In: *Int. J. Bif. Chaos* **3**, 765–771 (1993) LA-UR-92-2972

Statistical error in a chord estimator of correlation dimension: the "rule of five"

James Theiler
Center for Nonlinear Studies and Theoretical Division,
Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Turab Lookman
Department of Applied Mathematics,
University of Western Ontario, London, Ontario N6A 5B7, CANADA

(Revised September 1, 1992)

ABSTRACT

The statistical precision of a chord method for estimating fractal dimension from a correlation integral is derived. The optimal chord length is determined, and a comparison is made to other estimators. These calculations use the approximation that all pairwise distances between the points are statistically independent; the adequacy of this approximation is assessed numerically. The chord method provides a very quick and easy dimension estimate which is only slightly less precise than the optimal estimator.

Running title: a chord estimator of correlation dimension
Keywords: correlation dimension, statistical error
PACS: 02.60.+y, 05.45.+b

1. Introduction

Estimating the fractal dimension of a strange attractor from a finite sample of points is a problem that has attracted considerable interest over the last decade, particularly in the context of time series analysis. See Tong [1990], Grassberger *et al.* [1991], and Casdagli *et al.* [1992] for recent reviews of time series in general, and Theiler [1990a] for a review of dimension estimation in particular.

By embedding a time series into a higher dimensional space, the geometry of the underlying attractor can be reconstructed. This generally involves the use of time delay coordinates [Packard *et al.*, 1980] or their linear combinations [Broomhead and King, 1987], but the issues involved in making a good embedding are beyond the scope of this brief article. (See [Sauer *et al.*, 1991; Casdagli *et al.*, 1992b; Gibson *et al.*, 1992a] for fuller expositions. Instead, it will be assumed that we have a sample of N independent data points in an m dimensional phase space: $\{x_1, \ldots, x_N\} \in \mathbf{R}^m$.

One can define a *pointwise* mass function $B_x(N,r)$ which counts the fraction of points (not including x) that are within a distance r of the point x. The scaling of $B_x(N,r)$ with r for large N and small r defines a pointwise dimension: $B_x(N,r) \sim r^{d_x}$.

Averaging the pointwise mass function over all the points x leads to the correlation integral of Grassberger and Procaccia [1983] and Takens [1983]. Here $C(N,r) = \langle B_x(N,r)\rangle_x$ is the fraction of distances smaller than r between all pairs of points in an N point sample. If, in the limit of large N and small r, the correlation integral scales as r^ν, then the exponent ν defines the correlation dimension. Specifically,

$$\nu = \lim_{r\to 0}\lim_{N\to\infty} \frac{\log C(N,r)}{\log r} \tag{1}$$

In practice, ν is usually estimated as the slope in a log-log plot of $C(N,r)$ versus r, restricted to some range of r over which "good scaling" is observed. There is still a fair bit of "art" to this procedure, and numerous pitfalls await the unwary user; see [Rapp, 1992] for an anecdotal essay in the context of neuroscience applications.

One very common approach is to estimate the slope by taking a least squares fit through the points along some segments on the $C(N,r)$ curve. This is a reasonable approach, as long as it is done properly [Denker and Keller, 1986; Cutler, 1991], but the method is often abused. Some authors are led to believe that it is optimal because least squares was employed. Even more dangerous is a tendency to associate the standard error of the unweighted least-squares fit with the "error bar" on the dimension itself. The two are utterly different.

1.1 Takens Estimator

Takens [1985] has proposed a "best" estimator for ν directly from the set of pairwise distances. (Smith [1992a] has pointed out that the same estimator was derived much earlier in another context, before being "rediscovered" by Takens, but it was Takens who realized the relevance to fractal dimension estimation.) The Takens estimator requires the choice of a single free parameter, R_o, the upper cutoff distance. All pairwise distances larger than R_o are discarded, and all distances r which are less than R_o are averaged according to

$$\hat{\nu} = \frac{-1}{\langle \log(r/R_o)\rangle} \tag{2}$$

The Takens estimator is optimal in the situation that all distances are independent (which is not strictly true, but often a good hypothesis for small distances – more about this later), and the correlation integral is strictly proportional to r^ν. (this can fail from finite size effects, lacunarity, or noise) see [Theiler, 1988; Cutler, 1991; Smith, 1992b] for discussion of when this is not the case). See Cawley and Licht [1986] for some numerical experiments with this estimator.

An equivalent form of this estimator, in terms of the correlation integral, is given by

$$\hat{\nu} = \frac{C(R_o)}{\int_0^{R_o} [C(r)/r]\, dr} \tag{3}$$

1.2 Chord Estimator

Arguably the easiest way to estimate a slope of the correlation integral to choose two points on the curve, at say R_o and R_1, and measure the slope of the chord that is drawn through those two points. That is,

$$\hat{\nu} = \frac{\log C(N, R_o) - \log C(N, R_1)}{\log R_o - \log R_1} \tag{4}$$

This may seem terribly inefficient, because so much information is thrown away, but for a well chosen chord, the statistical error is only be 25% worse than for the optimal Takens estimator, which uses all the information in the $C(N, r)$ function.

In choosing the optimal chord, two effects are traded off against each other. First, the "length" of the chord should be as large as possible, so that the slope is less sensitive to small fluctuations in the endpoints. Second, the chord should not be so long as to reach too far into the small distances, where the fluctuations of the endpoint is particularly large. In what follows, the optimal choice of chord will be derived; the rule for optimal choice can be expressed independently of ν and the upper cutoff R_o.

2. Derivation of Statistical Error and Optimal Chord

Suppose the correlation integral is evaluated at two distances, R_o and $R_1 < R_o$. Define $\mathcal{R} = R_o/R_1 > 1$. Let $\mathcal{N}_o$ and $\mathcal{N}_1$ be the number of distances less than R_o and R_1 respectively. Finally define $n_o = \mathcal{N}_o - \mathcal{N}_1$. The chord estimate of dimension is given by

$$\hat{\nu} = \frac{\log \hat{\mathcal{N}}_o - \log \hat{\mathcal{N}}_1}{\log R_o - \log R_1} = \frac{\log(\hat{\mathcal{N}}_1 + \hat{n}_o) - \log \hat{\mathcal{N}}_1}{\log \mathcal{R}} \tag{5}$$

Now, $\mathcal{N}_1$ and $\mathcal{N}_o$ are not statistically independent, but $\mathcal{N}_1$ and $n_o = \mathcal{N}_o - \mathcal{N}_1$ are. Further, since the random variables $\mathcal{N}_1$ and n_o have a Poisson distribution, we can write the variances $\mathrm{Var}(\mathcal{N}_1) = \mathcal{N}_1$ and $\mathrm{Var}(n_o) = n_o$. Then, we can compute

$$\mathrm{Var}(\nu) = \left(\frac{\partial \nu}{\partial \mathcal{N}_1}\right)^2 \mathrm{Var}(\mathcal{N}_1) + \left(\frac{\partial \nu}{\partial n_o}\right)^2 \mathrm{Var}(n_o). \tag{6}$$

In particular,

$$\frac{\partial \nu}{\partial \mathcal{N}_1} = \frac{1}{\log \mathcal{R}} \left[\frac{1}{\mathcal{N}_1 + n_o} - \frac{1}{\mathcal{N}_1}\right] \tag{7}$$

$$\frac{\partial \nu}{\partial n_o} = \frac{1}{\log \mathcal{R}} \left[\frac{1}{\mathcal{N}_1 + n_o}\right] \tag{8}$$

Thus, we can write

$$\mathrm{Var}(\nu) = \frac{1}{(\log \mathcal{R})^2}\left(\left[\frac{1}{\mathcal{N}_1 + n_o} - \frac{1}{\mathcal{N}_1}\right]^2 \mathcal{N}_1 + \left[\frac{1}{\mathcal{N}_1 + n_o}\right]^2 n_o\right). \tag{9}$$

Let us introduce the parameter Θ defined by

$$\Theta = \frac{C(N, R_o)}{C(N, R_1)} = \mathcal{N}_o/\mathcal{N}_1 = (R_o/R_1)^\nu. \tag{10}$$

Then, use $\mathcal{N}_1 = \mathcal{N}_o/\Theta$ and $n_o = \mathcal{N}_o(\Theta - 1)/\Theta$, as well as $\nu \log \mathcal{R} = \log \Theta$ to rewrite Eq. (9).

$$\mathrm{Var}(\nu)/\nu^2 = \frac{1}{\mathcal{N}_o}\left[\frac{\Theta - 1}{(\log \Theta)^2}\right]. \tag{11}$$

This expression is independent of ν and of R_o, and is minimized with Θ approximately equal to 4.921554, giving a relative statistical error for ν as about $1.25/\sqrt{\mathcal{N}_o}$. By contrast, the Takens best estimator (which the reader will recall uses *all* distances less than the upper cutoff R_o instead of just the single distance R_1) has a relative error of $1.0/\sqrt{\mathcal{N}_o}$.

Having chosen an upper scale R_o, the optimal choice of lower scale R_1 is given by that R_1 for which

$$C(R_o)/C(R_1) = \Theta_{\text{optimal}} \approx 5. \tag{12}$$

Using this "rule of five," it is possible to very rapidly estimate a fractal dimension which is only slightly less efficient than optimal.

Note that there is no penalty for large dimension. This analysis also suggests that the optimal choice of R_o is as large as possible. This is true from the point of view of statistical precision, but if systematic errors are included, smaller R_o are generally desired (see Theiler [1990b] for a discussion of the tradeoff between statistical and systematic error). If lacunarity is taken into account, then there is an advantage to larger $\mathcal{R}$. See Smith [1992c] for a fuller discussion of best estimators in the presence of lacunarity.

3. Comparison to Other Estimators

The Takens method uses all distances less than R_o, whereas the chord method uses $C(N, r)$ at only two values of r. Qualitatively speaking, there are two different sources of the increased imprecison in the chord method. One results from ignoring distances below the lower cutoff R_1, and the other results from ignoring distances between R_1 and R_o. It is possible to assess the relative contributions of these two effects by considering two other estimators. The first was suggested by Ellner [1988] (see also Olofsen *et al.* [1992] for an extension to multiple embedding dimensions) who derived the maximum likelihood estimator that uses all distances between R_1 and R_o. Here,

$$\hat{\nu} = \frac{-1}{\frac{\mathcal{N}_1}{\mathcal{N}_o - \mathcal{N}_1}\log(R_1/R_o) + \langle \log(r/R_o)\rangle} \tag{13}$$

where the average is over all $\mathcal{N}_o - \mathcal{N}_1$ distances between R_1 and R_o. Note that as $R_1 \to 0$, $\mathcal{N}_1/\mathcal{N}_o \to 0$ and the original Takens estimator is retrieved. In terms of the correlation integral, this estimator can be written

$$\hat{\nu} = \frac{C(R_o) - C(R_1)}{\int_{R_1}^{R_o} [C(r)/r]\; dr}. \tag{14}$$

Ellner [1988] has shown that the variance of this estimator is given by

$$\begin{aligned} \mathrm{Var}(\nu)/\nu^2 &= \frac{1/[1-(R_1/R_o)^\nu]}{\mathcal{N}_o} && (15) \\ &= \frac{1}{\mathcal{N}_o}\left[\frac{\Theta}{\Theta - 1}\right] && (16) \end{aligned}$$

In particular, at $\Theta = 4.921554$, the relative statistical error is $1.12/\sqrt{\mathcal{N}_o}$.

A second alternative estimator was suggested by Smith [1992b]; it is the maximum likelihood estimator based on K+2 distances equally spaced on a logarithmic scale: $R_k = R_o/\mathcal{R}^k$ (note that the special case $K = 0$ is the chord estimator):

$$\hat{\nu} = -\frac{\log[C(R_o) + C(R_1) + \cdots + C(R_K)] - \log[C(R_1) + \cdots + C(R_{K+1})]}{\log \mathcal{R}} \tag{17}$$

It is interesting to consider the $K \to \infty$ limit for this estimator, because — in opposition to Ellner's estimator — the estimator does *not* throw away information in $C(N, r)$ for $r < R_1$, but *does* throw away information in $C(N, r)$ for distances between R_o and R_1. Ellner's estimator arguably makes more sense, because there may be good reason to throw away the $r < R_1$ information (noise may have corrupted the small distance scales), whereas the only case for ignoring $C(N, r)$ at the intermediate distance scales is ease of implementation (though this is not a trivial consideration). Smith's estimator is illustrative because it does well just when Ellner's does poorly (at small Θ), and vice-versa (for large Θ).

For large K, it can be shown that

$$\mathrm{Var}(\nu)/\nu^2 = \left[\frac{(\Theta - 1)^2}{\Theta(\log\Theta)^2}\right]\frac{1}{\mathcal{N}_o} \tag{18}$$

In the limit $\Theta \to 1$, this approaches $\mathrm{Var}(\nu)/\nu^2 = 1/\mathcal{N}_o$, which is the same as the Takens estimator. In fact, in the $\Theta \to 1$ limit, the $K \to \infty$ version of Eq. 17 is exactly equivalent to the Takens estimator. (If $\Theta \to 1$ and $K \to \infty$ in such a way that $K \log\Theta$ is fixed, then Ellner's estimator is approached.)

4. The Independent Distances Hypothesis

All of the estimators described so far were derived to be optimal under the assumption that all distances less than R_o are independent. In general this is not true. One "solution" is to randomly choose $N/2$ pairs using all N data points; as long as the data points are independent, so will the pairs be. This may be philosophically correct (PC), but it is very inefficient because it ignores most of the short distances. If there are N independent data points, each with m independent coordinates, one might expect that the $O(mN)$ shortest distances are "nearly" independent; though of course they will not be strictly independent since they will be constrained by, for instance, various triangle inequalities. [1]

In Fig. 2, we plot statistical error in a chord estimator of dimension for data uniformly distributed on the unit segment. For small R_o, we find that this error agrees with the prediction of Eq. (11), which assumes that distances are independent. For larger R_o, however, we find that the statistical error is larger than predicted by Eq. (11), at least for smaller Θ. (For larger Θ, the error is dominated by statistical error of $C(N, R_1)$, where R_1 is small, so distances less than R_1 are nearly independent; thus Eq. (11) predicts error at large Θ well even when R_o is large.) When R_o is large enough that the independent distances hypothesis leads to a poor approximation, the optimal Θ becomes larger than five.

4.1 Anomalous Precision of Dimension Estimator

It was shown by Theiler [1990b] that except for very special cases, the precision with which the correlation integral $C(N, r)$ is estimated scales as $1/\sqrt{N}$ for sufficiently large N. It was assumed that dimension estimates based on the correlation integral would similarly scale, but some recent numerical observations suggests that this may not be so, and that the estimated dimension for smooth integer-dimensional attractors may have a precision which scales as $1/N$. We have no analytical (or even reasonable heuristic) explanations for this anomalous phenomenon.

The effect is shown in Fig. 3. Two different data sets were generated, both randomly, but according to different distributions. In Fig. 3(ab), the statistical error for the correlation integral (a) and the chord dimension (b) are shown as a function of N for data produced by a unit-variance gaussian random number generator. The correlation integral is evaluated at $R_o = 0.1$, and the chord dimension is estimated with $\Theta = 5$. Although the error in the correlation integral scales as $1/\sqrt{N}$, the error in the chord dimension appears to scale as $1/N$.

That this is not a general property is shown in Fig. 3(cd). Here data was generated using an iterated function system [Barnsley and Demko, 1985]: $x_{n+1} = (1/4)x_n$ with probability $1/2$; and $x_{n+1} = (3/4)x_n + (1/4)$ with probability $1/2$. The data fills the unit segment [0,1], but with a highly nonuniform Cantor-set-like measure. [2] The correlation dimension of this set is less than one. Again the correlation integral is estimated at $R_o = 0.1$, and the chord estimator uses $\Theta = 5$. What we see in this case is that estimators for both the correlation integral and the correltion dimension have errors that scale like $1/\sqrt{N}$.

This anomalously precise estimate of dimension when the attractor dimension is equal to the embedding dimension may account for the observation in [Theiler *et al.*, 1992 (Fig. 2a)] that the significance of a dimension-based test for nonlinearity in a low-dimensional time series of length N scales linearly with N. In that case, the given time series is compared surrogate data sets which are linear correlated noise (surrogate data) with the same auto-

[1] Another possibility is to construct a spanning tree, and compute distances between points only on the tree; this is a kind of compromise which sidesteps the triangle inequality constraints, and still provides mostly small distances, but it would still be incorrect to consider those distances as truly independent.

[2] The unit segment [0,1] has half of its measure concentrated in the interval [0,1/4], and the other half in [1/4,1]; each of these two intervals has half of its measure concentrated in the first quarter of its length, and the other half in the last three quarters; the measure in each of these four intervals is similarly partitions, *ad infinitum.*

correlation as the original data. The significance is defined by the difference between the estimated dimension for the real and the surrogate data, that quantitity divided by the standard deviation of the the dimension estimates for the many surrogate data sets. Since as we have seen, this standard deviation scales as $1/N$, this explains why the significance increases so rapidly for N.

5. Discussion

In the case where all distances are known to full resolution, and there is no noise in the data, the most precise estimator is that of Takens. By throwing information in the $C(N,r)$ curve at distances less than a lower cutoff R_1 (as one might due if these distances are contaminated by noise), the best estimator is Ellner's. For very small R_1 (large Θ), not much information is lost by ignoring these small distances, but for large $R_1 \approx R_o$, the effect of a small lever-arm is evident, and the estimator has a large statistical error.

If on the other hand one only knows the correlation integral $C(N,r)$ at discrete values of r that are equally spaced on a logarithmic scale, then Smith's estimator is best. In the $K \to \infty$ limit, it keeps information that Ellner's estimator discards, namely the behavior of $C(N,r)$ in the small r limit. It also discards information that Ellner's estimator uses in the behavior of $C(N,r)$ for distances between R_1 and R_o. This estimator works best when $R_1 \approx R_o$ (small Θ), because less information is lost.

The chord estimator is less precise than either Smith's or Ellner's estimator, because it considers only two points, and ignores everything else. However, when those two points are chosen so that the ratio of their correlation integrals is about five, then a very rapid estimate of dimension is obtained which is only 25% less precise than the optimal estimate that uses the full resoultion and full range of the correlation integral.

Acknowledgements

The derivation of Eq. (11) arose out of a discussion between the authors during the first SIAM Conference on Dynamical Systems, Orlando, Florida, in May 1990. We are grateful to the Los Alamos Advanced Computing Laboratory for the computer resources used to perform the numerical experiments shown here. Work by JT is partially supported by a grant from the National Institute of Mental Health 1-R01-MH47184-01 and is performed under the auspices of the Department of Energy.

Figure captions

Figure 1. Coefficient of relative statistical error, defined by $\sqrt{\mathcal{N}_o \mathrm{Var}(\nu)}/\nu$, for four estimators as a function of $\Theta = C(N,R_o)/C(N,R_1) = (R_o/R_1)^\nu$. The relative statistical error itself is a factor of $1/\sqrt{\mathcal{N}_o}$ smaller than these curves. The solid line is the chord estimator, which estimates dimension from the $C(N,r)$ curve at only two values of r. The optimum chord estimator is given at $\Theta \approx 5$. At this value, it is 25% worse than the Takens estimator (dashed line), and about 11% worse than the estimators of Ellner (dotted line) and Smith (dashed dotted line). Note that the minimum is quite shallow; for $1.35 \leq \Theta \leq 75$ the statistical error is still no more than twice the optimal Takens error.

Figure 2. Theoretical (dashed line) and numerically estimated (solid line) statistical error for the chord estimator using data generated randomly and uniformly on the unit segment. In all cases $N = 300$ points were used; the three curves correspond (from top to bottom) to $R_o = 0.02, 0.10, 0.50$. The agreement with theory breaks down for large R_o because the approximation of independent distances fails.

Figure 3. Anomalous scaling of statistical error as a function of the number N of points in the data set. Panels (a,c) are for the correlation integral itself, evaluated at $R_o = 0.1$ for (a) unit variance gaussian noise, and (b) random data with a highly nonuniform Cantor-like distribution. Theory predicts that these should scale as $1/\sqrt{N}$ Panels (b,d) are for the estimate of chord dimension using $R_o = 0.1$ and $\Theta = 5$. Here, we observe the anomalous $1/N$ scaling for the gaussian distribution, while we see the expected $1/\sqrt{N}$ scaling for the Cantor distribution.

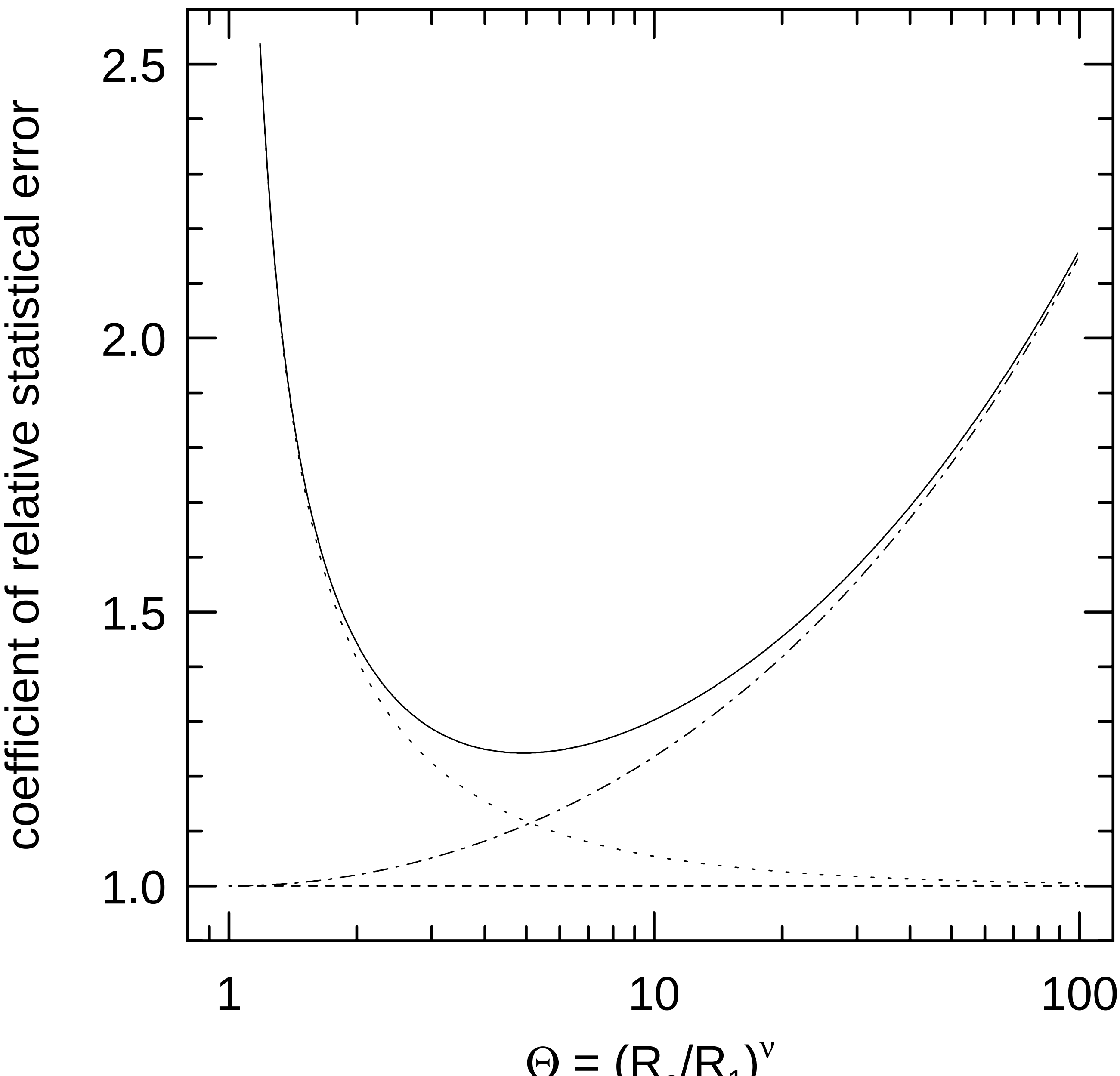

Figure 1: Theiler and Lookman

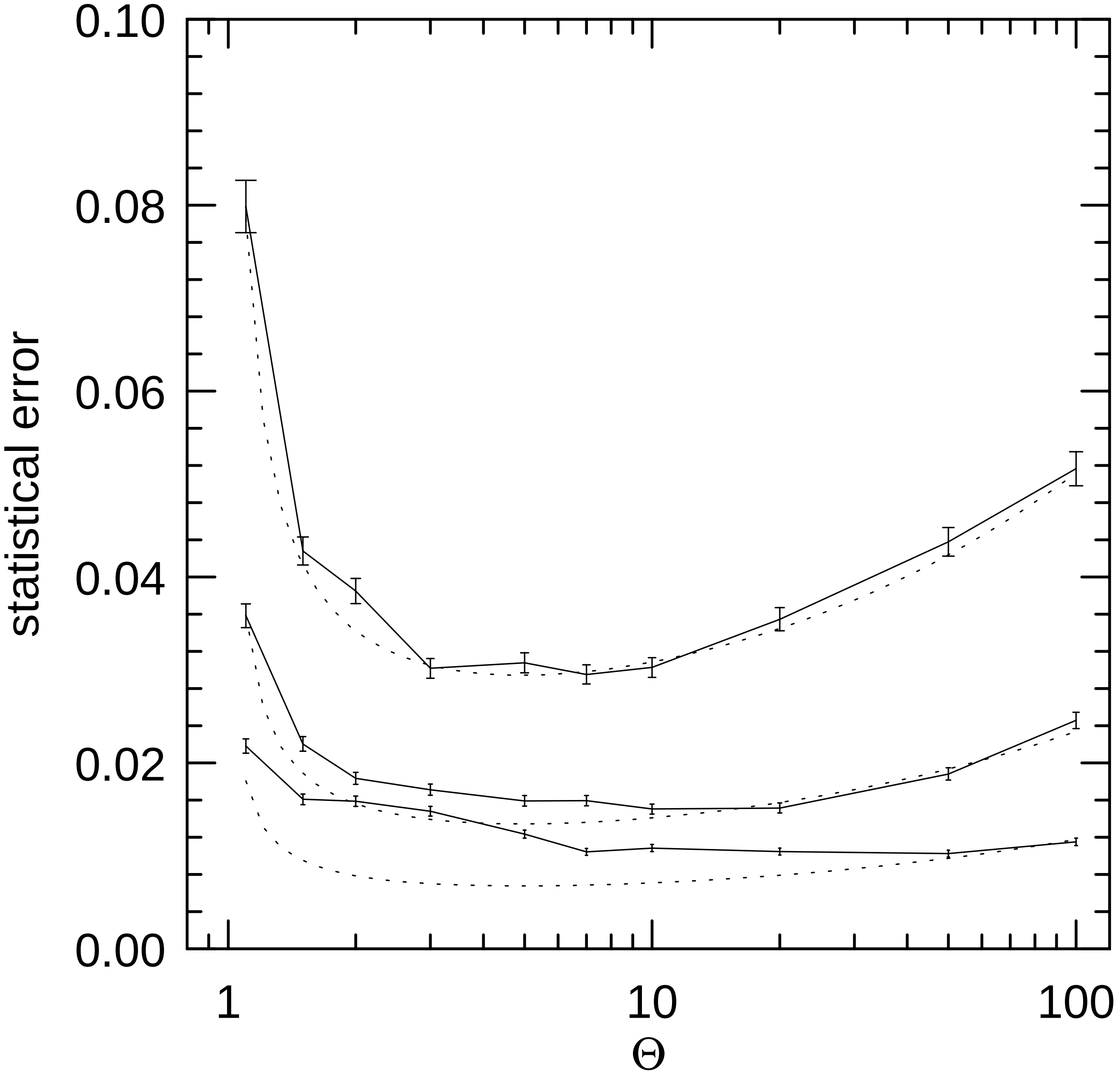

Figure 2: Theiler and Lookman

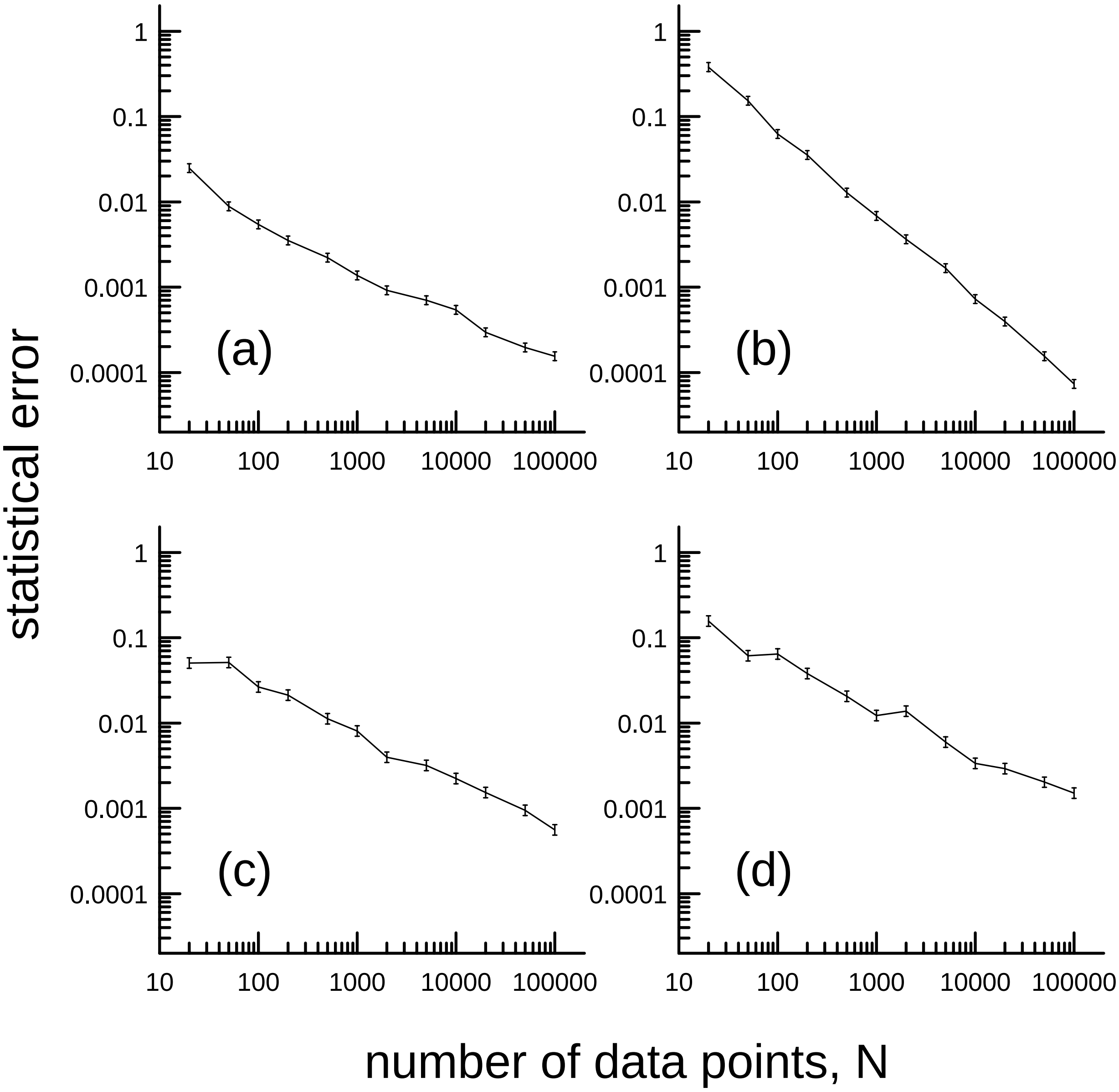

Figure 3: Theiler and Lookman