MASTERARBEIT / MASTER THESIS

Titel der Masterarbeit / Title of the Master Thesis

"Law of Series"

verfasst von / submitted by

Michael Unterlercher, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2022 / Vienna, 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 821
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Mathematik UG2002
Betreut von / Supervisor:	Assoz. Prof. Mag. Dr. Roland Zweimüller, Privatdoz.

Abstract

In der vorliegenden Arbeit behandeln wir das Gesetz der Serie, eine Eigenschaft bestimmter stochastischer Prozesse, in welchen Ereignisse nicht unabhängig voneinander, sondern tendentiell in Gruppen auftreten.
Wir erwähnen zunächst einige bekannte Resultate aus der Ergodentheorie, bevor wir das Gesetz der Serie als mögliche Eigenschaft homogener Signalprozesse einführen. Für den Hauptteil betrachten wir eine ergodische Abbildung mit positiver dynamischer Entropie hinsichtlich einer endlichen Partition. Basierend auf [2] beweisen wir, dass für die meisten von der Partition erzeugten kleinen Zylindermengen B die Verteilungsfunktion der normalisierten Wartezeit bis zum ersten Auftreten von B annähernd (mit kleinem Fehler Epsilon) von der Funktion $1 - e^{-t}$ majorisiert wird. Dies bedeutet, dass die Besuche eines solchen B entlang der Zeitachse entweder (beinahe) unabhängig voneinander auftreten, oder einander anziehen und Serien erzeugen. Ein ausführlicher Beweis des obigen Resultats ist dabei das Hauptziel dieser Arbeit.

Abstract [EN]

The present thesis covers the Law of Series, a property of specific stochastic processes in which random events do not arise independently from each other, but tend to occur in groups separated by longer periods of absence.
We begin by recalling some well known results of ergodic theory before introducing the Law of Series as possible property of homogeneous signal processes. In the main part of this work we consider an ergodic transformation with positive dynamical entropy with respect to a finite partition. Based on [2], we are going to show that for most small cylinder sets B generated by the partition the distribution function of the normalized waiting time to B is dominated by $1 - e^{-t}$ up to a small error of epsilon. This means, that the visits to such a B along the time axis are either (nearly) independent from one another (as in an i.i.d. process) or they tend to attract each other and create series. A detailed elaboration of the above result is hereby the main goal of this thesis.

Contents

1. Introduction

This thesis addresses a phenomena known as the *Law of Series*. We say that a *series* happens, whenever a random event, typically extremely rare, happens surprisingly often throughout a period of time. There are a lot of serial events which can be perfectly understood as a result of physical dependence, such as volcanic eruptions appearing in series during periods of increased tectonic activity. The Law of Series is concerned with serial events, for which no obvious reason can be found causing these series. It is the belief, that some series happen more often than they should by pure chance, usually in combination with the belief that there exists some unexplained force or law causing the series. The best known examples of such series, which are believed to be caused by a Law of Series, are runs of good luck in the casino (and a casino is most likely the place one comes across the term Law of Series).
Although there have been approaches of proving the existence of a Law of Series, where the most notable one is the book *Das Gesetz der Serie* by the Austrian biologist Paul Kammerer [6], the debate about the Law of Series has avoided strict scientific language so far. In this work, we will describe a rigorous approach to the Law of Series, embedded in ergodic theory.
Starting from the Poisson Process $(X_t)_{t\geq 0}$, in which random events (which we call *signals*) occur independently from each other, we consider two deviations: *Attracting*, which is the tendency of signals to occur in groups, and *repelling*, which is the tendency of signals to occur more evenly distributed. The Law of Series is a deviation towards attracting. Its definition only depends on the distribution of the waiting time W until the first signal occurs, which is for the Poisson Process given by the standard exponential law with distribution function $F_W(t) = 1 - e^{-t}$. In particular, a process $(X_t)_{t\geq 0}$ obeys the Law of Series, if it satisfies the following two postulates:

1. There is no repelling, meaning that $F_W(t) \leq 1 - e^{-t}$ for all $t \geq 0$.

2. There exists attracting, meaning that $F_W(t) > 1 - e^{-t}$ for a least one $t \geq 0$.

There have been published two major results concerning the Law of Series, the paper *Law of Series* [2] by Downarowicz and Lacroix in 2011, and *Spontaneous clustering in theoretical and some empirical stationary processes* [3] by Downarowicz, Lacroix and L´eandri in 2010.
Both works consider an ergodic measure-preserving transformation $T : X \to X$ on a standard probability space $(X, \mathcal{A}, \mu)$, together with a finite partition $\mathcal{P}$ of X. Each partition gives rise to the partitions $\mathcal{P}^n = \{\cap_{i=0}^{n-1} T^{-i}(A_{j_i}) : A_{j_i} \in \mathcal{P}\}$, whose elements are called *cylinders* of length n. The main theorem in [2], addressing the first postulate of the Law of Series, states, that whenever the dynamical entropy of the partition $\mathcal{P}$

is positive, the process $(X_t)_{t\geq 0}$ counting the visits of a point $x \in X$ to a cylinder B, reveals for almost every long cylinder nearly no repelling, or equivalently the normalized waiting time until the first visit to such a cylinder B is bounded from above by $1 - e^{-t}$ (up to a small error ϵ).
In this work we are going to provide a detailed elaboration of this theorem (i.e. of [2, Theorem 1]), as well as the results leading to the theorem, with the goal of making it accessible to a wider audience.
The attracting postulate of the Law of Series is covered in [3], where it is proven, that indeed many of the processes counting the visits to a small cylinder show attracting behaviour. We are going to mention this result in the final section 3.3.
In Chapter 2 we are going to prepare for the Law of Series, assuming basic knowledge in probability and measure theory. In the sections 2.1-2.5 we will introduce the basic concepts and theorems of ergodic theory, as well as the Shannon Entropy of partitions, symbolic dynamical systems and the dynamical entropy. Section 2.6 provides concepts of independence of partitions and processes, which will play an important role proving our main result, Theorem 3.1. In section 2.7 we state two important theorems, namely the Shannon-McMillan-Breiman Theorem and the Ornstein-Weiss Return Times Theorem. Throughout this chapter, we will skip the proofs of the arguments and refer to [4, Chapter 1-3] for further details.
The third chapter, in which we follow [4, Chapter 5], consists of three sections:
In the first section, we are going to introduce the Law of Series as a property of a *homogeneous signal processes*, as well as the terms *attracting* and *repelling*.
In the second section, we will formulate and prove in full detail the repelling postulate of the Law of Series([2, Therorem 1]).
The third section finishes the picture of the Law of Series, covering the result of the attracting postulate of the Law of Series ([3, Therorem 1]) without going into detail.

All figures appearing in this work are taken from [4].

2. Preliminaries

In this chapter we provide all the requirements in order to conveniently treat the Law of Series, introducing ergodic theory, symbolic dynamical systems and entropy theory. We assume familiarity of the reader with basics in probability theory, measure theory and dynamical systems. The sections 2.2-2.7 loosely follow [4, Chapter 1-3], where most of the proofs of the cited theorems can be found.

2.1. Measure preserving maps and ergodic theory

Our first topic will be ergodic theory. We are interested in the evolution of points x of a set X, which are iterated by a transformation $T : X \to X$. Especially, we want to understand the behaviour of $T^n(x)$ when n goes to infinity.
We are going to cover the core definitions and theorems of ergodic theory. A more detailed introduction into ergodic theory, as well as proofs of the Theorems 2.1,2.2 and 2.3 can be found in [7].

Definition 2.1. *A transformation $T : X \to X$ on a measure space $(X, \mathcal{A}, \mu)$ is called* measure preserving*, if it is measurable (i.e. $T^{-1}(A) \in \mathcal{A}$ for every $A \in \mathcal{A}$) and if $\mu(T^{-1}(A)) = \mu(A)$ holds for all $A \in \mathcal{A}$.*
We say that $T : X \to X$ is an invertible *measure-preserving transformation, if T is measure preserving, almost surely bijective and T^{-1} is also measure preserving.*
If $(X, \mathcal{A}, \mu)$ is a probability space, then T is called a (invertible) probability preserving transformation.

Definition 2.2. *If $T : X \to X$ is measure preserving (probability preserving) on $(X, \mathcal{A}, \mu)$, then the quartet $(X, \mathcal{A}, \mu, T)$ is called a* measure preserving (probability preserving) dynamical system.
If T is additionally invertible, then we call $(X, \mathcal{A}, \mu, T)$ an invertible *measure preserving (probability preserving) dynamical system.*

The first famous result in ergodic theory is the following:

Theorem 2.1 (Poincaré's Recurrence Theorem)**.** *Suppose $(X, \mathcal{A}, \mu, T)$ to be a probability preserving dynamical system. Then for every measurable set B and for almost every $x \in B$, there is a sequence $(n_k) \to \infty$ such that $T^{n_k}(x) \in B$.*

In words, Poincaré's Recurrence Theorem states, that almost every $x \in B$ returns to B infinitely often.

Definition 2.3. *A measure preserving transformation $T : X \to X$ is called* ergodic, *if for every T-*invariant set *A (i.e. those A satisfying $T^{-1}(A) = A$) either $\mu(A) = 0$ or $\mu(X \setminus A) = 0$.*
The quartet $(X, \mathcal{A}, \mu, T)$ is then an ergodic dynamical system.

This definition allows us to state the main theorem of ergodic theory:

Theorem 2.2 (Ergodic Theorem)**.** *Let $(X, \mathcal{A}, \mu, T)$ be a probability preserving dynamical system. Then, for any $f \in L_1(\mu)$, there is a function $f^* \in L_1(\mu)$, such that*

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \overset{\mu\ a.e.}{\longrightarrow} f^* \quad as\ n \to \infty.$$

If T is additionally ergodic, then

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \overset{\mu\ a.e.}{\longrightarrow} E(f) = \int f d\mu \quad as\ n \to \infty.$$

We now consider any measurable set $B \subset X$ satisfying $\mu(B) > 0$, where $(X, \mathcal{A}, \mu, T)$ is a probability preserving dynamical system. The *return time to B* is defined on B as

$$\varphi_B(x) = \min\{n > 0 : T^n(x) \in B\}.$$

Note that by Poincaré's Recurrence Theorem, the return time to B is finite for almost every $x \in B$. Therefore, we can define almost everywhere the *induced map on B*, given by

$$T_B(x) = T^{\varphi_B(x)}(x).$$

Associated with the induced map T_B is the *induced dynamical system* $(B, B\cap\mathcal{A}, \mu_B, T_B)$, where $B\cap\mathcal{A} = \{B\cap A : A \in \mathcal{A}\}$ and $\mu_B(A) = \frac{\mu(A\cap B)}{\mu(B)}$ denotes the conditional probability of A given B. The induced system shares many important properties with the original system $(X, \mathcal{A}, \mu, T)$:

Theorem 2.3. *Let $(X, \mathcal{A}, \mu, T)$ be a probability preserving system, and let $B \in \mathcal{A}$ be a set with positive measure. Then*

(i) T_B is probability preserving with respect to μ_B.

(ii) if T is ergodic, then T_B is ergodic.

(iii) (Kac Formula) If T is ergodic, then for every $f \in L_1(\mu)$ it holds that

$$\int_X f d\mu = \int_B \sum_{k=0}^{\varphi_B - 1} f \circ T^k d\mu. \tag{2.1}$$

In particular,

$$\int_B \varphi_B d\mu_B = \frac{1}{\mu(B)}. \tag{2.2}$$

2.2. Partitions and sigma-algebras

From now on, we restrict $(X, \mathcal{A}, \mu)$ to be a *standard probability space*. A standard probability space is a probability space $(X, \mathcal{A}, \mu)$, where X is a complete and separable metric space, $\mathcal{A}$ is the Borel σ-algebra, and μ is a Borel probability measure. We also assume the measure μ to be *complete* (or equivalently the σ-algebra $\mathcal{A}$ to be *completed*), which means that every subset of an $\mathcal{A}$-measurable set with measure μ zero is again measurable.

By a *partition* $\mathcal{P}$ of $(X, \mathcal{A}, \mu)$ we will understand an at most countable partition of X into measurable sets A_i, i.e. $\mathcal{P} = \{A_i : i \in I \subseteq \mathbb{N}\}$, where $\mu(\cup_{i\in I} A_i) = 1$ and $i \neq j \Longrightarrow A_i \cap A_j = \emptyset$.

We say that a partition is *finite*, if all but finitely many elements of the partition have measure zero. We will identify two partitions with each other, and write $\mathcal{P} = \mathcal{Q}$, whenever $\mathcal{P} = \mathcal{Q} \pmod{\mu}$, i.e. if

$$\sum_{\substack{i\geq 1,\\ A_i\in\mathcal{P}, C_i\in\mathcal{Q}}} \mu(A_i \triangle C_i) = 0 \tag{2.3}$$

holds, where $\triangle$ denotes the symmetric difference of sets (i.e. $A\triangle C = (A \setminus C) \cup (C \setminus A)$). Note that we consider finite element partitions as countable in (2.3) by attaching countably many copies of the empty set.

A partition $\mathcal{P}$ is *finer* than (or equivalently is a *refinement* of) $\mathcal{Q}$, writing $\mathcal{Q} \preccurlyeq \mathcal{P}$, when each element of $\mathcal{Q}$ is a union of elements of the partition $\mathcal{P}$.

The *join* of two partitions is defined as the partition

$$\mathcal{P} \vee \mathcal{Q} = \{A \cap B : A \in \mathcal{P}, B \in \mathcal{Q}\}.$$

We say that a sequence of partitions $(\mathcal{Q}_k)_{k\in\mathbb{N}}$ is said to *generate* a σ-algebra $\mathcal{B}$, where $\mathcal{B}$ always denotes a sub σ-algebra of $\mathcal{A}$, if $\mathcal{B}$ is the smallest σ-algebra which contains all elements of each partition $\mathcal{Q}_k$. In this case, we will denote $\mathcal{B}$ as

$$\mathcal{B} = \bigvee_{k=1}^{\infty} \mathcal{Q}_k.$$

A sequence of partitions $(\mathcal{Q}_k)_{k\in\mathbb{N}}$ is called *refining*, if $\mathcal{Q}_k \preccurlyeq \mathcal{Q}_{k+1}$ holds for every $k \geq 1$.

For two σ-algebras $\mathcal{B}$ and $\mathcal{C}$, the join $\mathcal{B} \vee \mathcal{C}$ will be the σ-algebra generated by the union of $\mathcal{B}$ and $\mathcal{C}$. Similarly, the join of a family of σ-algebras $\mathcal{B}_k$, $k \in \mathbb{N}$, is the σ-algebra generated by the union of all $\mathcal{B}_k$, denoted by $\bigvee_{k\in\mathbb{N}} \mathcal{B}_k$.

The standard lemma below provides a family of sets which generates the join:

Lemma 2.1. *Let $\mathcal{B}_k, k \in \mathbb{N}$, be a collection of σ-algebras. Then, the join $\bigvee_{k\in\mathbb{N}} \mathcal{B}_k$ is generated by the π-system*

$$\Pi = \{ \bigcap_{k=1}^{n} B_k : B_k \in \mathcal{B}_{\alpha_k}, \alpha_k \in \mathbb{N}, n \geq 1\}.$$

Sometimes we will join partitions with σ-algebras. In such cases, we will replace a partition $\mathcal{P}$ by the σ-algebra generated by the elements of the partition, and we will denote this generated σ-algebra again by $\mathcal{P}$.
Whenever a partition $\mathcal{P}$ is measurable with respect to a σ-algebra $\mathcal{B}$, we will denote this as $\mathcal{P} \subseteq \mathcal{B}$.

2.3. Shannon entropy of partitions

We assume all that logarithms in this chapter are to base 2 and that $0 * \log 0 = 0$.
Let $\mathcal{P}$ be a partition of the probability space $(X, \mathcal{A}, \mu)$. For $x \in X$, the set A_x denotes the set of the partition $\mathcal{P}$ which contains x.

Definition 2.4. *The* information function *of a partition $\mathcal{P}$ is a map $I_{\mu,\mathcal{P}} : X \to [0, \infty]$, where*

$$I_{\mu,\mathcal{P}}(x) = I_{\mathcal{P}}(x) = -\log \mu(A_x).$$

Definition 2.5. *The* Shannon *(or* static) entropy *of a partition $\mathcal{P}$ with respect to μ is the expected value of the information function:*

$$H(\mu, \mathcal{P}) = H(\mathcal{P}) = \int I_{\mathcal{P}}(x) d\mu(x) = -\sum_{A \in \mathcal{P}} \mu(A) \log \mu(A)$$

For a finite partition $\mathcal{P} = \{A_1, A_2, ..., A_k\}$, we can interpret the Shannon entropy as follows: If $A_1, A_2, ..., A_k$ denote the outcomes of a random experiment (where the probability of each outcome equals $\mu(A_i)$), then $H(\mathcal{P})$ measures the uncertainty removed (or information gained) by performing the experiment, i.e. $H(\mathcal{P})$ measures the uncertainty about which set A_i of the partition a point $x \in X$ will belong to. A detailed interpretation of Shannon entropy can be found in [4, Section 0.3.2].

It follows directly from the definition that the Shannon entropy of a partition is non-negative and that every finite partition $\mathcal{P}$ satisfies $H(\mathcal{P}) < \infty$. Additionally, the following fundamental estimate holds :

$$H(\mathcal{P}) \leq \log \#\mathcal{P}. \tag{2.4}$$

Note that for finite partitions we obtain equality in (2.4), when the partition $\mathcal{P}$ is evenly distributed, i.e. if $\mu(A_i) = 1/k$ holds for every set A_i whenever $\mathcal{P}$ has k elements.

Lemma 2.2. *For each element A of a finite partition $\mathcal{P}$ it holds that*

$$H(\mathcal{P}) \leq (1 - \mu(A)) \log \#\mathcal{P} + 1. \tag{2.5}$$

For two given partitions $\mathcal{P}$ and $\mathcal{Q}$, the *conditional information function* is defined as

$$I_{\mu,\mathcal{P}|\mathcal{Q}}(x) = I_{\mathcal{P}|\mathcal{Q}}(x) = I_{\mathcal{P} \vee \mathcal{Q}}(x) - I_{\mathcal{Q}}(x) = -\log\Big(\frac{\mu(A_x \cap B_x)}{\mu(B_x)}\Big) = -\log \mu_{B_x}(A_x).$$

Definition 2.6. *The* conditional Shannon entropy of $\mathcal{P}$ given $\mathcal{Q}$ *is defined as the expected value of the conditional information function:*

$$H(\mu, \mathcal{P}|\mathcal{Q}) = H(\mathcal{P}|\mathcal{Q}) = \int I_{\mathcal{P}|\mathcal{Q}}(x) d\mu(x).$$

The conditional entropy $H(\mathcal{P}|\mathcal{Q})$ can be interpreted as the expected information gain by performing the random experiment associated with $\mathcal{P}$, given the outcome of the experiment associated with $\mathcal{Q}$.
It holds that

$$\begin{aligned} H(\mathcal{P}|\mathcal{Q}) &= -\sum_{B\in\mathcal{Q}, A\in\mathcal{P}} \mu(A\cap B)\log \mu_B(A) \\ &= -\sum_{B\in\mathcal{Q}} \mu(B) \sum_{A\in\mathcal{P}} \mu_B(A)\log\mu_B(A) = \sum_{B\in\mathcal{Q}} \mu(B) H_B(\mathcal{P}), \end{aligned} \tag{2.6}$$

where $H_B(\mathcal{P}) = H(\mu_B, \mathcal{P})$ denotes the entropy of $\mathcal{P}$ with respect to the measure μ_B. We will be frequently using this expression of conditional entropy in calculations.

Lemma 2.3. *(Properties of conditional entropy) Let $\mathcal{P}, \mathcal{Q}$ and $\mathcal{R}$ denote countable partitions of $(X, \mathcal{A}, \mu)$. Then*

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

(ii) $H(\mathcal{P} \vee \mathcal{Q}|\mathcal{R}) = H(\mathcal{P}|\mathcal{Q} \vee \mathcal{R}) + H(\mathcal{Q}|\mathcal{R})$

(iii) $\mathcal{Q} \preccurlyeq \mathcal{P} \Rightarrow H(\mathcal{Q}) \le H(\mathcal{P})$

(iv) $\mathcal{Q} \preccurlyeq \mathcal{P} \Rightarrow H(\mathcal{Q}|\mathcal{R}) \le H(\mathcal{P}|\mathcal{R})$

(v) $\mathcal{R} \preccurlyeq \mathcal{Q} \Rightarrow H(\mathcal{P}|\mathcal{R}) \ge H(\mathcal{P}|\mathcal{Q})$

(vi) $H(\mathcal{P}|\mathcal{Q}) \le H(\mathcal{P})$

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

(viii) $H(\mathcal{P} \vee \mathcal{Q}|\mathcal{R}) \le H(\mathcal{P}|\mathcal{R}) + H(\mathcal{Q}|\mathcal{R})$.

As a direct consequence of (vii) in the above lemma, we have:

Corollary 2.1. *For $k \ge 1$, and partitions $\mathcal{P}_i, 1 \le i \le k$, it holds that*

$$H(\bigvee_{i=1}^{k} \mathcal{P}_i) \le \sum_{i=1}^{k} H(\mathcal{P}_i).$$

We will now generalize the concept of conditioning by introducing the conditional entropy given a σ-algebra $\mathcal{B}$, where $\mathcal{B} \subseteq \mathcal{A}$ denotes a sub σ-algebra of $\mathcal{A}$.

Definition 2.7. *The* conditional entropy of $\mathcal{P}$ given $\mathcal{B}$ *is the value*

$$H(\mu, \mathcal{P}|\mathcal{B}) = H(\mathcal{P}|\mathcal{B}) = \inf\{H(\mathcal{P}|\mathcal{Q}) : \mathcal{Q} \subseteq \mathcal{B}\},$$

where $\mathcal{Q}$ ranges over all countable partitions which are measurable with respect to $\mathcal{B}$.

Lemma 2.4. *Let $\mathcal{P}, \mathcal{Q}$ and $\mathcal{R}$ denote countable partitions of $(X, \mathcal{A}, \mu)$, and let $\mathcal{B}$ and $\mathcal{C}$ be sub σ-algebras of $\mathcal{A}$. Then*

(i) $H(\mathcal{P}|\mathcal{Q} \vee \mathcal{B}) = \sum_{B \in \mathcal{Q}} \mu(B) H_B(\mathcal{P}|\mathcal{B})$

(ii) $H(\mathcal{P} \vee \mathcal{Q}|\mathcal{B}) = H(\mathcal{P}|\mathcal{Q} \vee \mathcal{B}) + H(\mathcal{Q}|\mathcal{B})$

(iii) $\mathcal{P} \subseteq \mathcal{B} \Rightarrow H(\mathcal{P}|\mathcal{B}) = 0$

(iv) $\mathcal{C} \subseteq \mathcal{B} \Rightarrow H(\mathcal{P}|\mathcal{C}) \geq H(\mathcal{P}|\mathcal{B})$.

(v) $\mathcal{Q} \preccurlyeq \mathcal{P} \Rightarrow H(\mathcal{Q}|\mathcal{B}) \leq H(\mathcal{P}|\mathcal{B})$

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

Subsequent use of (ii) in Lemma 2.4 leads to the identity below:

Lemma 2.5. *Let, for $k \geq 1$, $\mathcal{P}_i, 1 \leq i \leq k$ be Partitions and $\mathcal{B}$ be a σ-algebra. Then,*

$$H(\bigvee_{i=1}^{k} \mathcal{P}_i|\mathcal{B}) = H(\mathcal{P}_1|\mathcal{B}) + \sum_{i=2}^{k} H(\mathcal{P}_i| \bigvee_{j=1}^{i-1} \mathcal{P}_j \vee \mathcal{B}).$$

The Shannon entropy is invariant under measure preserving transformations:

Lemma 2.6. *Let $T : X \to X$ be a probability preserving transformation, let $\mathcal{P}$ and $\mathcal{Q}$ denote partitions and let $\mathcal{B}$ be a σ-algebra. Then, for every $k \geq 1$,*

(i) $H(\mathcal{P}) = H(T^{-k}(\mathcal{P}))$.

(ii) $H(\mathcal{P}|\mathcal{Q}) = H(T^{-k}(\mathcal{P})|T^{-k}(\mathcal{Q}))$.

(iii) $H(\mathcal{P}|\mathcal{B}) = H(T^{-k}(\mathcal{P})|T^{-k}(\mathcal{B}))$.

If T is an invertible probability preserving transformation, then the identities above hold for all $k \in \mathbb{Z}$.

Let now $\mathcal{B}_{\mathbb{N}_0}$ denote the space of of all equivalence classes of countable partitions (i.e. $\mathcal{P}$,$\mathcal{P}' \in [\mathcal{P}]$ if $\mathcal{P} = \mathcal{P}'(\mathrm{mod}\ \mu)$). Then

$$d_1([\mathcal{P}], [\mathcal{Q}]) = d_1(\mathcal{P}, \mathcal{Q}) = \inf_{\pi} \sum_{i=1}^{\infty} \mu(A_i \triangle C_{\pi(i)}),$$

where the infimum is taken over of all permutations π of $\mathbb{N}$, is a metric on $\mathcal{B}_{\mathbb{N}_0}$.
Similarly, for $k \geq 2$, the map

$$d_k([\mathcal{P}], [\mathcal{Q}]) = d_k(\mathcal{P}, \mathcal{Q}) = \inf_{\pi} \sum_{i=1}^{k} \mu(A_i \triangle C_{\pi(i)}),$$

where the infimum ranges over all permutations of $\{1, ..., k\}$, is a metric on the space $\mathcal{B}_k$ of all (equivalence classes of) partitions with exactly k elements.

Lemma 2.7. *[4, Fact 1.7.3] Let $(\mathcal{Q}_k)_{k\geq 1}$ be a generating and refining sequence of a sigma-algebra $\mathcal{B}$ and let $\mathcal{P}$ be measurable with respect to $\mathcal{B}$. Then for every $\epsilon > 0$ there exists a k and a partition $\mathcal{P}_k$ satisfying $\mathcal{P}_k \preccurlyeq \mathcal{Q}_k$, such that $d_1(\mathcal{P}, \mathcal{P}_k) < \epsilon$.*

Lemma 2.8. *Let $(X, \mathcal{A}, \mu)$ be a standard probability space. Then, the space $\mathcal{B}_{\mathbb{N}_0}$ equipped with the metric d_1 is a complete and separable metric space. For $k \geq 2$, the spaces $(\mathcal{B}_k, d_k)$ are complete and separable metric spaces as well.*

We finish this section with a useful statement about the entropy conditional on a sequence of refining partitions (or more general on a sequence of increasing σ-algebras):

Lemma 2.9. *[4, Lemma 1.7.11] Let $\mathcal{P}$ be a partition such that $H(\mathcal{P}) < \infty$. If $(\mathcal{Q}_k)_{k\geq 1}$ is a refining sequence of partitions, then*

$$H(\mathcal{P}| \bigvee_{k\in\mathbb{N}} \mathcal{Q}_k) = \lim_{k\to\infty} \downarrow H(\mathcal{P}|\mathcal{Q}_k).$$

If $(\mathcal{B}_k)_{k\geq 1}$ is a sequence of monotone σ-algebras, i.e. $\mathcal{B}_k \subseteq \mathcal{B}_{k+1}$ for all $k \geq 1$, we have

$$H(\mathcal{P}| \bigvee_{k\in\mathbb{N}} \mathcal{B}_k) = \lim_{k\to\infty} \downarrow H(\mathcal{P}|\mathcal{B}_k).$$

2.4. Symbolic dynamical systems and processes

We start by introducing a very important class of measure preserving dynamical systems: A *symbolic dynamical system* is a quartet $(\Lambda^{\mathbb{S}}, \mathcal{C}, \nu, \sigma)$, where the elements of the space $\Lambda^{\mathbb{S}}$ (with $\mathbb{S} \in \{\mathbb{N}, \mathbb{Z}\}$) are sequences $(x_i)_{i\in\mathbb{S}}$ over a countable set Λ. The σ-algebra $\mathcal{C}$ denotes the completed product σ-algebra in $\Lambda^{\mathbb{S}}$, where each copy of Λ is equipped with the σ-algebra of all subsets. The map σ typically is the left shift map, and ν denotes an arbitrary measure such that the left shift is measure preserving.
We say that a measure preserving dynamical system $(Y, \mathcal{B}, \nu, S)$ is a *factor* of the system $(X, \mathcal{A}, \mu, T)$, if there exists a measurable map $\pi : X \to Y$, such that $\pi \circ T = S \circ \pi$ and $\pi\mu = \nu$ (meaning that $\mu(\pi^{-1}(B)) = \nu(B)$ for every $B \in \mathcal{B}$).

Let us now consider a finite measurable partition $\mathcal{P} = \{A_1, ..., A_k\}$ of an invertible probability preserving system $(X, \mathcal{A}, \mu, T)$. We define the *coding map* of $\mathcal{P}$ as

$$c(x) = i \Leftrightarrow x \in A_i,$$

where $c : X \to \{1, ..., k\}$. Then, the map

$$\begin{aligned} \pi : \quad & X \to \{1, ..., k\}^{\mathbb{Z}} \\ & x \mapsto (c(T^n(x))_{n\in\mathbb{Z}} \end{aligned}$$

is a factor map from $(X, \mathcal{A}, \mu, T)$ to $(\{1, ..., k\}^{\mathbb{Z}}, \mathcal{C}, \sigma, \nu)$, where the measure $\nu = \pi\mu$ and $\mathcal{C}$ is the completed product σ-algebra in $\{1, ..., k\}^{\mathbb{Z}}$ (where each copy of $\{1, ..., k\}$ has the σ-algebras of all subsets). The map $\sigma : \{1, ..., k\}^{\mathbb{Z}} \to \{1, ..., k\}^{\mathbb{Z}}$ denotes the left shift map, given by

$$\sigma((x_n)_{n\in\mathbb{Z}}) = (x_{n+1})_{n\in\mathbb{Z}}.$$

Based on the choice of the partition $\mathcal{P}$, this procedure allows us to transform the dynamical system $(X, \mathcal{A}, \mu, T)$ into a symbolic dynamical system $(\Lambda^{\mathbb{Z}}, \mathcal{C}, \sigma, \nu)$, where $\Lambda = \{1, ..., k\}$. Note that the same procedure can be done also for countably infinite partitions, resulting in a symbolic system in which $\Lambda = \mathbb{N}$.
From now on, we will denote this symbolic system as $(X, \mathcal{P}, \mu, T)$ and call it the *process generated by* $\mathcal{P}$ or just process. The sequence $(x_n)_{n\in\mathbb{Z}}$ corresponding to a point $x \in X$ will be called the *$\mathcal{P}$-name of x*.
For a given sequence $(x_n)_{n\in\mathbb{Z}}$, the elements $B = (x_i, x_{i+1}, ..., x_j) \in \Lambda^{\{i,i+1,...,j\}} = \Lambda^{[i,j]}$, where $i, j \in \mathbb{Z}$, $i \leq j$, are called *blocks* of *length* $j - i + 1$. Associated with each block B is the cylinder set

$$U_B = \{(x_n)_{n\in\mathbb{Z}} : (x_i, x_{i+1}, ..., x_j) = B\} \subset \Lambda^{\mathbb{Z}},$$

which we call a *cylinder* of *length* $j-i+1$. The product σ-algebra $\mathcal{C}$ is generated by the collection of all finite-length cylinders, formally

$$\mathcal{C} = \sigma(\{U_B : B = (x_i, x_{i+1}, ..., x_j) \in \Lambda^{[i,j]}, i, j \in \mathbb{Z}, i \leq j\}).$$

The left shift map is is clearly measure preserving on all such finite cylinders, which implies that, for an arbitrary measurable partition $\mathcal{P}$, the symbolic system $(X, \mathcal{P}, \mu, T)$ is an invertible measure preserving dynamical system whenever the original system $(X, \mathcal{A}, \mu, T)$ is. Similarly, if the original system is ergodic, the symbolic system is ergodic as well.
Associated with each cylinder U_B is the set $\pi^{-1}(U_B)$, which we call again a *cylinder*.
For $n \geq 1$, the join

$$\mathcal{P}^n = \bigvee_{i=0}^{n-1} T^{-i}(\mathcal{P}) = \{\pi^{-1}(U_B) : B \in \Lambda^n\} \tag{2.7}$$

is the partition into the preimages $\pi^{-1}(U_B)$ over all blocks $B \in \Lambda^{[0,n-1]} = \Lambda^n$. Equivalently, each block $B \in \Lambda^n$ in the symbolic system determines exactly one element $\pi^{-1}(U_B)$ of the partition $\mathcal{P}^n$. For $n = 0$, the partition $\mathcal{P}^0$ in (2.7) denotes the trivial partition $\mathcal{P}$.
From now on, we will identify a block B with the associated cylinder U_B (in the symbolic system) or with the cylinder $\pi^{-1}(U_B)$ (in the original system), and B will denote the

block as well as the cylinders. We are not going to use U_B and $\pi^{-1}(U_B)$ again. It will be clear from the context, whether B is meant to be a block, the associated cylinder in the symbolic system, or the associated cylinder in the original system. In different words, we will be identifying the Cartesian product Λ^n with the join $\mathcal{P}^n$.

Similarly to (2.7),we denote for any finite subset $\mathbb{D} \subset \mathbb{Z}$ the partition

$$\mathcal{P}^{\mathbb{D}} = \bigvee_{i \in \mathbb{D}} T^{-i}(\mathcal{P}).$$

If $\mathbb{D}$ is (countably) infinite, then $\mathcal{P}^{\mathbb{D}}$ denotes the smallest completed σ-algebra containing all partitions $T^{-i}(\mathcal{P})$, where $i \in \mathbb{D}$.
We will call the (completed) σ-algebra $\mathcal{P}^- = \mathcal{P}^{(-\infty,-1]}$ the *full past* of the process. The (completed) σ-algebra $\mathcal{P}^+ = \mathcal{P}^{[1,\infty)}$ will be called the *full future* of the process. In addition to $\mathcal{P}^n = \mathcal{P}^{[0,n-1]}$ we will use the convention $\mathcal{P}^{-n} = \mathcal{P}^{[-n,-1]}$.

2.5. Dynamical entropy of partitions

Again, let $(X, \mathcal{A}, \mu, T)$ denote a probability preserving system, and let $(X, \mathcal{P}, \mu, T)$ be the corresponding symbolic system for an at most countable partition $\mathcal{P}$.
We start this section with an auxiliary lemma about the join $\mathcal{P}^n$:

Lemma 2.10. *For every $n \geq 1$,*

$$\frac{1}{n+1} H(\mathcal{P}^{n+1}) \leq \frac{1}{n} H(\mathcal{P}^n).$$

This allows us to state a core definition of entropy theory, the *dynamical entropy of a partition* $\mathcal{P}$. The dynamical entropy can be interpreted as the average information gain per iterate.

Definition 2.8. *Let $H(\mathcal{P}) < \infty$. The* dynamical entropy *of (the process generated by) $\mathcal{P}$ is defined as*

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

Note that by Lemma 2.10 the limit in the definition exists as the Shannon entropy of the partition is finite.
The formula below gives rise to another interpretation of dynamical entropy as expected information gain in one step given all information from the future or the past:

Lemma 2.11. *[4, Fact 2.3.4] Let $(X, \mathcal{P}, \mu, T)$ be the process generated by $\mathcal{P}$, where $H(\mathcal{P}) < \infty$. Then*

$$h(\mathcal{P}) = H(\mathcal{P} | \mathcal{P}^+).$$

If $(X, \mathcal{P}, \mu, T)$ is additionally invertible, it also holds that

$$h(\mathcal{P}) = H(\mathcal{P} | \mathcal{P}^-).$$

Lemma 2.12. *Let $\mathcal{P}, \mathcal{Q}$ denote countable partitions on a probability preserving system $(X, \mathcal{A}, \mu, T)$. Then*

(i) $h(\mathcal{P}) \leq H(\mathcal{P})$.

(ii) $h(\mathcal{P} \vee \mathcal{Q}) \leq h(\mathcal{P}) + h(\mathcal{Q})$.

(iii) $h(T^{-k}(\mathcal{P})) = h(\mathcal{P})$ *for* $k \in \mathbb{N}$.

(iv) $h(\mathcal{P}^k) = h(\mathcal{P})$ *for* $k \in \mathbb{N}$.

If $(X, \mathcal{A}, \mu, T)$ is additionally invertible, then (iii) and (iv) hold for $k \in \mathbb{Z}$.

Our next result, called the *power rule*, allows us to calculate the dynamical entropy of a process $(X, \mathcal{P}, \mu, T)$ when T and $\mathcal{P}$ are replaced by their n-th iterate:

Lemma 2.13. *(Power Rule) [4, Fact 2.4.19]*
Let $(X, \mathcal{P}, \mu, T)$ be an invertible process generated by $\mathcal{P}$ and $n \in \mathbb{Z}$. Then

$$h(\mu, T^n, \mathcal{P}^{|n|}) = |n| h(\mu, T, \mathcal{P}).$$

As a consequence of the above lemma, the following identity, which we will frequently use (and also call power rule), can be derived:

Corollary 2.2. *In an invertible process $(X, \mathcal{P}, \mu, T)$ for every $n \in \mathbb{N}$ it holds that*

$$H(\mathcal{P}^n | \mathcal{P}^-) = n h(\mathcal{P}).$$

We can also define the entropy of a probability preserving system $(X, \mathcal{A}, \mu, T)$:

Definition 2.9. *The* Kolmogorov-Sinai entropy *of a system $(X, \mathcal{A}, \mu, T)$ is the value*

$$h(\mu, T) = \sup\{h(\mathcal{P}) : H(\mathcal{P}) < \infty\}.$$

2.6. Independence properties of partitions and processes

We continue by introducing some terms concerning the independence of two countable partitions $\mathcal{P}$ and $\mathcal{Q}$:

Definition 2.10. *Two partitions $\mathcal{P}$ and $\mathcal{Q}$ are called* independent *(written as $\mathcal{P} \perp \mathcal{Q}$), if*

$$\mu(A \cap B) = \mu(A)\mu(B)$$

for every $A \in \mathcal{P}$ and $B \in \mathcal{Q}$.

If $H(\mathcal{P}) < \infty$, it easily follows from the above definition that

$$\mathcal{P} \perp \mathcal{Q} \Leftrightarrow H(\mathcal{P}|\mathcal{Q}) = H(\mathcal{P}). \tag{2.8}$$

Definition 2.11. *Two countable partitions $\mathcal{P}$ and $\mathcal{Q}$ are called* ϵ-independent *(we write it as $\mathcal{P} \perp^\epsilon \mathcal{Q}$), if*

$$\sum_{A\in\mathcal{P},C\in\mathcal{Q}} |\mu(A\cap C) - \mu(A)\mu(C)| \leq \epsilon. \tag{2.9}$$

A result similar to (2.8) also holds for ϵ-independent partitions:

Lemma 2.14. *[4, Fact 3.1.3] For every $M > 0$ and $\epsilon > 0$ there exists a $\delta > 0$ such that any two countable partitions $\mathcal{P}$ and $\mathcal{Q}$, where $H(\mathcal{P}) < M$, satisfy*

$$\mathcal{P} \perp^\delta \mathcal{Q} \implies H(\mathcal{P}|\mathcal{Q}) \geq H(\mathcal{P}) - \epsilon$$

and

$$H(\mathcal{P}|\mathcal{Q}) \geq H(\mathcal{P}) - \delta \implies \mathcal{P} \perp^\epsilon \mathcal{Q}.$$

Definition 2.12. *Let $\mathcal{P}$ be a countable partition with finite entropy. We say $\mathcal{P}$ is* ϵ-entropy independent *of a countable partition $\mathcal{Q}$, if*

$$H(\mathcal{P}|\mathcal{Q}) \geq H(\mathcal{P}) - \epsilon.$$

Definition 2.13. *A partition $\mathcal{P}$ with finite entropy is called* ϵ-independent (ϵ-entropy independent) *of a σ-algebra $\mathcal{A}$, if it is ϵ-independent (ϵ-entropy independent) of any countable partition $\mathcal{Q}$ which is $\mathcal{A}$-measurable.*

Next, we are going to apply the above concepts of independence to processes:

Definition 2.14. *A process $(X, \mathcal{P}, \mu, T)$ is called* independent, *if $h(\mathcal{P}) = H(\mathcal{P})$ or equivalently, if $H(\mathcal{P}|\mathcal{P}^+) = H(\mathcal{P})$.*
In an independent invertible process, it additionally holds that $H(\mathcal{P}|\mathcal{P}^-) = H(\mathcal{P})$.

Definition 2.15. *A process $(X, \mathcal{P}, \mu, T)$ is called* ϵ-independent (ϵ-entropy independent), *if the partition $\mathcal{P}$ is ϵ-independent (ϵ-entropy independent) of the future $\mathcal{P}^+$.*

If $\mathcal{P}$ hat finite entropy, then ϵ-entropy independence can be stated as

$$h(\mathcal{P}) \geq H(\mathcal{P}) - \epsilon.$$

A good source of ϵ-entropy independent processes are induced systems, which is shown in the next theorem:

Theorem 2.4. *[4, Theorem 3.1.13] Let $(X, \mathcal{P}, \mu, T)$ be a process with finite positive entropy $h(\mathcal{P})$, and let $\epsilon > 0$. Then, for sufficiently large n, there is a set $X_n \subset X$ with measure μ of at least $1 - \epsilon$, where X_n is a union of cylinders B of the partition $\mathcal{P}^{[1,n]}$, with the property that for every $\mathcal{P}^{[1,n]}$-cylinder $B \subset X_n$ the process $(B, \mathcal{P}, \mu_B, T_B)$ is ϵ-entropy independent.*

Figure 2.1 shows an illustration of an induced invertible process $(B, \mathcal{P}, \mu_B, T_B)$ generated by $\mathcal{P}$ for some $B \in \mathcal{P}^{[1,n]}$ and $x \in B$. The $\mathcal{P}$-name of x in the induced process is given by $(...a_{-1}a_0a_1a_2a_3...)$.

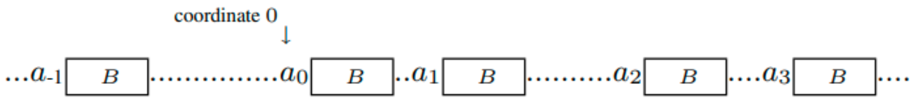


Figure 2.1.: Symbolic representation of the induced process $(B, \mathcal{P}, \mu_B, T_B)$ for some $x \in B$ and $B \in \mathcal{P}^{[1,n]}$: The symbol $a_0 = x_0$ determines the position of x in the partition $\mathcal{P}$. The succeeding block $B \in \Lambda^{[1,n]}$ determines that x is an element of B. The symbol $a_1 = x_{\varphi_B(x)}$ shows us the position of a_1 in $\mathcal{P}$, where the block $B \in \Lambda^{[a_1+1,a_1+n]}$ proves that a_1 indeed lies in B. Similarly, all a_k lie in B and for positive k, a_k determines in which cylinder of $\mathcal{P}$ the k-th return to B is located.

2.7. Entropy theorems in processes

We will now cover two of the most important theorems concerning the dynamical entropy of process a $(X, \mathcal{P}, \mu, T)$.
Recall that for every $n \in \mathbb{N}$ and $x \in X$, the information function of the partition $\mathcal{P}^n$ is given by $I_{\mathcal{P}^n}(x) = -\log \mu(A_x^n)$, where A_x^n is the unique cylinder of $\mathcal{P}^n$ containing x.

Theorem 2.5. *(Shannon-McMillan-Breiman Theorem) [4, Theorem 3.3.1]*
Let $(X, \mathcal{P}, \mu, T)$ be an ergodic process, where $\mathcal{P}$ is an at most countable partition and $H(\mathcal{P}) < \infty$. Then

$$\lim_{n\to\infty} \frac{1}{n} I_{\mathcal{P}^n}(x) = h(\mu, T, \mathcal{P}) \quad \mu\text{-a.e.}$$

As almost sure convergence implies convergence in probability, as a direct consequence of Shannon-McMillan-Breiman Theorem, we can deduce: For every $\alpha > 0$ and for every $\epsilon > 0$ exists an $n_0 \in \mathbb{N}$, such that for all $n \geq n_0$ it holds that

$$\mu\{x \in X : 2^{-n(h+\alpha)} < \mu(A_x^n) < 2^{-n(h-\alpha)}\} > 1 - \epsilon. \tag{2.10}$$

In order to treat the second important entropy theorem, the *Ornstein-Weiss-Theorem*, we define the *first return time to the n-th cylinder* as

$$\varphi_{A_x^n}(x) = \min\{j > 0 : T^j(x) \in A_x^n\}.$$

Similarly, the *k-th return time to the n-th cylinder* is defined as

$$\varphi_{A_x^n}^{(k)}(x) = \min\{j > 0 : \sum_{i=1}^{j} \mathbb{1}_{A_x^n}(T^i x) = k\}.$$

Theorem 2.6. *(Ornstein-Weiss-Theorem) [4, Theorem 3.4.1]*
Let $(X, \mathcal{P}, \mu, T)$ be an ergodic process. Then

$$\lim_{n\to\infty} \frac{1}{n} \log \varphi_{A_x^n}(x) = h(\mu, T, \mathcal{P}) \;\; \mu\text{-a.e.}.$$

If $(X, \mathcal{P}, \mu, T)$ is additionally invertible, then

$$\lim_{n\to\infty} \frac{1}{n} \log \varphi_{A_x^{-n}}(x) = h(\mu, T, \mathcal{P}) \;\; \mu\text{-a.e.}.$$

Ornstein-Weiss-Theorem still holds when we replace the first return time by the k-th return time:

Corollary 2.3. *If $(X, \mathcal{P}, \mu, T)$ is ergodic, then for every $k \in \mathbb{N}$ it holds that*

$$\lim_{n\to\infty} \frac{1}{n} \log \varphi^{(k)}_{A_x^n}(x) = h(\mu, T, \mathcal{P}) \;\; \mu\text{-a.e.}$$

If $(X, \mathcal{P}, \mu, T)$ is additionally invertible, then

$$\lim_{n\to\infty} \frac{1}{n} \log \varphi^{(k)}_{A_x^{-n}}(x) = h(\mu, T, \mathcal{P}) \;\; \mu\text{-a.e.}.$$

Applying Ornstein-Weiss Theorem for convergence in probability allows us to bound the return times in an ergodic process $(X, \mathcal{P}, \mu, T)$ by the dynamical entropy as follows: For every $k \geq 1$, for every $\alpha > 0$ and for every $\epsilon > 0$ exists an $n_0 \in \mathbb{N}$, such that for all $n \geq n_0$ we have

$$\mu\{x \in X : 2^{n(h-\alpha)} < \varphi^{(k)}_{A_x^n}(x) < 2^{n(h+\alpha)}\} > 1 - \epsilon. \tag{2.11}$$

3. Law of Series

3.1. Attracting and repelling in signal processes

In this section we will introduce the *Law of Series* in ergodic theory, including all terms which allow us to formally define it, especially those of a *homogeneous signal process*, *attracting* and *repelling*. We follow [4, Section 5.2].

Definition 3.1. *As* signal process *we will understand a continuous time (and, for very small time increments also a discrete time) stochastic process* $(X_t)_{t\geq 0}$ *on a probability space* $(X, \mathcal{A}, \mu)$ *satisfying:*

(i) $X_0 = 0$ *a.s.*

(ii) X_t *is an integer*

(iii) $(X_t)_{t\geq 0}$ *has non-decreasing and right-continuous trajectories* $t \to X_t(x)$.

We say that, for given $x \in X$, *a* signal occurs at time t, *if the trajectory* $X_t(x)$ *jumps by a unit (or several units) at time* t.

Definition 3.2. *We call a signal process* homogeneous *if, for every* $n \in \mathbb{N}$ *and for every* $t_0, t_1, t_2, ..., t_n \geq 0$ *with* $0 \leq t_1 < t_2 < \cdots < t_n$, *the joint distribution of the increments*

$$X_{t_2} - X_{t_1}, X_{t_3} - X_{t_2}, ..., X_{t_n} - X_{t_{n-1}} \tag{3.1}$$

is equal to that of

$$X_{t_2+t_0} - X_{t_1+t_0}, X_{t_3+t_0} - X_{t_2+t_0}, ..., X_{t_n+t_0} - X_{t_{n-1}+t_0}.$$

Definition 3.3. *The real number* $\lambda \in (0, \infty)$, *which satisfies* $E(X_1) = \lambda$ *for a signal process* $(X_t)_{t\geq 0}$, *is called the* intensity *of the signals.*

The expected value of a homogeneous signal process $(X_t)_{t\geq 0}$ is linear with respect to the time t, which is stated in the following lemma:

Lemma 3.1. *If a homogeneous signal process has intensity* λ, *then for all* $t \geq 0$ *we have*

$$E(X_t) = \lambda t.$$

Proof. By homogeneity, for nonnegative t_1 and t_2 it holds that

$$\begin{aligned}E(X_{t_1}) &= \sum_{k\geq 1} k\mu\{X_{t_1} = k\} = \sum_{k\geq 1} k\mu\{X_{t_1} - X_0 = k\} \\ &= \sum_{k\geq 1} k\mu\{X_{t_1+t_2} - X_{t_2} = k\} = E(X_{t_1+t_2}) - E(X_{t_2}).\end{aligned}$$

We claim that

$$E(X_{t_1+t_2+...+t_n}) = E(X_{t_1}) + E(X_{t_2}) + ... + E(X_{t_n}) \tag{3.2}$$

holds for every $n \geq 2$ and for all nonnegative real numbers $0 \leq t_1 < t_2 < \cdots < t_n$. We have already shown that this is true for $n = 2$. Suppose (3.2) holds for $n-1$. Then

$$E(X_{t_1+t_2+...+t_n}) = E(X_{t_1+t_2+...+t_{n-1}}) + E(X_{t_n}) = E(X_{t_1}) + E(X_{t_2}) + ... + E(X_{t_n}),$$

which proves our claim.
Now, by taking $t_1 = t_2 = ... = t_n = 1$, (3.2) implies for all integers $n \geq 0$ that $E(X_n) = nE(X_1) = \lambda n$. Similarly, for every rational $t = \frac{p}{q}$, we obtain that $E(X_{\frac{p}{q}}) = pE(X_{\frac{1}{q}}) = \frac{p}{q}E(X_1) = \lambda\frac{p}{q}$.
Finally, the statement holds as well for irrational t, as $(X_t)_{t\geq 0}$ is almost surely nondecreasing and therefore $E(X_t)$ is monotonically increasing. □

Associated with a homogeneous signal process is a random variable called the *waiting time* until the first signal occurs, given by

$$W(x) = \min\{t : X_t(x) \geq 1\}.$$

The most famous example of a homogeneous signal process is the *Poisson Process*, which we define as follows:

Definition 3.4. *The Poisson Process with intensity λ is a homogeneous signal process with jump unit 1, satisfying*

(i) The increments in (3.1) *are independent.*

(ii) The random variables X_t are Poisson distributed with expectation λ, i.e.

$$\mu\{X_t = k\} = e^{-\lambda t}\frac{(\lambda t)^k}{k}$$

holds for all integers $k \geq 0$ and for all $t \in [0, \infty)$.

A further characterization of Poisson processes can be found in [5, Satz 3.1.3].
Property (ii) of the Poisson process implies, that the waiting time distribution of such a process is given by

$$F_W(t) = \mu\{W \leq t\} = \mu\{X_t \geq 1\} = 1 - \{X_t = 0\} = 1 - e^{-\lambda t}.$$

The independence of the increments in the Poisson process mean, that for any fixed time t, signals arriving before t do not influence future signals, or equivalently, that the signals arrive independently from one another.
The Poisson process will therefore be our reference point of a process with independent signals, from which we consider two deviations, called *attracting* and *repelling*. Informally, signals attract each other when they are likely to occur in groups (or series), separated by long periods of absence, whereas signals repel each other, when they tend to arrive evenly distributed.
In mathematical terms, attracting and repelling are defined as follows:

Definition 3.5. *The signals of a homogeneous signal process with intensity λ* attract *each other from a distance $t > 0$, if*

$$F_W(t) < 1 - e^{-\lambda t}.$$

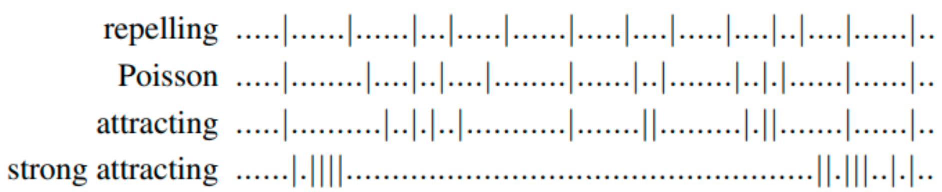
The signals of a homogeneous signal process repel *each other from a distance $t > 0$, if*

$$F_W(t) > 1 - e^{-\lambda t}.$$

Note that both terms depend only on the distribution of the waiting time. This gets more comprehensive, when we consider the fact that for each random variable X_t (which equals the number of signals within the time period $(0, t]$), it holds that $\mu\{X_t > 0\} = \mu\{W \leq t\} = F_W(t)$. Then, any homogeneous signal process with intensity λ satisfies

$$E(X_t | X_t > 0) = \frac{E(X_t)}{\mu\{X_t > 0\}} = \frac{\lambda t}{F_W(t)}.$$

The quantity $E(X_t | X_t > 0)$ gives us the expected number of signals observed within the time span $(0, t]$, in the case that at least one signal occurs. If we observe at least one signal, then attracting from the distance t means, that we expect more signals to occur in $(0, t]$ compared to the Poisson process with the same intensity λ. Similarly, the signals repel from each other from a distance t, whenever the first signal lowers the expected number of occurring signals in the observation period.
In the figure below we give a naive illustration how signals typically occur in different processes with the same intensity:



The difference $|1 - e^{-\lambda t} - F_W(t)|$ is called the *force of attracting/repelling* at time t. The force of attracting can be arbitrarily close to 1, which happens in processes where the distribution function of the waiting time stays near zero until large t. Such processes are characterized by long absences of signals, including a high probability of a

long waiting time, followed by a huge amount of signals within a short period of time, again followed by a long time span without signals. Such a behaviour of a process is called *strong attracting* (note that we will not give a formal definition of this relatively new term in literature).
In contrast, maximal repelling is obtained, when signals occur periodically (with gaps equal $1/\lambda$). The distribution function then satisfies $F_W(t) = \min\{\lambda t, 1\}$, and the maximal force of repelling is given at $t = 1/\lambda$ and equals $1/e$.

Figure 3.1.: The distribution function of the waiting time F_W in the Poisson process, a strongly attracting process and a strongly repelling process.

If a process reveals attracting from one distance, and repelling from another distance (which means that the distribution function of the process satisfies $F_W(t) > 1 - e^{-\lambda t}$ as well as $F_W(t) < 1 - e^{-\lambda t}$, respectively for at least one $t > 0$), the behaviour of the signals depends on the applied time perspective, and the tendency of clustering is unclear. In order to treat a Law of Series, we want stronger conditions:

Definition 3.6. *A homogeneous signal process obeys the* Law of Series*, if*

$$F_W(t) \leq 1 - e^{-\lambda t}$$

holds for all $t \geq 0$*, and*

$$F_W(t) < 1 - e^{-\lambda t}$$

holds for at least one $t \geq 0$*.*

In written words, the Law of Series can be characterized by the following two postulates:

(1) There is no repelling from any distance.

(2) There is attracting from at least one distance.

3.2. The repelling postulate of the Law of Series

In this section, which is the main part of this work, we are going to provide a proof of [2, Theorem 1], the repelling postulate of the Law of Series. We rely on [2] as well as on [4, Section 5.3].

Let $(X, \mathcal{A}, \mu, T)$ be an invertible ergodic probability preserving dynamical system on a standard probability space $(X, \mathcal{A}, \mu)$, and let $B \subset X$ denote any measurable set of positive measure. We consider the discrete time process

$$X_t(x) = \#\{n \in (0, t] : T^n(x) \in B\}, \tag{3.3}$$

whose signals are the occurrences of the event B, where $t \geq 0$ takes on integer values. Applying the Ergodic Theorem (for $f = 1_B$), we can see that the intensity λ of the signals equals $\mu(B)$. Using that T preserves μ, it easily follows that the process is homogeneous, hence by Lemma 3.1 we have $E(X_t) = \lambda t$ for all nonnegative integers t.
Associated with this process is the *waiting time until the first occurrence of* B, given by

$$W_B(x) = \min\{t : X_t(x) \geq 1\} = \min\{n \geq 1 : T^n(x) \in B\}.$$

There is a relation between the distribution F_B of the waiting time W_B and the distribution G_B of the return time φ_B to the set B, which is stated in the following lemma:

Lemma 3.2. *In an ergodic probability preserving system* $(X, \mathcal{A}, \mu, T)$ *it holds that*

$$F_B(t) = \mu(B) \sum_{i=0}^{t-1} (1 - G_B(i)) \tag{3.4}$$

Proof. Using Kac Formula (Theorem 2.3(iii)), we have

$$\begin{aligned}
\mu\{W_B = n\} &= \int_X 1_{\{W_B = n\}} d\mu = \int_B \sum_{k=0}^{\varphi_B - 1} 1_{\{\varphi_B = n\}} \circ T^k d\mu = \int_B \sum_{k=0}^{\varphi_B - 1} 1_{\{\varphi_B = n+k\}} d\mu \\
&= \sum_{k \geq 1} \mu(B \cap \{\varphi_B = k\} \cap \{\varphi_B \in \{n, n+1, ..., n+k-1\}\}) \\
&= \mu(B \cap \{\varphi_B \geq n\}) = \mu(B)\mu_B\{\varphi_B \geq n\}.
\end{aligned}$$

This implies

$$\begin{aligned}
F_B(t) = \mu\{W_B \leq t\} &= \sum_{i=1}^{t} \mu\{W_B = i\} = \sum_{i=1}^{t} \mu(B)\mu_B\{\varphi_B \geq i\} \\
&= \sum_{i=1}^{t} \mu(B)(1 - \mu_B\{\varphi_B \leq i-1\}) = \sum_{i=0}^{t-1} \mu(B)(1 - G_B(i)).
\end{aligned} \tag{3.5}$$

□

The above lemma shows, that the waiting time and the return time determine each other, which will allow us to replace the waiting time by the return time in our calculations.

Our next step is to normalize the process (3.3) by changing the time unit to $1/\lambda = 1/\mu(B)$. Our normalized process $\overline{X}_t$ is then given by

$$\overline{X}_t(x) = X_{\frac{t}{\lambda}}(x)$$

with time steps $t \in T = \{k\mu(B) : k \in \mathbb{N}\}$, and it is homogeneous for all $t \in T$.

The waiting time of this process $\overline{W}_B$, called the *normalized waiting time*, then satisfies $\overline{W}_B = \mu(B)W_B$. Similarly, for the *normalized return time* $\overline{\varphi}_B$ it holds that $\overline{\varphi}_B = \mu(B)\varphi_B$.

This normalization has many advantages. First of all, as the sets B we are going to look at have very small measure, the time increments $\lambda = \mu(B)$ are small such that the process $\overline{X}_t(x)$ becomes nearly continuous. Therefore, we can confidently call it a (homogeneous) signal process in the sense of Definition 3.1. Additionally, the parameter λ disappears from the calculations, as by homogeneity for all $t \in T$ and $k \in \mathbb{N}$ it now holds that

$$E(\overline{X}_t) = E(\overline{X}_{\lambda k}) = E(X_k) = \lambda k = t.$$

Similar to identity (3.4), the normalized distribution functions $\overline{F}_B$ of $\overline{W}_B$ and $\overline{G}_B$ of $\overline{\varphi}_B$ are related by the following integral estimate:

Lemma 3.3. *Let B denote a set of positive measure in an ergodic probability preserving system $(X, \mathcal{A}, \mu, T)$. Then*

$$\overline{F}_B(t) \leq \int_0^t 1 - \overline{G}_B(s)ds \leq \overline{F}_B(t) + \mu(B).$$

Proof. Using (3.5) in the proof of the previous lemma and integration by substitution, we obtain

$$\begin{aligned}
\overline{F}_B(t) &= \mu\{W_B \leq \frac{t}{\mu(B)}\} = \sum_{i=1}^{\lfloor \frac{t}{\mu(B)} \rfloor} \mu\{W_B = i\} = \sum_{i=1}^{\lfloor \frac{t}{\mu(B)} \rfloor} \mu(B)\mu_B\{\varphi_B \geq i\} \\
&= \int_0^{\lfloor \frac{t}{\mu(B)} \rfloor} \mu(B)\mu_B\{\varphi_B \geq r\}dr = \int_0^{\lfloor \frac{t}{\mu(B)} \rfloor \mu(B)} \mu_B\{\varphi_B \geq \frac{s}{\mu(B)}\}ds \\
&\leq \int_0^t \mu_B\{\varphi_B \geq \frac{s}{\mu(B)}\}ds = \int_0^t \mu_B\{\overline{\varphi}_B > s\}ds = \int_0^t 1 - \overline{G}_B(s)ds.
\end{aligned}$$

The second inequality follows from the fact that

$$\begin{aligned}
\int_0^t 1 - \overline{G}_B(s)ds &= \int_0^{\frac{t}{\mu(B)}} \mu(B)\mu_B\{\varphi_B \geq r\}dr \\
&= \int_0^{\lfloor \frac{t}{\mu(B)} \rfloor} \mu(B)\mu_B\{\varphi_B \geq r\}dr + \int_{\lfloor \frac{t}{\mu(B)} \rfloor}^{\frac{t}{\mu(B)}} \mu(B)\mu_B\{\varphi_B \geq r\}dr \\
&\leq \overline{F}_B(t) + \mu(B).
\end{aligned}$$

□

Let now $\mathcal{P}$ be a finite partition of X. We are interested in the waiting time distributions of the cylinders $B \in \mathcal{P}^m$ for large $m \in \mathbb{N}$.
For each $x \in X$, let

$$Rep_m(x) = \sup_{t \geq 0}(\overline{F}_{A_x^m}(t) - 1 + e^{-t})$$

define the maximal force of repelling of the cylinder $A_x^m = B \in \mathcal{P}^m$ containing x. With this, we can state our main theorem:

Theorem 3.1. *(The Law of Series, repelling postulate)[2, Theorem 1]*
Let $(X, \mathcal{P}, \mu, T)$ be an invertible ergodic process generated by a finite partition $\mathcal{P}$ on a standard probability space $(X, \mathcal{A}, \mu)$, and let the dynamical entropy of the process satisfy $h(\mathcal{P}) > 0$. Then

$$Rep_m \xrightarrow[m \to \infty]{} 0 \;\; in \;\; L^1(\mu).$$

Note that by Lemma A.1, L_1 convergence of Rep_m is equivalent to convergence in measure. Thus, Theorem 3.1 can be stated equivalently as follows: For every $\epsilon > 0$, the measure of the union of all cylinders of length m which repel with force ϵ converges to zero as m goes to infinity. In particular, we want to show that for every $\epsilon > 0$

$$\mu\{\bigcup B \in \mathcal{P}^m : \sup_{t \geq 0}(\overline{F}_B(t) - 1 + e^{-t}) \geq \epsilon\} \leq \epsilon \tag{3.6}$$

holds for sufficiently large m.

The proof of Theorem 3.1

Before we start with the formal proof of Theorem 3.1, we will briefly discuss the idea behind it.
Note that from now on, we will replace ϵ by β in Definition 2.11, i.e. we will talk about β-independence instead of ϵ-independence, as ϵ will be reserved for the force of repelling in Theorem 3.1.
Assuming a finite partition $\mathcal{P}$, and a positive dynamical entropy $h(\mathcal{P})$ of the process, we want to obtain, for fixed $\epsilon > 0$, the estimate

$$\overline{F}_B(t) \leq 1 - e^{-t} + \epsilon$$

for all $t \geq 0$, and for most cylinders B with sufficiently long length $m \geq m_0$. Instead of cylinders of the partition $\mathcal{P}^m$ we will work with a concatenation of cylinders $BA \in \mathcal{P}^{[-n,r)}$ (i.e. the cylinders $B \cap A$ with $B \in \mathcal{P}^{-n}$, $A \in \mathcal{P}^r$) with $n + r = m$. These cylinders consist of a 'positive' part A with fixed length r, and a 'negative' part B (as it is part of the past $\mathcal{P}^-$) with sufficiently long length n.
Let us now recall Theorem 2.4. Applying this theorem in reverse time, we obtain, together with the power rule, rectangle rule and Lemma 2.14: For every chosen $\beta > 0$ and $r \geq 1$, for sufficiently large n and for nearly all of the cylinders $B \in \mathcal{P}^{-n}$, the process

induced on B generated by the partition $\mathcal{P}^r$, given by $(B, \mathcal{P}^r, \mu_B, T_B)$, is β-independent of the past $\mathcal{P}^-$.
We will show in Lemma 3.9, that for most cylinders $B \in \mathcal{P}^{-n}$, the induced process $(B, \mathcal{P}^r, \mu_B, T_B)$ is not only β-independent of $\mathcal{P}^-$, but additionally of a large part of the future, which contains a finite number of return times $\varphi_B^{(k)}$ of the set B.

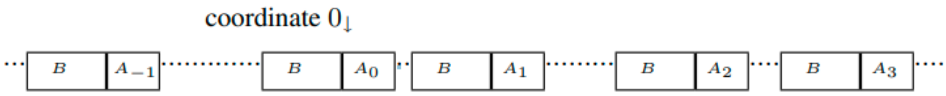


Figure 3.2.: Illustration of the induced process $...A_{-1}A_0A_1A_2A_3...$ of blocks of length r following the returns of B for some $x \in B$. For most B, the induced process is β-independent of the past with additional β-independence properties concerning the position of the returns of B.

This β-independence allows us to estimate with high accuracy the distribution function $\overline{G}_{BA}$ of the normalized return time to BA as follows:

$$\begin{aligned}
\overline{G}_{BA}(t) = \mu_{BA}\{\overline{\varphi}_{BA} \leq t\} &= \mu_{BA}\Big\{\varphi_{BA} \leq \frac{t}{\mu(BA)}\Big\} \\
&= \sum_{k=1}^{\infty} \mu_{BA}\Big\{\varphi_A^{(B)} = k, \varphi_B^{(k)} \leq \frac{t}{p\mu(B)}\Big\} \\
&\approx \sum_{k=1}^{\infty} \mu_{BA}\{\varphi_A^{(B)} = k\}\mu_B\Big\{\overline{\varphi}_B^{(k)} \leq \frac{t}{p}\Big\} \\
&\approx \sum_{k=1}^{\infty} p(1-p)^{k-1}\overline{G}_B^{(k)}\Big(\frac{t}{p}\Big)
\end{aligned} \tag{3.7}$$

where $\varphi_A^{(B)}$ denotes the return time of A in the process generated by $\mathcal{P}^r$ induced on B, and $p = \mu_B(A)$.
With use of Lemma 3.3, it then follows from (3.7) that

$$\overline{F}_{BA}(t) \approx \sum_{k=1}^{\infty} p(1-p)^{k-1} \int_0^t 1 - \overline{G}_B^{(k)}(\frac{s}{p})ds. \tag{3.8}$$

Lemma 3.4 will provide a function, which dominates the right hand side of (3.8) (and therefore $\overline{F}_{BA}$) for all t.
In Lemma 3.6 we will prove, that this function (which dominates the distribution $\overline{F}_{BA}$) converges from above to $1 - e^{-t}$, uniformly for all $t \in [0, \infty)$ when p converges to 0^+, which implies that $\overline{F}_{BA}(t) \leq 1 - e^{-t} + \epsilon$ holds for small enough p.
The smallness of $p = \mu_B(A)$ will be regulated by the length r of the cylinder A in Lemma 3.7.

We will now start with the rigorous proof of Theorem 3.1. For simplicity, we will first assume equality in (3.8) and in the left inequality in the statement of Lemma 3.3. Note that in contrast to [2, Lemma 0], in our calculation an extra term $1-p$ appears in the result, i.e. in the denominator of the function bounding the distribution $\overline{F}(t)$ from above.

Lemma 3.4. *[2, Lemma 0] Let $\overline{G}^{(k)}$, $k \geq 1$, be a sequence of distribution functions on $[0,\infty)$ satisfying $\int_0^\infty 1-\overline{G}^{(k)}(s)ds = k$. For $p \in (0,1)$, let*

$$\overline{G}(t) = \sum_{k=1}^{\infty} p(1-p)^{k-1}\overline{G}^{(k)}(\frac{t}{p}) \quad and \quad \overline{F}(t) = \int_0^t 1-\overline{G}(s)ds.$$

Then, for $e_p = (1-p)^{-\frac{1}{p}}$,

$$\overline{F}(t) \leq \frac{1}{(1-p)\ln(e_p)}(1-e_p^{-t}).$$

Proof. Let $p \in (0,1)$. Using the identity $p\sum_{k=1}^{\infty}(1-p)^{k-1} = 1$, we have

$$\overline{F}(t) = \sum_{k=1}^{\infty} p(1-p)^{k-1}\int_0^t 1-\overline{G}^{(k)}(\frac{s}{p})ds. \tag{3.9}$$

For every $k \geq 1$, the indicator function $\mathbb{1}_{[k,\infty)}$ is a distribution function for which $\int_0^\infty 1-\overline{G}^{(k)}(s)ds = k$ holds. We claim that this function maximizes the value of the integral $\int_0^t 1-\overline{G}^{(k)}(\frac{s}{p})ds$ for every $t \in [0,\infty)$and $k \geq 1$. We have

$$\int_0^t 1-\mathbb{1}_{[k,\infty)}(\frac{s}{p})ds = \int_0^t 1-\mathbb{1}_{[kp,\infty)}(s)ds = \int_0^t \mathbb{1}_{[0,kp)}(s)ds = \begin{cases} t & if\ t \leq kp \\ kp & if\ t > kp \end{cases}$$

As $1-\overline{G}^{(k)}(\frac{s}{p})$ lies inside the interval $[0,1]$ for every $s \in [0,t]$, the function $\mathbb{1}_{[k,\infty)}$ maximizes the corresponding integral for $t \leq kp$. When t is larger than kp, we obtain by using integration by substitution

$$\int_0^t 1-\overline{G}^{(k)}(\frac{s}{p})ds = p\int_0^t (1-\overline{G}^{(k)}(\frac{s}{p}))\frac{1}{p}ds = p\int_0^{\frac{t}{p}} 1-\overline{G}^{(k)}(x)dx \leq p\int_0^\infty 1-\overline{G}^{(k)}(x)dx = kp,$$

which proves our claim. Now, applying monotone convergence theorem, this implies

$$\begin{aligned}\overline{F}(t) &= \sum_{k=1}^{\infty} p(1-p)^{k-1}\int_0^t 1-\overline{G}^{(k)}(\frac{s}{p})ds \\ &\leq \sum_{k=1}^{\infty} p(1-p)^{k-1}\int_0^t \mathbb{1}_{[0,kp)}(s)ds = \int_0^t \sum_{k=1}^{\infty} p(1-p)^{k-1}\mathbb{1}_{[0,kp)}(s)ds. \end{aligned}\tag{3.10}$$

Let now n be the integer such that $np \leq t < (n+1)p$ holds. Splitting the integral into np parts of size p (and one part of size $t - np$), the last term in (3.10) reads

$$\sum_{j=1}^{n} \int_{(j-1)p}^{jp} \sum_{k=1}^{\infty} p(1-p)^{k-1} \mathbb{1}_{[0,kp)}(s) ds + \int_{np}^{t} \sum_{k=1}^{\infty} p(1-p)^{k-1} \mathbb{1}_{[0,kp)}(s) ds.$$

As the indicator functions $\mathbb{1}_{[0,kp)}$ vanish on the integrals if k is smaller than j, this is the same as

$$\sum_{j=1}^{n} \int_{(j-1)p}^{jp} \sum_{k=j}^{\infty} p(1-p)^{k-1} ds + \int_{np}^{t} \sum_{k=n+1}^{\infty} p(1-p)^{k-1} ds. \tag{3.11}$$

For each j, the integral terms remain the same value when we substitute j by $\lceil \frac{s}{p} \rceil$ inside the sum, therefore (3.11) equals

$$\sum_{j=1}^{n} \int_{(j-1)p}^{jp} \sum_{k=\lceil \frac{s}{p} \rceil}^{\infty} p(1-p)^{k-1} ds + \int_{np}^{t} \sum_{k=\lceil \frac{s}{p} \rceil}^{\infty} p(1-p)^{k-1} ds = \int_{0}^{t} \sum_{k=\lceil \frac{s}{p} \rceil}^{\infty} p(1-p)^{k-1} ds.$$

For every $s \in [0, t]$ it holds that

$$\sum_{k=\lceil \frac{s}{p} \rceil}^{\infty} p(1-p)^{k-1} = 1 - \sum_{k=1}^{\lceil \frac{s}{p} \rceil - 1} p(1-p)^{k-1} = 1 - p\Big(\frac{1-(1-p)^{\lceil \frac{s}{p} \rceil - 1}}{1-(1-p)}\Big) = (1-p)^{\lceil \frac{s}{p} \rceil - 1}.$$

Finally, this implies

$$\begin{aligned} \overline{F}(t) &\leq \int_0^t \sum_{k=\lceil \frac{s}{p} \rceil}^{\infty} p(1-p)^{k-1} ds = \int_0^t (1-p)^{\lceil \frac{s}{p} \rceil - 1} ds \\ &\leq \int_0^t (1-p)^{\frac{s}{p} - 1} ds = \frac{p((1-p)^{\frac{t}{p}} - 1)}{(1-p)\ln(1-p)} = \frac{1 - e_p^{-t}}{(1-p)\ln(e_p)} \end{aligned}$$

□

Note that we have proven, recalling (3.8), that the distribution $\overline{F}_{BA}$ of the normalized waiting time $\overline{W}_{BA}$ is maximal, when $\overline{G}_B^{(k)} = 1_{[k,\infty)}$ holds for all k. This is exactly the case in a process where B is visited periodically. This implies, that the repelling of BA is strongest, when the repelling of B is strongest.
Next, we are going to state the version of the above lemma which will be required in the proof of the main theorem, i.e. allowing an error of δ in (3.7) and taking into account the left inequality in Lemma 3.3.

Corollary 3.1. *Let $\overline{G}^{(k)}$, $k \geq 1$, be a sequence of distribution functions on $[0, \infty)$ satisfying $\int_0^\infty 1 - \overline{G}^{(k)}(s) ds = k$. For $p \in (0,1)$ and $\delta > 0$, let*

$$\overline{G}(t) \leq \sum_{k=1}^{\infty} p(1-p)^{k-1} \overline{G}^{(k)}(\frac{t}{p}) + \delta \quad \text{and} \quad \overline{F}(t) \leq \int_0^t 1 - \overline{G}(s) ds.$$

Then, for $e_p = (1-p)^{-\frac{1}{p}}$,

$$\overline{F}(t) \leq \frac{1}{(1-p)\ln(e_p)}(1-e_p^{-t}) + \delta t.$$

Our next lemma is an auxiliary lemma, providing some properties of the functions e_p:

Lemma 3.5. *For* $p \in (0,1)$*, let* $e_p = (1-p)^{-\frac{1}{p}}$*. Then, the following hold:*

(i) $e_p > e$.

(ii) $(1-p)\ln(e_p) < 1$.

(iii) $(1-p)\ln(e_p) \to 1$ *when* $p \to 0^+$.

(iv) $p \to e_p$ *is increasing.*

(v) $e_p \to e$ *when* $p \to 0^+$.

Proof. (i) Using the inequality $1-p < e^{-p}$, we obtain

$$e_p = (1-p)^{-\frac{1}{p}} = \frac{1}{(1-p)^{\frac{1}{p}}} > \frac{1}{(e^{-p})^{\frac{1}{p}}} = e.$$

(ii) Series expansion of the logarithm gives us

$$\ln(e_p) = \ln((1-p)^{-\frac{1}{p}}) = -\frac{1}{p}\ln(1-p) = -\frac{1}{p}\sum_{k\geq 1} -\frac{p^k}{k} = \sum_{k\geq 1}\frac{p^{k-1}}{k} = 1 + \sum_{k\geq 1}\frac{p^k}{k+1}. \tag{3.12}$$

This implies

$$(1-p)\ln(e_p) = (1-p)(1+\sum_{k\geq 1}\frac{p^k}{k+1}) = 1+\sum_{k\geq 1}\frac{p^k}{k+1}-\sum_{k\geq 1}\frac{p^k}{k} = 1-\sum_{k\geq 1}\frac{p^k}{k(k+1)} < 1. \tag{3.13}$$

(iii) Follows immediately from (3.13).

(iv) Follows from (3.12) as the logarithm is monotonically increasing.

(v) We substitute p by $\frac{1}{n}$. Then, $e_p = e_{\frac{1}{n}} = (1-\frac{1}{n})^{-n}$. By monotonicity of e_p (stated in (iv)), this gives us

$$\lim_{p\to 0^+} e_p = \lim_{n\to\infty} e_{\frac{1}{n}} = \lim_{n\to\infty}\Big(1-\frac{1}{n}\Big)^{-n} = e.$$

□

With help of this lemma, we can now show that the estimate for the distribution function $\overline{F}_{BA}$ of the normalized waiting time to BA provided in Corollary 3.1 implies uniform convergence to $1-e^{-t}$, when p converges to 0 from above:

Lemma 3.6. *Let $e_p = (1-p)^{-\frac{1}{p}}$. Then, for $t \in [0,\infty)$ the functions*

$$g_p(t) = min\{1, \frac{1}{(1-p)\ln(e_p)}(1-e_p^{-t}) + pt\}$$

converge uniformly to $1-e^{-t}$ from above, as p decreases to zero.

Proof. We first want to show that $g_p(t)$ is not smaller than $1-e^{-t}$. It is clear that the constant function 1 satisfies this condition for all $t \geq 0$. Taking the difference of the second function of $g_p(t)$ and $1-e^{-t}$, we have, using *(i)* and *(ii)* of Lemma 3.5,

$$\begin{aligned}\frac{1}{(1-p)\ln(e_p)}(1-e_p^{-t}) + pt - (1-e^{-t}) &\geq \frac{1}{(1-p)\ln(e_p)}(1-e^{-t}) + pt - (1-e^{-t}) \\ &= \Big(\frac{1}{(1-p)\ln(e_p)} - 1\Big)\Big(1-e^{-t}\Big) + pt \geq 0.\end{aligned}$$

We now fix $\epsilon > 0$. We intend to find a $\delta > 0$ such that

$$|g_p(t) - 1 + e^{-t}| \leq \epsilon$$

holds for all $0 < p < \delta$ and for all $t \in [0,\infty)$. As $|1-(1-e^{-t})| \leq \epsilon$ holds for all t being greater than or equal to $-\ln(\epsilon)$, the functions $g_p(t)$ decrease uniformly to $1-e^{-t}$ for all $t \in [-\ln(\epsilon),\infty)$. It remains to be shown that

$$\frac{1}{(1-p)\ln(e_p)}(1-e_p^{-t}) + pt - 1 + e^{-t} \leq \epsilon$$

holds for all small enough p and for all $t \in [0,-\ln(\epsilon))$. Note that we omitted the absolute value on the left side as it is positive for all $t \geq 0$. Let us now consider the auxiliary function $h_p(t) = e^{-t} - e_p^{-t}$. This function is nonnegative for $t \geq 0$ and has a unique positive maximum at $t_{max}(p) = \frac{\ln(\ln(e_p))}{\ln(e_p)-1}$. By (i) of Lemma 3.5, we know that $y_p = \ln(e_p) > 1$ is satisfied for any positive p. Recalling the fact that $\ln(1+p) < p$ holds for all $p > 0$, we obtain that $\ln(y_p) < y_p - 1$. This implies, for arbitrary positive p

$$t_{max}(p) = \frac{\ln(y_p)}{y_p - 1} < 1. \tag{3.14}$$

We now assume $\delta > 0$ to satisfy

$$\delta \leq \frac{\epsilon}{3(-\ln(\epsilon))}, \quad e_\delta - e \leq \frac{\epsilon}{3} \quad and \quad \frac{1}{(1-\delta)\ln(e_\delta)} - 1 \leq \frac{\epsilon}{3}.$$

Note that the third assumption on δ is possible by (ii) and (iii) of Lemma 3.5. Considering again our auxiliary function $h_p(t)$, using (3.14), we see that

$$h_p(t) = e^{-t} - e_p^{-t} = \frac{e_p^t - e^t}{e^t e_p^t} \leq \frac{e_p^{t_{max}(p)} - e^{t_{max}(p)}}{e^{t_{max}(p)} e_p^{t_{max}(p)}} \leq e_p^{t_{max}(p)} - e^{t_{max}(p)} < e_p - e \leq \frac{\epsilon}{3}$$

holds for all positive p smaller than δ (and for all $t \geq 0$). Now, for arbitrary $p \in (0, \delta)$ and $t \in [0, -ln(\epsilon))$ we obtain

$$\begin{aligned}
\frac{1}{(1-p)\ln(e_p)}(1 - e_p^{-t}) + pt - 1 + e^{-t} &\leq \frac{1}{(1-p)\ln(e_p)}(1 - e_p^{-t}) + pt - 1 + e_p^{-t} + \frac{\epsilon}{3} \\
&= \Big(\frac{1}{(1-p)\ln(e_p)} - 1\Big)\Big(1 - e_p^{-t}\Big) + pt + \frac{\epsilon}{3} \\
&\leq \frac{1}{(1-\delta)\ln(e_\delta)} - 1 + \delta t + \frac{\epsilon}{3} \leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon.
\end{aligned}$$

□

We continue by defining the term *μ-tolerance*, which we will frequently use:

Definition 3.7. *Let $\mathcal{P}$ be a finite partition in a probability space $(X, \mathcal{A}, \mu)$, and let $\delta > 0$. We say that a property $\Phi(B)$ holds with μ-tolerance δ, if*

$$\mu\Big(\bigcup\{B \in \mathcal{P} : \Phi(B)\}\Big) \geq 1 - \delta.$$

Let us also recall the inequality stated in Lemma 2.2, i.e. that

$$H(\mathcal{P}) \leq (1 - \mu(A)) \log \#\mathcal{P} + 1 \tag{3.15}$$

holds for every element A of a finite partition $\mathcal{P}$.

Before continuing with the proof of Theorem 3.1, we quickly remind ourselves of the properties of our process, i.e. that $(X, \mathcal{P}, \mu, T)$ is an ergodic invertible process generated by a finite partition $\mathcal{P}$ with positive dynamical entropy $h(\mathcal{P}) = h$. We will use these properties in the upcoming arguments of this section without recalling them each time.

Lemma 3.7. *[2, Lemma 1] For every $\delta > 0$ there is an $r \in \mathbb{N}$ such that for every $n \in \mathbb{N}$ the following holds for $B \in \mathcal{P}^{-n}$ with μ-tolerance δ and for all $A \in \mathcal{P}^r$:*

$$\mu_B(A) \leq \delta.$$

Proof. Let $\delta > 0$ be given. We assume α to be small enough to satisfy

$$\sqrt{\alpha} \leq \delta \quad \textit{and} \quad \frac{h + \alpha - 3\sqrt{\alpha}}{h + \alpha} \geq 1 - \frac{\delta}{2}$$

and we set $\gamma = \alpha / \log \#\mathcal{P}$. We choose r so large that

$$\frac{1}{r} \leq \alpha, \quad \frac{1}{r(h+\alpha)} \leq \frac{\delta}{2},$$

and, using Shannon-McMillan-Breilman Theorem, that

$$\mu\{x \in X : 2^{-r(h+\alpha)} < \mu(A_r^x) < 2^{-r(h-\alpha)}\} > 1 - \gamma.$$

This implies, that most of the mass of the partition $\mathcal{P}^r$ (up to probability γ) consists of sets whose probability lies within the interval $(2^{-r(h+\alpha)}, 2^{-r(h-\alpha)})$. Let$\overline{\mathcal{P}^r}$ denote the collection of sets of $\mathcal{P}^r$ with the above property. Then, as the probability μ of each element is larger than $2^{-r(h+\alpha)}$, $\overline{\mathcal{P}^r}$ has at most $2^{r(h+\alpha)} - 1$ elements.
Let $\widetilde{\mathcal{P}^r}$ be the partition of X into the elements of $\overline{\mathcal{P}^r}$ and the complement of their union, and let $\mathcal{R}$ be the partition of the remaining elements of $\mathcal{P}^r$ and the union of the elements of $\overline{\mathcal{P}^r}$, such that $\mathcal{P}^r = \widetilde{\mathcal{P}^r} \vee \mathcal{R}$. By the power rule (Corollary 2.2) and by applying (3.15) for the largest element of $\mathcal{R}$ (whose measure exceeds $1-\gamma$) in the last line, we have

$$\begin{aligned} rh &= H(\mathcal{P}^r|\mathcal{P}^-) \leq H(\mathcal{P}^r|\mathcal{P}^{-n}) = H(\widetilde{\mathcal{P}^r} \vee \mathcal{R}|\mathcal{P}^{-n}) \\ &= H(\widetilde{\mathcal{P}^r}|\mathcal{R} \vee \mathcal{P}^{-n}) + H(\mathcal{R}|\mathcal{P}^{-n}) \leq H(\widetilde{\mathcal{P}^r}|\mathcal{P}^{-n}) + H(\mathcal{R}) \\ &\leq \sum_{B \in \mathcal{P}^{-n}} \mu(B) H_B(\widetilde{\mathcal{P}^r}) + \gamma r \log \#\mathcal{P} + 1. \end{aligned}$$

Note that in the last part we also used the fact that $\mathcal{P}^r$ has not less elements than $\mathcal{R}$, i.e. that $\#\mathcal{R} \leq \#\mathcal{P}^r \leq (\#\mathcal{P})^r$. Dividing by r, we deduce

$$\sum_{B \in \mathcal{P}^{-n}} \mu(B) \frac{1}{r} H_B(\widetilde{\mathcal{P}^r}) \geq h - \gamma \log \#\mathcal{P} - \frac{1}{r} \geq h - 2\alpha. \tag{3.16}$$

For every $B \in \mathcal{P}^{-n}$, $\frac{1}{r} H_B(\widetilde{\mathcal{P}^r})$ is not greater than $\frac{1}{r} \log \#\widetilde{\mathcal{P}^r}$, which is at most $h + \alpha$ (as $\widetilde{\mathcal{P}^r}$ has not more than $2^{r(h+\alpha)}$ elements). This estimate, together with (3.16), leads to

$$h + \alpha - \sum_{B \in \mathcal{P}^{-n}} \mu(B) \frac{1}{r} H_B(\widetilde{\mathcal{P}^r}) = \int_{\mathcal{P}^{-n}} \Big(h + \alpha - \frac{1}{r} H_B(\widetilde{\mathcal{P}^r})\Big) d\mu(B) \leq 3\alpha = (3\sqrt{\alpha})(\sqrt{\alpha})$$

and applying rectangle rule, we obtain that

$$\frac{1}{r} H_B(\widetilde{\mathcal{P}^r}) \geq h + \alpha - 3\sqrt{\alpha}$$

with μ-tolerance $\sqrt{\alpha}$, and by choice of α also with μ-tolerance δ. On the other hand, again by (3.15), for each $A \in \widetilde{\mathcal{P}^r}$ and $B \in \mathcal{P}^{-n}$ we have

$$\frac{1}{r} H_B(\widetilde{\mathcal{P}^r}) \leq \frac{1}{r}(1 - \mu_B(A)) \log \#\widetilde{\mathcal{P}^r} + \frac{1}{r} \leq (1 - \mu_B(A))(h + \alpha) + \frac{1}{r}$$

Combining these two estimates for $H_B(\widetilde{\mathcal{P}^r})$, we observe that, for $B \in \mathcal{P}^{-n}$ with μ-tolerance δ and for all $A \in \widetilde{\mathcal{P}^r}$, it holds that

$$1 - \mu_B(A) \geq \frac{h + \alpha - 3\sqrt{\alpha}}{h + \alpha} - \frac{1}{r(h + \alpha)} \geq 1 - \delta,$$

which gives us $\mu_B(A) \leq \delta$. Finally, as $\mathcal{P}^r$ is a refinement of $\widetilde{\mathcal{P}^r}$, the same holds for every element of $\mathcal{P}^r$. $\square$

For $\alpha > 0$ and $M \in \mathbb{N}$ we define the periodic subset of $\mathbb{Z}$

$$\mathbb{D}(M, \alpha) = \bigcup_{m \in \mathbb{Z}} [mM + \alpha M, (m+1)M - \alpha M) \cap \mathbb{Z}.$$

By the power rule, we know that the entropy of the partition $\mathcal{P}^r$ conditional on the past $\mathcal{P}^-$ equals rh. The join of $\mathcal{P}^-$and the σ-algebra $\mathcal{P}^{\mathbb{D}(M,\alpha)}$ covers the full past and a large part of the future $\mathcal{P}^+$ (with periodic gaps, see Figure 3.3).
The lemma below shows, that for sufficiently large M, the entropy of $\mathcal{P}^r$ conditional on that join remains close to the value rh.

***********oo..************.........************.........************.........

Figure 3.3.: The coordinates 0 through $r-1$ of the partition $\mathcal{P}^r$ are marked by circles. The stars denote the coordinates of the conditioning σ-algebra $\mathcal{P}^- \vee \mathcal{P}^{\mathbb{D}(M,\alpha)}$, including the entire past and a large part of the future with periodic gaps of size $\lceil 2\alpha M \rceil$.

Lemma 3.8. *[2, Lemma 2] For every $\alpha \in (0,1)$ and $r \in \mathbb{N}$, there exists some $M_0 \in \mathbb{N}$, such that for every $M \geq M_0$ it holds that*

$$H(\mathcal{P}^r | \mathcal{P}^- \vee \mathcal{P}^{\mathbb{D}(M,\alpha)}) \geq rh - \alpha. \tag{3.17}$$

Proof. We first consider the case where $r = 1$. Let

$$\mathbb{D}'(M, \alpha) = \bigcup_{m \in \mathbb{Z}} [mM + \alpha M, (m+1)M) \cap \mathbb{Z}$$

and $\gamma = \frac{\alpha^2}{2(1-\alpha)}$. Recalling the definition of dynamical entropy, we can choose M_0 large enough, such that for every $M \geq M_0$ it holds that $\frac{1}{\lfloor (1-\alpha)M \rfloor} H(\mathcal{P}^{\lfloor (1-\alpha)M \rfloor}) - h < \gamma$. This implies

$$H(\mathcal{P}^{\lfloor (1-\alpha)M \rfloor}) < \lfloor (1-\alpha)M \rfloor (h + \gamma).$$

For every $m \geq 1$,the periodic set $[0, mM) \cap \mathbb{D}'(M, \alpha)$ contains all integers inside the interval $[0, mM)$, except for periodic gaps containing the integers $\{iM, iM+1, .., iM +$

$\lceil \alpha M \rceil - 1\}$, for $i \in \{0, 1, ..., m-1\}$. Using Lemma 2.6, Corollary 2.1, and the identity $\lfloor (1-\alpha)M \rfloor = M - \lceil \alpha M \rceil$, the entropy of the partition $\mathcal{P}^{[0,mM)\cap \mathbb{D}'(M,\alpha)}$ satisfies

$$\begin{aligned}
H(\mathcal{P}^{[0,mM)\cap \mathbb{D}'(M,\alpha)}) &= H\left(\bigvee_{i=0}^{m-1} \bigvee_{j=\lceil \alpha M \rceil + iM}^{(i+1)M-1} T^{-j}(\mathcal{P})\right) = H\left(\bigvee_{i=0}^{m-1} \bigvee_{j=iM}^{(i+1)M-1-\lceil \alpha M \rceil} T^{-j}(\mathcal{P})\right) \\
&\leq \sum_{i=0}^{m-1} H\left(\bigvee_{j=iM}^{(i+1)M-1-\lceil \alpha M \rceil} T^{-j}(\mathcal{P})\right) \\
&= \sum_{i=0}^{m-1} H\left(T^{iM}\left(\bigvee_{j=iM}^{(i+1)M-1-\lceil \alpha M \rceil} T^{-j}(\mathcal{P})\right)\right) \\
&= \sum_{i=0}^{m-1} H\left(\bigvee_{j=0}^{\lfloor (1-\alpha)M \rfloor - 1} T^{-j}(\mathcal{P})\right) \\
&= mH(\mathcal{P}^{\lfloor (1-\alpha)M \rfloor}) < m\lfloor (1-\alpha)M \rfloor (h+\gamma). \qquad (3.18)
\end{aligned}$$

For the entropy of the complementary partition $\mathcal{P}^{[0,mM)\setminus \mathbb{D}'(M,\alpha)}$ we have

$$H(\mathcal{P}^{[0,mM)\setminus \mathbb{D}'(M,\alpha)} | \mathcal{P}^- \vee \mathcal{P}^{[0,mM)\cap \mathbb{D}'(M,\alpha)}) = \\ H(\mathcal{P}^{[0,mM)\setminus \mathbb{D}'(M,\alpha)} \vee \mathcal{P}^{[0,mM)\cap \mathbb{D}'(M,\alpha)} | \mathcal{P}^-) - H(\mathcal{P}^{[0,mM)\cap \mathbb{D}'(M,\alpha)} | \mathcal{P}^-).$$

As the first term on the right hand side equals $H(\mathcal{P}^{[0,mM)} | \mathcal{P}^-)$, and the second term is smaller than $H(\mathcal{P}^{[0,mM)\cap \mathbb{D}'(M,\alpha)})$, we obtain with inequality (3.18) and the power rule

$$\begin{aligned}
H(\mathcal{P}^{[0,mM)\setminus \mathbb{D}'(M,\alpha)} | \mathcal{P}^- \vee \mathcal{P}^{[0,mM)\cap \mathbb{D}'(M,\alpha)}) > mMh - m\lfloor (1-\alpha)M \rfloor (h+\gamma) = \\
m\lceil \alpha M \rceil (h + \gamma - \frac{\gamma M}{\lceil \alpha M \rceil}) \geq m\lceil \alpha M \rceil (h + \gamma - \frac{\gamma M}{\alpha M}) = m\lceil \alpha M \rceil (h - \frac{\alpha}{2}). \qquad (3.19)
\end{aligned}$$

Applying Lemma 2.5, we can express the above entropy term as

$$\sum_{j \in [0,mM)\setminus \mathbb{D}'(M,\alpha)} H(T^{-j}(\mathcal{P}) | \bigvee_{\substack{i \in [0,mM)\setminus \mathbb{D}'(M,\alpha), \\ i<j}} T^{-i}(\mathcal{P}) \vee \mathcal{P}^- \vee \mathcal{P}^{[0,mM)\cap \mathbb{D}'(M,\alpha)}). \qquad (3.20)$$

Each entropy term inside the sum contains the full past of each partition $T^{-j}(\mathcal{P})$, as

$$\begin{aligned}
(T^{-j}(\mathcal{P}))^- &= \bigvee_{i=1}^{j-1} T^{-i}(\mathcal{P}) \vee \mathcal{P}^- \subseteq \bigvee_{i=1}^{j-1} T^{-i}(\mathcal{P}) \vee \mathcal{P}^- \vee \mathcal{P}^{[0,mM)\cap \mathbb{D}'(M,\alpha)} \\
&= \bigvee_{\substack{i \in [0,mM)\setminus \mathbb{D}'(M,\alpha), \\ i<j}} T^{-i}(\mathcal{P}) \vee \mathcal{P}^- \vee \mathcal{P}^{[0,mM)\cap \mathbb{D}'(M,\alpha)}.
\end{aligned}$$

This implies for every $j \in [0, mM) \setminus \mathbb{D}'(M, \alpha)$ that

$$H(T^{-j}(\mathcal{P}) | \bigvee_{\substack{i \in [0,mM)\setminus \mathbb{D}'(M,\alpha), \\ i<j}} T^{-i}(\mathcal{P}) \vee \mathcal{P}^- \vee \mathcal{P}^{[0,mM)\cap \mathbb{D}'(M,\alpha)}) \leq H(T^{-j}(\mathcal{P}) | (T^{-j}(\mathcal{P}))^-) = h. \qquad (3.21)$$

The sum (3.20) consists of $m\lceil\alpha M\rceil$ entropy terms, which are, by (3.21) all not larger than h. More than half of these terms must be greater than or equal to $h-\alpha$, else the sum would be at most

$$\frac{m\lceil\alpha M\rceil}{2}h+\frac{m\lceil\alpha M\rceil}{2}(h-\alpha)=m\lceil\alpha M\rceil(h-\frac{\alpha}{2}),$$

which contradicts inequality (3.19). As at least half of the elements of $[0,Mm)\backslash\mathbb{D}'(M,\alpha)$ lie inside the left half of $[0,mM)$, there is at least one $j\in[0,\frac{mM}{2})$, where the left hand side of (3.21) is larger than or equal to $h-\alpha$. We can express this j as $kM+i$, where i is an element of the set $\{0,1,...,\lceil\alpha M\rceil-1\}$ and k lies within $\{0,...,\lceil\frac{m}{2}\rceil-1\}$. Shifting by j, this gives us

$$H(T^{-j}(\mathcal{P})|(T^{-j}(\mathcal{P}))^{-}\vee\mathcal{P}^{[0,mM)\cap\mathbb{D}'(M,\alpha)})=H(\mathcal{P}|\mathcal{P}^{-}\vee T^{kM+i}(\mathcal{P}^{[0,mM)\cap\mathbb{D}'(M,\alpha)})=$$
$$H(\mathcal{P}|\mathcal{P}^{-}\vee T^{i}(\mathcal{P}^{[0,(m-k)M)\cap\mathbb{D}'(M,\alpha)})\leq H(\mathcal{P}|\mathcal{P}^{-}\vee T^{i}(\mathcal{P}^{[0,\frac{mM}{2})\cap\mathbb{D}'(M,\alpha)}).$$

So, for every $m\geq 1$, there exists an $i\in\{0,1,...,\lceil\alpha M\rceil-1\}$, such that

$$H(\mathcal{P}|\mathcal{P}^{-}\vee T^{i}(\mathcal{P}^{[0,\frac{mM}{2})\cap\mathbb{D}'(M,\alpha)})\geq h-\alpha. \tag{3.22}$$

As the set $\{0,1,...,\lceil\alpha M\rceil-1\}$ is finite, at least one i repeats in this role for infinitely many m. This implies, that for this i, and the subsequence (m_k) of such m, inequality (3.22) holds whenever $m=m_k$. Now, let $\mathcal{B}_k$ denote the σ-algebras generated by $T^{i}(\mathcal{P}^{[0,\frac{m_kM}{2})\cap\mathbb{D}'(M,\alpha)})$. For all $k\geq 1$, it holds that $\mathcal{B}_k\subseteq\mathcal{B}_{k+1}$. Using Lemma 2.1, the join $\bigvee_{k\in\mathbb{N}}\mathcal{B}_k$ is generated by the family $\{T^{i}\bigcap_{j=1}^{n}T^{-j}(\mathcal{P}):j\in[0,\frac{m_kM}{2})\cap\mathbb{D}'(M,\alpha),n\geq 1,k\geq 1\}$, which also generates the σ-algebra $T^{i}(\mathcal{P}^{\mathbb{D}'(M,\alpha)})$. This implies that $\bigvee_{k\in\mathbb{N}}\mathcal{B}_k=T^{i}(\mathcal{P}^{\mathbb{D}'(M,\alpha)})$ and therefore $\mathcal{P}^{-}\vee\bigvee_{k\in\mathbb{N}}\mathcal{B}_k=\mathcal{P}^{-}\vee T^{i}(\mathcal{P}^{\mathbb{D}'(M,\alpha)})$. Applying Lemma 2.9, we obtain

$$H(\mathcal{P}|\mathcal{P}^{-}\vee T^{i}(\mathcal{P}^{\mathbb{D}'(M,\alpha)}))=\lim_{k\to\infty}\downarrow H(\mathcal{P}|\mathcal{P}^{-}\vee T^{i}(\mathcal{P}^{[0,\frac{m_kM}{2})\cap\mathbb{D}'(M,\alpha)}))\geq h-\alpha.$$

For every $j\in\mathbb{D}(M,\alpha)$ and $i\in\{0,1,...,\lceil\alpha M\rceil-1\}$, the sum $i+j$ lies inside $\mathbb{D}'(M,\alpha)$. This implies, that $T^{-j}(\mathcal{P})=T^{i}(T^{-i-j}(\mathcal{P}))\subseteq T^{i}(\mathcal{P}^{\mathbb{D}'(M,\alpha)})$. As the σ-algebra $\mathcal{P}^{\mathbb{D}(M,\alpha)}$ is the smallest σ-algebra containing all partitions $T^{-j}(\mathcal{P})$, it lies within $T^{i}(\mathcal{P}^{\mathbb{D}'(M,\alpha)})$, and we have

$$H(\mathcal{P}|\mathcal{P}^{-}\vee\mathcal{P}^{\mathbb{D}(M,\alpha)})\geq H(\mathcal{P}|\mathcal{P}^{-}\vee T^{i}(\mathcal{P}^{\mathbb{D}'(M,\alpha)}))\geq h-\alpha.$$

For the case $r>1$, we refer to [2, Lemma 2]. □

Next, we recall that the k-th return time to a measurable set B with $\mu(B)>0$ is given by

$$\varphi_B^{(k)}(x)=\min\Big\{j>0:\sum_{i=1}^{j}1_B(T^{i}(x))=k\Big\}.$$

For such a set B, and $k \in \mathbb{N}$, let

$$\begin{aligned} \mathcal{Q}_B^{(k)} &= \mathcal{Q}^{(k)} = \{\mathcal{Q}_j^{(k)}\} \\ \mathcal{Q}_{B,j}^{(k)} &= \mathcal{Q}_j^{(k)} = \{x \in B : \varphi_B^{(k)} = j\} \end{aligned} \tag{3.23}$$

denote the countable partition of B given by the k-th return times to B.

The next lemma is the most difficult and elaborate argument in the proof of Theorem 3.1:

Lemma 3.9. *[2, Lemma 3] For every $\beta > 0$, $r \in \mathbb{N}$ and $K \in \mathbb{N}$, there exists an n_0, such that for every $n \geq n_0$, for $B \in \mathcal{P}^{-n}$ with μ-tolerance β, and for every $k \in \{1, ..., K\}$, the partition $\mathcal{P}^r$ is, with respect to μ_B, β-independent jointly of the past $\mathcal{P}^-$ and the partition $\mathcal{Q}^{(k)}$ (written as $\mathcal{P}^r \perp^\beta \mathcal{P}^- \vee \mathcal{Q}^{(k)}$). That is, for every $\mathcal{P}^- \vee \mathcal{Q}^{(k)}$-measurable partition $\mathcal{R}$:*

$$\sum_{A \in \mathcal{P}^r, C \in \mathcal{R}} |\mu_B(A \cap C) - \mu_B(A)\mu_B(C)| \leq \beta.$$

Proof. First, applying Lemma 2.14, we choose $\gamma > 0$ small enough such that

$$H(\mathcal{P}^r|\mathcal{Q}) \geq H(\mathcal{P}^r) - \gamma \implies \mathcal{P}^r \perp^{\frac{\beta}{2}} \mathcal{Q}$$

holds for any partition $\mathcal{Q}$. Let α be small enough to satisfy

$$0 < \frac{2\alpha}{h - \alpha} < 1, \;\; 16K\sqrt{\alpha} < 1, \;\; \sqrt{2\alpha} < \frac{\beta}{2}, \;\; \sqrt{2\alpha} < \gamma, \;\; K\sqrt[4]{\alpha} < \frac{\beta}{8}.$$

Now, we set n_0 large enough, such that the following hold:

(i) $H(\mathcal{P}^r|\mathcal{P}^{-n}) < rh + \alpha$ for all $n \geq n_0$.

(ii) $\mu_B\{2^{n(h-\alpha)} \leq \varphi_B^{(k)} \leq 2^{n(h+\alpha)}\} > 1 - \alpha$ for all $n \geq n_0$, for all $k \in \{1, ..., K\}$ and for $B \in \mathcal{P}^{-n}$ with μ-tolerance α.

(iii) The assertion of Lemma 3.8 holds for r, α, and $M_0 = \lfloor 2^{n_0(h-\alpha)} \rfloor$.

(iv) $(M + 1)^{1+\frac{2\alpha}{h-\alpha}} < \alpha M^2$ and $\frac{\log(M+1)}{M(h-\alpha)} < \alpha$ for all $M \geq M_0 = \lfloor 2^{n_0(h-\alpha)} \rfloor$.

Note that assumption (i) holds as $H(\mathcal{P}^r|\mathcal{P}^{-n})$ decreases to rh when $n \to \infty$ (by power rule). To see that condition (ii) on n_0 is possible, we observe that

$$\begin{aligned} \mu\{x \in X : 2^{n(h-\alpha)} \leq \varphi_{A_x^{-n}}^{(k)} \leq 2^{n(h+\alpha)}\} &= \sum_{B \in \mathcal{P}^{-n}} \mu(B)\mu_B\{2^{n(h-\alpha)} \leq \varphi_B^{(k)} \leq 2^{n(h+\alpha)}\} \\ &= \int_{\mathcal{P}^{-n}} \mu_B\{2^{n(h-\alpha)} \leq \varphi_B^{(k)} \leq 2^{n(h+\alpha)}\} d\mu(B). \end{aligned}$$

Then, by Ornstein-Weiss Theorem in reverse time (in form of estimate (2.11)), we can choose n_0 large enough such that

$$\int_{\mathcal{P}^{-n}} \mu_B\{2^{n(h-\alpha)} \leq \varphi_B^{(k)} \leq 2^{n(h+\alpha)}\} d\mu(B) > (1-\alpha)^2 = (1-\alpha)(1-\alpha) \tag{3.24}$$

holds for all $k \in \{1, ..., K\}$ and $n \geq n_0$. By rectangle rule, this implies

$$\begin{aligned} &\mu\{B \in \mathcal{P}^{-n} : \mu_B\{\varphi_B^{(k)} \notin [2^{n(h-\alpha)}, 2^{n(h+\alpha)}]\} \geq \alpha\} < \alpha \\ \implies &\mu\{B \in \mathcal{P}^{-n} : \mu_B\{2^{n(h-\alpha)} \leq \varphi_B^{(k)} \leq 2^{n(h+\alpha)}\} > 1-\alpha\} > 1-\alpha. \end{aligned}$$

Now, similarly to $M_0 = \lfloor 2^{n_0(h-\alpha)} \rfloor$, we set $M_n = \lfloor 2^{n(h-\alpha)} \rfloor$ for all $n \geq n_0$. By (ii), for arbitrary $n \geq n_0$ and for $B \in \mathcal{P}^{-n}$ chosen with μ-tolerance α, for most points $x \in B$ (up to μ_B-probability α) the first K returns to B lie within the interval $[2^{n(h-\alpha)}, 2^{n(h+\alpha)}]$. By (iv) and because $M_n + 1 > 2^{n(h-\alpha)}$, we have

$$\alpha M_n^2 > (M_n+1)^{1+\frac{2\alpha}{h-\alpha}} = (M_n+1)^{\frac{h+\alpha}{h-\alpha}} = 2^{\frac{h+\alpha}{h-\alpha}\log(M_n+1)} > 2^{\frac{h+\alpha}{h-\alpha}n(h-\alpha)} = 2^{n(h+\alpha)}. \tag{3.25}$$

Therefore, for every $n \geq n_0$, it holds that

$$[2^{n(h-\alpha)}, 2^{n(h+\alpha)}] \subset [M_n, \alpha M_n^2]. \tag{3.26}$$

We now fix $n \geq n_0$. We intend to find an M $\in \{M_n, ..., 2M_n\}$, such that, for nearly every $B \in \mathcal{P}^{-n}$, the first K returns to B happen most likely inside the set $\mathbb{D}(M, \alpha)$. Let $\alpha' = \alpha + \frac{n}{M_n}$. We will first show, that for carefully chosen M (and for most B), the first K return times of nearly every $x \in B$ are contained in $\mathbb{D}(M, \alpha')$. As it holds that $n < \frac{\log(M_n+1)}{h-\alpha}$, we have, again using (iv),

$$\alpha' = \alpha + \frac{n}{M_n} < \alpha + \frac{\log(M_n+1)}{M_n(h-\alpha)} < 2\alpha.$$

We fix $k \in \{1, ..., K\}$, and consider the product space

$$(\mathcal{P}^{-n} \times \mathbb{N}) \times [M_n, 2M_n]$$

equipped with the product σ-algebra $\mathcal{A}_1 \otimes \mathcal{A}_2$ and unique product probability measure $\mu_1 \otimes \mu_2$. The first factor of the product space is the probability space $(\mathcal{P}^{-n} \times \mathbb{N}, \mathcal{A}_1, \mu_1)$, equipped with the σ-algebra $\mathcal{A}_1$ of all subsets of $\mathcal{P}^{-n} \times \mathbb{N}$. For a product set $B' \times N \in \mathcal{A}_1$, the probability μ_1 is given by

$$\mu_1(B', N) = \sum_{B \in B'} \mu(B)\mu_B\{\varphi_B^{(k)} \in N\}.$$

The probability measure μ_2 on the space $([M_n, 2M_n], \mathcal{A}_2)$ is given by the uniform distribution on the integers in $[M_n, 2M_n]$, where $\mathcal{A}_2$ is the Borel σ-algebra on the interval

$[M_n, 2M_n]$. This means, that for a set $M \in \mathcal{A}_2$, we have $\mu_2(M) = \frac{\#\{M \cap \mathbb{Z}\}}{M_n+1}$.

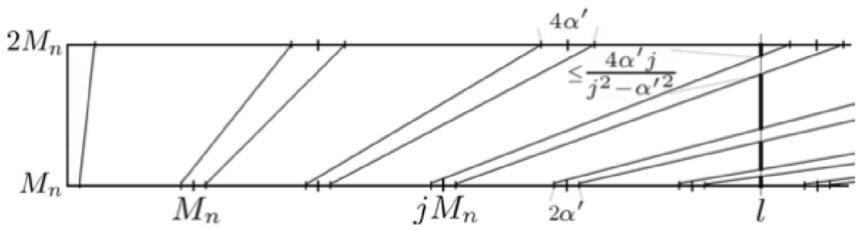


Figure 3.4.: The complement of $\mathbb{D}$ is given by thin skew strips j, starting around jM_n. For fixed $l \in [M_n, \alpha M_n^2]$, the uniform measure on the integers of the union of all those strips does not exceed 14α.

Let now $\mathbb{D}$ denote the set in the product space, whose $\mathbb{N}$-section, for fixed M and B equals $\mathbb{D}(M, \alpha')$, so we have

$$\mathbb{D} = \{(B, \mathbb{D}(M, \alpha'), M) : B \in \mathcal{P}^{-n}, M \in [M_n, 2M_n]\}$$

Our next goal is to prove that

$$\mu_1 \otimes \mu_2(\mathbb{D}) > 1 - 16\alpha$$

The complement of $\mathbb{D}$ is given by the thin skew strips shown in Figure 3.4. Note that, for each B, the union of the strips consists of all $M \in [M_n, 2M_n]$ and $l \in \mathbb{N}$ where l is not inside $\mathbb{D}(M, \alpha')$. We first take a look at the $[M_n, 2M_n]$ -section of $\mathbb{D}$, which, for a fixed pair (B, l), is given by

$$\mathbb{D}_{B,l} = \{M \in [M_n, 2M_n] : (B, l, M) \in \mathbb{D}\} = \{M \in [M_n, 2M_n] : l \in \mathbb{D}(M, \alpha')\} \subseteq [M_n, 2M_n].$$

For fixed l (and B), the $[M_n, 2M_n]$-section of our product space is given by a vertical line in Figure 3.4. The set $\mathbb{D}_{B,l}$ is then given by the vertical line, excluding the vertical section of each strip j which intersects l.
We now want to show, that for any $l \in [M_n, \alpha M_n^2] \cap \mathbb{Z}$ and any $B \in \mathcal{P}^{-n}$, it holds that

$$\mu_2(\mathbb{D}_{B,l}) > 1 - 14\alpha.$$

To prove this claim, we make two observations:
The first one is that not too many strips j intersect each vertical line l. The j-th strip starts at jM_n and intersects all $l \in [jM_n - \alpha', 2jM_n + 2\alpha'] \cap \mathbb{Z}$. If the j-th strip intersects some l, the $2j + 1$-th strip cannot contain l, as it intersects only integers larger than or equal to $(2j + 1)M_n - \alpha' > 2jM_n + \alpha' \geq l$. This implies, that each l is intersected by strips $j, j + 1, j + 2, ...$ up to at most $2j$ for some $j \geq 1$. Note that the 0-th strip cannot

intersect l as $l \geq M_n$.
The second observation is that the measure of each strip intersecting l is not too large. We first consider the Lebesgue measure of $\mathbb{D}_{B,l}$. The maximal Lebesgue measure of the $j-th$ strip on $[M_n, 2M_n]$ (which equals the largest vertical line that can be drawn inside each strip in Figure 3.4) is given, when l (which we allow here to be non-integer) equals $2jM_n - 2M_n\alpha'$. Then, l satisfies $l = jm_j + m_j\alpha'$ for some $m_j \in [M_n, 2M_n]$, and the maximal length of the j-th strip is given by the interval $[m_j, 2M_n]$. With the above boundary conditions, we have $m_j = 2M_n\frac{j-\alpha'}{j+\alpha'}$ and, for $\mathcal{S}_j = [m_j, 2M_n]$, calculation leads to

$$\lambda_{[M_n,2M_n]}(\mathcal{S}_j) = \lambda_{[M_n,2M_n]}\Big[2M_n\frac{j-\alpha'}{j+\alpha'}, 2M_n\Big] = 2M_n\Big(1 - \frac{j-\alpha'}{j+\alpha'}\Big) = 2M_n\frac{2\alpha'}{j+\alpha'}.$$

This implies, for the normalised Lebesgue measure λ'

$$\lambda'_{[M_n,2M_n]}(\mathcal{S}_j) = \frac{1}{M_n}\lambda_{[M_n,2M_n]}(\mathcal{S}_j) = \frac{4\alpha'}{j+\alpha'} < \frac{4\alpha'}{j} < \frac{8\alpha}{j}.$$

Now, each vertical line is intersected by skew strips with indices $j, j+1, j+2, \ldots$ up to at most $2j$, and the vertical measure of each strip is less than $\frac{8\alpha}{j}$. For $j \geq 2$, this gives us

$$\begin{aligned}\lambda'_{[M_n,2M_n]}(\mathbb{D}_{B,l}) &\geq 1 - \bigcup_{i=j}^{2j}\lambda'_{[M_n,2M_n]}(\mathcal{S}_i) > 1 - 8\alpha\Big(\frac{1}{j} + \frac{1}{j+1} + \ldots + \frac{1}{2j}\Big)\\ &\geq 1 - 8\alpha\Big(\frac{j+1}{j}\Big) \geq 1 - 12\alpha.\end{aligned}$$

For $j = 1$, the inequality holds as well, as $8\alpha(1 + 1/2) = 12\alpha$.

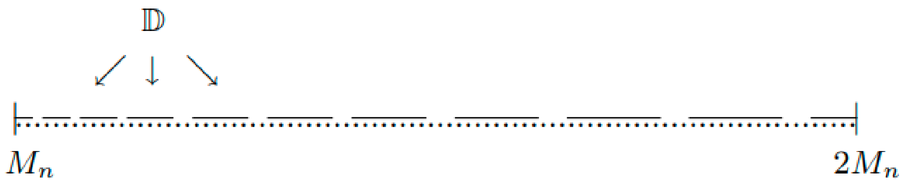


Figure 3.5.: For fixed $l \in \mathbb{N}$ with $M_n \leq l \leq \alpha M_n^2$, the uniform measure on $[M_n, 2M_n]$ (here shown on the horizontal axis) does differ from the normalized Lebesgue measure by not more that 2α, as there is a only bounded amount of intersections (depending on M_n) between strips j and sections belonging to $\mathbb{D}$.

It remains to be shown, that the measure of $\mathbb{D}_{B,l}$ is still large when we consider the uniform measure on the integers μ_2. As l is not larger than αM_n^2, not more than $\lceil \alpha M_n \rceil$ strips are small enough to possibly intersect l. Additionally, up to half of these strips (all strips j where $2jM_n + 2\alpha' < l$) are too small to intersect l. Therefore, it clearly holds that not more than $\lfloor \alpha M_n \rfloor$ strips intersect l. These strips alternate with not more than $\lfloor \alpha M_n \rfloor$ intervals which belong to $\mathbb{D}_{B,l}$ (see Figure 3.5). A fixed interval I of $\mathbb{D}_{B,l}$ with Lebesgue measure $\lambda(I)$ contains at least $\lfloor \lambda(I) \rfloor$ integers, and therefore satisfies $\mu_2(I) \geq \frac{\lfloor \lambda(I) \rfloor}{M_n+1}$. This implies

$$\mu_2(I) + \frac{2}{M_n} \geq \mu_2(I) + \frac{2}{M_n+1} \geq \frac{\lfloor \lambda(I) \rfloor + 2}{M_n+1} \geq \frac{\lfloor \lambda(I) \rfloor + 1}{M_n} \geq \frac{\lambda(I)}{M_n} = \lambda'_{[M_n,2M_n]}(I).$$

As this inequality holds for all intervals, and because $\sum_{j=1}^{k} \lambda'_{[M_n,2M_n]}(I_j) = \lambda'_{[M_n,2M_n]}(\mathbb{D}_{B,l})$ holds for $k \leq \lfloor \alpha M_n \rfloor$ being the number of intervals which belong to $\mathbb{D}_{B,l}$, we deduce

$$\mu_2(\mathbb{D}_{B,l}) = \sum_{j=1}^{k} \mu_2(I_j) \geq \sum_{j=1}^{k} \Big(\lambda'_{[M_n,2M_n]}(I_j) - \frac{2}{M_n} \Big) \geq \lambda'_{[M_n,2M_n]}(\mathbb{D}_{B,l}) - \lfloor \alpha M_n \rfloor \frac{2}{M_n} > 1 - 14\alpha.$$

Now, using this and the observations made in (3.24) and (3.25), our set $\mathbb{D}$ satisfies

$$\begin{aligned}
\mu_1 \otimes \mu_2(\mathbb{D}) &= \int_{\mathcal{P}^{-n} \times \mathbb{N}} \mu_2(\mathbb{D}_{B,l}) d\mu_1 \geq \int_{\mathcal{P}^{-n} \times [M_n, \alpha M_n^2] \cap \mathbb{Z}} \mu_2(\mathbb{D}_{B,l}) d\mu_1 \\
&> (1 - 14\alpha) \int_{\mathcal{P}^{-n} \times [M_n, \alpha M_n^2] \cap \mathbb{Z}} 1 d\mu_1 \\
&= (1 - 14\alpha) \int_{\mathcal{P}^{-n}} \mu_B \{ M_n \leq \varphi_B^{(k)} \leq \alpha M_n^2 \} d\mu(B) \\
&\geq (1 - 14\alpha) \int_{\mathcal{P}^{-n}} \mu_B \{ 2^{n(h-\alpha)} \leq \varphi_B^{(k)} \leq 2^{n(h+\alpha)} \} d\mu(B) \\
&> (1 - 14\alpha)(1 - \alpha)(1 - \alpha) > 1 - 16\alpha.
\end{aligned}$$

Considering the M-section of $\mathbb{D}$, which, for fixed $M \in [M_n, 2M_n]$, is given by $\mathbb{D}_M = \{(B, \mathbb{D}(M, \alpha'), M) : B \in \mathcal{P}^{-n}\} \subseteq \mathcal{P}^{-n} \times \mathbb{N}$, it now holds that

$$(\mu_1 \otimes \mu_2)(\mathbb{D}^{\mathrm{C}}) = \int_{[M_n, 2M_n]} \mu_1({\mathbb{D}_{\mathbb{M}}}^{\mathrm{C}}) d\mu_2 < 16\alpha = (16\sqrt{\alpha})(\sqrt{\alpha}).$$

With use of rectangle rule, it follows that for every M of a set of probability μ_2 of at least $1 - 16\sqrt{\alpha}$, the measure μ_1 of $\mathbb{D}_M$ is larger than $1 - \sqrt{\alpha}$. Then, for every such M, we have

$$\mu_1(\mathbb{D}_M) = \int_{\mathcal{P}^{-n}} \mu_B \{ \varphi_B^{(k)} \in \mathbb{D}(M, \alpha') \} d\mu > 1 - \sqrt{\alpha} > (1 - \sqrt[4]{\alpha})(1 - \sqrt[4]{\alpha}),$$

and, again by rectangle rule, we deduce

$$\mu\{ B \in \mathcal{P}^{-n} : \mu_B \{ \varphi_B^{(k)} \in \mathbb{D}(M, \alpha') \} \geq 1 - \sqrt[4]{\alpha} \} \geq 1 - \sqrt[4]{\alpha}.$$

In written words, this means: For every M of a set with probability μ_2 of at least $1 - 16\sqrt{\alpha}$, with μ-tolerance $\sqrt[4]{\alpha}$ for $B \in \mathcal{P}^{-n}$, the probability μ_B that the k-th return

time to B, $\varphi_B^{(k)}$, falls inside $\mathbb{D}(M,\alpha')$ is not smaller than $1-\sqrt[4]{\alpha}$. This implies, that every M of a set with probability μ_2 being not smaller than $1-16K\sqrt{\alpha}$ satisfies the above for all $k \in \{1,...,K\}$. As $16K\sqrt{\alpha} < 1$, there is at least one M, for which the above holds for all k. This is our final choice of M and remains fixed from now on.
For this M, and for cylinders $B \in \mathcal{P}^{-n}$ chosen with μ-tolerance $K\sqrt[4]{\alpha}$, each of the first K return times, $\varphi_B^{(k)}$, happens inside $\mathbb{D}(M,\alpha')$ with probability μ_B not smaller than $1-\sqrt[4]{\alpha}$.
We now want to show, that if for some $B \in \mathcal{P}^{-n}$ and $x \in B$, it holds that $\varphi_B^{(k)}(x)$ is inside $\mathbb{D}(M,\alpha')$, then the k-th return of the whole block B is contained in $\mathbb{D}(M,\alpha)$. Note that $\varphi_B^{(k)}(x)$ lies within $\mathbb{D}(M,\alpha')$, exactly if $T^{-\varphi_B^{(k)}(x)}(\mathcal{P})$ is contained in the σ-algebra $\mathcal{P}^{\mathbb{D}(M,\alpha')}$. Similarly, we say that the k-th return of a block B is inside $\mathbb{D}(M,\alpha)$, whenever

$$T^{-\varphi_B^{(k)}(x)}(B) \in \mathcal{P}^{\mathbb{D}(M,\alpha)}. \tag{3.27}$$

Note that, as B is an element of the partition $\mathcal{P}^{-n}$, it holds that

$$T^{-\varphi_B^{(k)}(x)}(B) \in T^{-\varphi_B^{(k)}(x)}(\mathcal{P}^{-n}) = \bigvee_{i=1}^{n} T^{-\varphi_B^{(k)}(x)+i}(\mathcal{P}).$$

Therefore, the k-th return of B is inside $\mathbb{D}(M,\alpha)$, if and only if the integers $\varphi_B^{(k)}(x)-1, \varphi_B^{(k)}(x)-2, ..., \varphi_B^{(k)}(x)-n$ are elements of $\mathbb{D}(M,\alpha)$. By choice of α', the set $\mathbb{D}(M,\alpha')$ is given by

$$\begin{aligned}\mathbb{D}(M,\alpha') &= \bigcup_{m\in\mathbb{Z}} [mM+\alpha'M,(m+1)M-\alpha'M)\cap\mathbb{Z}\\ &= \bigcup_{m\in\mathbb{Z}} [mM+\alpha M+n,(m+1)M-\alpha M-n)\cap\mathbb{Z}.\end{aligned}$$

It is fully contained within $\mathbb{D}(M,\alpha)$, and $\mathbb{D}(M,\alpha)$ additionally consists of n integers on the left and on the right side of each periodic subset of $\mathbb{D}(M,\alpha')$. Therefore, if for some $B \in \mathcal{P}^{-n}$ and $x \in B$, it holds that $\varphi_B^{(k)}(x) \in \mathbb{D}(M,\alpha')$, the integers $\varphi_B^{(k)}(x)-1, \varphi_B^{(k)}(x)-2, ...$ up to $\varphi_B^{(k)}(x)-n$ lie inside $\mathbb{D}(M,\alpha)$, which implies that the k-th return of B is contained in $\mathbb{D}(M,\alpha)$. Therefore, from now on, we will consider the k-th return of a block B inside $\mathbb{D}(M,\alpha)$ instead of the k-th return time being inside $\mathbb{D}(M,\alpha')$.
Note that, for every $B \in \mathcal{P}^{-n}$, (3.27) still holds, when we substitute $\varphi_B^{(k)}(x)$ by any $i \in \mathbb{D}(M,\alpha')$, which implies that $T^{-i}(\mathcal{P}^{-n}) \in \mathcal{P}^{\mathbb{D}(M,\alpha)}$ for every $i \in \mathbb{D}(M,\alpha')$. As the σ-algebra $(\mathcal{P}^{-n})^{\mathbb{D}(M,\alpha')}$ is the smallest σ-algebra containing all the partitions $T^{-i}(\mathcal{P}^{-n})$, we conclude

$$(\mathcal{P}^{-n})^{\mathbb{D}(M,\alpha')} \subseteq \mathcal{P}^{\mathbb{D}(M,\alpha)}. \tag{3.28}$$

We continue our proof of the desired β-independence of $\mathcal{P}^r$ and $\mathcal{P}^- \vee \mathcal{Q}^{(k)}$ (for $k \in \{1,...,K\}$), by showing that, for most B, $\mathcal{P}^r$ is $\frac{\beta}{2}$-independent of the join $\mathcal{P}^- \vee \tilde{\mathcal{Q}}^{(k)}$. Before we define the partition $\tilde{\mathcal{Q}}^{(k)}$, we consider the map $\tilde{\varphi}_B^{(k)}$, which is defined on B as the k-th return time to B fully visible inside $\mathbb{D}(M, \alpha')$, given by

$$\tilde{\varphi}_B^{(k)}(x) = \min\Big\{j \in \mathbb{D}(M, \alpha') : \sum_{\substack{i \in \mathbb{D}(M,\alpha') \cap \mathbb{N} \\ i \leq j}} \mathbb{1}_B(T^i x) = k\Big\}.$$

Now, the countable partition $\tilde{\mathcal{Q}}_B^{(k)}$ is defined as

$$\tilde{\mathcal{Q}}_B^{(k)} = \tilde{\mathcal{Q}}^{(k)} = \{\tilde{\mathcal{Q}}_j^{(k)}\},$$

where

$$\tilde{\mathcal{Q}}_j^{(k)} = \{x \in B : \tilde{\varphi}_B^{(k)}(x) = j\} = \Big\{x \in B : \sum_{\substack{i \in \mathbb{D}(M,\alpha') \cap \mathbb{N} \\ i \leq j}} \mathbb{1}_B(T^i x) = k\Big\}.$$

Each set $\tilde{\mathcal{Q}}_j^{(k)}$ is defined by a finite sum of functions $\mathbb{1}_B \circ T^i$ (where $i \in \mathbb{D}(M, \alpha')$), and for each of these functions it holds that

$$\{\mathbb{1}_B \circ T^i = 1\} = T^{-i}(B) \quad \textit{and} \quad \{\mathbb{1}_B \circ T^i = 0\} = T^{-i}(B^{\mathrm{C}}). \tag{3.29}$$

This implies, that every function $\mathbb{1}_B \circ T^i$ is $(\mathcal{P}^{-n})^{\mathbb{D}(M,\alpha')}$ measurable, and therefore, for every k and every $j \in \mathbb{N}$, also $\tilde{\mathcal{Q}}_j^{(k)}$ (defined by a finite sum of measureable functions) is $(\mathcal{P}^{-n})^{\mathbb{D}(M,\alpha')}$ measurable. Then, with use of (3.28), we obtain that $\tilde{\mathcal{Q}}^{(k)}$ is fully covered by $\mathcal{P}^{\mathbb{D}(M,\alpha)}$ (restricted on B), such that

$$\tilde{\mathcal{Q}}_j^{(k)} \subset \mathcal{P}^{\mathbb{D}(M,\alpha)}. \tag{3.30}$$

In order to use this fact, we will now consider some entropy estimates concerning $\mathcal{P}^{\mathbb{D}(M,\alpha)}$. Using assumptions (i) and (iii) at the beginning of the proof, it holds that

$$\sum_{B \in \mathcal{P}^{-n}} \mu(B) H_B(\mathcal{P}^r | \mathcal{P}^- \vee \mathcal{P}^{\mathbb{D}(M,\alpha)}) = H(\mathcal{P}^r | \mathcal{P}^{-n} \vee \mathcal{P}^- \vee \mathcal{P}^{\mathbb{D}(M,\alpha)}) =$$

$$H(\mathcal{P}^r | \mathcal{P}^- \vee \mathcal{P}^{\mathbb{D}(M,\alpha)}) \geq rh - \alpha > H(\mathcal{P}^r | \mathcal{P}^{-n}) - 2\alpha =$$

$$\sum_{B \in \mathcal{P}^{-n}} \mu(B) H_B(\mathcal{P}^r) - 2\alpha$$

Written as integrals, the above inequality can be stated as

$$\int_{\mathcal{P}^{-n}} H_B(\mathcal{P}^r) - H_B(\mathcal{P}^r | \mathcal{P}^- \vee \mathcal{P}^{\mathbb{D}(M,\alpha)}) d\mu(B) < 2\alpha = \sqrt{2\alpha}\sqrt{2\alpha}$$

As for every B it holds that $H_B(\mathcal{P}^r) - H_B(\mathcal{P}^r | \mathcal{P}^- \vee \mathcal{P}^{\mathbb{D}(M,\alpha)})$ is nonnegative, we can again use rectangle rule to deduce: For $B \in \mathcal{P}^{-n}$ with μ-tolerance $\sqrt{2\alpha}$, we have

$$H_B(\mathcal{P}^r | \mathcal{P}^- \vee \mathcal{P}^{\mathbb{D}(M,\alpha)}) \geq H_B(\mathcal{P}^r) - \sqrt{2\alpha}.$$

Recalling (3.30), this gives us for every $k \in \{1, ..., K\}$, and for $B \in \mathcal{P}^{-n}$ with μ-tolerance $\sqrt{2\alpha}$,

$$H_B(\mathcal{P}^r | \mathcal{P}^- \vee \tilde{\mathcal{Q}}^{(k)}) \geq H_B(\mathcal{P}^r | \mathcal{P}^- \vee \mathcal{P}^{\mathbb{D}(M,\alpha)}) \geq H_B(\mathcal{P}^r) - \sqrt{2\alpha} \geq H_B(\mathcal{P}^r) - \gamma.$$

By choice of γ at the beginning of the proof, this implies $\frac{\beta}{2}$-independence of $\mathcal{P}^r$ and the σ-algebra $\mathcal{P}^- \vee \tilde{\mathcal{Q}}^{(k)}$, i.e. for every countable $\mathcal{P}^- \vee \tilde{\mathcal{Q}}^{(k)}$-measurable partition $\tilde{\mathcal{R}}^{(k)}$ (and for every $k \in \{1, ..., K\}$) it holds that

$$\sum_{\substack{A \in \mathcal{P}^r \\ \tilde{C}^{(k)} \in \tilde{\mathcal{R}}^{(k)}}} |\mu_B(A \cap \tilde{C}^{(k)}) - \mu_B(A)\mu_B(\tilde{C}^{(k)})| \leq \frac{\beta}{2}.$$

In order to derive from this the β-independence of $\mathcal{P}^r$ and $\mathcal{P}^- \vee \mathcal{Q}^{(k)}$ (for $B \in \mathcal{P}^{-n}$ with μ-tolerance β and $k \in \{1, ..., K\}$), we need to show that the partitions $\mathcal{Q}^{(k)}$ and $\tilde{\mathcal{Q}}^{(k)}$ are close. Therefore, we now consider the $\frac{\beta}{2}$-independence of $\mathcal{P}^r$ and $\tilde{\mathcal{Q}}^{(k)}$. We aim to first show β-independence of $\mathcal{P}^r$ and $\mathcal{Q}^{(k)}$, and obtain from this the β-independence of $\mathcal{P}^r$ and $\mathcal{P}^- \vee \mathcal{Q}^{(k)}$. The corresponding $\frac{\beta}{2}$-independence of $\mathcal{P}^r$ and $\tilde{\mathcal{Q}}^{(k)}$ is given by

$$\sum_{\substack{A \in \mathcal{P}^r \\ \tilde{\mathcal{Q}}_j^{(k)} \in \tilde{\mathcal{Q}}^{(k)}}} |\mu_B(A \cap \tilde{\mathcal{Q}}_j^{(k)}) - \mu_B(A)\mu_B(\tilde{\mathcal{Q}}_j^{(k)})| \leq \frac{\beta}{2}. \tag{3.31}$$

Let us recall the fact that for our chosen M, and for cylinders $B \in \mathcal{P}^{-n}$ chosen with μ-tolerance $K\sqrt[4]{\alpha}$, each of the first K returns of B happen inside $\mathbb{D}(M, \alpha)$ with probability μ_B not smaller than $1 - \sqrt[4]{\alpha}$. Then, for such a B, the probability μ_B that all the first K returns of B satisfy this, is larger than or equal to $1 - K\sqrt[4]{\alpha}$. Equivalently, for most $B \in \mathcal{P}^{-n}$ (up to probability μ of $K\sqrt[4]{\alpha}$), there is a set $\tilde{B}$ with measure $\mu_B(\tilde{B}) \leq K\sqrt[4]{\alpha}$, outside of which it holds that $\varphi_B^{(k)} = \tilde{\varphi}_B^{(k)}$ for all $k \in \{1, ..., K\}$. Therefore, on the complement $\tilde{B}^{\mathrm{C}}$, the partitions given by the return times satisfy

$$\mathcal{Q}_j^{(k)} \cap \tilde{B}^{\mathrm{C}} = \tilde{\mathcal{Q}}_j^{(k)} \cap \tilde{B}^{\mathrm{C}} \tag{3.32}$$

for all $j \in \mathbb{N}$ and for all $k \in \{1, ..., K\}$.

Combining the above results, with μ-tolerance $\sqrt{2\alpha} + K\sqrt[4]{\alpha} < \beta$ for $B \in \mathcal{P}^{-n}$, and for all $k \in \{1, .., K\}$, we have the $\frac{\beta}{2}$ independence of $\mathcal{P}^r$ and $\tilde{\mathcal{Q}}^{(k)}$ (as well as the one of $\mathcal{P}^r$ and $\mathcal{P}^- \vee \tilde{\mathcal{Q}}^{(k)}$) stated in (3.31), and the equality in (3.32) holds, where the measure μ_B of $\tilde{B}$ does not exceed $K\sqrt[4]{\alpha} \leq \frac{\beta}{8}$. Now, for any B for which the above holds, the relative complement of $\tilde{\mathcal{Q}}^{(k)}$ in $\mathcal{Q}^{(k)}$ satisfies, for $j \in \mathbb{N}$ and by using (3.32)

$$\begin{aligned}
\sum_{\substack{\mathcal{Q}_j^{(k)} \in \mathcal{Q}^{(k)} \\ \tilde{\mathcal{Q}}_j^{(k)} \in \tilde{\mathcal{Q}}^{(k)}}} \mu_B(\mathcal{Q}_j^{(k)} \setminus \tilde{\mathcal{Q}}_j^{(k)}) &= \sum_{\substack{\mathcal{Q}_j^{(k)} \in \mathcal{Q}^{(k)} \\ \tilde{\mathcal{Q}}_j^{(k)} \in \tilde{\mathcal{Q}}^{(k)}}} \mu_B(\tilde{B}^{\mathrm{C}} \cap (\mathcal{Q}_j^{(k)} \setminus \tilde{\mathcal{Q}}_j^{(k)})) \\
&+ \sum_{\substack{\mathcal{Q}_j^{(k)} \in \mathcal{Q}^{(k)} \\ \tilde{\mathcal{Q}}_j^{(k)} \in \tilde{\mathcal{Q}}^{(k)}}} \mu_B(\tilde{B} \cap (\mathcal{Q}_j^{(k)} \setminus \tilde{\mathcal{Q}}_j^{(k)})) \\
&= \sum_{\substack{\mathcal{Q}_j^{(k)} \in \mathcal{Q}^{(k)} \\ \tilde{\mathcal{Q}}_j^{(k)} \in \tilde{\mathcal{Q}}^{(k)}}} \mu_B\{\tilde{B} \cap (\mathcal{Q}_j^{(k)} \setminus \tilde{\mathcal{Q}}_j^{(k)})\} \\
&= \mu_B(\tilde{B}) \sum_{\substack{\mathcal{Q}_j^{(k)} \in \mathcal{Q}^{(k)} \\ \tilde{\mathcal{Q}}_j^{(k)} \in \tilde{\mathcal{Q}}^{(k)}}} \mu_{\tilde{B}}(\mathcal{Q}_j^{(k)} \setminus \tilde{\mathcal{Q}}_j^{(k)}) \leq \frac{\beta}{8}. \qquad (3.33)
\end{aligned}$$

Switching the roles of $\mathcal{Q}^{(k)}$ and $\tilde{\mathcal{Q}}^{(k)}$, the above inequality holds as well, and we obtain

$$\begin{aligned}
&\sum_{\substack{\mathcal{Q}_j^{(k)} \in \mathcal{Q}^{(k)} \\ \tilde{\mathcal{Q}}_j^{(k)} \in \tilde{\mathcal{Q}}^{(k)}}} \max\{\mu_B(\mathcal{Q}_j^{(k)} \setminus \tilde{\mathcal{Q}}_j^{(k)}), \mu_B(\tilde{\mathcal{Q}}_j^{(k)} \setminus \mathcal{Q}_j^{(k)})\} \\
&\leq \sum_{\substack{\mathcal{Q}_j^{(k)} \in \mathcal{Q}^{(k)} \\ \tilde{\mathcal{Q}}_j^{(k)} \in \tilde{\mathcal{Q}}^{(k)}}} \mu_B(\mathcal{Q}_j^{(k)} \setminus \tilde{\mathcal{Q}}_j^{(k)}) + \mu_B(\tilde{\mathcal{Q}}_j^{(k)} \setminus \mathcal{Q}_j^{(k)}) \\
&\leq \frac{\beta}{8} + \frac{\beta}{8} = \frac{\beta}{4}. \qquad (3.34)
\end{aligned}$$

Taking a closer look at the structure of our partitions, we can see that either $\mathcal{Q}_j^{(k)} \subset \tilde{\mathcal{Q}}_j^{(k)}$ (when $j \in \mathbb{D}(M, \alpha')$) holds, or $\tilde{\mathcal{Q}}_j^{(k)} \subset \mathcal{Q}_j^{(k)}$ (when $j \notin \mathbb{D}(M, \alpha')$). Therefore, for all $j \in \mathbb{N}$, we deduce the the equality

$$|\mu_B(\mathcal{Q}_j^{(k)}) - \mu_B(\tilde{\mathcal{Q}}_j^{(k)})| = \max\{\mu_B(\mathcal{Q}_j^{(k)} \setminus \tilde{\mathcal{Q}}_j^{(k)}), \mu_B(\tilde{\mathcal{Q}}_j^{(k)} \setminus \mathcal{Q}_j^{(k)})\} \qquad (3.35)$$

With this we now can show, that

$$\sum_{\substack{A \in \mathcal{P}^r \\ \mathcal{Q}_j^{(k)} \in \mathcal{Q}^{(k)}}} |\mu_B(A \cap \mathcal{Q}_j^{(k)}) - \mu_B(A)\mu_B(\mathcal{Q}_j^{(k)})| \leq \beta \qquad (3.36)$$

holds for $B \in \mathcal{P}^{-n}$ with μ-tolerance β and for all $k \in \{1, ..., K\}$. Adding the terms $\mu_B(A \cap \tilde{\mathcal{Q}}_j^{(k)})$ and $\mu_B(A)\mu_B(\tilde{\mathcal{Q}}_j^{(k)})$ inside the vertices, using triangle inequality as well as

(3.34) and (3.35), the left term in (3.36) satisfies

$$\sum_{\substack{A\in\mathcal{P}^r\\ \mathcal{Q}_j^{(k)}\in\mathcal{Q}^{(k)}}} |\mu_B(A\cap\mathcal{Q}_j^{(k)})-\mu_B(A)\mu_B(\mathcal{Q}_j^{(k)})| \leq \sum_{\substack{A\in\mathcal{P}^r\\ \mathcal{Q}_j^{(k)}\in\mathcal{Q}^{(k)}\\ \tilde{\mathcal{Q}}_j^{(k)}\in\tilde{\mathcal{Q}}^{(k)}}} |\mu_B(A\cap\mathcal{Q}_j^{(k)})-\mu_B(A\cap\tilde{\mathcal{Q}}_j^{(k)})|+$$

$$\sum_{\substack{A\in\mathcal{P}^r\\ \mathcal{Q}_j^{(k)}\in\mathcal{Q}^{(k)}\\ \tilde{\mathcal{Q}}_j^{(k)}\in\tilde{\mathcal{Q}}^{(k)}}} |\mu_B(A\cap\tilde{\mathcal{Q}}_j^{(k)})-\mu_B(A)\mu_B(\tilde{\mathcal{Q}}_j^{(k)})| + \sum_{\substack{A\in\mathcal{P}^r\\ \mathcal{Q}_j^{(k)}\in\mathcal{Q}^{(k)}\\ \tilde{\mathcal{Q}}_j^{(k)}\in\tilde{\mathcal{Q}}^{(k)}}} |\mu_B(A)\mu_B(\tilde{\mathcal{Q}}_j^{(k)})-\mu_B(A)\mu_B(\mathcal{Q}_j^{(k)})|$$

$$\begin{aligned}
&\leq \sum_{\substack{A\in\mathcal{P}^r\\ \mathcal{Q}_j^{(k)}\in\mathcal{Q}^{(k)}\\ \tilde{\mathcal{Q}}_j^{(k)}\in\tilde{\mathcal{Q}}^{(k)}}} \max\{\mu_B(A\cap(\mathcal{Q}_j^{(k)}\setminus\tilde{\mathcal{Q}}_j^{(k)})),\mu_B(A\cap(\tilde{\mathcal{Q}}_j^{(k)}\setminus\mathcal{Q}_j^{(k)}))\}\\
&+\frac{\beta}{2}+\sum_{\substack{A\in\mathcal{P}^r\\ \mathcal{Q}_j^{(k)}\in\mathcal{Q}^{(k)}\\ \tilde{\mathcal{Q}}_j^{(k)}\in\tilde{\mathcal{Q}}^{(k)}}} \mu_B(A)\max\{\mu_B(\mathcal{Q}_j^{(k)}\setminus\tilde{\mathcal{Q}}_j^{(k)}),\mu_B(\tilde{\mathcal{Q}}_j^{(k)}\setminus\mathcal{Q}_j^{(k)})\}\\
&\leq \frac{\beta}{4}+\frac{\beta}{2}+\frac{\beta}{4}=\beta. \qquad (3.37)
\end{aligned}$$

It remains to be shown that, for given $k\in\{1,...,K\}$, this still holds when we replace the partition $\mathcal{Q}^{(k)}$ by any (at most countable) partition $\mathcal{R}^{(k)}=\{C_j^{(k)}: j\in\mathbb{N}\}$ which is measurable with respect to the σ-algebra $\mathcal{P}^-\vee\mathcal{Q}^{(k)}$. As $\mathcal{P}^-$ is generated by the partitions $\{T^i(\mathcal{P}): i\geq 1\}$, we obtain, together with Lemma 2.1 and 2.7, that each element $C_j^{(k)}$ of $\mathcal{R}^{(k)}$ is (up to a set with arbitrary small measure) the union of elements of the partition $\mathcal{P}^{-m}$ (for some $m\in\mathbb{N}$) intersected with the union of elements of the partition $\mathcal{Q}^{(k)}$. When we consider the partition $\tilde{\mathcal{R}}^{(k)}=\{\tilde{C}_j^{(k)}: j\in\mathbb{N}\}$, where each element $\tilde{C}_j^{(k)}$ equals $C_j^{(k)}$ except for replacing the $\mathcal{Q}^{(k)}$ part by $\tilde{\mathcal{Q}}^{(k)}$, we see that $\tilde{\mathcal{R}}^{(k)}\in\mathcal{P}^-\vee\tilde{\mathcal{Q}}^{(k)}$, and replacing $\mathcal{Q}^{(k)}$ by $\mathcal{R}^{(k)}$ (and $\mathcal{Q}_j^{(k)}$ by $C_j^{(k)}$), as well as $\tilde{\mathcal{Q}}^{(k)}$ by $\tilde{\mathcal{R}}^{(k)}$, the results (3.32), (3.33), (3.34), (3.35) and (3.37) still hold, which implies (for $B\in\mathcal{P}^{-n}$ with μ-tolerance β and $k\in\{1,...,K\}$) β-independence of $\mathcal{P}^r$ and the join $\mathcal{P}^-\vee\mathcal{Q}^{(k)}$.

□

We now replace the past of the partition $\mathcal{P}$ by the past of the partition $\mathcal{P}^r$ in the conditioning σ-algebra of the result of Lemma 3.9, i.e. assuming that $\mathcal{P}^r\perp^\beta(\mathcal{P}^r)^-\vee\mathcal{Q}^{(k)}$ holds instead of $\mathcal{P}^r\perp^\beta\mathcal{P}^-\vee\mathcal{Q}^{(k)}$. Note that we could not verify this assumption, which is stronger than the result provided in Lemma 3.9 (as $\mathcal{P}^-\subseteq(\mathcal{P}^r)^-$). Nevertheless, we consider it to hold from now on as it leads to an elegant and comprehensible proof of Theorem 3.1.

Applying the (measure preserving) induced map T_B on both sides, the β-independence

$\mathcal{P}^r \perp^\beta (\mathcal{P}^r)^- \vee \mathcal{Q}^{(k)}$ implies that

$$T_B^{-i}(\mathcal{P}^r) \perp^\beta (T_B^{-i}(\mathcal{P}^r))^- \vee T_B^{-i}(\mathcal{Q}^{(k)}) \tag{3.38}$$

holds for all $i \in \mathbb{N}$ and for every $k \in \{1, ..., K\}$. The partition $T_B^{-i}(\mathcal{P}^r)$ consists of the cylinders of length r following the i-th return of B. Note that its past $(T_B^{-i}(\mathcal{P}^r))^-$ is not a fixed σ-algebra. It depends on the return time of each $x \in B$ and is formally given by

$$(T_B^{-i}(\mathcal{P}^r))^- = \bigvee_{j=1}^{\infty} T^{-\varphi_B^{(i)}(x)+j}(\mathcal{P}^r). \tag{3.39}$$

For $x \in B \cap A$, the return time to A on the induced process on B is given by

$$\varphi_A^{(B)}(x) = \min\{j > 0 : T_B^j(x) \in A\}.$$

In order to show the attracting postulate of the Law of Series, it remains to put the items together:

Proof of Theorem 3.1. We fix $\epsilon > 0$. By Lemma 3.6, for $e_p = (1-p)^{-\frac{1}{p}}$, the functions

$$g_p(t) = min\{1, \frac{1}{(1-p)\ln(e_p)}(1 - e_p^{-t}) + pt\}$$

converge uniformly to $1 - e^{-t}$ from above on $t \in [0, \infty)$, when $p \to 0^+$. Therefore, we can choose $\delta > 0$ small enough such that $g_\delta(t) \leq 1 - e^{-t} + \epsilon$ holds for all $t \geq 0$. Let δ also satisfy $(1-2\delta)(1-\delta) \geq 1-\epsilon$. We now fix an $r \in \mathbb{N}$ such that Lemma 3.7 is satisfied. This means, that $\mu_B(A) \leq \delta$ holds for every $A \in \mathcal{P}^r$, for every $n \geq 1$ and for $B \in \mathcal{P}^{-n}$ with μ-tolerance δ. The partition $\mathcal{P}^r$ has at most $(\#\mathcal{P})^r$ elements. Therefore, for any B, the joint μ_B-measure of all $A \in \mathcal{P}^r$ satisfying $\mu_B(A) < \delta(\#\mathcal{P})^{-r}$ is smaller than δ. Equivalently, for $A \in \mathcal{P}^r$ with μ_B-tolerance δ, it holds that $\mu_B(A) \geq \delta(\#\mathcal{P})^{-r}$. Let $\mathcal{A}_B$ be the subfamily of $\mathcal{P}^r$ (depending on B), where this inequality holds. We now set K large enough, such that for any $p \geq \delta(\#\mathcal{P})^{-r}$ the inequality

$$\sum_{k=K+1}^{\infty} p(1-p)^k < \frac{\delta}{2}$$

is satisfied. Let $\beta < \delta$ be so small such that for any $p \geq \delta(\#\mathcal{P})^{-r}$ it holds that

$$\frac{1}{p}(2K^2 + K)\beta < \frac{\delta}{2}.$$

Now, applying Lemma 3.9 for our chosen r, β and K, there is an n_0, such that for every $n \geq n_0$, for $B \in \mathcal{P}^{-n}$ with μ-tolerance β and for every $k \in \{1, ..., K\}$, the partition $\mathcal{P}^r$ is, with respect to μ_B, β-independent of its past $(\mathcal{P}^r)^-$ joined with the return time partition $\mathcal{Q}^{(k)}$. Thus, with μ-tolerance $\delta + \beta < 2\delta$ for $B \in \mathcal{P}^{-n}$, we have both, the above β-independence as well as the estimate $\mu_B(A) < \delta$ for all $A \in \mathcal{P}^r$. Let $\mathcal{B}_n$ be the

subfamily of $\mathcal{P}^{-n}$ containing all B for which these two conditions hold.
We now fix $n \geq n_0$, and consider the partition $\mathcal{P}^{[-n,r)}$. Each cylinder BA of this partition is an intersection $B \cap A$ of one cylinder $B \in \mathcal{P}^{-n}$ and one cylinder $A \in \mathcal{P}^r$.
Note that the measure of the cylinders $BA \in \mathcal{P}^{[-n,r)}$, where $B \in \mathcal{B}_n$ and $A \in \mathcal{A}_B$, is at least $1-\epsilon$, as

$$\mu\Big(\bigcup_{\substack{B\in\mathcal{B}_n\\ A\in\mathcal{A}_B}} BA\Big) = \sum_{\substack{B\in\mathcal{B}_n\\ A\in\mathcal{A}_B}} \mu(B\cap A) = \sum_{B\in\mathcal{B}_n} \mu(B)\mu_B(\mathcal{A}_B) \geq (1-2\delta)(1-\delta) \geq 1-\epsilon.$$

Next, we take a look at the distribution $\overline{G}_{BA}$ of the normalized return time to a cylinder BA, which satisfies $B \in \mathcal{B}_n$ and $A \in \mathcal{A}_B$. We recall the map $\varphi_A^{(B)}$, which is defined on BA, and denotes the return time to A on the induced process on B. Note that $\varphi_A^{(B)}$ and the return time to BA are related to each other, as

$$\varphi_{BA}(x) = \sum_{k=0}^{\varphi_A^{(B)}-1} \varphi_B \circ T_B^k(x) = \varphi_B^{\varphi_A^{(B)}(x)}(x).$$

In words, the return time to BA is exactly the k-th return time to B, whenever the induced return time to A equals k. From this we can see that

$$\{\varphi_{BA} = j\} = \{\varphi_B^{\varphi_A^{(B)}} = j\} = \bigcup_{k=1}^{\infty} \{\varphi_A^{(B)} = k, \varphi_B^{(k)} = j\}$$

holds for all positive integers j. Let $p = \mu_B(A)$, where p is, by the assumptions on our cylinder BA, not smaller than $\delta(\#\mathcal{P})^{-r}$. Then for the normalized return time $\overline{G}_{BA}$ it holds that

$$\begin{aligned}
\overline{G}_{BA}(t) &= \mu_{BA}\{\overline{\varphi}_{BA} \leq t\} = \mu_{BA}\Big\{\varphi_{BA} \leq \frac{t}{\mu(BA)}\Big\} = \sum_{j=1}^{\lfloor \frac{t}{p\mu(B)} \rfloor} \mu_{BA}\{\varphi_{BA} = j\} \\
&= \sum_{j=1}^{\lfloor \frac{t}{p\mu(B)} \rfloor} \sum_{k=1}^{\infty} \mu_{BA}\{\varphi_A^{(B)} = k, \varphi_B^{(k)} = j\} = \sum_{k=1}^{\infty} \mu_{BA}\Big\{\varphi_A^{(B)} = k, \varphi_B^{(k)} \leq \frac{t}{p\mu(B)}\Big\}.
\end{aligned}$$

The k-th term of this sum equals

$$\mu_{BA}\Big\{\varphi_A^{(B)} = k, \varphi_B^{(k)} \leq \frac{t}{p\mu(B)}\Big\} = \frac{1}{p}\mu_B\Big(\Big\{\varphi_A^{(B)} = k\Big\} \cap A \cap \Big\{\varphi_B^{(k)} \leq \frac{t}{p\mu(B)}\Big\}\Big) =$$

$$\frac{1}{p}\mu_B\Big(\{A_k = A\} \cap \{A_{k-1} \neq A\} \cap \cdots \cap \{A_1 \neq A\} \cap \{A_0 = A\} \cap \Big\{\varphi_B^{(k)} \leq \frac{t}{p\mu(B)}\Big\}\Big), \tag{3.40}$$

where $A_i \in T_B^{-i}(\mathcal{P}^r)$ denotes the cylinder of length r following the i-th copy of B. We want to show, that for all $k \in \{1, ..., K\}$, each term in (3.40) is β-independent of the intersection to its right. Note that, recalling the observations made in (3.38), for fixed $k \in \{1, ..., K\}$,

$$T_B^{-i}(\mathcal{P}^r) \perp^{\beta} (T_B^{-i}(\mathcal{P}^r))^- \vee T_B^{-i}(\mathcal{Q}^{(k-i)}) \tag{3.41}$$

holds for every $i \in \{0, ..., k\}$. Starting with $i = k$, we aim to prove, that the intersection right to the set A_i (which is a set of the partition $T_B^{-i}(\mathcal{P}^r)$), is measurable with respect to the past of A_i (represented by the σ-algebra $(T_B^{-i}(\mathcal{P}^r))^-$), joined with the partition $T_B^{-i}(\mathcal{Q}^{(k-i)})$.
For $j \in \{0, 1, ..., i-1\}$, the sets A_j are measurable with respect to the σ-algebra $(T_B^{-i}(\mathcal{P}^r))^-$. This follows from (3.39) and the fact that, for $j < i$, the j-th return to B happens before the i-th return to B. Note that we could not verify the sets A_j to be measurable with respect to the σ-algebra $(T_B^{-i}(\mathcal{P}))^-$, which is why we use an assumption (i.e. $\mathcal{P}^r \perp^\beta (\mathcal{P}^r)^- \vee \mathcal{Q}^{(k)}$) in this proof which is stronger than the result of Lemma 3.9 ($\mathcal{P}^r \perp^\beta \mathcal{P}^- \vee \mathcal{Q}^{(k)}$).
It remains to be shown, that also the set given by the k-th return time satisfies the β-independence in the sense of (3.41). In particular, we aim to show that for all $t \geq 0$, for all $k \in \{1, ..., K\}$ and for all $i \in \{0, ..., k\}$ it holds that

$$\Big\{\varphi_B^{(k)} \leq \frac{t}{p\mu(B)}\Big\} \subseteq (T_B^{-i}(\mathcal{P}^r))^- \vee T_B^{-i}(\mathcal{Q}^{(k-i)}). \tag{3.42}$$

To proof this, we make the observation that we can split the set $\{\varphi_B^{(k)} \leq \frac{t}{p\mu(B)}\}$ into two parts, one part given by the i-th return to B, and one part following the i-th return to B. This follows from the fact that

$$\begin{aligned}
\{\varphi_B^{(k)} = j\} &= \Big\{\sum_{m=1}^{j} 1_B(T^m(x) = k\Big\} = \bigcup_{l\in\mathbb{N}} \Big\{\sum_{m=1}^{l} 1_B(T^m(x) = i\Big\} \cap \Big\{\sum_{m=l+1}^{j} 1_B(T^m(x) = k-i\Big\} \\
&= \bigcup_{l\in\mathbb{N}} \{\varphi_B^{(i)} = l\} \cap \Big\{\sum_{m=1}^{j-l} 1_B(T^m(T^{\varphi_B^{(i)}(x)}(x))) = k-i\Big\} \\
&= \bigcup_{l\in\mathbb{N}} \{\varphi_B^{(i)} = l\} \cap T_B^{-i} \circ \{\varphi_B^{(k-i)} = j-l\}.
\end{aligned}$$

From this, we obtain

$$\Big\{\varphi_B^{(k)} \leq \frac{t}{p\mu(B)}\Big\} = \bigcup_{l\in\mathbb{N}} \{\varphi_B^{(i)} = l\} \cap T_B^{-i} \circ \Big\{\varphi_B^{(k-i)} \leq \frac{t}{p\mu(B)} - l\Big\}. \tag{3.43}$$

For every l, the right hand side of each intersection is a union of cylinders of the partition $T_B^{-i}(\mathcal{Q}^{(k-i)})$ and is therefore measurable with respect to that partition.
In order to obtain (3.42), it remains to be shown that that the left hand side of every intersection in (3.43), given by $\mathcal{Q}_l^{(i)}$ for $l \in \mathbb{N}$, is an element of the σ-algebra $(T_B^{-i}(\mathcal{P}^r))^-$, meaning that the i-th return of a block $B \in \mathcal{P}^{-n}$ always happens before the return of the set A_i (recall that this is visualised in Figure 3.2). By definition, we have $\mathcal{Q}_l^{(i)} = \Big\{\sum_{m=1}^{l} 1_B(T^m(x)) = i\Big\}$. Every function occurring in this sum is measurable with respect to the partition $T^{-m}(\mathcal{P}^{-n})$ (recall (3.29) in the proof of Lemma 3.9). For $m < l$, every partition $T^{-m}(\mathcal{P}^{-n})$ is clearly measurable with respect to the past of $T^{-l}(\mathcal{P}^{-n})$. For the partition $T^{-l}(\mathcal{P}^{-n})$, it holds that

$$T^{-l}(\mathcal{P}^{-n}) = \bigvee_{m'=1}^{n} T^{-l+m'}(\mathcal{P}) = \bigvee_{m'=1}^{n} T^{-\varphi_B^{(i)}+m'}(\mathcal{P}) \subseteq (T_B^{-i}(\mathcal{P}))^- \subseteq (T_B^{-i}(\mathcal{P}^r))^-,$$

which implies that $\mathcal{Q}_l^{(i)}$ is measurable with respect to the past of $T_B^{-i}(\mathcal{P}^r)$.
Recalling expression (3.43), now for every $t \geq 0$, $k \in \{1, ..., K\}$ and $i \in \{0, ..., k\}$, the set $\{\varphi_B^{(k)} \leq \frac{t}{p\mu(B)}\}$ is a finite union of elements of the σ- algebra $(T_B^{-i}(\mathcal{P}^r))^- \vee T_B^{-i}(\mathcal{Q}^{(k-i)})$, which proves that (3.42) is correct.
This proves that for all $i \in \{0, ..., k\}$ and for each A_i, the intersection right to A_i in (3.40) is measurable with respect to the σ-algebra $(T_B^{-i}(\mathcal{P}^r))^- \vee T_B^{-i}(\mathcal{Q}^{(k-i)})$, which implies the β-independence property.
Starting with the set $\{A_k = A\}$ and the intersection to its right, (3.40) equals

$$\frac{1}{p}\mu_B\{A_k = A\}\mu_B\Big(\{A_k - 1 \neq A\} \cap \cdots \cap \{A_1 \neq A\} \cap \{A_0 = A\} \cap \Big\{\varphi_B^{(k)} \leq \frac{t}{p\mu(B)}\Big\}\Big),$$

up to a difference of at most $\frac{1}{p}\beta$. Proceeding from the left (i.e. continuing with $i = k-1$), we can replace the probabilities of the intersections by the products of the probabilities, until (3.40) is almost

$$\frac{1}{p}\mu_B\{A_k = A\}\mu_B\{A_{k-1} \neq A\} \cdots \mu_B\{A_1 \neq A\}\mu_B\{A_0 = A\}\mu_B\Big\{\varphi_B^{(k)} \leq \frac{t}{p\mu(B)}\Big\}, \quad (3.44)$$

with an error of not more than $\frac{1}{p}(k+1)\beta$. As T_B is measure preserving, and $\{A_k = A\}$ equals $T_B^{-k}(A)$, it holds that $\mu_B\{A_k = A\} = \mu_B(A) = p$. Similarly, we have $\mu_B\{A_0 = A\} = p$. For $i \in \{1, ..., k-1\}$ we obtain $\mu_B\{A_i \neq A\} = \mu_B(A^{\mathrm{c}}) = 1 - p$, because $\{A_i \neq A\}$ is the same set as $T_B^{-i}(A^{\mathrm{c}})$. With this, (3.44) is the same as

$$p(1-p)^{k-1}\mu_B\Big\{\varphi_B^{(k)} \leq \frac{t}{p\mu(B)}\Big\} = p(1-p)^{k-1}\overline{G}_B^{(k)}\Big(\frac{t}{p}\Big).$$

Now, for every $k \in \{1, ..., K\}$, the k-th term of the sum $\sum_{k=1}^{\infty} \mu_{BA}\{\varphi_A^{(B)} = k, \varphi_B^{(k)} \leq \frac{t}{p\mu(B)}\}$ satisfies the estimate

$$\Big|\mu_{BA}\Big\{\varphi_A^{(B)} = k, \varphi_B^{(k)} \leq \frac{t}{p\mu(B)}\Big\} - p(1-p)^{k-1}\overline{G}_B^{(k)}\Big(\frac{t}{p}\Big)\Big| \leq \frac{1}{p}(k+1)\beta \leq \frac{1}{p}(K+1)\beta. \quad (3.45)$$

From this, we deduce for the first K terms of the sum

$$\begin{aligned}\sum_{k=1}^{K} \mu_{BA}\Big\{\varphi_A^{(B)} = k, \varphi_B^{(k)} \leq \frac{t}{p\mu(B)}\Big\} &\leq \sum_{k=1}^{K}\Big(p(1-p)^{k-1}\overline{G}_B^{(k)}\Big(\frac{t}{p}\Big) + \frac{1}{p}(K+1)\beta\Big) \\ &= \sum_{k=1}^{K} p(1-p)^{k-1}\overline{G}_B^{(k)}\Big(\frac{t}{p}\Big) + \frac{1}{p}(K^2 + K)\beta. \quad (3.46)\end{aligned}$$

Similarly to (3.45), by excluding the set $\{\varphi_B^{(k)} \leq \frac{t}{p\mu(B)}\}$, β-independence also leads to the inequality $|\mu_{BA}\{\varphi_A^{(B)} = k\} - p(1-p)^{k-1}| \leq \frac{1}{p}K\beta$, again for all $k \in \{1, ..., K\}$. This implies, for the tail of the geometric series $\mu_{BA}\{\varphi_A^{(B)} = k\}$ above K,

$$\sum_{k=K+1}^{\infty} \mu_{BA}\{\varphi_A^{(B)} = k\} = 1 - \sum_{k=1}^{K} \mu_{BA}\{\varphi_A^{(B)} = k\} \leq 1 - \sum_{k=1}^{K}\Big(p(1-p)^{k-1} - \frac{1}{p}K\beta\Big)$$
$$= 1 - \sum_{k=1}^{K} p(1-p)^{k-1} + \frac{1}{p}K^2\beta = \sum_{k=K+1}^{\infty} p(1-p)^{k-1} + \frac{1}{p}K^2\beta. \tag{3.47}$$

The results obtained in (3.46) and (3.47) now allow us to estimate the normalized distribution function of the return time to BA from above:

$$\begin{aligned}
\overline{G}_{BA}(t) &= \sum_{k=1}^{K} \mu_{BA}\Big\{\varphi_A^{(B)} = k, \varphi_B^{(k)} \leq \frac{t}{p\mu(B)}\Big\} + \sum_{k=K+1}^{\infty} \mu_{BA}\Big\{\varphi_A^{(B)} = k, \varphi_B^{(k)} \leq \frac{t}{p\mu(B)}\Big\} \\
&\leq \sum_{k=1}^{K} (p(1-p)^{k-1}\overline{G}_B^{(k)}\Big(\frac{t}{p}\Big) + \frac{1}{p}(K^2+K)\beta + \sum_{k=K+1}^{\infty} \mu_{BA}\{\varphi_A^{(B)} = k\} \\
&\leq \sum_{k=1}^{K} p(1-p)^{k-1}\overline{G}_B^{(k)}\Big(\frac{t}{p}\Big) + \sum_{k=K+1}^{\infty} p(1-p)^{k-1} + \frac{1}{p}(2K^2+K)\beta \\
&\leq \sum_{k=1}^{K} p(1-p)^{k-1}\overline{G}_B^{(k)}\Big(\frac{t}{p}\Big) + \sum_{k=K+1}^{\infty} p(1-p)^{k-1}\overline{G}_B^{(k)}\Big(\frac{t}{p}\Big) + \frac{1}{p}(2K^2+K)\beta + \frac{\delta}{2} \\
&\leq \sum_{k=1}^{\infty} p(1-p)^{k-1}\overline{G}_B^{(k)}\Big(\frac{t}{p}\Big) + \delta.
\end{aligned}$$

Note that the inequality in the penultimate line holds because $\sum_{k=K+1}^{\infty} p(1-p)^{k-1} < \delta/2$ and as $\overline{G}_B^{(k)}$ are distribution functions for all $k \geq 1$.

For every positive integer k, the distribution function $\overline{G}_B^{(k)}$ satisfies

$$\int_0^\infty 1 - \overline{G}_B^{(k)}(s)ds = \int_0^\infty \mu_B\{\overline{\varphi}_B^{(k)} > s\}ds = \sum_{j=1}^{\infty} \mu_B\{\overline{\varphi}_B^{(k)} \geq j\} = \sum_{j=1}^{\infty} j\mu_B\{\overline{\varphi}_B^{(k)} = j\} = k.$$

Recalling Lemma 3.3, the distribution function of the normalized waiting time to BA, $\overline{F}_{BA}(t)$ is bounded from above by $\int_0^t 1 - \overline{G}_{BA}(s)ds$ for every $t \geq 0$. Therefore, we can apply Corollary 3.1 on $\overline{F}_{BA}$, such that

$$\overline{F}_{BA}(t) \leq \frac{1}{(1-p)\ln(e_p)}(1 - e_p^{-t}) + \delta t$$

for all $t \geq 0$ and for $p = \mu_B(A) \leq \delta$. As e_p increases with p (see (iv) of Lemma 3.5), we have $1 - e_p^{-t} \leq 1 - e_\delta^{-t}$. Additionally, by (3.13), it holds that $\frac{1}{(1-p)\ln(e_p)} \leq \frac{1}{(1-\delta)\ln(e_\delta)}$. By these inequalities and the fact that $\overline{F}_{BA}$ is a distribution function (i.e. it cannot exceed one), we conclude that

$$\begin{aligned}\overline{F}_{BA}(t) &\le \min\Big\{1, \frac{1}{(1-p)\ln(e_p)}(1-e_p^{-t}) + \delta t\Big\} \\ &\le \min\Big\{1, \frac{1}{(1-\delta)\ln(e_\delta)}(1-e_\delta^{-t}) + \delta t\Big\} = g_\delta(t) \le 1 - e^t + \epsilon.\end{aligned}$$

This implies, that for our chosen ϵ, with μ-tolerance ϵ for cylinders BA of the partition $\mathcal{P}^{[-n,r)}$ with length $m = n + r \ge n_0 + r$, for all $t \ge 0$ the force of repelling is at most ϵ. Equivalently, for $A_x^{[-n,r)}$ being the cylinder of the partition $\mathcal{P}^{[-n,r)}$ containing x, we have proven that

$$\mu\Big\{x \in X : \sup_{t\ge 0}(\overline{F}_{A_x^{[-n,r)}}(t) - 1 + e^{-t}) \ge \epsilon\Big\} \longrightarrow 0 \;\; as \;\; m \to \infty.$$

To show that the cylinders of the partition $\mathcal{P}^m$ also repel with a force of not more than ϵ, we observe that the same decay of repelling also holds for the process generated by the partition $T^{-n}(\mathcal{P})$, which implies that almost all cylinders $C \in T^{-n}(\mathcal{P}^{[-n,r)}) = \mathcal{P}^m$ repel with force of at most ϵ. Applying Lemma A.1, this finishes the proof of the theorem. $\square$

3.3. The attracting postulate of the Law of Series

In this final section, we are going to complete the picture of the Law of Series. We will not dicsuss the stated theorems in detail and refer to [3] and [4, Section 5.4] for further reading.

In order obtain the second postulate of the Law of Series, we need the presence of attracting. This will not be possible in an as general setting as in the repelling case, as not all processes $(X, \mathcal{P}, \mu, T)$ with positive dynamical entropy reveal attracting behaviour for sufficiently long cylinders. Nevertheless, we will see that indeed in most positive entropy processes (we will call these processes *typical*) many sufficiently long cylinders show even strong attracting behaviour.

Recall that $(X, \mathcal{A}, \mu)$ denotes a standard probability space. By Lemma 2.8, for every $k \ge 2$, the space $\mathcal{B}_k$ of all measurable partitions $\mathcal{P}$ with k elements endowed with the metric d_k, given by

$$d_k(\mathcal{P}, \mathcal{Q}) = \inf_\pi \sum_{i=1}^{k} \mu(A_i \triangle C_{\pi(i)}),$$

is a complete and separable metric space.
In a complete metric space a subset is called *residual*, if it contains a dense G_δ set, or equivalently, if its complement is a countable union of nowhere dense sets. A residual set can be considered as a very large subset of the space. Given a fixed residual set, its elements are called *typical*. Similarly, we call a process $(X, \mathcal{P}, \mu, T)$ typical, where $\mathcal{P} \in \mathcal{B}_k$, whenever $\mathcal{P}$ is an element of a residual subset of $\mathcal{B}_k$.

To state the attracting postulate of the Law of Series, we require the following definitions: We say that a measure preserving dynamical system is *aperiodic*, if the measure μ of all periodic points of X (those $x \in X$ for which $T^n(x) = x$ for some $n \in \mathbb{N}$) equals zero. The *upper density* of a set $N \subseteq \mathbb{N}$ is defined as

$$\overline{D}(N) = \limsup_{n\to\infty} \frac{\#(N \cap [1, n])}{n}.$$

Whenever a property holds for a set of upper density 1, we say that the property holds for the *majority of lengths*.

We state two versions of the main theorem of this section, one version addressing an intuitive understanding of the theorem and one rigorous version:

Theorem 3.2. *(The Law of Series, attracting postulate)[3, Theorem 1] In a typical aperiodic ergodic process $(X, \mathcal{P}, \mu, T)$, for majority of lengths n, all cylinders reveal strong attracting.*

Theorem. *Let $(X, \mathcal{A}, \mu, T)$ be an ergodic aperiodic system on a standard probability space $(X, \mathcal{A}, \mu)$. Then, for every fixed $k \geq 2$, in the space $\mathcal{B}_k$ of all k-element partitions of X endowed with the metric d_k there exists a residual subset $\mathcal{C} \subseteq \mathcal{B}_k$, such that for every $\mathcal{P} \in \mathcal{C}$ the generated process $(X, \mathcal{P}, \mu, T)$ has the following property: There exists a set $N \subset \mathbb{N}$ of upper density one, such that for every $\epsilon > 0$ there is an n_0, such that for every $n \in N$ where $n \geq n_0$, every cylinder $B \in \mathcal{P}^n$ attracts with force $1 - \epsilon$.*

In every process $(X, \mathcal{P}, \mu, T)$ satisfying the above theorem, all cylinders of the partition $\mathcal{P}^n$ (where $n \in N$ is sufficiently large) attract with arbitrary strong force $1 - \epsilon$ and therefore exhibit arbitrary small repelling (up to ϵ). This implies, that for every such cylinder $B \in \mathcal{P}^n$ the process counting the occurrences of B obeys the Law of Series.
Note that Theorem 3.2 does not depend on the entropy of the process $(X, \mathcal{P}, \mu, T)$, i.e. it also holds for processes with zero entropy. Nevertheless, in the zero entropy case, repelling behaviour of cylinders of length n, where $n \notin N$, is not excluded (as it is, other than marginal, in the positive entropy case by Theorem 3.1).

Lemma 3.10. *[4, Theorem 5.4.4] Let $(X, \mathcal{A}, \mu, T)$ denote a probability preserving dynamical system with positive Kolmogorov-Sinai entropy $h(\mu, T)$, where $(X, \mathcal{A}, \mu)$ is standard, and let, for fixed $k \geq 2$, $\mathcal{B}_k$ be equipped with the metric d_k. Then the subset $\mathcal{D} \subset \mathcal{B}_k$ of all k-element partitions, for which the dynamical entropy $h(\mathcal{P})$ of the generated process $(X, \mathcal{P}, \mu, T)$ is positive, is residual in the space $\mathcal{B}_k$.*

In words, in any system with positive Kolmogorov-Sinai entropy, the majority of the k-element partitions $\mathcal{P}$ generate a positive dynamical entropy $h(\mathcal{P})$.

Combining the above lemma and the Theorems 3.2 and 3.1, and as the countable intersection of residual sets is again residual, we conclude:
In the class of ergodic aperiodic systems $(X, \mathcal{A}, \mu, T)$ with positive Kolmgorov-Sinai entropy $h(\mu, T)$, in a typical process $(X, \mathcal{P}, \mu, T)$ generated by a finite partition $\mathcal{P}$, long

cylinders show almost no repelling, while many of them reveal strong attracting, i.e. obey the Law of Series.

This means, that the presence of attracting and the absence of repelling correspond with the natural behaviour of a process generated by a finite partition, whenever the underlying dynamical system is ergodic, aperiodic and has positive Kolmogorov-Sinai entropy.

A. Appendix

Lemma A.1. *Let $(X, \mathcal{A}, \mu)$ be a measure space, and let f, f_1, f_2,... be elements of the space $\mathcal{L}^1(X, \mathcal{A}, \mu, \mathbb{R})$. If there is a nonnegative real valued integrable function g such that $|f| \leq g$ and $|f_n| \leq g$ holds for all $n \geq 1$, then*

$$f_n \to f \;\; in \;\; measure \Longleftrightarrow f_n \to f \;\; in \;\; L^1(\mu)$$

Lemma A.2. *(Rectangle rule) Let $(X, \mathcal{A}, \mu)$ be a probability space and let $f \geq 0$ be a measurable function on X such that $\int f d\mu < \gamma\delta$ for some $\gamma > 0$ and $\delta > 0$. Then*

$$\mu\{x \in X : f(x) \geq \gamma\} < \delta.$$

Corollary A.1. *If the assumptions of Lemma A.2 are satisfied, then*

$$\mu\{x \in X : f(x) > \gamma\} < \delta$$
$$\mu\{x \in X : f(x) > \gamma\} \leq \delta$$
$$\mu\{x \in X : f(x) \geq \gamma\} \leq \delta.$$

Bibliography

[1] Cohn, Donald L.: *Measure Theory.* Second Edition. Birkhäuser, 2013

[2] Downarowicz, T. ; Lacroix, Y.: The law of series. In: *Ergodic Theory and Dynamical Systems* 31 (2011), S. 351–367

[3] Downarowicz, T. ; Lacroix, Y. ; L´eandri, D.: Spontaneous clustering in theoretical and some empirical stationary processes. In: *ESAIM: Probability and Statistics* 14 (2010), S. 256–262

[4] Downarowicz, Tomasz: *Entropy in Dynamical Systems.* Cambridge University Press, 2011

[5] Černý, Jiří: *Stochastische Prozesse.* lecture notes, winter term 2012

[6] Kammerer, Paul: *Das Gesetz der Serie. Eine Lehre von den Wiederholungen im Lebens- und im Weltgeschehen.* Deutsche Verlagsanstalt, 1919

[7] Sarig, Omri: *Lecture Notes on Ergodic Theory.* 2020

[8] Walters, Peter: *Ergodic Theory - Introductory Lectures.* Springer-Verlag, 1975

[9] Zweimüller, Roland: *Ergodic Theory and Dynamical Systems II.* lecture, summer term 2019

[10] Zweimüller, Roland: *Maß- und Integrationstheorie.* lecture notes, winter term 2015