UNIVERSITÄT ZU LÜBECK

From the Institute of Mathematics
of the University of Lübeck
Director: Prof. Dr. Jürgen Prestin

Determining the Kolmogorov-Sinai entropy by using ordinal patterns

Bestimmen der Kolmogorov-Sinai Entropie mit Hilfe ordinaler Muster

Dissertation

for Fullfillment of Requirements
for the Doctoral Degree of the University of Lübeck
- from the Department of Computer Sciences

submitted by
Tim Gutjahr
from Lübeck, Germany
Lübeck 2021

First referee: Prof. Dr. Karsten Keller

Second referee: Prof. Dr. Alexander Schnurr

Date of oral examination: 17.03.2022

Approved for printing. Lübeck, 21.03.2022

Abstract

In this thesis, we investigate how ordinal patterns can be used to determine the complexity of measure-preserving dynamical systems. Roughly speaking, ordinal patterns describe how a sequence of values can be ordered by size. The permutation entropy, a quantity based on the distribution of ordinal patterns, is compared to the Kolmogorov-Sinai entropy, a measure for the complexity of dynamical systems. One of the main results of this dissertation states that the permutation entropy and Kolmogorov-Sinai entropy are equal for one-dimensional dynamical functions whose domain of definition can be partitioned into a countable number of intervals on which the function is monotone. This is a generalization of a previously existing result, which only allowed for a partition into a finite number of intervals and required continuity.

We further explore if the above mentioned results still hold true when generalizing the permutation entropy by replacing the Shannon entropy used in its definition with the Rényi entropy. We are able to show that this is generally not the case.

Additionally, two types of conditional variants of the permutation entropy are investigated. We establish conditions under which they are equal to the Kolmogorov-Sinai entropy.

Finally, we establish specific conditions for multidimensional dynamical systems under which the permutation and Kolmogorov-Sinai entropy are equal. The methods used to achieve this result also provide an upper bound for the speed of converge of the entropy of ordinal patterns towards the Kolmogorov-Sinai entropy for some specific dynamical systems.

Zusammenfassung

In dieser Dissertation untersuchen wir, wie ordinale Muster genutzt werden können, um die Komplexität von maßerhaltenden dynamischen Systemen zu bestimmen. Ordinale Muster beschreiben, grob gesagt, wie eine Sequenz von Werten nach Größe sortiert werden kann. Die Permutationsentropie, ein Größe basierend auf der Verteilung von ordinalen Mustern, wird mit der Kolmogorov-Sinai Entropie, ein Maß für die Komplexität von dynamischen Systemen, verglichen. Eines der Hauptresultate dieser Dissertation besagt, dass die Permutationsentropie und Kolmogorov-Sinai Entropie für eindimensionale Funktionen gleich sind, wenn sich dessen Definitionsbereich in eine abzählbare Anzahl von Intervallen aufteilen lässt, auf denen die Funktion monoton ist. Dies ist eine Verallgemeinerung eines vorher existierenden Resultates, welches nur eine Partition in endlich viele Intervalle zuließ und Stetigkeit erforderte.

Des Weiteren erforschen wir, ob die oben genannten Resultate weiterhin gelten, wenn man die Permutationsentropie verallgemeinert, indem man die in der Definition verwendete Shannon Entropie durch die Rényi Entropie ersetzt. Wir können zeigen, dass dies im Allgemeinen nicht der Fall ist.

Zusätzlich werden zwei Typen bedingter Permutationsentropie untersucht. Wir stellen Bedingungen auf, unter denen diese gleich der Kolmogorov-Sinai Entropie sind.

Schließlich führen wir bestimmte Bedingungen für mehrdimensionale dynamische Systeme ein, unter denen die Permutationsentropie und Kolmogorov-Sinai Entropie gleich sind. Die Methoden, die dafür verwendet wurden, liefern für bestimmte dynamische Systeme außerdem eine obere Schranke für die Geschwindigkeit, mit der die Entropie ordinaler Muster gegen die Kolmogorov-Sinai Entropie konvergiert.

Acknowledgment

I would like to thank my family for their support throughout all the years. In particular my parents Claudia and Uwe who have always been there for me.

I'm grateful for all the people I have met during my time working at the Institute for Mathematics in Lübeck. There are too many to name individually. I will miss the times eating lunch together in the student cafeteria.

Finally, I want to thank Karsten for being my supervisor. He was never impatient with me and gave me the time I needed but was always there for advice or support when I asked for it. Without him, I would not have started writing this thesis in the first place and I would not have finished it either.

Contents

1 Introduction

Determining the *Kolmogorov-Sinai entropy* (*K-S entropy*) of a measure-preserving dynamical system is a key problem in the analysis of a system's complexity. Roughly speaking, this entropy is a measure for the amount of information that is gained on average by observing the system's dynamics.

Measure-preserving dynamical systems can be used to model different kinds of physical or abstract systems. They describe how ”points” in some state space change over discrete time steps. The ”points” in this state space are weighted by some measure which does not change over time.

Complexity is an important aspect of dynamical systems. Depending on the point of view, there are different interpretations of complexity. From a modeling viewpoint, complexity is related to how fast small changes to the current state of system can grow over time. From a stochastic viewpoint, complexity of a system can describe how difficult it is to predict a new outcome of a dynamical system based on previous outcomes. This difficulty is directly related to the amount of information that is gained on average by observing the system's dynamics.

In all cases, the idea of complexity is connected to some kind of uncertainty. To quantify this uncertainty and therefore to quantify the complexity of dynamical system, one can use the Shannon entropy, which is a concept from information theory [28]. This leads to the definition of the so called 'Kolmogorov-Sinai entropy' as a measure for the complexity of dynamical system which was first introduced by Kolmogorov [20] and later refined by Sinai [30]. Originally, the Kolmogorov-Sinai (K-S) entropy was not directly introduced to measure complexity but because it is invariant under certain isomorphism. This invariance allowed Kolmogorov to show that two specific measure-preserving dynamical systems are not equivalent because their Kolmogorov-Sinai entropy is different. This was an open question until that point.

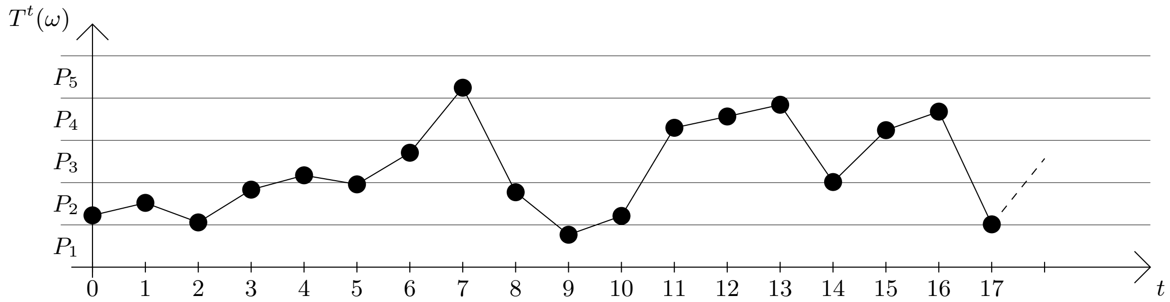
Besides the Kolmogorov-Sinai entropy, there exist alternative methods to measure the complexity of a dynamical system. One example is the so called *Lyapunov exponent*. This concept is applicable if the considered system is smooth enough. Under specific conditions, one can show that Kolmogorov-Sinai entropy can be directly calculated using the Lyapunov exponent [26]. However, in this thesis we will only consider the Kolmogorov-Sinai entropy as the 'classical' measure for complexity because it can be defined in more general situations.

While the Kolmogorov-Sinai entropy has a precise mathematical definition and interesting properties, its computation can be difficult. Therefore, in 2002, Bandt and Pompe introduced the so called *permutation entropy* as an alternative measure for the complexity of a one-dimensional dynamical system which is easier to evaluate numerically than the K-S entropy [5]. The permutation entropy is a measure for the information contained in the ordinal structure of a dynamical system. One can extend the definition of the permutation entropy to multi-dimensional systems by introducing a number of real-valued random variables as observables that each project the multi-dimensional dynamics into the real numbers in which the ordinal structure is then considered.

Given data generated by an underlying but not necessarily known dynamical system, the value of the permutation entropy is usually easier to estimate from this data than the value of Kolmogorov-Sinai entropy. This led to many practical applications of the permutation entropy (e.g. [21],[24],[29]). While determining the value of the permutation entropy, one would still like to know the value of the Kolmogorov-Sinai entropy as a measure for the complexity of the corresponding system. Therefore, the relation between the Kolmogorov-Sinai and the permutation entropy is a question of interest. Bandt, Pompe and G. Keller showed that those entropies are equal for one-dimensional interval maps if there exists a finite partition of the domain of definition into intervals such that the considered map is monotone and continuous on each of those intervals [4]. It was also shown that the permutation entropy is an upper bound for the K-S entropy for all one-dimensional systems (see [4] or [2]). This upper bound can be generalized to multidimensional maps under sufficiently general conditions [16]. It is still an open question if and how the condition of piecewise monotony for the equality of the entropies can be generalized to a larger class of one-dimensional maps.

1.1 Basic concepts

As mentioned above, to determine the complexity of a dynamical system, one can analyze the uncertainty of possible outcomes of the system using the Shannon entropy. The Shannon entropy can only be applied to a finite (or countable infinite) distribution of outcomes but the considered systems typically take values in the real numbers, so uncountably many outcomes are possible. Therefore, one needs to discretize the system first



$$\omega \in P((2,2,2,2,3,3,3,5,2,1,2,4,4,4,3,4,4,2))$$

Figure 1.1:

When determining the Kolmogorov-Sinai entropy, the discretization is achieved by using a finite (or countable infinite) partition of the system. Instead of considering the exact outcome of the system, one only takes into account in what subset of the given partition a point is located (Figure 1.1). Considering this for each point in a sequence of outcomes, creates a sequence of sets in which each specific point is located in. We call such a sequence *symbolic pattern*. Each symbolic pattern can, for example, be encoded by the individual indices of the sets the different points are located in.

One motivation of the permutation entropy is to achieve a discretization without needing to use a given partition. This is done by considering all possible ways different values in a sequence of outcomes can be ordered (see Figure 1.2). We call such an ordering of a sequence of points an *ordinal pattern*. Since all the different ordinal patterns can be indexed using permutations, the resulting entropy is called permutation entropy.

The fact that the permutation entropy only depends on the relative order of points can provide some practical advantages over the Kolmogorov-Sinai entropy when working with real-world data. For example, the permutation entropy is invariant with regard to monotone transformation of the data. This is relevant in practice because monotone transformations of data can, for example, occur if the some underlying system is observed using different measuring devices. But ideally, a measure for complexity should only depend on the underlying system and not be affected by these transformations.

Additionally, small amounts of noise are less likely to change the ordering of a sequence than the values of the sequence itself. Therefore permutation entropy can be less affected by noise (see e.g. [27]). Higher amounts of noise can, however, effect the permutation entropy more significantly. This means that permutation could theoretically also be used to differentiate between noisy random data and deterministic data.

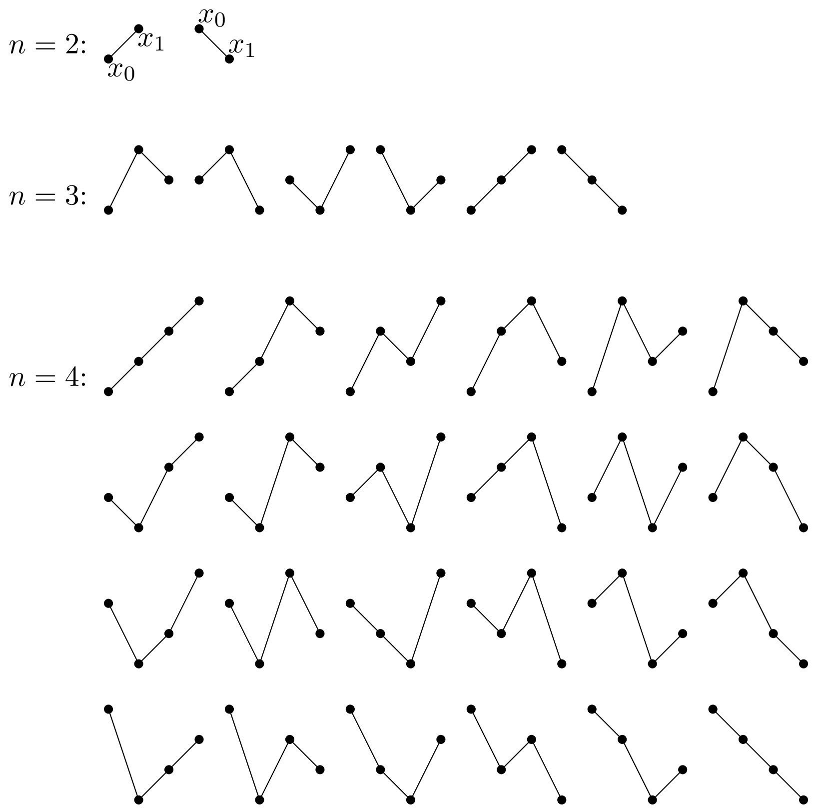


Figure 1.2: Possible orderings of sequences of length n

1.2 Previous results

Ordinal patterns as a measure for the complexity of dynamical systems were first introduced by Band and Pompe in 2002 [5]. In the same year, Band, Pompe and G. Keller showed that the permutation entropy is equal to the Kolmogorov-Sinai entropy for a specific class of dynamical system. This was the first result that related the permutation entropy to some other measure of complexity. Motivated by this result, other theoretical properties of permutation entropy and ordinal patterns and possible applications of them were investigated in the following years.

Band, Pompe and G. Keller also defined a topological variant of the permutation entropy and showed that it is equal to the topological entropy of dynamical system for some dynamical systems. It was later shown by Misiurewicz that there exist some dynamical systems for which the topological permutation is not equal to the topological entropy [23]. It is not known to this date whether there exist dynamical systems for which the permutation entropy is not equal to the Kolmogorov-Sinai entropy.

A. Antoniouk, K. Keller and S. Maksymenko investigated the separation properties of ordinal patterns [3]. Roughly speaking, they showed that, under certain circumstances, all the information of a dynamical system can be captured by looking at ordinal patterns. As a consequence of this result, one can use partitions into ordinal patterns to generate the Kolmogorov-Sinai entropy. This approach using ordinal patterns to generate the Kolmogorov-Sinai is similar to the use of ordinal patterns in the calculation of the permutation entropy. It is not known whether these approaches are always yield the same value. One difference between them is that this approach requires an additional limit for the computation compared to the one limit needed to calculate the permutation entropy. Theoretically though, both approaches could yield the same result. One attempt to establish a connect between these approaches was done in a paper from 2013 [38]. However, as we will discuss in section 3.5.1, the methods used in the mentioned paper are not sufficient to establish such a connection.

One can define conditional variants of the permutation entropy. These variants were first investigated in [37]. This paper made some theoretical statements about the conditional permutation entropy and compared how the complexity of some simulated data can be estimated using the original or the conditional variant of the permutation entropy. It was noticed that the conditional variant typically converges faster than the original permutation entropy which makes it the preferable for practical applications.

In 2017, I. Stolz and K. Keller introduced a generalization of the concept of ordinal patterns [33]. They observed that the order relations between different points in a sequence, which determines the ordinal pattern of this sequence, can be replaced by more general relations. This leads to a generalization of the concept of ordinal patterns. The basic and general ordinal patterns have in common that they are both determined by comparing pairs of points in a sequence of points. Considering the distribution of pairs instead of single points is what provides the separation property of ordinal patterns. Stolz and Keller showed that, under some conditions, certain generalized patterns also exhibit this separation property [33].

Amigo et. al. defined an alternative type of permutation entropy which is based on analyzing the ordinal patterns of coarse-grained versions of the dynamical system. This approach can be seen as a hybrid between the original version of the permutation entropy and the Kolmogorov-Sinai entropy. It was shown that this alternative definition of the permutation entropy is equal to the K-S entropy [2]. However, it is unknown how this alternative type of permutation entropy is related to the original definition of the permutation entropy by Bandt and Pompe. Therefore, we will only focus on the original definition in this thesis. Additionally, the approach of Amigo et. al. has the disadvantage that an additional limit is required for its calculation. An overview about the differences between these two approaches can be found in [1].

1.3 Practical applications

When analyzing data using permutation entropy, it is often interesting to find out how the value of the permutation entropy changes over time. So instead of calculating one value for the entropy of the entire signal, one calculates different values for the entropy of the signal within a specific time window and slights this window along the signal.

An example where permutation entropy has shown to be a useful tool is the classification of data from electroencephalogram (EEG):

Permutation entropy has been used to extract features from EEG data which were then used to automatically detect epileptic seizures (see e.g [25] and [41]). Seizures could be detected with a high accuracy. This approach utilizes the fact the EEG during seizures is characterized by a lower permutation entropy. By determining the entropy of the EEG, the multidimensional data of each channel is represented by single numbers. This reduces the complexity of the data and causes the subsequent classification to be better at generalizing to new data. Additionally, PE is robust to small noise and has low computational complexity which allows this method to be used in real-time.

Keller et al. showed that conditional variants of the permutation entropy can be useful for data classification by testing those methods on EEG data [17]. In another paper, the performance of generalizations of the permutation entropy based on Rényi entropies were tested [18]. These generalized permutation entropies can potentially provide some new information about the underlying data.

Nicolaou and Georgiou used permutation entropy to identify different sleep stages [24]. They recorded EEG signals during sleep and calculated the permutation entropy for different points in time. It was noticed that the permutation entropy would yield significantly different values for different stages, which can be explained by the fact that the complexity of the brain activity is different for different sleep stages. Based on these different values, the different sleep stages could be classified.

In [29], the performance of anesthesia was investigated using permutation entropy. So called local field potentials were measured, which is a more precise method to investigate the brain activity of specific areas than measuring an electroencephalogram. Besides the permutation entropy, various other methods were applied to the recorded data and their correlation to the concentration of anesthesia after a certain period of time was determined. Out of all tried methods, the permutation entropy showed the highest correlation.

In this thesis, we will focus on investigating the distribution of ordinal patterns when the length of the patterns tends to infinity. However, there are also applications that are based on analyzing the distribution of ordinal patterns of short length. For example, one can establish an estimator for the so called Hurst exponent of fractional Brownian motion based on analyzing the distribution of ordinal patterns of length 3 [32]. The Hurst exponent is, roughly speaking, a measure for the roughness of data that is modeled by a fractional Brownian motion. Estimating the Hurst exponent of rock joint surfaces in different directions, which is an important but difficult task in geology, is an example where methods based on short ordinal patterns have shown good results [13].

As another example, analyzing the distribution of short ordinal patterns can be used to detect change points in time series [31]. A change point is a point in time where the distribution of the underlying process changes. Changes in the distribution of the process typically also imply changes in the distribution of ordinal patterns. Analyzing the changes in the distribution of ordinal patterns has the advantage of being more robust with regard to monotone transformation of the data. Theoretically, one can use ordinal patterns of any length for this approach but patterns of length 4 (see Figure 1.2) have shown to be a good compromise between being complex enough to reflect different patterns in the data and easy enough to estimate.

1.4 Overview of results

In chapter 2, the basic definitions are established and some often used concepts are introduced.

Chapter 3 focuses on the relationship between Kolmogorov-Sinai and permutation entropy in the one-dimensional case. We investigate generalized versions of the Kolmogorov-Sinai and the permutation entropy that are based on the Rényi entropies, which are a generalization of the Shannon entropy. These generalized entropies depend on some parameter $q \in \mathbb{R}$ and we will demonstrate that the relationship between K-S and permutation entropy is also different for different $q \in \mathbb{R}$. For $q = 1$ we will show that the considered entropies are equal for (countably) piecewise monotone functions. This is a generalization of an existing result by Bandt, Pompe and G. Keller who showed the the above equality holds true if the functions are monotone and, additionally, continuous on a finite number of intervals. [4].

In section 3.5 we investigate two different variants of conditional permutation entropy and analyze their relationship. Additionally we establish some condition under which these conditional variants and the original permutation entropy coincide.

In chapter 4 we introduce conditions under which a multidimensional variant of the permutation entropy is equal to the Kolmogorov-Sinai entropy. This is the first time such an equality is shown in more the one dimension, however, the conditions are relatively restrictive. Essentially, we require that the considered maps are expanding in every direction, Lipschitz continuous and the measure is not too spread out. The methods used in this chapter differ from the ones used in the previous chapter. They could potentially be further weakened and these methods allow for a different perspective on the question of equality between permutation and Kolmogorov-Sinai entropy. In special cases, they can be used to obtain an theoretical upper bound for the speed of convergence for the permutation entropy.

2 Preliminaries

2.1 Often used spaces

We start by introducing the mathematical spaces that will be used within this thesis.

Definition 2.1 (Complete probability space)**.** A triple $(\Omega, \mathscr{A}, \mu)$ is called probability space if $\mathscr{A}$ is a σ-algebra defined on Ω and $\mu : \mathscr{A} \to [0, 1]$ a probability measure. $(\Omega, \mathscr{A}, \mu)$ is called complete if for all $A \in \mathscr{A}$

$$\mu(A) = 0 \quad \Rightarrow \quad B \in \mathscr{A} \qquad \text{for all } B \subseteq A$$

holds true. If $(\Omega, \mathscr{A}, \mu)$ is not complete, one can add all subsets of sets A with $\mu(A) = 0$ to the σ-algebra $\mathscr{A}$ and extend the domain of definition of μ accordingly. Such an extension is called completion.

Because the value of the entropy is not affected by sets of measure zero, working with the completion of the probability space instead of the original space has no effect on the results we are interested in. Therefore, our results are not less general if, for example for practical application, we do not assume completeness from the beginning.

We often require that the sample space Ω of our considered probability space $(\Omega, \mathscr{A}, \mu)$ is a complete separable metric space and that $\mathscr{A}$ is the Borel σ-algebra based on the metric defined on Ω. This provides a connection between the metric and the events given by the σ-algebra $\mathscr{A}$. Roughly speaking, the countability related to the separability of the metric space is transferable to the Borel σ-algebra $\mathscr{A}$. This guarantees that the considered σ-algebra is not too 'complex'.

Definition 2.2 (Complete separable metric space)**.** Given a set $\Omega \neq \emptyset$, a map $d : \Omega \times \Omega \to \mathbb{R}$ is called metric if for all $\omega_1, \omega_2, \omega_3 \in \Omega$

(i) $d(\omega_1, \omega_2) = 0 \Leftrightarrow \omega_1 = \omega_2$,

(ii) $d(\omega_1, \omega_2) = d(\omega_2, \omega_1)$ and

(iii) $d(\omega_1, \omega_3) \leq d(\omega_1, \omega_2) + d(\omega_2, \omega_3)$

hold true. (Ω, d) is called metric space.

This space is called complete if for every Cauchy sequence $(\omega_n)_{n\in\mathbb{N}}$ in Ω there exists $\omega \in \Omega$ with $\lim_{n\to\infty} d(\omega_n, \omega) = 0$.

(Ω, d) is called separable, if there exists a countable set $A \subseteq \Omega$ such that for all $\omega \in \Omega$ and $\varepsilon > 0$ there exists $a \in A$ with $d(\omega, a) < \varepsilon$.

Notice that the meaning of completeness for a metric space is different from the meaning of completeness for a probability space.

Definition 2.3 (Borel σ-algebra). Let (Ω, d) be a metric space and τ the topology generated by the sets

$$\{\omega \in \Omega \mid d(\omega, \omega_0) < \varepsilon\}$$

for all $\omega_0 \in \Omega$ and $\varepsilon > 0$. Then the Borel σ-algebra $\mathscr{B}(\Omega)$ on Ω is the smallest σ-algebra containing all sets in τ. If it arises out of the context on what space Ω the Borel σ-algebra is defined on, we simply write $\mathscr{B}$ instead. Sometimes we extend $\mathscr{B}$ to its completion (see eg. Definition 2.4) and use the some notation $\mathscr{B}$.

In this thesis and for many applications, we typically consider the Borel σ-algebra $\mathscr{B}(\Omega)$ for the metric space (Ω, d) where Ω is a subset of $\mathbb{R}^n$ for $n \in \mathbb{N}$ and d the Euclidian distance. In this case, $\mathscr{B}(\Omega)$ is simply the smallest σ-algebra containing all sets of the type

$$[a_1, b_1[\times[a_2, b_2[\times \ldots \times [a_n, b_n[\cap \Omega$$

with $a_1, b_1, a_2, b_2 \ldots, a_n, b_2 \in \mathbb{R}$.

2.1.1 Standard probability spaces

Some statements within this thesis, which will mainly appear in section 3.4.5, cannot be proven on arbitrary probability space but only on so called *Standard probability spaces*.

Definition 2.4 (Standard probability space). A probability $(\Omega, \mathscr{A}, \mu)$ is called *Standard probability space*, if there exists a complete separable metric space $(\widetilde{\Omega}, d)$ such that $(\Omega, \mathscr{A}, \mu)$ is isomorphic to $(\widetilde{\Omega}, \mathscr{B}_{\widetilde{\mu}}(\widetilde{\Omega}), \widetilde{\mu})$, where $\mathscr{B}_{\widetilde{\mu}}(\widetilde{\Omega})$ denotes the completion of the Borel σ-algebra of $\widetilde{\Omega}$ with regard to $\widetilde{\mu}$.

Isomorphisms between probability spaces are defined the following:

Definition 2.5 (Isomorphism between probability spaces). Let $(\Omega_1, \mathscr{A}_1, \mu_1)$ and $(\Omega_2, \mathscr{A}_2, \mu_2)$ be two probability spaces. A map $\varphi : \Omega_1 \to \Omega_2$ is called isomorphism, if there exist sets $A_1 \in \mathscr{A}_1$ and $A_2 \in \mathscr{A}_2$ with $\mu_1(A_1) = \mu_2(A_2) = 1$ such that

(i) $\varphi' : A_1 \to A_2$ with $\varphi'(\omega) = \varphi(\omega)$ for all $\omega \in A_1$ is invertible,

(ii) $\varphi(A \cap A_1) \in \mathscr{A}_2$ for all $A \in \mathscr{A}_1$,

(iii) $\varphi^{-1}(A \cap A_2) \in \mathscr{A}_1$ for all $A \in \mathscr{A}_2$,

(iv) $\mu_1(\varphi^{-1}(A)) = \mu_2(A)$ for all $A \in \mathscr{A}_2$.

If there exists such an isomorphism between two probability spaces, these spaces are called isomorphic.

The statements that can only be proven on Standard probability spaces are basically all related to the so called ergodic decomposition of measures. To be able to decompose any probability measure into an uncountable collection of measures, the assumption of being standard is essential for the probability space.

Since for practical purposes we are mostly interested in Borel probability spaces on some subset of $\mathbb{R}^d$ anyway, the assumption of being Standard is not a significant restriction. Most Borel probability spaces, whose dimension is not too big in a certain sense, can be made isomorphic to a complete separable metric space by adding a set of points with measure zero.

2.2 Measure-preserving dynamical systems

Given a probability space $(\Omega, \mathscr{A}, \mu)$, adding a measurable and measure-preserving map T to this space creates a so called *measure-preserving dynamical system.*

Definition 2.6 (Measure-preserving dynamical systems)**.** Let $(\Omega, \mathscr{A}, \mu)$ be a probability space. A map $T : \Omega \to \Omega$ is called measure-preserving with regard to μ or μ is called invariant under T if T is measurable, i.e. $T^{-1}(A) \in \mathscr{A}$ for all $A \in \mathscr{A}$, and $\mu(T^{-1}(A)) = \mu(A)$ for all $A \in \mathscr{A}$. If T is measure-preserving with regard to μ, the tuple $(\Omega, \mathscr{A}, \mu, T)$ is called measure-preserving dynamical system.

2.2.1 Ergodicity

Ergodic dynamical systems play a special role in the theory of measure-preserving dynamical systems because they have many nice properties.

Definition 2.7 (Ergodicity)**.** Let $(\Omega, \mathscr{A}, \mu, T)$ be a measure-preserving dynamical systems. The map T is called ergodic with regard to μ if

$$T^{-1}(A) = A \Rightarrow \mu(A) \in \{0, 1\}$$

holds true for all $A \in \mathscr{A}$.

So the T-invariant sets of an ergodic dynamical system are only the trivial sets with measure 0 or 1. One of the most important consequences of this fact is Birkhoff's (pointwise) ergodic theorem:

Theorem 2.8 (Birkhoff's (pointwise) ergodic theorem)**.** Let $(\Omega, \mathscr{A}, \mu, T)$ be a measure-preserving dynamical system. Then for every integrable $\mathbb{R}$-valued random variable X there exists an integrable $\mathbb{R}$-valued random variable $\widehat{X}$ with

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} X(T^i(\omega)) = \widehat{X}(\omega)$$

satisfying $E(X) = E(\widehat{X})$ and $\widehat{X} \circ T = \hat{X}$ μ-almost surely. If T is ergodic, than $\widehat{X} = E(X)$ μ-almost surely.

Birkhoff's theorem will be crucial for some of the arguments in the following chapters. It can be used to deduce statements about the asymptotic behavior of dynamical systems, which is relevant when investigating the complexity of such systems.

On a practical level, Birkhoff's theorem allows us to estimate probabilities based on only one realization of the dynamical system. Given a dynamical system $(\Omega, \mathscr{A}, \mu, T)$, the exact distribution of the probability measure μ is typically not known in practice. Instead, one can only observes a realization $\omega, T(\omega), T^2(\omega), \ldots$ for some fixed $\omega \in \Omega$. According to Birkhoff's Theorem, for any set $A \in \mathscr{A}$ the quantity

$$\#\{t \in \{0, 1, \ldots, n-1\} \mid T^t(\omega) \in A\}/n$$

converges to $E(1_A) = \mu(A)$ for μ-almost all $\omega \in \Omega$. So for any set $A \in \mathscr{A}$, we can μ-almost surely estimate its probability based on a single realization of the dynamical system.

There exists equivalent formulation of ergodicity that can be easier to use or the verify:

Lemma 2.9 ([40])**.** Let $(\Omega, \mathscr{A}, \mu, T)$ be a measure-preserving dynamical system. The following statements are equivalent:

(i) T is ergodic with regard to μ.

(ii) $\mu(T^{-1}(A) \triangle A) = 0$ implies $\mu(A) \in \{0, 1\}$ for all $A \in \mathscr{A}$.

(iii) $\bigcup_{n=0}^{\infty} T^{-n}(A) = \Omega$ for all $A \in \mathscr{A}$ with $\mu(A) > 0$

(iv) $\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \mu(A \cap T^{-t}(B)) = \mu(A)\mu(B)$ for all $A, B \in \mathscr{A}$.

We use the notation

$$A \triangle B := (A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A)$$

to denote the symmetric difference between sets A and B.

2.2.2 Periodicity

Definition 2.10 (Aperiodicity)**.** Given a measure-preserving dynamical system $(\Omega, \mathscr{A}, \mu, T)$, let

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

be the set of periodic points of T. The map T is called (μ-almost surely) aperiodic, if

$$\mu(\Pi) = 0$$

holds true.

With

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

we denote the set of all points with period smaller or equal to $k \in \mathbb{N} \cup \{\infty\}$. These sets are μ-almost surely T-invariant:

Lemma 2.11. Let $(\Omega, \mathscr{A}, \mu, T)$ be a measure-preserving dynamical system. Then

$$\mu\left(T^{-1}(\Pi_k)\triangle\Pi_k\right) = \mu\left(T^{-1}(\Pi_k)\setminus\Pi_k\right) = 0$$

holds true for al $k \in \mathbb{N} \cup \{\infty\}$.

Proof. We will first show that

$$\Pi_k \subseteq T^{-1}(\Pi_k) \quad \text{für alle } k \in \mathbb{N} \cup \{\infty\} \tag{2.1}$$

holds true. Let $k \in \mathbb{N} \cup \{\infty\}$ and $\omega \in \Pi_k$. So there exists some $n \leq k$ with $T^n(\omega) = \omega$. This implies

$$T^n(T(\omega)) = T(T^n(\omega)) = T(\omega).$$

Hence, $T(\omega) \in \Pi_k$ which is equivalent $\omega \in T^{-1}(\Pi_k)$. Therefore, (2.1) holds true.

The fact that T is measure-preserving then provides

$$\begin{aligned}\mu\left(\Pi_k \triangle T^{-1}(\Pi_k)\right) =& \mu\left(T^{-1}(\Pi_k)\setminus\Pi_k\right) + \mu\left(\Pi_k \setminus T^{-1}(\Pi_k)\right) \overset{(2.1)}{=} \mu\left(T^{-1}(\Pi_k)\setminus\Pi_k\right)\\ &\overset{(2.1)}{=} \mu(T^{-1}(\Pi_k)) - \mu(\Pi_k) = \mu(\Pi_k) - \mu(\Pi_k) = 0.\end{aligned}$$ □

In particular, if T is ergodic with regard to μ, we have

$$\mu(\Pi_k) \in \{0, 1\} \quad \text{for all } k \in \mathbb{N} \cup \{\infty\}$$

according to Lemma 2.9.

Definition 2.12 (Non-atomic measure)**.** Let $(\Omega, \mathscr{A}, \mu)$ be a probability space. The measure μ is called non-atomic, if $\mu(\{\omega\}) = 0$ holds true for all $\omega \in \Omega$.

Periodicity is tied to the concept of *non-atomic* measures as shown in the following lemma.

Lemma 2.13. Let $(\Omega, \mathscr{A}, \mu, T)$ be a measure-preserving dynamical system. If T is μ-almost surely aperiodic, then μ is non-atomic.

Proof. We perform a proof by contradiction. Suppose, T is μ-almost surely aperiodic but there exists some $\omega \in \Omega$ and some $\varepsilon > 0$ with $\mu(\{\omega\}) = \varepsilon$. We have

$$\{\omega\} \subseteq T^{-n}(T^n(\{\omega\})).$$

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

$$\mu(T^n(\{\omega\})) = \mu(T^{-n}(T^n(\{\omega\}))) \geq \mu(\{\omega\}) = \varepsilon$$

because T is measure-preserving. Since $\{\omega\}$ has positive measure and T was assumed to be aperiodic, ω cannot be a periodic point. This implies $T^n(\{\omega\}) \cap T^m(\{\omega\}) = \emptyset$ for all $n \neq m \in \mathbb{N}_0$. Therefore,

$$\mu\left(\bigcup_{n=0}^{N-1} T^n(\{\omega\})\right) = \sum_{n=0}^{N-1} \mu\left(T^n(\{\omega\}\right) \geq N\varepsilon$$

holds true for all $N \in \mathbb{N}$. Choosing $N > 1/\varepsilon$ yields the contradiction

$$\mu\left(\bigcup_{n=0}^{N-1} T^n(\{\omega\})\right) \geq N\varepsilon > 1. \qquad \square$$

Some of the statements in the following chapter only hold if the considered measure is non-atomic, which is why we often consider aperiodic dynamical systems.

Convexity Many of the functions involved in the definition of the Shannon entropy in the next section involve convex function. A subset X of a real-valued vector space is called convex if

$$\alpha x_1 + (1-\alpha)x_2 \in X \quad \text{for all } x_1, x_2 \in X \text{ and } \alpha \in [0,1]$$

holds true. A function $f : X \to \mathbb{R}$ defined on a convex set X is called convex if

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2) \quad \text{for all } x_1, x_2 \in X \text{ and } \alpha \in [0,1]$$

and concave if

$$f(\alpha x_1 + (1-\alpha)x_2) \geq \alpha f(x_1) + (1-\alpha) f(x_2) \quad \text{for all } x_1, x_2 \in X \text{ and } \alpha \in [0,1]$$

holds true. Using induction, one can extend the above definition to convex combinations involving finitely many elements. This leads to the so called *Jensen's inequality* which will be used multiple times in this thesis to prove certain properties of the Shannon entropy and related concepts.

Lemma 2.14 (Jensen's inequality [14])**.** Let X be a convex subset of a real-valued vector space, $n \in \mathbb{N} \cup \{\infty\}$, $x_i \in X$ and $\alpha_i \geq 0$, $i \in \{1, 2, \ldots, n\}$, with $\sum_{i=1}^n \alpha_i = 1$. If $f : X \to \mathbb{R}$ is a convex function, then

$$f\left(\sum_{i=1}^n \alpha_i x_i\right) \leq \sum_{i=1}^n \alpha_i f(x_i)$$

holds true. If f is concave function,

$$f\left(\sum_{i=1}^n \alpha_i x_i\right) \geq \sum_{i=1}^n \alpha_i f(x_i)$$

holds true.

If $X \subseteq \mathbb{R}$ is a convex subset of the real number and $f : X \to \mathbb{R}$ is two times continuously differentiable, then f is convex if $f''(x) \geq 0$ for all $x \in X$ and concave if $f''(x) \leq 0$ for all $x \in X$.

2.3 Entropy

2.3.1 Shannon and Rényi entropy

A vector $p = (p_1, p_2, ..., p_n) \in \mathbb{R}^n$ with $n \in \mathbb{N} \cup \{\infty\}$ is called stochastic, if $p_i \geq 0$ for all $i \in \{1, 2, ..., n\}$ and $\sum_{i=1}^p p_i = 1$. Entropy in its simplest form is defined for the set of all stochastic vectors. The Shannon or Rényi entropy quantify how close to the equidistribution $(1/n, 1/n, \dots, 1/n)$ a stochastic vector $p \in [0,1]^n$ is. The corresponding entropy of p is maximal if p is equal to the equidistribution. Since the outcome of random events is most difficult to predict if the distribution of the events is equal to the equidistribution, the Shannon or Rényi entropy can be seen as a measure of 'unpredictability'.

Definition 2.15 (Rényi entropy)**.** Let $p = (p_1, p_2, ..., p_n)$ a stochastic vector of length $n \in \mathbb{N} \cup \{\infty\}$. The Rényi entropy $H(p, q)$ of p and parameter $q \in \mathbb{R}$ is defined as

$$H(p,q) = \begin{cases} \frac{-1}{q-1} \log(\sum_{i=1}^n p_i^q), & \text{falls } q \neq 1, \\ -\sum_{i=1}^n p_i \log(p_i), & \text{falls } q = 1. \end{cases}$$

We set $0^q := 0$ for al $q \in \mathbb{R}$ and $0 \cdot \log(0) := 0$.
$H(p) := H(p, 1)$ is called Shannon entropy of p.

So the Rényi entropy is a generalization of the Shannon entropy. The following Lemma shows that the Rényi entropy is continuous in q.

Lemma 2.16. For all stochastic vectors p

$$\lim_{q \to 1} H(p,q) = H(p,1)$$

holds true.

An important property of the Rényi entropy is the monotonicity decreasing as a function of q. The monotonicity of the Rényi entropy as a function of q is an important property that will be used multiple times in chapter 3 to establish certain inequalities.

Lemma 2.17 (Monotony in q)**.** For all stochastic vectors p

$$q_1 \leq q_2 \Rightarrow H(p, q_1) \geq H(p, q_2)$$

holds true.

The proves of the two previous lemmas can be found in the appendix.

We now define the Rényi entropy of a partition of Ω. Given a measure space $(\Omega, \mathscr{A})$, we call a collection of sets $\mathcal{P} = \{P_i\}_{i \in I}$ partition of Ω if the following holds true:

(i) $P_i \in \mathscr{A}$ for all $i \in I$,

(ii) $P_i \neq \emptyset$ for all $i \in I$,

(iii) $P_i \cap P_j = \emptyset$ for all $i, j \in I$ with $i \neq j$,

(iv) $\bigcup_{i \in I} P_i = \Omega$.

Definition 2.18 (Rényi entropy of a partition)**.** Let $(\Omega, \mathscr{A}, \mu)$ be a probability space and $\mathcal{P} = \{P_1, P_2, ..., P_n\}$ a partition of Ω of length $n \in \mathbb{N} \cup \{\infty\}$. Then

$$H_\mu(\mathcal{P}, q) := H((\mu(P_1), \mu(P_2), ..., \mu(P_n)), q)$$

defines the Rényi entropy of the partition $\mathcal{P}$.
$H_\mu(\mathcal{P}) := H_\mu(\mathcal{P}, 1)$ is called Shannon entropy of the partition $\mathcal{P}$.
If the choice of the considered measure μ is obvious from the context, we simply write $H(\mathcal{P}, q)$ or $H(\mathcal{P})$ instead of $H_\mu(\mathcal{P}, q)$ or $H_\mu(\mathcal{P})$

We want to calculate the entropy of specific kind of partitions. In the following, the classes of partitions are defined that are of interest in this thesis.

Definition 2.19 (Special classes of partitions)**.** Let $(\Omega, \mathscr{A}, \mu)$ be a probability space. We define

$$\mathbb{P}(\mathscr{A}) := \{\mathcal{P} \subseteq \mathscr{A} \mid \mathcal{P} \text{ is a finite partition of } \Omega\}$$

and

$$\mathbb{P}^c(\mathscr{A}) := \{\mathcal{P} \subseteq \mathscr{A} \mid \mathcal{P} \text{ is a finite or countable partition of } \Omega \text{ with } H(\mathcal{P}) < \infty\}.$$

There are different ways one can combine multiple partitions. The following definition describes one method of combining partitions that is relevant for calculating the entropy of a dynamical system.

Definition 2.20 (Refinement of collections of sets)**.** Let $\mathcal{P} = \{P_1, P_2, ..., P_n\}$ and $\mathcal{Q} = \{Q_1, Q_2, ..., Q_m\}$ be two collections of sets in Ω. Then we define the refinement of these collections as

$$\mathcal{P} \vee \mathcal{Q} := \{P_i \cap Q_j \mid i \in \{1, 2, ..., n\}, j \in \{1, 2, ..., m\} \text{ and } P_i \cap P_j \neq \emptyset\}.$$

The refinement of a finite or countably infinite number of collections $\mathcal{P}_1, \mathcal{P}_2, \ldots$ is defined as

$$\bigvee_{k=1}^{l} \mathcal{P}_k = \left\{ \bigcap_{k=1}^{l} P_k \neq \emptyset \mid P_k \in \mathcal{P}_k \text{ for all } k \in \{1, 2, \ldots, l\} \right\}$$

for $l \in \mathbb{N} \cup \{\infty\}$.

In particular, if $\mathcal{P}$ and $\mathcal{Q}$ are partitions of Ω, the refinement $\mathcal{P} \vee \mathcal{Q}$ is a partition of Ω as well.
One of the main differences between the Rényi entropy for $q \neq 1$ and the Shannon entropy is the so called 'Subadditivity'. While the Shannon entropy exhibits this property the Rényi entropy does not in general.

Lemma 2.21 (Subadditivity)**.** Let $(\Omega, \mathscr{A}, \mu)$ be a probability space. Then for all finite or countable partitions $\mathcal{P}$ and $\mathcal{Q}$ of Ω.

$$H(\mathcal{P} \vee \mathcal{Q}) \leq H(\mathcal{P}) + H(\mathcal{Q})$$

holds true

The proof of this Lemma can be found in the appendix.

2.3.2 Kolmogorov-Sinai entropy

The iterates $T^t(\omega)$ are recursively defined as $T^t(\omega) = T(T^{(t-1)}(\omega))$ for $t \in \mathbb{N}$ and $\omega \in \Omega$ with $T^0(\omega) = \omega$. The value of the Kolmogorov-Sinai entropy depends on the position of the elements of orbits $(\omega, T(\omega), T^2(\omega), \ldots)$ with respect to a finite or countable partition. All elements $\omega \in \Omega$ whose orbit of some length $n \in \mathbb{N}$ is the same with respect to same partition partition build a so called *symbolic pattern.*

Definition 2.22 (Symbolic patterns)**.** Let $(\Omega, \mathscr{A}, \mu, T)$ be a measure-preserving dynamical system and $\mathcal{P} = \{P_i\}_{i \in I}$ a partition of Ω with some finite or countable index set I. For $n \in \mathbb{N}$ and a multi index $\boldsymbol{i} = (i_0, i_1, \ldots i_{n-1}) \in I^n$ we define the symbolic pattern

$$P(\boldsymbol{i}) := \bigcap_{t=0}^{n-1} T^{-t}(P_{i_t}) = P_{i_0} \cap T^{-1}(P_{i_1}) \cap \ldots T^{-n+1}(P_{i_{n-1}})$$

and the partition into symbolic patterns

$$\mathcal{P}^{(n)} := \{P(\boldsymbol{i}) \mid \boldsymbol{i} \in I^n \text{ and } P(\boldsymbol{i}) \neq \emptyset\}.$$

The partition into symbolic patterns can also be written as

$$\mathcal{P}^{(n)} = \bigvee_{t=0}^{n-1} T^{-t}(\mathcal{P}),$$

where $T^{-t}(\mathcal{P}) = \{T^{-t}(P_i)\}_{i \in I}$.

When determining the complexity of a dynamical system, we consider the probabilities of $\omega, T(\omega), T^2(\omega), \ldots, T^{(n-1)}(\omega)$ lying within specific sequences of sets. This leads to the definition of the Kolmogorov-Sinai entropy.

Definition 2.23 ((generalized)Kolmogorov-Sinai entropy)**.** The*(generalized) Kolmogorov-Sinai entropy* (or *entropy rate*) of T *with regard to the partition* $\mathcal{P}$ is defined as

$$h_\mu(T, \mathcal{P}, q) := \liminf_{n \to \infty} \frac{1}{n} H_\mu(\mathcal{P}^{(n)}, q), \tag{2.2}$$

The (standard) entropy rate is given by

$$h_\mu(T, \mathcal{P}) := h_\mu(T, \mathcal{P}, 1)$$

By
$$h_\mu(T,q) := \sup_{\mathcal{P}\in\mathbb{P}(\mathscr{A})} h(T,\mathcal{P},q),$$
the *(generalized) Kolmogorov-Sinai entropy* of T is defined, where the supremum is taken over all finite or over all countable partitions with finite entropy. The (standard) Kolmogorov-Sinai entropy is given by
$$h_\mu(T) := h_\mu(T,1)$$
If the choice of the considered measure μ is obvious from the context, we simply write $h(T,\mathcal{P},q)$ or $h(T,q)$ instead of $h_\mu(T,\mathcal{P},q)$ or $h_\mu(T,q)$.

Originally, the Kolmogorov-Sinai entropy was defined as the supremum of the entropy rates over finite partitions, disregarding countable partitions. For $q = 1$ however, according to Abramov's Theorem, the supremum of the entropy rates over all countable partitions with finite entropy is not larger than the supremum of the entropy rates over all finite partitions (see e.g. [15]):
$$h_\mu(T,1) = \sup_{\mathcal{P}\in\mathbb{P}(\mathscr{A})} h(T,\mathcal{P},1) = \sup_{\mathcal{P}\in\mathbb{P}^c(\mathscr{A})} h(T,\mathcal{P},1).$$

2.3.3 Generating partitions

One can define an order relation for the set of all partitions:

Definition 2.24 (Relationship between collections of sets)**.** Let $(\Omega,\mathscr{A},\mu)$ be a probability space and $\mathcal{P},\mathcal{Q}$ be two collections of sets in Ω. The collection $\mathcal{Q}$ is called *finer* than $\mathcal{P}$ or a *refinement* of $\mathcal{P}$ (with regard to μ) if for all sets $Q \in \mathcal{Q}$ there exists a set $P \in \mathcal{P}$ with $\mu(Q \setminus P) = 0$. We write
$$\mathcal{P} \underset{\mu}{\prec} \mathcal{Q} \qquad \text{or} \qquad \mathcal{Q} \underset{\mu}{\succ} \mathcal{P}$$
to denote that $\mathcal{Q}$ is a refinement of $\mathcal{P}$. If it is clear from the context what measure μ is considered, we simply write
$$\mathcal{P} \prec \mathcal{Q} \qquad \text{or} \qquad \mathcal{Q} \succ \mathcal{P}.$$

The relative order of two given partitions also affects the relative order of the Rényi entropies of those partitions.

Lemma 2.25. Let $(\Omega,\mathscr{A},\mu)$ be a probability space and $\mathcal{P}$ and $\mathcal{Q}$ two countable partitions of Ω. Then $\mathcal{P} \prec \mathcal{Q}$ implies
$$H(\mathcal{P},q) \leq H(\mathcal{Q},q)$$
for all $q \in \mathbb{R}$.

The proof of the above lemma can be found in the appendix.

Given some a measure space $(\Omega, \mathscr{A})$ and let $\mathcal{P}$ be some collection of element of $\mathscr{A}$, we denote by

$$\sigma(\mathcal{P}) := \bigcap_{\substack{\mathcal{P}\subseteq\mathscr{S}\subseteq\mathscr{A}:\\ \mathscr{S} \text{ is a } \sigma-\text{algebra}}} \mathscr{S}$$

the smallest σ-algebra containing all elements of $\mathcal{P}$. Notice that $\sigma(\mathcal{P})$ is indeed a σ-algebra because the intersection of σ-algebras is a σ-algebra again.

Definition 2.26 (Almost-sure equality between σ-algebras)**.** Let $(\Omega, \mathscr{A}_1, \mu)$ and $(\Omega, \mathscr{A}_2, \mu)$ be two probability spaces. We write

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

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

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

if $\mathscr{A}_1 \underset{\mu}{\subseteq} \mathscr{A}_2$ and $\mathscr{A}_2 \underset{\mu}{\subseteq} \mathscr{A}_1$ are true.

Given two finite or countable partitions $\mathcal{P}$ and $\mathcal{Q}$, it is easy to see that $\sigma(\mathcal{P}) \underset{\mu}{\subseteq} \sigma(\mathcal{Q})$ is equivalent to $\mathcal{P} \underset{\mu}{\prec} \mathcal{Q}$.

Instead of considering σ-algebras based on a partitions one can also generate σ-algebras based on random variables. Given a random variable $\boldsymbol{X} = (X_1, X_2, \ldots, X_d) : \Omega \to \mathbb{R}^d$, we define

$$\sigma(\boldsymbol{X}) := \sigma\left(\{X_i^{-1}(B) \mid B \in \mathscr{B}(\mathbb{R}),\ i \in \{1, 2, \ldots, d\}\}\right)$$

Given a collection of random variables $\{\boldsymbol{X}^{(j)} \mid j \in J\}$, where J is some index set, we define

$$\sigma\left(\left\{\boldsymbol{X}^{(j)} \mid j \in J\right\}\right) := \sigma\left(\left\{\bigcap_{j\in J} A_j \mid A_j \in \sigma\left(\boldsymbol{X}^{(j)}\right),\ j \in J\right\}\right)$$

It is often difficult to determine the Kolmogorov-Sinai entropy of a dynamical system explicitly because one has to calculate the entropy rate of all finite partitions to determine the supremum. However, under specific circumstances, there exist partitions or sequences of partitions that can fully capture the dynamic of the system and are sufficient to determine the Kolmogorov-Sinai entropy.

Definition 2.27 (Generating partition)**.** Let $(\Omega, \mathscr{A}, \mu, T)$ be a measure-preserving dynamical system. A finite partition $\mathcal{P}$ of Ω is called generating partition if

$$\sigma\left(\bigvee_{t=0}^{\infty} T^{-t}(\mathcal{P})\right) \underset{\mu}{=} \mathscr{A}$$

The sequence of partitions $\{\mathcal{P}_k\}_{k\in\mathbb{N}}$ is called generating sequence of partitions, if $\mathcal{P}_k \prec P_{k+1}$ for all $k \in \mathbb{N}$ and

$$\sigma\left(\bigvee_{k=0}^{\infty}\bigvee_{t=0}^{\infty} T^{-t}\left(\mathcal{P}_k\right)\right) \underset{\mu}{=} \mathscr{A}$$

holds true.

Example 1 (Interval partitions)**.** Given a dynamical system $(\Omega, \mathscr{B}, \mu, T)$ with $\Omega \subseteq \mathbb{R}^d$ and $\mathscr{B}$ being the Borel σ-algebra defined on Ω, one can define a sequence of generating partitions based on intervals. Define

$$P_j^k := \begin{cases} [j \cdot 2^{-k}, (j+1) \cdot 2^{-k}[& \text{if } j \in \{-4^k, \ldots, 4^k - 1\} \\ ]-\infty, 2^{-k}[& \text{if } j = -4^k - 1 \\ [2^k, \infty[& \text{if } j = 4^k \end{cases}$$

for all $j \in \{-4^k - 1, 4^k, \ldots, 4^k - 1, , 4^k\}$ and $k \in \mathbb{N}$ and set

$$J_k := \{(j_1, j_2, \ldots, j_d) \in \{-4^k - 1, 4^k, \ldots, 4^k - 1, 4^k\}^d \mid (P_{j_1}^k \times P_{j_2}^k \times \ldots \times P_{j_d}^k) \cap \Omega \neq \emptyset\}.$$

Then $\{\mathcal{P}_k\}_{k \in \mathbb{N}}$ with

$$\mathcal{P}_k := \{(P_{j_1}^k \times P_{j_2}^k \times \ldots \times P_{j_d}^k) \cap \Omega \mid (j_1, j_2, \ldots, j_d) \in J_k\}$$

is a sequence of generating partitions because the Borel σ-algebra is the smallest σ-algebra containing all (multidimensional) intervals.

Theorem 2.28 (Entropy of generating partition [40])**.** Let $\mathcal{P}$ be a generating partition. Then

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

holds true. Let $\{\mathcal{P}_k\}_{k \in \mathbb{N}}$ be a generating sequence of partitions. Then

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

holds true

So generating partitions can be used to determine the entropy of a dynamical system.

2.3.4 Bernoulli-Shifts

Fix $N \in \mathbb{N}$. (One-sided) Bernoulli shifts are maps that have a relatively simple dynamic. They are defined on

$$\Sigma^{\mathbb{N}} = \{\omega = (\omega_i)_{i=1}^{\infty} \mid \omega_i \in \Sigma\} \tag{2.3}$$

where $\Sigma = \{0, 1, 2, ..., N-1\}$.

Given $s, t \in \mathbb{N}$ with $s \leq t$ and $(a_s, ..., a_t) \in \Sigma^{t-s+1}$,

$$C_s^t(a_s, ..., a_t) := \{\omega \in \Sigma^{\mathbb{N}} : \omega_i = a_i \text{ for } i \in \{s, ..., t\}\}$$

defines a so called cylinder set in $\Sigma^{\mathbb{N}}$. Let

$$\mathscr{C} := \sigma\left(\{C_s^t(a_s, ..., a_t) : s, t \in \mathbb{N}, s \leq t, (a_s, ..., a_t) \in \Sigma^{t-s+1}\}\right)$$

be the σ-algebra generated by all cylinder sets.

Given a stochastic vector $p = (p_1, p_2, \ldots, p_N)$, we define the measure ν on the measure space $(\Sigma^{\mathbb{N}}, \mathscr{C})$ as the product measure generated by p, i.e.

$$\nu(C_s^t(a_s, ..., a_t)) = \prod_{i=s}^{t} p_{a_i}$$

for all cylinder sets $C_s^t(a_s, ..., a_t) \in \mathscr{C}$.

Define a map $B : \Sigma^{\mathbb{N}} \to \Sigma^{\mathbb{N}}$ by

$$\omega \mapsto \widetilde{\omega} = B(\omega), \quad \text{where} \quad \widetilde{\omega}_i = \omega_{i+1} \quad \text{for all } i \in \mathbb{N}$$

It is easy to see that $(\Sigma^{\mathbb{N}}, \mathscr{C}, \nu, B)$ forms a measure-preserving dynamical system. We call $(\Sigma^{\mathbb{N}}, \mathscr{C}, \nu, B)$ *Bernoulli shift generated by the stochastic vector* p.

Let

$$\mathcal{C}_k := \{C_1^k(a_1, ..., a_k) : (a_1, ..., a_k) \in \Sigma^k\} \tag{2.4}$$

be the partition into cylinder sets of length $k \in \mathbb{N}$. Then

$$\mathcal{C}_k^{(n)} = \bigvee_{t=0}^{n-1} B^{-t}(\mathcal{C}_k) = \{C_1^{k+n-1}(a_1, ..., a_{k+n-1}) : (a_1, ..., a_{k+n-1}) \in \Sigma^{k+n-1}\} = C_{k+n-1} \tag{2.5}$$

for all $n \in \mathbb{N}$. One can use this to directly calculate the entropy rate $h_\nu(B, \mathcal{C}_k)$.

$$\begin{aligned}
H_\nu\left(\bigvee_{t=0}^{n-1} B^{-t}(\mathcal{C}_k), q\right) &= \frac{-1}{q-1} \log\left(\sum_{C \in \bigvee_{t=0}^{n-1} B^{-t}(\mathcal{C}_k)} \nu(C)^q\right) \\
&= \frac{-1}{q-1} \log\left(\sum_{(a_1,...,a_{k+n-1}) \in \Sigma^{k+n-1}} \nu(C_1^{k+n-1}(a_1, ..., a_{k+n-1}))^q\right) \\
&= \frac{-1}{q-1} \log\left(\sum_{(a_1,...,a_{k+n-1}) \in \Sigma^{k+n-1}} \left(\prod_{i=1}^{k+n-1} p_{a_i}\right)^q\right) \\
&= \frac{-1}{q-1} \log\left(\sum_{(a_1,...,a_{k+n-1}) \in \Sigma^{k+n-1}} \prod_{i=1}^{k+n-1} p_{a_i}^q\right) \\
&= \frac{-1}{q-1} \log\left(\left(\sum_{a \in \Sigma} p_a^q\right)^{k+n-1}\right) \\
&= (k+n-1) \cdot \frac{-1}{q-1} \log\left(\sum_{a \in \Sigma} p_a^q\right) \\
&= (k+n-1) \cdot H(p, q)
\end{aligned} \tag{2.6}$$

for all $n \in \mathbb{N}$ and $q \neq 1$. Due to the continuity in q of the Rényi entropy (Lemma 2.16) the above statements also holds true for $q = 1$. Therefore

$$\begin{aligned}
h_\nu(B, \mathcal{C}_k, q) &= \liminf_{n\to\infty} \frac{1}{n} H\left(\bigvee_{t=0}^{n-1} B^{-t}(\mathcal{C}_k), q\right) \\
&= \liminf_{n\to\infty} \frac{k+n-1}{n} \cdot H(p, q) = H(p, q)
\end{aligned}$$

holds true for all $q \in \mathbb{R}$.

Now consider the function $\varphi : \Sigma^N \to [0,1[$ with

$$\varphi((\omega_1, \omega_2, \dots)) = \sum_{i=1}^{\infty} \omega_i \cdot N^{-i} \tag{2.7}$$

One can easily show that this function is invertible ν-almost everywhere. Only elements $\omega \in \Sigma^N$ ending in $(\dots, N-1, N-1, N-1, \dots)$ can be mapped to the same value as other element:

$$\varphi((\omega_1, \omega_2, \dots, \omega_k, N-1, N-1, N-1, \dots)) = \varphi((\omega_1, \omega_2, \dots, \omega_k + 1, 0, 0, 0, \dots))$$

So this map is invertible if we restrict the domain of definition to the elements $\omega \in \Sigma^{\mathbb{N}}$ that are not ending in repeats of $N-1$. This is not a problem because the set of elements ending in repeats of $N-1$ has measure 0.

We consider the function $T : [0,1[\to [0,1[$ with $T(\omega) = N \cdot \omega \mod 1$. Define the measure μ on the the Borel σ-algebra $\mathscr{B}$ on $[0,1[$ by

$$\mu(A) := \nu(\varphi^{-1}(A)) \tag{2.8}$$

for all $A \in \mathscr{B}$. Then $([0,1[, \mathscr{B}, T, \mu)$ is a measure preserving dynamical system which we will also call a Bernoulli shift generated by the stochastic vector p. The map φ is an isomorphism between $(\Sigma^{\mathbb{N}}, \mathscr{C}, \nu)$ and $([0,1[, \mathscr{B}, \mu)$ and can be used to transform B into T because

$$\varphi \circ B = T \circ \varphi \tag{2.9}$$

holds true. Now consider the partitions

$$\mathcal{P}_k := \{[i \cdot N^{-k}, (i+1) \cdot N^{-k}[\mid i \in \{0, 1, \dots, N^k - 1\}\} \tag{2.10}$$

for $k \in \mathbb{N}$. It is easy to see that

$$\varphi(\mathcal{C}_k) = \mathcal{P}_k$$

holds true. Using (2.5), this implies

$$\mathcal{P}_k^{(n)} = \mathcal{P}_{k+n-1} \tag{2.11}$$

for all $k, n \in \mathbb{N}$. Additionally, given $k \in \mathbb{N}$ and $P_1, P_2 \in \mathcal{P}_k$, there exists $(a_1^1, a_2^1, \dots, a_k^1), (a_1^2, a_2^2, \dots, a_k^2) \in \Sigma^k$ with

$$P_i = \varphi(C_1^k(a_1^i, a_2^i, \dots, a_k^i))$$

for $i \in \{1, 2\}$. Therefore, for all $t \geq k$, the fact that the measure μ is based on the product measure ν yields

$$\begin{aligned}
\mu(P_1 \cap T^{-t}(P_2)) &= \nu(\varphi^{-1}(P_1 \cap T^{-t}(P_2))) = \nu(\varphi^{-1}(P_1) \cap \varphi^{-1}(T^{-t}(P_2))) \\
&\overset{(2.9)}{=} \nu(\varphi^{-1}(P_1) \cap B^{-t}(\varphi^{-1}(P_2))) \\
&= \nu(\varphi^{-1}(\varphi(C_1^k(a_1^1, a_2^1, \ldots, a_k^1))) \cap B^{-t}(\varphi^{-1}(\varphi(C_1^k(a_1^2, a_2^2, \ldots, a_k^2))))) \\
&= \nu(C_1^k(a_1^1, a_2^1, \ldots, a_k^1) \cap B^{-t}(C_1^k(a_1^2, a_2^2, \ldots, a_k^2))) \\
&= \nu(C_1^{2k+t}(a_1^1, a_2^1, \ldots, a_k^1, b_{k+1}, b_{k+2}, \ldots, b_t, a_1^2, a_2^2, \ldots, a_k^2)) \\
&= \sum_{(b_{k+1},\ldots,b_t)\in\Sigma^{t-k}} \nu(C_1^{2k+t}(a_1^1, a_2^1, \ldots, a_k^1, b_{k+1}, b_{k+2}, \ldots, b_t, a_1^2, a_2^2, \ldots, a_k^2)) \\
&= \sum_{(b_{k+1},\ldots,b_t)\in\Sigma^{t-k}} \left(\prod_{i=1}^{k} p_{a_i^1}\right) \cdot \left(\prod_{i=k}^{t} p_{b_i}\right) \cdot \left(\prod_{i=1}^{k} p_{a_i^2}\right) \\
&= \left(\prod_{i=1}^{k} p_{a_i^1}\right) \cdot \left(\prod_{i=1}^{k} p_{a_i^2}\right) \\
&= \nu(C_1^k(a_1^1, a_2^1, \ldots, a_k^1)) \cdot \nu(C_1^k(a_1^2, a_2^2, \ldots, a_k^2))) \\
&= \mu(P_1) \cdot \mu(P_2).
\end{aligned} \tag{2.12}$$

Thus, iteratively applying the function T creates independence between cylinder sets, which is the main reason why many properties of Bernoulli shifts can be calculated analytically, as can be seen in section 4.2.2 for example.

As another consequence of (2.9), the entropy rates of B and T with regard to the correct partitions are equal:

Lemma 2.29. Let $([0, 1[, \mathscr{B}, T, \mu)$ be a Bernoulli shift generated by the stochastic vector $p \in [0, 1]^N$ and $\mathcal{P}_k$ as in (2.10). Then

$$H\left(\mathcal{P}_k^{(n)}, q\right) = (k + n - 1) \cdot H(p, q).$$

In particular,

$$h(T, \mathcal{P}_k, q) = H(p, q)$$

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

Proof. Let be $([0, 1[, \mathscr{B}, T, \mu)$ a Bernoulli shift generated by the stochastic vector $p \in [0, 1]^N$ and $(\Sigma^{\mathbb{N}}, \mathscr{C}, \nu, B)$ the Bernoulli shift based on the set $\Sigma = \{1, 2, \ldots, N\}$ and generated by the same stochastic vector p. Let further $\varphi : \Sigma^{\mathbb{N}} \to [0, 1]$ be given as in (2.7) and $\mathcal{C}_k$ the partition into cylinder sets of length $k \in \mathbb{N}$ as defined in (2.4). Since $\varphi(\mathcal{C}_k) = \mathcal{P}_k$ holds true for all $k \in \mathbb{N}$, we have

$$T^{-t}(\mathcal{P}_k) \overset{(2.9)}{=} \varphi(\varphi^{-1}(T^{-t}(\mathcal{P}_k))) = \varphi(B^{-t}(\varphi^{-1}(\mathcal{P}_k))) = \varphi(B^{-t}(\mathcal{C}_k)). \tag{2.13}$$

Therefore

$$\begin{aligned}
H_\mu\left(\bigvee_{t=0}^{n-1} T^{-1}(\mathcal{P}_k), q\right) &= \frac{-1}{q-1}\log\left(\sum_{P\in\bigvee_{t=0}^{n-1}T^{-1}(\mathcal{P}_k)} \mu(P)^q\right)\\
&\overset{(2.8)}{=} \frac{-1}{q-1}\log\left(\sum_{P\in\bigvee_{t=0}^{n-1}T^{-t}(\mathcal{P}_k)} \nu(\varphi^{-1}(P))^q\right)\\
&\overset{(2.13)}{=} \frac{-1}{q-1}\log\left(\sum_{P\in\bigvee_{t=0}^{n-1}\varphi B^{-t}(\mathcal{C}_k)} \nu(\varphi^{-1}(P))^q\right)\\
&= \frac{-1}{q-1}\log\left(\sum_{C\in\bigvee_{t=0}^{n-1}B^{-t}(\mathcal{C}_k)} \nu(C)^q\right)\\
&= H_\nu\left(\bigvee_{t=0}^{n-1} B^{-t}(\mathcal{C}_k), q\right)\\
&\overset{(2.6)}{=} (k+n-1)\cdot H(p,q)
\end{aligned}$$

for all $q \in \mathbb{R}$. Consequently

$$\begin{aligned}
h_\mu(T,\mathcal{P}_k,q) &= \liminf_{n\to\infty}\frac{1}{n}H\left(\bigvee_{t=0}^{n-1} B^{-t}(\mathcal{P}_k), q\right)\\
&= \liminf_{n\to\infty}\frac{k+n-1}{n}\cdot H(p,q) = H(p,q)
\end{aligned}$$

holds true for all $q \in \mathbb{R}$. □

Since $(\mathcal{P}_k)_{k\in\mathbb{N}}$ as given in (2.10) is a generating sequence of partitions, the above Lemma together with Theorem 2.28 implies in particular

$$h(T) = \lim_{k\to\infty} h(T,\mathcal{P}_k) = H(p).$$

The map φ does not only preserve the entropy but also the order relation within the respective spaces, more precisely

$$\omega \leq_{\text{lex}} \omega' \Leftrightarrow \varphi(\omega) \leq \varphi(\omega') \tag{2.14}$$

holds true for all $\omega, \omega' \in \Sigma^N$, where "$\leq_{\text{lex}}$" is the so called lexicographic order and "$\leq$" is the standard order on $\mathbb{R}$. The lexicographic order "$\leq_{\text{lex}}$" is a total order defined on $\Sigma^{\mathbb{N}}$. Given $\omega = (\omega_i)_{i=1}^\infty \in \Sigma^{\mathbb{N}}$ and $\omega' = (\omega'_i)_{i=1}^\infty \in \Sigma^{\mathbb{N}}$, then $\omega \leq_{\text{lex}} \omega'_i$ holds true if $\omega = \omega'$ or if $\omega_{i_*} < \omega'_{i_*}$ where $i_* := \min\{n \in \mathbb{N} \mid \omega_i \neq \omega'_i\}$.

Statement (2.14) is especially interesting when considering ordinal patterns, which will be defined in the next section. The equivalence (2.14) implies that, roughly speaking, the ordinal patterns of $(\Sigma^{\mathbb{N}}, \mathscr{C}, \nu, B)$ and $([0,1[, \mathscr{B}, \mu, T)$ are equivalent.

2.4 Ordinal patterns

Definition 2.30 (Ordinal patterns)**.** Let $(\Omega, \mathscr{A}, \mu, T)$ be a measure-preserving dynamical system, $\boldsymbol{X} = (X_1, X_2, \dots, X_d) : \Omega \to \mathbb{R}^d$ a vector of random variables for $d \in \mathbb{N}$ and

$$\mathfrak{S}_n := \{(\pi_0, \pi_1, \dots \pi_{n-1}) \in \{0, 1, \dots n-1\}^n \mid \pi_i \neq \pi_j \text{ for } i \neq j\}$$

the set of all permutations of length n. Given a permutation $\pi = (\pi_0, \pi_1, ..., \pi_{n-1}) \in \mathfrak{S}_n$, we say that $(x_0, x_1, \dots, x_{n-1}) \in \mathbb{R}^n$ has ordinal pattern π if

$$x_{\pi_0} \leq x_{\pi_1} \leq \dots \leq x_{\pi_{n-1}}$$

holds true and if $x_{\pi_t} = x_{\pi_{t+1}}$ implies $\pi_t < \pi_{t+1}$ for all $t \in \{0, 1, \dots, n-2\}$.

We denote the set of points with ordinal pattern π by

$$P_\pi^{X_i} := \{\omega \in \Omega \mid (X_i(T^{\pi_0}(\omega)), X_i(T^{\pi_1}(\omega)), \dots, X_i(T^{\pi_{n-1}}(\omega))) \text{ has ordinal pattern } \pi\}$$

and by

$$OP^{\boldsymbol{X}}(n) := \{P_\pi^{X_i} \neq \emptyset \mid i \in \{1, 2, \dots, d\}, \pi \in \mathfrak{S}_n\}$$

the partition of Ω into these sets.

Given a set Ω, let $\mathrm{id} : \Omega \to \Omega$ with $\mathrm{id}(\omega) = \omega$ for all $\omega \in \Omega$ be the identity map.

For $\Omega \subseteq \mathbb{R}$ and $\boldsymbol{X} = \mathrm{id}$ we simply write $OP(n)$ instead of $OP^{\mathrm{id}}(n)$.

In the above definition, an ordinal pattern of length n was encoded by a permutation $\pi = (\pi_0, \pi_1, \dots, \pi_{n-1}) \in \mathfrak{S}_n$, where $T^{\pi_i}(\omega)$ is the i-th smallest element in the sequence $\omega, T(\omega), \dots T^{n-1}(\omega)$. It can often be helpful to consider a different way of encoding ordinal patterns:

If we know for all $s, t \in \{0, 1, \dots, n-1\}$ whether $T^s(\omega) < T^t(\omega)$ is true, we can determine to which ordinal pattern P_π the point ω belongs to. Therefore, we can encode an ordinal pattern by all pairwise comparisons of elements of the orbit of length n. The set

$$R := \{(x, y) \in \mathbb{R}^2 \mid x < y\} \tag{2.15}$$

will be used throughout this thesis to describe the order relation between two points.

Given two function $f_1, f_2 : \Omega \to \mathbb{R}$, they can be combined to $(f_1 \times f_2) : \Omega^2 \to \mathbb{R}^2$ with

$$(f_1 \times f_2)(\omega_1, \omega_2) := (f_1(\omega_1), f_2(\omega_2))$$

or $(f_1, f_2) : \Omega \to \mathbb{R}^2$ with

$$(f_1, f_2)(\omega) := (f_1(\omega), f_2(\omega)).$$

In particular, this provides

$$(f_1 \times f_2)^{-1}(R) = \{(\omega_1, \omega_2) \in \Omega^2 \mid f_1(\omega_1) < f_2(\omega_2)\} \subseteq \Omega^2$$

and

$$(f_1, f_2)^{-1}(R) = \{\omega \in \Omega \mid f_1(\omega) < f_2(\omega)\} \subseteq \Omega$$

Given a random variable $X : \Omega \to \mathbb{R}$ we define

$$\mathcal{R}_X := (X \times X)^{-1}\{R, \mathbb{R}^2 \setminus R\}.$$

which is a partition of Ω^2. For $\Omega \subseteq \mathbb{R}$ and $X = \mathrm{id}$ we simply write

$$\mathcal{R} := \mathcal{R}_{\mathrm{id}}$$

instead.

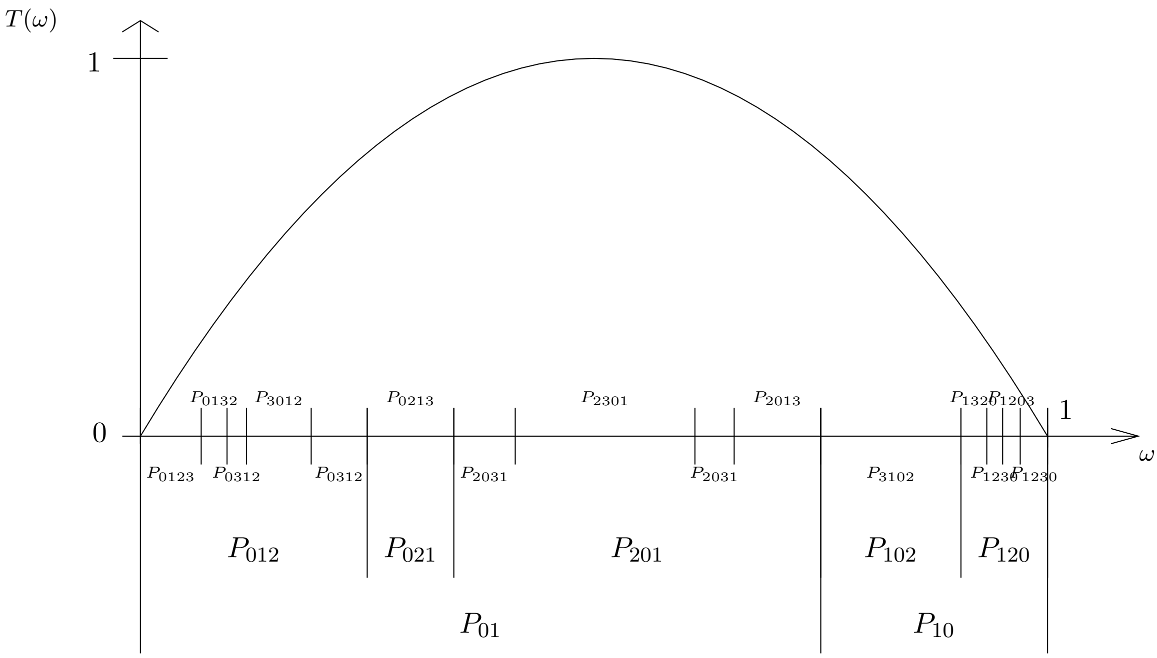
Using this notation, we can alternatively describe the partition into ordinal patterns by

$$OP^{\boldsymbol{X}}(n) = \bigvee_{i=1}^{d} \bigvee_{s=0}^{n-1} \bigvee_{t=s+1}^{n-1} (T^s, T^t)^{-1} (X_i \times X_i)^{-1} (\{R, \mathbb{R}^2 \setminus R\}). \tag{2.16}$$

Example 2 (Example: Ordinal patterns of logistic map)**.** Consider the dynamical system $([0,1], \mathscr{B}, \mu, T)$ where $T : [0,1] \to [0,1]$ is defined by $T(\omega) = 4\omega(1-\omega)$ and $\mu(A) = \int_A (\pi\sqrt{x(1-x)})^{-1} \,\mathrm{d}x$ for all $A \in \mathscr{B}$.

Figure 2.4 shows the partitions into ordinal patterns of length 2, 3 and 4. One can make some observation about ordinal patterns when looking at this Figure: The first obvious observation is that the partition $OP(3)$ of ordinal patterns of length 3 is finer than $OP(2)$ and $OP(4)$ is finer than $OP(3)$. Secondly, not every ordinal pattern is realized. The partition $OP(3)$ could theoretically contain $6 = 3!$ sets, but $P_{210} = \emptyset$ holds true. How many sets $P_\pi \in OP(n)$ are non-empty depending on $n \in \mathbb{N}$ is an interesting question itself (see for example [10]) and is related to the definition of 'topological variants' of the permutation entropy [4].

Additionally, not every pattern is a connected interval. For example, P_{2031} is the union of two non-adjacent intervals.



2.4.1 Permutation entropy

Since taking the supremum over all finite partitions is necessary to calculate the Kolmogorov-Sinai entropy, its determining can be difficult. There are theoretical results that ensure, under some conditions, the existence of a partition, such that the entropy rate with regards to this partition yields the K-S entropy. However, in practice, one does not know how such partitions look like. Additionally, this partition depends on the dynamics T whose precise description might be unknown as well for practical applications.

The permutation entropy has the advantage that it can be calculated without having to find such partitions. The ordinal patterns necessary for calculating the permutation entropy automatically partition the space Ω in a way that can capture the information of a system, independently of the considered map T.

Definition 2.31 (Permutation entropy)**.** Let $(\Omega, \mathscr{A}, \mu, T)$ be a measure-preserving dynamical system. Let $\boldsymbol{X} = (X_1, X_2, \dots X_d)$ be a vector of real-valued random variables defined on Ω and $q \in \mathbb{R}$. Then the lower and upper permutation entropy are defined as

$$\underline{\mathrm{PE}}_\mu(\boldsymbol{X}, T, q) = \liminf_{n\to\infty} \frac{1}{n} H_\mu \left(OP^{\boldsymbol{X}}(n), q\right)$$

and

$$\overline{\mathrm{PE}}_\mu(\boldsymbol{X}, T, q) = \limsup_{n\to\infty} \frac{1}{n} H_\mu \left(OP^{\boldsymbol{X}}(n), q\right), \tag{2.17}$$

respectively. If the choice of the considered measure μ is obvious from the context, we simply write $\underline{\mathrm{PE}}(\boldsymbol{X}, T, q)$ and $\overline{\mathrm{PE}}(\boldsymbol{X}, T, q)$ instead of $\overline{\mathrm{PE}}_\mu(\boldsymbol{X}, T, q)$ and $\underline{\mathrm{PE}}_\mu(\boldsymbol{X}, T, q)$.

Choosing $\boldsymbol{X} = id$ and $q = 1$ in definition 2.31 provides the definition of the one-dimensional permutation entropy as originally defined by Bandt and Pompe [5]:

$$\underline{\mathrm{PE}}(T) := \underline{\mathrm{PE}}(id, T, 1)$$

Remark 2.32. Unlike in Definition 2.23 of the Kolmogorov-Sinai entropy, one does not know whether $\frac{1}{n}H(OP(n))$ converges for $n \to \infty$. Therefore, we differentiate between the permutation entropy defined using the limit inferior and the entropy using the limit superior. However, in many of the here considered cases the limit inferior will behave equivalent to the limit superior in the definition of the permutation entropy.

Different definitions can be found in the literature: Amigó, Kennel and Kocarev used the limit inferior for the definition of the permutation entropy [2]. Bandt, Pompe and G. Keller used the limit directly while showing that the limit exists [4]. K. Keller, A. M. Unakafov and V. A. Unakafova used the limit superior in [16].

3 Permutation entropy (based on Rényi entropies) for one-dimensional systems

3.1 Overview

In this chapter, we will focus on one-dimensional dynamical systems $(\Omega, \mathscr{B}, \mu, T)$, meaning $\Omega \subseteq \mathbb{R}$. The σ-algebra $\mathscr{B}$ is the Borel σ-algebra defined on Ω, i.e. the smallest σ-algebra containing the sets $\{\omega \in \Omega \mid \omega_1 \leq \omega < \omega_2\}$ for all $\omega_1, \omega_2 \in \mathbb{R} \cup \{-\infty, +\infty\}$. We define the one-dimensional permutation entropies

$$\underline{\mathrm{PE}}(T, q) := \underline{\mathrm{PE}}(id, T, q)$$

and

$$\overline{\mathrm{PE}}(T, q) := \overline{\mathrm{PE}}(id, T, q)$$

for parameters $q \in \mathbb{R}$ and simply write $\underline{\mathrm{PE}}(T)$ and $\overline{\mathrm{PE}}(T)$ for $q = 1$.

The following theorem is due to Takens and Verbitskiy. They originally proved the statement of this theorem for invertible systems only and it was later mentioned in Verbitskiy's doctoral thesis that the same holds true for non-invertible systems [11]. There also exists a statement for non-ergodic systems (see [35]), but we will not consider this here.

Theorem 3.1 (Takens and Verbitskiy [34])**.** Let $(\Omega, \mathscr{A}, \mu)$ be a Standard probability space and $T : \Omega \to \Omega$ an aperiodic and ergodic measure-preserving function. Then

$$h(T, q) = h(T)$$

holds true for all $q \geq 1$. Additionally, $h(T) > 0$ implies

$$h(T, q) = \infty$$

for all $q < 1$.

It is natural to wonder whether a statement similar to the above theorem holds also true for the permutation entropies $\underline{\mathrm{PE}}(T, q)$ or $\overline{\mathrm{PE}}(T, q)$ for $q \geq 1$ and how these entropies are related to the (generalized) Kolmogorov-Sinai entropy. We will study this question for general $q \in \mathbb{R}$ in this chapter whenever possible. However, we can only show some statements for $q = 1$. In this case we will prove the specific statement and outline the difficulties of generalizing the specific statements to arbitrary $q \in \mathbb{R}$.

Table 3.1 gives an overview over the statements that are considered in this chapter. Theorem 3.15 was first proven by Bandt Pompe and G. Keller and the first theoretical result about the permutation entropy [4]. Theorem 3.17 is the main result of this chapter and generalizes Theorem 3.15 to more general dynamical systems.

The statement of Theorem 3.10 for $q = 1$ was first explicitly stated and proven by Amigo [2]. In this chapter we state with Theorem 3.10 an inequality that holds true also for $q \neq 1$. In case of $q \neq 1$ we can, however, only relate $\underline{\mathrm{PE}}(T, q)$ to $\sup_{P \in \mathbb{P}_o(\mathscr{B})} h(T, \mathcal{P}, q)$, which is the

<table>
<tr><th>ergodic systems</th><th>$q < 0$</th><th>$q \in [0, 1[$</th><th>$q = 1$</th><th>$q > 1$</th></tr>
<tr><td>piecewise monotone & contin.</td><td></td><td>Theorem 3.13: $\overline{\mathrm{PE}}(T, q) \leq \log \#\mathcal{M}$</td><td>Theorem 3.15: $\overline{\mathrm{PE}}(T) = h(T)$</td><td rowspan="2">Theorem 3.10: $\underline{\mathrm{PE}}(T, q) \geq \sup_{P \in \mathbb{P}_o(\mathscr{B})} h(T, \mathcal{P}, q)$
$\overline{\mathrm{PE}}(T, q) < h(T)$ possible (Example 5)</td></tr>
<tr><td>countably piecewise monotone</td><td colspan="2" rowspan="2">Theorem 3.10: $\underline{\mathrm{PE}}(T, q) \geq \sup_{P \in \mathbb{P}_o(\mathscr{B})} h(T, \mathcal{P}, q)$</td><td>Theorem 3.17 $\overline{\mathrm{PE}}(T) = h(T)$</td></tr>
<tr><td>general dynamical system</td><td>Theorem 3.10: $\underline{\mathrm{PE}}(T) \geq h(T)$</td><td>Theorem 3.10: $\underline{\mathrm{PE}}(T, q) \geq \sup_{P \in \mathbb{P}_o(\mathscr{B})} h(T, \mathcal{P}, q)$</td></tr>
</table>

Table 3.1: Overview of different results in this chapter for one-dimensional dynamical systems. For simplicity, only the results for the ergodic case are displayed. The partition into sets on which T is monotone is denoted by $\mathcal{M}$.

supremum of the entropy rates over all ordered partitions. It is in general not known for what dynamical systems and parameters $q \neq 1$ this supremum is equal to $h(T, q)$, which equals the supremum of the entropy rates over all partitions. Corollary 3.14 shows that for $q < 1$ there exist systems for which those two suprema can take different values. This corollary is a consequence of Theorem 3.13 which shows that $\underline{\mathrm{PE}}(T, q)$ and $\overline{\mathrm{PE}}(T, q)$ are both equal to the Kolmogorov-Sinai entropy $h(T)$ for $q \in [0, 1]$ if the dynamical is piecewise monotone and continuous.

For $q > 1$ we give an example of a piecewise monotone and continuous ergodic dynamical system for where $\overline{\mathrm{PE}}(T, q)$ is strictly smaller than $h(T)$. This is not possible for $q = 1$ which shows that, unlike the generalized Kolmogorov-Sinai entropy, the permutation entropy based on Rényi entropies behaves differently for $q > 1$ than for $q = 1$.

There are still many questions that are not answered in this thesis and to which there is no known answer so far. For example, it is still an open question whether $\underline{\mathrm{PE}}(T) \leq h(T)$ holds true in general. There exists no proof but also no counterexample to this statement. For $q \neq 1$ we can typically only establish an inequality.

In Section 3.5 we investigate conditional variants of the permutation entropy. It was first noticed in [37] that a conditional permutation entropy can converge faster than the original permutation entropy which makes them attractive for practical applications.

3.2 Conditional entropies

One can quantify the relationship between two kinds of entropy by looking at the conditional entropy of the partitions used in the definition of the corresponding entropies. Conditional entropies are defined as follows:

Definition 3.2 (Conditional entropy)**.** Let $(\Omega, \mathscr{A}, \mu)$ be a probability space and $\mathcal{P}$ and $\mathcal{Q}$ finite or countable partitions of Ω. Then

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

defines the conditional Rényi entropy of $\mathcal{P}$ with respect to $\mathcal{Q}$ for all $q \in \mathbb{R}$.

It follows from Lemma 2.25 that

$$H(\mathcal{P}, q) - H(\mathcal{Q}, q) \leq H(\mathcal{P} \vee \mathcal{Q}, q) - H(\mathcal{Q}, q) = H(\mathcal{P}|\mathcal{Q}, q)$$

holds true.

By choosing $\mathcal{P}$ and $\mathcal{Q}$ as the partition into symbolic and into ordinal patterns, we can use the above equation to quantify the entropy difference of the Kolmogorov-Sinai and permutation entropy.

It is, however, often difficult to determine the conditional entropy exactly. Therefore, we establish an upper bound for the conditional entropy that depends on the 'amount of intersection' between two partitions. Let $(\Omega, \mathscr{A}, \mu)$ be a probability space and $\mathcal{P}$ and $\mathcal{Q}$ two partitions of Ω. We define

$$\Delta(\mathcal{P}|Q) := \{P \in \mathcal{P} : \mu(P \cap Q) > 0\} \tag{3.1}$$

as the collection of all subsets of $\mathcal{P}$ that are intersecting the set $Q \in \mathcal{Q}$ μ-almost surely.

Lemma 3.3. Let $(\Omega, \mathscr{A}, \mu)$ be a probability space and $\mathcal{P}$ and $\mathcal{Q}$ two finite or countable partitions of Ω. Then

$$H(\mathcal{P}|\mathcal{Q}, q) \leq \begin{cases} \sum_{Q \in \mathcal{Q}} \mu(Q) \cdot \log(\#\Delta(\mathcal{P}|Q)) & \text{for } q = 1 \\ \max_{Q \in \mathcal{Q}} \log(\#\Delta(\mathcal{P}|Q)) & \text{for } q \neq 1 \end{cases}$$

holds true, where $\#A$ denotes the number of elements contained in a set A.

Proof. Consider the function $f :]0, \infty[\to \mathbb{R}$ with $f(x) = x^q$. Since $f''(x) = q(q-1)x^{q-2}$, we have $f''(x) < 0$ for $q < 1$ and $f''(x) > 0$ for $q > 1$ for all $x > 0$. Thus, f is concave for $q < 1$ and convex for $q > 1$. Therefore, Jensen's inequality (Lemma 2.14) implies

$$\sum_{P \in \Delta(\mathcal{P}|Q)} \frac{f(\mu(P \cap Q))}{\#\Delta(P|Q)} \begin{cases} \leq f\left(\sum\limits_{P \in \Delta(\mathcal{P}|Q)} \frac{\mu(P \cap Q)}{\#\Delta(\mathcal{P}|Q)}\right) & \text{if } q < 1 \\ \geq f\left(\sum\limits_{P \in \Delta(\mathcal{P}|Q)} \frac{\mu(P \cap Q)}{\#\Delta(\mathcal{P}|Q)}\right) & \text{if } q > 1 \end{cases} \tag{3.2}$$

for all $Q \in \mathcal{Q}$. Using this and the fact that the function $\frac{-1}{q-1} \log$ is monotonically increasing for $q < 1$ and decreasing for $q > 1$ provides

$$\begin{aligned}
H(\mathcal{P} \vee \mathcal{Q}, q) &= \frac{-1}{q-1} \log \left(\sum_{Q\in\mathcal{Q}} \sum_{P\in\mathcal{P}} \mu(P \cap Q)^q \right) \\
&= \frac{-1}{q-1} \log \left(\sum_{Q\in\mathcal{Q}} \sum_{P\in\Delta(\mathcal{P}|Q)} \mu(P \cap Q)^q \right) \\
&= \frac{-1}{q-1} \log \left(\sum_{Q\in\mathcal{Q}} \#\Delta(\mathcal{P}|Q) \sum_{P\in\Delta(\mathcal{P}|Q)} \frac{f(\mu(P\cap Q))}{\#\Delta(\mathcal{P}|Q)} \right) \\
&\overset{(3.2)}{\leq} \frac{-1}{q-1} \log \left(\sum_{Q\in\mathcal{Q}} \#\Delta(\mathcal{P}|Q) \cdot f\left(\sum_{P\in\Delta(\mathcal{P}|Q)} \frac{\mu(P\cap Q)}{\#\Delta(\mathcal{P}|Q)} \right) \right) \\
&= \frac{-1}{q-1} \log \left(\sum_{Q\in\mathcal{Q}} \#\Delta(\mathcal{P}|Q) \cdot f\left(\frac{\mu(Q)}{\#\Delta(\mathcal{P}|Q)} \right) \right) \\
&= \frac{-1}{q-1} \log \left(\sum_{Q\in\mathcal{Q}} \#\Delta(\mathcal{P}|Q)^{1-q} \cdot \mu(Q)^q \right) \\
&\leq \frac{-1}{q-1} \log \left(\left(\max_{Q\in\mathcal{Q}} \#\Delta(\mathcal{P}|Q) \right)^{1-q} \cdot \sum_{Q\in\mathcal{Q}} \mu(Q)^q \right) \\
&= \frac{-1}{q-1} \log \left(\left(\max_{Q\in\mathcal{Q}} \#\Delta(\mathcal{P}|Q) \right)^{1-q} \right) + \frac{-1}{q-1} \log \left(\sum_{Q\in\mathcal{Q}} \mu(Q)^q \right) \\
&= \max_{Q\in\mathcal{Q}} \log(\#\Delta(\mathcal{P}|Q)) + H(\mathcal{Q}, q)
\end{aligned}$$

for all $q \neq 1$. This is equivalent to $H(\mathcal{P}|\mathcal{Q}, q) \leq \max_{Q\in\mathcal{Q}} \log(\#\Delta(\mathcal{P}|Q))$.

For $q = 1$, consider the function $g :]0,\infty] \to]0,\infty]$ with $g(x) = x\log(x)$. This function is convex because $g''(x) = x^{-1} > 0$. So we can use Jensen's inequality to show

$$\begin{aligned}
&\sum_{P\in\Delta(\mathcal{P}|Q)} \mu(P\cap Q)\log(\mu(P\cap Q)) \\
=&\#\Delta(\mathcal{P}|Q) \sum_{P\in\Delta(\mathcal{P}|Q)} \frac{1}{\#\Delta(\mathcal{P}|Q)} \cdot g(\mu(P\cap Q)) \\
\geq&\#\Delta(\mathcal{P}|Q) \cdot g\left(\sum_{P\in\Delta(\mathcal{P}|Q)} \frac{1}{\#\Delta(\mathcal{P}|Q)} \cdot \mu(P\cap Q) \right) \\
=&\#\Delta(\mathcal{P}|Q) \cdot g\left(\frac{\mu(Q)}{\#\Delta(\mathcal{P}|Q)} \right) \\
=&\#\Delta(\mathcal{P}|Q) \cdot \frac{\mu(Q)}{\#\Delta(\mathcal{P}|Q)} \cdot \log\left(\frac{\mu(Q)}{\#\Delta(\mathcal{P}|Q)} \right) \\
=&\mu(Q) \cdot (\log(\mu(Q)) - \log(\#\Delta(\mathcal{P}|Q)))\,.
\end{aligned}$$

Using the above inequality provides

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

This is equivalent to $H(\mathcal{P}|\mathcal{Q}) \leq \sum_{Q \in \mathcal{Q}} \mu(Q) \cdot \log(\#\Delta(\mathcal{P}|Q))$. □

So the conditional entropy is bounded by the average logarithmic 'amount of intersection' for $q = 1$ and the maximal logarithmic 'amount of intersection' for $q \neq 1$. Notice that the upper bound for $q \neq 1$ in the above lemma cannot be smaller than the upper bound for $q = 1$.

3.3 Permutation entropy as an upper bound for K-S entropy

The aim of this section is to investigate if the permutation entropy is an upper bound for the Kolmogorov-Sinai entropy, i.e. if

$$\underline{\mathrm{PE}}(T, q) \geq h(T, q)$$

holds true for all $q \geq 1$. We will focus on the lower permutation entropy because the obvious inequality $\overline{\mathrm{PE}}(T, q) \geq \underline{\mathrm{PE}}(T, q)$ implies that the upper permutation entropy will be an upper bound if the lower permutation entropy is.

3.3.1 Combinatorial approach using Rényi entropy

In this subsection, we want to show that the permutation entropy is an upper bound for the Kolmogorov-Sinai entropy by using combinatorial arguments. To achieve this, it is necessary that the order information given by the partition into ordinal patterns $OP(n)$ can be used to infer information about the possible symbolic patterns $\mathcal{P}^{(n)}$. This is only possible, if the elements of $\mathcal{P}$ can be ordered in some sense. We will call partitions whose elements can be ordered *ordered partitions*.

Definition 3.4 (Ordered partitions)**.** Let Ω be a subset of $\mathbb{R}$ and $\mathscr{B}$ be the Borel σ-algebra on Ω and μ be a probability measure on $(\Omega, \mathscr{B})$. Then we call a partition $\mathcal{P} = \{P_i\}_{i \in I}$ of Ω *ordered* (with regard to μ), if

$$\mu^2\left(\{(\omega_1, \omega_2) \in P_{i_1} \times P_{i_2} \mid \omega_1 < \omega_2\}\right) \in \left\{0, \mu^2(P_{i_1} \times P_{i_2})\right\} \tag{3.3}$$

holds true for all $i_1, i_2 \in I$ with $i_1 \neq i_2$. Here μ^2 denotes the product measure of μ with itself on the product σ-algebra $\mathscr{B} \otimes \mathscr{B}$. We define

$$\mathbb{P}_o(\mathscr{B}) := \{\mathcal{P} \in \mathbb{P}(\mathscr{B}) \mid \mathcal{P} \text{ is an ordered partition}\}$$

and

$$\mathbb{P}_o^c(\mathscr{B}) := \{\mathcal{P} \in \mathbb{P}^c(\mathscr{B}) \mid \mathcal{P} \text{ is an ordered partition}\}.$$

The product measure μ^2 is defined by

$$\mu^2(A \times B) = \mu(A) \cdot \mu(B)$$

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

Remark 3.5. Notice that

$$\{(\omega_1, \omega_2) \in P_{i_1} \times P_{i_2} \mid \omega_1 < \omega_2\} = (P_{i_1} \times P_{i_2}) \cap \bigcup_{q \in \mathbb{Q}}]-\infty, q] \times]q, \infty[\quad \in \mathscr{B} \otimes \mathscr{B}$$

holds true. Therefore, the probability $\mu^2\left((P_{i_1} \times P_{i_2}) \cap R\right)$ is well defined. This might not necessarily be true if the Borel σ-algebra $\mathscr{B}$ is replaced with an arbitrary σ-algebra.

Given two disjoint intervals P_{i_1} and P_{i_2} for $i_1, i_2 \in I$ with $i_1 \neq i_2$, one of the intervals has to be located to the left of the other intervals. Without restriction of generality we can assume that P_{i_1} is the left interval. Then any point $\omega_1 \in P_{i_1}$ is smaller than any point $\omega_2 \in P_{i_2}$. Hence

$$\{(\omega_1, \omega_2) \in P_{i_1} \times P_{i_2} \mid \omega_1 < \omega_2\} = P_{i_1} \times P_{i_2}$$

and

$$\{(\omega_1, \omega_2) \in P_{i_2} \times P_{i_1} \mid \omega_1 < \omega_2\} = \emptyset$$

hold true, which implies

$$\mu^2\left(\{(\omega_1, \omega_2) \in P_{i_1} \times P_{i_2} \mid \omega_1 < \omega_2\}\right) \in \left\{0, \mu^2(P_{i_1} \times P_{i_2})\right\}$$

for all $i_1, i_2 \in I$ with $i_1 \neq i_2$. So every partition $\mathcal{P} = \{P_i\}_{i \in I}$ into intervals is an ordered partition. Therefore, the partitions in $\mathbb{P}_o(\mathscr{B})$ can be seen as probabilistic versions of interval partitions.

In the following lemma, we will prove the existence of an upper bound for the 'amount of intersection'

$$\Delta(\mathcal{P}^{(n)} | P_\pi) = \{P \in \mathcal{P}^{(n)} \mid \mu(P \cap P_\pi) > 0\}$$

between ordinal and symbolic patterns, given an ordered partition $\mathcal{P}$.

Lemma 3.6. Let $(\Omega, \mathscr{A}, \mu, T)$ be a measure-preserving dynamical system, $X : \Omega \to \mathbb{R}$ a random variable and $\mathcal{U}$ a finite ordered partition of $\mathbb{R}$ with regard to μ_X. Then for $\mathcal{P} := X^{-1}(\mathcal{U})$, $n \in \mathbb{N}$ and all $P_\pi \in OP^X(n)$

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

holds true.

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

$$i_1 < i_2 \Rightarrow \mu^2(\{(\omega_1, \omega_2) \in (X^{-1}(U_{i_1}) \times X^{-1}(U_{i_2})) \mid X(\omega_1) > X(\omega_2)\}) = 0 \tag{3.4}$$

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

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

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

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

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

For all $\omega \in P_\pi$,

$$X(T^{\pi_0}(\omega)) \leq X(T^{\pi_1}(\omega)) \leq \dots \leq X(T^{\pi_{n-1}}(\omega))$$

holds true. Using (3.4), this implies

$$i_{\pi_0} \leq i_{\pi_1} \leq \dots \leq i_{\pi_{n-1}}$$

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

$$\begin{aligned}\#\Delta(\mathcal{P}^{(n)}|P_\pi) &= \#\{(i_0, i_1, \dots i_{n-1}) \in I^n \mid \mu(P(\boldsymbol{i}) \cap P_\pi) > 0\} \\ &\leq \#\{(i_0, i_1, \dots i_{n-1}) \in I^n \mid i_{\pi_0} \leq i_{\pi_1} \leq \dots \leq i_{\pi_{n-1}}\} \\ &= \binom{n + \#\mathcal{U} - 1}{\#\mathcal{U} - 1}\end{aligned}$$

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

This lemma can be used to directly proof the following two corollaries 3.7 and 3.9.

Corollary 3.7. Let $(\Omega, \mathscr{A}, \mu, T)$ be a measure-preserving dynamical system, $\boldsymbol{X} = (X_1, X_2, \dots, X_d) : \Omega \to \mathbb{R}^d$ a vector of random variables and $\mathcal{U}$ a finite partition of $\mathbb{R}$ into intervals. Then

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

holds true.

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

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

holds true for all $i \in \{1, 2, \ldots, d\}$. Together with Lemma 3.6 and 3.3, this provides

$$
\begin{aligned}
h\left(T, \bigvee_{i=1}^{d} \mathcal{P}_i\right) &= \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=1}^{d} \mathcal{P}_i^{(n)}\right) = \lim_{m\to\infty} \frac{1}{mk} H\left(\bigvee_{i=1}^{d} \mathcal{P}_i^{(mk)}\right) \\
&\leq \lim_{m\to\infty} \frac{1}{mk} H\left(\bigvee_{i=1}^{d} \mathcal{P}_i^{(mk)} \vee (OP^{\boldsymbol{X}}(k))^{(mk)}\right) \\
&= \lim_{m\to\infty} \frac{1}{mk} \left[H((OP^{\boldsymbol{X}}(k))^{(mk)}) + H\left(\bigvee_{i=1}^{d} \mathcal{P}_i^{(mk)} | (OP^{\boldsymbol{X}}(k))^{(mk)}\right)\right] \\
&\leq \lim_{m\to\infty} \frac{1}{mk} \left[H((OP^{\boldsymbol{X}}(k))^{(mk)}) + \sum_{i=1}^{d} H\left(\mathcal{P}_i^{(mk)} | (OP^{\boldsymbol{X}}(k))^{(mk)}\right)\right] \\
&\leq \lim_{m\to\infty} \frac{1}{mk} \left[H((OP^{\boldsymbol{X}}(k))^{(mk)}) + \sum_{i=1}^{d} H\left(\mathcal{P}_i^{(mk)} | (OP^{X_i}(k))^{(mk)}\right)\right] \\
&\leq \lim_{m\to\infty} \left[\frac{1}{mk} H((OP^{\boldsymbol{X}}(k))^{(mk)}) + \sum_{i=1}^{d} \frac{1}{k} H(\mathcal{P}_i^{(k)} | OP^{X_i}(k))\right] \\
&= h(T, OP^{\boldsymbol{X}}(k)) + \sum_{i=1}^{d} \frac{1}{k} H(\mathcal{P}_i^{(k)} | OP^{X_i}(k)) \\
&\leq h(T, OP^{\boldsymbol{X}}(k)) + \sum_{i=1}^{d} \frac{1}{k} \sum_{P_\pi^{X_i} \in OP^{X_i}(k)} \mu(P_\pi^{X_i}) \log(\#\Delta(\mathcal{P}_i^{(k)} | P_\pi^{X_i})) \\
&\leq h(T, OP^{\boldsymbol{X}}(k)) + \sum_{i=1}^{d} \frac{1}{k} \sum_{P_\pi^{X_i} \in OP^{X_i}(k)} \mu(P_\pi^{X_i}) \log\left(\binom{k + \#\mathcal{U} - 1}{\#\mathcal{U} - 1}\right) \\
&= h(T, OP(k)) + \frac{d}{k} \log\left(\binom{k + \#\mathcal{U} - 1}{\#\mathcal{U} - 1}\right) \\
&\leq h(T, OP(k)) + \frac{d}{k} \log\left((k + \#\mathcal{U} - 1)^{\#\mathcal{U}-1}\right) \\
&= h(T, OP(k)) + d(\#\mathcal{U} - 1) \frac{\log(k + \#\mathcal{U} - 1)}{k}
\end{aligned}
$$

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

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

The following theorem can be found in [3]. We give here a different and shorter proof than the one shown in [3] that relies mainly on the combinatorial argument of the above corollary.

Theorem 3.8. Let $(\Omega, \mathscr{A}, \mu, T)$ be a measure-preserving dynamical system and $\boldsymbol{X} = (X_1, X_2, \ldots, X_d) : \Omega \to \mathbb{R}^d$ a vector of random variables satisfying

$$
\sigma\left(\left\{X_i \circ T^t \mid t \in \mathbb{N}_0, i \in \{1, 2, \ldots, d\}\right\}\right) = \mathscr{A}. \tag{3.5}
$$

Then

$$
h(T) = \lim_{k\to\infty} h(T, OP^{\boldsymbol{X}}(k))
$$

holds true.

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

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

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

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

holds true. Using (3.5), this implies

$$\begin{aligned} \mathscr{A} &= \sigma\left(\left\{T^{-t}(\boldsymbol{X}^{-1}(p_i^{-1}(\mathcal{U}_l))) \mid t \in \mathbb{N}_0, l \in \mathbb{N}, i \in \{1, 2, \ldots, d\}\right\}\right) \\ &= \sigma\left(\left\{T^{-t}(X_i^{-1}(\mathcal{U}_l)) \mid t \in \mathbb{N}_0, l \in \mathbb{N}, i \in \{1, 2, \ldots, d\}\right\}\right) \\ &= \sigma\left(\bigvee_{l=1}^{\infty} \bigvee_{t=1}^{\infty} T^{-t}\left(\left\{X_i^{-1}(\mathcal{U}_l) \mid i \in \{1, 2, \ldots, d\}\right\}\right)\right) \end{aligned}$$

Thus, $(\mathcal{P}_l)_{l\in\mathbb{N}}$ with

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

is a generating sequence of finite partitions, which implies

$$h(T) = \lim_{l\to\infty} h(T, \mathcal{P}_l)$$

according to Theorem 2.28. Corollary 3.7 provides

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

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

$$h(T) = \lim_{l\to\infty} h(T, \mathcal{P}_l) \leq \lim_{k\to\infty} h(T, OP^{\boldsymbol{X}}(k)). \tag{3.6}$$

On the other hand,

$$h(T) = \sup_{\mathcal{P}\in\mathbb{P}(\mathscr{A})} h(T, \mathcal{P}) \geq \lim_{k\to\infty} h(T, OP^{\boldsymbol{X}}(k))$$

holds true, which, together with (3.6), finishes the proof. □

Corollary 3.9. Let $(\Omega, \mathscr{B}, \mu, T)$ be a measure-preserving dynamical system with $\Omega \subseteq \mathbb{R}$. Then for all finite ordered partitions $\mathcal{P} \in \mathbb{P}_o(\mathscr{B})$

$$\underline{\mathrm{PE}}(T, q) \geq h(T, \mathcal{P}, q)$$

is true for all $q \in \mathbb{R}$.

Proof. Using Lemma 2.25, Lemma 3.6 and Lemma 3.3 yields

$$\begin{aligned}
H(\mathcal{P}^{(n)},q) \leq& H(\mathcal{P}^{(n)} \vee OP(n),q) = H(OP(n),q) + H(\mathcal{P}^{(n)}|OP(n),q)\\
\leq& H(OP(n),q) + \max_{\pi\in\mathcal{S}_n} \log(\#\Delta(\mathcal{P}^{(n)}|P_\pi))\\
\leq& H(OP(n),q) + \max_{\pi\in\mathcal{S}_n} \log\left(\binom{n+\#\mathcal{P}-1}{\#\mathcal{P}-1}\right)\\
\leq& H(OP(n),q) + \max_{\pi\in\mathcal{S}_n} \log\left((n+\#\mathcal{P}-1)^{\#\mathcal{P}-1}\right)\\
\leq& H(OP(n),q) + (\#\mathcal{P}-1)\log(n+\#\mathcal{P}-1)
\end{aligned}$$

for all $q \in \mathbb{R}$. Hence,

$$\begin{aligned}
\underline{\mathrm{PE}}(T,q) &= \liminf_{n\to\infty} \frac{1}{n} H\left(OP(n),q\right)\\
&\geq \liminf_{n\to\infty} \frac{1}{n}\left[H(\mathcal{P}^{(n)},q) - (\#\mathcal{P}-1)\log(n+\#\mathcal{P}-1)\right]\\
&= \liminf_{n\to\infty} \frac{1}{n} H\left(\mathcal{P}^{(n)},q\right) = h(T,\mathcal{P},q).
\end{aligned}$$

□

Theorem 3.10. Let $(\Omega, \mathscr{B}, \mu, T)$ be a measure-preserving dynamical system with $\Omega \subseteq \mathbb{R}$. Then

$$\underline{\mathrm{PE}}(T,q) \geq \sup_{\mathcal{P}\in\mathbb{P}_o(\mathscr{B})} h(T,\mathcal{P},q)$$

holds true for all $q \in \mathbb{R}$. In particular,

$$\underline{\mathrm{PE}}(T) \geq h(T)$$

is true.

Proof. The first inequality is an immediate consequence of Corollary 3.9. Since the Borel σ-algebra is generated by intervals, Theorem 2.28 provides

$$\sup_{\mathcal{P}\in\mathbb{P}_o(\mathscr{B})} h(T,\mathcal{P}) = h(T)$$

This implies the second inequality. □

3.3.2 Generating partitions for Rényi entropy

As an alternative approach to using combinatorial arguments one can use the generating properties of ordinal patterns to show that the permutation entropy is an upper bound for the Kolmogorov-Sinai entropy using Rényi entropy.

Recall that a sequence of partitions $\{\mathcal{P}_k\}_{k\in\mathbb{N}}$ is called generating sequence of partitions, if $\mathcal{P}_k \prec P_{k+1}$ for all $k \in \mathbb{N}$ and

$$\sigma\left(\bigvee_{k=0}^{\infty} \mathcal{P}_k\right) \underset{\mu}{=} \mathscr{A}$$

holds true (see Definition 2.27).

One of the key properties of ordinal patterns is the fact that the sequence $(OP(k))_{k\in\mathbb{N}}$ of ordinal patterns is generating for ergodic systems:

Theorem 3.11 (Ordinal patterns are generating [3])**.** Let $(\Omega, \mathscr{B}, \mu, T)$ be an ergodic measure-preserving dynamical system with $\Omega \subseteq \mathbb{R}$. Then $(OP(k))_{k\in\mathbb{N}}$ is a generating sequence of partitions, i.e.

$$\sigma\left(OP(k)\right)_{k\in\mathbb{N}}) \underset{\mu}{=} \mathscr{B}$$

holds true. This implies

$$\lim_{k\to\infty} h(T, OP(k)) = h(T)$$

A proof of this theorem can be found in [3]. One can also show that in the one-dimensional case the limit of the entropy rates based on ordinal patterns is a lower bound for the permutation entropy. To show this, one does not even need to require that the underlying system is ergodic.

Lemma 3.12. Let $(\Omega, \mathscr{B}, \mu, T)$ be a measure-preserving dynamical system with $\Omega \subseteq \mathbb{R}$. Then

$$\underline{\mathrm{PE}}(T, q) \geq \lim_{k\to\infty} h(T, OP(k), q)$$

holds true for all $q \in \mathbb{R}$.

Proof. For all $k, n \in \mathbb{N}$

$$OP(n+k) \succ OP(k)^{(n)}$$

holds true. According to Lemma 2.25, this implies

$$\mathrm{PE}(T, q) = \liminf_{n\to\infty} \frac{1}{n} H(OP(n+k), q) \geq \liminf_{n\to\infty} \frac{1}{n} H(OP(k)^n, q) = h(T, OP(k), q)$$

for all $k \in \mathbb{N}$ and $q \in \mathbb{R}$. Thus,

$$\mathrm{PE}(T, q) \geq \lim_{k\to\infty} h(T, OP(k), q)$$ □

We would like to show that $\underline{\mathrm{PE}}(T, q) \geq h(T, q)$ holds true for all $q \in \mathbb{R}$ using the above lemma. Therefore, we need to investigate, whether

$$\lim_{k\to\infty} h(T, OP(k), q) = h(T, q) \tag{3.7}$$

holds true for all $q \in \mathbb{R}$. For $q = 1$ and ergodic systems, this follows from Theorems 3.11 and 2.28, using the fact that the sequence of ordinal patterns is generating. For $q \neq 1$ we can not show that (3.7) is true because there exists no statement like the one given in Theorem 2.28 when using Rényi entropy for $q \neq 1$. In the following, we will give an example of a dynamical system and a generating sequence of partitions that does not generate the entropy for $q \neq 1$:

Choose the stochastic vector p with $H(p) \neq H(p, q)$ for $q > 1$, for example $p = (1/3, 2/3)$. Let $([0, 1[, \mathscr{B}, T, \mu)$ be the Bernoulli shift generated by the stochastic vector p and

$$\mathcal{P}_K := \{[i \cdot N^{-k}, (i+1) \cdot N^{-k}[\mid i \in \{0, 1, \ldots, N^k - 1\}\}$$

Lemma 2.29 implies

$$\lim_{k\to\infty} h(T, \mathcal{P}_k, q) = H(p, q)$$

for all $q \in \mathbb{R}$. For $q = 1$, this provides

$$h(T) = \lim_{k\to\infty} h(T, \mathcal{P}_k) = H(p) \tag{3.8}$$

due to Theorem 2.28 and because $(\mathcal{P}_k)_{k\in\mathbb{N}}$ is a sequence of generating partitions (see example 1).

Since Bernoulli shifts are aperiodic and ergodic (see for example [7]), Theorem 3.1 provides

$$h(T, q) = h(T)$$

for all $q > 1$. Using the fact that $h(T) = H(p)$ holds true for a Bernoulli shift T, one can conclude

$$\lim_{k\to\infty} h(T, \mathcal{P}_k, q) = H(p, q) \neq H(p) = h(T) = h(T, q)$$

holds true for all $q > 1$.

So we have shown that a generating sequence of partitions does in general not generate the entropy for $q \neq 1$. As a consequence, we cannot use properties of generating partitions for $q \neq 1$ to determine the entropy and are, therefore, not able to show that (3.7) holds true.

3.3.3 Permutation entropy for $q < 1$

Takens and Verbitsky showed that

$$h(T, q) = \infty$$

holds true for all $q < 1$ and aperiodic maps T [36]. So it is natural to ask whether $\mathrm{PE}(T, q) = \infty$ is true as well for $q < 1$. This question is directly related to the following question:

Does

$$\sup_{P\in\mathbb{P}_o(\mathscr{B})} h(T, \mathcal{P}, q) \stackrel{?}{=} h(T, q) \tag{3.9}$$

hold true for $q < 1$?

Theorem 3.10 provides

$$\mathrm{PE}(T, q) \geq \sup_{P\in\mathbb{P}_o(\mathscr{B})} h(T, \mathcal{P}, q)$$

for all $q \in \mathbb{R}$, so a positive answer to (3.9) would imply $\mathrm{PE}(T, q) = \infty$. However, this is in general not true as we will see in the following Theorem.

Theorem 3.13. Let $(\Omega, \mathscr{B}, \mu, T)$ be a measure-preserving dynamical system with $\Omega \subseteq \mathbb{R}$ being an interval. Suppose there exists a finite partition $\mathcal{M}$ of Ω into intervals such that T is monotone and continuous on each of those intervals. Then

$$\overline{\mathrm{PE}}(T, q) \leq \lim_{n\to\infty} \frac{1}{n} \log \#\{M \in \mathcal{M}^{(n)} \mid M \neq \emptyset\}$$

holds true for all $q \geq 0$.

Proof. Let $\mathcal{M}$ be the partition into intervals on which T is monotone and continuous. It was shown in [4] that

$$\lim_{n\to\infty} \frac{1}{n} \log \#\{P_\pi \in OP(n) \mid P_\pi \neq \emptyset\} = \lim_{n\to\infty} \frac{1}{n} \log \#\{M \in \mathcal{M}^{(n)} \mid M \neq \emptyset\}$$

holds true for the here considered class of functions. Therefore,

$$\begin{aligned}\overline{\mathrm{PE}}(T,0) &= \limsup_{n\to\infty} \frac{1}{n} \log\left(\sum_{P_\pi \in OP(n)} \mu(P_\pi)^0\right) \\ &= \limsup_{n\to\infty} \frac{1}{n} \log \#\{P_\pi \in OP(n) \mid \mu(P_\pi) > 0\} \\ &\leq \lim_{n\to\infty} \frac{1}{n} \log \#\{P_\pi \in OP(n) \mid P_\pi \neq \emptyset\}\end{aligned}$$

holds true. The quantity $\limsup_{n\to\infty} \frac{1}{n} \log \#\{P_\pi \in OP(n) \mid P_\pi \neq \emptyset\}$ could be considered a topological version of the permutation entropy. Using the fact that the Réniy entropy is monotonically decreasing in q (Lemma 2.17) then implies

$$\overline{\mathrm{PE}}(T,q) \leq \overline{\mathrm{PE}}(T,0) \leq \lim_{n\to\infty} \frac{1}{n} \log \#\{M \in \mathcal{M}^{(n)} \mid M \neq \emptyset\}$$

for all $q \geq 0$. □

Given a dynamical system satisfying the conditions in Theorem 3.13, the above theorem implies in particular

$$\sup_{P\in\mathbb{P}_o(\mathscr{B})} h(T,\mathcal{P},q) \overset{Thm.\ 3.10}{\leq} \overline{\mathrm{PE}}(T,q) \leq \lim_{n\to\infty} \frac{1}{n} \log \#\{M \in \mathcal{M}^{(n)} \mid M \neq \emptyset\} \leq \log(\#\mathcal{M}) < \infty$$

for all $q \in [0,1[$. In contrast, we have

$$\sup_{P\in\mathbb{P}(\mathscr{B})} h(T,\mathcal{P},q) = h(T,q) = \infty$$

for all $q \in [0,1[$ and aperiodic functions.

If T is continuous on all of Ω, it was shown by Misiurewicz and Szlenk that $\lim_{n\to\infty} \frac{1}{n} \log \#\{M \in \mathcal{M}^{(n)} \mid M \neq \emptyset\}$ is equal to the topological entropy of the dynamical system [22]. A measure μ is called *measure of maximal entropy* if the Kolmogorov-Sinai entropy $h_\mu(T)$ is equal to the topological entropy. The Lebesgue measure as an invariant measure for the tent map would be an example for such a measure. Many of the practical relevant measures are measures of maximal entropy. So if $(\Omega, \mathscr{B}, \mu, T)$ is an continuous system and μ is a measure of maximal entropy, then $\lim_{n\to\infty} \frac{1}{n} \log \#\{M \in \mathcal{M}^{(n)} \mid M \neq \emptyset\}$ is equal to the Kolmogorov-Sinai entropy. In this case, one can exactly determine the value of $\overline{\mathrm{PE}}(T,q)$ for $q \in [0,1]$ as shown in the following Corollary:

Corollary 3.14. Let $(\Omega, \mathscr{B}, \mu, T)$ be an ergodic measure-preserving dynamical system with $\Omega \subseteq \mathbb{R}$ being an interval. Suppose there exists a finite partition of Ω into intervals such that T is monotone and continuous on each of those intervals. If

$$\lim_{n\to\infty} \frac{1}{n} \log \#\{M \in \mathcal{M}^{(n)} \mid M \neq \emptyset\} = h(T),$$

then

$$\underline{\mathrm{PE}}(T,q) = \overline{\mathrm{PE}}(T,q) = \sup_{P\in\mathbb{P}_o(\mathscr{B})} h(T,\mathcal{P},q) = h(T)$$

holds true for all $q \in [0,1]$.

Proof. Fix $q \in [0,1]$. Theorem 3.13 provides

$$\overline{\mathrm{PE}}(T,q) \leq \lim_{n\to\infty} \frac{1}{n} \log \#\{M \in \mathcal{M}^{(n)} \mid M \neq \emptyset\} = h(T)$$

and Theorem 3.10 yields

$$\underline{\mathrm{PE}}(T,q) \geq \sup_{\mathcal{P}\in\mathbb{P}_o(\mathscr{B})} h(T,\mathcal{P},q).$$

Combining those statements with the fact that the Rényi entropy is monotonically decreasing in q (Lemma 2.17) results in

$$h(T) = \overline{\mathrm{PE}}(T,q) \geq \underline{\mathrm{PE}}(T,q) \sup_{\mathcal{P}\in\mathbb{P}_o(\mathscr{B})} h(T,\mathcal{P},q) \geq \sup_{\mathcal{P}\in\mathbb{P}_o(\mathscr{B})} h(T,\mathcal{P},1) = h(T). \qquad \square$$

3.4 K-S entropy for piecewise monotone functions

3.4.1 Overview

The aim of this section is to investigate if the Kolmogorov-Sinai is an upper bound for the permutation entropy entropy, i.e. if

$$\overline{\mathrm{PE}}(T) \leq h(T)$$

holds true. We will focus on the upper permutation entropy because the obvious inequality $\underline{\mathrm{PE}}(T) \leq \overline{\mathrm{PE}}(T)$ implies that the lower permutation entropy will be bounded by the Kolmogorov-Sinai entropy as well.

If we manage to show that $\overline{\mathrm{PE}}(T) \leq h(T)$ holds true, we can combine this with the inequality $\underline{\mathrm{PE}}(T) \geq h(T)$ obtained in the previous section to conclude

$$\underline{\mathrm{PE}}(T) = \overline{\mathrm{PE}}(T) = h(T).$$

Unlike in the previous section, we will restrict our considerations to the case $q = 1$. The reason for this becomes clear in Section 3.4.8 where we will give an example of a piecewise monotone function for which $\overline{\mathrm{PE}}(T,q) = h(T,q)$ does not hold true for $q \neq 1$.

Bandt, Pompe and G. Keller proved the following statement in 2002:

Theorem 3.15 ([4])**.** Let $(\Omega, \mathscr{B}, \mu, T)$ be a measure-preserving dynamical system for some interval $\Omega \subseteq \mathbb{R}$. If there exists a finite partition of Ω into intervals such that T is piecewise monotonically increasing or decreasing and continuous on each of these intervals, then

$$\overline{\mathrm{PE}}(T) = h(T) \tag{3.10}$$

holds true.

A map $T : \Omega \to \mathbb{R}$ is called monotone on an interval $M \subseteq \Omega$ if

$$\omega_1 \leq \omega_2 \quad \text{implies} \quad T(\omega_1) \leq T(\omega_2) \quad \text{for all } \omega_1, \omega_2 \in M$$

or

$$\omega_1 \leq \omega_2 \quad \text{implies} \quad T(\omega_1) \geq T(\omega_2) \quad \text{for all } \omega_1, \omega_2 \in M$$

holds true. Bandt and Pompe called a map T *piecewise monotone* if there exists a finite partition $\mathcal{M}$ of Ω into intervals such that T is monotone and continuous on each set $M \in \mathcal{M}$.

Here, we want to consider functions T that are piecewise monotone in a more general and probabilistic sense. Therefore, we give a different definition for piecewise monotony, which is similar to Definition 3.4 of ordered partitions.

Definition 3.16 (Piecewise monotone function)**.** We call a map $T : \Omega \to \Omega$ (*countably*) *piecewise monotone* (with regard to μ) if there exists a partition $\mathcal{M} = \{M_i\}_{i\in I} \in \mathbb{P}_o^c$ such that

$$\mu^2 \left((M_i \times M_i) \cap R \cap (T \times T)^{-1}(R)\right) \in \left\{0, \mu^2 \left((M_i \times M_i) \cap R\right)\right\} \tag{3.11}$$

holds true for all $i \in I$, where $R = \{(x, y) \in \mathbb{R}^2 \mid x < y\}$. We will call such a partition $\mathcal{M}$ *partition into monotony parts* or *partition into monotony sets* of T.

It was not known whether (3.10) is true for a more general case than piecewise monotony. In the following paragraph, we consider the well studied Gauss function as an example of a countably piecewise monotone map. While the value of Kolmogorov-Sinai entropy of this function can be determined analytically, the value of the permutation entropy was unknown so far.

In this section, we are able to show that the equality of K-S and permutation entropy still holds true if we omit the condition of continuity and if there exists a countable partition of the domain of definition into intervals such that the one-dimensional map is monotone on each of those intervals. Unlike in the paper of Bandt, Pompe and G. Keller [4], we do not require that this partition into intervals is finite. Since continuity and the finiteness of the partitions into monotony intervals was crucial to the approach used in [4], we are forced to apply different methods here, and use a more measure-theoretic approach.

Example 3 (Gauss function)**.** The map $T : [0, 1] \to [0, 1]$ with

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

is called Gauss function (see Figure 3.1). This map is measure-preserving with regard to the measure μ, which is defined by $\mu(A) = \frac{1}{\log 2} \int_A \frac{1}{1+x}\, dx$ for all $A \in \mathscr{B}$ [7]. The partition $\mathcal{M} = \{[\frac{1}{n+1}, \frac{1}{n}[\mid n \in \mathbb{N}\} \cup \{\{0\}\}$ of $[0, 1]$ is a partition into monotony intervals of T. The map T is countably piecewise monotone but not piecewise monotone. Therefore, we cannot use Theorem 3.15 to decide whether $\mathrm{PE}(T)$ and $h(T)$ are equal. However, we can use our new theorem below to show the equality as explained in Section 3.4.4.

We will show here that (3.10) is true for countably piecewise monotone maps T as well. Our main result can be formulated as follows:

Theorem 3.17 (Main result)**.** Let $(\Omega, \mathscr{B}, \mu)$ be a Standard probability space with $\Omega \subseteq \mathbb{R}$ and $T : \Omega \to \Omega$ a measure-preserving map. Suppose that T is countably piecewise monotone and $\mathcal{M}$ a countable partition into monotony parts of T with $H(\mathcal{M}) < \infty$. Then

$$\underline{\mathrm{PE}}(T) = \overline{\mathrm{PE}}(T) = h(T)$$

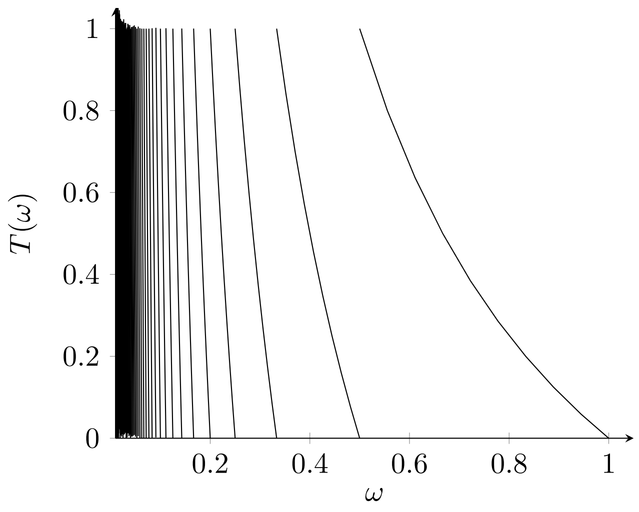


Figure 3.1: Graph of the Gauss function T.

holds true.

We have already given a proof of the above theorem in a paper together with K. Keller [12]. The proof given in this thesis is more comprehensive, especially the generalization to non-aperiodic systems. The different arguments used to proof the results are singled out more precisely. Additionally, the proof of the non-ergodic part presented in this dissertation uses completely different techniques compared to the proof in the before mentioned paper.

Theorem 3.17 is a consequence of the following more general theorem:

Theorem 3.18. Let $(\Omega, \mathscr{B}, \mu, T)$ be a measure-preserving dynamical system with $\Omega \subseteq \mathbb{R}$, and assume that the following conditions are satisfied:

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

$$\mathcal{M}^{(m)} \otimes \mathcal{M}^{(m)} \vee \{R, \Omega^2 \setminus R\} \prec \mathcal{M}^{(m)} \otimes \mathcal{M}^{(m)} \vee \bigvee_{u=1}^{m} (T \times T)^{-u} \left(\{R, \Omega^2 \setminus R\}\right). \tag{3.12}$$

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

$$\sum_{Q \in \mathcal{Q}} \limsup_{n \to \infty} \frac{1}{n} \sum_{l=1}^{n} \mu(Q \cap T^{-l}(Q)) < \varepsilon. \tag{3.13}$$

Then

$$\overline{\mathrm{PE}}(T) \leq h(T)$$

holds true.

As already mentioned, the above theorem is a generalization of the result of Bandt, Pompe and G. Keller (Theorem 3.15 here). Note that in the simpler case of piecewise monotony the restriction $H(\mathcal{M}) < \infty$ is not necessary because $H(\mathcal{M})$ is always finite for a finite partition $\mathcal{M}$ into monotony intervals.

To prove our results, we begin with Lemma 3.19 by reducing our problem to a combinatorial one. This is analogue to the approach used in [4]. While Bandt et al. followed this by an examination of periodic points, utilizing the piecewise monotony and continuity, we use the piecewise monotony more directly and then apply measure-theoretic arguments.

Rough outline of the proof We first start by proving Theorem 3.18. To achieve this, it will be shown in subsection 3.4.2 that the entropy difference $\overline{\mathrm{PE}}(T) - h(T, \mathcal{P})$ is bounded from above by a term depending on the number of intersections between sets of points with an ordinal pattern and the sets of the partition $\mathcal{P}^{(n)}$. This number of intersections depends on the chosen partition $\mathcal{P}$.

The partition $\mathcal{P}$ will be chosen as a combination of the partitions $\mathcal{M}$ and $\mathcal{Q}$ given in the conditions 1 and 2 of Theorem 3.18. Using condition 1, we show that for this choice of $\mathcal{P}$ the upper bound for the entropy difference is finite. Condition 2 allows us to create an upper bound that can be smaller than any $\varepsilon > 0$.

After proving Theorem 3.18, we will use this result to prove Theorem 3.17. We will first show, that condition 1 in Theorem 3.18 is satisfied for (countably) piecewise monotone functions. Then it will be shown, that condition 2 is satisfied for aperiodic and ergodic functions.

In subsection 3.4.5, ergodic decomposition will be used to generalize the above result for ergodic and aperiodic functions to non-ergodic and aperiodic functions.

Finally, in subsection 3.4.7, the assumption of aperiodicity will be removed.

3.4.2 Upper bound for entropy difference

Recall that we defined

$$\Delta(OP(n)|P(\boldsymbol{i})) = \{P_\pi \in OP(n) \mid \mu(P(\boldsymbol{i}) \cap P_\pi) > 0\}$$

for $\boldsymbol{i} \in I^n$ as the set of all permutations whose ordinal patterns are intersecting the set $P(\boldsymbol{i})$. Roughly speaking, if the size of $\Delta(OP(n)|P(\boldsymbol{i}))$ does not grow too fast on average for increasing n, then the partition $OP(n)$ does not add a lot of new information to $\mathcal{P}^{(n)}$, so $H(OP(n))$ and $H(\mathcal{P}^{(n)})$ will be similar in size.

We give an upper bound on the difference between $\overline{\mathrm{PE}}(T)$ and $h(T)$ based on $\#\Delta(OP(n)|P(\boldsymbol{i}))$.

Lemma 3.19. Let $(\Omega, \mathscr{B}, \mu, T)$ be a measure-preserving dynamical system with $\Omega \subseteq \mathbb{R}$ and $\mathcal{P} = \{P_i\}_{i\in I}$ a finite or countable partition of Ω with $H(\mathcal{P}) < \infty$. Then

$$\overline{\mathrm{PE}}(T) \leq h(T, \mathcal{P}) + \limsup_{n\to\infty} \frac{1}{n} \sum_{\boldsymbol{i}\in I^n} \mu(P(\boldsymbol{i})) \log(\#\Delta(OP(n)|P(\boldsymbol{i})))$$

and, for all $q > 1$,

$$\overline{\mathrm{PE}}(T, q) \leq h(T, \mathcal{P}, q) + \limsup_{n\to\infty} \frac{1}{n} \max_{\boldsymbol{i}\in I^n} \log(\#\Delta(OP(n)|P(\boldsymbol{i})))$$

hold true.

Proof. Let $\mathcal{P} = \{P_i\}_{i\in I}$ be a finite or countable partition of Ω with $H(\mathcal{P}) < \infty$. Lemma 3.3 provides

$$H(OP(n)|\mathcal{P}, q) \leq \begin{cases} \sum_{\boldsymbol{i}\in \mathcal{I}^n} \mu(P(\boldsymbol{i})) \cdot \log(\#\Delta(OP(n)|P(\boldsymbol{i}))) & \text{for } q = 1 \\ \max_{\boldsymbol{i}\in I^n} \log(\#\Delta(OP(n)|P(\boldsymbol{i}))) & \text{for } q > 1. \end{cases}$$

This implies

$$\begin{aligned} H(OP(n)) &\leq H(\mathcal{P}^{(n)} \vee OP(n)) = H(\mathcal{P}^{(n)}) + H(OP(n)|\mathcal{P}^{(n)}) \\ &= H(\mathcal{P}^{(n)}) + \sum_{\boldsymbol{i}\in I^n} \mu(P(\boldsymbol{i})) \log(\#S_n^{\mathcal{P}}(\boldsymbol{i})) \end{aligned}$$

and

$$H(OP(n), q) \leq H(\mathcal{P}^{(n)}, q) + \max_{\boldsymbol{i}\in I^n} \log(\#\Delta(OP(n)|P(\boldsymbol{i})))$$

for $q > 1$. Dividing both sides by n and taking the limit superior $n \to \infty$ finishes the proof. □

The above lemma allows to reduce the task of determining the difference between $H(OP(n))$ and $H(\mathcal{P}^{(n)})$ to determining the quantity $\Delta(OP(n)|P(\boldsymbol{i}))$ in dependence of $n \in \mathbb{N}$. Determining the latter can be done using combinatorial arguments and exploiting the monotony of T. This will be done in the following two lemmas 3.20 and 3.21.

Lemma 3.20. Let $(\Omega, \mathscr{B}, \mu, T)$ be a measure-preserving dynamical system with $\Omega \subseteq \mathbb{R}$. Suppose, there exists a finite or countable ordered partition $\mathcal{M} = \{M_i\}_{i\in I}$ of Ω and $m \in \mathbb{N}$ such that

$$\mathcal{M}^{(m)} \otimes \mathcal{M}^{(m)} \vee \{R, \mathbb{R}^2 \setminus R\} \;\prec\; \mathcal{M}^{(m)} \otimes \mathcal{M}^{(m)} \vee \bigvee_{u=1}^{m} (T \times T)^{-u}(\{R, \mathbb{R}^2 \setminus R\}) \tag{3.14}$$

holds true, where $R := \{(\omega_1, \omega_2) \in \Omega^2 : \omega_1 < \omega_2\}$ and $\mathcal{M}^{(m)} \otimes \mathcal{M}^{(m)} := \{M(\boldsymbol{i}) \times M(\boldsymbol{j}) : \boldsymbol{i}, \boldsymbol{j} \in I^n\}$. Then for all $n \in \mathbb{N}$ with $n \geq m$ and multi indices $\boldsymbol{i} = (i_0, i_1, \dots, i_{n-1}) \in I^n$

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

is true.

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

$$\begin{aligned} &\mathcal{M}^{(s)} \otimes \mathcal{M}^{(s)} \vee \bigvee_{t=0}^{s-1} (T \times T)^{-t}(\{R, \mathbb{R}^2 \setminus R\}) \\ &\prec \mathcal{M}^{(s)} \otimes \mathcal{M}^{(s)} \vee \bigvee_{t=s-m}^{s-1} (T \times T)^{-t}(\{R, \mathbb{R}^2 \setminus R\}) \end{aligned} \tag{3.15}$$

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

The above statement is trivial for $s = m$. Suppose, (3.15) holds true for some $s \in \mathbb{N}$ with $s \geq m$. We will now show that (3.15) holds for $s + 1$:

$$\begin{aligned}
&\mathcal{M}^{(s+1)} \otimes \mathcal{M}^{(s+1)} \vee \bigvee_{t=0}^{s} (T \times T)^{-t} \left(\{R, \mathbb{R}^2 \setminus R\}\right) \\
&= (T \times T)^{-1} \left(\mathcal{M}^{(s)} \otimes \mathcal{M}^{(s)} \vee \bigvee_{t=0}^{s-1} (T \times T)^{-t} \left(\{R, \mathbb{R}^2 \setminus R\}\right) \right) \\
&\quad \vee \mathcal{M}^{(m)} \otimes \mathcal{M}^{(m)} \vee \{R, \mathbb{R}^2 \setminus R\} \\
&\overset{(3.14)}{\prec} (T \times T)^{-1} \left(\mathcal{M}^{(s)} \otimes \mathcal{M}^{(s)} \vee \bigvee_{t=0}^{s-1} (T \times T)^{-t} \left(\{R, \mathbb{R}^2 \setminus R\}\right) \right) \\
&\quad \vee \mathcal{M}^{(m)} \otimes \mathcal{M}^{(m)} \vee \bigvee_{u=1}^{m} (T \times T)^{-u} \left(\{R, \mathbb{R}^2 \setminus R\}\right) \\
&= (T \times T)^{-1} \left(\mathcal{M}^{(s)} \otimes \mathcal{M}^{(s)} \vee \bigvee_{t=0}^{s-1} (T \times T)^{-t} \left(\{R, \mathbb{R}^2 \setminus R\}\right) \right) \vee \mathcal{M} \otimes \mathcal{M} \\
&\prec (T \times T)^{-1} \left(\mathcal{M}^{(s)} \otimes \mathcal{M}^{(s)} \vee \bigvee_{t=s-m}^{s-1} (T \times T)^{-t} \left(\{R, \mathbb{R}^2 \setminus R\}\right) \right) \vee \mathcal{M} \otimes \mathcal{M} \qquad (3.16) \\
&= \mathcal{M}^{(s+1)} \otimes \mathcal{M}^{(s+1)} \vee \bigvee_{t=s+1-m}^{s} (T \times T)^{-t} \left(\{R, \mathbb{R}^2 \setminus R\}\right),
\end{aligned}$$

In (3.16), the induction hypotheses was used.

Notice that

$$\begin{aligned}
\mathcal{M}^{(n)} &= \bigvee_{s=1}^{n} (\mathrm{id}, T)^{-n+s} \left(\mathcal{M}^{(s)} \otimes \mathcal{M}^{(s)} \right), \\
OP(n) &= \bigvee_{s=1}^{n} (\mathrm{id}, T)^{-n+s} \left(\bigvee_{t=0}^{s-1} (T \times T)^{-t} (\{R, \mathbb{R}^2 \setminus R\}) \right)
\end{aligned}$$

hold true (see (2.16)). This implies

$$\begin{aligned}
\mathcal{M}^{(n)} \vee OP(n) = &\bigvee_{s=1}^{n} (\mathrm{id}, T)^{-n+s} \left(\mathcal{M}^{(s)} \otimes \mathcal{M}^{(s)} \vee \bigvee_{t=0}^{s-1} (T \times T)^{-t} (\{R, \mathbb{R}^2 \setminus R\}) \right) \\
= &\bigvee_{s=1}^{m-1} (\mathrm{id}, T)^{-n+s} \left(\mathcal{M}^{(s)} \otimes \mathcal{M}^{(s)} \vee \bigvee_{t=0}^{s-1} (T \times T)^{-t} (\{R, \mathbb{R}^2 \setminus R\}) \right) \\
&\vee \bigvee_{s=m}^{n} (\mathrm{id}, T)^{-n+s} \left(\mathcal{M}^{(s)} \otimes \mathcal{M}^{(s)} \vee \bigvee_{t=0}^{s-1} (T \times T)^{-t} (\{R, \mathbb{R}^2 \setminus R\}) \right)
\end{aligned}$$

$$
\begin{aligned}
&\overset{(3.15)}{\prec} \bigvee_{s=1}^{m-1} (\mathrm{id}, T)^{-n+s} \left(\mathcal{M}^{(s)} \otimes \mathcal{M}^{(s)} \vee \bigvee_{t=0}^{s-1} (T \times T)^{-t}(\{R, \mathbb{R}^2 \setminus R\}) \right) \\
&\quad \vee \bigvee_{s=m}^{n} (\mathrm{id}, T)^{-n+s} \left(\mathcal{M}^{(s)} \otimes \mathcal{M}^{(s)} \vee \bigvee_{t=s-m}^{s-1} (T \times T)^{-t}(\{R, \mathbb{R}^2 \setminus R\}) \right) \\
&= \mathcal{M}^{(n)} \vee \bigvee_{s=0}^{n-1} \bigvee_{t=n-m}^{n-1} (T^s, T^t)^{-1}(\{R, \mathbb{R}^2 \setminus R\}).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\#\Delta(OP(n)|M(\boldsymbol{i})) &= \#\Delta(\mathcal{M}^{(n)} \vee OP(n)|M(\boldsymbol{i})) \\
&\leq \#\Delta\left(\mathcal{M}^{(n)} \vee \bigvee_{s=0}^{n-1} \bigvee_{t=n-m}^{n-1} (T^s, T^t)^{-1}(\{R, \mathbb{R}^2 \setminus R\}) | M(\boldsymbol{i}) \right) \\
&= \#\Delta\left(\bigvee_{s=0}^{n-1} \bigvee_{t=n-m}^{n-1} (T^s, T^t)^{-1}(\{R, \mathbb{R}^2 \setminus R\}) | M(\boldsymbol{i}) \right) \\
&\leq \prod_{s=0}^{n-1} \#\Delta\left(\bigvee_{t=n-m}^{n-1} (T^s, T^t)^{-1}(\{R, \mathbb{R}^2 \setminus R\}) | M(\boldsymbol{i}) \right) \\
&\leq \prod_{s=0}^{n-1} \prod_{t=n-m}^{n-1} \#\Delta\left((T^s, T^t)^{-1}(\{R, \mathbb{R}^2 \setminus R\}) | M(\boldsymbol{i}) \right) \qquad (3.17)
\end{aligned}
$$

holds true. Note that

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

For $s = t$, we have

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

If $i_s \neq i_t$ is true, $T^s(\omega)$ and $T^t(\omega)$ are located in different sets M_{i_s} and M_{i_t} for all $\omega \in M(\boldsymbol{i})$. Since μ-almost every point in the left interval is smaller than any point in the other interval,

$$M(\boldsymbol{i}) \vee (T^s, T^t)^{-1}(\{R, \mathbb{R}^2 \setminus R\}) = M(\boldsymbol{i})$$

holds true, which implies

$$
\begin{aligned}
&\#\Delta\left((T^s, T^t)^{-1}(\{R, \mathbb{R}^2 \setminus R\}) | M(\boldsymbol{i}) \right) \leq \#\Delta\left(M(\boldsymbol{i}) \vee (T^s, T^t)^{-1}(\{R, \mathbb{R}^2 \setminus R\}) | M(\boldsymbol{i}) \right) \\
&= \#\Delta\left(M(\boldsymbol{i}) | M(\boldsymbol{i}) \right) = 1.
\end{aligned}
$$

In all other cases, we have

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

The above considerations can be summarized as

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

In combination with (3.17), this provides

$$
\begin{aligned}
\#\Delta\left((T^s, T^t)^{-1}(\{R, \mathbb{R}^2 \setminus R\}) | M(\boldsymbol{i}) \right) &\leq 2^{\sum_{t=n-m}^{n-1} \#\{s \in \{0,1,\ldots,n-1 | i_s = i_t \text{ and } s \neq t\}} \\
&= 2^{\sum_{u=1}^{m} \#\{s \in \{0,1,\ldots,n-1 | i_s = i_{n-u} \text{ and } s \neq n-u\}}. \qquad \square
\end{aligned}
$$

In the following subsection we will show that the conditions of the previous lemma can be satisfied for $m = 1$ if the function T is countably piecewise monotone.

One could also state the previous Lemma for $m = 0$ by setting $\mathcal{M}^{(0)} := \{\Omega\}$ and $\bigvee_{t=1}^{0}(T \times T)^{-t}(\{R, \mathbb{R}^2 \setminus R\}) := \{\Omega\}$ as trivial partitions and $\bigcup_{l=1}^{0}\{s \in \{0, 1, ..., n-1\} \mid i_s = i_{n-l}$ and $s \neq n-l\} := \emptyset$. The conditions for $m = 0$ are satisfied if the function is monotone on Ω itself, since then $\{R, \mathbb{R}^2 \setminus R\} \prec \{\Omega\}$ holds true.

In subsection 3.4.6, we will give an example of a dynamical system where the requirements of Lemma 3.20 can not be fulfilled for $m \leq 1$ but for $m = 2$.

Roughly speaking, the number m determines for how far in the future the ordinal structure needs be observed to determine the ordinal structure in the present.

Proof of Theorem 3.18. Take $\varepsilon > 0$. According to Conditions 1 and 2, there exist finite or countably infinite ordered partitions $\mathcal{M} = \{M_i\}_{i \in I}$, $\mathcal{Q} = \{Q_j\}_{j \in J}$ and $m \in \mathbb{N}$ with $H(\mathcal{M}) < \infty$ and $H(\mathcal{Q}) < \infty$ satisfying (3.12) and (3.13). Consider the partition

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

Notice that $\mathcal{P}$ is again a finite or countably infinite ordered partition with $H(\mathcal{P}) < H(\mathcal{M}) + H(\mathcal{Q}) < \infty$. Using Lemma 3.19, this implies

$$\begin{aligned} \overline{\mathrm{PE}} &\leq h(T, \mathcal{P}) + \limsup_{n \to \infty} \frac{1}{n} \sum_{(\boldsymbol{i},\boldsymbol{j}) \in (I \times J)^n} \mu(P((\boldsymbol{i},\boldsymbol{j}))) \log(\#\Delta(OP(n)|P((\boldsymbol{i},\boldsymbol{j})))) \\ &\leq \mathrm{KS} + \limsup_{n \to \infty} \frac{1}{n} \sum_{(\boldsymbol{i},\boldsymbol{j}) \in (I \times J)^n} \mu(P((\boldsymbol{i},\boldsymbol{j}))) \log(\#\Delta(OP(n)|P((\boldsymbol{i},\boldsymbol{j})))), \end{aligned} \tag{3.18}$$

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

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

for all $(\boldsymbol{i},\boldsymbol{j}) = ((i_0, j_0), (i_1, j_1), \ldots, (i_{n-1}, j_{n-1})) \in (I \times J)^n$. Lemma 3.20 provides

$$\begin{aligned} &\sum_{(\boldsymbol{i},\boldsymbol{j}) \in (I \times J)^n} \mu(P((\boldsymbol{i},\boldsymbol{j}))) \cdot \log(\#\Delta(OP(n)|P((\boldsymbol{i},\boldsymbol{j})))) \\ &\leq \log 2 \sum_{(\boldsymbol{i},\boldsymbol{j}) \in (I \times J)^n} \mu(P((\boldsymbol{i},\boldsymbol{j}))) \left(\sum_{u=1}^{m} \#\{s \in \{0, 1, ..., n-1\} \mid (i_s, j_s) = (i_{n-u}, j_{n-u}) \text{ and } s \neq n-u\} \right) \\ &\leq \log 2 \sum_{(\boldsymbol{i},\boldsymbol{j}) \in (I \times J)^n} \mu(P((\boldsymbol{i},\boldsymbol{j}))) \left(\sum_{u=1}^{m} \#\{s \in \{0, 1, ..., n-1\} \mid j_s = j_{n-u} \text{ and } s \neq n-u\} \right) \\ &= \log 2 \sum_{u=1}^{m} \sum_{\boldsymbol{j} \in J^n} \left(\sum_{\boldsymbol{i} \in I^n} \mu(P((\boldsymbol{i},\boldsymbol{j}))) \right) \cdot \#\{s \in \{0, 1, ..., n-1\} \mid j_s = j_{n-u} \text{ and } s \neq n-u\} \\ &= \log 2 \sum_{u=1}^{m} \sum_{\boldsymbol{j} \in J^n} \mu(Q(\boldsymbol{j})) \cdot \#\{s \in \{0, 1, ..., n-1\} \mid j_s = j_{n-u} \text{ and } s \neq n-u\} \\ &\leq \log 2 \sum_{u=1}^{m} \sum_{\boldsymbol{j} \in J^n} \mu(Q(\boldsymbol{j})) \cdot (\#\{s \in \{0, 1, ..., n-u-1\} \mid j_s = j_{n-u}\} + u - 1) \\ &= \log 2 \cdot \left(m(m-1)/2 + \sum_{u=1}^{m} \sum_{\boldsymbol{j} \in J^n} \mu(Q(\boldsymbol{j})) \cdot \#\{s \in \{0, 1, ..., n-u-1\} \mid j_s = j_{n-u}\} \right) \end{aligned} \tag{3.19}$$

Combining (3.18) and (3.19) yields

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

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

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \sum_{\boldsymbol{j}\in J^n} \mu(Q(\boldsymbol{j})) \cdot \#\{s \in \{0,1,...,n-u-1\} \mid j_s = j_{n-u}\} \\
&= \limsup_{n\to\infty} \frac{1}{n} \sum_{j_{n-u}\in J} \sum_{\boldsymbol{j}\in J^{n-u}} \mu(Q((\boldsymbol{j}, j_{n-u}))) \cdot \#\{s \in \{0,1,...,n-u-1\} \mid j_s = j_{n-u}\} \\
&= \limsup_{n\to\infty} \frac{1}{n-u} \sum_{j_{n-u}\in J} \sum_{l=0}^{n-u} \mu(\{\omega \in T^{-n+u}(Q_{j_{n-u}}) \mid \#\{s \in \{0,1,...,n-u-1\} \mid T^s(\omega) \in Q_j\} = l\}) \cdot l \\
&= \sum_{Q\in\mathcal{Q}} \limsup_{n\to\infty} \frac{1}{n-u} \sum_{l=0}^{n-u} \mu(T^{-n+u}(Q) \cap T^{-l}(Q)) \\
&= \sum_{Q\in\mathcal{Q}} \limsup_{n\to\infty} \frac{1}{n-u} \sum_{l=0}^{n-u} \mu(Q \cap T^{-l}(Q)) \overset{(3.13)}{<} \varepsilon.
\end{aligned}$$

Hence,

$$\overline{\mathrm{PE}}(T) \leq h(T) + \log 2 \cdot m \cdot \varepsilon.$$

The statement of the theorem is then proven by choosing ε arbitrarily close to 0. □

3.4.3 Using monotony

In this subsection we show that condition 1 of Theorem 3.18 is satisfied for $m = 1$ if the function T is (countably) piecewise monotone.

Lemma 3.21. Let $T : \Omega \to \Omega$ be a countably piecewise monotone map on $\Omega \subseteq \mathbb{R}$ and $\mathcal{M} = \{M_i\}_{i\in I}$ a countable partition into monotony parts of T. Then (3.14) holds true for $m = 1$, that is

$$\mathcal{M} \otimes \mathcal{M} \vee \{R, \mathbb{R}^2 \setminus R\} \;\prec\; \mathcal{M} \otimes \mathcal{M} \vee (T \times T)^{-1}(\{R, \mathbb{R}^2 \setminus R\})$$

is true.

Proof. We will show that for all $i, j \in I$

$$M_i \times M_j \vee \{R, \mathbb{R}^2 \setminus R\} \;\prec\; M_i \times M_j \vee (T \times T)^{-1}(\{R, \mathbb{R}^2 \setminus R\})$$

holds true, which implies the statement of this lemma.

If $i \neq j$ holds true, ω_1 and ω_2 are located in different sets M_i and M_j for all $(\omega_1, \omega_2) \in M_i \times M_j$. In Figure 3.2, this corresponds to the fact that $M_i \times M_j$ either completely lies in the triangle R or in the triangle $\mathbb{R}^2 \setminus R$. Since μ-almost every point in the left interval is smaller than μ-almost every point in the other interval,

$$M_i \times M_j \vee \{R, \mathbb{R}^2 \setminus R\} \prec M_i \times M_j$$

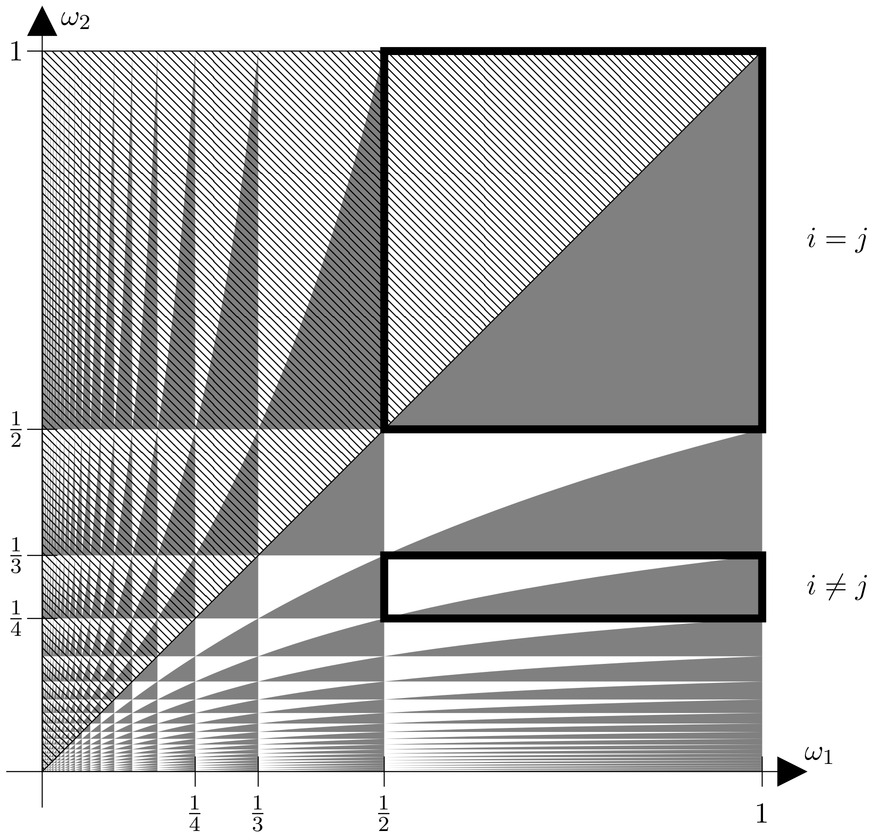


Figure 3.2: The striped area corresponds to the set $R = \{(\omega_1, \omega_2) \in \Omega^2 \mid \omega_1 < \omega_2\}$ and the gray area to $(T \times T)^{-1}(R)$ for the Gauss function T.

holds true, which yields

$$M_i \times M_j \vee \{R, \mathbb{R}^2 \setminus R\} \prec M_i \times M_j \prec M_i \times M_j \vee (T \times T)^{-1}(\{R, \mathbb{R}^2 \setminus R\}).$$

If $i = j$ holds true ω_1 and ω_2 are located in the same set $M_i = M_j$. Since the function T is monotone on M_i, the order relation between ω_1 and ω_2 is either the same as the order relation between $T(\omega_1)$ and $T(\omega_2)$ for all $(\omega_1, \omega_2) \in M_i \times M_j$ or the the same as the order relation between $T(\omega_2)$ and $T(\omega_1)$ for all $(\omega_1, \omega_2) \in M_i \times M_j$. This implies

$$M_i \times M_j \vee \{R, \mathbb{R}^2 \setminus R\} \prec M_i \times M_j \vee (T \times T)^{-1}(\{R, \mathbb{R}^2 \setminus R\}).$$

In terms of Figure 3.2, this means that in each square $M_i \times M_i$ the set R is equal to $(T \times T)^{-1}(R)$ if T is monotonically increasing in M_i and equal to $(T \times T)^{-1}(R^c)$ if T is monotonically decreasing in M_i. □

Lemma 3.22. Let $T : \Omega \to \Omega$ be a countably piecewise monotone map on $\Omega \subseteq \mathbb{R}$ and $\mathcal{M} = \{M_i\}_{i \in I}$ a countable partition into monotony parts of T. Then for all $n \in \mathbb{N}$ and multi indices $\boldsymbol{i} = (i_0, i_1, \ldots, i_{n-1}) \in I^n$

$$\#\Delta(OP(n)|M(\boldsymbol{i})) \leq 2^{\#\{s \in \{0,1,\ldots,n-2\} | i_s = i_{n-1}\}}$$

holds true.

Proof. This is an immediate consequence of Lemma 3.21 and Lemma 3.20 for $m = 1$. □

Corollary 3.23. Let $(\Omega, \mathscr{B}, \mu, T)$ be a measure-preserving dynamical system with $\Omega \subseteq \mathbb{R}$. Let T be countably piecewise monotone and $\mathcal{M}$ a countable partition into monotony parts of T. If $H(\mathcal{M}) < \infty$ holds true, then

$$\frac{1}{n} H(OP(n), q) \leq \frac{1}{n} H(\mathcal{M}^{(n)}, q) + \log 2$$

for all $q \geq 1$. In particular,

$$\overline{\mathrm{PE}}(T) \leq h(T, \mathcal{M}) + \log 2 \leq h(T) + \log 2.$$

holds true.

Proof. We have

$$\overline{\mathrm{PE}}(T) \leq h(T, \mathcal{M}) + \limsup_{n\to\infty} \frac{1}{n} \sum_{\boldsymbol{i}\in I^n} \mu(M(\boldsymbol{i})) \log(\#\Delta(OP(n)|M(\boldsymbol{i})))$$

according to Lemma 3.19. Lemma 3.22 provides

$$\#\Delta(OP(n)|M(\boldsymbol{i})) \leq 2^{\#\{s\in\{0,1,\ldots,n-2\}|i_s=i_{n-1}\}} \leq 2^{n-1}$$

for all $n \in \mathbb{N}$. Combining the above statements yields

$$\begin{aligned}\overline{\mathrm{PE}}(T) &\leq h(T, \mathcal{M}) + \limsup_{n\to\infty} \frac{1}{n} \sum_{\boldsymbol{i}\in I^n} \mu(M(\boldsymbol{i}))(n-1)\log 2 \\ &= h(T, \mathcal{M}) + \limsup_{n\to\infty} (n-1)/n \log 2 = h(T, \mathcal{M}) + \log 2 \\ &\leq h(T) + \log 2.\end{aligned}$$

□

Remark 3.24. The upper bound $\log 2$ on the difference between the entropies in corollary 3.23 does not depend on the map T, as long as T is countably piecewise monotone. So we have, in particular, $\overline{\mathrm{PE}}(T^m) \leq h(T^m) + \log 2$ for all $m \in \mathbb{N}$, which leads together with $\overline{\mathrm{PE}}(T^m) \geq h(T^m)$ to

$$h(T) = \lim_{m\to\infty} \frac{\overline{\mathrm{PE}}(T^m)}{m}. \tag{3.20}$$

However, it is, to our knowledge, not known whether $\lim_{m\to\infty} \overline{\mathrm{PE}}(T^m)/m = \overline{\mathrm{PE}}(T)$ holds true in general.

We will now show that the conditions in the above lemma can be satisfied if the considered dynamical system is ergodic. In a later subsection we will show that those conditions can be fulfilled for non-ergodic systems as well.

3.4.4 The ergodic case

Theorem 3.25. Let $(\Omega, \mathscr{B}, \mu, T)$ be a measure-preserving dynamical system with $\Omega \subseteq \mathbb{R}$. Let T be aperiodic, ergodic and countably piecewise monotone and $\mathcal{M}$ a countable partition into monotony parts of T. If $H(\mathcal{M}) < \infty$ holds true, then

$$\underline{\mathrm{PE}}(T) = \overline{\mathrm{PE}}(T) = h(T).$$

Proof. To proof this theorem, we need to show that the two conditions of Theorem 3.18 are satisfied. It follow from Lemma 3.21 that the first conditions is true for the given partition into monotony parts $\mathcal{M}$. So it remains to show that for any $\varepsilon > 0$ there exists a partition $\mathcal{Q}$ with

$$\sum_{Q \in \mathcal{Q}} \limsup_{n\to\infty} \frac{1}{n} \sum_{l=1}^{n} \mu(Q \cap T^{-l}(Q)) < \varepsilon \tag{3.21}$$

Define

$$Q_j^k := \begin{cases} [j \cdot 2^{-k}, (j+1) \cdot 2^{-k}[\cap \Omega & \text{if } j \in \{-4^k, \dots, 4^k - 1\} \\ ]-\infty, 2^{-k}[\cap \Omega & \text{if } j = -4^k - 1 \\ [2^k, \infty[\cap \Omega & \text{if } j = 4^k \end{cases}$$

for all $j \in \{-4^k - 1, 4^k, \dots, 4^k - 1, , 4^k\}$ and $k \in \mathbb{N}$ and set

$$J_k := \{j \in \{-4^k - 1, 4^k, \dots, 4^k - 1, 4^k\} \mid Q_j^k \neq \emptyset\}.$$

The collection of sets $\mathcal{Q}_k := \{Q_j^k\}_{j \in J_k}$ is a finite ordered partitions of Ω satisfying

$$Q_k \prec Q_{k+1} \tag{3.22}$$

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

$$\bigcap_{k=1}^{\infty} \mathcal{Q}_k(\omega) = \{\omega\}, \tag{3.23}$$

for all $\omega \in \Omega$, where $\mathcal{Q}_k(\omega)$ denotes the subset of $\mathcal{Q}_k$ that contains the element ω. According to Lemma 2.9, the ergodicity of T implies

$$\sum_{j \in J_k} \limsup_{n\to\infty} \frac{1}{n} \sum_{l=1}^{n} \mu(Q_j^k \cap T^{-l}(Q_j^k)) = \sum_{j \in J_k} \mu(Q_j^k)^2 = \int \mu(\mathcal{Q}_k(\omega))\, \mathrm{d}\omega$$

for all $k \in \mathbb{N}$. Since the aperiodicity of T implies

$$\lim_{k\to\infty} \mu(\mathcal{Q}_k(\omega)) \overset{(3.22)}{=} \mu\left(\bigcap_{k=1}^{\infty} \mathcal{Q}_k(\omega)\right) \overset{(3.23)}{=} \mu(\{\omega\}) = 0,$$

applying the dominated convergence theorem yields

$$\begin{aligned}\lim_{k\to\infty} \sum_{j \in J_k} \limsup_{n\to\infty} \frac{1}{n} \sum_{l=1}^{n} \mu(Q_j^k \cap T^{-l}(Q_j^k)) &= \lim_{k\to\infty} \int \mu(\mathcal{Q}_k(\omega))\, \mathrm{d}\omega \\ &= \int \lim_{k\to\infty} \mu(\mathcal{Q}_k(\omega))\, \mathrm{d}\omega = 0.\end{aligned}$$

Hence, for all $\varepsilon > 0$ there exists a partition $\mathcal{Q}_k$ that satisfies (3.21). □

Continuing example 3 One can show that the Gauss map T considered in example 3 is ergodic (see e.g. [8]). In order to apply Theorem 3.25 it remains to show that $H(\mathcal{M}) < \infty$ is true for $\mathcal{M} = \{[\frac{1}{n+1}, \frac{1}{n}[\mid n \in \mathbb{N}\} \cup \{\{0\}\}$:

Consider the map $\Phi :]0,\infty] \to]0,\infty]$ with $\Phi(x) = -x\log(x)$ for all $x > 0$. The map Φ is monotonically increasing on $]0, 1/\mathrm{e}[$. Choose $N \in \mathbb{N}$ such that

$$\frac{1}{n} - \frac{1}{n+1} = \frac{1}{n(n+1)} < \frac{\log 2}{\mathrm{e}}$$

is true for all $n \geq N$. Recall that the invariant measure μ of the Gauss function T is defined by $\mu([a, b[) = \frac{1}{\log 2}\int_a^b \frac{1}{1+x}\, dx$ for $0 \leq a < b \leq 1$. We have

$$\mu\left(\left[\frac{1}{n+1}, \frac{1}{n}\right[\right) = \frac{1}{\log 2}\int_{\frac{1}{n+1}}^{\frac{1}{n}} \frac{1}{1+x}\, dx \leq \frac{1}{\log 2}\int_{\frac{1}{n+1}}^{\frac{1}{n}} 1\, dx = \frac{1}{\log 2}\cdot\frac{1}{n(n+1)}$$

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

$$\begin{aligned}\Phi\left(\mu\left(\left[\frac{1}{n+1}, \frac{1}{n}\right[\right)\right) &\leq \Phi\left(\frac{1}{\log 2}\cdot\frac{1}{n(n+1)}\right)\\ &\leq \Phi\left(\frac{1}{\log 2}\cdot\frac{1}{n^2}\right) = \frac{\log\log 2 + 2\log n}{(\log 2)n^2}\end{aligned} \tag{3.24}$$

for all $n \geq N+1$. So we can conclude

$$\begin{aligned}H(\mathcal{M}) &= \sum_{n=1}^{\infty}\Phi\left(\mu\left(\left[\frac{1}{n+1}, \frac{1}{n}\right[\right)\right)\\ &= \sum_{n=1}^{N}\Phi\left(\mu\left(\left[\frac{1}{n+1}, \frac{1}{n}\right[\right)\right) + \sum_{n=N+1}^{\infty}\Phi\left(\mu\left(\left[\frac{1}{n+1}, \frac{1}{n}\right[\right)\right)\\ &\overset{(3.24)}{\leq} \sum_{n=1}^{N}\Phi\left(\mu\left(\left[\frac{1}{n+1}, \frac{1}{n}\right[\right)\right) + \sum_{n=N+1}^{\infty}\frac{\log\log 2 + 2\log n}{(\log 2)n^2}\\ &\leq \sum_{n=1}^{N}\Phi\left(\mu\left(\left[\frac{1}{n+1}, \frac{1}{n}\right[\right)\right) + \sum_{n=1}^{\infty}\frac{\log\log 2}{(\log 2)n^2} + \sum_{n=1}^{\infty}\frac{2\log n}{(\log 2)n^2} < \infty,\end{aligned}$$

which allows us to use Theorem 3.25 to conclude that $\underline{\mathrm{PE}}(T) = \overline{\mathrm{PE}}(T) = h(T)$ holds true.

It remains to show that the second condition (3.13) in Theorem 3.18 holds true. As shown in the previous section, this is true if the system is ergodic. But as we will show with Lemma 3.37 in the next section, ergodicity is not even required for (3.13) to hold true.

3.4.5 Using ergodic decomposition

To generalize the results of the ergodic case in section 3.4.4 to non-ergodic systems, we use the so called *ergodic decomposition*. The idea is to use the fact that, under some technical conditions, every T-invariant measure can be written as a convex combination of ergodic measures.

To be able to write an invariant measure as a convex combination of ergodic invariant measures, it is necessary to guarantee that a convex combination of invariant measures is, in fact, an invariant measure again. This can be easily seen:

Let μ_1, μ_2 be two T-invariant measures and $\alpha \in [0,1]$. The sum $\mu := \mu_1 + \mu_2$ is a probability measure and for all $A \in \mathscr{A}$ we have

$$\begin{aligned} &(\alpha\mu_1 + (1-\alpha)\mu_2)(T^{-1}(A)) = \alpha\mu_1(T^{-1}(A)) + (1-\alpha)\mu_2(T^{-1}(A)) \\ &= \ \alpha\mu_1(A) + (1-\alpha)\mu_2(A) = (\alpha\mu_1 + (1-\alpha)\mu_2)(A), \end{aligned}$$

hence $\alpha\mu_1 + (1-\alpha)\mu_2$ is T-invariant as well.

Conditional expectation One way to obtain a convex combination of measures is by using conditional expectation. We will follow here the approach used in [9] and extend the results of interest to the permutation entropy.

Definition 3.26 (Conditional expectation)**.** Let $(\Omega, \mathscr{B}, \mu)$ be a Borel probability space, $\mathscr{A} \subseteq \mathscr{B}$ a sub-σ-algebra and $f : \Omega \to \mathbb{R}$ a $\mathscr{B}$-measurable and integrable function. Then the conditional expectation $E(f|\mathscr{A})$ of f with respect to $\mathscr{A}$ is (μ-almost everywhere uniquely) defined by the properties

(i) $E(f|\mathscr{A})$ is a $\mathscr{A}$-measurable real-valued function defined on Ω,

(ii) $\int_A E(f|\mathscr{A})\,\mathrm{d}\mu = \int_A f\,\mathrm{d}\mu$ for all $A \in \mathscr{A}$.

Property (i) and (ii) characterize the conditional expectation $E(f|\mathscr{A})(\omega)$, roughly speaking, as an average that is dependent on the element of the σ-algebra $\mathscr{A}$ in which ω is located. This 'dependent averaging' with respect to some ω and selected sets in $\mathscr{A}$ can be expressed as an integral with respect to specific measures $\mu_\omega^{\mathscr{A}}$:

Theorem 3.27 (Conditional measures [8])**.** Let $(\Omega, \mathscr{B}, \mu)$ be a Standard probability space, $\mathscr{A} \subseteq \mathscr{B}$ a sub-σ-algebra and $f : \Omega \to \mathbb{R}$ a $\mathscr{B}$-measurable and integrable function. Then there exists a set $\Omega_0 \in \mathscr{B}$ with $\mu(\Omega_0) = 1$ and a collection of measures $\{\mu_\omega^{\mathscr{A}} : \omega \in \Omega_0\}$ on $(\Omega, \mathscr{B})$ such that

$$E(f|\mathscr{A})(\omega) = \int_{\Omega_0} f(\nu)\,\mathrm{d}\mu_\omega^{\mathscr{A}}(\nu)$$

for all $\mathscr{B}$-measurable and integrable functions $f : \Omega \to \mathbb{R}$. The measures $\mu_\omega^{\mathscr{A}}$ are called conditional measures and are μ-almost surely unique.

Combining property (ii) of the conditional expectation and Theorem 3.27 yields

$$\begin{aligned} \mu(A) &= \int_\Omega 1_A(\omega)\,\mathrm{d}\mu(\omega) = \int_\Omega E(1_A|\mathscr{A})(\omega)\,\mathrm{d}\mu(\omega) \\ &= \int_\Omega \left(\int_{\Omega_0} 1_A(\nu)\,\mathrm{d}\mu_\omega^{\mathscr{A}}(\nu)\right)\mathrm{d}\mu(\omega) = \int_\Omega \mu_\omega^{\mathscr{A}}(A)\,\mathrm{d}\mu(\omega), \end{aligned}$$

for all σ-algebras $\mathscr{A} \subseteq \mathscr{B}$ and $A \in \mathscr{B}$, where 1_A is the indicator function of A. So, conditioning on a σ-algebra $\mathscr{A}$ generates a decomposition of the measure μ into components $\mu_\omega^{\mathscr{A}}$. We write

$$\mu = \int_\Omega \mu_\omega^{\mathscr{A}}\,\mathrm{d}\mu(\omega)$$

to abbreviate this fact.

We are interested in a decomposition of the measure μ such that the components $\mu_\omega^{\mathscr{A}}$ are ergodic with regard to T. To achieve this, we need to choose the σ-algebra $\mathscr{A}$ accordingly:

Consider

$$\mathscr{E} := \{E \in \mathscr{B} \mid T^{-1}(E) = E\}. \tag{3.25}$$

This collection of sets is a σ-algebra:

We have $T^{-1}(\Omega) = \Omega$, so $\Omega \in \mathscr{E}$. Additionally,

$$T^{-1}(\Omega \setminus E) = T^{-1}(\Omega) \setminus T^{-1}(E) = \Omega \setminus E$$

holds true for all $E \in \mathscr{E}$, so $\Omega \setminus E \in \mathscr{E}$. Finally, for all $E_1, E_2, \ldots \in \mathscr{E}$

$$T^{-1}\left(\bigcup_i E_i\right) = \bigcup_i T^{-1}(E_i) = \bigcup_i E_i,$$

is fulfilled, which implies $\bigcup_i E_i \in \mathscr{E}$.

By choosing $\mathscr{A} = \mathscr{E}$, the decomposition

$$\mu = \int_\Omega \mu_\omega^{\mathscr{E}} \, \mathrm{d}\mu(\omega)$$

is exactly the ergodic decomposition of the measure μ according to the following theorem:

Theorem 3.28 (Ergodic decomposition of invariant measure [8])**.** Let $(\Omega, \mathscr{B}, \mu)$ be a Standard probability space. Let

$$\mu = \int \mu_\omega^{\mathscr{E}} \, \mathrm{d}\mu(\omega)$$

be a decomposition of μ with $\mathscr{E}$ defined as in (3.25) and $\mu_\omega^{\mathscr{E}}$ given by Theorem 3.27. Then T is measure-preserving and ergodic with regard to $\mu_\omega^{\mathscr{E}}$ for μ-almost every $\omega \in \Omega$.

The proof of this Theorem is relatively technical. For example, it is necessary to use the fact that each sub-σ-algebra of a Standard probability space can be considered to be generated by a countable collection of subsets of Ω.

3.4.5.1 Convex combination of entropy

If μ is a convex combination of finitely many measures, the same is asymptotically true for the entropy with regard to these measures, as shown in the following lemma:

Lemma 3.29 (Entropy of convex combination)**.** Let $(\Omega, \mathscr{A}, \mu, T)$ be a measure-preserving dynamical system and $\alpha \in (0, 1)$. Let $\mu = \alpha\mu_1 + (1-\alpha)\mu_2$ where μ_1 and μ_2 are probability measures on $(\Omega, \mathscr{A})$. Then

$$\alpha \cdot H_{\mu_1}(\mathcal{P}) + (1-\alpha) \cdot H_{\mu_2}(\mathcal{P}) \leq H_\mu(\mathcal{P})$$

and

$$H_\mu(\mathcal{P}) \leq \alpha \cdot H_{\mu_1}(\mathcal{P}) + (1-\alpha) \cdot H_{\mu_2}(\mathcal{P}) - \alpha \log(\alpha) - (1-\alpha) \log(1-\alpha)$$

hold true for all finite or countable partitions $\mathcal{P} \in \mathbb{P}^c(\mathscr{A})$. In particular, if μ_1 and μ_2 are invariant with regard to T,

$$h_\mu(T,\mathcal{P}) = \alpha \cdot h_{\mu_1}(T,\mathcal{P}) + (1-\alpha) \cdot h_{\mu_2}(T,\mathcal{P})$$

holds true. If, additionally, $\underline{\mathrm{PE}}_{\mu_1}(T) = \overline{\mathrm{PE}}_{\mu_1}(T)$ and $\underline{\mathrm{PE}}_{\mu_2}(T) = \overline{\mathrm{PE}}_{\mu_2}(T)$ holds true, then

$$\overline{\mathrm{PE}}_\mu(T,\mathcal{P}) = \alpha \cdot \overline{\mathrm{PE}}_{\mu_1}(T,\mathcal{P}) + (1-\alpha) \cdot \overline{\mathrm{PE}}_{\mu_2}(T,\mathcal{P}).$$

Proof. Let $\mathcal{P} \in \mathbb{P}^{\mathrm{C}}(\mathscr{A})$ be a finite or countable partition of Ω and $\alpha \in (0,1)$. We will show the first inequality:

Consider $f : (0,\infty) \to \mathbb{R}$ with $f(x) = -x \log x$. The function f is concave because $f''(x) = -\frac{1}{x} < 0$ for all $x > 0$. This implies

$$\begin{aligned}
H_\mu(\mathcal{P}) &= -\sum_{P\in\mathcal{P}} \mu(P)\log\mu(P) = \sum_{P\in\mathcal{P}} f(\alpha\mu_1(P) + (1-\alpha)\mu_2(P)) \\
&\geq \sum_{P\in\mathcal{P}} \alpha f(\mu_1(P)) + (1-\alpha) f(\mu_2(P)) \\
&= -\alpha \sum_{P\in\mathcal{P}} \mu_1(P)\log(\mu_1(P)) - (1-\alpha)\sum_{P\in\mathcal{P}} \mu_2(P)\log(\mu_2(P)) \\
&= \alpha H_{\mu_1}(\mathcal{P}) + (1-\alpha) H_{\mu_2}(\mathcal{P}).
\end{aligned}$$

We will now show the second inequality:

For all $P \in \mathcal{P}$ we have

$$\mu(P) = \alpha\mu_1(P) + (1-\alpha)\mu_2(P) \geq \alpha\mu_1(P).$$

Using the monotony of the logarithm provides

$$-\log(\mu(P)) \leq -\log(\alpha\mu_1(P))$$

and, analogously,

$$-\log(\mu(P)) \leq -\log((1-\alpha)\mu_2(P)).$$

Thus,

$$\begin{aligned}
H_\mu(\mathcal{P}) &= -\sum_{P\in\mathcal{P}} \mu(P)\log\mu(P) \\
&= -\sum_{P\in\mathcal{P}} (\alpha\mu_1(P) + (1-\alpha)\mu_2(P))\log\mu(P) \\
&= -\alpha\sum_{P\in\mathcal{P}} \mu_1(P)\log\mu(P) - (1-\alpha)\sum_{P\in\mathcal{P}^{(n)}} \mu_2(P)\log\mu(P) \\
&\leq -\alpha\sum_{P\in\mathcal{P}} \mu_1(P)\log(\alpha\mu_1(P)) - (1-\alpha)\sum_{P\in\mathcal{P}} \mu_2(P)\log((1-\alpha)\mu_2(P)) \\
&= -\alpha\sum_{P\in\mathcal{P}} \mu_1(P)\log(\mu_1(P)) - (1-\alpha)\sum_{P\in\mathcal{P}} \mu_2(P)\log(\mu_2(P)) \\
&\quad -\alpha\sum_{P\in\mathcal{P}} \mu_1(P)\log(\alpha) - (1-\alpha)\sum_{P\in\mathcal{P}} \mu_2(P)\log(1-\alpha) \\
&= \alpha H_{\mu_1}(\mathcal{P}) + (1-\alpha)H_{\mu_2}(\mathcal{P}) - \alpha\log(\alpha) - (1-\alpha)\log(1-\alpha).
\end{aligned}$$

Now consider a sequence of partitions $(\mathcal{P}_n)_{n\in\mathbb{N}} \subseteq \mathbb{P}_{\mathrm{C}}(\mathscr{A})$. Using the above inequalities yields on the one hand

$$\lim_{n\to\infty} \frac{1}{n} H_\mu(\mathcal{P}_n) \geq \alpha \lim_{n\to\infty} \frac{1}{n} H_{\mu_1}(\mathcal{P}_n) + (1-\alpha) \lim_{n\to\infty} \frac{1}{n} H_{\mu_2}(\mathcal{P}_n)$$

and on the other hand

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} H_\mu(\mathcal{P}_n) &\leq \lim_{n\to\infty} \frac{1}{n} \left[\alpha H_{\mu_1}(\mathcal{P}_n) + (1-\alpha) H_{\mu_2}(\mathcal{P}_n) - \alpha\log(\alpha) - (1-\alpha)\log(1-\alpha)\right] \\ &= \alpha \lim_{n\to\infty} \frac{1}{n} H_{\mu_1}(\mathcal{P}_n) + (1-\alpha) \lim_{n\to\infty} \frac{1}{n} H_{\mu_2}(\mathcal{P}_n),\end{aligned}$$

assuming all considered limits exist. This implies

$$\lim_{n\to\infty} \frac{1}{n} H_\mu(\mathcal{P}_n) = \alpha \lim_{n\to\infty} \frac{1}{n} H_{\mu_1}(\mathcal{P}_n) + (1-\alpha) \lim_{n\to\infty} \frac{1}{n} H_{\mu_2}(\mathcal{P}_n).$$

Choosing $\mathcal{P}_n$ as $\bigvee_{i=0}^{n-1} T^{-i}(\mathcal{P})$ for a fixed partition $\mathcal{P} \in \mathbb{P}^{\mathrm{C}}$ or as $OP(n)$, respectively, finishes the proof. □

The above Lemma can be inductively modified to the case where the measure μ is a combination of finitely many measures. However, we are interested in the ergodic decomposition of μ, which can be a convex combination of uncountably many measures. Therefore, a different approach is necessary.

This approach will make use of the concept of conditional measures introduced in Theorem 3.27. The conditional measures provide a way to generalize the definition of the conditional entropy in a way that allows for conditioning on arbitrary sub-σ-algebras.

Definition 3.30 (Shannon entropy conditioned on σ-algebra)**.** Let $(\Omega, \mathscr{B}, \mu)$ be a Standard probability space, $\mathscr{A}, \mathscr{C} \subseteq \mathscr{B}$ sub-σ-algebras and $\mathcal{P}$ a finite or countable partition of Ω

$$H_\mu(\mathcal{P}|\mathscr{A}) := \int_\Omega H_{\mu_\omega^{\mathscr{A}}}(\mathcal{P})\,\mathrm{d}\mu(\omega),$$

defines the entropy of $\mathcal{P}$ conditioned on the σ-algebra $\mathscr{A}$, where $\mu_\omega^{\mathscr{A}}$ is given as in Theorem 3.27. If the choice of the considered measure μ is obvious from the context, we simply write $H(\mathcal{P}|\mathscr{A})$ instead of $H_\mu(\mathcal{P}|\mathscr{A})$.

If $\mathscr{A} = \sigma(\mathcal{Q})$ for a finite or countable partition $\mathcal{Q}$ with $H_\mu(\mathcal{Q}) < \infty$, then

$$H_\mu(\mathcal{P}|\mathscr{A}) = H_\mu(\mathcal{P}|\mathcal{Q})$$

corresponds to the entropy conditioned on a partition of Ω as defined in 3.2.

The properties of the Shannon entropy conditioned on a σ-algebra are analogue to the properties of the Shannon entropy conditioned on a partition:

Lemma 3.31 (Properties of conditional entropy [9])**.** Let $(\Omega, \mathscr{A}, \mu)$ be a Standard probability space, $\mathcal{P}$ and $\mathcal{Q}$ finite or countable partitions of Ω and $\mathscr{C} \subseteq \mathscr{A}$ a σ-algebra. Then

$$H(\mathcal{P}|\mathscr{C}) \leq H(\mathcal{P})$$

and

$$H(\mathcal{P} \vee \mathcal{Q}|\mathscr{C}) = H(\mathcal{P}|\mathcal{Q} \vee \mathscr{C}) + H(\mathcal{Q}|\mathscr{C})$$

holds true. Let $(\mathcal{Q}_k)_{k\in\mathbb{N}}$ a sequence of partitions of Ω with $\mathcal{Q}_k \prec \mathcal{Q}_{k+1}$ for all $k \in \mathbb{N}$. Then

$$\lim_{k\to\infty} H(\mathcal{P}|\mathscr{C} \vee \mathcal{Q}_k) = H(\mathcal{P}|\mathscr{C} \vee \bigvee_{k=1}^{\infty} \mathcal{Q}_k)$$

If $T^{-1}(\mathscr{C}) = \mathscr{C}$, then

$$\lim_{n\to\infty} \frac{1}{n} H(\mathcal{P}^{(n)}|\mathscr{C}) = H(\mathcal{P}|\mathscr{C} \vee \bigvee_{t=1}^{\infty} T^{-i}(\mathcal{P})) \tag{3.26}$$

holds true.

Property (3.26) is sometimes called "future formula" because in $H(\mathcal{P}|\mathscr{A} \vee \bigvee_{t=1}^{\infty} T^{-t}(\mathcal{P}))$, the term $\bigvee_{t=1}^{\infty} T^{-i}(\mathcal{P})$ can be interpreted as the information about $\mathcal{P}$ starting at $t = 1$, so for the future, and $\mathcal{P} = T^{-0}(\mathcal{P})$ describes the information about $\mathcal{P}$ for $t = 0$, so in the present.

There exist some properties of measure-preserving dynamical systems $(\Omega, \mathscr{B}, \mu, T)$ that remain true μ-almost surely for every subsystem $(\Omega, \mathscr{B}, T, \mu_\omega^{\mathscr{A}})$, where $\mu = \int \mu_\omega^{\mathscr{A}}$ is a decomposition of μ:

Lemma 3.32. Let $(\Omega, \mathscr{B}, \mu)$ be a Standard probability space and

$$\mu = \int \mu_\omega^{\mathscr{A}} \, \mathrm{d}\mu(\omega)$$

a decomposition of μ for a σ-algebra $\mathscr{A} \subseteq \mathscr{B}$. If T is μ-almost surely aperiodic and there exists a finite or countable partition $\mathcal{P}$ of Ω with $H_\mu(\mathcal{P}) < \infty$, then there exists a set $\Omega_0 \in \mathscr{B}$ with $\mu(\Omega_0) = 1$ such that

(i) T is $\mu_\omega^{\mathscr{A}}$-almost surely aperiodic and

(ii) $H_{\mu_\omega^{\mathscr{A}}}(\mathcal{P}) < \infty$ holds true

for all $\omega \in \Omega_0$.

Proof. Consider the set

$$\Pi := \bigcup_{n=1}^{\infty} \{\omega' \in \Omega : T^n(\omega') = \omega'\}$$

of all periodic points. Since T is μ-almost surely aperiodic, we have $\mu(\Pi) = 0$. Set

$$\Omega_1 := \{\omega \in \Omega \mid \mu_\omega^{\mathscr{A}}(\Pi) = 0\}.$$

Suppose, $\mu(\Omega_1) < 1$ holds true. Then there exist $\varepsilon > 0$ and a set $A \subseteq \Omega \setminus \Omega_1$ with $\mu(A) > 0$ and $\mu_\omega^{\mathscr{A}}(\Pi) > \varepsilon$ for all $\omega \in A$. This would imply

$$\mu(\Pi) = \int \mu_\omega^{\mathscr{A}}(\Pi) \, \mathrm{d}\mu(\omega) \geq \int_A \mu_\omega^{\mathscr{A}}(\Pi) \, \mathrm{d}\mu(\omega) \geq \varepsilon \cdot \mu(A) > 0,$$

which is a contradiction to $\mu(\Pi) = 0$. So $\mu(\Omega_1) = 1$ must be true. Now set

$$\Omega_2 := \{\omega \in \Omega \mid H_{\mu_\omega^{\mathscr{A}}}(\mathcal{M}) < \infty\}.$$

Suppose, $\mu(\Omega_2) < 1$ holds true. Using Lemma 3.31, this would imply

$$H_\mu(\mathcal{M}) \geq H_\mu(\mathcal{M}|\mathscr{A}) = \int_\Omega H_{\mu_\omega^{\mathscr{A}}}(\mathcal{M})\,\mathrm{d}\mu(\omega) \geq \int_{\Omega\setminus\Omega_2} H_{\mu_\omega^{\mathscr{A}}}(\mathcal{M})\,\mathrm{d}\mu(\omega) = \infty,$$

which is a contradiction to the assumption $H_\mu(\mathcal{M}) < \infty$. So we must have $\mu(\Omega_2) = 1$. Set

$$\Omega_0 := \Omega_1 \cap \Omega_2$$

which has measure 1 according to the above arguments. □

The next Lemma can be found in [9] but with the additional assumption that the system is invertible.

Lemma 3.33. Let $(\Omega, \mathscr{B}, \mu)$ be a Standard probability space and $T : \Omega \to \Omega$ a measure-preserving map. Then for any finite partition $\mathcal{P}$

$$h(T, \mathcal{P}) = \lim_{n\to\infty} \frac{1}{n} H(\mathcal{P}^{(n)}|\mathscr{E})$$

holds true.

Proof. Let $(\mathcal{Q}_k)_{k\in\mathbb{N}}$ be a sequence of finite partitions of Ω with $\mathcal{Q}_k \subseteq \mathscr{E}$ and $\mathcal{Q}_k \prec \mathcal{Q}_{k+1}$ for all $k \in \mathbb{N}$ and $\sigma\left(\bigvee_{k=1}^{\infty} \mathcal{Q}_k\right) = \mathscr{E}$. Notice that, according to the definition of $\mathscr{E}$,

$$T^{-i}(\mathcal{Q}_k) = \mathcal{Q}_k$$

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

Take $\varepsilon > 0$ and any finite partition $\mathcal{P}$ of Ω. Lemma 3.31 provides

$$\lim_{n\to\infty} \frac{1}{n} H(\mathcal{P}^{(n)}|\mathscr{E}) = H(\mathcal{P}|\bigvee_{i=1}^{\infty} T^{-i}(\mathcal{P}) \vee \mathscr{E}) = \lim_{k\to\infty} H(\mathcal{P}|\bigvee_{i=1}^{\infty} T^{-i}(\mathcal{P}) \vee \mathcal{Q}_k),$$

so there exists $k_0 \in \mathbb{N}$ with

$$H\left(\mathcal{P}|\bigvee_{i=1}^{\infty} T^{-i}(\mathcal{P}) \vee \mathcal{Q}_{k_0}\right) \leq H\left(\mathcal{P}|\bigvee_{i=1}^{\infty} T^{-i}(\mathcal{P}) \vee \mathscr{E}\right) + \varepsilon.$$

Now define $\mathcal{P}_0 := \mathcal{P} \vee \mathcal{Q}_{k_0}$. Then

$$\begin{aligned}
h(T,\mathcal{P}) \leq& h(T,\mathcal{P}_0) = H(\mathcal{P}_0|\bigvee_{i=1}^{\infty} T^{-i}(\mathcal{P}_0)) = H(\mathcal{P} \vee \mathcal{Q}_{k_0}|\bigvee_{i=1}^{\infty} T^{-i}(\mathcal{P}) \vee \bigvee_{i=1}^{\infty} T^{-i}(\mathcal{Q}_{k_0}))\\
=& H(\mathcal{P} \vee \mathcal{Q}_{k_0}|\bigvee_{i=1}^{\infty} T^{-i}(\mathcal{P}) \vee \mathcal{Q}_{k_0})\\
\leq& H(\mathcal{P}|\bigvee_{i=1}^{\infty} T^{-i}(\mathcal{P}) \vee \mathcal{Q}_{k_0}) + H(\mathcal{Q}_{k_0}|\bigvee_{i=1}^{\infty} T^{-i}(\mathcal{P}) \vee \mathcal{Q}_{k_0})\\
=& H(\mathcal{P}|\bigvee_{i=1}^{\infty} T^{-i}(\mathcal{P}) \vee \mathcal{Q}_{k_0}) \leq H(\mathcal{P}|\bigvee_{i=1}^{\infty} T^{-i}(\mathcal{P}) \vee \mathscr{E}) + \varepsilon\\
=& \lim_{n\to\infty} \frac{1}{n} H(\mathcal{P}^{(n)}|\mathscr{E}) + \varepsilon.
\end{aligned}$$

The first statement of this Lemma follows from choosing ε arbitrarily small. □

The following theorem is a consequence of the above lemma.

Theorem 3.34 (Ergodic decomposition of Kolmogorov-Sinai entropy [9])**.** Let $(\Omega, \mathscr{B}, \mu)$ be a Standard probability space and $T : \Omega \to \Omega$ a measure-preserving map. Then

$$h_\mu(T) = \int_\Omega h_{\mu_\omega^{\mathscr{E}}}(T)\,\mathrm{d}\mu(\omega)$$

holds true, with $\mathscr{E}$ defined as in (3.25).

This statement can not easily be extended to the permutation entropy. This is mainly due to the fact that it is unknown whether the future formula (3.26) holds true when using ordinal patterns instead of symbolic patterns.

However, in the special case of piecewise monotone maps, we can prove a statement similar to the one given by Lemma 3.33:

Lemma 3.35. Let $(\Omega, \mathscr{B}, \mu)$ be a Standard probability space with $\Omega \subseteq \mathbb{R}$ and $T : \Omega \to \Omega$ an aperiodic measure-preserving map. Suppose that T is countably piecewise monotone and $\mathcal{M}$ a countable partition into monotony parts of T with $H(\mathcal{M}) < \infty$. Then

$$\lim_{n\to\infty} \frac{1}{n} H_\mu(OP(n)|\mathscr{E}) = \int h_{\mu_\omega^{\mathscr{E}}}(T)\,\mathrm{d}\mu(\omega)$$

holds true.

Proof. According to Theorem 3.28 and Lemma 3.32, there exists a set $\Omega_0 \in \mathscr{B}$ with $\mu(\Omega_0) = 1$ such that T is measure-preserving, aperiodic and ergodic with respect to $\mu_\omega^{\mathscr{E}}$ and $H_{\mu_\omega^{\mathscr{E}}}(\mathcal{M}) < \infty$ holds true for all $\omega \in \Omega_0$.

Set

$$f_n(\omega) := \frac{1}{n} H_{\mu_\omega^{\mathscr{E}}}(OP(n))$$

for all $\omega \in \Omega$ and

$$g(\omega) := H_{\mu_\omega^{\mathscr{E}}}(\mathcal{M}) + \log 2.$$

For all $\omega \in \Omega_0$, Corollary 3.23 implies

$$f_n(\omega) \leq \frac{1}{n} H_{\mu_\omega^{\mathscr{E}}}(\mathcal{M}^{(n)}) + \log 2 \leq H_{\mu_\omega^{\mathscr{E}}}(\mathcal{M}) + \log 2 = g(\omega).$$

Since $\mu_\omega^{\mathscr{E}}$ is ergodic with regard to T for every $\omega \in \Omega_1$, the pointwise convergence

$$\lim_{n\to\infty} f_n(\omega) = h_{\mu_\omega^{\mathscr{E}}}(T)$$

follows from Theorem 3.25. Additionally,

$$\int g\,\mathrm{d}\mu = H(\mathcal{M}|\mathscr{E}) + \log(2) \leq H_\mu(\mathcal{M}) + \log 2 < \infty$$

and g and f_n are measurable. This allows us to use the dominated convergence theorem to show

$$\begin{aligned}\int h_{\mu_\omega^{\mathscr{E}}}(T)\,\mathrm{d}\mu(\omega) &= \int \lim_{n\to\infty} f_n(\omega)\,\mathrm{d}\mu(\omega)\\ &= \lim_{n\to\infty}\int f_n(\omega)\,\mathrm{d}\mu(\omega)\\ &= \lim_{n\to\infty}\frac{1}{n}\int H_{\mu_\omega^{\mathscr{E}}}(OP(n))\,\mathrm{d}\mu(\omega)\\ &= \lim_{n\to\infty}\frac{1}{n}H_\mu(OP(n)|\mathscr{E}).\end{aligned}$$

The last equality is simply the definition of the conditional entropy as given in 3.30. □

Using the above lemma, we can now show that, for the here considered class of measure-preserving dynamical system, equality of permutation entropy and Kolmogorov-Sinai entropy is equivalent to being able to apply the methods used in the ergodic decomposition of the Kolmogorov-Sinai entropy in Theorem 3.34 analogously to the permutation entropy.

Theorem 3.36 (Ergodic decomposition of permutation entropy)**.** Let $(\Omega, \mathscr{B}, \mu)$ be a Standard probability space with $\Omega \subseteq \mathbb{R}$ and $T : \Omega \to \Omega$ an aperiodic measure-preserving map. Suppose that T is countably piecewise monotone and $\mathcal{M}$ a countable partition into monotony parts of T with $H(\mathcal{M}) < \infty$. Then the following two statements are equivalent:

(i) $\lim_{n\to\infty} \frac{1}{n} H(OP(n)|\mathscr{E}) = \overline{\mathrm{PE}}(T)$ with $\mathscr{E}$ defined as in (3.25),

(ii) $\underline{\mathrm{PE}}(T) = \overline{\mathrm{PE}}(T) = h(T)$.

Proof. Suppose that (i) holds true. Then

$$\overline{\mathrm{PE}}(T) \overset{(i)}{=} \lim_{n\to\infty}\frac{1}{n}H(OP(n)|\mathscr{E}) \overset{Lem.3.35}{=} \int h_{\mu_\omega^{\mathscr{E}}}(T)\,\mathrm{d}\mu(\omega) \overset{Thm.3.34}{=} h_\mu(T).$$

Now, suppose that (ii) holds true. Lemma 3.31 implies

$$\lim_{n\to\infty}\frac{1}{n}H(OP(n)|\mathscr{E}) \le \overline{\mathrm{PE}}(T).$$

It remains to show the reverse inequality. Take $\varepsilon > 0$. According to Theorem 2.28, there exists a finite partition $\mathcal{P}$ into intervals satisfying $h(T, \mathcal{P}) \ge h(T) - \varepsilon$. Lemma 3.31 provides

$$\begin{aligned}H(\mathcal{P}^{(n)}|\mathscr{E}) \le& H(\mathcal{P}^{(n)} \vee OP(n)|\mathscr{E})\\ =& H(\mathcal{P}^{(n)}|OP(n) \vee \mathscr{E}) + H(OP(n)|\mathscr{E})\\ \le& H(\mathcal{P}^{(n)}|OP(n)) + H(OP(n)|\mathscr{E})\end{aligned}$$

Lemma 3.3 combined with Lemma 3.6 yields

$$\lim_{n\to\infty}\frac{1}{n}H(\mathcal{P}^{(n)}|OP(n)) \le \lim_{n\to\infty}\frac{1}{n}(\#\mathcal{P}-1)\log(n+\#\mathcal{P}-1) = 0$$

and Lemma 3.33 implies

$$\lim_{n\to\infty}\frac{1}{n}H(\mathcal{P}^{(n)}|\mathscr{E}) = h(T, \mathcal{P}).$$

Therefore,

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}H(OP(n)|\mathscr{E}) &\geq \lim_{n\to\infty}\frac{1}{n}\left[H(\mathcal{P}^{(n)}|\mathscr{E}) - H(\mathcal{P}^{(n)}|OP(n))\right]\\ &= h(T,\mathcal{P})\\ &\geq h(T)-\varepsilon\\ &\overset{\text{(ii)}}{=} \overline{\mathrm{PE}}(T)-\varepsilon.\end{aligned}$$

This implies (i) because ε can be chosen arbitrarily close to 0. □

Unfortunately, we are not able to show that

$$\lim_{n\to\infty}\frac{1}{n}H(OP(n)|\mathscr{E}) = \overline{\mathrm{PE}}(T)$$

holds true in general. However, we can still use ergodic decomposition, but not by applying this concept to the permutation entropy directly but by applying ergodic decomposition to Theorem 3.25 which contained a condition for the equality of the permutation and Kolmogorov-Sinai entropy:

Lemma 3.37. Let $(\Omega,\mathscr{B},\mu)$ be a Standard probability space and $T:\Omega\to\Omega$ an aperiodic measure-preserving map. Then for all $\varepsilon>0$, there exists a finite ordered partition $\mathcal{Q}$ that satisfies (3.13), i.e.

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

holds true.

Proof. Define

$$Q_j^k := \begin{cases}[j\cdot 2^{-k},(j+1)\cdot 2^{-k}[\,\cap\,\Omega & \text{if } j\in\{-4^k,\dots,4^k-1\}\\ ]-\infty,2^{-k}[\,\cap\,\Omega & \text{if } j=-4^k-1\\ [2^k,\infty[\,\cap\,\Omega & \text{if } j=4^k\end{cases}$$

for all $j\in\{-4^k-1,4^k,\dots,4^k-1,,4^k\}$ and $k\in\mathbb{N}$ and set

$$J_k := \{j\in\{-4^k-1,4^k,\dots,4^k-1,4^k\}\mid Q_j^k\neq\emptyset\}.$$

The collection of sets $\mathcal{Q}_k := \{Q_j^k\}_{j\in J_k}$ is a finite ordered partitions of Ω satisfying

$$Q_k \prec Q_{k+1} \tag{3.27}$$

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

$$\bigcap_{k=1}^{\infty}\mathcal{Q}_k(\omega)=\{\omega\}, \tag{3.28}$$

for all $\omega\in\Omega$, where $\mathcal{Q}_k(\omega)$ denotes the subset of $\mathcal{Q}_k$ that contains the element ω.

According to Theorem 3.28 and Lemma 3.32, there exists a decomposition

$$\mu=\int\mu_\omega^{\mathscr{E}}$$

of μ and a set $\Omega_0 \in \mathscr{B}$ with $\mu(\Omega_0) = 1$ such that T is measure-preserving, aperiodic and ergodic with respect to $\mu_\omega^{\mathscr{E}}$ for all $\omega \in \Omega_0$.

Now define

$$f_{n,k}(\omega) := \frac{1}{n} \sum_{j \in J_k} \sum_{l=1}^{n} \mu_\omega^{\mathscr{E}}(Q_j^k \cap T^{-l}(Q_j^k))$$

for all $\omega \in \Omega$ and $n, k \in \mathbb{N}$. Since T is ergodic with regard to $\mu_\omega^{\mathscr{E}}$ for all $\omega \in \Omega_0$, Lemma 2.9 (iv) implies

$$\lim_{n\to\infty} f_{n,k}(\omega) = \sum_{j \in J_k} \mu_\omega^{\mathscr{E}}(Q_j^k)^2$$

for all $k \in \mathbb{N}$ and $\omega \in \Omega_0$. Because

$$f_{n,k}(\omega) \leq \frac{1}{n} \sum_{j \in J_k} \sum_{l=1}^{n} \mu_\omega^{\mathscr{E}}(Q_j^k) = \sum_{j \in J_k} \mu_\omega^{\mathscr{E}}(Q_j^k) = 1$$

holds true for all $n, k \in \mathbb{N}$ and $\omega \in \Omega$, we can use the dominated convergence theorem to deduce

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j \in J_k} \sum_{l=1}^{n} \mu(Q_j^k \cap T^{-l}(Q_j^k)) = \lim_{n\to\infty} \frac{1}{n} \sum_{j \in J_k} \sum_{l=1}^{n} \int \mu_\omega^{\mathscr{E}}(Q_j^k \cap T^{-l}(Q_j^k))\, \mathrm{d}\mu(\omega)$$
$$= \lim_{n\to\infty} \int f_{n.k}(\omega)\, \mathrm{d}\mu(\omega) = \int \lim_{n\to\infty} f_{n.k}(\omega)\, \mathrm{d}\mu(\omega) = \int \sum_{j \in J_k} \mu_\omega^{\mathscr{E}}(Q_j^k)^2\, \mathrm{d}\mu(\omega).$$

The sequence $\left(\sum_{j \in J_k} \mu_\omega^{\mathscr{E}}(Q_j^k)^2\right)_{k \in \mathbb{N}}$ is monotonically decreasing for $k \to \infty$ for all $\omega \in \Omega$, so it is dominated and its limit exists and we can again use the dominated convergence theorem to show

$$\lim_{k\to\infty} \lim_{n\to\infty} \frac{1}{n} \sum_{j \in J_k} \sum_{l=1}^{n} \mu(Q_j^k \cap T^{-l}(Q_j^k))$$
$$= \lim_{k\to\infty} \int \sum_{j \in J_k} \mu_\omega^{\mathscr{E}}(Q_j^k)^2\, \mathrm{d}\mu(\omega) = \int \lim_{k\to\infty} \sum_{j \in J_k} \mu_\omega^{\mathscr{E}}(Q_j^k)^2\, \mathrm{d}\mu(\omega). \tag{3.29}$$

Set

$$g_{k,\omega}(\omega') := \mu_\omega^{\mathscr{E}}(\mathcal{Q}_k(\omega'))$$

for all $\omega, \omega' \in \Omega$. The sequence $(g_{k,\omega}(\omega'))_{k \in \mathbb{N}}$ is monotonically decreasing due to (3.27) and bounded from below for all $\omega, \omega' \in \Omega$, therefore $\lim_{k\to\infty} g_{k,\omega}(\omega')$ exists for all $\omega, \omega' \in \Omega$.

Now take any $\varepsilon > 0$. According to Egorov's theorem (see e.g. [6]), for all $\omega \in \Omega$ there exists a set $B_\omega \in \mathscr{B}$ with

$$\mu_\omega^{\mathscr{E}}(\Omega \setminus B_\omega) < \varepsilon/2, \tag{3.30}$$

such that $g_{k,\omega}$ converges uniformly for $k \to \infty$ on B_ω, i.e.

$$\lim_{k\to\infty} \sup_{\omega' \in B_\omega} \left| g_{k,\omega}(\omega') - \lim_{k\to\infty} g_{k,\omega}(\omega') \right| = 0. \tag{3.31}$$

Since $\mu_\omega^{\mathscr{E}}$ is aperiodic for all $\omega \in \Omega_0$, we have

$$\lim_{k\to\infty} g_{k,\omega}(\omega') = \lim_{k\to\infty} \mu_\omega^{\mathscr{E}}(\mathcal{Q}_k(\omega')) = \mu_\omega^{\mathscr{E}}\left(\bigcap_{k=1}^{\infty} \mathcal{Q}_k(\omega')\right) \overset{(3.28)}{=} \mu_\omega^{\mathscr{E}}(\{\omega'\}) = 0 \tag{3.32}$$

for all $\omega \in \Omega_0$ and $\omega' \in \Omega$ due to Lemma 2.13. Therefore, (3.31) is equivalent to

$$\lim_{k\to\infty} \sup_{\omega'\in B_\omega} g_{k,\omega}(\omega') = 0. \tag{3.33}$$

Notice that $g_{k,\omega}(\omega')$ is constant on the subsets of $\mathcal{Q}_k$, i.e.

$$g_{k,\omega}(\omega_1') = \mu_\omega^{\mathscr{E}}(\mathcal{Q}_k(\omega_1')) = \mu_\omega^{\mathscr{E}}(\mathcal{Q}_k(\omega_2')) = g_{k,\omega}(\omega_2')$$

for all $\omega_1', \omega_2' \in Q_j^k \in \mathcal{Q}_k$, $k \in \mathbb{N}$ and $\omega \in \Omega$. This implies

$$\sup_{\omega'\in B_\omega} g_{k,\omega}(\omega') = \sup_{\omega'\in B_\omega} \mu_\omega^{\mathscr{E}}(\mathcal{Q}_k(\omega')) = \sup_{\substack{j\in J_k:\\ Q_j^k\cap B_\omega\neq\emptyset}} \mu_\omega^{\mathscr{E}}(Q_j^k). \tag{3.34}$$

Hence, we can conclude that

$$\begin{aligned}
&\lim_{k\to\infty}\lim_{n\to\infty}\frac{1}{n}\sum_{j\in J_k}\sum_{l=1}^{n}\mu(Q_j^k\cap T^{-l}(Q_j^k))\\
&\overset{(3.29)}{=} \int \lim_{k\to\infty}\sum_{j\in J_k}\mu_\omega^{\mathscr{E}}(Q_j^k)^2\,\mathrm{d}\mu(\omega) = \int_{\Omega_0}\lim_{k\to\infty}\sum_{j\in J_k}\mu_\omega^{\mathscr{E}}(Q_j^k)^2\,\mathrm{d}\mu(\omega)\\
&= \int_{\Omega_0}\lim_{k\to\infty}\left[\sum_{\substack{j\in J_k:\\ Q_j^k\cap B_\omega=\emptyset}}\mu_\omega^{\mathscr{E}}(Q_j^k)^2 + \sum_{\substack{j\in J_k:\\ Q_j^k\cap B_\omega\neq\emptyset}}\mu_\omega^{\mathscr{E}}(Q_j^k)^2\right]\mathrm{d}\mu(\omega)\\
&\le \int_{\Omega_0}\lim_{k\to\infty}\left[\mu_\omega^{\mathscr{E}}(\Omega\setminus B_\omega) + \left(\sup_{\substack{j\in J_k:\\ Q_j^k\cap B_\omega\neq\emptyset}}\mu_\omega^{\mathscr{E}}(Q_j^k)\right)\left(\sum_{\substack{j\in J_k:\\ Q_j^k\cap B_\omega\neq\emptyset}}\mu_\omega^{\mathscr{E}}(Q_j^k)\right)\right]\mathrm{d}\mu(\omega)\\
&\overset{(3.30)}{\le} \int_{\Omega_0}\varepsilon/2 + \lim_{k\to\infty}\sup_{\substack{j\in J_k:\\ Q_j^k\cap B_\omega\neq\emptyset}}\mu_\omega^{\mathscr{E}}(Q_j^k)\,\mathrm{d}\mu(\omega)\\
&\overset{(3.34)}{=} \int_{\Omega_0}\varepsilon/2 + \lim_{k\to\infty}\sup_{\omega''\in B_\omega}\mu_\omega^{\mathscr{E}}(Q_j^k)\,\mathrm{d}\mu(\omega)\\
&\overset{(3.33)}{=} \int_{\Omega_0}\varepsilon/2\,\mathrm{d}\mu(\omega) = \varepsilon/2
\end{aligned}$$

holds true. Therefore, we can choose a $k \in \mathbb{N}$ such that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j\in J_k}\sum_{l=1}^{n}\mu(Q_j^k\cap T^{-l}(Q_j^k)) \le \varepsilon$$

holds true. Then $\mathcal{Q} = \mathcal{Q}_k$ satisfies (3.13). □

An alternative method to prove the above lemma can be found in a paper by T. Gutjahr and K. Keller [12]. They used so called Rokhlin towers instead: Let $(\Omega, \mathscr{B}, \mu, T)$ be a measure-preserving dynamical system as given Lemma 3.37, then for all $\varepsilon > 0$ and $n \in \mathbb{N}$, there exists a set $B \in \mathscr{B}$ such that

$$\mu\left(\bigcup_{t=0}^{n-1} T^{-t}(B)\right) > 1 - \varepsilon$$

holds true and the sets $T^{-t}(B)$,$t \in \{0, 1, \dots, n-1\}$ are pairwise disjoint. The collection of sets $\{T^{-t}(B) \mid t \in \{0, 1, \dots, n-1\}\}$ is called Rokhlin tower of height n with base B. The sets in the finite ordered partition $\mathcal{Q}$ in Lemma 3.37 were then constructed based on the different sets in the Rokhlin tower.

Using Theorem 3.18 and the previous lemmas, we can now prove a version of the main Theorem 3.17 for aperiodic dynamical system.

Theorem 3.38 (Main result for aperiodic systems)**.** Let $(\Omega, \mathscr{B}, \mu)$ be a Standard probability space with $\Omega \subseteq \mathbb{R}$ and $T : \Omega \to \Omega$ an aperiodic measure-preserving map. Suppose that T is countably piecewise monotone and $\mathcal{M}$ a countable partition into monotony parts of T with $H(\mathcal{M}) < \infty$. Then

$$\underline{\mathrm{PE}}(T) = \overline{\mathrm{PE}}(T) = h(T)$$

holds true

Proof. It follows from Lemma 3.37 and Lemma 3.21 that the conditions of Theorem 3.18 are fulfilled which directly provides $\overline{\mathrm{PE}}(T) \leq h(T)$. Together with Theorem 3.10, this implies

$$h(T) \leq \underline{\mathrm{PE}}(T) \leq \overline{\mathrm{PE}}(T) \leq h(T).$$ □

3.4.6 Example of piecewise monotone function with $H(\mathcal{M}) = \infty$

In this subsection we will illustrate that the condition $H(\mathcal{M}) < \infty$ for the partition into monotony parts $\mathcal{M}$ of a piecewise monotone function T is sufficient but not necessary for $\mathrm{PE}(T) = h(T)$ to hold true.

Given a piecewise monotone function T and its partition into monotony parts $\mathcal{M}$, the partition $\mathcal{M} \vee \mathcal{Q}$ is a partition into monotony parts for every ordered partition $\mathcal{Q}$. So, many different partitions $\mathcal{M}$ into monotony parts can be found for one function T. To find a possible necessary condition it is, therefore, reasonable to only consider the smallest partition into monotony parts. We call the partition $\mathcal{M}$ smallest partition into monotony parts, if it is a partition into monotony parts and

$$\mathcal{M} \prec \mathcal{M}'$$

is true for every partition into monotony parts $\mathcal{M}'$. Technically, the smallest partition $\mathcal{M}$ is generally not unique because we can add or remove sets with measure 0 to the individual subsets of $\mathcal{M}$. If we consider two partitions $\mathcal{M}_1$ and $\mathcal{M}_2$ equivalent, when $\mathcal{M}_1 \prec M_2$ and $\mathcal{M}_2 \prec \mathcal{M}_1$ holds true, the smallest partition into monotony will then be unique modulo this equivalence relation.

But even when considering the smallest partition into monotony parts $\mathcal{M}$, the condition $H(\mathcal{M}) > \infty$ is not always necessary for $\mathrm{PE}(T) = h(T)$ to hold true. In the following example we will consider a measure-preserving dynamical system $(\Omega, \mathscr{B}, \mu, T)$ with $H(\mathcal{M}) = \infty$ for the smallest partition into monotony parts and $\mathrm{PE}(T) = h(T)$.

Example 4. Consider $\Omega = [0,1[$ and the Borel σ-algebra $\mathscr{B}$ on Ω. Set

$$\begin{aligned} S &:= \sum_{i=0}^{\infty} \frac{1}{(i+1)(\log(i+1))^2}, \\ m_i &:= \frac{1}{S} \cdot \frac{1}{(i+1)(\log(i+1))^2} \qquad \text{for } i \in \mathbb{N}, \\ a_0 &:= 0, \\ a_1 &:= m_1, \\ a_i &:= 1 - \frac{m_{i+1}}{m_1} \qquad \text{for } i \geq 2, \\ M_i &:= [a_{i-1}, a_i[\qquad \text{for } i \in \mathbb{N}. \end{aligned}$$

Notice that $\{M_i\}_{i\in\mathbb{N}}$ is a countable partition of Ω into intervals.

The function $T : \Omega \to \Omega$ is defined piecewise linear on each set M_i (see figure 3.3) by

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

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

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

for all $A \in \mathscr{B}$. The measure μ is defined in such a way that

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

holds true for all $i \in \mathbb{N}$. Since $\{M_i\}_{i\in\mathbb{N}}$ is a countable partition of Ω, we have

$$\mu(\Omega) = \sum_{i=1}^{\infty} \mu(M_i) = \sum_{i=1}^{\infty} m_i = \frac{1}{S} \sum_{i=1}^{\infty} \frac{1}{(i+1)(\log(i+1))^2} = \frac{S}{S} = 1,$$

so μ is, in fact, a probability measure.

By construction of T, we have $T(M_i) = M_{i-1}$ for all $i \in \mathbb{N} \setminus \{1\}$. This implies

$$T^{-1}(M_{i-1}) \cap [a_1, 1[= M_i$$

for all $i \in \mathbb{N} \setminus \{1\}$. In each interval M_i, the function T has the constant derivative $\frac{a_{i-2} - a_{i-1}}{a_i - a_{i-1}}$ for $i \in \mathbb{N} \setminus \{1\}$ and $\frac{1}{m_1}$ for $i = 1$. Thus, for all $i \in \mathbb{N} \setminus \{1\}$ and all intervals $[a, b[\subseteq M_{i-1}$

$$\lambda(T^{-1}([a, b[) \cap M_i) = \frac{a_i - a_{i-1}}{a_{i-1} - a_{i-2}} \cdot \lambda([a, b[)$$

holds true, which yields

$$\begin{aligned}
\mu(T^{-1}([a,b[)) &= \mu(T^{-1}([a,b[\cap[0,a_1[) + \mu(T^{-1}([a,b[\cap[a_1,1[) \\
&= \mu(T^{-1}([a,b[\cap M_1) + \mu(T^{-1}([a,b[\cap M_i) \\
&= \frac{m_1\cdot\lambda(T^{-1}([a,b[\cap M_1)}{\lambda(M_1)} + \frac{m_i\cdot\lambda(T^{-1}([a,b[\cap M_i)}{\lambda(M_i)} \\
&= \frac{m_1\cdot m_1\cdot\lambda([a,b[)}{a_1-a_0} + \frac{m_i\cdot\frac{a_i-a_{i-1}}{a_{i-1}-a_{i-2}}\cdot\lambda([a,b[)}{a_i-a_{i-1}} \\
&= \lambda([a,b[)\cdot\left(\frac{m_1^2}{m_1} + \frac{m_i\cdot(a_i-a_{i-1})}{(a_{i-1}-a_{i-2})\cdot(a_i-a_{i-1})}\right) \\
&= \lambda([a,b[)\cdot\left(\frac{m_1\cdot(a_{i-1}-a_{i-2})}{a_{i-1}-a_{i-2}} + \frac{m_i}{a_{i-1}-a_{i-2}}\right) \\
&= \lambda([a,b[)\cdot\left(\frac{m_1\cdot\left(\frac{m_{i-1}}{m_1}-\frac{m_i}{m_1}\right)}{a_{i-1}-a_{i-2}} + \frac{m_i}{a_{i-1}-a_{i-2}}\right) \\
&= \lambda([a,b[)\cdot\left(\frac{m_{i-1}-m_i+m_i}{\lambda(M_{i-1})}\right) = \frac{m_{i-1}\cdot\lambda([a,b[)}{\lambda(M_{i-1})} \\
&= \frac{m_{i-1}\cdot\lambda([a,b[\cap M_{i-1})}{\lambda(M_{i-1})} = \mu([a,b[)
\end{aligned}$$

for all $[a,b[\in M_{i-1}$. Hence, μ is invariant with regard to T.

It is easy to see that the smallest partition $\mathcal{M}$ into monotony parts for the function given in example 4 is given by

$$\mathcal{M} := \{M_i\}_{i\in\mathbb{N}},$$

where M_i is defined as above. This partition satisfies

$$\begin{aligned}
H(\mathcal{M}) &= -\sum_{i=1}^{\infty}\mu(M_i)\log\mu(M_i) = -\sum_{i=1}^{\infty} m_i\log(m_i) \\
&= \log(S) + \frac{1}{S}\cdot\sum_{i=2}^{\infty}\frac{\log(i\cdot(\log(i))^2)}{i\cdot(\log(i))^2} \geq \log(S) + \frac{1}{S}\cdot\sum_{i=2}^{\infty}\frac{\log(i)}{i\cdot(\log(i))^2} \\
&= \log(S) + \frac{1}{S}\cdot\sum_{i=2}^{\infty}\frac{1}{i\cdot\log(i)} = \infty.
\end{aligned}$$

So we cannot use Theorem 3.17 to show that

$$\overline{\mathrm{PE}}(T) = h(T)$$

holds true. However, we can use the more general Theorem 3.18 to show this fact. But first, one needs to verify that the conditions of Theorem 3.18 are fulfilled. We will first show that there exists a partition $\mathcal{M}$ of $[0,1[$ into intervals, such that

$$\mathcal{M}^{(m)}\otimes\mathcal{M}^{(m)}\vee\{R,\mathbb{R}^2\setminus R\} \ \prec\ \mathcal{M}^{(m)}\otimes\mathcal{M}^{(m)}\vee\bigvee_{u=1}^{m}(T\times T)^{-u}(\{R,\mathbb{R}^2\setminus R\})$$

holds true for some $m\in\mathbb{N}$.

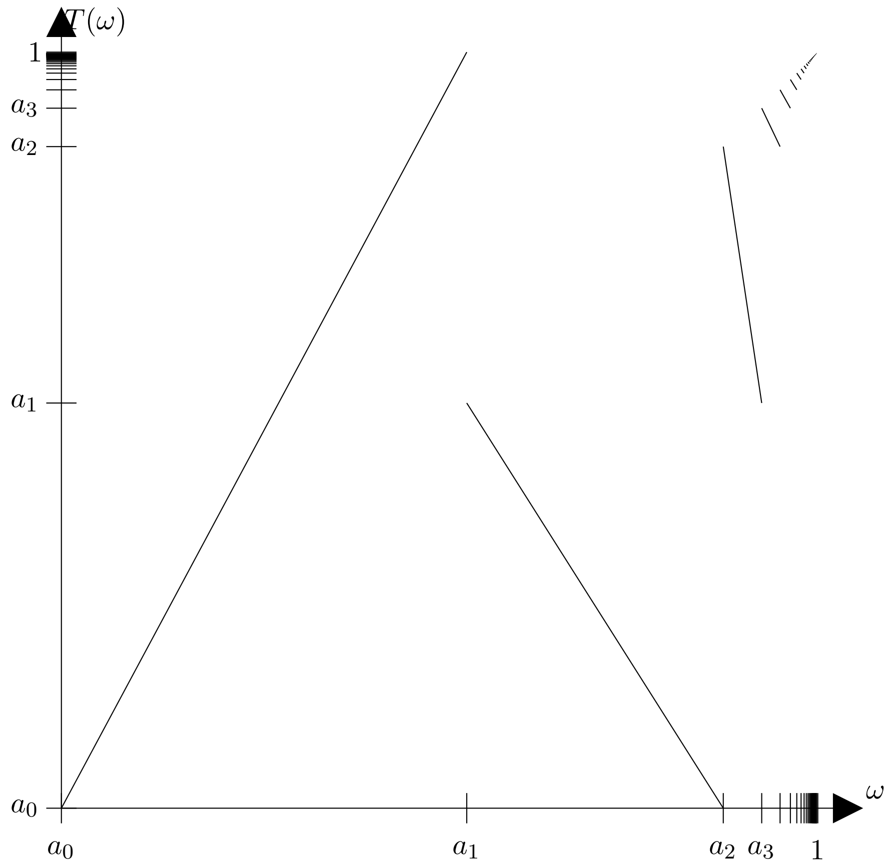


Figure 3.3: Graph of the function T given in example 4

Lemma 3.39. Let $(\Omega, \mathscr{B}, \mu, T)$ be given as in example 4 and $\mathcal{M} := \{M_1, M_2\}$ with $M_1 := [0, a_1[$ and $M_2 := [a_1, 1[$. Then (3.14) holds true with $m = 2$, that is

$$\mathcal{M}^{(2)} \otimes \mathcal{M}^{(2)} \vee \{R, \mathbb{R}^2 \setminus R\} \prec \mathcal{M}^{(2)} \otimes \mathcal{M}^{(2)} \vee \bigvee_{u=1}^{2} (T \times T)^{-u}(\{R, \mathbb{R}^2 \setminus R\})$$

is true.

Proof. Every set $C \in \mathcal{M}^{(2)} \otimes \mathcal{M}^{(2)}$ can be written as

$$C = C_1^1 \cap T^{-1}(C_2^1) \times C_1^2 \cap T^{-1}(C_2^2)$$

with $C_i^j \in \mathcal{M}$ for all $i, j \in \{1, 2\}$. We will show that for all 16 possible sets $C \in \mathcal{M}^{(2)} \otimes \mathcal{M}^{(2)}$

$$\{C\} \vee \{R, \mathbb{R}^2 \setminus R\} \prec \{C\} \vee \bigvee_{u=1}^{2} (T \times T)^{-u}(\{R, \mathbb{R}^2 \setminus R\})$$

holds true, which then implies the statement of this lemma.

If $C_1^1 \neq C_1^2$ holds true, we have $\omega_1 < \omega_2$ for all $(\omega_1, \omega_2) \in C$ or $\omega_1 \geq \omega_2$ for all $(\omega_1, \omega_2) \in C$, which implies

$$\{C\} \vee \{R, \mathbb{R}^2 \setminus R\} = \{C\} \prec \{C\} \vee \bigvee_{u=1}^{2} (T \times T)^{-u}(\{R, \mathbb{R}^2 \setminus R\}).$$

If $C_1^1 = C_1^2 = M_1$ holds true, we have $\omega_1, \omega_2 \in M_1$ for all $(\omega_1, \omega_2) \in C$. Since T is monotonically increasing on M_1, the order relation between ω_1 and ω_2 is the same as the order relation between $T(\omega_1)$ and $T(\omega_2)$ for all $(\omega_1, \omega_2) \in C$. This implies

$$\{C\} \vee \{R, \mathbb{R}^2 \setminus R\} = \{C\} \vee (T \times T)^{-1}(\{R, \mathbb{R}^2 \setminus R\}) \prec \{C\} \vee \bigvee_{u=1}^{2} (T \times T)^{-u}(\{R, \mathbb{R}^2 \setminus R\}).$$

It remains to consider the four cases where $C_1^1 = C_1^2 = M_2$ holds true. Notice that

$$\begin{aligned} M_2 \cap T^{-1}(M_1) &= [a_1, a_2[, \\ M_2 \cap T^{-1}(M_2) &= [a_2, 1[\end{aligned} \tag{3.35}$$

is true. So, if additionally $C_2^1 \neq C_2^2$ is satisfied, we have either $\omega_1 \in [a_1, a_2[$ and $\omega_2 \in [a_2, 1[$ or vice-versa. Since $[a_1, a_2[$ and $[a_2, 1[$ are two disjoint intervals, this implies

$$\{C\} \vee \{R, \mathbb{R}^2 \setminus R\} = \{C\} \prec \{C\} \vee \bigvee_{u=1}^{2} (T \times T)^{-u}(\{R, \mathbb{R}^2 \setminus R\}).$$

Now consider the cases where $C_1^1 = C_1^2 = M_2$ and $C_2^1 = C_2^2 = M_1$ holds true. According to (3.35), we have $(\omega_1, \omega_2) \in [a_1, a_2[$ for all $(\omega_1, \omega_2) \in C$. Since T is monotonically decreasing on $[a_1, a_2[$, the order relation between ω_1 and ω_2 is the same as the order relation between $T(\omega_2)$ and $T(\omega_1)$ for all $(\omega_1, \omega_2) \in C$. This implies

$$\{C\} \vee \{R, \mathbb{R}^2 \setminus R\} = \{C\} \vee (T \times T)^{-1}(\{R, \mathbb{R}^2 \setminus R\}) \prec \{C\} \vee \bigvee_{u=1}^{2} (T \times T)^{-u}(\{R, \mathbb{R}^2 \setminus R\}).$$

Finally, consider the case where $C_1^1 = C_1^2 = M_2$ and $C_2^1 = C_2^2 = M_2$ holds true. Set

$$D := \bigcup_{i=1}^{\infty} \{(\omega_1, \omega_2) \in \Omega^2 : \omega_1, \omega_2 \in [a_i, a_{i+1}[\}$$

as the set of all pairs (ω_1, ω_2) where ω_1 and ω_2 are located in the same interval $[a_i, a_{i+1}[$ for some $i \in \mathbb{N}$. If $(\omega_1, \omega) \in D$ is fulfilled for $(\omega_1, \omega_2) \in C$, the points ω_1 and ω_2 are located in the same interval $[a_i, a_{i+1}[$ where the function T is monotonically decreasing. This implies

$$\begin{aligned} \{C\} \vee \{R, \mathbb{R}^2 \setminus R\} &\prec \{C\} \vee \{R, \mathbb{R}^2 \setminus R\} \vee \{D\} \\ &= \{C\} \vee (T \times T)^{-1}(\{R, \mathbb{R}^2 \setminus R\}) \vee \{D\} \\ &\prec \{C\} \vee \bigvee_{u=1}^{2} (T \times T)^{-u}(\{R, \mathbb{R}^2 \setminus R\}) \vee \{D\}. \end{aligned} \tag{3.36}$$

If $(\omega_1, \omega) \in D$ is not fulfilled for $(\omega_1, \omega_2) \in C$, the points ω_1 and ω_2 are located in different intervals $[a_i, a_{i+1}[$ and $[a_j, a_{j+1}[$. Since $T([a_i, a_{i+1}[) = [a_{i-1}, a_i[$ for all $i \in \mathbb{N}$, the order relation between $[a_i, a_{i+1}[$ and $[a_j, a_{j+1}[$ is the same as the order relation between the intervals $T([a_i, a_{i+1}[)$ and $T([a_j, a_{j+1}[)$. This implies

$$\begin{aligned}\{C\} \vee \{R, \mathbb{R}^2 \setminus R\} &\prec \{C\} \vee \{R, \mathbb{R}^2 \setminus R\} \vee \{\Omega \setminus D\} \\ &= \{C\} \vee (T \times T)^{-1}(\{R, \mathbb{R}^2 \setminus R\}) \vee \{\Omega \setminus D\} \\ &\prec \{C\} \vee \bigvee_{u=1}^{2} (T \times T)^{-u}(\{R, \mathbb{R}^2 \setminus R\}) \vee \{\Omega \setminus D\}. \qquad (3.37)\end{aligned}$$

Combining (3.36) and (3.37) provides

$$\{C\} \vee \{R, \mathbb{R}^2 \setminus R\} \prec \{C\} \vee \bigvee_{u=1}^{2} (T \times T)^{-u}(\{R, \mathbb{R}^2 \setminus R\}) \vee \{D, \Omega \setminus D\} \qquad (3.38)$$

for $C_1^1 = C_1^2 = C_2^1 = C_2^2 = M_2$. To determine whether $(\omega_1, \omega_2) \in D$ holds true, we need the information of $(T \times T)^{-2}(\{R, \mathbb{R}^2 \setminus R\})$, which has not been used so far in this proof.

For $(\omega_1, \omega_2) \in C \cap D$, we have $(T(\omega_1), T(\omega_2)) \in D$. Therefore, ω_1 and ω_2 are located in the same interval $[a_i, a_{i+1}[$ for some $i \in \mathbb{N}$ on which the function T is monotonically decreasing. This implies that the order relation between $T(\omega_1)$ and $T(\omega_2)$ is the same as the order relation between $T^2(\omega_2)$ and $T^2(\omega_1)$.

For $(\omega_1, \omega_2) \in C \setminus D$, we have $(T(\omega_1), T(\omega_2)) \notin D$. Therefore, ω_1 and ω_2 are located in the different intervals $[a_i, a_{i+1}[$ and $[a_j, a_{j+1}[$ with $i, j \in \mathbb{N}$ and $i \neq j$. Since $T([a_i, a_{i+1}[) = [a_{i-1}, a_i[$ for all $i \in \mathbb{N}$, the order relation between $T(\omega_1)$ and $T(\omega_2)$ is the same as the order relation between $T^2(\omega_1)$ and $T^2(\omega_2)$.

To summarize, we can infer whether (ω_1, ω_2) is located in D by comparing the order relation between $T(\omega_1)$ and $T(\omega_2)$ and the order relation between $T^2(\omega_1)$ and $T^2(\omega_2)$. This implies

$$\bigvee_{u=1}^{2} (T \times T)^{-u}(\{R, \mathbb{R}^2 \setminus R\}) \vee \{D, \Omega \setminus D\} \prec \bigvee_{u=1}^{2} (T \times T)^{-u}(\{R, \mathbb{R}^2 \setminus R\})$$

Together with (3.38), this yields

$$\{C\} \vee \{R, \mathbb{R}^2 \setminus R\} \prec \{C\} \vee \bigvee_{u=1}^{2} (T \times T)^{-u}(\{R, \mathbb{R}^2 \setminus R\}). \qquad \square$$

It is easy to see that the function T considered in example 4 is aperiodic. Together with the above lemma and the fact that $H(\mathcal{M}) < \infty$ holds true, Theorem 3.38 then implies $\underline{\mathrm{PE}}(T) = \overline{\mathrm{PE}}(T) = h(T)$.

3.4.7 Generalization to non-aperiodic maps

Recall that we call a dynamical system $(\Omega, \mathscr{A}, \mu, T)$ aperiodic if, $\mu(\bigcup_{n=1}^{\infty}\{\omega \in \Omega \mid T^n(\omega) = \omega\}) = 0$ holds true. In particular, aperiodicity implies that $T^s(\omega)$ and $T^t(\omega)$ cannot be equal for $s \neq t \in \mathbb{N}_0$ and μ-almost all $\omega \in \Omega$. This simplifies the definition of the permutation entropy because then the inequalities $T^s(\omega) \leq T^t(\omega)$ and $T^t(\omega) \leq T^s(\omega)$ are mutually exclusive. Additionally, aperiodicity was necessary in (3.32) for the proof of Lemma 3.37.

We will now show that the condition of aperiodicity in Theorem 3.17 is not needed to show that $\underline{\mathrm{PE}}(T) = \overline{\mathrm{PE}}(T) = h(T)$ holds true. Recall that

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

is the set of points with period less or equal to $k \in \mathbb{N}$ and

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

the set of all periodic points. Thus, a map T is aperiodic (with regard to μ) if

$$\mu(\Pi) = 0$$

holds true.

Lemma 3.40. Let $(\Omega, \mathscr{B}, \mu, T)$ be a measure-preserving dynamical system. If there exists $k \in \mathbb{N}$ with $\mu(\Pi_k) = 1$, then

$$h(T) = \underline{\mathrm{PE}}(T) = \overline{\mathrm{PE}}(T) = 0.$$

Proof. It is enough to show $\overline{\mathrm{PE}}(T) = 0$ because Theorem 3.10 implies $h(T) \leq \underline{\mathrm{PE}}(T) \leq \overline{\mathrm{PE}}(T)$.

Suppose $\mu(\Pi_k) = 1$ for some $k \in \mathbb{N}$. Set $\Pi_0 := \emptyset$. Notice that for all $i \in \{1, 2, \ldots, k\}$ and $\omega \in \Pi_i \setminus \Pi_{i-1}$

$$T^t(\omega) = T^{t \bmod i}(\omega) \tag{3.39}$$

holds true. Consider now the partition

$$\mathcal{P} := \{\Pi_i \setminus \Pi_{i-1}\}_{i=1}^{k}$$

of Ω. Then Lemma 3.3 provides

$$\begin{aligned}
H(OP(n)) &\leq H(\mathcal{P}) + H(OP(n)|\mathcal{P}) \\
&\leq H(\mathcal{P}) + \sum_{i=1}^{k} \mu(\Pi_i \setminus \Pi_{i-1}) \cdot \log\left(\#\Delta\left(OP(n)|\Pi_i \setminus \Pi_{i-1}\right)\right) \\
&= H(\mathcal{P}) + \sum_{i=1}^{k} \mu(\Pi_i \setminus \Pi_{i-1}) \cdot \log\left(\#\Delta\left(\bigvee_{s,t=0}^{n-1} (T^s, T^t)^{-1}(R, \mathbb{R}^2 \setminus \mathbb{R})|\Pi_i \setminus \Pi_{i-1}\right)\right) \\
&\overset{(3.39)}{=} H(\mathcal{P}) + \sum_{i=1}^{k} \mu(\Pi_i \setminus \Pi_{i-1}) \cdot \log\left(\#\Delta\left(\bigvee_{s,t=0}^{i-1} (T^s, T^t)^{-1}(R, \mathbb{R}^2 \setminus \mathbb{R})|\Pi_i \setminus \Pi_{i-1}\right)\right) \\
&\leq H(\mathcal{P}) + \log\left(\#\Delta\left(\bigvee_{s,t=0}^{k-1} (T^s, T^t)^{-1}(R, \mathbb{R}^2 \setminus \mathbb{R})\right)\right).
\end{aligned}$$

This implies

$$\begin{aligned}
\overline{\mathrm{PE}}(T) &= \limsup_{n\to\infty} \frac{1}{n} H(OP(n)) \\
&\leq \limsup_{n\to\infty} \frac{1}{n} \left[H(\mathcal{P}) + \log\left(\#\Delta\left(\bigvee_{s,t=0}^{k-1} (T^s, T^t)^{-1}(R, \mathbb{R}^2 \setminus \mathbb{R})\right)\right)\right] \\
&= 0.
\end{aligned}$$

□

Given a measure μ on a measure space $(\Omega, \mathscr{A})$ and a set $A \in \mathscr{A}$ with $\mu(A) > 0$, the measure μ_A defined by

$$\mu_A(B) := \frac{\mu(B \cap A)}{\mu(A)}$$

for all $B \in \mathscr{A}$ is the restriction of μ on A.

The following two lemmas show how the entropy and the T-invariance are effected by restricting the measure μ on some set A with $\mu(A) > 0$.

Lemma 3.41. Let $(\Omega, \mathscr{A}, \mu)$ be a probability space and $\mathcal{M}$ a countable partition with $H_\mu(\mathcal{M}) < \infty$. Then for all sets $A \in \mathscr{A}$ with $\mu(A) > 0$

$$H_{\mu_A}(\mathcal{M}) < \infty$$

holds true, where μ_A is the probability measure on $(\Omega, \mathscr{A})$ defined by $\mu_A(B) := \frac{\mu(B\cap A)}{\mu(A)}$ for all $B \in \mathscr{A}$.

Proof. Consider $\Phi : [0,1] \to \mathbb{R}$ with $\Phi(x) := -x\log(x)$ for $x \in (0,1]$ and $\Phi(0) = 0$. The function Φ is monotonically increasing for $x \in]0, 1/e[$ because $\Phi'(x) = -\log(x) - 1 > 0$ for $x \in]0, 1/\mathrm{e}[$. Therefore,

$$\begin{aligned}
H_{\mu_A}(\mathcal{M}) &= -\sum_{M\in\mathcal{M}} \frac{\mu(M\cap A)}{\mu(A)} \log\left(\frac{\mu(M\cap A)}{\mu(A)}\right)\\
&= \log(\mu(A)) - \sum_{M\in\mathcal{M}} \frac{\mu(M\cap A)}{\mu(A)} \log\left(\mu(M\cap A)\right)\\
&= \log(\mu(A)) + \frac{1}{\mu(A)} \sum_{M\in\mathcal{M}} \Phi\left(\mu(M\cap A)\right)\\
&= \log(\mu(A)) + \frac{1}{\mu(A)} \sum_{\substack{M\in\mathcal{M}:\\ \mu(M)<e^{-1}}} \Phi\left(\mu(M\cap A)\right) + \frac{1}{\mu(A)} \sum_{\substack{M\in\mathcal{M}:\\ \mu(M)\geq e^{-1}}} \Phi\left(\mu(M\cap A)\right)\\
&\leq \log(\mu(A)) + \frac{1}{\mu(A)} \sum_{\substack{M\in\mathcal{M}:\\ \mu(M)<e^{-1}}} \Phi\left(\mu(M)\right) + \frac{1}{\mu(A)} \sum_{\substack{M\in\mathcal{M}:\\ \mu(M)\geq e^{-1}}} \Phi\left(\mu(M\cap A)\right)\\
&\leq \log(\mu(A)) + \frac{1}{\mu(A)} \sum_{M\in\mathcal{M}} \Phi\left(\mu(M)\right) + \frac{1}{\mu(A)} \sum_{\substack{M\in\mathcal{M}:\\ \mu(M)\geq e^{-1}}} \Phi\left(\mu(M\cap A)\right)\\
&= \log(\mu(A)) + \frac{1}{\mu(A)} H_\mu(\mathcal{M}) + \frac{1}{\mu(A)} \sum_{\substack{M\in\mathcal{M}:\\ \mu(M)\geq e^{-1}}} \Phi\left(\mu(M\cap A)\right).
\end{aligned}$$

Since $\#\{M \in \mathcal{M} : \mu(M) \geq 1/\mathrm{e}\} \leq \mathrm{e}$, the sum

$$\sum_{\substack{M\in\mathcal{M}:\\ \mu(M)\geq 1/\mathrm{e}}} \Phi\left(\mu(M\cap A)\right)$$

has to be finite. Therefore, the assumption $H_\mu(\mathcal{M}) < \infty$ implies $H_{\mu_A}(\mathcal{M}) < \infty$ according to the above inequality. □

Lemma 3.42. Let $(\Omega, \mathscr{A}, \mu, T)$ be a measure preserving dynamical system and $A \in \mathscr{A}$ with $\mu(A) > 0$ and

$$\mu(A \triangle T^{-1}(A)) = 0. \tag{3.40}$$

Then $(\Omega, \mathscr{A}, \mu_A, T))$ is a measure-preserving dynamical system.

Proof. Take any $A, B \in \mathscr{A}$. Then

$$\begin{aligned}
\mu(T^{-1}(B) \cap A) &\geq \mu(T^{-1}(B) \cap A \cap T^{-1}(A)) \\
&= \mu(T^{-1}(B) \cap A) - \mu(T^{-1}(B) \cap A \setminus T^{-1}(A)) \\
&\geq \mu(T^{-1}(B) \cap A) - \mu(A \setminus T^{-1}(A)) \\
&\geq \mu(T^{-1}(B) \cap A) - \mu(A \triangle T^{-1}(A)) \\
&\overset{(3.40)}{=} \mu(T^{-1}(B) \cap A).
\end{aligned} \tag{3.41}$$

This implies

$$\begin{aligned}
\mu_A(B) &= \frac{\mu(B \cap A)}{\mu(A)} = \frac{\mu\left(T^{-1}(B \cap A)\right)}{\mu(A)} = \frac{\mu\left(T^{-1}(B) \cap T^{-1}(A)\right)}{\mu(A)} \\
&= \frac{\mu\left(T^{-1}(B) \cap \left[\left(T^{-1}(A) \setminus A\right) \cup \left(A \cap T^{-1}(A)\right)\right]\right)}{\mu(A)} \\
&= \frac{\mu\left(T^{-1}(B) \cap \left(T^{-1}(A) \setminus A\right)\right) + \mu\left(T^{-1}(B) \cap A \cap T^{-1}(A)\right)}{\mu(A)} \\
&\overset{(3.40)}{=} \frac{\mu\left(T^{-1}(B) \cap A \cap T^{-1}(A)\right)}{\mu(A)} \\
&\overset{(3.41)}{=} \frac{\mu\left(T^{-1}(B) \cap A\right)}{\mu(A)} \\
&= \mu_A(T^{-1}(B)).
\end{aligned}$$

Thus, $(\Omega, \mathscr{A}, \mu_A, T))$ is a measure-preserving dynamical system. □

In the following Lemma we consider periodic systems where the length of the period is not necessarily uniformly bounded by some constant $k \in \mathbb{N}$, unlike in Lemma 3.40 where this was the case. To still be able to show that the permutation entropy of such systems is 0, we additionally need to require that the entropy of the system is bounded in some sense. This can be achieved if the considered system is (countably) piecewise monotone.

Lemma 3.43. Let $(\Omega, \mathscr{B}, \mu, T)$ be a measure-preserving dynamical system with $\Omega \subseteq \mathbb{R}$. Let T be countably piecewise monotone and $\mathcal{M} \in \mathbb{P}_o^c(\mathscr{B})$ a partition into monotony parts of T. Suppose $\mu(\Pi) = 1$ holds true. Then

$$\underline{\mathrm{PE}}(T) = \overline{\mathrm{PE}}(T) = h(T) = 0$$

holds true.

Proof. As in Lemma 3.40, it is enough to show $\overline{\mathrm{PE}}(T) = 0$ because Theorem 3.10 implies $h(T) \leq \underline{\mathrm{PE}}(T) \leq \overline{\mathrm{PE}}(T)$.

Since

$$1 = \mu(\Pi) = \lim_{k \to \infty} \mu(\Pi_k),$$

for all $m \in \mathbb{N}$ there exists $k(m) \in \mathbb{N}$ with

$$\mu_p\left(\Pi_{k(m)}\right) \geq 1 - \frac{1}{m}. \tag{3.42}$$

Now define for all $m \in \mathbb{N}$ the measure $\mu_p^{(m)}$ on $(\Omega, \mathscr{B})$ by

$$\mu_p^{(m)}(A) := \frac{\mu\left(A \cap \Pi_{k(m)}\right)}{\mu\left(\Pi_{k(m)}\right)}$$

for all $A \in \mathscr{B}$. According to Lemma 2.11,

$$\mu\left(T^{-1}\left(\Pi_{k(m)}\right) \triangle \Pi_{k(m)}\right) = 0.$$

holds true. Due to Lemma 3.42, this implies that $(\Omega, \mathscr{B}, \mu_p^{(m)}, T)$ is a measure-preserving dynamical system for all $m \in \mathbb{N}$. So we can apply Lemma 3.40 to $\overline{\mathrm{PE}}_{\mu_p^{(m)}}(T)$, which provides

$$\overline{\mathrm{PE}}_{\mu_p^{(m)}}(T) = 0. \tag{3.43}$$

We introduce a further probability measure $\mu_q^{(m)}$ on $(\Omega, \mathscr{B})$ defined by

$$\mu_q^{(m)}(A) := \begin{cases} \frac{\mu(A \setminus \Pi_{k(m)})}{1 - \mu(\Pi_{k(m)})}, & \text{if } \mu\left(\Pi_{k(m)}\right) < 1, \\ \mu_p^{(m)}(A), & \text{if } \mu_p\left(\Pi_{k(m)}\right) = 1 \end{cases}$$

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

$$\mu = \mu\left(\Pi_{k(m)}\right) \cdot \mu_p^{(m)} + \left(1 - \mu\left(\Pi_{k(m)}\right)\right) \cdot \mu_q^{(m)}.$$

Notice that the measure $\mu_q^{(m)}$ has to be invariant with regard to T because μ and $\mu_p^{(m)}$ are T-invariant. Additionally, Lemma 3.41 provides that $H_{\mu_q^{(m)}}(\mathcal{M}) < \infty$ holds true for all $m \in \mathbb{N}$. So we can use Corollary 3.23, which implies

$$\overline{\mathrm{PE}}_{\mu_q^{(m)}}(T) \leq h_{\mu_q^{(m)}}(T, \mathcal{M}) + \log(2). \tag{3.44}$$

Now take $\varepsilon > 0$. We have $h_\mu(T, \mathcal{M}) < H_\mu(\mathcal{M}) < \infty$, so there exists a finite partition $\mathcal{P}$ with $h_\mu(T, \mathcal{M}) \leq h_\mu(T, \mathcal{P}) + \varepsilon$. Using Lemma 3.29 and 3.40 yields

$$\begin{aligned} &\left(1 - \mu\left(\Pi_{k(m)}\right)\right) \cdot h_{\mu_q^{(m)}}(T, \mathcal{M}) \\ =&h_\mu(T, \mathcal{M}) - \mu\left(\Pi_{k(m)}\right) \cdot h_{\mu_p^{(m)}}(T, \mathcal{M}) = h_\mu(T, \mathcal{M}) \leq h_\mu(T, \mathcal{P}) + \varepsilon \\ =&\mu\left(\Pi_{k(m)}\right) \cdot h_{\mu_p^{(m)}}(T, \mathcal{P}) + \left(1 - \mu\left(\Pi_{k(m)}\right)\right) \cdot h_{\mu_q^{(m)}}(T, \mathcal{P}) + \varepsilon \\ =&\left(1 - \mu\left(\Pi_{k(m)}\right)\right) \cdot h_{\mu_q^{(m)}}(T, \mathcal{P}) + \varepsilon \\ \leq&\frac{1}{m} \cdot h_{\mu_q^{(m)}}(T, \mathcal{P}) + \varepsilon \\ \leq&\frac{1}{m} \cdot \log \#(\mathcal{P}) + \varepsilon. \end{aligned} \tag{3.45}$$

Consequently, Lemma 3.29 provides

$$\begin{aligned}
\overline{\mathrm{PE}}_\mu(T) \leq& \limsup_{n\to\infty} \frac{1}{n} \Big[\mu\left(\Pi_{k(m)}\right) \cdot H_{\mu_p^{(m)}}(OP(n)) + \left(1-\mu\left(\Pi_{k(m)}\right)\right) \cdot H_{\mu_q^{(m)}}(OP(n)) \\
&- \mu\left(\Pi_{k(m)}\right) \log\left(\mu\left(\Pi_{k(m)}\right)\right) - \left(1-\mu\left(\Pi_{k(m)}\right)\right) \log\left(\left(1-\mu\left(\Pi_{k(m)}\right)\right)\right)\Big] \\
\leq& \mu\left(\Pi_{k(m)}\right) \overline{\mathrm{PE}}_{\mu_p^{(m)}}(T) + \left(1-\mu\left(\Pi_{k(m)}\right)\right) \overline{\mathrm{PE}}_{\mu_q^{(m)}}(T) \\
\overset{(3.43)}{=}& \left(1-\mu\left(\Pi_{k(m)}\right)\right) \limsup_{n\to\infty} \frac{1}{n} H_{\mu_q^{(m)}}(OP(n)) \\
\overset{(3.44)}{\leq}& \left(1-\mu_p\left(\Pi_{k(m)}\right)\right) \cdot \left(h_{\mu_q^{(m)}}(T,\mathcal{M}) + \log(2)\right) \\
\overset{(3.45)}{\leq}& \frac{1}{m} \cdot (\log(\#P) + \log(2)) + \varepsilon
\end{aligned}$$

for all $m \in \mathbb{N}$ and $\varepsilon > 0$. This implies

$$\overline{\mathrm{PE}}(T) = \overline{\mathrm{PE}}_\mu(T) \leq \lim_{m\to\infty} \frac{1}{m} \cdot (\log(\#P) + \log(2)) + \varepsilon = \varepsilon \tag{3.46}$$

and thus

$$\overline{\mathrm{PE}}(T) = 0$$

because ε can be chosen arbitrarily close to 0. □

Proof of Theorem 3.17. If $\mu(\Pi) = 1$ holds true, the proof of this Theorem directly follows from Lemma 3.43. For $\mu(\Pi) = 0$, the proof of this Theorem directly follows from Theorem 3.38. So it remains to consider the case $\mu(\Pi) \in]0,1[$. Define the measure μ_p on the periodic part of the system as

$$\mu_p(A) := \frac{\mu(A \cap \Pi)}{\mu(\Pi)}$$

for all $A \in \mathscr{B}$ and the measure μ_a on the aperiodic part as

$$\mu_a(A) := \frac{\mu(A \setminus \Pi)}{1-\mu(\Pi)}$$

for all $A \in \mathscr{B}$. Lemma 3.42 provides that μ_a and μ_p are invariant with regard to T. So we can apply Theorem 3.38, which implies

$$\underline{\mathrm{PE}}_{\mu_a}(T) = \overline{\mathrm{PE}}_{\mu_a}(T) = h_{\mu_a}(T) \tag{3.47}$$

because $H_{\mu_a}(\mathcal{M}) < \infty$ holds true due to Lemma 3.41. Additionally, Lemma 3.43 provides

$$\underline{\mathrm{PE}}_{\mu_p}(T) = \overline{\mathrm{PE}}_{\mu_p}(T) = h_{\mu_p}(T) = 0. \tag{3.48}$$

According to Lemma 3.29, we have

$$\begin{aligned}
\overline{\mathrm{PE}}_\mu(T) =& (1-\mu(\Pi)) \cdot \overline{\mathrm{PE}}_{\mu_a}(T) + \mu(\Pi) \cdot \overline{\mathrm{PE}}_{\mu_p}(T) \\
\overset{(3.48)}{=}& (1-\mu(\Pi)) \cdot \overline{\mathrm{PE}}_{\mu_a}(T) \\
\overset{(3.47)}{=}& (1-\mu(\Pi)) \cdot h_{\mu_a}(T) \\
\overset{(3.48)}{=}& \mu(\Pi) \cdot h_{\mu_p}(T) + (1-\mu(\Pi)) \cdot h_{\mu_a}(T) \\
=& h_\mu(T).
\end{aligned}$$

Analogously, one can show

$$\underline{\mathrm{PE}}_\mu(T) = h(T).$$

□

3.4.8 Permutation entropy of piecewise monotone functions using Rényi entropy

In this section we show that the equality $\overline{\mathrm{PE}}(T) = h(T)$ for piecewise monotone functions does not hold true in general when using Rényi entropy instead of Shannon entropy. To achieve this we consider one-sided Bernoulli shifts. Remember that the Kolmogorov-Sinai entropy $h(T)$ of a Bernoulli shift T generated by a stochastic vector p is equal to $H(p)$. We will make use of the following fact about Rényi entropies of stochastic vectors:

Lemma 3.44. For all $q > 1$ and all $c > 0$ there exists a stochastic vector p with

$$H(p) - H(p, q) \geq c.$$

Proof. Let $c > 0$ and $q > 1$. Choose $N \in \mathbb{N}$ such that

$$N \geq 2 \cdot \left(\frac{c}{\log(2)} + \frac{q}{q-1} \right)$$

holds true. Consider the stochastic vector $p = (p_1, p_2, \ldots, p_{2^N+1}) \in [0,1]^{2^N+1}$ with

$$p_1 = \frac{1}{2}$$

and

$$p_i = \frac{1}{2^{N+1}} \qquad \text{for } i \in \{2, 3, \ldots, 2^N + 1\}.$$

Then

$$\begin{aligned} H(p) &= -\sum_{i=1}^{2^N+1} p_i \log(p_i) \\ &= -2^N \cdot \frac{1}{2^{N+1}} \log\left(\frac{1}{2^{N+1}}\right) - \frac{1}{2} \log\left(\frac{1}{2}\right) \\ &= \frac{N+1}{2} \log(2) + \frac{1}{2} \log(2) \\ &\geq \frac{N}{2} \log(2) \end{aligned}$$

and

$$\begin{aligned} H(p, q) &= \frac{-1}{q-1} \log\left(\sum_{i=1}^{2^N+1} p_i^q \right) \\ &= \frac{-1}{q-1} \log\left(2^N \cdot \frac{1}{2^{(N+1)q}} + \frac{1}{2^q} \right) \\ &\leq \frac{-1}{q-1} \log\left(\frac{1}{2^q}\right) \\ &= \frac{q}{q-1} \log(2). \end{aligned}$$

Hence,

$$H(p) - H(p,q) \geq \log(2) \cdot \left(\frac{N}{2} - \frac{q}{q-1}\right) \geq c. \qquad \square$$

Example 5. Take any $q > 1$. Choose $N \in \mathbb{N}$ and a stochastic vector $p = (p_1, p_2, \dots, p_N) \in [0,1]^N$ such that

$$H(p,q) + \log 2 < H(p) \tag{3.49}$$

holds true. This is always possible due to Lemma 3.44. Let $([0,1[, \mathscr{B}, T, \mu)$ be the Bernoulli shift generated by p. Since T is aperiodic and ergodic (see for example [7]), Theorem 3.1 implies

$$h(T) = h(T,q) \tag{3.50}$$

Let $\mathcal{M} = \{[(i-1)/N, i/N[\mid i \in \{1,2,\dots,N\}\}$ be the partition into monotone parts of T. Then

$$\mathrm{PE}(T,q) \overset{Cor.3.23}{\leq} h(T,\mathcal{M},q) + \log 2 \overset{Lem.2.29}{=} H(p,q) + \log 2 \overset{(3.49)}{<} H(p) \overset{(3.8)}{=} h(T) \overset{(3.50)}{=} h(T,q)$$

holds true.

3.5 Conditional permutation entropy

As stated in Lemma 3.31, the entropy rate of a partition can be calculated by taking the limit of a sequence of conditional entropies instead. More precisely, given a finite partition $\mathcal{P}$, we have

$$\lim_{n\to\infty} \frac{1}{n} H(\mathcal{P}^{(n)}) = \lim_{n\to\infty} H(\mathcal{P}|T^{-1}(\mathcal{P}^{(n)})) = \lim_{n\to\infty} H(\mathcal{P}^{(n)}|T^{-1}(\mathcal{P}^{(n)})) = \lim_{n\to\infty} H(\mathcal{P}^{(n+1)}|\mathcal{P}^{(n)}) \tag{3.51}$$

The first equality directly follows from Lemma 3.31. The second and third equality follow from

$$\begin{aligned}
H(\mathcal{P}|T^{-1}(\mathcal{P}^{(n)})) &\leq H(\mathcal{P}^{(n)}|T^{-1}(\mathcal{P}^{(n)})) \\
&\leq H(\mathcal{P}^{(n+1)}|\mathcal{P}^{(n)}) \\
&= H(\mathcal{P}^{(n+1)}) - H(\mathcal{P}^{(n)}) \\
&= H(\mathcal{P}^{(n+1)}) - H(T^{-1}(\mathcal{P}^{(n)})) \\
&= H(\mathcal{P}^{(n+1)}|T^{-1}(\mathcal{P}^{(n)})) \\
&= H(\mathcal{P} \vee T^{-1}(\mathcal{P}^{(n)})|T^{-1}(\mathcal{P}^{(n)})) \\
&\leq H(\mathcal{P}|T^{-1}(\mathcal{P}^{(n)})) + H(T^{-1}(\mathcal{P}^{(n)})|T^{-1}(\mathcal{P}^{(n)})) \\
&= H(\mathcal{P}|T^{-1}(\mathcal{P}^{(n)})).
\end{aligned}$$

Analogously to (3.51), one can define conditional variants of the permutation entropy, simply by replacing the partition into symbolic patterns $\mathcal{P}^{(n)}$ by the partition into ordinal patterns $OP(n)$. This provides the quantities

$$H(OP(n+1)|OP(n)) \qquad \text{and} \qquad H(OP(n)|T^{-1}(OP(n))).$$

Unlike in (3.51), it is not immediately obvious whether these quantities converges to the original permutation entropy $\mathrm{PE}(T)$ or even to the same value. This will be investigated in this section.

As observed in [37], simulations indicate that the conditional permutation entropy converges faster than the standard permutation entropy. It is not clear why this is the case on a theoretical level but one can get an idea by considering the following situation:

Let $(a_n)_{n\in\mathbb{N}}$ be an monotonically decreasing and bounded sequence of real numbers. Then there exists $a \in \mathbb{R}$ with $\lim_{n\to\infty} a_n = a$. Set

$$b_n := \frac{1}{n}\sum_{i=1}^{n} a_n$$

According to the Stolz-Cesàro theorem, $(b_n)_{n\in\mathbb{N}}$ also converges to a. However, the speed of convergence is typically different. The monotony of $(a_n)_{n\in\mathbb{N}}$ provides

$$|b_n - a| = \left|\frac{1}{n}\sum_{i=1}^{n} a_i - a\right| = \frac{1}{n}\sum_{i=1}^{n} |a_i - a| \geq \frac{1}{n}\sum_{i=1}^{n} |a_n - a| = |a_n - a|\,. \tag{3.52}$$

Therefore, $(a_n)_{n\in\mathbb{N}}$ converges at least as fast to a as $(b_n)_{n\in\mathbb{N}}$ does.

Now let $\mathcal{P}$ be a finite partition of Ω and set $\mathcal{P}^{(0)} := \{\Omega\}$. By choosing

$$a_n := H(\mathcal{P}^{(n)}|\mathcal{P}^{(n-1)})$$

for $n \in \mathbb{N}$, it follows form (3.52) that the sequence of conditional entropies converges at least as fast to $h(T, \mathcal{P})$ as

$$\begin{aligned} b_n &= \frac{1}{n}\sum_{i=0}^{n} a_i = \frac{1}{n}\sum_{i=0}^{n} H(\mathcal{P}^{(i)}|\mathcal{P}^{(i-1)}) \\ &= \frac{1}{n}\sum_{i=0}^{n} H(\mathcal{P}^{(i)}) - H(\mathcal{P}^{(i-1)}) = \frac{1}{n}H(\mathcal{P}^{(n)}) - H(\mathcal{P}^{(0)}) = \frac{1}{n}H(\mathcal{P}^{(n)}). \end{aligned}$$

Notice that this is only true because $H(\mathcal{P}^{(n)}|\mathcal{P}^{(n-1)}) = H(\mathcal{P}|T^{-1}(\mathcal{P}^{(n-1)}))$ is monotonically decreasing in $n \in \mathbb{N}$.

Analogously, one could try to argue that $H(OP(n)|OP(n-1))$ converges faster to $\mathrm{PE}(T)$ than $\frac{1}{n}H(OP(n))$. However, $H(OP(n)|OP(n-1))$ is typically not monotonically decreasing in n and we do not even know whether $H(OP(n)|OP(n-1))$ converges in general. So we can not theoretically guarantee that the conditional permutation will converge faster. Nevertheless, simulations show that it typically still does. This could indicate that $(H(OP(n)|OP(n-1)))_{n\in\mathbb{N}}$ does not differ that much from a monotone sequence.

The faster convergence speed of the conditional permutation entropy makes it useful for practical applications. Therefore, it is interesting to analyze the relation between the conditional permutation entropy and the standard permutation entropy.

The following three statements are a generalization of a statement given in [19].

Lemma 3.45. Let $(\Omega, \mathscr{A}, \mu, T)$ be a measure-preserving dynamical system and $(\mathcal{P}_n)_{n\in\mathbb{N}}$ a sequence of finite partitions of Ω satisfying

$$\mathcal{P}_n \vee T^{-1}(\mathcal{P}_n) \prec \mathcal{P}_{n+1}. \tag{3.53}$$

Then

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

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

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

$$\begin{aligned}
&\liminf_{n\to\infty} H\left(\mathcal{P}_n \,\middle|\, T^{-1}\left(\mathcal{P}_n^{(k)}\right)\right) = \liminf_{n\to\infty} \left[H\left(\mathcal{P}_n^{(k+1)}\right) - H\left(T^{-1}\left(\mathcal{P}_n^{(k)}\right)\right)\right] \\
= \ &\liminf_{n\to\infty} \left[H\left(\mathcal{P}_n^{(k+1)}\right) - H\left(\mathcal{P}_n^{(k)}\right)\right] \overset{(3.53)}{\leq} \liminf_{n\to\infty} \left[H\left(\mathcal{P}_{n+1}^{(k)}\right) - H\left(\mathcal{P}_n^{(k)}\right)\right] \\
= \ &\liminf_{n\to\infty} H\left(\mathcal{P}_{n+1}^{(k)} \,\middle|\, \mathcal{P}_n^{(k)}\right).
\end{aligned}$$

The Stolz-Cesàro theorem further provides

$$\begin{aligned}
&\liminf_{n\to\infty} H\left(\mathcal{P}_{n+1}^{(k)} \,\middle|\, \mathcal{P}_n^{(k)}\right) \leq \liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \left[H\left(\mathcal{P}_{i+1}^{(k)} \,\middle|\, \mathcal{P}_i^{(k)}\right)\right] \\
= \ &\liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \left[H\left(\mathcal{P}_{i+1}^{(k)}\right) - H\left(\mathcal{P}_i^{(k)}\right)\right] = \liminf_{n\to\infty} \frac{1}{n} \left[H\left(\mathcal{P}_{n+1}^{(k)}\right) - H\left(\mathcal{P}_1^{(k)}\right)\right] \\
= \ &\liminf_{n\to\infty} \frac{1}{n} H\left(\mathcal{P}_{n+1}^{(k)}\right) \overset{(3.53)}{\leq} \liminf_{n\to\infty} \frac{1}{n} H(\mathcal{P}_{n+k}) = \liminf_{n\to\infty} \frac{1}{n} H(\mathcal{P}_n).
\end{aligned}$$

□

Notice that (3.53) is fulfilled for $\mathcal{P}_n := OP^{\boldsymbol{X}}(n)$.

Lemma 3.46. Let $(\Omega, \mathscr{A}, \mu, T)$ be a measure-preserving dynamical system and $\boldsymbol{X} = (X_1, X_2, \ldots, X_d) : \Omega \to \mathbb{R}^d$ a vector of random variables satisfying

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

Then

$$h(T) \leq \liminf_{n\to\infty} H\left(OP^{\boldsymbol{X}}(n) \,\middle|\, T^{-1}\left((OP^{\boldsymbol{X}}(n))^{(k)}\right)\right)$$

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

Proof. According to Theorem 3.11, we have

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

Using the future formula for the entropy rate (see e.g [39]), we can write

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

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

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

for all $k \in \mathbb{N}$. □

Theorem 3.47. Let $(\Omega, \mathscr{A}, \mu, T)$ be a measure-preserving dynamical system and $\boldsymbol{X} = (X_1, X_2, \ldots, X_d) : \Omega \to \mathbb{R}^d$ a vector of random variables satisfying

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

If $h(T) \geq \overline{\mathrm{PE}}(\boldsymbol{X}, T)$ is true, then

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

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

Proof. Lemma 3.45 and 3.46 provide

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

The assumption $h(T) \geq \overline{\mathrm{PE}}(\mathbf{X}, T)$ then implies

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

for all $k \in \mathbb{N}$. □

The above theorem implies that the conditional permutation entropy is equal to the standard permutation entropy for piecewise monotone functions T.

3.5.1 No-in-between sets V_n

Theorem 3.47 in particular implies that the two different variants of conditional permutation entropy are equal for piecewise monotone functions. We want to investigate whether this equality is true for more general one-dimensional systems. To achieve this, we consider the sets

$$\begin{aligned} V_n :=& \{\omega \in \Omega \mid \omega \leq T^n(\omega) \text{ and } T^s(\omega) \notin (\omega, T^n(\omega)) \text{ for all } s \in \{1, 2, \ldots, n-1\}\} \\ &\cup \{\omega \in \Omega \mid T^n(\omega) \leq \omega \text{ and } T^s(\omega) \notin (T^n(\omega), \omega) \text{ for all } s \in \{1, 2, \ldots, n-1\}\} \end{aligned} \tag{3.54}$$

for $n \in \mathbb{N}$. They contain all points $\omega \in \Omega$ for which no point in the set $\{T(\omega), T^2(\omega), \ldots, T^{n-1}(\omega)\}$ is located between ω and $T^n(\omega)$. These sets were first considered in [38]. They can be used to determine the relationship between different kind of entropies based on ordinal patterns. The following lemma gives the main reason why this is possible.

Lemma 3.48. Let $(\Omega, \mathscr{B}, \mu, T)$ be a measure-preserving dynamical system with $\Omega \subseteq \mathbb{R}$. Then

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

holds true for all $n \in \mathbb{N}$, where V_n is defined as in (3.54).

Proof. Fix some $n \in \mathbb{N}$. The set V_n can be written as a union of sets in $OP(n) \vee T^{-1}(OP(n))$ because

$$V_n = \bigcup_{(A_1,A_2,\ldots,A_{n-1}) \in \{R,\Omega^2 \setminus R\}^{n-1}} \underbrace{\bigcap_{s=1}^{n-1} (\mathrm{id}, T^s)^{-1}(A_s)}_{\in \sigma(OP(n))} \cap \underbrace{\bigcap_{s=1}^{n-1} (T^s, T^n)^{-1}(\Omega^2 \setminus A_s)}_{\in \sigma(T^{-1}(OP(n)))} \tag{3.55}$$

holds true. Notice that

$$OP(n+1) = OP(n) \vee T^{-1}(OP(n)) \vee (\mathrm{id}, T^n)^{-1}(\{R, \Omega^2 \setminus R\}). \tag{3.56}$$

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

$$\#\Delta((\mathrm{id}, T^n)^{-1}(\{R, \Omega^2 \setminus R\})|Q) = 1 \tag{3.57}$$

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

$$\begin{aligned}
&H(OP(n+1)|OP(n) \vee T^{-1}(OP(n))) \\
&\overset{(3.56)}{\leq} H(OP(n) \vee T^{-1}(OP(n))|OP(n) \vee T^{-1}(OP(n))) \\
&\quad + H((\mathrm{id}, T^n)^{-1}(\{R, \Omega^2 \setminus R\})|OP(n) \vee T^{-1}(OP(n))) \\
&= H((\mathrm{id}, T^n)^{-1}(\{R, \Omega^2 \setminus R\})|OP(n) \vee T^{-1}(OP(n))) \\
&\overset{\textit{Lem. } 3.3}{\leq} \sum_{Q \in OP(n) \vee T^{-1}(OP(n))} \mu(Q) \cdot \log(\#\Delta((\mathrm{id}, T^n)^{-1}(\{R, \Omega^2 \setminus R\})|Q)) \\
&= \sum_{\substack{Q \in OP(n) \vee T^{-1}(OP(n)) \\ Q \subseteq V_n}} \mu(Q) \cdot \log(\#\Delta((\mathrm{id}, T^n)^{-1}(\{R, \Omega^2 \setminus R\})|Q)) \\
&\quad + \sum_{\substack{Q \in OP(n) \vee T^{-1}(OP(n)) \\ Q \not\subseteq V_n}} \mu(Q) \cdot \log(\#\Delta((\mathrm{id}, T^n)^{-1}(\{R, \Omega^2 \setminus R\})|Q))
\end{aligned}$$

$$
\begin{aligned}
&\overset{(3.55)}{=} \sum_{\substack{Q \in OP(n) \vee T^{-1}(OP(n)) \\ Q \subseteq V_n}} \mu(Q) \cdot \log(\#\Delta((\mathrm{id}, T^n)^{-1}(\{R, \Omega^2 \setminus R\})|Q)) \\
&\quad + \sum_{\substack{Q \in OP(n) \vee T^{-1}(OP(n)) \\ Q \subseteq \Omega \setminus V_n}} \mu(Q) \cdot \log(\#\Delta((\mathrm{id}, T^n)^{-1}(\{R, \Omega^2 \setminus R\})|Q)) \\
&\overset{(3.57)}{=} \sum_{\substack{Q \in OP(n) \vee T^{-1}(OP(n)) \\ Q \subseteq V_n}} \mu(Q) \cdot \log(\#\Delta((\mathrm{id}, T^n)^{-1}(\{R, \Omega^2 \setminus R\})|Q)) \\
&\leq \sum_{\substack{Q \in OP(n) \vee T^{-1}(OP(n)) \\ Q \subseteq V_n}} \mu(Q) \cdot \log(2) \\
&= \log(2) \cdot \mu(V_n).
\end{aligned}
$$

□

Lemma 3.49. Let $(\Omega, \mathscr{B}, \mu, T)$ be a measure-preserving dynamical system and $V_n \in \mathscr{B}$ as defined in (3.54). If $\lim_{n\to\infty} \mu(V_n) = 0$, then

$$\lim_{n\to\infty} \left[H\left(OP(n+1)|OP(n)\right) - H\left(OP(n)|T^{-1}(OP(n))\right) \right] = 0$$

holds true.

Proof. Since $T^{-1}(OP(n)) \prec OP(n)$, we have

$$H(OP(n+1)|OP(n)) \geq H(T^{-1}(OP(n))|OP(n)) = H\left(OP(n)|T^{-1}(OP(n))\right)$$

for all $n \in \mathbb{N}$. So it remains to show that

$$\lim_{n\to\infty} \left[H\left(OP(n+1)|OP(n)\right) - H\left(OP(n)|T^{-1}(OP(n))\right) \right] \leq 0 \tag{3.58}$$

holds true:

$$
\begin{aligned}
&H\left(OP(n+1)|OP(n)\right) - H\left(OP(n)|T^{-1}(OP(n))\right) \\
&= H(OP(n+1)) - H(OP(n)) - H(OP(n) \vee T^{-1}(OP(n))) + H(T^{-1}(OP(n))) \\
&= H(OP(n+1)) - H(OP(n)) - H(OP(n) \vee T^{-1}(OP(n))) + H(OP(n)) \\
&= H(OP(n+1)) - H(OP(n) \vee T^{-1}(OP(n))) = H\left(OP(n+1)|OP(n) \vee T^{-1}(OP(n)\right) \\
&\overset{Lem.3.48}{\leq} \log(2)\mu(V_n).
\end{aligned}
$$

Therefore, (3.58) follows from $\lim_{n\to\infty} \mu(V_n) = 0$. □

So, if $\mu(V_n)$ converges to 0 for $n \to \infty$, the two types of conditional entropy are equal. As the following lemma will show, $\lim_{n\to\infty} \mu(V_n) = 0$ holds true if, roughly speaking, the dependence between ω and $T^n(\omega)$ decreases to 0 for $n \to \infty$. This is, for example true if the considered system is mixing. A measure-preserving dynamical system $(\Omega, \mathscr{A}, \mu, T)$ is called mixing, if

$$\lim_{n\to\infty} \mu(A \cap T^{-n}(B)) = \mu(A) \cdot \mu(B) \tag{3.59}$$

holds true for all $A, B \in \mathscr{A}$.

Lemma 3.50. Let $(\Omega, \mathscr{B}, \mu, T)$ be an aperiodic and mixing measure-preserving dynamical system with $\Omega \subseteq \mathbb{R}$. If T is ergodic, then

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

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

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

holds true, where V_n is defined as in (3.54).

A proof for the 'mixing part' of the above lemma can be found in [38]. We give here an alternative proof, that includes a statement for the weaker ergodic case.

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

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

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

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

holds true.

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

$$\begin{aligned}
&\mu\left(V_t \cap \bigcap_{i=1}^{N} \bigcup_{s=1}^{n_0-1} T^{-s}(U_i)\right) \\
&\leq \mu(U_1 \cap T^{-t}(U_1 \cup U_2)) + \sum_{i=2}^{N-1} \mu(U_i \cap T^{-t}(U_{i-1} \cup U_i \cup U_{i+1})) + \mu(U_N \cap T^{-t}(U_{N-1} \cup U_N))
\end{aligned}$$

holds true for all $t \in \mathbb{N}$ with $t > n_0$. If T is mixing, this implies

$$\begin{aligned}
&\lim_{n\to\infty} \mu\left(V_n \cap \bigcap_{i=1}^{N} \bigcup_{s=1}^{n_0-1} T^{-s}(U_i)\right) \\
&\leq \mu(U_1)\mu(U_1 \cup U_2) + \sum_{i=2}^{N-1} \mu(U_i)\mu(U_{i-1} \cup U_i \cup U_{i+1}) + \mu(U_N)\mu(U_{N-1} \cup U_N) \\
&\leq \mu(U_1)2\varepsilon + \sum_{i=2}^{N-1} \mu(U_i)3\varepsilon + \mu(U_N)2\varepsilon \\
&\leq 3\varepsilon.
\end{aligned}$$

If T is not necessarily mixing but ergodic, we have

$$
\begin{aligned}
&\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mu\left(V_t\cap\bigcap_{i=1}^{N}\bigcup_{s=1}^{n_0-1}T^{-s}(U_i)\right)=\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mu\left(V_{t+n_0}\cap\bigcap_{i=1}^{N}\bigcup_{s=1}^{n_0-1}T^{-s}(U_i)\right)\\
&\leq \mu(U_1)\mu(U_1\cup U_2)+\sum_{i=2}^{N-1}\mu(U_i)\mu(U_{i-1}\cup U_i\cup U_{i+1})+\mu(U_N)\mu(U_{N-1}\cup U_N)\\
&\leq \mu(U_1)2\varepsilon+\sum_{i=2}^{N-1}\mu(U_i)3\varepsilon+\mu(U_N)2\varepsilon\\
&\leq 3\varepsilon.
\end{aligned}
$$

The Stolz-Cesàro theorem then provides

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

Hence

$$
\begin{aligned}
&\liminf_{n\to\infty}\mu\left(V_n\right)\\
&=\liminf_{n\to\infty}\left[\mu\left(V_n\cap\bigcap_{i=1}^{N}\bigcup_{s=1}^{n_0-1}T^{-s}(U_i)\right)+\mu\left(V_n\setminus\bigcap_{i=1}^{N}\bigcup_{s=1}^{n_0-1}T^{-s}(U_i)\right)\right]\\
&\leq\liminf_{n\to\infty}\mu\left(V_n\cap\bigcap_{i=1}^{N}\bigcup_{s=1}^{n_0-1}T^{-s}(U_i)\right)+\varepsilon\\
&\leq 4\varepsilon.
\end{aligned}
$$

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

After showing that $\liminf_{n\to\infty}\mu(V_n)=0$ holds true for ergodic systems, one might try to show that $\limsup_{n\to\infty}\mu(V_n)=0$ holds true as well. However, this is in general not the case as illustrated by the following example:

Example 6. Consider the rotation map $T:[0,1[\to[0,1[$ with $T(\omega)=\omega+a \mod 1$ for some irrational $a\in\mathbb{R}\setminus\mathbb{Q}$. One can show that the measure-preserving dynamical system $([0,1[,\mathscr{B},\lambda,T)$ is ergodic, where λ denotes the Lebesgue measure (see e.g. [7]). And in particular,

$$
\limsup_{d\to\infty}\lambda(V_d)\geq\frac{1}{2}
$$

holds true.

Proof. Define

$$
d_0(\omega):=\min\left\{k\in\mathbb{N}\mid\omega<T^k(\omega)<\omega+\frac{1}{2}\right\}.
$$

and, inductively,

$$
d_n(\omega):=\min\{k\in\mathbb{N}\mid\omega<T^k(\omega)<T^{d_{n-1}(\omega)}(\omega)\}.
$$

for all $\omega \in [0,1[$. Notice that $d_{n+1}(\omega) > d_n(\omega)$ holds true for all $n \in \mathbb{N}_0$ and $\omega \in [0,1[$. Since T is ergodic, the orbit $\{\omega, T(\omega), T^2(\omega), \ldots\}$ is dense in $[0,1[$ for (almost) all $\omega \in [0,1[$. Hence, $d_n(\omega)$ is well defined for all $n \in \mathbb{N}_0$ and $\omega \in \Omega$. Further,

$$d_n(\omega) = d_n(\omega')$$

for all $n \in \mathbb{N}_0$ and $\omega, \omega' \in [0, \frac{1}{2}[$. This implies

$$d_n := d_n\left(\frac{1}{4}\right) = d_n(\omega)$$

for all $\omega \in [0, \frac{1}{2}[$. Therefore,

$$\begin{aligned}
&\lambda\left(V_{d_n}\right) \geq \lambda\left(V_{d_n} \cap \left[0, \frac{1}{2}\right[\right) \\
&\geq \lambda\left(\left\{\omega \in \Omega \mid \omega < T^{d_n}(\omega), T^l(\omega) \notin (\omega, T^{d_n}(\omega)) \text{ for all } l \in \{1, 2, ..., d_n - 1\}\right\} \cap \left[0, \frac{1}{2}\right[\right) \\
&= \lambda\left(\left[0, \frac{1}{2}\right[\right) = \frac{1}{2}
\end{aligned}$$

for all $n \in \mathbb{N}_0$. Hence,

$$\limsup_{d\to\infty} \lambda(V_d) \geq \limsup_{n\to\infty} \lambda(V_{d_n}) = \frac{1}{2}.$$ □

Using Lemma 3.48, one can establish a relation between the permutation entropy and the Kolmogorov-Sinai entropy bases on ordinal patterns. This was first done in [38] and originally motivated the consideration of the sets V_n. We give here a different proof of the same statement, which is directly based on the properties of the conditional entropy.

Lemma 3.51. Let $(\Omega, \mathscr{B}, \mu, T)$ be a measure-preserving dynamical system with $\Omega \subseteq \mathbb{R}$. Then

$$\sum_{n=1}^{\infty} H\left(OP(n) \,\middle|\, OP(n-1) \vee T^{-1}(OP(n-1))\right) < \infty \tag{3.60}$$

implies

$$\overline{\mathrm{PE}}(T) = h(T).$$

Proof. Using Theorem 3.8, the statement of this theorem is equivalent to

$$\overline{\mathrm{PE}}(T) = \lim_{k\to\infty} h(OP(k)).$$

It follows from Lemma 3.12 that

$$\overline{\mathrm{PE}}(T) \geq \lim_{k\to\infty} h(OP(k))$$

holds true. So it remains to show

$$\overline{\mathrm{PE}}(T) \leq \lim_{k\to\infty} h(OP(k)).$$

The equality

$$OP(n) = \bigvee_{s=1}^{n} \bigvee_{t=0}^{n-s} T^{-t}(OP(s)) = \bigvee_{s=1}^{n} (OP(s))^{(n-s+1)}$$

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

$$\begin{aligned}
&H\left(OP(n)\,\middle|\,OP(k)^{(n)}\right)\\
&= H\left(\bigvee_{s=1}^{n}(OP(s))^{(n-s+1)}\,\middle|\,OP(k)^{(n)}\right)\\
&= H\left(\bigvee_{s=1}^{n}(OP(s))^{(n-s+1)}\,\middle|\,OP(k)^{(n)}\right)\\
&= H\left(\bigvee_{s=1}^{n}(OP(s))^{(n-s+1)} \vee OP(k)^{(n)}\right) - H\left(OP(k)^{(n)}\right)\\
&= \sum_{i=1}^{n} H\left(\bigvee_{s=1}^{i}(OP(s))^{(n-s+1)} \vee OP(k)^{(n)}\right) - H\left(\bigvee_{s=1}^{i-1}(OP(s))^{(n-s+1)} \vee OP(k)^{(n)}\right)\\
&= \sum_{i=1}^{n} H\left(\bigvee_{s=1}^{i}(OP(s))^{(n-s+1)}\,\middle|\,\bigvee_{s=1}^{i-1}(OP(s))^{(n-s+1)} \vee OP(k)^{(n)}\right)\\
&= \sum_{i=1}^{n} H\left((OP(i))^{(n-i+1)}\,\middle|\,(OP(i-1))^{(n-i+2)} \vee OP(k)^{(n)}\right)\\
&\leq \sum_{i=k+1}^{n} H\left((OP(i))^{(n-i+1)}\,\middle|\,(OP(i-1))^{(n-i+2)}\right)\\
&= \sum_{i=k+1}^{n} H\left(\bigvee_{t=0}^{n-i} T^{-t}(OP(i))\,\middle|\,\bigvee_{s=0}^{n-i+1} T^{-s}(OP(i-1))\right)\\
&\leq \sum_{i=k+1}^{n} \sum_{t=0}^{n-i} H\left(T^{-t}(OP(i))\,\middle|\,\bigvee_{s=0}^{n-i+1} T^{-s}(OP(i-1))\right)\\
&\leq \sum_{i=k+1}^{n} \sum_{t=0}^{n-i} H\left(T^{-t}(OP(i))\,\middle|\,T^{-t}\left(OP(i-1) \vee T^{-1}(OP(i-1))\right)\right)\\
&= \sum_{i=k+1}^{n} (n-i+1) \cdot H\left(OP(i)\,\middle|\,OP(i-1) \vee T^{-1}(OP(i-1))\right)\\
&\leq n \cdot \sum_{i=k+1}^{n} H\left(OP(i)\,\middle|\,OP(i-1) \vee T^{-1}(OP(i-1))\right)
\end{aligned}$$

for all $k \leq n$. Hence,

$$\begin{aligned}
\overline{\mathrm{PE}}(T) &\leq \lim_{k\to\infty} \limsup_{n\to\infty} \frac{1}{n} H\left(OP(n) \vee OP(k)^{(n)}\right) \\
&\leq \lim_{k\to\infty} \left[h(OP(k)) + \limsup_{n\to\infty} \frac{1}{n} H\left(OP(n) \middle| OP(k)^{(n)}\right)\right] \\
&\leq \lim_{k\to\infty} \left[h(OP(k)) + \limsup_{n\to\infty} \sum_{i=k+1}^{n} H\left(OP(i) \middle| OP(i-1) \vee T^{-1}(OP(i-1))\right)\right] \\
&\leq \lim_{k\to\infty} h(OP(k)) + \lim_{k\to\infty} \sum_{i=k+1}^{\infty} H\left(OP(i) \middle| OP(i-1) \vee T^{-1}(OP(i-1))\right) \\
&\overset{(3.60)}{=} \lim_{k\to\infty} h(OP(k)).
\end{aligned}$$

□

In combination with Lemma 3.48, the above lemma implies that the permutation entropy is equal to the Kolmogorov-Sinai entropy if

$$\sum_{n=1}^{\infty} \mu(V_n) \tag{3.61}$$

holds true. However, as we will show in the following lemma, this is impossible for basically every interesting dynamical system. In particular, this implies that the statement in Lemma 3.51 is not an equivalence.

Lemma 3.52. Let $(\Omega, \mathscr{B}, \mu, T)$ be an aperiodic and ergodic measure-preserving dynamical system with $\Omega \subseteq \mathbb{R}$. Then

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

holds true, where V_n is defined as in (3.54).

Proof. Recall that

$$\Pi := \bigcup_{s\in\mathbb{N}} \{\omega \in \Omega \mid \omega = T^s(\omega)\}$$

is the set of periodic points. Since T is μ-almost surely aperiodic, we have $\mu(\Pi) = 0$.

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

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

or, equivalently,

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

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

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

Set

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

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

$$\begin{aligned} 0 < \mu\left(\bigcap_{n=K}^{\infty} (\Omega \setminus V_n) \setminus \Pi\right) &= \mu\left(\bigcup_{i=1}^{\infty}\left\{\omega \in \bigcap_{n=K}^{\infty} (\Omega \setminus V_n) \setminus \Pi \mid \delta(\omega) > 1/i\right\}\right) \\ &= \lim_{i\to\infty} \mu\left(\left\{\omega \in \bigcap_{n=K}^{\infty} (\Omega \setminus V_n) \setminus \Pi \mid \delta(\omega) > 1/i\right\}\right). \end{aligned}$$

So there exists some $\delta > 0$ such that

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

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

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

we have

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

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

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

and, consequently,

$$\begin{aligned} &\mu\left(A_\delta \cap (\omega_0 - \delta/2, \omega_0 + \delta/2) \cap \bigcup_{n=K}^{\infty} T^{-n}((\omega_0 - \delta/2, \omega_0 + \delta/2))\right) \\ &= \mu(A_\delta \cap (\omega_0 - \delta/2, \omega_0 + \delta/2)) > 0. \end{aligned}$$

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

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

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

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

holds true. This implies that

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

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

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

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

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

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

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

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

The above lemma still holds true if we drop the condition of ergodicity:

Corollary 3.53. Let $(\Omega, \mathscr{B}, \mu)$ be a Standard probability space with $\Omega \subseteq \mathbb{R}$ and $T : \Omega \to \Omega$ an aperiodic measure-preserving map. Then

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

holds true, where V_n is defined as in (3.54).

Proof. According to Theorem 3.28 and Lemma 3.32, there exists a decomposition

$$\mu = \int \mu_\omega^{\mathscr{E}}$$

of μ and a set $\Omega_0 \in \mathscr{B}$ with $\mu(\Omega_0) = 1$ such that T is measure-preserving, aperiodic and ergodic with respect to $\mu_\omega^{\mathscr{E}}$ for all $\omega \in \Omega_0$. The monotone convergence theorem and Lemma 3.52 then provide

$$\begin{aligned}
\lim_{N\to\infty} \sum_{n=1}^{N} \mu(V_n) &= \lim_{N\to\infty} \sum_{n=1}^{N} \int \mu_\omega^{\mathscr{E}}(V_n)\, \mathrm{d}\mu(\omega) \\
&= \lim_{N\to\infty} \int \sum_{n=1}^{N} \mu_\omega^{\mathscr{E}}(V_n)\, \mathrm{d}\mu(\omega) \\
&= \int \lim_{N\to\infty} \sum_{n=1}^{N} \mu_\omega^{\mathscr{E}}(V_n)\, \mathrm{d}\mu(\omega) \\
&\geq \int_{\Omega_0} \lim_{N\to\infty} \sum_{n=1}^{N} \mu_\omega^{\mathscr{E}}(V_n)\, \mathrm{d}\mu(\omega) \\
&= \int_{\Omega_0} \infty \, \mathrm{d}\mu(\omega) = \infty.
\end{aligned}$$

□

To summarize, we have shown with Lemma 3.52 that the summation condition (3.61) that was establish in [38] as a way to prove that the Kolmogorov-Sinai and permutation entropy are equal, can not be satisfied. Therefore, a different kind of summation condition would be necessary. But as of now, such a better condition has not been found.

4 Permutation entropy of piecewise expanding and mixing maps

The aim of this chapter is to find conditions for multidimensional measure-preserving dynamical systems under which the multidimensional permutation entropy coincides with the Kolmogorov-Sinai entropy. As will be explained in the next section, different methods are necessary in the multidimensional case compared to the one-dimensional case. The downside of those methods is that they are only applicable to more specific dynamical systems. On the upside, those methods allow for explicit upper bounds on the rate of convergence of the permutation entropy for some dynamical systems (see Theorem 4.5).

4.1 Introduction

In the multidimensional setting, we consider measure-preserving dynamical system $(\Omega, \mathscr{B}, \mu, T)$ where $\Omega \subseteq \mathbb{R}^d$ with $d \in \mathbb{N}$ and $\mathscr{B}$ is the Borel σ-algebra on Ω. The case $d = 1$ corresponds to the one-dimensional setting considered in the previous chapter.

Recall that

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

and

$$\mathcal{R}_X = (X \times X)^{-1}\{R, \mathbb{R}^2 \setminus R\}$$

for any random variable $X : \Omega \to \mathbb{R}$. Let $p_i : \mathbb{R}^d \to \mathbb{R}$ be the projection on the i-th coordinate, i.e. $p_i((\omega_1, \omega_2, \ldots, \omega_d)) = \omega_i$. To abbreviate the notations, we set

$$\mathcal{R}_i := \mathcal{R}_{p_i} = \{(\omega_1, \omega_2) \in \Omega^2 \mid p_i(\omega_1) < p_i(\omega_2)\}.$$

and

$$OP^i(n) := OP^{p_i}(n) = \bigvee_{s=0}^{n-1} \bigvee_{t=0}^{n-1} (T^s, T^t)^{-1}(\mathcal{R}_i),$$

for all $i \in \{1, 2, \ldots, d\}$. Notice that

$$(T^s, T^t)^{-1}(\mathcal{R}_i) = \{\omega \in \Omega \mid p_i(T^s(\omega)) < p_i(T^t(\omega))\}.$$

The partition into ordinal patterns of length n is then given by

$$OP(n) := \bigvee_{i=1}^{d} OP^i(n).$$

and the (multidimensional) permutation entropy by

$$\mathrm{PE}(T) = \liminf_{n\to\infty} \frac{1}{n} H(OP(n)).$$

In the one-dimensional case ($d = 1$), piecewise monotony was the main condition for the permutation entropy to be equal to the Kolmogorov-Sinai entropy. A system is called piecewise monotone, if there exists a partition $\mathcal{M}$ of Ω, such that for all $M \in \mathcal{M}$ $\omega_1 \leq \omega_2$ either implies

$T(\omega_1) \leq T(\omega_2)$ for all $\omega_1, \omega_2 \in M$ or $T(\omega_1) \geq T(\omega_2)$ for all $\omega_1, \omega_2 \in M$. Such a condition cannot directly be generalized to $d > 1$ because $\omega_1 \leq \omega_2$ wouldn't be well defined.

One possible way to generalize the idea of piecewise monotony to higher dimensions could be as follows:

We call T piecewise monotone, if there exists a partition $\mathcal{M}$ of $\Omega \subseteq \mathbb{R}^d$ such that for all $M \in \mathcal{M}$ and $i \in \{1, 2, \ldots, d\}$

$$p_i(\omega_1) \leq p_i(\omega_2) \text{ implies } p_i(T(\omega_1)) \leq p_i(T(\omega_2)) \text{ for all } \omega_1, \omega_2 \in M$$

or

$$p_i(\omega_1) \leq p_i(\omega_2) \text{ implies } p_i(T(\omega_1)) \geq p_i(T(\omega_2)) \text{ for all } \omega_1, \omega_2 \in M.$$

However, this is very limiting. Consider for example the function $T : [0, 1[^2 \to [0, 1[^2$ with $T((\omega_1, \omega_2)) := (2\omega_1 \mod 1, 2\omega_2 \mod 2)$. This function suffices the above given generalized conditions for piecewise monotony. But even though it is technically a multidimensional map, practically it is only a product of two one-dimensional systems and can therefore easily be separated in multiple one-dimensional systems. With the above given generalized definition of piecewise monotony, it is difficult to find functions that fulfill those conditions and are more than just a product of one-dimensional piecewise monotone functions.

Hence, trying to find conditions that directly try to generalize the one-dimensional idea of piecewise monotony might not be a good approach. Instead, we try to find conditions that do not depend on the ordering between different points or the ordering between one-dimensional projections of those points.

Since this new multidimensional approach should work for the one-dimensional case as well, considering simple one-dimensional systems first to get an idea for useful conditions could be a good strategy. So in the next section, we try to prove the equality of permutation and Kolmogorov-Sinai entropy for a specific one-dimensional system without using the fact that this system is piecewise monotone. In the subsequent section we then try to establish conditions, under which the ideas of the following section can be generalized to higher dimensions.

4.2 α-permutation entropies

The central idea of this approach is the following: We split the partition of ordinal patterns $OP(n)$ into two parts $\underline{OP}(n, \alpha)$ and $\overline{OP}(n, \alpha)$, i.e.

$$OP^i(n) = \underline{OP}^i(n, \alpha) \vee \overline{OP}^i(n, \alpha),$$

and consider the entropy of each part individually. We define these partitions as

$$\overline{OP}^i(n, \alpha) := \bigvee_{\substack{(s,t) \in E_n: \\ |t-s| \geq n^\alpha}} (T^s, T^t)^{-1}(\mathcal{R}_i),$$
$$\underline{OP}^i(n, \alpha) := \bigvee_{\substack{(s,t) \in E_n: \\ |t-s| < n^\alpha}} (T^s, T^t)^{-1}(\mathcal{R}_i)$$

for $\alpha \in [0, 1]$, where $E_n := \{(s, t) \in \{0, 1, \ldots, n-1\}^2 \mid s < t\}$.

Based on these partitions one can define the *lower* and *upper α permutation entropies*

$$\begin{aligned} &\liminf_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=1}^{d} \underline{OP}^i(n,\alpha)\right) \\ \text{and} \quad &\liminf_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=1}^{d} \overline{OP}^i(n,\alpha)\right). \end{aligned}$$

It is easy to see that

$$\begin{aligned} \overline{OP}^i(n,1) &= \underline{OP}^i(n,0) = \{\Omega\}, \\ \overline{OP}^i(n,0) &= \underline{OP}^i(n,1) = OP^i(n), \\ \overline{OP}^i(n,\alpha_2) &\prec \overline{OP}^i(n,\alpha_1) \\ \text{and} \quad \underline{OP}^i(n,\alpha_1) &\prec \underline{OP}^i(n,\alpha_2) \end{aligned}$$

hold true for for all $\alpha_1, \alpha_2 \in [0,1]$ with $\alpha_1 \leq \alpha_2$ and all $n \in \mathbb{N}$. In particular, this implies

$$\begin{aligned} &\liminf_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=1}^{d} \underline{OP}^i(n,\alpha)\right) \leq \mathrm{PE}(T) \\ \text{and} \quad &\liminf_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=1}^{d} \overline{OP}^i(n,\alpha)\right) \leq \mathrm{PE}(T) \end{aligned}$$

for all $\alpha \in [0,1]$.

On the other hand, the α-permutation entropies are not that much smaller as the following lemma shows. This lemma is not necessary to prove the main statement of this chapter but shows that the α-permutation entropies are an upper bound for the Kolmogorov-Sinai entropy the same way the original permutation entropy is, as shown in Theorem 3.10.

Lemma 4.1. Let $(\Omega, \mathscr{B}, \mu, T)$ be a measure-preserving dynamical system with $\Omega \subseteq \mathbb{R}^d$. Then

$$\begin{aligned} &\liminf_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=1}^{d} \underline{OP}^i(n,\alpha)\right) \geq h(T) \\ \text{and} \quad &\liminf_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=1}^{d} \overline{OP}^i(n,\alpha)\right) \geq h(T) \end{aligned}$$

hold true for all $\alpha \in]0,1[$.

We will introduce another lemma to help with the proof of Lemma 4.1. Given a partition $\mathcal{P}$ and a set $Q \in \mathscr{B}$, recall that

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

is the collection of sets in $\mathcal{P}$ that are intersecting the set Q.

Lemma 4.2. Let $(\Omega, \mathscr{B}, \mu, T)$ be a measure-preserving dynamical system with $\Omega \subseteq \mathbb{R}^d$, $\mathcal{U} := \{U_i\}_{i \in I}$ a finite partition of $\mathbb{R}$ into intervals and $\mathcal{Q} := p_j^{-1}(\mathcal{U})$. Then for all $\alpha \in [0,1]$

$$\#\Delta(\mathcal{Q}^{(n)}|P_\pi) \leq \binom{\#\mathcal{Q} + \lceil n^{1-\alpha} \rceil}{\#\mathcal{Q} - 1}^{\lceil n^\alpha \rceil} \tag{4.1}$$

holds true for all $P_\pi \in \overline{OP}^i(n, \alpha)$ and

$$\#\Delta(\mathcal{Q}^{(n)}|P_\pi) \leq \binom{\#\mathcal{Q} + \lceil n^{\alpha} \rceil - 1}{\#\mathcal{Q} - 1}^{\lceil n^{1-\alpha} \rceil + 1} \tag{4.2}$$

holds true for all $P_\pi \in \underline{OP}^i(n, \alpha)$.

Proof. Fix $j \in \{1, 2, \dots, d\}$. Set $I = \{1, 2, \dots N\}$ with $N = \#\mathcal{Q} = \#\mathcal{U}$ and label the sets $Q_i \in \mathcal{U}$ with $i \in I$ in such a way that

$$i_1 < i_2 \Rightarrow p_j(\omega_1) < p_j(\omega_2) \quad \text{for all } \omega_1 \in Q_{i_1} \text{ and all } \omega_2 \in Q_{i_2} \tag{4.3}$$

holds true for all $i_1, i_2 \in I$. Fix $n \in \mathbb{N}$. Set

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

for all $\boldsymbol{i} = (i_0, i_1, \dots i_{n-1}) \in I^n$. We will first prove (4.1):

Take $P_\pi \in \overline{OP}^i(n, \alpha)$. We have

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

Define

$$\begin{aligned} S := \bigcup_{m=1}^{n-1} \{(s_1, s_2, \dots s_m) \in \{0, 1, \dots, n-1\}^m :& \\ & s_{i+1} - s_i = \lceil n^\alpha \rceil \text{ for all } i \in \{1, \dots, m-1\} \\ & \text{and } s_1 < \lceil n^\alpha \rceil \text{ and } s_m > n - 1 - \lceil n^\alpha \rceil\} \end{aligned}$$

as the set of all sequences in $\{0, 1, \dots, n-1\}$ whose consecutive elements have a difference equal to $\lceil n^\alpha \rceil$ and whose length is maximal. Since $|s_{i_1} - s_{i_2}| \geq n^\alpha$ for all $\boldsymbol{s} = (s_1, s_2, \dots s_m) \in S$ and $i_1, i_2, \in \{1, \dots m\}$ with $i_1 \neq i_2$, there exists a permutation $\pi^{\boldsymbol{s}} = (\pi_1^{\boldsymbol{s}}, \pi_2^{\boldsymbol{s}}, \dots, \pi_m^{\boldsymbol{s}})$ of $\boldsymbol{s}$ such that

$$p_j(T^{\pi_1^{\boldsymbol{s}}}(\omega)) \leq p_j(T^{\pi_2^{\boldsymbol{s}}}(\omega)) \leq \dots \leq p_j(T^{\pi_m^{\boldsymbol{s}}}(\omega))$$

holds true for all $\omega \in P_\pi$. Using (4.3), this implies

$$i_{\pi_1^{\boldsymbol{s}}} \leq i_{\pi_2^{\boldsymbol{s}}} \leq \dots \leq i_{\pi_m^{\boldsymbol{s}}}$$

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

$$\begin{aligned} & \#\{(i_{s_k})_{k=1}^m \in I^m \mid \mu\left(\bigcap_{k=1}^m T^{-s_k}(Q_{i_{s_k}}) \cap P_\pi\right) > 0\} \\ & \leq \#\{(i_{s_k})_{k=1}^m \in I^m \mid i_{\pi_1^{\boldsymbol{s}}} \leq i_{\pi_2^{\boldsymbol{s}}} \leq \dots \leq i_{\pi_m^{\boldsymbol{s}}}\} \\ & = \binom{N + m - 1}{N - 1} \end{aligned}$$

holds true for all $\boldsymbol{s} \in S$. Since the length m of a sequence $\boldsymbol{s} \in S$ is at most

$$\left\lceil \frac{n}{\lceil n^\alpha \rceil} \right\rceil + 1 \leq \left\lceil \frac{n}{n^\alpha} \right\rceil + 1 = \lceil n^{1-\alpha} \rceil + 1,$$

we have

$$\#\{(i_{s_k})_{k=1}^m \in I^m \mid \mu\left(\bigcap_{k=1}^m T^{-s_k}(Q_{i_{s_k}}) \cap P_\pi\right) > 0\} \leq \binom{N + \lceil n^{1-\alpha} \rceil}{N-1}$$

for all $\boldsymbol{s} \in S$. The set S is defined in such a way that for every $i \in \{0, 1, \ldots, n-1\}$ there exists a sequence $\boldsymbol{s} \in S$ and an index $k \in \{1, 2, \ldots, m\}$ such that $s_k = i$. This implies

$$\begin{aligned}
&\#\{\boldsymbol{i} \in I^n \mid \mu(Q(\boldsymbol{i}) \cap P_\pi) > 0\} \\
&\leq \prod_{\boldsymbol{s} \in S} \#\{(i_{s_k})_{k=1}^m \in I^m \mid \mu\left(\bigcap_{k=1}^m T^{-s_k}(Q_{i_{s_k}}) \cap P_\pi\right) > 0\} \\
&\leq \prod_{\boldsymbol{s} \in S} \binom{N + \lceil n^{1-\alpha} \rceil}{N-1} = \binom{N + \lceil n^{1-\alpha} \rceil}{N-1}^{\#S},
\end{aligned}$$

which finishes the proof of (4.1) by using $\#S = \lceil n^\alpha \rceil$. We will now prove (4.2):

Now take $P_\pi \in \underline{OP}^i(n, \alpha)$. For all $s \in \{0, 1, \ldots, n - \lceil n^\alpha \rceil\}$ there exists a permutation $\pi^s = (\pi_1^s, \pi_2^s, \ldots, \pi_{\lceil n^\alpha \rceil}^s)$ of $\{s, s+1, \ldots, s + \lceil n^\alpha \rceil - 1\}$ such that

$$p_j(T^{\pi_1^s}(\omega)) \leq p_j(T^{\pi_2^s}(\omega)) \leq \ldots \leq p_j(T^{\pi_m^s}(\omega))$$

holds true for all $\omega \in P_\pi$. Using (4.3), this implies

$$i_{\pi_1^s} \leq i_{\pi_2^s} \leq \ldots \leq i_{\pi_m^s}$$

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

$$\begin{aligned}
&\#\{(i_k)_{k=s}^{s+\lceil n^\alpha \rceil - 1} \in I^{\lceil n^\alpha \rceil} \mid \mu\left(\bigcap_{k=s}^{s+\lceil n^\alpha \rceil - 1} T^{-k}(Q_{i_k}) \cap P_\pi\right) > 0\} \\
&\leq \#\{(i_k)_{k=s}^{s+\lceil n^\alpha \rceil - 1} \in I^{\lceil n^\alpha \rceil} \mid i_{\pi_1^s} \leq i_{\pi_2^s} \leq \ldots \leq i_{\pi_m^s}\} \\
&= \binom{N + \lceil n^\alpha \rceil - 1}{N-1}
\end{aligned}$$

holds true for all $s \in \{0, 1, \ldots, n - \lceil n^\alpha \rceil\}$. This implies

$$\begin{aligned}
&\#\{\boldsymbol{i} \in I^n \mid \mu(Q(\boldsymbol{i}) \cap P_\pi) > 0\} \\
&\leq \prod_{u=0}^{\lceil n^{1-\alpha} \rceil} \#\{(i_k)_{k=u(\lceil n^\alpha \rceil - 1)}^{(u+1)(\lceil n^\alpha \rceil - 1)} \in I^{\lceil n^\alpha \rceil} \mid \mu\left(\bigcap_{k=s}^{s+\lceil n^\alpha \rceil - 1} T^{-k}(Q_{i_k}) \cap P_\pi\right) > 0\} \\
&\leq \prod_{u=0}^{\lceil n^{1-\alpha} \rceil} \binom{N + \lceil n^\alpha \rceil - 1}{N-1} = \binom{N + \lceil n^\alpha \rceil - 1}{N-1}^{\lceil n^{1-\alpha} \rceil + 1},
\end{aligned}$$

which finishes the proof of (4.2). □

Proof of Lemma 4.1. Because the Borel σ-algebra can be generated by finite partitions into sets of the form $\bigcap_{j=1}^{d} p_j^{-1}(A_j)$ where A_j are intervals, for all $\varepsilon > 0$ there exist finite partitions $\mathcal{U}_j$ of $\mathbb{R}$ into intervals, such that

$$h(T) \leq h\left(T, \bigvee_{j=1}^{d} p_j^{-1}(\mathcal{U}_j)\right) + \varepsilon$$

holds true. Set $\mathcal{Q}_j := p_j^{-1}(\mathcal{U}_j)$. Using properties of the Shannon entropy, we get

$$\begin{aligned} H\left(\bigvee_{j=1}^{d} \mathcal{Q}_j^{(n)}\right) &\leq H\left(\bigvee_{j=1}^{d} \mathcal{Q}_j^{(n)} \vee \overline{OP}^j(n,\alpha)\right) \\ &= H\left(\bigvee_{j=1}^{d} \mathcal{Q}_j^{(n)} \middle| \overline{OP}^j(n,\alpha)\right) + H\left(\bigvee_{j=1}^{d} \overline{OP}^j(n,\alpha)\right) \\ &\leq \sum_{j=1}^{d} H\left(\mathcal{Q}_j^{(n)} | \overline{OP}^j(n,\alpha)\right) + H\left(\bigvee_{j=1}^{d} \overline{OP}^j(n,\alpha)\right) \end{aligned}$$

for all $n \in \mathbb{N}$. Lemma 3.3 implies

$$H(\mathcal{Q}_j^{(n)} | \overline{OP}^j(n,\alpha)) \leq \sum_{P_\pi \in \overline{OP}^j(n,\alpha)} \mu(P_\pi) \cdot \log(\#\Delta(\mathcal{Q}_j^{(n)} | P_\pi)),$$

which together with Lemma 4.2 yields

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{n} H(\mathcal{Q}_j^{(n)} | \overline{OP}^j(n,\alpha)) \\ \leq &\lim_{n\to\infty} \frac{1}{n} \sum_{P_\pi \in \overline{OP}^j(n,\alpha)} \mu(P_\pi) \cdot \log\left(\binom{\#\mathcal{Q}_j + \lceil n^{1-\alpha} \rceil}{\#\mathcal{Q}_j - 1}^{\lceil n^\alpha \rceil}\right) \\ = &\lim_{n\to\infty} \frac{1}{n} \log\left(\binom{\#\mathcal{Q}_j + \lceil n^{1-\alpha} \rceil}{\lceil n^{1-\alpha} \rceil + 1}^{\lceil n^\alpha \rceil}\right) \end{aligned}$$

$$
\begin{aligned}
&= \lim_{n\to\infty} \frac{\lceil n^{\alpha}\rceil}{n} \log \binom{\#\mathcal{Q}_j + \lceil n^{1-\alpha}\rceil}{\lceil n^{1-\alpha}\rceil + 1} \\
&\leq \lim_{n\to\infty} \frac{\lceil n^{\alpha}\rceil}{n} \log \left((\#\mathcal{Q}_j + \lceil n^{1-\alpha}\rceil)^{\#Q_j - 1} \right) \\
&= \lim_{n\to\infty} \frac{\lceil n^{\alpha}\rceil \log(\#\mathcal{Q}_j + \lceil n^{1-\alpha}\rceil)}{n} \cdot (\#Q_j - 1) \\
&= \lim_{n\to\infty} \frac{\log(\#\mathcal{Q}_j + \lceil n^{1-\alpha}\rceil)}{n^{1-\alpha}} \cdot (\#Q_j - 1) = 0.
\end{aligned}
$$

So we can conclude

$$
\begin{aligned}
h(T) &\leq \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{j=1}^{d} \mathcal{Q}_j^{(n)} \right) + \varepsilon \\
&\leq \liminf_{n\to\infty} \frac{1}{n} \left[H(\bigvee_{j=1}^{d} \mathcal{Q}_j^{(n)} | \overline{OP}^j(n,\alpha)) + H\left(\bigvee_{j=1}^{d} \overline{OP}^j(n,\alpha) \right) \right] + \varepsilon \\
&= \liminf_{n\to\infty} \frac{1}{n} H\left(\bigvee_{j=1}^{d} \overline{OP}^j(n,\alpha) \right) + \varepsilon.
\end{aligned}
$$

Analogously, using Lemma 4.2, one can show

$$
\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} H(\mathcal{Q}_j^{(n)} | \underline{OP}^j(n,\alpha)) \\
&\leq \lim_{n\to\infty} \frac{1}{n} \log \left(\binom{\#\mathcal{Q}_j + \lceil n^{\alpha}\rceil - 1}{\#\mathcal{Q}_j - 1}^{\lceil n^{1-\alpha}\rceil + 1} \right) \\
&= \lim_{n\to\infty} \frac{\lceil n^{1-\alpha}\rceil + 1}{n} \log \binom{\#\mathcal{Q}_j + \lceil n^{\alpha}\rceil - 1}{\lceil n^{\alpha}\rceil} \\
&\leq \lim_{n\to\infty} \frac{\lceil n^{1-\alpha}\rceil + 1}{n} \log \left((\#\mathcal{Q}_j + \lceil n^{\alpha}\rceil - 1)^{\#Q_j - 1} \right) \\
&= \lim_{n\to\infty} \frac{(\lceil n^{1-\alpha}\rceil + 1) \log(\#\mathcal{Q}_j + \lceil n^{\alpha}\rceil - 1)}{n} \cdot (\#Q_j - 1) \\
&= \lim_{n\to\infty} \frac{\log(\#\mathcal{Q}_j + \lceil n^{\alpha}\rceil - 1)}{n^{\alpha}} \cdot (\#Q_j - 1) = 0,
\end{aligned}
$$

which implies

$$
\begin{aligned}
h(T) &\leq \liminf_{n\to\infty} \frac{1}{n} \left[H(\bigvee_{j=1}^{d} \mathcal{Q}_j^{(n)} | \underline{OP}^j(n,\alpha)) + H\left(\bigvee_{j=1}^{d} \underline{OP}^j(n,\alpha) \right) \right] + \varepsilon \\
&= \liminf_{n\to\infty} \frac{1}{n} H\left(\bigvee_{j=1}^{d} \underline{OP}^j(n,\alpha) \right) + \varepsilon.
\end{aligned}
$$

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

4.2.1 Combinatorial arguments

Given a finite partition $\mathcal{P} = \{P_i\}_{i \in I}$, recall that

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

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

The following lemma is a generalization of some of the results proved in Lemma 3.20 and is fundamental to most of the arguments used in this chapter.

Lemma 4.3. Let $(\Omega, \mathscr{B}, \mu, T)$ be a measure-preserving dynamical system with $\Omega \subseteq \mathbb{R}^d$, $\mathcal{U} = \{U_i\}_{i \in I}$ a partition of $[-b, b]$ into intervals and $\mathcal{P} = \{P_i\}_{i \in I} = p_j^{-1}(\mathcal{U})$. Then for all $j \in \{1, 2, \ldots, d\}$, $n \in \mathbb{N}$, $E \subseteq \{0, 1, \ldots, n-1\}^2$ and multi indices $\boldsymbol{i} = (i_0, i_1, \ldots, i_{n-1}) \in I^n$

$$\#\Delta\left(\bigvee_{(s,t) \in E} (T^s, T^t)^{-1}(\mathcal{R}_j) \,\middle|\, P(\boldsymbol{i}) \right) \leq 2^{\#\{(s,t) \in E \mid \, i_s = i_t\}}$$

holds true.

Proof. Fix $j \in \{1, 2, \ldots, d\}$, $n \in \mathbb{N}$, a multi index $\boldsymbol{i} = (i_0, i_1, \ldots, i_{n-1}) \in I^n$ and $E \subseteq \{0, 1, \ldots, n-1\}^2$. Then

$$\#\Delta\left(\bigvee_{(s,t) \in E} (T^s, T^t)^{-1}(\mathcal{R}_j) \,\middle|\, P(\boldsymbol{i}) \right) \leq \prod_{(s,t) \in E} \#\Delta((T^s, T^t)^{-1}(\mathcal{R}_j) | P(\boldsymbol{i})) \tag{4.4}$$

holds true. Note that

$$(T^s, T^t)^{-1}(\mathcal{R}_j) = \{\{\omega \in \Omega \mid p_j(T^s(\omega)) < p_j(T^t(\omega))\}, \{\omega \in \Omega \mid p_j(T^s(\omega)) \geq p_j(T^t(\omega))\}\}.$$

If $i_s \neq i_t$ is true, $p_j(T^s(\omega))$ and $p_j(T^t(\omega))$ are located in different intervals U_{i_s} and U_{i_t} for all $\omega \in P(\boldsymbol{i})$. Because every point in the left interval is smaller than any point in the other interval, we have either

$$\Delta((T^s, T^t)^{-1}(\mathcal{R}_j) | P(\boldsymbol{i})) = \{\omega \in \Omega \mid p_j(T^s(\omega)) < p_j(T^t(\omega))\}$$

or

$$\Delta((T^s, T^t)^{-1}(\mathcal{R}_j) | P(\boldsymbol{i})) = \{\omega \in \Omega \mid p_j(T^s(\omega)) \geq p_j(T^t(\omega))\}.$$

This implies

$$\#\Delta((T^s, T^t)^{-1}(\mathcal{R}_j) | P(\boldsymbol{i})) = 1.$$

If $i_s = i_t$,

$$\#\Delta((T^s, T^t)^{-1}(\mathcal{R}_j) | P(\boldsymbol{i})) \leq \#((T^s, T^t)^{-1}(\mathcal{R}_j) = 2$$

holds true. Applying (4.4) then yields

$$\#\Delta\left(\bigvee_{(s,t) \in E} (T^s, T^t)^{-1}(\mathcal{R}_j) \,\middle|\, P(\boldsymbol{i}) \right) \leq 2^{\#\{(s,t) \in E \mid \, i_s = i_t\}}. \qquad \square$$

4.2.2 Permutation entropy of Bernoulli shifts from a different perspective

Given $N \in \mathbb{N} \setminus \{1\}$ and a stochastic vector $\boldsymbol{q} = (q_1, q_2, \ldots, q_N)$ with $q_* := \max_{1 \le i \le N} q_i < 1$, let $([0,1[, \mathscr{B}, \mu, T)$ with

$$T(\omega) = N \cdot \omega \mod 1$$

be the Bernoulli shift generated by $\boldsymbol{q}$ as introduced in Section 2.3.4. To exemplify how comparing the Kolmogorov-Sinai entropy to the α-permutation entropies can potentially be easier than comparing it to the original permutation entropy, we compare the different entropies for this Bernoulli shifts. The relatively simple structure of Bernoulli shifts allows us to determine the difference between these entropies analytically. In Section 2.3.4, we explicitly determined the Kolmogorov-Sinai entropy of Bernoulli shifts. And since Bernoulli shifts are piecewise monotone, Theorem 3.17 implies that the permutation entropy of these system is equal to its Kolmogorov-Sinai entropy.

We will now show that the Kolmogorov-Sinai and permutation entropy are equal for Bernoulli shifts by considering α permutation entropies. Let

$$\mathcal{P} = \{[(i-1)/N, i/N[\mid i \in \{1, 2, \ldots, N\}\}$$

be the partition of $[0,1[$ into the N intervals on which T is monotone. Recall that

$$\mu\left(\bigcap_{t=0}^{n-1} P_{i_t}\right) = \prod_{t=0}^{n} q_{i_t}$$

holds true for all $(i_0, i_1, \ldots, i_{n-1}) \in \{1, \ldots, N\}^n$.

As in Section 3, we will use conditional entropies similar to

$$H(OP(n)|\mathcal{P}^{(n)})$$

to determine the entropy difference between symbolic and ordinal patterns. But instead of using $OP(n)$, we use $\underline{OP}(n,\alpha)$ or $\overline{OP}(n,\alpha)$ and instead of using a fixed partition $\mathcal{P}$, we consider partitions that additionally depend on the length n of the considered ordinal patterns. In the example of Bernoulli shifts, for fixed $\alpha \in]0,1[$, we consider $\mathcal{P}_{\lceil n^\alpha \rceil}$ for $n \in \mathbb{N}$ with

$$\mathcal{P}_k := \{[(i-1) \cdot N^{-k}, i \cdot N^{-k}[\mid i \in \{1, 2, \ldots, N^k\}\}$$

for $k \in \mathbb{N}$ as defined in (2.10). Even though these partitions change depending on n, $\frac{1}{n} H(\mathcal{P}_{\lceil n^\alpha \rceil}^{(n)})$ still converges to the Kolmogorov-Sinai entropy due to

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} H(\mathcal{P}_{\lceil n^\alpha \rceil}^{(n)}) &\overset{(2.11)}{=} \lim_{n\to\infty} \frac{1}{n} H(\mathcal{P}^{(n+\lceil n^\alpha \rceil - 1)}) \\
&= \lim_{n\to\infty} \frac{n + \lceil n^\alpha \rceil - 1}{n} \cdot \frac{1}{n + \lceil n^\alpha \rceil - 1} H(\mathcal{P}^{(n+\lceil n^\alpha \rceil - 1)}) \\
&= \lim_{n\to\infty} \frac{1}{n + \lceil n^\alpha \rceil - 1} H(\mathcal{P}^{(n+\lceil n^\alpha \rceil - 1)}) \\
&\overset{Lem. 2.29}{=} H(\boldsymbol{q}) = h(T).
\end{aligned} \tag{4.5}$$

Applying Lemma 3.3 and Lemma 4.3 to the here considered partitions into intervals $\mathcal{P}_{\lceil n^\alpha \rceil} =:$

$\{P_i^n\}_{i\in I_n}$ yields

$$H(\overline{OP}(n,\alpha)|\mathcal{P}^{(n)}_{\lceil n^\alpha\rceil}) \leq \log(2)\cdot \sum_{\boldsymbol{i}\in I_n^n} \mu(P^n(\boldsymbol{i}))\#\{(s,t)\in E_n \mid s+n^\alpha \leq t \text{ and } i_s = i_t\}$$

$$\leq \log(2)\cdot n^2\cdot \mu\left(\bigcup_{P\in\mathcal{P}_{\lceil n^\alpha\rceil}} \{\omega\in\Omega \mid \#\{(s,t)\in E_n \mid s+n^\alpha\leq t \text{ and } \omega\in T^{-s}(P)\cap T^{-t}(P)\}>1\}\right)$$

$$\leq \log(2)\cdot n^2\cdot \sum_{P\in\mathcal{P}_{\lceil n^\alpha\rceil}} \sum_{\substack{(s,t)\in E_n:\\ s+n^\alpha\leq t}} \mu(T^{-s}(P)\cap T^{-t}(P))$$

$$\leq \log(2)\cdot n^3\cdot \sum_{P\in\mathcal{P}_{\lceil n^\alpha\rceil}} \sum_{t=\lceil n^\alpha\rceil}^{n} \mu(P\cap T^{-t}(P)) \tag{4.6}$$

$$\overset{(2.12)}{=} \log(2)\cdot n^3\cdot \sum_{t=\lceil n^\alpha\rceil}^{n} \sum_{P\in\mathcal{P}_{\lceil n^\alpha\rceil}} \mu(P)^2$$

$$\leq \log(2)\cdot n^3\cdot \sum_{P\in P_{\lceil n^\alpha\rceil}} \mu(P) \sum_{t=\lceil n^\alpha\rceil}^{n} q_*^t \leq \log(2)\cdot n^{3+\alpha}\cdot q_*^{n^\alpha},$$

which converges to 0 for $n\to\infty$.

Analogously to (4.6), one can show

$$H(\underline{OP}(n,\alpha)|\mathcal{P}^{(n)}_{\lceil n^\alpha\rceil}) \leq \log(2)\cdot n^3 \sum_{P\in\mathcal{P}_{\lceil n^\alpha\rceil}} \sum_{t=1}^{\lceil n^\alpha\rceil-1} \mu(P\cap T^{-t}(P)).$$

To find an upper bound for the above quantity, we will use the following lemma.

Lemma 4.4. Let $([0,1[,\mathscr{B},\mu,T)$ be the Bernoulli shift generated by the stochastic vector $\boldsymbol{q}=(q_1,q_2,\ldots,q_N)$ and

$$\mathcal{P}=\{[(i-1)/N,i/N[\mid i\in\{1,2,\ldots,N\}\}$$

Then

$$\sum_{P\in\mathcal{P}^{(k)}} \mu(P\cap T^{-t}(P)) \leq \left(\sum_{i=1}^{N} q_i^{\lfloor k/t+1\rfloor}\right)^t$$

holds true for all $k,t\in\mathbb{N}$ with $t<k$.

Proof. Fix $k,t\in\mathbb{N}$ with $t<k$. Take $P\in\mathcal{P}^{(k)}$. There exists $\boldsymbol{i}=(i_0,i_1,\ldots,i_{k-1})$ with

$$P=\bigcap_{s=0}^{k-1} T^{-s}(P_{i_s}).$$

Let

$$D:=\{P\in\mathcal{P}^{(k)}\mid P\cap T^{-t}(P)\neq\emptyset\}$$

Suppose $P \in D$. Then $i_s = i_{s+t}$ must hold true for all $s \in \{0, 1, \ldots, k-1-t\}$. Hence

$$\begin{aligned} P \cap T^{-t}(P) &= \left(\bigcap_{u=0}^{\lfloor (k+t)/t \rfloor - 1} T^{-ut} \left(\bigcap_{s=0}^{t-1} T^{-s} P_{i_s} \right) \right) \cap \left(\bigcap_{s=t \cdot \lfloor (k+t)/t \rfloor}^{k+t-1} \right) \\ &\subseteq \bigcap_{u=0}^{\lfloor (k+t)/t \rfloor - 1} T^{-ut} \left(\bigcap_{s=0}^{t-1} T^{-s-u\cdot t} P_{i_s} \right) \end{aligned}$$

and

$$\begin{aligned} \mu\left(P \cap T^{-t}(P)\right) &\leq \mu \left(\bigcap_{u=0}^{\lfloor (k+t)/t \rfloor - 1} T^{-ut} \left(\bigcap_{s=0}^{t-1} T^{-s} P_{i_s} \right) \right) \\ &= \prod_{u=0}^{\lfloor (k+t)/t \rfloor - 1} \prod_{s=0}^{t-1} q_{i_s} = \prod_{s=0}^{t-1} q_{i_s}^{\lfloor (k+t)/t \rfloor}. \end{aligned}$$

Therefore,

$$\begin{aligned} &\sum_{P \in \mathcal{P}^{(k)}} \mu\left(P \cap T^{-t}(P)\right) = \sum_{P \in D} \mu\left(P \cap T^{-t}(P)\right) \\ &\leq \sum_{(i_0, \ldots, i_{t-1}) \in \{1, \ldots, N\}^t} \prod_{s=0}^{t-1} q_{i_s}^{\lfloor (k+t)/t \rfloor} = \left(\sum_{i=1}^{N} q_i^{\lfloor k/t+1 \rfloor} \right)^t. \end{aligned}$$

□

Using the above Lemma, we can show that for sufficiently large $n \in \mathbb{N}$

$$\begin{aligned} H(\underline{OP}(n, \alpha) | \mathcal{P}_{\lceil n^\alpha \rceil}^{(n)}) &\leq \log(2) \cdot n^3 \cdot \sum_{t=1}^{\lceil n^\alpha \rceil - 1} \sum_{P \in \mathcal{P}_{\lceil n^\alpha \rceil}} \mu(P \cap T^{-t}(P)) \\ &\leq \log(2) \cdot n^3 \cdot \sum_{t=1}^{\lceil n^\alpha \rceil - 1} \left(\sum_{i=1}^{N} q_i^{\lfloor \lceil n^\alpha \rceil / t + 1 \rfloor} \right)^t \\ &\leq \log(2) \cdot n^3 \cdot \sum_{t=1}^{\lceil n^\alpha \rceil - 1} \left(\sum_{i=1}^{N} q_i^{\lfloor n^\alpha / t + 1 \rfloor} \right)^t \\ &= \log(2) \cdot n^3 \cdot \left[\sum_{t=1}^{\lceil n^{\alpha/2} \rceil - 1} \left(\sum_{i=1}^{N} q_i^{\lfloor n^\alpha / t + 1 \rfloor} \right)^t + \sum_{t=\lceil n^{\alpha/2} \rceil}^{\lceil n^\alpha \rceil - 1} \left(\sum_{i=1}^{N} q_i^{\lfloor n^\alpha / t + 1 \rfloor} \right)^t \right] \\ &\leq \log(2) \cdot n^3 \cdot \left[\sum_{t=1}^{\lceil n^{\alpha/2} \rceil - 1} \left(N \cdot q_*^{\lfloor n^\alpha / t + 1 \rfloor} \right)^t + \sum_{t=\lceil n^{\alpha/2} \rceil}^{\lceil n^\alpha \rceil - 1} \left(\sum_{i=1}^{N} q_i^2 \right)^t \right] \\ &\leq \log(2) \cdot n^3 \cdot \left[\sum_{t=1}^{\lceil n^{\alpha/2} \rceil - 1} \left(N \cdot q_*^{\lfloor n^\alpha / n^{\alpha/2} + 1 \rfloor} \right) + \sum_{t=\lceil n^{\alpha/2} \rceil}^{\lceil n^\alpha \rceil - 1} \left(\sum_{i=1}^{N} q_* \cdot q_i \right)^{\lceil n^{\alpha/2} \rceil} \right] \\ &\leq \log(2) \cdot n^3 \cdot \left[(\lceil n^{\alpha/2} \rceil - 1) \cdot N \cdot q_*^{n^{\alpha/2}} + (\lceil n^\alpha \rceil - \lceil n^{\alpha/2} \rceil) \cdot q_*^{n^{\alpha/2}} \right] \\ &\leq \log(2) \cdot n^{3+\alpha} \cdot (N+1) \cdot q_*^{n^{\alpha/2}} \end{aligned}$$

holds true. Again, this converges to 0 for $n \to \infty$. Combining the above arguments implies

$$\begin{aligned}\frac{1}{n}H(OP(n)) &\leq \frac{1}{n}H(OP(n)|\mathcal{P}^{(n)}_{\lceil n^\alpha \rceil}) + \frac{1}{n}H(\mathcal{P}^{(n)}_{\lceil n^\alpha \rceil}) \\ &\leq \frac{1}{n}H(\overline{OP}(n,\alpha)|\mathcal{P}^{(n)}_{\lceil n^\alpha \rceil}) + \frac{1}{n}H(\underline{OP}(n,\alpha)|\mathcal{P}^{(n)}_{\lceil n^\alpha \rceil}) + \frac{1}{n}H(\mathcal{P}^{(n)}_{\lceil n^\alpha \rceil}) \\ &\leq \log(2) \cdot n^{2+\beta} \cdot (q_*^{n^\alpha} + (N+1) \cdot q_*^{n^{\alpha/2}} + \frac{1}{n}H(\mathcal{P}^{(n)}_{\lceil n^\alpha \rceil}) \end{aligned} \tag{4.7}$$

for all $\alpha \in]0,1[$. This immediately provides an upper bound for the entropy difference:

Theorem 4.5. Let $([0,1[, \mathscr{B}, \mu, T)$ be the Bernoulli shift generated by the stochastic vector $\boldsymbol{q} = (q_1, q_2, \ldots, q_N)$ and $q_* := \max_{1 \leq i \leq N} q_i$. Then

$$\frac{1}{n}H(OP(n)) - h(T) \leq \inf_{0<\alpha<1} \log(2) \cdot n^{2+\alpha} \cdot (q_*^{n^\alpha} + (N+1) \cdot q_*^{n^{\alpha/2}} + n^{\alpha-1} \cdot H(\mathbf{q}). \tag{4.8}$$

holds true.

Proof. Let $\alpha \in]0,1[$, $n \in \mathbb{N}$ and $\mathcal{P}_k$ as given above. Lemma 2.29 provides

$$\frac{1}{n}H(\mathcal{P}^{(n)}_{\lceil n^\alpha \rceil}) = \frac{\lceil n^\alpha \rceil + n - 1}{n} \cdot H(\mathbf{q}) \leq \frac{n^\alpha + n}{n} \cdot H(\mathbf{q}) = (n^{\alpha-1} + 1) \cdot H(\mathbf{q})$$

The proof of this Theorem then directly follows from (4.7) and the fact that $H(\mathbf{q}) = h(T)$ holds true. ☐

Notice that we have not used the piecewise monotony of the function T at any point to prove this result. Therefore, a method similar to the above approach could, theoretically, be used in higher dimensions. To prove the part of this result involving $H(\underline{OP}(n,\alpha)$ we, roughly speaking, used the fact that Bernoulli shifts are expanding, in the sense that the distance between a pair of close points grows exponentially under iteration. To prove the other part concerning $H(\overline{OP}(n,\alpha)$, we used some kind of mixing property of Bernoulli shifts.

Throughout this chapter, we will introduce the conditions (C1)-(C6) to generalize those ideas of expanding and mixing, together with some other necessary conditions, to higher dimensions.

Another advantage of the above approach compared to using monotony, is that the more constructive nature of this arguments yields an explicit upper bound for the entropy difference that does not only hold true asymptotically but for every $n \in \mathbb{N}$.

Different combinatorial arguments can give the lower bound

$$\frac{1}{n}H(OP(n)) - h(T) \geq (N-1)\frac{\log(n+N-1)}{n}.$$

So for Bernoulli shift maps, one can directly give an upper bound on how fast $\frac{1}{n}H(OP(n))$ converges to $h(T)$.

4.3 Generalization to higher dimensions

The general strategy of our approach can be summarized in the following theorem:

Theorem 4.6. Let $(\Omega, \mathscr{B}, T, \mu)$ be a measure-preserving dynamical system for $\Omega \subseteq \mathbb{R}^d$. Suppose for some $\alpha \in]0,1[$ there exist sequences $(\mathcal{P}_n)_{n\in\mathbb{N}}$ and $(\mathcal{Q}_n)_{n\in\mathbb{N}}$ of partitions of Ω with

(i) $\limsup_{n\to\infty} \frac{1}{n} H(\mathcal{P}_n^{(n)}) \leq h(T)$,

(ii) $\limsup_{n\to\infty} \frac{1}{n} H(\mathcal{Q}_n^{(n)} | \mathcal{P}_n^{(n)}) = 0$,

(iii) $\limsup_{n\to\infty} \frac{1}{n} H(\overline{OP}^j(n,\alpha) | \mathcal{Q}_n^{(n)}) = 0$ for all $j \in \{1,2,\ldots,d\}$ and

(iv) $\limsup_{n\to\infty} \frac{1}{n} H(\underline{OP}^j(n,\alpha) | \mathcal{P}_n^{(n)}) = 0$ for all $j \in \{1,2,\ldots,d\}$.

Then

$$\overline{\mathrm{PE}}(T) \leq h(T)$$

holds true.

Proof. It holds

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} H\left(OP(n)\right) \\
&\leq \limsup_{n\to\infty} \frac{1}{n} H\left(OP(n) \vee \mathcal{P}_n^{(n)}\right) \\
&\leq \limsup_{n\to\infty} \frac{1}{n} H\left(\mathcal{P}_n^{(n)}\right) + \frac{1}{n} H\left(OP(n) | \mathcal{P}_n^{(n)}\right) \\
&\overset{(i)}{\leq} h(T) + \limsup_{n\to\infty} \frac{1}{n} H\left(OP(n) | \mathcal{P}_n^{(n)}\right) \\
&\leq h(T) + \limsup_{n\to\infty} \frac{1}{n} \sum_{j=1}^{d} H\left(\underline{OP}^j(n) \vee \overline{OP}^j(n,\alpha) | \mathcal{P}_n^{(n)}\right) \\
&\leq h(T) + \limsup_{n\to\infty} \frac{1}{n} \sum_{j=1}^{d} H\left(\underline{OP}^j(n,\alpha) | \mathcal{P}_n^{(n)}\right) + H\left(\overline{OP}^j(n,\alpha) | \mathcal{P}_n^{(n)}\right) \\
&\overset{(iv)}{\leq} h(T) + \limsup_{n\to\infty} \frac{1}{n} \sum_{j=1}^{d} H\left(\overline{OP}^j(n,\alpha) | \mathcal{P}_n^{(n)}\right) \\
&\leq h(T) + \limsup_{n\to\infty} \frac{1}{n} \sum_{j=1}^{d} H\left(\overline{OP}^j(n,\alpha) | \mathcal{Q}_n^{(n)}\right) + H\left(\mathcal{Q}_n^{(n)} | \mathcal{P}_n^{(n)}\right) \\
&\overset{(ii)+(iii)}{=} h(T).
\end{aligned}$$

□

4.3.1 Upper α-permutation entropy

To prove (i), (ii) and (iii) in Theorem 4.6, we require that the measure preserving dynamical system $(\Omega, \mathscr{B}, \mu, T)$ fulfills the following conditions:

(C1) Ω is bounded, i.e. there exists $b > 0$ with

$$\Omega \subseteq \{x \in \mathbb{R}^d \mid \|x\| \leq b\},$$

(C2) the measure μ is not "too concentrated" on small stripes, i.e there exists a constant $C > 0$ and $\gamma \in]0,1[$ such that for all $x, y \in \mathbb{R}$ with $x < y$ and $i \in \{1, 2, \dots, d\}$

$$\mu(p_i^{-1}([x,y])) \leq C(y-x)^\gamma,$$

where $p_i : \Omega \to \mathbb{R}$ with $p_i((\omega_1, \omega_2, \dots, \omega_d)) = \omega_i$ is the projection on the i-th coordinate,

(C3) T is piecewise expanding, i.e. there exists a constant $c_1 > 1$ and a finite partition $\mathcal{M}$ of Ω such that for all $M \in \mathcal{M}$ and for all $\omega, \nu \in M$ we have

$$\|T(\omega) - T(\nu)\| \geq c_1 \|\omega - \nu\|,$$

where $\|\circ\| = \|\circ\|_2$ is the 2-norm,

(C4) the sequence of partitions $(T^{-n}(\mathcal{M}))_{n\in\mathbb{N}_0}$ is mixing with a mixing rate $\varphi(n)$ that is satisfies

$$\lim_{n\to\infty} n^{3/\alpha} \cdot \varphi(n) = 0$$

for some $\alpha \in]0,1[$.

Remark 4.7.

- Condition (C2) implies in particular that the measure μ is non-atomic, i.e $\mu(\{\omega\}) = 0$ for all $\omega \in \Omega$.
- If T is continuously differentiable on each $M \in \mathcal{M}$, condition (C3) is satisfied if every eigenvalue of the Jacobian is larger than c for every point in M.
- The mixing rate of the sequence of partitions $(T^{-n}(\mathcal{M}))_{n\in\mathbb{N}_0}$ in condition (C4) corresponds to the mixing rate of the stationary stochastic process $(X_n)_{n\in\mathbb{N}_0}$ with $X_n(\omega) = i$ if $T^n(\omega) \in M_i$.

Mixing rate Let $\varphi : \mathbb{N} \to [0,\infty[$ be a monotonically decreasing function. We call a sequence of partitions $(T^{-n}(\mathcal{M}))_{n\in\mathbb{N}_0}$ mixing with rate $\varphi(n)$ if there exists a $k \in \mathbb{N}$ such that the following holds true for all $n \in \mathbb{N}$:

There exists a collection $\mathcal{F}$ of sets in $\subseteq \sigma(\bigvee_{i=0}^{n-1} T^{-i}(\mathcal{M}))$ with

$$\mu\left(\Omega \setminus \bigcup_{F\in\mathcal{F}} F\right) \leq n^{-4} \tag{4.9}$$

such that

$$\sup_{\substack{A\in\mathcal{F} \\ B\in\sigma(\bigvee_{i=0}^{n-1} T^{-i}(\mathcal{M}))}} \frac{\mu(A \cap T^{-kn}(B)) - \mu(A)\mu(B)}{\mu(A)} \leq \varphi(n). \tag{4.10}$$

The mixing rate is small if the dependency between different iterations decreases fast enough. For example, if $(X_n)_{n\in\mathbb{N}_0}$ with $X_n(\omega) = i$ if $T^n(\omega) \in M_i$ is a Markov process (of order k), the mixing rate $\varphi(n)$ of $(X_n)_{n\in\mathbb{N}_0}$ is zero for all $n \in \mathbb{N}$.

To illustrate that there exist dynamical systems that satisfy the above conditions and that those systems are not too special, we will consider the following example.

Example 7. Consider the measure-preserving dynamical system $(\Omega, \mathscr{A}, \mu, T)$ with $\Omega = [0,1[^2$, $\mathscr{A}$ the Borel-σ-algebra of Ω, μ the 2-dimensional Lebesgue measure on Ω and $T : \Omega \to \Omega$ defined by

$$T(\omega) = T((\omega_1, \omega_2)) = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} (\omega_1, \omega_2)^T \mod 1.$$

One can show that T is measure-preserving with regard to μ. Conditions (C1) and (C2) are true for $C = \gamma = 1$ and (for example) $b = 2$. Condition (C3) is true for $c_1 = 2$ because 2 is the smallest absolute value of the eigenvalues of $A := \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$. The partition into expanding subsets is given by

$$\mathcal{M} := \{A^{-1}([i-1, i[\times[j-1, j[) \cap \Omega \mid i, j \in \{1, 2, 3, 4\}\}.$$

It is more difficult to verify condition (C4). We will show that condition (C4) holds true for $k = 1$ and $\varphi = 0$:

Diagonalizing A provides

$$A = \frac{1}{2} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}.$$

This implies

$$\mathcal{M}^{(n)} = \{A^{-n}([i-1, i[\times[j-1, j[) \cap \Omega \mid i, j \in \{1, 2, \dots, 4^n\}\}$$

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

$$\mathcal{F} := \{A^{-n}([i-1, i[\times[j-1, j[) \mid i, j \in \{1, 2, \dots, 4^n\} \text{ and } A^{-n}([i-1, i[\times[j-1, j[) \subseteq \Omega\}$$

The number of sets in $\mathcal{M}^{(n)} \backslash \mathcal{F}$ is proportional to the length of the circumference of the quadrangle $A^n([0,1[^2)$. Considering the diagonalization of A, one can see that this circumference is bounded by $c \cdot 4^n$ for some positive constant $c \in \mathbb{R}$. This implies

$$\begin{aligned} \mu\left(\Omega \setminus \bigcup_{F \in \mathcal{F}} F\right) &= \mu\left(\bigcup_{F \in \mathcal{M}^{(n)} \backslash \mathcal{F}} F\right) \leq \sum_{F \in \mathcal{M}^{(n)} \backslash \mathcal{F}} \mu(F) \\ &\leq \#\left(\mathcal{M}^{(n)} \setminus \mathcal{F}\right) \cdot \max_{F \in \mathcal{M}^{(n)}} \mu(F) \leq c \cdot 4^n \cdot 8^{-n} = c \cdot 2^{-n}. \end{aligned}$$

Therefore, (4.9) is fulfilled. Since $\mu(A) = 8^{-n}$ for all $A \in \mathcal{F}$, we have

$$\frac{\mu(A \cap T^{-n}(B)) - \mu(A)\mu(B)}{\mu(A)} = \frac{\mu(B)/det(\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}^n) - 8^{-n} \cdot \mu(B)}{\mu(A)} = 0.$$

for all $B \in \mathcal{A}$. So (4.10) holds true for $\varphi(n) = 0$ for all $n \in \mathbb{N}$.

To construct the partitions $\mathcal{Q}_n$ used in Theorem 4.6, we will set $n_\gamma := \lceil n^{4/\gamma} \rceil$ and $\mathcal{U}_n$ as

$$\mathcal{U}_n := \{[b(-1 + 2i/n_\gamma), b(-1 + 2(i+1)/n_\gamma)[\mid i \in \{0, 1, \dots n_\gamma - 1\}\} \tag{4.11}$$

Define

$$\mathcal{Q}_n := \bigvee_{j=1}^{d} p_j^{-1}(\mathcal{U}_n). \tag{4.12}$$

It becomes clear in the proof of Theorem 4.10 why this is a useful choice.

The sequence of partitions $\mathcal{P}_n^{(n)}$ in Theorem 4.6 will then be chosen as

$$\mathcal{P}_n := \mathcal{M}^{\lceil n^\beta \rceil} \tag{4.13}$$

with an appropriate $\beta \in]0,1[$. Looking at (4.5), it is easy to see that $(\mathcal{P}_n)_{n\in\mathbb{N}}$ satisfies (i) in Theorem 4.6. We will now show that (ii) in Theorem 4.6 holds true for the above choice of $\mathcal{Q}_n$ and $\mathcal{P}_n$.

Lemma 4.8. Let $(\Omega, \mathscr{B}, \mu, T)$ be a measure-preserving dynamical system satisfying conditions (C1),(C2) and (C3) and $(\mathcal{U}_n)_{n\in\mathbb{N}}$ a sequence of partitions as defined in (4.11). Then

$$\lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{j=1}^{d} \left(p_j^{-1}(\mathcal{U}_n)\right)^{(n)} \middle| \mathcal{P}_n^{(n)} \right) = 0$$

holds true, where $\mathcal{P}_n = \mathcal{M}^{\lceil n^\beta \rceil}$ for some $\beta \in [1/2, 1[$.

Proof. Using properties of the conditional entropy, we have

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{j=1}^{d} \left(p_j^{-1}(\mathcal{U}_n)\right)^{(n)} \middle| \mathcal{P}_n^{(n)} \right) \\
&= \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{j=1}^{d} \left(p_j^{-1}(\mathcal{U}_n)\right)^{(n)} \middle| \mathcal{M}^{(n+\lceil n^\beta \rceil)} \right) \\
&\le \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{j=1}^{d} \left(p_j^{-1}(\mathcal{U}_n)\right)^{(n)} \middle| \mathcal{M}^{(n+\lceil \sqrt{n} \rceil)} \right) \\
&\le \sum_{j=1}^{d} \lim_{n\to\infty} \frac{1}{n} H\left(\left(p_j^{-1}(\mathcal{U}_n)\right)^{(n)} \middle| \mathcal{M}^{(n+\lceil \sqrt{n} \rceil)} \right) \\
&\le \sum_{j=1}^{d} \lim_{n\to\infty} H\left(p_j^{-1}(\mathcal{U}_n) \middle| \mathcal{M}^{(\lceil \sqrt{n} \rceil)} \right).
\end{aligned}$$

Therefore, it remains to show that

$$\lim_{n\to\infty} H\left(p_j^{-1}(\mathcal{U}_n) \middle| \mathcal{M}^{(\lceil \sqrt{n} \rceil)} \right) = 0 \tag{4.14}$$

holds true for all $j \in \{1, 2, \dots, d\}$.

Take $j \in \{1, 2, \dots, d\}$ and set

$$\mathcal{U}_n^* := \left\{ \left] b(-1 + (2i + n^{4/\gamma - 5/\gamma^2})/n_\gamma), b(-1 + (2(i+1) - n^{4/\gamma - 5/\gamma^2})/n_\gamma) \right[\mid i \in \{0, 1, \dots n_\gamma - 1\} \right\}$$

and

$$W_n^* := \{M \in \mathcal{M}^{(\lceil\sqrt{n}\rceil)} \mid M \cap \bigcup_{U \in \mathcal{U}_n^*} p_j^{-1}(U) \neq \emptyset\}.$$

For a set $M \in \mathscr{B}$, define

$$V_n(M) := \{A \in p_j^{-1}(\mathcal{U}_n) \mid A \cap M \neq \emptyset\}.$$

Choose $N \in \mathbb{N}$ such that

$$2B \cdot c_1^{-\sqrt{N}} \leq b \cdot N^{-5/\gamma^2 - 1}$$

holds true. Condition (C1) and (C3) imply that the diameter of M is not larger than $2b \cdot c_1^{-\sqrt{n}}$ for all $M \in \mathcal{M}^{(\lceil\sqrt{n}\rceil)}$, so for all $n \geq N$ we have

$$\#V_n(M) \leq 2$$

for all $M \in \mathcal{M}^{(\lceil\sqrt{n}\rceil)}$ and

$$\#V_n(M) = 1$$

for all $M \in W_n^*$. This implies for all $n \geq N$

$$\begin{aligned}
& H\left(p_j^{-1}(\mathcal{U}_n) \middle| \mathcal{M}^{(\lceil\sqrt{n}\rceil)}\right) \\
\leq & \sum_{M \in \mathcal{M}^{(\lceil\sqrt{n}\rceil)}} \mu(M) \log(V_n(M)) \\
= & \sum_{M \in W_n^*} \mu(M) \log(V_n(M)) + \sum_{M \in \left(\mathcal{M}^{(\lceil\sqrt{n}\rceil)} \setminus W_n^*\right)} \mu(M) \log(V_n(M)) \\
\leq & \sum_{M \in W_n^*} \mu(M) \log(1) + \sum_{M \in \left(\mathcal{M}^{(\lceil\sqrt{n}\rceil)} \setminus W_n^*\right)} \mu(M) \log(2) \\
= & \log(2) \cdot \mu\left(\bigcup_{M \in \mathcal{M}^{(\lceil\sqrt{n}\rceil)} \setminus W_n^*} M\right) \\
\leq & \log(2) \cdot \mu\left(\Omega \setminus \bigcup_{U \in U_n^*} p_j^{-1}(U)\right) \\
= & \log(2) \cdot \Big[\mu\left(p_j^{-1}\left(\left[-b, b(-1 + n^{4/\gamma - 5/\gamma^2}/n_\gamma)\right]\right)\right) \\
& + \sum_{i=1}^{n_\gamma - 1} \mu\left(p_j^{-1}\left(\left[b(-1 + (2i - n^{4/\gamma - 5/\gamma^2})/n_\gamma), b(-1 + (2i + n^{4/\gamma - 5/\gamma^2})/n_\gamma)\right]\right)\right) \\
& + \mu\left(p_j^{-1}\left(\left[b - n^{4/\gamma - 5/\gamma^2}/n_\gamma, b\right]\right)\right)\Big] \\
\overset{(C2)}{\leq} & \log(2) \cdot \left[2C \cdot (b \cdot n^{-5/\gamma^2})^\gamma + (n_\gamma - 1) \cdot C \cdot (2b \cdot n^{-5/\gamma^2})^\gamma\right] \\
\leq & \log(2) \cdot C\left[2b^\gamma \cdot n^{-5/\gamma} + n^{4/\gamma} \cdot (2b)^\gamma \cdot n^{-5/\gamma}\right] \\
= & \log(2) \cdot C\left[2b^\gamma \cdot n^{-5/\gamma} + (2b)^\gamma \cdot n^{-1/\gamma}\right],
\end{aligned}$$

which converges to 0 for $n \to \infty$. So, one can conclude that (4.14) holds true. □

Given $\Omega \subseteq \mathbb{R}^d$, let

$$B(\omega, r) := \{\omega' \in \Omega \mid \|\omega' - \omega\| \leq r\}$$

be the ball with radius $r \geq 0$ centered at $\omega \in \Omega$ and

$$B(A, r) := \bigcup_{\omega \in A} B(\omega, r)$$

the r-parallel set around $A \subseteq \Omega$. In particular, if $\Omega \subseteq \mathbb{R}$ and $A = [a, b]$ is an interval, then

$$B(A, \delta) = [a - \delta, b + \delta].$$

We now want to show that (iii) in Theorem 4.6 holds true. To achieve this, we first introduce the following lemma.

Lemma 4.9. Let $(\Omega, \mathscr{B}, \mu, T)$ be a measure-preserving dynamical system satisfying conditions (C1),(C2),(C3) and (C4) with k and φ as in (4.10) and c_1 as in condition (C3). Let $\{U_i\}_{i\in I}$ be a collection of sets in $\mathbb{R}$. Then for all $\delta > 0$ and $t \geq -\frac{\log(\delta)}{\log(c_1)}$

$$\begin{aligned}
&\mu\left(\bigcup_{i\in I} p_j^{-1}(U_i) \cap T^{-kt}(p_j^{-1}(U_i))\right) \\
&\leq n^{-4} + \sum_{i\in I} \mu(p_j^{-1}(B(U_i, \delta))) \cdot (\mu(p_j^{-1}(B(U_i, \delta))) + \varphi(t))
\end{aligned}$$

holds true

Proof. Let c and $\mathcal{M}$ be as in condition (C3) and (C4). According to the definition of the mixing rate, there exists a $k \in \mathbb{N}$ such that for all $t \in \mathbb{N}$ there exists a collection of sets $\mathcal{F} \subseteq \mathcal{M}^{(t)}$ with

$$\mu\left(\Omega \setminus \bigcup_{F \in \mathcal{F}} F\right) \leq n^{-4}$$

and

$$\sup_{B \in \sigma(\bigvee_{i=0}^{n-1} T^{-i}(\mathcal{M}))} \mu(A \cap T^{-kn}(B)) - \mu(A)\mu(B) \leq \varphi(n) \cdot \mu(A)$$

for all $A \in \mathcal{F}$. We have

$$
\begin{aligned}
&\mu\left(\bigcup_{i\in I} p_j^{-1}(U_i) \cap T^{-kt}(p_j^{-1}(U_i))\right)\\
&\leq \mu\left(\bigcup_{i\in I}\left(\left(\bigcup_{\substack{M\in\mathcal{M}^{(t)}\\ M\cap p_j^{-1}(U_i)\neq\emptyset}} M\right)\cap T^{-kt}\left(\bigcup_{\substack{M\in\mathcal{M}^{(t)}\\ M\cap p_j^{-1}(U_i)\neq\emptyset}} M\right)\right)\right)\\
&= \mu\left(\bigcup_{i\in I}\left(\left(\left(\bigcup_{\substack{M\in\mathcal{F}\\ M\cap p_j^{-1}(U_i)\neq\emptyset}} M\right)\cup\left(\bigcup_{\substack{M\in\mathcal{M}^{(t)}\setminus\mathcal{F}\\ M\cap p_j^{-1}(U_i)\neq\emptyset}} M\right)\right)\right.\right.\\
&\qquad\left.\left.\cap T^{-kt}\left(\bigcup_{\substack{M\in\mathcal{M}^{(t)}\\ M\cap p_j^{-1}(U_i)\neq\emptyset}} M\right)\right)\right)\\
&\leq \mu\left(\bigcup_{i\in I}\left(\left(\bigcup_{\substack{M\in\mathcal{F}\\ M\cap p_j^{-1}(U_i)\neq\emptyset}} M\right)\cap T^{-kt}\left(\bigcup_{\substack{M\in\mathcal{M}^{(t)}\\ M\cap p_j^{-1}(U_i)\neq\emptyset}} M\right)\right.\right.\\
&\qquad\left.\left.\cup\left(\bigcup_{\substack{M\in\mathcal{M}^{(t)}\setminus\mathcal{F}\\ M\cap p_j^{-1}(U_i)\neq\emptyset}} M\right)\right)\right)\\
&\leq \mu\left(\left(\bigcup_{M\in\mathcal{M}^{(t)}\setminus\mathcal{F}} M\right)\cup\bigcup_{i\in I}\left(\left(\bigcup_{\substack{M\in\mathcal{F}\\ M\cap p_j^{-1}(U_i)\neq\emptyset}} M\right)\cap T^{-kt}\left(\bigcup_{\substack{M\in\mathcal{M}^{(t)}\\ M\cap p_j^{-1}(U_i)\neq\emptyset}} M\right)\right)\right)\\
&\leq \mu\left(\bigcup_{M\in\mathcal{M}^{(t)}\setminus\mathcal{F}} M\right)+\sum_{i\in I}\mu\left(\left(\bigcup_{\substack{M\in\mathcal{F}\\ M\cap p_j^{-1}(U_i)\neq\emptyset}} M\right)\cap T^{-kt}\left(\bigcup_{\substack{M\in\mathcal{M}^{(t)}\\ M\cap p_j^{-1}(U_i)\neq\emptyset}} M\right)\right)\\
&=\mu\left(\Omega\setminus\bigcup_{F\in\mathcal{F}} F\right)+\sum_{i\in I}\mu\left(\bigcup_{\substack{M\in\mathcal{F}\\ M\cap p_j^{-1}(U_i)\neq\emptyset}} M\cap T^{-kt}\left(\bigcup_{\substack{M'\in\mathcal{M}^{(t)}\\ M'\cap p_j^{-1}(U_i)\neq\emptyset}} M'\right)\right)\\
&\leq n^{-4}+\sum_{i\in I}\sum_{\substack{M\in\mathcal{F}\\ M\cap p_j^{-1}(U_i)\neq\emptyset}}\mu\left(M\cap T^{-kt}\left(\bigcup_{\substack{M'\in\mathcal{M}^{(t)}\\ M'\cap p_j^{-1}(U_i)\neq\emptyset}} M'\right)\right)
\end{aligned}
$$

$$\overset{(C4)}{\leq} n^{-4} + \sum_{i\in I} \sum_{\substack{M\in\mathcal{F} \\ M\cap p_j^{-1}(U_i)\neq\emptyset}} \mu(M) \cdot \left(\mu \left(\bigcup_{\substack{M'\in\mathcal{M}^{(t)} \\ M'\cap p_j^{-1}(U_i)\neq\emptyset}} M' \right) + \varphi(t) \right)$$

$$= n^{-4} + \sum_{i\in I} \mu \left(\bigcup_{\substack{M\in\mathcal{F} \\ M\cap p_j^{-1}(U_i)\neq\emptyset}} M \right) \cdot \left(\mu \left(\bigcup_{\substack{M'\in\mathcal{M}^{(t)} \\ M'\cap p_j^{-1}(U_i)\neq\emptyset}} M' \right) + \varphi(t) \right) \tag{4.15}$$

for all $t \in \mathbb{N}$. Now Choose $t \in \mathbb{N}$ with $t \geq -\frac{\log(\delta)}{\log(c_1)}$. This is equivalent to

$$\left(\frac{1}{c_1}\right)^t \leq \delta.$$

According to condition (C3), the diameter of the set M satisfies

$$\operatorname{diam}(M) \leq \left(\frac{1}{c_1}\right)^t$$

for all $M \in \mathcal{M}^{(t)}$, which implies

$$\operatorname{diam}(M) \leq \delta$$

for all $M \in \mathcal{M}^{(t)}$. Therefore,

$$\left(\bigcup_{\substack{M\in\mathcal{M}^{(t)} \\ M\cap p_j^{-1}(U_i)\neq\emptyset}} M \right) \subseteq p_j^{-1}(B(U_j, \delta))$$

holds true for all $i \in \{0, 1\}$, which, combined with (4.15) finishes the proof. □

Theorem 4.10. Let $(\Omega, \mathscr{B}, \mu, T)$ be a measure-preserving dynamical system satisfying conditions (C1),(C2),(C3) and (C4) and $(\mathcal{U}_n)_{n\in\mathbb{N}}$ a sequence of partitions as defined in (4.11). Then

$$\lim_{n\to\infty} \frac{H\left(\overline{OP}^j(n, \alpha) | \mathcal{Q}_n^{(n)}\right)}{n} = 0, \tag{4.16}$$

holds true for $\mathcal{Q}_n := p_j^{-1}(\mathcal{U}_n)$ and all $\alpha \in]0, 1[$ with $\lim_{n\to\infty} n^{3/\alpha} \cdot \varphi(n) = 0$.

Proof. For $n \in \mathbb{N}$ write $I_n := \{0, 1, \dots, n^4 - 1\}$ and

$$U_i^n := \left[b(-1 + 2i/n^4), b(-1 + 2(i+1)/n^4)\right[$$

for $i \in I_n$, so $\mathcal{U}_n = \{U_i^n\}_{i\in I_n}$. Further, $Q_i^n := p_j^{-1}(U_i^n)$, so $\mathcal{Q}_n = \{Q_i^n\}_{i\in I_n}$.
Take $\alpha \in]0, 1[$. Define for $n \in \mathbb{N}$, $i \in I_n$ and some $\varepsilon > 0$ the set

$$A_n(i) := \{\omega \in \Omega \mid \#\{(s, t) \in E_n \mid T^s(\omega), T^t(\omega) \in Q_i^n \text{ and } |t - s| \geq n^\alpha\} \leq n_0\}$$

where $n_0 := \lfloor \varepsilon \cdot \sqrt{n} \rfloor$. Notice that for all $n \in \mathbb{N}$ and $\boldsymbol{i} = (i_0, i_1, \dots, i_{n-1}) \in I_n^n$ with $Q^n(\boldsymbol{i}) \subseteq \bigcap_{i \in I_n} A_n(i)$

$$\#\{(s,t) \in E_n \mid i_s = i_t \text{ and } |t-s| \geq n^\alpha\} \leq n_0$$

holds true. Combining Lemma 3.3 and Lemma 4.3 provides

$$\begin{aligned}
&H(\overline{OP}^j(n,\alpha)|\mathcal{Q}^{(n)}) \leq \sum_{\boldsymbol{i} \in I^n} \mu(Q(\boldsymbol{i})) \cdot \log\left(\#\Delta\left(\overline{OP}^j(n,\alpha)|Q(\boldsymbol{i})\right)\right) \\
&\leq \log(2) \sum_{\boldsymbol{i} \in I^n} \mu(Q(\boldsymbol{i})) \cdot \#\{(s,t) \in \{0,1,...,n-1\}^2 | \ i_s = i_t \text{ and } |t-s| \geq n^\alpha\} \\
&= \log(2) \sum_{\substack{\boldsymbol{i} \in I^n: \\ Q(\boldsymbol{i}) \subseteq \bigcap_{i \in I_n} A_n(i)}} \mu(Q(\boldsymbol{i})) \cdot \#\{(s,t) \in \{0,1,...,n-1\}^2 | \ i_s = i_t \text{ and } |t-s| \geq n^\alpha\} \\
&\quad + \log(2) \sum_{\substack{\boldsymbol{i} \in I^n: \\ Q(\boldsymbol{i}) \not\subseteq \bigcap_{i \in I_n} A_n(i)}} \mu(Q(\boldsymbol{i})) \cdot \#\{(s,t) \in \{0,1,...,n-1\}^2 | \ i_s = i_t \text{ and } |t-s| \geq n^\alpha\} \\
&\leq \log(2) \sum_{\substack{\boldsymbol{i} \in I^n: \\ Q(\boldsymbol{i}) \subseteq \bigcap_{i \in I_n} A_n(i)}} \mu(Q(\boldsymbol{i})) \cdot n_0 + \log(2) \sum_{\substack{\boldsymbol{i} \in I^n: \\ Q(\boldsymbol{i}) \not\subseteq \bigcap_{i \in I_n} A_n(i)}} \mu(Q(\boldsymbol{i})) \cdot n^2 \\
&\leq \log(2) \cdot n_0 + \log(2) \cdot \left(1 - \mu\left(\bigcap_{i \in I_n} A_n(i)\right)\right) \cdot n^2.
\end{aligned}$$

After dividing by n and because ε can be arbitrarily close to 0, it remains to show that

$$\lim_{n\to\infty} \left(1 - \mu\left(\bigcap_{i \in I_n} A_n(i)\right)\right) \cdot n = \lim_{n\to\infty} \mu\left(\bigcup_{i \in I_n} A_n(i)^c\right) \cdot n = 0$$

holds true. We have

$$\begin{aligned}
A_n(i)^c &= \{\omega \in \Omega \mid \#\{(s,t) \in E_n \mid T^s(\omega), T^t(\omega) \in Q_i^n \text{ and } |t-s| \geq n^\alpha\} > n_0\} \\
&\subseteq \{\omega \in \Omega \mid \#\{(s,t) \in E_n \mid T^s(\omega), T^t(\omega) \in Q_i^n \text{ and } |t-s| \geq n^\alpha\} > 0\} \\
&= \bigcup_{\substack{(s,t) \in E_n \\ |t-s| \geq n^\alpha}} T^{-s}(Q_i^n) \cap T^{-t}(Q_i^n) \\
&\subseteq \bigcup_{s=1}^{n} T^{-s}\left(\bigcup_{t=\lceil n^\alpha \rceil}^{n-1} Q_i^n \cap T^{-t}(Q_i^n)\right),
\end{aligned}$$

which implies

$$
\begin{aligned}
\mu\left(\bigcup_{i\in I_n} A_n(i)^c\right) &\leq \mu\left(\bigcup_{i\in I_n}\bigcup_{s=1}^{n} T^{-s}\left(\bigcup_{t=\lceil n^\alpha\rceil}^{n-1} Q_i^n \cap T^{-t}(Q_i^n)\right)\right)\\
&= \mu\left(\bigcup_{s=1}^{n} T^{-s}\left(\bigcup_{t=\lceil n^\alpha\rceil}^{n-1}\bigcup_{i\in I_n} Q_i^n \cap T^{-t}(Q_i^n)\right)\right)\\
&\leq n\cdot\mu\left(\bigcup_{t=\lceil n^\alpha\rceil}^{n-1}\bigcup_{i\in I_n} Q_i^n \cap T^{-t}(Q_i^n)\right)\\
&\leq n\cdot\sum_{t=\lceil n^\alpha\rceil}^{n-1}\mu\left(\bigcup_{i\in I_n} Q_i^n \cap T^{-t}(Q_i^n)\right)
\end{aligned}
\tag{4.17}
$$

Let $c_1 > 1$ be as in condition (C3) and set $\delta_n := n^{-8/\gamma^2}$. Chose $N \in \mathbb{N}$ such that

$$
N^\alpha \geq -\frac{\log(\delta_N)}{\log(c_1)}\cdot k = \frac{8k\log(N)}{\gamma^2\log(c_1)}
$$

holds true. Applying Lemma 4.9 to (4.17) provides for all $n \geq N$

$$
\begin{aligned}
\mu\left(\bigcup_{i\in I_n} A_n(i)^c\right) &\leq n\cdot\sum_{t=\lceil n^\alpha\rceil}^{n-1}\mu\left(\bigcup_{i\in I_n} Q_i^n \cap T^{-t}(Q_i^n)\right)\\
&= n\cdot\sum_{t=\lceil n^\alpha\rceil}^{n-1}\mu\left(p_j^{-1}(U_i^n)\cap T^{-t}(p_j^{-1}(U_i^n))\right)\\
&\leq n\cdot\sum_{t=\lceil n^\alpha\rceil}^{n-1} n^{-4} + \sum_{i\in I_n}\mu\left(p_j^{-1}(B_{\delta_n}(U_i^n))\right)\cdot\left(\mu\left(p_j^{-1}(B_{\delta_n}(U_i^n))\right)+\varphi(\lfloor t/k\rfloor)\right)\\
&\leq n^{-2} + n\cdot\sum_{t=\lceil n^\alpha\rceil}^{n-1}\sum_{i\in I_n}\left(\mu\left(p_j^{-1}(U_i^n)\right)+\mu\left(p_j^{-1}(B_{\delta_n}(U_i^n)\setminus U_n^i)\right)\right)\\
&\quad\cdot\left(\mu\left(p_j^{-1}(B_{\delta_n}(U_i^n))\right)+\varphi(\lfloor t/k\rfloor)\right)\\
&\overset{(C2)}{\leq} n^{-2} + n\cdot\sum_{i\in I_n}\sum_{t=\lceil n^\alpha\rceil}^{n-1}\left(\mu\left(p_j^{-1}(U_i^n)\right)+2C\cdot(\delta_n)^\gamma\right)\cdot\left(\mu\left(p_j^{-1}(B_{\delta_n}(U_i^n))\right)+\varphi(\lfloor t/k\rfloor)\right)\\
&= n^{-2} + n\cdot\sum_{i\in I_n}\sum_{t=\lceil n^\alpha\rceil}^{n-1}\left(\mu\left(p_j^{-1}(U_i^n)\right)+2C\cdot n^{-8/\gamma}\right)\cdot\left(\mu\left(p_j^{-1}(B_{\delta_n}(U_i^n))\right)+\varphi(\lfloor t/k\rfloor)\right)
\end{aligned}
$$

$$
\begin{aligned}
&\leq n^{-2} + n^2 \cdot \sum_{i\in I_n} \left(\mu\left(p_j^{-1}(U_i^n)\right) + 2C\cdot n^{-8/\gamma}\right)\cdot\left(\mu\left(p_j^{-1}(B_{\delta_n}(U_i^n))\right) + \varphi(\lfloor n^\alpha/k\rfloor)\right)\\
&\leq n^{-2} + 2C\cdot n^{-6/\gamma}\cdot \# I_n \cdot (1+\varphi(\lfloor n^\alpha/k\rfloor))\\
&\quad + n^2\cdot\sum_{i\in I_n}\mu\left(p_j^{-1}(U_i^n)\right)\cdot\left(\max_{i'\in I_n}(\mu(p_j^{-1}(B_{\delta_n}(U_{i'}^n)))) + \varphi(\lfloor n^\alpha/k\rfloor)\right)\\
&\leq n^{-2} + 2C\cdot n^{-6/\gamma}\cdot\lceil n^{4/\gamma}\rceil\cdot(1+\varphi(\lfloor n^\alpha/k\rfloor)) + n^2(\cdot\max_{i'\in I_n}(\mu(p_j^{-1}(B_{\delta_n}(U_{i'}^n)))) + \varphi(\lfloor n^\alpha/k\rfloor))\\
&\leq n^{-2} + 2C\cdot n^{-6/\gamma}\cdot(n^{4/\gamma}+1)\cdot(1+\varphi(\lfloor n^\alpha/k\rfloor)) + n^2\cdot(C\cdot(2(1+\delta_n)/n_\gamma)^\gamma + \varphi(\lfloor n^\alpha/k\rfloor))\\
&\leq n^{-2} + 2C\cdot(n^{-2/\gamma}+n^{-6/\gamma})\cdot(1+\varphi(\lfloor n^\alpha/k\rfloor)) + n^2\cdot(C\cdot(4\cdot n^{-4/\gamma})^\gamma + \varphi(\lfloor n^\alpha/k\rfloor))\\
&\leq n^{-2} + 2C\cdot(n^{-2/\gamma}+n^{-6/\gamma})\cdot(1+\varphi(\lfloor n^\alpha/k\rfloor)) + C\cdot 4^\gamma\cdot n^{-2} + n^2\cdot\varphi(\lfloor n^\alpha/k\rfloor).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\lim_{n\to\infty}\mu\left(\bigcup_{i\in I_n} A_n(i)^c\right)\cdot n\\
\leq\ &\lim_{n\to\infty} n^{-1} + 2C\cdot(n^{1-2/\gamma}+n^{1-6/\gamma})\cdot(1+\varphi(\lfloor n^\alpha/k\rfloor)) + C\cdot 4^\gamma\cdot n^{-1} + n^3\cdot\varphi(\lfloor n^\alpha/k\rfloor)\\
=\ &\lim_{n\to\infty} n^3\cdot\varphi(\lfloor n^\alpha/k\rfloor) = k^{3/\alpha}\cdot\lim_{n\to\infty}\left(\left(\frac{n}{k^{1/\alpha}}\right)^\alpha\right)^{3/\alpha}\cdot\varphi\left(\left\lfloor\left(\frac{n}{k^{1/\alpha}}\right)^\alpha\right\rfloor\right) = 0.
\end{aligned}
$$

□

If φ decreases exponentially, like in example 7, (4.16) holds true for all $\alpha \in]0,1[$.

4.3.2 Lower α-permutation entropy

To show (iv) in Theorem 4.6, we require that our measure-preserving dynamical system fulfills two further conditions.

(C5) The map T Lipschitz continuous on each set M, i.e. there exists a constant $c_2 > 1$ such that

$$\|T(\omega) - T(\nu)\| \leq c_2\|\omega - \nu\|$$

holds true for all $\omega, \nu \in M \in \mathcal{M}$.

(C6) the measure μ is absolutely continuous with regard to the d-dimensional Lebesgue measure λ and its density $d\mu/d\lambda$ satisfies

$$\sup_{n\in\mathbb{N}}\mu(D_n)\cdot n < \infty$$

where

$$D_n := \{\omega\in\Omega \mid (d\mu/d\lambda)(\omega) \geq n\}.$$

Notice that those conditions are both true for dynamical system given in example 7.

Condition (C5) guarantees that the deformation caused by the map T is bounded. Condition (C3) guarantees the same for the inverse of T, which is defined piecewise on each set $M \in \mathcal{M}$. The following lemma is a consequence of this bounded deformation.

Lemma 4.11. Let $T : \Omega \to \Omega$ be a map that satisfies condition (C3) for a constant $c_1 > 1$ and a partition $\mathcal{M}$. Consider the map $S : \Omega \to \Omega$ with $S(\omega) := \omega - T(\omega)$ and the on $A \subseteq \Omega$ restricted map $S_{|A} : A \to S(A)$ with $S_{|A}(\omega) := S(\omega)$ for all $\omega \in A$. Then the following holds true:

(i) For all $M \in \mathcal{M}$
$$\|S_{|M}(\omega) - S_{|M}(\nu)\| \geq (c_1 - 1)\|\omega - \nu\|$$
holds true for all $\omega, \nu \in M$,

(ii) $S_{|M}$ is injective for all $M \in \mathcal{M}$,

(iii) $S_{|M}^{-1}$ is Lipschitz continuous for all $M \in \mathcal{M}$, i.e
$$\|S_{|M}^{-1}(\omega) - S_{|M}^{-1}(\nu)\| \leq \frac{1}{c_1 - 1}\|\omega - \mu\|$$
holds true for all $\omega, \nu \in S(M)$.

If, additionally, T satisfies condition (C5), then

(iv) for all $M \in \mathcal{M}$
$$\|S_{|M}(\omega) - S_{|M}(\nu)\| \leq (c_2 + 1)\|\omega - \mu\|$$
holds true for all $\omega, \nu \in M$,

(v) for all $M \in \mathcal{M}$
$$\|S_{|M}^{-1}(\omega) - S_{|M}^{-1}(\nu)\| \geq \frac{1}{c_2 + 1}\|\omega - \mu\|$$
holds true for all $\omega, \nu \in S(M)$.

Proof. Fix $M \in \mathcal{M}$. The triangle inequality implies

$$\begin{aligned}\|S_{|M}(\omega) - S_{|M}(\nu)\| &= \|\omega - T(\omega) - \nu + T(\nu)\| \\ &= \|(T(\nu) - T(\omega)) - (\nu - \omega)\| \\ &\geq \|T(\nu) - T(\omega)\| - \|\nu - \omega\| \\ &\overset{(C3)}{\geq} (c_1 - 1)\|\nu - \omega\|\end{aligned}$$

and

$$\begin{aligned}\|S_{|M}(\omega) - S_{|M}(\nu)\| &= \|\omega - T(\omega) - \nu + T(\nu)\| \\ &= \|(T(\nu) - T(\omega)) + (\omega - \nu)\| \\ &\leq \|T(\nu) - T(\omega)\| + \|\omega - \nu\| \\ &\overset{(C5)}{\leq} (c_2 + 1)\|\nu - \omega\|\end{aligned}$$

for all $\omega, \nu \in M$. Therefore, (i) and (iv) are true. Statement (ii) follows immediately from (i). We will now show (iii) and (v). Take $\omega, \nu \in S(M)$. Since $S_{|M}$ is injective according to (ii), there exist unique elements $\omega', \nu' \in M$ with $S_{|M}(\omega') = \omega$ and $S_{|M}(\nu') = \nu$. Thus,

$$\begin{aligned}\|S_{|M}^{-1}(\omega) - S_{|M}^{-1}(\nu)\| &= \|\omega' - \nu'\| \\ &\overset{(i)}{\leq} \frac{1}{c_1 - 1}\|S_{|M}(\omega') - S_{|M}(\nu')\| \\ &= \frac{1}{c_1 - 1}\|\omega - \nu\|\end{aligned}$$

and

$$\begin{aligned}\|S_{|M}^{-1}(\omega) - S_{|M}^{-1}(\nu)\| &= \|\omega' - \nu'\| \\ &\overset{(iv)}{\geq} \frac{1}{c_2 + 1}\|S_{|M}(\omega') - S_{|M}(\nu')\| \\ &= \frac{1}{c_2 + 1}\|\omega - \nu\|\end{aligned}$$

hold true □

Notice that the above lemma holds true for T^t instead of T as well by replacing the constants c_1 and c_2 with c_1^t and c_2^t and $\mathcal{M}$ with $\mathcal{M}^{(t)}$.

Given $t \in \mathbb{N}$ and two disjoint one-dimensional intervals $I_1, I_2 \subseteq \mathbb{R}$, set $M := I_1 \cap T^{-t}(I_2)$. Then either

$$\omega < T^t(\omega)$$

holds true for all $\omega \in \mathbb{R}$ with $\omega \in M$ or

$$\omega > T^t(\omega)$$

holds true for all $\omega \in M$. This exclusivity of the order relations allows to deduce in what ordinal pattern ω is located in based on the intervals that $T^t(\omega)$ is located in for different $t \in \mathbb{N}_0$. In the higher dimensional case of $\Omega \subseteq \mathbb{R}^d$, it is more difficult to construct a set $M \subseteq \Omega$ such that for $j \in \{1, 2, \ldots, d\}$ either

$$p_j(\omega) < p_j(T^t(\omega))$$

holds true for all $\omega \in M$ or

$$p_j(\omega) > p_j(T^t(\omega))$$

holds true for all $\omega \in M$. As we will show in the following lemma, this exclusivity of order relation for certain points in higher dimensions is guaranteed if those points are not too close to the set $\{\omega \in \Omega \mid p_j(\omega) = p_j(T^t(\omega))\}$. How far away from this sets those points need to be mainly depends on the constants c_1, c_2 in Conditions (C3) and (C5). The proof of this lemma makes use of the properties deduced from bounded distortion that were established in the above lemma.

Lemma 4.12. Let $(\Omega, \mathscr{B}, \mu, T)$ be a measure-preserving dynamical system satisfying conditions (C1), (C3) and (C5). Then for all $n, t \in \mathbb{N}$ with $n \geq t$ and all sets $M \in \mathcal{M}^{(n)}$ and $M' \in \mathcal{M}^{(t)}$ with

$$M \subseteq M'$$

and

$$M \nsubseteq B(\{\omega \in M' \mid p_j(\omega) = p_j(T^t(\omega))\}, 2b(c_2^t/(c_1-1)+1)c_1^{-n})$$

we have either

$$p_j(\omega) < p_j(T^t(\omega))$$

for all $\omega \in M$ or

$$p_j(\omega) > p_j(T^t(\omega))$$

for all $\omega \in M$.

Proof. Take $n, t \in \mathbb{N}$ with $t \leq n$, $M \in \mathbb{N}$ and $M' \in \mathcal{M}^{(t)}$ with

$$M \nsubseteq B(\{\omega \in M' \mid p_j(\omega) = p_j(T^t(\omega))\}, 2b(c_2^t/(c_1-1)+1)c_1^{-n}).$$

Let $c_1 > 1$ be the expansion constant in condition (C3). Notice that $\operatorname{diam}(M) \leq 2bc_1^{-n}$ holds true, which implies

$$M \cap B(\{\omega \in M' \mid p_j(\omega) = p_j(T^t(\omega))\}, 2bc_2^t c_1^{-n}/(c_1-1)) = \emptyset$$

and thus

$$M \cap B(\{\omega \in M' \mid p_j(\omega) = p_j(T^t(\omega))\}, 2bc_2^t c_1^{-n}/(c_1-1)) \cap M' = \emptyset. \tag{4.18}$$

Consider the function $S_{|M'} : M' \to S_{|M'}(M')$ with $S_{|M'}(\omega) = \omega - T^t(\omega)$ for all $\omega \in M'$. We have

$$\begin{aligned}
&B(\{\omega \in M' \mid p_j(\omega) = p_j(T^t(\omega))\}, 2bc_2^t c_1^{-n}/(c_1-1)) \cap M' \\
&= B(\{\omega \in M' \mid p_j(\omega - T^t(\omega)) = 0\}, 2bc_2^t c_1^{-n}/(c_1-1)) \cap M' \\
&= B(\{\omega \in M' \mid \omega - T^t(\omega) \in p_j^{-1}(\{0\})\}, 2bc_2^t c_1^{-n}/(c_1-1)) \cap M' \\
&= B(S_{|M'}^{-1}(p_j^{-1}(\{0\})), 2bc_2^t c_1^{-n}/(c_1-1)) \cap M'.
\end{aligned}$$

Lemma 4.11 (iii) provides

$$\begin{aligned}
&B(S_{|M'}^{-1}(p_j^{-1}(\{0\})), 2bc_2^t c_1^{-n}/(c_1-1)) \cap M' \\
&\supseteq B(S_{|M'}^{-1}(p_j^{-1}(\{0\})), 2bc_2^t c_1^{-n}/(c_1^t-1)) \cap M' \\
&\supseteq S_{|M'}^{-1}(B(p_j^{-1}(\{0\}), 2bc_2^t c_1^{-n})) \cap M' \\
&= S_{|M'}^{-1}(B(p_j^{-1}(\{0\}), 2bc_2^t c_1^{-n})) \\
&= S_{|M'}^{-1}(p_j^{-1}(B(\{0\}, 2bc_2^t c_1^{-n}))) \\
&= \{\omega \in \mathcal{M}' : \left|p_j(\omega) - p_j(T^t(\omega))\right| \leq 2bc_2^t c_1^{-n}\}
\end{aligned}$$

Using (4.18), this implies

$$M \cap \{\omega \in \mathcal{M}' \mid \left|p_j(\omega) - p_j(T^t(\omega))\right| \leq 2bc_2^t c_1^{-n}\} = \emptyset.$$

Because $M \subseteq M'$ holds true, this is equivalent to

$$M \subseteq \{\omega \in \mathcal{M}' \mid \left|p_j(\omega) - p_j(T^t(\omega))\right| > 2bc_2^t c_1^{-n}\}.$$

Now suppose that the conclusion of this lemma does not hold true, i.e. there exists $\omega_1, \omega_2 \in M$ with

$$p_j(\omega_1) \leq p_j(T^t(\omega_1)) \qquad \text{and} \qquad p_j(\omega_2) \geq p_j(T^t(\omega_2)).$$

Using $\omega_1, \omega_2 \in \{\omega \in \mathcal{M}' \mid \left|p_j(\omega) - p_j(T^t(\omega))\right| > 2bc_2^t c_1^{-n}\}$ provides

$$p_j(T^t(\omega_1)) - p_j(\omega_1) > 2bc_2^t c_1^{-n} \qquad \text{and} \qquad p_j(\omega_2) - p_j(T^t(\omega_2)) > 2bc_2^t c_1^{-n}.$$

This implies

$$\begin{aligned}
&\left|p_j(T^t(\omega_1)) - p_j(T^t(\omega_2))\right| \\
&= \left|(p_j(T^t(\omega_1)) - p_j(\omega_1) + p_j(\omega_2) - p_j(T^t(\omega_2))) - (p_j(\omega_2) - p_j(\omega_1))\right| \\
&\geq \left|p_j(T^t(\omega_1)) - p_j(\omega_1) + p_j(\omega_2) - p_j(T^t(\omega_2))\right| - |p_j(\omega_2) - p_j(\omega_1)| \\
&> 4bc_2^t c_1^{-n} - |p_j(\omega_2) - p_j(\omega_1)| \\
&\geq 4bc_2^t c_1^{-n} - \|\omega_2 - \omega_1\| \\
&\geq 4bc_2^t c_1^{-n} - \operatorname{diam}(M) \\
&\geq 4bc_2^t c_1^{-n} - 2bc_1^{-n} \\
&\geq 2bc_2^t c_1^{-n}.
\end{aligned} \tag{4.19}$$

On the other hand

$$\begin{aligned}
&\left|p_j(T^t(\omega_1)) - p_j(T^t(\omega_2))\right| \\
&\leq \|T^t(\omega_1) - T^t(\omega_2)\| \\
&\overset{(C5)}{\leq} c_2^t \cdot \|\omega_1 - \omega_2\| \\
&\leq c_2^t \cdot \operatorname{diam}(M) \\
&\leq 2bc_2^t c_1^{-n}
\end{aligned}$$

holds true, which is a contradiction to (4.19). □

Now we can establish an upper bound on the conditional lower α-permutation entropy based on the measure of the sets considered in the lemma above.

Lemma 4.13. Let $(\Omega, \mathscr{B}, \mu, T)$ be a measure-preserving dynamical system satisfying conditions (C1), (C3) and (C5). Then for all $0 < \alpha < \beta$

$$\begin{aligned}
&\frac{1}{n} H\left(\underline{OP}^j(n, \alpha) | \mathcal{M}^{(n+\lceil n^\beta \rceil)}\right) \\
&\leq \log(2) \sum_{t=1}^{\lfloor n^\alpha \rfloor} \sum_{M' \in \mathcal{M}^{(t)}} \mu(B(\{\omega \in M' \mid p_j(\omega) = p_j(T^t(\omega))\}, 2b(c_2^t/(c_1 - 1) + 1)c_1^{-n^\beta}) \cap M')
\end{aligned}$$

holds true, where c_1 is the expansion constant given in condition (C3).

Proof. For all $n \in \mathbb{N}$ and $0 < \alpha < \beta$, we have

$$\begin{aligned}
&\frac{1}{n} H\left(\underline{OP}^j(n,\alpha)|\mathcal{M}^{(n+\lceil n^\beta \rceil)}\right)\\
&\leq \frac{1}{n} H\left(\bigvee_{s=0}^{n-1} T^{-s}\left(\bigvee_{t=1}^{\lfloor n^\alpha \rfloor}(\mathrm{id},T)^{-t}(\mathcal{R}_j)\right)|\mathcal{M}^{(n+\lceil n^\beta \rceil)}\right)\\
&\leq \frac{1}{n}\sum_{s=0}^{n-1} H\left(T^{-s}\left(\bigvee_{t=1}^{\lfloor n^\alpha \rfloor}(\mathrm{id},T)^{-t}(\mathcal{R}_j)\right)|\mathcal{M}^{(n+\lceil n^\beta \rceil)}\right)\\
&\leq \frac{1}{n}\sum_{s=0}^{n-1} H\left(T^{-s}\left(\bigvee_{t=1}^{\lfloor n^\alpha \rfloor}(\mathrm{id},T)^{-t}(\mathcal{R}_j)\right)|T^{-s}\left(\mathcal{M}^{(\lceil n^\beta \rceil)}\right)\right)\\
&= H\left(\bigvee_{t=1}^{\lfloor n^\alpha \rfloor}(\mathrm{id},T)^{-t}(\mathcal{R}_j)|\mathcal{M}^{(\lceil n^\beta \rceil)}\right)\\
&\leq \sum_{t=1}^{\lfloor n^\alpha \rfloor} H\left((\mathrm{id},T)^{-t}(\mathcal{R}_j)|\mathcal{M}^{(\lceil n^\beta \rceil)}\right).
\end{aligned}$$

Consider now a fixed $t \in \{1,2,\ldots,\lfloor n^\alpha \rfloor\}$. Recall that

$$\Delta((\mathrm{id},T)^{-t}(\mathcal{R}_j)|M) = \{P \in (\mathrm{id},T)^{-t}(\mathcal{R}_j) \mid \mu(P\cap M) > 0\}$$

for all sets $M \in \mathscr{B}$ and set

$$B_{n,t}(M') := B(\{\omega \in M' \mid p_j(\omega) = p_j(T^t(\omega))\}, 2b(c_2^t/(c_1-1)+1)c_1^{-n^\beta})$$

for all $t \in \{1,2,\ldots,\lfloor n^\alpha \rfloor\}$ and $M' \in M^{(t)}$. If $M \not\subseteq B_{n,t}(M')\cap M'$ holds true for $M \in \mathcal{M}^{(\lceil n^\beta \rceil)}$ with $M \subseteq M'$, then Lemma 4.12 provides

$$\Delta((\mathrm{id},T)^{-t}(\mathcal{R}_j)|M) = 1.$$

Hence,

$$\begin{aligned}
&H\left((\mathrm{id},T)^{-t}(\mathcal{R}_j)|\mathcal{M}^{(\lceil n^\beta \rceil)}\right)\\
\leq &\sum_{M\in\mathcal{M}^{(\lceil n^\beta \rceil)}} \mu(M)\cdot\log(\#\Delta((\mathrm{id},T)^{-t}(\mathcal{R}_j)|M))\\
= &\sum_{M'\in\mathcal{M}^{(t)}}\sum_{M\in\mathcal{M}^{(\lceil n^\beta \rceil)}} \mu(M\cap M')\cdot\log(\#\Delta((\mathrm{id},T)^{-t}(\mathcal{R}_j)|M))\\
= &\sum_{M'\in\mathcal{M}^{(t)}}\sum_{\substack{M\in\mathcal{M}^{(\lceil n^\beta \rceil)}\\ M\subseteq B_{n,t}(M')\cap M'}} \mu(M\cap M')\cdot\log(\#\Delta((\mathrm{id},T)^{-t}(\mathcal{R}_j)|M))\\
&+\sum_{M'\in\mathcal{M}^{(t)}}\sum_{\substack{M\in\mathcal{M}^{(\lceil n^\beta \rceil)}\\ M\not\subseteq B_{n,t}(M')\cap M'}} \mu(M\cap M')\cdot\log(\#\Delta((\mathrm{id},T)^{-t}(\mathcal{R}_j)|M))\\
\leq &\log(2)\cdot\sum_{M'\in\mathcal{M}^{(t)}}\sum_{\substack{M\in\mathcal{M}^{(\lceil n^\beta \rceil)}\\ M\subseteq B_{n,t}(M')\cap M'}} \mu(M\cap M')\\
\leq &\log(2)\cdot\sum_{M'\in\mathcal{M}^{(t)}} \mu(B_{n,t}(M')\cap M') \qquad \square
\end{aligned}$$

So according to the above lemma, we now need to show that

$$\log(2)\sum_{t=1}^{\lfloor n^\alpha\rfloor}\sum_{M'\in\mathcal{M}^{(t)}}\mu(B(\{\omega\in M'\mid p_j(\omega)=p_j(T^t(\omega))\},2b(c_2^t/(c_1-1)+1)c_1^{-n^\beta})\cap M')$$

converges to 0 for $n\to\infty$ to show that the conditional entropy converges to 0. To achieve this, we establish an upper bound for the measure of those sets considered in the above sum by, additionally, using Condition (C6).

Lemma 4.14. Let $(\Omega,\mathscr{B},\mu,T)$ be a measure-preserving dynamical system satisfying condition (C1), (C3), (C5) and (C6). Then

$$\begin{aligned}&\mu(B(\{\omega\in M'\mid p_j(\omega)=p_j(T^t(\omega))\},2b(c_2^t/(c_1-1)+1)c_1^{-n})\cap M'\setminus D_k)\\&\le k(2b)^d(c_2^t+1)^2(c_2^t/(c_1-1)+1)c_1^{-n}\end{aligned}$$

holds true for all $k,n,t\in\mathbb{N}$, $M'\in\mathcal{M}^{(t)}$ and $\beta>0$ with

$$D_k=\{\omega\in\Omega\mid (d\mu/d\lambda)(\omega)\ge k\}$$

as defined in (C6).

Proof. We have

$$\begin{aligned}&B(\{\omega\in M'\mid p_j(\omega)=p_j(T^t(\omega))\},2b(c_2^t/(c_1-1)+1)c_1^{-n})\cap M'\\&=B(\{\omega\in M'\mid p_j(\omega-T^t(\omega))=0\},2b(c_2^t/(c_1-1)+1)c_1^{-n})\cap M'\\&=B(\{\omega\in M'\mid \omega-T^t(\omega)\in p_j^{-1}(\{0\})\},2b(c_2^t/(c_1-1)+1)c_1^{-n})\cap M'\\&=B(S_{|M'}^{-1}(p_j^{-1}(\{0\})),2b(c_2^t/(c_1-1)+1)c_1^{-n})\cap M'.\end{aligned}$$

Using Lemma 4.11 (v), this implies

$$\begin{aligned}&B(\{\omega\in M'\mid p_j(\omega)=p_j(T^t(\omega))\},2b(c_2^t/(c_1-1)+1)c_1^{-n})\cap M'\\&\subseteq S_{|M'}^{-1}(B(p_j^{-1}(\{0\}),2b(c_2^t/(c_1-1)+1)c_1^{-n}(c_2^t+1)))\cap M'\\&=S_{|M'}^{-1}(B(p_j^{-1}(\{0\}),2b(c_2^t/(c_1-1)+1)c_1^{-n}(c_2^t+1)))\\&=S_{|M'}^{-1}(p_j^{-1}(B(0,2b(c_2^t/(c_1-1)+1)c_1^{-n}(c_2^t+1)))).\end{aligned}$$

Therefore,

$$\begin{aligned}&\mu(B(\{\omega\in M'\mid p_j(\omega)=p_j(T^t(\omega))\},2b(c_2^t/(c_1-1)+1)c_1^{-n})\cap M'\setminus D_k)\\&\le\mu(S_{|M'}^{-1}(p_j^{-1}(B(0,2b(c_2^t/(c_1-1)+1)c_1^{-n}(c_2^t+1))))\setminus D_k)\\&\le k\lambda(S_{|M'}^{-1}(p_j^{-1}(B(0,2b(c_2^t/(c_1-1)+1)c_1^{-n}(c_2^t+1))))\setminus D_k)\\&\le k\lambda(S_{|M'}^{-1}(p_j^{-1}(B(0,2b(c_2^t/(c_1-1)+1)c_1^{-n}(c_2^t+1)))))\\&\le k(c_2^t+1)\lambda(p_j^{-1}(B(0,2b(c_2^t/(c_1-1)+1)c_1^{-n}(c_2^t+1))))\\&\le k(c_2^t+1)2b(c_2^t/(c_1-1)+1)c_1^{-n}(c_2^t+1)(2b)^{d-1}\\&=k(2b)^d(c_2^t+1)^2(c_2^t/(c_1-1)+1)c_1^{-n}\end{aligned}$$

holds true. ∎

The upper bound established in the above lemma now allows us to show that the measure of those sets considered in Lemma 4.13 converges to 0.

Corollary 4.15. Let $(\Omega, \mathscr{B}, \mu, T)$ be a measure-preserving dynamical system satisfying conditions (C1), (C3), (C5) and (C6). Then

$$\lim_{n\to\infty} \sum_{t=1}^{\lfloor n^\alpha \rfloor} \sum_{M' \in \mathcal{M}^{(t)}} \mu(B(\{\omega \in M' \mid p_j(\omega) = p_j(T^t(\omega))\}, 2b(c_2^t/(c_1-1)+1)c_1^{-n^\beta}) \cap M') = 0,$$

holds true for all $0 < \alpha < \beta$.

Proof. Let $0 < \alpha < \beta$ and choose some $\alpha' \in (\alpha, \beta)$. Set

$$c := 2(\#\mathcal{M})(c_1+1)^2(c_2/\min(c_1-1,1)+1).$$

Lemma 4.14 provides

$$\begin{aligned}
&\lim_{n\to\infty} \sum_{t=1}^{\lfloor n^\alpha \rfloor} \sum_{M' \in \mathcal{M}^{(t)}} \mu(B(\{\omega \in M' \mid p_j(\omega) = p_j(T^t(\omega))\}, 2b(c_2^t/(c_1-1)+1)c_1^{-n^\beta}) \cap M') \\
= &\lim_{n\to\infty} \sum_{t=1}^{\lfloor n^\alpha \rfloor} \sum_{M' \in \mathcal{M}^{(t)}} \Big[\mu(B(\{\omega \in M' \mid p_j(\omega) = p_j(T^t(\omega))\}, 2b(c_2^t/(c_1-1)+1)c_1^{-n^\beta}) \cap M' \setminus D_{c^{n^{\alpha'}}}) \\
&+ \mu(B(\{\omega \in M' \mid p_j(\omega) = p_j(T^t(\omega))\}, 2b(c_2^t/(c_1-1)+1)c_1^{-n^\beta}) \cap M' \cap D_{c^{n^{\alpha'}}})\Big] \\
\leq &\lim_{n\to\infty} \sum_{t=1}^{\lfloor n^\alpha \rfloor} \sum_{M' \in \mathcal{M}^{(t)}} \Big[c^{n^{\alpha'}}(2b)^d(c_2^t+1)^2(c_2^t/(c_1-1)+1)c_1^{-n^\beta} + \mu(D_{c^{n^{\alpha'}}})\Big] \\
\leq &\lim_{n\to\infty} \sum_{t=1}^{\lfloor n^\alpha \rfloor} (\#\mathcal{M})^t \Big[c^{n^{\alpha'}}(2b)^d(c_2^t+1)^2(c_2^t/(c_1-1)+1)c_1^{-n^\beta} + \mu(D_{c^{n^{\alpha'}}})\Big] \\
\leq &\lim_{n\to\infty} n^\alpha (\#\mathcal{M})^{n^\alpha} \Big[c^{n^{\alpha'}}(2b)^d(c_2^{n^\alpha}+1)^2(c_2^{n^\alpha}/(c_1-1)+1)c_1^{-n^\beta} + \mu(D_{c^{n^{\alpha'}}})\Big] \\
\leq &\lim_{n\to\infty} (2b)^d c^{(n^\alpha + n^{\alpha'})} c_1^{-n^\beta} + c^{n^\alpha}\mu(D_{c^{n^{\alpha'}}}) \\
= &\lim_{n\to\infty} c^{n^\alpha}\mu(D_{c^{n^{\alpha'}}}) \\
= &\lim_{n\to\infty} c^{(n^\alpha - n^{\alpha'})} c^{n^{\alpha'}}\mu(D_{c^{n^{\alpha'}}}) \\
\leq &(\sup_{n\in\mathbb{N}} c^{n^{\alpha'}}\mu(D_{c^{n^{\alpha'}}})) \lim_{n\to\infty} c^{(n^\alpha - n^{\alpha'})} \\
\overset{(C6)}{=} &\, 0.
\end{aligned}$$

□

We now can use Theorem 4.6 to prove the main result of this chapter:

Theorem 4.16. Let $(\Omega, \mathscr{B}, \mu, T)$ be a measure-preserving dynamical system satisfying conditions (C1)-(C6). Then

$$\limsup_{n\to\infty} \frac{1}{n} H(OP(n)) \leq h(T)$$

holds true

Proof. Choose some $\alpha \in]0,1[$ that satisfies the mixing condition (C4) and $\beta \in]0,1[$ with $\beta > \max(\alpha, 1/2)$. Take sequences of partitions $(\mathcal{Q}_n)_{n\in\mathbb{N}}$ and $(\mathcal{P}_n)_{n\in\mathbb{N}}$ as given in (4.12) and (4.13). Then (i)-(iv) in Theorem 4.6 are satisfied:

Lemma 4.8 implies (ii). Theorem 4.10 shows that (iii) holds true. And Lemma 4.13 together with Corollary 4.15 proves (iv). And finally, because of

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} H(\mathcal{P}_n^{(n)}) \\
&= \lim_{n\to\infty} \frac{1}{n} H(\mathcal{M}^{(n+\lfloor n^\beta \rfloor)}) \\
&= \lim_{n\to\infty} \frac{n+\lfloor n^\beta \rfloor}{n} \cdot \frac{1}{n+\lfloor n^\beta \rfloor} H(\mathcal{M}^{(n+\lfloor n^\beta \rfloor)}) \\
&\leq \lim_{n\to\infty} \frac{n+n^\beta}{n} \cdot \frac{1}{n+\lfloor n^\beta \rfloor} H(\mathcal{M}^{(n+\lfloor n^\beta \rfloor)}) \\
&\leq \lim_{n\to\infty} (1+n^{\beta-1}) \cdot \frac{1}{n+\lfloor n^\beta \rfloor} H(\mathcal{M}^{(n+\lfloor n^\beta \rfloor)}) \\
&= \lim_{n\to\infty} \frac{1}{n+\lfloor n^\beta \rfloor} H(\mathcal{M}^{(n+\lfloor n^\beta \rfloor)}) = h(T,\mathcal{M}) \leq h(T),
\end{aligned}$$

(i) holds true. □

5 Conclusion

In chapter 3, we showed that the permutation and Kolmogorov-Sinai entropy are equal for one-dimensional dynamical systems that are countably piecewise monotone. This is a generalization of a previous result which stated that these entropies are equal if the system is monotone and continuous on a finite number of intervals. It is still an open question how or whether the conditions that allow this to be true can be further generalized. The condition of monotony was crucial for the main arguments of the corresponding proofs. Removing it would therefore require a completely different approach. Theoretically, it could also be possible that the condition of monotony is strictly necessary and can not be removed. Unfortunately, no example of a dynamical system for which permutation and Kolmogorov-Sinai entropy are not equal has been found yet. Finding such an example could help to clarify what the most general conditions for one-dimensional systems are under which the K-S and permutation entropy coincide.

One method to find such an example, which has not been explored in this thesis, is to investigate how applying measure-theoretic isomorphisms to a dynamical system affects the permutation entropy of the system. One central property of the Kolmogorov-Sinai entropy is the fact that this entropy is invariant under such isomorphisms. Finding an isomorphism that changes the value of the permutation entropy would provide an example of a dynamical system for which the value of the K-S and permutation entropy are different. To construct such an isomorphism, one could try to use that fact that isomorphisms, which per definition preserve the measure-theoretic structure of a dynamical system, do not necessarily preserve the order structure of the system. However, considering the results of chapter 3, this isomorphism can not be a simple piecewise monotone function but needs to be more complex. Considering functions as isomorphisms whose graph have a fractal-like shape could potentially provide interesting examples.

Further, we investigated permutation entropy based on Rényi entropies. It turns out that the value of these entropies is typically different from the value of the permutation entropy based on the Shannon entropy or the K-S entropy. For most cases, it is still an open question how one can analytically determine the value of the permutation entropy based on Rényi entropies. In practice, the fact that the permutation entropy is possibly different when based on Rényi entropies can be an advantage since it allows to calculate a whole spectrum of potentially different permutation entropies for one dynamical system. This spectrum of permutation entropies could be used to better classify the complexity of such systems. How this advantage can be fully utilized in practice is subject to further research.

While the permutation entropy based on Rényi entropies is typically not equal to the Kolmogorov-Sinai entropy based on Rényi entropies, it is more directly related to the supremum of the entropy rates based on Rényi entropy over all interval partitions. As a consequence of this fact, we showed that taking the supremum of the entropy rates based on Rényi entropies over all interval partitions typically does not yield the same result compared to taking the supremum over all partitions. This is important to keep in mind for practical calculations since, due to limited precision, only interval partitions can be realized on a computer.

At the the end of chapter 3, we showed that the permutation entropy, conditional permutation entropy and K-S entropy are equal under certain conditions. This is an important result because the conditional permutation entropy typically converges faster to the K-S entropy of a dynamical system than the original permutation entropy does. Knowing that these entropies are all equal justifies using the conditional permutation entropy in practical application for which fast convergence is relevant.

As part of the proof of some results related to the conditional permutation entropy, we considered a sequence of measures of specific sets named V_n. It was previously shown that the K-S and permutation entropy would be equal if this sequence converged to zero fast enough. However, we were able to show that this sequence always converges too slowly. Therefore, a more refined approach is needed to show the equality between permutation and K-S when trying to use a similar approach. But it is more likely that simply considering the measure of the sets V_n is a too rough approach to show the equality between permutation and K-S entropy.

In chapter 4, we established conditions under which the multidimensional permutation and K-S entropy are equal. The established conditions are relatively specific but the first conditions of this kind that allow to prove such an equality. It is likely that these condition can be further generalized or simplified. For example, one could try to combine conditions (C2) and (C5) into one conditions that will only depend on how the measure of the systems is affected by the dynamic T without making any assumptions about how T is affecting the metric of the system. Additionally, it would be interesting to try to weaken the mixing condition given by (C4). It could possibly be enough te require that the system is simply mixing as defined in (3.59). Interestingly, the dynamical system being mixing also played a central role when proving certain results for the conditional permutation entropy (see Lemma 3.50).

As mentioned in chapter 4, if the considered map of the dynamical system is continuously differentiable, condition (C3) is satisfied if every eigenvalue of the Jacobian is larger than some constant $c > 1$. Therefore, it would be interesting to see if applying methods from the theory of smooth dynamical system can be helpful to generalize the results of this chapter. For smooth dynamical systems, the so called Lyapunov exponent can be used to quantify the system's complexity. One can show that under certain conditions the Lyapunov exponent and the Kolmogorov-Sinai entropy are equal [26]. Whether a similar direct connection between the Lyapunov exponent and the permutation entropy can be created is a question worth further research.

Finally, the methods discussed in chapter 4 can potentially be used to find results for one-dimensional systems that are not necessarily (countably) piecewise monotone since they do not make use of the fact that the system is, in some sense, order preserving.

Appendix

Proof of Lemma 2.16. Let p be a stochastic vector of length $n \in \mathbb{N} \cup \{\infty\}$. Then

$$H(p,q) = \frac{-1}{q-1} \log\left(\sum_{i=1}^{n} p_i^q\right) = -\frac{f(q)}{g(q)}$$

with $f(q) := \log(\sum_{i=1}^{n} p_i^q)$ and $g(q) := q-1$ for all $q \in \mathbb{R}$. The derivatives are

$$f'(q) = \frac{\sum_{i=1}^{n} \log(p_i) p_i^q}{\sum_{i=1}^{n} p_i^q}$$

and

$$g'(q) = 1.$$

L'Hospital's rule then implies

$$\lim_{q\to 1} H(p,q) = -\lim_{q\to 1} \frac{f(q)}{g(q)} = -\lim_{q\to 1} \frac{f'(q)}{g'(q)} = -\lim_{q\to 1} \frac{\sum_{i=1}^{n} \log(p_i) p_i^q}{\sum_{i=1}^{n} p_i^q} = -\sum_{i=1}^{n} \log(p_i) p_i = H(p).$$

□

Proof of Lemma 2.17. Let p be a stochastic vector of length $n \in \mathbb{N} \cup \{\infty\}$ and $q_1, q_2 \in \mathbb{R} \setminus \{1\}$. Consider $f : [0,\infty) \to \mathbb{R}$ with

$$f(x) = x^{\frac{q_1-1}{q_2-1}}.$$

The first and second derivative are given by

$$f'(x) = \frac{q_1-1}{q_2-1} x^{\frac{q_1-q_2}{q_2-1}}$$

and

$$f''(x) = \frac{(q_1-1)\cdot(q_1-q_2)}{(q_2-1)^2} x^{\frac{q_1-2q_2+1}{q_2-1}}.$$

Suppose $q_1 \le q_2 < 1$. Then $f''(x) \ge 0$ for all $x \in [0,\infty)$ which implies that f is convex. Using Jensen's inequality yields

$$\left(\sum_{i=1}^{n} p_i^{q_2}\right)^{\frac{q_1-1}{q_2-1}} = f\left(\sum_{i=1}^{n} p_i \cdot p_i^{q_2-1}\right) \le \sum_{i=1}^{n} p_i f\left(p_i^{q_2-1}\right) = \sum_{i=1}^{n} p_i \cdot p_i^{q_1-1} = \sum_{i=1}^{n} p_i^{q_1}$$

which is equivalent to

$$\begin{aligned}
&\left(\sum_{i=1}^{n} p_i^{q_2}\right)^{q_1-1} \geq \left(\sum_{i=1}^{n} p_i^{q_1}\right)^{q_2-1} \\
\Leftrightarrow\ & (q_1-1)\log\left(\sum_{i=1}^{n} p_i^{q_2}\right) \geq (q_2-1)\log\left(\sum_{i=1}^{n} p_i^{q_1}\right) \\
\Leftrightarrow\ & \frac{1}{q_2-1}\log\left(\sum_{i=1}^{n} p_i^{q_2}\right) \geq \frac{1}{q_1-1}\log\left(\sum_{i=1}^{n} p_i^{q_1}\right) \\
\Leftrightarrow\ & \frac{-1}{q_2-1}\log\left(\sum_{i=1}^{n} p_i^{q_2}\right) \leq \frac{-1}{q_1-1}\log(\sum_{i=1}^{n} p_i^{q_1}) \\
\Leftrightarrow\ & H(p,q_2) \leq H(p,q_1).
\end{aligned}$$

Thus, $H(p,q)$ is monotonically decreasing for $q < 1$.

Now suppose $1 < q_1 \leq q_2$ holds true. Then $f''(x) \leq 0$ for all $x \in [0,\infty)$, which implies that f is concave. Jensen's inequality implies

$$\left(\sum_{i=1}^{n} p_i^{q_2}\right)^{\frac{q_1-1}{q_2-1}} = f\left(\sum_{i=1}^{n} p_i \cdot p_i^{q_2-1}\right) \geq \sum_{i=1}^{n} p_i f\left(p_i^{q_2-1}\right) = \sum_{i=1}^{n} p_i \cdot p_i^{q_1-1} = \sum_{i=1}^{n} p_i^{q_1}$$

which is equivalent to

$$\begin{aligned}
&\left(\sum_{i=1}^{n} p_i^{q_2}\right)^{q_1-1} \geq \left(\sum_{i=1}^{n} p_i^{q_1}\right)^{q_2-1} \\
\Leftrightarrow\ & (q_1-1)\log\left(\sum_{i=1}^{n} p_i^{q_2}\right) \geq (q_2-1)\log\left(\sum_{i=1}^{n} p_i^{q_1}\right) \\
\Leftrightarrow\ & \frac{1}{q_2-1}\log\left(\sum_{i=1}^{n} p_i^{q_2}\right) \geq \frac{1}{q_1-1}\log\left(\sum_{i=1}^{n} p_i^{q_1}\right) \\
\Leftrightarrow\ & \frac{-1}{q_2-1}\log\left(\sum_{i=1}^{n} p_i^{q_2}\right) \leq \frac{-1}{q_1-1}\log\left(\sum_{i=1}^{n} p_i^{q_1}\right) \\
\Leftrightarrow\ & H(p,q_2) \leq H(p,q_1).
\end{aligned}$$

Thus, $H(p,q)$ is monotonically decreasing for $q < 1$. For $q_1 < 1 < q_2$, the continuity of $H(p,q)$ (Lemma 2.16) implies

$$H(p,q_1) \geq \lim_{q\uparrow 1} H(p,q) = H(p,1) = \lim_{q\downarrow 1} H(p,q) \geq H(p,q_2).$$

So $H(p,q)$ is monotonically decreasing for all $q \in \mathbb{R}$. □

Proof of Lemma 2.21. We will first show $H(\mathcal{P} \vee \mathcal{Q}) \leq H(\mathcal{P}) + H(\mathcal{Q})$ which is equivalent to

$H(\mathcal{P}) + H(\mathcal{Q}) - H(\mathcal{P} \vee \mathcal{Q}) \geq 0$:

$$\begin{aligned}
&H(\mathcal{P}) + H(\mathcal{Q}) - H(\mathcal{P} \vee \mathcal{Q}) \\
=& -\sum_{P\in\mathcal{P}} \mu(P) \log \mu(P) - \sum_{Q\in\mathcal{Q}} \mu(Q) \log \mu(Q) + \sum_{P\in\mathcal{P}} \sum_{Q\in\mathcal{Q}} \mu(Q\cap P) \log \mu(Q \cap P) \\
=& -\sum_{P\in\mathcal{P}} \sum_{Q\in\mathcal{Q}} \mu(Q\cap P) \log \left(\frac{\mu(P)\mu(Q)}{\mu(Q\cap P)} \right) \\
\geq& -\log \left(\sum_{P\in\mathcal{P}} \sum_{Q\in\mathcal{Q}} \mu(Q\cap P) \frac{\mu(P)\mu(Q)}{\mu(Q\cap P)} \right) \\
=& -\log(1) = 0.
\end{aligned}$$

where we used the convexity of $-\log$ combined with Jensen's inequality to show the above inequality. □

Proof of Lemma 2.25. Let $\mathcal{Q} = \{Q_i\}_{i\in I}$ be a refinement of $\mathcal{P} = \{P_j\}_{j\in J}$, so for every $j \in J$ there exists an index set $I_j \subseteq I$ with

$$\mu \left(P_j \triangle \bigcup_{i\in I_j} Q_i \right) = 0,$$

which implies

$$\mu(P_j) = \mu \left(\bigcup_{i\in I_j} Q_i \right).$$

We now investigate the cases $q < 1$, $q > 1$ and $q = 1$ individually:

For $q < 1$ we have

$$\sum_{j\in J} \mu(P_j)^q = \sum_{j\in J} \mu \left(\bigcup_{i\in I_j} Q_i \right)^q = \sum_{j\in J} \left(\sum_{i\in I_j} \mu(Q_i) \right)^q \leq \sum_{j\in J} \left(\sum_{i\in I_j} \mu(Q_i)^q \right) = \sum_{i\in I} \mu(Q_i)^q$$

Hence, using the monotony of the logarithm,

$$H(\mu, q, \mathcal{P}) = \frac{-1}{q-1} \log \left(\sum_{j\in J} \mu(P_j)^q \right) \leq \frac{-1}{q-1} \log \left(\sum_{i\in I} \mu(Q_i)^q \right) = H(\mu, q, \mathcal{P}).$$

For $q > 1$ we have

$$\sum_{j\in J} \mu(P_j)^q = \sum_{j\in J} \mu \left(\bigcup_{i\in I_j} Q_i \right)^q = \sum_{j\in J} \left(\sum_{i\in I_j} \mu(Q_i) \right)^q \geq \sum_{j\in J} \left(\sum_{i\in I_j} \mu(Q_i)^q \right) = \sum_{i\in I} \mu(Q_i)^q$$

which implies

$$H(\mu, q, \mathcal{P}) = \frac{-1}{q-1} \log \left(\sum_{j\in J} \mu(P_j)^q \right) \leq \frac{-1}{q-1} \log \left(\sum_{i\in I} \mu(Q_i)^q \right) = H(\mu, q, \mathcal{P}).$$

For $q = 1$ the statements follows from the continuity of the Rényi entropy in q (Lemma 2.16). □

List of Symbols

$C_s^t(a_s, ..., a_t)$	Cylinder set	18
$(f_1 \times f_2)(\omega_1, \omega_2)$	$= (f_1(\omega_1), f_2(\omega_2))$	23
$(f_1, f_2)(\omega)$	$= (f_1(\omega), f_2(\omega))$	23
$\mathscr{A}_1 \underset{\mu}{=} \mathscr{A}_2$	Equality between σ-algebras	17
$\mathscr{A}_1 \underset{\mu}{\subseteq} \mathscr{A}_2$	Relation between σ-algebras	17
$\bigvee_{k=1}^{l} \mathcal{P}_k$	Refinement of multiple partitions	14
$\mathbb{P}^c(\mathscr{A})$	Collection of finite or countable partitions with finite entropy	14
$\operatorname{diam}(A)$	Diameter of a set A	108
id	Identity maps	23
$\mathcal{P} \prec \mathcal{Q}$	$\mathcal{Q}$ is a refinement of $\mathcal{P}$	16
$\mathcal{P} \vee \mathcal{Q}$	Refinement of two partitions	14
$\mathcal{R}_X$	$= (X \times X)^{-1}\{R, \mathbb{R}^2 \setminus R\}$	24
$\mathscr{E}$	Collection of invariant sets, i.e $\{E \in \mathscr{B} \mid T^{-1}(E) = E\}$	54
μ^2	Product measure: $\mu^2(A \times B) = \mu(A) \cdot \mu(B)$	32
$\mu_\omega^{\mathscr{A}}$	Conditional measure	53
$\mathbb{P}_o^c(\mathscr{B})$	Collection of countable or finite ordered partitions with finite entropy	31
$OP^{\boldsymbol{X}}(n)$	Partition into ordinal patterns	23
$\mathbb{P}_o(\mathscr{B})$	Collection of finite ordered partitions	31
$\overline{OP}^i(n, \alpha)$	Upper ordinal patterns for parameter $\alpha \in [0, 1]$	90
$\underline{OP}^i(n, \alpha)$	Lower ordinal patterns for parameter $\alpha \in [0, 1]$	90
$\overline{\mathrm{PE}}_\mu(\boldsymbol{X}, T, q)$	Upper permutation entropy	25
$\mathbb{P}(\mathscr{A})$	Collection of finite partitions	14
Π	Set of periodic points	10
Π_n	Set of points with period $\leq n$	10
$\mathfrak{S}_n$	Set of all permutations of $(0, 1, \ldots, n-1)$	23
$\mathcal{P}^{(n)}$	Partition into symbolic patterns of length $n \in \mathbb{N}$	15

$\sigma(\boldsymbol{X})$	σ-algebra generated by random variable $\boldsymbol{X}$	17
$\sigma(\mathcal{P})$	σ-algebra generated by partition $\mathcal{P}$	17
$\underline{\mathrm{PE}}_\mu(\boldsymbol{X}, T, q)$	Lower permutation entropy	25
$A \triangle B$	Symmetric difference	10
$B(\omega, r)$	Ball centered at ω with radius r	106
$B(A, r)$	r-parallel set of A	106
$H(p)$	Shannon entropy	13
$H(p, q)$	Rényi entropy for parameter q	13
$h(T)$	Kolmogorov-Sinai entropy	16
$h(T, \mathcal{P})$	Entropy rate	16
$h(T, \mathcal{P}, q)$	(generalized) entropy rate for parameter $q \in \mathbb{R}$	15
$h(T, q)$	(generlized) Kolmogorov-Sinai entropy for parameter $q \in \mathbb{R}$	16
$P(\boldsymbol{i})$	Symbolic pattern given a partition $\{P_i\}_{i \in I}$ and multi index $\boldsymbol{i} \in I^n$	15
$P_\pi^{X_i}$	Set of points with ordinal pattern π	23
p_i	Projection on the i-th coordinate	89
R	$= \{(x, y) \in \mathbb{R}^2 \mid x < y\}$	23
V_n	No-in-between sets	79

Bibliography

[1] J. M. Amigó and K. Keller, Permutation entropy: One concept, two approaches, *European Physical Journal Special Topics* **222**.2 (June 2013), 263–273.

[2] J. M. Amigó, M. B. Kennel, and L. Kocarev, The permutation entropy rate equals the metric entropy rate for ergodic information sources and ergodic dynamical systems, *Physica D: Nonlinear Phenomena* **210**.1 (2005), 77–95.

[3] A. Antoniouk, K. Keller, and S. Maksymenko, *Kolmogorov-Sinai entropy via separation properties of order-generated σ-algebras*, 2014, DOI: `10.3934/dcds.2014.34.1793`, URL: `http://aimsciences.org//article/id/9e37dcdf-90e1-4262-965e-1c1d569ae434`.

[4] C. Bandt, G. Keller, and B. Pompe, Entropy of interval maps via permutations, *Nonlinearity* **15**.5 (2002), 1595.

[5] C. Bandt and B. Pompe, Permutation Entropy: A Natural Complexity Measure for Time Series, *Phys. Rev. Lett.* **88** (17 Apr. 2002), 174102.

[6] R. Beals, *Analysis: An Introduction*, Cambridge University Press, 2004, 216, DOI: `10.1017/CBO9780511755163`.

[7] M. Einsiedler and T. Ward, *Ergodic Theory: with a view towards Number Theory*, Graduate Texts in Mathematics, Springer London, 2010.

[8] M. Einsiedler and T. Ward, *Ergodic Theory: with a view towards Number Theory*, **259**, Graduate Texts in Mathematics, Springer London, 2010.

[9] M. Einsiedler, E. Lindenstrauss, and T. Ward, *Entropy in ergodic theory and homogeneous dynamics*, aufgerufen am July 5, 2017, URL: `https://tbward0.wixsite.com/books/entropy`.

[10] S. Elizalde, The Number of Permutations Realized By a Shift, *SIAM Journal on Discrete Mathematics* **23**.2 (2009), 765–786, DOI: `10.1137/080726689`, eprint: `https://doi.org/10.1137/080726689`, URL: `https://doi.org/10.1137/080726689`.

[11] E. Evgeny Verbitskiy, "Generalized entropies in dynamical systems", English, Relation: http://www.rug.nl/ date_submitted:2000 Rights: University of Groningen, PhD thesis, 2000.

[12] T. Gutjahr and K. Keller, Equality of Kolmogorov-Sinai and permutation entropy for one-dimensional maps consisting of countably many monotone parts, *Discrete & Continuous Dynamical Systems* **39**.7 (2019), 4207–4224.

[13] T. Gutjahr et al., Quantification of fracture roughness by change probabilities and Hurst exponents, *Mathematical Geoscience* (2021), submitted.

[14] J. L. W. V. Jensen, Sur les fonctions convexes et les inégalités entre les valeurs moyennes, *Acta Mathematica* **30**.none (1906), 175–193, DOI: `10.1007/BF02418571`, URL: `https://doi.org/10.1007/BF02418571`.

[15] G. Keller, *Equilibrium States in Ergodic Theory*, London Mathematical Society Student Texts, Cambridge University Press, 1998.

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

[17] K. Keller, A. M. Unakafov, and V. A. Unakafova, Ordinal Patterns, Entropy, and EEG, *Entropy* **16**.12 (2014), 6212–6239, ISSN: 1099-4300, DOI: 10.3390/e16126212, URL: https://www.mdpi.com/1099-4300/16/12/6212.

[18] K. Keller et al., Permutation Entropy: New Ideas and Challenges, *Entropy* **19**.3 (2017), ISSN: 1099-4300, DOI: 10.3390/e19030134, URL: https://www.mdpi.com/1099-4300/19/3/134.

[19] K. Keller et al., Permutation Entropy: New Ideas and Challenges, *Entropy* **19**.3 (2017), ISSN: 1099-4300, DOI: 10.3390/e19030134, URL: https://www.mdpi.com/1099-4300/19/3/134.

[20] A. N. Kolmogorov, A new metric invariant of transitive dynamical systems and of automorphisms of Lebesgue spaces (Russian), *Dokl. Akad. Nauk SSSR (N.S.)* **119**.5 (1958), 861–864.

[21] X. Li, G. Ouyang, and D. A. Richards, Predictability analysis of absence seizures with permutation entropy, *Epilepsy Research* **77**.1 (2007), 70–74.

[22] M. Misiurewicz and W. Szlenk, Entropy of piecewise monotone mappings, eng, *Studia Mathematica* **67**.1 (1980), 45–63, URL: http://eudml.org/doc/218304.

[23] M. Misiurewicz, Permutations and topological entropy for interval maps, *Nonlinearity* **16**.3 (2003), 971.

[24] N. Nicolaou and J. Georgiou, The Use of Permutation Entropy to Characterize Sleep Electroencephalograms, *Clinical EEG and Neuroscience* **42**.1 (2011), 24–28.

[25] N. Nicolaou and J. Georgiou, Detection of epileptic electroencephalogram based on Permutation Entropy and Support Vector Machines, *Expert Systems with Applications* **39**.1 (2012), 202–209, ISSN: 0957-4174, DOI: https://doi.org/10.1016/j.eswa.2011.07.008, URL: https://www.sciencedirect.com/science/article/pii/S0957417411009705.

[26] J. B. Pesin, Characteristic Lyapunov exponents and smooth ergodic theory, English, *Russ. Math. Surv.* **32**.4 (1977), 55–114, ISSN: 0036-0279, DOI: 10.1070/RM1977v032n04ABEH001639.

[27] C. Quintero-Quiroz et al., Numerical and experimental study of the effects of noise on the permutation entropy, *New Journal of Physics* **17**.9 (Sept. 2015), 093002, DOI: 10.1088/1367-2630/17/9/093002, URL: https://doi.org/10.1088/1367-2630/17/9/093002.

[28] C. E. Shannon, A mathematical theory of communication, *The Bell System Technical Journal* **27**.3 (1948), 379–423, DOI: 10.1002/j.1538-7305.1948.tb01338.x.

[29] A. Silva et al., Comparison of Anesthetic Depth Indexes Based on Thalamocortical Local Field Potentials in Rats, *Anesthesiology* **112**.2 (2010), 355–363.

[30] Y. G. Sinai, On the notion of entropy of a dynamical system, *(Doklady) Acad. Sci. URSS (N. S.)* **124**.4 (1959), 768–771.

[31] M. Sinn, A. Ghodsi, and K. Keller, Detecting Change-Points in Time Series by Maximum Mean Discrepancy of Ordinal Pattern Distributions, *Uncertainty in Artificial Intelligence - Proceedings of the 28th Conference, UAI 2012* (Oct. 2012).

[32] M. Sinn and K. Keller, Estimation of ordinal pattern probabilities in Gaussian processes with stationary increments, *Computational Statistics & Data Analysis* **55**.4 (2011), 1781–1790, ISSN: 0167-9473, DOI: https://doi.org/10.1016/j.csda.2010.11.009, URL: https://www.sciencedirect.com/science/article/pii/S0167947310004378.

[33] I. Stolz and K. Keller, A General Symbolic Approach to Kolmogorov-Sinai Entropy, *Entropy* **19**.12 (2017), ISSN: 1099-4300, DOI: `10.3390/e19120675`, URL: `https://www.mdpi.com/1099-4300/19/12/675`.

[34] F. Takens and E. Verbitski, Generalized entropies: Rényi and correlation integral approach, *Nonlinearity* **11**.4 (1998), 771.

[35] F. Takens and E. Verbitskiy, Renyi entropies of aperiodic dynamical systems, English, *Israel journal of mathematics* **127** (2002), Relation: http://www.rug.nl/informatica/organisatie/overorganisatie/iwi Rights: University of Groningen. Research Institute for Mathematics and Computing Science (IWI), 279–302, ISSN: 0021-2172.

[36] F. Takens and E. Verbitskiy, Rényi entropies of aperiodic dynamical systems, *Israel Journal of Mathematics* **127**.1 (Dec. 2002), 279–302.

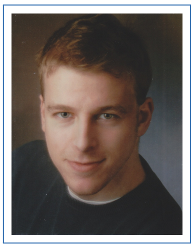
[37] A. M. Unakafov and K. Keller, Conditional entropy of ordinal patterns, *Physica D: Nonlinear Phenomena* **269** (2014), 94–102, ISSN: 0167-2789, DOI: `https://doi.org/10.1016/j.physd.2013.11.015`, URL: `http://www.sciencedirect.com/science/article/pii/S0167278913003242`.

[38] V. Unakafova, A. Unakafov, and K. Keller, An approach to comparing Kolmogorov-Sinai and permutation entropy, *The European Physical Journal Special Topics* **222**.2 (June 2013), 353–361, ISSN: 1951-6401, DOI: `10.1140/epjst/e2013-01846-7`, URL: `https://doi.org/10.1140/epjst/e2013-01846-7`.

[39] P. Walters, *An Introduction to Ergodic Theory*, Graduate Texts in Mathematics, Springer New York, 2000.

[40] P. Walters, *An introduction to ergodic theory*, **79**, Graduate Texts in Mathematics, Springer-Verlag, New York 1982.

[41] Y. Yang et al., Epileptic Seizure Prediction Based on Permutation Entropy, *Frontiers in Computational Neuroscience* **12** (2018), 55, DOI: `10.3389/fncom.2018.00055`.



Tim Gutjahr

Personal Life

1993 born Lübeck, Germany

Education

1999–2012 **Abitur**, *Thomas-Mann Schule*, Lübeck.

2012–2015 **Bachelor of Science, Mathematik in Medizin und Lebenswissenschaften**, *Universität zu Lübeck*, Lübeck.

2015–2017 **Master of Science, Mathematik in Medizin und Lebenswissenschaften**, *Universität zu Lübeck*, Lübeck.

Work Experience

2015-2016 **Student Research Assistant**, *Universität zu Lübeck*, Lübeck.

2017–2021 **Research Assistant**, *Universität zu Lübeck*, Lübeck.

Publications

2019 Gutjahr, T., Keller, K; *Equality of Kolmogorov-Sinai and permutation entropy for one-dimensional maps consisting of countably many monotone parts*; Discrete Contin. Dyn. Syst. A 39 (2019), 4207-4224

2020 Gutjahr, T., Keller, K; *Ordinal Pattern Based Entropies and the Kolmogorov–Sinai Entropy: An Update*; Entropy 2020, 22(1), 63

2021 Gutjahr, T., Keller, K.; *Generalized ordinal patterns and the KS-entropy Entropy*; Entropy 2021, 23(8), 1097

2021 Gutjahr, T., Hale, S., Keller, K., Blum., P., Winter, S.; *Quantification of fracture roughness by change probabilities and Hurst exponents*; Mathematical Geosciences, submitted