ENTROPY DETERMINATION BASED ON THE ORDINAL STRUCTURE OF A DYNAMICAL SYSTEM

KARSTEN KELLER
Universität zu Lübeck
Institut für Mathematik, Ratzeburger Allee 160
Lübeck, 23562, Germany

SERGIY MASYMENKO
Institute of Mathematics of NAS of Ukraine
Tereshchenkivs’ka str. 3
Kyiv 01601, Ukraine

INGA STOLZ
Universität zu Lübeck
Institut für Mathematik, Ratzeburger Allee 160
Lübeck, 23562, Germany

(Communicated by the associate editor name)

Abstract. The ordinal approach to evaluate time series due to innovative works of Bandt and Pompe has increasingly established itself among other techniques of nonlinear time series analysis. In this paper, we summarize and generalize the theory of determining the Kolmogorov-Sinai entropy of a measure-preserving dynamical system via increasing sequences of order generated partitions of the state space. Our main focus are measuring processes without information loss. Particularly, we consider the question of the minimal necessary number of measurements related to the properties of a given dynamical system.

1. Introduction

Since the invention of permutation entropy by Bandt and Pompe [8] and the proof of its coincidence with Kolmogorov-Sinai entropy for piecewise monotone interval maps by Bandt et al. in [7], there is some increasing interest in considering time series and dynamical systems from the pure ordinal point of view (see Amigó, [4]). The idea behind this viewpoint is that much information of a system is already contained in ordinal patterns describing the up and down of its orbits. This ordinal view can be particularly useful when having physical quantities for which the statement that a measuring value is larger than another one is well interpretable, but concrete purely given differences of measuring values are not. A prominent example is the (indirect)

1991 Mathematics Subject Classification. Primary: 58F15, 58F17; Secondary: 53C35.

Key words and phrases. Kolmogorov-Sinai entropy, permutation entropy, ordinal time series analysis, algebra reconstruction dimension.

The authors were supported by Marie Curie Actions - International Research Staff Exchange Scheme (IRSES) FP7-People-2011-IRSES. Project number 295164.
measurement of temperature as the mean kinetic energy of the particles of a system by a thermometer. One can make statements about what is warmer or colder, but, for example, the interpretation of an increase by 1°C with not knowing the baseline value is complicated.

This paper is generally discussing the Kolmogorov-Sinai entropy from the ordinal viewpoint. It reviews and particularly extends and generalizes former results given by Antoniouk et al. [6], Amigó [3], Keller [14], Keller and Sinn [15, 16] and Amigó et al. [5]. Aspects of entropy estimation are touched.

The framework. The basic model of our discussion is a measure-preserving dynamical system $(\Omega, \mathcal{A}, \mu, T)$, i.e. $\Omega$ is a non-empty set whose elements are interpreted as the states of a system, $\mathcal{A}$ is a sigma-algebra on $\Omega$, $\mu : \mathcal{A} \to [0, 1]$ is a probability measure, and $T : \Omega \to \Omega$ is a $\mathcal{A}$-$\mathcal{A}$-measurable $\mu$-preserving map describing the dynamics of the system. $\mu$-preserving means that $\mu(T^{-1}(A)) = \mu(A)$ for all $A \in \mathcal{A}$; the measure $\mu$ is then called $T$-invariant.

We want to have some kind of regularity of $T$ by assuming at least one of the following conditions:

- $T$ is ergodic with respect to $\mu$, i.e.
  
  \begin{align*}
  \mu(A) \in \{0,1\} \text{ for all } A \in \mathcal{A} \text{ with } T^{-1}(A) = A,
  \end{align*}

- $(1) \Omega$ can be embedded into some compact metrizable space so that $\mathcal{A} = \mathcal{B}(\Omega)$.

Here and in the whole paper, $\mathcal{B}(\Omega)$ denotes the Borel $\sigma$-algebra in the case that $\Omega$ is a topological space. As usual, equivalent to $T$ is ergodic with respect to $\mu$, we say that $\mu$ is ergodic for $T$.

Often the states of a system, whatever they are, cannot be accessed directly, but information on them can be obtained by measurements. In this paper such measurements are assumed to be given via observables $X_1, X_2, X_3, \ldots$ defined as $\mathbb{R}$-valued random variables on the probability space $(\Omega, \mathcal{A}, \mu)$. So the measurements are provided by a stochastic process - we say sequence of observables $X = (X_i)_{i \in \mathbb{N}}$ - whose realization has components $(X_i(T^t(\omega)))_{t \in \mathbb{N}_0}$. Here $X_i(T^t(\omega))$ is interpreted as the $i$-th measured value from the system at time $t$ when starting in state $\omega \in \Omega$.

A priori we have infinitely many observables providing more and more information, the finite case, however, is included by equality of all $X_i$: $i \geq n$ for some $n \in \mathbb{N}$. We will write $X = (X_i)_{i=1}^n$ in the case of finitely many observables and $X = X$ in the case of only one observable $X$.

Unless otherwise stated, in the following $(\Omega, \mathcal{A}, \mu, T)$ is a measure-preserving dynamical system and $X = (X_i)_{i \in \mathbb{N}}$ a sequence of observables.

Kolmogorov-Sinai entropy. In order to recall the Kolmogorov-Sinai entropy, let $q \in \mathbb{N}$ and $\mathcal{P} = \{P_1, P_2, \ldots, P_q\} \subset \mathcal{A}$ be a finite partition of $\Omega$, i.e. $\Omega = \bigcup_{l=1}^q P_l$, $P_l \neq \emptyset$ for $l = 1, 2, \ldots, q$, $P_{l_1} \cap P_{l_2} = \emptyset$ for different $l_1, l_2 \in \{1, 2, \ldots, q\}$, and let $A = \{1, 2, \ldots, q\}$ be the corresponding alphabet. Each word $a_1 a_2 \ldots a_t$ of length $t \in \mathbb{N}$ defines a set

$$P_{a_1 a_2 \ldots a_t} := \{\omega \in \Omega \mid (\omega, T^t(\omega), \ldots, T^{t-1}(\omega)) \in P_{a_1} \times P_{a_2} \times \ldots \times P_{a_t}\},$$
and the collection of all non-empty sets obtained for such words of length \( t \) provides a partition \( \mathcal{P}_t \subset \mathcal{A} \) of \( \Omega \). In particular, \( \mathcal{P}_1 = \mathcal{P} \).

The \textit{entropy rate} of \( T \) with respect to an initial partition \( \mathcal{P} \) is given by

\[
h_\mu(T, \mathcal{P}) = \lim_{t \to \infty} \frac{1}{t} H_\mu(\mathcal{P}_t),
\]

where \( H_\mu(\mathcal{C}) \) denotes the \textit{(Shannon) entropy} of a finite partition \( \mathcal{C} = \{C_1, C_2, \ldots, C_q\} \subset \mathcal{A} \) of \( \Omega \); \( q \in \mathbb{N} \), i.e.

\[
H_\mu(\mathcal{C}) = -\sum_{l=1}^{q} \mu(C_l) \ln(\mu(C_l))
\]

(with \( 0 \ln(0) := 0 \)), and the \textit{Kolmogorov-Sinai entropy} is defined by

\[
h_{\mu}^{KS}(T) = \sup_{\mathcal{P} \text{ finite partition}} h_\mu(T, \mathcal{P}).
\]

Although the Kolmogorov-Sinai entropy is well-defined, its determination is not easy. In some special cases one can find finite partitions already determining it, usually called generating partitions (see Definition 6.3), however, do not exist or are not accessible. As a substitute, we want to consider special sequences of partitions only depending on the ordinal structure of a dynamical system.

**Ordinal partitioning.** For a single observable \( X \) on \( (\Omega, \mathcal{A}, \mu, T) \) and \( s, t \in \mathbb{N}_0 \) with \( s < t \), consider the bisection

\[
\mathcal{P}_{X,T}^{s,t} = \{ \omega \in \Omega \mid X(T^s(\omega)) < X(T^t(\omega)) \} \cup \{ \omega \in \Omega \mid X(T^s(\omega)) \geq X(T^t(\omega)) \}
\]

of \( \Omega \) and, for observables \( X_1, X_2, \ldots, X_n \) on \( (\Omega, \mathcal{A}, \mu, T) \) and \( d, n \in \mathbb{N} \), the partition

\[
\mathcal{P}_{d}^{(X_i)_{i=1}^{n}} = \bigvee_{i=1}^{n} \bigvee_{0 \leq s < t \leq d} \mathcal{P}_{X_i,T}^{s,t},
\]

i.e. the coarsest partition refining all bisections \( \mathcal{P}_{X_i,T}^{s,t}; i = 1, 2, \ldots, n \). (If one of the sets of the right hand side of (3) is empty, \( \mathcal{P}_{X_i,T}^{s,t} \) is considered to consist of only one set.)

The partition \( \mathcal{P}_{d}^{(X_i)_{i=1}^{n}} \) is called \textit{ordinal partition of order} \( d \) \textit{associated to} \( (X_i)_{i=1}^{n} \). By definition its parts contain all states with equal ordinal measurement structure for an initial orbit part.

**A central statement.** Clearly, in order to preserve information of the given system, the observables should separate orbits of the system in a certain sense. In order to give a precise description, let in the following \( \sigma((X \circ T^t)_{t \in \mathbb{N}_0}) \) be the \( \sigma \)-algebra generated by all random variables \( X_i \circ T^t; i \in \mathbb{N}, t \in \mathbb{N}_0 \) and write \( \mathcal{F} \supseteq \mathcal{G} \) if for each \( G \in \mathcal{G} \) there exists some \( F \in \mathcal{F} \) with \( \mu(F \Delta G) = 0 \).

The following generalization of a statement in Antoniouk et al. [6] says that if there is no information loss by measuring with observables, all information is preserved also by only considering measurements from the ordinal viewpoint.
Theorem 1.1. Let \((\Omega, \mathcal{A}, \mu, T)\) be a measure-preserving dynamical system and \(X = (X_i)_{i \in \mathbb{N}}\) be a sequence of observables such that \(\sigma((X \circ T^t)_{t \in \mathbb{N}}) \supset \mathcal{A}\). Assume that (1) or (2) holds. Then

\[
    h_{KS}^{\mu}(T) = \lim_{d,n \to \infty} h_{\mu}(T, \mathcal{P}_d^{(X_i)_{i=1}^n}T) = \sup_{d,n \in \mathbb{N}} h_{\mu}(T, \mathcal{P}_d^{(X_i)_{i=1}^n}T).
\]

When Bandt and Pompe [8] invented the permutation entropy, they considered one-dimensional systems with coincidence of states and measurements. This fits into the given general approach as follows: \(\Omega\) is a Borel subset of \(\mathbb{R}\) and only one observable is considered to be the identity map \(id\) from \(\Omega\) to \(\mathbb{R}\). In this situation the assumptions of Theorem 1.1 are satisfied and so it holds

\[
    h_{KS}^{\mu}(T) = \lim_{d \to \infty} h_{\mu}(T, \mathcal{P}_d^{id}T) = \sup_{d \in \mathbb{N}} h_{\mu}(T, \mathcal{P}_d^{id}T)
\]

(compare [15, 16]).

Structure of the paper. The paper is organized as follows. In Section 2 we provide a proof of Theorem 1.1 on the basis of Antoniouk et al. [6]. We, moreover, discuss this statement from different perspectives in Section 3 by presenting its modifications and variants. Section 4 is devoted to the concept of permutation entropy, in particular to the two different approaches to it given by Bandt et al. in [7] and Amigó et al. in [5], respectively, and to its relation to the Kolmogorov-Sinai entropy. The ordinal approach to dynamical systems opens new perspectives to the estimation of system complexity. Advantages and limitations of this approach are discussed in Section 5. The natural question of how many observables are necessary for satisfying the assumptions of Theorem 1.1 is in the focus of Section 6. The corresponding discussion is strongly related to Takens’ delay embedding and similar ideas (see Takens [23] and Sauer [22]).

2. Kolmogorov-Sinai entropy from the ordinal viewpoint

This section is devoted to the proof of Theorem 1.1.

Preliminaries. In the following we write \(\mathcal{F} \overset{\mu}{=} \mathcal{G}\) if \(\mathcal{F} \supseteq \mathcal{G}\) and \(\mathcal{F} \subsetneq \mathcal{G}\), and denote by \(1_A\) the indicator function of a subset \(A \subset \Omega\). Moreover \(\sigma(\Diamond)\) denotes the \(\sigma\)-algebra generated by a set \(\Diamond\) of subsets of \(\Omega\), by a sequence or double sequence \(\Diamond\) of sets of subsets of \(\Omega\), or by a random variable \(\Diamond\) on \(\Omega\).

Given two finite partitions \(\mathcal{C}, \mathcal{D} \subset \mathcal{A}\) of \(\Omega\), we write \(\mathcal{C} \prec \mathcal{D}\) if \(\mathcal{D}\) is finer than \(\mathcal{C}\) or, equivalently, if \(\mathcal{C}\) is coarser than \(\mathcal{D}\), that is, each element \(C \in \mathcal{C}\) is a finite union of some elements of \(\mathcal{D}\). Note that \(\prec\) on the set of finite partitions of \(\Omega\) contained in \(\mathcal{A}\) is a partial order.

The join \(\bigvee_{r=1}^m \mathcal{C}_r\) of \(m \in \mathbb{N}\) finite partitions \(\mathcal{C}_r = \{C_r^{(1)}, C_r^{(2)}, \ldots, C_r^{(|\mathcal{C}_r|)}\} \subset \mathcal{A}\) of \(\Omega\) with \(r = 1, 2, \ldots, m\) is the coarsest partition refining all \(\mathcal{C}_r\); \(r = 1, 2, \ldots, m\), i.e.

\[
    \bigvee_{r=1}^m \mathcal{C}_r = \big\{ \bigcap_{r=1}^m C_r^{(l_r)} \neq \emptyset \mid l_r \in \{1, 2, \ldots, |\mathcal{C}_r|\} \big\} \text{ for } r = 1, 2, \ldots, m.
\]
For an observable $Y$ on $(\Omega, \mathcal{A}, \mu, T)$ we consider the finite partitions
\[
P^Y_d := \bigvee_{0 \leq s < t \leq d} P_{s,t}^{Y,T} \quad \text{and} \quad \tilde{P}^Y_d := \bigvee_{0 < t \leq d} P_{0,t}^{Y,T}
\]
(compare (3)) for $d \in \mathbb{N}$ and the $\sigma$-algebras $\Sigma^{Y,T}$ and $\tilde{\Sigma}^{Y,T}$ generated from all $P^Y_d$ and $\tilde{P}^Y_d$; $d \in \mathbb{N}$, respectively.

Besides $P^{(X)_{i=1}}_d = \bigvee_{i=1}^n P^{X_i,T}_d$ (compare (4)), for $d, n \in \mathbb{N}$ we are interested in the finite partitions
\[
\tilde{P}^{(X)_{i=1}}_d := \bigvee_{i=1}^n \tilde{P}^{X_i,T}_d.
\]

Furthermore, we need the following $\sigma$-algebras associated to these partitions:
\[
\Sigma^{X,T} := \sigma\left(\left(P^{(X)_{i=1}}_{d,n}\right)_{d,n \in \mathbb{N}}\right) = \sigma\left(\left(\Sigma_i^{X,T}\right)_{i \in \mathbb{N}}\right)
\]
and
\[
\tilde{\Sigma}^{X,T} := \sigma\left(\left(\tilde{P}^{(X)_{i=1}}_{d,n}\right)_{d,n \in \mathbb{N}}\right) = \sigma\left(\left(\tilde{\Sigma}_i^{X,T}\right)_{i \in \mathbb{N}}\right).
\]

The proof. Although we consider dynamical systems equipped with infinitely many observables, we can follow closely the argumentation in the paper Antoniouk et al. [6]. So let us first recall or modify those statements of that paper used in our proof.

**Lemma 2.1.** [6, Lemma 3.2] Let $F : \mathbb{R} \to [0, 1]$ be the distribution function of an observable $X$, that is $F(a) = \mu(\{\omega \in \Omega \mid X(\omega) \leq a\})$ for all $a \in \mathbb{R}$. Then
\[
\sigma(F \circ X) \overset{\mu}{=} \sigma(X).
\]

**Lemma 2.2.** [6, Lemma 3.3] Let $T : \Omega \leftrightarrow$ be an ergodic map and let $I_d : \Omega \to \mathbb{R}$ be defined by $I_d(\omega) := \sum_{t=1}^d 1_{\{X(T^t(\omega)) \leq X(\omega)\}}$ for all $d \in \mathbb{N}$ and $\omega \in \Omega$. Then
\[
F(X(\omega)) = \lim_{d \to \infty} \frac{I_d(\omega)}{d} \quad \text{for a.e. } \omega \in \Omega.
\]

By very slight modifications we can extend [6, Corollary 3.4 and Corollary 3.5] to countably many observables:

**Corollary 2.3.** Let $T : \Omega \leftrightarrow$ be an ergodic map. Then
\[
\sigma(X) \overset{\mu}{\subset} \tilde{\Sigma}^{X,T} \subset \Sigma^{X,T}.
\]

**Proof.** Compare to [6, Corollary 3.4]. The $\sigma$-algebra $\tilde{\Sigma}^{X,T}$ is generated by the $\sigma$-algebras $\tilde{\Sigma}^{X_i,T} := \sigma((\bigvee_{0 < t \leq d} P_{t}^{X_i,T})_{d \in \mathbb{N}})$; $i \in \mathbb{N}$. Therefore by $\sigma(X_i) \overset{\mu}{\subset} \tilde{\Sigma}^{X_i,T}$ for all $i \in \mathbb{N}$ it follows the assumption. This is true since $\frac{I_d}{d} : \Omega \to [0, 1]$ is $\tilde{\Sigma}^{X,T}$-$\mathcal{B}([0,1])$-measurable for all $d \in \mathbb{N}$ and hence so is $F \circ X$ and $X$ by Lemma 2.1 and Lemma 2.2. The inclusion $\tilde{\Sigma}^{X,T} \subset \Sigma^{X,T}$ is given by construction (compare (4) and (6)). \qed
Corollary 2.4. Let $T : \Omega \leftrightarrow$ be an ergodic map. Then

$$\sigma((X \circ T^t)_{t \in \mathbb{N}_0})^\mu \subset \Sigma X^T.$$  

Proof. For fixed $n \in \mathbb{N}$, in [6, Proof of Corollary 3.5] it is shown that

$$\mathcal{P}_d^{X_i \circ T^t, T} < \mathcal{P}_{d+1}^{X_i, T} \text{ for all } d \in \mathbb{N} \text{ and } i = 1, 2, \ldots, n$$

implying

$$\Sigma^{X_i \circ T^t, T} \subset \Sigma^{X_i, T} \text{ for all } i = 1, 2, \ldots, n \text{ and } t \in \mathbb{N}_0.$$  

Moreover, Corollary 2.3 gives

$$\sigma(X \circ T^t) \overset{\mu}{\subset} \Sigma^{X \circ T^t, T} \text{ for all } t \in \mathbb{N}_0.$$  

Consequently, $\sigma((X \circ T^t)_{t \in \mathbb{N}_0})^\mu \subset \Sigma X^T$. \hfill \qed

Lemma 2.5. $(\mathcal{P}_d^{(X_i)_{i=1}^n, T})_{d,n \in \mathbb{N}}$ is an increasing sequence in $n$ for fixed $d$, and for fixed $n$ it is an increasing sequence in $d$.

In particular, $(\mathcal{P}_{d_j}^{(X_{i_j})_{i=1}^{n_j}, T})_{d_j,n_j \in \mathbb{N}}$ is an increasing sequence in $j$ if $(d_j)_{j \in \mathbb{N}}$ and $(n_j)_{j \in \mathbb{N}}$ are increasing sequences in $\mathbb{N}$.

Proof. Given $d,n \in \mathbb{N}$, it holds

$$\mathcal{P}_d^{(X_i)_{i=1}^n, T} = \bigvee_{i=1}^n \bigvee_{0 \leq s < t \leq d} \mathcal{P}_{s,t}^{X_i, T},$$

$$\mathcal{P}_{d+1}^{(X_i)_{i=1}^n, T} = \bigvee_{i=1}^n \bigvee_{0 \leq s < t \leq d+1} \mathcal{P}_{s,t}^{X_i, T},$$

$$\mathcal{P}_d^{(X_i)_{i=1}^{n+1}, T} = \bigvee_{i=1}^{n+1} \bigvee_{0 \leq s < t \leq d} \mathcal{P}_{s,t}^{X_i, T},$$

implying $\mathcal{P}_d^{(X_i)_{i=1}^n, T} < \mathcal{P}_{d+1}^{(X_i)_{i=1}^n, T}, \mathcal{P}_d^{(X_i)_{i=1}^{n+1}, T}$ and so the above statements. \hfill \qed

For completing the proof of Theorem 1.1, we apply the following statement (see Walters [27, Theorem 4.22]):

Lemma 2.6. For a sequence $(\mathcal{C}_d)_{d \in \mathbb{N}}$ of finite partitions $\mathcal{C}_d \in \mathcal{A}$ of $\Omega$ increasing with respect to $\prec$ and satisfying $\sigma((\mathcal{C}_d)_{d \in \mathbb{N}})^{\mu} \subset \mathcal{A}$, it holds

$$h_{KS}^\mu(T) = \lim_{d \to \infty} h_\mu(T, \mathcal{C}_d).$$

First suppose that $T$ is an ergodic map. Then under the assumptions of Theorem 1.1 and by Corollary 2.4 it holds $\mathcal{A} \subset \sigma((X \circ T^t)_{t \in \mathbb{N}_0})^\mu \subset \Sigma X^T$. Since by Lemma 2.5 $(\mathcal{P}_{d_j}^{(X_{i_j})_{i=1}^{n_j}, T})_{d_j,n_j \in \mathbb{N}}$ is an increasing sequence in $j$ with respect to $\prec$ for increasing sequences $(d_j)_{j \in \mathbb{N}}$ and $(n_j)_{j \in \mathbb{N}}$ in $\mathbb{N}$, the assertion of Theorem 1.1 follows from Lemma 2.6.

In the non-ergodic case the ergodic decomposition theorem is consulted. For a thorough treatment we refer the reader to Einsiedler and Ward [10] and Einsiedler et al. [9].
ORDINAL ENTROPY DETERMINATION

In particular, the ergodic decomposition theorem claims that under certain conditions any \( T \)-invariant measure \( \mu \) can be decomposed into ergodic components and subsequently the entropy rate as well as the Kolmogorov-Sinai entropy of \( T \) with respect to \( \mu \) can be written as the integral of the entropies with respect to the decomposition.

In order to complete the proof of Theorem 1.1, we apply the following statement (see Einsiedler and Ward [10, Theorem 6.2], Einsiedler et al. [9, Theorem 5.27] and Keller and Sinn [15] for the case of a non-invertible \( T \)):

**Theorem 2.7.** Let \( (\Omega, \mathcal{A}, \mu, T) \) be a measure-preserving dynamical system satisfying (2). Then there exists a probability space \( (\Omega^*, \mathcal{A}^*, \nu) \) and a map \( \omega^* \mapsto \mu_{\omega^*} \) associating to each \( \omega^* \in \Omega^* \) a probability measure \( \mu_{\omega^*} \) on \( (\Omega, \mathcal{A}) \) such that the following is valid:

\( \Omega^* \) can be embedded into some compact metrizable space so that \( \mathcal{A} = \mathcal{B}(\Omega^*) \), the map \( \omega^* \in \Omega^* \rightarrow \int_{\Omega^*} f \, d\mu_{\omega^*} \) is \( \mathcal{A}^* \cdot \mathcal{B}((\mathbb{R})) \)-measurable for every essentially bounded measurable function \( f : \Omega \rightarrow \mathbb{R} \), the measure \( \mu_{\omega^*} \) is ergodic \( T \)-invariant for \( \nu \)-a.e. \( \omega^* \in \Omega^* \), and

\[
\mu = \int_{\Omega^*} \mu_{\omega^*} \, d\nu(\omega^*).
\]

Moreover, it holds

(10) \( h_{\mu}^{KS}(T) = \int_{\Omega^*} h_{\mu_{\omega^*}}^{KS}(T) \, d\nu(\omega^*) \)

and

(11) \( h_{\mu}(T, \mathcal{P}) = \int_{\Omega^*} h_{\mu_{\omega^*}}(T, \mathcal{P}) \, d\nu(\omega^*) \) for each finite partition \( \mathcal{P} \subset \mathcal{A} \) of \( \Omega \).

Altogether we obtain

\[
h_{\mu}^{KS}(T) \overset{(10)}{=} \int_{\Omega^*} h_{\mu_{\omega^*}}^{KS}(T) \, d\nu(\omega^*) \overset{\text{Theorem 1.1}}{=} \lim_{j \to \infty} \int_{\Omega^*} h_{\mu_{\omega^*}}(T, \mathcal{P}_{d_j}^{(X_i)_{i=1}^n ; T_j}) \, d\nu(\omega^*) \overset{\text{ergodic case}}{=} \lim_{j \to \infty} \int_{\Omega^*} h_{\mu_{\omega^*}}(T, \mathcal{P}_{d_j}^{(X_i)_{i=1}^n ; T_j}) \, d\nu(\omega^*) \overset{(11)}{=} \lim_{j \to \infty} h_{\mu}(T, \mathcal{P}_{d_j}^{(X_i)_{i=1}^n ; T_j}).
\]

Here \((n_j)_{j \in \mathbb{N}}\) and \((d_j)_{j \in \mathbb{N}}\) are strictly increasing sequences of natural numbers.

3. Modifications and consequences of Theorem 1.1.

We want to have a closer look at Theorem 1.1. For this recall that \( X \circ T^{tot} \) can be interpreted as a measurement of a system at time \( t \). As discussed in Section 1, there is no information loss when taking a pure ordinal viewpoint in the case that these measurements have ‘separating properties’.

**Less comparisons.** The main Theorem 1.1 can be given in a relaxed version if the considered observables provide a ‘separation’ from the outset (compare also [16, 17]). In order to determine the Kolmogorov-Sinai entropy, this means, in the case of ‘separating’
original observables, one does not need all comparisons between the elements of an orbit but only comparisons between points and their iterates.

**Theorem 3.1.** Let \((\Omega, \mathcal{A}, \mu, T)\) be a measure-preserving dynamical system and \(X = (X_i)_{i \in \mathbb{N}}\) be a sequence of observables such that \(\sigma(X) \supseteq \mathcal{A}\). Assume that (1) or (2) holds. Then

\[
h_{\mu}^{\text{KS}}(T) = \lim_{d,n \to \infty} h_{\mu}(T, \tilde{P}_d^{(X_i)_{i=1}^n, T}) = \sup_{d,n \in \mathbb{N}} h_{\mu}(T, \tilde{P}_d^{(X_i)_{i=1}^n, T}).
\]

For an ergodic map \(T\) we have that \(\mathcal{A} \supseteq \tilde{\Sigma}^{X,T}\), which follows from Corollary 2.3 and the assumption \(\sigma(X) \supseteq \mathcal{A}\). Moreover \((\tilde{P}_d^{(X_i)_{i=1}^n, T})_{d,n \in \mathbb{N}}\) is an increasing sequence in \(d\) and \(N\) with respect to \(\prec\), as it can be shown analogical to the proof of Lemma 2.5. Thus, for \(T\) ergodic the assertion follows by Lemma 2.6. To show the non-ergodic case one can use the ergodic decomposition theorem as in the proof of Theorem 1.1.

It seems that the assumption \(\sigma(X) \supseteq \mathcal{A}\) in Theorem 3.1 cannot be replaced by the assumption \(\sigma((X \circ T^t)_{t \in \mathbb{N}_0}) \supseteq \mathcal{A}\) in Theorem 1.1. At least, the argumentation of the proof of Corollary 2.4 cannot be adapted. Whereas

\[
(\tilde{\Sigma}^{X \circ T^t,T} \subset \tilde{\Sigma}^{X^T, T}\) for all \(t \in \mathbb{N}_0\)
\]

is true as (9) is, the analogue

\[
(\tilde{P}^{X \circ T^t,T}_d \prec \tilde{P}^{X^T,T}_{d+1}) \text{ for all } d \in \mathbb{N} \text{ and } i = 1, 2, \ldots, n
\]

of (7) is false. Therefore the analogue

\[
(\tilde{\Sigma}^{X \circ T^t,T} \subset \tilde{\Sigma}^{X^T, T}\) for all \(i = 1, 2, \ldots, n\) and \(t \in \mathbb{N}_0\)
\]

of (8) is not guaranteed. Let us give an example.

**Example 3.2.** Let \(\Omega = [0, 1]\) and \(T : \Omega \leftrightarrow \Omega\) be defined by

\[
T(\omega) = \frac{\omega}{2} \quad \text{for } \omega \leq \frac{1}{2} ,
\]

\[
T(\omega) = 2 - \frac{\omega}{2} \quad \text{else}
\]

\((T\) is the tent map preserving the equidistribution on \([0, 1]\).) Let

\[
Y = 2 \cdot 1_{[0, 1/3]} + 3 \cdot 1_{1/3, 2/3] + 1_{2/3, 1]},
\]

\(\omega_1 = 1\) and \(\omega_2 = \frac{5}{6}\). Then

\[
(Y(T^t(\omega_1)))_{t \in \mathbb{N}_0} = (1, 2, 2, 2, 2, 2, \ldots),
\]

\[
(Y(T^t(\omega_2)))_{t \in \mathbb{N}_0} = (1, 2, 3, 3, 3, 3, \ldots).
\]

It follows that \(\omega_1\) and \(\omega_2\) are separated by \(\mathcal{D}_{0,1}^{Y \circ T^t,T}\) and hence for all \(\tilde{\mathcal{P}}_{d}^{Y \circ T^t,T}; d \in \mathbb{N}\) but are not separated by \(\tilde{\mathcal{P}}_{d}^{Y^T,T}\) for all \(d \in \mathbb{N}\). Consequently, \(\tilde{\mathcal{P}}_{d}^{Y \circ T^t,T} \not\prec \tilde{\mathcal{P}}_{d+l}^{Y^T,T}\) for all \(d \in \mathbb{N}\) and \(l \in \mathbb{N}_0\).
Other partitions. For a single observable $X$ on a measure-preserving dynamical system $(\Omega, A, \mu, T)$ and $s, t \in \mathbb{N}_0$ with $s < t$, let

$$Q_{s,t}^{X,T} = \{ \omega \in \Omega \mid X(T^s) > X(T^t) \}, \quad \{ \omega \in \Omega \mid X(T^s) \leq X(T^t) \}$$

and

$$R_{s,t}^{X,T} = \{ \omega \in \Omega \mid X(T^s) < X(T^t) \}, \quad \{ \omega \in \Omega \mid X(T^s) > X(T^t) \}, \quad \{ \omega \in \Omega \mid X(T^s) = X(T^t) \}.$$

Further, for observables $X_1, X_2, \ldots, X_n$ on $(\Omega, A, \mu, T)$ and $d \in \mathbb{N}$, let

$$Q_d^{(X_i)_{i=1}^n,T} = \bigvee_{i=1}^n \bigvee_{0 \leq s < t \leq d} Q_{s,t}^{X_i,T}$$

and

$$R_d^{(X_i)_{i=1}^n,T} = \bigvee_{i=1}^n \bigvee_{0 \leq s < t \leq d} R_{s,t}^{X_i,T}.$$

(If one of the sets of the right hand side of (12) or (12) is empty, then it is not considered in order to have only nonempty sets.) Then the following is valid:

**Corollary 3.3.** The statement of Theorem 1.1 remains true when substituting $P_d^{(X_i)_{i=1}^n,T}$ by $Q_d^{(X_i)_{i=1}^n,T}$ or $R_d^{(X_i)_{i=1}^n,T}$.

**Proof.** Application of Theorem 1.1 to $-X = (-X_i)_{i \in \mathbb{N}}$ provides

$$h^{KS}_\mu(T) = \lim_{d,n \to \infty} h_\mu(T, P_d^{(-X_i)_{i=1}^n,T}) = h_\mu(T, Q_d^{(X_i)_{i=1}^n,T}).$$

Moreover, each $R_d^{(X_i)_{i=1}^n,T}$ is finer than $P_d^{(X_i)_{i=1}^n,T}$ implying

$$h_\mu(T, R_d^{(X_i)_{i=1}^n,T}) \geq h_\mu(T, P_d^{(X_i)_{i=1}^n,T}).$$

Therefore

$$h^{KS}_\mu(T) \geq \lim_{d,n \to \infty} h_\mu(T, R_d^{(X_i)_{i=1}^n,T}) \geq \lim_{d,n \to \infty} h_\mu(T, P_d^{(X_i)_{i=1}^n,T}) = 1.1 h^{KS}_\mu(T).$$

The existence of the limit

$$\lim_{d,n \to \infty} h_\mu(T, R_d^{(X_i)_{i=1}^n,T})$$

and its coincidence with the corresponding supremum is obvious (compare discussion for $P_d^{(X_i)_{i=1}^n,T}$ in Section 2).

Let us consider an order $\prec$ between observables $X, Y$ by $X \prec Y$ iff for all $\omega_1, \omega_2 \in \Omega$ the following holds (compare [3]):

$$Y(\omega_1) \leq Y(\omega_2) \text{ implies } X(\omega_1) \leq X(\omega_2).$$

One easily shows the following:
Lemma 3.4. For $X \prec Y$ it holds $\mathcal{R}_d^{X,T} \prec \mathcal{R}_d^{Y,T}$.

Note that for $X \prec Y$ not generally $\mathcal{P}_d^{X,T} \prec \mathcal{P}_d^{Y,T}$ and $\mathcal{Q}_d^{X,T} \prec \mathcal{Q}_d^{Y,T}$. After the following corollary being an immediate consequence of Theorem 1.1, we will illustrate this point by an example.

Corollary 3.5. Let $(\Omega, \mathcal{A}, \mu, T)$ be a measure-preserving dynamical system and $X = (X_i)_{i \in \mathbb{N}}$ be a sequence of observables with $X_1 \prec X_2 \prec X_3 \prec \ldots$ and $\sigma(\{X \circ T^t\}_{t \in \mathbb{N}}) \supseteq \mathcal{A}$. Assume that (1) or (2) holds. Then

$$h_{\mu}^{\text{KS}}(T) = \lim_{d, t \to \infty} h_{\mu}(T, \mathcal{R}_d^{X,t}) = \sup_{d, t \in \mathbb{N}} h_{\mu}(T, \mathcal{R}_d^{X,t}).$$

Example 3.6. See Example 3.2 and let

$$X = 2 \cdot 1_{[0, 5/8]} + 1_{[5/8, 1]}$$

and

$$Y = 4 \cdot 1_{[0, 1/8] \cup [3/8, 5/8]} + 3 \cdot 1_{1/8, 3/8} + 1_{[5/8, 1]}.$$n

Obviously, $X \prec Y$. Let $\omega_1 = \frac{1}{2}$ and $\omega_2 = \frac{3}{4}$. Then

$$(X(T^t(\omega_1)))_{t \in \mathbb{N}} = (2, 2, 1, 2, 2, 2, 2, 2, \ldots),$$

$$(X(T^t(\omega_2)))_{t \in \mathbb{N}} = (1, 2, 1, 2, 2, 2, 2, 2, \ldots),$$

$$(Y(T^t(\omega_1)))_{t \in \mathbb{N}} = (3, 4, 1, 4, 4, 4, 4, 4, \ldots),$$

and

$$(Y(T^t(\omega_2)))_{t \in \mathbb{N}} = (1, 4, 1, 4, 4, 4, 4, 4, \ldots).$$

From this, on one hand it follows that $\omega_1$ and $\omega_2$ are separated by $\mathcal{P}_{0,1}^{X,T}$, i.e. lie in different elements of $\mathcal{P}_{0,1}^{X,T}$, hence are separated by $\mathcal{P}_d^{X,T}$ for all $d \in \mathbb{N}$. On the other hand, this implies that $\omega_1$ and $\omega_2$ are not separated by $\mathcal{P}_d^{Y,T}$ for all $d \in \mathbb{N}$.

Therefore for no $d \in \mathbb{N}$ the partition $\mathcal{P}_d^{Y,T}$ is finer than $\mathcal{P}_d^{X,T}$. The similar is true for $\mathcal{Q}_d^{Y,T}$ and $\mathcal{Q}_d^{X,T}$, since $\mathcal{Q}_d^{Z,T} = \mathcal{P}_d^{Z,T}$ for an observable $Z$ on $(\Omega, \mathcal{A}, \mu, T)$.

Remark 3.7. Each finite partition $\mathcal{C} = \{C_1, C_2, \ldots, C_q\} \subset \mathcal{A}$; $q \in \mathbb{N}$ is generated by observables of the form $X = \sum_{l=1}^q \alpha_l \cdot 1_{C_l}$ in the sense that $C_l = X^{-1}(\alpha_l)$ for all $l = 1, 2, \ldots, q$, where $\alpha_l$; $l = 1, 2, \ldots, q$ are different real numbers. If a partition $\mathcal{D} \subset \mathcal{A}$ is finer than $\mathcal{C}$, than it can be written as

$$\mathcal{D} = \bigcup_{l=1}^q \{D_j^{(l)} \mid j = 1, 2, \ldots, m_l\}$$

with $m_1, m_2, \ldots, m_q \in \mathbb{N}$ and $C_l = \bigcup_{j=1}^{m_l} D_j^{(l)}$.

If $X = \sum_{l=1}^q \alpha_l \cdot 1_{C_l}$ for different $\alpha_l \in \mathbb{N}$ and if $m > m_l$ for all $l = 1, 2, \ldots, q$, then for

$$Y = \sum_{l=1}^q \sum_{j=1}^{m_l} (\alpha_l \cdot m + j) \cdot 1_{D_j^{(l)}}$$

it holds $X \prec Y$. This shows that an increasing sequence $(\mathcal{C}_d)_{d \in \mathbb{N}}$ can be ‘generated’ by a sequence $(X_d)_{d \in \mathbb{N}}$ of observables with $X_1 \prec X_2 \prec X_3 \prec \ldots$. 
Permutation entropy

The idea of considering dynamical systems from the ordinal viewpoint is strongly related to the invention of the permutation entropy, which we want to discuss now. We first give a definition of it in our general framework:

**Definition 4.1.** Given a sequence $X = (X_i)_{i \in \mathbb{N}}$ of observables on a measure-preserving dynamical system $(\Omega, \mathcal{A}, \mu, T)$, we define the permutation entropy $h_{\mu}^*(T, X)$ with respect to $X$ by

$$h_{\mu}^*(T, X) = \lim_{n \to \infty} \limsup_{d \to \infty} \frac{1}{d} H_{\mu}(T, \mathcal{P}_d^{(X_i)_{i=1}^n})$$

Originally, by Bandt et al. in [7] the definition of permutation entropy was given directly for one-dimensional systems. In our framework, this is $h_{\mu}^*(T, \text{id})$ with $T$ being an interval map.

**Permutation and Kolmogorov-Sinai entropy.** One reason for investigating the permutation entropy is its close relationship to the well-established Kolmogorov-Sinai entropy first observed by Bandt et al. in [7]. In their seminal paper they have shown that both entropies are coinciding for piecewise monotone interval maps $T$, i.e. for selfmaps $T$ on intervals splitting into finitely many subintervals on which $T$ is continuous and monotone.

Moreover, in the case that $\sigma((X \circ T^t)_{t \in \mathbb{N}_0}) \supseteq \mathcal{A}$ and that (1) or (2) holds, the Kolmogorov-Sinai entropy is not larger than permutation entropy. It holds for finitely many observables

$$\lim_{d \to \infty} h_{\mu}(T, \mathcal{P}_d^{(X_i)_{i=1}^n}) \leq \limsup_{d \to \infty} \frac{1}{d} H_{\mu}(T, \mathcal{P}_d^{(X_i)_{i=1}^n})$$

(see Keller et al. [18, Corollary 3]), hence the corresponding inequality for infinitely many ones follows by $n$ approaching to infinity. So let us summarize:

**Corollary 4.2.** Let $(\Omega, \mathcal{A}, \mu, T)$ be a measure-preserving dynamical system and $X = (X_i)_{i \in \mathbb{N}}$ be a sequence of observables such that $\sigma((X \circ T^t)_{t \in \mathbb{N}_0}) \supseteq \mathcal{A}$. Assume that (1) or (2) holds. Then

$$h_{\mu}^{KS}(T) \leq h_{\mu}^*(T, X).$$

The approach of Amigó et al. [3, 5]. This approach to permutation entropy different to the original is based on a refining sequence of finite partitions and is justified by the following statement due to Amigó et al. [3, 5]. We express the statement by finite-valued observables and refer here to Remark 3.7.

**Theorem 4.3.** For a measure-preserving dynamical system $(\Omega, \mathcal{A}, \mu, T)$ the following is valid:

(i) If $X$ is a finitely-valued observable, and $\mathcal{P}$ the finite partition generated by $X$, then

$$h_{\mu}(T, \mathcal{P}) = h_{\mu}^*(T, X).$$
(ii) If \((X_i)_{i \in \mathbb{N}}\) is a sequence of finitely-valued observables with \(X_1 \prec X_2 \prec X_3 \prec \ldots\) and the corresponding sequence of finite partitions generates \(\mathcal{A}\), then
\[
\lim_{i \to \infty} h^*_{\mu}(T, X_i).
\]

One immediately sees that by Lemma 2.6 assertion (ii) follows directly from statement (i). Amigó et al. took the right hand side of (15) as their modified concept of permutation entropy before showing its equality to Kolmogorov-Sinai entropy.

We want to finish this section by stating the following general problem, which is interesting on the different levels from the original one-dimensional definition of permutation entropy to the generalization for finitely or infinitely many observables.

**Problem.** Are the Kolmogorov-Sinai entropy and the permutation entropy coinciding and, if not, under which assumptions?

Note that the pure combinatorial part of the problem is relatively well understood (see Unakafova et al. [26], Keller et al. [18]).

5. Ordinal time series analysis

Ever since the idea of Bandt and Pompe [8] to consider the rank order of consecutive values of a time series instead of the values themselves, the ordinal approach attracts increasing attention and is applied in many scientific fields, for example in biomedical research, engineering and econophysics (see Amigó et al. [1, 2], Zanin et al. [28] and the references given there).

The reason is that the ordinal viewpoint brings with it many advantages especially for measuring complexity, such as robustness against small noise, simplicity of application and interpretation, and low computational costs. As mentioned, the determination of Kolmogorov-Sinai entropy is usually not easy, our discussion above, however, suggests that the ordinal approach can be used as a framework for estimating the Kolmogorov-Sinai entropy of dynamical systems and suchlike from real world data.

In the following we consider the theory developed in the previous sections in an applied context and discuss the pros and cons of using this approach in view of studying long and complex time series. A detailed exposition of this ordinal pattern approach is provided in Keller et al. [17].

**Ordinal patterns.** The task of gaining information about an underlying system via measurements is a common everyday problem. As already mentioned, this issue is increasingly addressed by using information lying in the ordinal structure of a system or a time series obtained from it. This leads to considering the up and downs in a time series, which can be described via so-called ordinal patterns.

**Definition 5.1.** For \(d \in \mathbb{N}\) denote the set of permutations of \(\{0, 1, \ldots, d\}\) by \(\Pi_d\). We say that a real vector \((x_s)_{s=0}^d\) has ordinal pattern \(\pi = (\pi_0, \pi_1, \ldots, \pi_d) \in \Pi_d\) of order \(d\) if
\[
x_{\pi_0} \geq x_{\pi_1} \geq \cdots \geq x_{\pi_{d-1}} \geq x_{\pi_d}
\]
and
\[
\pi_{u-1} > \pi_u \text{ if } x_{\pi_{u-1}} = x_{\pi_u} \text{ for any } u \in \{1, 2, \ldots, d\}.
\]
Given a time series \((x_t)_{t \in \mathbb{N}_0}\), the *ordinal pattern of order \(d\) at time \(t\) is defined as that of \((x_{t+s})_s=0^d\) and denoted by \(\pi_t\).

**Example 5.2.** In Figure 1 we consider a time series of 50 data points where exemplary the ordinal pattern \(\pi_{10} = (0, 5, 3, 4, 6, 1, 2) \in \Pi_6\) is emphasized, which corresponds to the order relation of the six successive values at \(t = 10\), that is

\[
x_t > x_{t+5} > x_{t+3} > x_{t+4} > x_{t+6} > x_{t+1} > x_{t+2}; \quad t = 10.
\]

![Figure 1](image_url)

**Figure 1.** Illustration of an ordinal pattern of order \(d = 6\) assigned to six successive values (plotted in the vertical direction) of a time series of 50 data points.

It is easily seen that, following the framework given in Section 1, two states \(\omega_1 \in \Omega\) and \(\omega_2 \in \Omega\) belong to the same part of some ordinal partition \(P_d^{(X_i)_{i=1}^n, T}\) iff the ordinal patterns of the vectors

\[(X_i(\omega_1), X_i(T(\omega_1)), \ldots, X(T^{\text{ord}}(\omega_2)))\quad \text{and} \quad (X_i(\omega_2), X_i(T(\omega_2)), \ldots, X(T^{\text{ord}}(\omega_2)))\]

coincide. Clearly, the other previous considered partitions (see Equations (6),(12) and (13)), despite some adjustments in terms of equality, can be coherent assimilated to this ordinal approach by redefining ordinal patterns in terms of the equality of values. The setting (16) is here in some sense arbitrary, however, the proposed definition of ordinal patterns has established itself. We will use it in the following to demonstrate how the previous covered theory provides interesting and promising tools for extracting the information saved in an ordinal pattern sequence or suchlike, for example, by estimating the permutation entropy (see Equation (14)) or by approximating the Kolmogorov-Sinai entropy.

In order to utilize ordinal patterns for the analysis of a system, sequential data \((x_t)_{t \in \mathbb{N}}\) obtained from a given measurement are transformed into a series \((\pi_t)_{t \in \mathbb{N}_0}\) of ordinal patterns. Distributions of ordinal patterns obtained from this approach are the central objects of exploration.

Note that ordinal patterns do not provide a symbolic representation as it is usually considered, since partitions of the state space are not given a priori, but are created on the basis of the given dynamics. However, the ordinal patterns as ‘symbols’ are very simple objects being directly obtained from the orbits of the system and containing intrinsic causal information. For the relationship of symbolic dynamics and representations and ordinal time series analysis see Amigó et al. [1].
For simplicity, we now restrict our exposition to the one-dimensional case with only one measurement. What we have in mind is a measure preserving dynamical system \((\Omega, \mathcal{A}, \mu, T)\), where \(\Omega\) is a Borel subset of \(\mathbb{R}\), acting as the model of a system, with a single observable \(X\) being the identity map. The extension of the ideas to the general case is obvious.

**Estimation of ordinal quantities.** The naive and mainly used estimator of ordinal pattern probabilities, so of the probability of the ordinal partition parts, is the relative frequency of ordinal patterns in an orbit of some length. For some \(t, d \in \mathbb{N}\), some ordinal pattern \(\pi\) of order \(d\) and some \(\omega \in \Omega\) the estimation is given by the number

\[
\hat{p}_\pi = \frac{1}{t - d + 1} \# \{s \in \{0, 1, \ldots, t - d\} \mid (X(T^{os}(\omega)), X(T^{os+1}(\omega)), \ldots, X(T^{os+d}(\omega))) \text{ has ordinal pattern } \pi \}.
\]

Here \(t + 1\) is the length of the considered orbit of \(\omega\). Clearly, the estimation only makes sense in the ergodic case. Then, by Birkhoff’s ergodic theorem, the corresponding estimator is consistent.

If in the ergodic case all \(\hat{p}_\pi; \pi \in \Pi_d\) are determined, it follows immediately that in the simple case considered a reasonable estimator for (14) is given by the empirical permutation entropy of order \(d \in \mathbb{N}\):

\[
\hat{h}^*_\mu(T, X) = -\frac{1}{d} \sum_{\pi \in \Pi_d} \hat{p}_\pi \ln \hat{p}_\pi.
\]

It gives furthermore also some information on the Kolmogorov-Sinai entropy.

**Assets and drawbacks.** Irrespective of the considered ordinal partition, the ordinal approach brings along some practical advantages and disadvantages. Note that most difficulties to overcome are common to any sort of time series analysis.

Considering the order relation between the values of a time series, small inaccuracies in measurements (e.g. errors between the state of a system and its observed value) are mostly negligible. Hence, the methods considered are relatively robust towards calibration differences of measuring instruments. Furthermore, the ordinal approach is easily interpretable and there already exist efficient methods to perform an ordinal time series analysis in real time. For a deeper discussion we refer to Riedl et al. [20] as well as Unakafova and Keller [25]. Last but not least, a foreknowledge of the data range when analyzing data is usually not necessary.

In contrast, the ordinal analysis of time series can be rather poor if the underlying system is so complex that such a large value \(d\) is needed that the computational capacity is insufficient. If, for example, the permutation entropy of a dynamical system is very large, its estimation by the empirical permutation entropy is problematic. Note that generally also for simple systems the convergency of empirical permutation entropies of order \(d\) to the permutation entropy can be rather slow, which is the reason for considering a conditional adaption of the permutation entropy (see Unakafov and Keller [24]).
In addition, the choice of a suitable order \( d \) with respect to the length of the original time series is affected by common problems. Large values of \( d \) are needed to evaluate encapsulated information as accurate as possible but a large \( d \) grants \( (d + 1)! \) possible ordinal patterns which have to be considered if nothing is known about the original time series. If one chooses an overlarge \( d \) relative to the length of a time series, it can happen that not all ordinal patterns which are substantial for describing the underlying dynamics are observed in the ordinal pattern distribution or suchlike. This is known as undersampling.

Moreover, ordinal time series analysis can lead to an arbitrary poor approximation of the Kolmogorov-Sinai entropy or poor representation of the underlying dynamics by the statistics, especially while working on wrong assumptions, e.g. a given system fails to be ergodic or the chosen observables cause information loss while measuring. The next section alludes to the latter problem.

6. Algebra reconstruction dimension

Theorems 1.1 claims that the Kolmogorov-Sinai entropy of \( T \) can be computed provided that we have sufficiently many observables “generating” \( \mathcal{A} \) up to \( \mu \)-measure zero. Essential for applications, the natural question arises how we can decrease the number of observables as much as possible. In this section we briefly review the known results in this direction.

Only one observable. The following example shows that theoretically in most real cases we can find only one such observable.

**Example 6.1.** Let \( I = [0, 1] \), \( Z \) be a separable complete metric space (such spaces are called Polish), \( \Omega \subset Z \) be its uncountable Borel subset, and \( \mathcal{A} := \mathcal{B}(\Omega) \) be the Borel \( \sigma \)-algebra of \( \Omega \). Then the pair \( (\Omega, \mathcal{B}(\Omega)) \) is called a standard Borel space. It is well known, e.g. see Kechris [13, Proposition 12.1], that then there exists a measurable isomorphism of \( (\Omega, \mathcal{B}(\Omega)) \) onto the space \( (I, \mathcal{B}(I)) \), that is a bijection \( X : \Omega \rightarrow I \) such that \( X^{-1}(\mathcal{B}(I)) = \mathcal{B}(\Omega) \).

Let \( \mu \) be a measure on \( (\Omega, \mathcal{B}(\Omega)) \) and \( T : \Omega \rightarrow \Omega \) be any \( \mu \)-preserving map. Then

\[
\mathcal{B}(\Omega) \supset \sigma((X \circ T^t)_{t \in \mathbb{N}_0}) \supset \sigma(X) = X^{-1}(\mathcal{B}(I)) = \mathcal{B}(\Omega),
\]

that is \( \sigma((X \circ T^t)_{t \in \mathbb{N}_0}) = \mathcal{B}(\Omega) \). Moreover, as every separable metric space \( Z \) can be embedded into a Hilbert cube being a compact space, compare to Hurewicz and Wallmann [11, Chapter V, §5, Theorem V4], we see that condition (2) holds for \( \Omega \subset Z \) as well, and therefore by Theorem 1.1 the Kolmogorov-Sinai entropy \( h^{KS}(T) \) of \( T \) can be computed via the formula (5).

Notice that the function \( X : \Omega \rightarrow [0, 1] \subset \mathbb{R} \) from Example 6.1 is not in general continuous and its explicit construction is very complicated. Therefore it is not useful for real applications. This leads to the following notion.

**Definition 6.2.** Let \( (\Omega, \mathcal{B}(\Omega)) \) be a standard Borel space with measure \( \mu \) on \( \mathcal{B}(\Omega) \), and \( T : \Omega \rightarrow \Omega \) be a \( \mathcal{B}(\Omega)-\mathcal{B}(\Omega) \)-measurable map. By the algebra reconstruction dimension of \( T \) with respect to \( \mu \) we will mean the minimal integer number \( n \geq 1 \) such that there
exists a continuous map $X : \Omega \to \mathbb{R}^n$ satisfying
\begin{equation}
\sigma((X \circ T^t)_{t \in \mathbb{N}}) \overset{\mu}{\supset} \mathcal{B}(\Omega).
\end{equation}
This number will be denoted by $\text{ard}_\mu(T)$. If such $n$ does not exist, then we will assume that $\text{ard}_\mu(T) = \infty$.

Thus $\text{ard}_\mu(T)$ is the minimal number of continuous observables needed to approximate the Kolmogorov-Sinai entropy via (5).

Given a map $T : \Omega \to \Omega$, a map $X : \Omega \to \mathbb{R}^n$ and $t \in \mathbb{N}$ one can define the following $t$-reconstruction map
\[ \Lambda_{X,T,t} = (X, X \circ T, \ldots, X \circ T^{t-1}) : \Omega \to \mathbb{R}^{nt} \]
and an $\infty$-reconstruction map
\[ \Lambda_{X,T,\infty} = (X, X \circ T, X \circ T^2, \ldots) : \Omega \to \mathbb{R}^\infty. \]
Evidently, $\Lambda_{X,T,1} = X$,
\[ \sigma((X \circ T^s)_{s=0}^{t-1}) = \sigma(\Lambda_{X,T,t}), \]
and
\[ \sigma(X) \subset \sigma(\Lambda_{X,T,t}) \subset \sigma(\Lambda_{X,T,t+1}) \subset \sigma(\Lambda_{X,T,\infty}); \ t \in \mathbb{N}. \]
In particular, (17) can be reformulated as follows:
\begin{equation}
\sigma(\Lambda_{X,T,\infty}) \overset{\mu}{\supset} \mathcal{B}(\Omega).
\end{equation}

Before discussing $\text{ard}_\mu(T)$ we will present an example for the existence of one separating observable, that is $X : \Omega \to \mathbb{R}$ satisfying (18), and therefore allowing to approximate the Kolmogorov-Sinai entropy by formula (5), see Theorem 6.5 below. However, now this observable is “discrete”, i.e. it takes at most countable many values.

**Definition 6.3.** Let $(\Omega, \mathcal{A}, \mu, T)$ be a measure-preserving dynamical system. An at most countable partition $\mathcal{C} = \{C_l\}_{l=1}^q \subset \mathcal{A}$ of $\Omega$ for some $q \in \mathbb{N} \cup \{\infty\}$, is called generating with respect to $T$, if
\[ \sigma((T^{-t}C)_{t \in \mathbb{N}_0}) = C, \]
where $T^{-t}C = \{(T^t)^{-1}C_l\}_{l=1}^q$.

The following lemma is evident.

**Lemma 6.4.** Suppose a measure-preserving dynamical system $(\Omega, \mathcal{A}, \mu, T)$ has a generating partition $\mathcal{C} = \{C_l\}_{l=1}^q$; $q \in \mathbb{N} \cup \{\infty\}$. Define a function $X : \Omega \to \mathbb{R}$ by $X = \sum_{l=1}^q l \cdot 1_{C_l}$ (compare Remark 3.7). Then $\sigma(X) = \sigma(\mathcal{C})$, whence
\[ \sigma(\Lambda_{X,T,\infty}) = \sigma((X \circ T^t)_{t \in \mathbb{N}_0}) = \sigma((T^{-t}C)_{t \in \mathbb{N}_0}) \overset{\mu}{=} \mathcal{A}. \]

In general, a $\mu$-preserving map does not have a generating partition. Nevertheless, for non-singular ergodic automorphisms of standard probability spaces such partitions do exist, what we discuss now. First we recall necessary definitions.

Let $(\Omega, \mathcal{A}, \mu)$ be a probability space. The measure $\mu$ is called complete if for any subset $A \in \mathcal{A}$ with $\mu(A) = 0$ every its subset $B$ also belongs to $\mathcal{A}$.

A countable family of sets $\{A_l\}_{l \in \mathbb{N}} \subset \mathcal{A}$ is called a complete basis of $(\Omega, \mathcal{A}, \mu)$ if
(a) for each \( A \in \mathcal{A} \) there exists a \( B \in \sigma(\{A_i\}_{i=1}^\infty) \) with \( A \subset B \) and \( \mu(B \setminus A) = 0 \);
(b) for any \( \omega_1, \omega_2 \in \Omega \) there exists an \( l \in \mathbb{N} \) such that \( \omega_1 \in A_l \) and \( \omega_2 \in \Omega \setminus A_l \);
(c) each intersection \( \bigcap_{l \in \mathbb{N}} B_l \), where every \( B_l \) is either \( A_l \) or \( \Omega \setminus A_l \), is non-empty.

A probability space \((\Omega, \mathcal{A}, \mu)\) is called standard if it has a complete basis and \( \mu \) is complete.

It has been proved by Rohlin [21] that every standard probability space with non-atomic measure is isomorphic with the probability space \((I, \mathcal{B}(I), \lambda)\), where \( \lambda \) is the Lebesgue measure on \( I \).

Recall also that a one-to-one transformation \( T : \Omega \rightarrow \Omega \) is non-singular with respect to a measure \( \mu \) if it is bi-measurable, i.e. \( T^{-1}\mathcal{A} = \mathcal{A} \) and \( T\mathcal{A} = \mathcal{A} \), and \( \mu(A) = 0 \) if and only if \( \mu(T(A)) = 0 \) for all \( A \in \mathcal{A} \).

The following theorem is a consequence of results by Rohlin [21], Parry [19] and Krieger [12] about the existence of countable and finite generating partitions of ergodic maps.

**Theorem 6.5.** [21, 19, 12] Let \((\Omega, \mathcal{B}(\Omega), \mu)\) be a standard probability space, and \( T : \Omega \rightarrow \Omega \) be a non-singular ergodic \( \mu \)-preserving map. Then \( \Omega \) has a countable generating partition with respect to \( T \). Hence there is a discrete measurable function \( X : \Omega \rightarrow \mathbb{R} \) taking at most countable distinct values and satisfying (18).

Moreover, if \( h_{KS}(T) < \infty \), then \( T \) admits a finite generating partition, and so \( X \) can be assumed to take only finitely many distinct values.

The continuous case. Notice that the function \( X \) from Theorem 6.5 is slightly better than the one from Example 6.1, as it takes a discrete set of values mutually distinct for distinct elements of the generating partition \( C \). Nevertheless, it is hard to construct as it requires to know a generating partition for \( T \), and so it is not useful for application as well.

Now we will consider the opposite situation when almost any continuous map \( X : \Omega \rightarrow \mathbb{R}^n \) satisfies (18).

**Lemma 6.6.** Let \( \Omega \) be a Polish space admitting an embedding \( X : \Omega \rightarrow \mathbb{R}^n \). Then for any measure \( \mu \) on \( \mathcal{B}(\Omega) \) and any \( \mu \)-preserving map \( T \), we have that \( \text{ard}_\mu(T) \leq n \). In particular, if \( \dim \Omega = k \); \( k \in \mathbb{N} \), then \( \text{ard}_\mu(T) \leq 2k + 1 \).

**Proof.** Since \( X \) is an embedding, we obtain that \( \sigma(X) = X^{-1}(\mathcal{B}(\mathbb{R}^n)) = \mathcal{B}(\Omega) \), whence \( \sigma(A_{X,T,\infty}) = \mathcal{B}(\Omega) \) as well.

The second statement follows from the well known fact that every \( k \)-dimensional separable metric space \( \Omega \) can be embedded into \( \mathbb{R}^{2k+1} \), [11, Chapter V, §4, Theorem V3]. Moreover, by the same theorem the set of embeddings \( \text{Emb}(\Omega, \mathbb{R}^{2k+1}) \) is residual (and, in particular, dense) in the space \( C(\Omega, \mathbb{R}^{2k+1}) \) of all continuous maps. Therefore almost every family of \( 2k + 1 \) continuous observables will allow to approximate the Kolmogorov-Sinai entropy of \( T \). \( \square \)

The next statement is a slight generalization of Theorem 2.2 from Keller [14].

**Theorem 6.7.** Let \( \Omega \) be a smooth manifold and \( \mathcal{D}(\Omega) \) be the group of its \( C^\infty \) diffeomorphisms. Then there exists a residual subset \( \mathcal{W} \) of \( \mathcal{D}(\Omega) \) such that \( \text{ard}_\mu(T) = 1 \) for each \( T \in \mathcal{W} \) and any measure \( \mu \) preserved by \( T \).
Proof. Let $\dim \Omega = k$. For each $n \in \mathbb{N}$ let
\[ \mathcal{E}_n = \{(X, T) \in C^\infty(\Omega, \mathbb{R}) \times \mathcal{D}(\Omega) \mid \Lambda_{X, T, n} : \Omega \to \mathbb{R}^n \text{ is an embedding}\}. \]

Thus if $(X, T) \in \mathcal{E}_n$, then $\text{ard}_\mu(T) = 1$.

It is proved by Takens [23] that if $n \geq 2k+1$, then $\mathcal{E}_n$ is residual (and in particular non-empty and everywhere dense) in $C^\infty(\Omega, \mathbb{R}) \times \mathcal{D}(\Omega)$. Thus we have that $\mathcal{E}_{2k+1} = \bigcap_{l=1}^\infty U_l$, where each $U_l$ is open and everywhere dense in the space $C^\infty(\Omega, \mathbb{R}) \times \mathcal{D}(\Omega)$. Let $p : C^\infty(\Omega, \mathbb{R}) \times \mathcal{D}(\Omega) \to \mathcal{D}(\Omega)$ be the natural projection, i.e. $p(X, T) = T$. It is a standard fact from general topology that $p$ is an open map, whence
\[ \mathcal{W} = p(\mathcal{E}_{2k+1}) = \bigcap_{l=1}^\infty p(U_l) \]
is a residual subset of $\mathcal{D}(\Omega)$. Then $\text{ard}_\mu(T) = 1$ for each $T \in \mathcal{W}$ and any measure $\mu$ preserved by $T$. \qed

Notice that the latter result does not guarantee that for any measure $\mu$ on $\mathcal{B}(\Omega)$ preserved by some diffeomorphism $T$ there exists some other $\mu$-preserving diffeomorphism $T'$ with $\text{ard}_\mu(T') = 1$.

The following notion allows to decrease the dimension $2k+1$ in Lemma 6.6 by putting some restrictions on $\mu$.

**Definition 6.8.** Let $X : \Omega \to R$ be a continuous map between topological spaces. Then the following subset of $\Omega$
\[ N_X = \{ \omega \in \Omega \mid X^{-1}(X(\omega)) \neq \{\omega\} \} \]
will be called the set of non-injectivity of $X$.

**Lemma 6.9.** (Antoniouk et al. [6, Theorem 4.2]) Let $X : \Omega \to R$ be a continuous map between Polish spaces and $\mu$ be a measure on $\mathcal{B}(\Omega)$. Suppose there exists a Borel subset $D$ such that $N_X \subset D$ and $\mu(D) = 0$. Then $\sigma(X) \equiv \mathcal{B}(\Omega)$.

Let $\Omega$ be a smooth manifold of dimension $k$. Say that a subset $Q \subset \Omega$ has Lebesgue measure zero, if for any local chart $\phi : \Omega \supset U \to \mathbb{R}^k$ in $\Omega$ the set $\phi(Q \cap U)$ has Lebesgue measure zero in $\mathbb{R}^k$. Notice that there is no natural definition of a set of fixed positive Lebesgue measure.

A measure $\mu$ on $\mathcal{B}(\Omega)$ will be said Lebesgue absolutely continuous if $\mu(Q) = 0$ for each subset $Q \subset \Omega$ of measure zero.

**Theorem 6.10.** (Antoniouk et al. [6, Theorem 2.13]) Let $\Omega$ be a smooth manifold of dimension $k$ and $\mu$ be a Lebesgue absolutely continuous measure on $\mathcal{B}(\Omega)$. For each $n \in \mathbb{N}$ let
\[ \mathcal{V}_n = \{ X \in C^\infty(\Omega, \mathbb{R}^n) \mid N_X \in \mathcal{B}(\Omega), \mu(N_X) = 0 \}. \]
If $n > k$, then $\mathcal{V}_n$ is residual in $C^\infty(\Omega, \mathbb{R}^n)$. Hence $\text{ard}_\mu(T) \leq k + 1$ for any (not necessarily continuous) $\mu$-preserving map $T : \Omega \to \Omega$. 

6.1. **Comparison of results.** It is convenient to compare these results in the following table, where it is assumed that $\Omega$ is a Polish space of dimension $k$.

| $\Omega$                  | $\mu$                      | $T$                           | $\text{ard}_\mu(T)$ | Statement |
|---------------------------|----------------------------|-------------------------------|----------------------|-----------|
| Borel space               | any measure                | any $\mu$-preserving measurable map | $\leq 2k + 1$ | Lemma 6.6 |
| Smooth manifold           | Lebesgue absolutely continuous | any $\mu$-preserving measurable map | $\leq k + 1$ | Theorem 6.10 |
| Smooth manifold           | any measure preserved by $T$ | generic diffeomorphism       | 1                    | Theorem 6.7 |

**REFERENCES**

[1] J. M. Amigó, K. Keller, V.A. Unakafova Ordinal symbolic analysis with application to biomedical recordings, to appear in Phil. Trans. Royal Soc. A.

[2] J. M. Amigó, K. Keller, J. Kurths (eds.) Recent progress in symbolic dynamics and permutation complexity. Ten years of permutation entropy, *The European Physical Journal Special Topics*, 222 (2013), 241–598.

[3] J. M. Amigó, The equality of Kolmogorov–Sinai entropy and metric permutation entropy generalized, *Physica D: Nonlinear Phenomena*, 241 (2012), 789–793.

[4] [978-3-642-04083-2] J. M. Amigó, *Permutation Complexity in Dynamical Systems: Ordinal Patterns, Permutation Entropy and All That*, Springer Series in Synergetics, Springer Complexity, Springer-Verlag, Berlin Heidelberg, 2010.

[5] J. M. Amigó, M. B. Kennel and L. Kocarev, The permutation entropy rate equals the metric entropy rate for ergodic information sources and ergodic dynamical systems, *Physica D: Nonlinear Phenomena*, 210 (2005), 77–95.

[6] A. Antoniouk, K. Keller and S. Maksymenko, Kolmogorov-Sinai entropy via separation properties of order-generated sigma-algebras, *Discrete and Continuous Dynamical Systems - A*, 34 (2014), 1793–1809.

[7] (MR1925429) [10.1088/0951-7715/15/5/312] C. Bandt, G. Keller and B. Pompe, Entropy of interval maps via permutations, *Nonlinearity*, 15 (2002), 1595–1602.

[8] C. Bandt and B. Pompe, Permutation entropy: a natural complexity measure for time series, *Physical Review Letters*, 88 (2002), 174102.

[9] M. Einsiedler, E. Lindenstrauss, T. Ward, Entropy in ergodic theory and homogeneous dynamics, unpublished, [access: July 02, 2014] Available from: http://maths.dur.ac.uk/~tpcc68/entropy/welcome.html.

[10] [978-1-4471-2591-4] M. Einsiedler and T. Ward, *Ergodic Theory With a View Towards Number Theory*, Graduate Texts in Mathematics, Springer-Verlag, New York, 2010.

[11] [9780691079479] W. Hurewicz and H. Wallman, *Dimension Theory*, Princeton Mathematical Series 4, Princeton University Press, Princeton, New Jersey, 1941.

[12] W. Krieger, On entropy and generators of measure-preserving transformations, *Transactions of the American Mathematical Society*, 149 (1970), 453–464.

[13] A. S. Kechris, *Classical descriptive set theory*. Graduate Texts in Mathematics, 156. Springer-Verlag, New York, 1995. 402 p.

[14] K. Keller, Permutations and the Kolmogorov-Sinai entropy, *Discrete and Continuous Dynamical Systems - A*, 32 (2012), 891–900.
[15] K. Keller and M. Sinn, Kolmogorov-Sinai entropy from the ordinal viewpoint, Physica D: Nonlinear Phenomena, 239 (2010), 997–1000.

[16] K. Keller and M. Sinn, A standardized approach to the Kolmogorov-Sinai entropy, Nonlinearity, 22 (2009), 2417–2422.

[17] K. Keller, M. Sinn and J. Emonds, Time series from the ordinal viewpoint, Stochastics and Dynamics, 2 (2007), 247–272.

[18] K. Keller, A. M. Unakafov, V. A. Unakafova, On the relation of KS entropy and permutation entropy, Physica D: Nonlinear Phenomena, 241 (2012), 1477–1481.

[19] W. Parry, Generators and strong generators in ergodic theory, Bulletin of the American Mathematical Society, 72 (1966), 294–296.

[20] M. Riedl, A. Müller and N. Wessel, Practical considerations of permutation entropy; a tutorial review, The European Physical Journal Special Topics, 222 (2013), 249–262.

[21] V. A. Rohlin, Exact endomorphisms of a Lebesgue space, Izvestiya Akademii Nauk SSSR Seriya Matematicheskaya, 25, (1961), 499–530, English translation: American Mathematical Society Translations, 39 (1964), 1–36.

[22] T. Sauer, J. A. Yorke, and M. Casdagli, Embeddology, Journal of Statistical Physics 65 (1991), 579–616.

[23] F. Takens, Detecting strange attractors in turbulence, in Dynamical Systems and Turbulence (eds. D. A. Rand and L. S. Young), Lecture Notes in Mathematics 898, Springer-Verlag, Berlin New York, (1981), 366–381.

[24] A. M. Unakafov and K. Keller, Conditional entropy of ordinal patterns, Physica D: Nonlinear Phenomena, 268 (2014), 94–102.

[25] V. A. Unakafova and K. Keller, Efficiently measuring complexity on the basis of real-world data, Entropy, 15 (2013), 4392–4415.

[26] V. A. Unakafova, A. M. Unakafov and K. Keller, An approach to comparing Kolmogorov-Sinai and permutation entropy, The European Physical Journal Special Topics, 222 (2013), 353–361.

[27] [10783040/0-387-95152-0] P. Walters, An Introduction to Ergodic Theory, Graduate Texts in Mathematics; 79, Springer-Verlag, New York, 2000.

[28] M. Zanin, L. Zunino, O. A. Rosso, and D. Papo, Permutation entropy and its main biomedical and econophysics applications: a review, Entropy, 14, (2012), 1553–1577.

Received xxxx 20xx; revised xxxx 20xx.

E-mail address: keller@math.uni-luebeck.de
E-mail address: maks@imath.kiev.ua
E-mail address: stolz@math.uni-luebeck.de