ON THE NUMERICAL DETERMINATION OF THE DIMENSION OF AN ATTRACTOR

by

Floris Takens

1. Introduction.

In [1] and independently in [4], a method was described to estimate a "dimension", or more precisely the correlation dimension, from experimental data. For a description of this method see also section 2 of the present paper. Various people asked me about the "convergence" of this algorithm, in other words about the errors to be expected in these estimates when only using a finite part of the data sequence. In this paper I discuss these errors and also extend this algorithm somewhat in order to minimize these errors.

In this discussion we use ideas from Mathematical Statistics, in particular Fischer's maximum likelihood rule, e.g. see [5]. In order to keep the paper self contained, I give complete proofs of the facts needed.

2. Description of the method.

The method of determining the correlation dimension of an attractor or ω-limit set in a physical experiment consists of two parts. First the "reconstruction of the phase space" in $\mathbb{R}^n$ and second the determination of the correlation dimension of a bounded subset in $\mathbb{R}^n$ with a probability measure. In the first part, a sequence $\{a_0,a_1,a_2,\ldots\}$, obtained as a time sequence of experimental data ($a_i \in \mathbb{R}$), is used to produce a sequence $\{A_0,A_1,A_2,\ldots\}$ in $\mathbb{R}^n$; this sequence is defined by $A_i = (a_i,a_{i+1},\ldots,a_{i+n-1})$. We assume that this sequence $\{A_i\}_{i\in\mathbb{N}}$ is bounded in $\mathbb{R}^n$ and define its limit set A by $A = \{x \in \mathbb{R}^n | x$ is an accumulation point of $\{A_i\}\}$. We also assume that $\{A_i\}$ defines a probability measure on A: for $K \subset A$ closed, and K_ε the ε-neighbourhood of K in $\mathbb{R}^n$, we define the measure of K as

$$m(K) = \lim_{\varepsilon\to 0} \left(\lim_{n\to\infty} \frac{1}{N} \cdot (\#\{0 \leq i \leq N | A_i \in K_\varepsilon\})\right)$$

(the assumption that this measure exists implies the existence of these limits). The experiment admits a finite dimensional deterministic description (or model) with phase space X, if the sequence $\{a_i\}$ is (in this model) obtained by composing the orbit $x(i) \in X$, describing the evolution during the experiment (with discrete time $i \in \mathbb{N}$), with a "read out" map $y: X \to \mathbb{R}$ (so $a_i = y(x(i))$. Then the sequence $\{A_i\}$ and the set A in $\mathbb{R}^n$ are the image of the orbit $\{x(i)\}$ and it ω-limit under a smooth transformation of X into $\mathbb{R}^n$. In [3] we proved that if n is sufficiently big, and if y, togehter with the dynamics in X satisfy certain generic conditions, then this

transformation from X to $\mathbb{R}^n$ is in fact an embedding. So then A is a faithfull representation of the ω-limit (or attractor).

If the collection of experimental data has started after the transient phenomena (i.e. phenomena due to particular initial conditions) have died out, then we may assume that the points A_i belong to A. This we assume from now on.

The correlation dimension for such a set $A \subset \mathbb{R}^n$ with probability measure m is defined by the following formulae, using the product measure m^2 on $A \times A$:

$$\text{correlation dimension } (A) = \lim_{\varepsilon \to o} \frac{\ln P(\varepsilon)}{\ln \varepsilon}$$

with $P(\varepsilon) = m^2(\{(a_1,a_2) \in A \times A \mid \rho(a_1,a_2) \leq \varepsilon\})$; ρ refers to the Euclidean distance in $\mathbb{R}^n$. If $\lim_{\varepsilon \to o}$ does not exist, the correlation dimension is not defined. This notion was introduced independently in [1] and [4]. For a numerical determination of the correlation dimension one generates a large number of (random) pairs (A_i,A_j) (from the sequence $\{A_i\}$ defining A and m), for example all pairs (A_i,A_j) with $0 \leq i < j \leq N$ for some N, and counts the number of pairs for which the distance is smaller than ε for various values of ε. In this way one obtains estimates for $P(\varepsilon)$, for these values of ε. In many cases, e.g. see [2], it then turns out that the estimated values of $\ln(P(\varepsilon))$ depend, to a good approximation, linearly on $\ln \varepsilon$. The slope of $\ln(P(\varepsilon))$, as function of $\ln \varepsilon$, determines the correlation dimension. In this way it is not easy to estimate the accuracy of this method. Especially not because, for smaller ε, the estimate of $\ln(P(\varepsilon))$ becomes more inaccurate.

For this reason we propose here a somewhat different way to proceed. This will lead to an estimate for the correlation dimension for which we know the standard error and for which this standard error is minimal in a sense to be discussed in section 4. In this procedure we also start with a (finite) sequence of distances $\rho_1,\rho_2,\ldots$ of (randomly) choosen pairs (A_i,A_j). Assuming that the limit, defining the correlation dimension, exists, we have $P(\varepsilon) = c.\varepsilon^{\alpha}$ + higher order terms (where α is the correlation dimension). We now take some $\varepsilon_o > 0$ fixed, and assume that for $\varepsilon \leq \varepsilon_o$ we have exactly $P(\varepsilon) = c.\varepsilon^{\alpha}$; we then try to estimate α.

This means that we disregard all distances ρ_i which are bigger than ε_o and use the remaining distances to estimate α. This problem, of estimating α from these distances, can be formulated independently of ε_o by dividing all distances by ε_o. In this way we get the following problem:

let $r_1,r_2,\ldots,r_m \in [0,1]$ be a random sample from a probability distribution (so we use R for the random variabele and $r_1,\ldots,r_m$ as the values of this variabele in particular experiments). What is the most likely value of α?

In the next section we deal with this problem.

In section 5 we discuss some cases where the distances ρ_i, as far as they are smaller than ε_o, are not distributed according to $P(\varepsilon) = c.\varepsilon^{\alpha}$.

3. The maximum likelihood rule.

According to Fischer's maximum likelihood rule, e.g. see [5], we should proceed as follows. If we find a sample $r_1,\ldots,r_m$ (see the end of section 2) then we observe that the probability of finding a sample in $(r_1,r_1+dr_1),\ldots,(r_m,r_m+dr_m)$, dependent on α is $\Pi\alpha.r^{\alpha-1}dr_i$. We calculate the value of α for which this is maximal: taking the logarithm of the last expression gives $m.\ln\alpha+(\alpha-1).\Sigma \ln r_i$ which attains its maximum for $\frac{m}{\alpha}+\Sigma \ln r_i = 0$ or for $\alpha^{-1} = \overline{\ln r_i}$ ($\overline{\ln r_i}$ indicates the average value of $\ln r_i$, $i=1,\ldots,m$). This means that $(-\overline{\ln r_i})^{-1}$ is the most likely value of α, given the sample $r_1,\ldots,r_m$. This is at least a simple rule to estimate α; in what sense it is "most likely" we shall discuss in section 4.
In this section we only derive the standard error of this estimate.

First we calculate, for a given distribution $\alpha.R^{\alpha-1}.dR$ on $[0,1]$ the mean- or expectation value of $\ln R$:

$$\mathcal{E}(\ln R) = \int_0^1 \ln r.\alpha.r^{\alpha-1}\,dr = r^\alpha \ln r \Big|_0^1 - \int_0^1 \frac{r^\alpha}{r}.\,dr = -\frac{1}{\alpha}.$$

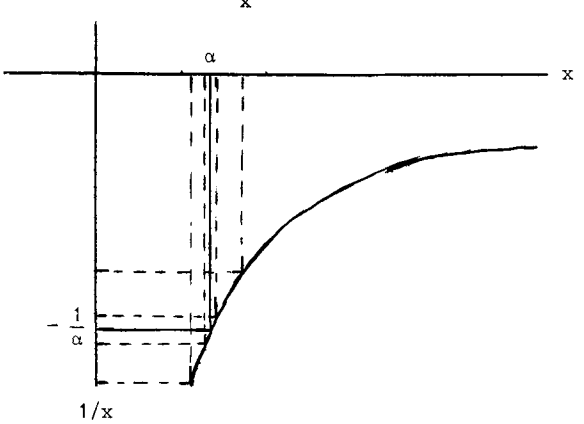
For this calculation one has to assume that $\alpha \neq 0$, but the case where $\alpha = 0$, or α is near zero, is the one giving the least problems in the numerical determination of the correlation dimension. From now on we assume that the parameter α is positive. For the expectation value of $(\ln R)^2$ we find:

$$\mathcal{E}((\ln R)^2) = \int_0^1 (\ln r)^2.\alpha.r^{\alpha-1}dr = r^\alpha(\ln r)^2 \Big|_0^1 - \int_0^1 \frac{2\ln r}{r}.r^\alpha dr = -2\int_0^1 \ln r.r^{\alpha-1}dr = \frac{2}{\alpha^2}.$$

This means that the variance of $\ln R$ equals $\sigma(\ln R) = \mathcal{E}((\ln R)^2) - (\mathcal{E}(\ln R))^2 = \frac{1}{\alpha^2}$.
So the standard error when estimating the average of $\ln R$, or $\mathcal{E}(\ln R) = -\frac{1}{\alpha}$, from a sample of size m is

$$\frac{1}{\sqrt{m}}.\frac{1}{\alpha}.$$

For m large, this is small. So, for large m we can obtain the standard error in the estimate of α by multiplying the above quantity with the absolute value of the derivative of the function $f(x) = -\frac{1}{x}$ (see the figure below). In this way we obtain $\frac{1}{\sqrt{m}}.\alpha$.



So for example , the estimate of the correlation dimension α will have a standard error of 10% if the sample size is 100.

The above considerations were carried out for known and fixed α. However if we only know the sample $r_1,\ldots,r_m$ they can also be used: we first estimate α and then use the estimated value of α to determine the standard error of the estimate.

Finally we observe that, especially in case of small sample sizes, the estimate of α will have a tendency of being in average somewhat high. For the estimate of $-\frac{1}{\alpha}$ or $\mathbb{E}(\ln R)$ this is not the case, but due to the non-linearity of $f(x) = -\frac{1}{x}$ a symmetric interval around $-\frac{1}{\alpha}$ gets transformed into a non-symmetrical interval around α, see the above figure.

4. Optimality.

We come back to the problem at the end of section 2. On $[0,1]$ we have a probability distribution $\alpha.R^{\alpha-1}.dR$, with α unknown. We have to estimate α on the basis of a random sample $r_1,\ldots,r_m$. In the previous section we saw how the maximum likelihood rule suggested to estimate α by taking the inverse of the average of $-\ln r_i$. Here we consider a larger class of procedures, namely we consider estimates of α, based on taking the average of $g(r_i)$ for some function $g: [0,1] \longrightarrow \mathbb{R}$ (of course we have to restrict to those functions for which the expectation value $\mathbb{E}_\alpha(G) = \int_0^1 g(r).\alpha.r^{\alpha-1}.dr$ exists for all $\alpha > 0$) writing G for $g(R)$. We shall show that we get the best results when taking $g(R) = \ln(R)$. We need to explain what we mean here by "best". For this we calculate the standard error when estimating α on the basis of the average of $g(r_i)$ when m, the sample size, is large.

We denote by $\mathbb{E}_\alpha(G)$, resp. $\mathbb{E}_\alpha(G^2)$, the expectation value of G, resp. G^2, with respect to the probability distrubution $\alpha.r^{\alpha-1}.dr$. We compute, just as in the case with $g(r) = \ln(r)$, the standard error when estimating $\mathbb{E}_\alpha(g)$ from a sample of size m. This gives

$$\frac{1}{\sqrt{m}}.\sqrt{\mathbb{E}_{\bar{\alpha}}(G^2) - (\mathbb{E}_{\bar{\alpha}}(G))^2}\ , \text{ when the actual value of } \alpha \text{ is } \bar{\alpha}.$$

For large m this is small (we assume that $\mathbb{E}_{\bar{\alpha}}(G^2)$ exists). In order to obtain an estimate for α, we have to use the inverse of the function of

$$\alpha \longmapsto \mathbb{E}_\alpha(G).$$

If this inverse does not exist, then we cannot estimate α from the average of $g(r_i)$ but then also g will not be acceptable for the criterion we give below. So from now on we assume that $\alpha \longmapsto \mathbb{E}_\alpha(G)$ is invertable. Again as in the case where $g(r) = \ln(r)$, we get the standard error, when estimating α, by multiplying the standard error of $\mathbb{E}_\alpha(G)$, see above, by the absolute value of the inverse of the derivative of $(\alpha \longmapsto \mathbb{E}_\alpha(G))$ in $\alpha = \bar{\alpha}$:

$$\frac{1}{\sqrt{m}} \cdot \frac{\sqrt{\mathcal{E}_{\bar{\alpha}}(G^2) - (\mathcal{E}_{\bar{\alpha}}(G))^2}}{|\partial_\alpha \mathcal{E}_{\bar{\alpha}}(G)|} .$$

This we want to be as small as possible for α and m fixed. So we define the quality of g, as function of $\bar{\alpha}$, by

$$Q(g,\bar{\alpha}) = \frac{(\partial_\alpha(\mathcal{E}_{\bar{\alpha}}(G)))^2}{\mathcal{E}_{\bar{\alpha}}(G^2) - \mathcal{E}_{\bar{\alpha}}(G))^2} .$$

So the main problem is to find those g, depending on $\bar{\alpha}$, for which this quantity attains its maximum value (for $\bar{\alpha}$ fixed). In fact we shall show that this optimal function g is independent of $\bar{\alpha}$ and equals $g(R) = \ln(R)$.

In order to prove this we calculate $Q(g,\bar{\alpha})$ for $g(r) = \ln(r) + \tilde{g}(r)$. We first show that it is no loss of generality to assume that

$$* \quad \int_0^1 \tilde{g}(r).\bar{\alpha}.r^{\bar{\alpha}-1}.dr = 0 \text{ and } \int_0^1 \ln(r).\tilde{g}(r).\bar{\alpha}.r^{\bar{\alpha}-1}dr = 0.$$

This follows from the fact that the quality of a function g does not change if we add a constant to g or if we multiply g with a non-zero scalar. So we may replace $g(r)$ by $\lambda_1(\ln(r) + \tilde{g}(r)) + \lambda_2 = \ln(r) + (\lambda_1 - 1).\ln r + \lambda_1.\tilde{g}(r) + \lambda_2$, i.e., we may replace $\tilde{g}(r)$ by $(\lambda_1 - 1)\ln r + \lambda_1.\tilde{g}(r) + \lambda_2$. By ajusting λ_1 and λ_2, we can get the above qualities except if $\tilde{g}$ satisfies the following quation:

$$-\frac{1}{\bar{\alpha}^2} - \frac{1}{\bar{\alpha}} .\mathcal{E}_{\bar{\alpha}}(\tilde{G}) - \mathcal{E}_{\bar{\alpha}}(\ln R.\tilde{G}) = 0 \qquad (1).$$

So if we prove that for all $\tilde{g}$ not satisfying (1) we have $Q(\ln + \tilde{g},\bar{\alpha}) \leq Q(\ln r,\bar{\alpha})$, then by continuity we have the result also for those $\tilde{g}$ satisfying (1).

Now we calculate $Q(\ln + \tilde{g}, \bar{\alpha})$ using the above equalities *. To follow these calculations it is useful to remember that

$$\int_0^1 r^\beta.\ln r.dr = -\frac{1}{(\beta+1)^2} , \int_0^1 r^\beta.(\ln r)^2.dr = \frac{2}{(\beta+1)^3} , \beta > -1.$$

$$\mathcal{E}_{\bar{\alpha}}(G) = \int_0^1 (\ln r + \tilde{g}(r)).\bar{\alpha}.r^{\bar{\alpha}-1}dr = -\frac{1}{\bar{\alpha}}$$

$$\partial_\alpha(\mathcal{E}_{\bar{\alpha}}(G)) = \int_0^1 \partial_\alpha((\ln r + \tilde{g}(\alpha))\alpha.r^{\alpha-1})dr = \int_0^1 (\ln r + \tilde{g}(r))(r^{\alpha-1} + \alpha.r^{\alpha-1}.\ln r)dr.$$

So $\partial_\alpha(\mathcal{E}_{\bar{\alpha}}(G)) = \int_0^1 \ln r(r^{\bar{\alpha}-1} + \bar{\alpha}.r^{\bar{\alpha}-1}.\ln r)dr = \frac{1}{\bar{\alpha}^2}$.

$$\mathfrak{E}_{\bar{\alpha}}(G^2) = \int_0^1 (\ln r + \tilde{g}(r))^2 . \bar{\alpha} . r^{\bar{\alpha}-1} dr = \int_0^1 (\ln r)^2 . \bar{\alpha} . r^{\bar{\alpha}-1} + \int_0^1 (\tilde{g}(r))^2 . \bar{\alpha} . r^{\bar{\alpha}-1} dr \geq \frac{2}{\bar{\alpha}^2} ;$$

the equality only holds if $\tilde{g}(r) \equiv 0$.

From this it follows that if $\tilde{g} \not\equiv 0$ and if $\tilde{g}$ satisfies $*$, then $Q(g,\bar{\alpha}) < Q(\ln, \bar{\alpha})$. This solves our main problem.

In the language of mathematical statistics, our results can be stated as follows: $\overline{-\ell n(r_i)}$ is an unbiased estimate for $\frac{1}{\alpha}$ with minimal variance (this can also be derived from [5,§38]); $-\ell n(r_i)^{-1}$ is an assymptotically unbiased estimate for α with assymptotically minimal variance.

5. Concluding remarks.

In the preceeding two sections we assumed that we were dealing with a probability distribution $\alpha . r^{\alpha-1} . dr$ on $[0,1]$ (corresponding to a probability distribution $c . \rho^{\alpha-1} . d\rho$ on $[0,\varepsilon_o]$, or corresponding to $P(\varepsilon) = c' . \varepsilon^{\alpha}$ for $\varepsilon \leq \varepsilon_o$). This assumption is not always justified. In fact, in a number of cases one has observed that for some ε_1, $P(\varepsilon)$ is approximately proportional to $c_1 . \varepsilon^{\alpha_1}$ for $\varepsilon < \varepsilon_1$ and is approximately proportional to $c_2 . \varepsilon^{\alpha_2}$ for $\varepsilon > \varepsilon_1$. This means that the "observed" or estimated dimension may depend on the scale (ε_o) of our observations. We calculate the result of such a probability distribution when applying our method of averaging $-\ln(\rho_i)$.

For convenience we transform again the interval $[0,\varepsilon_o]$ to $[0,1]$, replacing the variable $\rho \in [0,\varepsilon_o]$ by $r \in [0,1]$. If $P(r)$ denotes the probability of r' to belong to $[0,r]$, then we want

$$P(r) = \begin{cases} c_1 . r^{\alpha_1} & \text{for} \quad r \leq r_o \\ c_2 . r^{\alpha_2} & \text{for} \quad r \geq r_o \end{cases} ; \qquad 0 < r_o < 1.$$

Since $P(1) = 1$, $c_2 = 1$. Since the two definitions of $P(r_o)$ should agree, $c_1 = r_o^{\alpha_2 - \alpha_1}$. For this probability distribution we calculate the expectation value of $\ln(r)$:

$$\mathfrak{E}(\ln(R)) = \int_0^1 \ln r . dP(r) = \int_0^{r_o} \ln r . c_1 . \alpha_1 . r^{\alpha_1 - 1} \, dr + \int_{r_o}^1 \ln r . \alpha_2 . r^{\alpha_2 - 1} . dr =$$

$$c_1 \{ \ln r . r^{\alpha_1} \Big|_0^{r_o} - \int_0^{r_o} r^{\alpha_1 - 1} . dr \} + \{ \ln r . r^{\alpha_2} \Big|_{r_o}^1 - \int_{r_o}^1 r^{\alpha_2 - 1} . dr \} =$$

$$r_o^{\alpha_2 - \alpha_1} \{ r_o^{\alpha_1} . \ln r_o - \frac{r_o^{\alpha_1} - 1}{\alpha_1} \} + \{ - r_o^{\alpha_2} . \ln r_o - \frac{1}{\alpha_2}(1 - r_o^{\alpha_2}) \} = -\frac{1}{\alpha_2} + r_o^{\alpha_2} (- \frac{1}{\alpha_1} + \frac{1}{\alpha_2}) .$$

Some consequences of this formula are:

- In the case where these is a "random noise" in the signal $\{a_1, a_2, \ldots\}$ due to experimental (or round off) errors we may expect a probability distribution for r of the above type with $\alpha_1 = \infty$ and r_1, or rather $\varepsilon_o . r_1$ (see the transformation from ρ to r above) or the order of the experimental errors. In this case the random noise gives a correction term $\frac{1}{\alpha_2} . r_o^{\alpha_2}$. It is interesting to observe that this term decreases rapidly as α_2 takes higher values. So for higher dimensional attractors the effect of experimental errors becomes smaller.
- If α_1 is very near zero, $\xi(\ln(R))$ will be very strongly negative (leading to estimates of the correlation dimension near zero). So in this case our way of estimating the correlation dimension tends to ignore the α_2 part. Now however we cannot say much about the standard error in the determination of $\xi(\ln R)$.

In any case the above calculation shows that it is important to verify whether the distribution of distances in $[0,\varepsilon_o]$ really satisfies $P(\varepsilon) \sim c.\varepsilon^{\alpha}$. This can be done by repeating the estimation of the correlation dimension for various values of ε_o.

There is even an example of Zaslavski [2] [6] where numerical results indicate that $P(\varepsilon) \sim c.\varepsilon^{\alpha}$ does not hold on any interval of the form $[0,\varepsilon_o]$; with the present algorithm I found the estimates of the correlation dimension fluctuate from ± 1.1 to ± 1.5.

On the other hand there is an example (namely the interval mapping $x \longmapsto 1 - \mu x^2$ on $[-1,+1]$ for μ just below the Feigenbaum point 1.4011...) where we know that we should find the correlation dimension to be zero, but where we find, except for very low values of ε_o ($< 10^{-2}$), either an estimate for the correlation dimension around 0.5, or we find that some r_i is zero (leading to ln 0).

Finally some remarks on how to choose ε_o. Of course ε_o must be bigger than the experimental error; the ratio between ε_o and the experimental error should be bigger if one deals with lower dimensional attractors. Also ε_o should be sufficiently big so that the sample of distances $\rho \leq \varepsilon_o$ is not to small; however this sample size can be raised by considering more pairs (A_i, A_j).
Also ε_o should not be too big: A is bounded so $P(\varepsilon) \sim c.\varepsilon^{\alpha}$ will certainly be violated for large values of ε. Especially when the correlation dimension is big, the last two requirements may force one to consider very large numbers of pairs (A_i, A_j).

References.

1. P. Grassberger, I. Procaccia, On the characterization of strange attractors, Phys.Rev.Lett. 50 (1983), 346-349.

2. ———————— , Measuring the strangeness of strange attractors, Physica 9 D (1983), 189-208.

3. F. Takens, Detecting strange attractors in turbulence, in Dynamical systems and turbulence, Warwick 1980, Lecture Notes in Mathematics, 898, Springer-Verlag, Berlin.

4. _____ , Invariants related to dimension and entropy, in Atas do 13° colóqkio brasileiro de matemática, Rio de Janeiro, 1983.

5. B.L. van der Waerden, Mathematical Statistics, Grundlehren vol. 156, Springer Verlag, Berlin.

6. G.M. Zaslavskii, The simplest case of a strange attractor, Phys.Lett. 69 A (1978), 145-147.

F. Takens
Mathematisch Instituut
Postbus 800
9700 AV Groningen
The Netherlands