Ordinal Pattern Based Entropies and the Kolmogorov–Sinai Entropy: An Update

Tim Gutjahr * and Karsten Keller

Institute of Mathematics, University of Lübeck, D-23562 Lübeck, Germany; keller@math.uni-luebeck.de
* Correspondence: gutjahr@math.uni-luebeck.de; Tel.: +49-451-3101-6050; Fax: +49-451-3101-6004

Abstract: Different authors have shown strong relationships between ordinal pattern based entropies and the Kolmogorov–Sinai entropy, including equality of the latter one and the permutation entropy, the whole picture is however far from being complete. This paper is updating the picture by summarizing some results and discussing some mainly combinatorial aspects behind the dependence of Kolmogorov–Sinai entropy from ordinal pattern distributions on a theoretical level. The paper is more than a review paper. A new statement concerning the conditional permutation entropy will be given as well as a new proof for the fact that the permutation entropy is an upper bound for the Kolmogorov–Sinai entropy. As a main result, general conditions for the permutation entropy being a lower bound for the Kolmogorov–Sinai entropy will be stated. Additionally, a previously introduced method to analyze the relationship between permutation and Kolmogorov–Sinai entropies as well as its limitations will be investigated.

Keywords: ordinal patterns; Kolmogorov–Sinai entropy; permutation entropy; conditional entropy

1. Introduction

The Kolmogorov–Sinai entropy is a central measure for quantifying the complexity of a measure-preserving dynamical system. Although it is easy from the conceptional viewpoint, its determination and its estimation from given data can be challenging. Since Bandt, Keller, and Pompe showed the coincidence of Kolmogorov–Sinai entropy and permutation entropy for interval maps (see [1]), there have been different attempts to approach the Kolmogorov–Sinai entropy by ordinal pattern based entropies (see e.g., [2–6] and references therein), leading to a nice subject of study. In this paper, we want to discuss the relationship of the Kolmogorov–Sinai entropy to the latter kind of entropies. We respond to the state of the art and give some generalizations and new results, mainly emphasizing combinatorial aspects.

For this, let $(\Omega, \mathscr{A}, \mu, T)$ be a *measure-preserving dynamical system*, which we think to be *fixed in the whole paper*. Here, $(\Omega, \mathscr{A}, \mu)$ is a probability space equipped with a $\mathscr{A}$-$\mathscr{A}$- measurable map $T : \Omega \to \Omega$ satisfying $\mu(T^{-1}(A)) = \mu(A)$ for all $A \in \mathscr{A}$. Certain properties of the system will be specified at the places where they are of interest. It is suggested for the following to interpret Ω as the set of states of a system, μ as their distribution, and T as a description of the dynamics underlying the system and saying that the system is in state $T(\omega)$ at time $t + 1$ if it is in state $\omega \in \Omega$ at time t.

In the following, we give the definitions of the central entropies considered in this paper.

1.1. The Kolmogorov–Sinai Entropy

The base of quantifying dynamical complexity is to consider the development of partitions and their entropies under the given dynamics. Recall that the coarsest partitions refining given partitions $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_k$ and $\mathcal{P}, \mathcal{Q}$ of Ω are defined by

$$\bigvee_{s=1}^{k} \mathcal{P}_s := \left\{ \bigcap_{s=1}^{k} P_s \neq \emptyset \mid P_s \in \mathcal{P}_s \text{ for } s = 1, 2, \ldots, k \right\}$$

and

$$\mathcal{P} \vee \mathcal{Q} := \{ P \cap Q \neq \emptyset \mid P \in \mathcal{P}, Q \in \mathcal{Q} \},$$

respectively. The entropy of a finite or countably infinite partition $\mathcal{Q} \subset \mathscr{A}$ of Ω is given by

$$H(\mathcal{Q}) := - \sum_{Q \in \mathcal{Q}} \mu(Q) \log \mu(Q).$$

For a finite or countably infinite partition $\mathcal{P} := \{P_i\}_{i \in I} \subset \mathscr{A}$ of Ω and some $k \in \mathbb{N}$, consider the partition

$$\mathcal{P}^{(k)} := \bigvee_{t=0}^{k-1} T^{-t}(\mathcal{P}) = \{ P(i) \neq \emptyset \mid i \in I^k \},$$

where

$$P(i) := \bigcap_{t=0}^{k-1} T^{-t}(P_{i_t})$$

for each multiindex $i = (i_0, i_1, \ldots, i_{k-1}) \in I^k$. The *entropy rate* of T with regard to a finite or countably infinite partition $\mathcal{P} \subset \mathscr{A}$ with $H(\mathcal{P}) < \infty$ is defined by

$$h(\mathcal{P}) := \lim_{n \to \infty} \frac{1}{n} H(\mathcal{P}^{(n)}).$$

The *Kolmogorov–Sinai entropy* is then defined as

$$\mathrm{KS} := \sup_{\mathcal{P}} h(\mathcal{P}),$$

where the supremum is taken over all finite or over all countably finite partitions $\mathcal{P} \subset \mathscr{A}$ with $H(\mathcal{P}) < \infty$.

1.2. Ordinal Pattern Based Entropy Measures

As the determination and estimation of the Kolmogorov–Sinai entropy based on the given definition are often not easy, there are many different alternative approaches to it, among them the permutation entropy approach by Bandt and Pompe [7]. The latter is built up on the concept of ordinal patterns, which we describe in a general manner now.

For this, let $X = (X_1, X_2, \ldots, X_d) : \Omega \to \mathbb{R}^d$ be a random vector for $d \in \mathbb{N}$. Here, each of the random variables X_i can be interpreted as an observable measuring some quantity in the following sense: If the system is in state ω at time 0, then the arschvalue of the quantity mesured at time t provides $X_i(T^t(\omega))$. This general approach includes the one-dimensional case that states and measurements coincide, and this is that $\Omega \subseteq \mathbb{R}$ and $X = id$ is the identical map on Ω. This case, originally considered in [7] and subsequent papers, is discussed in Section 3. We refer to it as the *simple one-dimensional case*.

Let

$$\Pi_n := \{ (r_0, r_1, \ldots r_{n-1}) \in \{0, 1, \ldots n-1\}^n \mid r_i \neq r_j \text{ for } i \neq j \}$$

be the set of all permutations of length n. We say that a vector $(x_0, x_1, \ldots, x_{n-1}) \in \mathbb{R}^n$ has ordinal pattern $\pi = (r_0, r_1, \ldots r_{n-1}) \in \Pi_n$ if

$$x_{r_{i-1}} < x_{r_i} \quad \text{or} \quad x_{r_{i-1}} = x_{r_i} \text{ and } r_i < r_{i-1}$$

holds true for all $i \in \{1, 2 \ldots, n-1\}$. The $n!$ possible ordinal patterns (compare Figure 1) provide a classification of the vectors. We denote the set of points with ordinal pattern $\pi_1, \pi_2, \ldots, \pi_d \in \Pi_n$ with regard to $X_1, X_2, \ldots, X_d$, respectively, by

$$
\begin{aligned}
&P^{\mathbf{X}}_{\pi_1, \pi_2, \ldots, \pi_d} \\
&= \left\{ \omega \in \Omega \mid \left(X_i(\omega), X_i(T(\omega)), \ldots, X_i(T^{n-1}(\omega)) \right) \text{ has ordinal pattern } \pi_i \text{ for } i = 1, 2, \ldots, d \right\}
\end{aligned}
$$

and by

$$OP^{\mathbf{X}}(n) := \left\{ P^{\mathbf{X}}_{\pi_1, \pi_2, \ldots, \pi_d} :\neq \varnothing \mid \pi_1, \pi_2, \ldots, \pi_d \in \Pi_n \right\}$$

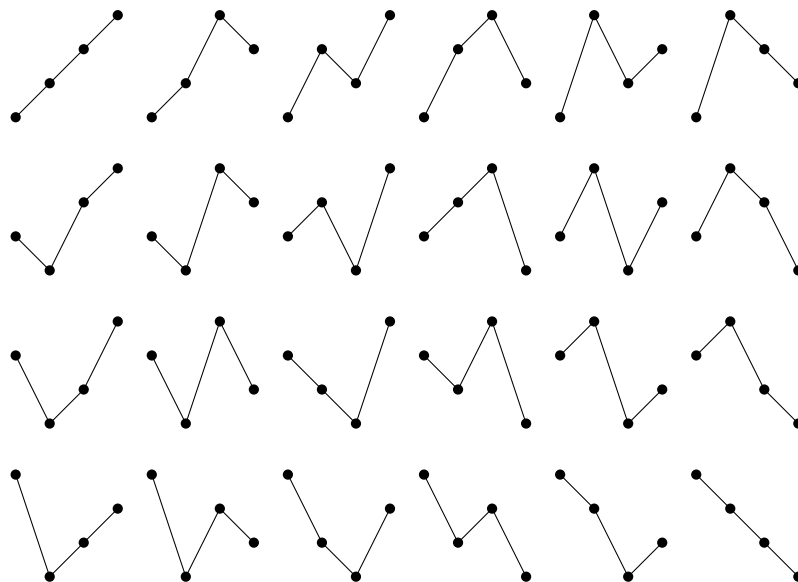the partition of Ω into those sets.

Figure 1. Abstract visualization of all 24 possible ordinal patterns of length 4.

We are especially interested in three ordinal pattern based entropy measures. These are the *lower* and *upper permutation entropies* defined as

$$\underline{\mathrm{PE}}^{\mathbf{X}} = \liminf_{n \to \infty} \frac{1}{n} H\left(OP^{\mathbf{X}}(n) \right)$$

and

$$\overline{\mathrm{PE}}^{\mathbf{X}} = \limsup_{n \to \infty} \frac{1}{n} H\left(OP^{\mathbf{X}}(n) \right),$$

respectively, and the *conditional entropy of ordinal patterns* defined by

$$\mathrm{CE}^{\mathbf{X}} = \liminf_{n \to \infty} \left[H\left(\left(OP^{\mathbf{X}}(n) \right)^{(2)} \right) - H\left(OP^{\mathbf{X}}(n) \right) \right].$$

We speak of the *permutation entropy* if the upper and lower permutation entropies coincide.

1.3. Outline of This Paper

In Section 2, we will focus on the relationship between permutation and Kolmogorov–Sinai entropies in the general setting. With Theorems 1 and 3, we will restate two known statements. A new proof of Theorem 1 will be given in Appendix A.2. Theorem 3 is stated for completeness. Theorem 2 establishes a new relationship between the conditional permutation entropy and the Kolmogorov–Sinai entropy.

In Section 3, the relationship between permutation and Kolmogorov–Sinai entropies in the one-dimensional case is investigated. Conditions are introduced, under which the permutation entropy is equal to the Kolmogorov–Sinai entropy. The given conditions allow for a generalization of previous results. We will explain why (countably) piecewise monotone functions satisfy these conditions and consider two examples.

In Section 4, we will investigate a method to analyze the relationship between permutation and Kolmogorov–Sinai entropies that was first introduced in [5]. We will use this method to relate two different kinds of conditional permutation entropies in the general setting. Theorem 5 shows that this method cannot be used directly to prove equality between permutation and Kolmogorov–Sinai entropies.

The results of the paper are summarized in Section 5. The proofs for all new results can be found in the Appendix A.

2. Relating Entropies

2.1. Partitions via Ordinal Patterns

Given some $d, n \in \mathbb{N}$ and some random vector $X = (X_1, X_2, \ldots, X_d)$, the partition described above can be defined in an alternative way, which is a bit more abstract but better fitting for the approach used in the proof of Theorem 4:

We can determine to which set $P_\pi \in OP^{X_i}(n)$ a point $\omega \in \Omega$ belongs to if we know whether $X_i(T^s(\omega)) < X_i(T^t(\omega))$ holds true for all $s, t \in \{0, 1, \ldots, n-1\}$ with $s < t$. Therefore, we can write

$$OP^{X_i}(n) = \bigvee_{s=0}^{n-1} \bigvee_{t=s+1}^{n-1} \left\{ \{\omega \in \Omega \mid X_i(T^s(\omega)) < X_i(T^t(\omega))\}, \{\omega \in \Omega \mid X_i(T^s(\omega)) \geq X_i(T^t(\omega))\} \right\}.$$

Throughout this paper, we will use the set

$$R := \{(x, y) \in \mathbb{R}^2 \mid x < y\}$$

to describe the order relation between two points. This notation allows us to write

$$OP^{X}(n) = \bigvee_{i=1}^{d} \bigvee_{s=0}^{n-1} \bigvee_{t=s+1}^{n-1} (T^s, T^t)^{-1} \left((X_i \times X_i)^{-1} \left(\{R, \mathbb{R}^2 \setminus R\} \right) \right).$$

2.2. Ordinal Characterization of the Kolmogorov–Sinai Entropy

To be able to reconstruct all information of the given system via quantities based on the random vector $X = (X_1, X_2, \ldots, X_d) \to \mathbb{R}^d$, we need to assume that the latter itself does not reduce information. From the mathematical viewpoint, this means that the σ-algebra generated by X is equivalent to the originally given σ-algebra $\mathscr{A}$, i.e., that

$$\sigma\left(\{X_i \circ T^t \mid t \in \mathbb{N}_0, i \in \{1, 2, \ldots, d\}\} \right) \overset{\mu}{=} \mathscr{A} \tag{1}$$

holds true, which is roughly speaking that orbits are separated by the given random vector. For definitions and some more details concerning σ-algebras and partitions, see Appendix A.1.

The following statement saying that, under (1), ordinal patterns entailing the complete information of the given system have been shown in [3] in a s slightly weaker form than given here.

Theorem 1. *Let* $X : \Omega \to \mathbb{R}^d$ *be a random vector satisfying* (1). *Then,*

$$\underline{\mathrm{PE}}^X \geq \lim_{k \to \infty} h\left(OP^X(k)\right) = \mathrm{KS} \tag{2}$$

holds true.

Note that the inequality in (2) is a relatively simple fact: Since the partition $OP^X(n)$ is finer than the partition $(OP^X(k))^{(n-k)}$ for all $n \geq k$, we have

$$H\left(OP^X(n)\right) \geq H\left((OP^X(k))^{(n-k)}\right).$$

Dividing both sides by n and taking n and subsequently k to infinity proves this inequality.

Proofs of the inequality $\underline{\mathrm{PE}}^X \geq \mathrm{KS}$ are also implicitly given in [1,8]. One-dimensional systems with direct observation as considered there are discussed in Section 3 in detail.

We will give a proof of the equality in (2) in Appendix A.2 being alternative to that in [3].

2.3. Conditional Entropies

In the case that (1) holds and that KS and $\underline{\mathrm{PE}}^X$ coincide, in Appendix A.3, we will prove different representations of the Kolmogorov–Sinai entropy by ordinal pattern based conditional entropies as they are given in the following theorem.

Theorem 2. *Let* $X : \Omega \to \mathbb{R}^d$ *be a random vector satisfying* (1). *If* $\mathrm{KS} \geq \overline{\mathrm{PE}}^X$ *is true, then*

$$\begin{aligned}
\mathrm{KS} &= \liminf_{n \to \infty} H\left(OP^X(n) \mid T^{-1}\left((OP^X(n))^{(k)}\right)\right) \\
&= \liminf_{n \to \infty} H\left((OP^X(n+1))^{(k)} \mid (OP^X(n))^{(k)}\right) \\
&= \underline{\mathrm{PE}}^X = \overline{\mathrm{PE}}^X
\end{aligned}$$

holds true for all $k \in \mathbb{N}$, *in particular, in the case* $k = 1$, *one has* $\mathrm{KS} = \mathrm{CE}^X = \underline{\mathrm{PE}}^X = \overline{\mathrm{PE}}^X$.

2.4. Amigó'S Approach

Amigó et al. [2,8] describe an alternative ordinal way to the Kolmogorov–Sinai entropy, which is based on a refining sequence of finite partitions. We present it in a slightly more general manner as originally given and in the language of finite-valued random variables. Note that the basic result behind Amigo's approach in [2,8] is that the Kolmogorov–Sinai entropy of a finite alphabet source and its permutation entropy given some order on the alphabet coincide (see also [9] for an alternative algebraic proof of the statement).

Theorem 3. *Given a sequence* $(X_k)_{k \in \mathbb{N}}$ *of* $\mathbb{R}$-*valued random variables satisfying*

(i) $\#(X_k(\Omega)) < \infty$ *for all* $k \in \mathbb{N}$,
(ii) $\sigma(X_k) \subseteq \sigma(X_{k+1})$ *for all* $k \in \mathbb{N}$,
(iii) $\sigma(\{X_k \mid k \in \mathbb{N}\}) \overset{\mu}{=} \mathscr{A}$,

then it holds

$$\lim_{k \to \infty} \underline{\mathrm{PE}}^{X_k} = \mathrm{KS}.$$

3. The Simple One-Dimensional Case

In the following, we consider the case that Ω is a subset of $\mathbb{R}$ with $\mathscr{A}$ coinciding with the Borel σ-algebra $\mathscr{B}$ on Ω, and with $X = id$ being the identical map on Ω. The X is superfluous here, which is why we leave out each superscript X. For example, we write $OP(n)$ instead of $OP^{id}(n)$.

3.1. (Countably) Piecewise Monotone Maps

We discuss some generalization of the results of Bandt, Keller, and Pompe that Kolmogorov–Sinai entropy and permutation entropy coincide for interval maps (see [1]) on the basis of a statement given in the paper [10]. The discussion sheds some light on structural aspects of the proofs given in that paper with some potential for further generalizations.

Definition 1. *Let Ω be a subset of $\mathbb{R}$ and $\mathscr{B}$ be the Borel σ-algebra on Ω and μ be a probability measure on $(\Omega, \mathscr{B})$. Then, we call a partition $\mathcal{M} = \{M_i\}_{i \in I}$ of Ω ordered (with regard to μ), if $\mathcal{M} \subset \mathscr{B}$ and*

$$\mu^2 \left((M_{i_1} \times M_{i_2}) \cap R \right) \in \left\{ 0, \mu^2(M_{i_1} \times M_{i_2}) \right\} \tag{3}$$

holds true for all $i_1, i_2 \in I$ with $i_1 \neq i_2$. Here, μ^2 denotes the product measure of μ with itself.

We call a map $T : \Omega \to \Omega$ (countably) piecewise monotone (with regard to μ) if there exists a finite (or countably infinite) ordered partition $\mathcal{M} = \{M_i\}_{i \in I}$ of Ω with $H(\mathcal{M}) < \infty$ such that

$$\mu^2 \left((M_i \times M_i) \cap R \cap (T \times T)^{-1}(R) \right) \in \left\{ 0, \mu^2 \left((M_i \times M_i) \cap R \right) \right\} \tag{4}$$

holds true for all $i \in I$.

Given a probability space $(\Omega, \mathscr{A}, \mu)$, for two families of sets $\mathcal{P}, \mathcal{Q} \subseteq \mathscr{A}$, we write

$$\mathcal{P} \prec \mathcal{Q}$$

if, for all $Q \in \mathcal{Q}$, there exists a $P \in \mathcal{P}$ with $\mu(Q \setminus P) = 0$. If $\mathcal{P} = \{P_i\}_{i \in I}$ and $\mathcal{Q} = \{Q_j\}_{j \in J}$ are partitions of Ω in $\mathscr{A}$, $\mathcal{P} \prec \mathcal{Q}$ is equivalent to the fact that for every $i \in I$ there exists a set $J_i \subseteq J$ such that P_i and $\bigcup_{j \in J_i} Q_j$ are equal up to some set with measure 0.

Moreover, given a partition $\mathcal{M} = \{M_i\}_{i \in I}$ of a set Ω, let

$$\mathcal{M}^{(m)} \otimes \mathcal{M}^{(m)} := \{M_{i_1} \times M_{i_2} \mid M_{i_1}, M_{i_2} \in \mathcal{M}^{(m)}\}.$$

In Appendix A.4, we will show the following statement:

Theorem 4. *Let Ω be a subset of $\mathbb{R}$ and $\mathscr{A} = \mathscr{B}$ be the Borel σ-algebra on Ω, and assume that the following conditions are satisfied:*

Condition 1: *There exists a finite or countably infinite ordered partition $\mathcal{M} = \{M_i\}_{i \in I} \subset \mathscr{B}$ with $H(\mathcal{M}) < \infty$ and some $m \in \mathbb{N}$ with*

$$\mathcal{M}^{(m)} \otimes \mathcal{M}^{(m)} \vee \{R, \Omega^2 \setminus R\} \prec \mathcal{M}^{(m)} \otimes \mathcal{M}^{(m)} \vee \bigvee_{u=1}^{m} (T \times T)^{-u} \left(\{R, \Omega^2 \setminus R\} \right). \tag{5}$$

Condition 2: *For all $\epsilon > 0$, there exists a finite or countably infinite ordered partition $\mathcal{Q}$ with $H(\mathcal{Q}) < \infty$ and*

$$\sum_{Q \in \mathcal{Q}} \limsup_{n \to \infty} \frac{1}{n} \sum_{l=1}^{n} \mu(Q \cap T^{-l}(Q)) < \epsilon. \tag{6}$$

Then,

$$\overline{PE} \leq KS$$

holds true.

Theorem 4 extracts the two central arguments in proving the main statement of [10] in the form of Conditions 1 and 2. This statement is given in a slightly stronger form in Corollary 1. In the proof of [10], the m in Condition 1 is equal to 1. We will discuss in Section 3.2 a situation where Condition 1 with $m = 2$ is of interest.

Corollary 1. *Let Ω be a compact subset of $\mathbb{R}$ and $\mathscr{A} = \mathscr{B}$ be the Borel σ-algebra on Ω. If T is (countably) piecewise monotone, then*

$$\overline{PE} \leq KS$$

holds true.

Since below we directly refer to the main statement in [10], which assumes compactness, and for simplicity, the Theorem is formulated under this assumption, we however will discuss a relaxation of the assumption in Remark A1.

To prove the above corollary, one needs to verify that Conditions 1 and 2 are satisfied for one-dimensional systems if T is piecewise monotone. It is easy to see that Condition 2 holds true for T being aperiodic and ergodic: If T is aperiodic, for any $\epsilon > 0$, one can choose a finite ordered partition $\mathcal{Q}$ such that $\mu(Q) < \epsilon$ holds true for all $Q \in \mathcal{Q}$. The ergodicity then implies

$$\sum_{Q \in \mathcal{Q}} \limsup_{n \to \infty} \frac{1}{n} \sum_{l=1}^{n} \mu(Q \cap T^{-l}(Q)) = \sum_{Q \in \mathcal{Q}} \mu(Q)^2 < \sum_{Q \in \mathcal{Q}} \mu(Q) \cdot \epsilon = \epsilon.$$

One can also show that Condition 2 is true for non-ergodic aperiodic compact systems (see Remark A1 and [10]).

If T is (countably) piecewise monotone, there exists a finite (or countable infinite) ordered partition $\mathcal{M} = \{M_i\}_{i \in I}$ with $H(\mathcal{M}) < \infty$ satisfying (4), which is equivalent to

$$\mu^2 \left(M_i \times M_i \cap R \cap (T \times T)^{-1}(R) \right) \in \left\{ 0, \mu^2 \left(M_i \times M_i \cap (T \times T)^{-1}(R) \right) \right\}$$

for all $i \in I$. Therefore,

$$\{M_i \times M_i\} \vee \{R, \Omega^2 \setminus R\} \vee (T \times T)^{-1} \left(\{R, \Omega^2 \setminus R\} \right) \stackrel{(4)}{\prec} \{M_i \times M_i\} \vee (T \times T)^{-1} \left(\{R, \Omega^2 \setminus R\} \right) \quad (7)$$

is true for all $i \in I$. Because $\mathcal{M}$ is an ordered partition, we have

$$\{M_{i_1} \times M_{i_2}\} \vee \{R, \Omega^2 \setminus R\} \stackrel{(3)}{\prec} \{M_{i_1} \times M_{i_2}\} \quad (8)$$

for all $i_1 \neq i_2 \in I$. This implies

$$\mathcal{M} \otimes \mathcal{M} \vee \{R, \Omega^2 \setminus R\}$$

$$= \left(\bigvee_{\substack{(i_1, i_2) \in I^2: \\ i_1 \neq i_2}} \{M_{i_1} \times M_{i_2}\} \vee \{R, \Omega^2 \setminus R\} \right) \vee \left(\bigvee_{i \in I} \{M_i \times M_i\} \vee \{R, \Omega^2 \setminus R\} \right)$$

$$\overset{(8)}{\prec} \bigvee_{\substack{(i_1, i_2) \in I^2: \\ i_1 \neq i_2}} \{M_{i_1} \times M_{i_2}\} \vee \left(\bigvee_{i \in I} \{M_i \times M_i\} \vee \{R, \Omega^2 \setminus R\} \right)$$

$$\prec \bigvee_{\substack{(i_1, i_2) \in I^2: \\ i_1 \neq i_2}} \{M_{i_1} \times M_{i_2}\} \vee \left(\bigvee_{i \in I} \{M_i \times M_i\} \vee \{R, \Omega^2 \setminus R\} \vee (T \times T)^{-1}(\{R, \Omega^2 \setminus R\}) \right)$$

$$\overset{(7)}{\prec} \bigvee_{\substack{(i_1, i_2) \in I^2: \\ i_1 \neq i_2}} \{M_{i_1} \times M_{i_2}\} \vee \left(\bigvee_{i \in I} \{M_i \times M_i\}(T \times T)^{-1}(\{R, \Omega^2 \setminus R\}) \right)$$

$$\prec \bigvee_{(i_1, i_2) \in I^2} \{M_{i_1} \times M_{i_2}\}(T \times T)^{-1}(\{R, \Omega^2 \setminus R\})$$

$$= \mathcal{M} \otimes \mathcal{M} \vee (T \times T)^{-1}(\{R, \Omega \setminus R\}).$$

Hence, Condition 1 holds true if T is (countably) piecewise monotone. To show that Corollary 1 holds true if the dynamical system is not aperiodic, one splits the system into a periodic part and an aperiodic part in the following way:

Let

$$\Theta := \bigcup_{t=1}^{\infty} \{\omega \in \Omega \mid T^t(\omega) = \omega\}$$

be the set of periodic points. Assume that $\mu(\Theta) \notin \{0, 1\}$ is true. Then,

$$\overline{PE} \leq \limsup_{n \to \infty} \frac{1}{n} H(OP(n) \vee \{\Theta, \Omega \setminus \Theta\})$$

$$= \limsup_{n \to \infty} \frac{1}{n} \left[H(OP(n) \vee \{\Theta, \Omega \setminus \Theta\}) - H(\Theta, \Omega \setminus \Theta) \right]$$

$$= \mu(\Theta) \cdot \limsup_{n \to \infty} \frac{1}{n} - \sum_{P_\pi \in OP(n)} \frac{\mu(P_\pi \cap \Theta)}{\mu(\Theta)} \log \left(\frac{\mu(P_\pi \cap \Theta)}{\mu(\Theta)} \right) \qquad (9)$$

$$+ \mu(\Omega \setminus \Theta) \cdot \limsup_{n \to \infty} \frac{1}{n} - \sum_{P_\pi \in OP(n)} \frac{\mu(P_\pi \setminus \Theta)}{\mu(\Omega \setminus \Theta)} \log \left(\frac{\mu(P_\pi \setminus \Theta)}{\mu(\Omega \setminus \Theta)} \right) \qquad (10)$$

holds true, where (9) is the periodic part of the upper permutation entropy and (10) the aperiodic part. One can use the aperiodic version of Corollary 1 to show that the Kolmogorov–Sinai entropy is an upper bound for (10). The proof of Corollary 1 for non-aperiodic dynamical systems is complete with Lemma A5 in Appendix A.4, which shows that (9) is equal to 0.
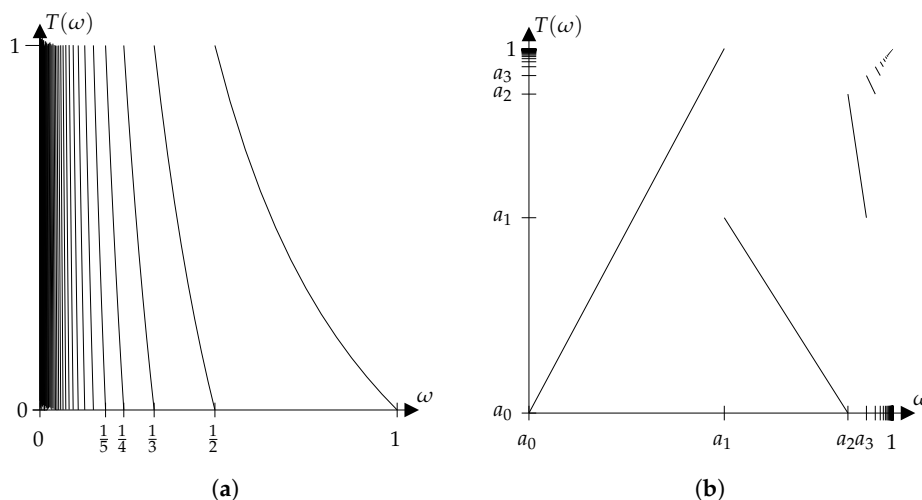
3.2. Examples

In order to illustrate the discussion in Section 3.1, we consider two examples. The first one reflects the situation in Corollary 1, and the second one discusses the case $m = 2$ in Condition 1 in Theorem 4.

Example 1 (Gaussian map). *The map $T : [0,1] \to [0,1]$ with*

$$T(\omega) = \begin{cases} 1/\omega \bmod 1, & \text{if } \omega > 0, \\ 0, & \text{if } \omega = 0 \end{cases}$$

is called a Gaussian map (see Figure 2a). This map is measure-preserving with regard to the measure μ, which is defined by $\mu(A) = \frac{1}{\log 2} \int_A \frac{1}{1+x} \, dx$ for all $A \in \mathscr{B}$ [11]. The partition $\mathcal{M} = \{[\frac{1}{n+1}, \frac{1}{n}[| \ n \in \mathbb{N}\} \cup \{\{0\}\}$ of $[0,1]$ is a countably infinite partition into monotony intervals of T satisfying $H(\mathcal{M}) < \infty$. This map is countably piecewise monotone and ergodic. Thus, its Kolmogorov–Sinai entropy is equal to its permutation entropy.

Figure 2. (**a**) Graph of the Gaussian map. (**b**) Graph of the map given in Example 2.

Example 2. *Consider $\Omega = [0,1[$ and the Borel σ-algebra $\mathscr{B}$ on Ω. Set*

$$S := \sum_{i=0}^{\infty} \frac{1}{(i+1)(\log(i+1))^2},$$

$$m_i := \frac{1}{S} \cdot \frac{1}{(i+1)(\log(i+1))^2} \qquad \text{for } i \in \mathbb{N},$$

$$a_0 := 0,$$

$$a_1 := m_1,$$

$$a_i := 1 - \frac{m_{i+1}}{m_1} \qquad \text{for } i \geq 2,$$

$$M_i := [a_{i-1}, a_i[\qquad \text{for } i \in \mathbb{N}.$$

The map $T : \Omega \to \Omega$ is defined as piecewise linear on each set M_i (see Figure 2b) by

$$T(\omega) = \begin{cases} \frac{\omega}{m_1}, & \text{if } \omega \in M_1, \\ a_{i-1} \cdot \frac{a_i - \omega}{a_i - a_{i-1}} + a_{i-2} \cdot \frac{\omega - a_{i-1}}{a_i - a_{i-1}}, & \text{if } \omega \in M_i \text{ with } i \in \mathbb{N} \setminus \{1\}. \end{cases}$$

Let λ be the one-dimensional Lebesgue measure. Define a measure μ on $(\Omega, \mathscr{B})$ by

$$\mu(A) := \sum_{j=1}^{\infty} \frac{m_j \cdot \lambda(A \cap M_j)}{\lambda(M_j)}$$

for all $A \in \mathscr{B}$.

One can verify that T is measure-preserving and ergodic with regard to μ. The partition $\mathcal{M} := \{M_i\}_{i \in \mathbb{N}}$ does satisfy (5) for m = 1, but $H(\mathcal{M}) = \infty$ holds true. Therefore, Condition 1 does not hold true for m = 1. However, one can show that Condition 1 holds true for m = 2 and the partition $\mathcal{M}' := \{M_1, \bigcup_{m=2}^{\infty} M_i\}$, which implies that the Kolmogorov–Sinai entropy is equal to the permutation entropy of this map due to Theorem 4.

4. A Supplementary Aspect

To determine under what conditions the Kolmogorov–Sinai entropy and the upper or lower permutation entropies coincide remains an open problem in the general case, and in the simple one-dimensional case of maps not being (countably) piecewise monotone the relation of Kolmogorov–Sinai and upper and lower permutation entropies is not completely understood. There is not even known an example where the entropies differ. Finally, we shortly want to discuss a further approach for discussing the relationship of Kolmogorov–Sinai entropy and upper and lower permutation entropies.

In [12], it was shown that under (1) the Kolmogorov–Sinai entropy is equal to the permutation entropy if roughly speaking the information contents of 'words' of k 'successive' ordinal patterns of large length n is not too far from the information contents of ordinal patterns of length $n + k - 1$. We want to explain this for the simple one-dimensional case and $k = 2$.

The ordinal pattern of some $(x_0, x_1, \ldots, x_n)$ contains all information on the order relation between the points $x_0, x_1, \ldots, x_n$. When considering the 'overlapping' ordinal patterns of $(x_0, x_1, \ldots, x_{n-1})$ and $(x_1, x_2, \ldots, x_n)$, one has the same information with one exception: The order relation between x_0 and x_n is not known a priori. Looking at the related partitions, the missing information is quantified by the conditional entropy $H(OP(n+1)|OP(n) \vee T^{-1}(OP(n)))$. There is one situation reducing this missing information, namely that one of the $x_i; i = 1, 2, \ldots, n-1$ lies between x_0 and x_n. Then, the order relation between x_0 and x_n is known by knowing the ordinal patterns of $(x_0, x_1, \ldots, x_{n-1})$ and $(x_1, x_2, \ldots, x_n)$. Therefore, the following set is of some special interest:

$$V_n := \{\omega \in \Omega \mid \omega \leq T^n(\omega) \text{ and } T^s(\omega) \notin (\omega, T^n(\omega)) \text{ for all } s \in \{1, 2, \ldots, n-1\}\}$$
$$\cup \{\omega \in \Omega \mid T^n(\omega) \leq \omega \text{ and } T^s(\omega) \notin (T^n(\omega), \omega) \text{ for all } s \in \{1, 2, \ldots, n-1\}\}. \tag{11}$$

The following is shown in Appendix A.5:

Lemma 1. *Let Ω be a subset of $\mathbb{R}$ and $\mathscr{A} = \mathscr{B}$ be the Borel σ-algebra on Ω. Then,*

$$H\left(OP(n+1)|OP(n) \vee T^{-1}(OP(n))\right) \leq \log(2) \cdot \mu(V_n) \tag{12}$$

holds true for all $n \in \mathbb{N}$.

This indicates that analyzing the measure of V_n as defined in (11) can be a useful approach to gain inside into the relationship between different kinds of entropies based on ordinal patterns. In particular, the behavior of $\mu(V_n)$ for $n \to \infty$ is of interest.

Lemma 2. *Let Ω be a subset of $\mathbb{R}$ and $\mathscr{A} = \mathscr{B}$ be the Borel σ-algebra on Ω. If T is ergodic, then*

$$\liminf_{n \to \infty} \mu(V_n) = 0$$

holds true, and, if (stronger) T is mixing, then

$$\lim_{n \to \infty} \mu(V_n) = 0$$

holds true.

The statement under the assumption of mixing has been shown in [5], and the proof in the ergodic case is given in Appendix A.5.

One can show that in the simple one-dimensional case the Kolmogorov–Sinai entropy is equal to the permutation entropy if

$$\sum_{n=1}^{\infty} H\left(OP(n+1)|OP(n) \vee T^{-1}(OP(n))\right) < \infty \tag{13}$$

holds true. Using (12), this is the case when $\sum_{n=1}^{\infty} \mu(V_n)$ is finite, providing a fast decay of the $\mu(V_n)$. However, we have $\sum_{n=1}^{\infty} \mu(V_n) = \infty$ as stated in Theorem 5, which will be proved in Appendix A.6.

Theorem 5. *Let Ω be a subset of $\mathbb{R}$, $\mathscr{A} = \mathscr{B}$ be the Borel σ-algebra on Ω and T be aperiodic and ergodic. Then,*

$$\sum_{n=1}^{\infty} \mu(V_n) = \infty$$

holds true.

Although formula $\sum_{n=1}^{\infty} \mu(V_n) < \infty$ is false, we cannot answer the question of whether or when (13) is valid. Possibly, an answer to this question, and a better understanding of the kind of decay of the $\mu(V_n)$, could be helpful in further investigating the relationship of Kolmogorov–Sinai entropy and upper and lower permutation entropies, at least in the simple one-dimensional ergodic case.

5. Conclusions

With Theorem 1, we have slightly generalized a statement given in [3] by removing a technical assumption and using more basic combinatorial arguments. The remaining assumption (1) on the random vector **X** cannot be weakened in general.

In Section 2.3, we have shown that the equality of the permutation entropy and the Kolmogorov–Sinai entropy implies the equality of conditional permutation entropy and Kolmogorov–Sinai entropies as well. We considered two different kinds of conditional permutation entropy, which have turned out to be equal in the cases considered in Section 2.3; it is however not clear whether these two kinds of conditional permutation entropy are equal in the general.

In Section 4, we have established some condition under which these two kinds of conditional entropy are equal, independently from of the equality between permutation and Kolmogorov–Sinai entropies. This condition is based on a concept introduced in [5] that was originally introduced as a tool for better understanding the relationship between permutation and Kolmogorov–Sinai entropies in a general setting. However, with Theorem 5, we have shown that this tool cannot directly be used to show the equality between permutation and Kolmogorov–Sinai entropies. It is an interesting question of whether and how a clever adaption and improvement of it can allow for new insights in the relationship between permutation and Kolmogorov–Sinai entropies.

In Section 3, we considered the simpler one-dimensional case. With Theorem 4, we have given two conditions under which the permutation entropy is a lower bound for the Kolmogorov–Sinai entropy. This theorem generalizes previous statements in [1] and slightly generalizes a statement in [10]. One of the conditions (Condition 2) holds true for a large class of dynamical systems, while, for the other one (Condition 1) to hold true, it is necessary that the system is in some sense 'order preserving'. It is still an unsolved and interesting question, whether Condition 1 can be weakened, especially since, to the best of our knowledge, there does not exist a counterexample to the equality of permutation entropy and Kolmogorov–Sinai entropies. Finding a generalization of Theorem 4 to a multidimensional setting is a further interesting question one could ask.

Appendix A. Proofs

Appendix A.1. Preliminaries

Given a probability space $(\Omega, \mathscr{A}, \mu)$ and two σ-algebras $\mathscr{A}_1, \mathscr{A}_2 \subseteq \mathscr{A}$, we write

$$\mathscr{A}_1 \overset{\mu}{\subseteq} \mathscr{A}_2$$

if for all $A_1 \in \mathscr{A}_1$ there exists $A_2 \in \mathscr{A}_2$ with $\mu(A_1 \setminus A_2) + \mu(A_2 \setminus A_1) = 0$. We write

$$\mathscr{A}_1 \overset{\mu}{=} \mathscr{A}_2$$

if both $\mathscr{A}_1 \overset{\mu}{\subseteq} \mathscr{A}_2$ and $\mathscr{A}_2 \overset{\mu}{\subseteq} \mathscr{A}_1$ hold true.

For a collection of $\mathbb{R}$-valued random variables $\{X_i\}_{i \in I}$ defined on some measure space $(\Omega, \mathscr{A})$, we denote by

$$\sigma\left(\{X_i \mid i \in I\}\right)$$

the smallest σ-algebra containing all sets in $\{X_i^{-1}(B) \mid i \in I, B \in \mathscr{B}\}$, where $\mathscr{B}$ is the Borel σ-algebra on $\mathbb{R}$.

Given a family of disjoint sets $\mathcal{P} \subseteq \mathscr{A}$ and some set $Q \in \mathscr{A}$, we define

$$\Delta(\mathcal{P}|Q) := \{P \in \mathcal{P} \mid \mu(P \cap Q) > 0\}$$

as those sets in $\mathcal{P}$ that are intersecting Q almost surely.

For two partitions $\mathcal{P}$ and $\mathcal{Q}$ of Ω in $\mathscr{A}$, the number of elements in $\Delta(\mathcal{P}|Q)$ can be used for an upper bound of the conditional entropy $H(\mathcal{P}|\mathcal{Q}) := H(\mathcal{P} \vee \mathcal{Q}) - H(\mathcal{Q})$:

Consider the function $f : [0,1] \to \mathbb{R}$ with $f(x) = x \log(x)$. Since f is convex, Jensen's inequality provides

$$\sum_{P \in \Delta(\mathcal{P}|Q)} \mu(P \cap Q) \log(\mu(P \cap Q))$$

$$= \#\Delta(\mathcal{P}|Q) \sum_{P \in \Delta(\mathcal{P}|Q)} \frac{1}{\#\Delta(\mathcal{P}|Q)} \cdot f(\mu(P \cap Q))$$

$$\geq \#\Delta(\mathcal{P}|Q) \cdot f\left(\sum_{P \in \Delta(\mathcal{P}|Q)} \frac{1}{\#\Delta(\mathcal{P}|Q)} \cdot \mu(P \cap Q)\right)$$

$$= \#\Delta(\mathcal{P}|Q) \cdot f\left(\frac{\mu(Q)}{\#\Delta(\mathcal{P}|Q)}\right)$$

$$= \#\Delta(\mathcal{P}|Q) \cdot \frac{\mu(Q)}{\#\Delta(\mathcal{P}|Q)} \cdot \log\left(\frac{\mu(Q)}{\#\Delta(\mathcal{P}|Q)}\right)$$

$$= \mu(Q) \cdot (\log(\mu(Q)) - \log(\#\Delta(\mathcal{P}|Q)))$$

for all $Q \in \mathcal{Q}$. Using the above inequality implies

$$
\begin{aligned}
H(\mathcal{P} \vee \mathcal{Q}) &= - \sum_{Q \in \mathcal{Q}} \sum_{P \in \mathcal{P}} \mu(P \cap Q) \log(\mu(P \cap Q)) \\
&= - \sum_{Q \in \mathcal{Q}} \sum_{P \in \cdot (\mathcal{P}|Q)} \mu(P \cap Q) \log(\mu(P \cap Q)) \\
&\leq - \sum_{Q \in \mathcal{Q}} \mu(Q) \cdot (\log(\mu(Q)) - \log(\#\Delta(\mathcal{P}|Q))) \\
&= H(\mathcal{Q}) + \sum_{Q \in \mathcal{Q}} \mu(Q) \cdot \log(\#\Delta(\mathcal{P}|Q)).
\end{aligned}
$$

This is equivalent to

$$
H(\mathcal{P}|\mathcal{Q}) \leq \sum_{Q \in \mathcal{Q}} \mu(Q) \cdot \log(\#\Delta(\mathcal{P}|Q)). \tag{A1}
$$

Appendix A.2. Proof of the Equality in Formula (2)

The proof is based on the following Lemma A1 and Corollary A1.

Lemma A1. *Let $X : \Omega \to \mathbb{R}$ be a random variable and $\mathcal{U}$ a finite ordered partition of $\mathbb{R}$ with regard to the image measure μ_X. Then, for $\mathcal{P} := X^{-1}(\mathcal{U})$, $n \in \mathbb{N}$ and all $P_\pi \in OP^X(n)$*

$$
\#\Delta(\mathcal{P}^{(n)}|P_\pi) \leq \binom{n + \#\mathcal{U} - 1}{\#\mathcal{U} - 1}
$$

holds true.

Proof. Set $I = \{1, 2, \ldots, \#\mathcal{U}\}$ and label the sets $U_i \in \mathcal{U}$ with $i \in I$ in such a way that

$$
i_1 < i_2 \Rightarrow \mu^2(\{(\omega_1, \omega_2) \in (X^{-1}(U_{i_1}) \times X^{-1}(U_{i_2})) : X(\omega_1) > X(\omega_2)\}) = 0 \tag{A2}
$$

holds true for all $i_1, i_2 \in I$. Since $\mathcal{U}$ is assumed to be an ordered partition, this is always possible. Set $P_i := X^{-1}(U_i)$ for all $i \in I$ so that $\mathcal{P} = \{P_i\}_{i \in I}$.

Fix $n \in \mathbb{N}$ and $P_\pi \in OP^X(n)$. Using

$$
P(\boldsymbol{i}) = \bigcap_{t=0}^{n-1} T^{-t}(P_{i_t})
$$

for all $\boldsymbol{i} = (i_0, i_1, \ldots i_{n-1}) \in I^n$, we have

$$
\#\Delta(\mathcal{P}^{(n)}|P_\pi) = \#\{\boldsymbol{i} \in I^n : \mu(P(\boldsymbol{i}) \cap P_\pi) > 0\}.
$$

There exists a permutation $\pi = (r_0, r_1, \ldots, r_{n-1})$ of $\{0, 1, \ldots, n-1\}$ such that

$$
X(T^{r_0}(\omega)) \leq X(T^{r_1}(\omega)) \leq \ldots \leq X(T^{r_{n-1}}(\omega))
$$

holds true for all $\omega \in P_\pi$. Using (A2), this implies

$$
i_{r_0} \leq i_{r_1} \leq \ldots \leq i_{r_{n-1}}
$$

for all $i = (i_0, i_1, \ldots i_{n-1}) \in I^n$ with $\mu(P(i) \cap P_\pi) > 0$. Therefore,

$$
\begin{aligned}
\#\Delta(\mathcal{P}^{(n)}|P_\pi) &= \#\{(i_0, i_1, \ldots i_{n-1}) \in I^n : \mu(P(i) \cap P_\pi) > 0\} \\
&\leq \#\{(i_0, i_1, \ldots i_{n-1}) \in I^n : i_{r_0} \leq i_{r_1} \leq \ldots \leq i_{r_{n-1}}\} \\
&= \binom{n + \#\mathcal{U} - 1}{\#\mathcal{U} - 1}
\end{aligned}
$$

holds true for all $P_\pi \in OP^X(n)$.　□

The lemma can be used to directly prove the following result.

Corollary A1. *Let* $X = (X_1, X_2, \ldots, X_d) : \Omega \to \mathbb{R}^d$ *be a random vector and* $\mathcal{U}$ *a finite partition of* $\mathbb{R}$ *into intervals. Then,*

$$
h\left(\bigvee_{i=1}^d X_i^{-1}(\mathcal{U})\right) \leq \lim_{k \to \infty} h\left(OP^X(k)\right)
$$

holds true.

Proof. Take $k, m \in \mathbb{N}$ and set $\mathcal{P}_i := X_i^{-1}(\mathcal{U})$. Then,

$$
\begin{aligned}
&H(\mathcal{P}_i^{(mk)}|(OP^{X_i}(k))^{(mk)}) \\
&\leq \sum_{t=0}^{m-1} H(T^{-kt}(\mathcal{P}_i^{(k)})|(OP^{X_i}(k))^{(mk)}) \\
&\leq \sum_{t=0}^{m-1} H(T^{-kt}(\mathcal{P}_i^{(k)})|T^{-kt}(OP^{X_i}(k))) \\
&= mH(\mathcal{P}_i^{(k)}|OP^{X_i}(k))
\end{aligned}
$$

holds true for all $i \in \{1, 2, \ldots, d\}$. Together with (A1) and Lemma A1, this provides

$$
\begin{aligned}
h\left(\bigvee_{i=1}^d \mathcal{P}_i\right) &= \lim_{n \to \infty} \frac{1}{n} H\left(\bigvee_{i=1}^d \mathcal{P}_i^{(n)}\right) = \lim_{m \to \infty} \frac{1}{mk} H\left(\bigvee_{i=1}^d \mathcal{P}_i^{(mk)}\right) \\
&\leq \lim_{m \to \infty} \frac{1}{mk} H\left(\bigvee_{i=1}^d \mathcal{P}_i^{(mk)} \vee (OP^X(k))^{(mk)}\right) \\
&= \lim_{m \to \infty} \frac{1}{mk}\left[H((OP^X(k))^{(mk)}) + H\left(\bigvee_{i=1}^d \mathcal{P}_i^{(mk)}|(OP^X(k))^{(mk)}\right)\right] \\
&\leq \lim_{m \to \infty} \frac{1}{mk}\left[H((OP^X(k))^{(mk)}) + \sum_{i=1}^d H\left(\mathcal{P}_i^{(mk)}|(OP^X(k))^{(mk)}\right)\right]
\end{aligned}
$$

$$\leq \lim_{m\to\infty} \frac{1}{mk}\left[H((OP^{\mathbf{X}}(k))^{(mk)}) + \sum_{i=1}^{d} H\left(\mathcal{P}_i^{(mk)}|(OP^{X_i}(k))^{(mk)}\right)\right]$$

$$\leq \lim_{m\to\infty}\left[\frac{1}{mk}H((OP^{\mathbf{X}}(k))^{(mk)}) + \sum_{i=1}^{d}\frac{1}{k}H(\mathcal{P}_i^{(k)}|OP^{X_i}(k))\right]$$

$$= h(OP^{\mathbf{X}}(k)) + \sum_{i=1}^{d}\frac{1}{k}H(\mathcal{P}_i^{(k)}|OP^{X_i}(k))$$

$$\leq h(OP^{\mathbf{X}}(k)) + \sum_{i=1}^{d}\frac{1}{k}\sum_{P_\pi\in OP^{X_i}(k)}\mu(P_\pi)\log(\#\Delta(\mathcal{P}_i^{(k)}|P_\pi))$$

$$\leq h(OP^{\mathbf{X}}(k)) + \sum_{i=1}^{d}\frac{1}{k}\sum_{P_\pi\in OP^{X_i}(k)}\mu(P_\pi)\log\left(\binom{k+\#\mathcal{U}-1}{\#\mathcal{U}-1}\right)$$

$$\leq h(OP(k)) + \frac{d}{k}\log\left((k+\#\mathcal{U}-1)^{\#\mathcal{U}-1}\right)$$

$$= h(OP(k)) + d(\#\mathcal{U}-1)\cdot\frac{\log(k+\#\mathcal{U}-1)}{k}$$

for all $k \in \mathbb{N}$, which implies

$$h(\mathcal{P}) \leq \lim_{k\to\infty}\left[h(T,OP(k)) + d(\#\mathcal{U}-1)\cdot\frac{\log(k+\#\mathcal{U}-1)}{k}\right] = \lim_{k\to\infty}h(OP(k)). \quad \square$$

We are now able to prove of the equality in (2). Let $p_i : \mathbb{R}^d \to \mathbb{R}$ with $p_i((x_1,x_2,\ldots,x_d)) = x_i$ be the projection on the i-th coordinate, and let $\mathscr{B}(\mathbb{R}^d)$ denote the Borel σ-algebra on $\mathbb{R}^d$. Since this σ-algebra is generated by sets of the type

$$I_1 \times I_2 \times \ldots I_d,$$

where I_i are intervals, there exists an increasing sequence of finite partition $(\mathcal{U}_l)_{l\in\mathbb{N}}$ of $\mathbb{R}$ into intervals, such that

$$\mathscr{B}(\mathbb{R}^d) = \sigma\left(\left\{p_i^{-1}(\mathcal{U}_l) \mid l \in \mathbb{N}, i \in \{1,2,\ldots,d\}\right\}\right)$$

holds true. Using (1), this implies

$$\mathscr{A} = \sigma\left(\left\{T^{-t}\left(\mathbf{X}^{-1}(p_i^{-1}(\mathcal{U}_l))\right) \mid t \in \mathbb{N}_0, l \in \mathbb{N}, i \in \{1,2,\ldots,d\}\right\}\right)$$

$$= \sigma\left(\left\{T^{-t}\left(X_i^{-1}(\mathcal{U}_l)\right) \mid t \in \mathbb{N}_0, l \in \mathbb{N}, i \in \{1,2,\ldots,d\}\right\}\right)$$

$$= \sigma\left(\left\{T^{-t}\left(\bigvee_{i=1}^{d}X_i^{-1}(\mathcal{U}_l)\right) \mid t \in \mathbb{N}_0, l \in \mathbb{N}\right\}\right).$$

Thus, $(P_l)_{l\in\mathbb{N}}$ with

$$\mathcal{P}_l := \bigvee_{i=1}^{d}X_i^{-1}(\mathcal{U}_l)$$

is a generating sequence of finite partitions, which implies (see e.g., [13])

$$\mathrm{KS} = \lim_{l\to\infty}h(\mathcal{P}_l).$$

Corollary A1 provides

$$h(\mathcal{P}_l) \leq \lim_{k\to\infty}h(OP^{\mathbf{X}}(k))$$

for all $l \in \mathbb{N}$. Combining the two previous statements yields

$$\text{KS} = \lim_{l \to \infty} h(\mathcal{P}_l) \leq \lim_{k \to \infty} h(OP^X(k)). \tag{A3}$$

On the other hand,

$$\text{KS} = \sup_{\mathcal{P}} h(\mathcal{P}) \geq \lim_{k \to \infty} h(OP^X(k))$$

holds true, which, together with (A3), finishes the proof of the equality in (2).

Appendix A.3. Proof of Theorem 2

For preparing the proof of Theorem 2, let us first give two lemmata.

Lemma A2. *Let $(\mathcal{P}_n)_{n \in \mathbb{N}}$ be a sequence of finite partitions of Ω in $\mathscr{A}$ satisfying*

$$\mathcal{P}_n \vee T^{-1}(\mathcal{P}_n) \prec \mathcal{P}_{n+1}. \tag{A4}$$

Then,

$$\liminf_{n \to \infty} H\left(\mathcal{P}_n \,\Big|\, T^{-1}\left(\mathcal{P}_n^{(k)}\right)\right) \leq \liminf_{n \to \infty} H\left(\mathcal{P}_{n+1}^{(k)} \,\Big|\, \mathcal{P}_n^{(k)}\right) \leq \liminf_{n \to \infty} \frac{1}{n} H(\mathcal{P}_n)$$

holds true for all $k \in \mathbb{N}$.

Proof. Take $k \in \mathbb{N}$. We have

$$\liminf_{n \to \infty} H\left(\mathcal{P}_n \,\Big|\, T^{-1}\left(\mathcal{P}_n^{(k)}\right)\right) = \liminf_{n \to \infty} \left[H\left(\mathcal{P}_n^{(k+1)}\right) - H\left(T^{-1}\left(\mathcal{P}_n^{(k)}\right)\right)\right]$$

$$= \liminf_{n \to \infty} \left[H\left(\mathcal{P}_n^{(k+1)}\right) - H\left(\mathcal{P}_n^{(k)}\right)\right] \overset{(A4)}{\leq} \liminf_{n \to \infty} \left[H\left(\mathcal{P}_{n+1}^{(k)}\right) - H\left(\mathcal{P}_n^{(k)}\right)\right] = \liminf_{n \to \infty} H\left(\mathcal{P}_{n+1}^{(k)} \,\Big|\, \mathcal{P}_n^{(k)}\right).$$

The Stolz–Cesàro theorem further provides

$$\liminf_{n \to \infty} H\left(\mathcal{P}_{n+1}^{(k)} \,\Big|\, \mathcal{P}_n^{(k)}\right) \leq \liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \left[H\left(\mathcal{P}_{i+1}^{(k)} \,\Big|\, \mathcal{P}_i^{(k)}\right)\right]$$

$$= \liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \left[H\left(\mathcal{P}_{i+1}^{(k)}\right) - H\left(\mathcal{P}_i^{(k)}\right)\right] = \liminf_{n \to \infty} \frac{1}{n} \left[H\left(\mathcal{P}_{n+1}^{(k)}\right) - H\left(\mathcal{P}_1^{(k)}\right)\right]$$

$$= \liminf_{n \to \infty} \frac{1}{n} H\left(\mathcal{P}_{n+1}^{(k)}\right) \overset{(A4)}{\leq} \liminf_{n \to \infty} \frac{1}{n} H(\mathcal{P}_{n+k}) = \liminf_{n \to \infty} \frac{1}{n} H(\mathcal{P}_n). \quad \square$$

Notice that (A4) is fulfilled for $\mathcal{P}_n := OP^X(n)$.

Lemma A3. *Let $X = (X_1, X_2, \ldots, X_d) : \Omega \to \mathbb{R}^d$ be a random vector satisfying (1). Then,*

$$\text{KS} \leq \liminf_{n \to \infty} H\left(OP^X(n) \,\Big|\, T^{-1}\left((OP^X(n))^{(k)}\right)\right)$$

holds true for all $k \in \mathbb{N}$.

Proof. According to Theorem 1, we have

$$\text{KS} = \lim_{n \to \infty} h(T, OP^X(n)).$$

Using the future formula for the entropy rate (see e.g., [13]), we can write

$$h(OP^X(n)) = \lim_{l \to \infty} H\left(OP^X(n) \,\Big|\, T^{-1}\left((OP^X(n))^{(l)}\right)\right)$$

for all $n \in \mathbb{N}$. This implies

$$
\begin{aligned}
\mathrm{KS} &= \lim_{n \to \infty} \lim_{l \to \infty} H\left(OP^X(n) \,\middle|\, T^{-1}\left((OP^X(n))^{(l)}\right)\right) \\
&\le \liminf_{n \to \infty} H\left(OP^X(n) \,\middle|\, T^{-1}\left((OP^X(n))^{(k)}\right)\right)
\end{aligned}
$$

for all $k \in \mathbb{N}$.　□

Now, Lemmas A2 and A3 provide

$$
\begin{aligned}
\mathrm{KS} &\le \liminf_{n \to \infty} H\left(OP^X(n) \,\middle|\, T^{-1}\left((OP^X(n))^{(k)}\right)\right) \\
&\le \liminf_{n \to \infty} H\left((OP^X(n+1))^{(k)} \,\middle|\, (OP^X(n))^{(k)}\right) \\
&\le \underline{\mathrm{PE}}^X \le \overline{\mathrm{PE}}^X.
\end{aligned}
$$

The assumption $\mathrm{KS} \ge \overline{\mathrm{PE}}^X$ then implies

$$
\begin{aligned}
\mathrm{KS} &= \liminf_{n \to \infty} H\left(OP^X(n) \,\middle|\, T^{-1}\left((OP^X(n))^{(k)}\right)\right) \\
&= \liminf_{n \to \infty} H\left((OP^X(n+1))^{(k)} \,\middle|\, (OP^X(n))^{(k)}\right) \\
&= \underline{\mathrm{PE}}^X = \overline{\mathrm{PE}}^X
\end{aligned}
$$

for all $k \in \mathbb{N}$.

Appendix A.4. Proofs for the Simple One-Dimensional Case

This subsection is mainly devoted to the proofs of Theorem 4 and to the proof of Lemma A5 mentioned at the end of Section 3.1. Recall the assumption that Ω is a subset of $\mathbb{R}$ and $\mathscr{A} = \mathscr{B}$ the Borel σ-algebra on Ω. The following lemma is a step to the proof of Theorem 4.

Lemma A4. *Let Ω be a subset of $\mathbb{R}$ and $\mathscr{A} = \mathscr{B}$ be the Borel σ-algebra on Ω. Suppose, there exist an ordered partition $\mathcal{M} = \{M_i\}_{i \in I}$ and an $m \in \mathbb{N}$ satisfying (5). Then, for all $n \in \mathbb{N}$ with $n \ge m$ and multiindices $i = (i_0, i_1, \dots, i_{n-1}) \in I^n$*

$$
\#\Delta(OP(n)|M(i)) \le 2^{\sum_{u=1}^m \#\{s \in \{0,1,\dots,n-1\} \,|\, i_s = i_{n-u} \text{ and } s \ne n-u\}}
$$

holds true.

Proof. Fix $m \in \mathbb{N}$ and $n \in \mathbb{N}$ with $n \ge m$ and $i = (i_0, i_1, \dots, i_{n-1}) \in I^n$. We will show that

$$
\begin{aligned}
&\mathcal{M}^{(s)} \otimes \mathcal{M}^{(s)} \vee \bigvee_{t=0}^{s-1} (T \times T)^{-t}(\{R, \Omega^2 \setminus R\}) \\
&\prec \mathcal{M}^{(s)} \otimes \mathcal{M}^{(s)} \vee \bigvee_{t=s-m}^{s-1} (T \times T)^{-t}(\{R, \Omega^2 \setminus R\})
\end{aligned}
\tag{A5}
$$

holds true for all $s \in \mathbb{N}$ with $s \ge m$ using induction over s:

The above statement is trivial for $s = m$. Suppose that (A5) holds true for some $s \in \mathbb{N}$ with $s \geq m$. We will show that (A5) then holds true for $s + 1$:

$$\mathcal{M}^{(s+1)} \otimes \mathcal{M}^{(s+1)} \vee \bigvee_{t=0}^{s} (T \times T)^{-t} \left(\{R, \Omega^2 \setminus R\} \right)$$

$$= (T \times T)^{-1} \left(\mathcal{M}^{(s)} \otimes \mathcal{M}^{(s)} \vee \bigvee_{t=0}^{s-1} (T \times T)^{-t} \left(\{R, \Omega^2 \setminus R\} \right) \right)$$
$$\vee \mathcal{M}^{(m)} \otimes \mathcal{M}^{(m)} \vee \{R, \Omega^2 \setminus R\}$$

$$\overset{(5)}{\prec} (T \times T)^{-1} \left(\mathcal{M}^{(s)} \otimes \mathcal{M}^{(s)} \vee \bigvee_{t=0}^{s-1} (T \times T)^{-t} \left(\{R, \Omega^2 \setminus R\} \right) \right)$$
$$\vee \mathcal{M}^{(m)} \otimes \mathcal{M}^{(m)} \vee \bigvee_{u=1}^{m} (T \times T)^{-u} \left(\{R, \Omega^2 \setminus R\} \right)$$

$$= (T \times T)^{-1} \left(\mathcal{M}^{(s)} \otimes \mathcal{M}^{(s)} \vee \bigvee_{t=0}^{s-1} (T \times T)^{-t} \left(\{R, \Omega^2 \setminus R\} \right) \right) \vee \mathcal{M} \otimes \mathcal{M}$$

$$\prec (T \times T)^{-1} \left(\mathcal{M}^{(s)} \otimes \mathcal{M}^{(s)} \vee \bigvee_{t=s-m}^{s-1} (T \times T)^{-t} \left(\{R, \Omega^2 \setminus R\} \right) \right) \vee \mathcal{M} \otimes \mathcal{M} \qquad \text{(A6)}$$

$$= \mathcal{M}^{(s+1)} \otimes \mathcal{M}^{(s+1)} \vee \bigvee_{t=s+1-m}^{s} (T \times T)^{-t} \left(\{R, \Omega^2 \setminus R\} \right).$$

In (A6), the induction hypotheses was used.

Notice that

$$\mathcal{M}^{(n)} = \bigvee_{s=1}^{n} (id, T)^{-n+s} \left(\mathcal{M}^{(s)} \otimes \mathcal{M}^{(s)} \right),$$

$$OP(n) = \bigvee_{s=1}^{n} (id, T)^{-n+s} \left(\bigvee_{t=0}^{s-1} (T \times T)^{-t} (\{R, \Omega^2 \setminus R\}) \right)$$

hold true. This implies

$$\mathcal{M}^{(n)} \vee OP(n) = \bigvee_{s=1}^{n} (id, T)^{-n+s} \left(\mathcal{M}^{(s)} \otimes \mathcal{M}^{(s)} \vee \bigvee_{t=0}^{s-1} (T \times T)^{-t} (\{R, \Omega^2 \setminus R\}) \right)$$

$$= \bigvee_{s=1}^{m-1} (id, T)^{-n+s} \left(\mathcal{M}^{(s)} \otimes \mathcal{M}^{(s)} \vee \bigvee_{t=0}^{s-1} (T \times T)^{-t} (\{R, \Omega^2 \setminus R\}) \right)$$
$$\vee \bigvee_{s=m}^{n} (id, T)^{-n+s} \left(\mathcal{M}^{(s)} \otimes \mathcal{M}^{(s)} \vee \bigvee_{t=0}^{s-1} (T \times T)^{-t} (\{R, \Omega^2 \setminus R\}) \right)$$

$$\overset{(A5)}{\prec} \bigvee_{s=1}^{m-1} (id, T)^{-n+s} \left(\mathcal{M}^{(s)} \otimes \mathcal{M}^{(s)} \vee \bigvee_{t=0}^{s-1} (T \times T)^{-t} (\{R, \Omega^2 \setminus R\}) \right)$$
$$\vee \bigvee_{s=m}^{n} (id, T)^{-n+s} \left(\mathcal{M}^{(s)} \otimes \mathcal{M}^{(s)} \vee \bigvee_{t=s-m}^{s-1} (T \times T)^{-t} (\{R, \Omega^2 \setminus R\}) \right)$$

$$= \mathcal{M}^{(n)} \vee \bigvee_{s=0}^{n-1} \bigvee_{t=n-m}^{n-1} (T^s, T^t)^{-1} (\{R, \Omega^2 \setminus R\}).$$

Therefore,

$$\#\Delta(OP(n)|M(\boldsymbol{i})) = \#\Delta(\mathcal{M}^{(n)} \vee OP(n)|M(\boldsymbol{i}))$$

$$\leq \#\Delta\left(\mathcal{M}^{(n)} \vee \bigvee_{s=0}^{n-1}\bigvee_{t=n-m}^{n-1}(T^s,T^t)^{-1}(\{R,\Omega^2\setminus R\})|M(\boldsymbol{i})\right)$$

$$= \#\Delta\left(\bigvee_{s=0}^{n-1}\bigvee_{t=n-m}^{n-1}(T^s,T^t)^{-1}(\{R,\Omega^2\setminus R\})|M(\boldsymbol{i})\right)$$

$$\leq \prod_{s=0}^{n-1}\prod_{t=n-m}^{n-1}\#\Delta\left((T^s,T^t)^{-1}(\{R,\Omega^2\setminus R\})|M(\boldsymbol{i})\right) \tag{A7}$$

holds true. Notice that

$$(T^s,T^t)^{-1}(\{R,\Omega^2\setminus R\}) = \{\{\omega\in\Omega:T^s(\omega)<T^t(\omega)\},\{\omega\in\Omega:T^s(\omega)\geq T^t(\omega)\}\}.$$

For $s=t$, we have

$$\#\Delta\left((T^s,T^t)^{-1}(\{R,\Omega^2\setminus R\})|M(\boldsymbol{i})\right) = \#\Delta\left(\{\varnothing,\Omega\}|M(\boldsymbol{i})\right) = 1.$$

If $i_s\neq i_t$ is true, using the fact that $\mathcal{M}$ is an ordered partition yields

$$\#\Delta\left((T^s,T^t)^{-1}(\{R,\Omega^2\setminus R\})|M(\boldsymbol{i})\right)$$

$$= \#\Delta\left((T^s,T^t)^{-1}((M_{i_s}\times M_{i_t})\vee\{R,\Omega^2\setminus R\})|M(\boldsymbol{i})\right)$$

$$\overset{(3)}{=} \#\Delta\left((T^s,T^t)^{-1}(M_{i_s}\times M_{i_t})|M(\boldsymbol{i})\right) = 1.$$

For all other cases, we have

$$\#\Delta\left((T^s,T^t)^{-1}(\{R,\Omega^2\setminus R\})|M(\boldsymbol{i})\right) \leq \#\left((T^s,T^t)^{-1}(\{R,\Omega^2\setminus R\})\right) = 2.$$

The above observations can be summarized as

$$\#\Delta\left((T^s,T^t)^{-1}(\{R,\Omega^2\setminus R\})|M(\boldsymbol{i})\right)\begin{cases}=1 & \text{if } s=t \text{ or } i_s\neq i_t,\\ \leq 2 & \text{if } s\neq t \text{ and } i_s=i_t.\end{cases}$$

In combination with (A7), this provides

$$\#\Delta\left((T^s,T^t)^{-1}(\{R,\Omega^2\setminus R\})|M(\boldsymbol{i})\right)$$

$$\leq 2^{\sum_{t=n-m}^{n-1}\#\{s\in\{0,1,\dots,n-1\}|i_s=i_t \text{ and } s\neq t\}}$$

$$= 2^{\sum_{u=1}^{m}\#\{s\in\{0,1,\dots,n-1\}|i_s=i_{n-u} \text{ and } s\neq n-u\}}. \qquad \square$$

Notice that the above Lemma immediately implies

$$H(OP(n)) \leq H(OP(n)\vee\mathcal{M}^{(n)}) = H(\mathcal{M}^{(n)}) + H(OP(n)|\mathcal{M}^{(n)})$$

$$\overset{(A1)}{\leq} H(\mathcal{M}^{(n)}) + \sum_{\boldsymbol{i}\in I^n}\mu(M(\boldsymbol{i}))\cdot\log(\#\Delta(OP(n)|M(\boldsymbol{i})))$$

$$\leq H(\mathcal{M}^{(n)}) + \log(2)\cdot nm. \tag{A8}$$

We come now to the proof of Theorem 4, which slightly generalizes a proof given in [10], where the case $m = 1$ was considered. For better readability, we restate this proof with the generalization to arbitrary $m \in \mathbb{N}$ at the appropriate places within the proof.

Take $\epsilon > 0$. According to Conditions 1 and 2, there exist finite or countably infinite ordered partitions $\mathcal{M} = \{M_i\}_{i \in I}$, $\mathcal{Q} = \{Q_j\}_{j \in J}$ and $m \in \mathbb{N}$ with $H(\mathcal{M}) < \infty$ and $H(\mathcal{Q}) < \infty$ satisfying (5) and (6). Consider the partition

$$\mathcal{P} := \mathcal{M} \vee \mathcal{Q} = \{M_i \cap Q_j\}_{(i,j) \in I \times J} =: \{P_{(i,j)}\}_{(i,j) \in I \times J}.$$

Notice that $\mathcal{P}$ is again a finite or countably infinite ordered partition with $H(\mathcal{P}) < H(\mathcal{M}) + H(\mathcal{Q}) < \infty$. Using (A1), this implies

$$\begin{aligned}
\overline{\text{PE}} &\leq \limsup_{n \to \infty} \frac{1}{n} H(OP(n) \vee \mathcal{P}^{(n)}) \\
&= h(\mathcal{P}) + \limsup_{n \to \infty} \frac{1}{n} H(OP(n) | \mathcal{P}^{(n)}) \\
&\leq \text{KS} + \limsup_{n \to \infty} \frac{1}{n} H(OP(n) | \mathcal{P}^{(n)}) \\
&\leq \text{KS} + \limsup_{n \to \infty} \frac{1}{n} \sum_{(\boldsymbol{i}, \boldsymbol{j}) \in (I \times J)^n} \mu(P((\boldsymbol{i}, \boldsymbol{j}))) \log(\#\Delta(OP(n) | P((\boldsymbol{i}, \boldsymbol{j})))),
\end{aligned} \tag{A9}$$

where we consider $(\boldsymbol{i}, \boldsymbol{j})$ itself as one multiindex and $I \times J$ as one index set. Thus,

$$P((\boldsymbol{i}, \boldsymbol{j})) = \bigcap_{t=0}^{n-1} T^{-t}(M_{i_t} \cap Q_{j_t})$$

for all $(\boldsymbol{i}, \boldsymbol{j}) = ((i_0, j_0), (i_1, j_1), \ldots, (i_{n-1}, j_{n-1})) \in (I \times J)^n$. Lemma A4 provides

$$\begin{aligned}
&\sum_{(\boldsymbol{i}, \boldsymbol{j}) \in (I \times J)^n} \mu(P((\boldsymbol{i}, \boldsymbol{j}))) \cdot \log(\#\Delta(OP(n) | P((\boldsymbol{i}, \boldsymbol{j})))) \\
&\leq \log 2 \sum_{(\boldsymbol{i}, \boldsymbol{j}) \in (I \times J)^n} \mu(P((\boldsymbol{i}, \boldsymbol{j}))) \left(\sum_{u=1}^{m} \#\{s \in \{0, 1, \ldots, n-1\} \mid (i_s, j_s) = (i_{n-u}, j_{n-u}) \text{ and } s \neq n - u\} \right) \\
&\leq \log 2 \sum_{(\boldsymbol{i}, \boldsymbol{j}) \in (I \times J)^n} \mu(P((\boldsymbol{i}, \boldsymbol{j}))) \left(\sum_{u=1}^{m} \#\{s \in \{0, 1, \ldots, n-1\} \mid j_s = j_{n-u} \text{ and } s \neq n - u\} \right) \\
&= \log 2 \sum_{u=1}^{m} \sum_{\boldsymbol{j} \in J^n} \left(\sum_{\boldsymbol{i} \in I^n} \mu(P((\boldsymbol{i}, \boldsymbol{j}))) \right) \cdot \#\{s \in \{0, 1, \ldots, n-1\} \mid j_s = j_{n-u} \text{ and } s \neq n - u\} \\
&= \log 2 \sum_{u=1}^{m} \sum_{\boldsymbol{j} \in J^n} \mu(Q(\boldsymbol{j})) \cdot \#\{s \in \{0, 1, \ldots, n-1\} \mid j_s = j_{n-u} \text{ and } s \neq n - u\} \\
&\leq \log 2 \sum_{u=1}^{m} \sum_{\boldsymbol{j} \in J^n} \mu(Q(\boldsymbol{j})) \cdot (\#\{s \in \{0, 1, \ldots, n-u-1\} \mid j_s = j_{n-u}\} + u - 1) \\
&= \log 2 \cdot \left(m(m-1)/2 + \sum_{u=1}^{m} \sum_{\boldsymbol{j} \in J^n} \mu(Q(\boldsymbol{j})) \cdot \#\{s \in \{0, 1, \ldots, n-u-1\} \mid j_s = j_{n-u}\} \right).
\end{aligned} \tag{A10}$$

Combining (A9) and (A10) yields

$$\overline{\text{PE}} \leq \text{KS} + \log 2 \cdot \sum_{u=1}^{m} \limsup_{n \to \infty} \frac{1}{n} \sum_{\boldsymbol{j} \in J^n} \mu(Q(\boldsymbol{j})) \cdot \#\{s \in \{0, 1, \ldots, n-u-1\} \mid j_s = j_{n-u}\}.$$

For each $u \in \{1, \ldots, m\}$, we have

$$
\limsup_{n\to\infty} \frac{1}{n} \sum_{j\in J^n} \mu(Q(j)) \cdot \#\{s \in \{0, 1, ..., n-u-1\} \mid j_s = j_{n-u}\}
$$

$$
= \limsup_{n\to\infty} \frac{1}{n} \sum_{j_{n-u}\in J} \sum_{j\in J^{n-u}} \mu(Q((j, j_{n-u}))) \cdot \#\{s \in \{0, 1, ..., n-u-1\} \mid j_s = j_{n-u}\}
$$

$$
= \limsup_{n\to\infty} \frac{1}{n-u} \sum_{j_{n-u}\in J} \sum_{l=0}^{n-u} \mu(\{\omega \in T^{-n+u}(Q_{j_{n-u}}) \mid \#\{s \in \{0, 1, ..., n-u-1\} \mid T^s(\omega) \in Q_j\} = l\}) \cdot l
$$

$$
= \sum_{Q\in\mathcal{Q}} \limsup_{n\to\infty} \frac{1}{n-u} \sum_{l=0}^{n-u} \mu(T^{-n+u}(Q) \cap T^{-l}(Q))
$$

$$
= \sum_{Q\in\mathcal{Q}} \limsup_{n\to\infty} \frac{1}{n-u} \sum_{l=0}^{n-u} \mu(Q \cap T^{-l}(Q)) \overset{(6)}{<} \epsilon.
$$

Hence,

$$
\overline{PE} \le KS + \log 2 \cdot m \cdot \epsilon.
$$

The statement of the theorem follows from the fact that ϵ can be chosen arbitrarily close to 0.

Lemma A5. *Let Ω be a subset of $\mathbb{R}$ and $\mathscr{A} = \mathscr{B}$ be the Borel σ-algebra on Ω. If T is (countably) piecewise monotone and completely periodic, i.e.,*

$$
\mu\left(\bigcup_{t=1}^{\infty} \{\omega \in \Omega \mid T^t(\omega) = \omega\}\right) = 1,
$$

then

$$
\overline{PE} = 0
$$

holds true.

Proof. Let

$$
\Theta_k := \bigcup_{t=1}^{k} \{\omega \in \Omega : T^t(\omega) = \omega\}
$$

be the set of all points with period smaller or equal to $k \in \mathbb{N}$ and

$$
\Theta = \bigcup_{t=1}^{\infty} \{\omega \in \Omega : T^t(\omega) = \omega\}
$$

the set of all periodic points. Since $\mu(\Theta) = 1$, for all $\epsilon > 0$ there exists $k \in \mathbb{N}$ with

$$
\mu(\Theta_k) > 1 - \epsilon.
$$

This implies

$$
\overline{PE} \le \limsup_{n\to\infty} \frac{1}{n} H(OP(n) \vee \{\Theta_k, \Omega \setminus \Theta_k\})
$$

$$
= \limsup_{n\to\infty} \frac{1}{n} \left[H(OP(n) \vee \{\Theta_k, \Omega \setminus \Theta_k\}) - H(\Theta_k, \Omega \setminus \Theta_k) \right]
$$

$$= \mu(\Theta_k) \cdot \limsup_{n\to\infty} \frac{1}{n} - \sum_{P_\pi \in OP(n)} \frac{\mu(P_\pi \cap \Theta_k)}{\mu(\Theta_k)} \log\left(\frac{\mu(P_\pi \cap \Theta_k)}{\mu(\Theta_k)}\right)$$

$$+ \mu(\Omega \setminus \Theta_k) \cdot \limsup_{n\to\infty} \frac{1}{n} - \sum_{P_\pi \in OP(n)} \frac{\mu(P_\pi \setminus \Theta_k)}{\mu(\Omega \setminus \Theta_k)} \log\left(\frac{\mu(P_\pi \setminus \Theta_k)}{\mu(\Omega \setminus \Theta_k)}\right)$$

$$\overset{(A8)}{\leq} \mu(\Theta_k) \cdot \limsup_{n\to\infty} \frac{1}{n} - \sum_{P_\pi \in OP(n)} \frac{\mu(P_\pi \cap \Theta_k)}{\mu(\Theta_k)} \log\left(\frac{\mu(P_\pi \cap \Theta_k)}{\mu(\Theta_k)}\right)$$

$$+ \epsilon \cdot \limsup_{n\to\infty} \frac{1}{n} \cdot \left[H(\mathcal{M}^{(n)}) + \log(2) \cdot n\right]$$

$$\leq \mu(\Theta_k) \cdot \limsup_{n\to\infty} \frac{1}{n} - \sum_{P_\pi \in OP(n)} \frac{\mu(P_\pi \cap \Theta_k)}{\mu(\Theta_k)} \log\left(\frac{\mu(P_\pi \cap \Theta_k)}{\mu(\Theta_k)}\right) + (H(\mathcal{M}) + \log(2)) \cdot \epsilon$$

$$= \mu(\Theta_k) \cdot \limsup_{n\to\infty} \frac{1}{n} - \sum_{P_\pi \in OP(k)} \frac{\mu(P_\pi \cap \Theta_k)}{\mu(\Theta_k)} \log\left(\frac{\mu(P_\pi \cap \Theta_k)}{\mu(\Theta_k)}\right) + (H(\mathcal{M}) + \log(2)) \cdot \epsilon$$

$$= (H(\mathcal{M}) + \log(2)) \cdot \epsilon.$$

This provides $\overline{PE} = 0$ because ϵ can be chosen arbitrarily close to 0. $\square$

Remark A1. *To be able to show that Condition 2 holds true under the assumptions of Corollary 1 for non-ergodic systems via ergodic decomposition, one needs to require that $(\Omega, \mathcal{B}, \mu)$ is a Lebesgue space. A probability space $(\Omega, \mathcal{B}, \mu)$ is called a Lebesgue space if $(\Omega, \mathcal{B}, \mu)$ is isomorph to some probability space $(\widetilde{\Omega}, \widetilde{\mathcal{B}}, \widetilde{\mu})$, where $\widetilde{\Omega}$ is a complete separable metric space and $\widetilde{\mathcal{B}}$ the completion of the Borel σ-algebra on $\widetilde{\Omega}$, i.e., $\widetilde{B}$ contains additionally all subsets of Borel sets with measure 0. If $\Omega \subseteq \mathbb{R}$ is a Borel subset, $(\Omega, \mathcal{B}, \mu)$ is a Lebesgue space if $\mathcal{B}$ is complete with regard to μ (see e.g., [14]).*

Alternatively, one can use Rokhlin–Halmos towers to show that Condition 2 holds true for non-ergodic systems (see [10]). For this approach, it is only necessary to require that Ω is a separable metric space, $\mathcal{B}$ the Borel σ-algebra on Ω, and $T : \Omega \to \Omega$ an aperiodic map [15].

Moreover, notice that, in [10], it was required that Ω is a compact metric space so that μ is regular, which allowed for approximating any set of $\mathcal{B}$ by a finite union of intervals. However, this is not necessary because the Borel σ-algebra is generated by the algebra containing all sets of the type $I \cap \Omega$, where I is an open or closed interval, and every set of a σ-algebra can be approximated by a set of the algebra that generates that σ-algebra (see, e.g., [16]).

Appendix A.5. Proof of Lemma 1 and the 'Ergodic Part' of Lemma 2

Let Ω be a subset of $\mathbb{R}$ and $\mathscr{A} = \mathscr{B}$ be the Borel σ-algebra on Ω. We start with showing Lemma 1. For this, fix some $n \in \mathbb{N}$. By its definition, the set V_n can be written as a union of sets in $OP(n) \vee T^{-1}(OP(n))$. Notice that

$$OP(n+1) = OP(n) \vee T^{-1}(OP(n)) \vee (id, T^n)^{-1}(\{R, \Omega^2 \setminus R\}). \tag{A11}$$

For $Q \in OP(n) \vee T^{-1}(OP(n))$, consider some $\omega \in Q$. If $\omega \notin V_n$ is true, we can use the transitivity of the order relation to determine the order relation of ω and $T^n(\omega)$ from the ordering given by Q. This implies

$$\#\Delta((id, T^n)^{-1}(\{R, \Omega^2 \setminus R\})|Q) = 1 \tag{A12}$$

for all $Q \subseteq \Omega \setminus V_n$. Thus,

$$H(OP(n+1)|OP(n) \vee T^{-1}(OP(n)))$$

$$\overset{(A11)}{\leq} H(OP(n) \vee T^{-1}(OP(n))|OP(n) \vee T^{-1}(OP(n))) +$$
$$H((id, T^n)^{-1}(\{R, \Omega^2 \setminus R\})|OP(n) \vee T^{-1}(OP(n)))$$

$$= H((id, T^n)^{-1}(\{R, \Omega^2 \setminus R\})|OP(n) \vee T^{-1}(OP(n)))$$

$$\overset{(A1)}{\leq} \sum_{Q \in OP(n) \vee T^{-1}(OP(n))} \mu(Q) \cdot \log(\#\Delta((id, T^n)^{-1}(\{R, \Omega^2 \setminus R\})|Q))$$

$$= \sum_{\substack{Q \in OP(n) \vee T^{-1}(OP(n)) \\ Q \subseteq V_n}} \mu(Q) \cdot \log(\#\Delta((id, T^n)^{-1}(\{R, \Omega^2 \setminus R\})|Q))$$

$$+ \sum_{\substack{Q \in OP(n) \vee T^{-1}(OP(n)) \\ Q \nsubseteq V_n}} \mu(Q) \cdot \log(\#\Delta((id, T^n)^{-1}(\{R, \Omega^2 \setminus R\})|Q))$$

$$= \sum_{\substack{Q \in OP(n) \vee T^{-1}(OP(n)) \\ Q \subseteq V_n}} \mu(Q) \cdot \log(\#\Delta((id, T^n)^{-1}(\{R, \Omega^2 \setminus R\})|Q))$$

$$+ \sum_{\substack{Q \in OP(n) \vee T^{-1}(OP(n)) \\ Q \subseteq \Omega \setminus V_n}} \mu(Q) \cdot \log(\#\Delta((id, T^n)^{-1}(\{R, \Omega^2 \setminus R\})|Q))$$

$$\overset{(A12)}{=} \sum_{\substack{Q \in OP(n) \vee T^{-1}(OP(n)) \\ Q \subseteq V_n}} \mu(Q) \cdot \log(\#\Delta((id, T^n)^{-1}(\{R, \Omega^2 \setminus R\})|Q))$$

$$\leq \sum_{\substack{Q \in OP(n) \vee T^{-1}(OP(n)) \\ Q \subseteq V_n}} \mu(Q) \cdot \log(2)$$

$$= \log(2) \cdot \mu(V_n).$$

This shows Lemma 1.

To prove the 'ergodic part' of Lemma 2, take $\epsilon > 0$. Choose an ordered partition $\mathcal{U} = \{U_i\}_{i=1}^N$ of Ω such that $0 < \mu(U_i) < \epsilon$ holds true for all $i \in I$. This is always possible because μ was assumed to be aperiodic. Label the sets $U_i \in \mathcal{U}$ with $i \in \{1, 2, \ldots, N\}$ in such a way that

$$i_1 < i_2 \Rightarrow \mu^2(\{(\omega_1, \omega_2) \in U_{i_1} \times U_{i_2} : \omega_1 > \omega_2\}) = 0$$

holds true for all $i_1, i_2 \in \{1, 2, \ldots, N\}$. Since T is ergodic, there exists an $n_o \in \mathbb{N}$ such that

$$\mu\left(\bigcap_{i=1}^N \bigcup_{s=1}^{n_0-1} T^{-s}(U_i)\right) > 1 - \epsilon$$

holds true.

The set $\bigcap_{i=1}^N \bigcup_{s=1}^{n_0-1} T^{-s}(U_i)$ consists of all $\omega \in \Omega$ with orbit $(\omega, T(\omega), \ldots, T^{n_0-1})$ visiting each of the sets in $\mathcal{U}$. Thus, if such ω lies in U_i with $1 < i < N$ and in V_t for $t \geq n_0$, by definition of V_t, the point $T^t(\omega)$ must belong to $U_{t-1} \cup U_t \cup U_{t+1}$. With a similar argumentation for $\omega \in U_1$ or $\omega \in U_N$, one obtains the following:

$$\mu\left(V_t \cap \bigcap_{i=1}^N \bigcup_{s=1}^{n_0-1} T^{-s}(U_i)\right)$$

$$\leq \mu(U_1 \cap T^{-t}(U_1 \cup U_2)) + \sum_{i=2}^{N-1} \mu(U_i \cap T^{-t}(U_{i-1} \cup U_i \cup U_{i+1})) + \mu(U_N \cap T^{-t}(U_{N-1} \cup U_N))$$

holds true for all $t \in \mathbb{N}$ with $t > n_0$. Using the ergodicity of T implies

$$\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \mu\left(V_t \cap \bigcap_{i=1}^{N}\bigcup_{s=1}^{n_0-1} T^{-s}(U_i)\right) = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mu\left(V_{t+n_0}\cap\bigcap_{i=1}^{N}\bigcup_{s=1}^{n_0-1}T^{-s}(U_i)\right)$$

$$\leq \mu(U_1)\mu(U_1\cup U_2) + \sum_{i=2}^{N-1}\mu(U_i)\mu(U_{i-1}\cup U_i\cup U_{i+1}) + \mu(U_N)\mu(U_{N-1}\cup U_N)$$

$$\leq \mu(U_1)2\epsilon + \sum_{i=2}^{N-1}\mu(U_i)3\epsilon + \mu(U_N)2\epsilon$$

$$\leq 3\epsilon.$$

The Stolz–Cesàro theorem then provides

$$\liminf_{n\to\infty}\mu\left(V_n\cap\bigcap_{i=1}^{N}\bigcup_{s=1}^{n_0-1}T^{-s}(U_i)\right) \leq \liminf_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mu\left(V_t\cap\bigcap_{i=1}^{N}\bigcup_{s=1}^{n_0-1}T^{-s}(U_i)\right) \leq 3\epsilon.$$

Hence,

$$\liminf_{n\to\infty}\mu\left(V_n\right)$$

$$= \liminf_{n\to\infty}\left[\mu\left(V_n\cap\bigcap_{i=1}^{N}\bigcup_{s=1}^{n_0-1}T^{-s}(U_i)\right) + \mu\left(V_n\setminus\bigcap_{i=1}^{N}\bigcup_{s=1}^{n_0-1}T^{-s}(U_i)\right)\right]$$

$$\leq \liminf_{n\to\infty}\mu\left(V_n\cap\bigcap_{i=1}^{N}\bigcup_{s=1}^{n_0-1}T^{-s}(U_i)\right) + \epsilon$$

$$\leq 4\epsilon.$$

Since ϵ can be choosen arbitrarily close to 0, this implies $\liminf_{n\to\infty}\mu\left(V_n\right) = 0$.

Appendix A.6. Proof of Theorem 5

Let Ω be a subset of $\mathbb{R}$ and $\mathscr{A} = \mathscr{B}$ be the Borel σ-algebra on Ω and consider the set

$$\Theta = \bigcup_{t=1}^{\infty}\{\omega\in\Omega \mid T^t(\omega) = \omega\}.$$

Since T is μ-almost surely aperiodic, we have $\mu(\Theta) = 0$.

We will prove the statement of the theorem by contradiction. Suppose $\sum_{n=1}^{\infty}\mu(V_n) < \infty$ holds true. Using the Borel–Cantelli lemma, this implies

$$\mu\left(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}V_n\right) = 0$$

or, equivalently,

$$\lim_{K\to\infty}\mu\left(\bigcup_{k=1}^{K}\bigcap_{n=k}^{\infty}(\Omega\setminus V_n)\right) = \mu\left(\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}(\Omega\setminus V_n)\right) = 1.$$

Therefore, there exists a $K\in\mathbb{N}$ with

$$\mu\left(\left(\bigcap_{n=K}^{\infty}(\Omega\setminus V_n)\right)\setminus\Theta\right) = \mu\left(\bigcap_{n=K}^{\infty}(\Omega\setminus V_n)\right) = \mu\left(\bigcup_{k=1}^{K}\bigcap_{n=k}^{\infty}(\Omega\setminus V_n)\right) > 0.$$

Set

$$\delta(\omega) := \min_{1 \le s \le K-1} |\omega - T^s(\omega)|$$

for all $\omega \in \Omega$. Notice that every aperiodic point $\omega \notin \Theta$ satisfies $\delta(\omega) > 0$. Thus,

$$0 < \mu \left(\bigcap_{n=K}^{\infty} (\Omega \setminus V_n) \setminus \Theta \right) = \mu \left(\bigcup_{i=1}^{\infty} \left\{ \omega \in \bigcap_{n=K}^{\infty} (\Omega \setminus V_n) \setminus \Theta \mid \delta(\omega) > 1/i \right\} \right)$$

$$= \lim_{i \to \infty} \mu \left(\left\{ \omega \in \bigcap_{n=K}^{\infty} (\Omega \setminus V_n) \setminus \Theta \mid \delta(\omega) > 1/i \right\} \right).$$

Thus, there exists some $\delta > 0$ such that

$$A_\delta := \{ \omega \in \bigcap_{n=K}^{\infty} (\Omega \setminus V_n) \setminus \Theta \mid \delta(\omega) > \delta \}$$

has a strictly positive measure. Because there exists a countable set $\Omega_\delta \subseteq \Omega$ with

$$A_\delta = \bigcup_{\omega \in \Omega_\delta} A_\delta \cap (\omega - \delta/2, \omega + \delta/2),$$

we have

$$\mu(A_\delta \cap (\omega_0 - \delta/2, \omega_0 + \delta/2)) > 0$$

for some $\omega_0 \in \Omega$. Using the ergodicity of T, this implies

$$\mu \left(\bigcup_{n=K}^{\infty} T^{-n}((\omega_0 - \delta/2, \omega_0 + \delta/2)) \right) = \mu \left(\bigcup_{n=0}^{\infty} T^{-n}(T^{-K}((\omega_0 - \delta/2, \omega_0 + \delta/2))) \right) = 1$$

and, consequently,

$$\mu \left(A_\delta \cap (\omega_0 - \delta/2, \omega_0 + \delta/2) \cap \bigcup_{n=K}^{\infty} T^{-n}((\omega_0 - \delta/2, \omega_0 + \delta/2)) \right)$$

$$= \mu(A_\delta \cap (\omega_0 - \delta/2, \omega_0 + \delta/2)) > 0.$$

Thus, in particular, $A_\delta \cap (\omega_0 - \delta/2, \omega_0 + \delta/2) \cap \bigcup_{n=K}^{\infty} T^{-n}((\omega_0 - \delta/2, \omega_0 + \delta/2))$ is not empty. Now take some

$$\omega \in A_\delta \cap (\omega_0 - \delta/2, \omega_0 + \delta/2) \cap \bigcup_{n=K}^{\infty} T^{-n}((\omega_0 - \delta/2, \omega_0 + \delta/2)).$$

We have $|\omega - \omega_0| < \delta/2$. Additionally, there exists $n_0 \in \mathbb{N}$ with $n_0 \ge K$ such that $T^{n_0}(\omega) \in (\omega_0 - \delta/2, \omega_0 + \delta/2)$ holds true, which is equivalent to $|\omega_0 - T^{n_0}(\omega)| < \delta/2$. As a consequence,

$$|\omega - T^{n_0}(\omega)| \le |\omega - \omega_0| + |\omega_0 - T^{n_0}(\omega)| < \delta$$

holds true. This implies that

$$m := \min\{ n \in \mathbb{N} \mid |\omega - T^n(\omega)| < \delta \}$$

is smaller or equal to n_0. In particular, $m \in \mathbb{N}$ is well defined and not infinite. On the other hand,

$$\omega \in A_\delta \subseteq \{ \omega \in \Omega \mid \min_{1 \le s \le K-1} |\omega - T^s(\omega)| > \delta \}$$

implies $m \geq K$. By construction of m, we have

$$|\omega - T^s(\omega)| \geq \delta > |\omega - T^m(\omega)|$$

for all $s \in \{1, 2, \ldots, m-1\}$. Hence, $\omega \in V_m$ holds true, which is a contradiction to

$$\omega \in A_\delta \subseteq \bigcap_{n=K}^{\infty} \Omega \setminus V_n.$$

Therefore, $\sum_{n=1}^{\infty} \mu(V_n) < \infty$ cannot be true.

References

1. Bandt, C.; Keller, G.; Pompe, B. Entropy of interval maps via permutations. *Nonlinearity* **2002**, *15*, 1595. [CrossRef]
2. Amigó, J.M. The equality of Kolmogorov–Sinai entropy and metric permutation entropy generalized. *Physica D* **2012**, *241*, 789–793. [CrossRef]
3. Antoniouk, A.; Keller, K.; Maksymenko, S. Kolmogorov–Sinai entropy via separation properties of order-generated σ-algebras. *Discret. Cont. Dyn.-A* **2014**, *34*, 1793.
4. Fouda, E.; Koepf, W.; Jacquir, S. The ordinal Kolmogorov–Sinai entropy: A generalized approximation. *Commun. Nonlinear Sci.* **2017**, *46*, 103–115. [CrossRef]
5. Unakafova, V.; Unakafov, A.; Keller, K. An approach to comparing Kolmogorov–Sinai and permutation entropy. *Eur. Phys. J. Spec. Top.* **2013**, *222*, 353–361. [CrossRef]
6. Watt, S.; Politi, A. Permutation entropy revisited. *Chaos Soliton Fract.* **2019**, *120*, 95–99. [CrossRef]
7. Bandt, C.; Pompe, B. Permutation Entropy: A Natural Complexity Measure for Time Series. *Phys. Rev. Lett.* **2002**, *88*, 174102. [CrossRef] [PubMed]
8. Amigó, J.M.; Kennel, M.B.; Kocarev, L. The permutation entropy rate equals the metric entropy rate for ergodic information sources and ergodic dynamical systems. *Physica D* **2005**, *210*, 77–95. [CrossRef]
9. Haruna, T.; Nakajima, K. Permutation complexity via duality between values and orderings. *Physica D* **2011**, *240*, 1370–1377. [CrossRef]
10. Gutjahr, T.; Keller, K. Equality of Kolmogorov–Sinai and permutation entropy for one-dimensional maps consisting of countably many monotone parts. *Discret. Cont. Dyn.-A* **2019**, *39*, 4207. [CrossRef]
11. Einsiedler, M.; Ward, T. *Ergodic Theory: With a View Towards Number Theory*; Graduate Texts in Mathematics; Springer: London, UK, 2010.
12. Keller, K.; Unakafov, A.M.; Unakafova, V.A. On the relation of KS entropy and permutation entropy. *Physica D* **2012**, *241*, 1477–1481. [CrossRef]
13. Walters, P. *An Introduction to Ergodic Theory*; Graduate Texts in Mathematics; Springer: New York, NY, USA, 2000.
14. Rokhlin, V.A. On the fundamental ideas of measure theory. *Am. Math. Soc. Transl.* **1952**, *71*, 55.
15. Heinemann, S.; Schmitt, O. Rokhlin's Lemma for Non-invertible Maps. *Dyn. Syst. Appl.* **2001**, *2*, 201–213.
16. Halmos, P.R. Extension of Measures. In *Measure Theory*; Springer: New York, NY, USA, 1950; pp. 49–72.