

Article

Permutation Entropy: New Ideas and Challenges

Karsten Keller^{1*}, Teresa Mangold¹, Inga Stolz¹, and Jenna Werner¹

Institute of Mathematics, University of Lübeck, Lübeck D-23562, Germany

* Correspondence: keller@math.uni-luebeck.de; Tel.: +49-451-3101-6030

Abstract: During the last years some new variants of Permutation entropy have been introduced and applied to EEG analysis, among them a conditional variant and variants using some additional metric information or being based on entropies different from the Shannon entropy. In some situations it is not completely clear what kind of information the new measures and their algorithmic implementations provide. We discuss the new developments and illustrate them for EEG data.

Keywords: ordinal patterns; Permutation entropy; Approximate entropy; Sample entropy; Conditional entropy of ordinal patterns; Kolmogorov-Sinai entropy; classification

1. Introduction

The concept of Permutation entropy introduced by Bandt and Pompe [1] in 2002 has been applied to data analysis in various disciplines (compare e.g. the collection [2], and [3], [4]). The Permutation entropy of a time series is the Shannon entropy of the distribution of ordinal patterns in the time series (see also [5]). Such ordinal patterns, describing order types of vectors, are coded by permutations. Denoting the set of permutations of $\{0, 1, \dots, d\}$ for d in the natural numbers \mathbb{N} by Π_d , we say that a vector $(v_0, v_1, \dots, v_d) \in \mathbb{R}^{d+1}$ has *ordinal pattern* $\pi = (r_0, r_1, \dots, r_d) \in \Pi_d$ if

$$v_{r_0} \geq v_{r_1} \dots \geq v_{r_{d-1}} \geq v_{r_d}$$

and

$$r_{l-1} > r_l \text{ if } v_{r_{l-1}} = v_{r_l}.$$

Definition 1. The empirical Permutation entropy (ePE) of order $d \in \mathbb{N}$ and of delay $\tau \in \mathbb{N}$ of a time series $(x_t)_{t=0}^{N-1}$ with $N \in \mathbb{N}$ is given by

$$\text{ePE} \left(d, \tau, (x_t)_{t=0}^{N-1} \right) = -\frac{1}{d} \sum_{\pi \in \Pi_d} p_{\pi}^{\tau} \ln p_{\pi}^{\tau}, \quad (1)$$

where

$$p_{\pi}^{\tau} = \frac{\#\{t \in \{d\tau, d\tau + 1, \dots, N - 1\} \mid (x_t, x_{t-\tau}, \dots, x_{t-d\tau}) \text{ has ordinal pattern } \pi\}}{N - d\tau} \quad (2)$$

is the relative frequency of ordinal patterns π in the time series and $0 \ln 0$ is defined by 0. The vectors $(x_t, x_{t-\tau}, \dots, x_{t-d\tau})$ related to t, d and τ are called delay vectors.

In contrast to some other presentations of Permutation entropy in literature the defining formula contains the term $\frac{1}{d}$. This is justified by an interesting statement shown by Bandt et al. [6], which is roughly saying that for time series from a special class of systems, for $d \rightarrow \infty$, (1) approaches to a nonnegative real number, it is however open how large the corresponding class is. Some further discussion of this point is following in this paper.

In order to be more flexible in data analysis, various modifications of Permutation entropy have been developed during the last years. One class of such modifications is based on adding some metric information of the corresponding delay vectors to the ordinal patterns considered. Entropy

measures of that type are the *fine-grained Permutation entropy* proposed by Liu and Wang [7], the *weighted Permutation entropy* and the *robust Permutation entropy* introduced by Fadlallah et al. [8] and Keller et al. [9], respectively, Bian et al. [10] have adapted Permutation entropy to the situation of delay vectors with equal components which are relevant when the number of possible values in a time series is small. Other variants are to consider Tsallis or Renyi entropies instead of the Shannon entropy which goes back to the work of Zunino et al. [11] and Liang et al. [12], or to integrate information of different scales as done by Li et al. [13], Ouyang et al. [14] and Azami and Ascordero [15]. Unakafov and Keller [16] have proposed the Conditional entropy of (successive) ordinal patterns as a measure which has been shown to perform better than the Permutation entropy itself in many cases.

In this paper we discuss some of the new ('non-weighting') developments on Permutation entropy. We first give some theoretical background in order to justify and to motivate conditional variants of Permutation entropy. Then we have a closer look at Renyi and Tsallis modifications of Permutation entropy. The last part of the paper is aimed to the classification of data from the viewpoint of epileptic activity. Here we point out that ordinal time series analysis combined with other methods and automatic learning could be promising for data analysis. We rely on EEG data reported in [17] and [18] and use algorithms developed in Unakafova and Keller [19].

2. Some theoretical background

The Kolmogorov-Sinai entropy. In order to get a better understanding of Permutation entropy, we take a modelling approach. For this, we choose a discrete dynamical system defined on a probability space (Ω, \mathcal{B}, P) . The elements $\omega \in \Omega$ are considered as the (hidden) states of the system, the elements of the σ -algebra \mathcal{B} as the events one is interested in, and the probability P measures the chances of such events taking place. The dynamics on the system is given by a map $T : \Omega \rightarrow \Omega$ being $\mathcal{B} - \mathcal{B}$ -measurable, i.e. satisfying $T^{-1}(B) \in \mathcal{B}$ for all $B \in \mathcal{B}$. The map T is assumed to be P -invariant saying that $P(T^{-1}(A)) = P(A)$ for all $A \in \mathcal{B}$ in order to guarantee that the distribution of the states of the system does not change under the dynamics.

For simplicity, during the whole paper we assume that (Ω, \mathcal{B}, P) and T are fixed; specifications are given directly where they are needed.

In dynamical systems the Kolmogorov-Sinai entropy based on entropy rates of finite partitions is a well established theoretical complexity measure. For explaining this concept, consider a finite partition $\mathcal{C} = \{C_1, C_2, \dots, C_l\} \subset \mathcal{B}$ and note that the (Shannon) entropy of such a partition is defined by

$$H(\mathcal{C}) = - \sum_{i=1}^l P(C_i) \ln P(C_i).$$

For $A = \{1, 2, \dots, l\}$ the set A^k of words $a_1 a_2 \dots a_k$ of length k over A provides a partition \mathcal{C}_k of Ω into pieces

$$C_{a_1 a_2 \dots a_k} := \{\omega \in \Omega \mid (\omega, T(\omega), \dots, T^{k-1}(\omega)) \in C_{a_1} \times C_{a_2} \times \dots \times C_{a_k}\}.$$

(In this paper $T^{\circ l}$ denotes the l -fold concatenation of T , where $T^{\circ 0}$ is the identity map.) The distribution of word probabilities $P(C_{a_1 a_2 \dots a_k})$ contains some information on the complexity of the system, which, as the word length k approaches to ∞ , provides the *entropy rate*

$$\text{EntroRate}(\mathcal{C}) = \lim_{k \rightarrow \infty} \frac{1}{k} H(\mathcal{C}_k) = \lim_{k \rightarrow \infty} (H(\mathcal{C}_k) - H(\mathcal{C}_{k-1})) \quad (3)$$

of \mathcal{C} . The latter can be interpreted as the mean information per symbol. Here let $H(\mathcal{C}_0) := 0$. Note that both the sequences $(\frac{1}{k} H(\mathcal{C}_k))_{k=1}^{\infty}$ and $(H(\mathcal{C}_k) - H(\mathcal{C}_{k-1}))_{k=1}^{\infty}$ are monotonically non-increasing.

In order to have a complexity not depending only on a fixed partition and measuring the overall complexity, the *Kolmogorov-Sinai entropy* (*KS entropy*) is defined by

$$KS = \sup_{\mathcal{C} \text{ finite partition}} \text{EntroRate}(\mathcal{C}).$$

For more information, in particular concerning formula (3), see for example Walters [20].

By the nature of the concept of KS entropy, it is not easy to determine and to estimate from time dependent measurements of a system. One important point is to find finite partitions supporting as much information on the dynamics as possible. If there is no feasible so-called *generating partition* containing all information (for a definition, see Walters [20]), this is not easy. The approach of Bandt and Pompe builds up appropriate partitions based on the ordinal structure of a system. Let us explain the idea in a general context.

Observables and ordinal partitioning. In the modeling, an observed time series x_0, x_1, \dots, x_{N-1} is considered as the sequence of ‘outcoming’ values $X(\omega), X(T(\omega)), \dots, X(T^{N-1}(\omega))$ for some $\omega \in \Omega$, where X is a real-valued random variable on (Ω, \mathcal{B}, P) . Here X is interpreted as an observable establishing the outreading process.

Since it is no additional effort to consider more than one observable, let $\mathbf{X} = (X_1, X_2, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$ be a random vector with observables X_1, X_2, \dots, X_n as components. Originally, Bandt and Pompe [1] discussed the case that the measuring values are coinciding with the states of a one-dimensional system, which in our language means that $\Omega \subset \mathbb{R}$, $n = 1$, and there is only one observable which coincides with the identity map.

For each order $d \in \mathbb{N}$ we are interested in the partition

$$\mathcal{C}^{\mathbf{X}}(d) = \{C_{(\pi_1, \pi_2, \dots, \pi_n)} \mid \pi_i \text{ ordinal pattern for } i = 1, 2, \dots, n\}$$

with

$$C_{(\pi_1, \pi_2, \dots, \pi_n)} = \{\omega \in \Omega \mid (X_i(T^{d\pi_1}(\omega)), \dots, X_i(T^{d\pi_n}(\omega)), X_i(\omega)) \\ \text{has ordinal pattern } \pi_i \text{ for } i = 1, 2, \dots, n\}$$

called *ordinal partition of order d*. Here ordinal patterns are of order d and the delay is one.

No information loss. For the rest of this section assume that the measurements by the observables do not loose information on the modelling system roughly meaning that the random variables $X_1, X_1 \circ T, X_1 \circ T^2, \dots, X_2, X_2 \circ T, X_2 \circ T^2, \dots, X_n, X_n \circ T, X_n \circ T^2, \dots$ separate states of Ω and, more precisely, that for each event C in the σ -algebra generated by these random variables there exists some $B \in \mathcal{B}$ with $P(C \Delta B) = 0$. Note that already for one observable the separation property described is mostly satisfied in a certain sense (see Takens [21], Gutman [22]).

Moreover, assume that T is *ergodic* meaning that $P(T^{-1}(B)) = P(B)$ implies $P(B) \in \{0, 1\}$ for all $B \in \mathcal{B}$. Ergodicity says that the given system does not separate into proper subsystems and, by Birkhoff’s ergodic theorem (see [20]), allows to estimate properties of the system on the base of orbits $\omega, T(\omega), T^2(\omega), \dots$

Under these assumptions it holds

$$KS = \lim_{d \rightarrow \infty} \text{EntroRate}(\mathcal{C}^{\mathbf{X}}(d)) = \sup_{d \in \mathbb{N}} \text{EntroRate}(\mathcal{C}^{\mathbf{X}}(d)). \quad (4)$$

This statement shown in Antoniouk et al. [23] and generalized in [24] can be interpreted as follows: If there is no loss of information caused by the measuring process, the whole information of the process can be obtained by only considering ordinal relations between the measurements for each observable.

$$\begin{array}{ccccccccc}
H(\mathcal{C}^{\mathbf{X}}(1)_1) & H(\mathcal{C}^{\mathbf{X}}(1)_2)/2 & H(\mathcal{C}^{\mathbf{X}}(1)_3)/3 & H(\mathcal{C}^{\mathbf{X}}(1)_4)/4 & H(\mathcal{C}^{\mathbf{X}}(1)_5)/5 & \dots & \searrow & \text{EntroRate}(\mathcal{C}^{\mathbf{X}}(1)) \\
H(\mathcal{C}^{\mathbf{X}}(2)_1) & H(\mathcal{C}^{\mathbf{X}}(2)_2)/2 & H(\mathcal{C}^{\mathbf{X}}(2)_3)/3 & H(\mathcal{C}^{\mathbf{X}}(2)_4)/4 & H(\mathcal{C}^{\mathbf{X}}(2)_5)/5 & \dots & \searrow & \text{EntroRate}(\mathcal{C}^{\mathbf{X}}(2)) \\
H(\mathcal{C}^{\mathbf{X}}(3)_1) & H(\mathcal{C}^{\mathbf{X}}(3)_2)/2 & H(\mathcal{C}^{\mathbf{X}}(3)_3)/3 & H(\mathcal{C}^{\mathbf{X}}(3)_4)/4 & H(\mathcal{C}^{\mathbf{X}}(3)_5)/5 & \dots & \searrow & \text{EntroRate}(\mathcal{C}^{\mathbf{X}}(3)) \\
H(\mathcal{C}^{\mathbf{X}}(4)_1) & H(\mathcal{C}^{\mathbf{X}}(4)_2)/2 & H(\mathcal{C}^{\mathbf{X}}(4)_3)/3 & H(\mathcal{C}^{\mathbf{X}}(4)_4)/4 & H(\mathcal{C}^{\mathbf{X}}(4)_5)/5 & \dots & \searrow & \text{EntroRate}(\mathcal{C}^{\mathbf{X}}(4)) \\
H(\mathcal{C}^{\mathbf{X}}(5)_1) & H(\mathcal{C}^{\mathbf{X}}(5)_2)/2 & H(\mathcal{C}^{\mathbf{X}}(5)_3)/3 & H(\mathcal{C}^{\mathbf{X}}(5)_4)/4 & H(\mathcal{C}^{\mathbf{X}}(5)_5)/5 & \dots & \searrow & \text{EntroRate}(\mathcal{C}^{\mathbf{X}}(5)) \\
& & & & & & & \dots \nearrow \text{KS}
\end{array}$$

Figure 1. Approximation of KS entropy: the direct way

Formula (4) to be read together with formula (3) is illustrated by Figures 1 and 2, where directions of arrows indicate the direction of convergence.

Conditional entropy of ordinal patterns. In order to motivate a complexity measure based on considering successive ordinal patterns, we deduce some useful inequalities from the above statement. First fix some $d \in \mathbb{N}$ and note that, since $(H(\mathcal{C}^{\mathbf{X}}(d)_k) - H(\mathcal{C}^{\mathbf{X}}(d)_{k-1}))_{k=1}^{\infty}$ is monotonically non-increasing, it holds for all $k \in \mathbb{N}$

$$\begin{aligned}
H(\mathcal{C}^{\mathbf{X}}(d)_k) - H(\mathcal{C}^{\mathbf{X}}(d)_{k-1}) &\leq \frac{1}{k} \sum_{i=1}^k (H(\mathcal{C}^{\mathbf{X}}(d)_i) - H(\mathcal{C}^{\mathbf{X}}(d)_{i-1})) \\
&= \frac{1}{k} H(\mathcal{C}^{\mathbf{X}}(d)_k).
\end{aligned}$$

Because $\text{EntroRate}(\mathcal{C}^{\mathbf{X}}(d))$ is converging to the KS entropy in a monotonically non-decreasing way, the following inequality is valid:

$$\begin{aligned}
\text{KS} &\leq \liminf_{d \rightarrow \infty} (H(\mathcal{C}^{\mathbf{X}}(d)_{k_d}) - H(\mathcal{C}^{\mathbf{X}}(d)_{k_d-1})) \leq \liminf_{d \rightarrow \infty} \frac{1}{k_d} H(\mathcal{C}^{\mathbf{X}}(d)_{k_d}) \\
&\quad \text{for each sequence } (k_d)_{d=1}^{\infty} \text{ of natural numbers.}
\end{aligned}$$

Note that the faster the sequence $(k_d)_{d=1}^{\infty}$ increases, the nearer are the terms in the inequality, and good choices of the sequence provide equality of all terms in the last inequality. In particular, it holds

$$\text{KS} \leq \liminf_{d \rightarrow \infty} (H(\mathcal{C}^{\mathbf{X}}(d)_2) - H(\mathcal{C}^{\mathbf{X}}(d))),$$

being the background for the following definition (compare Unakafov and Keller [16]):

Definition 2. $H(\mathcal{C}^{\mathbf{X}}(d)_2) - H(\mathcal{C}^{\mathbf{X}}(d))$ is called the Conditional entropy of ordinal patterns of order d .

Indeed the concept given is a conditional entropy, since however $\mathcal{C}^{\mathbf{X}}(d)_2$ is a finer partition than $\mathcal{C}^{\mathbf{X}}(d)$, it reduces to a difference of entropies.

Permutation entropy. The description of KS entropy given by formula (4) includes a double-limit where the inner one is non-increasing and the outer is non-decreasing. Bandt et al. [6] (see also [1])

$$\begin{array}{ccccccccc}
H(\mathcal{C}^{\mathbf{X}}(1)_2) - H(\mathcal{C}^{\mathbf{X}}(1)_1) & H(\mathcal{C}^{\mathbf{X}}(1)_3) - H(\mathcal{C}^{\mathbf{X}}(1)_2) & H(\mathcal{C}^{\mathbf{X}}(1)_4) - H(\mathcal{C}^{\mathbf{X}}(1)_3) & \dots & \searrow & \text{EntroRate}(\mathcal{C}^{\mathbf{X}}(1)) \\
H(\mathcal{C}^{\mathbf{X}}(2)_2) - H(\mathcal{C}^{\mathbf{X}}(2)_1) & H(\mathcal{C}^{\mathbf{X}}(2)_3) - H(\mathcal{C}^{\mathbf{X}}(2)_2) & H(\mathcal{C}^{\mathbf{X}}(2)_4) - H(\mathcal{C}^{\mathbf{X}}(2)_3) & \dots & \searrow & \text{EntroRate}(\mathcal{C}^{\mathbf{X}}(2)) \\
H(\mathcal{C}^{\mathbf{X}}(3)_2) - H(\mathcal{C}^{\mathbf{X}}(3)_1) & H(\mathcal{C}^{\mathbf{X}}(3)_3) - H(\mathcal{C}^{\mathbf{X}}(3)_2) & H(\mathcal{C}^{\mathbf{X}}(3)_4) - H(\mathcal{C}^{\mathbf{X}}(3)_3) & \dots & \searrow & \text{EntroRate}(\mathcal{C}^{\mathbf{X}}(3)) \\
H(\mathcal{C}^{\mathbf{X}}(4)_2) - H(\mathcal{C}^{\mathbf{X}}(4)_1) & H(\mathcal{C}^{\mathbf{X}}(4)_3) - H(\mathcal{C}^{\mathbf{X}}(4)_2) & H(\mathcal{C}^{\mathbf{X}}(4)_4) - H(\mathcal{C}^{\mathbf{X}}(4)_3) & \dots & \searrow & \text{EntroRate}(\mathcal{C}^{\mathbf{X}}(4)) \\
H(\mathcal{C}^{\mathbf{X}}(5)_2) - H(\mathcal{C}^{\mathbf{X}}(5)_1) & H(\mathcal{C}^{\mathbf{X}}(5)_3) - H(\mathcal{C}^{\mathbf{X}}(5)_2) & H(\mathcal{C}^{\mathbf{X}}(5)_4) - H(\mathcal{C}^{\mathbf{X}}(5)_3) & \dots & \searrow & \text{EntroRate}(\mathcal{C}^{\mathbf{X}}(5)) \\
& & & & & \dots \nearrow \text{KS}
\end{array}$$

Figure 2. Approximation of KS entropy: the conditional way

have proposed the concept of Permutation entropy only needing one limit and being the starting point of using ordinal pattern methods. Here it is given in our general context.

Definition 3. For $d \in \mathbb{N}$, the quantity $\text{PE}^{\mathbf{X}}(d) = \frac{1}{d}H(\mathcal{C}^{\mathbf{X}}(d))$ is called the Permutation entropy of order d with respect to \mathbf{X} . Moreover, by the Permutation entropy with respect to \mathbf{X} we understand $\text{PE}^{\mathbf{X}} = \limsup_{d \rightarrow \infty} \text{PE}^{\mathbf{X}}(d)$.

The definition of Permutation entropy given is justified by the result of Bandt et al. [6] that if T is a piecewise monotone interval map, then it holds $\text{PE}^{\text{id}} = \text{KS}$. Here Ω is an interval and $\mathbf{X} = \text{id}$, where id denotes the identity on Ω . We don't want to say more about this result, but mention that general equality of KS entropy and Permutation entropy is an open problem, however as shown in [23] (under the assumptions above) it holds

$$\text{KS} \leq \text{PE}^{\mathbf{X}}.$$

Assuming that the increments of entropy of successive ordinal partitions are well-behaved in the sense that

$$\lim_{d \rightarrow \infty} (H(\mathcal{C}^{\mathbf{X}}(d+1)) - H(\mathcal{C}^{\mathbf{X}}(d)))$$

exists, by the Stolz-Cesàro theorem one obtains the inequality

$$\begin{aligned} \text{KS} &\leq \liminf_{d \rightarrow \infty} (H(\mathcal{C}^{\mathbf{X}}(d)_2) - H(\mathcal{C}^{\mathbf{X}}(d))) \\ &\leq \lim_{d \rightarrow \infty} (H(\mathcal{C}^{\mathbf{X}}(d+1)) - H(\mathcal{C}^{\mathbf{X}}(d))) \\ &= \lim_{d \rightarrow \infty} (H(\mathcal{C}^{\mathbf{X}}(d)) - H(\mathcal{C}^{\mathbf{X}}(d-1))) \\ &= \lim_{d \rightarrow \infty} \frac{1}{d} \sum_{i=1}^d (H(\mathcal{C}^{\mathbf{X}}(i)) - H(\mathcal{C}^{\mathbf{X}}(i-1))) \\ &= \lim_{d \rightarrow \infty} \frac{1}{d} H(\mathcal{C}^{\mathbf{X}}(d)) = \text{PE}^{\mathbf{X}} \end{aligned}$$

(compare [16]). This inequality is shading some light on the relationship of KS entropy, Permutation entropy and the various conditional entropy differences (see Figure 2) considered.

Summarizing, all quantities we have considered are related to the entropies $H(\mathcal{C}^{\mathbf{X}}(d)_k)$. For simplicity, we now switch to only one observable X . The generalization is simple but not necessary for the following. In the restricted case, for $d, k \in \mathbb{N}$ it holds

$$H(\mathcal{C}^X(d)_k) = - \sum_{(\pi_l)_{l=1}^k \in \Pi_d^k} P_{(\pi_l)_{l=1}^k} \ln P_{(\pi_l)_{l=1}^k},$$

where

$$P_{(\pi_l)_{l=1}^k} = P(\{\omega \in \Omega \mid (X(T^{ol+d-1}(\omega)), \dots, X(T^{ol}(\omega)), X(T^{ol-1}(\omega))) \text{ has ordinal pattern } \pi_l \text{ for } l = 1, 2, \dots, k\}).$$

The practical viewpoint. By the above considerations, despite of the the general relation of KS entropy and Permutation entropy, asymptotics of entropies taken from ordinal pattern distributions is relatively well understood, allowing a good interpretation of what they measure. The practical problem however is that asymptotics can be very slow and thus causing problems in the estimation

of $P(\mathcal{C}^X(d)_k)$ from orbits of a system when d or k is high. Here for $n, k \in \mathbb{N}$ a naive estimator of $P_{(\pi_l)_{l=1}^k}$ is

$$p_{(\pi_l)_{l=1}^k} = \frac{\#\{t \in \{d+k-1, \dots, N-1\} \mid (x_{t-i}, \dots, x_{t-i-d}) \text{ has ord. patt. } \pi_i \text{ for } i=0, \dots, k-1\}}{N-d-k+1}$$

and a naive estimator of $H(\mathcal{C}^X(d)_k)$ is

$$h(d, k) = - \sum_{(\pi_l)_{l=1}^k \in \Pi_d^k} p_{(\pi_l)_{l=1}^k} \ln p_{(\pi_l)_{l=1}^k}.$$

The problem is that one needs too many measurements for a reliable estimation if d and k are large. This is demonstrated in the one dimensional case for the logistic map with ‘maximal chaos’. Here $\Omega = [0, 1]$, T is defined by $T(\omega) = 4\omega(1 - \omega)$ for $x \in [0, 1]$, P is the Lebesgue measure, and X is the identity map. Note that this map and maps of the whole logistic family are often used for testing complexity measures. (For more information, see [25] and [26].)

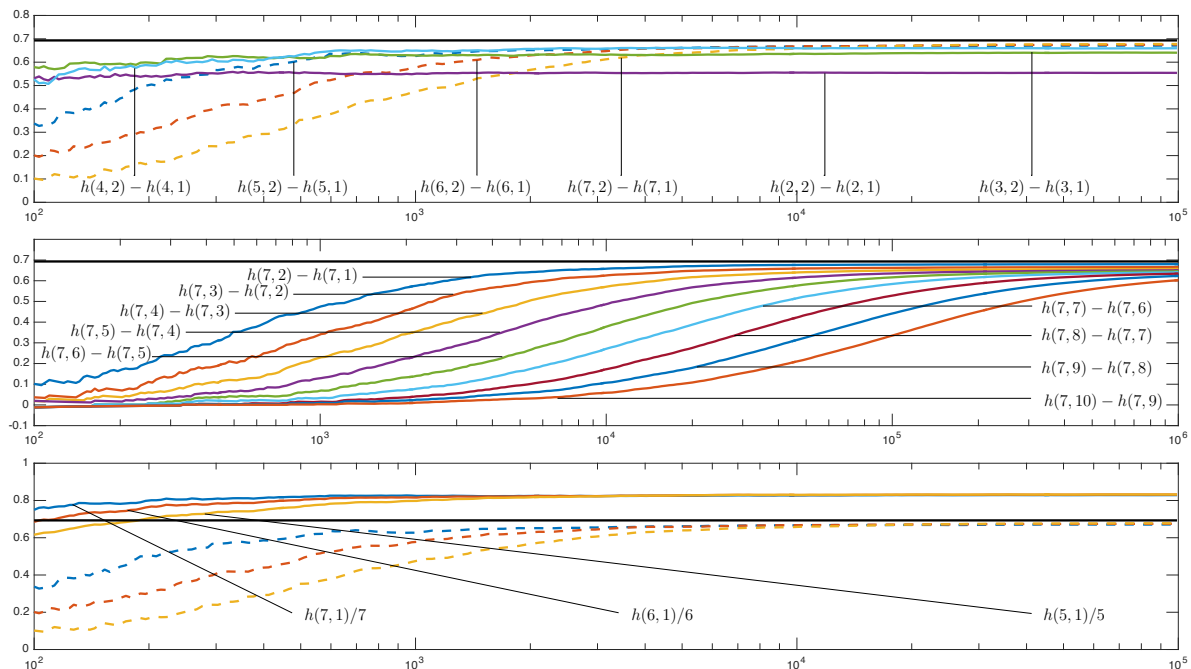


Figure 3. Different estimates of the KS entropy of $T(\omega) = 4\omega(1 - \omega)$; $\omega \in [0, 1]$ for different orbit lengths. On the top: $h(d, 2) - h(d, 1)$ for $d = 2, 3, \dots, 7$; in the middle: $h(7, k) - h(7, k - 1)$ for $k = 2, 3, \dots, 10$; on the bottom: $h(d, 2) - h(d, 1)$ versus $\frac{h(d, 1)}{d}$ for $d = 5, 6, 7$.

Figure 3 shows $h(d, k) - h(d, k - 1)$ and $\frac{h(d, 1)}{d}$ for different $d, k \in \mathbb{N}$ in dependence of orbits lengths of T between 10^2 , and 10^5 and 10^6 , respectively. The horizontal black line indicates the KS entropy of T which is $\ln 2$. The graphic on the top provides curves for $h(d, k) - h(d, k - 1)$ for $k = 2$ and $d = 2, 3, \dots, 7$ being estimates of the Conditional entropy of ordinal patterns of corresponding orders d . The approximation of the KS entropy is rather good for $d = 7$, but a long orbit is needed. In general, the higher d is the longer a stabilization of the estimate takes. A similar situation is demonstrated in the graphic in the middle, now with fixed $d = 7$ but increasing k . In the graphic on the bottom three of the curves from the upper graphic are drawn again in contrast to curves showing $\frac{h(d, 1)}{d}$ for $d = 5, 6, 7$. The latter can be considered as estimates of the Permutation entropy, which, however, are bad since d is not large enough. Note that by the result of Bandt et al. estimates for very high d must be near to the KS entropy. The results suggest that using Conditional entropy of ordinal patterns is a good

compromise. Good performance of Conditional entropy of ordinal patterns for the logistic family is reported in Unakafov and Keller [16].

3. Generalizations based on the families of Renyi and Tsallis entropies

The concept. It is natural to generalize Permutation entropy to Tsallis and Renyi entropy variants, which has first been done by Zunino et al. [11]. Liang et al. [12] discuss performances of a large list of complexity measures, among them the classical as well as Tsallis and Renyi Permutation entropies in tracking changes of EEG during different anesthesia states. They report that the class of Permutation entropies in some features show good performance relative to the other measures, with best results for the Renyi variant. Let us have a closer look at the new concepts.

Definition 4. For some given positive $\alpha \neq 1$, the empirical Renyi Permutation entropy (eRPE) and empirical Tsallis Permutation entropy (eTPE) of order $d \in \mathbb{N}$ and of delay $\tau \in \mathbb{N}$ of a time series $(x_t)_{t=0}^{N-1}$ with $N \in \mathbb{N}$ is defined by

$$\text{eRPE}(\alpha, d, \tau, (x_t)_{t=0}^{N-1}) = -\frac{1}{d(1-\alpha)} \ln \sum_{\pi \in \Pi_d} (p_\pi^\tau)^\alpha$$

and

$$\text{eTPE}(\alpha, d, \tau, (x_t)_{t=0}^{N-1}) = -\frac{1}{d(\alpha-1)} \left(1 - \sum_{\pi \in \Pi_d} (p_\pi^\tau)^\alpha \right),$$

respectively, with p_π as given in (2). (We include the factor $\frac{1}{d}$ into the entropy formulas only for reasons of comparability with the classical Permutation entropy.)

Some properties. As in general the Renyi and the Tsallis entropy of a distribution for $\alpha \rightarrow 1$ converge to the Shannon entropy, convergence of eRPE and eTPE to the ePE holds. The two concepts principally can be used in data analysis to more emphasize the role of small ordinal pattern probabilities if $\alpha < 1$ or of large ones if $\alpha > 1$ (compare the graphs of the functions $x \mapsto x \ln x$, $x \mapsto x^\alpha$ for different α and $x \in [0, 1]$). The consequences of this weighting gets obvious for the eRPE when considering limits for $\alpha \rightarrow 0$ and $\alpha \rightarrow \infty$. One easily sees that

$$\begin{aligned} \lim_{\alpha \rightarrow \infty} \text{eRPE}(\alpha, d, \tau, (x_t)_{t=0}^{N-1}) &= -\frac{\ln(\max_{\pi \in \Pi_d} p_\pi^\tau)}{d} \\ \lim_{\alpha \rightarrow 0} \text{eRPE}(\alpha, d, \tau, (x_t)_{t=0}^{N-1}) &= \frac{\ln(\#\{\pi \in \Pi_d \mid p_\pi^\tau \neq 0\})}{d}, \end{aligned}$$

meaning that for large α the eRPE mainly measures the largest relative ordinal pattern frequency, with low entropy for high relative frequency, and for small α the number of occurring ordinal patterns. Since

$$\text{eTPE}(\alpha, d, \tau, (x_t)_{t=0}^{N-1}) = \frac{e^{d(1-\alpha) \text{eRPE}(\alpha, d, \tau, (x_t)_{t=0}^{N-1})} - 1}{d(1-\alpha)},$$

the eTPE is only a monotone functional of the eRPE for fixed α ; despite of a different scale it has similar properties as the eRPE.

For $\alpha = 2$ the eRPE has a nice interpretation. Having $N_\pi = (N - d\tau)p_\pi^\tau$ in the time series, then there are $\frac{N_\pi(N_\pi-1)}{2}$ pairs of different times providing ordinal pattern π . Since all in all we have $\frac{(N-d\tau)(N-d\tau-1)}{2}$ different time pairs, the quantity

$$\frac{\sum_{\pi \in \Pi_d} N_\pi(N_\pi - 1)}{(N - d\tau)(N - d\tau - 1)} = \frac{\sum_{\pi \in \Pi_d} N_\pi^2 - (N - d\tau)}{(N - d\tau)^2 - (N - d\tau)} \approx \sum_{\pi \in \Pi_d} (p_\pi^\tau)^2$$

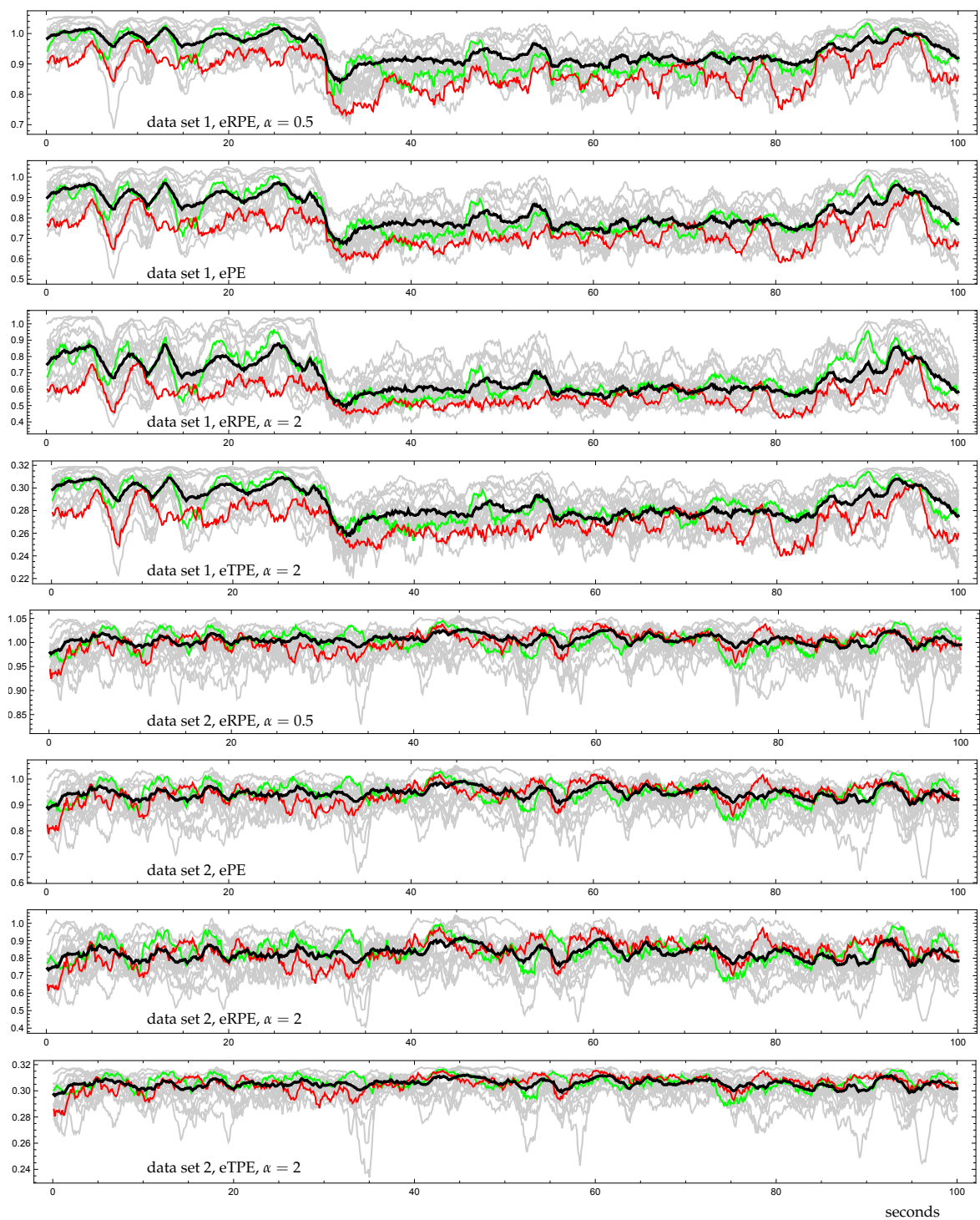


Figure 4. Comparison of eRPE for $\alpha = 0.5, 2$, eTPE for $\alpha = 2$ and ePE, computed from EEG recordings before stimulator implantation (data set 1) and after stimulator implantation (data set 2) for 19 channels using a shifted time window, order $d = 3$ and delay $\tau = 4$. Highlighting, in particular, the channels T3 and P3 as well as the entropy over all channels by a fat black line. The sampling rate is 256 Hz.

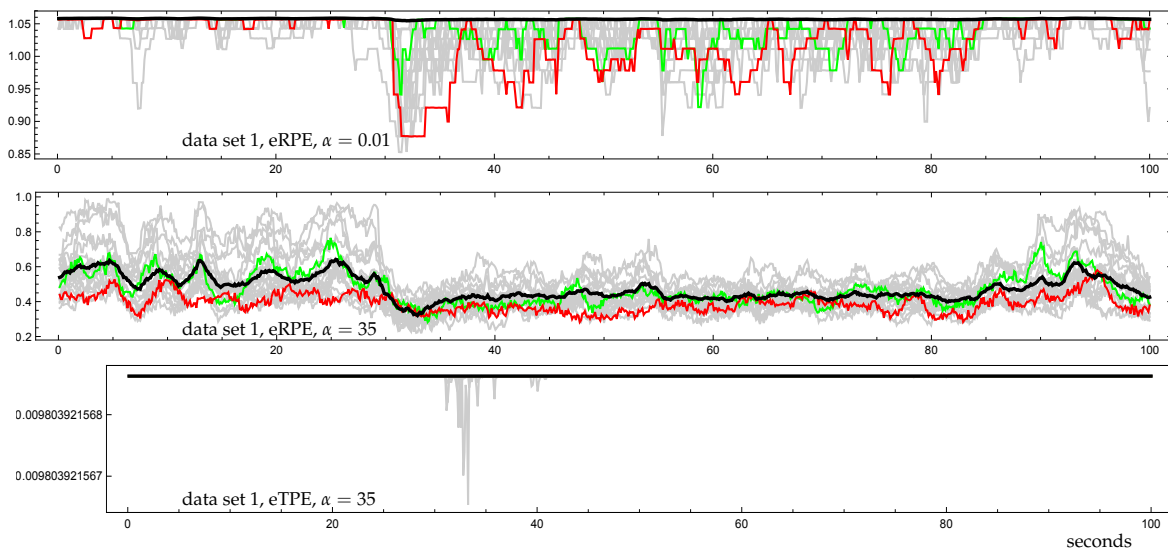


Figure 5. eRPE for $\alpha = 0.01, 35$ and eTPE for $\alpha = 35$ computed from data set 1 for 19 channels using a shifted time window, $d = 3$ and $\tau = 4$ (cf. Figure 4).

Table 1. Concordance of sign of entropy differences of ePE and eRPE for given α

α	0.5	0.8	0.9	1.1	1.2	1.5	2	250
Fp2	98.69 %	99.52 %	99.77 %	99.78 %	99.58 %	99.05 %	98.38 %	94.85 %
T3	95.41 %	98.39 %	99.22 %	99.26 %	98.59 %	96.92 %	95.20 %	89.31 %
P3	93.18 %	97.71 %	98.93 %	99.05 %	98.21 %	96.30 %	94.33 %	83.66 %

can be interpreted as the degree of recurrence in the time series. This quantity is in fact the *symbolic correlation integral* recently introduced by Caballero et al. [27].

Demonstration. We demonstrate the performance of eRPE and eTPE for different α on the base of EEG data discussed in [17]. For this, we consider two parts each of a 19-channel scalp EEG of a boy with lesions predominantly in the left temporal lobe caused by a connatal toxoplasmosis. The data were sampled with a rate of 256 Hz, meaning that 256 measurements were obtained each second. The first data part (data set 1) was taken from an EEG derived at an age of eight years and the second one (data set 2) was derived at an age of 11 years, four months after the implantation of a vagus stimulator. Note that epileptic activity was significant before vagus stimulation and was the reason for the implantation. (For some more details on the data, see [17]).

Figure 4 shows ePE, eRPE for $\alpha = 0.5, 2$ and eTPE for $\alpha = 2$ for the two data sets in dependence on a shifted time window, where $d = 3$ and $\tau = 4$. Each graphic represents the 19 channels for a fixed entropy by 19 entropy curves. Among the channels are T3 and P3, which are interesting in the following. The curve related to a fixed channel contains pairs (t, h_t) , where h_t is the entropy of the segment of the related time series ending at time t and containing $2 \cdot 256 + 3 \cdot 4$ successive measurements. (Each segment provides $512 = 2 \cdot 256$ ordinal patterns representing a time segment of two seconds.) The t are chosen from a time segment of 100 seconds, where the beginning time is set to 0 for simplicity. We have added a fat black curve representing the entropy of the whole brain instead of the single channels. Here the relative frequencies used for entropy determination were obtained by counting ordinal patterns in the considered time segment pooled over all channels.

The two EEG parts reflect changes of brain dynamics. Whereas the first EEG part shows low entropies relative to the other channels of P3 and, partially, of T3, entropies of both channels are relatively higher in the second EEG part. More information with additional data sets derived directly before and after the vagus stimulator implantation is available from [17], where only the ePE was

considered. Here we want to note that **P3** and **T3** are from the left part of the brain with the lesions, and mainly **P3** seems to reflect some kind of irregular behavior related to them. The most interesting point is that before vagus stimulator implantation **P3** and partially **T3** are conspicuous both in phases with and without epileptic activity. For orientation: The graphic given in Figure 4 for data set 1 and in Figure 5 show a part from around 30 to 90 seconds with low entropies for pooled ordinal patterns (fat black curve) related to a generalized epileptic seizure meaning epileptic activity in the whole brain.

Figure 4 suggests that a visual inspection of the data using eRPE and eTPE instead of ePE does not seem to get further insights when α is chosen close to 1. Here it is important to note that, for the parameters considered, all ordinal patterns are accessed at many times (with some significant frequency). Our guess is supported by Table 1. For each of the channels FP2, **T3**, and **P3** and given α , the relative frequency of concordant pairs of the observations ePE, eRPE at time s and ePE, eRPE at time t among all pairs (s, t) ; $s < t$ is shown. (s, t) is said to be concordant if the difference between ePE at times s and t and the difference between eRPE at times s and t have the same sign.

The results particularly reflect the fact that for channel Fp2, providing measurements from the front of the brain, the ordinal patterns are more equally distributed than for **T3**, as well as that for **P3** the distribution of ordinal patterns is the farthest from equidistribution. For contrast, we also consider $\alpha = 250$. Figure 5 related to data set 1 indicates that extreme choices of α could be useful in order to analyze and visualize changes in the brain dynamics more forceful. The upper graphic of eRPE for $\alpha = 0.01$ shows that at the beginning of an epileptic seizure the number of ordinal patterns abruptly decreases for nearly all channels and, after some increasing, stays on a relatively low level until the end of the seizure. For $\alpha = 35$ it is interesting to look at the whole ensemble of the entropies.

Here eRPE indicates much more variability of largest ordinal pattern frequencies of the channels in the seizure-free parts than in the seizure epochs. The very low entropy at the beginning of the seizure mainly goes back to much more only increasing or decreasing ordinal patterns. Here the special kind of scaling given by the eTPE allows to emphasize the special situation at the beginning of a seizure and can therefore be interesting for an automatic detection of epileptic seizures, whereby a right tuning of α is important.

4. Classification on the base of different entropies

Classification is an important issue in the analysis of EEG data, for which in many cases entropy measures can be exploited. Here often it is not clear which of these measures are the best and how much they are 'overlapping'. In view of the discussion in Section 2, we note that although the nature of Shannon entropy based measures, which we considered in this paper, is nearly similar for low orders, their performance can be rather different in dependence on the delay.

Here we want to discuss classification of EEG data using ePE, empirical Conditional entropies of ordinal patterns (eCE, see below) and, additionally, Approximate entropy (AppEn) (see Pincus [28]) and Sample entropy (SampEn) (see Richmann and Moormann [29], and extending work in Keller et al. [9]). For definition and usage of AppEn and SampEn, in particular for the parameter choice ("AppEn(2, 0.2 σ , x)", "SampEn(2, 0.2 σ , x)"), we directly refer to [9]. The eCE, motivated and defined for dynamical systems in Section 2, is given by the following definition (compare also [9]):

Definition 5. Given a time series $(x_t)_{t=0}^{N-1}$ with $N \in \mathbb{N}$, the quantity

$$\text{eCE} \left(d, \tau, (x_t)_{t=0}^{N-1} \right) = \sum_{\pi \in \Pi_d} p_{\pi}^{\tau} \ln p_{\pi}^{\tau} - \sum_{\pi_1, \pi_2 \in \Pi_d} p_{\pi_1, \pi_2}^{\tau} \ln p_{\pi_1, \pi_2}^{\tau}$$

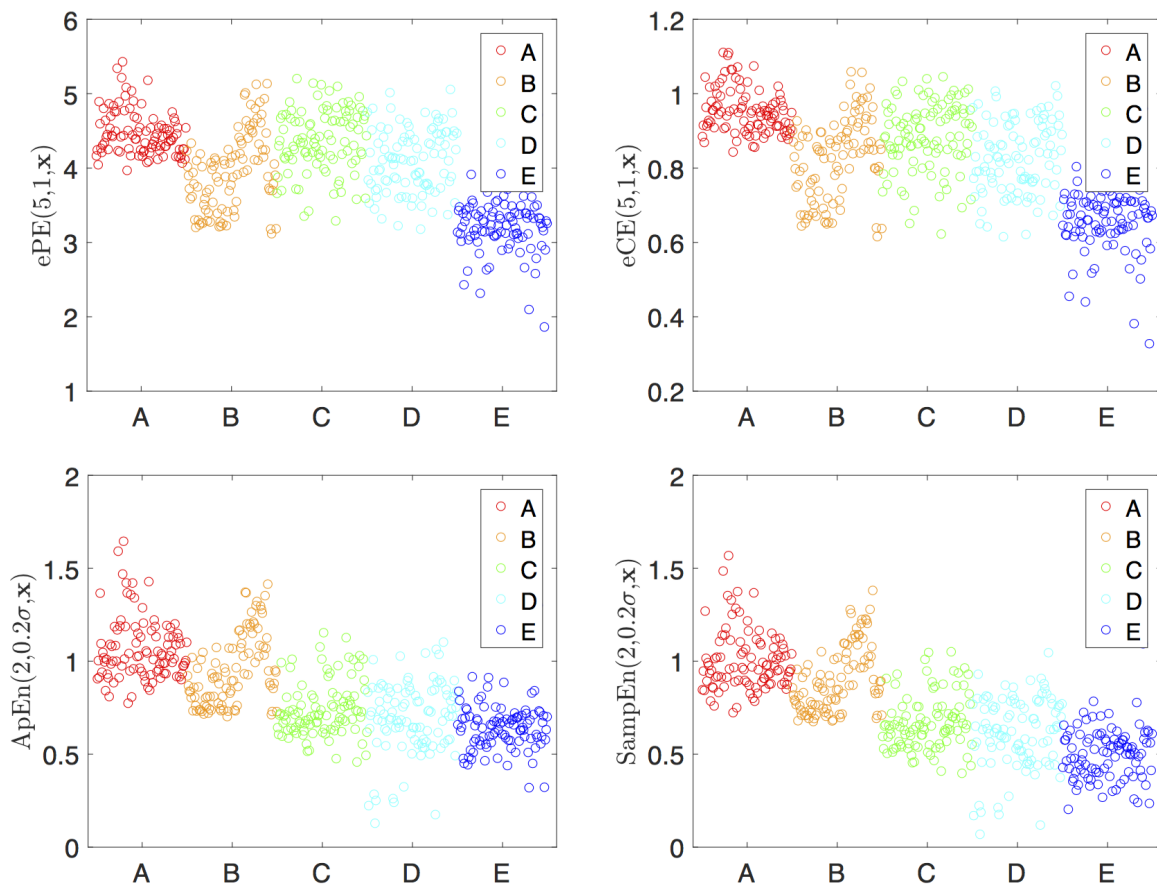


Figure 6. Entropy of the EEG data sorted by groups for four different entropy measures

is called empirical Conditional entropy of ordinal patterns (eCE) of order $d \in \mathbb{N}$ and of delay $\tau \in \mathbb{N}$, where p_{π}^{τ} is defined by (2) and p_{π_1, π_2}^{τ} by

$$p_{\pi_1, \pi_2}^{\tau} = \frac{1}{N - d\tau - 1} \# \{t \in \{d\tau, d\tau + 1, \dots, N - 2\} \mid (x_t, x_{t-\tau}, \dots, x_{t-d\tau}), (x_{t+1}, x_{t+1-\tau}, \dots, x_{t+1-d\tau}) \text{ have ordinal patterns } \pi_1, \pi_2\}$$

Note that for each time series $(x_t)_{t=0}^{N-1}$ and each $d \in \mathbb{N}$, it holds $\text{eCE}(d, 1, (x_t)_{t=0}^{N-1}) = h(d, 2) - h(d, 1)$. So $\text{eCE}(d, 1, (x_t)_{t=0}^{N-1})$ is an estimate of the Conditional entropy of ordinal patterns defined in Section 2.

Table 2. Results for classification on the base of one entropy

entropy	classification accuracy (in %)
ApEn	31.0
SampEn	37.8
ePE	32.0
eCE	30.0

The data. The considered EEG data are recordings from the Bonn EEG Database [18], each of a length of 23.6 seconds recorded at a sampling rate of 173.61 Hz. The data consists of five groups of each 100 recordings:

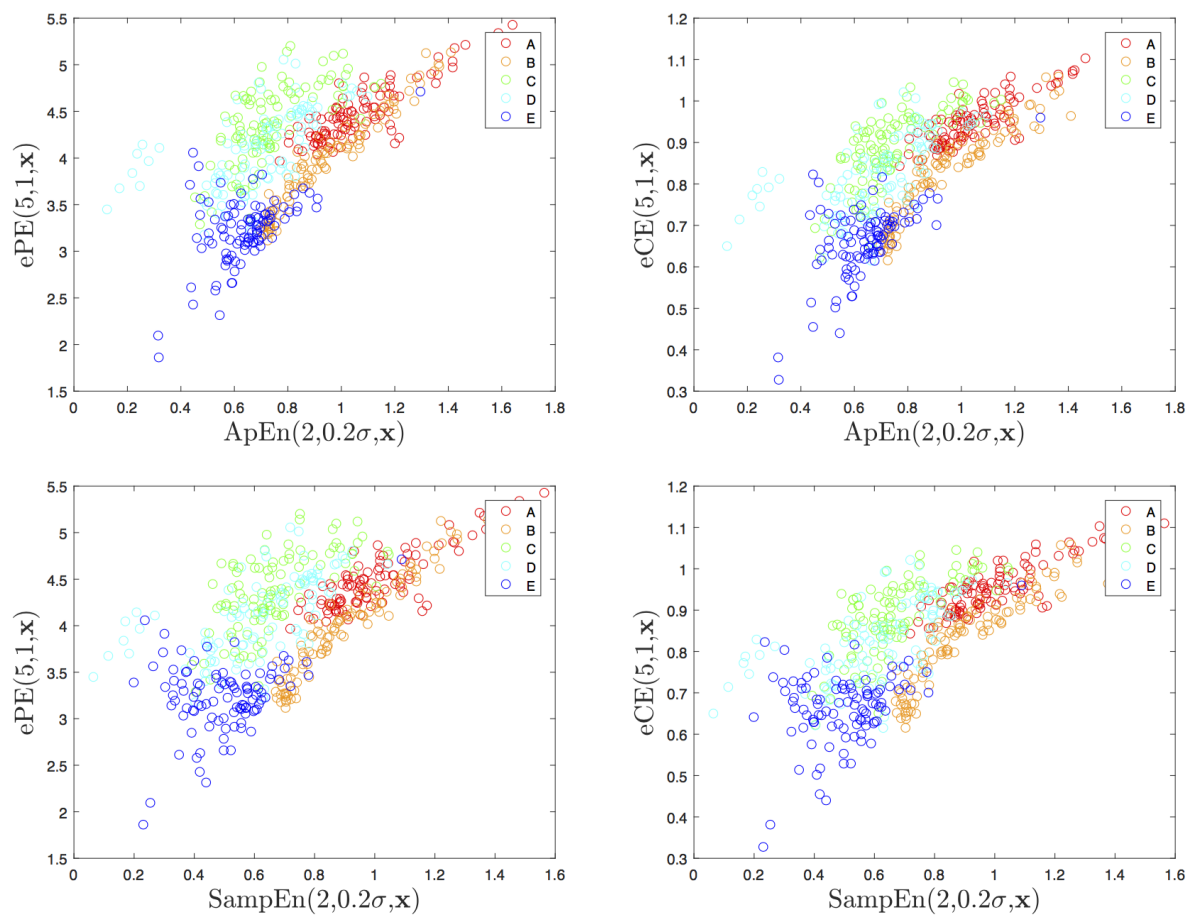


Figure 7. One entropy versus another one for four entropy combinations

- *group A*: surface EEG's recorded from healthy subjects with open eyes,
- *group B*: surface EEG's recorded from healthy subjects with closed eyes,
- *group C*: intracranial EEG's recorded from subjects with epilepsy during a seizure-free period from within the epileptogenic zone,
- *group D*: intracranial EEG's recorded from subjects with epilepsy during a seizure-free period from hippocampal formation of the opposite hemisphere of the brain,
- *group E*: intracranial EEG's recorded from subjects with epilepsy during a seizure period.

In contrast to [9], where the groups A and B and the groups C and D were pooled, each of the groups is considered separately in the following.

Table 3. Results for classification on the base of two entropies

entropy	classification accuracy (in %)
ApEn & SampEn	51.0
ApEn & ePE	58.0
ApEn & eCE	61.8
SampEn & ePE	64.0
SampEn & eCE	64.6
ePE & eCE	48.2

In order to give a visual impression of how the considered entropies separate the data, we provide three figures. For the ordinal pattern based measures we have chosen order $d = 5$ and delay $\tau = 1$. Figure 6 shows the values of the four considered entropies for all data sets. The values

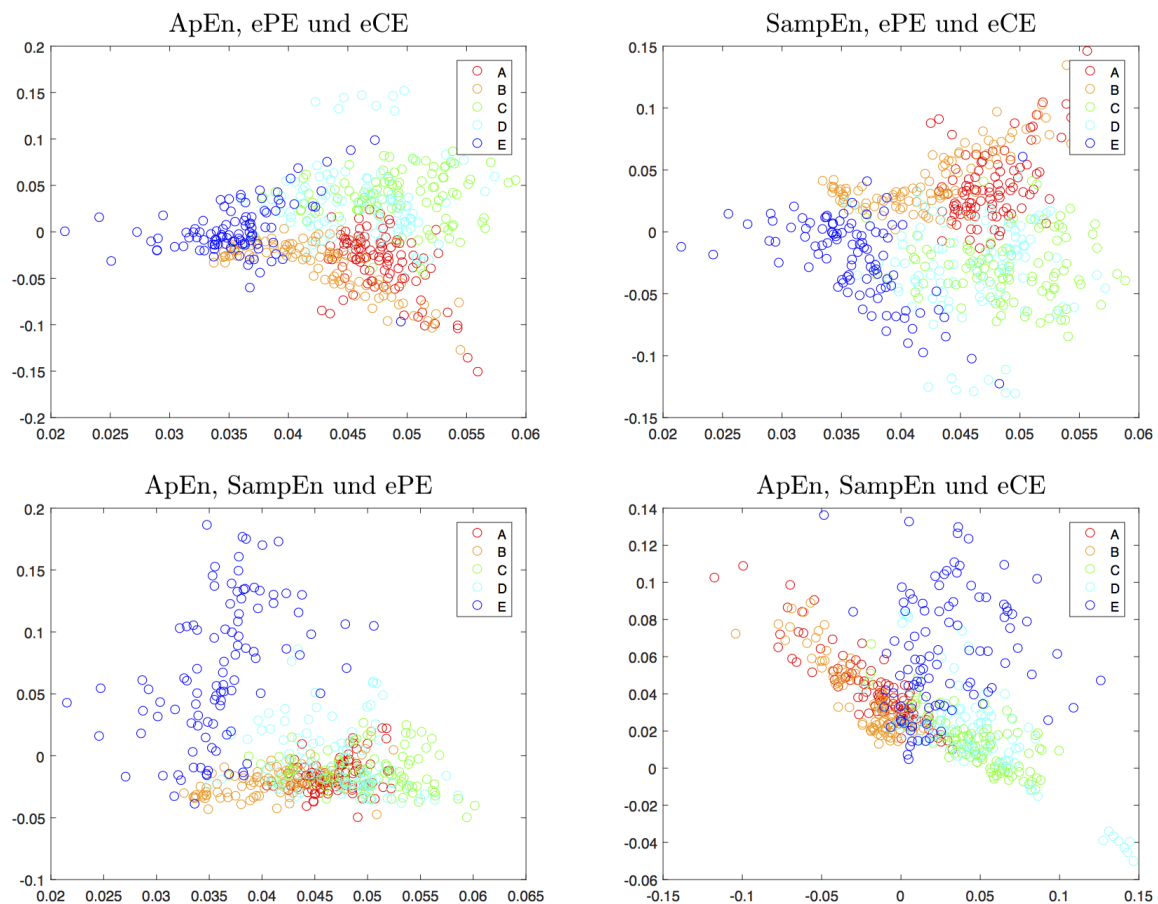


Figure 8. Second principal component versus first one obtained from principal component analysis on three entropy variables

obtained are drawn from the left to the right starting from those from group A and ending with those from group E. One sees that each of the considered entropies is not much separating, however, one can see different kinds of separation properties for the ordinal pattern based entropies and the two other entropies. A better separation is seen in Figure 7 where one entropy measure is plotted versus another one, in four different combinations. Here the discrimination between E, the union of A and B, and the union of C and D is rather good, confirming the results in [9], but both A, B and C, D are strongly overlapping. The general separation seems to be slightly better using three entropies, which is illustrated by Figure 8. Here, however, a two-dimensional representation is chosen by plotting the second principal component versus the first one, both obtained by principal component analysis from the three entropy components variables.

Table 4. Results for classification on the base of three entropy

entropy	classification accuracy (in %)
ApEn & SampEn & ePE	67.4
ApEn & SampEn & eCE	66.8
ApEn & ePE & eCE	65.4
SampEn & ePE & eCE	71.8

In order to get more objective and reliable results, we have tested a classification of the data on the base of the chosen entropies and of Random Forests (see Breiman [30]), a popular efficient machine learning method. For this purpose the data were randomly divided into a learning group consisting

of 80% of the data sets and a testing group consisting of the remaining 20%. The obtained accuracy of the classification, i.e. the relative frequency of correctly classified data sets, was averaged over 1000 testing trials for each entropy combination considered. The results of this procedure, summarized in Tables 2, 3 and 4, show that including a further entropy results in a higher accuracy, but that the way of combining entropies is crucial for the results. Note that combining all four entropies provides an accuracy of only 70.6%.

Clearly, accuracy of classification using ePE and eCE depends on the choice of the parameters of the entropies. Whereas the choice of higher orders doesn't make sense by statistical reasons as already mentioned in Section 2, testing different delays τ is useful since the delay contains important scale information. Note that already for both, ePE and eCE, the classification accuracy varies by more than 14%. Considering only delays $\tau = 1, 2, \dots, 9$ for $d = 5$, the maximal accuracies for ePE and eCE is 45.01% and 48.0% for $\tau = 6$ and $\tau = 5$, respectively. Note that both results are better than those for the SampEn (see Table 2) and that $\tau = 1$ provides the worst results. Combining two ePE's and eCE's for delays, in $\{1, 2, \dots, 9\}$ one reaches accuracy of 61.79% (for delays 1 and 2) and 62.28% (for delays 1 and 9), respectively.

5. Resume

Throughout this paper, we have discussed ordinal pattern based complexity measures both from the viewpoint of their theoretical foundation and their application in EEG analysis, in the center the Permutation entropy and conditional variants of it. We have pointed out that, as in many situations in (model-based) data analysis, one has to attract attention to the discrepancy of theoretical asymptotics and statistical requirements, here in view of estimating KS entropy. In the case of moderately but not extremely long data sets, the concept of Conditional entropy of ordinal patterns as discussed could be a compromise. It has been shown to have a better performance than the classical Permutation entropy in many situations.

A good way of further investigating performances of ordinal pattern based measures is an extensive testing of these measures for data classification. In this direction, the results of this paper for a restricted parameter choice are already promising, however further studies are required. For this manner, based on the considerations in Sections 4 and 3, the authors also propose including Renyi and Tsallis variants of Permutation entropy (with an extreme parameter choice) and classical concepts like Approximate entropy and Sample entropy. The latter are interesting in combination with ordinal complexity measures since they possibly address other features. The most important challenge, however, is to deal with the great number of entropy measures and parameters, in the opinion of the authors this can be faced by using machine learning ideas in a sophisticated way.

Acknowledgements

We would like to thank Heinz Lauffer from the Clinic and Outpatient Clinic for Pediatrics of the University Greifswald for providing the EEG data discussed in Section 3. Moreover, the first author thanks Manuel Ruiz Marín from the Technical University of Cartagena and his colleagues for some nice days in South Spain discussing their concept of Symbolic correlation integral.

Conflicts of Interest

The authors declare no conflict of interest.

Bibliography

1. Bandt, C.; Pompe, B. Permutation entropy—A natural complexity measure for time series. *Phys. Rev. E* **2002**, *66*, 174102.

2. Amigó, J.M.; Keller, K.; Kurths, J. (eds) Recent progress in symbolic dynamics and permutation complexity. Ten years of permutation entropy. *Eur. Phys. J. Spec. Top.* **2013**, *222*, 247–257.
3. Zanin, M.; Zunino, L.; Rosso, O.A.; Papo, D. Permutation entropy and its main biomedical and econophysics applications: A review. *Entropy* **2012**, *14*, 1553–1577.
4. Amigó J.M.; Keller, K.; Unakafova, V.A. Ordinal symbolic analysis and its application to biomedical recordings. *Phil. Trans. Royal. Soc. A*, **2015**, *373*, 20140091.
5. Amigó, J.M. *Permutation Complexity in Dynamical Systems*. Springer-Verlag: Berlin-Heidelberg, 2010.
6. Bandt, C.; Keller, G.; Pompe, B. Entropy of interval maps via permutations. *Nonlinearity* **2002**, *15*(5), 1595–1602.
7. Liu, X-F.; Wang, Y. Fine-grained permutation entropy as a measure of natural complexity for time series. *Chin. Phys. B* **2009**, *18*, 2690–2695.
8. Fadlallah, B.; Chen, B.; Keil, A.; Príncipe, J. Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information. *Phys. Rev. E* **2013**, *87*, 022911.
9. Keller, K.; Unakafov, A.M.; Unakafova, V.A. Ordinal Patterns, Entropy, and EEG. *Entropy* **2014**, *16*, 6212–6239.
10. Bian, C.; Qin, C.; Ma, Q.D.Y.; Shen, Q. Modified permutation-entropy analysis of heartbeat dynamics. *Phys. Rev. E* **2012**, *85*, 021906.
11. Zunino, L.; Perez, D.G.; Kowalski, A.; Martín, M.T.; Garavaglia, M.; Plastino, A.; Rosso, O.A. Brownian motion, fractional Gaussian noise, and Tsallis permutation entropy. *Physica A* **2008**, *387*, 6057–6068.
12. Liang, Z.; Wang, Y.; Sun, X.; Li, D.; Voss, L.J.; Sleight, J.W.; Hagihira, S.; Li, X. EEG entropy measures in anesthesia. *Front. Comput. Neurosci.* **2015**, *9*, 00016.
13. Li, D.; Li, X.; Liang, Z.; Voss, L.J.; Sleight, J.W. Multiscale permutation entropy analysis of EEG recordings during sevoflurane anesthesia. *J. Neural Eng.* **2010**, *7*, 046010.
14. Ouyang, G.; Li, J.; Liu, X.; Li, X. Dynamic characteristics of absence EEG recordings with multiscale permutation entropy analysis. *Epilepsy Res.*, **2013**, *104*, 246–252.
15. Azami, H.; Escudero, J. Improved multiscale permutation entropy for biomedical signal analysis: Interpretation and application to electroencephalogram recordings. *Biomed. Signal Proces.* **2016**, *23*, 28–41.
16. Unakafov, A.M.; Keller, K. Conditional entropy of ordinal patterns. *Physica D* **2013**, *269*, 94–102.
17. Keller, K.; Lauffer, H. Symbolic analysis of high-dimensional time series. *Int. J. Bifurc. Chaos* **2003**, *13*, 2657–2668.
18. Andrzejak, R.G.; Lehnertz, .K; Rieke, C.; Mormann, F; David, P; Elger, C.E. Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Phys. Rev. E* **2001**, *64*, 061907.
19. Unakafova, V.A.; Keller, K. Efficiently Measuring Complexity on the Basis of Real-World Data. *Entropy* **2013**, *15*, 4392–4415.
20. Walters, P. *An Introduction to Ergodic Theory*. Springer-Verlag: New York, 1982.
21. Takens, F. Detecting strange attractors in turbulence. in: “Dynamical Systems and Turbulence” (eds.: Rand, D.A.; Young, L.S.), *Lecture Notes in Mathematics* 898, Springer-Verlag: Berlin-New York, 1981, 366–381.
22. Gutman, J. Takens’ embedding theorem with a continuous observable. preprint, math/1510.05843.
23. Antoniouk, A.; Keller, K.; Maksymenko, S. Kolmogorov-Sinai entropy via separation properties of order-generated σ -algebras *Discrete Contin. Dyn. Syst. A* **2014**, *34*, 1793–1809.
24. Keller, K.; Maksymenko, S.; Stolz, I. Entropy determination based on the ordinal structure of a dynamical system. *Discrete Contin. Dyn. Syst. B* **2015**, *20*, 3507–3524.
25. Sprott, J.C. *Chaos and time-series analysis*. Oxford University Press: Oxford, 2003.
26. Young, L.-S. Mathematical theory of Lyapunov exponents. *J. Phys. A: Math. Theor.* **2013**, *46*, 1–17.
27. Caballero, M.V.; Mariano, M.; Ruiz, M. Draft: Symbolic correlation integral. Getting rid of the proximity parameter, http://data.leo-univ-orleans.fr/media/seminars/175/WP_208.pdf (accessed: 14-Feb-2017).
28. Pincus, S.M. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences* **1991**, *88*, 2297–2301.
29. Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology* **2000**, *278*, 2039–2049.
30. Breiman, L. Random forests. *Machine Learning* **2001**, *45*, 5–32.



© 2017 by the authors; licensee *Preprints*, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).