

## Article

# A General Symbolic Approach to Kolmogorov-Sinai Entropy

Inga Stolz \* and Karsten Keller

Institute of Mathematics, University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany;  
keller@math.uni-luebeck.de

\* Correspondence: stolz@math.uni-luebeck.de; Tel.: +49-451-3101-6050

Received: 31 October 2017; Accepted: 5 December 2017; Published: 9 December 2017

**Abstract:** It is popular to study a time-dependent nonlinear system by encoding outcomes of measurements into sequences of symbols following certain symbolization schemes. Mostly, symbolizations by threshold crossings or variants of it are applied, but also, the relatively new symbolic approach, which goes back to innovative works of Bandt and Pompe—ordinal symbolic dynamics—plays an increasing role. In this paper, we discuss both approaches novelly in one breath with respect to the theoretical determination of the Kolmogorov-Sinai entropy (KS entropy). For this purpose, we propose and investigate a unifying approach to formalize symbolizations. By doing so, we can emphasize the main advantage of the ordinal approach if no symbolization scheme can be found that characterizes KS entropy directly: the ordinal approach, as well as generalizations of it provide, under very natural conditions, a direct route to KS entropy by default.

**Keywords:** symbolization; KS entropy; generating partitions;  $\sigma$ -algebras

## 1. Introduction

Using symbolizations to study observed data plays an important role in today's time series analysis (see for instance the review papers of Daw et al. [1], Zanin et al. [2], Amigó et al. [3], and the examples in biology, medicine, artificial intelligence and data mining, just to mention a few, given therein). Thereby, it is assumed that time series, given by measurements of a real-world time-dependent system, store information about the complexity of the underlying system, which can be accessed by symbolic dynamics. In this paper, we assume further that measurements provide  $n$ -dimensional real-valued outcomes, that is a measuring process provides  $n$  time series.

Knowing the complexity is a key to classify systems and to predict future developments. A data analyst can for instance quantify complexity by empirical entropy measures, in particular by estimating the well-defined Kolmogorov-Sinai entropy (KS entropy). In order to estimate the KS entropy, however, a data analyst is always faced with the problem of choosing an adequate symbolization scheme.

Symbolizing a time series could be done in a “classical” manner for example by subdividing the data range into a finite number of intervals (see Section 2.1, often called the threshold crossing method in symbolic dynamics) or in an ordinal manner for example by considering the up and down behavior of subsequent measured values (see Section 2.2). The most ideal, however unrealistic, case is given if the analyst knows the underlying dynamics and picks a generating (under the dynamics) partition (see Sections 1.1 and 1.4 for the mathematical formulation of the general problem, as well as for instance Crutchfield and Packard [4], Bollt et al. [5] and Kennel and Buhl [6]).

In the present paper, we show, by proposing a unifying approach to formalize symbolizations, that under relatively weak assumptions, the search for a generating partition can be skipped if one chooses a symbolization scheme that regards a dependency between two measured values (see Section 2.2). In fact, following some rules by picking such a symbolization scheme, a generating sequence of finite partitions (see Sections 1.1 and 1.4) is provided by default and needs no further attention (see Section 2.3

for an overview and Sections 3 and 4.1, as well as the Appendix for the mathematics behind this). Moreover, the unifying approach allows one to consider “classical” and the relatively new ordinal symbolic dynamics [3] hand in hand and therefore to study respective assets and drawbacks.

In terms of the analyst, we propose a supplementing pool of complexity measures, which are in a certain sense approximations of the KS entropy and may be worth being compared in the finite setting of application (see Figure 7). Moreover, the relatively new ordinal approach could benefit from results achieved in “classical” symbolic dynamics, for instance to estimate a good symbolization scheme (see our ending remarks of the paper in Section 5 and for instance Steuer et al. [7], Letellier [8] and, published most recently, Li and Ray [9], as well as the references given therein). However, such topics exceed the scope of this paper.

### 1.1. Mathematical Formulation of the General Problem

Let us describe the central problems of determining KS entropy and give the main concepts of the paper without going into too much detail. The mathematical formulation is necessary at this point in order to state the results of the paper adequately.

We model a real-world time-dependent system by a state space  $\Omega$ , that is states  $\omega$  of the system are taken from the set  $\Omega$  and events on the system from a  $\sigma$ -algebra  $\mathcal{A}$  on  $\Omega$ . We assume that the states are distributed according to a probability distribution  $\mu$  on  $(\Omega, \mathcal{A})$ . Moreover, considering states of the system at times in  $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ , the dynamics of the system is described by a map  $T$  with the interpretation that the system is in state  $T(\omega)$  at time  $t + 1$  if it is in state  $\omega$  at time  $t$ . For mathematical correctness,  $T$  is required to be measurable with respect to  $\mathcal{A}$ . We assume that the distribution of the states does not change in time, meaning  $T$  is  $\mu$ -invariant, which is defined by  $T^{-1}(A) \in \mathcal{A}$  for all sets  $A \in \mathcal{A}$ .

The KS entropy is based on entropy rates of finite partitions of the state space. Given a finite partition  $\mathcal{C} = \{C^{(1)}, C^{(2)}, \dots, C^{(q)}\} \subset \mathcal{A}$  of  $\Omega$ , the entropy rate  $h_\mu(T, \mathcal{C})$ , roughly speaking, measures the complexity of possible symbolic paths (see Section 4.1). A symbolic path is given by assigning to each state of the orbit:

$$\omega, T(\omega), T^{\circ 2}(\omega), T^{\circ 3}(\omega), \dots$$

a symbol  $a$  when the state is contained in  $C^{(a)}$ . Here,  $T^{\circ t}(\omega)$  denotes the  $t$ -th iterate of  $\omega$  under  $T$ . We emphasize that starting with a partition  $\mathcal{C} = \{C^{(1)}, C^{(2)}, \dots, C^{(q)}\}$  is equivalent to a start where to each state in  $\Omega$  a symbol in  $\{1, 2, \dots, q\}$  is assigned (in a measurable way). That is why we use the term symbolic approach.

In order to obtain a complexity measure that is independent of the discretization determined by a finite partition, one takes the supremum of the entropy rate  $h_\mu(T, \mathcal{C})$  over all finite partitions  $\mathcal{C} \subset \mathcal{A}$  of  $\Omega$ , that is the KS entropy  $h_\mu^{\text{KS}}(T)$  of  $T$ :

$$h_\mu^{\text{KS}}(T) := \sup_{\mathcal{C} \text{ finite partition}} h_\mu(T, \mathcal{C}).$$

Since usually there are uncountably many finite partitions, the determination of KS entropy on the basis of the definition is not feasible, so one is interested in finding natural partitions “carrying” the KS entropy.

In the case of a generating partition  $\mathcal{C}$  under  $T$  (see Section 4.1), KS entropy is already characterized by this partition, meaning that:

$$h_\mu^{\text{KS}}(T) = h_\mu(T, \mathcal{C}) \quad (1)$$

(see, e.g., Walters [10], Theorem 4.18). Finding such suitable partitions, however, is impossible in most cases. A more realistic way of approaching KS entropy is to look for a generating and increasing, i.e., refining (see Section 4.1), sequence  $(\mathcal{C}_d)_{d \in \mathbb{N}}$  of finite partitions  $\mathcal{C}_d \subset \mathcal{A}$  of  $\Omega$ , where:

$$h_{\mu}^{\text{KS}}(T) = \lim_{d \rightarrow \infty} h_{\mu}(T, \mathcal{C}_d) = \sup_{d \in \mathbb{N}} h_{\mu}(T, \mathcal{C}_d) \quad (2)$$

(see, e.g., Walters [10], Theorem 4.22).

In the present paper, we discuss this countable increasing route to KS entropy in a framework where all partitions considered are derived from a natural real-valued “measuring process” and a symbolization scheme determined by a finite partition of the two-dimensional Euclidean space. The discussion includes and generalizes ideas from “classical” symbolic dynamics and from ordinal symbolic dynamics related to permutation entropy and sheds some new light on the latter one.

### 1.2. Observables and the Measuring Process

The modeling is completed by assuming that an  $n$ -dimensional outcome (here,  $n \in \mathbb{N} = \{1, 2, 3, \dots\}$ ) of the system for each time is provided by observables  $X_1, X_2, \dots, X_n$ , which mathematically are random variables on the probability space  $(\Omega, \mathcal{A}, \mu)$  with values in the real numbers  $\mathbb{R}$ . It provides the link between the dynamical model and the given  $n$ -dimensional time series data.

Fixing some state  $\omega \in \Omega$ , we interpret the real numbers:

$$X_i(\omega), X_i(T(\omega)), X_i(T^{\circ 2}(\omega)), X_i(T^{\circ 3}(\omega)), \dots$$

as the values measured by  $X_i$  at times  $0, 1, 2, 3, \dots$  when the given system is in state  $\omega$  at the beginning.

Therefore, the random vector  $\mathbf{X} = (X_i)_{i=1}^n$  for the time-developing system provides random vectors:

$$\mathbf{X} = (X_i)_{i=1}^n, \mathbf{X} \circ T = (X_i \circ T)_{i=1}^n, \mathbf{X} \circ T^{\circ 2} = (X_i \circ T^{\circ 2})_{i=1}^n, \mathbf{X} \circ T^{\circ 3} = (X_i \circ T^{\circ 3})_{i=1}^n, \dots$$

forming the measuring process:

$$(\mathbf{X} \circ T^{\circ t})_{t \in \mathbb{N}_0} = \left( (X_i \circ T^{\circ t})_{i=1}^n \right)_{t \in \mathbb{N}_0} \quad (3)$$

with the  $n$  time series  $(X_i(T^{\circ t}(\omega)))_{t \in \mathbb{N}_0}$  for  $i = 1, 2, \dots, n$  as outcomes. Note that the symbolizations we consider in the following are given at the observational level, i.e., with respect to the values of  $X_i$ ; this complies with symbolizing a time series in real-world data analysis.

Let us regard  $(\Omega, \mathcal{A}, \mu)$ ,  $T$ ,  $n \in \mathbb{N}$  and  $\mathbf{X} = (X_i)_{i=1}^n$  as fixed in the following.

### 1.3. Information Contents in the Language of Event Systems

It is a central question of the given paper whether a description of a system, for instance by a measurement or by a symbolization, provides the same information as another one. In information theory, this is a matter of the richness of the event systems associated with the descriptions, more precisely a relation between sub- $\sigma$ -algebras  $\mathcal{F}$  and  $\mathcal{F}'$  of  $\mathcal{A}$  defined by (compare to Walters [10], Definition 4.5):

$$\mathcal{F}' \stackrel{\mu}{\subset} \mathcal{F} \text{ if for each } F' \in \mathcal{F}' \text{ there exists some } F \in \mathcal{F} \text{ with } \mu(F \triangle F') = 0.$$

The inclusion  $\mathcal{F}' \stackrel{\mu}{\subset} \mathcal{F}$  means that for each event in  $\mathcal{F}'$ , there exists an event in  $\mathcal{F}$  being distinct from the first one with probability zero and that is interpreted as meaning that  $\mathcal{F}$  preserves all information contained in  $\mathcal{F}'$ .

The  $\sigma$ -algebra  $\mathcal{A}$  on  $\Omega$  consists of all events related to the given system, those events accessed by the given observables and the whole measuring process (3) form the sub- $\sigma$ -algebras  $\sigma(\mathbf{X})$  and

$\sigma\left((\mathbf{X} \circ T^{\circ t})_{t \in \mathbb{N}}\right)$  of  $\mathcal{A}$ , respectively. Mathematically,  $\sigma(\mathbf{X})$  is the smallest  $\sigma$ -algebra built from all preimages of Borel sets in  $\mathbb{R}$  for:

$$X_1, X_2, \dots, X_n, \dots,$$

and  $\sigma\left((\mathbf{X} \circ T^{\circ t})_{t \in \mathbb{N}}\right)$  is the smallest  $\sigma$ -algebra built from all preimages of Borel sets in  $\mathbb{R}$  for:

$$X_1, X_2, \dots, X_n, X_1 \circ T, X_2 \circ T, \dots, X_n \circ T, X_1 \circ T^{\circ 2}, X_2 \circ T^{\circ 2}, \dots, X_n \circ T^{\circ 2}, \dots$$

In these definitions, it is enough to take only intervals  $I$  instead of Borel sets. Here,  $(X_i \circ T^{\circ t})^{-1}(I)$  describes the event that the value of the  $i$ -th measurement at time  $t$  is in  $I$ .

The sub- $\sigma$ -algebra  $\sigma((\mathcal{C}_d)_{d \in \mathbb{N}})$ , which is the smallest  $\sigma$ -algebra built from all events contained in some of the partitions  $\mathcal{C}_d$  for  $d \in \mathbb{N}$ , provides the events accessed by the corresponding symbolization (see Section 4.1). Our goal is to construct an increasing sequence  $(\mathcal{C}_d)_{d \in \mathbb{N}}$  of finite partitions, i.e.,  $\mathcal{C}_{d+1}$  refines  $\mathcal{C}_d$  (see Section 4.1), which preserves the information given by the measuring process (3), i.e.,

$$\sigma\left((\mathbf{X} \circ T^{\circ t})_{t \in \mathbb{N}}\right) \overset{\mu}{\subset} \sigma((\mathcal{C}_d)_{d \in \mathbb{N}}), \quad (4)$$

or weaker by the observables themselves, i.e.,

$$\sigma(\mathbf{X}) \overset{\mu}{\subset} \sigma((\mathcal{C}_d)_{d \in \mathbb{N}}). \quad (5)$$

If (4) holds and the measuring process preserves the information of the original system, i.e., if:

$$\mathcal{A} \overset{\mu}{\subset} \sigma\left((\mathbf{X} \circ T^{\circ t})_{t \in \mathbb{N}}\right), \quad (6)$$

or if just (5) holds, but the observables preserve already the information of the original system, i.e., if:

$$\mathcal{A} \overset{\mu}{\subset} \sigma(\mathbf{X}), \quad (7)$$

then:

$$\mathcal{A} \overset{\mu}{\subset} \sigma((\mathcal{C}_d)_{d \in \mathbb{N}}) \subset \mathcal{A},$$

meaning that  $(\mathcal{C}_d)_{d \in \mathbb{N}}$  is generating (see Section 4.1 and compare to Walters [10]), which provides (2).

Conditions (6) and (7) are not as artificial as they appear at first glance:

1. There is a very natural set of observables satisfying (7), hence (6). If  $\Omega$  is a Borel subset of  $\mathbb{R}^n$ , it is very plausible to assume that states and vectors of measured values are coinciding. This can be modeled by observables  $X_i$ ;  $i = 1, 2, \dots, n$  with  $X_i$  being the  $i$ -th coordinate projection, i.e.,  $X_i(\omega) = x_i$  for  $\omega = (x_1, x_2, \dots, x_n)$ . Clearly, in this simplest variant of modeling measurements, observables basically are superfluous in the modeling.
2. In the case of only one observable, the separation of states is natural in a certain sense according to Takens' theory (see Takens [11] and Gutman [12]).

#### 1.4. A "Two-Dimensional" Way of Symbolizations

The partitions  $\mathcal{C}_d$  with  $d \in \mathbb{N}$  that we want to study in the following are formed on the basis of a finite partition  $\mathcal{R}$  of the two-dimensional Euclidean space  $\mathbb{R}^2$  and finite sets  $E_d \subset \mathbb{N}_0 \times \mathbb{N}_0$  of time pairs:

$$\mathcal{C}_d = \mathcal{C}^{\mathcal{R}, E_d}(T, \mathbf{X}) := \bigvee_{i=1}^n \bigvee_{(s,t) \in E_d} (X_i \circ T^{\circ s}, X_i \circ T^{\circ t})^{-1}(\mathcal{R}) \subset \mathcal{A}, \quad (8)$$

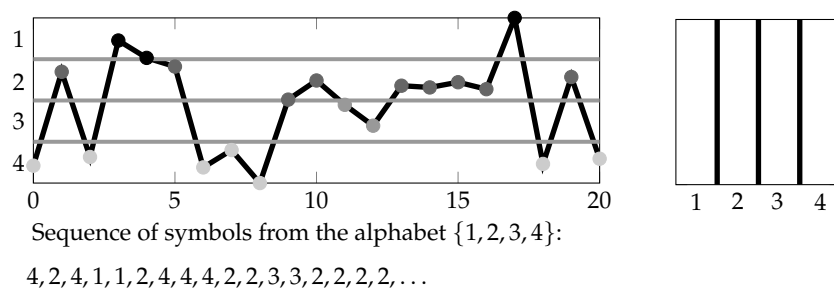
i.e.,  $\mathcal{C}_d$  is the coarsest partition refining all partitions:

$$(X_i \circ T^{os}, X_i \circ T^{ot})^{-1}(\mathcal{R}) = \{ \{ \omega \in \Omega \mid (X_i(T^{os}(\omega)), X_i(T^{ot}(\omega))) \in R \} \mid R \in \mathcal{R} \}$$

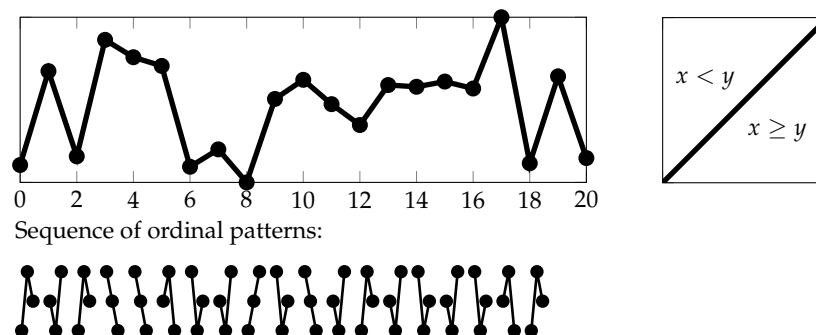
for  $(s, t) \in E_d$  (see Section 4.1 for the definition of the join  $\bigvee_{r=1}^m \mathcal{C}_r$  of finite partitions  $\mathcal{C}_r \subset \mathcal{A}$  of  $\Omega$ ).

Here,  $\mathcal{R}$  specifies the symbolization scheme for classifying the mutual position of measurements by  $X_i$  at two times  $s$  and  $t$  (see Figures 1 and 2, as well as the next section). We call  $\mathcal{R}$  the basic symbolization scheme in the following. Note that we display the two-dimensional Euclidean space  $\mathbb{R}^2$  by a square for illustrative purposes.

Further, the choice of  $(E_d)_{d \in \mathbb{N}}$  complies with Definition 1, in particular in order to realize that  $\mathcal{C}_{d+1}$  refines  $\mathcal{C}_d$  (see Section 4.1),  $\mathcal{C}_d$  is finite (see Definition 1(i)), and each time point that is relevant for the symbolization is accessed (see Definition 1(ii)).



**Figure 1.** Transforming a time series in the “classical” way. Left: intervals are turned into symbols from the alphabet  $\{1, 2, 3, 4\}$ . Right: two-dimensional view of the symbolization (see Sections 1.4 and 2.1).



**Figure 2.** Transforming a time series in the “ordinal” way. Left: vectors are transformed into ordinal patterns. Right: the basic symbolization scheme (see Sections 1.4 and 2.2).

**Definition 1.** We call a sequence  $(E_d)_{d \in \mathbb{N}}$  of sets  $E_d$  with  $E_1 \subset E_2 \subset \dots \subset \{(s, t) \mid s, t \in \mathbb{N}_0, s < t\}$  a timing if there exists a set  $\{v_0, v_1, \dots\} \subseteq \mathbb{N}_0$  with  $v_0 < v_1 < v_2 < \dots$  such that:

- (i) for each  $d \in \mathbb{N}$ , it holds  $E_d \subset \{v_0, v_1, \dots, v_d\}^2$ , and
- (ii) for each  $s \in \{v_0, v_1, \dots, v_d\}$ , there exists some  $t \in \{v_0, v_1, \dots, v_d\}$  such that  $(s, t) \in E_d$  or  $(t, s) \in E_d$ .

A timing is for instance given by the sets:

$$E_d = \{(s, t) \mid s, t \in \{0, 1, 2, \dots, d\} \text{ with } s < t\}; d \in \mathbb{N}, \quad (9)$$

or by (11) and (17) (see below). It is suggestive to call the timing defined by (9) full timing in the following.

Subsequently, we discuss the following two questions:

1. Why is the approach given natural and sufficiently general?
2. Under which conditions on the basic symbolization scheme  $\mathcal{R}$  and the timing  $E_d$  does the sequence  $(\mathcal{C}_d)_{d \in \mathbb{N}} = (\mathcal{C}^{\mathcal{R}, E_d}(T, \mathbf{X}))_{d \in \mathbb{N}}$  satisfy Statement (5) or even (4)?

Section 2 is devoted to the first question. In the first part of Section 3, we summarize our results to the second question and give some examples of basic symbolization schemes. Sufficient conditions that answer the second question including known results are presented for the interested reader in the second part of Section 3 and proven in Section 4 (see also the Appendix). We close this paper with some remarks about further theoretical and practical scientific issues (see Section 5).

## 2. Two Examples

At first glance, the above approach gives a rather exaggerated impression. The aim of the following examples is to convince the reader that sequences  $(\mathcal{C}_d)_{d \in \mathbb{N}}$  formed on the basis of basic symbolization schemes  $\mathcal{R}$  and timings  $(E_d)_{d \in \mathbb{N}}$  are natural and are unifying known symbolic approaches.

### 2.1. “Classical” Symbolic Dynamics

First, we discuss “classical” symbolic dynamics with a fixed partition (see for instance Daw et al. [1], Kurths et al. [13] and the references given therein). For convenience, we assume that the dynamics is living on the real line, i.e.,  $\Omega = \mathbb{R}$ , and restrict ourselves to the simple case that  $\mathbb{R}$  is subdivided into a finite number of intervals  $I_1, I_2, \dots, I_k$  (see for instance Figure 1).

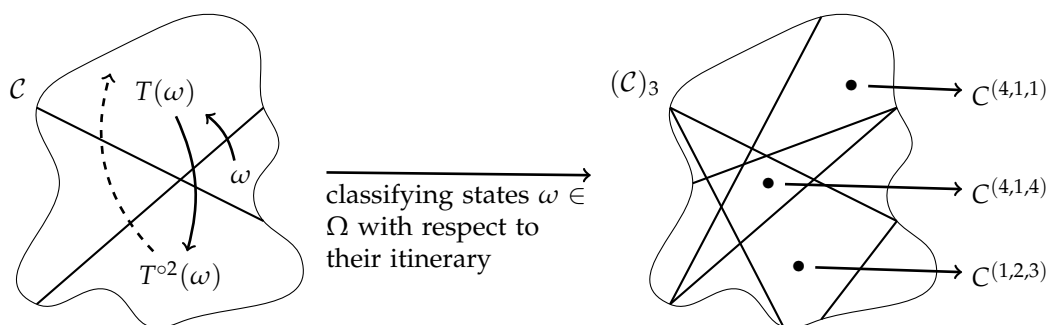
The determination of the entropy rate of the partition  $\mathcal{C} = \{I_1, I_2, \dots, I_k\}$  is based on the partitions:

$$(\mathcal{C})_t = \{C^{(a_1, a_2, \dots, a_t)} \mid a_1, a_2, \dots, a_t \in \{1, 2, \dots, k\}\}$$

increasing with  $t \in \mathbb{N}$ , where:

$$C^{(a_1, a_2, \dots, a_t)} = \{\omega \in \Omega \mid \omega \in I_{a_1}, T(\omega) \in I_{a_2}, \dots, T^{o t-1}(\omega) \in I_{a_t}\}$$

consists of those states  $\omega \in \Omega$  successively visiting  $I_{a_1}, I_{a_2}, \dots, I_{a_t}$ , i.e., having the symbolic itinerary  $a_1, a_2, \dots, a_t$ . Compare to Section 4.1, in particular for the deliberate notation used in this subsection, and see Figure 3, where we illustrate the symbolization process underlying the determination of the entropy rate for  $t = 3$ .



**Figure 3.** Symbolization process underlying the determination of the entropy rate  $h_\mu(T, \mathcal{C})$  for  $t = 3$  (see Section 4.1).

For all  $t \in \mathbb{N}$ , it holds:

$$(\mathcal{C})_t = \bigvee_{s=0}^{t-1} (T^{os})^{-1}(\mathcal{C}).$$

In order to rewrite the “classical” approach into a form compatible with the proposed one, we need an artificial two-dimensional “blow-up” of the partitions  $(\mathcal{C})_t$ ;  $t \in \mathbb{N}$ , which is given by the partition:

$$\mathcal{R} := \{I_1 \times \mathbb{R}, I_2 \times \mathbb{R}, \dots, I_k \times \mathbb{R}\} \text{ (see Figure 1),} \quad (10)$$

of  $\mathbb{R}^2$  and the sets of time pairs:

$$E_t := \{(0, 1), (1, 2), (2, 3), \dots, (t-1, t)\}; t \in \mathbb{N}, \quad (11)$$

that is:

$$(\mathcal{C})_t = \bigvee_{s=0}^{t-1} (X \circ T^{\circ s})^{-1}((\mathcal{C})_1) = \bigvee_{(s,u) \in E_t} (X \circ T^{\circ s}, X \circ T^{\circ u})^{-1}(\mathcal{R}) = \mathcal{C}^{\mathcal{R}, E_t}(T, X).$$

Here, we consider, motivated by our general procedure, the single observable  $X$  (meaning  $\mathbf{X} = X$  in the general framework) with  $X(\omega) = \omega$  for all  $\omega \in \Omega$ , which fits the situation described at the end of Section 1.3.

This means in the other direction that the partitions  $\mathcal{C}_d = \mathcal{C}^{\mathcal{R}, E_d}(T, X)$  are coinciding with the partitions  $(\mathcal{C})_d$ . In particular, it holds:

$$h_\mu(T, \mathcal{C}_d) = h_\mu(T, (\mathcal{C})_d)$$

for all  $d \in \mathbb{N}$ , and since  $h_\mu(T, (\mathcal{C})_t) = h_\mu(T, \mathcal{C})$  for all  $t \in \mathbb{N}$  (see for instance Einsiedler and Schmidt [14], Satz 3.13), we obtain:

$$h_\mu(T, \mathcal{C}_d) = h_\mu(T, \mathcal{C})$$

for all  $d \in \mathbb{N}$ . This fact implying that the sequence  $(\mathcal{C}_d)_{d \in \mathbb{N}} = (\mathcal{C}^{\mathcal{R}, E_d}(T, X))_{d \in \mathbb{N}}$  is generating iff  $\mathcal{C}$  is generating under  $T$  (see the final remarks in Section 4.1) says that  $\mathcal{R}$  as given by (10) has no generating potential when  $\mathcal{C}$  fails to be generating under  $T$ . This is not surprising since  $\mathcal{R}$  is no more than a two-dimensional “blow-up” of  $\mathcal{C}$ . The second example shows the existence of good choices of  $\mathcal{R}$  with generating properties under certain assumptions.

In Figure 4, we study (compare to Section 4.1) two different initial partitions, that is  $\mathcal{C} = \{[0, 1/2), [1/2, 1)\}$  and  $\mathcal{D} = \{[0, 1/4), [1/4, 1)\}$ , under the transformation  $T : ([0, 1)) \leftrightarrow$ , defined by:

$$T(\omega) = \begin{cases} 2\omega & \text{if } 0 \leq \omega < \frac{1}{2}, \\ 2 - 2\omega & \text{if } \frac{1}{2} \leq \omega < 1, \end{cases} \quad (12)$$

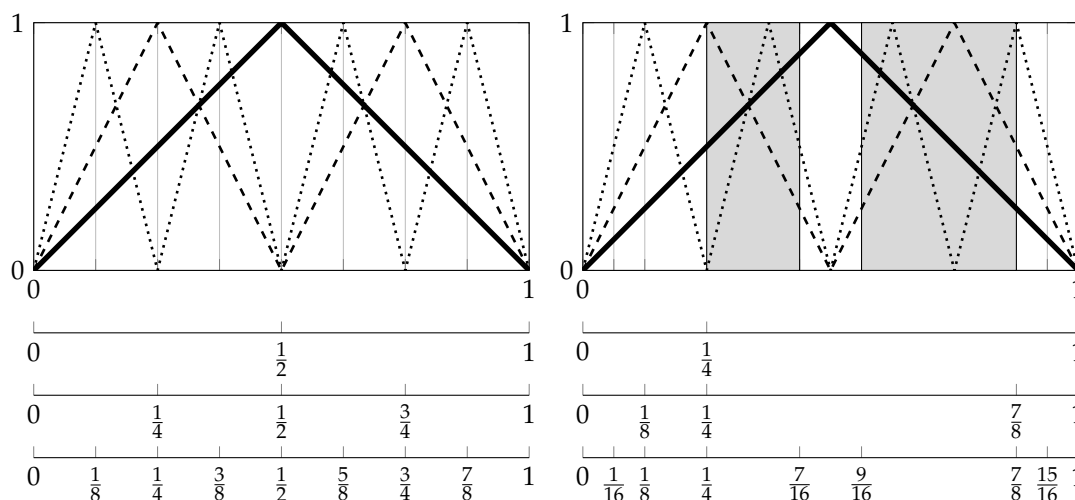
i.e.,  $T$  is the full tent map on  $\Omega := [0, 1)$ . The KS entropy is  $\ln(2)$  (see for instance Bollt et al. [5] and the references given therein), and in fact, it holds:

$$H_\mu(\mathcal{C}) = \frac{1}{2}H_\mu((\mathcal{C})_2) = \frac{1}{3}H_\mu((\mathcal{C})_3) = \dots = \frac{1}{t}H_\mu((\mathcal{C})_t) = \dots = \ln(2),$$

whereby:

$$H_\mu(\mathcal{D}) = -\frac{1}{4}\ln\left(\frac{1}{4}\right) - \frac{3}{4}\ln\left(\frac{3}{4}\right) < \ln(2)$$

(see Section 4.1). Since  $\frac{1}{t}H_\mu((\mathcal{C})_t)$  decreases to  $h_\mu(T, \mathcal{C})$  for any finite partition  $\mathcal{C} \subset \mathcal{A}$  of  $\Omega$  (see Walters [10], Chapter 4) and by Theorem 4.18 of [10] (see Equation (1)), we have by  $\mathcal{C}$  a generating and by  $\mathcal{D}$  a non-generating (misplaced) partition under  $T$  (see Bollt et al. [5] and Steuer et al. [7] for detailed information about possible consequences if a non-generating partition is used in time series analysis, as well as Figure 7).



**Figure 4.** Different partition sequences on  $[0, 1)$  with respect to the full tent map (see Equation (12)). Left: initial partition  $\mathcal{C} = (\mathcal{C})_1 = \{[0, \frac{1}{2}), [\frac{1}{2}, 1)\}$ ,  $(\mathcal{C})_2$  and  $(\mathcal{C})_3$ , which are generating under  $T$ . Right: initial partition  $\mathcal{D} = (\mathcal{D})_1 = \{[0, \frac{1}{4}), [\frac{1}{4}, 1)\}$ ,  $(\mathcal{D})_2$  and  $(\mathcal{D})_3$ , which are not generating under  $T$  (see Section 2.1). Detailed information about possible consequences if a non-generating partition is used in time series analysis is given by Bollt et al. [5].

## 2.2. Ordinal Symbolic Dynamics

Ordinal symbolic dynamics is a relatively new symbolic approach going back to Bandt and Pompe [15] and applied in various fields (see for instance Zanin et al. [2], Amigó et al. [3] and the references given therein). The idea of the symbolization scheme is to partition the state space according to ordinal patterns of orders  $d \in \mathbb{N}$ . For fixed  $d$  and a random vector  $\mathbf{X} = (X_i)_{i=1}^n$ , two states  $\omega_1, \omega_2 \in \Omega$  belong to the same part of a partition if for each  $i = 1, 2, \dots, n$ , the observable  $X_i$  provides the same order relations on the orbits of length  $d + 1$  of  $\omega_1$  and  $\omega_2$ :

For all  $s, t$  with  $0 \leq s < t \leq d$ , it holds:

$$X_i(T^{\circ s}(\omega_1)) \geq X_i(T^{\circ t}(\omega_1)) \text{ iff } X_i(T^{\circ s}(\omega_2)) \geq X_i(T^{\circ t}(\omega_2)).$$

One easily sees that the obtained partitions can be written in the form  $\mathcal{C}_d = \mathcal{C}^{\mathcal{R}, E_d}(T, \mathbf{X})$  with:

$$\mathcal{R} = \{ \{(x, y) \in \mathbb{R} \times \mathbb{R} \mid x \geq y\}, \{(x, y) \in \mathbb{R} \times \mathbb{R} \mid x < y\} \} \text{ (see Figure 2)} \quad (13)$$

and the full timing (see Equation (9)).

The sequence  $(\mathcal{C}_d)_{d \in \mathbb{N}}$  is obviously increasing. Antoniouk et al. [16] show the following statement for  $\mathcal{R}$  as given by (13) and the full timing  $(E_d)_{d \in \mathbb{N}}$  given by (9), here formulated in the language of our general approach:

If  $T$  is ergodic and  $\mathbf{X}$  satisfies (6) or weaker (7), then

$$(\mathcal{C}_d)_{d \in \mathbb{N}} = (\mathcal{C}^{\mathcal{R}, E_d}(T, \mathbf{X}))_{d \in \mathbb{N}} \text{ is generating, hence } h_\mu^{\text{KS}}(T) = \lim_{d \rightarrow \infty} h_\mu(T, \mathcal{C}_d) = \sup_{d \in \mathbb{N}} h_\mu(T, \mathcal{C}_d). \quad (14)$$

Unlike the “classical” approach, the basic symbolization scheme  $\mathcal{R}$ , as given by (13), regards a kind of dependency between two measurements by  $X_i$ . The statement (14) shows the substantial difference between “classical” and ordinal symbolic dynamics: by using  $\mathcal{R}$  as given in (13), we obtain a generating sequence  $(\mathcal{C}_d)_{d \in \mathbb{N}}$ , regardless of whether  $\mathcal{C}_1 = \mathcal{C}$  is generating under  $T$ .

### 2.3. An Extension of Ordinal Symbolic Dynamics

In the rest of the paper, we discuss for which  $\mathcal{R}$  and  $(E_d)_{d \in \mathbb{N}}$  the statement (14) remains true; whereby, this section is dedicated to those readers who are mainly interested in the idea and results of our study. Since the corresponding considerations to validate our subsequent statement are fairly technical and due to even more general results, we refer here only to the statements and proofs in the later discussion (Section 3 and the Appendix). First of all, obviously, (14) remains true when  $\mathcal{R}$  is substituted by a refinement of (13) (see Figure 5). Moreover, in the case that  $\Omega$  is a Borel subset of  $\mathbb{R}^n$  and  $X_i, i = 1, 2, \dots, n$  is the  $i$ -th coordinate projection (see the closing remarks of Section 1.3), as well as  $\mu(A) > 0$  for all open subsets  $A$  of  $\Omega$ , Statement (14) remains true if we modify  $\mathcal{R}$  even more (see the closing remarks of Section 3.1):

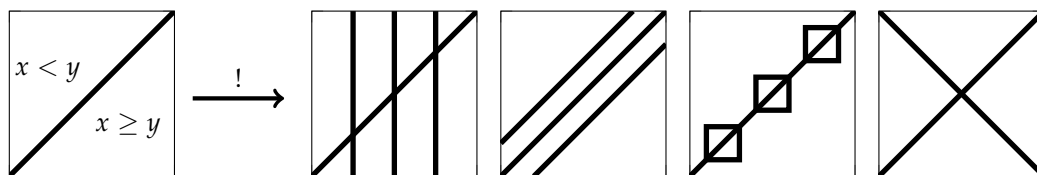


Figure 5. Statement (14) obviously remains true if  $\mathcal{R}$  is substituted by a refinement of (13).

**Theorem 1.** Let  $\Omega$  be a Borel subset of  $\mathbb{R}^n$  and  $X_i; i = 1, 2, \dots, n$  be the  $i$ -th coordinate projection. Further, let  $\mathcal{R}$  be a basic symbolization scheme defined by:

$$\mathcal{R} = \{ \{ (x, y) \in \mathbb{R} \times \mathbb{R} \mid g(x) \geq y \}, \{ (x, y) \in \mathbb{R} \times \mathbb{R} \mid g(x) < y \} \}, \quad (15)$$

or finer, where  $g : \mathbb{R} \leftrightarrow \mathbb{R}$  is a one-to-one  $\mathcal{B}(\mathbb{R})$ - $\mathcal{B}(\mathbb{R})$  measurable map with  $\mathcal{B}(\mathbb{R})$  being the Borel  $\sigma$ -algebra on  $\mathbb{R}$  (see for instance Figure 6). If  $\mu(A) > 0$  for all open subsets  $A$  of  $\Omega$ , then for the full timing  $(E_d)_{d \in \mathbb{N}}$  (see Equation (9)), Statement (14) is fulfilled.

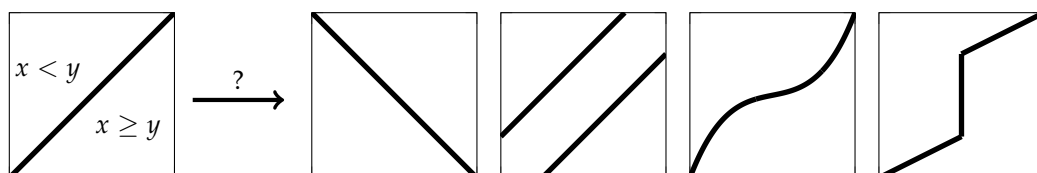


Figure 6. Does Statement (14) remain true if arbitrary symbolization schemes are considered along with the timing given by (9) (see Sections 2.3 and 3)?

## 3. Main Mathematical Results

Antoniouk et al. [16] show that the search for a generating partition under  $T$  can be bypassed by choosing ordinal symbolic dynamics. It namely provides, in the ergodic case and if  $\mathbf{X}$  satisfies (6) or weaker (7), a generating sequence of finite partitions by default, that is the generating property is valid regardless of the properties of the original system considered (see Statement (14) in Section 2.2).

The question arises if other symbolic approaches deliver similar results. In fact, by generalizing the ideas and results of Antoniouk et al. [16], we give sufficient conditions on  $\mathcal{R}$  and  $(E_d)_{d \in \mathbb{N}}$  for (14). This we present the interested reader in the next sections.

### 3.1. Preserving the Information of Observables

The following is quite technical, but shows under which conditions the information given by the observables is preserved if basic symbolization schemes  $\mathcal{R}$  such as given by (15), or finer, are considered. Hence, if Theorem 2 holds and  $\mathcal{A} \overset{\mu}{\subset} \sigma(\mathbf{X})$ , then  $(\mathcal{C}_d)_{d \in \mathbb{N}}$  is generating.

**Theorem 2.** Let  $T$  be ergodic,  $\mathbf{X} = (X_i)_{i=1}^n$  a random vector,  $\mathcal{R}$  a basic symbolization scheme,  $(E_d)_{d \in \mathbb{N}}$  a timing and  $(\mathcal{C}_d)_{d \in \mathbb{N}}$  a sequence of finite partitions constructed from  $\mathcal{R}$  and  $(E_d)_{d \in \mathbb{N}}$  (see Equation (8)). If further:

- (i)  $g$  is admissible with respect to  $X_i$  for all  $i = 1, 2, \dots, n$ ,
- (ii)  $F_{X_i}$  is admissible with respect to  $g \circ X_i$  for all  $i = 1, 2, \dots, n$  and
- (iii)  $\{g \circ X_i \geq X_i \circ T^{ot}\} \in \sigma\left((\mathcal{C}^{\mathcal{R}, E_d}(T, X_i))_{d \in \mathbb{N}}\right)$  for all  $i = 1, 2, \dots, n$  and  $t \in \mathbb{N}_0$ ,

then:

$$\sigma(\mathbf{X}) \stackrel{\mu}{\subset} \sigma((\mathcal{C}_d)_{d \in \mathbb{N}}).$$

We call a function  $\phi : \mathbb{R} \leftarrow$  admissible with respect to a random variable  $Y$  on  $\Omega$  if  $\sigma(Y) \stackrel{\mu}{\subset} \sigma(\phi \circ Y)$ ; this is for example the case if  $\phi$  is a one-to-one  $\mathcal{B}(\mathbb{R})$ - $\mathcal{B}(\mathbb{R})$  measurable map (see the closing remarks of the Appendix on one-to-one maps and Lemma A3 for general conditions on  $\phi$  such that  $\phi$  is admissible). Requiring that  $F_{X_i}$  has to be admissible with respect to  $g \circ X_i$  means that  $g$  has to be constructed in such a way that:

$$\sigma(g \circ X_i) \stackrel{\mu}{\subset} \sigma(F_{X_i} \circ g \circ X_i)$$

holds (compare to the proof of Lemma A3 and subsequent remarks). This assumption is redundant if  $\Omega$  is a Borel subset of  $\mathbb{R}^n$ ; each  $X_i$ ;  $i = 1, 2, \dots, n$  is the  $i$ -th coordinate projection (see the closing remarks of Section 1.3); and  $\mu(A) > 0$  for all open subsets  $A$  of  $\Omega$ , because then,  $F_{X_i}$  is one-to-one (see the closing remarks of the Appendix on one-to-one maps). Finally, note that symbolizations based on  $\mathcal{R}$  as given by (15), or finer, have the property (iii) of Theorem 2 (in particular, compare (iii) of Theorem 2 to the structure of (15)). Summarizing, the assumptions of Theorem 2 are generalizations of the assumptions of Theorem 1, and thus, Theorem 1 follows by Theorem 2.

### 3.2. Preserving the Information of the Measuring Process

In this section, we state sufficient conditions such that the information given by the measuring process is preserved. Therefore, if these conditions are fulfilled and  $\mathcal{A} \stackrel{\mu}{\subset} \sigma\left((\mathbf{X} \circ T^{ot})_{t \in \mathbb{N}}\right)$ , then  $(\mathcal{C}_d)_{d \in \mathbb{N}}$  is generating. Define:

$$\mathcal{C}^{\mathcal{R}, E_d}(T, \mathbf{Y}) := \bigvee_{i=1}^n \bigvee_{(s,t) \in E_d} (Y_i \circ T^{os}, Y_i \circ T^{ot})^{-1}(\mathcal{R}) \subset \mathcal{A},$$

where  $\mathbf{Y} = (Y_i)_{i=1}^n$  is an arbitrary random vector (compare to Equation (8)), and consider the special case  $\mathbf{Y} = (Y_i)_{i=1}^n = \mathbf{X} \circ T^{ol} = (X_i \circ T^{ol})_{i=1}^n$  for some  $l \in \mathbb{N}$  in the following.

**Definition 2.** Let  $\mathcal{R}$  be a basic symbolization scheme and  $(E_d)_{d \in \mathbb{N}}$  be a timing. We call the tuple  $(\mathcal{R}, (E_d)_{d \in \mathbb{N}})$  consistent if for all  $t > 1$  and  $d \in \mathbb{N}$ , it holds:

$$\bigvee_{s=0}^{t-1} \mathcal{C}^{\mathcal{R}, E_d}(T, \mathbf{X} \circ T^{os}) \prec \bigvee_{s=0}^{t-2} \mathcal{C}^{\mathcal{R}, E_{d+1}}(T, \mathbf{X} \circ T^{os}). \quad (16)$$

Compare to Keller et al. [17], who regard the ordinal approach: observe that, in the consistent case, by applying (16) repeatedly, one shows that:

$$\bigvee_{s=0}^{t-1} \mathcal{C}^{\mathcal{R}, E_d}(T, \mathbf{X} \circ T^{os}) \prec \mathcal{C}^{\mathcal{R}, E_{d+t-1}}(T, \mathbf{X})$$

for all  $t, d \in \mathbb{N}$ . Consistency ensures that for all  $t \in \mathbb{N}_0$ , it holds:

$$\sigma\left(\left(\mathcal{C}^{\mathcal{R}, E_d}(T, \mathbf{X} \circ T^{ot})\right)_{d \in \mathbb{N}}\right) \subset \sigma((\mathcal{C}_d)_{d \in \mathbb{N}}).$$

In other words, if the information given by a measurement at a time  $t \in \mathbb{N}$  is preserved by  $(\mathcal{C}^{\mathcal{R}, E_d}(T, \mathbf{X} \circ T^{\circ t}))_{d \in \mathbb{N}}$ , then it also is preserved by  $(\mathcal{C}_d)_{d \in \mathbb{N}}$  (compare to the proof of Theorem 3).

Consistency depends on the interplay of the underlying system, the considered random vector  $\mathbf{X} = (X_i)_{i=1}^n$ ,  $\mathcal{R}$  and  $(E_d)_{d \in \mathbb{N}}$ ; however, a skillful choice of the timing guarantees that  $(\mathcal{R}, (E_d)_{d \in \mathbb{N}})$  is consistent independent of the system and  $\mathbf{X}$ . This we discuss in Section 4.3. Note here that  $(\mathcal{R}, (E_d)_{d \in \mathbb{N}})$  is always consistent if  $(E_d)_{d \in \mathbb{N}}$  is the full timing (see Equation (9)); however, the tuple is not consistent in general if a timing given by:

$$E_d = \{(0, t) \mid t \in \{0, 1, 2, \dots, d\}\}; d \in \mathbb{N} \quad (17)$$

is considered.

**Theorem 3.** Let  $\mathbf{X} = (X_i)_{i=1}^n$  be a random vector,  $\mathcal{R}$  a basic symbolization scheme,  $(E_d)_{d \in \mathbb{N}}$  a timing and  $(\mathcal{C}_d)_{d \in \mathbb{N}}$  a sequence of finite partitions constructed by  $\mathcal{R}$  and  $(E_d)_{d \in \mathbb{N}}$  (see Equation (8)). If further:

- (i)  $\sigma(\mathbf{X}) \stackrel{\mu}{\subset} \sigma((\mathcal{C}_d)_{d \in \mathbb{N}})$  and
- (ii)  $(\mathcal{R}, (E_d)_{d \in \mathbb{N}})$  is consistent,

then:

$$\sigma((\mathbf{X} \circ T^{\circ t})_{t \in \mathbb{N}}) \stackrel{\mu}{\subset} \sigma((\mathcal{C}_d)_{d \in \mathbb{N}}).$$

Recall that (i) of Theorem 3 particularly holds if Conditions (i)–(iii) of Theorem 2 are fulfilled.

#### 4. Proofs

We begin by summarizing some basic notations and concepts. Thereafter, we prove Theorem 2 and Theorem 3, whereby quite technical and complicated lemmas can be found in the Appendix.

##### 4.1. Preliminaries

Our whole discussion is concerned with finite partitions  $\mathcal{C} \subset \mathcal{A}$  of  $\Omega$  and with sequences of finite partitions  $(\mathcal{C}_d)_{d \in \mathbb{N}}$ . A finite partition  $\mathcal{C}$  of  $\Omega$  is a set system:

$$\mathcal{C} := \{C^{(1)}, C^{(2)}, \dots, C^{(q)}\} \subset \mathcal{A}; q \in \mathbb{N},$$

where:

$$\bigcup_{l=1}^q C^{(l)} = \Omega \text{ and } C^{(l)} \cap C^{(k)} = \emptyset \text{ for any } l \neq k \in \{1, 2, \dots, q\}.$$

Particularly, we are interested in such sequences that are increasing, meaning that  $\mathcal{C}_{d+1}$  is finer than  $\mathcal{C}_d$  for all  $d \in \mathbb{N}$ :

A partition  $\mathcal{D} = \{D^{(1)}, D^{(2)}, \dots, D^{(p)}\}$  is finer than a partition  $\mathcal{C} = \{C^{(1)}, C^{(2)}, \dots, C^{(q)}\}$ ; we write  $\mathcal{C} \prec \mathcal{D}$ , or, equivalently,  $\mathcal{C}$  is coarser than  $\mathcal{D}$  if for all  $l \in \{1, 2, \dots, q\}$  there exists a nonempty  $K \subset \{1, 2, \dots, p\}$  such that:

$$C^{(l)} = \bigcup_{k \in K} D^{(k)}.$$

Moreover, we consider the join  $\bigvee_{r=1}^m \mathcal{C}_r$  of finite partitions  $\mathcal{C}_r \subset \mathcal{A}$  of  $\Omega$  with  $m \in \mathbb{N}$  and  $r = 1, 2, \dots, m$ , which is defined by:

$$\bigvee_{r=1}^m \mathcal{C}_r = \left\{ \bigcap_{r=1}^m C_r^{(l_r)} \neq \emptyset \mid l_r \in \{1, 2, \dots, |\mathcal{C}_r|\} \text{ for } r = 1, 2, \dots, m \right\},$$

that is the coarsest partition refining all  $\mathcal{C}_r$ .

Furthermore, sub- $\sigma$ -algebras of  $\mathcal{A}$  are central to us, especially the ones  $\sigma(\mathcal{M})$  generated by subsets  $\mathcal{M}$  of  $\mathcal{A}$ , i.e., the smallest sub- $\sigma$ -algebra containing  $\mathcal{M}$ :

$$\sigma(\mathcal{M}) = \bigcap_{\substack{\mathcal{F} \subset \mathcal{A} \text{ is } \sigma\text{-algebra} \\ \text{and } \mathcal{M} \subset \mathcal{F}}} \mathcal{F}.$$

We consider also the join of  $\sigma$ -algebras  $\mathcal{F}_i \subset \mathcal{A}; i \in I$  defined by:

$$\bigvee_{i \in I} \mathcal{F}_i = \sigma \left( \bigcup_{i \in I} \mathcal{F}_i \right).$$

Overall, we have a special interest in the following sub- $\sigma$ -algebras of  $\mathcal{A}$ :

$$\sigma((\mathcal{C}_d)_{d \in \mathbb{N}}) = \bigvee_{d \in \mathbb{N}} \sigma(\mathcal{C}_d)$$

for  $(\mathcal{C}_d)_{d \in \mathbb{N}}$  being a finite partition sequence of  $\Omega$ ,

$$\begin{aligned} \sigma(\mathbf{X}) &= \bigvee_{i=1}^n \sigma(X_i) = \bigvee_{i=1}^n \left\{ X_i^{-1}(B) \mid B \in \mathcal{B}(\mathbb{R}) \right\}, \\ \sigma(\mathbf{X} \circ T^{\circ s}) &= \bigvee_{i=1}^n \sigma(X_i \circ T^{\circ s}) = \bigvee_{i=1}^n \left\{ (X_i \circ T^{\circ s})^{-1}(B) \mid B \in \mathcal{B}(\mathbb{R}) \right\}, \\ \sigma\left((\mathbf{X} \circ T^{\circ t})_{t \in \mathbb{N}_0}\right) &= \bigvee_{t \in \mathbb{N}_0} \sigma(\mathbf{X} \circ T^{\circ t}) = \bigvee_{i=1}^n \bigvee_{t \in \mathbb{N}_0} \sigma(X_i \circ T^{\circ t}) \end{aligned}$$

for  $\mathbf{X} = (X_i)_{i=1}^n$  being a random vector on  $\Omega$  and  $s \in \mathbb{N}$ .

We close this subsection by giving an exact definition of the entropy rate  $h_\mu(T, \mathcal{C})$  of a finite partition  $\mathcal{C}$  (see Figure 3): one assigns to every part  $\mathcal{C}^{(i)}$  of the partition  $\mathcal{C} = \{\mathcal{C}^{(1)}, \mathcal{C}^{(2)}, \dots, \mathcal{C}^{(q)}\}$  the letter  $i$  of the alphabet  $A = \{1, 2, \dots, q\}$ . Each word  $(a_1, a_2, \dots, a_t)$  of length  $t \in \mathbb{N}$  over  $A$  defines a set:

$$\mathcal{C}^{(a_1, a_2, \dots, a_t)} = \left\{ \omega \in \Omega \mid \left( \omega, T(\omega), \dots, T^{\circ t-1}(\omega) \right) \in \mathcal{C}^{(a_1)} \times \mathcal{C}^{(a_2)} \times \dots \times \mathcal{C}^{(a_t)} \right\}.$$

All non-empty sets  $\mathcal{C}^{(a_1 a_2 \dots a_t)}$  provide a partition  $(\mathcal{C})_t \subset \mathcal{A}$  of  $\Omega$ . We use the notation  $(\mathcal{C})_t$  to emphasize that the partition is constructed with respect to  $T$ . In particular,  $(\mathcal{C})_1 = \mathcal{C}$ . The entropy rate of  $T$  with respect to  $\mathcal{C}$  is given by:

$$h_\mu(T, \mathcal{C}) = \lim_{t \rightarrow \infty} \frac{1}{t} H_\mu((\mathcal{C})_t) = \lim_{t \rightarrow \infty} (H_\mu((\mathcal{C})_t) - H_\mu((\mathcal{C})_{t-1})), \quad (18)$$

where  $H_\mu((\mathcal{C})_t)$  is the Shannon entropy of  $(\mathcal{C})_t$ , that is for a finite partition  $\mathcal{D} = \{D^{(1)}, D^{(2)}, \dots, D^{(p)}\}$ :

$$H_\mu(\mathcal{D}) = - \sum_{l=1}^p \mu(D^{(l)}) \ln(\mu(D^{(l)})) \text{ (with } 0 \ln(0) := 0).$$

For a fuller treatment, e.g., for statements that the limit in Equation (18) exists and  $\frac{1}{t} H_\mu((\mathcal{C})_t)$ , as well as  $H_\mu((\mathcal{C})_t) - H_\mu((\mathcal{C})_{t-1})$  decreases to  $h_\mu(T, \mathcal{C})$ , we refer the reader to Chapter 4 of Walters [10]. Note that for all  $t \in \mathbb{N}$ , it holds  $(\mathcal{C})_{t-1} \prec (\mathcal{C})_t$ , and:

$$H_\mu((\mathcal{C})_t) - H_\mu((\mathcal{C})_{t-1}) \leq \frac{1}{t} \sum_{s=1}^t H_\mu((\mathcal{C})_s) - H_\mu((\mathcal{C})_{s-1}) = \frac{1}{t} H_\mu((\mathcal{C})_s). \quad (19)$$

Moreover, we say that  $\mathcal{C}$  is generating under  $T$  if:

$$\mathcal{A} \overset{\mu}{\subset} \sigma((\mathcal{C})_{t \in \mathbb{N}}).$$

If we consider instead an arbitrary sequence of finite partitions  $(\mathcal{C}_d)_{d \in \mathbb{N}}$  for which:

$$\mathcal{A} \overset{\mu}{\subset} \sigma((\mathcal{C}_d)_{d \in \mathbb{N}})$$

holds, then we call  $(\mathcal{C}_d)_{d \in \mathbb{N}}$  just generating.

#### 4.2. Proof of Theorem 2

In order to prove Theorem 2, we generalize the results of Antoniouk et al. [16] (Lemmas 3.2, 3.3 and Corollary 3.4) and extend their proofs. Thereby, we utilize properties of the distribution function  $F_{X_i} : \mathbb{R} \rightarrow [0, 1]$  of a random variable  $X_i : \Omega \rightarrow \mathbb{R}$ , i.e.,

$$F_{X_i}(a) = \mu(\{\omega \in \Omega \mid X_i(\omega) \leq a\})$$

for all  $a \in \mathbb{R}$  (see Lemmas A1 and A2 in the Appendix) and show:

$$\sigma(X_i) \overset{\mu}{\subset} \sigma\left(\left(\mathcal{C}^{\mathcal{R}, E_d}(T, X_i)\right)_{d \in \mathbb{N}}\right)$$

by the detour:

$$\sigma(X_i) \overset{\mu}{\subset} \sigma(g \circ X_i) \overset{\mu}{\subset} \sigma(F_{X_i} \circ g \circ X_i) \overset{\mu}{\subset} \sigma\left(\left(\mathcal{C}^{\mathcal{R}, E_d}(T, X_i)\right)_{d \in \mathbb{N}}\right).$$

**Proof of Theorem 2.** Note that it is enough to show that:

$$\sigma(X_i) \overset{\mu}{\subset} \sigma\left(\left(\mathcal{C}^{\mathcal{R}, E_d}(T, X_i)\right)_{d \in \mathbb{N}}\right)$$

holds for all  $i = 1, 2, \dots, n$  since the sub- $\sigma$ -algebras  $\sigma\left(\left(\mathcal{C}^{\mathcal{R}, E_d}(T, X_i)\right)_{d \in \mathbb{N}}\right)$  generate the sub- $\sigma$ -algebra:

$$\sigma((\mathcal{C}_d)_{d \in \mathbb{N}}) = \sigma\left(\left(\mathcal{C}^{\mathcal{R}, E_d}(T, \mathbf{X})\right)_{d \in \mathbb{N}}\right).$$

Thus, let us regard  $i \in \mathbb{N}$  as fixed. By the assumptions Theorem 2(i) and (ii) (compare also to Lemma A3), we obtain:

$$\sigma(X_i) \overset{\mu}{\subset} \sigma(g \circ X_i) \overset{\mu}{\subset} \sigma(F_{X_i} \circ g \circ X_i).$$

Moreover, by Theorem 2(iii) and Lemma A2, we have that:

$$\sigma(F_{X_i} \circ g \circ X_i) \overset{\mu}{\subset} \sigma\left(\left(\mathcal{C}^{\mathcal{R}, E_d}(T, X_i)\right)_{d \in \mathbb{N}}\right).$$

Hence:

$$\sigma(X_i) \overset{\mu}{\subset} \sigma\left(\left(\mathcal{C}^{\mathcal{R}, E_d}(T, X_i)\right)_{d \in \mathbb{N}}\right),$$

which completes the proof.  $\square$

### 4.3. Proof of Theorem 3

The main additional property needed in Theorem 3 is that  $(\mathcal{R}, (E_d)_{d \in \mathbb{N}})$  is consistent (see Definition 2). In order to see how a timing has to be constructed such that  $(\mathcal{R}, (E_d)_{d \in \mathbb{N}})$  is generally consistent, let us regard  $t, d \in \mathbb{N}$  as fixed and consider states  $\omega_1, \omega_2 \in \Omega$ , which are in different parts of:

$$\bigvee_{s=0}^{t-1} \mathcal{C}^{\mathcal{R}, E_d}(T, \mathbf{X} \circ T^{\circ s}).$$

Then, there exists at least one  $s \in \{0, 1, \dots, t-1\}$ , one  $i = 1, 2, \dots, n$  and a time pair  $(u, v) \in E_d$  such that:

$$(X_i(T^{\circ s+u}(\omega_1)), X_i(T^{\circ s+v}(\omega_1))) \text{ and } (X_i(T^{\circ s+u}(\omega_2)), X_i(T^{\circ s+v}(\omega_2)))$$

are in different parts of  $\mathcal{R}$ . If now for any  $s \in \{0, 1, \dots, t-1\}$  and  $(u, v) \in E_d$ , it holds that  $(s+u, s+v) \in E_{d+t-1}$ , then:

$$(X_i \circ T^{\circ s+u}, X_i \circ T^{\circ s+v})^{-1}(\mathcal{R}) \prec \mathcal{C}_{d+t-1}.$$

Hence,  $(\mathcal{R}, (E_d)_{d \in \mathbb{N}})$  is consistent if the previous holds for all  $t, d \in \mathbb{N}$ .

**Proof of Theorem 3.** Compare to [16] (Corollary 3.5). By Theorem 3(ii), i.e.,  $(\mathcal{R}, (E_d)_{d \in \mathbb{N}})$  is consistent, it holds:

$$\mathcal{C}^{\mathcal{R}, E_d}(T, \mathbf{X} \circ T) \prec \mathcal{C}^{\mathcal{R}, E_{d+1}}(T, \mathbf{X})$$

for all  $d \in \mathbb{N}$ . These refinements imply:

$$\sigma\left(\left(\mathcal{C}^{\mathcal{R}, E_d}(T, \mathbf{X} \circ T^{\circ t})\right)_{d \in \mathbb{N}}\right) \subset \sigma((\mathcal{C}_d)_{d \in \mathbb{N}})$$

for all  $t \in \mathbb{N}_0$  (see Walters [10], Chapter 4, Section 1). Moreover, by Theorem 3(i), it holds:

$$\sigma(\mathbf{X} \circ T^{\circ t}) \stackrel{\mu}{\subset} \sigma\left(\left(\mathcal{C}^{\mathcal{R}, E_d}(T, \mathbf{X} \circ T^{\circ t})\right)_{d \in \mathbb{N}}\right)$$

for all  $t \in \mathbb{N}_0$ . Hence:

$$\sigma\left(\left(\mathbf{X} \circ T^{\circ t}\right)_{t \in \mathbb{N}_0}\right) \stackrel{\mu}{\subset} \sigma((\mathcal{C}_d)_{d \in \mathbb{N}}),$$

which proves the theorem.  $\square$

## 5. Some Remarks

In proposing and studying our universal symbolic approach toward KS entropy, we have restricted ourselves to the kernel ideas. In particular, we have attached importance to point out that a “two-dimensional” symbolization, that is linking two observations, can provide a better basic symbolization scheme than symbolizing only on a one-dimensional observational level. The obtained results can be simply generalized in two directions:

On the one hand, infinitely many observables instead of finitely many ones can be considered. Here, the results obtained by Keller et al. [18] (see also the references given in the paper) can directly be adapted, which leads to a description of the KS entropy by a double limit substituting the limit in (2). On the other hand, some of our results remain true when relaxing ergodicity by some rather general conditions on the dynamics considered. For this, the ergodic decomposition theorem can be utilized. We refer to the discussion in Keller et al. [18] and the references given therein.

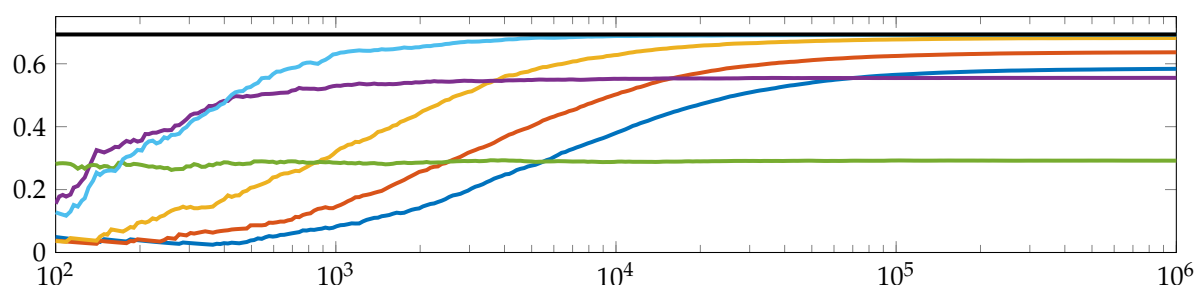
Our study does not touch aspects as the speed of convergence in (2), as a general comparison of basic symbolic schemes and as an entropy estimation, which are, incontestably, very interesting both from the theoretical and practical viewpoint. Our approach provides some theoretical framework within concrete methods for time series and system analysis and can be specified in accordance with requirements given in practice.

In order to give a brief perspective on the matter, we decoded a finite orbit  $(x_t)_{t=0}^T$ ;  $T \in \mathbb{N}$  of the tent map (see Section 2.1), that is:

$$x_{t+1} = \begin{cases} 2x_t & \text{if } 0 \leq x_t < \frac{1}{2}, \\ 2 - 2x_t & \text{if } \frac{1}{2} \leq x_t < 1 \end{cases} \quad (20)$$

for all  $t = 0, 1, \dots, T-1$  and  $x_0$  uniformly distributed, into a sequence of symbols, fixed a word length  $t \in \mathbb{N}$  and naively estimated the difference  $H_\mu((C)_t) - H_\mu((C)_{t-1})$  by replacing the probabilities by relative frequencies of symbol word occurrences.

Figure 7 shows the results for different  $t \in \mathbb{N}$  and symbolization schemes in dependence of the orbit length  $T$ , here between  $10^2$  and  $10^6$ . We chose to take the difference because of (19), i.e., the difference for fixed  $t$  is a better approximation of the entropy rate  $h_\mu(T, C)$  than  $\frac{1}{t}H_\mu((C)_t)$ . For a fuller treatment, we refer the reader to Keller et al. [19], in particular for more information of how to construct a symbol sequence with respect to ordinal symbolic dynamics with fixed order  $d \in \mathbb{N}$  and word length  $t \in \mathbb{N}$ .



**Figure 7.** Different estimates of the Kolmogorov-Sinai entropy (black line:  $\ln(2)$ ) of the tent map (see Equation (20)) for different orbit lengths by naively estimating  $H_\mu((C)_t) - H_\mu((C)_{t-1})$ . Dark blue, red, yellow: ordinal symbolization scheme with respect to (13) and (9) for  $d = 3, 6, 8$  and  $t = 12, 6, 2$  (see Keller et al. [19] for a fuller treatment). Green, purple, light blue: “classical” symbolization scheme with respect to misplaced partitions  $\{[0, 0.9), [0.9, 1)\}$ ,  $\{[0, 0.4), [0.4, 1)\}$  and to the generating (under the dynamics) partition  $\{[0, 0.5), [0.5, 1)\}$  for  $d = 1$  and  $t = 8$ .

Clearly, Figure 7 emphasizes common problems of time series analysis, which have to be faced, as, for example, the trade-off between computational capacity and computation accuracy, which includes undersampling problems, the choice of parameters, stationarity assumptions, and so forth (see for instance Keller et al. [19]). A discussion of all these aspects would be beyond the scope of this paper, but is planned for the future.

**Acknowledgments:** We acknowledge financial support by Land Schleswig-Holstein within the funding program Open Access Publikationsfonds.

**Author Contributions:** Inga Stolz and Karsten Keller jointly conceived and discussed the presented ideas. Inga Stolz developed the theory and performed the numerical simulations. Karsten Keller supervised the course of action and findings. Both authors contributed materially to the paper and read as well as approved the final manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix

**Lemma A1.** Let  $X, Y : \Omega \rightarrow \mathbb{R}$  be two random variables,  $F_X : \mathbb{R} \rightarrow [0, 1]$  the distribution function of  $X$  and  $I_d^{X,Y} : \Omega \rightarrow [0, 1]$  be the counting map of  $X$  and  $Y$ , given by:

$$I_d^{X,Y}(\omega) := \frac{1}{d} \sum_{i=1}^d \mathbf{1}_{\{X \circ T^{oi} \leq Y\}}(\omega)$$

for all  $\omega \in \Omega$ . If  $T$  is ergodic, then:

$$\lim_{d \rightarrow \infty} I_d^{X,Y} = F_X \circ Y \text{ } \mu \text{ almost everywhere.}$$

**Proof of Lemma A1.** Let  $A_a = X^{-1}((-\infty, a])$  for any  $a \in \mathbb{R}$ . By Birkhoff's ergodic theorem (see for instance Walters [10]), there exists a set  $N_a \subset \Omega$  such that  $\mu(N_a) = 0$  and:

$$F_X(a) = \mu(A_a) = \lim_{d \rightarrow \infty} \frac{1}{d} \sum_{t=1}^d \mathbf{1}_{\{X \circ T^{ot} \leq a\}}(\omega) \quad (\text{A1})$$

for any  $a \in \mathbb{R}$  and  $\omega \in \Omega \setminus N_a$ . Let  $B$  be a countable dense subset of  $\mathbb{R}$  such that it includes all  $a \in \mathbb{R}$  for which  $F_X$  is discontinuous, and let  $N = \bigcup_{a \in B} N_a$ . Then,  $\mu(N) = 0$ , and Equation (A1) holds for each  $a \in B$  and  $\omega \in \Omega \setminus N$ . Our next claim is that for all  $\omega \in \Omega \setminus N$ , it holds:

$$\lim_{d \rightarrow \infty} I_d^{X,Y}(\omega) = F_X(Y(\omega)).$$

By (A1), this is true if  $\omega \in \Omega \setminus N$  satisfies  $a := Y(\omega) \in B$ . It is moreover true if  $\omega \in \Omega \setminus N$  and  $a := Y(\omega) \in \mathbb{R} \setminus B$ , which we show in the following:

Let  $(b_i)_{i \in \mathbb{N}}$  and  $(c_i)_{i \in \mathbb{N}}$  be two sequences converging to  $a$  with:

$$b_i \in B \cap (-\infty, a) \text{ and } c_i \in B \cap (a, \infty)$$

for all  $i \in \mathbb{N}$ . Hence,  $b_i < a < c_i$ , and for all  $i \in \mathbb{N}$ , it holds:

$$F_X(b_i) = \lim_{d \rightarrow \infty} \frac{1}{d} \sum_{t=1}^d \mathbf{1}_{\{X \circ T^{ot} \leq b_i\}}(\omega) \text{ and } F_X(c_i) = \lim_{d \rightarrow \infty} \frac{1}{d} \sum_{t=1}^d \mathbf{1}_{\{X \circ T^{ot} \leq c_i\}}(\omega)$$

since  $\omega \in \Omega \setminus N$ . Thus, for all  $d \in \mathbb{N}$ , we have:

$$\sum_{t=1}^d \mathbf{1}_{\{X \circ T^{ot} \leq b_i\}}(\omega) \leq \sum_{t=1}^d \mathbf{1}_{\{X \circ T^{ot} \leq Y\}}(\omega) \leq \sum_{t=1}^d \mathbf{1}_{\{X \circ T^{ot} \leq c_i\}}(\omega).$$

Furthermore, since  $F_X$  is continuous at  $a$ , we obtain:

$$F_X(a) = \lim_{i \rightarrow \infty} F_X(b_i) \leq \liminf_{d \rightarrow \infty} I_d^{X,Y} \leq \limsup_{d \rightarrow \infty} I_d^{X,Y} \leq \lim_{i \rightarrow \infty} F_X(c_i) = F_X(a).$$

Hence, we can summarize that for all  $\omega \in \Omega \setminus N$ , it holds  $\lim_{d \rightarrow \infty} I_d^{X,Y}(\omega) = F_X(Y(\omega))$ , which is the desired conclusion.  $\square$

**Lemma A2.** Let  $\mathbf{X} = (X_i)_{i=1}^n$  and  $\mathbf{Y} = (Y_i)_{i=1}^n$  be two random vectors. If  $T$  is ergodic and:

$$\{Y_i \geq X_i \circ T^{ot}\} \in \sigma\left(\left(\mathcal{C}^{\mathcal{R}, E_d}(T, X_i)\right)_{d \in \mathbb{N}}\right)$$

for all  $i = 1, 2, \dots, n$  and  $t \in \mathbb{N}_0$ , then:

$$\sigma((F_{X_i} \circ Y_i)_{i=1}^n) \stackrel{\mu}{\subset} \sigma((\mathcal{C}_d)_{d \in \mathbb{N}}).$$

**Proof of Lemma A2.** Since the sub- $\sigma$ -algebras  $\sigma\left(\left(\mathcal{C}^{\mathcal{R}, E_d}(T, X_i)\right)_{d \in \mathbb{N}}\right)$  generate the sub- $\sigma$ -algebra:

$$\sigma((\mathcal{C}_d)_{d \in \mathbb{N}}) = \sigma\left(\left(\mathcal{C}^{\mathcal{R}, E_d}(T, \mathbf{X})\right)_{d \in \mathbb{N}}\right),$$

it is enough to show that for all  $i = 1, 2, \dots, n$ , it holds:

$$\sigma(F_{X_i} \circ Y_i) \stackrel{\mu}{\subset} \sigma\left(\left(\mathcal{C}^{\mathcal{R}, E_d}(T, X_i)\right)_{d \in \mathbb{N}}\right).$$

Hence, let us regard  $i$  as fixed. By the assumption given, that is:

$$\{Y_i \geq X_i \circ T^{ot}\} \in \sigma\left(\left(\mathcal{C}^{\mathcal{R}, E_d}(T, X_i)\right)_{d \in \mathbb{N}}\right)$$

for all  $t \in \mathbb{N}_0$ , it holds that  $I_d^{X_i, Y_i}$  is  $\sigma\left(\left(\mathcal{C}^{\mathcal{R}, E_d}(T, X_i)\right)_{d \in \mathbb{N}}\right)$ - $\mathcal{B}([0, 1])$ -measurable for any  $d \in \mathbb{N}$  (see for instance Billingsley [20], remarks on simple real functions in Section 13). Hence:

$$\sigma\left(\left(I_d^{X_i, Y_i}\right)_{d \in \mathbb{N}}\right) \subset \sigma\left(\left(\mathcal{C}^{\mathcal{R}, E_d}(T, X_i)\right)_{d \in \mathbb{N}}\right).$$

Moreover, the limit of  $I_d^{X_i, Y_i}$  as  $d$  approaches infinity exists for each  $\omega \in \Omega$  since  $I_d^{X_i, Y_i} \leq I_{d+1}^{X_i, Y_i}$  and  $0 \leq I_d^{X_i, Y_i} \leq 1$ , hence:

$$\sigma\left(\lim_{d \rightarrow \infty} I_d^{X_i, Y_i}\right) \subset \sigma\left(\left(I_d^{X_i, Y_i}\right)_{d \in \mathbb{N}}\right)$$

(see for instance Billingsley [20], Theorem 13.4.(ii)). Furthermore, by Lemma A1, there exists a set  $N \subset \Omega$  with  $\mu(N) = 0$  such that:

$$\lim_{d \rightarrow \infty} I_d^{X_i, Y_i}(\omega) = F_{X_i}(Y_i(\omega))$$

for all  $\omega \in \Omega \setminus N$ . Hence, for any  $B \in \mathcal{B}([0, 1])$ , it holds:

$$\mu\left(\left(F_{X_i} \circ Y_i\right)^{-1}(B) \triangle \left(\lim_{d \rightarrow \infty} I_d^{X_i, Y_i}\right)^{-1}(B)\right) \leq \mu(N) = 0,$$

which gives:

$$\sigma(F_{X_i} \circ Y_i) \stackrel{\mu}{\subset} \sigma\left(\left(I_d^{X_i, Y_i}\right)_{d \in \mathbb{N}}\right),$$

and the lemma follows.  $\square$

The following lemma yields sufficient conditions for (i) and (ii) of Theorem 2.

**Lemma A3.** Let  $X : (\Omega, \mathcal{A}, \mu) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  be a random variable,  $\phi : \mathbb{R} \leftarrow \mathcal{B}(\mathbb{R})$ - $\mathcal{B}(\mathbb{R})$  a measurable map and  $\mathcal{G}$  a family of subsets of  $\mathbb{R}$  that generates the Borel  $\sigma$ -algebra  $\mathcal{B}(\mathbb{R})$ . If  $\phi$  has the two properties:

- (i)  $\phi(G) \in \mathcal{B}(\mathbb{R})$  and
- (ii)  $\mu(X^{-1}((\phi^{-1}\phi(G)) \setminus G)) = 0$

for all  $G \in \mathcal{G}$ , then  $\sigma(X) \stackrel{\mu}{\subset} \sigma(\phi \circ X)$ .

**Proof of Lemma A3.** Since  $\mathcal{G}$  generates  $\mathcal{B}(\mathbb{R})$ , it holds that  $\sigma(X)$  is generated by the sets  $X^{-1}(\mathcal{G})$  (see for instance Elstrodt [21], Kapitel 1, Satz 4.4). Hence, the Lemma is proven if for any  $G \in \mathcal{G}$ , there exists some  $G' \in \sigma(\phi \circ X)$  such that  $\mu(X^{-1}(G) \triangle G') = 0$ . In order to show this, choose:

$$G' = X^{-1}(\phi^{-1}\phi(G)) = (\phi \circ X)^{-1}\phi(G).$$

By Lemma A3(i), it holds that  $G' \in \sigma(\phi \circ X)$ , and by (ii), we see that:

$$\mu(X^{-1}(G) \triangle G') = \mu(X^{-1}((\phi^{-1}\phi(G)) \setminus G)) = 0,$$

which completes the proof.  $\square$

Note that since  $\phi$  is a  $\mathcal{B}(\mathbb{R})$ - $\mathcal{B}(\mathbb{R})$  measurable map in Lemma A3, it holds in particular:

$$\sigma(X) \supset \sigma(\phi \circ X)$$

for any random variable  $X : \Omega \rightarrow \mathbb{R}$ : Let  $A \in \sigma(\phi \circ X)$ , then:

$$A = (\phi \circ X)^{-1}(B) = X^{-1}\phi^{-1}(B)$$

for some  $B \in \mathcal{B}(\mathbb{R})$ . Hence, by  $\phi^{-1}(B) \in \mathcal{B}(\mathbb{R})$ , it follows that  $A \in \sigma(X)$ .

Lemma A3 is evident if  $\phi$  is a one-to-one  $\mathcal{B}(\mathbb{R})$ - $\mathcal{B}(\mathbb{R})$  measurable map since then:

$$(\phi^{-1}\phi(B)) \setminus B = \emptyset$$

and  $\phi(B) \in \mathcal{B}(\mathbb{R})$  for all  $B \in \mathcal{B}(\mathbb{R})$  (see for instance Cantón et al. [22]); nevertheless, it also includes self-maps such as the distribution function  $F_X$  of a random variable  $X$  (see Antoniouk et al. [16], Lemma 3.2), i.e.,

$$\sigma(X) \stackrel{\mu}{\subset} \sigma(F_X \circ X) :$$

Let  $\mathcal{G} = \{(-\infty, a) \mid a \in \mathbb{R}\}$ ; since  $F_X$  is increasing, Lemma A3(i) holds for all  $G \in \mathcal{G}$ . Assumption Lemma A3(ii) is proven by Antoniouk et al. [16] (Lemma 3.1(3)) by showing firstly that:

$$F_X^{-1}F_X((-\infty, a)) \setminus (-\infty, a)$$

coincides either with the interval  $[a, a^*]$  or with  $[a, a^*)$  for any  $a \in \mathbb{R}$ , where  $a^* = \sup(F_X^{-1}F_X(a))$  and subsequently that  $\mu(X^{-1}([a, a^*])) = 0$ . Hence, the inclusion (compare to Theorem 2(ii)):

$$\sigma(g \circ X) \stackrel{\mu}{\subset} \sigma(F_X \circ g \circ X),$$

where  $g : \mathbb{R} \leftarrow$  is a self-map, holds if we obtain:

$$\mu(X^{-1}g^{-1}([a, a^*])) = 0$$

for any  $a \in \mathbb{R}$ . This is true if, for instance, either  $F_X$  is one-to-one or  $g(\omega) = \omega$  for all  $\omega \in \Omega$  where  $F_X$  is not one-to-one.

## References

1. Daw, C.S.; Finney, C.E.A.; Tracy, E.R. A review of symbolic analysis of experimental data. *Rev. Sci. Instrum.* **2003**, *74*, 915–930.
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.
3. Amigó, J.M.; Keller, K.; Unakafova, V.A. Ordinal symbolic analysis and its application to biomedical recordings. *Phil. Trans. R. Soc. A* **2015**, *373*, 20140091.
4. Crutchfield, J.P.; Packard, N.H. Symbolic dynamics of noisy chaos. *Phys. D: Nonlinear Phenom.* **1983**, *7*, 201–223.
5. Bollt, E.M.; Stanford, T.; Lai, Y.-C.; Życzkowski, K. What symbolic dynamics do we get with a misplaced partition?: On the validity of threshold crossings analysis of chaotic time-series. *Phys. D: Nonlinear Phenom.* **2001**, *154*, 259–286.
6. Kennel, M.B.; Buhl, M. Estimating Good Discrete Partitions from Observed Data: Symbolic False Nearest Neighbors. *Phys. Rev. Lett.* **2003**, *91*, 084102.
7. Steuer, R.; Molgedey, L.; Ebeling, W.; Jiménez-Montaño, M.A. Entropy and optimal partition for data analysis. *Eur. Phys. J. B* **2001**, *19*, 265–269.

8. Letellier, C. Symbolic sequence analysis using approximated partition. *Chaos, Solitons & Fractals* **2008**, *36*, 32–41.
9. Li, Y.; Ray, A. Unsupervised Symbolization of Signal Time Series for Extraction of the Embedded Information. *Entropy* **2017**, *19*, 148.
10. Walters, P. *An Introduction to Ergodic Theory*; Springer: New York, NY, USA, 1981.
11. Takens, F. Detecting strange attractors in turbulence. In *Lecture Notes in Mathematics, Proceedings of a Symposium Held at the University of Warwick 1979/80, Coventry, UK, 1980*; Rand, D., Young, L.S., Eds.; Springer: Berlin/Heidelberg, Germany, 1981; Volume 898, pp. 366–381.
12. Gutman, Y. Takens' embedding theorem with a continuous observable. In *Ergodic Theory: Advances in Dynamical Systems*; Assani, I., Ed.; De Gruyter: Berlin/Heidelberg, Germany, 2016.
13. Kurths, J.; Schwarz, U.; Witt, A.; Krampe, R.T.; Abel, M. Measures of complexity in signal analysis. In Proceedings of the AIP Conference: Chaotic Fractal and Nonlinear Signal Processing, 3rd Technical Conference on Nonlinear Dynamics (Chaos) and Full Spectrum Processing, New London, CT, USA, 10–13 July 1995; Volume 375, pp. 33–54.
14. Einsiedler, M.; Schmidt, K. *Dynamische Systeme: Ergodentheorie und Topologische Dynamik*; Springer: Basel, Switzerland, 2013. (In German)
15. Bandt, C.; Pompe, B. Permutation Entropy: A Natural Complexity Measure for Time Series. *Phys. Rev. Lett.* **2002**, *88*, 174102.
16. Antoniouk, A.; Keller, K.; Maksymenko, S. Kolmogorov-Sinai entropy via separation properties of order-generated sigma-algebras. *Discrete Contin. Dyn. Syst. A* **2014**, *34*, 1793–1809.
17. Keller, K.; Unakafov, A.M.; Unakafova, V.A. On the relation of KS entropy and permutation entropy. *Phys. D: Nonlinear Phenom.* **2012**, *241*, 1477–1481.
18. Keller, K.; Maksymenko, S.; Stolz, I. Entropy determination based on the ordinal structure of a dynamical system. *Discret. Contin. Dyn. Syst. B* **2015**, *20*, 3507–3524.
19. Keller, K.; Mangold, T.; Stolz, I.; Werner, J. Permutation Entropy: New Ideas and Challenges. *Entropy* **2017**, *19*, 134.
20. Billingsley, P. *Probability and Measure*, 3rd ed.; John Wiley & Sons, Inc.: New York, NY, USA, 1995.
21. Elstrodt, J. *Maß- und Integrationstheorie*, 4. korrigierte Auflage; Springer-Verlag: Berlin/Heidelberg, Germany, 2005. (In German)
22. Cantón, A.; Grandos, A.; Pommerenke, C. Borel Images and Analytic Functions. *Mich. Math. J.* **2004**, *52*, 279–287.



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