# Maximum-likelihood estimation of the entropy of an attractor

Jaap C. Schouten
*Department of Chemical Process Technology, Delft University of Technology, Julianalaan 136, 2628 BL Delft, The Netherlands*

Floris Takens
*Department of Mathematics and Computer Science, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands*

Cor M. van den Bleek
*Department of Chemical Process Technology, Delft University of Technology, Julianalaan 136, 2628 BL Delft, The Netherlands*
(Received 15 July 1993)

In this paper, a maximum-likelihood estimate of the (Kolmogorov) entropy of an attractor is proposed that can be obtained directly from a time series. Also, the relative standard deviation of the entropy estimate is derived; it is dependent on the entropy and on the number of samples used in the estimation.

PACS number(s): 05.45.+b

## I. INTRODUCTION

Invariants of attractors, such as the correlation dimension and the Kolmogorov entropy, are useful quantitative tools in the characterization of chaotic systems. During the last decade, many practical examples of chaotic attractors in physical and chemical systems have been shown in the literature. Also, various methods have been suggested, using the correlation integral, to calculate the correlation dimension and the Kolmogorov entropy from experimental data. When these invariants are used for further applications, such as the classification, characterization, or modeling of the underlying chaotic phenomena, it is necessary to have a reliable measure of the accuracy with which the invariant was obtained. Of course, in principle the level of uncertainty can be estimated by repeating the random drawing of a subset of the points on the attractor and the following calculation of the invariant many times, and subsequently calculating the average and the variance of the invariant. However, it would be much more convenient to obtain an estimate of the uncertainty from calculating the invariant only once. This can be achieved by applying a maximum-likelihood approach in which the correlation integral is treated as a probability distribution. Takens [1] has already derived a maximum-likelihood estimator for the correlation dimension together with an estimate of its standard error. A maximum-likelihood estimator for the Kolmogorov entropy was derived by Olofsen, De Goede, and Heijungs [2], which is strongly based on the method proposed by Grassberger and Procaccia [3] of calculating the entropy from the quotient of two correlation integrals at large embeddings. In this paper, we propose a different maximum-likelihood approach to the estimation of the Kolmogorov entropy and its standard error. We believe that this approach is simple and unambiguous, while the calculation can be done quickly when an efficient algorithm is used.

The method of determining the Kolmogorov entropy of an attractor in a physical experiment consists of two parts. First the attractor should be reconstructed in the state space whereafter the entropy can be estimated using the maximum-likelihood approach. Here, we will assume that the attractor has already been reconstructed using Takens's reconstruction technique [4] with specific choices of the reconstruction parameters (viz., embedding dimension and delay time). Of course, it may be possible that the correlation integrals, from which the entropy is estimated, are dependent on the choice of the reconstruction parameters. That means that also the calculated entropy will depend on this choice. In principle, this can be checked by repeating the estimation for various combinations of the reconstruction parameters.

## II. THE ENTROPY OF AN ATTRACTOR

The Kolmogorov entropy of an attractor can be considered as a measure for the rate of information loss along the attractor or as a measure for the degree of predictability of points along the attractor given an (arbitrary) initial point. In general, a positive, finite entropy is considered as the conclusive proof that the time series and its underlying dynamic phenomenon are chaotic. A zero entropy represents a constant or a regular, cyclic phenomenon that can be represented in the state space by a fixed point, a periodic attractor, or a multiperiodic attractor. An infinite entropy refers to a stochastic, nondeterministic phenomenon.

Here we will apply the definitions of the order-2 Kolmogorov entropy as suggested by Takens [5] and by Grassberger and Procaccia [3] (see also [6]). According to these definitions, we will estimate the entropy from the average time required for two orbits of the attractor, which are initially very close together, to diverge. More precisely, we calculate the entropy from the average of time $t_0$ that is needed for two points on the attractor, which are initially within a specified maximum distance $l_0$, to separate until the distance between these points has become larger than $l_0$. In this way, the entropy can be considered as an invariant, quantifying the rate of separation of nearby points on the attractor. We define two points to be initially nearby when these points are on different orbits and within a distance that is less than the specified maximum distance $l_0$.

According to Takens [5] and Grassberger and Procaccia [3], the separation of nearby points on different orbits is assumed to be exponential, and the time interval $t_0$ required for two initially nearby points to separate by a distance larger than $l_0$ will be exponentially distributed according to

$$C(t_0) \sim e^{-Kt_0} , \tag{1}$$

where $K$ is the Kolmogorov entropy. In most practical cases, this assumption is justified, tests of which have been given, for example, by Grassberger and Procaccia [7].

Generally, the points in an experimental time series are measured at discrete, constant time intervals with a time step $\tau_s$ between two sampled data points that equals $1/f_s$, where $f_s$ is the sampling frequency. Consequently, for practical purposes, $C(t_0)$ should be transformed into a *discrete* distribution function that is defined as

$$C(b) = e^{-Kb\tau_s} , \tag{2}$$

with $b = 1,2,3,\ldots$.

This cumulative distribution function describes the exponential decrease as a function of $b$. This variable $b$ equals the number of *sequential* pairs of points on the attractor, given an initial pair of independent points within a distance $l_0$, in which the interpoint distance is for the first time bigger than the specified maximum interpoint distance $l_0$. In other words, $b$ is obtained from the number of times that

$$\|\mathbf{X}_{i+b-1} - \mathbf{X}_{j+b-1}\| \le l_0 \tag{3}$$

with $b = 1,2,3,\ldots$, provided that

$$\|\mathbf{X}_i - \mathbf{X}_j\| \le l_0 , \tag{4}$$

while

$$\|\mathbf{X}_{i+b} - \mathbf{X}_{j+b}\| > l_0 . \tag{5}$$

In this case, the points on the (reconstructed) attractor are represented by their state vectors (using a delay time of unity) that are described by

$$\mathbf{X}_i = (x_i, x_{i+1}, \ldots, x_{i+m-1})^T \tag{6}$$

and

$$\mathbf{X}_j = (x_j, x_{j+1}, \ldots, x_{j+m-1})^T , \tag{7}$$

where $x_i, \ldots, x_{i+m-1}$, $x_j, \ldots, x_{j+m-1}$ are the respective data points in the time series [4]. In principle, the maximum distance $l_0$ can be based on any norm definition, e.g., the Euclidean norm or the maximum norm. However, when the maximum norm is used and when the attractor is reconstructed straightforwardly from the time series, then $b$ is readily obtained by counting the number of times that the absolute difference between sequential pairs of data points in the time series is smaller than $l_0$, given an initial pair of independent data points $(x_i, x_j)$, according to

$$|x_{i+m-1+b-1} - x_{j+m-1+b-1}| \le l_0 \quad \text{for } b = 1,2,3,\ldots , \tag{8}$$

and provided that

$$\max|x_{i+k} - x_{j+k}| \le l_0 \tag{9}$$

with $0 \le k \le m-1$, while

$$|x_{i+m-1+b} - x_{j+m-1+b}| > l_0 . \tag{10}$$

In this way, using an efficient algorithm, $b$ can be determined very quickly and directly from the time series.

The rate of the exponential decrease with $b$ is thus measured by the invariant entropy $K$. In principle, this definition of the entropy is only valid in the limit case, that means for small distances $l_0$ and large embedding dimensions $m$, so that when

$$l_0 \to 0 \wedge m \to \infty . \tag{11}$$

Consequently, not only the parameters used in the reconstruction of the attractor (viz., embedding dimension and delay time) but also the choices of the parameters used in the calculation of the correlation integral $C(b)$ (viz., embedding dimension $m$ and the length scale $l_0$) may affect the estimation of the Kolmogorov entropy. As also mentioned in the Introduction, in principle, this can be checked by repeating the estimation of the entropy for various values of these (reconstruction) parameters.

## III. MAXIMUM-LIKELIHOOD ESTIMATION OF THE ENTROPY

For convenience, we now introduce $k = K\tau_s$. So the distribution function $C(b)$ now reads as

$$C(b) = e^{-kb} . \tag{12}$$

The probability of finding a distance bigger than $l_0$ after exactly $b$ interpoint distances is

$$p(b) = C(b-1) - C(b) = e^{-k(b-1)} - e^{-kb} . \tag{13}$$

From this, it is determined that

$$p(b) = (e^k - 1)e^{-kb} . \tag{14}$$

This probability density function is known as the *geometric* probability density function (see, for example, Ref. [8]); it has been correctly normalized as can be deduced from

$$\sum_{b=1}^{\infty} p(b) = (e^k - 1) \sum_{b=1}^{\infty} e^{-kb} = 1 , \tag{15}$$

using that

$$a + a^2 + a^3 + \cdots = \frac{a}{1-a} \quad \text{when } |a| < 1 . \tag{16}$$

Using the probability density distribution of $b$, we will now derive an expression for the entropy based on a maximum-likelihood estimation. The probability of finding exactly the sample $(b_1, b_2, \ldots, b_M)$, depending on $k$, from a random drawing of $M$ pairs of independent points on the attractor, is

$$p_k = p(b_1, b_2, \ldots, b_M; k) = \prod_{i=1}^{M} p(b_i) = (e^k - 1)^M \exp\left(-k \sum_{i=1}^{M} b_i\right) . \tag{17}$$

First, we take the logarithm of both sides; this results in the so-called log-likelihood function $L(k)$:

$$L(k) = \ln[p(b_1, b_2, \ldots, b_M; k)] = M \ln(e^k - 1) - k \sum_{i=1}^{M} b_i . \tag{18}$$

Finding the maximum of this function means that we want to find the value of $k$ (or entropy $K$) that leads to the largest probability of finding the sample $(b_1, b_2, \ldots, b_M)$. So we want to find the maximum of $L(k)$, which follows from

$$\frac{\partial L(k)}{\partial k} = \frac{M}{1 - e^{-k}} - \sum_{i=1}^{M} b_i = 0 . \tag{19}$$

From this equation we can derive the *maximum-likelihood estimate* of the entropy, $K_{\mathrm{ML}}$:

$$k_{\mathrm{ML}} = -\ln\left(1 - \frac{1}{\bar{b}}\right) \Longrightarrow K_{\mathrm{ML}} = -\frac{1}{\tau_s} \ln\left(1 - \frac{1}{\bar{b}}\right) , \tag{20}$$

with

$$\bar{b} = \frac{1}{M} \sum_{i=1}^{M} b_i , \tag{21}$$

which is the average value of the $b$'s in the sample $(b_1, b_2, \ldots, b_M)$, with sample size $M$. We can easily prove that this value of $K$ is indeed leading to a maximum of $L(k)$ by calculating the second derivative of $L(k)$ in $k = k_{\mathrm{ML}}$:

$$\frac{\partial^2 L(k)}{\partial k^2} = -\frac{M e^{-k}}{(1 - e^{-k})^2} = -M\bar{b}(\bar{b} - 1) < 0 . \tag{22}$$

Because this second derivative is always negative [the average of $b$ is always larger than *1* (so positive)], the value of $k_{\mathrm{ML}}$ is indeed leading to a maximum of $L(k)$. From the derivation of $K_{\mathrm{ML}}$ given above, it follows that the maximum-likelihood estimate of $K$ is only a function of the sample average of $b$. Using this result, the same expression for the entropy estimate can be derived by calculating the expectation value $E(b)$ of $b$, using the probability density $p(b)$, which leads to

$$E(b) = \frac{e^k}{e^k - 1} = \frac{1}{1 - e^{-k}} . \tag{23}$$

When we use in Eq. (23) for $E(b)$ the sample average of $b$, we obtain the same expression for $k$ as was derived with the maximum-likelihood method [Eq. (20)].

## IV. STANDARD DEVIATION OF THE ENTROPY ESTIMATE

The standard deviation of the maximum-likelihood estimate of $K$ can be obtained from the variance of $b$, $\mathrm{var}(b) = E(b^2) - [E(b)]^2$, which reads for this geometric distribution as

$$\mathrm{var}(b) = \frac{e^k}{(e^k - 1)^2} . \tag{24}$$

The standard deviation in the estimate of the sample average of $b$ follows from

$$\sigma(\bar{b}) = \sqrt{\mathrm{var}(\bar{b})} = \sqrt{\mathrm{var}(b)/M} = \frac{e^{k_{\mathrm{ML}}/2}}{\sqrt{M}\,(e^{k_{\mathrm{ML}}} - 1)} . \tag{25}$$

For large sample sizes $M$, the standard deviation of the average of $b$ will be small. In that case, we can use the derivative of the function $k = -\ln(1 - 1/b)$ in the point $k = k_{\mathrm{ML}}$ to estimate the standard deviation of $k$ (or $K$), because we can write for small values of the standard deviation of the average of $b$ that

$$\frac{[k_{\mathrm{ML}} + \sigma(k_{\mathrm{ML}})] - [k_{\mathrm{ML}} - \sigma(k_{\mathrm{ML}})]}{[\bar{b} + \sigma(\bar{b})] - [\bar{b} - \sigma(\bar{b})]} = \frac{\sigma(k_{\mathrm{ML}})}{\sigma(\bar{b})} \simeq \left| \left( \frac{\partial k}{\partial b} \right)_{k = k_{\mathrm{ML}}} \right| , \tag{26}$$

as can be seen in Fig. 1. From the latter equation, the standard deviation of $k$ is obtained as

$$\sigma(k_{\mathrm{ML}}) \simeq \sigma(\bar{b}) \left| \left( \frac{\partial k}{\partial b} \right)_{k = k_{\mathrm{ML}}} \right| = \sigma(\bar{b}) \frac{(e^{k_{\mathrm{ML}}} - 1)^2}{e^{k_{\mathrm{ML}}}} . \tag{27}$$

The combination of Eqs. (25) and (27) gives the following result for the standard deviation of $K$:

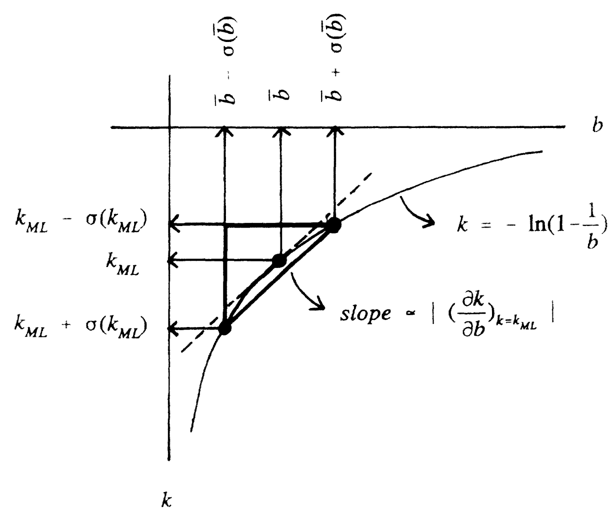


FIG. 1. Approximation of the standard deviation of $k_{\mathrm{ML}}$ from the sample average of $b$.

$$\sigma(K_{\mathrm{ML}})=(\tau_s\sqrt{M})^{-1}(e^{K_{\mathrm{ML}}\tau_s/2}-e^{-K_{\mathrm{ML}}\tau_s/2}) \quad (28)$$

$$=(\tau_s\sqrt{M})^{-1}2\sinh(K_{\mathrm{ML}}\tau_s/2)$$

$$=[\tau_s\sqrt{M}\sqrt{\bar{b}(\bar{b}-1)}]^{-1} . \quad (29)$$

Consequently, the relative standard deviation $s(K_{\mathrm{ML}})$ of the estimate of the entropy is obtained from

$$s(K_{\mathrm{ML}})=\frac{\sigma(K_{\mathrm{ML}})}{K_{\mathrm{ML}}}\simeq(\tau_s\sqrt{M}K_{\mathrm{ML}})^{-1}2\sinh(K_{\mathrm{ML}}\tau_s/2)$$

$$=[\tau_s\sqrt{M}K_{\mathrm{ML}}\sqrt{\bar{b}(\bar{b}-1)}]^{-1} . \quad (30)$$

This means that $s(K_{\mathrm{ML}})$ is not only dependent on the sample size $M$ but also on the maximum-likelihood estimate $K_{\mathrm{ML}}$ itself. However, one can show that for sufficiently small values of $k_{\mathrm{ML}}=K_{\mathrm{ML}}\tau_s$ (i.e., $k_{\mathrm{ML}} \ll 1$), the relative standard error becomes nearly independent of $k_{\mathrm{ML}}$ and thus can be approximated by

$$s(K_{\mathrm{ML}})\simeq\frac{1}{\sqrt{M}} . \quad (31)$$

For example, at a sampling frequency of 200 Hz and an entropy of 10 bits/sec, $k_{\mathrm{ML}}$ will be 0.05. In that case, a relative standard error of 1% or smaller is obtained with a sample size of at least 10 000 values of $b$.

## V. CONCLUDING REMARKS

A maximum-likelihood approach has been proposed for the estimation of the Kolmogorov entropy [Eqs. (20) and (21)] and of the relative standard error of the entropy estimate [Eq. (30)]. The possibility of calculating a (statistical) estimate of the standard error will be very useful, especially for the determination of how many pairs of points should be taken into account in the analysis to obtain a required accuracy. For example, one could define the requirement that $s(K_{\mathrm{ML}})$ should always be at least smaller than 0.1%, which means that the calculation of $K_{\mathrm{ML}}$ should be based on at least a sample size of the order of $10^6$ values of $b$. The determination of $K_{\mathrm{ML}}$ and its standard error from a time series can be done rather quickly when an efficient algorithm is used [9].

Obviously, the values of $b$ obtained from the time series can also be used to determine quantitatively the cumulative distribution function $C(b)$ of $b$. It is useful to compare this function with the calculated exponential function obtained from the maximum-likelihood estimate of $K$. This comparison will provide an indication of how well the assumed exponential decrease of $C(b)$ is described by the experimental data.

The largest possible value of the entropy that may be computed from a time series can be directly deduced from Eq. (20): the largest value of $K_{\mathrm{ML}}$ is obtained for the smallest average value of $b$ that is larger than 1. This situation is encountered when $(M-1)$ times a value of $b=1$ is obtained and just 1 times a value of $b=2$. In that case, the average value of $b$ equals

$$\bar{b}=\frac{[(M-1)\times 1+1\times 2]}{M}=\frac{(M+1)}{M} . \quad (32)$$

The maximum entropy reads then as

$$K_{\mathrm{ML,max}}=f_s\ln(M+1) , \quad (33)$$

which equals $\ln(M+1)$ per time step. From this, it is observed that the maximum possible value of the entropy increases when the sample size $M$ is increased.

A last remark will be made about the units of entropy. The Kolmogorov entropy is usually expressed in *bits per unit of time*. For that reason, the base of the distribution function $C(b)$ should be 2 instead of $e$. When the base $e$ is used, as in this paper, the entropy is usually expressed as *nats per unit of time*.

[1] F. Takens, in *Dynamical Systems and Bifurcations,* edited by B. L. J. Braaksma, H. W. Broer, and F. Takens, Lecture Notes in Mathematics Vol. 1125 (Springer, Berlin, 1984), pp. 99–106.

[2] E. Olofsen, J. De Goede, and R. Heijungs, Bull. Math. Biol. **54**, 45 (1992).

[3] P. Grassberger and I. Procaccia, Phys. Rev. A **28**, 2591 (1983).

[4] F. Takens, in *Dynamical Systems and Turbulence,* edited by D. A. Rand and L.-S. Young, Lecture Notes in Mathematics Vol. 898 (Springer, Berlin, 1981), p. 366.

[5] F. Takens, *Atas do 13° Coloqkio Brasileiro de Matematica* (Instituto de Matematica Pura e Aplicada, Rio de Janeiro, 1983).

[6] P. Grassberger, in Chaos, Nonlinear Science: Theory and Applications, edited by Arun V. Holden (Manchester University Press, Manchester, 1987).

[7] P. Grassberger and I. Procaccia, Physica **13D**, 34 (1984).

[8] A. B. Clarke and R. L. Disney, *Probability and Random Processes for Engineers and Scientists* (Wiley, New York, 1970), pp. 169–174.

[9] J. C. Schouten and C. M. van den Bleek, *RRCHAOS: A Menu-driven Software Package for Chaotic Time Series Analysis,* Reactor Research Foundation, Delft, The Netherlands (1993) (unpublished).