Article

Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems

ECKMANN, Jean-Pierre, RUELLE, D.

Abstract

We show that values of the correlation dimension estimated over a decade from the Grassberger-Procaccia algorithm cannot exceed the value 2 log10N if N is the number of points in the time series. When this bound is saturated it is thus not legitimate to conclude that low dimensional dynamics is present. The estimation of Lyapunov exponents is also discussed.

Reference

UNIVERSITÉ
DE GENÈVE

PHYSICA D

Fundamental limitations
for estimating dimensions and Lyapunov exponents
in dynamical systems

J.-P. Eckmann[1] and D. Ruelle

I.H.E.S., 91440 Bures-sur-Yvette, France

We show that values of the correlation dimension estimated over a decade from the Grassberger–Procaccia algorithm cannot exceed the value $2\log_{10} N$ if N is the number of points in the time series. When this bound is saturated it is thus not legitimate to conclude that low dimensional dynamics is present. The estimation of Lyapunov exponents is also discussed.

The purpose of this note is to question the validity of a number of recently published estimates of dimensions of attractors which are based on rather short time series. The values obtained are like 6 or 7, and we shall argue that they are probably a reflection of the small number of data points rather than of the dimension of a hypothetical attractor. Our conclusions go in the same direction as those of Grassberger [1] discussing work of Nicolis and Nicolis [2], and Procaccia [3] discussing work of Tsonis and Elsner [4]. Our analysis is however more precise, and somewhat more optimistic than that of Procaccia (we believe dimension estimates twice as large as those allowed by ref. [3]). See also the more pessimistic bounds of Smith [5].

While it is obvious that a short time series of low precision must lead to spurious results, we wish to argue that – even with good precision

data – wrong (too low) dimensions will be obtained. A similar analysis will apply to estimates of Lyapunov exponents.

Let (u_i) be a (scalar) time series with $i = 1, \ldots, N$ (the choice of sampling time unit will be discussed below). Using an embedding dimension m, we first reconstruct a trajectory in $\mathbb{R}^m$, with $x_n = (u_n, u_{n+1}, \ldots, u_{n+m-1})$. (This method, advocated by one of us (DR), was first documented in ref. [6].) Then, according to the Grassberger–Procaccia algorithm (GP) [7], we count the number $\mathcal{N}(r)$ of pairs of points with mutual distance $\leq r$. Note now that $\mathcal{N}(r)$ varies from 0 to $\frac{1}{2}(N-m)(N-m+1) \approx \frac{1}{2}N^2$. The algorithm next calls for plotting $\log \mathcal{N}(r)$ versus $\log r$. For small r, the slope of this plot is an estimate of the correlation dimension d [7]. (For larger r, the plot is not expected to be linear.) Thus, the method assumes

$$\mathcal{N}(r) \approx \text{const.} \times r^d, \tag{1}$$

and, if D is the diameter of the reconstructed

[1] Permanent address: Département de Physique Théorique, Université de Genève, 1211 Geneva 4, Switzerland.

attractor, we should have

$$\eta(D) \approx \tfrac{1}{2}N^2 \tag{2}$$

so that

$$\mathscr{N}(r) \approx \frac{N^2}{2}\left(\frac{r}{D}\right)^d. \tag{3}$$

The determination of the slope of $\log \mathscr{N}(r)$ requires using several values of r, and these should be "small" compared to D. But, obviously, we also need $\mathscr{N}(r)$ large with respect to 1, for statistical reasons. This forces

$$\tfrac{1}{2}N^2\left(\frac{r}{D}\right)^d \gg 1 \quad \text{and} \quad \frac{r}{D} = \rho \ll 1. \tag{4}$$

Taking logarithms, we find the requirement

$$2\log N > d\log(1/\rho). \tag{5}$$

From this it is clear that the GP-algorithm will *not produce dimensions larger than*

$$d_{\max} = \frac{2\log N}{\log(1/\rho)}. \tag{6}$$

Using decimal logarithms, and $\rho = 0.1$, we see that if $N = 1000$, then $d \le 6$, and if $N = 100000$, then $d \le 10$. Values of ρ larger than 0.1 might be adequate but, since the method is interesting mainly in very nonlinear situations, this would have to be justified. Thus, if the GP method yields a dimension of 6 for $N = 1000$ points, the result is probably worthless.

In case (u_i) is obtained by discretizing a continuous time signal, we have $N = T/\tau$ where T is the total recording time and τ the sampling time. One can of course try to make N large in (6) by taking τ small, but an easy geometric argument shows that τ should not be so small that consecutive points x_n, x_{n+1} on a reconstructed orbit are closer than the typical distance of points x_n, x_p

which are close on the attractor, but for which $|n - p|$ is large.

When can then a dimension estimate be considered reliable? First of all, the GP plots should be displayed and their linearity at small $\log r$ should be verified, as well as equality of slopes for different embedding dimensions. Next, the estimated dimension should be well below the quantity (6), obtained by using an honest value of N (not one artificially boosted by interpolation). A trick introduced by Scheinkman and Le Baron [8] may be of use to check the value of $d_{\max}$ in (6): they perform the GP algorithm both on the original time series (u_i) and on a "scrambled" series obtained by randomly permuting the u_i. The "scrambled" dimension is expected to be approximately equal to $d_{\max}$ and should be well above the "true" dimension.

We now briefly discuss the estimation of Lyapunov exponents. The situation is here somewhat worse than for the dimension. Any method to determine a Lyapunov exponent from an experimental time series requires that near a sequence of points x_n one finds other points x_{n+k} (for some k) so that the rate of divergence of orbits can be estimated. The number of points in a ball of radius r around a point x is

$$\mathscr{N}'(r) \approx \text{const.}' \times r^d, \tag{1'}$$

with

$$\mathscr{N}'(D) \approx N \tag{2'}$$

so that

$$\mathscr{N}'(r) \approx N\left(\frac{r}{D}\right)^d. \tag{3'}$$

(Strictly speaking, the information dimension rather than the correlation dimension should be used here, but the difference is not expected to be significant for present purposes.) As before,

this forces

$$N\left(\frac{r}{D}\right)^d \gg 1 \quad \text{and} \quad \rho = \frac{r}{D} \ll 1 \qquad (4')$$

so that we must have

$$\log N > d \log(1/\rho).$$

This says that the number of points N needed to estimate Lyapunov exponents is about the *square* of that needed to estimate the dimension.

The conclusion of what we have said above is obvious, but worth repeating: to extract useful dynamical information from time series (dimensions, Lyapunov exponents, etc.), *long* time series of high quality are necessary. We hope that the challenge of providing more such time series can be met.

References

[1] P. Grassberger, Nature 323 (1986) 609–612; 326 (1987) 524.
[2] C. Nicolis and G. Nicolis, Nature 326 (1987) 523.
[3] I. Procaccia, Nature 333 (1988) 498–499.
[4] A.A. Tsonis and J.B. Elsner, Nature 333 (1988) 545–547.
[5] L. Smith, Phys. Lett. A 133 (1988) 283.
[6] N.H. Packard, J.P. Crutchfield, J.D. Farmer and R.S. Shaw, Phys. Rev. Lett. 45 (1980) 712–716.
[7] P. Grassberger and I. Procaccia, Phys. Rev. Lett. 58 (1987) 2387–2389.
[8] J.A. Scheinkman and B. Le Baron, Nonlinear dynamics and stock returns, J. Business, to appear.