Permutation entropy — a natural complexity measure for time series

Christoph Bandt and Bernd Pompe
Institute of Mathematics and Institute of Physics, University of Greifswald, Germany
(December 17, 2001)

We introduce complexity parameters for time series based on comparison of neighboring values. The definition directly applies to arbitrary real-world data. For some well-known chaotic dynamical systems it is shown that our complexity behaves similar as Lyapunov exponents, and is particularly useful in the presence of dynamical or observational noise. The advantages of our method are its simplicity, extremely fast calculation, its robustness and invariance with respect to non-linear monotonous transformations.

05.45.Tp, 05.45.-a, 02.50.-r

I. COMPLEXITY MEASURES FOR TIME SERIES

Various measures of complexity were developed to compare time series and distinguish regular, (e.g. periodic), chaotic and random behavior. Among others, it has been reported that complexity of heart and brain data can distinguish healthy and sick subjects and sometimes even predict heart attack or epileptic seizure [1]. The main types of complexity parameters are entropies, fractal dimensions, Lyapunov exponents. They are all defined for typical orbits of presumably ergodic dynamical systems, and there are profound relations between these quantities which allow to compute one from the other [2,3].

Problem. The basic conceptual problem is that these definitions are not made for an arbitrary series of observations \( \{x_1, x_2, \ldots\} \). As a consequence, there is also a computational problem. Many ingenious algorithms, tricks and recipes have been developed during the last 20 years in order to estimate complexity measures from real-world time series [4–7]. They work wonderfully when the time series is simulated from a low-dimensional dynamical system, but most of them break down as soon as noise is added to the series. For real-world series, “noise elimination” requires careful preprocessing of the data and fine-tuning of parameters, and the results cannot be reproduced without specifying details of the method.

One idea to overcome these problems is application of Kolmogorov–Chaitin algorithmic complexity to orbits of dynamical systems [8,9,7].

Our approach. We go another way, defining simple complexity measures which are easily calculated for any type of time series, be it regular, chaotic, noisy or from reality. A practical and a theoretical example were chosen to compare our complexity with established concepts.

In Sec. III, we recognize voiced sounds in a speech signal.

In Sec. IV, we show for the well-known family of logistic maps that our permutation entropy is very similar to Lyapunov exponents over the range of several thousand parameter values. Most importantly, it yields meaningful results in the presence of observational and dynamical noise (Sec. V).

Source entropy. Given a stationary time series \( \{x_t\} \), most complexity parameters measure in one or the other way the multitude of different vectors \((x_{t+1}, \ldots, x_{t+n})\) for various \( n \), passing then with \( n \) to \( \infty \). If the \( x_t \) attain a finite number \( M \) of values, the classical source entropy \( h \) of Shannon measures the mean conditional uncertainty of the future \( x_{t+1} \) given the whole past \( \ldots x_{t-1}, x_t \). We have \( 0 \leq h \leq \log M \), with \( h = 0 \) if the series is perfectly predictable from the past and \( h = \log M \) if all values are independent and uniformly distributed. Large \( h \) indicates high complexity. Shannon’s entropy can be generalized by Rényi’s \( \alpha \)-entropies [10,4,2].

Partitions. If the \( x_t \) attain infinitely many values, it is common to replace them by a symbol sequence \( \{s_t\} \) with finitely many symbols, and calculate source entropy for the \( s_t \). We use a partition \( X = P_1 \cup \ldots \cup P_m \) of the set of values and define \( s_t = i \) if \( x_t \in P_i \). Then we can increase the number \( m \) of pieces and get a limit for fine resolution which is the Kolmogorov–Sinai entropy \( h_K \), an isomorphism invariant of the dynamical system [2,11,3]. For special generating partitions, no such limit is needed. These partitions are difficult to find, however, even for two-dimensional examples like the Henon system [12]. For unimodal maps, the critical point defines a generating partition [13] but any misplacement of the partition point gives an entropy smaller than \( h_K \) [14]. Our viewpoint (cf. [15]) is that the symbol sequence must come naturally from the \( x_t \), without further model assumptions. Thus we suggest to take partitions given by comparison of neighboring values \( x_t \). Research in progress [16] shows that this is similar to using generating partitions.

II. BASIC DEFINITIONS

The following definitions apply to an arbitrary time series, with a weak stationarity assumption explained below. For simplicity, we shall assume that the values \( x_t \) have a continuous distribution so that equal values \( x_t, x_t^* = x_t, t^* \neq t \), are very rare. Thus we can neglect equal values and consider only inequalities between the \( x_t \). Our entropies are calculated for different embedding dimensions \( n \), but we do not attempt to determine a limit for large \( n \) although this is an interesting theoretical problem.
For each ered here as possible order types of consider all uments with chaotic and real–world time series, increase at most linearly with $n$. We define the permutation entropy of order $n = 2$ as a measure of the probabilities of appearance of the permutations 01 and 10. So,

$$H(2) = -\frac{4}{6} \log \frac{4}{6} - \frac{2}{6} \log \frac{2}{6} \approx 0.918.$$ 

As usual, log is with base 2, thus $H$ is given in bit. Next, we compare three consecutive values. $(4,7,9)$ and $(7,9,10)$ represent the permutation 01 since they are in crescent order. $(9,10,6)$ and $(6,11,3)$ correspond to the permutation 120 while $(10,6,11)$ has the permutation type 102, with smallest value in the middle. The permutation entropy of order $n = 3$ is

$$H(3) = -2 \cdot \frac{2}{5} \log \frac{2}{5} - \frac{1}{5} \log \frac{1}{5} \approx 1.522.$$ 

Definition. Consider a time series $\{x_t\}_{t=1..T}$. We consider all $n!$ permutations of order $n$ which are considered here as possible order types of $n$ different numbers. For each $\pi$ in $S_n$ we determine the relative frequency

$$p(\pi) = \frac{\# \{ t \mid 0 \leq t \leq T-n, (x_{t+1},...,x_{t+n}) has type \pi \}}{T-n+1}.$$ 

This estimates the frequency of $\pi$ as good as possible for a finite series of values. To determine $p(\pi)$ exactly, we have to assume an infinite time series $\{x_1,x_2,...\}$ and take the limit for $T \to \infty$ in the above formula. This limit exists with probability 1 when the underlying stochastic process fulfills a very weak stationarity condition: for $k \leq n$, the probability for $x_t < x_{t+k}$ should not depend on $t$.

The permutation entropy of order $n \geq 2$ is defined as

$$H(n) = -\sum p(\pi) \log p(\pi)$$

where the sum runs over all $n!$ permutations $\pi$ of order $n$. This is the information contained in comparing $n$ consecutive values of the time series. It is clear that $0 \leq H(n) \leq \log n!$ where the lower bound is attained for an increasing or decreasing sequence of values, and the upper bound for a completely random system (i.i.d. sequence) where all $n!$ possible permutations appear with the same probability. The time series presents some sort of dynamics when $H(n) < \log n!$. Actually, in our experiments with chaotic and real–world time series, $H(n)$ did increase at most linearly with $n$ [16]. So it seems useful to define the permutation entropy per symbol of order $n$, dividing by $n-1$ since comparisons start with the second value:

$$h_n = H(n)/(n-1).$$

Related entropies. We can also determine the information contained in sorting the $n$th value among the previous $n-1$ when their order is already known:

$$d_n = H(n) - H(n-1), \quad d_2 = H(2).$$

This could be called sorting entropy of order $n$. Clearly, $h_n = (d_2 + ... + d_n)/(n-1)$ which means that $h_n$ is a stable parameter characterizing the overall behaviour of the time series while the $d_n$ describe more special properties connected for instance with points of period $n$. In some cases, $d_n$ can increase with $n$.

Instead of $H(n)$, we also studied $H^0(n) = \log \# \{ \pi \in S_n \mid p(\pi) > 0 \}$ and defined a corresponding $h_n^0 = H^0(n)/(n-1)$. This definition is simpler, but $h_n^0$ is not robust under noise, is influenced by outliers, and needs more data for accurate estimates.

III. RESULTS FOR A SPEECH SIGNAL

To demonstrate how permutation entropy applies to real data, we consider in Fig. 1a the utterance “permutation entropy measures complexity” spoken by a male speaker in a modal register. The signal was sampled at 11 kHz, using a low cost PC sound card with standard data preprocessing (anti–aliasing low pass), without special filtering. The signal was analyzed using a sliding window of $T_{win} = 512$ samples, that is 46ms, and a one sample window shift. The 43000 windows of the 4s signal were processed on a PC in less than a second, which means that permutation entropy can be used in real-time applications. Due to the shortness of the window we chose only low order $n = 3, 4, 5$ for Fig. 1b).

The signal starts and ends with noise, and the corresponding normalized entropies $H(n)/\log n!$ are close to 1, as well as for unvoiced sounds, while all voiced sounds can be recognized by the decrease of entropy.

A well–known complexity parameter for short–time speech analysis is the zero–crossing rate [17]

$$ZCR = \# \{ t \mid x(t)x(t+1) < 0 \}/(T_{win} - 1).$$

Assuming zero mean of the signal, ZCR is expected to be close to 0.5 for noisy data like pure fricatives (unvoiced sounds) or noisy breaks, and close to zero for voiced speech (Fig. 1c). Therefore it is recommended for detection of voiced segments.

In Fig. 1a), at first glance, ZCR and permutation entropy (PE) behave similarly. However, at some places PE seems to be a better indicator. During silence (e.g., 0.0..0.3s) ZCR misleading indicates low complexity due to a tiny 100Hz component (first harmonic of 50Hz mains hum, not visible in the figure). PE is more robust.
against such perturbations. Moreover, PE sometimes better indicates transitions, e.g., at 1.6s, from voiced segment to a break, or at 2.9 s where the two first syllables of “complexity” are not properly separated by ZCR. The ‘u’ in “permutation” is also better detected by PE.

Various experiments have shown that PE, in contrast to ZCR, is robust with respect to the choice of window length (even 128 samples were sufficient), to the sampling frequency (3.11 kHz gives very similar results) and to observational noise. The order \( n = 3 \ldots 7 \) of the PE had also little influence on the results. The computational effort is only about \( 2n \) times the effort for the ZCR.

IV. RESULTS FOR CHAOTIC TIME SERIES

To get a feeling for our entropy, we processed not only single time series, but orbits of several parametric families of dynamical systems. For the present paper we took 5001 parameters values of the logistic map \( x_{t+1} = rx_t(1 - x_t) \) with \( 3.5 \leq r \leq 4 \) since the reader is well familiar with all peculiarities of this family [13]. Some two-dimensional dynamical systems, like Henon and the dissipative standard maps, were also investigated. It always turned out that PE is very similar to the positive Lyapunov exponent. In order to compare with periodic windows, band–merging points etc., Feigenbaum diagram and Lyapunov exponents are given in Figs. 1 a,b). In a more theoretical paper [16], it was shown that for piecewise monotone maps \( h_n \) converges to the Kolmogorov–Sinai entropy \( h_k \) with \( n \rightarrow \infty \), but this convergence is rather slow.

Proper range for \( n \) and \( T \). We used an extremely fast algorithm where each pair of values need to be compared only once, so time was not a problem even on a PC. For larger \( n \), the number \( n! \) of permutations which can appear in the time series causes memory restrictions. More seriously, \( T \) should be considerably larger than this number in order to estimate \( H(n) \) accurately. We took \( T = 10^6 \) which gave accurate results for \( n \leq 15 \) in case of all Henon and most logistic maps, and for \( n \leq 12 \) in case of logistic maps with \( r \) very close to 4.

For real–world time series it hardly makes sense to go beyond order \( n = 6 \) or 7. Thus it is good to see the similarity of the functions \( h_6 \) and \( h_{12} \) in Fig. 2c,d), and it is also good to know that \( h_6 \) can be reliably estimated already from \( T = 1000 \) values (Fig. 3a).

Permutation entropy and Lyapunov exponent in Fig. 2 c,d), and b) clearly have a very similar appearance over the whole chaotic regime. Differences appearing within periodic windows will be explained now. Suppose \( x_t \) approaches a stable orbit of period \( m \), and \( C \) is the product of derivatives of the generating map along this orbit. Then the Lyapunov exponent \( \frac{1}{m} \log |C| \) is negative. Moreover, inside a window of a stable period, \( C \) runs from +1 through 0 to -1 ( [13], cf. Fig. 2a,b). Now for \( C > 0 \) there are at most \( m \) permutations in the time series which leads to \( h_n = (\log m)/(n - 1) \) for \( n \geq m \).

For \( C < 0 \) we have \( 2m \) permutations, and \( h_n \) increases by \( 1/(n - 1) \). The difference is 1/5 in Fig. 2e) and 1/11 in Fig. 2d). Unfortunately, numerical rounding produces \( 2m \) permutations already for some \( r \) with \( C > 0 \). For large \( n \), the entropy \( h_n \) tends to zero in periodic windows. Generally, it seems that \( \lambda \leq h_{12} \leq h_6 \).

The tent map. For \( r = 4 \), conjugate to the tent map [13], we studied averages of 1000 time series of size \( T = 10^6 \) with \( k = 2.6 \). Results for \( T = 10^6 \) were confirmed by an exact study of the invariant measure, giving differences less than \( 10^{-4} \) for \( n \leq 10 \) [16]. In Fig. 3a), we see that \( h_n \) increases up to \( n = 7, 8 \) and then decreases. The limit for \( n \rightarrow \infty \) exists, which is 1, the Kolmogorov–Sinai entropy [16]. The variance is small for estimates of \( h_n \) (Fig. 3b). However, the estimates of \( h_n \) are biased as \( T \) is small compared to \( n! \) which should be compensated by a correction term as in [18].

V. THE EFFECTS OF NOISE

Permutation entropies have a practically important invariance property. If \( y_t = f(x_t) \), where \( f \) is an arbitrary strictly increasing (or decreasing) real function, then \( h_n \) is the same for \( x_t \) and \( y_t \). Such non–linear functions \( f \) occur, for example, when measuring physiological data with different equipment. However, the invariance of \( h_n \) also implies its discontinuity near the constant time series \( x_t = c \) where \( h_n = 0 \). If this series is disturbed by an i.i.d. noise, no matter how small, then \( h_n = (\log n!)/(n - 1) \), the largest possible value.

Observational noise. For time series of period \( m \) and \( n = km \) the disturbed time series admits \( m \) times \( k! \) permutations so that \( h_n = (\log m + m \cdot \log k!)/(n - 1) \) as long as the noise preserves the order of the periodic orbit. In Fig. 2e), where Gaussian noise was added to the deterministic time series, we have \( m = 4 \) with \( h_{12} \approx 1.12 \) on the left and \( m = 3 \) with value 1.39 in the middle. Outside the windows with small periods, however, the observational noise causes only a small increase of entropy. For the higher noise level, some peaks appear when \( r \) is just below a band–merging point, for example 3.5748, 3.5925, 3.6785 (compare Figs. 2a,e). It is clear that the noise anticipates the band–merging point but it seems surprising that the peak is larger than the entropy in the chaotic regime just behind band–merging.

Dynamical noise, added to \( x_t \) during each step of the iteration, gives still better results. The entropy function \( h_n(r) \) becomes smooth, approximating the entropy of the undisturbed time series for small noise level (Fig. 2f). Actually, Fig. 3a) does not change at all for \( s \leq 0.004 \).

Near the period 3 window in Fig. 2f) there are examples of noise–induced order where larger noise gives smaller \( h_n \) but in general \( h_n \) increases with noise level. The effects of small periods disappear for larger noise
VI. CONCLUSION

Permutation entropies $h_n$ are appropriate complexity measures for chaotic time series, in particular in the presence of dynamical and observational noise. In contrast with all known complexity parameters, a small noise does not essentially change the complexity of a chaotic signal. Permutation entropies can be calculated for arbitrary real-world time series. Since the method is extremely fast and robust, it seems preferable when there are huge data sets and no time for preprocessing and fine-tuning of parameters.

FIG. 2. Logistic equations for varying control parameter $r$ (step $\Delta r = 10^{-4}$). a) Bifurcation diagram, b) Lyapunov exponent $\lambda$, c, d) permutation entropies $h_6$, $h_{12}$, e) $h_{12}$ with Gaussian observational noise at standard deviations $s = 0.00025$ (lower line), and $s = 0.004$ (upper line), f) $h_{12}$ with additive Gaussian dynamical noise at standard deviations $s = 0.00025$, $s = 0.001$, $s = 0.004$ (from thin to thick lines) ($T = 10^6$ data for each $r$-value)

FIG. 3. Logistic map ($r = 4$). a) Mean $\langle h_n \rangle$ of our estimator of order $n$ permutation entropies $h_n$ for varying length $T$ of time series, b) corresponding standard deviation $\sigma$ of $h_n$
学霸图书馆
www.xuebalib.com

本文献由“学霸图书馆-文献云下载”收集自网络，仅供学习交流使用。

学霸图书馆（www.xuebalib.com）是一个“整合众多图书馆数据库资源，提供一站式文献检索和下载服务”的24小时在线不限IP图书馆。

图书馆致力于便利、促进学习与科研，提供最强文献下载服务。

图书馆导航：
图书馆首页 文献云下载 图书馆入口 外文数据库大全 疑难文献辅助工具