Journal of Statistical Physics, Vol. 78, Nos. 3/4, 1995

A Monte Carlo Method for Estimating the Correlation Exponent

T. Mikosch[1] and Qiang Wang[2]

Received March 22, 1993; final July 25, 1994

We propose a Monte Carlo method for estimating the correlation exponent of a stationary ergodic sequence. The estimator can be considered as a bootstrap version of the classical Hill estimator. A simulation study shows that the method yields reasonable estimates.

KEY WORDS: Correlation exponent; correlation dimension; Hill estimator; Monte Carlo method; bootstrap.

1. INTRODUCTION

Over the past decade there has been much interest in the asymptotic behavior of dynamical systems, in particular in detecting "chaotic" behavior of these systems and testing for the existence of "strange" or "fractal" attracting sets. Dimensions of different type such as the Hausdorff, Renýi, or correlation dimensions are appropriate quantitative measures.

The present paper is entirely devoted to the statistical estimation of the correlation exponent. It is our aim to propose a Monte Carlo method for estimating it.

We start with a stationary ergodic sequence $\mathbf{X}, \mathbf{X}_1, \mathbf{X}_2, \ldots$ of d-dimensional random vectors with marginal distribution $\mathbf{F}$ which is supposed to be continuous and which coincides with the invariant probability measure of the sequence. We will suppose in the sequel that $d \geqslant 1$ is fixed. For example, the random variables $\mathbf{X}_i$ can be thought of as the vectors $\mathbf{X}_i = (X_i, \ldots, X_{i+d-1})$, $i = 1, 2, \ldots$, which are constructed from a one-dimensional

[1] University of Groningen, Faculty of Mathematics and Physics, P.O. Box 800, NL-9700 AV Groningen, The Netherlands.

[2] National Hydrology Institute, Hydrometeorological Processes Division, Saskatoon, Saskatchewan S7N 3H5, Canada.

0022-4715/95/0200-0799$07.50/0

stationary ergodic sequence (X_i), i.e., the $\mathbf{X}_i$ are obtained by embedding the sequence (X_i) in a d-dimensional phase space. In our simulation studies we will construct the sequence of the $\mathbf{X}_i$ in this way.

Let $\mathbf{Y}$ and $\mathbf{X}$ be independent and identically distributed (iid) with common distribution $\mathbf{F}$,

$$C(\varepsilon) = P(|\mathbf{X} - \mathbf{Y}| \leqslant \varepsilon), \qquad \varepsilon > 0$$

Here $|\cdot|$ denotes any norm in the Euclidean space $\mathscr{R}^d$. We will suppose that the limit

$$D = \lim_{\varepsilon \to 0} \frac{\log C(\varepsilon)}{\log \varepsilon}$$

exists finite or infinite and call it the *correlation exponent of the sequence* $(\mathbf{X}_i)$ *or of* $\mathbf{F}$.

The estimation of D would be a standard statistical problem if we had a sequence of iid random variables $Z_1, Z_2, \ldots$ with the same distribution as $|\mathbf{X} - \mathbf{Y}|$ and such that $P(|\mathbf{X} - \mathbf{Y}| \leqslant x)$ is a regularly varying (at zero) function of x. This can be handled by so-called tail estimates which are well studied in the literature (see Section 2). In many cases of interest the observations $\mathbf{X}_1, \mathbf{X}_2, \ldots$ are dependent, so that it is not possible to construct iid Z_i from them.

In the literature, the usual statistical procedure for estimating D is the Grassberger–Procaccia method.[8] Since one does not know the probabilities $C(\varepsilon)$, one has to work with an approximation. Usually $C(\varepsilon)$ is replaced by the nearest-neighbor statistic

$$C_n(\varepsilon) = \frac{1}{n^2} \sum_{i,j=1}^{n} I(|\mathbf{X}_i - \mathbf{X}_j| \leqslant \varepsilon)$$

where $I(\cdot)$ denotes the indicator function. For ε fixed this is a U-statistic with kernel $I(|x - y| \leqslant \varepsilon)$ and for fixed n it is a special empirical distribution function of U-statistic structure. It is known[17] that

$$C_n(\varepsilon) \to C(\varepsilon) \quad \text{a.s.}$$

(a.s. stands for almost surely) and using the same ideas as for the proof of the classical Glivenko–Cantelli theorem, it is not difficult to see that $C_n(\varepsilon)$ even converges uniformly for $\varepsilon > 0$ with probability one. Denker and Keller[5] proved the asymptotic normality of a vector $C_n(\varepsilon_i)$, $i = 1, \ldots, k$, at distinct ε_i provided the $\mathbf{X}_i$ satisfy certain mixing conditions. The Grassberger–Procaccia method determines an estimate of D by plotting the values $\log C_n(\varepsilon)$ against $\log \varepsilon$ for a variety of ε values and then by calculating the slope of the so-defined curve in a sufficiently broad linear region. This can be done by linear regression techniques, and so we will sometimes refer to the Grassberger–Procaccia method as a linear regression procedure. Cutler[3] showed the asymptotic normality of a least squares estimator of D based on the vector $\log C_n(\varepsilon_i)$, $i=1, ..., k$, at distinct ε_i and under the same conditions as in Denker and Keller.[5]

Since $C(\varepsilon)$ is replaced by $C_n(\varepsilon)$, one would need an estimate for the error in this approximation. For $\mathbf{X}_i$ iid the central limit theorem for general U-statistics gives an estimate of the error of order $1/\sqrt{n}$. In the case of dependence in which we are mainly interested nothing seems to be known. But one is not actually interested in an error estimate for ε fixed, but for an ε which tends to zero in dependence on n. Since D is defined as a limit for $\varepsilon \to 0$, one has to apply the same limiting procedure for $C_n(\varepsilon)$. But $C_n(\varepsilon)$ is naturally equal to zero for sufficiently small ε. So one has to find a reliable region of ε values (depending on n) such that $C_n(\varepsilon)$ can be taken as a surrogate of $C(\varepsilon)$. For some work concerning the choice of the ε region we refer to Grassberger.[9]

The statistical estimator of D which we propose is based on the observation that we can write

$$C_n(\varepsilon)=\int_{|x-y|\leqslant\varepsilon} d\mathbf{F}_n(x)\, d\mathbf{F}_n(y), \qquad C(\varepsilon)=\int_{|x-y|\leqslant\varepsilon} d\mathbf{F}(x)\, d\mathbf{F}(y)$$

where $\mathbf{F}_n$ is the empirical distribution based on the observations $\mathbf{X}_1, ..., \mathbf{X}_n$. On the one hand, we know that $C_n(\varepsilon) \to C(\varepsilon)$ a.s. as a consequence of the fact that $\mathbf{F}_n \to \mathbf{F}$ a.s. in the uniform metric (which is a corollary of the ergodic theorem). Thus it might be worthwhile to work with random variables which are (conditionally) iid with distribution $\mathbf{F}_n$. This is the basic idea of the celebrated *bootstrap* which has become popular in statistics since Efron's[7] paper. We propose to apply the bootstrap idea to the well-known Hill estimator, which is a special tail estimator.

Our paper is organized as follows: In Section 2 we recall the notion of the Hill estimator. In Section 3 we introduce a Monte Carlo method for estimating the correlation exponent D. In Section 4 we consider a small simulation study in order to show the consistency of the estimator. In Section 5 we give some general comments about the method.

2. THE HILL ESTIMATOR

Suppose that $Z, Z_1, Z_2, ...$ are iid nonnegative real-valued random variables such that

$$P(Z \leqslant x) = L(x^{-1})x^{\alpha} \tag{2.1}$$

for a slowly varying (at infinity) function L [i.e., $L(cx)/L(x) \to 1$, $x \to \infty$, $c > 0$] and a positive number a. It is an important problem to estimate a given a sample

$$Z_1, ..., Z_n$$

A standard estimator of a is due to Hill:[14] Let

$$Z_{(1)} \leqslant \cdots \leqslant Z_{(n)}$$

denote the order statistics corresponding to the sample. For $m \geqslant 1$ define

$$H_{mn} = \left(-\frac{1}{m} \sum_{i=1}^{m} \log Z_{(i)} + \log Z_{(m+1)} \right)^{-1}$$

where log denotes the natural logarithm. Under the conditions above and if $m = m(n) \to \infty$, $m = o(n)$, it is shown, e.g., in Mason,[16] that the Hill estimator is a weakly consistent estimator of a. Moreover, if $m = n^{\gamma}$ for some $\gamma \in (0, 1)$ one can show the strong consistency and asymptotic normality of H_{mn} (see, e.g., refs. 16, 10, 4, 13, 1, and 6). Hall[10] derived an optimal γ for certain Z with a special probability behavior at zero. Moreover, it is known from the literature that H_{mn} is a maximum-likelihood estimator based on the m smallest z-values, and it is therefore optimal in a certain sense. For dependent Z_i the only paper we are aware of is due to Hsing,[15] who proves the consistency and asymptotic normality for moving average type Z_i. In this case and in the iid one it is known that one needs huge data sets in order to derive a reasonable estimate of a. This is natural because only the smallest order statistics of the Z_i will contribute to the estimation of the parameter a; in view of (2.1) only very small x will determine the value of a, and in a sequence of idd (Z_i) it is quite a rare event that one of the Z_i is smaller than a given level x.

It is known from the literature that, in general, the estimator cannot be improved by taking more and more order statistics into account ("let the small values speak for themselves").

In order to explain the particular form of the Hill estimate we give two arguments:

A. Suppose we know precisely that

$$P(Z \leqslant x) = x^a, \qquad x \leqslant \varepsilon \tag{2.2}$$

for a known value of ε. Then the maximum-likelihood estimator for $1/a$ is given by

$$\frac{-1}{m(\varepsilon)} \sum_{i=1}^{m(\varepsilon)} \log Z_{(i)} = \frac{-1}{m(\varepsilon)} \sum_{i=1}^{n} \log(Z_i)\, I(Z_i \leqslant \varepsilon)$$

where

$$m(\varepsilon)=\sum_{i=1}^{n} I(Z_i\leqslant\varepsilon)=\text{card}\{i: Z_i\leqslant\varepsilon\}$$

Notice that this estimator has been proposed by Takens.[19] In practice we rarely know the value of ε. Therefore we have to "search" for an appropriate ε such that (2.2) approximately holds. Thus it seems reasonable to let ε tend to zero together with the sample size n, i.e., $\varepsilon_n\to 0$. From the strong law of large numbers we know that

$$m(\varepsilon)\sim nP(Z\leqslant\varepsilon)\qquad\text{a.s.,}\qquad n\to\infty$$

and if $\varepsilon=\varepsilon_n\to 0$ then we have to expect that $m_n/n=m(\varepsilon_n)/n\to 0$. But we still have to assume that $m_n\to\infty$. This explains the two conditions on m above. We mention that $1/H_{mn}$ is the precise maximum-likelihood estimator of $1/a$ if the likelihood function is based on the m smallest order statistics $Z_{(1)}, ..., Z_{(m)}$.

B. To justify the random centering for $1/H_{mn}$ we give a standard argument based on regular variation. Under (2.1) we have for $y>0$

$$1-G(y)=P(1/Z\geqslant y)=P(Z\leqslant 1/y)=L(y)\,y^{-a}$$

Observe that dominated convergence and the uniform convergence theorem for regularly varying functions[2] imply that

$$\int_t^{\infty}\frac{\log y-\log t}{1-G(t)}\,dG(y)=\int_1^{\infty}\frac{1-G(tu)}{1-G(t)}\frac{du}{u}$$

$$\to\int_1^{\infty}u^{-a}\frac{du}{u}=\frac{1}{a},\qquad t\to\infty \tag{2.3}$$

Now, let

$$G_n(y)=\frac{1}{n}\sum_{i=1}^{n} I\left(\frac{1}{Z_i}\leqslant y\right),\qquad y\in\mathscr{R}$$

denote the empirical distribution function corresponding to the sample $1/Z_1, ..., 1/Z_n$. The function G_n is a nonparametric estimate of G. If we replace G in (2.3) by G_n and if we take $t=1/Z_{(m)}$, then we obtain

$$\int_{1/Z_{(m)}}^{\infty}\frac{\log y-\log(1/Z_{(m)})}{1-G_n(1/Z_{(m)})}\,dG_n(y)$$

which is precisely $1/H_{mn}$. Thus regular variation of $P(Z\leqslant x)$ is another argument for Hill's estimator.

Dependence of the observations $\mathbf{X}_i$ is a major difficulty for applying the Hill estimator to the data $Z_i = |\mathbf{X}_{2i} - \mathbf{X}_{2i-1}|$. Even if one does not consider the differences of two consecutive observations, the quality of the estimate cannot be expected to improve essentially. This can also be seen by simulations.

In the following section we will introduce a sequence of pseudo-independent data constructed from the $\mathbf{X}_i$ and apply the Hill estimator to them.

3. THE MONTE CARLO METHOD

We now consider a sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ as introduced in Section 1. As mentioned in Section 2 we need iid data in order to apply the Hill estimator. For dependent $\mathbf{X}_i$ we cannot expect that the Hill estimator works. Thus we will introduce a sequence

$$\mathbf{X}_1^*, \mathbf{X}_2^*, \ldots, \mathbf{X}_{2n}^*$$

of (conditionally) iid random variables with common distribution

$$\mathbf{F}_n = n^{-1} \sum_{i=1}^{n} \delta_{\mathbf{X}_i}$$

where δ denotes the Dirac measure. This means that we draw a random sample with replacement from $\mathbf{X}_1, \ldots, \mathbf{X}_n$. As mentioned in Section 1, the ergodic theorem implies that $\mathbf{F}_n$ converges weakly to $\mathbf{F}$ for almost all sample paths of $(\mathbf{X}_i)$ so that $\mathbf{F}_n$ is "close" to $\mathbf{F}$. We also have that

$$\begin{aligned} P^*(|\mathbf{X}_1^* - \mathbf{X}_2^*| \leqslant \varepsilon) &= \int_{|x-y| \leqslant e} d\mathbf{F}_n(x)\, d\mathbf{F}_n(y) \\ &= C_n(\varepsilon) \\ &\to C(\varepsilon) = P(|X - Y| \leqslant \varepsilon) = \int_{|x-y| \leqslant \varepsilon} d\mathbf{F}(x)\, d\mathbf{F}(y) \end{aligned}$$

with probability one. Here

$$P^*(\cdot) = P(\cdot \mid \mathbf{X}_1, \mathbf{X}_2, \ldots)$$

We conclude that if $C(\varepsilon)$ has power law behavior at zero, $C_n(\varepsilon)$ has approximately the same behavior.

Now we apply the idea of Hill estimation from Section 2 to the new sample $(\mathbf{X}_i^*)$ and define

$$H_{mn}^* = \left(-\frac{1}{m} \sum_{i=1}^{m} \log Z_{(i)}^* + \log Z_{(m+1)}^* \right)^{-1}$$

where

$$Z_i^* = |\mathbf{X}_{2i}^* - \mathbf{X}_{2i-1}^*|, \qquad i = 1, ..., n$$

and

$$Z_{(1)}^* \leqslant \cdots \leqslant Z_{(n)}^*$$

denote the corresponding order statistics. For m and $\mathbf{F}_n$ fixed we can repeat the above construction for independent samples of the Z_i^* as often as we wish (at least theoretically), say B times. Then we get a vector of B (conditionally) iid random variables. We can construct its empirical distribution, calculate its mean value, variance, etc., and determine confidence intervals.

The arguments for this Monte Carlo method are heuristic ones. We will show via simulations in Section 4 that the method works with a reasonable quality.

4. SOME SIMULATION STUDIES

In this section we provide some simulation examples of both independent and strongly dependent data. As mentioned in Section 1, we start with a one-dimensional sequence $X_1, X_2, ...$ which is supposed to be stationary and ergodic. For $d \geqslant 1$ fixed we embed the sequence in a d-dimensional phase space:

$$\mathbf{X}_i = (X_i, ..., X_{i+d-1}), \qquad i = 1, ..., n$$

with marginal distribution $\mathbf{F}$. We also suppose that the correlation exponent

$$D = D(d) = \lim_{\varepsilon \to 0} \frac{\log C(\varepsilon)}{\log \varepsilon}$$

exists. We will say that $(\mathbf{X}_i)$ or $\mathbf{F}$ has *correlation dimension* D if the $D(d)$ assume the constant value D for sufficiently large d.

As described in Section 3, we construct the estimators H_{mn}^* for a fixed sample of size n and a given value m. We repeat here the algorithm:

- From the original sample $\mathbf{X}_1, ..., \mathbf{X}_n$ draw a random sample (with replacement): $\mathbf{X}_1^*, ..., \mathbf{X}_{2n}^*$.
- Form the set of n distances $Z_k^* = |\mathbf{X}_{2k}^* - \mathbf{X}_{2k-1}^*|$.

- Take the order statistics $Z^*_{(1)}, \ldots, Z^*_{(n)}$ from the set of n distances Z^*_k.
- Calculate the corresponding Hill estimate H^*_{mn} from the values $Z^*_{(1)}, \ldots, Z^*_{(m+1)}$.
- Repeat this procedure 1000 times independently and average over the 1000 values of the Hill estimates to obtain the final estimate for the correlation exponent $D(d)$.

For fixed m, the 1000 values of Hill estimates generate an empirical distribution function with mean value which coincides with our proposed Monte Carlo estimate of $D(d)$. This allows for the construction of confidence intervals around the estimate. The number of 1000 replications is very high; simulations show that 200 or 300 are already enough for a stable estimation.

Example 1. *The Lorenz Map.* We consider a realization (X_t) of the Lorenz map (e.g., Smith[18]) with sample size $n = 10^5$. It is known that this system has correlation dimension 2.05. We compare the quality of the Grassberger–Procaccia method and of the averaged estimators H^*_{mn}. Both estimators slightly underestimate the true value even for higher dimensions d. The Hill-type estimator outperforms the linear regression estimator slightly.

We see from Fig. 1 that the estimator H^*_{mn} yields reasonable results in a sufficiently broad region of m values. It is our experience that the estimator provides the best results for m between $n^{1/3}$ and $n^{2/3}$, which is motivated by the conditions $m = m(n) \to \infty$ but $m = o(n)$. The Hill estimator overcomes the problem with the choice of the linear ε region in the Grassberger–Procaccia method. On the other hand, we have now to choose a reasonable value of m. Analogously to the Grassberger–Procaccia method, we propose to plot H^*_{mn} against a variety of m values and then to choose an m from a region where the graph of the curve is almost parallel to the x axis or to average over this region of H^*_{mn} values.

The Grassberger–Procaccia method and the Hill estimator yield very similar results (see Fig. 1). An advantage of the Monte Carlo procedure for the Hill estimator is that it yields an empirical distribution of the 1000 generated H^*_{mn} values for fixed m. Thus one can construct confidence bands from this distribution. Since we were not able to prove the consistency of the method under general dependence conditions, these confidence intervals have to be handled with care.

We also mention that these confidence bands are comparable with the Hill estimate in the iid situation where one uses a normal approximation. In particular, in all our examples, for m fixed, the empirical distribution function of the H^*_{mn} values is close to a normal distribution (see also Fig. 4).

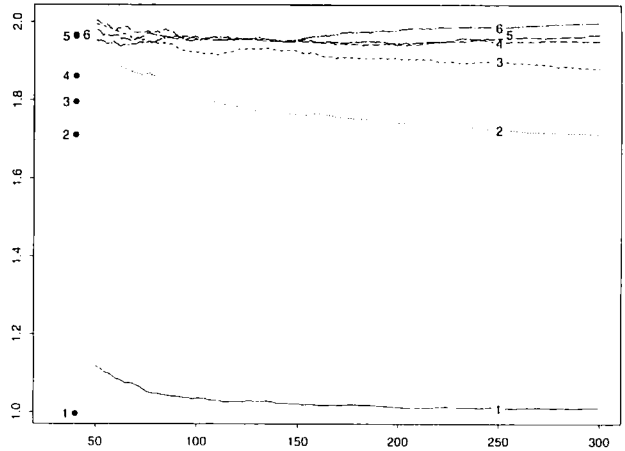

Fig. 1. Lorenz map: The averaged estimator of the 1000 H^*_{mn} (y axis) for m between 50 and 300 (x axis) and with phase space dimensions $d = 1, \ldots, 6$. The bullets show the estimates (for the same sample) of $D(d)$ obtained by the Grassberger–Procaccia method. The correlation dimension is 2.05.

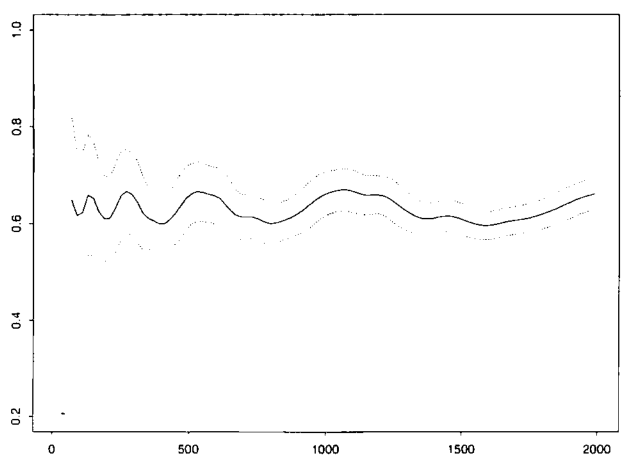

Fig. 2. Cantor set: The averaged estimator of the 1000 H^*_{mn} (solid line) (y axis) for m between 100 and 2000 (x axis) and with phase space dimension $d = 1$. For each m the dotted lines give a 95% confidence interval around the estimated mean value. The true value is 0.6309298.

Example 2. *The Cantor Distribution.* Let C_0 denote the standard Cantor set in [0,1] (Fig. 2). A dynamical system can be constructed by the ternary shift map $T(x) = 3x \pmod 1$ where the initial value is selected randomly and uniformly from C_0. We suppose that (X_i) is a realisation of the shift map, hence $d = 1$. The sample size is $n = 10^4$. It is known that $D = D(1) = \log 2/\log 3 = 0.6309298$ (e.g., Cutler[3]).

For comparison we consider the corresponding linear regression approach. In Fig. 3 we plot $\log C_n(\varepsilon)$ against $\log \varepsilon$ for a variety of ε values and for $n = 10^4$ values X_i. Linear regression in the $\log \varepsilon$ region $[-15, -2]$ yields an estimate of 0.631 for $D = D(1)$. This precision is quite striking.

We might ask why the Hill estimate seems less effective than the linear regression estimator. To understand this phenomenon we go back to Argument B in Section 2 (we also use the same notation): The Hill estimate is motivated by the relation

$$\int_1^\infty \frac{1 - G(tu)}{1 - G(t)} \frac{du}{u} \to \int_1^\infty u^{-D} \frac{du}{u} = \frac{1}{D}, \qquad t \to \infty \tag{4.1}$$

which holds if $1 - G(y) = P(Z \leqslant 1/y) = L(y)\, y^{-D}$ for a slowly varying function L. However, the situation for the Cantor set is different. In this case calculation shows that $1 - G(y) = y^{-D} g(\log y)$ for a positive, periodic, continuous function g which is bounded from below and from above by positive constants and such that $g(k \log 3) = 1$ for nonnegative integers k. (This fact and the following argument we learnt from David Vere-Jones.)

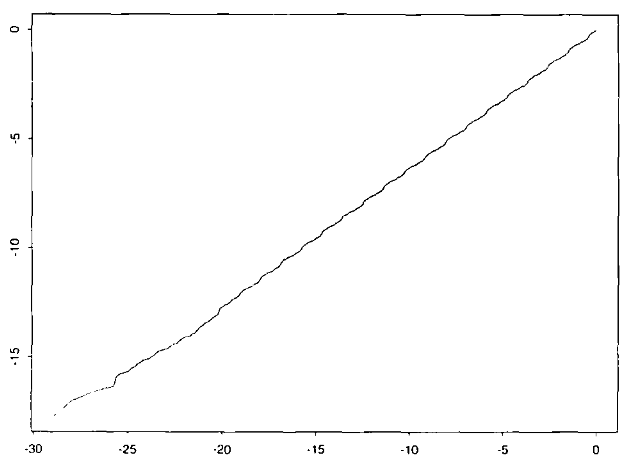

Fig. 3. Cantor set: Plot of $\log C_{10^4}(\varepsilon)$ (y axis) against $\log \varepsilon$ (x axis) for $\varepsilon \in [e^{-30}, e^{-2}]$ for the random Cantor set.

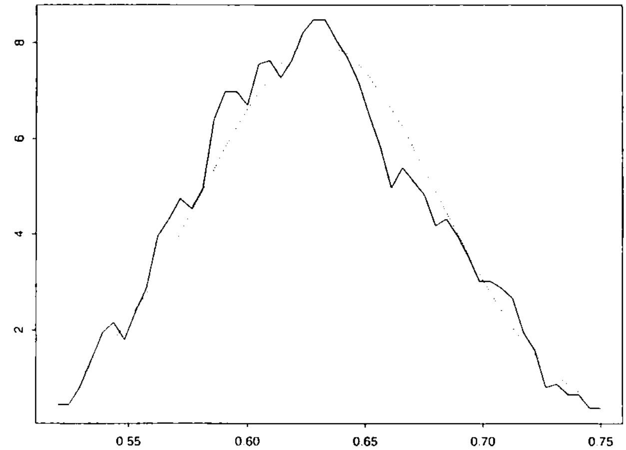

Fig. 4. Cantor set: The probability density function of the vector of the 1000 H^*_{mn} (solid line) (y axis) in the interval [0.5, 0.75] (x axis) for $m = 200$. The dotted line is the normal density function with the same mean value and variance as the vector of the H^*_{mn}.

In this case, the Hill estimate for different m will reflect ("estimate") the "lacunary" behavior of

$$\int_1^\infty u^{-D} \frac{g(\log(tu))}{g(\log(u))} \frac{du}{u}$$

as a function of t. This explains the oscillations of the estimation curve in Fig. 2. The Grassberger–Procaccia estimator yields such a precise value since it can be based on a wide span of ε values which is possible by the self-similarity of this example. A similar value could be obtained for Hill estimation by appropriately averaging over a range of Hill estimates for a region of m values.

Finally, in Fig. 4 we illustrate the concentration of the 1000 simulated H^*_{mn} around the true value log 2/log 3 by its probability density (smoothed histogram).

Example 3. *IID Noise.* The (X_i) are iid with a standard Gaussian distribution. Theoretically, $D(d) = d$. In Fig. 5 we show the influence of both the sample size n (x axis) and the phase space dimension d (y axis) on the estimation of $D(d)$ (z axis). The sample size is increased in 20 steps by 5000 from 5000 to 10^5. The estimates are again taken as an average over 1000 realizations of H^*_{mn}.

In Fig. 6 we fix the sample $n = 10^5$ and consider the dependence of the estimator on the phase space dimension d. For $d > 5$ the estimator is systematically biased. The bias is caused by the fact that different

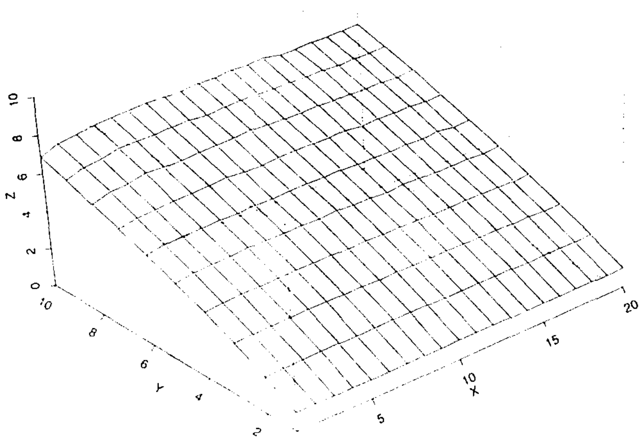

Fig. 5. IID Gaussian noise: The averaged estimator of the 1000 H^*_{mn} (z axis) for $m = 300$ with phase space dimensions $d = 1, \ldots, 10$ (y axis) for Gaussian standard white noise and with different sample size $n = 5000, 10000, \ldots, 100000$ (x axis). The numbers k on the x axis indicate the sample size $n = k * 5000$. The true values are $D(d) = d$.

phase space dimensions d require different choices of ε regions in the Grassberger–Procaccia method and, since ε regions can be translated into m regions, this means for the Hill estimate that m has to be chosen in dependence on d as well. A detailed study of the relationship between ε and m and their dependence on d is given by Harte[11] and Harte and Vere-Jones.[12]

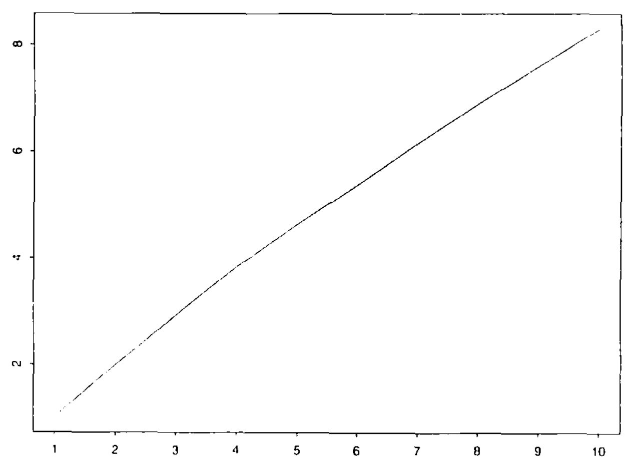

Fig. 6. IID Gaussian noise: The values of the averaged estimator of the 1000 H^*_{mn} (y axis) plotted against d (x axis) for $d = 1, \ldots, 10$ and $n = 10^5$, $m = 300$.

5. CONCLUDING REMARKS

The Monte Carlo method described above arose from attempts to detect fractal dimensions from data sets of meteorological and seismic activity (see Vere-Jones *et al.*[20]). We compared the results with those from regression-type estimators. Our conclusions is that both methods deliver similar results given that the underlying model does not exhibit lacunary behavior as in the Cantor case (see Example 2 above). At the moment we do not know how to overcome this problem. However, we found it particularly more satisfactory to work with the Monte Carlo method. In the regression approach one has to choose an ε region for every data set in a specific way which depends very much on the observations. This requires plenty of skill and experience, even more if one applies correction techniques for "lifting up" the graph of $C_n(\varepsilon)$ for small ε.

The alternative Monte Carlo method proposes a way of directly estimating the correlation exponent. As pointed out before, one has to choose an appropriate $m = o(n)$, but the rule that m is between $n^{1/3}$ and $n^{2/3}$ works in our cases sufficiently well. Moreover, we propose to determine the Hill estimate simultaneously for a sequence of m values which corresponds to a region of ε values. This makes Hill's estimate also more comparable with the Grassberger–Procaccia method, where one can take the whole region of ε values into account. The method is easily realized on a computer which is equipped with one of the modern statistical packages (we used Splus and a SPARC station). The calculations are computer-intensive concerning memory and speed. But we do not consider this as a disadvantage for those who have access to powerful computer facilities. We would also like to mention that the method does not very much depend on the embedding dimension d, since we only deal with the scalar quantities Z_i^*.

The Monte Carlo version of the Hill estimate cannot be expected to be a precise point estimate; bootstrap methods in general increase the variability of the estimator in order to get an empirical distribution function (that is the original motivation for the bootstrap). Thus the Monte Carlo method delivers a confidence bound around the estimated value. This is difficult to derive for the Grassberger–Procaccia method, although also for this approach statistical error estimates are known.[3] They require a special dependence structure (mixing) of the observations and knowledge about the covariance structure of the $C_n(\varepsilon)$ for different ε. This is hard to check in reality.

Our arguments for the Hill estimate are of heuristic nature. They are based on the fact that the Hill estimate for iid observations is the best available tail estimate in the sense of maximum-likelihood estimation and also the commonly used tail estimate in statistics. However, we are aware of the problems which arise from dependent data, even if they are randomized in the above way.

We mention that Harte[11] and Harte and Vere-Jones[12] informed us that they have made progress in the study of the relationship between the choice of ε (in regression estimates) and m (in Hill's estimate). They also propose bias corrections to the Hill procedure which are basically weighted Hill (i.e., kernel-type) estimates, and they study the influence of roundoff effects (which are inherent to any real data set) on the estimation of fractal dimensions.

ACKNOWLEDGMENTS

We are very much indebted to a referee whose criticism led to a substantial improvement of the present paper, as well as to a better understanding of the Hill estimator. Both authors take pleasure in thanking David Vere-Jones and David Harte for their everyday readiness to discuss the topic of the paper. In particular, we learnt from David Vere-Jones about the reasons for the behavior of Hill's estimate under lacunarity. This paper was written during the authors' stay at the Institute of Statistics and Operations Research of Victoria University Wellington. The research of T. M. was supported in part by New Zealand FRST grant 93-VIC-36-039.

REFERENCES

1. J. Beirlant and J. K. Teugels, The asymptotic behavior of Hill's estimator, *Theor. Prob. Appl.* **13**:463–469 (1986).
2. N. H. Bingham, C. M. Goldie, and J. K. Teugels, *Regular Variation* (Cambridge University Press, Cambridge, 1987).
3. C. D. Cutler, Some results on the behaviour and estimation of the fractal dimensions of distributions on attractors, *J. Stat. Phys.* **62**:651–708 (1991).
4. S. Csörgő, P. Deheuvels, and D. Mason, Kernel estimates of the tail index of a distribution, *Ann. Stat.* **13**:1050–1077 (1985).
5. M. Denker and G. Keller, Rigorous statistical procedures for data from dynamical systems, *J. Stat. Phys.* **44**:67–94 (1986).
6. P. Deheuvels, E. Haeusler, and D. M. Mason, Almost sure convergence of the Hill estimator, *Math. Proc. Camb. Phil. Soc.* **104**:371–381 (1988).
7. B. Efron, Bootstrap methods: Another look at the jackknife, *Ann. Stat.* **7**:1–26 (1979).
8. P. Grassberger and I. Procaccia, Measuring the strangeness of strange attractors, *Physica D* **9**:189:208 (1983).
9. P. Grassberger, Finite sample corrections to entropy and dimension estimates, *Phys. Lett. A*:369–373 (1988).
10. P. Hall, On simple estimates of an exponent of regular variation, *J. R. Stat. Soc. B* **44**:37–42 (1982).
11. D. Harte, Dimension estimates of earthquake epicentres and hypocentres, Technical Report, Victoria University Wellington (1994), in preparation.

12. D. Harte and D. Vere-Jones, Sources of bias affecting dimension estimates, Technical Report, Victoria University Wellington (1994), in preparation.
13. E. Häusler and J. L. Teugels, On asymptotic normality of Hill's estimator for the exponent of regular variation, *Ann. Stat.* **13**:743–756 (1985).
14. B. M. Hill, A simple general approach to inference about the tail of a distribution, *Ann. Stat.* **3**:1163–1174 (1975).
15. T. Hsing, On tail index estimation using dependent data, *Ann. Stat.* **19**:1547–1569 (1991).
16. D. Mason, Laws of large numbers for sums of extreme values, *Ann. Prob.* **10**:754–764 (1982).
17. T. Mikosch and Q. Wang, Some results on estimating Renýi dimensions, Technical Report, Victoria University Wellington (1993).
18. R. L. Smith, Estimating dimension in noisy chaotic time series, *J. R. Stat. Soc. B* **54**:329–351 (1992).
19. F. Takens, On the numerical determination of the dimension of an attractor, in *Dynamical Systems and Turbulence* (Springer, Berlin, 1985), pp. 99–106.
20. D. Vere-Jones, R. B. Davies, D. Harte, T. Mikosch, and Q. Wang, Problems and examples in the estimation of fractal dimension from meteorological and earthquake data, in *Proceedings of the International Conference on Applications of Time Series Analysis in Astronomy and Meteorology* (Padua, 1993).