PHYSICAL REVIEW E, VOLUME 64, 035202

Divergence measure between chaotic attractors

L. Diambra*
Departamento de Fisiologia e Biofísica, Universidade de São Paulo, Avenue Professor Lineu Prestes 1524, ICB1 CEP 05508-900, São Paulo, SP Brazil

(Received 9 March 2001; published 30 August 2001)

We propose a measure of divergence of probability distributions for quantifying the dissimilarity of two chaotic attractors. This measure is defined in terms of a generalized entropy. We illustrate our procedure by considering the effect of additive noise in the well known Hénon attractor. Finally, we show how our approach allows one to detect nonstationary events in a time series.

DOI: 10.1103/PhysRevE.64.035202 PACS number(s): 05.45.Tp, 05.40.Ca, 05.45.Df

Through the appropriate embedding procedures, strange attractors can be numerically approximated by a large set of points, either from experimental time series (TS) or from numerical simulation of chaotic systems. Advances in nonlinear analysis of TS have made it possible to identify and classify chaotic dynamical systems, determine if a signal is deterministic or not, and establish correlations where the traditional linear analyses were not sensitive (see Ref. [1] for a review). However, there are many situations where we do not need a complete characterization of an attractor, but rather a quantitative way of comparing attractors. For instance, recently several authors have proposed using some measures of dissimilarity of attractors to analyze nonstationary signals [2,3] and for TS classification [4]. In some situations it could be important to quantify the difference of two attractors corresponding to slightly different parameters of a chaotic dynamical system. The computation of the hierarchy of generalized dimensions does not help, because even if all dimensions of two fractal sets are equal, this does not guarantee that the two fractal objects are identical. Furthermore, for a reliable estimation of nonlinear dynamical measures used to characterize chaotic dynamical systems, large quantities of precise data are necessary to achieve accurate description of the small scale structure in different regions of the attractor and these structures are easily damaged by noise.

In order to give a quantitative answer to these issues, a number of dissimilarity measures have been proposed in the literature. Kantz [5] introduced a cross-correlation (CC) integral to evaluate the similarity of attractors. He found that for small scales the decay rate of the CC is greater than the decay rate of the autocorrelation (AC). Alternatively, Albano *et al.* [6] use the Kolmogorov-Smirnov test for comparing quantitatively two sets of AC integrals. More recently, Schreiber [3] proposed that the nonlinear cross-prediction error can be used for measuring the similarity of short sequences.

The quantitative comparison of attractors can be relevant in many different problems, such as numerical taxonomy of TS, to establish a criterion for stationarity, to study the numerical convergence of chaotic solutions, to evaluate the effect of nonlinear noise reduction of noisy chaotic attractors, among other applications. For the above-mentioned purposes we need a reliable way of comparing attractors rather than their detailed characterization. In this Rapid Communication, we propose a divergence measure based on a generalized entropy function for quantifying the similarity of attractors. The procedure introduced here takes advantage of coarse-grained information without losing sensibility to high-order correlations in the data. We remark that it makes it possible to compare attractors, even in some situations where the more commonly used nonlinear measures are not computable.

From the information theory viewpoint, the amount of uncertainty of the probability distribution (PD), p_i, is defined in a general way by $H_f[p_i]=-\Sigma_i f[p_i]$ [7]. There is not a unique information measure H_f. The more commonly used information measure or entropy function was introduced by Shannon and Weaver [8], where $f(p)=p\ln p$. Generalized entropy f_q has been postulated by Rényi [9] and Havrda and Charvát [10]. Rényi's generalized entropy has been used to define a hierarchy of generalized dimensions [11]. Tsallis introduced the Havrda-Charvát entropy function to elaborate a nonextensive thermodynamics [12]. We have a divergence measure $D_f(p:\hat{p})$ associated with an entropy function *f*, between two PD p_i and $\hat{p}_i$. A general divergence measure form associated to the *f*-entropy was given by Csiszár [13],

$$D_f(p:\hat{p})=\sum_i \left(\hat{p}_i f\left[\frac{p_i}{\hat{p}_i}\right]+p_i f\left[\frac{\hat{p}_i}{p_i}\right]\right), \tag{1}$$

where *f* is a convex function and one imposes the condition $f(1)=0$, which guarantees $D_f(p:p)=0$. Rényi's generalized entropy does not satisfy convexity. Fortunately, Havrda-Charvát entropy function fulfills this property. For this reason we will work here with the Havrda-Charvát entropy function. From here on we shall refer to the generalized divergence measure associated with the Havrda-Charvát entropy function, simply as the *q*-divergence, and will be denoted by D_q. If we replace *f* by the function corresponding to the Shannon entropy, we obtain the well known Kullback-Leibler distance [14]

$$D_1(p:\hat{p})=\sum_i \left(p_i \ln\left[\frac{p_i}{\hat{p}_i}\right]+\hat{p}_i \ln\left[\frac{\hat{p}_i}{p_i}\right]\right). \tag{2}$$

*Electronic address: diambra@fisio.icb.usp.br

The function f corresponding to the Havrda-Charvát entropy is given by $f_q(p)=(q-1)^{-1}(p^q-p)$. Then, the associated q-divergence is given by

$$D_q(p:\hat{p})=(q-1)^{-1}\left(\sum_i p_i^q\hat{p}_i^{1-q}+\sum_i \hat{p}_i^q p_i^{1-q}-2\right). \tag{3}$$

It is easy to show that $D_q(p:\hat{p})\to D_1(p:\hat{p})$ when $q\to 1$. The q-divergence measure is positive definite, has been made symmetric, and fulfills $D_q(p:p)=0$. Also $D_q(p:\hat{p})$ considered as a function of p_i and $\hat{p}_i$ is convex. We remark that D_q is semimetric, since it may not satisfy the triangular inequality.

After these definitions, let us consider the divergence between two finite time series embedded in $\mathcal{R}^d$, $X=(\mathbf{x}_0,\ldots,\mathbf{x}_N)$ and $Y=(\mathbf{y}_0,\ldots,\mathbf{y}_N)$. There are two well known ways of estimating quantities (2) and (3) from X and Y. The most straightforward, but also more expensive, is to use a box counting (BC) approach that counts the number of points n_i in the box i. The probabilities p_i can be estimated as $p_i=n_i/N$. On the other hand, a more efficient method of estimating Eqs. (2) and (3) is by correlation sums [15]. Instead of taking a fixed mesh, one can calculate the probability $P(\mathbf{x}_j,\varepsilon)$ of finding a point within a sphere of radius ε centered at $\mathbf{x}_j$, with $j=1,\ldots,M$ randomly chosen from the trajectory X. $P(\mathbf{x}_j,\varepsilon)$ is estimated by counting the number n_j of points falling in the sphere of radius ε centered at $\mathbf{x}_j$,

$$P(\mathbf{x}_j,\varepsilon)=N^{-1}\sum_{i=1}^{N}\Theta(\varepsilon-|\mathbf{x}_i-\mathbf{x}_j|)\quad j=1,\ldots,M.$$

Using this definition, expressions (2) and (3) are replaced by

$$D_1(p:\hat{p},\varepsilon)=M^{-1}\left(\sum_{j=1}^{M}\ln\left[\frac{\sum_{i=1}^{N'}\Theta(\varepsilon-|\mathbf{y}_i-\mathbf{x}_j|)}{\sum_{i=1}^{N'}\Theta(\varepsilon-|\mathbf{x}_i-\mathbf{x}_j|)}\right]+\sum_{j=1}^{M}\ln\left[\frac{\sum_{i=1}^{N'}\Theta(\varepsilon-|\mathbf{x}_i-\mathbf{y}_j|)}{\sum_{i=1}^{N'}\Theta(\varepsilon-|\mathbf{y}_i-\mathbf{y}_j|)}\right]\right), \tag{4}$$

$$D_q(p:\hat{p},\varepsilon)=\frac{M^{-1}}{(q-1)}\left(\sum_{j=1}^{M}\left(\frac{\sum_{i=1}^{N'}\Theta(\varepsilon-|\mathbf{y}_i-\mathbf{x}_j|)}{\sum_{i=1}^{N'}\Theta(\varepsilon-|\mathbf{x}_i-\mathbf{x}_j|)}\right)^q+\sum_{j=1}^{M}\left(\frac{\sum_{i=1}^{N'}\Theta(\varepsilon-|\mathbf{x}_i-\mathbf{y}_j|)}{\sum_{i=1}^{N'}\Theta(\varepsilon-|\mathbf{y}_i-\mathbf{y}_j|)}\right)^q-2M\right), \tag{5}$$

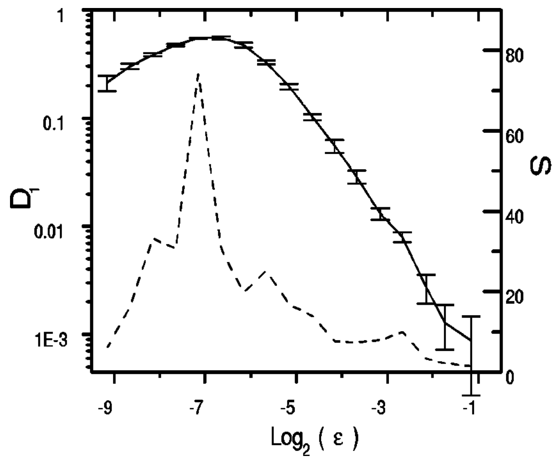

FIG. 1. Solid line: $D_1(\varepsilon)$ for the TS X generated by the Hénon system and the TS Y generated by the same system contaminated with noise ($\eta=20$ dB), computed with the BC scheme using $N=10\,000$ points and $d=2$ (left axis). Dashed line: the statistic $S(\varepsilon)$ (right axis).

where $\Theta(x)$ is the step function which has the value 1 if $x\geqslant 0$ and is 0 otherwise, and $|\mathbf{x}_i-\mathbf{y}_j|$ is the distance between $\mathbf{x}_i$ and $\mathbf{y}_j$. The sum is taken only for those i's and j's that are separated in time by more than B samples to avoid artifactural correlations [16], thus $N'=N-d-B$. Notice that quantities (2) and (3) are defined for two finite discrete probability distributions p_i and $\hat{p}_i$ only if $p_i>0$, and $\hat{p}_i>0$, and if there is a one-to-one correspondence between the elements i [9]. In order to satisfy these requirements, we perform the summation in the BC approach [Eqs. (2) and (3)] only over the boxes i that contain points from both X and Y (i.e., $p_i>0$ and $\hat{p}_i>0$). In the sphere counting (SC) scheme [Eqs. (4) and (5)], we include in the summation over j only spheres that contain points both in X and Y, and for this reason, M decreases with ε. In our numerical examples we shall estimate D_q using these two methods.

As our first example, we take the trajectory of 10 000 points of the Hénon model $x_{n+1}=1-ax_n^2+bx_{n-1}$, with parameters $a=1.4$ and $b=0.3$; cf. Kantz [5]. Set X corresponds to the clean attractor, while set Y consists of the same set of points plus an additive Gaussian noise. In Fig. 1 we present the mean value of D_1 versus ε computed with the BC method. We computed the mean value over five realizations of Y with signal-to-noise ratio $\eta=20$ dB [17]. Of course, the q-divergence depends on the length scale ε. By choosing a relatively small ε, D_q will pick up local differences between X and Y. However, taking ε too small leads to poor statistics. For large ε, we lose the small scale structure of the attractors so that they become indistinguishable. We can see that D_1 reaches a maximum at a value of ε that will be denoted by ε_0. The dashed curve corresponds to $S=\langle D_1\rangle/\sigma(D_1)$, where $\langle\,\rangle$ denotes the mean value over five realizations and σ denotes the standard deviation. We can see that the maximum value of D_1 presents very good statistics.

In Fig. 2 we display the q dependence of $D_q(\varepsilon_0)$ for different noise levels ($\eta=30$ dB, solid line; $\eta=20$ dB, dashed line; and $\eta=10$ dB, dotted line). The divergence of

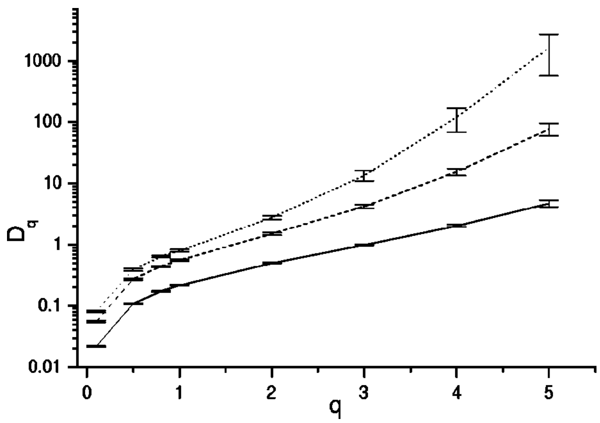

FIG. 2. $D_q(\varepsilon_0)$ between the TS X generated by the Hénon system and the TS Y generated by the same system contaminated with several level of noise as a function of the parameter q (solid line, 30 dB of SNR; dashed line, 20 dB of SNR; and dotted line, 10 dB of SNR). The calculations were performed with the BC scheme using $N=10\,000$ points and $d=2$.

the two attractors increases with q, which could be interpreted as a gain control parameter. This fact can be used to detect small divergence, as we will discuss in the last example. In Fig. 3 we illustrate the behavior of the mean value of D_2 versus ε computed with the SC method. We computed the mean value over five realizations of Y for each level of noise. We can see that D_2 presents a maximum at ε_0 that depends on the characteristic length scale at which the attractors differ. In particular in Fig. 1, $\varepsilon_0 \approx 0.01$ and the variance of the noise is the one-hundredth of the attractor. The ε dependence of D_q characterizes the relationship between X and Y. In Fig. 4 we display $D_2(\varepsilon_0)$ versus η. This figure shows that $D_2(\varepsilon_0)$ scales exponentially with the level of noise added to the signal. Thus, in the comparison of a clean and a noisy time series, the physical interpretation of ε_0 is closely related to the level of added noise.

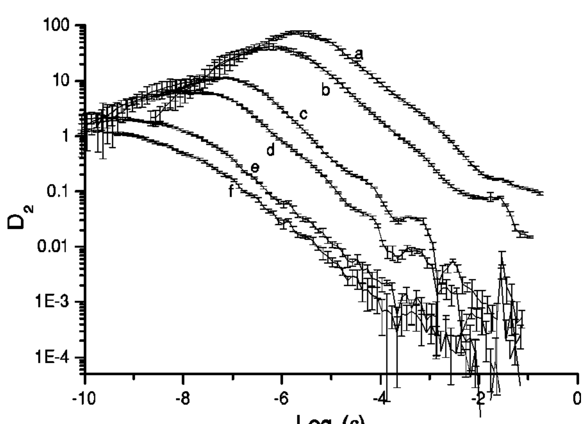

FIG. 3. Divergence $D_2(\varepsilon)$ between the Hénon system and itself contaminated with several levels of noise (a, $\eta=30$ dB; b, $\eta=27$ dB; c, $\eta=20$ dB; d, $\eta=17$ dB; e, $\eta=10$ dB; and f, $\eta=7$ dB). The calculations were performed with the SC scheme using $N=10\,000$ points, $B=20$, and $d=2$.

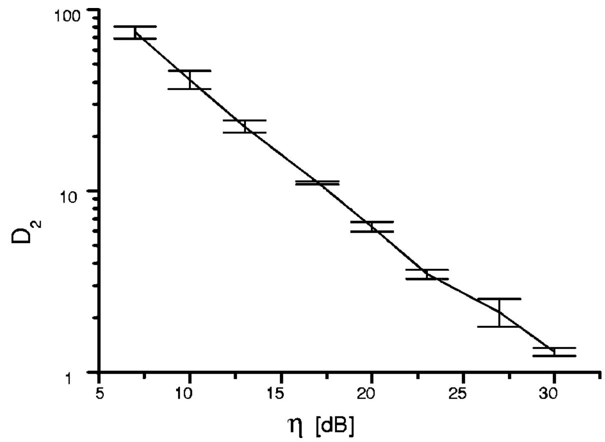

FIG. 4. Divergence $D_2(\varepsilon_0)$ between Hénon system and itself contaminated with noise as a function of the level of noise η.

Finally, let us show how the q-divergence D_q can be used to detect nonstationarity events in a TS. As a numerical example, let us consider a generalized baker map defined by

$$v_n \leqslant \alpha: \quad u_{n+1} = \beta u_n, \quad v_{n+1} = v_n/\alpha,$$

$$v_n > \alpha: \quad u_{n+1} = 0.5 + \beta u_n, \quad v_{n+1} = \frac{v_n - \alpha}{1-\alpha}.$$

For this map, the parameter β can be varied without changing the positive Lyapunov exponent. We generate a nonstationary TS of length 8192 points with $\alpha=0.4$ and two values of β. In the first 4096 iterations we use $\beta=0.6$, and in the second 4096 iterations we set $\beta=0.8$. We record $u+v$, then we subtract the mean value and normalize to unit variance separately each one of the two parts; cf. Schreiber [3]. The total signal of 8192 points was divided in several segments S_i with 1000 points (there are an overlap of 900 points between two consecutive segments). Thus, we have a nonstationary event in the middle of the TS that is very hard to detect, because observables like mean, variance, and maximal Lyapunov exponent are constant by construction. Figure

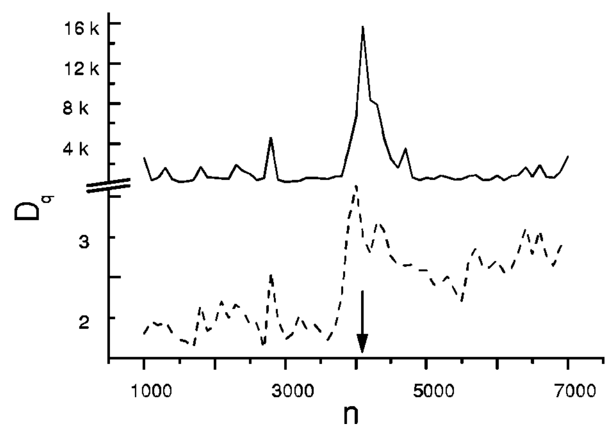

FIG. 5. D_6 (solid line) and D_2 (dashed line) between two nonoverlapping subsequent segments of TS with a nonstationary event at 4096.

5 shows $D_2(S_i:S_j)$ and $D_6(S_i:S_j)$, between the two nearest nonoverlapping segments S_i and S_j $(j=i+10)$. This example shows also that the parameter q plays a role of a nonlinear gain parameter. We can see that for $q=6$ the divergence was able to detect precisely when the small change in β occur, while for $q=2$ we have a poor discrimination power. We used resolution $\varepsilon=0.1155$ in the computation. We have a window resolution with similar results for ε in $[0.075,0.15]$.

We want to mention that, as any other methods, the q-divergence sensitively depends on any translation or rotation of the same data set with respect to the other. However, D_q remains invariant when the same transformation is applied over both data sets.

We have introduced the q-divergence as a measure of dissimilarity of two finite sets. Comparing clean and noisy attractors, the q-divergence has a characteristic length scale in a clear correspondence with the level of noise. Furthermore, the q-divergence decreases exponentially when η increases. Also this tool promises to be useful for detecting a nonstationary event in a TS, even in very hard conditions. Thus, it will be seen that some interesting physical insight is gained by recourse to this type of dissimilarity measure.

The author acknowledges the financial support of FAPESP Grant No. 99/07186-3. The author thanks C.P. Malta for a careful reading of the manuscript and her useful comments.

[1] H. D. I. Abarbanel, *Analysis of Observed Chaotic Data* (Springer-Verlag, New York, 1996); H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, England, 1997).

[2] R. Manuca and R. Savit, Physica D **99**, 134 (1996).

[3] T. Schreiber, Phys. Rev. Lett. **78**, 843 (1997).

[4] T. Schreiber and A. Schmitz, Phys. Rev. Lett. **79**, 1475 (1997).

[5] H. Kantz, Phys. Rev. E **49**, 5091 (1994); C. Diks, W. R. van Zwet, F. Takens, and J. DeGoede, *ibid.* **53**, 2169 (1996).

[6] A.M. Albano, P.E. Rapp, and A. Passamante, Phys. Rev. E **52**, 196 (1995).

[7] J. Aczél and Z. Daròczy, *On Measures of Information and Their Characterizations*, Mathematics in Science and Engineering, Vol. 115 (Academic Press, 1975).

[8] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Chicago, 1949).

[9] A. Rényi, *Probability Theory* (North-Holland, Amsterdam, 1970).

[10] M. E. Havrda and F. Charvát, Kybernetika (Praha) **3**, 30 (1997).

[11] P. Grassberger, Phys. Lett. A **97**, 227 (1983).

[12] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).

[13] I. Csiszár, Periodica Mathematica Hungarica **2**, 191 (1972).

[14] S. Kullback, *Information Theory and Statistics* (Wiley, New York, 1959).

[15] P. Grassberger and I. Procaccia, Phys. Rev. Lett. **50**, 346 (1983).

[16] J. Theiler, Phys. Rev. A **34**, 2427 (1986).

[17] The signal-to-noise ratio η in decibels (dB) is defined as $10\log_{10}[\sigma(\text{signal})/\sigma(\text{noise})]$.