MDPI

Article

Generalized Ordinal Patterns and the KS-Entropy

Tim Gutjahr and Karsten Keller *

Institute of Mathematics, University of Lübeck, D-23562 Lübeck, Germany; gutjahr@math.uni-luebeck.de
* Correspondence: keller@math.uni-luebeck.de; Tel.: +49-451-3101-6030; Fax: +49-451-3101-6004

Abstract: Ordinal patterns classifying real vectors according to the order relations between their components are an interesting basic concept for determining the complexity of a measure-preserving dynamical system. In particular, as shown by C. Bandt, G. Keller and B. Pompe, the permutation entropy based on the probability distributions of such patterns is equal to Kolmogorov–Sinai entropy in simple one-dimensional systems. The general reason for this is that, roughly speaking, the system of ordinal patterns obtained for a real-valued "measuring arrangement" has high potential for separating orbits. Starting from a slightly different approach of A. Antoniouk, K. Keller and S. Maksymenko, we discuss the generalizations of ordinal patterns providing enough separation to determine the Kolmogorov–Sinai entropy. For defining these generalized ordinal patterns, the idea is to substitute the basic binary relation $\leq$ on the real numbers by another binary relation. Generalizing the former results of I. Stolz and K. Keller, we establish conditions that the binary relation and the dynamical system have to fulfill so that the obtained generalized ordinal patterns can be used for estimating the Kolmogorov–Sinai entropy.

Keywords: ordinal patterns; measure-preserving dynamical system; Kolmogorov–Sinai entropy; permutation entropy; ergodic theory

1. Introduction

In 2002, Bandt and Pompe introduced so-called permutation entropy [1]. This entropy has been established in non-linear dynamical system theory and time series analysis, including applications in many fields from biomedicine to econophysics (compare with Zanin et al. [2]). It is a crucial point that permutation entropy is theoretically justified by asymptotic results relating it to Kolmogorov–Sinai entropy (KS entropy, also called metric entropy) which is the central complexity measure for dynamical systems. The important relationship of permutation entropy and KS entropy was first observed and mathematically founded for piece-wise monotone dynamical systems by Bandt et al. [3].

The (empirical) concept of permutation entropy is based upon analyzing the distribution of ordinal patterns in a time series or the underlying system. In this paper, we concentrate on a *measure-preserving dynamical system* $(\Omega, \mathscr{A}, \mu, T)$, i.e., a probability space $(\Omega, \mathscr{A}, \mu)$ equipped with a measurable map $T : \Omega \to \Omega$ satisfying $\mu(T^{-1}(A)) = \mu(A)$ for all $A \in \mathscr{A}$.

Given a random variable $X : \Omega \to \mathbb{R}$, in this paper, an *ordinal pattern of length* $n \in \mathbb{N}$ with respect to X is considered as a subset of the state space Ω. It is indicated by a permutation $\pi = (\pi_0, \pi_1, \ldots, \pi_{n-1})$ of $\{0, 1, \ldots, n-1\}$ and defined by

$$P_\pi := \{\omega \in \Omega \mid X(T^{\pi_0}(\omega)) \leq X(T^{\pi_1}(\omega)) \leq \ldots \leq X(T^{\pi_{n-1}}(\omega))\}. \qquad (1)$$

(Usually, ordinal patterns are defined in the range of X, i.e., for the vectors $(X(T(\omega)), X(T^1(\omega)), \ldots, X(T^{n-1}(\omega)))$). The collection of ordinal patterns:

$$OP(n) := \{P_\pi \mid \pi \text{ is a permutation of length } n\}$$

is a partition of Ω.

In the rest of this section, we assume that X preserves enough information about the given system in a certain sense. This is particularly the case if Ω is contained in $\mathbb{R}$ and X is the identity map. A precise general description of the assumption is given when presenting the results of this paper. It was shown in [4] that, under not too restrictive further conditions, the probability distribution on the partitions $OP(n)$ for $n \in \mathbb{N}$ can be used for determining the KS entropy of the given system. The reason is that, roughly speaking, under these conditions, $OP(n)$ is able to separate the orbits of the system if $n \to \infty$.

In order to address the problem that this paper is concerned with, we give a description of ordinal patterns being slightly different from the above. One can determine to which ordinal pattern P_π of length n a point ω belongs to if, for all (s,t) in:

$$E_n = \{(s,t) \in \mathbb{N}_0^2 \mid 0 \le s < t \le n-1\}, \tag{2}$$

one knows whether $X(T^s(\omega)) \le X(T^t(\omega))$ holds true or not. In other words, there exists a set $A \subseteq E_n$ such that:

$$P_\pi = \bigcap_{(s,t) \in A} \{\omega \in \Omega \mid (X(T^s(\omega)), X(T^t(\omega))) \in R\}$$

$$\cap \bigcap_{(s,t) \in E_n \setminus A} \{\omega \in \Omega \mid (X(T^s(\omega)), X(T^t(\omega))) \in \mathbb{R}^2 \setminus R\}, \tag{3}$$

where:

$$R := \{(x,y) \in \mathbb{R}^2 \mid x > y\}. \tag{4}$$

The above set contains all the points $\omega \in \Omega$ that satisfy $X(T^s(\omega)) > X(T^t(\omega))$ for $(s,t) \in A$ and $X(T^s(\omega)) \le X(T^t(\omega))$ for $(s,t) \in E_n \setminus A$. Note that, given some arbitrary $A \subseteq E_n$, the set on the right hand side of (3) can be empty. In the case that it is non-empty, it coincides with some ordinal pattern P_π of length n.

While Equation (3) might be a bit more abstract than (1), it shows a way to generalize the concept of ordinal patterns on the basis of replacing the set R in (4) by some arbitrary Borel subset R of $\mathbb{R}^2$, also to investigate why ordinal patterns are so successful.

Definition 1. *We call a non-empty Borel subset R of $\mathbb{R}^2$ discriminating relation.*

The figures given in this paper show different discriminating relations R. In each case, only the part of R contained in $[0,1[^2$ is presented. Note that in the case that X maps Ω into $[0,1[$, the restriction of R itself to this part would not change anything. Figure 1a illustrates R as given in (4), again only on $[0,1[^2$. In the case of such an R, note that $\tan(\pi(X - 1/2))$ mapping $[0,1[$ into $[-\infty, \infty[$ would not make a difference to a given X for our considerations, since order relations and associated partitions are preserved.

Given some discriminating relation, *generalized ordinal patterns of length n with respect to X* are given as the non-empty sets defined by the right hand side of (3) for some $A \subseteq E_n$. Obviously, they also form a partition of Ω. The question that arises is what a discriminating relation R should look like, such that those generalized ordinal patterns inhibit the same nice properties the original ordinal patterns had. More precisely, we ask the following question:
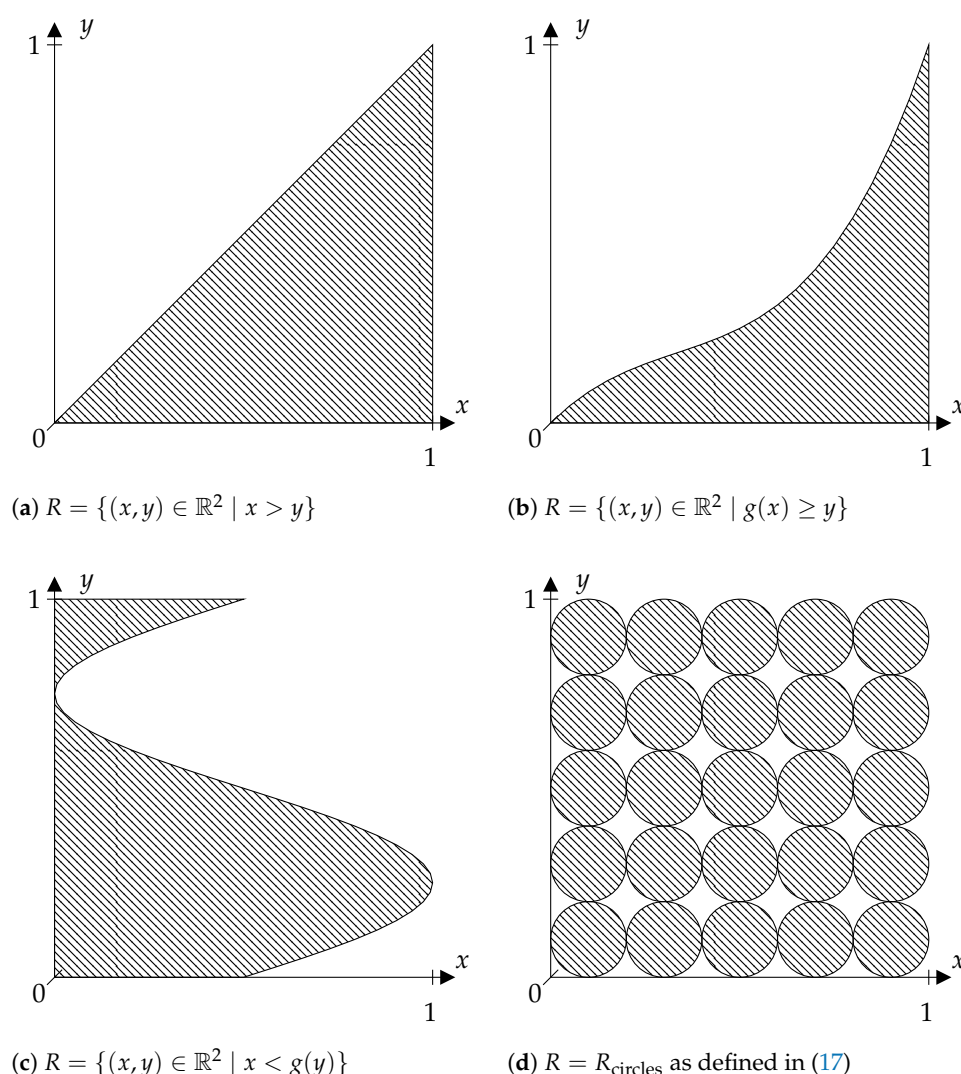
(**a**) $R = \{(x,y) \in \mathbb{R}^2 \mid x > y\}$

(**b**) $R = \{(x,y) \in \mathbb{R}^2 \mid g(x) \geq y\}$

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

(**d**) $R = R_{\text{circles}}$ as defined in (17)

Figure 1. This figure illustrates some special discriminating relations R (striped areas) considered in Section 4. Only the part of R contained in $[0,1[^2$ is shown (compare the corresponding remarks in Section 4).

Main Question. *Under what conditions on a discriminating relation R the partitions given by the generalized ordinal patterns determine the KS entropy of a dynamical system?*

Why is this determination of entropy, which is precisely described by formula (10) in Theorem 1 interesting? For answering this question, interpret X as an observable, ω as the initial state of the given system and $X(\omega), X(T(\omega)), X(T^2(\omega)), X(T^3(\omega)), \ldots$ as the measured values at times $0, 1, 2, 3, \ldots$. Determining (generalized) ordinal patterns on the basis of those values is a symbolization, where a symbol obtained is the (generalized) ordinal pattern containing ω. Generally, symbolization means a coarse-graining of the state space underlying a system, where each point is assigned one of finitely many given symbols. Instead of considering the precise development of the system, one is interested in the change of symbols in the course of time, justifying the naming of the method symbolic dynamics. Note that a symbolization is equivalent to partitioning the state space into classes of states (with the same symbol).

The reason for obtaining the full entropy from the (generalized) ordinal patterns is, roughly speaking, that the symbol system obtained has high potential for separating orbits. Such kinds of successful symbolizations are important, for example, in big data analysis, see, e.g., Smith et al. [5].

The above question was first considered in [6], where the authors basically showed that sets of the form:

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

lead to generalized ordinal patterns that, under some conditions, can be used to determine the entropy if $g : \mathbb{R} \to \mathbb{R}$ is measurable and one-to-one. Such an R is shown in Figure 1b and will be discussed in Section 4 as well as another R illustrated in Figure 1.
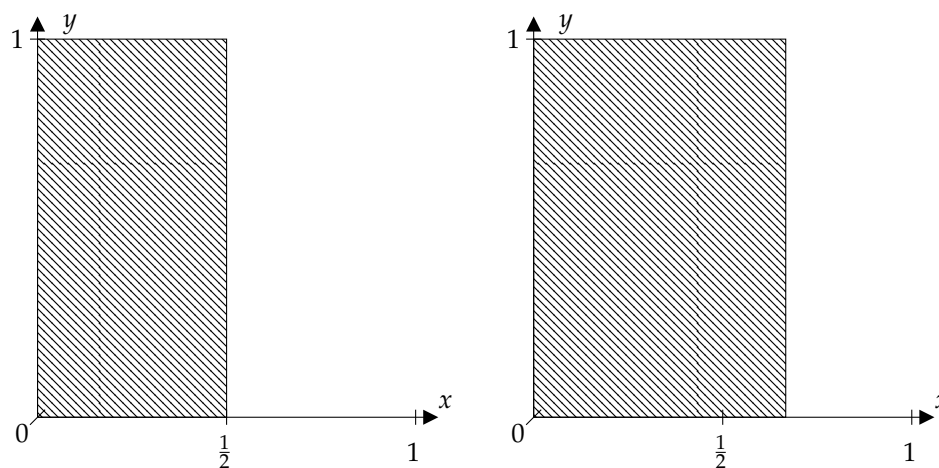
In this paper, we consider general sets $R \subseteq \mathbb{R}^2$ that cannot necessarily be described by functions and inequalities and establish some conditions under which the entropy can be determined using those sets. As in [6], the discussion also includes a generalization of the sets E_n given by (2) and is conducted in a multidimensional framework. In particular, the results give insights as to why the basic ordinal approach and generalizations are working.

It is instructive to discuss the partition of $\mathbb{R}^2$ into R and $\mathbb{R}^2 \setminus R$ from the viewpoint of symbolic dynamics. In contrast to classical symbolization approaches with symbolizing only in the range of single "measurements" x, the symbolization of pairs (x, y) via the partition $\{R, \mathbb{R}^2 \setminus R\}$ also regards some kind of link between x and y if R lies "diagonal" in a certain sense. We will discuss this constellation, which explains the success of ordinal patterns in a wider context, more precisely in Section 5.

A completely different constellation is given for the sets R shown in Figure 2. Here, R is obtained as a half-plane from a "horizontal" division of $\mathbb{R}^2$. If, for example, $\Omega = [0, 1[$, $\mathscr{A}$ is the Borel σ-algebra and μ the Lebesgue measure on Ω, and if T is the tent map on Ω, meaning that:

$$T(\omega) = \begin{cases} 2\omega & \text{if } 0 \leq \omega < \frac{1}{2}, \\ 2 - 2\omega & \text{if } \frac{1}{2} \leq \omega < 1, \end{cases} \tag{5}$$

and X is the identity map, then the location of the horizontal cut is substantial.

(a) $R = \{(x, y) \in \mathbb{R}^2 \mid x \leq 1/2\}$ **(b)** $R = \{(x, y) \in \mathbb{R}^2 \mid x \leq 2/3\}$

Figure 2. "Non-diagonal" discriminating relations.

On the one hand, $R = \{(x, y) \in \mathbb{R}^2 \mid x \leq 2/3\}$ (Figure 2b) does not discriminate enough to obtain the KS entropy of the given system, and on the other hand, there is enough discrimination by $R = \{(x, y) \in \mathbb{R}^2 \mid x \leq 1/2\}$ (Figure 2a) due to the fact that $\{[0, \frac{1}{2}[, [\frac{1}{2}, 1[\}$ is a generating partition for T. In the situation considered, there is no additional information given by the measurements (x, y) relative to measurement x, hence R provides nothing more than a classical symbolization. For a detailed discussion of these facts, see [6].

The rest of this paper is organized as follows. Section 2 provides the notions and concepts being necessary for formulating the main statement of this paper in Section 3. This statement is rather abstract and general and has to be considered in relation to some special

cases discussed in Section 4 and making our ideas and findings transparent. Section 5 is devoted to the proof of our main statement.

2. Preliminaries

Throughout this paper, $(\Omega, \mathscr{A}, \mu, T)$ will be a measure-preserving dynamical system.

2.1. Some Notions

We will write $\mathscr{B} = \mathscr{B}(\mathbb{R})$ or $\mathscr{B}(\mathbb{R}^d)$ for the Borel σ-algebra on $\mathbb{R}$ or $\mathbb{R}^d; d \in \mathbb{N}$, respectively. Given a random variable $X : \Omega \to \mathbb{R}$, by μ_X we denote the push-forward measure of μ with regard to X, i.e., $\mu_X(A) := \mu(X^{-1}(A))$ for all $A \in \mathscr{B}(\mathbb{R})$. The measure $\mu_X \times \mu_X = \mu_X^2$ is the product measure, i.e., $\mu_X^2(A \times B) = \mu_X(A)\mu_X(B)$ for all $A, B \in \mathscr{B}(\mathbb{R})$.

For some Borel set $R \in \mathscr{B}(\mathbb{R}^2)$, we define the function $f_X^R : \mathbb{R} \to [0, 1]$ by

$$f_X^R(x) := \mu(\{\omega \in \Omega \mid (x, X(\omega)) \in R\}). \tag{6}$$

If it is clear from the context which set $R \in \mathscr{B}(\mathbb{R}^2)$ is considered, we simply write f_X instead of f_X^R. The function f_X can be represented as the integral:

$$f_X(x) = \int 1_R(x, y) \, d\mu_X(y).$$

Since 1_R is integrable with regard to μ_X^2, f_X is integrable and therefore, also measurable by Fubini's Theorem.

The complement $\mathbb{R}^2 \setminus R$ of a set $R \subseteq \mathbb{R}^2$ will be denoted by R^c. The notation ∂R will be used for the boundary of a set R, i.e., the closure of R without its interior.

2.2. Entropy

The *Shannon entropy* of a finite partition $\mathcal{P} \subset \mathscr{A}$ of Ω is defined as

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

The refinement of two partitions $\mathcal{P}, \mathcal{Q} \subset \mathscr{A}$ of Ω is given by

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

For a finite collection of partitions $\mathcal{P}_i \subset \mathscr{A}, i \in \{1, 2, \dots, n\}$ of Ω, one analogously defines:

$$\bigvee_{i=1}^{n} \mathcal{P}_i := \left\{ \bigcap_{i=1}^{n} P_i \neq \varnothing \mid P_i \in \mathcal{P}_i \text{ for all } i \in \{1, 2, \dots, n\} \right\}.$$

The *entropy rate* of a finite partition $\mathcal{P} \subset \mathscr{A}$ of Ω is defined as

$$h(T, \mathcal{P}) := \lim_{n \to \infty} \frac{1}{n} H\left(\bigvee_{t=0}^{n-1} T^{-t}(\mathcal{P}) \right),$$

where $T^{-t}(\mathcal{P}) = \{T^{-t}(P) \mid P \in \mathcal{P}\}$. For the existence of the limit in the formula, see, e.g., [7]. We are interested in determining the *Kolmogorov–Sinai entropy* of a system, which is defined as

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

where the supremum is taken over all finite partitions of Ω in $\mathscr{A}$.

Note that the Kolmogorov–Sinai entropy serves as the central complexity measure for dynamical systems and can be considered as a reference for other complexity measures, including in data analysis. Roughly speaking, it measures the mean information obtained by each iteration step. Since the Kolmogorov–Sinai entropy is the supremum of the entropy

rates of all finite partitions, its determination and its estimation in a practical context are not easy and it is of some interest to find natural finite partitions for which the entropy rate is near the Kolmogorov–Sinai entropy. This is also a motivation for considering ordinal patterns and its generalization in this paper.

2.3. σ-Algebras

Given a family of sets $A_i \in \mathscr{A}, i \in I$, by $\sigma(A_i \mid i \in I)$ we denote the smallest σ-algebra containing all sets A_i. Analogously, for a family of partitions $\mathcal{P}_i \subset \mathscr{A}, i \in I$ of Ω, we define $\sigma(\mathcal{P}_i \mid i \in I)$ as the smallest σ-algebra containing all partitions $\mathcal{P}_i$ as subsets. Given d-dimensional random vectors $(Y_{i,1}, Y_{i,2}, \ldots, Y_{i,d}) : \Omega \to \mathbb{R}^d, i \in I$, we define:

$$\sigma((Y_{i,1}, Y_{i,2}, \ldots, Y_{i,d}) \mid i \in I) := \sigma((Y_{i,1}, Y_{i,2}, \ldots, Y_{i,d})^{-1}(B) \mid i \in I, B \in \mathscr{B}(\mathbb{R}^d))$$

as the smallest σ-algebra containing all preimages of Borel sets. When comparing two σ-algebras $\mathscr{A}_1, \mathscr{A}_2 \subseteq \mathscr{A}$, we ignore sets with measure 0, i.e., we write:

$$\mathscr{A}_1 \subseteq_\mu \mathscr{A}_2$$

if for all $A_1 \in \mathscr{A}_1$ there exists an $A_2 \in \mathscr{A}_2$ with $\mu((A_1 \setminus A_2) \cup (A_2 \setminus A_1)) = 0$. In this sense, $\mathscr{A}_1 =_\mu \mathscr{A}_2$ means that $\mathscr{A}_1 \subseteq_\mu \mathscr{A}_2$ and $\mathscr{A}_2 \subseteq_\mu \mathscr{A}_1$.

3. The Main Statement

Recall that a measure-preserving dynamical system $(\Omega, \mathscr{A}, \mu, T)$ is *ergodic* if for each $B \in \mathscr{A}$ with $T^{-1}(B) = B$ it holds $\mu(B) \in \{0, 1\}$, meaning that the system does not divide into proper parts.

Referring to Section 1, we give some preparation for stating our main result. Recall that for defining (generalized) ordinal patterns it was basic to know whether $(X(T^s(\omega)), X(T^s(\omega))) \in R$ or $(X(T^s(\omega)), X(T^s(\omega))) \in \mathbb{R}^2 \setminus R$ for $\omega \in \Omega$ and a random variable X on Ω, where "time" pairs (s, t) were taken from the sets E_n (see (2)). In order to also allow reducing the number of necessary "comparisons", we relax the definition of the sets E_n leading to the following concept.

Definition 2. *We call a sequence $(E_n)_{n \in \mathbb{N}}$ with $E_1 \subseteq E_2 \subseteq E_3 \subseteq \ldots \subseteq \mathbb{N}_0^2$ timing, if E_n contains finitely many elements and if there exists a sequence $(a_n)_{n \in \mathbb{N}} \in \mathbb{N}_0^\mathbb{N}$ with:*

$$\limsup_{N \to \infty} \frac{\#\{n \in \{1, 2, \ldots, N\} \mid (a_n, a_n + n) \in \bigcup_{k=1}^\infty E_k\}}{N} = 1. \tag{7}$$

Formula (7) roughly says that nearly each "temporal" distance is available for "comparisons". It guarantees that enough time-pairs are considered to not have any loss of information contained in the "thinned out" generalized ordinal patterns relative to the "full" generalized ordinal patterns.

Remark 1. *In the first paper on generalized ordinal patterns ([6]), a timing $(E_n)_{n \in \mathbb{N}}$ was differently defined: the authors of that paper called a sequence of finite sets $E_1 \subseteq E_2 \subseteq E_3 \subseteq \ldots \subseteq \mathbb{N}_0^2$ timing, if there exists an increasing sequence $(a_n)_{n \in \mathbb{N}}$ such that for all $n \in \mathbb{N}$:*

(i) $E_n \subseteq \{a_0, a_1, \ldots, a_n\}^2$,

(ii)   for all $s \in \{a_0, a_1, \ldots, a_n\}$, there exists a $t \in \{a_0, a_1, \ldots, a_n\}$ with $s \neq t$ and $(s, t) \in E_n$,

(iii) $(id, T^n)^{-1}(\mathcal{R}_i) \subseteq_\mu \sigma\left(\bigvee_{(s,t) \in E_k}(T^s, T^t)^{-1}(\mathcal{R}_i) \mid k \in \mathbb{N}\right)$ for all $i \in \{1, 2, \ldots, d\}$

hold true. Note that the last condition does not only depend on the timing $(E_n)_{n \in \mathbb{N}}$ but also on T and $X = (X_1, X_2, \ldots, X_d)$. Instead of those three conditions, we instead simply require that almost all differences can be found in the timing.

Given a random vector $(X_1, X_2, \ldots, X_d)$ and $R \in \mathscr{B}(\mathbb{R}^2)$, we define the partition:

$$\mathcal{R}_i := (X_i, X_i)^{-1}(\{R, R^c\}) = \{(X_i, X_i)^{-1}(R), (X_i, X_i)^{-1}(R^c)\}$$

for all $i \in \{1, 2, \ldots, d\}$, which is equal to:

$$\mathcal{R}_i = \{\{(\omega_1, \omega_2) \in \Omega^2 \mid (X_i(\omega_1), X_i(\omega_2)) \in R\},$$
$$\{(\omega_1, \omega_2) \in \Omega^2 \mid (X_i(\omega_1), X_i(\omega_2)) \notin R\}\}.$$

Then, for E_n as given in (2) the partition $\bigvee_{(s,t) \in E_k} (T^s, T^t)^{-1}(\mathcal{R}_i)$ is no more than the partition of generalized ordinal patterns with respect to X_i defined in Section 1 and $\bigvee_{i=1}^{d} \bigvee_{(s,t) \in E_k} (T^s, T^t)^{-1}(\mathcal{R}_i)$ can be considered as the partition of generalized ordinal patterns with respect to $(X_1, X_2, \ldots, X_d)$.

The proof of the following main theorem of the paper is given in Section 5.

Theorem 1. *Let $(\Omega, \mathscr{A}, \mu, T)$ be an ergodic measure-preserving dynamical system, $\mathbf{X} = (X_1, X_2, \ldots, X_d) : \Omega \to \mathbb{R}^d$ be a random vector, $R \in \mathscr{B}(\mathbb{R}^2)$ be a discriminating relation and $(E_n)_{n \in \mathbb{N}}$ be a timing. Assume that the following conditions are valid:*

There exists a countable set $C \subseteq \mathbb{R}^2$ with:

$$\mu_{X_i}^2 (\partial R \setminus C) = 0 \text{ for all } i \in \{1, 2, \ldots, d\}. \tag{8}$$

There exists a random variable $Y : \Omega \to \mathbb{R}$ with:

$$\#Y(\Omega) < \infty \text{ and } \mathscr{A} =_\mu \sigma\Big((f_{X_i}^R \circ X_i \circ T^t, Y) \mid t \in \mathbb{N}_0, i \in \{1, 2, \ldots, d\}\Big). \tag{9}$$

Then:

$$h(T) = \lim_{k \to \infty} h\left(T, \bigvee_{i=1}^{d} \bigvee_{(s,t) \in E_k} (T^s, T^t)^{-1}(\mathcal{R}_i)\right). \tag{10}$$

holds true.

At first glance, conditions (8) and (9), being sufficient for (10), are looking very special. The considerations in the following section will, however, elucidate their role and show that they are relatively general. Roughly speaking, (8) says that the distribution of pairs of "independent measurements" with respect to X_i is discrete on the boundary of R. Condition (9) is an orbit separation condition based on the involved "measurements" and the functions $f_{X_i}^R$. In general:

$$\mathscr{A} \supseteq_\mu \sigma\Big((f_{X_i}^R \circ X_i \circ T^t, Y) \mid t \in \mathbb{N}_0, i \in \{1, 2, \ldots, d\}\Big)$$

holds true because all functions involved in (9) are $\mathscr{A}$ measurable. Therefore, (9) is equivalent to

$$\mathscr{A} \subseteq_\mu \sigma\Big((f_{X_i}^R \circ X_i \circ T^t, Y) \mid t \in \mathbb{N}_0, i \in \{1, 2, \ldots, d\}\Big)$$

The inclusion of the random variable Y provides some further separation and allows the above inclusion to hold true for a wider class of dynamical systems than, for example, the ones considered in [6]. In the case that Y is constant, it can also be omitted. In theory, Y should be chosen to take different values on those sets on which $f_{X_i}^R \circ X_i \circ T^t$ takes the same values for $i \in \{1, 2, \ldots, d\}$ and $t \in \mathbb{N}_0$. In practice, the fact that such a random variable Y exists is sufficient and Y does not need to be explicitly specified. An example is given in Section 4.5.

4. Special Cases

In the following, we discuss some special situations where the assumptions of Theorem 1, i.e., (8) and (9), are satisfied. Lemma 2 provides an easy-to-check condition, that of when (8) holds true. It is more difficult to see, when the condition (9) is satisfied. Roughly speaking, this condition is fulfilled if $X = (X_1, X_2, \ldots, X_d)$ together with Y can uniquely describe the outcomes of the whole dynamical system and applying f_{X_i} to the results of X is, in some sense, "reversible" for all $i \in \{1, 2, \ldots, d\}$. In other words, $X = (X_1, X_2, \ldots, X_d)$ together with Y preserve the information of the system and there is no information loss for the symbolization. The first means that:

$$\mathscr{A} =_\mu \sigma\big((X_i \circ T^t, Y) \mid t \in \mathbb{N}_0, i \in \{1, 2, \ldots, d\}\big),$$

which obviously follows from

$$\mathscr{A} =_\mu \sigma\Big((f_{X_i}^R \circ X_i \circ T^t, Y) \mid t \in \mathbb{N}_0, i \in \{1, 2, \ldots, d\}\Big).$$

To describe the range of outcomes of the random variables X on a probability space $(\Omega, \mathscr{A}, \mu)$, we will use its cumulative distribution functions $F_X : \mathbb{R} \to [0, 1]$ defined by

$$F_X(x) := \mu_X(] - \infty, x]).$$

When applying the cumulative distribution functions F_X to the outcomes X of a system, we do not lose any essential information about the system, according to the following lemma. This lemma is a simple modification of Lemma A.3 in [6].

Lemma 1. *Let $(\Omega, \mathscr{A}, \mu)$ be a probability space, $X : \Omega \to \mathbb{R}$ be a random variable and $g : \mathbb{R} \to \mathbb{R}$ be a $\mathscr{B} - \mathscr{B}$ measurable function which maps Borel sets to Borel sets and satisfies the following property:*

$$\text{For all } x \in \mathbb{R} \text{ it holds } \mu(X^{-1}(g^{-1}(g(] - \infty, x]))) \setminus] - \infty, x])) = 0 \qquad (11)$$

Then, $\sigma(g \circ X) =_\mu \sigma(X)$. In particular, $\sigma(g \circ X) =_\mu \sigma(X)$ if $g = F_X$ or g is injective on $X(\Omega)$.

Condition (11) in the above Lemma is a slightly weaker condition on g than injectivity. If g is injective, then $g^{-1}(g(] - \infty, x])) =] - \infty, x]$ will hold true for all $x \in \mathbb{R}$ and condition (11) will be satisfied. More general, condition (11) can still be true if g is not necessarily injective but if all sets on which g is not injective, which are given by $X^{-1}(g^{-1}(g(] - \infty, x]))) \setminus] - \infty, x])$ for all $x \in \mathbb{R}$, have measure 0. For example, this is true if g is equal to the cumulative distribution function.

4.1. On the Boundary of R

The condition (8) in Theorem 1, that the boundary of R apart from countably many points has measure 0, holds true for all "simple" sets R. In the following lemma, we specify what we mean by "simple".

Lemma 2. *Let $(\Omega, \mathscr{A}, \mu)$ be a probability space, $(X_1, X_2, \ldots, X_d)$ be a random vector and $R \in \mathscr{B}(\mathbb{R}^2)$. If, for all $i \in \{1, 2, \ldots, d\}$:*

$$\partial R \cap (\{x\} \times \mathbb{R}) \text{ is countable for } \mu_{X_i}\text{-almost all } x \in \mathbb{R}$$

$$\text{or} \quad \partial R \cap (\mathbb{R} \times \{y\}) \text{ is countable for } \mu_{X_i}\text{-almost all } y \in \mathbb{R},$$

then R satisfies (8), i.e., there exists a countable set $C \subseteq \mathbb{R}^2$ with:

$$\mu_{X_i}^2(\partial R \setminus C) = 0 \quad \text{for all } i \in \{1, 2, \ldots, d\}.$$

Proof. Consider the sets:

$$A_i := \{x \in \mathbb{R} \mid \mu_{X_i}(\{x\}) > 0\}$$

for $i \in \{1, 2, \ldots, d\}$, which, obviously, are countable. Set:

$$C := \bigcup_{i=1}^{d} A_i \times A_i.$$

Let $i \in \{1, 2, \ldots, d\}$. If $\{y \in \mathbb{R} \mid (x, y) \in \partial R\}$ is countable for μ_{X_i}-almost all $x \in \mathbb{R}$, Fubini's theorem implies:

$$
\begin{aligned}
\mu_{X_i}^2(\partial R \setminus C) &= \int \int 1_{\partial R \setminus C}(x, y) \, d\mu_{X_i}(y) \, d\mu_{X_i}(x) \\
&= \int \mu_{X_i}(\{y \in \mathbb{R} \mid (x, y) \in \partial R \setminus C\}) \, d\mu_{X_i}(x) \\
&= \int \sum_{\substack{y \in \mathbb{R}: \\ (x,y) \in \partial R \setminus C}} \mu_{X_i}(\{y\}) \, d\mu_{X_i}(x) \\
&= \int 0 \, d\mu_{X_i}(x) = 0.
\end{aligned}
$$

Analogously, one can show the same if $\{x \in \mathbb{R} \mid (x, y) \in \partial R\}$ is countable for μ_{X_i}-almost all $y \in \mathbb{R}$. $\square$

Remark 2. *The patterns visualized in Figure 1 could also be defined on the whole real axis instead of a bounded interval by, for example, applying the transformation $\varphi(x) = \tan(\pi(x - 1/2))$. Then, $\widetilde{R} = (\varphi \times \varphi)(R)$ is a pattern defined on $\mathbb{R}^2$ if R is defined on $[0, 1[^2$.*

In the following three subsections, Y is assumed to be constant, and hence can be omitted.

4.2. Basic Ordinal Patterns

If:

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

(see Figure 1a), then $f_{X_i}^R$ is just the distribution function of μ_{X_i}, i.e., $f_{X_i}^R = F_{X_i}$. Since $\partial R \cap \{x\} \times \mathbb{R} = \{(x, x)\}$ is finite for all $x \in \mathbb{R}$, (8) holds true by Lemma 2. According to Lemma 1, one has:

$$\sigma(X_i \circ T^t) \subseteq_\mu \sigma(F_{X_i} \circ X_i \circ T^t) \tag{12}$$

for all $t \in \mathbb{N}_0$ and $i = 1, 2, \ldots, n$. Therefore:

$$
\begin{aligned}
\mathscr{A} &=_\mu \sigma\left((f_{X_i}^R \circ X_i \circ T^t) \mid t \in \mathbb{N}_0, i \in \{1, 2, \ldots, d\}\right) \\
&= \sigma\left((F_{X_i} \circ X_i \circ T^t) \mid t \in \mathbb{N}_0, i \in \{1, 2, \ldots, d\}\right)
\end{aligned}
\tag{13}
$$

is equivalent to:

$$\mathscr{A} =_\mu \sigma\left((X_i \circ T^t) \mid t \in \mathbb{N}_0, i \in \{1, 2, \ldots, d\}\right). \tag{14}$$

By Theorem 1, for ergodic systems, condition (14) implies (10). A more general statement also includes a large class of non-ergodic systems which was shown in [4]. Condition (14) is, for example, satisfied if $\Omega \in \mathscr{B}(\mathbb{R}^d)$ and X_i is the projection on the i-th coordinate for all $i \in \{1, 2, \ldots, d\}$, or if Ω is a compact Hausdorff space and $X = (X_1, X_2, \ldots, X_d)$ is injective and continuous. One can also use Taken's theorem to argue that the set of maps $X : \Omega \to \mathbb{R}^d$ that satisfy (14) is large in a certain topological sense. For both, see Keller [8].

4.3. Patterns Defined by "Injective" Functions

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)$ be a random vector and consider now:

$$R = \{(x, y) \in \mathbb{R} \mid g(x) \geq y\} \tag{15}$$

for a $\mathscr{B} - \mathscr{B}$ measurable function $g : \mathbb{R} \to \mathbb{R}$ (see Figure 1b). Since $\partial R \cap (\{x\} \times \mathbb{R}) = \{(x, g(x))\}$ is finite for all $x \in \mathbb{R}$, (8) holds true by Lemma 2. Moreover, one easily sees that $f_{X_i}^R = F_{X_i} \circ g$.

Now, suppose that:

$$\sigma(F_{X_i}) \subseteq \sigma(F_{X_i} \circ g) \tag{16}$$

holds true for all $i \in \{1, 2, \ldots, d\}$. This directly yields:

$$\sigma(F_{X_i} \circ X_i \circ T^t) \subseteq \sigma(F_{X_i} \circ g \circ X_i \circ T^t)$$

for all $i \in \{1, 2, \ldots, d\}$ and $t \in \mathbb{N}$. Remember that $\sigma(X_i \circ T^t) \subseteq_\mu \sigma(F_{X_i} \circ X_i \circ T^t)$ holds true according to Lemma 1. Thus, (14) and (16) imply (10). When considering basic ordinal patterns in Section 4.2, we stated some conditions under which (14) holds true. It remains to consider when (16) is satisfied.

Assume that g maps Borel sets to Borel sets and is injective. This implies:

$$\sigma(g \circ X_i \circ T^t) = \sigma(X_i \circ T^t)$$

for all $t \in \mathbb{N}_0$ and $i \in \{1, 2, \ldots, d\}$. Now, suppose that:

$$\mathscr{A} = \sigma\big(F_{X_i} \circ X_i \circ T^t \mid t \in \mathbb{N}_0, i \in \{1, 2, \ldots, d\}\big).$$

holds true. This would then imply (16). However, the above equation only holds true μ-almost surely (see (13)). This can be a problem when applying the function g because there could exist sets $B \in \mathscr{B}$ with $\mu_{X_i}(B) = 0$ but $\mu_{X_i}(g(B)) > 0$. Additionally, we therefore need to require that $\mu_{X_i}(g^{-1}(B)) = 0$ implies $\mu_{X_i}(B) = 0$ for all $B \in \mathscr{B}$.

Theorem 1 then provides the following statement:

Corollary 1. *Let* $(\Omega, \mathscr{A}, \mu, T)$ *be an ergodic measure-preserving dynamical system,* $\boldsymbol{X} = (X_1, X_2, \ldots, X_d) : \Omega \to \mathbb{R}^d$ *be a random vector and* $(E_n)_{n \in \mathbb{N}}$ *be a timing. Let further* $g : \mathbb{R} \to \mathbb{R}$ *be a* $\mathscr{B} - \mathscr{B}$ *measurable function which maps Borel sets to Borel sets, is injective on* $X_i(\Omega)$ *and satisfies* $\mu_{X_i}(g^{-1}(B)) = 0 \Rightarrow \mu_{X_i}(B) = 0$ *for all* $B \in \mathscr{B}$ *and* $i \in \{1, 2, \ldots, d\}$. *Let* $R = \{(x, y) \in \mathbb{R}^2 \mid g(x) \geq y\}$.

Then, (14) *implies* (10). *Moreover,* (10) *holds true if* $\Omega \in \mathscr{B}(\mathbb{R}^d)$ *and* X_i *is the projection on the i-th coordinate for all* $i \in \{1, 2, \ldots, d\}$ *or if* Ω *is a compact Hausdorff space and* $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)$ *is injective and continuous.*

Note that the statements in Corollary 1, in principle, were shown in [6]. The case of basic ordinal patterns is included by $g(x) = x$ for all $x \in \mathbb{R}$.

4.4. Patterns Defined by "Surjective" Functions

Swapping coordinates in (15) yields the set:

$$R = \{(x, y) \in \mathbb{R} \mid x < g(y)\}$$

(see Figure 1c) with (8) following from Lemma 2 and with:

$$f_{X_i}^R(x) = \mu(\{\omega \in \Omega \mid (x, X_i(\omega)) \in R\}) = \mu(\{\omega \in \Omega \mid x < g(X_i(\omega))\})$$
$$= \mu(\{\omega \in \Omega \mid g(X_i(\omega)) \in \,]x, \infty[\}) = 1 - F_{g \circ X_i}(x).$$

Corollary 2. *Let* $(\Omega, \mathscr{A}, \mu, T)$ *be an ergodic measure-preserving dynamical system,* $\boldsymbol{X} = (X_1, X_2, \ldots, X_d) : \Omega \to \mathbb{R}^d$ *be a random vector and* $(E_n)_{n \in \mathbb{N}}$ *be a timing. Let further* $g : \mathbb{R} \to \mathbb{R}$ *be a* $\mathscr{B} - \mathscr{B}$ *measurable function and let* $R = \{(x, y) \in \mathbb{R}^2 \mid x < g(y)\}$. *Then, the following holds:*

(i)   *If* $F_{g \circ X_i}$ *is injective on* $X_i(\Omega)$ *for* $i \in \{1, 2, \ldots, d\}$, (14) *implies* (10).

(ii)   *If* $\Omega \in \mathscr{B}(\mathbb{R}^d)$ *and* X_i *is the projection on the i-th coordinate for all* $i \in \{1, 2, \ldots, d\}$ *or if* Ω *is a compact Hausdorff space and* $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)$ *is injective and continuous, if further* $\mu(U) > 0$ *for every non-empty open set* $U \subseteq \Omega$, *g is continuous and* $X_i(\Omega) \subseteq g(X_i(\Omega))$, *then* (10) *is valid in each of the following two cases:*

    (1)   *For each* $i \in \{1, 2, \ldots, d\}$ *and all* $x_1, x_2 \in X_i(\Omega)$ *with* $x_1 < x_2$, *there exists some* $y \in X_i(\Omega)$ *with* $x_1 < y < x_2$,

    (2) Ω *is connected.*

Proof. (i): If the above assumptions are satisfied and $F_{g \circ X_i}$ is injective on $X_i(\Omega)$ for all $i \in \{1, 2, \ldots, d\}$, then by Lemma 1 it holds that $\sigma(X_i \circ T^t) \subseteq_\mu \sigma(F_{g \circ X_i} \circ X_i \circ T^t)$ for all $t \in \mathbb{N}_0$ and $i \in \{1, 2, \ldots, d\}$. This implies (9), hence, by Theorem 1 the statement (10).

(ii): Given the assumptions of (ii), we have to show that $F_{g \circ X_i}$ is injective on $X_i(\Omega)$ for all $i \in \{1, 2, \ldots, d\}$. If Ω is connected, then (1) is obviously satisfied. We can thus start from (1). Take $x_1, x_2 \in X_i(\Omega)$ with $x_1 < x_2$. Then, $g^{-1}(]x_1, x_2[)$ is non-empty and because $g \circ X_i$ is continuous, $X_i^{-1}(g^{-1}(]x_1, x_2[))$ contains a non-empty open set. This implies that $F_{g \circ X_i}(x_1) < F_{g \circ X_i}(x_2)$. because every non-empty open set was assumed to have a strictly positive measure. $\square$

Notice that, unlike in (15), it is not necessary that g is one-to-one.

4.5. Piecewise Patterns

The previous subsection illustrates that (9) is fulfilled if, roughly speaking, $(X_1, X_2, \ldots, X_d)$ preserves all information and if $f_{X_i}^R$ is a μ_{X_i} almost surely invertible function for all $i \in \{1, 2, \ldots, d\}$. The finite-valued random variable Y in (9) can be used to weaken the condition of invertibility in the sense that only piecewise invertibility is needed where the different pieces are induced by the random variable Y.

For $\Omega = [0, 1[$ and an absolutely continuous measure μ, one could, for example, consider:

$$R_{\text{circles}} = \{(x, y) \in \Omega^2 \mid \|(kx \mod 1, ky \mod 1) - (0.5, 0.5)\|_2 \leq 0.5\} \qquad (17)$$

for any $k \in \mathbb{N}$, as shown for $k = 5$ in Figure 1d. The set R satisfies condition (9) with $Y(\omega) = i$ for $\omega \in [(i-1)/(2k), i/(2k)[$ and $i \in \{1, 2, \ldots, 2k\}$. The set R is a pattern with k^2 circles of diameter $1/k$ distributed in $[0, 1]^2$ on a square grid.

4.6. A Remark on the Work of Amigó et al.

Consider the discriminating relation:

$$R_k = \{(x, y) \in \mathbb{R}^2 \mid \lceil k \cdot x \rceil \geq \lceil k \cdot y \rceil\}$$

shown in Figure 3 for $k \in \mathbb{N}$. Assume for simplicity that the dynamical system is defined on $\Omega = [0, 1[$ and that X is the identity map *id*. It is easy to see that:

$$\sigma\left(\bigvee_{t=0}^{n} T^{-t}(\mathcal{P}_k) \mid n \in \mathbb{N} \right) =_\mu \sigma\left(f_{id}^{R_k} \circ T^t \mid t \in \mathbb{N}_0 \right)$$

holds true, where $\mathcal{P}_k := \{[(i-1)/k, i/k[\}_{i=1}^{k}$. Therefore, (9) in Theorem 1 holds true if $\mathcal{P}_k$ is a generating partition.
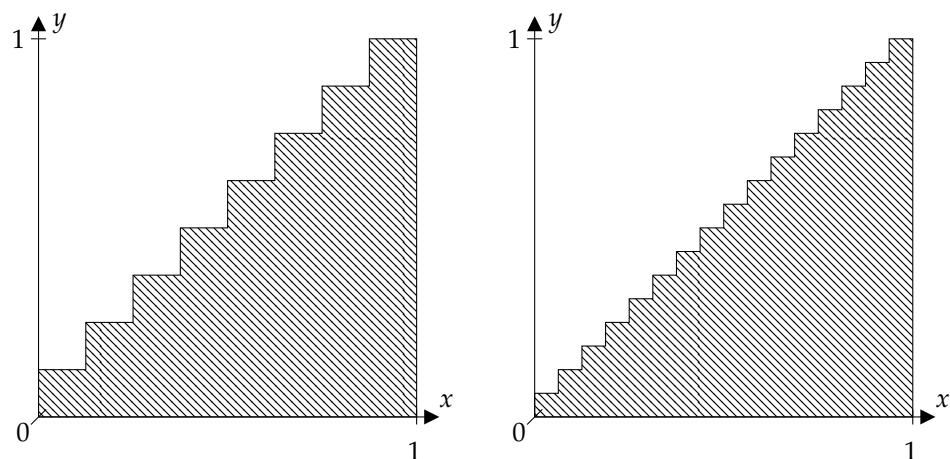
Figure 3. $R_k := \{(x,y) \in \mathbb{R}^2 \mid \lfloor k \cdot x \rfloor \geq \lfloor k \cdot y \rfloor\}$ for $k = 8$ (left side) and $k = 16$ (right side), only shown in $[0.1]^2$

Additionally, one could consider the quantity:

$$\lim_{k\to\infty} \liminf_{n\to\infty} \frac{1}{n} H\left(\bigvee_{s=0}^{n-1} \bigvee_{t=s+1}^{n-1} (T^s, T^t)^{-1}(\{R_k, \mathbb{R}^2 \setminus R_k\})\right) \tag{18}$$

which was introduced by Amigó et. al. [9]. They used finite-valued random variables to quantize the dynamical system into k parts and considered the ordinal patterns of the quantized systems while we directly apply the quantization to the discriminating relation. Both approaches only differ in their notation. They showed in their paper that the limit in (18) is equal to the Kolmogorov–Sinai entropy.

5. Proof of the Main Statement

We first recall some definitions and statements related to partitions and the conditional entropy. For two partitions $\mathcal{P}, \mathcal{Q} \subset \mathscr{A}$ of Ω, the *conditional entropy* is defined as

$$H(\mathcal{P}|\mathcal{Q}) := H(\mathcal{P} \vee \mathcal{Q}) - H(\mathcal{Q}).$$

Roughly speaking, the conditional entropy $H(\mathcal{P}|\mathcal{Q})$ describes how much uncertainty is left in the outcomes described by the sets given in $\mathcal{P}$ if one already has information about the outcomes described by the sets given in $\mathcal{Q}$. For example, if $\mathcal{P} = \mathcal{Q}$, then $H(\mathcal{P}|\mathcal{Q}) = 0$. However, if $\mathcal{P}$ and $\mathcal{Q}$ are independent, meaning that $\mu(P \cap Q) = \mu(P) \cdot \mu(Q)$ for all $P \in \mathcal{P}$ and $Q \in \mathcal{Q}$, and $H(\mathcal{P}|\mathcal{Q}) = H(\mathcal{Q})$.

Without explicitly referencing them, we will use the following properties of the conditional entropy:

(i) $H(T^{-1}(\mathcal{P})|T^{-1}(\mathcal{Q})) = H(\mathcal{P}|\mathcal{Q})$,
(ii) $H(\bigvee_{i=1}^n \mathcal{P}_i|\mathcal{Q}) \leq \sum_{i=1}^n H(\mathcal{P}_i|\mathcal{Q})$,
(iii) $H(\mathcal{P}|\mathcal{Q}_1 \vee \mathcal{Q}_2) \leq H(\mathcal{P}|\mathcal{Q}_1)$.

See, for examples, [7] for proofs.

A sequence of partitions $(P_i)_{i\in\mathbb{N}}$ in $\mathscr{A}$ of Ω is said to be generating (the σ-algebra $\mathscr{A}$), if $\sigma(P_i) \subseteq \sigma(\mathcal{P}_{i+1})$ for all $i \in \mathbb{N}$ and:

$$\sigma(\mathcal{P}_i \mid i \in \mathbb{N}) =_\mu \mathscr{A}$$

holds true. As a consequence of this property:

$$\lim_{n\to\infty} H(\mathcal{P} \mid \mathcal{P}_n) = 0 \tag{19}$$

holds true for all partitions $\mathcal{P} \subset \mathscr{A}$ of Ω. Using the properties of the conditional entropy implies:

$$h(T) = \lim_{n \to \infty} h(T, \mathcal{P}_n).$$

For $N \in \mathbb{N}$:

$$\mathcal{U}_N = \left\{ \left[(i-1)/2^N, i/2^N \right[\mid i \in \{1, 2, \ldots, 2^N - 1\} \right\} \cup \left\{ \left[(2^N - 1)/2^N, 1 \right] \right\}$$

will denote the partition of $[0, 1]$ in 2^N equally sized intervals.

We start the proof of Theorem 1 with two basic lemmata.

Lemma 3. *Let $(\Omega, \mathscr{A}, \mu, T)$ be a measure-preserving dynamical system, $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ be a random vector and Y be a random variable satisfying* (9). *Then, there exists some constant $c \in \mathbb{R}$ with:*

$$h(T^m) \leq \lim_{N \to \infty} h\left(T^m, \bigvee_{i=1}^{d} \bigvee_{t=0}^{m-1} T^{-t} \left(X_i^{-1} \left(f_{X_i}^{-1}(\mathcal{U}_N) \right) \right) \right) + c.$$

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

Proof. Fix $m \in \mathbb{N}$. Set:

$$\mathcal{M} := \left\{ Y^{-1}(y) \mid y \in Y(\Omega) \right\}$$

Since Y was assumed to attain only a finite number of different values, $\mathcal{M}$ is a finite partition of Ω. Because the Borel σ-algebra of $[0, 1]$ is generated by the partitions $\mathcal{U}_N$ and due to (9), we have:

$$\begin{aligned}
\mathscr{A} =_\mu \ & \sigma\left((f_{X_i} \circ X_i \circ T^t, Y) \mid t \in \mathbb{N}_0, i \in \{1, 2, \ldots, d\} \right) \\
= \ & \sigma\left(T^{-t} \left(X_i^{-1} \left(f_{X_i}^{-1}(\mathcal{U}_N) \right) \right) \vee \mathcal{M} \mid N \in \mathbb{N}, t \in \mathbb{N}_0, i \in \{1, 2, \ldots, d\} \right).
\end{aligned}$$

Thus, for any $\varepsilon > 0$ and any finite partition $\mathcal{P} \subset \mathscr{A}$ of Ω, there exists an $N_\varepsilon \in \mathbb{N}$ and a $t_\varepsilon \in \mathbb{N}$ with:

$$\begin{aligned}
h(T^m, \mathcal{P}) \leq \ & h\left(T^m, \bigvee_{i=1}^{d} \bigvee_{t=0}^{t_\varepsilon - 1} T^{-t} \left(X_i^{-1} \left(f_{X_i}^{-1}(\mathcal{U}_N) \right) \right) \vee \mathcal{M} \right) + \varepsilon \\
\leq \ & h\left(T^m, \bigvee_{i=1}^{d} \bigvee_{t=0}^{t_\varepsilon - 1} T^{-t} \left(X_i^{-1} \left(f_{X_i}^{-1}(\mathcal{U}_N) \right) \right) \right) + h(T^m, \mathcal{M}) + \varepsilon \\
\leq \ & h\left(T^m, \bigvee_{i=1}^{d} \bigvee_{t=0}^{t_\varepsilon - 1} T^{-t} \left(X_i^{-1} \left(f_{X_i}^{-1}(\mathcal{U}_N) \right) \right) \right) + H(\mathcal{M}) + \varepsilon \\
\leq \ & h\left(T^m, \bigvee_{i=1}^{d} \bigvee_{t=0}^{m-1} T^{-t} \left(X_i^{-1} \left(f_{X_i}^{-1}(\mathcal{U}_N) \right) \right) \right) + H(\mathcal{M}) + \varepsilon
\end{aligned}$$

for all $N \geq N_\varepsilon$. Hence:

$$h(T^m, \mathcal{P}) \leq \lim_{N \to \infty} h\left(T^m, \bigvee_{i=1}^{d} \bigvee_{t=0}^{m-1} T^{-t} \left(X_i^{-1} \left(f_{X_i}^{-1}(\mathcal{U}_N) \right) \right) \right) + H(\mathcal{M}) + \varepsilon$$

for any $\varepsilon > 0$, which implies:

$$\begin{aligned}
h(T^m) = \ & \sup_{\mathcal{P}} h(T^m, \mathcal{P}) \\
\leq \ & \lim_{N \to \infty} h\left(T^m, \bigvee_{i=1}^{d} \bigvee_{t=0}^{m-1} T^{-t} \left(X_i^{-1} \left(f_{X_i}^{-1}(\mathcal{U}_N) \right) \right) \right) + H(\mathcal{M}). \quad \square
\end{aligned}$$

Lemma 4. *Let $(\Omega, \mathscr{A}, \mu)$ be a probability space, $X : \Omega \to \mathbb{R}$ be a random variable and $A, B \in \mathscr{B}(\mathbb{R}^2)$. Then:*

$$\int \left| f_X^A - f_X^B \right| \, d\mu_X \leq \mu_X^2(A \triangle B)$$

holds true.

Proof. For all $x \in \mathbb{R}$:

$$\mu(\{\omega \in \Omega \mid (x, X(\omega)) \in A \triangle B\})$$
$$= \mu(\{\omega \in \Omega \mid (x, X(\omega)) \in A \setminus B\}) + \mu(\{\omega \in \Omega \mid (x, X(\omega)) \in B \setminus A\})$$
$$\geq \mu(\{\omega \in \Omega \mid (x, X(\omega)) \in A \setminus B\})$$
$$\geq \mu(\{\omega \in \Omega \mid (x, X(\omega)) \in A\}) - \mu(\{\omega \in \Omega \mid (x, X(\omega)) \in B\})$$
$$= f_X^A(x) - f_X^B(x)$$

holds true. Analogously, one can show:

$$\mu(\{\omega \in \Omega \mid (x, X(\omega)) \in A \triangle B\}) \geq f_X^B(x) - f_X^A(x).$$

This implies:

$$\mu(\{\omega \in \Omega \mid (x, X(\omega)) \in A \triangle B\}) \geq \left| f_X^A(x) - f_X^B(x) \right|$$

and, by Fubini's theorem:

$$\int \left| f_X^A - f_X^B \right| \, d\mu_X(x)$$
$$\leq \int \mu(\{\omega \in \Omega \mid (x, X(\omega)) \in A \triangle B\}) \, d\mu_X(x)$$
$$= \int \int 1_{\{\omega \in \Omega \mid (x, X(\omega)) \in A \triangle B\}}(\omega') \, d\mu(\omega') \, d\mu_X(x)$$
$$= \int \int 1_{\{y \in \mathbb{R} \mid (x, y) \in A \triangle B\}}(y') \, d\mu_X(y') \, d\mu_X(x)$$
$$= \int_{\mathbb{R}^2} 1_{A \triangle B}(x, y') \, d\mu_X^2(x, y')$$
$$= \mu_X^2(A \triangle B). \quad \square$$

Therefore, in particular, the above lemma implies that, if $(R_j)_{j \in \mathbb{N}}$ is a sequence of sets in $\mathscr{B}(\mathbb{R}^2)$ with $\lim_{j \to \infty} \mu_X^2(R_j \triangle R) = 0$, then $f_X^{R_j}$ converges to f_X^R in L^1 for $j \to \infty$.

Given $R \subseteq \mathbb{R}^2$ and a random variable $X : \Omega \to \mathbb{R}$, consider the function $f_{X,n}^R : \Omega \times \mathbb{R} \to [0, 1]$ with:

$$f_{X,n}^R(x, \omega) := \frac{1}{n} \#\{t \in \{1, 2, \dots, n\} \mid (x, X(T^t(\omega))) \in R\}.$$

We want to show that $f_{X,n}^R(x, \omega)$ converges to $f_X^R(x)$ for all $x \in \mathbb{R}$ and μ-almost all $\omega \in \Omega$. If $f_{X,n}^R(x, \omega)$ is monotone in x for all $\omega \in \Omega$ and $n \in \mathbb{N}$, this can be shown relatively easily using the pointwise ergodic theorem and the monotonicity of the considered functions. Monotonicity is guaranteed, if $x_1 \leq x_2$ implies:

$$\{y \in \mathbb{R} \mid (x_1, y) \in R\} \subseteq \{y \in \mathbb{R} \mid (x_2, y) \in R\}.$$

For example, if $R = \{(x, y) \in \mathbb{R}^2 \mid x > y\}$, the above implication holds true. For this special case, a proof of the statement in Lemma 5 can be found in [4].

However, we are interested in general sets $R \in \mathscr{B}(\mathbb{R}^2)$ and therefore, cannot use the monotonicity. Therefore, we have to prove this statement differently.

Lemma 5. *Let $(\Omega, \mathcal{A}, \mu, T)$ be an ergodic measure-preserving dynamical system, $X = (X_1, X_2, \ldots, X_d)$ be a random vector and $R \in \mathcal{B}(\mathbb{R}^2)$ satisfy (8). Then, for all $i \in \{1, 2, \ldots, d\}$, there exist sets $\widetilde{\Omega} \in \mathcal{A}$ and $B \in \mathcal{B}(\mathbb{R})$ with $\mu(\widetilde{\Omega}) = \mu_{X_i}(B) = 1$ satisfying:*

$$\lim_{n \to \infty} f_{X_i, n}^R(x, \omega) = f_{X_i}^R(x)$$

for all $\omega \in \widetilde{\Omega}$ and $x \in B$.

Proof. Fix $i \in \{1, 2, \ldots, d\}$. According to (8), there exists a countable set $C = \{(a_k, b_k) \mid k \in \mathbb{N}\}$ with:

$$\mu_{X_i}^2(\partial R \setminus C) = 0.$$

By the pointwise ergodic theorem (see, e.g., [10]), for all $j, k \in \mathbb{N}$, there exists $\Omega_{j,k}^* \in \mathcal{A}$ with $\mu(\Omega_{j,k}^*) = 1$ and:

$$\lim_{n \to \infty} f_{X_i, n}^{\{(a_k, b_k)\}}(a_j, \omega) = f_{X_i}^{\{(a_k, b_k)\}}(a_j)$$

for all $\omega \in \Omega_{j,k}^*$. It is easy to see that:

$$f_{X_i, n}^{\{(a_k, b_k)\}}(x, \omega) = 0 = f_{X_i}^{\{(a_k, b_k)\}}(x)$$

holds true for all $n \in \mathbb{N}$ and $\omega \in \Omega$ if $x \neq a_k$. Hence:

$$\lim_{n \to \infty} f_{X_i, n}^{\{(a_k, b_k)\}}(x, \omega) = f_{X_i}^{\{(a_k, b_k)\}}(x) \tag{20}$$

for all $x \in \mathbb{R}$ and $\omega \in \bigcap_{j=1}^{\infty} \Omega_{j,k}^*$. Using Fatou's lemma and the fact that C is countable implies:

$$\begin{aligned}
\liminf_{n \to \infty} f_{X_i, n}^{R \cap C}(x, \omega) &= \liminf_{n \to \infty} \sum_{\substack{k \in \mathbb{N}: \\ (a_k, b_k) \in R}} f_{X_i, n}^{\{(a_k, b_k)\}}(x, \omega) \\
&\geq \sum_{\substack{k \in \mathbb{N}: \\ (a_k, b_k) \in R}} \liminf_{n \to \infty} f_{X_i, n}^{\{(a_k, b_k)\}}(x, \omega) \\
&\stackrel{(20)}{=} \sum_{\substack{k \in \mathbb{N}: \\ (a_k, b_k) \in R}} f_{X_i}^{\{(a_k, b_k)\}}(x) = f_{X_i}^{R \cap C}(x)
\end{aligned} \tag{21}$$

for all $x \in \mathbb{R}$ and $\omega \in \bigcap_{j=1}^{\infty} \bigcap_{k=1}^{\infty} \Omega_{j,k}^*$. We will use this fact later.

Since $R \setminus \partial R$ is open, there exists a countable collection of pairwise disjoint rectangles $A_j \subseteq \mathbb{R}^2$ with:

$$R \setminus \partial R = \bigcup_{j=1}^{\infty} A_j. \tag{22}$$

Take $(x_j, y_j) \in A_j$ for all $j \in \mathbb{N}$. Using the pointwise ergodic theorem, for all $j \in \mathbb{N}$ there exists a set $\Omega_j \in \mathcal{A}$ with $\mu(\Omega_j) = 1$ and:

$$\lim_{n \to \infty} f_{X_i, n}^{A_j}(x_j, \omega) = f_{X_i}^{A_j}(x_j) \tag{23}$$

for all $\omega \in \Omega_j$. Because A_j is a rectangle, for all $\omega \in \Omega$:

$$f_{X_i, n}^{A_j}(x, \omega) = f_{X_i, n}^{A_j}(x_j, \omega)$$
$$\text{and} \quad f_{X_i}^{A_j}(x) = f_{X_i}^{A_j}(x_j)$$

holds true for all $x \in \mathbb{R}$ with $\{x\} \times \mathbb{R} \cap A_i \neq \varnothing$ and:

$$f_{X_i,n}^{A_j}(x,\omega) = f_{X_i}^{A_j}(x) = 0$$

holds true for all $x \in \mathbb{R}$ with $\mathbb{R} \times \{x\} \cap A_i = \varnothing$. Together with (23), this implies:

$$\lim_{n \to \infty} f_{X_i,n}^{A_j \setminus C}(x,\omega) = f_{X_i}^{A_j \setminus C}(x) \tag{24}$$

for all $x \in \mathbb{R}$ and $\omega \in \Omega_j$.

Set $R_J := \bigcup_{j=1}^{J} A_j$. Lemma 4 provides:

$$\lim_{J \to \infty} \int \left| f_{X_i}^{R_J \setminus C}(x) - f_{X_i}^{R \setminus C}(x) \right| d\mu_X(x)$$

$$= \lim_{J \to \infty} \mu_{X_i}^2 ((R_J \setminus C) \triangle (R \setminus C))$$

$$= \lim_{J \to \infty} \mu_{X_i}^2 ((R \setminus C) \setminus (R_J \setminus C))$$

$$= \lim_{J \to \infty} \mu_{X_i}^2 (R \setminus (R_J \cup C))$$

$$= \lim_{J \to \infty} \mu_{X_i}^2 (((R \cap \partial R) \cup (R \setminus \partial R)) \setminus (R_J \cup C))$$

$$= \lim_{J \to \infty} \mu_{X_i}^2 ((R \cap \partial R) \setminus (R_J \cup C)) + \mu_{X_i}^2 ((R \setminus \partial R) \setminus (R_J \cup C))$$

$$\leq \lim_{J \to \infty} \mu_{X_i}^2 (\partial R \setminus C) + \mu_{X_i}^2 ((R \setminus \partial R) \setminus R_J)$$

$$\overset{(8)}{=} \lim_{J \to \infty} \mu_{X_i}^2 ((R \setminus \partial R) \setminus R_J)$$

$$\overset{(22)}{=} \mu_{X_i}^2 ((R \setminus \partial R) \setminus (R \setminus \partial R)) = 0.$$

Therefore, there exists a set B_1 with $\mu_{X_i}(B_1) = 1$ and a sequence $(J_k)_{k \in \mathbb{N}}$ with:

$$\lim_{k \to \infty} f_{X_i}^{R_{J_k} \setminus C}(x) = f_{X_i}^{R \setminus C}(x) \tag{25}$$

for all $x \in B_1$. Thus:

$$\liminf_{n \to \infty} f_{X_i,n}^R(x,\omega) \geq \liminf_{n \to \infty} f_{X_i,n}^{R \cap C}(x,\omega) + \liminf_{n \to \infty} f_{X_i,n}^{R \setminus C}(x,\omega)$$

$$\overset{(21)}{\geq} f_{X_i}^{R \cap C}(x) + \liminf_{n \to \infty} f_{X_i,n}^{R_{J_k} \setminus C}(x,\omega)$$

$$\geq f_{X_i}^{R \cap C}(x) + \lim_{k \to \infty} \liminf_{n \to \infty} f_{X_i,n}^{R_{J_k} \setminus C}(x,\omega)$$

$$= f_{X_i}^{R \cap C}(x) + \lim_{k \to \infty} \liminf_{n \to \infty} \sum_{j=1}^{J_k} f_{X_i,n}^{A_j \setminus C}(x,\omega)$$

$$\geq f_{X_i}^{R \cap C}(x) + \lim_{k \to \infty} \sum_{j=1}^{J_k} \liminf_{n \to \infty} f_{X_i,n}^{A_j \setminus C}(x,\omega)$$

$$\overset{(24)}{=} f_{X_i}^{R \cap C}(x) + \lim_{k \to \infty} \sum_{j=1}^{J_k} f_{X_i}^{A_j \setminus C}(x)$$

$$= f_{X_i}^{R \cap C}(x) + \lim_{k \to \infty} f_{X_i}^{R_{J_k} \setminus C}(x)$$

$$\overset{(25)}{=} f_{X_i}^{R \cap C}(x) + f_{X_i}^{R \setminus C}(x)$$

$$= f_{X_i}^R(x)$$

for all $x \in B_1$ and $\omega \in \widetilde{\Omega}_1 := \bigcap_{j=1}^{\infty} \bigcap_{k=1}^{\infty} \Omega_j \cap \Omega_{j,k}^*$. Because $R^c \setminus \partial R$ is open as well, one can analogously show that there exist sets $\widetilde{\Omega}_2 \in \mathscr{A}$ and $B_2 \in \mathscr{B}(\mathbb{R})$ with $\mu(\widetilde{\Omega}_2) = \mu_{X_i}(B_2) = 1$ and:

$$\liminf_{n \to \infty} f_{X_i,n}^{R^c}(x,\omega) \geq f_{X_i}^{R^c}(x)$$

for all $x \in B_2$ and $\omega \in \widetilde{\Omega}_2$. This implies:

$$\limsup_{n \to \infty} f_{X_i,n}^{R}(x,\omega) = 1 - \liminf_{n \to \infty} f_{X_i,n}^{R^c}(x,\omega) \leq 1 - f_{X_i}^{R^c}(x) = f_{X_i}^{R}(x).$$

Hence:

$$f_{X_i}^{R}(x) \leq \liminf_{n \to \infty} f_{X_i,n}^{R}(x,\omega) \leq \limsup_{n \to \infty} f_{X_i,n}^{R}(x,\omega) \leq f_{X_i}^{R}(x) \tag{26}$$

for all $x \in B := B_1 \cap B_2$ and $\omega \in \widetilde{\Omega} := \widetilde{\Omega}_1 \cap \widetilde{\Omega}_2$. $\square$

Given a random vector $(X_1, X_2, \ldots, X_d)$ and $R \in \mathscr{B}(\mathbb{R}^2)$, we define the partition:

$$\mathcal{R}_i := (X_i, X_i)^{-1}(\{R, R^c\})$$

for all $i \in \{1, 2, \ldots, d\}$, which is equal to:

$$\mathcal{R}_i = \{\{(\omega_1, \omega_2) \in \Omega^2 \mid (X_i(\omega_1), X_i(\omega_2)) \in R\},$$
$$\{(\omega_1, \omega_2) \in \Omega^2 \mid (X_i(\omega_1), X_i(\omega_2)) \notin R\}\}.$$

Lemma 1. *Let $(\Omega, \mathscr{A}, \mu, T)$ be an ergodic measure-preserving dynamical system, $\boldsymbol{X} = (X_1, X_2, \ldots, X_d) : \Omega \to \mathbb{R}^d$ be a random vector, $(E_n)_{n \in \mathbb{N}}$ be a timing and $R \in \mathscr{B}(\mathbb{R}^2)$ satisfying (8). Then, there exists a sequence $(n_k)_{k \in \mathbb{N}} \subseteq \mathbb{N}_0^{\mathbb{N}}$ with:*

$$\lim_{k \to \infty} H\left(T^{-n_k}\left((f_{X_i} \circ X_i)^{-1}(\mathcal{U}_N) \right) \Bigg| \bigvee_{v=0}^{n_k} T^{-v}\left(\bigvee_{(s,t) \in E_k} (T^s, T^t)^{-1}(\mathcal{R}_i) \right) \right) = 0$$

for all $N \in \mathbb{N}$ and $i \in \{1, 2, \ldots, d\}$.

Proof. Because $(E_n)_{n \in \mathbb{N}}$ is a timing, there exist a sequence $(a_n)_{n \in \mathbb{N}} \in \mathbb{N}_0^{\mathbb{N}}$ with:

$$\limsup_{N \to \infty} \frac{\#\{n \in \{1, 2, \ldots, N\} \mid (a_n, a_n + n) \in \bigcup_{k=1}^{\infty} E_k\}}{N} = 1.$$

So one can find a strictly increasing sequence $(N_n)_{n \in \mathbb{N}} \in \mathbb{N}^{\mathbb{N}}$ with:

$$(a_{N_n}, a_{N_n} + N_n) \in \bigcup_{k=1}^{\infty} E_k$$

for all $n \in \mathbb{N}$ and:

$$\limsup_{n \to \infty} \frac{n}{N_n} = 1. \tag{27}$$

Now, fix $i \in \{1, 2, \ldots, d\}$. According to Lemma 5, there exist sets $\widetilde{\Omega} \in \mathscr{A}$ and $B \in \mathscr{B}(\mathbb{R})$ with $\mu(\widetilde{\Omega}) = \mu_{X_i}(B) = 1$ satisfying:

$$\lim_{n \to \infty} f_{X_i,n}^{R}(\omega, x) = f_{X_i}^{R}(x)$$

for all $\omega \in \widetilde{\Omega}$ and $x \in B$. Set

$$\Omega_0 := \widetilde{\Omega} \cap X_i^{-1}(B).$$

Consider the function $\phi_n : \Omega \to [0,1]$ with:

$$\phi_n(\omega) := \frac{1}{N_n} \#\{t \in \{N_1, N_2, \ldots, N_n\} \mid (X_i(T^t(\omega)), X_i(\omega)) \in R\}.$$

Then:

$$
\begin{aligned}
& f^R_{X_i}(X_i(\omega)) \\
&= \lim_{n \to \infty} f^R_{X_i,n}(\omega, X_i(\omega)) \\
&= \limsup_{n \to \infty} f^R_{X_i,N_n}(\omega, X_i(\omega)) \\
&\geq \limsup_{n \to \infty} \phi_n(\omega) \\
&= \limsup_{n \to \infty} \Big[f^R_{X_i,N_n}(\omega, X_i(\omega)) \\
&\quad - \frac{1}{N_n} \#\{t \in \{1, 2, \ldots, N_n\} \setminus \{N_1, N_2, \ldots, N_n\} \mid (X_i(T^t(\omega)), X_i(\omega)) \in R\} \Big] \\
&\geq \limsup_{n \to \infty} \Big[f^R_{X_i,N_n}(\omega, X_i(\omega)) - \frac{1}{N_n} \#\{1, 2, \ldots, N_n\} \setminus \{N_1, N_2, \ldots, N_n\} \Big] \\
&= \limsup_{n \to \infty} \Big[f^R_{X_i,N_n}(\omega, X_i(\omega)) - \frac{N_n - n}{N_n} \Big] \\
&= \lim_{n \to \infty} f^R_{X_i,N_n}(\omega, X_i(\omega)) - 1 + \limsup_{n \to \infty} \frac{n}{N_n} \\
&\overset{(27)}{=} f^R_{X_i}(X_i(\omega))
\end{aligned}
\tag{28}
$$

for all $\omega \in \Omega_0$.

It is easy to see that:

$$\sigma(\phi_n) \subseteq \sigma\left(\bigvee_{t=1}^{k} \left(id, T^{N_t}\right)^{-1}(\mathcal{R}_i) \mid k \in \mathbb{N} \right)$$

holds true for all $n \in \mathbb{N}$. This implies (see for instance [11], Theorem 13.4 (i)):

$$\sigma\left(f^R_{X_i} \circ X_i \right) \overset{(28)}{=_{\mu}} \sigma\left(\limsup_{n \to \infty} \phi_n \right) \subseteq \sigma\left(\bigvee_{t=1}^{k} \left(id, T^{N_t}\right)^{-1}(\mathcal{R}_i) \mid k \in \mathbb{N} \right).$$

Therefore $\bigvee_{t=1}^{k} \left(id, T^{N_t}\right)^{-1}(\mathcal{R}_i)$ is a sequence of partitions generating $\sigma\left(f^R_{X_i} \circ X_i \right)$. By (19), this implies:

$$\lim_{k \to \infty} H\left((f_{X_i} \circ X_i)^{-1}(\mathcal{U}_N) \,\Big|\, \bigvee_{t=1}^{k} \left(id, T^{N_t}\right)^{-1}(\mathcal{R}_i) \right) = 0 \tag{29}$$

for all $N \in \mathbb{N}$. Set:

$$n_k := \max_{1 \leq n \leq k} a_{N_n}.$$

Notice that:

$$\sigma\left(T^{-n_k} \left(\bigvee_{t=1}^{k} \left(id, T^{N_t}\right)^{-1}(\mathcal{R}_i) \right) \right) \subseteq \sigma\left(\bigvee_{v=0}^{n_k} T^{-v} \left(\bigvee_{(s,t) \in E_k} (T^s, T^t)^{-1}(\mathcal{R}_i) \right) \right)$$

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

$$
\begin{aligned}
&\lim_{k \to \infty} H\left(T^{-n_k}\left((f_{X_i} \circ X_i)^{-1}(\mathcal{U}_N) \right) \middle| \bigvee_{v=0}^{n_k} T^{-v}\left(\bigvee_{(s,t) \in E_k} (T^s, T^t)^{-1}(\mathcal{R}_i) \right) \right) \\
&\leq \lim_{k \to \infty} H\left(T^{-n_k}\left((f_{X_i} \circ X_i)^{-1}(\mathcal{U}_N) \right) \middle| T^{-n_k}\left(\bigvee_{t=1}^{k} \left(id, T^{N_t} \right)^{-1}(\mathcal{R}_i) \right) \right) \\
&= \lim_{k \to \infty} H\left((f_{X_i} \circ X_i)^{-1}(\mathcal{U}_N) \middle| \bigvee_{t=1}^{k} \left(id, T^{N_t} \right)^{-1}(\mathcal{R}_i) \right) \overset{(29)}{=} 0
\end{aligned}
$$

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

We can now finalize the proof of Theorem 1.

Proof of Theorem 1. Let $N \in \mathbb{N}$ and $m \in \mathbb{N}$. Set:

$$
\mathcal{P}_N^i := X_i^{-1}\left(f_{X_i}^{-1}(\mathcal{U}_N) \right)
$$

and:

$$
\mathcal{Q}_k^i := \bigvee_{(s,t) \in E_k} (T^s, T^t)^{-1}(\mathcal{R}_i)
$$

for all $i \in \{1, 2, \ldots, d\}$ and $k \in \mathbb{N}$. According to Lemma 1, there exists a sequence $(n_k)_{k \in \mathbb{N}} \subseteq \mathbb{N}_0^{\mathbb{N}}$ with:

$$
\lim_{k \to \infty} H\left(T^{-n_k}\left(\mathcal{P}_N^i \right) \middle| \bigvee_{v=0}^{n_k} T^{-v}\left(\mathcal{Q}_k^i \right) \right) = 0 \tag{30}
$$

for all $i \in \{1, 2, \ldots, d\}$. We have:

$$
\begin{aligned}
&\lim_{n \to \infty} \frac{1}{n} H\left(\bigvee_{i=1}^{d} \bigvee_{u=0}^{nm-1} T^{-u}\left(\mathcal{P}_N^i \right) \middle| \bigvee_{i=1}^{d} \bigvee_{u=0}^{nm-1} T^{-u}\left(\mathcal{Q}_k^i \right) \right) \\
&\leq \sum_{i=1}^{d} \lim_{n \to \infty} \frac{1}{n} H\left(\bigvee_{u=0}^{nm-1} T^{-u}\left(\mathcal{P}_N^i \right) \middle| \bigvee_{u=0}^{nm-1} T^{-u}\left(\mathcal{Q}_k^i \right) \right) \\
&= \sum_{i=1}^{d} \lim_{n \to \infty} \frac{1}{n} H\left(\bigvee_{u=0}^{nm-1-n_k} T^{-u}\left(T^{-n_k}\left(\mathcal{P}_N^i \right) \right) \middle| \bigvee_{u=0}^{nm-1-n_k} T^{-u}\left(\bigvee_{v=0}^{n_k} T^{-v}\left(\mathcal{Q}_k^i \right) \right) \right) \\
&\leq \sum_{i=1}^{d} \lim_{n \to \infty} \frac{1}{n} \sum_{u=0}^{nm-1-n_k} H\left(T^{-u}\left(T^{-n_k}\left(\mathcal{P}_N^i \right) \right) \middle| \bigvee_{v=0}^{n_k} T^{-v}\left(\mathcal{Q}_k^i \right) \right) \\
&= \sum_{i=1}^{d} \lim_{n \to \infty} \frac{nm-1-n_k}{n} H\left(T^{-n_k}\left(\mathcal{P}_N^i \right) \middle| \bigvee_{v=0}^{n_k} T^{-v}\left(\mathcal{Q}_k^i \right) \right) \\
&\leq \sum_{i=1}^{d} m \cdot H\left(T^{-n_k}\left(\mathcal{P}_N^i \right) \middle| \bigvee_{v=0}^{n_k} T^{-v}\left(\mathcal{Q}_k^i \right) \right)
\end{aligned}
$$

for all $k, m, N \in \mathbb{N}$. This implies:

$$h\left(T^m, \bigvee_{i=1}^{d} \bigvee_{u=0}^{m-1} T^{-u}\left(\mathcal{P}_N^i\right)\right) - \lim_{k \to \infty} h\left(T^m, \bigvee_{i=1}^{d} \bigvee_{u=0}^{m-1} T^{-u}\left(\mathcal{Q}_k^i\right)\right)$$

$$\leq \lim_{k \to \infty} \lim_{n \to \infty} \frac{1}{n} H\left(\bigvee_{i=1}^{d} \bigvee_{u=0}^{nm-1} T^{-u}\left(\mathcal{P}_N^i\right) \middle| \bigvee_{i=1}^{d} \bigvee_{u=0}^{nm-1} T^{-u}\left(\mathcal{Q}_k^i\right)\right)$$

$$\leq \lim_{k \to \infty} \sum_{i=1}^{d} m \cdot H\left(T^{-n_k}\left(\mathcal{P}_N^i\right) \middle| \bigvee_{v=0}^{n_k} T^{-v}\left(\mathcal{Q}_k^i\right)\right) \overset{(30)}{=} 0.$$

Using Lemma 3, we can conclude that there exists a constant $c \in \mathbb{R}$ with:

$$h(T^m) \leq \lim_{k \to \infty} h\left(T^m, \bigvee_{i=1}^{d} \bigvee_{u=0}^{m-1} T^{-u}\left(\bigvee_{(s,t) \in E_k} (T^s, T^t)^{-1}(\mathcal{R}_i)\right)\right) + c$$

$$= \lim_{k \to \infty} m \cdot h\left(T, \bigvee_{i=1}^{d} \bigvee_{(s,t) \in E_k} (T^s, T^t)^{-1}(\mathcal{R}_i)\right) + c$$

for all $m \in \mathbb{N}$. Thus:

$$h(T) - \lim_{k \to \infty} h\left(T, \bigvee_{i=1}^{d} \bigvee_{(s,t) \in E_k} (T^s, T^t)^{-1}(\mathcal{R}_i)\right)$$

$$= \lim_{m \to \infty} \frac{1}{m} \cdot h(T^m) - \lim_{k \to \infty} h\left(T, \bigvee_{i=1}^{d} \bigvee_{(s,t) \in E_k} (T^s, T^t)^{-1}(\mathcal{R}_i)\right)$$

$$\leq \lim_{m \to \infty} \frac{1}{m} \cdot c = 0,$$

which is equivalent to:

$$h(T) \leq \lim_{k \to \infty} h\left(T, \bigvee_{i=1}^{d} \bigvee_{(s,t) \in E_k} (T^s, T^t)^{-1}(\mathcal{R}_i)\right). \tag{31}$$

On the other hand:

$$h(T) = \sup_{\mathcal{P}} h(T, P) \geq \lim_{k \to \infty} h\left(T, \bigvee_{i=1}^{d} \bigvee_{(s,t) \in E_k} (T^s, T^t)^{-1}(\mathcal{R}_i)\right),$$

which, together with (31), finishes the proof. $\square$

6. Conclusions

We discussed a special "two-dimensional" approach to symbolic dynamics differing from many usual approaches which was introduced in [6]. From the practical viewpoint, the difference can be illustrated as follows: given the time-dependent measurements of a real-valued quantity, a symbolization is not conducted for the measurements themselves as in usual approaches, but for pairs of measurements at two different times. This means that to each pair of possible measured values, a symbol from a finite symbol set is assigned. Here, we only considered two symbols which lead to a partitioning of the two-dimensional real space $\mathbb{R}^2$ into a set R and its complement $\mathbb{R}^2 \setminus R$. In usual approaches, partitions of $\mathbb{R}$ are considered. (Advantages of the "two-dimensional" approach are described in [6]).

The set R, called a discriminating relation, was considered as a basic building block for constructing partitions of the state space of a given dynamical system, having time-dependent measurements of finitely many quantities in mind. In addition to the discrimination relation, the second central concept was the concept of a timing which roughly describes which pairs of times are included in the symbolization process and guarantees that there are not too few such pairs. The central question of the paper was that of under which conditions on a discriminating relation R the partitions constructed from R determine the KS entropy of a measure-preserving dynamical system. With Theorem 1, we gave a relatively general statement partially answering this question. Some specifications of the theorem in Section 4 illustrate the nature of "successful" discriminating relations.

Although the statement of Theorem 1 appears relatively natural when looking at the proofs a little closer, we do not expect that all cases where the K-S entropy can be constructed based on a discriminating relation is covered by the statement; however, we have no counterexample. The main tool used in the proofs of the results is the pointwise ergodic theorem. It allows to establish a connection between the generalized ordinal patterns and the shape of the discriminating relation.

The results of this paper, being on a rather abstract level, give some insights as to why the idea of ordinal patterns is working well, as reported by several applied papers, with extracting those advantageous features being more general than in the original ordinal approach. Having many choices for a discriminating relation, for practical purposes such as, for example, in a classification context, one needs methods and criteria for finding good discrimination relations, adapted to given data and problems. This is an important challenge for further research related to the given approach to symbolic dynamics. A further aspect is to discuss the approach for partitioning the $\mathbb{R}^2$ into more than two pieces.

References

1. Bandt, C.; Pompe, B. Permutation Entropy: A Natural Complexity Measure for Time Series. *Phys. Rev. Lett.* **2002**, *88*, 174102. [CrossRef]
2. Zanin, M.; Zunino, L.; Rosso, O.A.; Papo, D. Permutation Entropy and Its Main Biomedical and Econophysics Applications: A Review. *Entropy* **2012**, *14*, 1553–1577. [CrossRef]
3. Bandt, C.; Keller, G.; Pompe, B. Entropy of interval maps via permutations. *Nonlinearity* **2002**, *15*, 1595–1602. [CrossRef]
4. Antoniouk, A.; Keller, K.; Maksymenko, S. Kolmogorov-Sinai entropy via separation properties of order-generated σ-algebras. *Discret. Contin. Dyn. Syst. A* **2014**, *34*, 1793–1809. [CrossRef]
5. Smith, G.; Goulding, J.; Barrack, D. Towards Optimal Symbolization for Time Series Comparisons. In Proceedings of the 2013 IEEE 13th International Conference on Data Mining Workshops, Dallas, TX, USA, 7–10 December 2013; pp. 646–653. [CrossRef]
6. Stolz, I.; Keller, K. A General Symbolic Approach to Kolmogorov-Sinai Entropy. *Entropy* **2017**, *19*, 675. [CrossRef]
7. Walters, P. An introduction to ergodic theory. In *Graduate Texts in Mathematics*; Springer: New York, NY, USA, 1982; Volume 79.
8. Keller, K. Permutations and the Kolmogorov-Sinai entropy. *Discret. Contin. Dyn. Syst.* **2012**, *32*, 891–900. [CrossRef]
9. 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. *Phys. D* **2005**, *210*, 77–95. [CrossRef]
10. Cornfeld, I.P.; Fomin, S.V.; Sinai, Y.G. *Ergodic Theory*, 1st ed.; Springer: New York, NY, USA, 1982.
11. Billingsley, P. *Probability and Measure*, 2nd ed.; John Wiley and Sons: Hoboken, NJ, USA, 1986.