Vol 18 No 9, September 2009
1674-1056/2009/18(09)/4042-07

Chinese Physics B

© 2009 Chin. Phys. Soc.
and IOP Publishing Ltd

An improvement on measure methods of the complexity theory and its applications*

Wang Fu-Lai(王福来)† and Yang Hui-Huang(杨辉煌)

Department of Mathematics and Statistics, Zhejiang University of Finance and Economics, Hangzhou 310012, *China*

(Received 26 July 2008; revised manuscript received 27 April 2009)

A new method is proposed to transform the time series gained from a dynamic system to a symbolic series which extracts both overall and local information of the time series. Based on the transformation, two measures are defined to characterize the complexity of the symbolic series. The measures reflect the sensitive dependence of chaotic systems on initial conditions and the randomness of a time series, and thus can distinguish periodic or completely random series from chaotic time series even though the lengths of the time series are not long. Finally, the logistic map and the two-parameter Henón map are studied and the results are satisfactory.

Keywords: complexity theory, complexity, dynamic system, chaos
PACC: 9260X

1. Introduction

In a study of the information transmission of the human brain cortex, Xu[1,2] pointed out that the length of the series of information transmission is short while the estimation of dynamic parameters such as correlation dimensions (D_2) depends on series of long length. A series of long length may present a false appearance because many parameters might be a mixed result of the parameters in different periods. In fact, for other time series, Hsieh[3] also pointed out that the above phenomena existed in economic time series.

Now that the theory of chaos has not been mature and the algorithm of parameters such as correlation dimensions, Lyapunov exponent and K-entropy rely on the length of time series, many papers have appeared to define complexity from different aspects.[4−10] However, all these have some deficiencies, so some improvements still wait to be done.[1]

We can extract least information from a time series if it is completely random, so Xu[1] proposed a new definition of complexities C_1 and C_2 which are different from the hierarchical approach by D'Alessandro and Politi,[6] where the first-order C_1 represents the randomness of the sequence and second-order C_2 seems to be correlated to the fractal dimension. According to D'Alessandro and Politi,[6] the bigger C_1 is, the more random the dynamic system is; and the bigger C_2 is, the more complex the system is. However, some improvements of the definition and algorithm suggested by Xu[1] still need to be made because different kinds of time series cannot be distinguished by the algorithm which still depends on long series.

No significant changes have been made with the definition of complexities in Ref.[1] but there are some new definitions of forbidden words. In Refs.[11] and [12], a discrete information source is considered emitting a series of observable values $[x_1, x_2, \ldots, x_N]$, in order of time; it is possible to split the data into overlapping sets of length d, and study their order patterns, e.g., $x_1 < x_2 < \ldots < x_d$. Every group of d adjacent values forms a certain permutation Π. When the series corresponds to a deterministic variable, there will be some patterns that cannot be encountered in the data due to the underlying deterministic structure. In all cases of deterministic systems the number of forbidden patterns is much higher (two orders of magnitude) than that expected in the random case. Compared with the complexities suggested by Xu, Amigö's algorithm[11] shows the following weak points: (A) The forbidden patterns correspond to no clear physical index or meaning while Xu's complexities correspond to correlative dimensions. (B) Periodic and chaotic series cannot be distinguished from each other because

*Project supported by the Scientific Research Fund of Zhejiang Provincial Education Department of China (Grant No 20070814) and the National Natural Science Foundation of China (Grant No 10871168).
†Corresponding author. E-mail: flyerwon@sina.com

they are both considered as only deterministic. (C) The forbidden patterns in Ref.[5] cannot measure the complexities or determinism of different time series because the forbidden patterns are only an indicator of the deterministic nature of the series.

In this paper, a new method is proposed to transform the time series gained from a dynamic system to a symbolic series which extracts both the overall and the local information of the time series. Based on the transformation, the C_1 and C_2 complexities are redefined. Based on the new method, different kinds of time series can be distinguished and the algorithm depends on much shorter series than the algorithm in Ref.[1].

2. New measure method and its applications

Let $[S_i]_{i=1}^N$ be a symbolic sequence obtained from the dynamic system $[x_i]_{i=1}^N$ (or the time series) and suppose that N is large enough. For simplicity, let S_i be 0 or 1. Because there are 2^n words with length n, we may take subsequences $[S_i]_{i=1}^{n2^n}$. Let $N_{\rm a}(n)$ be the numbers of words with length n in the subsequences $[S_i]_{i=1}^{n2^n}$. Let $N_{\rm f}(n)$ be the numbers of forbidden words with word length n, where forbidden words mean that some subsequences $[S_i]_{i=1}^{n2^n}$ with length $n-1(s_1, s_2, \ldots, s_{n-1})$ can exist in the subsequences but the subsequences with length $n(s_1, s_2, \ldots, s_{n-1}, s_n)$. Here s_n equals 0 or 1. Complexities C_1 and C_2 are defined as follows:[1]

$$C_1 = \lim_{n\to\infty} \frac{\log_2 N_{\rm a}(n)}{n}, \quad C_2 = \lim_{n\to\infty} \frac{\log_2 N_{\rm f}(n)}{n}.$$

In Ref.[1], C_1 represents the randomness of the sequence and C_2 seems to be correlated to the fractal dimension. Consequently, if the dynamic system is completely random, we have $C_1 = 1$ and $C_2 = 0$, which means that random motion is random but not complex. According to the definitions of C_1 and C_2, when we put $C_1 = 1$, we mean that this measure takes values that converge to 1 over the period under study. If the dynamic system is in a periodic motion, we have $C_1 = 0$ and $C_2 = 0$, which means that the normal periodic motion is neither random nor complex. Complexity C_2 might vary with N and two complexities C_2's can be compared only when the lengths of the two sequences are the same. For convenience, we call the algorithm of C_1 and C_2 in Ref.[1] algorithm XJH.

In order to compare the new method with the algorithm XJH where the logistic map is analysed because of the fact that its dynamic properties vary with a parameter, we would include the logistic map as one of the examples in our analysis with the new method in this paper. The logistic map is given as follows:

$$x_{n+1} = ux_n(1 - x_n), \tag{1}$$

where u is a parameter and $u \in [0, 4]$.

2.1. The deficiencies of algorithm XJH

The deficiencies of algorithm XJH are as follows:

(A) The information of the dynamic system $[x_i]_{i=1}^N$ is too coarse-grained when $[x_i]_{i=1}^N$ is transformed into a symbolic sequence.

It is well known that it is the first and key step to transform a time series into a symbolic sequence when we calculate the complexity of the time series. Much information, especially the local information, would be lost if the information of the time series is too coarse-grained and to make the matter worse the dynamic behaviour of the system may be changed. The present methods of transformation can be classified into the mean method and the extreme value method. By the mean method, we compare each point of the time series with the mean. If a point is no less than the mean, we have a symbol 1 and a 0 otherwise; therefore we obtain a symbolic sequence of values 0 and 1. In the extreme value method, we compare every point of the time series with its next point. If a point is no less than the next one we obtain a symbol 1 and a 0 otherwise, and therefore we obtain a symbolic sequence of values 0 and 1. The former method neglects the local information of the time series while the latter neglects the overall information. With the mean method, Xu obtained $C_1 = 1$ and that C_2 approaches 0 both in the logistic map (1) where $u = 4$ and in a completely random system. The logistic system is a chaotic one when $u = 4$, which is quite different from a completely random system because a chaotic system has close returns[13−15] and can be predicted accurately in a short time.

(B) As far as C_2 is concerned, a periodic system and a chaotic system are not distinguished from each other efficiently. According to Ref.[1], C_2 should be equal or near to 0 in a periodic system, but sometimes C_2 is near to the value of C_2 in a chaotic system. Moreover, by algorithm XJH $C_2 = 0$ holds in a completely random system but the facts are that

$C_2 = 0.2902$ when $N = 1000$ and $C_2 = 0$ only when n is no less than about 1500 (in Ref.[1], $N = 2048$ and $N = 4096$), indicating that algorithm XJH requires a time series with a long length.

(C) According to Ref.[1], C_2 seems to be correlated to the fractal dimension. But in the logistic system (1), when $u = 4$, C_2 is significantly decreased by using algorithm XJH, and $C_2 \neq 0$ in a periodic system. The result is not satisfactory.

(D) If a periodic system is disturbed by weak noises, all sequences are possible as $N \to \infty$ in the disturbed system. By algorithm XJH, $C_1 = 1$ and $C_2 = 0$ and thus the disturbed system will be regarded as a completely random system.

2.2. New definitions of complexities C_1 and C_2

In this section a new method of transformation as well as a new definition of C_1 and C_2 is proposed.

2.2.1. Transformation of time series into symbolic sequences

An efficient method of transformation should meet the following requirements:

(A) The symbolic sequence can reflect the transformation process of the dynamic system as a whole and in detail.

(B) The symbolic sequence can reflect the sensitive dependence of the chaotic time series under initial circumstances and the predictability during a short period.

(C) The C_2 complexity can indicate the dimensions of regular fractals and the change of dimensions when the properties change with the change of parameters.

Let $(S_1S_2\ldots)$ be a semi-dynamic system composed of r symbols as $A_r = \{1, 2, \ldots, r\}$, where $S_i \in A_r$.

Definition 1 A symbolic series such as $(a_1a_2\ldots a_n)$ will be called a word with length of n, where $a_i \in A_r$ and n can be finite or infinite.

Definition 2 A word is called an allowed word if and only if it appears in $(S_1S_2\ldots)$, where $(S_1S_2\ldots)$ is obtained by transformation from time series of a dynamic system according to some certain rule. From symbolic dynamics the distance between two words $(a_1a_2\ldots a_n)$ and $(b_1b_2\ldots b_n)$ is

$$r = \sum_{i=1} \frac{1}{2^i} h(i), \tag{2}$$

where

$$h(i) = \begin{cases} 0, & \text{if } a_i = b_i, \\ 1, & \text{if } a_i \neq b_i. \end{cases}$$

In dynamics, if coarse-grained, the two words with a certain number (say n) of the same front symbols can be classified. Because of the probabilistic stability of chaotic attractors and the asymptotical stability[16] the number of points in every classification will increase with increasing n. Under the same patterns of coarse grain, the greater the numbers of classifications, the more complex the dynamic system is.

2.2.2. The steps of transformation from a time series to a symbolic sequence

The steps of transformation from a time series to a symbolic sequence are as follows:

Step 1 Divide the time series $[x_i]_{i=1}^N$ obtained from a dynamic system into two parts A and B by the mean and suppose that the points in B are no less than the mean and the points in A are less than the mean.

Step 2 Take the symbol of x_1 as $s_1 = 1$ ($s_1 = 3$) if $x_1 \in A$ ($x_1 \in B$), where $s_1 = 1$ is the initial value of the symbolic sequence.

Step 3 The symbol of $x_{i+1} (i = 1, 2, \ldots, N-1)$ is defined according to the rules as shown in Table 1, where N_1 is the number of symbols.

Table 1. Rules for transformation ($N_1 = 4$).

x_{i+1} in time series	relations of x_i and x_{i+1}	symbols of x_{i+1}
A	$x_{i+1} \geq x_i$	2
A	$x_{i+1} < x_i$	1
B	$x_{i+1} \geq x_i$	4
B	$x_{i+1} < x_i$	3

Step 4 Classify A into A_1 and A_2 and B into B_1 and B_2 by their means and suppose $B_2 > B_1 > A_2 > A_1$. Similarly, we obtain Table 2.

Step 5 The above steps 3 and 4 can be continued if N is large enough. The greater the N and N_1 are, the more accurately the symbolic sequence represents the information of the time series.

From the above steps, we see that the transformation method extracts the information of the dynamic system both as a whole and in detail.

Table 2. Rules for transformation ($N_1 = 8$).

x_{i+1} in time series	relations of x_i and x_{i+1}	symbols of x_{i+1}
A_1	$x_{i+1} \geq x_i$	2
A_1	$x_{i+1} < x_i$	1
A_2	$x_{i+1} \geq x_i$	4
A_2	$x_{i+1} < x_i$	3
B_1	$x_{i+1} \geq x_i$	6
B_1	$x_{i+1} < x_i$	5
B_1	$x_{i+1} \geq x_i$	8
B_1	$x_{i+1} < x_i$	7

2.2.3. Proposal of algorithm SD (the algorithm of symbolic dynamics)

Based on the above transformation method, the C_1 and C_2 complexities in algorithm SD are defined as follows:

Definition 3 Let $N_{\rm a}(n)$ be the number of allowed words with length n in sequences $[S_i]_{i=1}^{n2^n}$ and N_1 the number of symbols in the symbolic sequences. We have

$$C_1 = \lim_{n\to\infty} \frac{\log_{N_1} N_{\rm a}(n)}{n}. \tag{3}$$

To define C_2, in a different sense we determine $(s_1, s_2, \ldots, s_n)$ as a bifurcation word if both the words $(s_1, s_2, \ldots, s_n, s_{n+1})$ and $(s_1, s_2, \ldots, s_n, s'_{n+1})$ appear in $[S_i]_{i=1}^{n2^n}$, where s_{n+1} and s'_{n+1} are different from each other.

Definition 4 Let $B_{\rm a}(n)$ be the number of the bifurcation words without repetition. We have

$$C_2 = \lim_{n\to\infty} \frac{\log_{N_1} B_{\rm a}(n)}{n}. \tag{4}$$

Obviously, a bifurcation word is one of the allowed words in the system and thus $B_{\rm a}(n) \leq N_{\rm a}(n)$ holds. $B_{\rm a}(n)$ is the sum of bifurcation words (or subsequences) with length n in a system. Bifurcation is among the most important factors to describe the complexity of a system.

For a random system, though bifurcation subsequences exist, the bifurcations take place with very small probability because no deterministic factor functions in the system.

For a periodic system, no bifurcation takes place so we have $C_2 = 0$.

For a periodic system disturbed by weak noises, the system can be considered as periodic if the noise is weak enough. From the above analysis, we know $C_2 = 0$.

An example of a periodic system disturbed with weak noises is as follows:

$$x_{n+1} = ux_n(1 - x_n) + \xi_n, \tag{5}$$

where u is a parameter with $u \in [0, 3.5]$, and $\xi_n \sim N(0, 0.0005)$.

For completely random sequences, because all subsequences with length n will appear in the sequence if N is long enough, we have $M_{\rm a}(n) = N_1^n$ and thus $C_1 = 1$. The probability for two of the same subsequences to occur in the same time series is no greater than $\frac{1}{N_1^n}$. According to probability theory we have $\lim\limits_{n\to\infty} \frac{\log_{N_1} B_{\rm a}(n)}{n} = 0$.

2.3. The application of algorithm SD to the logistic map and other time series

(i) As mentioned in Subsection 2.1(B), in algorithm XJH $C_2 = 0$ holds in a completely random system but the facts are that $C_2 = 0.2902$ when $N = 1000$ and $C_2 = 0$ only when n is no less than about 1500 (in Ref.[1], $N = 2048$ and $N = 4096$), which means algorithm XJH requires a time series with long lengths. In order to make a comparison with Ref.[1], we take the logistic system as an example and let N be a number much less than 2048, say $N = 1000$. We may as well suppose $N_1 = 8$ and we must say that it makes little difference if we suppose $N_1 = 4$. Generally, we may choose a larger value for N_1 if N is large. When calculating C_1, we assume $n = 8$. For convenience in comparison, we have all the C_1 complexities divided by the C_1 of a completely random system and have all the C_2 complexities by the maximum C_2 in the logistic map with u ranging from 0 to 4.

In order to make a comparison with algorithm XJH, we take 2048 points and delete 1000 transient points for computation. By programming with Matlab 7.0, we have the following results which are shown in Figs.1(a)–1(d).

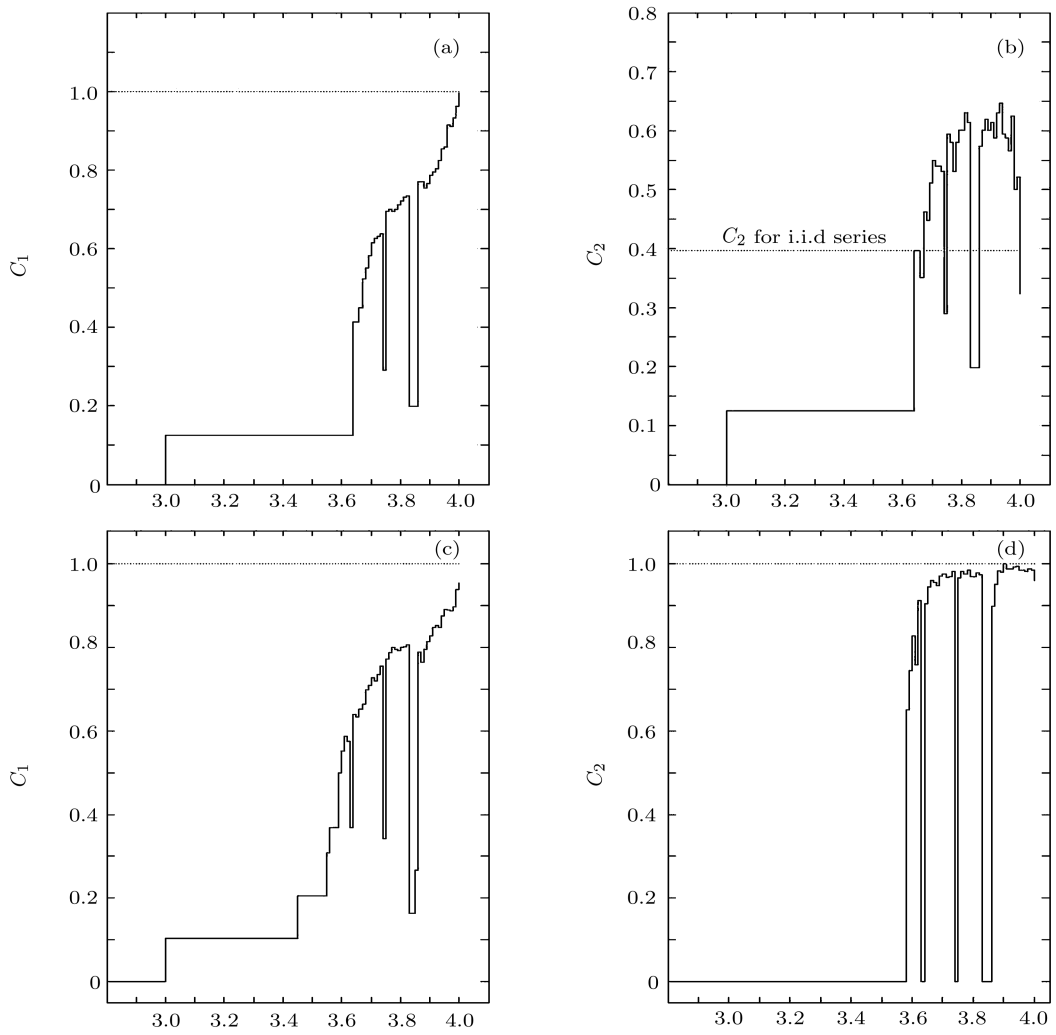


Fig.1. The C_1 complexities (a) and C_2 complexities (b) of the logistic map obtained by algorithm XJH ($N = 1000$); The C_1 complexities (c) and C_2 complexities (d) of the logistic map obtained by algorithm SD ($N = 1000$).

For logistic maps (1), the systems given by logistic maps (1) are periodic when $u < u_\infty = 3.56\ldots$, where u_∞ is the infinite periodic point. Logistic systems begin to enter the region of chaos where, however, periodic windows exist. At periodic windows, the logistic systems become periodic again. Values of u such as 3.73, 3.83, 3.84, 3.85 are periodic windows. Because algorithm XJH cannot extract efficiently local or overall information when transforming time series into symbolic series, we can see from Figs.1(a) and 1(b) that the algorithm is unable to distinguish periodic systems from chaotic systems when $3 \leq u \leq 3.63$. It is known that logistic maps (1) are periodic when $u < 3.56\ldots$ and chaotic when $3.56\ldots \leq u \leq 3.63$. In algorithm XJH, either C_1's or C_2's remain the same when $3 \leq u \leq 3.63$. It is intended by Xu[1] that for periodic systems, the more C_2's approach zero, the more desirable. However, C_2's for periodic systems by algorithm XJH do not come near enough to zero until N is no less than about 1500.

(ii) In another example, we discuss a two-parameter Henón map as follows:

$$\begin{aligned} x_{n+1} &= p - x_n^2 + qy_n, \\ y_{n+1} &= x_n, \end{aligned} \tag{6}$$

where p, q, x and y are real numbers, and p and q are the two parameters.

By programming with Matlab 7.0, we have the results which are shown in Figs.2(a)–2(e).

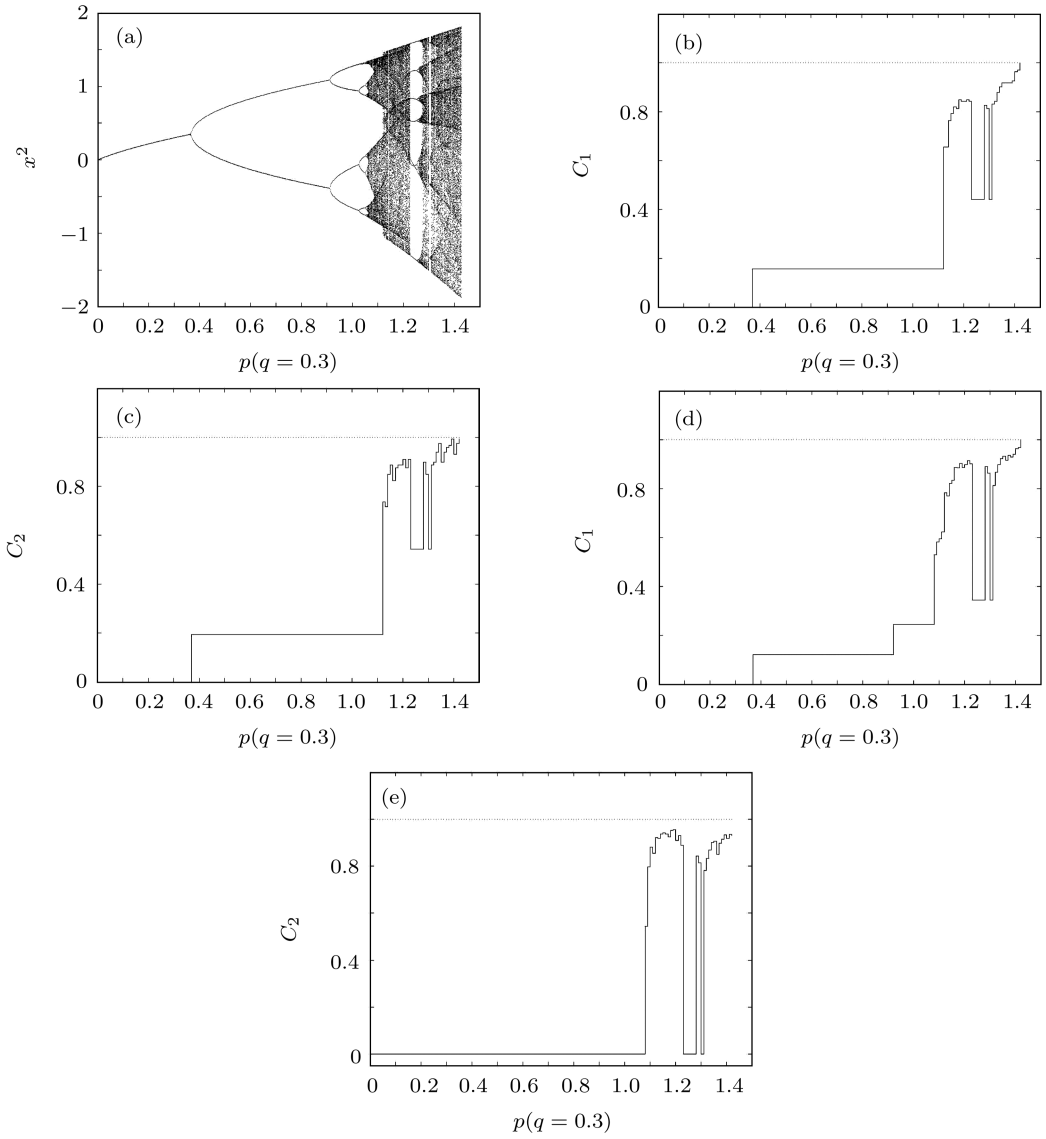


Fig.2. (a) Plot of Henón maps (6) for different values of parameter p; (b) The C_1 complexities of Henón maps (6) by algorithm XJH ($N = 1000$); (c) The C_2 complexities of Henón maps (6) by algorithm XJH ($N = 1000$); (d) The C_1 complexities of Henón maps (6) by algorithm SD ($N = 1000, N_1 = 8, n = 8$); (e) The C_2 complexities of Henón maps (6) by algorithm SD ($N = 1000, N_1 = 8, n = 6$).

For two-parameter Henón maps (6), the systems are periodic when $p < p_\infty = 1.07\ldots$, where p_∞ is the infinite periodic point. Henón systems begin to enter the region of chaos where, however, periodic windows exist. At periodic windows, the Henón systems become periodic again. Values of p such as 1.23, 1.24, 1.27, 1.30 are periodic windows. we can see from Figs.2(b) and 2(c) that the algorithm is unable to distinguish periodic systems from chaotic systems when $0.37 \leq p \leq 1.11$. In fact, Henón maps (6) ($q = 0.3$) are periodic when $p < 1.07\ldots$ and chaotic when $1.07\ldots < p \leq 1.11$. By algorithm XJH, either C_1's or C_2's remain the same when $0.37 \leq p \leq 1.11$.

Both in logistic maps (1) and Henon maps (6) ($q = 0.3$), algorithm XJH is not very successful in measuring the complexity of the systems, while with algorithm SD, the C_2's for periodic and random systems are definitely 0, so the C_2 complexity can distinguish chaotic systems from periodic and random ones. With algorithm SD, the C_1's are definitely 1 for random systems and far less than 1 for periodic ones, so the complexity C_1 can distinguish periodic systems from random ones. With C_1 and C_2, the complexity of any system can be measured.

From the definitions of C_1 and C_2, we know that the longer the length of the time series is, the greater the values N_1 and n can take are and the more accurately the C_1 and C_2 given by algorithm SD can characterize the complexity of any dynamic system.

3. Conclusions

We conclude that one could use the primitive history of a sequence, rather than (or in addition to) the exhaustive one, as a means for evaluating the complexity of a given sequence. With the algorithm SD, we do not demand a series with length longer than 1000 to evaluate the complexity. Furthermore, with data becoming longer, N_1 and n can take greater values and the complexity of the system will be better characterized. The algorithm SD is effective and easy to carry out in practice.

Acknowledgements

The author thanks Professor Yu Bohua (Department of Mathematics, Zhejing University, China) for his help in programming.

References

[1] Xu J H 1994 *Chaos, Solitons and Fractals* **4** 2111
[2] Xu J H and Wu X B 1994 *Science in China* B **1** 57
[3] Hsieh D A 1991 *The Journal of Finance* **46** 1839
[4] Auerbach D and Procaccia I 1990 *Phys. Rev.* A **41** 6602
[5] Crutchfield J P and Young K 1989 *Phys. Rev. Lett.* **63** 105
[6] D'Alessandro G and Politi A 1990 *Phys. Rev. Lett.* **64** 1609
[7] Hao B L 1991 *Physica* D **51** 161
[8] Kasper F and Schuster H G 1987 *Phys. Rev.* A **36** 842
[9] Liu Y P, Chen Z Y, Fan Y W, Ba W Z, Guo Q, Lu W, Tang X Q and Du Y Z 2008 *Chin. Phys.* B **17** 3156
[10] Zhang D H 2008 *Acta Phys. Sin.* **57** 3359 (in Chinese)
[11] Amigö J M, Zambrano S and Sanjuan M A F 2007 *Europhys. Lett.* **79** 50001
[12] Massimiliano Z 2008 *Chaos* 013119 1
[13] Gilmore C G 2001 *Global Finance Journal* **12** 35
[14] Gilmore C G 1993 *Journal of Economic Behavior and Organization* **22** 209
[15] Gilmore C G 1996 *Journal of Business and Finance Accounting* **23** 1357
[16] Lasota A and Mackey M 1994 *Chaos, Fractals and Noise,* second ed (New York: Springer-Verlag) p311