned} P(|T_n - \theta(\mu)| > \varepsilon) &= P\left(\left|\frac{1}{n}\sum_{j=1}^{n}(\log(\tilde{h}_1(\mathbf{X}_j)+\eta_{j,n}) - \theta(\mu))\right| > \varepsilon\right) \\ = & \ P\left(\frac{1}{n}\left|\sum_{j=1}^{n}(\log(\tilde{h}_1(\mathbf{X}_j)+\eta_{j,n}) - \theta(\mu))\right| > \varepsilon;\ \max_{k=1,\ldots,n}|\eta_{k,n}| > \frac{A}{2}\right) \\ + & \ P\left(\frac{1}{n}\left|\sum_{j=1}^{n}(\log(\tilde{h}_1(\mathbf{X}_j)+\eta_{j,n}) - \theta(\mu))\right| > \varepsilon;\ \max_{k=1,\ldots,n}|\eta_{k,n}| \le \frac{A}{2}\right) \\ = & \ W_{g1} + W_{g2}. \end{aligned} \tag{10}$$

By Lemma 2.1, it follows that

$$\begin{aligned} W_{g1} &\le P\left(\max_{k=1,\ldots,n}|\eta_{k,n}| > \frac{A}{2}\right) \le \sum_{k=1}^{n} P\left(|\eta_{k,n}| > \frac{A}{2}\right) \\ &= nP\left(|\eta_{1,n}| > \frac{A}{2}\right) \to 0, \quad \text{as} \quad n \to \infty. \end{aligned} \tag{11}$$

Consider W_{g2}. Using the Taylor expansion for $\log(b+x)$ with the remainder term in Lagrange form, we find

$$\begin{aligned} W_{g2} &= P\left(\left|\frac{1}{n}\sum_{j=1}^{n}\left[\log\tilde{h}_1(\mathbf{X}_j) - \theta(\mu) + \frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j)+\theta_{j,n}\eta_{j,n}}\right]\right| \right. \\ &\qquad\qquad \left. \mathbb{1}\left\{\max_{k=1,\ldots,n}|\eta_{k,n}| \le \frac{A}{2}\right\} > \varepsilon\right) \\ &\le P\left\{\left|\frac{1}{n}\sum_{j=1}^{n}\left[\log\tilde{h}_1(\mathbf{X}_j) - \theta(\mu)\right]\right| \mathbb{1}\left\{\max_{k=1,\ldots,n}|\eta_{k,n}| \le \frac{A}{2}\right\} > \frac{\varepsilon}{2}\right\} \\ &+ P\left(\left|\frac{1}{n}\sum_{j=1}^{n}\frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j)+\theta_{j,n}\eta_{j,n}}\right| \mathbb{1}\left\{\max_{k=1,\ldots,n}|\eta_{k,n}| \le \frac{A}{2}\right\} > \frac{\varepsilon}{2}\right) \\ &= W_{h1} + W_{h2}, \end{aligned} \tag{12}$$

where $\theta_{j,n}$, $j = 1, \ldots, n$ are $(0,1)$-valued random variables depending on $\tilde{h}_1(\mathbf{X}_j)$ and $\eta_{j,n}$.

From (9) and the law of large numbers for $\{\log\tilde{h}_1(\mathbf{X}_j)\}_{j\in\mathbb{N}}$, we deduce that

$$W_{h1} \le P\left(\left|\frac{1}{n}\sum_{j=1}^{n}\left[\log\tilde{h}_1(\mathbf{X}_j) - \theta(\mu)\right]\right| > \frac{\varepsilon}{2}\right) \to 0 \quad \text{as} \quad n \to \infty. \tag{13}$$

Using Chebychev's and Cauchy-Schwarz' inequalities, W_{h2} is estimated by

$$\begin{aligned} W_{h2} &\le \frac{4}{\varepsilon^2} E\left(\frac{1}{n}\sum_{j=1}^{n} \frac{\eta_{j,n}\mathbb{1}\left(\max_{k=1,\ldots,n}|\eta_{k,n}| \le \frac{A}{2}\right)}{\tilde{h}_1(\mathbf{X}_j)+\theta_{j,n}\eta_{j,n}}\right)^2 \\ &\le \frac{4}{\varepsilon^2 n}\sum_{j=1}^{n} E\left(\frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j)+\theta_{j,n}\eta_{j,n}}\right)^2 \mathbb{1}\left(\max_{k=1,\ldots,n}|\eta_{k,n}| < \frac{A}{2}\right). \end{aligned}$$

Furthermore, note that if $|\eta_{jn}| \le A/2$, then by (2)

$$\left(\frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j)+\theta_{j,n}\eta_{j,n}}\right)^2 \le \frac{4}{A^2}\eta_{j,n}^2 \qquad \mu - a.s. \tag{14}$$

and hence,

$$W_{h2} \le \frac{16}{A^2\varepsilon^2 n}\sum_{j=1}^{n} E\eta_{j,n}^2 = \frac{16}{A^2\varepsilon^2}\left(\frac{Eh_1^2(\mathbf{X}_1)}{n-1} + \frac{Eh_2^2(\mathbf{X}_1,\mathbf{X}_2)}{n-1}\right) \to 0. \tag{15}$$

The last equality holds since

$$\begin{aligned} E\eta_{j,n}^2 &= E\left(\frac{1}{n-1}\sum_{i=1:i\neq j}^{n} h_1(\mathbf{X}_i) + \frac{1}{n-1}\sum_{i=1:i\neq j}^{n} h_2(\mathbf{X}_i,\mathbf{X}_j)\right)^2 \\ &= \frac{Eh_1^2(\mathbf{X}_1)}{n-1} + \frac{Eh_2^2(\mathbf{X}_1,\mathbf{X}_2)}{n-1} \quad \text{for any} \quad j = 1, 2, \ldots, n. \end{aligned} \tag{16}$$

Inserting (13) and (15) into (12), (11) and (12) into (10) proves the theorem. □

3 Asymptotic distribution

In this section, we prove the asymptotic normality of T_n. We begin introducing some further notations:

$$A_1 = E\left(\frac{1}{\tilde{h}_1(\mathbf{X}_1)}\right); \tag{17}$$

$$\Phi(\mathbf{x},\mathbf{y}) = \frac{1}{2}\left(\frac{1}{\tilde{h}_1(\mathbf{x})}h_2(\mathbf{y},\mathbf{x}) + \frac{1}{\tilde{h}_1(\mathbf{y})}h_2(\mathbf{x},\mathbf{y})\right); \tag{18}$$

$$\psi(\mathbf{x}) = E(\Phi(\mathbf{X}_1,\mathbf{X}_2)|\mathbf{X}_1 = \mathbf{x}) - E\Phi(\mathbf{X}_1,\mathbf{X}_2); \tag{19}$$

$$Z_j = \log(\tilde{h}_1(\mathbf{X}_j)) - \theta(\mu) + A_1 h_1(\mathbf{X}_j) + 2\psi(\mathbf{X}_j); \tag{20}$$

Remark 3.1. *Note that, the random variables Z_j can also be written in the form*

$$Z_j = \log(\tilde{h}_1(\mathbf{X}_j)) - \theta(\mu) + \int \frac{h(\mathbf{X}_j, \mathbf{y})}{\tilde{h}_1(\mathbf{y})} \mu(\mathbf{dy}) - 1 \tag{21}$$

and that $\sigma^2 = \mathrm{Var}(Z_j)$. We shall make use of this in the next section to construct a consistent estimator for σ^2.

Proof of Theorem 1.2. First we give some simple consequences of the assumptions (4) and (5):

$$E\left|\frac{h_1(\mathbf{X}_1)}{\tilde{h}_1(\mathbf{X}_1)}\right| < \infty \quad ; \quad \mathrm{Var}\left(\frac{1}{\tilde{h}_1(\mathbf{X}_1)}\right) < \infty \quad ; \quad E\Phi^2(\mathbf{X}_1, \mathbf{X}_2) < \infty. \tag{22}$$

As in the beginning of the proof of Theorem 1.1 one can show that

$$\theta(\mu) = E\log(\tilde{h}_1(\mathbf{X}_1)) < \infty \quad \text{and} \quad \sigma^2 < \infty,$$

and consequently, σ^2 is well defined under the assumptions of the theorem.
Consider $\sqrt{n}\sigma^{-1}(T_n - \theta(\mu))$. As in the proof of Theorem 1.1 we use Taylor expansion, but we need one additional term in the expansion of T_n, namely

$$\begin{aligned}
\sqrt{n}\sigma^{-1}(T_n - \theta(\mu)) &= \frac{1}{\sqrt{n}\sigma} \sum_{j=1}^{n} [\log(\tilde{h}_1(\mathbf{X}_j) + \eta_{j,n}) - \theta(\mu)] \\
&= \frac{1}{\sqrt{n}\sigma} \sum_{j=1}^{n} \left[\log(\tilde{h}_1(\mathbf{X}_j)) - \theta(\mu) + \frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j)}\right] \\
&- \frac{1}{\sqrt{n}\sigma} \sum_{j=1}^{n} \frac{\eta_{j,n}^2}{\left(\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n}\right)^2} \\
&= W_1 - W_2,
\end{aligned} \tag{23}$$

where $\theta_{j,n}$, $j = 1, \ldots, n$ are $(0,1)$-valued random variables depending on $\tilde{h}_1(\mathbf{X}_j)$ and $\eta_{j,n}$.
The first step in the proof is to show that

$$W_2 \to 0 \quad \text{in probability.} \tag{24}$$

Write

$$
\begin{aligned}
W_2 &= P\left(\frac{1}{\sqrt{n}\sigma}\sum_{j=1}^{n}\frac{\eta_{j,n}^2}{\left(\tilde{h}_1(\mathbf{X}_j)+\theta_{j,n}\eta_{j,n}\right)^2} > \varepsilon\ ;\ \max_{k=1,\ldots,n}|\eta_{k,n}| \le \frac{A}{2}\right) \\
&+ P\left(\frac{1}{\sqrt{n}\sigma}\sum_{j=1}^{n}\frac{\eta_{j,n}^2}{\left(\tilde{h}_1(\mathbf{X}_j)+\theta_{j,n}\eta_{j,n}\right)^2} > \varepsilon\ ;\ \max_{k=1,\ldots,n}|\eta_{k,n}| > \frac{A}{2}\right) \\
&= W_{21} + W_{22}.
\end{aligned}
$$

By virtue of (14), (16), (4) and Chebyshev's inequality, we have

$$
\begin{aligned}
W_{21} &\le P\left(\frac{1}{\sqrt{n}\sigma}\sum_{j=1}^{n}\frac{4}{A^2}\eta_{j,n}^2 > \varepsilon\ ;\ \max_{k=1,\ldots,n}|\eta_{k,n}| \le \frac{A}{2}\right) \\
&\le \frac{4}{\sqrt{n}\sigma\varepsilon A^2}\sum_{j=1}^{n}E\eta_{j,n}^2 = \frac{4n}{\sqrt{n}\sigma\varepsilon A^2}E\eta_{1,n}^2 \to 0.
\end{aligned}
$$

By Lemma 2.1 it follows that

$$
W_{22} \le P\left(\max_{k=1,\ldots,n}|\eta_{k,n}| > \frac{A}{2}\right) \le nP\left(|\eta_{1,n}| > \frac{A}{2}\right) \to 0.
$$

Thus, (24) follows from these relations.

The second step in the proof is to establish the asymptotic equivalence of the distributions of W_1 and S_n, where

$$
S_n = \frac{1}{\sqrt{n}\sigma}\sum_{j=1}^{n}Z_j,
$$

and where Z_j and σ are defined in (20) and (6), respectively. To verify this, it is enough to show that

$$
E(W_1 - S_n)^2 \to 0 \quad \text{as} \quad n \to \infty. \tag{25}
$$

Note that

$$
\begin{aligned}
\frac{1}{\sqrt{n}\sigma}\sum_{j=1}^{n}\frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j)} &= \frac{1}{\sqrt{n}\sigma(n-1)}\sum_{j=1}^{n}\frac{1}{\tilde{h}_1(\mathbf{X}_j)}\sum_{i=1:i\neq j}^{n}h_1(\mathbf{X}_i) \\
&+ \frac{1}{\sqrt{n}\sigma(n-1)}\sum_{j=1}^{n}\frac{1}{\tilde{h}_1(\mathbf{X}_j)}\sum_{i=1:i\neq j}^{n}h_2(\mathbf{X}_i,\mathbf{X}_j) \\
&= SS_1 + SS_2.
\end{aligned}
$$

Define the following random variables

$$SS = \frac{A_1}{\sqrt{n}\sigma}\sum_{i=1}^{n} h_1(\mathbf{X}_i) \text{ and } \frac{2\sqrt{n}}{\sigma}U_1 = \frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n} 2\psi(\mathbf{X}_i),$$

where A_1 and $\psi(x)$ are defined in (17) and (19), respectively.

A simple calculation yields

$$E(W_1 - S_n)^2 = E\left(SS_1 + SS_2 - SS - \frac{2\sqrt{n}}{\sigma}U_1\right)^2,$$

and it is therefore sufficient to show that

$$E\left(SS_1 - SS\right)^2 \to 0 \tag{26}$$

and

$$E\left(SS_2 - \frac{2\sqrt{n}}{\sigma}U_1\right)^2 \to 0. \tag{27}$$

Rewrite SS_1 in the form

$$SS_1 = \frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n} h_1(\mathbf{X}_i)\frac{1}{n-1}\sum_{j=1:j\neq i}^{n}\frac{1}{\tilde{h}_1(\mathbf{X}_j)}.$$

It follows that

$$\begin{aligned}
E(SS - SS_1)^2 &= \frac{1}{n\sigma^2}\sum_{i=1}^{n} E\left(h_1(\mathbf{X}_i)\frac{1}{n-1}\sum_{j=1:j\neq i}^{n}\left[\frac{1}{\tilde{h}_1(\mathbf{X}_j)} - A_1\right]\right)^2 \\
&+ \frac{1}{n\sigma^2}E\left(\sum_{i=1}^{n}\sum_{m=1:m\neq i}^{n} h_1(\mathbf{X}_i)\frac{1}{n-1}\left[\sum_{j=1:j\neq i}^{n}\left(\frac{1}{\tilde{h}_1(\mathbf{X}_j)} - A_1\right)\right]\right. \\
&\times \left. h_1(\mathbf{X}_m)\frac{1}{n-1}\left[\sum_{l=1:l\neq m}^{n}\left(\frac{1}{\tilde{h}_1(\mathbf{X}_l)} - A_1\right)\right]\right) \\
&= QQ_1 + QQ_2.
\end{aligned} \tag{28}$$

By (5) and (22) we obtain

$$\begin{aligned}
QQ_1 &= \frac{1}{n\sigma^2}\sum_{i=1}^{n} E\left(h_1(\mathbf{X}_i)\frac{1}{n-1}\sum_{j=1:j\neq i}^{n}\left[\frac{1}{\tilde{h}_1(\mathbf{X}_j)} - A_1\right]\right)^2 \\
&= \frac{1}{\sigma^2}\frac{1}{(n-1)^2}Eh_1^2(\mathbf{X}_1)E\left(\sum_{j=2}^{n}\left[\frac{1}{\tilde{h}_1(\mathbf{X}_j)} - A_1\right]\right)^2 \\
&= \frac{1}{\sigma^2}\frac{1}{(n-1)}Eh_1^2(\mathbf{X}_1)E\left(\frac{1}{\tilde{h}_1(\mathbf{X}_2)} - A_1\right)^2 \to 0.
\end{aligned} \tag{29}$$

Using (22) and the degeneracy of the function $h_1(\mathbf{x})$, we find

$$
\begin{aligned}
QQ_2 &= \frac{1}{n\sigma^2}\sum_{i=1}^{n}\sum_{m=1:m\neq i}^{n} E\left(h_1(\mathbf{X}_i)\frac{1}{n-1}\left[\sum_{j=1:j\neq i}^{n}\left(\frac{1}{\tilde{h}_1(\mathbf{X}_j)}-A_1\right)\right]\right. \\
&\times \left. h_1(\mathbf{X}_m)\frac{1}{n-1}\left[\sum_{l=1:l\neq m}^{n}\left(\frac{1}{\tilde{h}_1(\mathbf{X}_l)}-A_1\right)\right]\right) \\
&= \frac{1}{n(n-1)^2\sigma^2}\sum_{i=1}^{n}\sum_{m=1:m\neq i}^{n} E\left(h_1(\mathbf{X}_m)\left[\sum_{j=1:j\neq i}^{n}\left(\frac{1}{\tilde{h}_1(\mathbf{X}_j)}-A_1\right)\right]\right. \\
&\times \left. \mathbf{E_i}\, h_1(\mathbf{X}_i)\left[\sum_{l=1:l\neq m}^{n}\left(\frac{1}{\tilde{h}_1(\mathbf{X}_l)}-A_1\right)\right]\right) \\
&= \frac{1}{n(n-1)^2\sigma^2}\sum_{i=1}^{n}\sum_{m=1:m\neq i}^{n} E\left(h_1(\mathbf{X}_m)\left[\sum_{j=1:j\neq i}^{n}\left(\frac{1}{\tilde{h}_1(\mathbf{X}_j)}-A_1\right)\right]\mathbf{E_i}\,\frac{h_1(\mathbf{X}_i)}{\tilde{h}_1(\mathbf{X}_i)}\right) \\
&= \frac{1}{n(n-1)^2\sigma^2}\sum_{i=1}^{n}\sum_{m=1:m\neq i}^{n} E\frac{h_1(\mathbf{X}_m)}{\tilde{h}_1(\mathbf{X}_m)}E\,\frac{h_1(\mathbf{X}_i)}{\tilde{h}_1(\mathbf{X}_i)} \\
&= \frac{1}{(n-1)\sigma^2}E\frac{h_1(\mathbf{X}_1)}{\tilde{h}_1(\mathbf{X}_1)}E\,\frac{h_1(\mathbf{X}_2)}{\tilde{h}_1(\mathbf{X}_2)} \to 0, \qquad (30)
\end{aligned}
$$

where $\mathbf{E_i}$ denotes the conditional expectation with respect to $\mathbf{X}_i$.

(26) follows now from (28),(29) and (30).

In order to establish (27), we rewrite SS_2 in the form

$$
\begin{aligned}
SS_2 &= \frac{1}{\sqrt{n}(n-1)\sigma}\sum_{1\leq i\neq j\leq n}\frac{1}{2}\left(\frac{1}{\tilde{h}_1(\mathbf{X}_j)}h_2(\mathbf{X}_i,\mathbf{X}_j)+\frac{1}{\tilde{h}_1(\mathbf{X}_i)}h_2(\mathbf{X}_j,\mathbf{X}_i)\right) \\
&= \frac{1}{\sqrt{n}(n-1)\sigma}\sum_{1\leq i\neq j\leq n}\Phi(\mathbf{X}_i,\mathbf{X}_j),
\end{aligned}
$$

where $\Phi(x,y)$ has been defined in (18).

Now we see that $\sigma n^{-1/2}SS_2$ is a U-statistic with kernel $\Phi(\mathbf{x},\mathbf{y})$.

Noticing that $E\Phi(\mathbf{X}_1,\mathbf{X}_2)=0$, the relation (27) follows from (22) and Hoeffding's decomposition of U-statistics.

Finally, the theorem follows applying the central limit theorem to S_n, since $\sigma^2>0$.

□

4 Consistent estimator of the variance

The random variables

$$Z_j = \log(\tilde{h}_1(\mathbf{X}_j)) - \theta(\mu) + \int \frac{h(\mathbf{X}_j, \mathbf{y})}{\tilde{h}_1(\mathbf{y})} \mu(\mathbf{dy}) - 1$$

can be used to estimate $\sigma^2 = \mathrm{Var}(Z_1)$, if, for example, $h(\mathbf{x}, \mathbf{y}) = \mathbb{1}\{\|\mathbf{x} - \mathbf{y}\| \le \varepsilon\}$. First we define random variables $\hat{Z}_j$ by replacing $\mu(\mathbf{dy})$ by $\hat{\mu}(\mathbf{dy})$ and the expectation by the sample mean in (21), i.e.

$$\hat{Z}_j = \log(\hat{\mu}(B(\mathbf{X}_j, \varepsilon)) - \hat{\theta}(\mu) + \frac{1}{n-1} \sum_{i=1:i \neq j}^{n} \frac{\mathbb{1}\{\|\mathbf{X}_j - \mathbf{X}_i\| \le \varepsilon\}}{\hat{\mu}(B(\mathbf{X}_i, \varepsilon))} - 1,$$

where

$$\hat{\mu}(B(\mathbf{X}_j, \varepsilon) = \frac{1}{n-1} \sum_{i=1:i \neq j}^{n} \mathbb{1}\{\|\mathbf{X}_j - \mathbf{X}_i\| \le \varepsilon\}$$

and

$$\hat{\theta}(\mu) = \frac{1}{n} \sum_{j=1}^{n} \log\left(\hat{\mu}(B(\mathbf{X}_j, \varepsilon))\right).$$

Then we take the sample second moment of $\{\hat{Z}_j\}_{j=1}^n$ which we denote by $\hat{\sigma}^2$ as an estimator for σ^2, since $EZ_j = 0$ for $j = 1, \ldots, n$. The consistency of $\hat{\sigma}^2$ will be proved in the next proposition. Since we are going to apply Theorem 1.1 to $h(\mathbf{x}, \mathbf{y}) = \mathbb{1}\{\|\mathbf{x} - \mathbf{y}\| \le \varepsilon\}$, we assume further that

$$P(0 \le h(\mathbf{X}_1, \mathbf{X}_2) \le C_1) = 1. \tag{31}$$

Proposition 4.1. *Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be i.i.d. $\mathbb{R}^d$-valued random variables with a probability distribution μ and T_n be the statistic defined in* (1) *such that* $\liminf_n T_n > -\infty$ *μ-a.s. Assume that* (2) *and* (31) *hold. Then the following statistic*

$$\begin{aligned}
\hat{\sigma}_n^2 &= \frac{1}{n} \sum_{j=1}^{n} \log^2 \left(\frac{1}{n-1} \sum_{i=1:i \neq j}^{n} h(\mathbf{X}_i, \mathbf{X}_j) \right) - T_n^2 \\
&+ \frac{1}{n^3} \sum_{s=1}^{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{h(\mathbf{X}_s, \mathbf{X}_i)}{\frac{1}{n-1} \sum_{u \neq i} h(\mathbf{X}_u, \mathbf{X}_i)} \cdot \frac{h(\mathbf{X}_s, \mathbf{X}_j)}{\frac{1}{n-1} \sum_{v \neq j} h(\mathbf{X}_v, \mathbf{X}_j)} + 1 \\
&- \frac{2}{n^2} \sum_{s=1}^{n} \sum_{i=1}^{n} \frac{h(\mathbf{X}_s, \mathbf{X}_i)}{\frac{1}{n-1} \sum_{u \neq i} h(\mathbf{X}_u, \mathbf{X}_i)} - 2T_n \\
&+ \frac{2}{n^2} \sum_{s=1}^{n} \sum_{j=1}^{n} \left[\log \left(\frac{1}{n-1} \sum_{u \neq s} h(\mathbf{X}_u, \mathbf{X}_s) \right) \right] \cdot \frac{h(\mathbf{X}_s, \mathbf{X}_j)}{\frac{1}{n-1} \sum_{v \neq j} h(\mathbf{X}_v, \mathbf{X}_j)}
\end{aligned} \tag{32}$$

is a consistent estimator for σ^2 *defined in* (6).

Proof. The proof of the proposition is straight forward, so we only sketch it. A simple calculation shows that

$$\begin{aligned}
\sigma^2 &= \mathrm{Var}(Z_1) = EZ_1^2 \\
&= E\left(\log(\tilde{h}_1(\mathbf{X}_1)) - \theta(\mu)\right)^2 + E\left(\int \frac{h(\mathbf{X}_1,\mathbf{y})}{\tilde{h}_1(\mathbf{y})}\mu(\mathbf{dy}) - 1\right)^2 \\
&+ 2E\left(\log(\tilde{h}_1(\mathbf{X}_1)) - \theta(\mu)\right)\left(\int \frac{h(\mathbf{X}_1,\mathbf{y})}{\tilde{h}_1(\mathbf{y})}\mu(\mathbf{dy}) - 1\right) \\
&= E_1 + E_2 + 2E_3
\end{aligned}$$

whence it is sufficient to find consistent estimators for E_1, E_2 and E_3.
Consider E_1. It is enough to find a consistent estimator for $E\log^2(\tilde{h}_1(\mathbf{X}_1))$ since $\theta(\mu)^2$ can be consistently estimated by T_n^2 (this follows from Slutzky's Theorem and Theorem 1.1). As before we can write

$$\begin{aligned}
\hat{E}_1 &:= \frac{1}{n}\sum_{j=1}^{n}\left(\log\tilde{h}_1(\mathbf{X}_j) + \frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n}}\right)^2 \\
&= \frac{1}{n}\sum_{j=1}^{n}\log^2\tilde{h}_1(\mathbf{X}_j) \\
&+ \frac{2}{n}\sum_{j=1}^{n}\log\tilde{h}_1(\mathbf{X}_j)\cdot\frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n}} \\
&+ \frac{1}{n}\sum_{j=1}^{n}\left(\frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n}}\right)^2 \\
&= S_{1n} + 2S_{2n} + S_{3n}
\end{aligned}$$

where $\eta_{j,n}$ are defined in (7) and $\theta_{j,n}$, $j = 1,\ldots,n$ are $(0,1)$-valued random variables depending on $\eta_{j,n}$ and $\tilde{h}_1(\mathbf{X}_j)$.
Clearly,

$$S_{1n} \xrightarrow{P} E\log^2\tilde{h}_1(\mathbf{X}_1) \qquad \text{as } n \to \infty.$$

By the conditions imposed on $\tilde{h}_1$ and h, we find that $|\log\tilde{h}_1(X_1)| < C_2$ μ-a.s., where

$C_2 = \max\{|\log A|, |\log C_1|\}$. Furthermore it is easy to see that

$$
\begin{aligned}
P(|S_{2n}| > \varepsilon) &\le \sum_{j=1}^{n} P\left(\left|\frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n}}\right| > \frac{\varepsilon}{C_2}\ ;\ |\eta_{j,n}| \le \frac{A}{2}\right) \\
&+ \sum_{j=1}^{n} P\left(\left|\frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n}}\right| > \frac{\varepsilon}{C_2}\ ;\ |\eta_{j,n}| > \frac{A}{2}\right) \\
&= P_{2an} + P_{2bn}.
\end{aligned}
$$

By Lemma 2.1 it follows that each of the terms P_{2an} and P_{2bn} tends to zero. Therefore S_{2n} tends to zero in probability. Likewise one shows that S_{2n} tends to zero, whence $\hat{E}_1$ is a consistent estimator of E_1.

Simple calculation yields that

$$
\begin{aligned}
E_2 &= E\left(\int \frac{h(\mathbf{X}_1, \mathbf{y})}{\tilde{h}_1(\mathbf{y})}\mu(\mathbf{dy}) - 1\right)^2 \\
&= E\left(\int \frac{h(\mathbf{X}_1, \mathbf{y})}{\tilde{h}_1(\mathbf{y})}\mu(\mathbf{dy})\right)^2 + 1 - 2E\left(\int \frac{h(\mathbf{X}_1, \mathbf{y})}{\tilde{h}_1(\mathbf{y})}\mu(\mathbf{dy})\right) \\
&= E_{2a} + 1 - 2E_{2b}.
\end{aligned}
$$

The natural estimator for the E_{2a} is

$$
\begin{aligned}
\hat{E}_{2a} &= \frac{1}{n^3}\sum_{s=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{h(\mathbf{X}_s, \mathbf{X}_i)}{\frac{1}{n-1}\sum_{u\neq i} h(\mathbf{X}_u, \mathbf{X}_i)} \cdot \frac{h(\mathbf{X}_s, \mathbf{X}_j)}{\frac{1}{n-1}\sum_{v\neq j} h(\mathbf{X}_v, \mathbf{X}_j)} \\
&= \frac{1}{n^3}\sum_{s=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{h(\mathbf{X}_s, \mathbf{X}_i)}{\tilde{h}_1(\mathbf{X}_i)} \cdot \frac{h(\mathbf{X}_s, \mathbf{X}_j)}{\tilde{h}_1(\mathbf{X}_j)} \\
&+ \frac{1}{n^3}\sum_{s=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{h(\mathbf{X}_s, \mathbf{X}_i)\eta_{i,n}}{(\tilde{h}_1(\mathbf{X}_i) + \theta_{i,n}\eta_{i,n})^2} \cdot \frac{h(\mathbf{X}_s, \mathbf{X}_j)\eta_{j,n}}{(\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n})^2} \\
&- \frac{1}{n^3}\sum_{s=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{h(\mathbf{X}_s, \mathbf{X}_i)}{\tilde{h}_1(\mathbf{X}_i)} \cdot \frac{h(\mathbf{X}_s, \mathbf{X}_j)\eta_{j,n}}{(\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n})^2} \\
&- \frac{1}{n^3}\sum_{s=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{h(\mathbf{X}_s, \mathbf{X}_i)\eta_{i,n}}{(\tilde{h}_1(\mathbf{X}_i) + \theta_{i,n}\eta_{i,n})^2} \cdot \frac{h(\mathbf{X}_s, \mathbf{X}_j)}{\tilde{h}_1(\mathbf{X}_j)} \\
&= \hat{E}_{2a1} + \hat{E}_{2a2} + \hat{E}_{2a3} + \hat{E}_{2a4},
\end{aligned}
$$

where $\theta_{j,n}$, $j = 1, \ldots, n$ are $(0,1)$-valued random variables depending on $\tilde{h}_1(\mathbf{X}_j)$ and $\eta_{j,n}$.

It is not difficult to see that

$$\hat{E}_{2a1} \xrightarrow{P} E_{2a} \qquad \text{as } n \to \infty$$

by the law of large numbers for U-statistics and (similarly as before) that

$$\hat{E}_{2a2}, \quad \hat{E}_{2a3} \ \text{ and } \ \hat{E}_{2a4} \xrightarrow{P} 0 \qquad \text{as } n \to \infty.$$

It follows that $\hat{E}_{2a}$ is a consistent estimator of E_{2a}.

Analogously, one can show that

$$\hat{E}_{2b} = \frac{1}{n^2} \sum_{i,s=1}^{n} \frac{h(\mathbf{X}_s, \mathbf{X}_i)}{(n-1)^{-1} \sum_{u \neq i} h(\mathbf{X}_u, \mathbf{X}_i)}$$

is consistent for E_{2b}. It follows that $\hat{E}_{2a} + 1 - 2\hat{E}_{2b}$ is a consistent estimator for E_2.

Note that E_3 can be written in the form

$$\begin{aligned} E_3 &= E\left(\log(\tilde{h}_1(\mathbf{X}_1))\right) \cdot \int \frac{h(\mathbf{X}_1, \mathbf{y})}{\tilde{h}_1(\mathbf{y})} \mu(\mathbf{dy}) - \theta(\mu) \\ &= E_{3a} - \theta(\mu). \end{aligned}$$

By Theorem 1.1, $\theta(\mu)$ can be consistently estimated by T_n and hence, it remains only to find a consistent estimator for E_{3a}. The same technique, used for proving consistency of $\hat{E}_1$ and $\hat{E}_{2a}$, will show that the estimator

$$\hat{E}_{3a} = \frac{1}{n^2} \sum_{s=1}^{n} \sum_{j=1}^{n} \left[\log\left(\frac{1}{n-1} \sum_{u \neq s} h(\mathbf{X}_u, \mathbf{X}_s) \right) \right] \cdot \frac{h(\mathbf{X}_s, \mathbf{X}_j)}{\frac{1}{n-1} \sum_{v \neq j} h(\mathbf{X}_v, \mathbf{X}_j)} .$$

is a consistent estimator for E_{3a}.

Finally, note that the following equality for $\hat{\sigma}_n^2$ defined in (32) holds

$$\hat{\sigma}_n^2 = \hat{E}_1 - T_n^2 + \hat{E}_{2a} + 1 - 2\hat{E}_{2b} + 2\hat{E}_3 - 2T_n$$

and therefore, $\hat{\sigma}_n^2$ consistently estimates σ^2. □

REFERENCES

Cutler, C.D. (1991). Some results on the behavior and estimation of the fractal dimension of distributions on attractors. *J. Statist. Phys.* **82** 651–707.

CUTLER, C.D. (1993). A review of the theory and estimation of fractal dimension. In *Nonlinear Time Ser. Chaos* **1** (H. Tong, ed.) 1–107. World Sci. Publishing, Singapore.

CUTLER, C.D. and DAWSON, D.A. (1989). Estimation of dimension for spatially-distributed data and related limit theorems. *J. Multivariate Anal.* **28** 115–148.

CUTLER, C.D. and DAWSON, D.A. (1990). Nearest neighbor analysis of a family of fractal distributions. *Ann. Probab.* **18** 256–271.

DEHLING, H. (1989). Complete convergence of triangular arrays and the law of the iterated logarithm for U-statistics. *Statist. Probab. Lett.* **7** 319-321.

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

GRASSBERGER, P. and PROCACCIA, I. (1983). Measuring the strangeness of strange attractors. *Phys. D* **9** 189–208.

HSU P. L. and ROBBINS H. (1947). Complete convergence and the law of large numbers. *Proc. Natl. Acad. Sci. USA* **33** 229-240.

KELLER, G. (1997). A new estimator for the information dimension with standard errors and confidence intervals, *Stochastic Processes Appl.* **71** 187–206.

Min, A. (2004), Limit theorems for statistical functionals with applications to dimension estimation. *Ph.D. Dissertation*, University of Göttingen.

Georg-August-Universität Göttingen
Institut für Mathematische Stochastik
Maschmühlenweg 8-10
37073 Göttingen
Germany
E-mail: denker@math.uni-goettingen.de

Technische Universität München
Zentrum Mathematik
Boltzmannstr. 3
85747 Garching
Germany
E-mail: aleksmin@ma.tum.de