

Measuring complexity using FuzzyEn, ApEn, and SampEn

Weiting Chen^{a,*}, Jun Zhuang^b, Wangxin Yu^b, Zhizhong Wang^b

^a Software Engineering Institute, East China Normal University, 3663 North Zhongshan Road, Shanghai 200062, China

^b Department of Biomedical Engineering, Shanghai Jiaotong University, 800 Dong Chuan Road, Shanghai 200240, China

Received 6 July 2007; received in revised form 17 April 2008; accepted 20 April 2008

Abstract

This paper compares three related measures of complexity, ApEn, SampEn, and FuzzyEn. Since vectors' similarity is defined on the basis of the hard and sensitive boundary of Heaviside function in ApEn and SampEn, the two families of statistics show high sensitivity to the parameter selection and may be invalid in case of small parameter. Importing the concept of fuzzy sets, we developed a new measure FuzzyEn, where vectors' similarity is defined by fuzzy similarity degree based on fuzzy membership functions and vectors' shapes. The soft and continuous boundaries of fuzzy functions ensure the continuity as well as the validity of FuzzyEn at small parameters. The more details obtained by fuzzy functions also make FuzzyEn a more accurate entropy definition than ApEn and SampEn. In addition, similarity definition based on vectors' shapes, together with the exclusion of self-matches, earns FuzzyEn stronger relative consistency and less dependence on data length. Both theoretical analysis and experimental results show that FuzzyEn provides an improved evaluation of signal complexity and can be more conveniently and powerfully applied to short time series contaminated by noise.

© 2008 IPEM. Published by Elsevier Ltd. All rights reserved.

Keywords: Complexity; Nonlinear; ApEn; SampEn; FuzzyEn

1. Introduction

Recent years have seen rapid development of various complexity measures such as dimensions, Lyapunov exponents and entropies. The calculations, however, are frequently confronted with the problem of insufficient number of data points. Also, most dimension and entropy definitions show application limitations to experimental time series since all recorded data are to a certain degree contaminated by noise when a 2% noise is serious enough to prevent accurate estimation [1]. To solve the problems of short and noisy recordings in physiological signals, Pincus [2] presented approximate entropy (ApEn) as a measure of complexity that is applicable to noisy, medium-sized datasets. Once introduced, the family of statistics has been widely applied to a variety of physiological and clinical datasets such as genetic sequences, hormone pulsatility, respiratory patterns, heart rate variability, electrocardiogram, electroencephalography, and electromyography [3–14], and has shown its superiority to most complex-

ity measures such as fractal dimension, Kolmogorov–Sinai entropy, and spectral entropy [15,16]. However, ApEn(m , r , N) is a biased statistic [2,17]. It lacks relative consistency and the result shows much dependence on data length. Richman and Moorman [18] investigated the mechanism responsible for the bias and proposed another statistic, sample entropy (SampEn), to relieve the bias caused by self-matching. SampEn displays relative consistency and less dependence on data length. Nevertheless, the similarity definition of vectors in SampEn is based on Heaviside function as in ApEn. Due to the inherent flaws of Heaviside function, problems still exist in the validity of the entropy definition, especially when small parameters are involved [18].

We [19] developed most recently a new related family of statistics, fuzzy entropy (FuzzyEn), as a measure of complexity. FuzzyEn derives from the concept of Zadeh's fuzzy sets and that is also why the statistic is so named. In this paper, we discussed certain flaws caused by Heaviside function in ApEn and SampEn, and then demonstrated the performance of FuzzyEn compared with ApEn and SampEn through analysis and experiments. Results show that FuzzyEn is a more accurate complexity measure. It not only owns stronger rel-

* Corresponding author. Tel.: +86 139 1828 3439.

E-mail address: waiting.yi@gmail.com (W. Chen).

ative consistency and less dependence on data length, which means less bias, but also achieves continuity, freer parameter selections and more robustness to noise.

2. Description of ApEn, SampEn, and FuzzyEn

2.1. ApEn [2]

For an N sample time series $\{u(i):1 \leq i \leq N\}$, given m , form vector sequences \mathbf{X}_1^m through \mathbf{X}_{N-m+1}^m as

$$\mathbf{X}_i^m = \{u(i), u(i+1), \dots, u(i+m-1)\}, \quad i = 1, \dots, N-m+1 \quad (1)$$

where m is the length of compared window. For each $i \leq N-m+1$, let $C_i^m(r)$ be $(N-m+1)^{-1}$ times the number of vectors \mathbf{X}_j^m within r of \mathbf{X}_i^m . By defining

$$\phi^m(r) = (N-m+1)^{-1} \sum_{i=1}^{N-m+1} \ln C_i^m(r) \quad (2)$$

where \ln is the natural logarithm, Pincus defined the parameter:

$$\text{ApEn}(m, r) = \lim_{N \rightarrow \infty} [\phi^m(r) - \phi^{m+1}(r)] \quad (3)$$

2.2. SampEn [18]

SampEn differs from ApEn mainly in two ways: (1) SampEn does not count self-matches; (2) SampEn does not use a template-wise approach. Define $B_i^m(r)$ as $(N-m-1)^{-1}$ times the number of vectors \mathbf{X}_j^m within r of \mathbf{X}_i^m , where j ranges from 1 to $N-m$, and $j \neq i$ to exclude self-matches, and then define

$$B^m(r) = (N-m)^{-1} \sum_{i=1}^{N-m} B_i^m(r) \quad (4)$$

Similarly, define $A_i^m(r)$ as $(N-m-1)^{-1}$ times the number of vectors \mathbf{X}_j^{m+1} within r of \mathbf{X}_i^{m+1} , where j ranges from 1 to $N-m$ ($j \neq i$), and set

$$A^m(r) = (N-m)^{-1} \sum_{i=1}^{N-m} A_i^m(r) \quad (5)$$

The parameter $\text{SampEn}(m, r)$ is then defined as $\lim_{N \rightarrow \infty} \{-\ln[A^m(r)/B^m(r)]\}$, which can be estimated by the statistic:

$$\text{SampEn}(m, r, N) = -\ln \left[\frac{A^m(r)}{B^m(r)} \right] \quad (6)$$

Set $B = \{[(N-m-1)(N-m)]/2\} B^m(r)$ and $A = \{[(N-m-1)(N-m)]/2\} A^m(r)$, so that B is the total number of template matches of length m and A is the total number of forward matches of length $m+1$, and then SampEn can be expressed as $-\ln(A/B)$.

2.3. FuzzyEn [19]

Like SampEn, FuzzyEn excludes self-matches and considers only the first $N-m$ vectors of length m to ensure that \mathbf{X}_i^m and \mathbf{X}_i^{m+1} are defined for all $1 \leq i \leq N-m$. For time series $\{u(i):1 \leq i \leq N\}$, form vectors:

$$\{\mathbf{X}_i^m = \{u(i), u(i+1), \dots, u(i+m-1)\} - u_0(i), \quad i = 1, \dots, N-m+1\} \quad (7)$$

where \mathbf{X}_i^m represents m consecutive u values, commencing with the i th point and generalized by removing a baseline:

$$u_0(i) = m^{-1} \sum_{j=0}^{m-1} u(i+j) \quad (8)$$

Given vector \mathbf{X}_i^m , calculate the similarity degree D_{ij}^m of its neighboring vector \mathbf{X}_j^m to it through the similarity degree defined by a fuzzy function¹:

$$D_{ij}^m = \mu(d_{ij}^m, r) \quad (9)$$

where d_{ij}^m is the maximum absolute difference of the corresponding scalar components of \mathbf{X}_i^m and \mathbf{X}_j^m . For each vector \mathbf{X}_i^m ($i = 1, \dots, N-m+1$), averaging all the similarity degree of its neighboring vectors \mathbf{X}_j^m ($j = 1, \dots, N-m+1$, and $j \neq i$), we get:

$$\phi_i^m(r) = (N-m-1)^{-1} \sum_{j=1, j \neq i}^{N-m} D_{ij}^m \quad (10)$$

Construct

$$\varphi^m(r) = (N-m)^{-1} \sum_{i=1}^{N-m} \phi_i^m(r) \quad (11)$$

and

$$\varphi^{m+1}(r) = (N-m)^{-1} \sum_{i=1}^{N-m} \phi_i^{m+1}(r) \quad (12)$$

and then we can define the parameter FuzzyEn(m, r) of the time series as

$$\text{FuzzyEn}(m, r) = \lim_{N \rightarrow \infty} [\ln \varphi^m(r) - \ln \varphi^{m+1}(r)] \quad (13)$$

which, for finite datasets, can be estimated by the statistic:

$$\text{FuzzyEn}(m, r, N) = \ln \varphi^m(r) - \ln \varphi^{m+1}(r) \quad (14)$$

¹ Different from that in [19], here we give no detailed fuzzy membership function $\mu(d_{ij}^m, r)$, because any function that possesses the following desired properties can be chosen: (1) being continuous so that the similarity does not change abruptly; (2) being convex so that self-similarity is the maximum.

3. Comparison of FuzzyEn, ApEn and SampEn

When judging whether a vector \mathbf{X}_j is similar to \mathbf{X}_i within a tolerance r , either in ApEn or in SampEn, the rule can be expressed simply as follows:

$$\theta(d_{ij}, r) = \begin{cases} 1, & \text{if } d_{ij} \leq r \\ 0, & \text{if } d_{ij} > r \end{cases} \quad (15)$$

where d_{ij} represents the distance between the two vectors, and $\theta(\cdot)$ is Heaviside function serving as a two-state classifier. The boundary of Heaviside function is rigid: the contributions of all the data points inside it are treated equally, while the data points just outside the boundary are left out. Take the similarity of the three points $d1$, $d2$, and $d3$ to the original point (Fig. 1) for example. The two points $d2$ and $d3$ within the boundary of Heaviside function are considered the same and both similar to the original point; whereas another point $d1$, very close to $d2$, is considered dissimilar because it falls just outside the boundary. A slight increase in r will make the boundary enclose the point $d1$, and then $d1$ will be identified with the other two points. So the contribution of Heaviside function depends totally on the tolerance r and shows high sensitivity to the change of r or data point position. Consequently, both ApEn and SampEn may vary significantly with a slight change of the parameter r , and thus may be discontinuous.

Let \mathbf{X} be a space. In a conventional two-state classifier, any input pattern $x \in \mathbf{X}$ is judged its belongingness to a given class C by whether it satisfies precise properties required for

membership. In the real physical world, however, boundaries between classes may be ambiguous, and the membership of an input pattern to a given class is always imprecise and uncertain. To describe such input–output relations which can hardly be achieved by the conventional two-state classifier, Zadeh [20] proposed the concept of fuzzy sets. By introducing the “membership degree” with a fuzzy function μ_C , Zadeh’s theory provided a mechanism for measuring the degree to which a pattern belongs to a given class: the nearer the value of $\mu_C(x)$ to unity, the higher the membership grade of x in the class. In FuzzyEn, we imported the concept of fuzzy sets to measure the similarity degree of two vectors.

The family of exponential function $\exp(-(d_{ij}/r)^n)$ is a kind of fuzzy functions easy to be understood, and it functions well as a fuzzy membership function. In this paper, we employed the exponential function as the fuzzy function in FuzzyEn. In the exponential function there is no rigid boundary and all the data points, whose membership is decided by point positions together with the parameter r and n , are considered as its members. For each vector \mathbf{X}_i , the contribution of each exponential function around it can be viewed as the fuzzy membership to indicate the similarity of its neighboring vector: the closer the neighboring vector \mathbf{X}_j is, the more similar \mathbf{X}_j is to \mathbf{X}_i ; and the similarity degree of \mathbf{X}_j to \mathbf{X}_i is almost zero when \mathbf{X}_j is far away from \mathbf{X}_i . Under the rule given by the exponential function, the similarity of the three points to the original point in Fig. 1 is considered differently according to their closeness, and a slight change of r will not influence the results much. So the contribution of exponential function is not so sensitive to the change of r and point position, and will change more gently if r changes. Accordingly, FuzzyEn is continuous and will not change drastically when there are slight changes of r .

Another difference of FuzzyEn from ApEn and SampEn is the construction of m -dimensional vectors from one-dimensional time series. In both ApEn and SampEn, vectors $\{\mathbf{X}_i^m, i = 1, \dots, N - m + 1\}$ are formed directly from the original m consecutive u values. However, some short-term physiology time series may cause fluctuation in vector coordinates and hide the similarity between certain vectors [13]. To catch such similarity, we generalized vectors in FuzzyEn by removing a baseline. After the generalization of vector series, it is vectors’ shapes rather than their absolute coordinates that determine their similarity, and thus we can get vectors’ similarity more accurately through fuzzy set.

As in SampEn and ApEn, parameters N , m and r in FuzzyEn are the length of the time series to be analyzed, the length of the compared window, and the width of the boundary for similarity measurement, respectively. The selections of these parameters are also like that in ApEn and SampEn [19]. When the family of exponential function $\exp(-(d_{ij}/r)^n)$ is selected as the fuzzy function, there is one more parameter n , the gradient of the boundary of the exponential function, which needs to be fixed for each calculation of FuzzyEn(m, n, r, N). The gradient n acts as the weight of vectors’ similarity.

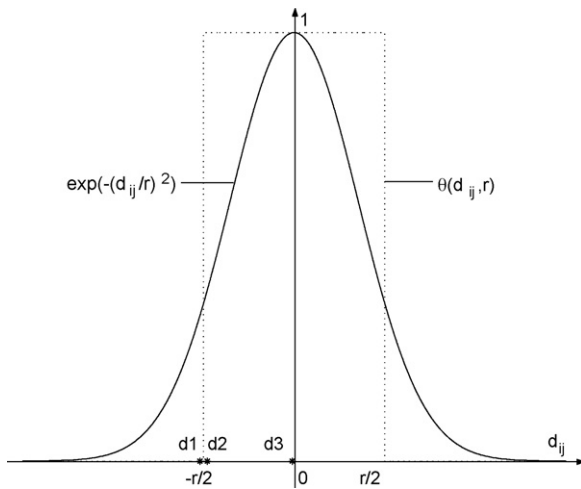


Fig. 1. Influences of parameter r and data point position on the contributions of Heaviside function $\theta(d_{ij}, r)$ and exponential function $\exp(-(d_{ij}/r)^2)$. According to Heaviside function, the two points $d2$ and $d3$ are considered the same and both similar to the original point, whereas another point $d1$, very close to the point $d2$, is considered dissimilar. The similarity of the points to the original point will be greatly changed if there is a slight change in r . However, under the rule of exponential function, the similarity of the three points to the original point depends on their closeness, and a slight change in r will not influence the results much.

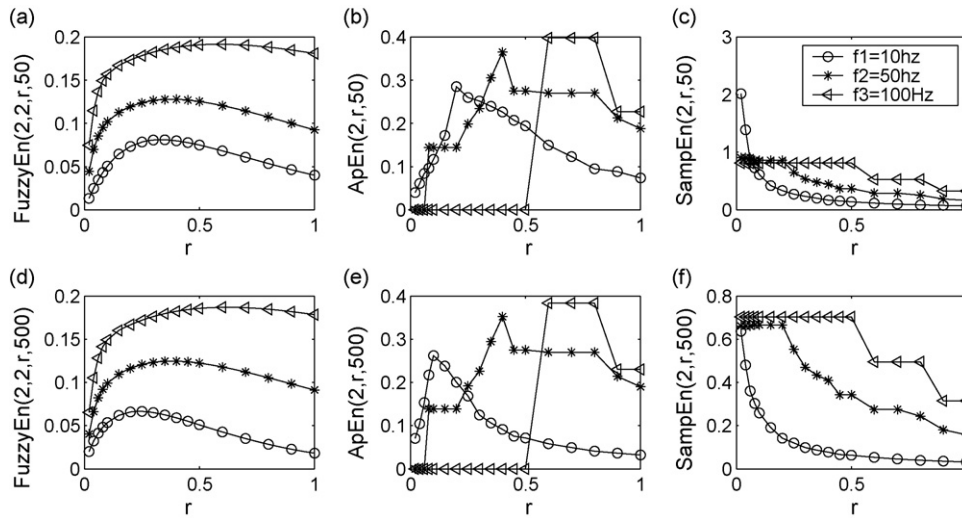


Fig. 2. The performances of FuzzyEn(2, 2, r , N), ApEn(2, r , N), and SampEn(2, r , N) statistics on periodical sinusoidal signals at different frequencies $f_1 = 10$ Hz, $f_2 = 50$ Hz, and $f_3 = 100$ Hz. ApEn of the signals are reversed when r is small. SampEn shows better results, but fails to distinguish the signals at small r values when N is also small. In addition, both ApEn and SampEn vary abruptly and discontinuously when r changes. Comparatively, FuzzyEn changes with the increasing r smoothly and continuously, and successfully characterizes the signals at small r values for short data length $N = 50$.

A value of $n > 1$ weights the similarity degree of the close vectors and unweights that of the far ones, while a value of $n < 1$ functions contrarily. The larger n is, the more the closer vectors and the less the further vectors are weighted. But a large n leads to the loss of detailed information. When n increases to infinity, the exponential function is reduced to Heaviside function, in which the detailed information is lost to the largest extent. For the sake of convenience and to catch as much detailed information as possible, small integers greater than one are recommended for the choice of n .

4. Performances of the three entropy statistics

Complexity is a concept that has multiple descriptions. The complexity of a signal can refer to the unpredictability of a signal, and it can also refer to the difficulties one has in describing or understanding a signal. For example, irregular signals are more complex than regular ones because they are more unpredictable; and regular signals varying quickly appear to be more complex than those varying slowly because quick-varying ones present more variations in a given period of time. This description of complexity means that random numbers are more complex than periodical sinusoidal signals, and that periodical signals with higher frequencies are more complex than those with lower frequencies.

In this section, we demonstrated the performances of the three entropy definitions on measuring signal complexity by applying them to experimental EMG signals as well as several typical datasets, namely, uniform independent, identically distributed (i.i.d.) random numbers, periodical sinusoidal signals, the MIX stochastic model, and the Logistic map. During each computation, both m and n are fixed to 2, and the width

of the boundary is set as r multiplied by the standard deviation (S.D.) of the original dataset. Results show that, compared with ApEn and SampEn statistics, FuzzyEn statistics go with theory more closely, and possess better properties of being continuous, less biased, capable of free parameter selections, and robust to noise. Experiments on the feature extraction from surface EMG signals also show that FuzzyEn can characterize different motions more efficiently.

4.1. Continuity and relative consistency

Fig. 2 depicts the performances of FuzzyEn, ApEn, and SampEn on three periodical sinusoidal signals $\sin 1$, $\sin 2$, and $\sin 3$ at different frequencies $f_1 = 10$ Hz, $f_2 = 50$ Hz, and $f_3 = 100$ Hz. As sinusoids with higher frequencies vary more quickly than those with lower frequencies, sinusoidal signals with higher frequencies appear to be more complex. So entropy values of $\sin 3$ should be larger than those of $\sin 2$, which are larger than those of $\sin 1$. Values of FuzzyEn can correctly characterize the complexity of the three sinusoidal signals, even at small values of r when the signals have very short data length $N = 50$. ApEn can distinguish the three sinusoidal signals when r is large enough, but when r falls below certain value, the entropy values of the signals reverse. The shorter the data length is, the larger the minimum r is needed to get correct results. So ApEn holds poor relative consistency when identifying periodical sinusoidal signals. Though SampEn shows better relative consistency than ApEn, the problem still exists. In addition, values of ApEn and SampEn change abruptly when there is a slight change in r , so both of them exhibit discontinuity. This problem does not bother FuzzyEn, whose values change with the increasing r smoothly and continuously.

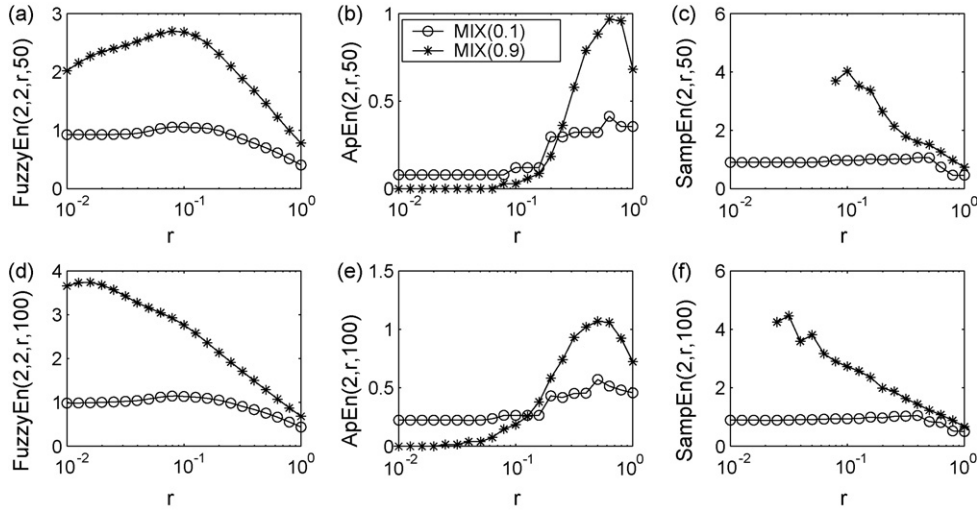


Fig. 3. The performances of FuzzyEn(2, 2, r , N), ApEn(2, r , N), and SampEn(2, r , N) statistics on measuring the complexity of MIX(0.1) and MIX(0.9). The abscissa is r , displayed logarithmically. ApEn of MIX(0.9) as a function of r crosses over that of MIX(0.1). SampEn gets larger values of MIX(0.9) than MIX(0.1) but gives no value at small values of r when N is also small. FuzzyEn not only gets larger values of MIX(0.9) than MIX(0.1), but also gives entropy definition at small r values and N values.

4.2. More freedom of parameter selection and less dependence on the data length

Experiments on MIX processes demonstrate that FuzzyEn is more flexible when it comes to parameter selection. The MIX(p) process [2] is a composite of deterministic and stochastic components. For fixed $0 < p < 1$, MIX(p) is defined as $\text{MIX}(p)_j = (1 - Z_j)X_j + Z_jY_j$, where $X_j = \sqrt{2} \sin(2\pi j/12)$ for all j ; Y_j is i.i.d. uniform random variables on $[-\sqrt{3}, \sqrt{3}]$; and Z_j is i.i.d. random variables, $Z_j = 1$ with probability p , $Z_j = 0$ with probability $1 - p$. The larger the p is, the more irregular the process becomes. In Fig. 3, ApEn statistics of MIX(0.9) as a function of r crosses over the plot of ApEn statistics of MIX(0.1), which implies the poor relative consistency of ApEn when characterizing MIX(0.1) and MIX(0.9). To get a correct judgment of the complexity of the two processes, appropriate parameter r must be chosen with caution. Though SampEn shows good relative consistency for the characterization of MIX(0.1) and MIX(0.9), no value is given at small r values when N is also small. The smaller N is, the

larger the minimum r is needed to ensure the definition of SampEn. So cautions are still needed for the selection of appropriate r . In contrast, FuzzyEn is much freer to select parameters due to its stronger relative consistency and its validity of the entropy definition in the case of small parameters of r and N .

Tests on i.i.d. uniform random numbers show that FuzzyEn statistics have the least dependence on data length N and agree with theory most closely among the three statistic families. Fig. 4 illustrates how the three statistics go with the expected value of i.i.d. uniform random numbers, which is $\log(r/\sqrt{3})$ [2], for different data length. All the abscissas in the figure are displayed logarithmically. ApEn(2, r , N) differs noticeably from the theoretical values for $N < 1000$ and $r < 0.2$; SampEn(2, r , N) satisfies the expectations better than ApEn, but it gets no value at small r values for small N values (e.g. in Fig. 4(a) when $r < 0.1$ for $N = 50$); FuzzyEn(2, 2, r , N) very closely matches the expected values for $N > 1000$ and $r > 0.01$. Even at small r values for very small N , such as $N = 50$, FuzzyEn still works well. So a valid estimation of the

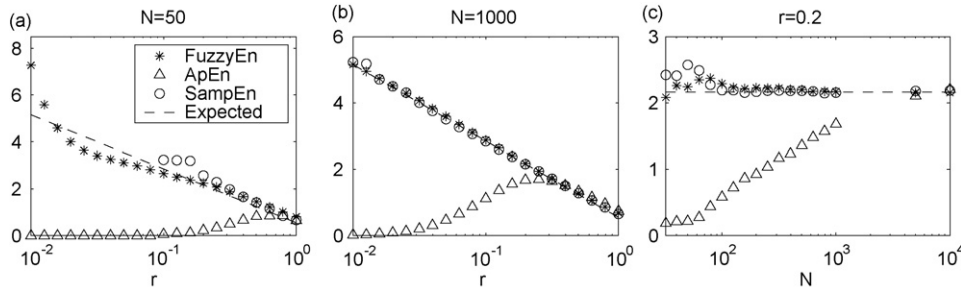


Fig. 4. The performances of FuzzyEn(2, 2, r , N), ApEn(2, r , N), and SampEn(2, r , N) statistics on i.i.d. uniform random numbers. (a) As functions of r for $N = 50$; (b) as functions of r for $N = 1000$; (c) as functions of N for $r = 0.2$. All the abscissas are displayed logarithmically. ApEn differs noticeably from the theoretical values ($\log(r/\sqrt{3})$) for $N < 1000$ and $r < 0.2$. SampEn satisfies the expectations better than ApEn but gets no value when $r < 0.1$ for $N = 50$. Among the three measures, FuzzyEn works the best.

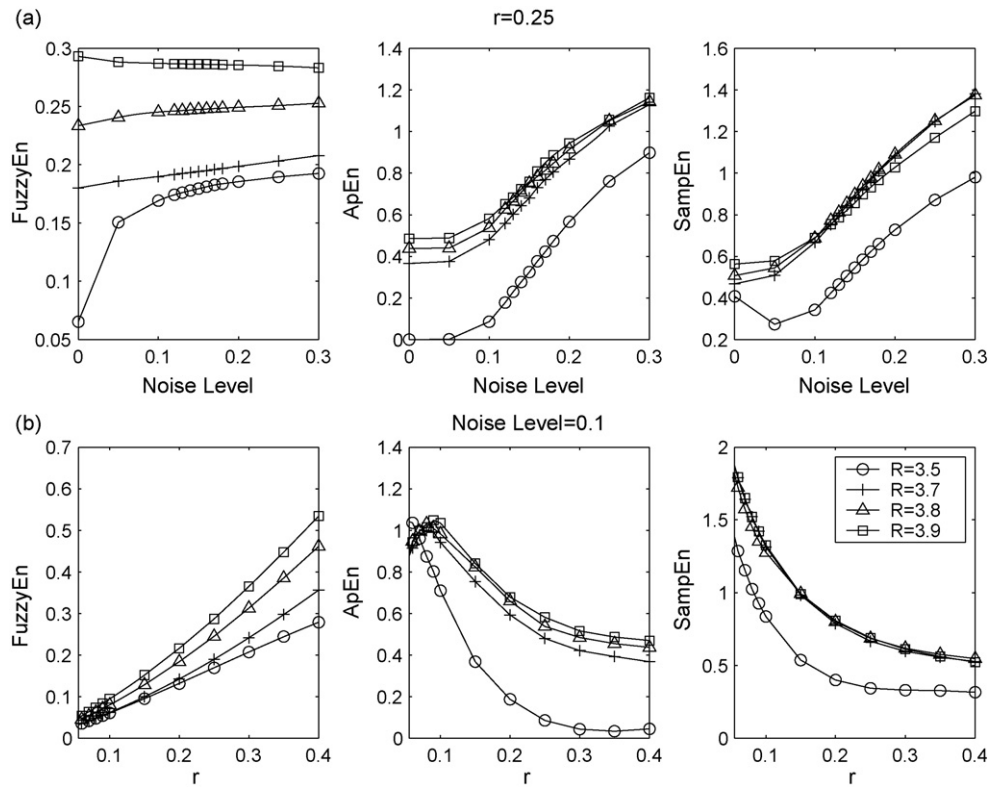


Fig. 5. The performances of the three entropy statistics on distinguishing Logistic systems at (a) $r=0.25$ and (b) noise level=0.1. The parameters used for the calculation here are $N=500$, $m=2$, and $n=2$. All the three entropies can correctly distinguish the clean datasets of different control parameter R . However, it becomes difficult for ApEn and SampEn to distinguish the system contaminated by noise with certain noise level. A larger r allows ApEn and SampEn to distinguish Logistic maps with noise of higher level, but it does not solve the problem ultimately. When noise level comes up to 0.3, both ApEn and SampEn fail in the system distinction, whereas FuzzyEn still functions, even at small r values.

theoretical value can be obtained more easily and freely with flexible parameter selections. Also, among the three entropy statistics, the differences of FuzzyEn values between i.i.d. uniform random numbers and periodical sine signals (Fig. 2) are the greatest, which indicates that FuzzyEn can most efficiently characterize the complexity of random signals and periodical signals.

4.3. Robustness to noise

Clean Logistic datasets $\{x|x_{i+1}=Rx_i(1-x_i)\}$ were obtained for $R=3.5, 3.7, 3.8$, and 3.9 . $R=3.5$ produces periodic (period four) dynamics, and $R=3.7-3.9$ produce chaotic dynamics with increasing complexity. For each system, series were generated after a transient period of 500 points. And the noisy time series were then got by adding Gaussian noise components with different noise levels (the noise level (NL) is the standard deviation of the noise divided by the standard deviation of the noise-free Logistic datasets [21]).

More complex Logistic systems produce larger entropy values, which is true to FuzzyEn, ApEn, and SampEn when the datasets are clean. However, when noises are superimposed, the system complexity measured may reverse; cases get even worse when small r values are chosen. Fig. 5 shows the performances of the three entropy statistics on distin-

guishing Logistic systems at different noise levels and r values. The data length used for the calculation is $N=500$. All the three entropies can correctly distinguish the clean datasets of different control parameter R . However, as noise is superimposed and NL increases, the distinction between the systems turns out to be different. It becomes difficult for both ApEn and SampEn to distinguish the system contaminated by noise with a level of 0.1 for $r=0.01$. A larger r improves the performances of ApEn and SampEn on the system distinction with noise of higher level. For example, for $r=0.25$, both ApEn and SampEn manage to identify Logistic maps with different R when NL=0.1. But larger tolerance cannot deal with the interference of noise with much higher level. When NL comes up to 0.3, both ApEn and SampEn fail to measure system complexity. Among the three statistics, FuzzyEn shows the best robustness to noise. The superimposed noise with a high level up to 0.3 on the systems still does not interfere with the ability of FuzzyEn to establish system distinction, even at a small r value of 0.01.

4.4. Performance on characterizing EMG signal

The performance of the three entropies is also tested on characterizing surface EMG signals as signal features. Complexity of EMG signals varies when muscles are involved

Table 1
DB indexes for the three entropy features

Entropy	DB index
FuzzyEn	0.8121
ApEn	8.4127
SampEn	0.8733

in different movements, and can thus be used as a characteristic feature of EMG signals for different motions. 80 sets of two-channel surface EMG signals were analyzed for four different motions: hand grasping (HG), hand opening (HO), forearm supination (FS), and forearm pronation (FP). The signal collection was completed in the EMG room of Shanghai Huashan Hospital in China, with informed consents provided by all the subjects. Skin surface of interested area was abraded with alcohol beforehand, and two sets of discs bipolar Ag/AgCl electrodes with diameters of 5 mm were placed over the flexor carpi radialis and the extensor carpi radialis longus on the right forearm. The sample rate was set to 1000 Hz and the bandwidth of the amplifier-filter was 10–500 Hz.

For the calculation of the three entropies, the parameter r was chosen on the basis of experimental data. We calculated the entropy values for 20 sets of surface EMG signals of forearm pronation with different r values, and then evaluated the S.D. of the entropy values. Generally, the larger the r is, the smaller the S.D. is. But too wide a boundary will lead to information loss. Therefore, appropriate selections of r can be picked at the point where the decrease in S.D. becomes slow. Here $r=0.5$ was set for the calculation of the three entropies in that the increase in r has little influence on the decrease in S.D. values of the statistics with $r>0.5$.

To measure the classification ability of the signal features, we chose the Davies–Bouldin (DB) criterion [22], which has been proved to be effective in many applications when used to evaluate the classification ability of feature space [23–25]. The DB criterion is defined as a function of the ratio of the sum of within-cluster scatter to between-cluster separation:

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} R_{ij} \quad (16)$$

with

$$R_{ij} = \frac{D_{ii} + D_{jj}}{D_{ij}} \quad (17)$$

where k is the total number of clusters, D_{ii} and D_{jj} are the dispersions of the i th and j th clusters, respectively, and D_{ij} is the distance between their mean values. As a low scatter and a high distance between clusters lead to low values of R_{ij} , a lower value of the DB index indicates a higher degree of cluster separability.

Table 1 shows the DB indexes of the three entropies characterizing the four motions. From the table we can see that the DB index of FuzzyEn is the smallest, which means features of FuzzyEn can most efficiently characterize the four motions among the three entropies.

5. Conclusions

All the three entropy definitions of FuzzyEn, ApEn, and SampEn measure the probability of similar points (vectors) within a boundary remaining similar in their evolvement from m dimensions toward $m+1$ dimensions. However, the similarity definitions between vectors are different, with the latter two based on the conventional two-state classifier and the former based on the concept of fuzzy sets.

Firstly, FuzzyEn employs fuzzy membership functions instead of Heaviside function to make a fuzzy boundary. As a simplification of exponential function when n is infinite, Heaviside function gets its simplicity at the cost of continuity and information details. And the discontinuous boundary of Heaviside function determined totally by the parameter r results in the sensitivity of both ApEn and SampEn to the parameter. In contrast, the soft and continuous boundaries of fuzzy functions not only ensure FuzzyEn to be well defined at small parameters but also enable it to change continuously. Besides, the more details taken into account by fuzzy functions make FuzzyEn a more accurate entropy definition than ApEn and SampEn.

Secondly, FuzzyEn decides whether two vectors are similar by their shapes rather than their absolute coordinates. Similarity based on vectors' shapes can capture the similarity of vectors varying greatly in numerical values but approximating in shapes. Especially in the case of signals with slow fluctuation, such a similarity definition can discover similar vectors submerged by the curve fluctuation. Thus the summation of similarity degree based on vectors' shapes is usually larger than that based on vectors' absolute coordinates. The larger summation of similarity degree, together with the exclusion of self-matches, ensures the stronger relative consistency of FuzzyEn and its less dependence on data length.

Taking advantage of the concept of fuzzy sets, FuzzyEn yields more satisfying results when characterizing signals with different complexity. Both theoretical analysis and experimental tests show that FuzzyEn provides an improved evaluation of complexity, and can thus serve as a convenient and powerful tool for short noisy experimental time series.

Acknowledgements

This work was supported by the National Basic Research Program (973) of China under Grant No. 2005CB724303. We thank Ms. Chen and Ms. Thurgood for their help in polishing this paper.

Conflict of interest

This work was supported by the National Basic Research Program (973) of China under Grant No. 2005CB724303.

References

- [1] Yu D, Small M, Harrison RG, Diks C. Efficient implementation of the Gaussian kernel algorithm in estimating invariants and noise level from noisy time series data. *Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 2000;61(4A):3750–6.
- [2] Pincus SM. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences of the United States of America* 1991;88(6):2297–301.
- [3] Bruhn J, Röpcke H, Hoeft A. Approximate entropy as an electroencephalographic measure of anesthetic drug effect during desflurane anesthesia. *Anesthesiology* 2000;92(3):715–26.
- [4] Koskinen M, Seppänen T, Tong S, Mustola S, Thakor NV. Monotonicity of approximate entropy during transition from awareness to unresponsiveness due to propofol anesthetic induction. *IEEE Transactions on Biomedical Engineering* 2006;53(4):669–75.
- [5] Burioka N, Cornélissen G, Maegaki Y, Halberg F, Kaplan DT, Miyata M, et al. Approximate entropy of the electroencephalogram in healthy awake subjects and absence epilepsy patients. *Clinical EEG and Neuroscience* 2005;36(3):188–93.
- [6] Pincus SM, Mulligan T, Iranmanesh A, Gheorghiu S, Godschalk M, Veldhuis JD. Older males secrete luteinizing hormone and testosterone more irregularly, and jointly more asynchronously, than younger males. *Proceedings of the National Academy of Sciences of the United States of America* 1996;93(24):14100–5.
- [7] Christen JA, Torres JL, Barrera J. A statistical feature of genetic sequences. *Biometrical Journal* 1998;40(7):855–63.
- [8] Pincus SM, Cummins TR, Haddad GG. Heart rate control in normal and aborted-SIDS infants. *American Journal of Physiology: Regulatory Integrative and Comparative Physiology* 1993;264(3 33-3):R638–46.
- [9] Pincus SM, Keefe DL. Quantification of hormone pulsatility via an approximate entropy algorithm. *American Journal of Physiology: Endocrinology and Metabolism* 1992;262(5 25-5):E741–54.
- [10] Caldirola D, Bellodi L, Caumo A, Migliaresi G, Perna G. Approximate entropy of respiratory patterns in panic disorder. *American Journal of Psychiatry* 2004;161(1):79–87.
- [11] Storella RJ, Wood HW, Mills KM, Kanters JK, Høgaard MV, Holstein-Rathlou NH. Approximate entropy and point correlation dimension of heart rate variability in healthy subjects. *Integrative Physiological and Behavioral Science* 1998;33(4):315–20.
- [12] Shin DG, Yoo CS, Yi SH, Bae JH, Kim YJ, Park JS, et al. Prediction of paroxysmal atrial fibrillation using nonlinear analysis of the R–R interval dynamics before the spontaneous onset of atrial fibrillation. *Circulation Journal* 2006;70(1):94–9.
- [13] Ning X, Xu Y, Wang J, Ma X. Approximate entropy analysis of short-term HFECG based on wave mode. *Physica A: Statistical Mechanics and its Applications* 2005;346(3–4):475–83.
- [14] Radhakrishnan N, Wilson JD, Lowery C, Murphy P, Eswaran H. Testing for nonlinearity of the contraction segments in uterine electromyography. *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering* 2000;10(12):2785–90.
- [15] Pincus SM. Approximate entropy (ApEn) as a complexity measure. *Chaos* 1995;5(1):110–7.
- [16] Grouven U, Beger FA, Schultz B, Schultz A. Correlation of Narcotrend Index, entropy measures, and spectral parameters with calculated propofol effect-site concentrations during induction of propofol-remifentanyl anaesthesia. *Journal of Clinical Monitoring and Computing* 2005;18(4):231–40.
- [17] Pincus SM, Goldberger AL. Physiological time-series analysis: what does regularity quantify? *American Journal of Physiology: Heart and Circulatory Physiology* 1994;266(4 35-4):H1643–56.
- [18] Richman JS, Moorman JR. Physiological time-series analysis using approximate and sample entropy. *American Journal of Physiology: Heart and Circulatory Physiology* 2000;278(6 47-6):H2039–49.
- [19] Chen W, Wang Z, Xie H, Yu W. Characterization of surface EMG signal based on fuzzy entropy. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 2007;15(2):266–72.
- [20] Zadeh LA. Fuzzy sets. *Information and Control* 1965;8:338–53.
- [21] Leontitis A, Bountis T, Paggi J. An adaptive way for improving noise reduction using local geometric projection. *Chaos* 2004;14(1):106–10.
- [22] Davies DL, Bouldin DW. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1979;1(4):224–7.
- [23] Zardoshti-Kermani M, Wheeler BC, Badie K, Hashemi RM. EMG feature evaluation for movement control of upper extremity prostheses. *IEEE Transactions on Rehabilitation Engineering* 1995;3(4):324–33.
- [24] Boostani R, Moradi MH. Evaluation of the forearm EMG signal features for the control of a prosthetic hand. *Physiological Measurement* 2003;24(2):309–19.
- [25] Micera S, Sabatini AM, Dario P, Rossi B. A hybrid approach to EMG pattern analysis for classification of arm movements using statistical and fuzzy techniques. *Medical Engineering and Physics* 1999;21(5):303–11.